IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 3 April 2024, accepted 23 April 2024, date of publication 26 April 2024, date of current version 6 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3393968

RESEARCH ARTICLE

Data-Driven Network Connectivity Analysis: An Underestimated Metric

JUNXIANG XU AND DIVYA JAYAKUMAR NAIR
Research Centre for Integrated Transport Innovation (rCITI), School of Civil and Environmental Engineering, University of New South Wales (UNSW), Sydney, NSW 2052, Australia

Corresponding authors: Junxiang Xu (junxiang.xu@unsw.edu.au) and Divya Jayakumar Nair (d.jayakumarnair@unsw.edu.au)

This work was supported by the 2023 Humanitarian Engineering Higher Degree Research (HDR) Grant from the Humanitarian Engineering Champions Committee, University of New South Wales (UNSW).

ABSTRACT In the network structure analysis, we explore an underestimated key metric, the Relative Size of Largest Connected Component (RSLCC) and demonstrate its importance in post-disaster network connectivity assessment. RSLCC was first investigated in the study of complex network structures but remains largely unexplored in terms of analysis within a specific application domain such as scenarios in transportation networks, wireless networks, communication networks, power networks, etc. Through the research presented in this paper, we not only prove that this metric is underestimated, but also design 7 methods to predict the value of this metric, with a Deep Neural Network (DNN) prediction accuracy of more than 99%. This study focuses on the assessment and analysis of post-disaster network connectivity, by exploring how the RSLCC, a key metric of network connectivity, can be used to efficiently predict and assess network connectivity in a disaster scenario, specifically, the approximate network connectivity value can be predicted simply by knowing the number of connected edges in the pre-disaster network and the number of connected edges in the post-disaster network. To achieve this, firstly, a sufficiently large-scale 100,000 datasets containing the values of attributes related to the network structure is prepared. Secondly, based on the preprocessing of the data, principal component analysis and variance contribution analysis are carried out, and the metric with the highest contribution to the principal component is approximated as the network connectivity. The next step is the prediction process, Network Disruption Degree (NDD) is chosen as the independent variable. since it is best to choose an extremely simple metric as the independent variable for prediction, rather than all network structure-related metrics, this paper demonstrates that it is possible to get satisfactory prediction results with this metric. It is found that NDD prediction methods have the highest prediction accuracy but take the longest run time and require training data of a sufficiently large size. If the prediction is done in small-size data, then Random Forest Regression (RFR) is proven to have the highest prediction accuracy. Although the network connectivity metric proposed in this paper is only an approximation, it provides good directions for simplifying the network connectivity analysis and the use of this metric for the study of practical modelling problems is also highly interpretable.

INDEX TERMS Network connectivity metric, network structure analysis, large-scale data, relative size of largest connected component (RSLCC), network disruption degree (NDD), deep neural network (DNN).

I. INTRODUCTION

In recent years, the frequency and complexity of disaster events have been on the rise globally, particularly in the case of natural disasters such as earthquakes, floods, and hurricanes. These events have inflicted significant impacts and threats upon society, economies, and the environment. To cite a few poignant examples, on December 26, 2004, an earthquake triggered a tsunami in the Indian Ocean, affecting 14 countries and regions and resulting in substantial loss of life and property damage [1]. On May 12, 2008, an 8.0 magnitude earthquake struck Wenchuan in

The associate editor coordinating the review of this manuscript and approving it for publication was Jiajie Fan.

China's Sichuan Province, causing nearly 70,000 fatalities and affecting a total of 46.256 million people [2]. On March 11, 2011, a magnitude 9.0 earthquake triggered a massive tsunami that severely impacted Japan's northeastern region, also leading to the Fukushima nuclear incident [3]. In late 2019 and early 2020, Australia experienced severe wildfires that devastated extensive forest areas, causing the loss of flora and fauna and contributing to deteriorating air quality [4]. According to ABC Australia, In July 2022, the Australian state of New South Wales was struck by heavy rainfall and subsequent flooding for the fourth time in 18 months, compelling approximately 50,000 people to evacuate and resulting in an estimated economic loss of around 3.5 billion US dollars. In this context, network connectivity has become particularly crucial, as networks play essential roles in post-disaster tasks such as rescue operations, information dissemination, and resource allocation. However, due to the unpredictability and uncertainty of disasters, network connectivity often faces challenges that can result in disruptions, delayed information flow, and inefficiencies in disaster response. Hence, understanding the changes in network connectivity is of profound significance for predicting network robustness, studying network resilience, responding to emergencies, optimizing network design, and analysing network disaster propagation.

The study of network connectivity problems is part of network reliability analysis, network reliability research is concerned with how to make the network maintain stable functionality and connectivity in the face of a variety of internal and external disturbances, attacks, failures, etc. [5], and network connectivity, along with network robustness and network resilience, are considered to be the three main components of network reliability research [6], [7]. We give **Figure 1** to reflect the correlation between these three concepts.

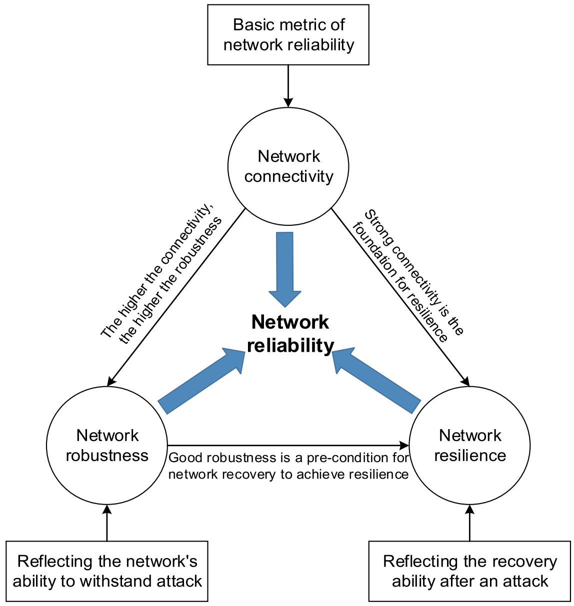


FIGURE 1. Relationship between network connectivity, network robustness, and network resilience. (This figure responds to three research perspectives on network reliability research and belongs to the macroscopic knowledge structure combing.)

Figure 1 illustrates the three key areas of network reliability research, whereas we will focus on network connectivity in this study and will not explore research problems about network resilience and network robustness. In practice, network connectivity is affected by many factors, such as node failure, edge disruption, and external interference. Therefore, understanding the changes in network connectivity under different scenarios and attacks is of important significance for analysing network robustness, studying network resilience, coping with emergencies, optimizing network design, and exploring network disaster propagation, etc. [8], [9]. However, it is worth stating that the network connectivity focused on in this study does not refer to a specific network, such as a transportation network, power network, energy network, information network, neural network, etc. Instead, from the perspective of network structural attributes, using techniques such as network structural analysis methodology, statistical analysis methodology, artificial intelligence algorithms, etc., we are attempting to find an approximate solution for predicting the network connectivity, regardless of whether it is a physical network or a virtual network. More specifically, we provide **Figure 2** to explain more clearly the motivation of this study. It should be further clarified that **Figure 2** is representative of our research idea for this study, is not based on previous research work, and is given only to clearly interpret our motivation for this study.

We conjecture that when the data of network structural attributes is large enough, the proximity metrics related to network connectivity can be found through data analysis to accomplish the prediction, this idea seems simple, but nothing has been attempted in academics so far, and we strive to get some universal insights through the research in this paper. The challenges of this research are threefold: (1) While network structure metrics values can be computationally obtained, there is no exact value for network connectivity, so how can training and prediction be done in the absence of historical connectivity values? (2) How to find the metrics that are close to correlating with network connectivity, is it a network structure metric or a combination of multiple metrics? (3) Which prediction approach to choose and how to prove that approach has high prediction accuracy?

Based on the above introduction, the rest of the paper is organized as follows. Section II is the literature review, in which the network connectivity concepts, metrics and quantitative approaches are reviewed. Section III is experimental design and preparation, which will introduce the selection of metrics, data preparation and the overall research framework. Section IV is based on a statistical research approach to find an approximate network connectivity metric. Section V uses several prediction methods to predict the approximate network connectivity metric and compare the prediction accuracy. Section VI replaces the dataset

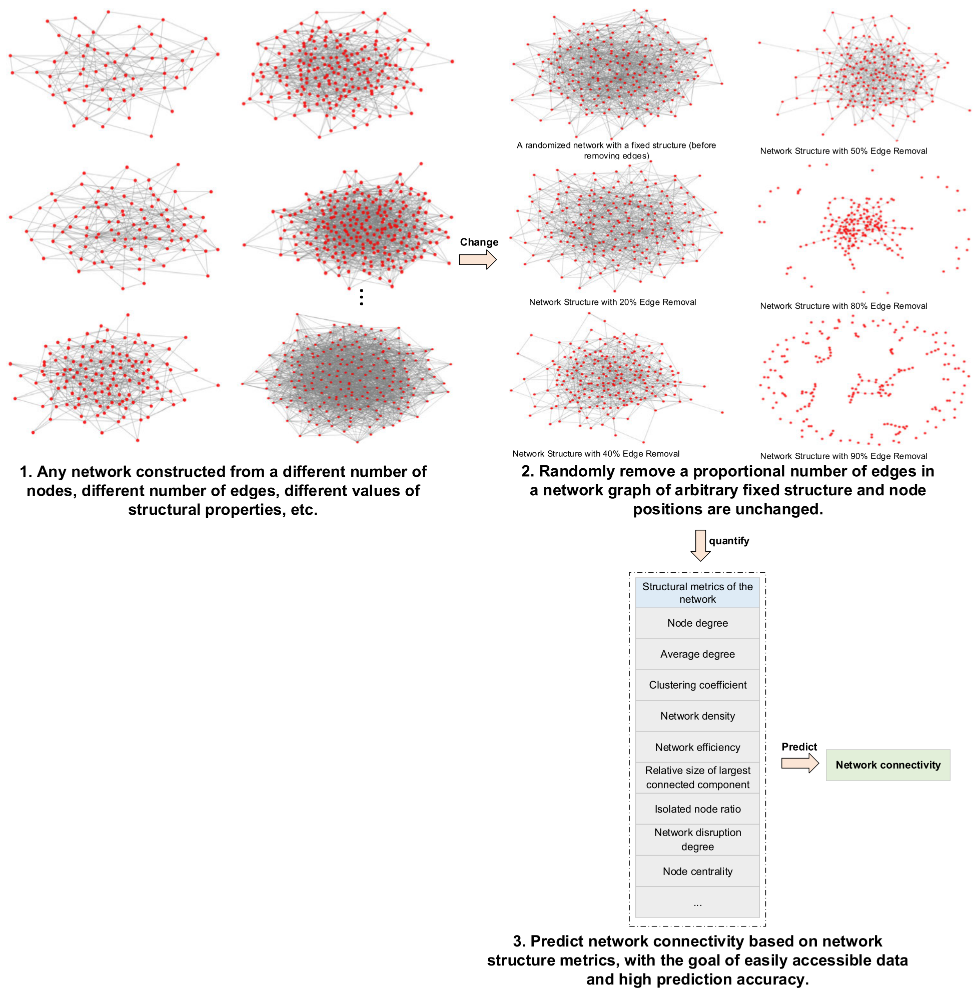


FIGURE 2. Clarification chart of the motivation for this study. (Generate network structure graphs using NetworkX in Python, and set the random seed np.random.seed() makes the structure of the generated network unique in step 2, and this figure shows the ideas and motivation for our research)

for secondary validation of the prediction methodology. Section VI is the conclusion of this paper.

