Graph Learning-Based Spatial Structural Identification of Drought Regions

Jingxin Zhang[1]

[1]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China - 2108160323005@stu.bucea.edu.cn

Keywords: Complex Networks; GCN; Spatial Structure Identification; Drought Regions.

Abstract

Under the backdrop of global climate change, the frequency and severity of drought events are continuously increasing, posing significant challenges to human society, ecosystems, and economic development. Traditional drought simulation methods often overlook the interactions among meteorological, hydrological, and geographic information. Complex network theory offers a new perspective for exploring these interconnections. Graph Neural Networks (GNNs), as a deep learning technique capable of handling geospatial data and complex structures, have advantages in capturing geographic correlation information and network topology. Therefore, combining complex networks and GNNs for drought simulation is of great significance. This study proposes a framework that integrates complex networks and graph learning to identify the spatial structure of drought regions. Using latitude-longitude grids as nodes, topological indicators such as degree, betweenness centrality, and clustering coefficient describe node features, which are combined with SPEI time series statistical features to form multidimensional vectors. These are input into a Graph Convolutional Network (GCN) to obtain low-dimensional embeddings, and clustering is used to divide the space into subregions. Results show that the clustering based on multi-feature combinations exhibits stronger spatial continuity and clearer boundaries. Regions with high degree and betweenness centrality and low clustering coefficient serve as network hubs and information bridges, while medium-feature regions are intermediate connecting zones, and regions with low feature values are peripheral isolated areas. This method offers a novel approach to analyzing drought system structures and regional risk management.

1. Introduction

In the context of accelerating global climate change, drought events are threatening human societies, ecosystems and regional sustainable development with significantly increased frequency and intensity. Unlike other sudden-onset natural disasters, droughts permeate the hydrological-ecological social system with ‘slow-onset and insidious’ characteristics, and their multi-layer driving mechanisms (e.g., anomalies in atmospheric circulation, imbalances in land-surface processes, and human intervention in water use) and nonlinear evolution patterns have always been a difficult issue in geoscientific research. Its multi-circle driving mechanism (e.g., atmospheric circulation anomaly, land surface process balance, human water intervention) and nonlinear evolution law have always been the difficult problems in geoscientific research. Against the background of the gradual severity of multi-circle composite extreme events, droughts show new characteristics, with the traditional slow-onset droughts shifting to sudden droughts on the one hand (Yuan et al., 2023), and the frequency and intensity of composite droughts gradually increasing on the other (Hao Zengchao and Chen Yang, 2024). Although the related aspects have become a hot issue in domestic and international research, the research on the new characteristics of droughts and the spatial and temporal propagation patterns of complex droughts is still limited due to the insufficient sample size and the increase in the uncertainty of climate model simulations caused by the dominance of dynamical processes (Singh et al, 2021; Zscheischler and Lehner, 2022; Mondal et al., 2023). Exploring new effective means to further grasp the new characteristics of drought and the spatial and temporal propagation laws is the key to achieving drought prediction and early warning, improving the scientific and technological level of drought management, and reducing drought losses.

The most primitive driving factor of drought is the water balance imbalance phenomenon caused by the anomalies of the climate system, which, as a typical nonlinear dissipative giant system, has cross-scale coupling characteristics that are particularly prominent in drought remote correlation studies. (Liu Shida et al., 2002; Jia Xiaojing et al., 2003; Ren Hongli and U Jifan, 2005; Ding Ruiqiang and Li Jianping, 2009; Fu et al., 2016). On a large regional scale, complex networks provide an effective means to analyse drought remote correlations by abstracting drought impact elements into a ‘node-edge’ topology. In recent years, the complex network method has been widely used by the geosciences in studies related to extreme climate events, including extreme precipitation events, anomalous SST patterns (Agarwal et al., 2019; Lu et al., 2022), and extreme high-temperature heatwaves (Mondal and Mishra, 2021; Zhang et al., 2023), etc., and the results of the studies have provided a new methodology and new methodology for the climatic extreme event The results of this study provide new methods and perspectives for characterising climate extreme events (Feng Guolin et al., 2006). Complex network theory provides a powerful framework for modelling the interdependencies among environmental, climatic and geographical factors in drought-prone regions.

