MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Assessing the Importance of Rain Gauge Stations through Network Theory

Muhamad Luqman Sapini[a,b]*, Mohd Salmi Md Noorani[a], Mohd Almie Alias[a], Fatimah Abdul Razak[a], Nurul Zahirah Abd Rahim[a,c], Muhamad Imran Sapini[a], Norliza Muhamad Yusof[b]

[a]Department of Mathematical Sciences, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; [b]Universiti Teknologi MARA (UiTM), Cawangan Negeri Sembilan, Kampus Seremban, Seremban 70300 Negeri Sembilan, Malaysia; [c]Universiti Teknologi MARA (UiTM), Cawangan Melaka, Kampus Jasin, Jalan Lembah Kesang 1/1-2, Kampung Seri Mendapat, 77300 Merlimau, Melaka, Malaysia

Abstract Understanding rainfall's spatial and temporal dynamics is pivotal for effective water resource management, especially in regions susceptible to extreme weather events. This study employs network theory to investigate rainfall distribution and variability across 21 Sungai Pahang, Malaysia monitoring stations. Utilising Gephi and GeoLayout for network visualisation and analysis, the paper reveals that the monitoring network possesses both small-world and scale-free characteristics, which offer insights into its robustness and vulnerability. Centrality measures, including degree, betweenness, and closeness, are examined to identify key nodes critical for information flow within the network. Specifically, stations with high centrality measures emerge as pivotal points for monitoring, while those with low Centrality offer valuable localised data. The study fills a crucial gap by addressing the challenges and limitations of existing rainfall monitoring networks in Malaysia, providing a nuanced understanding with broader applications in climate science and hydrology. The findings have significant implications for policymakers and practitioners involved in climate monitoring and disaster preparedness. Future work may extend this model to include additional environmental variables or adapt the methodology to other types of monitoring networks.

Keywords: Rainfall monitoring, network theory, centrality analysis, small-world, scale-free.

Introduction

The dynamics of rainfall patterns have long been a subject of academic interest, driven by the need to understand natural climatic variations and predict and manage water resources effectively [1]. These patterns, which can vary significantly in both time and space, have critical implications for agriculture [2], flood risk management [3], [4], and even urban planning [5]. Traditional methods of analysing rainfall and its impact often involve the use of statistical or computational methods that focus on the stationarity, periodicity, mutation, and trends of hydrological time series, such as the Mann–Kendall trend test [6], [7] [8], [9] and Entropy [10], [11] that rely on point measurements or area-averaged data. However, these conventional approaches sometimes fail to capture the complexity and interconnectedness inherent in hydrological systems.

Recently, there has been a growing trend in utilising complex network theory in hydrology, which provides a comprehensive approach to examining and comprehending data. Network theory provides tools for capturing the topology and dynamics of inherently linked systems. It has been successfully applied in various disciplines such as social science [12], [13], finance [14], transportation [15], [16], biology [17] [18], environment [19], [20], [21] and computer science [22] [23]. Within the context of hydrology, network theory allows us to look beyond isolated data points and consider the intricate connections between different rainfall stations or measurement points. These connections could be manifested as solid correlations between rainfall data, indicating a form of climatic or hydrological dependency that can be critically important for resource management.

***For correspondence:**
luqman0211@uitm.edu.my

Received: 20 Jan. 2024
Accepted: 17 June 2025

Recent years have seen a burgeoning interest in applying network theory to hydrological research. For example, [24] discusses how network theory has successfully characterised the interaction among constituents of complex systems, including hydrological systems. [25] highlight the application of complex network theory in hydrology and water resources, particularly in studying connections in rainfall, stream flow, river networks, and virtual water trade networks. [26] use complex network theory to learn hydrological systems and understand the connections and complexities of hydrological networks. Studies have utilised network analysis to probe the underlying architecture of hydrological systems, revealing new insights into catchment organisation, river networks, and water distribution infrastructure. Specifically, in the realm of rainfall studies, network analysis has been instrumental in decoding spatiotemporal correlations between rain gauge stations [27], [28], [29], identifying regions susceptible to flooding or drought [30], [31], and even in facilitating real-time monitoring of rainfall events [32]. This fresh perspective has enriched our understanding by complementing traditional hydrological models, offering a more nuanced and interconnected view of how rainfall patterns emerge, interact, and influence broader climatic and environmental settings.

An intriguing characteristic exhibited by networks in the physical world, including but not limited to biological systems, the internet, and social networks, is their propensity to manifest universal properties. In particular, numerous studies have demonstrated that these networks exhibit small-world or scale-free properties. [33] state that the structure of rainfall networks can vary spatially, and this is an important aspect of the small-world characteristic in rainfall networks. Due to their brief average path length and high clustering coefficient, small-world networks are effective at information flow and localised interactions [34]. In contrast, scale-free networks exhibit a degree distribution according to the power law, wherein a limited number of vastly interconnected nodes (hubs) substantially influence the network's connectivity [35] [36]. In the context of rainfall networks, complex network analysis has been used to quantify the roles of individual stations within homogeneous regions [37]. This approach identifies spatial connections in rainfall networks, crucial for various purposes such as interpolation and network design [29]. By utilising these characteristics as a framework, one can gain a more comprehensive understanding of the complex and fundamental composition of hydrological systems such as rainfall networks; consequently, network analysis has become an ever more essential instrument in this field.

The application of network theory in hydrological research has provided valuable insights into the distribution and interaction of rainfall. However, there remain knowledge gaps in our comprehension of regional contexts, namely in the case of Malaysia. A critical concern pertains to the insufficient investigation of the constraints and limits associated with Malaysia's existing rainfall monitoring network. Furthermore, there is a noticeable gap in the existing literature regarding systematically comparing the selection of thresholds in relation to the scale-free and small-world properties of rainfall networks. The impact of threshold values on network topology sensitivity suggests that centrality measurements may exhibit significant variations across different thresholds [38]. Therefore, selecting a threshold without careful consideration or based on arbitrary criteria may result in a skewed comprehension of network centrality, a crucial factor in assessing the significance and impact of individual nodes. This gap indicates the necessity for a rigorous technique to identify the best representative threshold value corresponding to rainfall networks' inherent small-world and scale-free characteristics. This is crucial to ensure that centrality analysis accurately captures the underlying dynamics of the system.

Considering these issues and opportunities, the primary objective of this study is to apply complex network analysis to a regional rainfall network in Malaysia, utilising a dataset that spans seven years and includes 21 rain gauge stations. Specifically, the study aims to (1) construct correlation-based rainfall networks at different thresholds, (2) investigate whether these networks exhibit universal properties like small-world and scale-free characteristics, and (3) identify and analyse central nodes within the network using various centrality metrics.

This research holds multiple significant implications. Initially, the study endeavours to fill a knowledge need by employing network theory within the framework of a regional hydrological system that has received limited attention in previous research. Additionally, this study presents a methodological framework for developing and analysing rainfall networks, which can be easily applied to other hydrological or environmental networks. Furthermore, the identification of core nodes within the network holds practical significance. These nodes possess the potential to play a crucial role in monitoring and predictive modelling since their Centrality often signifies a more significant influence on the overall behaviour of the system. Finally, a comprehensive comprehension of the topological characteristics of the network can provide valuable insights for developing better and more robust techniques in water resource management. Consequently, this can significantly enhance flood prediction capabilities and optimise agricultural planning processes.

The subsequent sections of this work are structured in the following manner: Section 2 of this paper provides an overview of the research methodology, outlining the procedure employed for doing computations within networked structures. Section 3 presents the findings and interpretations obtained from the collected data. The findings are further expanded upon in Section 4, while Section 5 encompasses the conclusions drawn from the study and recommendations for further research.

Research Methodology

A graph is a fundamental structure in network theory composed of two primary components: nodes (or vertices) and edges (or links). A node represents an individual entity, while an edge denotes a relationship or interaction between two nodes.

A network can be defined as a collection of nodes that are interconnected by edges [39]. The components within a graph structure are commonly referred to as nodes or vertices, while the connections between these components are known as links or edges. Mathematically, a network also known as a graph-based model, can thus be formally expressed as $G = \{P, E\}$, where P is the set of N nodes $\{P_1, P_2, ..., P_N\}$ and E is the set of n edges (with n defined as the total number of links in the network). In the context of this study, each node corresponds to a rainfall monitoring station, and an edge is established between two stations if their correlation exceeds a certain threshold. Figure 1 shows an example of a network.

There are multiple methods for constructing the edges of a network, depending on the nature of the data and the desired analysis. Common methods include using an adjacency matrix**,** edge lists**,** weighted connections**,** and threshold-based correlations. In this study, we employ a thresholding technique, where edges are created based on the strength of pairwise Pearson correlation values between rainfall stations. If the correlation between two stations surpasses a specified threshold, an edge is formed between them.

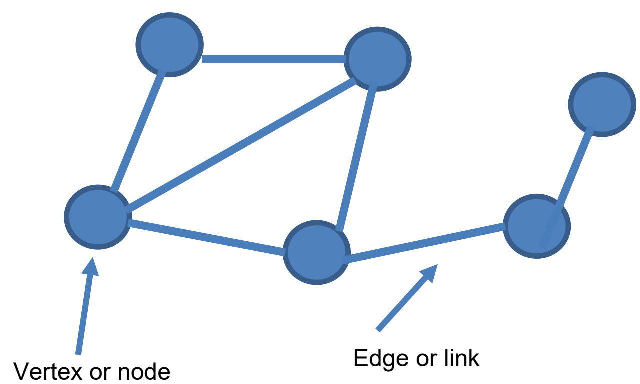

Figure 1. Example of graph or network with six nodes and seven edges

There are multiple strategies for constructing the edges of a network, depending on the nature of the data and the desired analysis. Common methods include using an adjacency matrix [40], [41], edge list [42], [43], weighted edges [44], [45], and thresholding [46], [47]. In this study, we employ a thresholding technique, where edges are created based on the strength of pairwise Pearson correlation values between rainfall stations. If the correlation between two stations surpasses a specified threshold, an edge is formed between them. Figure 2 visually represents the research framework, illustrating the critical components and their interrelationships in the study.