II. LITERATURE REVIEW

The concept of network connectivity is not subordinate to any particular discipline, as the concept has a wide range of applications, focusing on dozens of research areas such as neuro medicine [10], information and communication sciences [11], transport network planning [12], complex network systems [13], etc. Network connectivity has different definitions in different disciplines. The core issue is that the research approaches are different due to different metrics. The network connectivity calculation is not based on a specific mathematical equation or parameter but is generally reflected through one or a series of metrics. In this section, we organize the classification of network connectivity metrics in all the existing disciplines as shown in **Figure 3**.

Although the RSLCC metric has potential applications in a variety of fields such as the above, our study focuses on exploring the network connectivity metric application in disaster scenarios to enrich the current research insights in assessing network connectivity after a disaster.

The existing literature exploring network connectivity in disaster scenarios covers a wide range of research areas including disaster management, network science, and emergency response. Barabási et al. [14] explored the topology of worldwide networks, and their proposed scale-free network model is an important insight for understanding and analysing network connectivity under the influence of disasters. Little [15] investigated how to control cascading failures and understand the vulnerability of interconnected infrastructures. Their important contribution is to provide support for proposing strategies to maintain network connectivity in disaster scenarios. Bruneau et al. [16] provided a framework to quantitatively assess and enhance the resilience of communities after earthquakes, which includes network connectivity as one of the key factors in assessing community resilience. Boccaletti et al. [17] conducted a review study to introduce the structural properties of complex networks with several key network metrics, including network connec-

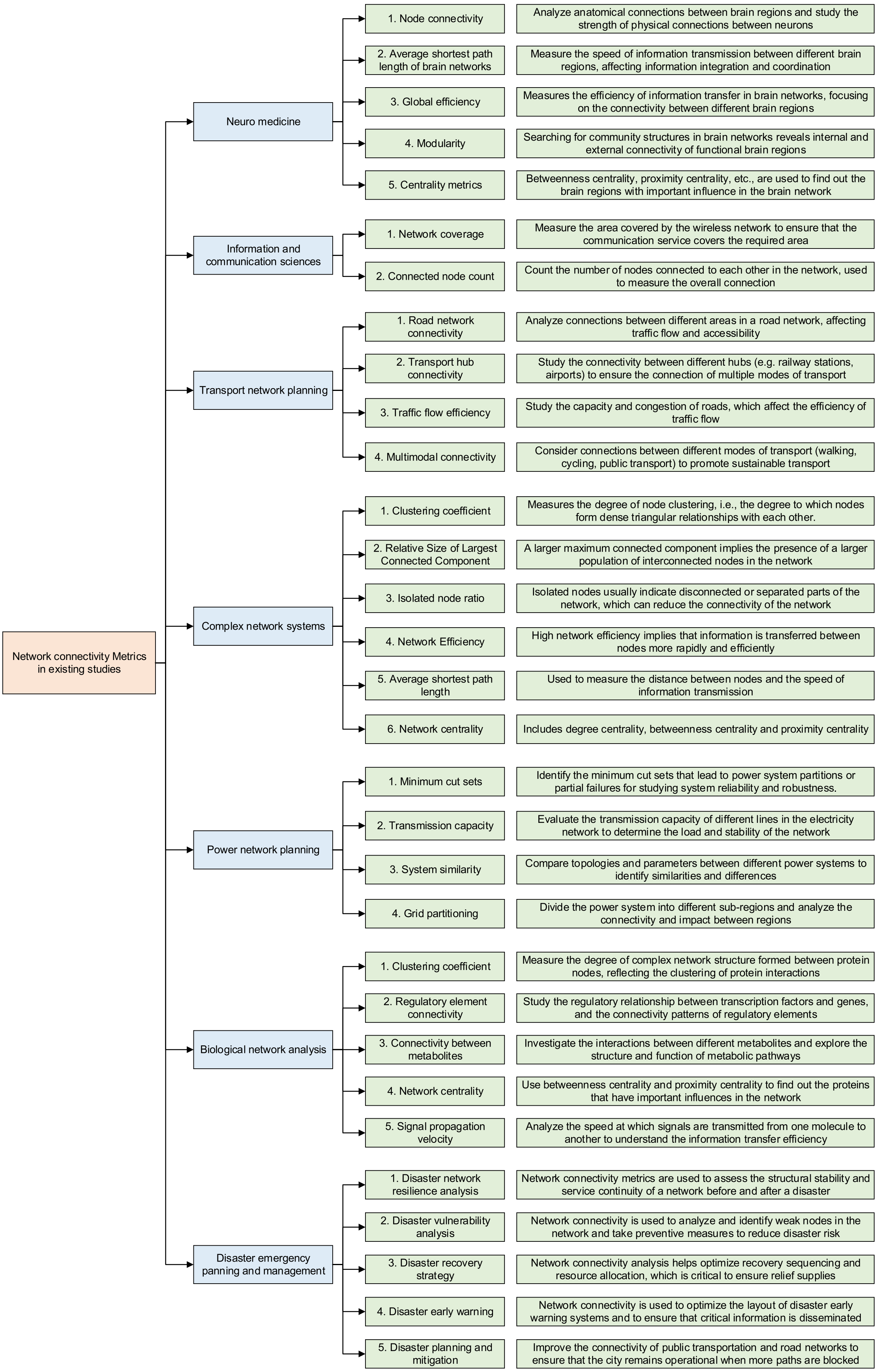


FIGURE 3. **Different network connectivity metrics in different disciplines. (This figure gives a metrics perspective on network connectivity across disciplines and research areas)**

tivity metrics, pointing out the potential value of studying network connectivity metrics in network disaster scenarios. Cimellaro et al. [18] also developed a framework for analytically quantifying disaster resilience that focuses on assessing and improving the ability of infrastructure networks to recover after a disaster, in particular providing insights into how network connectivity metrics affect network resilience. Vugrin et al. [19] provided a framework for resilience assessment of infrastructures and economic systems, analysed a case study of a supply chain under the impact of a hurricane, and discussed the assessment of network connectivity and recovery strategies. Simonovic and Peck [20] focused on assessing the dynamic resilience of climate change-induced natural disasters on coastal megacities across multiple dimensions including network connectivity, network modality, network clustering coefficients, and so on. Gao et al. [21] investigated the impact of a single network structure change on network connectivity, in particular the vulnerability and resilience analysis of network systems under the influence of disasters. Panteli and Mancarella [22] explored the impact of extreme weather and climate change on power system resilience and discussed possible mitigation strategies by maintaining a post-disaster network connectivity perspective. Bhatia et al. [23] discussed the example of the Indian railway network involving how to assess and improve the resilience of the system through network connectivity in the event of extreme weather events. Hazra et al. [24] pointed out how to improve communication network connectivity after a disaster through multi-information source infrastructure restoration, as well as maximizing network restoration using the resources of existing communication facilities. Kameshwar et al. [25] analysed the impact of disaster intensity, duration, and bridge damage on network connectivity, suggesting that multi-disaster and multi-infrastructure analyses are necessary to understand disaster network connectivity. Chang et al. [26] innovatively utilized an interdisciplinary approach to propose a data-driven measure of network connectivity using percolation theory, and gave simulation results of network connectivity changes under earthquakes. Li and Yan [27] developed a framework for improving network resilience with polycentricity, hierarchical networks, and modular collaboration, which was based on the integration of population, residential areas, critical facilities, and road network connectivity.

Network connectivity in disaster scenarios has been the focus of an interdisciplinary research field for more than 20 years. Starting from the scale-free network model proposed by Barabási et al. [14], researchers have gradually explored in depth the structural properties, resilience, and vulnerability of networks under the impact of natural or man-made disasters, as well as how network connectivity can be maintained or restored through various strategies. These studies have not only covered critical infrastructure networks such as power systems, transportation networks and supply chains but also focused on the long-term impact of climate change on urban network connectivity, reflecting the centrality of network connectivity in disaster management and emergency response. Building on the existing literature, our study proposes a new perspective and methodology focusing on predicting the values of the metrics that are approximately correlated with network connectivity in a disaster situation, i.e., RSLCC. Unlike previous studies that mainly started from qualitative descriptions or case studies, this study utilizes a data-driven approach using machine learning algorithms that have been developed to provide a new quantitative tool for disaster response and recovery decision-making and a new research focus.

To support our research work in this study, it is crucial to know about network connectivity metrics, and in fact, the most important aspect of network connectivity research is the network connectivity quantification. We will sort out the research content and measurement approaches of network connectivity in existing studies, as shown in **Table 1**. Most of the existing studies on network connectivity quantification tend to measure from the perspective of network structure analysis metrics, and some quantify connectivity through self-defined coefficients. The metrics used in existing studies mainly include network centrality metrics, network efficiency, network clustering coefficient, isolated node ratio, the relative size of largest connected component, etc. We also noticed that in real physical networks, such as transport networks, ecological landscape networks, and power networks, centrality metrics, especially betweenness centrality and proximity centrality are commonly used to quantify the network connectivity. In abstract networks, such as biological networks, neural networks, and information networks, network clustering coefficient, isolated node ratio, and network efficiency are commonly used to quantify network connectivity.

Table 1 summarizes several key indicators of network connectivity metrics, focusing on network centrality, clustering coefficients, network efficiency, and other metrics, especially since almost all the network connectivity metrics proposed by existing studies are combinations or improvements of these indicators, which also supports the selection of metrics for this study. We summarize the following three metrics that reflect the structural characteristics and functional performance of the network from different perspectives. (1) Network Centrality (Degree Centrality, Closeness Centrality, Betweenness Centrality): reveals the importance of nodes in the network in network connectivity and information transfer. Degree centrality focuses on the number of direct connections of a node, proximity centrality considers the length of paths to reach other nodes in the network, and betweenness centrality emphasizes the frequency of a node in acting as a path mediator. (2) Clustering Coefficient: measures how densely the neighbours of a node are connected, reflecting the presence of cliques or tightly connected subgroups in the network. (3) Network Efficiency: evaluates the efficiency of information transfer in a network, in particular the network's ability to respond and adapt quickly in the face of node or link failure. By fully applying these metrics, they can provide a scientific

TABLE 1. Research literature on quantification of network connectivity metrics.