Meanwhile, Graph Neural Network (GNN) has emerged as an effective tool for learning from spatially structured data (Scarselli et al., 2009), as a method that can directly deal with most of the actual graph structures, it is able to capture topological information and attribute interactions, build effective data-driven dynamic models on the network by integrating geospatial and temporal information to effectively capture complex spatio-temporal dynamic trends of events (Scarselli et al., 2009; Yu et al., 2023). Currently graph neural networks have achieved some results in the analysis and extrapolation of time-series data, such as the prediction of crop yields (Fung et al., 2019), the simulation and extrapolation of the spread of Corona Virus Disease 2019 (COVID-19) (Murphy et al., 2021), and the traffic flow prediction, etc.

Taking advantage of these complementary strengths, this study proposes an integrated framework that combines complex network modelling with graph-based learning to identify spatial structural patterns within drought-affected regions. By constructing drought-related complex networks and extracting representative topological features, joint topology-attribute characterization is achieved through the combination of the neighbourhood feature propagation mechanism of graph neural networks, and ultimately, spatial sub-regions with different drought propagation functions are delineated by clustering. This method introduces the concept of 'hub-node-bridge-edge-community structure' of network science into the spatial analysis of drought, which provides new insights into the structural organisation of drought systems, and provides theoretical support for analysing the network emergence mechanism of the rapid evolution of sudden drought and constructing a composite drought early warning model.

2. Data

In this study, the Standardized Precipitation Evapotranspiration Index (SPEI) data employed span the period from 1960 to 2023, with a spatial resolution of 1° × 1°, covering the entire globe. This dataset quantifies the balance of water supply and demand, thereby providing multidimensional and standardized metrics for drought definition. By setting different threshold values, drought severity can be classified. The classification criteria are typically based on the standardized probability distribution. The following thresholds are internationally accepted and commonly used in related research:

SPEI Range	Drought Severity Level
SPEI ≤ -2	Extreme Drought
-2 < SPEI ≤ -1.5	Severe Drought
-1.5 < SPEI ≤ -1	Moderate Drought
-1 < SPEI ≤ -0.5	Mild Drought
-0.5 < SPEI ≤ 0.5	No Drought
0.5 < SPEI ≤ 1	Mild Wet
1 < SPEI ≤ 1.5	Moderate Wet
1.5 < SPEI ≤ 2	Severe Wet
SPEI ≥ 2	Extreme Wet

Table 1 SPEI Drought Severity Classification

3. Methodology

This study establishes a multi-scale analytical framework that integrates complex network theory with graph neural networks to investigate the spatiotemporal dynamics of global drought evolution comprehensively. First, based on global SPEI data spanning 1901-2023, the event synchronization method is employed to construct a drought complex network and extract its topological properties, including degree centrality, betweenness centrality, and clustering coefficient. The network metrics and adjacency matrix are then used as input information for the spatial layer of the GCN model. This approach ultimately reveals the hierarchical structural characteristics of the global drought network, identifies critical drought hub regions, and quantifies the spatiotemporal scales of drought propagation.