The following section provides a comprehensive outline of the research methodology, detailing the steps in data collection, preprocessing, and subsequent analysis. The study adopts a multidimensional approach due to its interdisciplinary nature, combining hydrological science and complex network theory.

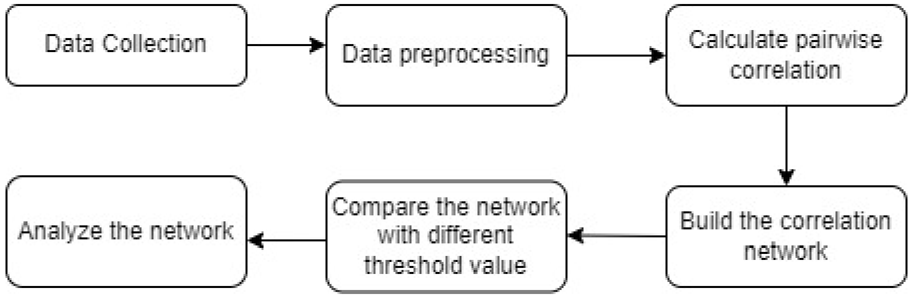

Figure 2. Research framework

Study Area and Data Collection

This study aims to investigate the spatial patterns of rainfall in the Sungai Pahang region, recognised as the longest river in Peninsular Malaysia. Figure 3 depicts the geographical representation of Sungai Pahang, illustrating the precise placement of 21 monitoring stations designated for rainfall observation. The climatic conditions in the surrounding area exhibit tropical nature, featuring consistently elevated temperatures and humidity levels throughout the year. Significantly, the geographical region experiences a monsoon period spanning from November to January, characterised by substantial rainfall that frequently results in flooding in the downstream sections of the river and adjacent regions [48]. The annual precipitation in the Pahang state exhibits variation, with values ranging from 2000mm to 2500mm. This variability is attributed to several factors, including elevation and proximity to the coast [48].

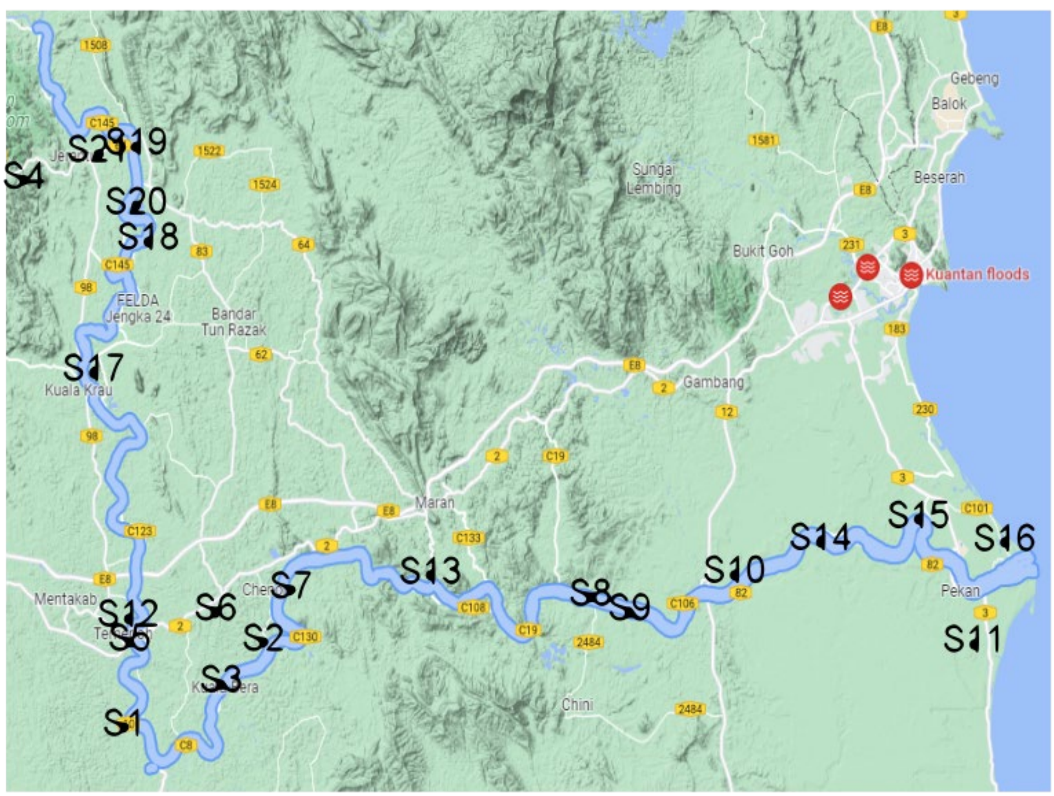

Figure 3. Geographical map of sungai Pahang with the location of 21 rainfall monitoring stations

Various aspects, such as the research aims and the need for adequate water resources planning and management, determine the period scale selection in this study. For example, the utilisation of daily or weekly data may be deemed sufficient for capturing short-term fluctuations in rainfall [49]. In contrast, monthly or yearly data are better suited for analysing long-term trends [50]. This study emphasises the spatial dynamic aspects of monthly rainfall, providing a comprehensive viewpoint encompassing both short-term and long-term variability [51], [52].

The dataset comprises monthly rainfall data collected from 21 specially chosen stations throughout the Sungai Pahang basin. The potential advantages of a more comprehensive network of stations in the region are evident, as it would likely yield more dependable outcomes. However, it is essential to acknowledge that the existing station selection is predominantly shaped by the accessibility of long-term, high-quality data. The data utilised in this analysis, covering the period from 2013 to 2020, was acquired from the Department of Irrigation and Drainage Malaysia, located in Ampang, Selangor, Malaysia. Table 1 presents a concise statistical overview of the monthly rainfall data collected from the 21 rain gauge stations.

Table 1. Summary rainfall statistics at 21 gauging stations at a monthly temporal scale

Statistic	Value
Length of data	83 months (August 2013 – June 2020)
Range of mean (mm)	100.7289 – 200.9614
Range of Standard deviation (mm)	67.8377 – 225.0994
Range of Minimum Rainfall (mm)	0-9.6
Range of Maximum Rainfall (mm)	321.5 - 1589

Data Preprocessing

Data preprocessing serves as a crucial step in the data analysis pipeline. This phase encompasses cleaning, transforming, and conditioning the data to facilitate subsequent analyses [53]. The overarching aim is to ensure that the dataset lacks errors, outliers, or missing values that might compromise the study's accuracy [54], [55].

A commonplace issue in real-world data is the presence of missing values, which can be attributed to various factors, including measurement errors, limitations in data collection, or data loss [56]. To tackle this issue, the utilisation of data imputation becomes imperative. For this study, mean imputation was employed as the imputation method of choice. This method involves replacing missing values with the mean of the available data, assuming that the missingness is random [57].

The decision to use mean imputation was made after considering multiple considerations. Initially, it is essential to note that the percentage of missing data observed throughout the 21 stations included in our analysis is below 5%. Consequently, imputation techniques are expected to have a limited effect on the overall findings. Furthermore, considering the substantial temporal scope of our dataset, which encompasses seven years from 2013 to 2020, the sample size is sufficiently big to mitigate the influence of mean imputation on the overall results. The method has gained significant recognition and has been effectively employed in previous research endeavours, validating its suitability for the present study's environment [58] [59].

Construction of Pairwise Correlation Matrix

A correlation network is a graphical representation of the relationships between variables in a dataset. To construct a correlation network, one can calculate the pairwise correlations across variables using the preprocessed data. The correlations can be represented visually as an adjacency matrix or a network graph [60]. Pairwise correlation is a statistical measure that quantifies the strength and direction of the linear association between two variables. Pearson's correlation coefficient is widely regarded as the prevailing metric for assessing pairwise correlation [61]. The correlation between the given data can be quantified by employing equations (1) and (2).

$$p(X,Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)} \tag{1}$$

$$\text{cov}(X,Y) = \mathrm{E}\left[(X-\mu(X))(Y-\mu(Y))\right] \tag{2}$$

In Equation (1) and (2), $p(X,Y)$ denotes the Pearson correlation between two variables X and Y. The terms $\mu(X)$ and $\mu(Y)$ represents the mean of *X* and *Y*, respectively, while $\sigma(X)$ and $\sigma(Y)$ denote their corresponding standard deviation of *X* and *Y*. The function E indicates the expectation operator, and $\text{cov}(X,Y)$ the covariance between *X* and *Y*.

The resulting value is Pearson's correlation coefficient, which ranges from -1 to 1. A value of -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 shows a perfect positive linear relationship [62].

In this study, the pairwise correlations are computed between monthly rainfall time series for each pair of the 21 stations. The resulting values form a symmetric correlation matrix, which serves as the foundation for constructing the rainfall network via thresholding.