Author/s	Publication time	Research content	Network properties	Network connectivity metric	Quantification method
Ricotta et al. [28]	2000	A landscape connectivity measure was investigated, which was based on graph theory.	Ecological network	Harary index	$\overline{H} = (H - H_{chain}) / (H_{planar} - H_{chain})$ where $\overline{H}$ is a standardized Harary index, H is an actual Harary index, H_{chain} is a chained Harary index, and H_{planar} is a planar Harary index.
Barzel and Biham [29]	2009	Based on the network topology analysis method, the quantification of network connectivity was defined, and the relationship between the strengths and weaknesses of connectivity was analysed.	Complex network	The ratio between the characteristic path length (The average of the shortest path lengths between all pairs of nodes in the network) and the average path length of the network	$C = L / \langle L \rangle$ where L is the average of the shortest path lengths between all pairs of nodes in the network, and $\langle L \rangle$ is the average path length of the network.
Rubinov and Sporns [30]	2010	Based on the complex network theory to study the brain network connection problem, the centrality metrics were used to quantify the brain connectivity.	Neural network	Centrality metrics, including closeness centrality and betweenness centrality	1. $(L_i^w)^{-1} = (n-1) / \sum_{j \in N, j \neq i} d_{ij}^w$ (Closeness centrality) 2. $b_i = 1/(n-1)(n-2)$ (Betweenness centrality) where L_i^w is the closeness centrality metric, n is the number of network nodes, d_{ij}^w is the Euclidean distance between any pair of nodes ij, b_i is the betweenness centrality metric.
Sullivan et al. [31]	2010	Explored network robustness from different link interruption situations and analysed the impact of network interruption on robustness.	Transport network	Network Robustness Index (The change in total travel time for a given time interval due to the redistribution of traffic in the system when a particular link is removed from the network)	$NRI_a = c_a - c$ $c_a = \sum_{i \in N} t_i^a x_i^a$ $c = \sum_{i \in N} t_i x_i$ where NRI_a is the value of Network Robustness Index, c_a is the total network travel time before removing links, c is the total network travel time after removing links, t_i^a is the original travel time for each link, t_i is the travel time for each link after removing links, x_i^a is the total number of original network links, x_i is the total number of original network links after removing links.
Zalesky et al. [32]	2012	The brain network was abstracted into a complex network system, and a quantification method to measure network connectivity by correlation was proposed and proved.	Neural network	Connectivity is quantified by correlation, which actually refers to the ratio of relative clustering coefficients of any two nodes	$\rho_{i,j} = C_i / C_j = \frac{2M_i}{k_i(k_i - 1)} \Big/ \frac{2M_j}{k_j(k_j - 1)}$ where $\rho_{i,j}$ is the ratio of the clustering coefficients of any two nodes, C_i is the clustering coefficient of node i, C_j is the clustering coefficient of node j, k_i, k_j are the node degrees of node i and node j, respectively, M_i, M_j are the number of edges that actually exist between the neighboring nodes of node i and node j, respectively.
Mishra et al. [12]	2012	Developed a unique method for measuring traffic connectivity, especially for traffic assignment models.	Transport network	A network connectivity metric measuring method based on network structure analysis	$\theta_w = \frac{1}{\lvert s_w \rvert - 1}, \theta_l = \frac{1}{\lvert s_l \rvert - 1}, \theta_r = \frac{1}{\lvert s_r \rvert - 1}$ where s_w, s_l, s_r are the values of node degree centrality, θ corresponding to the network centrality metric respectively.

TABLE 1. *(Continued.)* **Research literature on quantification of network connectivity metrics.**

Barzel and Barabási [33]	2013	Studied the physical connection and functional connection between cell biological networks and proposed a global physical connection prediction method.	Biological network	Isolated node ratio	$\Delta N = \left(1 - \frac{N^*}{N}\right) \times 100\%$ where N is the total number of nodes in the original network, N^* is the total number of nodes remaining after removing nodes.
Bolaños et al. [34]	2013	A small-world metric connectivity was proposed for analysing functional connectivity patterns in complex brain networks	Neural network	Weighted clustering coefficient	$C^w(i) = 3\sum_{j,h\neq i} w_{ij} w_{hi} w_{jh} \Big/ \left(\sum_{j,h\neq i} w_{hi} w_{jh} + \sum_{j,h\neq i} w_{ij} w_{hi} w_{jh}\right)$ where $C^w(i)$ is the value of weighted clustering coefficient, and w_{ij}, w_{hi}, w_{jh} are the weight coefficients of each node pair.
Milli et al. [35]	2015	A method for quantifying the degree of connectivity between individuals in a social network was proposed	Social network	Isolated node ratio	$\Delta N = \left(1 - \frac{N^*}{N}\right) \times 100\%$ where N is the total number of nodes in the original network, N^* is the total number of nodes remaining after removing nodes.
Cohen and D'Esposito [36]	2016	The functional connectivity was studied from the network isolation metric and the network clustering metric.	Neural network	Network efficiency	$E = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} h_{ij}, i = 1,2,...,N, i \neq j$ where N is the total number of nodes in the original network, E is the value of network efficiency, h_{ij} is the average of the inverse shortest path lengths between pairs of nodes.
Qian et al. [37]	2017	Focused on the global connectivity of the network, a link entropy was proposed to quantify the connectivity	Physical network	Link entropy	$LE_{ij} = \frac{\left[H(X_i) + H(X_j)\right]\big/2 + JSD(X_i \mid X_j)}{2}$ where $H(X_i)$, $H(X_j)$ are the information entropy of the nodes, X_i, X_j are the probability distribution of the nodes, and LE_{ij} is the value of link entropy.
Zhang and Ng [38]	2021	Used the four metrics in the network structure analysis, and proposed a fuzzy weighting method to integrate those four to form connectivity quantification	Transport network	Closeness centrality Betweenness centrality Relative size of largest connected component Network efficiency	1. $\left(L_i^w\right)^{-1} = (n-1) \Big/ \sum_{j\in N, j\neq i} d_{ij}^w$ 2. $b_i = 1\big/(n-1)(n-2)$ 3. $G = \frac{N^{*}}{N}$ 4. $E = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} h_{ij}, i = 1,2,...,N, i \neq j$ where L_i^w is the closeness centrality metric, n is the number of network nodes, d_{ij}^w is the Euclidean distance between any pair of nodes ij, b_i is the betweenness centrality metric. N is the total number of nodes in the original network, N^* is the total number of nodes remaining after removing nodes. h_{ij} is the average of the inverse shortest path lengths between pairs of nodes.
Cumming et al. [39]	2022	A cross-scale centrality metric was proposed to measure the ecological network connectivity	Ecological network	Betweenness centrality	$b_i = 1\big/(n-1)(n-2)$ where b_i is the betweenness centrality metric.

TABLE 1. ***(Continued.)*** **Research literature on quantification of network connectivity metrics.**

		problem, which was an improvement on the betweenness centrality metric.			
Sharma and Ram [40]	2023	Quantified the road network connectivity in airport areas based on centrality metrics	Transport network	Centrality metrics, including closeness centrality and betweenness centrality	1. $\left(L_i^w\right)^{-1} = (n-1)\Big/\sum_{j\in N, j\neq i} d_{ij}^w$ (Closeness centrality) 2. $b_i = 1\big/(n-1)(n-2)$ (Betweenness centrality) where L_i^w is the closeness centrality metric, n is the number of network nodes, d_{ij}^w is the Euclidean distance between any pair of nodes ij, b_i is the betweenness centrality metric.

basis and decision support for network design, optimization, and disaster recovery planning.

A. SUMMARY OF LITERATURE REVIEW

The existing metrics design and quantification approaches on network connectivity can be drawn on for this study to explore network connectivity quantification. However, unlike the existing studies, this study does not analyse a specific discipline area, nor does it use real network data, but only tries to find an approximate network connectivity index from the perspective of structural analysis and achieve a high accuracy of prediction.

The main contributions of this paper are summarized as follows:

(1) A data-driven statistical approach is used to find an approximate network connectivity metric, which is mainly obtained through principal component analysis and explains variance contribution analysis.

(2) Seven prediction methods are designed to predict the value of the approximate metric, and a comparative analysis of prediction accuracy is also conducted, The Deep Neural Network (DNN) prediction method is proved to have higher prediction accuracy.

(3) Accurate prediction of 1 independent variable for 1 dependent variable is achieved, and the independent variable data is easy to obtain, i.e., network disruption degree, which is determined by the ratio of the number of edges after an attack to the number of edges before the attack, and the approximate network connectivity metric can be predicted using only this independent variable.

III. EXPERIMENT DESIGN AND PREPARATION

The approach studied in this paper is to first select some metrics related to network connectivity based on the network structure analysis, which can be calculated by specific formulas. Then, 100,000 datasets are randomly generated using NetworkX in Python, and the data are cleaned, including removing duplicate values and outliers, as well as multiple covariance checking. Next, 80% of these datasets are set as training data, and 20% of them are set as test data to find an approximate network connectivity metric based on the metrics' weight contribution in principal component analysis. Finally, seven prediction methods are designed to predict the metric and analyse the prediction accuracy.

A. WHAT POTENTIAL METRICS ARE CHOSEN TO PREDICT NETWORK CONNECTIVITY (NC) AND WHY?

1) CLUSTERING COEFFICIENT (CC)

CC measures the degree of node clustering in the network, i.e., the degree to which nodes form dense triangular relationships with each other. A high clustering coefficient indicates that nodes are more tightly connected to each other, forming more local communities or clusters. Thus, higher clustering coefficients are usually associated with better network connectivity.

2) RELATIVE SIZE OF LARGEST CONNECTED COMPONENT (RSLCC)

RSLCC indicates the proportion of the largest connected subgraph in the network to the overall network. A larger maximum connected component implies the presence of a larger population of interconnected nodes in the network, which is important for the overall connectivity of the network. Therefore, a higher proportion of maximum connected components is associated with better network connectivity.

3) ISOLATED NODE RATIO (INR)

INR indicates the percentage of nodes in the network that are isolated and have no connected edges. Isolated nodes usually indicate disconnected or separated parts of the network, which can reduce the connectivity of the network. Therefore, a lower percentage of isolated nodes is associated with better network connectivity.

4) NETWORK EFFICIENCY (NE)

NE measures the efficiency of information transmission between nodes in a network. High network efficiency implies that information is transferred between nodes more rapidly and efficiently. Therefore, higher network efficiency is associated with better network connectivity.

5) NETWORK DISRUPTION DEGREE (NDD)

NDD is a measure of the degree of disruption of a network for a given removal ratio. A lower NDD value means that the connectivity of the network is less affected for a given percentage of edge removal, i.e., the network is still able to maintain relatively good connectivity. This implies that the network has a high degree of resilience and toughness and that the network can remain connected and functional even in the face of a certain degree of edge removal or failure.

The reason for choosing the above five potential metrics in this paper is that the existing studies on network connectivity quantification have already used these five metrics, but it is worth stating that we do not choose network centrality as a potential metric in this study, for three reasons:(1) This study focuses on the connectivity of the network as a whole, rather than node importance or influence. (2) Network centrality metrics require more information about node attributes and relationships, which are less accessible or not applicable in this study. (3) This study needs to avoid the interference of node importance on the findings and enhance the generalizability of the research insights.

B. EXPRESSIONS FOR 5 POTENTIAL METRICS

(1) CC- -^{C}

Cluster coefficient of a node i

$$C_i = \frac{2M_i}{k_i\,(k_i - 1)}, i = 1, 2, \ldots, N \tag{1}$$

where M_i is the number of edges that exist between the neighbouring nodes of the node i. k_i is the node degree of the node i.

The clustering coefficient of the network

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{2}$$

where C_i is the value of the clustering coefficient of the node i, N is the total number of nodes in the original network.

(2) RSLCC- -^{G}

$$G = \frac{N^{''}}{N} \tag{3}$$

where $N^{''}$ is the total number of nodes in the maximum connected subgraph after an attack on the network.

(3) INR- -$^{\Delta N}$

$$\Delta N = \left(1 - \frac{N^*}{N}\right) \times 100\% \tag{4}$$

where N^* is the total number of nodes in the network after an attack.