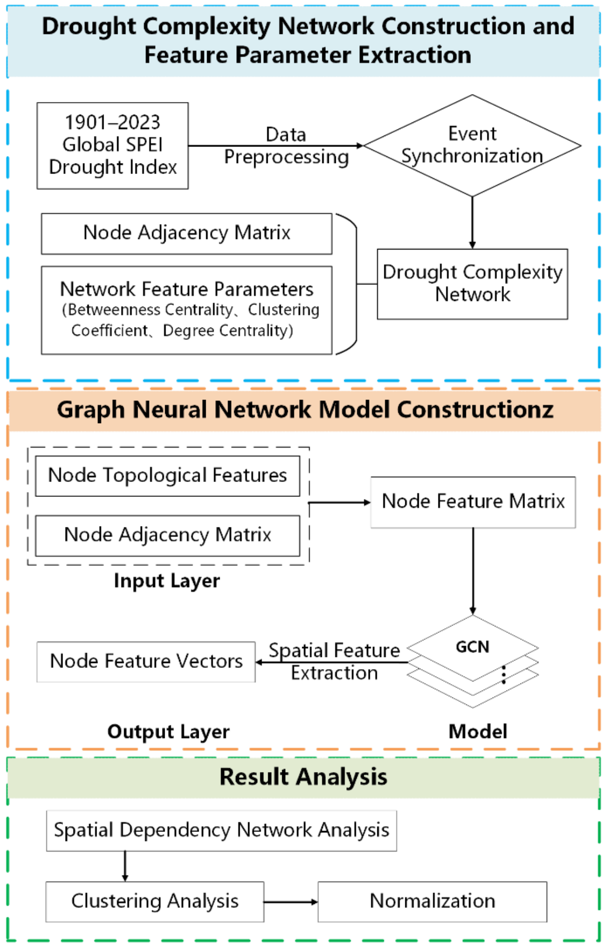

Figure 1 Schematic Diagram of the Methodological Framework

3.1 Construction of the Complex Network

The global drought complex network is constructed using the Event Synchronization (ES) method to explore the spatiotemporal connections among extreme drought events occurring at different locations. By analyzing time series data, the event synchronization approach effectively mitigates analytical complexity arising from confounding factors, making it particularly well-suited for investigating extreme climate events characterized by nonlinear and nonparametric properties.

The following equation can express the time interval of event synchronization:

$$d_{ij}^{lm} = t_i^l - t_j^m, \quad (1)$$

here, t_j^l and t_j^m denote the occurrence times of the l-th drought event at grid point i and the m-th drought event at grid point j, respectively. $t_i^l - t_i^{l-1}$ represents the time interval between the l-th and the (l−1)-th drought events at location i. $t_i^l - t_j^m$ is defined as the dynamic delay between two drought events. To ensure that the dynamic delay remains as reasonable as possible, a delay threshold τ_{max} is specified so that the dynamic delay is constrained within the interval $[0, \tau_{max}]$. When the condition $0 < t_i^l - t_j^m < \tau_{ij}^{lm}$ and $0 < t_i^l - t_j^m < \tau_{max}$ is satisfied, it is considered that the l-th event at grid point i and the m-th drought event at grid point j are synchronized, and the l-th event precedes the m-th event.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W14-2025-395-2025 | © Author(s) 2025. CC BY 4.0 License.

In this framework, each latitude-longitude grid cell is treated as a node in the drought complex network. The graph is defined as $G(V,L)$, where $V=\{v_1,v_2,\cdots v_N\}$ denotes the set of nodes (each corresponding to a spatial grid cell), and $L=\{l_1,l_2,\cdots l_K\}$ represents the set of edges. An edge is established between two nodes v_i and v_j.

3.2 Extraction of Network Feature Parameters

The resulting complex network captures the global patterns of drought co-variability. From this network, several topological indicators are extracted-such as degree, betweenness centrality, and clustering coefficient-which describe the connectivity, intermediary role, and local cohesiveness of each node(Murphy et al., 2021). These structural features are used as inputs to the subsequent graph learning framework, providing topological priors for spatial representation learning.

3.2.1 Degree Centrality: In a weighted directed network, the node degree reflects not only the number of connections but also integrates the directionality and the weight of those connections. In the drought complex network, the drought propagation strength can be calculated using the following equations:
The outward drought strength originating from node i, denoted as $\mathrm{Sr}_i^{\mathrm{out}}$, is defined as:

$$Sr_i^{out}=\sum_{j=1}^{N}C_{ij}\,, \tag{2}$$

the inward drought propagation strength received by node i, denoted as Sr_i^{in}, is computed as:

$$Sr_i^{in}=\sum_{j=1}^{N}C_{ji}\,, \tag{3}$$

where N represents the total number of grid cells within the study area, and C is the $N\times N$ adjacency matrix. Each element Cij indicates the probability of drought propagating from grid cell i to grid cell j. A higher value of Sr_i^{out} implies a stronger ability of grid cell i to transmit drought to other locations, while a higher Sr_i^{in} indicates that grid cell i is more susceptible to receiving drought influence from others.