Building the Networks

The next step is thresholding the correlation matrix to define the network's edges. The best threshold value for creating a correlation network depends on the specific application and the characteristics of the data. There is no definitive answer to the best threshold value, as it can vary depending on the data set. A common approach is to use a threshold based on the correlation coefficient, such as a threshold of 0.5 or higher, which means that only variables with a correlation coefficient of 0.5 or higher are strongly correlated [63], [46]. This threshold can be adjusted depending on the specific context [64]. For example, if we want to capture a more detailed network with more edges, a lower threshold can be used, while if we want a more simplified network with fewer edges, a higher threshold can be used.

An alternative methodology involves employing a threshold based on the p-value of the correlation coefficient, which quantifies the statistical significance of the observed connection [65]. This method can be utilised to ascertain the statistical significance of the association, thereby distinguishing it from random chance [66].

It is noteworthy to emphasise that in some instances, networks can be constructed utilising several thresholds. Subsequently, doing a comparative analysis of the networks can facilitate the extraction of more information and identify the most crucial edges. Six correlation thresholds are considered to establish the network, specifically 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. A connection between two nodes (stations) is set if their Pearson's correlation surpasses the specified threshold.

In constructing the network for spatial rainfall dynamics, we utilised Gephi, a visualisation and exploration software for various kinds of networks and complex systems [56] . Gephi provides a comprehensive environment for network data analysis, including tools for statistical computations, clustering, and community detection, among other functionalities [67].

One of the standout features of Gephi for our study is its ability to incorporate geographic information via plugins, notably the GeoLayout plugin. This plugin allows for integrating latitude and longitude information, making it possible to position the nodes (in our case, the 21 rainfall stations) on a geographical map [68]. This geo-referencing adds another layer of realism to our network model and facilitates a more intuitive understanding of spatial relationships and dependencies among the stations.

Selection of Optimal Threshold Value for Network Analysis

Six different thresholds (T) are considered for creating the correlation network: 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. One of the pivotal decisions in constructing a complex network is selecting an appropriate threshold value. The threshold value profoundly impacts the network's structure, affecting its properties and, consequently, the insights drawn from its analysis. Particularly in hydrological systems like rainfall distribution, the selected threshold is instrumental in shaping a network that embodies small-world and scale-free properties—two fundamental characteristics often present in real-world networks.

Small-World Network

Initially, our analysis aims to identify whether the constructed network adheres to the small-world property, a type of network characterised by a high clustering coefficient and relatively short average path lengths between nodes. The small-world property allows for efficient information transfer [69] and closely mimics many natural [70] and social systems [71]. We evaluate the clustering coefficient and average path length at various threshold levels to determine if our network exhibits this characteristic.

The clustering coefficient quantifies the tendency of a network to cluster and, therefore, is a measure of local density [72]. For a single node in an undirected network, the clustering coefficient C_i is defined as

$$C_i = \frac{2E_i}{k_i\left(k_i - 1\right)} \tag{3}$$

where E_i is the number of existing links between all neighbors of node (i) with degree k_i. The procedure is repeated for each node of the network. The average of the clustering coefficients of all the individual nodes, *n* is the clustering coefficient of the whole network *C*.

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i \quad (4)$$

The average path length is a concept used in network theory to measure the average number of steps along the shortest paths for all possible pairs of network nodes [73]. Equation (5) is the formulation to calculate the average path length for a simple, connected, undirected graph.

$$L = \frac{1}{N(N-1)}\sum_{i \neq j} d_{ij} \quad (5)$$

In Equation (5), d_{ij} denotes the shortest path (geodesic distance) between nodes i and j, and *N* is the total number of nodes in the network. The formula computes the average number of steps required to traverse the shortest paths between all distinct node pairs in a simple, connected, undirected graph. Small values of *L* ensure that information or resources are easily spread throughout the network. This property makes distributed information processing possible on technological networks and supports the six degrees of separation often reported in complex networks.

The small-world measurement, ω, is defined by comparing the clustering of the network to that of an equivalent lattice network, C_{latt}, and comparing path length to that of an equivalent random network, L_{rand}; the relationship is simply the difference of two ratios defined as:

$$\omega = \frac{L_{rand}}{L} - \frac{C}{C_{latt}} \quad (6)$$

In using the clustering of an equivalent lattice network rather than a random network [74], this metric is less susceptible to the fluctuations seen with C_{rand}. Moreover, values ω are restricted to intervals -1 to 1 regardless of network size. Values close to zero are considered the small world; positive values indicate a graph with more random characteristics, and negative values indicate a graph with more regular or lattice-like characteristics [75].

Scale-free Network

After confirming small-world characteristics, we focus on establishing whether the network exhibits scale-free properties. Scale-free networks follow a power-law degree distribution, where a few nodes (often called 'hubs') have significantly more connections than others [76], [77]. This property is expected in various real-world systems and provides insights into network robustness and vulnerability.

Networks with power-law properties are called scale-free networks [76]. The degree distribution of a scale-free graph is a power-law distribution with a peak at *p*(k), and is given by:

$$p(k) = k^{-\alpha} \quad (7)$$

Power laws are linear on a log scale, with slope α equal to their exponent. Figure 4 shows the power-law and exponential distributions graph in Normal and Log-log scales. Figure 4 serves as a conceptual illustration to help the reader distinguish between power-law and exponential degree distributions. It supports the interpretation of Equation (7) and provides visual context for identifying scale-free characteristics in empirical network data.

In scale-free networks, an *α* value between 2 and 3 is often cited as the "classic" range that indicates a heavy-tailed degree distribution where a few nodes serve as highly connected hubs [78]. However, if *α* falls outside this range, it can have specific implications for the network's topology and characteristics. A very low *α* suggests an even more extreme level of heterogeneity among the nodes, meaning a few ultra-high-degree nodes dominate the network. In many real-world scenarios, an *α* value below two can be seen as somewhat unrealistic, as it implies an infinite average degree and an infinite variance [78]. As α increases beyond 3, the network looks less and less scale-free. Hubs may still exist but are not as dominant. The network starts to resemble more of an exponential or a random network in terms of degree distribution [79].

To examine the scale-free nature of our network, we analyse the degree distribution across a range of threshold values. In most real-world applications, the value of α falls in the range $2 < \alpha \leq 3$, although values outside this range occasionally occur.

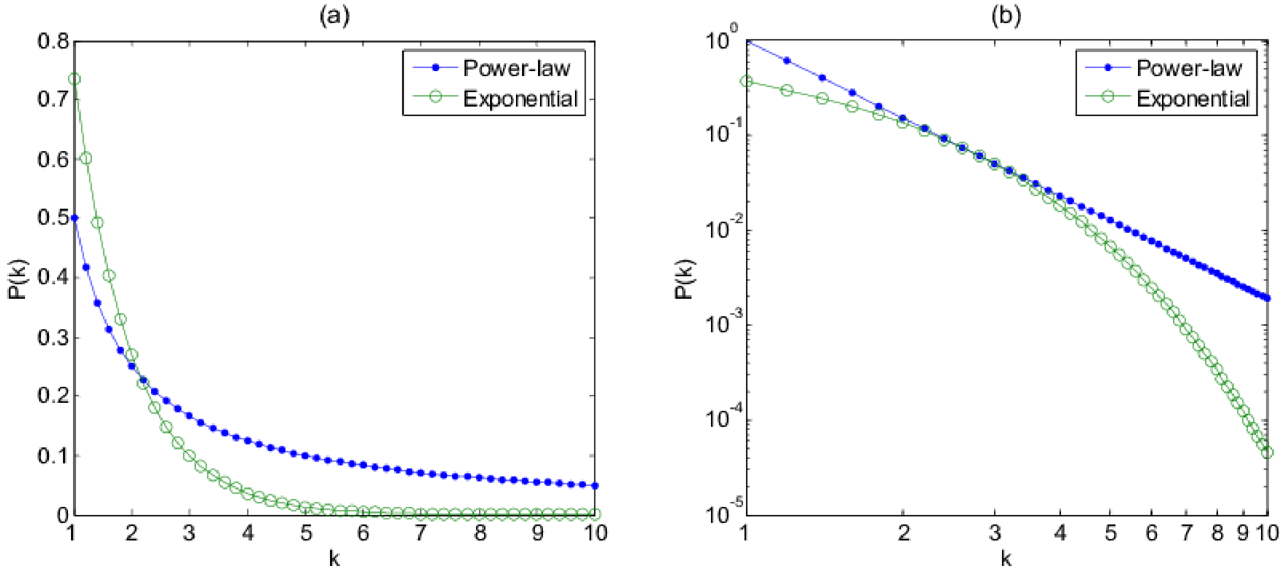

Figure 4. Power-law and exponential distributions: (a): Normal scale, (b): Log-log scale. Source image from [80]

The transition from assessing small-world properties to scale-free characteristics is logical and essential for a comprehensive network analysis. The small-world properties offer insights into local connectivity and short relational distances. In contrast, the scale-free properties provide a macroscopic view of the network, focusing on its resilience and susceptibility to changes. By iteratively evaluating these properties for different thresholds, we identify an optimal value that results in a network exhibiting both small-world and scale-free characteristics.

Through this systematic approach, we arrive at a threshold value that fits our observed data well and adds robustness to our subsequent analyses, thus ensuring that our study on the spatial dynamics of rainfall across Sungai Pahang is rigorous and insightful.