(4) NE- -^{E}

$$E = \frac{1}{N\,(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} h^{ij}, i = 1, 2, \ldots, N, i \neq j \tag{5}$$

where h^{ij} is the average of the inverse shortest path lengths between pairs of nodes,$h^{ij} = 1/d^{ij}$. d^{ij} is the Euclidean distance between any pair of nodes ij.

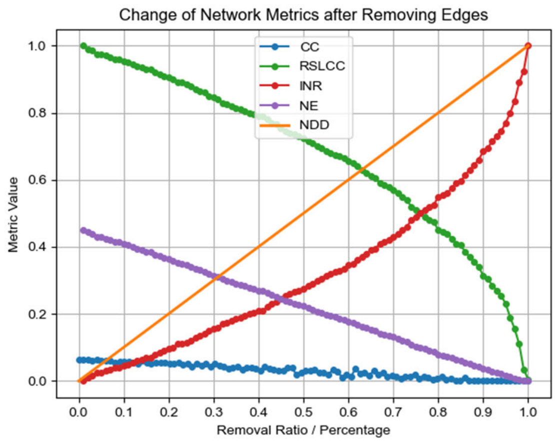


FIGURE 4. Change process diagram of the 5 metrics. (This figure shows the trend in the values of the metrics when the network edge is removed randomly)

(5) NDD- -^{D}

Network Disruption Degree in this study simply means Edge Removal Ratio p.

$$D = p = \left(1 - \frac{S^*}{S}\right) \times 100\% \tag{6}$$

where S is the total number of edges in the original network, S^* is the total number of edges in the network after the attack.

C. OBSERVING CHANGES IN 5 METRICS AFTER RANDOMLY REMOVING EDGES

A network structure graph with 200 nodes and 1200 edges is randomly generated using NetworkX, and the changes in the five metrics are observed as shown in **Figure 4** after gradually removing edges according to a certain percentage.

Figure 4 displays how the different network metrics change with the removal ratio or removal ratio after removing the edges from the network. The following are explanations and observations for each metric.

(1) CC: indicates the degree of node clustering in the network. The blue line represents the relationship between the removal ratio and the clustering coefficient. As the removal ratio increases, the clustering coefficient decreases, indicating that the nodes in the network are less clustered.

(2) RSLCC: indicates the relative size of the largest connected component. The green line represents the relationship between the removal ratio and the RSLCC. As the removal ratio increases, the RSLCC decreases, and the connectivity of the network becomes weaker.

(3) INR: indicates the ratio of isolated nodes. The red line represents the relationship between the removal ratio and the isolated node ratio. As the removal ratio increases, the isolated node ratio gradually increases, indicating that there are more and more isolated nodes in the network.

(4) NE: indicates the efficiency of the network. The purple line represents the relationship between the removal ratio

TABLE 2. Table showing a small portion of the data set.

No.	Nodes	Edges	Edges (after being attacked)	CC	RSLCC	INR	NE	NDD
1	10	20	20	0.2500	1.0000	0.0000	0.7190	0.0000
2	10	20	18	0.2400	1.0000	0.0000	0.6815	0.1000
3	10	20	16	0.4070	0.9000	0.1000	0.5740	0.2000
4	10	20	14	0.3270	0.9000	0.1000	0.5310	0.3000
5	10	20	12	0.3730	0.7000	0.3000	0.3670	0.4000
6	20	60	60	0.3853	1.0000	0.0000	0.6333	0.0000
7	20	60	54	0.3513	0.9500	0.0500	0.5750	0.1000
8	20	60	51	0.4180	0.9000	0.1000	0.5220	0.1500
9	20	60	42	0.2170	0.9000	0.1000	0.4890	0.3000
10	20	60	30	0.1930	0.7500	0.2500	0.3350	0.5000
11	40	200	200	0.2216	1.0000	0.0000	0.6227	0.0000
12	40	200	150	0.2290	0.8500	0.1500	0.4500	0.2500
13	40	200	186	0.2030	0.9750	0.0250	0.5850	0.0700
14	40	200	172	0.2220	0.9500	0.0500	0.5450	0.1400
15	40	200	147	0.2120	0.8500	0.1500	0.4450	0.2650
16	40	200	122	0.1840	0.8000	0.2000	0.3850	0.3900
17	40	200	98	0.1700	0.7000	0.3000	0.2960	0.5100
18	80	300	300	0.0958	1.0000	0.0000	0.4715	0.0000
19	80	300	283	0.0760	0.9870	0.0125	0.4560	0.0567
20	80	300	264	0.0781	0.9500	0.0500	0.4180	0.1200
21	80	300	221	0.0963	0.8625	0.1375	0.3430	0.2633
22	80	300	167	0.0653	0.7625	0.2375	0.2575	0.4433
23	80	300	124	0.0409	0.6375	0.3625	0.1794	0.5867
24	80	300	102	0.0585	0.6125	0.3875	0.1534	0.6600
25	100	500	500	0.0935	1.0000	0.0000	0.4965	0.0000
26	100	500	480	0.0923	0.9800	0.0200	0.4740	0.0400
27	100	500	422	0.1057	0.9300	0.0700	0.4198	0.1560
28	100	500	378	0.0873	0.8900	0.1100	0.3803	0.2440
29	100	500	313	0.0771	0.7800	0.2200	0.2952	0.3740
30	100	500	296	0.0860	0.7700	0.2300	0.2812	0.4080
31	100	500	271	0.0773	0.7600	0.2400	0.2679	0.4580
32	100	500	253	0.0691	0.7200	0.2800	0.2431	0.4940
33	100	500	205	0.0514	0.6700	0.3300	0.2030	0.5900
34	100	500	187	0.0579	0.6300	0.3700	0.1799	0.6260
35	100	500	144	0.0484	0.5400	0.4600	0.1321	0.7120
36	180	700	700	0.0476	1.0000	0.0000	0.3993	0.0000
37	180	700	694	0.0486	1.0000	0.0000	0.3976	0.0086
38	180	700	612	0.0427	0.9333	0.0667	0.3449	0.1257
39	180	700	550	0.0298	0.8944	0.1056	0.3105	0.2143
40	180	700	523	0.0262	0.8611	0.1389	0.2880	0.2529
41	180	700	501	0.0319	0.8556	0.1444	0.2808	0.2843

TABLE 2. *(Continued.)* Table showing a small portion of the data set.

42	180	700	472	0.0261	0.8167	0.1833	0.2564	0.3257
43	180	700	425	0.0313	0.7833	0.2167	0.2318	0.3929
44	180	700	416	0.0432	0.7556	0.2444	0.2185	0.4057
45	180	700	400	0.0495	0.7444	0.2556	0.2110	0.4286
46	180	700	346	0.0335	0.7000	0.3000	0.1829	0.5057
47	180	700	312	0.0220	0.6667	0.3333	0.1626	0.5543
48	180	700	222	0.0233	0.5500	0.4500	0.1077	0.6829
49	180	700	214	0.0220	0.5333	0.4667	0.1020	0.6943
50	180	700	167	0.0070	0.4611	0.5389	0.0753	0.7614

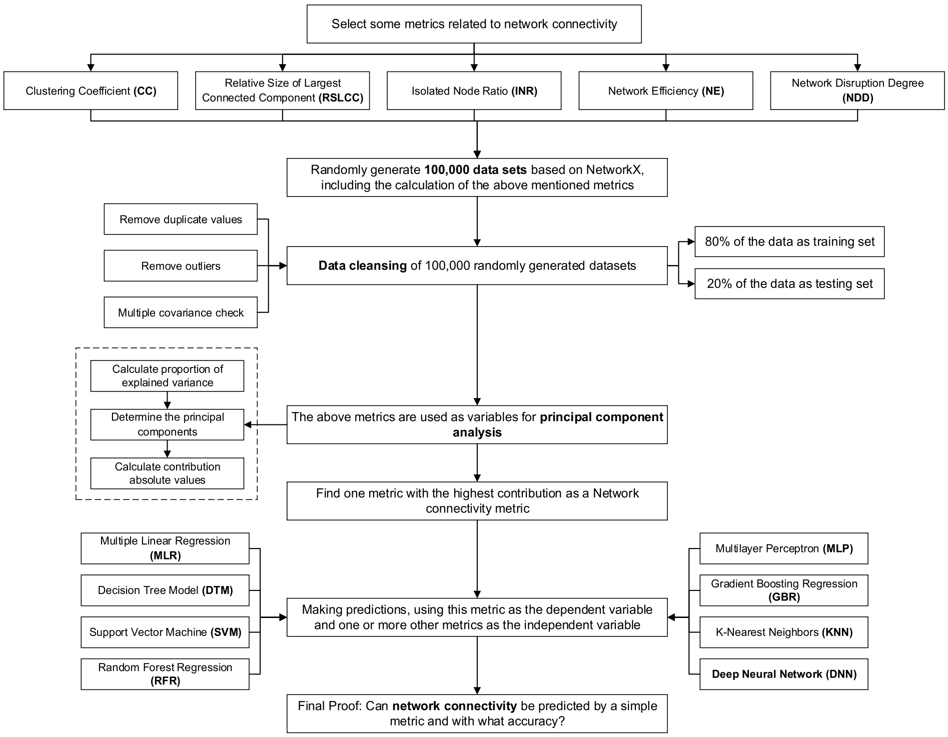


FIGURE 5. Overall framework diagram for experimental implementation. (This figure shows our technical framework for carrying out predictions of network connectivity)

and the network efficiency. As the removal ratio increases, the network efficiency gradually decreases, indicating that the efficiency of information transmission in the network decreases.

(5) NDD: indicates the degree of network disruption. The orange line represents the relationship between the removal ratio and the network disruption degree. The relationship between the removal ratio and the network disruption degree is linear, and as the removal ratio increases, the network disruption degree gradually increases.

D. DATA PREPARATION

NetworkX for Python is utilized to generate 100,000 data sets, each containing the number of nodes, number of edges, number of post-disaster edges (after being attacked), CC values, RSLCC values, INR values, NE values, and NDD values. Data can be obtained from:

https://github.com/JUNXIANGXU666/connectivity. Some of the data in the dataset are presented as shown in **Table 2**.

E. EXPERIMENTAL FRAMEWORK

Based on the above description of the experimental design, the subsection provides a general framework for the experiment implementation, as shown in **Figure 5**.

IV. MINING APPROXIMATE METRICS RELATED TO NETWORK CONNECTIVITY

A. DATA CLEANING

1) MISSING AND DUPLICATE VALUE CHECK

The missing value check was first carried out and the results are shown in **Table 3**.

TABLE 3. Missing value check results.

	Missing value check
Nodes	0
Edges	0
Edges (after being attacked)	0
CC	0
RSLCC	0
INR	0
NE	0
NDD	0

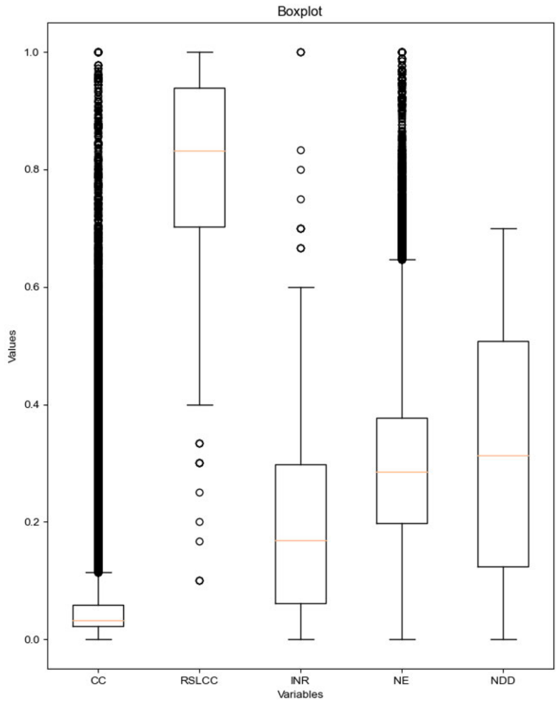


FIGURE 6. Data outlier test box line diagram.