To further analyze the patterns of drought propagation within the complex network, the study calculated the outward and inward strengths for each grid cell in the study area. The difference between them was used to determine whether a grid cell acts primarily as a source or a sink of drought propagation within the network. This difference is defined as:

$$\Delta Sr_i=Sr_i^{in}-Sr_i^{out}, \tag{4}$$

when ΔSr is positive, it indicates that the drought influence received by grid cell i exceeds the influence it transmits outward, and therefore this grid cell can be considered a convergence point (sink region) of drought propagation. Conversely, if ΔSr is negative, it implies that the outward propagation strength exceeds the inward strength, indicating that this grid cell serves as a source region of drought events.

3.2.2 Betweenness Centrality: Betweenness centrality quantifies the influence of a node on information transmission between other nodes. It characterizes the importance of a node by counting the number of shortest paths that pass through it. In the context of the drought complex network, betweenness centrality reflects the likelihood that drought propagates through a given grid location. A node with higher betweenness centrality plays a more significant role in facilitating drought transmission.

The betweenness centrality of node i is defined as:

$$BC_i=\sum_{s\neq i\neq t}\frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{5}$$

where σ_{st} denotes the total number of shortest paths between node s and node t, and $\sigma_{st}(i)$ represents the number of those paths that pass through node i. Nodes with higher betweenness centrality serve as critical bridges in the process of drought propagation.

3.2.3 Clustering Coefficient: The clustering coefficient measures the degree to which the neighbors of a node are interconnected, reflecting the tendency of a node's neighboring nodes to form a tightly knit cluster. In the drought complex network, it indicates the consistency of drought occurrences among the grid cells connected to a given grid cell.

For an undirected weighted network, the clustering coefficient of node iii, denoted as CC_i, can be defined as :

$$CC_i=\frac{1}{k_i(k_i-1)}\sum_{j,k}\frac{(w_{ij}+w_{ik})}{2w_i}\,a_{ij}\,a_{ik}\,a_{jk}, \tag{6}$$

where k_is the degree of node i, w_i denotes the average weight of all edges connected to node$i$, w_{ij} and w_{ik} are the weights of the edges between node i and its neighboring nodes j and k, respectively, and a_{ij}, a_{ik}, and a_{jk} are the elements of the adjacency matrix.

A higher clustering coefficient indicates a greater degree of consistency in drought occurrences among the grid cells connected to the given grid cell.

3.3 Construction of the Graph Neural Network

These structural indicators are then combined with statistical characteristics of the SPEI time series (e.g., mean, variance) to construct a multi-dimensional feature vector for each node. The resulting feature matrix serves as input to a Graph Convolutional Network (GCN) model. After training and hyperparameter tuning, low-dimensional node embeddings are obtained. Finally, K-means clustering is applied to the embeddings to identify spatial heterogeneity and functional subregions in the global drought network.

The core idea of Graph Convolutional Networks (GCNs) is to update the representation of each node by aggregating the feature information of its neighboring nodes through a weighted average. Specifically, for each node, GCN combines its own features with those of its neighbors using a linear transformation, followed by a non-linear activation function. The core propagation rule is expressed as:

$$\mathrm{H}^{(l+1)}=\sigma\left(\widehat{\mathrm{D}}^{-0.5}\widehat{\mathrm{A}}\widehat{\mathrm{D}}^{-0.5}\mathrm{H}^{(\mathrm{l})}\mathrm{W}^{(\mathrm{l})}\right), \tag{7}$$

where $\mathrm{H}^{(\mathrm{l})}$ denotes the node feature matrix at the l-th layer, $\widehat{\mathrm{A}}$ is the adjacency matrix with added self-loops, $\widehat{\mathrm{D}}$ is the degree matrix corresponding to A, and σ is an activation function, typically the ReLU function.

This convolution operation captures spatial dependencies through weighted summation over each node and its neighbors.

After obtaining the node embeddings, the K-Means algorithm was applied to cluster nodes, mapping similar nodes in the high-dimensional embedding space back to the geographic domain, and comparing the clustering outcomes across different input configurations. Multi-feature composite inputs-combining local connectivity strength (degree), global propagation control (betweenness centrality), and small-scale aggregation characteristics (clustering coefficient)-enabled a more comprehensive characterization of the multi-scale topological mechanisms driving drought propagation. As a result, the clustering outputs demonstrated stronger spatial continuity and clearer boundaries. In contrast, single-feature inputs were limited in their information dimensions and exhibited ambiguity in boundary delineation, thereby validating the added value of multi-feature integration for interpreting drought spatial patterns within the graph learning framework.