Centrality Network Analysis

There exist numerous methodologies for examining the attributes of networks. For example, networks can be assessed based on qualities such as Clustering, Topology, Adjacency, Centrality, and Entropy. Likewise, various metrics and methodologies exist for representing these features. The present work uses centrality analysis as a methodological approach to investigate the spatial dynamics of rainfall. Centrality analysis is a methodology employed within network analysis to ascertain the nodes of utmost significance within a given network [81]. It involves measuring the structural Centrality of nodes and evaluating their consistency with intuition and interpretability. There are different methodologies for computing the global importance of nodes such as betweenness centrality, degree centrality, closeness centrality, Cross-face Centrality and PageRank algorithm [82], [83]. This study employs three distinct centrality measures: Degree Centrality, Betweenness Centrality, and Closeness Centrality.

Degree Centrality

The Degree Centrality of the node in a network is the total number of edges connected to it [84]. It can be defined as

$$DC(v_i) = \sum_{j=1}^{N} A_{i,j} \tag{8}$$

where $DC(v_i)$ is the degree centrality of node, *A* is the adjacency matrix, $A_{i,j}=1$ if there is a link between nodes i and j, 0 otherwise, and N is the total number of nodes.

It shows that the nodes that have a lot of connections can be considered as the main node in a network. In a rainfall station network, degree centrality can be used to identify which stations are most connected to other stations. These stations may play an essential role in the flow of information and resources within the network, as they have more direct connections to other stations.

Betweenness Centrality

Betweenness Centrality (BC) of the node in a network is the shortest path (geodesics) available between a pair of nodes [85]. If the vertex or node is in the only direction other nodes must go through, this node should be essential and have a strong betweenness centrality. The higher the value of BC of the node in a network, the more critical a node is for the information flow. BC can be defined as

$$BC(v)=\sum_{j\neq k}\frac{g_{jk}(v)}{g_{jk}} \tag{9}$$

In equation 9, g_{jk} represents the total number of shortest paths between nodes *j* and *k*, while $g_{jk}(v)$ denotes the number of those shortest paths that pass-through node *v*. The betweenness centrality of node v is obtained by summing the fraction $\frac{g_{jk}(v)}{g_{jk}}$ over all distinct pairs of nodes *j* and *k*, where $j\neq k\neq v$.

$g_{jk}(v)$ is the shortest route connecting the nodes j and k with the nodes v and g_{jk} the shortest path connecting nodes *j* and *k*. In a rainfall station network, BC can be used to identify which stations are essential for connectivity and information flow within the network.

Closeness Centrality

Closeness centrality measures how close a node is to all other nodes in the network [86]. Let $d(v,v_j)$ denote the shortest path distance between node v and another node v_j. The total distance from node v to all other nodes is then given by:

$$l_v=\frac{1}{n}\sum_{j\in v}d(v,v_j) \tag{10}$$

where v is the set of all nodes in the network. The closeness centrality of node v, denoted as $CC(v)$ is defined as the inverse of this total distance:

$$CC(v)=\frac{n}{\sum_{j\in v}d(v,v_j)} \tag{11}$$

where n is the total number of nodes in the network. A higher CC(v) value indicates that the node is, on average, closer to all other nodes.

In a rainfall station network, closeness centrality can be used to identify which stations are central and have easy access to other stations in the network.

Result and Analysis

The piecewise correlation is calculated for all 21 rainfall stations in Sungai Pahang following the procedure in 2.2 and 2.3. Next, six different correlation thresholds (CT: 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8) are considered for creating the correlation network. We could integrate geographical information by employing Gephi along with GeoLayout, effectively superimposing the network graph onto the actual geographical locations. Visualizations of the six resulting networks for different correlation thresholds (0.3 to 0.8) are provided in the supplementary materials.

To choose the CT that fits with the rainfall data in Sungai Pahang, the resulting CT for the rainfall network should exhibit the characteristics of a small-world network and scale-free network.

Small-World Network

To assess the network type corresponding to different correlation thresholds, we evaluated three critical parameters: the Average Clustering Coefficient, the Average Shortest Path Length, and the Small-world Coefficient using equations (3) to (6). Table 2 summarizes these parameters.

Table 2. Small-world network properties across different correlation thresholds

Correlation Threshold	Average Clustering Coefficient	Average Shortest Path length	Small world Coefficient	Type of Network
0.3	0.968	1.328	-0.2150	Regular
0.4	0.902	1.372	-0.1731	Regular
0.5	0.813	1.420	-0.1088	Regular
0.6	0.742	1.522	-0.0850	Small World
0.7	0.638	1.703	-0.0508	Small World
0.8	0.286	1.519	0.3723	Random

The results indicate that the networks corresponding to correlation thresholds of 0.6 and 0.7 exhibit small-world properties, as evidenced by their small-world coefficient value close to zero. Conversely, networks formed at lower correlation thresholds (0.3, 0.4, 0.5) are predominantly of the 'Regular Network' type, characterised by high clustering but the negative value of the small-world attribute. Interestingly, the network at a threshold of 0.8 emerges as a 'Random' network, as indicated by its positive small-world coefficient, substantially low clustering coefficient, and high average shortest path length.

Scale-Free Network

In addition to assessing the small-world nature of the networks, we examined their scale-free properties. To accomplish this, the power-law coefficient (α) was computed for every correlation threshold. Table 3 presents a comprehensive summary of the given findings.

Our analysis suggests that the network at a correlation threshold of 0.7 exhibits scale-free properties, as indicated by its α value falling within the range of 2 to 3. All other networks did not conform to the scale-free model, as evidenced by their α values outside this range.

Table 3. Small-world network properties across different correlation thresholds

Correlation Threshold	α	Type of network
0.3	9.3374	Non-scale free
0.4	0.2971	Non-scale free
0.5	0.2724	Non-scale free
0.6	1.7255	Non-scale free
0.7	2.579	Scale-free
0.8	1.2955	Non-scale free

By integrating the findings on the small-world and scale-free attributes, it can be argued that the network, characterized by a correlation threshold 0.7, possesses notable potential in accurately portraying the rainfall network in the Sungai Pahang basin. This is primarily due to its manifestation of both small-world and scale-free features. This observation is consistent with previous scholarly publications that indicate hydrological networks frequently exhibit both types of features. Small-world networks are characterized by localized clusters of interactions and the ability to efficiently transmit information globally [69] [87]. Scale-free features imply that specific nodes, such as rainfall stations in our study, function as hubs, exhibiting significantly higher connections than other nodes [88]. The discoveries above possess significant significance for managing water resources within the region.

Network Analysis

A network analysis was conducted on the selected network corresponding to a correlation threshold of 0.7, which was identified as both a small-world and scale-free network. The network graph generated by Gephi is shown in Figure 5. The analysis focused on three key metrics: Degree Centrality, Closeness Centrality, and Betweenness Centrality. Table 4 outlines the centrality values for each of the 21 rainfall stations.

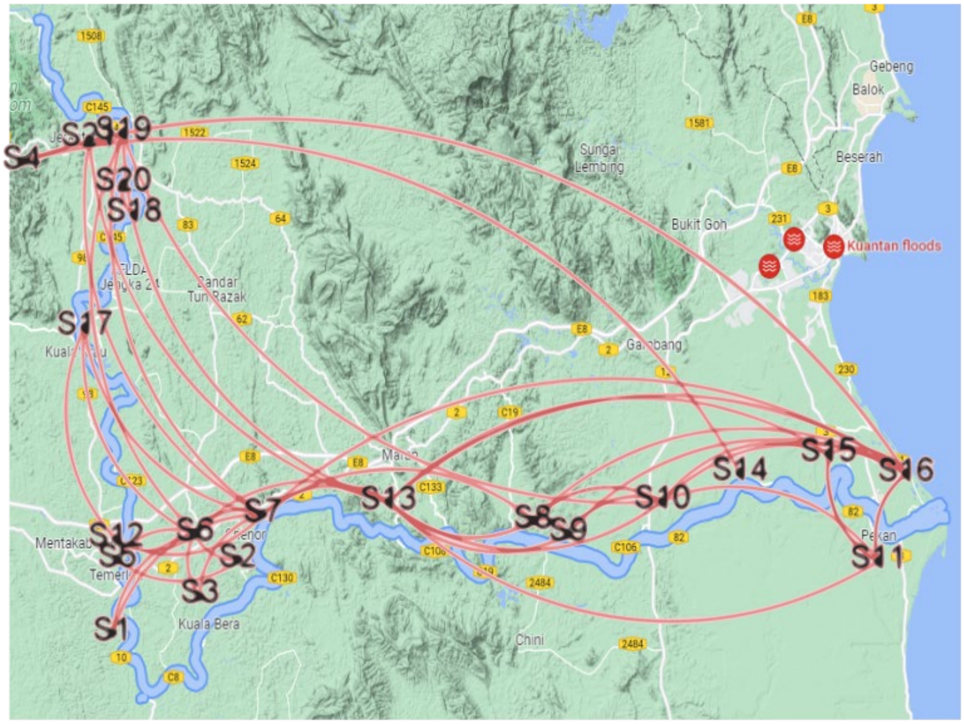

Figure 5. Network of sungai Pahang with threshold value 0.7

Table 4. Centrality metrics for rainfall stations in network at threshold 0.7

Station	Degree Centrality	Closeness Centrality	Betweenness Centrality
7	17	0.8696	24.018
13	16	0.8333	15.8662
19	15	0.7692	19.7735
6	14	0.7692	12.1561
16	12	0.7143	3.897
21	12	0.6897	4.97
10	10	0.6667	1.7317
11	10	0.6667	1.7317
8	10	0.6667	1.5579
20	10	0.6452	13.5972
9	9	0.625	8.2449
3	8	0.625	1.5673
18	8	0.6061	4.4314
12	8	0.6061	2.8688
14	8	0.6061	2.2309
17	8	0.6061	1.9306
15	8	0.5882	0.565
2	7	0.5882	2.4452
5	5	0.5263	0.4167
1	3	0.4762	0
4	2	0.4255	0

In a network of rainfall monitoring stations, the degree centrality of a station is a measure of how many other stations it is directly connected to [89]. This means that stations with a high degree centrality are connected to many other stations, while stations with low degree centrality are linked to only a few other stations [90]. Degree centrality is essential for rainfall monitoring stations because it is directly related to the information a station can receive and share. Stations with high degree centrality can receive information from many other stations, which helps improve the overall accuracy of their rainfall data.