Then the duplicate data check is performed and the result shows 'No duplicate data found'.

2) OUTLIER CHECK

Outlier checking is performed next and the results are obtained as shown in the box-and-line diagram displayed in **Figure 6**. The result of the re-check after the outliers are removed is displayed as shown in **Figure 7**.

(Figures 6 and 7 show the results of processing the raw data, visualized using box-and-line diagrams)

After removing the outliers, the training set data became 92,192 data sets instead of 100,000 data sets.

3) CHECKING FOR MULTICOLLINEARITY

a: CORRELATION ANALYSIS

The correlation analysis check of the metrics (variables) is shown in **Figure 8**.

b: VARIANCE INFLATION FACTOR (VIF)

The results of the variance inflation factor (VIF) checks are organized as shown in **Table 4**.

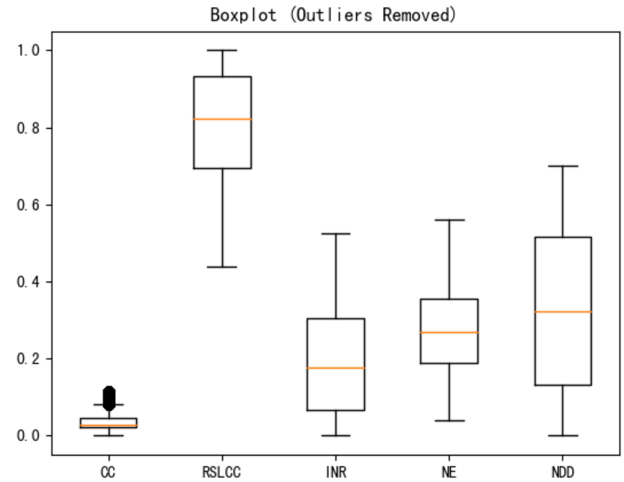


FIGURE 7. Box line chart after removing outliers.

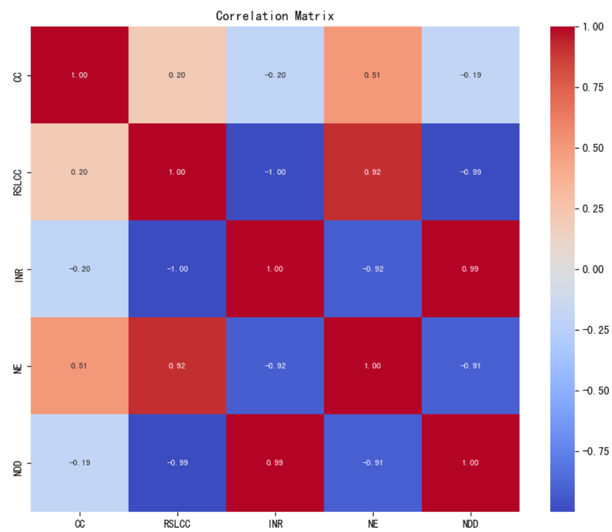


FIGURE 8. Correlation check for multicollinearity. (By observing these correlation coefficients, we can infer that there may be a multicollinearity problem between CC, INR, RSLCC, NE, and NDD because of the strong correlation between them, which may lead to the existence of redundant information in the model.)

TABLE 4. VIF check results.

	Metrics	VIF
1	CC	12.588685
2	RSLCC	165.128563
3	INR	291.506854
4	NE	171.490996
5	NDD	333.029683

Taken (1) and (2) together, the four variables, RSLCC, INR, NE, and NDD, have strong multicollinearity problems with each other, while the co-collinearity between CC and other variables is relatively weak.

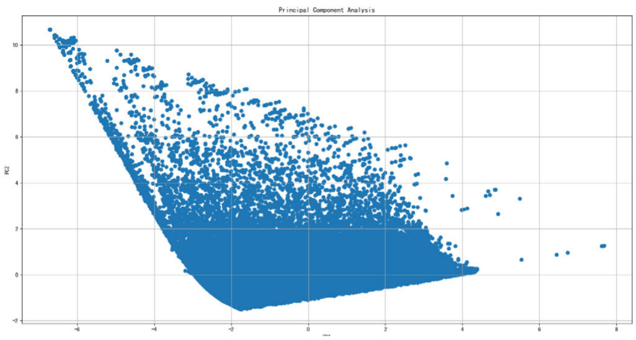


FIGURE 9. Principal component scatter plot.

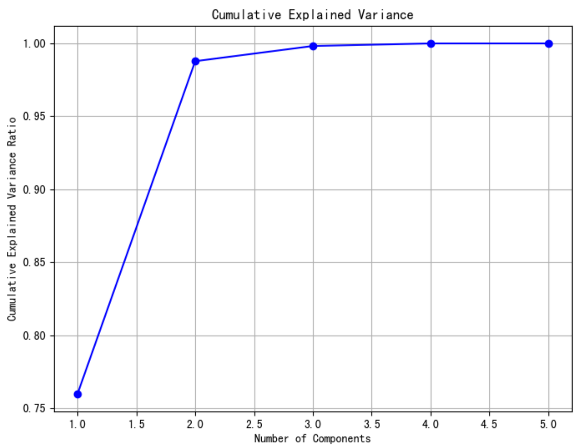


FIGURE 10. Cumulative explained variance ratio chart.

TABLE 5. Results of principal component analysis of 5 metrics.

No.	Name	Cumulative explained variance ratio
1	Principal Component 1	75.98%
2	Principal Component 2	22.80%
3	Principal Component 3	1.04%
4	Principal Component 4	0.18%
5	Principal Component 5	0.00%

B. PRINCIPAL COMPONENT ANALYSIS

The principal component scatter plot and the cumulative explained variance ratio chart are in **Figure 9** and **Figure 10**.

The specific results of the principal component analysis are shown in **Table 5**.

According to **Table 5**, it can be cleanly seen that principal component 1 explains more than 75% of the raw data, so only principal component 1 can be used to represent the overall data after dimensionality reduction. Then, the results of the contribution of metrics (variables) to the principal component 1 in **Table 6** are given.

Finally, the absolute values of the contribution of the variables to the principal components were transformed into weights and the results after normalization are shown in **Table 7**.

TABLE 6. Summary table of the results of principal component 1 analysis.

Name	Proportion of explained variance	CC	RSLCC	INR	NE	NDD
Principal Component 1	75.98%	0.0715	0.8898	-0.0146	-0.3293	0.0154

TABLE 7. Results of contribution weights of 5 metrics.

No.	Metrics	Principal Component 1 (contribution absolute values)	Weight values	Normalized weight values
1	CC	0.0715	0.0698	0.0626
2	RSLCC	0.8898	0.8692	0.7799
3	INR	0.0146	0.0143	0.0128
4	NE	0.3293	0.3217	0.2887
5	NDD	0.0154	0.1611	0.1446

According to the results shown in **Table 7**, the weight contribution of the five metrics to the principal component 1 is RSLCC, NE, NDD, CC and INR in descending order. In other words, network connectivity can be characterized by these five metrics, and there are three research approaches: (1) According to the results of the weight contribution degree, weights can be assigned to the five metrics and then combined to characterize network connectivity; (2) Choose both RSLCC and NE to characterize network connectivity, because these two metrics contribute close to 98% of the interpretability of network connectivity; (3) Choose only RSLCC to characterize network connectivity, because it has the largest weight contribution, treat it as a dependent variable, and the rest of one or more metrics as independent variables to be researched by using prediction methods.

The purpose of this paper is to provide an approach for finding a metric of approximate network connectivity, and for simplicity, research approach (3) has been chosen, where we use RSLCC as the approximate network connectivity metric, and then focus on the prediction methods for the metric and the analysis of the prediction results.

V. MULTIPLE PREDICTION METHODS FOR PREDICTING RSLCC METRIC VALUES

A. APPLICABILITY ANALYSIS OF PREDICTING METHODS

This section first analyses the applicability of the eight prediction methods in the case of multicollinearity, and the results of the analysis are shown in **Table 8**.

TABLE 8. Applicability analysis of 8 prediction methods in the face of multicollinearity.

No.	Prediction methods	Introduction	Applicability
1	Multiple Linear Regression (MLR)	MLR relies on the independence between predictor variables. When multicollinearity exists in the predictor variables, it negatively affects the estimation of the coefficients, causing the estimated coefficients to become unstable and increasing the uncertainty in the interpretation of the model.	×
2	Decision Tree Regression (DTR)	DTR is a nonparametric forecasting method that is not subject to multicollinearity. It can cope with highly correlated independent variables and generate nonlinear predictive models.	√
3	Support Vector Machine (SVM)	SVM is a nonlinear regression method based on kernel functions, which can provide better prediction performance when dealing with multiple covariance problems.	√
4	Random Forest Regression (RFR)	RFR is an integrated learning method that performs prediction by constructing multiple decision tree models. It can handle multiple covariance problems and has better prediction accuracy.	√
5	Gradient Boosting Regression (GBR)	GBR is an integrated learning method that performs regression analysis by constructing multiple weak prediction models step by step. It has some robustness in dealing with multicollinearity problems.	√
6	Multilayer Perceptron Regression (MLP)	MLP is an artificial neural network model that can handle multicollinearity problems through its deep structure and nonlinear activation functions. This allows the MLP to maintain predictive performance and stability in the face of multicollinearity.	√
7	K-Nearest Neighbours (KNN)	KNN can make predictions based on the characteristics of nearest neighbour samples, when dealing with multicollinearity problems, KNN can provide a nonparametric prediction solution.	√
8	Deep Neural Network (DNN)	DNN can perform predictive analysis by adjusting the neurons and activation functions in the hidden layer. It is particularly suitable for tasks that require complex correlations between multiple variables.	√

TABLE 9. Comparison of prediction accuracy of 7 prediction methods.

No.	Prediction method	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R^2 score
1	DTR	0.0004374688000382171	0.031724708953241316	0.006410052500763294	0.9226882861950656
2	SVM	0.005056047730584825	0.08534366251842505	0.06337161995206519	0.8606425852910482
3	RFR	0.0010948004670154251 5	0.01036553138324935	0.008805261809332523	0.9449577367140892
4	GBR	0.005552302816808898 6	0.010451377602033666	0.010297727818835617	0.9510468317797527
5	MLP	0.0004992727595136276	0.022344412266014687	0.01707512287376529	0.9684823061514242
6	KNN	0.00719848797786798	0.009484390359871462	0.009220180893764646	0.9561712560298919
7	DNN	0.0000652523110318101 3	0.008077890258712984	0.004801136900771813	0.9965293490359811

According to **Table 8**, seven prediction methods DTR, SVM, RFR, GBR, MLP, KNN and DNN will be selected for RSLCC metric prediction in this paper.

B. INDEPENDENT AND DEPENDENT VARIABLE SELECTION

Undoubtedly, this paper has already identified RSLCC as the predicted metric, which is the dependent variable, while the choice of the independent variable is considered for the following reasons:

(1) The remaining metrics are not considered all as independent variables because once this is done, the prediction would be meaningless since the RSLCC can be calculated directly when the values of all other independent variables can be determined.

(2) Prioritize one metric as the independent variable, and the rest of the metric values except RSLCC are used as training data only, and then add variables one by one if the prediction accuracy is not high.

(3) This prioritized metric is chosen to be as close as possible to the practical significance and is not recommended to be a metric that needs to be calculated by more than 1 variable.

(4) Since the sum of INR and RSLCC is always equal to 1, INR is not considered an independent variable.

Combining the above reasons for independent variables selection, we finally select NDD as the independent variable, and if the prediction accuracy is not high, another CC will be added as the independent variable, and so on.

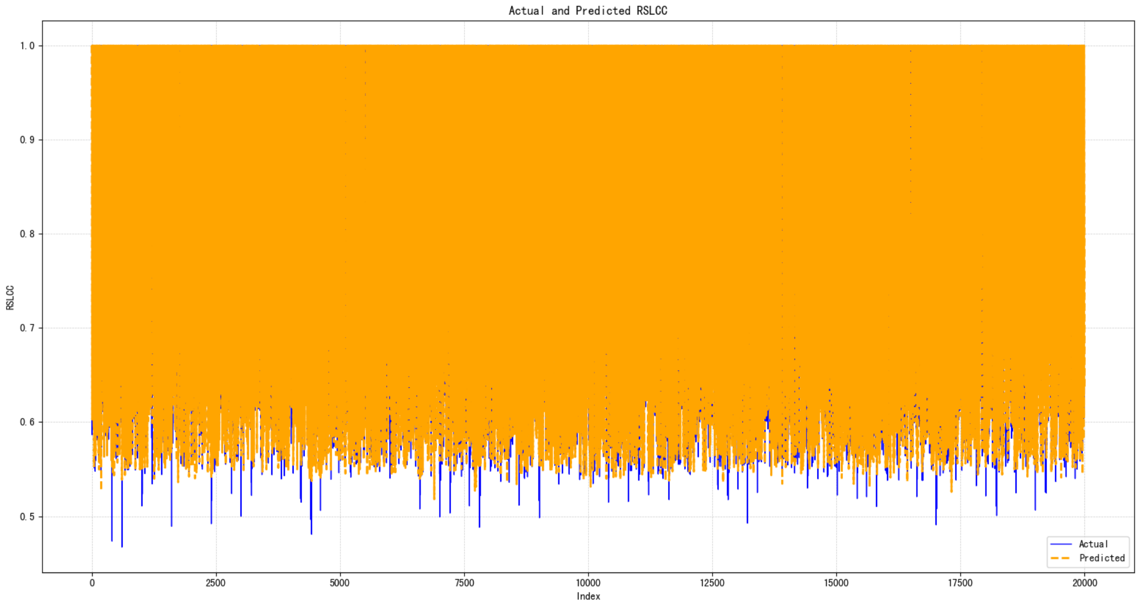


FIGURE 11. Predicting the dependent variable RSLCC from the independent variable NDD. (This figure shows a comparison of the difference between the actual and predicted RSLCC values, and because of the very large amount of data involved, the predicted differences can only be seen in terms of coverage)

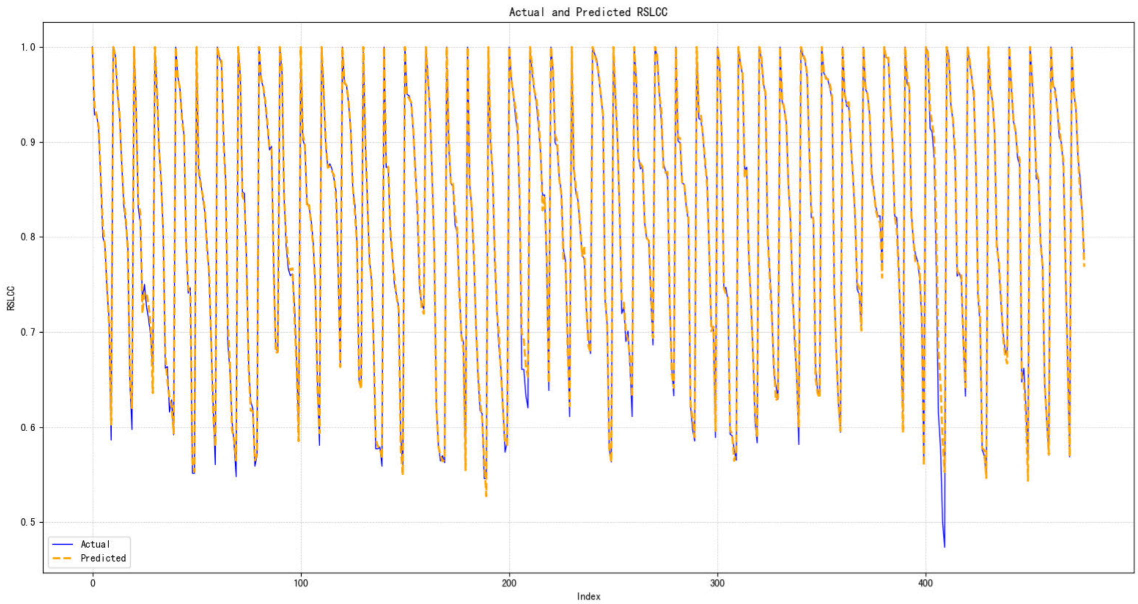


FIGURE 12. Predicting the dependent variable RSLCC from the independent variable NDD. (Reducing the size of the test set, to more clearly see the accuracy of the results of the comparison between predicted and actual values)

C. COMPARISON OF PREDICTION ACCURACY OF 7 PREDICTION METHODS

A comparison of the prediction results of the seven prediction methods is shown in **Table 9**.

Fortunately, according to **Table 9**, it can be seen that all 7 prediction methods can achieve 1 independent variable (NDD) to predict 1 dependent variable (RSLCC), and all of them have satisfactory prediction accuracy, especially DNN has as high as 99.65% accuracy in predicting RSLCC values (According to R^2 score in **Table 9**, R^2 score is a measure of the predictive effectiveness of a regression model, reflecting the fit degree between the predicted values of the model and the actual observations. The value of the R^2 score ranges from 0 to 1, and the closer the value is to 1, the better the predictive ability of the model, and the more variance it explains.). Intuitively, we use **Figure 11** to visualize the results of the DNN prediction method comparing the actual and predicted values.

We reduced the test data to 500 and then generated the predicted results again as shown in **Figure 12**.

However, although DNN has significant advantages over other prediction approaches in terms of prediction performance, the runtime of DNN prediction will take a very

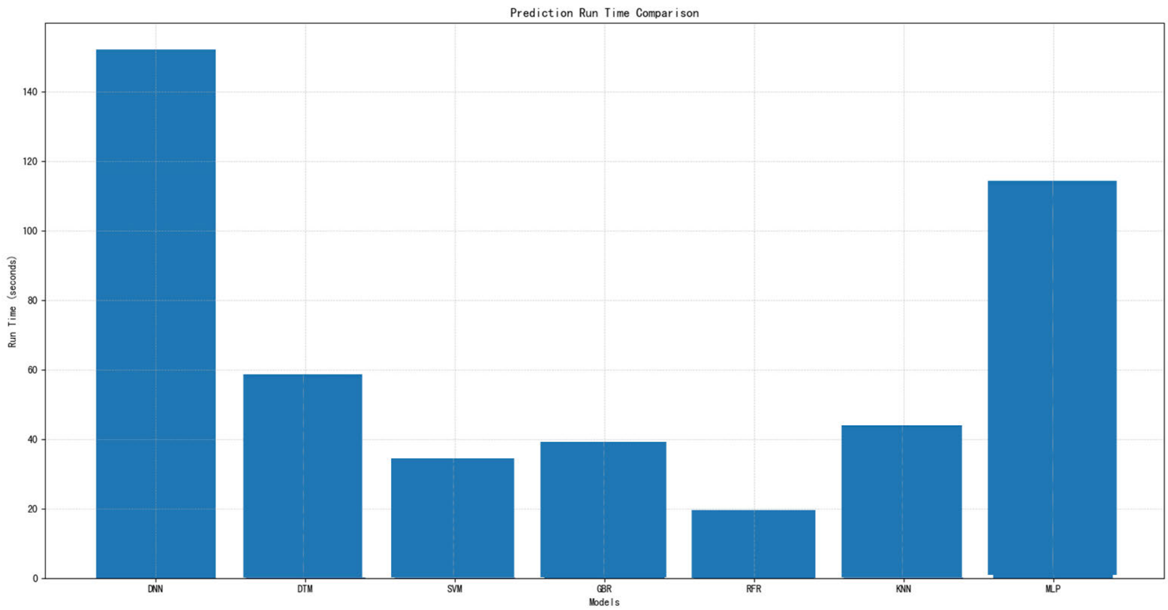


FIGURE 13. Comparison of the running time of 7 prediction methods. (This figure compares the performance of the seven algorithms from the perspective of prediction algorithm runtime)

TABLE 10. Test set with the same number of nodes, number of edges and NDD settings.

No.	Nodes	Edges	Edges (after being attacked)	CC	RSLCC	NE	NDD
1	200	1800	1305	0.0741	0.8650	0.3764	0.2750
2	200	1800	1305	0.0822	0.8500	0.3665	0.2750
3	200	1800	1305	0.0789	0.8600	0.3761	0.2750
4	200	1800	1305	0.0784	0.9000	0.3726	0.2750
5	200	1800	1305	0.0760	0.8500	0.3666	0.2750
6	200	1800	1305	0.0783	0.8550	0.3701	0.2750
7	200	1800	1305	0.0770	0.8550	0.3699	0.2750
8	200	1800	1305	0.0757	0.8600	0.3732	0.2750
9	200	1800	1305	0.0732	0.8550	0.3697	0.2750
10	200	1800	1305	0.0779	0.8450	0.3634	0.2750

TABLE 11. DNN prediction accuracy information table. (Set the same number of nodes, number of edges and NDD)

Prediction method	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R^2 score
DNN	0.00024542511977291746	0.01566604990969062	0.009003913879394532	-0.156302095514332

long time compared to other prediction approaches. Simply can be seen in **Figure 13**.

D. A POTENTIAL PROBLEM NEEDS TO BE DEMONSTRATED

According to the code for randomly generating data sets, the CC, NE, INR, and RSLCC are different in the network data generated by each run with the same NDD. This is because the network structure is not determined by these metrics mentioned in this study alone, but also takes into account metrics such as node degree, average node degree, average shortest path length, node strength, etc. Therefore, does this have an impact on the prediction results? We continue to use the DNN prediction model under large-scale data training, but the test data is based on the control variable method, and the number of nodes, the number of edges, and the NDD are set to the same for the study, and the specific test data set is shown in **Table 10**.

TABLE 12. Prediction results table in the small-scale datasets (RFR).