Additionally, these stations can share their data with many different stations, which helps ensure that all stations in the network have access to the most up-to-date information. Based on the data presented in Table 4, Station 7 has the highest Degree Centrality, establishing it as a critical node within the network since it would have the most connections to other stations. Stations 13, 19, and 6 also demonstrate a high to moderate Degree Centrality, making them essential for the network's overall structure. On the other end of the spectrum, Stations 1 and 4 have the lowest Degree Centrality located in remote or less accessible areas (as shown in Figure 5), providing valuable data on rainfall patterns in these underrepresented regions. These stations can act as sentinels, detecting and recording localized weather events that the more extensive network may not capture.

The betweenness centrality of a station can be used to measure how important it is for transferring information about rainfall between different parts of the network. Stations with high betweenness centrality are more critical because they are at the crossroads of many different paths [91]. This means that they are more likely to be the first to receive information about rainfall in a particular area, and they are also more likely to be able to share this information with other stations. Regarding Betweenness Centrality, Station 7 again stands out as the most central node. A node with the highest betweenness centrality often acts as a bridge or mediator within the network. In this case, it would mean that this station is critical for connecting different parts of the network, effectively serving as a prominent 'route' for the flow of water or information. Stations 13 and 19 show moderate betweenness centrality, suggesting their importance as mediators or bridges within the network. Stations 1 and 4, like their low degree centrality, also have the lowest betweenness centrality.

Closeness centrality measures how close a station is to other stations in the network, on average [92]. Rainfall monitoring stations with high closeness centrality are well-connected to other stations, meaning they will likely have access to much rainfall data from different network parts. For closeness centrality, Station 7 once more reigns supreme, further solidifying its vital role in the network. This station would be ideal for quickly disseminating information or responses across the network. Since it is 'close' to all other stations, changes in hydrological variables detected here can be quickly relayed to other nodes. Moreover, other stations like 13 and 19 exhibit high closeness centrality, indicating their importance in rapid information dissemination within the network. Stations 1 and 4, consistent with the other metrics, show the lowest closeness centrality. This station is relatively distant from the other stations in the network, indicating that it might be isolated or less connected. Data from this station could be less effective in influencing or representing the overall network.

In addition, we explored the correlations between three measures of Centrality: degree centrality, betweenness centrality, and closeness centrality, by performing regression fit analyses [93]. The results shown in Figure 6, Figure 7, and Figure 8 indicate the interrelationships and potential predictive power between these centrality measures within the rainfall network.

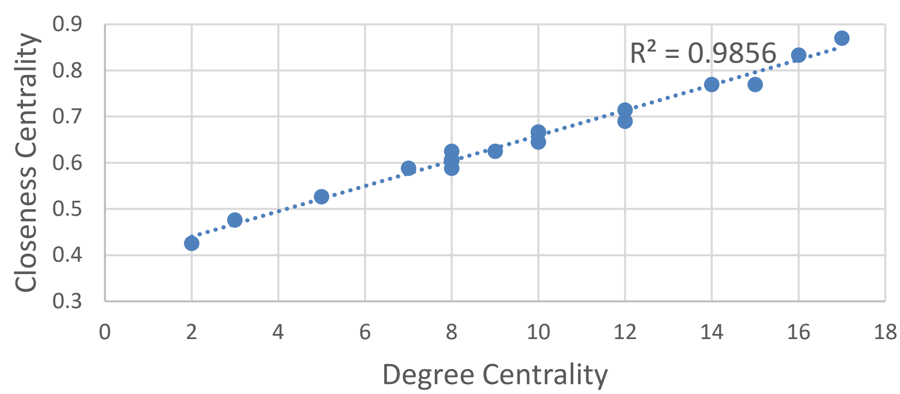

Figure 6. Relationship between degree centrality and closeness centrality for threshold network 0.7

Figure 6 shows a regression fit between degree and closeness centrality, yielding a very high R-square value of 0.9856, suggesting a strong positive correlation between these two measures. This implies that nodes with a higher number of direct connections (degree centrality) tend to be closer to all other nodes in the network (closeness centrality). This could be interpreted as highly connected stations having a more central role in the rainfall distribution network, being more influential in the overall connectivity of the system. In addition, no outlier exists since all points lie in a straight line due to the R square close to 1.

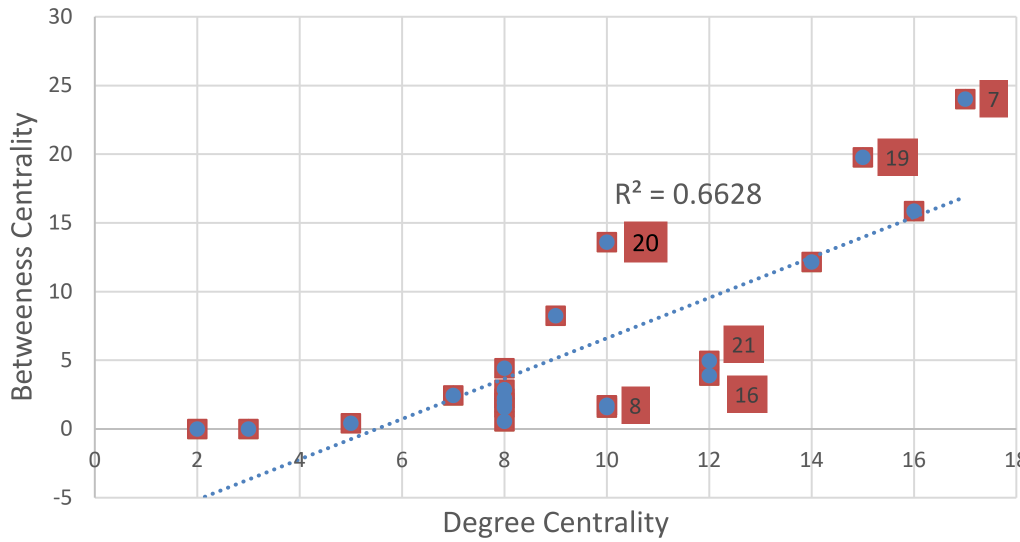

Figure 7. Relationship between degree centrality and betweenness centrality for threshold network 0.7

Figure 7 shows that a regression fit between degree and betweenness centrality produced an R-square value of 0.6628. While this value indicates a moderate positive correlation, it is significantly lower than the correlation between degree and closeness centrality. This suggests that although nodes with a higher degree centrality tend to have higher betweenness centrality, the predictive power is less specific [94]. This indicates that stations with many connections are not always the most significant in controlling the flow of rainfall distribution between other stations. In addition, five outlier stations exist where these values fall outside of the straight line. Stations were identified as outliers based on their divergence from the linear trend observed in the regression fit. This includes nodes with significantly different values in one centrality measure relative to others, suggesting unique roles in the network structure (e.g., locally central but globally peripheral).

Station 7 and 19 have a high degree centrality and high betweenness centrality. This implies that these stations can be key network dynamics rainfall stations. Another outlier is detected in Station 20, which has a moderate degree centrality and betweenness centrality. Unlike typical central nodes (high Centrality) or peripheral nodes (low Centrality), it neither dominates the network nor remains insignificant. This balance indicates its involvement in the network is more nuanced. Its centrality measures might reflect regional variability in rainfall distribution, where Station 20 is strategically important for specific areas or patterns but is not a primary driver of the overall network's connectivity.

Another outlier on Stations 16,8 and 21 has a high degree centrality with a low betweenness centrality. These three stations are directly connected to many other stations but do not lie on many of the shortest paths between stations. This means the station is a central hub for information exchange but is not crucial for routing information between different network parts [95]. The low betweenness centrality of the station suggests that there may be opportunities to optimize the network's communication protocols or station placement to make information flow more efficient. This could involve identifying alternative routes or prioritizing the maintenance of stations with higher betweenness centrality. While the node is important locally due to its high number of connections, it may have limited influence on the overall dynamics of the network. Its removal or failure might not significantly disrupt the network's connectivity, as it is not crucial for maintaining its cohesion.

The relationship between closeness centrality and betweenness centrality was evaluated, resulting in an R-square value of 0.6346, as shown in Figure 8. This moderate correlation suggests that nodes closer to all other nodes in the network do not always exhibit the highest control over the rainfall distribution. In addition, three outlier station exists where these values fall outside the straight line. Station 20 has a moderate betweenness centrality with moderate closeness centrality. The moderate betweenness centrality implies that Station 20 is somewhat influential in controlling the flow of information or water through the network, but not to the extent of being a critical juncture. Similarly, its moderate closeness centrality suggests it is relatively accessible to other stations but not the quickest node to disseminate information.