No.	Nodes	Edges	Edges (after being attacked)	CC	RSLCC(Actual)	RSLCC(Prediction)	INR	NE	NDD
1	58	328	267	0.1526	0.9138	0.9007	0.0690	0.4976	0.1860
2	106	500	400	0.0813	0.8962	0.9000	0.1132	0.3729	0.2000
3	120	600	462	0.0801	0.8833	0.8831	0.1167	0.3634	0.2300
4	658	2400	1789	0.0121	0.8587	0.8565	0.1337	0.2198	0.2546
5	884	4215	4016	0.0098	0.9751	0.9816	0.0249	0.3076	0.0472
6	512	2596	1543	0.0161	0.7656	0.7627	0.2285	0.2047	0.4056
7	244	1000	896	0.0286	0.9508	0.9468	0.0491	0.3426	0.1040
8	388	1967	1212	0.0241	0.7887	0.7873	0.2088	0.2261	0.3838
9	2000	8968	6954	0.0030	0.8830	0.8818	0.1160	0.2129	0.2246
10	1200	6000	4000	0.0055	0.8117	0.8161	0.1941	0.1973	0.3333
11	1642	5869	5000	0.0042	0.9184	0.9207	0.0816	0.2178	0.1481
12	2436	12000	8678	0.0038	0.8498	0.8493	0.1502	0.1976	0.2768
13	2209	8645	8200	0.0029	0.9737	0.9770	0.0262	0.2482	0.0515
14	2347	10987	7645	0.0036	0.8394	0.8334	0.1606	0.1881	0.3042
15	3200	10000	8212	0.0010	0.9053	0.9035	0.0946	0.1791	0.1788
16	3489	12000	6543	0.0017	0.7340	0.7305	0.2654	0.1130	0.4548
17	5200	40096	34768	0.0025	0.9315	0.9271	0.0684	0.2582	0.1329
18	5500	46000	40000	0.0028	0.9335	0.9272	0.0666	0.2647	0.1304

The results after prediction using DNN are shown in **Figure 14** and the prediction method accuracy is shown in **Table 11**.

The same number of nodes, number of edges, and NDD values do not have a great impact on the predicted RSLCC values, and an NDD value can only correspond to the prediction of an identical RSLCC value. This value is compared with the actual value generated by the control variables method, and although the R2 is poor, the MSE, RMSE, and MAE all prove to have high predictive power and accuracy.

It can be demonstrated that although the generated network attribute values will vary for the same number of nodes, edges, and NDD values, the DNN prediction model after large-scale data training still has good accuracy for predicting RSLCC values.

E. WHAT HAPPENS IF THE EXPERIMENT WITH SMALLER DATASETS?

This paper next needs to demonstrate that this training and test data is not coincidental, but that the predictive tools in this paper can be used in any network structure analysis problem related to these 5 metrics. Next, we discard the 100,000 datasets and re-generate another 363 sets of datasets using NetworkX (https://github.com/JUNXIANGXU666/connectivity), 18 of the datasets are used as the test set and the rest of the data are used as the training data. A comparison of the actual and predicted value results after the 7 prediction methods are run is shown in **Figure 15a** to **Figure 15g** below.

What can be seen is that DTR, RFR, GBR and KNN can still maintain good prediction accuracy in small-scale datasets with the research methods in this paper, indicating that they do not require a high data training scale and are more flexible to use. However, SVM, MLP and DNN perform poorly in small-scale data prediction, especially DNN, which can be seen that it needs large enough training data to predict with high accuracy. RFR gives the best results in small-scale data prediction and the table of results of RFR-based small-scale data prediction is given here as shown in **Table 12**.

F. RESEARCH RESULTS DISCUSSION

The finding of this study that the Relative Maximum Connectivity Component Size (RSLCC) has a decisive effect on network connectivity after a disaster not only confirms our initial assumptions (**Figure 2**) but also reveals the importance of RSLCC in the analysis of network structure. By analysing 100,000 datasets, we demonstrate that deep neural networks (DNNs) can achieve more than 99% accuracy in predicting RSLCC values, a finding that underscores the effectiveness of network connectivity prediction using advanced machine

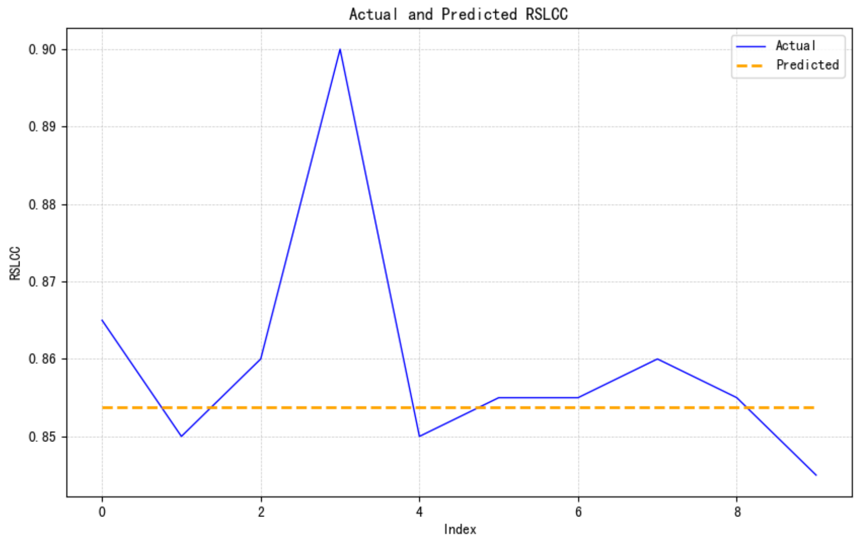


FIGURE 14. Predicting the dependent variable RSLCC from the independent variable NDD. (The purpose of this figure is to see the extent of change in the difference between actual and predicted values after replacing the dataset)

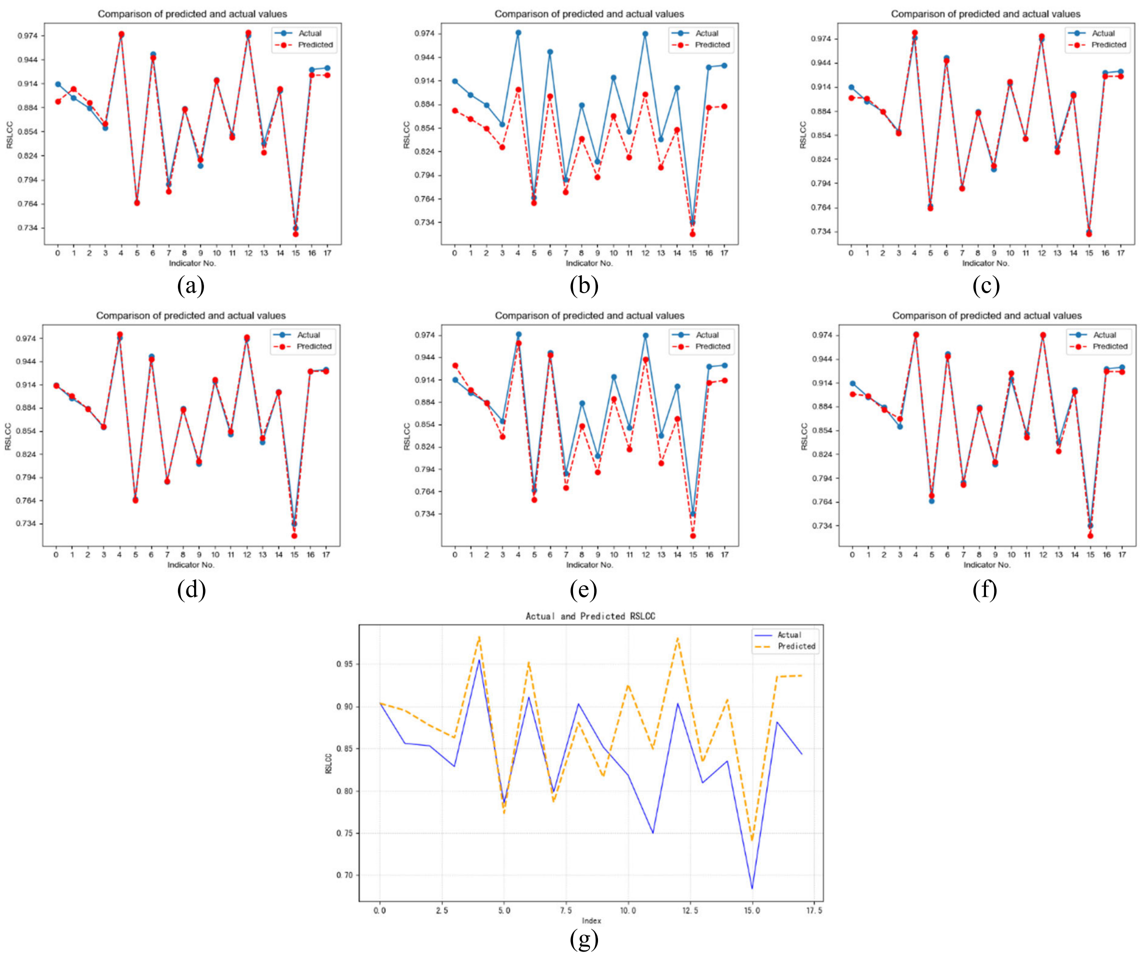


FIGURE 15. a.Comparison of predicted and actual values (DTR). b. Comparison of predicted and actual values (SVM) c. Comparison of predicted and actual values (RFR). d. Comparison of predicted and actual values (GBR). e. Comparison of predicted and actual values (MLP). f. Comparison of predicted and actual values (KNN). g. Comparison of predicted and actual values (DNN).

learning techniques. This echoes the review study of Boccaletti et al. [17] on the structural properties of complex networks, who pointed out the potential value of studying network connectivity metrics in disaster contexts. In addition, Vugrin et al. [19] discussed the assessment of network connectivity and recovery strategies by analysing the case of supply chains under the impact of hurricanes, which fit with our approach of simplified analysis of network connectivity through NDD metrics. Our findings are further supported by the study of Hazra et al. [24], who pointed out the importance of improving communication network connectivity through multi-information source infrastructure restoration after a disaster. In summary, our study complements and extends the existing literature both theoretically and practically. By deeply exploring the application of RSLCC in post-disaster network connectivity assessment, we not only validate the existing theoretical framework but also provide new analytical tools and perspectives for disaster management. Future research can further explore the application of RSLCC and other network connectivity metrics in different types of disasters and networks to enhance the effectiveness of disaster response and recovery strategies.

VI. CONCLUSION

In this study, five metrics related to network connectivity are given based on the literature review from the perspective of network structure analysis. The 100,000 datasets generated by NetworkX based on Python are analysed using statistical analysis and machine learning prediction algorithms, and we not only find one metric that is approximately related to network connectivity, but also use 7 prediction methods to predict the metrics, and specifically conclusions can be seen as follows:

(1) The Relative Size of the Largest Connected Component (RSLCC) is a good representation of network connectivity, in other words, it can be used as a proximity metric for measuring network connectivity.

(2) The Deep Neural Network (DNN) model trained by this study is more than 99% accurate in predicting RSLCC values and it can be proven to be able to be used for practical prediction.

(3) The relationship between RSLCC and network connectivity found in this study can be used for any network mathematical modelling, simulation applications, etc.

(4) Although RSLCC does not characterize 100% of network connectivity, and only comes close, this study provides a good research direction to continue exploring network connectivity from a data-driven perspective, making pioneering contributions to both prediction model improvement and metrics quantification.

(5) The findings of this study can be applied to transport networks, logistics networks, supply chain networks, energy networks, power networks, information networks, virtual networks, physical networks, complex networks, social networks, and other types of networks.

There are still some key issues that need to be addressed through the research in this paper, firstly, when training deep neural networks, if the backpropagation algorithm can be added on top of this, the predicted output of the model is first calculated through feedforward propagation, and then the gradient is calculated through backpropagation, the parameters are updated, and the model is continuously optimized iteratively. Feedforward propagation and backpropagation work closely together, feedforward propagation calculates the output, backpropagation calculates the gradient, and the two work together to complete the training process of the model, which may further improve the prediction accuracy. Then, there are also some important research techniques and methods that we should further learn from, for example, Chiarion et al. [41] introduced how to effectively use EEG data to understand the interactions and connectivity between brain regions, which provides us with ideas to learn from for data-driven network connectivity. In addition, Nave [42] proposed an improved semi-analytic method for solving the problem of Ordinary Differential Equation (ODE) systems can inspire us to explore the quantitative study of network connectivity under complex network systems. Finally, what this paper has done so far is just an exploration of the metric prediction, and whether the prediction result can be used in practical research and application needs to be further proved, so in the future, it can be tried combined with modelling to solve practical problems.

AUTHOR CONTRIBUTIONS

Junxiang Xu: introduction, method, literature review, writing-original draft.

Divya Jayakumar Nair: writing—review and editing, supervision.

DECLARATION OF INTEREST STATEMENT

The authors declare that they have no conflict of interest. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

REFERENCES

[1] T. Rossetto, N. Peiris, A. Pomonis, S. M. Wilkinson, D. Del Re, R. Koo, and S. Gallocher, "The Indian Ocean tsunami of December 26, 2004: Observations in Sri Lanka and Thailand," *Natural Hazards*, vol. 42, no. 1, pp. 105–124, 2007.

[2] Y. Song, Z. Li, X. Zhang, and M. Zhang, "Study on indirect economic impacts and their causes of the 2008 wenchuan earthquake," *Natural Hazards*, vol. 108, no. 2, pp. 1971–1995, Sep. 2021.

[3] O. Norio, T. Ye, Y. Kajitani, P. Shi, and H. Tatano, "The 2011 eastern Japan great earthquake disaster: Overview and comments," *Int. J. Disaster Risk Sci.*, vol. 2, no. 1, pp. 34–42, Mar. 2011.

[4] G. Wittwer and R. Waschik, "Estimating the economic impacts of the 2017–2019 drought and 2019–2020 bushfires on regional NSW and the rest of Australia," *Austral. J. Agricult. Resource Econ.*, vol. 65, no. 4, pp. 918–936, Oct. 2021.

[5] M. O. Ball, C. J. Colbourn, and J. S. Provan, "Network reliability," in *Handbooks in Operations Research and Management Science*. Amsterdam, The Netherlands: Elsevier, 1995, ch. 11, pp. 673–762.

[6] M. J. F. Alenazi and J. P. G. Sterbenz, "Comprehensive comparison and accuracy of graph metrics in predicting network resilience," in *Proc. 11th Int. Conf. Design Reliable Commun. Netw. (DRCN)*, Mar. 2015, pp. 157–164.

[7] D. Fan, B. Sun, H. Dui, J. Zhong, Z. Wang, Y. Ren, and Z. Wang, "A modified connectivity link addition strategy to improve the resilience of multiplex networks against attacks," *Rel. Eng. Syst. Saf.*, vol. 221, May 2022, Art. no. 108294.

[8] Y. Wang, O. Zhao, and L. Zhang, "Modeling urban rail transit system resilience under natural disasters: A two-layer network framework based on link flow," *Rel. Eng. Syst. Saf.*, vol. 241, Jan. 2024, Art. no. 109619.

[9] W. Wei, Q. Hu, and Q. Zhang, "Improving node connectivity by optimized dual tree-based effective node consolidation," *Rel. Eng. Syst. Saf.*, vol. 242, Feb. 2024, Art. no. 109747.

[10] M. J. Jafri, G. D. Pearlson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1666–1681, 2008.

[11] X. Zhou, R. K. Ganti, and J. G. Andrews, "Secure wireless network connectivity with multi-antenna transmission," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 425–430, Feb. 2011.

[12] S. Mishra, T. F. Welch, and M. K. Jha, "Performance indicators for public transit connectivity in multi-modal transportation networks," *Transp. Res. A, Policy Pract.*, vol. 46, no. 7, pp. 1066–1085, Aug. 2012.

[13] L. Wu, Q. Tan, and Y. Zhang, "Network connectivity entropy and its application on network connectivity reliability," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 21, pp. 5536–5541, Nov. 2013.

[14] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: The topology of the world-wide web," *Phys. A, Stat. Mech. Appl.*, vol. 281, nos. 1–4, pp. 69–77, Jun. 2000.

[15] R. G. Little, "Controlling cascading failure: Understanding the vulnerabilities of interconnected infrastructures," *J. Urban Technol.*, vol. 9, no. 1, pp. 109–123, Apr. 2002.

[16] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. von Winterfeldt, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthq. Spectra*, vol. 19, no. 4, pp. 733–752, Nov. 2003.

[17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, 2006.

[18] G. P. Cimellaro, A. M. Reinhorn, and M. Bruneau, "Framework for analytical quantification of disaster resilience," *Eng. Struct.*, vol. 32, no. 11, pp. 3639–3649, Nov. 2010.

[19] E. D. Vugrin, D. E. Warren, and M. A. Ehlen, "A resilience assessment framework for infrastructure and economic systems: Quantitative and qualitative resilience analysis of petrochemical supply chains to a hurricane," *Process Saf. Prog.*, vol. 30, no. 3, pp. 280–290, Sep. 2011.

[20] S. P. Simonovic and A. Peck, "Dynamic resilience to climate change caused natural disasters in coastal megacities quantification framework," *Brit. J. Environ. Climate Change*, vol. 3, no. 3, pp. 378–401, Sep. 2013.

[21] J. Gao, D. Li, and S. Havlin, "From a single network to a network of networks," *Nat. Sci. Rev.*, vol. 1, no. 3, pp. 346–356, Sep. 2014.

[22] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electr. Power Syst. Res.*, vol. 127, pp. 259–270, Oct. 2015.

[23] U. Bhatia, D. Kumar, E. Kodra, and A. R. Ganguly, "Network science based quantification of resilience demonstrated on the Indian railways network," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0141890.

[24] K. Hazra, V. K. Shah, S. Silvestri, V. Aggarwal, S. K. Das, S. Nandi, and S. Saha, "Designing efficient communication infrastructure in post-disaster situations with limited availability of network resources," *Comput. Commun.*, vol. 164, pp. 54–68, Dec. 2020.

[25] S. Kameshwar, H. Park, D. T. Cox, and A. R. Barbosa, "Effect of disaster debris, floodwater pooling duration, and bridge damage on immediate post-tsunami connectivity," *Int. J. Disaster Risk Reduction*, vol. 56, Apr. 2021, Art. no. 102119.

[26] K.-H. Chang, J.-B. Sheu, Y. J. Chen, C.-H. Chang, and C.-H. Liu, "Practice-based post-disaster road network connectivity analysis using a data-driven percolation theory-based method," *Transp. Res. E, Logistics Transp. Rev.*, vol. 171, Mar. 2023, Art. no. 103022.

[27] Z. Li and W. Yan, "Service flow changes in multilayer networks: A framework for measuring urban disaster resilience based on availability to critical facilities," *Landscape Urban Planning*, vol. 244, Apr. 2024, Art. no. 104996.

[28] C. Ricotta, A. Stanisci, G. C. Avena, and C. Blasi, "Quantifying the network connectivity of landscape mosaics: A graph-theoretical approach," *Community Ecol.*, vol. 1, no. 1, pp. 89–94, Jul. 2000.

[29] B. Barzel and O. Biham, "Quantifying the connectivity of a network: The network correlation function method," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 4, 2009, Art. no. 046104.

[30] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010.

[31] J. L. Sullivan, D. C. Novak, L. Aultman-Hall, and D. M. Scott, "Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 323–336, Jun. 2010.

[32] A. Zalesky, A. Fornito, and E. Bullmore, "On the use of correlation as a measure of network connectivity," *NeuroImage*, vol. 60, no. 4, pp. 2096–2106, May 2012.

[33] B. Barzel and A.-L. Barabási, "Network link prediction by global silencing of indirect correlations," *Nature Biotechnol.*, vol. 31, no. 8, pp. 720–725, Aug. 2013.

[34] M. Bolaños, E. M. Bernat, B. He, and S. Aviyente, "A weighted small world network measure for assessing functional connectivity," *J. Neurosci. Methods*, vol. 212, no. 1, pp. 133–142, Jan. 2013.

[35] L. Milli, A. Monreale, G. Rossetti, D. Pedreschi, F. Giannotti, and F. Sebastiani, "Quantification in social networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1–10.

[36] J. R. Cohen and M. D'Esposito, "The segregation and integration of distinct brain networks and their relationship to cognition," *J. Neurosci.*, vol. 36, no. 48, pp. 12083–12094, Nov. 2016.

[37] Y. Qian, Y. Li, M. Zhang, G. Ma, and F. Lu, "Quantifying edge significance on maintaining global connectivity," *Sci. Rep.*, vol. 7, no. 1, p. 45380, Mar. 2017.

[38] Y. Zhang and S. T. Ng, "Identification and quantification of node criticality through EWM–TOPSIS: A study of Hong Kong's MTR system," *Urban Rail Transit*, vol. 7, no. 3, pp. 226–239, Sep. 2021.

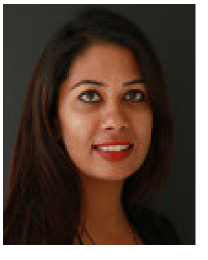
[39] G. S. Cumming, R. A. Magris, and K. Maciejewski, "Quantifying cross-scale patch contributions to spatial connectivity," *Landscape Ecol.*, vol. 37, no. 9, pp. 2255–2272, Sep. 2022.

[40] S. Sharma and S. Ram, "Investigation of road network connectivity and accessibility in less accessible airport regions: The case of India," *Sustainability*, vol. 15, no. 7, p. 5747, Mar. 2023.

[41] G. Chiarion, L. Sparacino, Y. Antonacci, L. Faes, and L. Mesin, "Connectivity analysis in EEG data: A tutorial review of the state of the art and emerging trends," *Bioengineering*, vol. 10, no. 3, p. 372, Mar. 2023.

[42] O. Nave, "Modification of semi-analytical method applied system of ODE," *Mod. Appl. Sci.*, vol. 14, no. 6, p. 75, May 2020.

JUNXIANG XU is currently pursuing the Ph.D. degree with the School of Civil and Environmental Engineering, University of New South Wales (UNSW). His research interests include emergency transport planning, relief transport network design, evacuation network planning, and travel behavior modeling.



DIVYA JAYAKUMAR NAIR is currently a Senior Lecturer and an Advance HE Fellow with the School of Civil and Environmental Engineering, University of New South Wales (UNSW). Her research interests include the logistics of non-profit organizations, vehicle routing problems, optimization problems, and traffic assignment.