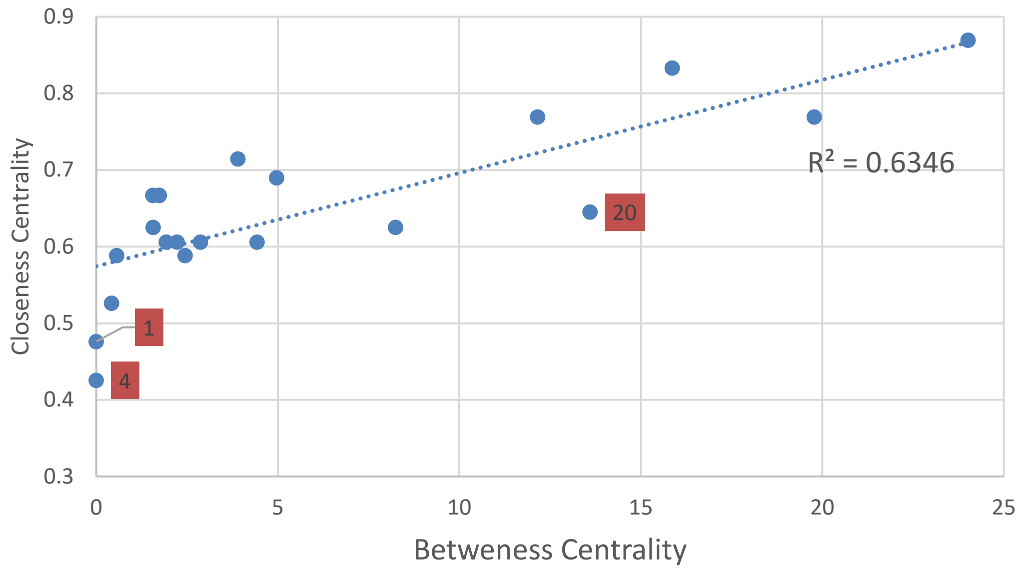

Figure 8. Relationship between closeness centrality and betweenness centrality for threshold network 0.7

Stations 1 and 4 are other outliers with the lowest betweenness and closeness centrality, as shown in Figure 8. The low values in betweenness and closeness centrality indicate that these stations are on the network's periphery. They are neither major connectors nor central in information dissemination within the network. Given their peripheral role, disruptions or failures at these stations would likely have a minimal impact on the overall network's functionality. While they may not contribute significantly to comprehensive network connectivity, they could be important for localized or specific monitoring purposes. For instance, they might capture unique microclimatic data or specific hydrological phenomena pertinent to their locations. Their status as outliers with low centrality measures may highlight areas where the network could be optimized. This could involve improving connectivity or reassessing the placement and role of these stations in the network. When analyzing data from the network, it's crucial to consider the unique positioning of these stations. Their data might need to be interpreted differently from that of more central stations because peripheral stations often capture localized microclimatic or hydrological phenomena that are not reflected in the more interconnected core of the network.

Discussion

The centrality measures revealed that Station 7 is a central hub within the network, with the highest values for degree, closeness, and betweenness centrality. This indicates its significance in the rainfall network for the Sungai Pahang basin. Understanding the central nodes is essential for practical management and invaluable for mathematical modelling.

Highly central nodes like Station 7 can be key network dynamics players. Their significant role can be leveraged in mathematical models to simulate scenarios such as extreme rainfall events, helping to fine-tune flood prediction algorithms and risk assessment tools. Researchers can potentially develop more accurate and computationally efficient models crucial for real-time flood forecasting and management strategies by focusing on these central nodes. Flood prediction is crucial for governments to develop reliable and accurate flood risk maps and plan for sustainable flood risk management [96].

Not only does Station 7 rank highly regarding various centrality measures, but its geographical location also seems somewhat central in Sungai Pahang, as shown in Figure 5. This enhances its importance as a node for capturing and predicting regional rainfall patterns. Geographical Centrality may reinforce its topological Centrality, making it a crucial point for advanced mathematical models that aim to represent real-world complexity in a networked system like a rainfall gauge network.

On the other end, Stations 1 and 4 had the lowest centrality values, suggesting they have a more peripheral role within this network. Looking at the latitude and longitude, stations 1 and 4 appear to be geographically farther from the rest, as shown in Figure 5. Even if these stations show the lowest centrality measures, their spatial dispersion can make them critical for capturing the variability in rainfall patterns across broader geographical regions. While highly central nodes like Station 7 are undeniably crucial for network robustness and predictive modelling, peripheral stations like Station 1 and Station 4 also have unique roles and should not be overlooked. These peripheral nodes can serve as sentinel stations, providing localized data that can capture anomalies or localized events not immediately apparent at more central nodes.

Regarding mathematical modelling, incorporating data from peripheral stations can add granularity and depth to simulations. While central nodes can be vital in capturing the 'big picture' dynamics, peripheral nodes contribute to the fine details, providing a more complete and accurate system representation. This is especially important for localized intervention strategies, where data from these stations may be vital in understanding micro-climates and smaller-scale hydrological features within the Sungai Pahang basin. Additionally, these stations can be crucial for capturing the initial conditions of less frequent but highly impactful events, like flash floods in specific sub-basins, which may not significantly affect the central nodes but have devastating local impacts.

Stations with mid-level centrality measures, such as Station 16 and Station 19, serve a unique role within the Sungai Pahang basin's rainfall network. They often act as bridges or gateways between the highly central and peripheral stations, facilitating the flow of information and influence throughout the network. Regarding network resilience, these mid-central stations are essential because they can act as secondary hubs if a primary, highly central node like Station 7 fails or is compromised. Their bridging role makes them vital for maintaining the integrity and efficiency of the network.

In mathematical modelling, including mid-level centrality nodes provides a more nuanced picture of network dynamics. They can be crucial in multi-scale models attempting to capture global and local behaviours within the network. By acknowledging the unique roles played by stations with varying degrees of Centrality, researchers can create more robust and accurate models to simulate real-world scenarios better.

Station 20 appears to be a specialized asset within the monitoring network. It offers the advantages of quick and representative data dissemination due to its high closeness centrality. Still, its low degree and betweenness centrality indicate that it functions more as an isolated, efficient unit rather than a central hub or control point. Its unique position could make it crucial for specific analyses or operational tasks despite its limited direct influence on the overall network.

This nuanced understanding of nodes with different centrality measures can help optimize the allocation of monitoring resources and inform more effective disaster management strategies across different scales, from local to basin-wide interventions.

Conclusion

This study sought to investigate the spatial and temporal dynamics of rainfall in Sungai Pahang, Malaysia, through the lens of network theory. Utilising a range of advanced analytical techniques and tools like Gephi and GeoLayout, we have constructed a network model that captures the complexities and interdependencies between 21 different monitoring stations. By evaluating the network's properties, particularly its small-world and scale-free characteristics and centrality measures for each station, we have unearthed insights that could significantly enhance our understanding of rainfall distribution and variability across the region.

Our findings indicate that the network displays small-world and scale-free properties, underlining its robustness and vulnerability. High centrality nodes are pivotal in information flow and key stations for effective monitoring. For instance, stations with high Centrality are pivotal for information dissemination and crucial network hubs. Conversely, nodes with low centrality measures exhibit limited influence on the overall system but may offer vital localised data.

Specifically, the unique characteristics of Station 20—a high closeness centrality but low degree and betweenness centrality—highlight its specialised role in the network. It acts as an efficient, isolated unit that can quickly relay information across the network, making it an asset for specific analyses and operational tasks.

The research fills an existing gap by thoroughly investigating the challenges and limitations of the current state of the rainfall monitoring network in Malaysia. It provides a foundation for future studies and offers actionable insights for policymakers and practitioners involved in climate monitoring, flood prediction, and water resource management.

In closing, this study contributes to the academic understanding of spatial rainfall dynamics and has significant practical implications. Future work can extend this model to include additional environmental variables, enhancing its predictive capabilities. Furthermore, the methodology can be adapted to other ecological monitoring networks, underlining its versatility and broader relevance.

The paper also opens the door for future research to investigate other characteristics of complex networks, such as resilience and modularity, and their applications in climate science and hydrology. This multifaceted approach offers a robust framework for comprehending and managing the intricate dynamics of rainfall and its associated risks in the region.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

We would like to express our sincere appreciation to Jabatan Pengairan dan Saliran (JPS) Malaysia for their kind support in providing the data used in this research. The access to their data has been crucial in enabling us to carry out this study effectively.

References

[1] Bartzke, G. S., Ogutu, J. O., Mukhopadhyay, S., Mtui, D., Dublin, H. T., & Piepho, H. P. (2018). Rainfall trends and variation in the Maasai Mara ecosystem and their implications for animal population and biodiversity dynamics. *PLoS One, 13*(9), e0202814. https://doi.org/10.1371/journal.pone.0202814

[2] Jale, J. da S., Xavier Júnior, S. F. A., Xavier, É. F. M., Stošić, T., Stošić, B., & Ferreira, T. A. E. (2019). Application of Markov chain on daily rainfall data in Paraíba-Brazil from 1995–2015. *Acta Scientiarum. Technology, 41*(1), 1–10. https://doi.org/10.4025/actascitechnol.v41i1.37186

[3] Qin, H. P., Li, Z. X., & Fu, G. (2013). The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management, 129*, 577–585. https://doi.org/10.1016/j.jenvman.2013.08.026

[4] Hettiarachchi, S., Wasko, C., & Sharma, A. (2017). Increase in urban flood risk resulting from climate change – The role of storm temporal patterns. *Hydrology and Earth System Sciences Discussions*, 1–28.

[5] Huong, H. T. L., & Pathirana, A. (2013). Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam. *Hydrology and Earth System Sciences, 17*(1), 379–394. https://doi.org/10.5194/hess-17-379-2013

[6] Hamed, K. H. (2008). Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. *Journal of Hydrology, 349*(3–4), 350–363. https://doi.org/10.1016/j.jhydrol.2007.11.009

[7] Hamed, K. H. (2009). Exact distribution of the Mann-Kendall trend test statistic for persistent data. *Journal of Hydrology, 365*(1–2), 86–94. https://doi.org/10.1016/j.jhydrol.2008.11.024

[8] Rahman, M. A., Yunsheng, L., & Sultana, N. (2017). Analysis and prediction of rainfall trends over Bangladesh using Mann–Kendall, Spearman's rho tests and ARIMA model. *Meteorology and Atmospheric Physics, 129*(4), 409–424. https://doi.org/10.1007/s00703-016-0479-4

[9] Hu, Y., Yue, S., & Wang, C. Y. (2016). Water Tower of the Yellow River in a changing climate: Toward an integrated assessment. *Water Resources Management, 13*(3), 245–259. http://repository.tudelft.nl/assets/uuid:f53620b8-3ce5-4f4b-9b48-a27d8c0496d5/2014_UNESCO-IHE_PHD_Thesis_YURONG_HU.pdf

[10] Kwon, T., Park, J., & Park, J. (2021). Design of optimal rainfall monitoring network using radar and road networks. *Entropy, 23*(378), 1–11.

[11] Yeh, H. C., Chen, Y. C., Chang, C. H., Ho, C. H., & Wei, C. (2017). Rainfall network optimization using radar and entropy. *Entropy, 19*(10), 1–13. https://doi.org/10.3390/e19100553

[12] Zhao, X., Liu, F., Wang, J., & Li, T. (2017). Evaluating influential nodes in social networks by local centrality with a coefficient. *ISPRS International Journal of Geo-Information, 6*(2), 35. https://doi.org/10.3390/ijgi6020035

[13] Bahari, N. N., Expert, P., & Razak, F. A. (2023). An analysis of actors in Malay films: Small worlds, centralities and genre diversity. *Mathematics, 11*(5), 1252. https://doi.org/10.3390/math11051252

[14] Zuhud, D. A. Z., Musa, M. H., Ismail, M., Bahaludin, H., & Razak, F. A. (2022). The causality and uncertainty of the COVID-19 pandemic to Bursa Malaysia Financial Services Index's constituents. *Entropy, 24*(8), 1–30. https://doi.org/10.3390/e24081100

[15] Deng, Y., Li, Q., & Lu, Y. (2015). A research on subway physical vulnerability based on network theory and FMECA. *Safety Science, 80*, 127–134. https://doi.org/10.1016/j.ssci.2015.07.019

[16] Wang, Z., Bhamra, R. S., Wang, M., Xie, H., & Yang, L. (2020). Critical hazards identification and prevention of cascading escalator accidents at metro rail transit stations. *International Journal of Environmental Research*

and Public Health, 17(10), 3400. https://doi.org/10.3390/ijerph17103400
[17] Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5–13. https://doi.org/10.1002/wps.20375
[18] Biane, C., Delaplace, F., & Klaudel, H. (2016). Networks and games for precision medicine. *BioSystems, 150*, 52–60. https://doi.org/10.1016/j.biosystems.2016.08.006
[19] Rasidi, N. M., Bakar, S. A., & Razak, F. A. (2016). Modelling the correlation of PM10 concentration and location of air quality monitoring stations in Malaysia using network method. *AIP Conference Proceedings, 1750*(1), 020010. https://doi.org/10.1063/1.4954560
[20] Rodríguez-Alarcón, R., & Lozano, S. (2019). A complex network analysis of Spanish river basins. *Journal of Hydrology, 578*, 124065. https://doi.org/10.1016/j.jhydrol.2019.124065
[21] Rodríguez-Alarcón, R., & Lozano, S. (2022). Complex network modeling of a river basin: An application to the Guadalquivir River in Southern Spain. *Journal of Hydroinformatics, 24*(3), 559–573. https://doi.org/10.2166/HYDRO.2022.148
[22] Majeed, A., & Rauf, I. (2020). Graph theory: A comprehensive survey about graph theory applications in computer science and social networks. *Inventions, 5*(1), 10. https://doi.org/10.3390/inventions5010010
[23] Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47
[24] Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports, 544*(1), 1–122. https://doi.org/10.1016/j.physrep.2014.07.001
[25] Tiwari, S., Jha, S. K., & Singh, A. (2020). Quantification of node importance in rain gauge network: Influence of temporal resolution and rain gauge density. *Scientific Reports, 10*(1), 1–17. https://doi.org/10.1038/s41598-020-66363-5
[26] Qi, M., Feng, M., Sun, T., & Yang, W. (2016). Resilience changes in watershed systems: A new perspective to quantify long-term hydrological shifts under perturbations. *Journal of Hydrology, 539*, 281–289. https://doi.org/10.1016/j.jhydrol.2016.05.039
[27] Xu, Y., Lu, F., Zhu, K., Song, X., & Dai, Y. (2020). Exploring the clustering property and network structure of a large-scale Basin's precipitation network: A complex network approach. *Water, 12*(6), 1739. https://doi.org/10.3390/W12061739
[28] Sivakumar, B., & Woldemeskel, F. M. (2015). A network-based analysis of spatial rainfall connections. *Environmental Modelling & Software, 69*, 55–62. https://doi.org/10.1016/j.envsoft.2015.02.020
[29] Jha, S. K., Zhao, H., Woldemeskel, F. M., & Sivakumar, B. (2015). Network theory and spatial rainfall connections: An interpretation. *Journal of Hydrology, 527*, 13–19. https://doi.org/10.1016/j.jhydrol.2015.04.035
[30] Boers, N., Bookhagen, B., Marwan, N., Kurths, J., & Marengo, J. (2013). Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System. *Geophysical Research Letters, 40*(16), 4386–4392. https://doi.org/10.1002/grl.50681
[31] Yoon, S. S., & Lee, B. (2017). Effects of using high-density rain gauge networks and weather radar data on urban hydrological analyses. *Water, 9*(12), 931. https://doi.org/10.3390/w9120931
[32] Chwala, C., & Kunstmann, H. (2019). Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges. *Wiley Interdisciplinary Reviews: Water, 6*(2), e1337. https://doi.org/10.1002/WAT2.1337
[33] Scarsoglio, S., Laio, F., & Ridolfi, L. (2013). Climate dynamics: A network-based approach for the analysis of global precipitation. *PLoS One, 8*(8), e71129. https://doi.org/10.1371/journal.pone.0071129
[34] Su, Y., & Jiang, X. (2023). Simulation research on knowledge flow in a collaborative innovation network. *Expert Systems, 40*(7), e13280. https://doi.org/10.1111/exsy.13280
[35] Holme, P., & Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical Review E, 65*(2), 026107. https://doi.org/10.1103/PhysRevE.65.026107
[36] Pereira, T. (2010). Hub synchronization in scale-free networks. *Physical Review E, 82*(3), 036201. https://doi.org/10.1103/PhysRevE.82.036201
[37] Agarwal, A., Marwan, N., Maheswaran, R., Merz, B., & Kurths, J. (2018). Quantifying the roles of single stations within homogeneous regions using complex network analysis. *Journal of Hydrology, 563*, 802–810. https://doi.org/10.1016/j.jhydrol.2018.06.050
[38] Naufan, I., Sivakumar, B., Woldemeskel, F. M., Raghavan, S. V., Vu, M. T., & Liong, S. Y. (2018). Spatial connections in regional climate model rainfall outputs at different temporal scales: Application of network theory. *Journal of Hydrology, 556*, 1232–1243. https://doi.org/10.1016/j.jhydrol.2017.05.029
[39] Konapala, G., & Mishra, A. (2017). Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *Journal of Hydrology, 555*, 600–620. https://doi.org/10.1016/j.jhydrol.2017.10.033
[40] Estrada, E., Hatano, N., & Benzi, M. (2012). The physics of communicability in complex networks. *Physics Reports, 514*(3), 89–119. https://doi.org/10.1016/j.physrep.2012.01.006
[41] De la Cruz Cabrera, O., Jin, J., Noschese, S., & Reichel, L. (2022). Communication in complex networks. *Applied Numerical Mathematics, 172*, 186–205. https://doi.org/10.1016/j.apnum.2021.10.005
[42] Emad, A., & Milenkovic, O. (2014). CaSPIAN: A causal compressive sensing algorithm for discovering directed interactions in gene networks. *PLoS One, 9*(3), e90781. https://doi.org/10.1371/journal.pone.0090781
[43] Zemla, J. C., & Austerweil, J. L. (2018). Rogers-T. *Computational Brain & Behavior*.
[44] Rahim, N. Z. A., Razak, F. A., Noorani, M. S. M., & Yusof, N. M. (2023). Comparing friends and peer tutors amidst COVID-19 using social network analysis. *Mathematics, 11*(4), 1053. https://doi.org/10.3390/math11041053
[45] Sapini, M. L., Noorani, M. S. M., Razak, F. A., Alias, M. A., & Yusof, N. M. (2022). Understanding published literatures on persistent homology using social network analysis. *Malaysian Journal of Fundamental and Applied Sciences, 18*(4), 413–429. https://doi.org/10.11113/mjfas.v18n4.2418
[46] Toubiana, D., & Maruenda, H. (2021). Guidelines for correlation coefficient threshold settings in metabolite

correlation networks exemplified on a potato association panel. *BMC Bioinformatics, 22*(1), 1–14. https://doi.org/10.1186/s12859-021-03994-z

[47] Kumar, S., & Deo, N. (2012). Correlation and network analysis of global financial indices. *Physical Review E, 86*(2), 026101. https://doi.org/10.1103/PhysRevE.86.026101

[48] Suhaila, J., Deni, S. M., Zin, W. Z. W., & Jemain, A. A. (2010). Spatial patterns and trends of daily rainfall regime in Peninsular Malaysia during the southwest and northeast monsoons: 1975–2004. *Meteorology and Atmospheric Physics, 110*(1), 1–18. https://doi.org/10.1007/s00703-010-0108-6

[49] Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics, 52*(3), 522–555. https://doi.org/10.1002/2014RG000464

[50] Rahman, M., & Abdullah, S. A. (2023). Long-term trend and variability of rainfall in Bangladesh. *Journal of Mathematics, Technology and Computer Mathematics, 2*(1). https://doi.org/10.33140/jmtcm.02.01.03

[51] Plouffe, C. C. F., Robertson, C., & Chandrapala, L. (2015). Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka. *Environmental Modelling & Software, 67*, 57–71. https://doi.org/10.1016/j.envsoft.2015.01.011

[52] Quilty, J., & Adamowski, J. (2021). A maximal overlap discrete wavelet packet transform integrated approach for rainfall forecasting – A case study in the Awash River Basin (Ethiopia). *Environmental Modelling & Software, 144*, 105119. https://doi.org/10.1016/j.envsoft.2021.105119

[53] Famili, A., Shen, W.-M., Weber, R., & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. *Intelligent Data Analysis, 1*(1), 3–23. https://doi.org/10.3233/IDA-1997-1102

[54] Milani, A. M. P., Loges, L. A., Paulovich, F. V., & Manssour, I. H. (2021). PrAVA: Preprocessing profiling approach for visual analytics. *Information Visualization, 20*(2–3), 101–122. https://doi.org/10.1177/14738716211021591

[55] Zhu, C., & Gao, D. (2016). Influence of data preprocessing. *Journal of Computational Science and Engineering, 10*(2), 51–60.

[56] Park, J., Park, J., & Choi, Y. (2022). Identification of research communities of environmental engineering and their evolution using coauthor network analysis. *Environmental Modelling & Software, 149*, 105320. https://doi.org/10.1016/j.envsoft.2022.105320

[57] Hardouin, J.-B., Conroy, R., & Sébille, V. (2011). Imputation by the mean score should be avoided when validating a patient reported outcomes questionnaire by a Rasch model in presence of informative missing data. *BMC Medical Research Methodology, 11*(1), 105. https://doi.org/10.1186/1471-2288-11-105

[58] Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

[59] Cheema, J. (2014). A review of missing data handling methods in education research. *Review of Educational Research, 84*(4), 487–508. https://doi.org/10.3102/0034654314532697

[60] Gu, J., Jia, Z., Cai, T., Song, X., & Mahmood, A. (2023). Dynamic correlation adjacency-matrix-based graph neural networks for traffic flow prediction. *Sensors, 23*(6), 2897. https://doi.org/10.3390/s23062897

[61] Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A, 77*(8), 733–742. https://doi.org/10.1002/cyto.a.20896

[62] Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., & Faloutsos, C. (2008). Epidemic thresholds in real networks. *ACM Transactions on Information and System Security, 10*(4), 1–26. https://doi.org/10.1145/1284680.1284681

[63] Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology, 4*(1), Article 17. https://doi.org/10.2202/1544-6115.1128

[64] Yan, X., Jeub, L. G. S., Flammini, A., Radicchi, F., & Fortunato, S. (2018). Weight thresholding on complex networks. *Physical Review E, 98*(4), 042304. https://doi.org/10.1103/PhysRevE.98.042304

[65] Wang, H. Q., & Tsai, C. J. (2013). CorSig: A general framework for estimating statistical significance of correlation and its application to gene co-expression analysis. *PLoS One, 8*(10), e77429. https://doi.org/10.1371/journal.pone.0077429

[66] Baudewig, J., Dechent, P., Merboldt, K. D., & Frahm, J. (2003). Thresholding in correlation analyses of magnetic resonance functional neuroimaging. *Magnetic Resonance Imaging, 21*(10), 1121–1130. https://doi.org/10.1016/j.mri.2003.08.013

[67] Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media, 3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

[68] Pavlopoulos, G. A., Paez-Espino, D., Kyrpides, N. C., & Iliopoulos, I. (2017). Empirical comparison of visualization tools for larger-scale network analysis. *Advances in Bioinformatics, 2017*, 1278932. https://doi.org/10.1155/2017/1278932

[69] Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters, 87*(19), 198701. https://doi.org/10.1103/PhysRevLett.87.198701

[70] Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Computational Biology, 3*(2), e17. https://doi.org/10.1371/journal.pcbi.0030017

[71] Nandi, A. K., Sumana, A., & Bhattacharya, K. (2014). Social insect colony as a biological regulatory system: Modelling information flow in dominance networks. *Journal of the Royal Society Interface, 11*(101), 20140951. https://doi.org/10.1098/rsif.2014.0951

[72] Masuda, N., Sakaki, M., Ezaki, T., & Watanabe, T. (2018). Clustering coefficients for correlation networks. *Frontiers in Neuroinformatics, 12*, 7. https://doi.org/10.3389/fninf.2018.00007

[73] Pan, R. K., & Saramäki, J. (2011). Path lengths, correlations, and centrality in temporal networks. *Physical Review E, 84*(1), 016105. https://doi.org/10.1103/PhysRevE.84.016105

[74] Humphries, M. D., Gurney, K., & Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not

scale-free, network. *Proceedings of the Royal Society B: Biological Sciences, 273*(1585), 503–511. https://doi.org/10.1098/rspb.2005.3354

[75] Humphries, M. D., & Gurney, K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One, 3*(4), e0002051. https://doi.org/10.1371/journal.pone.0002051

[76] Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

[77] Chen, A. X., *et al*. (2020). Scale-free structure of cancer networks and their vulnerability to hub-directed combination therapy. *bioRxiv*, 2020.07.01.159657. https://doi.org/10.1101/2020.07.01.159657

[78] Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications, 10*(1), 1017. https://doi.org/10.1038/s41467-019-08746-5

[79] Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science, 325*(5939), 412–413. https://doi.org/10.1126/science.1173299

[80] Herdagdelen, A., Aygun, E., & Bingol, H. (2023). Measuring preferential attachment in a hyper-textual dictionary reference network: Eksi Sözlük. *arXiv*.

[81] Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

[82] Ibrahim, M. H., Missaoui, R., & Vaillancourt, J. (2020). Cross-face centrality: A new measure for identifying key nodes in networks based on formal concept analysis. *IEEE Access, 8*, 206901–206913. https://doi.org/10.1109/ACCESS.2020.3038302

[83] Lin, S., Li, S., & Su, Y. (2022). Global importance analysis of nodes based on complex network theory. *Journal of Physics: Conference Series, 2290*(1), 012094. https://doi.org/10.1088/1742-6596/2290/1/012094

[84] Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

[85] Borrett, S. R. (2013). Throughflow centrality is a global indicator of the functional importance of species in ecosystems. *Ecological Indicators, 32*, 182–196. https://doi.org/10.1016/j.ecolind.2013.03.014

[86] Wei, X., & Chen, W. (2019). How does a firm's previous social network position affect innovation? Evidence from Chinese listed companies. *Sustainability, 11*(4), 1191. https://doi.org/10.3390/su11041191

[87] Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics, 1*(1), 3. https://doi.org/10.1186/1753-4631-1-3

[88] Wang, Y., Hu, Y., Di, Z., & Fan, Y. (2011). The effect of hub nodes on the community structure in scale-free networks. *Physica A: Statistical Mechanics and Its Applications, 390*(21–22), 4027–4033. https://doi.org/10.1016/j.physa.2011.06.031

[89] Giancarlo, R., Greco, D., Landolina, F., & Rombo, S. E. (2018). Network centralities and node ranking. *Encyclopedia of Bioinformatics and Computational Biology, 1*, 950–957. https://doi.org/10.1016/B978-0-12-809633-8.20425-1

[90] Bródka, P., Skibicki, K., Kazienko, P., & Musiał, K. (2011). A degree centrality in multi-layered social network. *Proceedings of the 2011 International Conference on Computational Aspects of Social Networks (CASoN)*, 237–242. https://doi.org/10.1109/CASON.2011.6085951

[91] Gregersen, I. B., Madsen, H., Rosbjerg, D., & Arnbjerg-Nielsen, K. (2013). A spatial and nonstationary model for the frequency of extreme rainfall events. *Water Resources Research, 49*(1), 127–136. https://doi.org/10.1029/2012WR012570

[92] Okamoto, K., Chen, W., & Li, X. Y. (2008). Ranking of closeness centrality for large-scale social networks. *Lecture Notes in Computer Science, 5059*, 186–195. https://doi.org/10.1007/978-3-540-69311-6_21

[93] Li, C., Li, Q., Van Mieghem, P., Stanley, H. E., & Wang, H. (2015). Correlation between centrality metrics and their application to the opinion model. *European Physical Journal B, 88*(3), 65. https://doi.org/10.1140/epjb/e2015-50671-y

[94] van Elteren, C., Quax, R., & Sloot, P. (2022). Dynamic importance of network nodes is poorly predicted by static structural features. *Physica A: Statistical Mechanics and Its Applications, 593*, 126889. https://doi.org/10.1016/j.physa.2022.126889

[95] Younis, M., Bangad, M., & Akkaya, K. (2003). Base-station repositioning for optimized performance of sensor networks. *IEEE Vehicular Technology Conference, 58*(5), 2956–2960. https://doi.org/10.1109/vetecf.2003.1286165

[96] Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536. https://doi.org/10.3390/w10111536