4. Results

4.1 Global Distribution of Drought Network Topological Features

Following the construction of the drought-related complex network, three representative structural features were extracted to characterize the topological role and structural position of each node within the network: node degree, betweenness centrality, and clustering coefficient. As illustrated in Figure 2, the global distribution of node degree reveals the connectivity of drought-related nodes; higher values indicate stronger connections to other nodes.

It can be clearly observed that in regions such as the Eurasian continent and parts of North America, node degree values are relatively high, reflecting that these areas exhibit tighter drought linkages with other regions and act as critical hubs for drought signal propagation. In contrast, in some high-latitude zones or tropical rainforest regions, node degree values are comparatively lower, suggesting a more isolated pattern of drought connectivity.

Notably, regions of high drought connectivity are mainly concentrated in three core belts: The Mediterranean-Central Asia drought corridor, which is alternately influenced by the subtropical high and westerlies, serving as a key hub for drought transmission across Eurasia; The interior of Australia, closely linked to Indian Ocean Dipole anomalies triggered by El Niño events; The Sahel region of Africa, where node degree centrality is also prominent, reflecting its role as a trans-equatorial bridge for drought signal transmission.

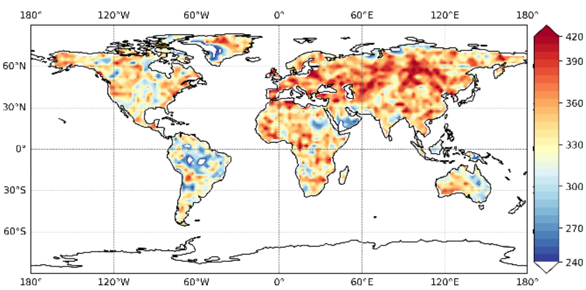

Figure 2 Global Distribution of Node Degree Centrality in the Drought-Related Complex Network

4.2 Clustering Analysis under Multi-Feature Inputs

To evaluate the impact of different structural feature inputs on graph neural network embeddings and subsequent clustering performance, this study designed four input configurations: using node degree alone, betweenness centrality alone, clustering coefficient alone, and a composite input combining all three features. These features were fed into the GCN model to generate node embeddings, which were then clustered using the K-Means algorithm. The clustering labels were mapped back to the geographic space, and the results are shown in Figure 3.

The results demonstrate that the multi-feature composite input (Figure 3d) yields clustering outcomes characterized by stronger spatial continuity and clearer boundaries, with distinct differentiation among categories. In Figure 3d, major clustered regions are observed to form strip-like or patch-like spatial patterns. Notably, the Mediterranean-Central Asia drought corridor, the interior of Australia, and the Sahel region of Africa show high consistency between the clusters and the areas with elevated topological feature values, indicating significant homogeneity and spatial aggregation of drought network characteristics. In addition, northeastern South America and the central and western United States also exhibit clustering regions that are markedly different from their surroundings, further illustrating that multi-feature inputs effectively enhance the discriminative capacity of node representations.

In contrast, while clustering results based on single-feature inputs broadly exhibit similar spatial trends to those of the composite input, there are clear differences in category boundaries in many areas. For example, Figure 3a (node degree input) successfully identifies high-connectivity regions across Eurasia, but the transitions between these regions and their surroundings are relatively blurred, and mid-latitude zones exhibit fragmented classification patterns. Figure 3b (betweenness centrality input) highlights the core hubs in Central Asia and North America but produces more scattered category distributions across the Southern Hemisphere. Figure 3c (clustering coefficient input) shows less distinct category boundaries in Africa and South America, resulting in substantial discontinuities.

Overall, single structural features are limited in their ability to represent the complex topological relationships of the drought network comprehensively and accurately. In contrast, the multi-feature composite input, by integrating local connectivity strength (node degree), global propagation control (betweenness centrality), and neighborhood aggregation (clustering coefficient), not only improves the expressive power of node embeddings but also produces more coherent and interpretable spatial partitioning. This provides a richer and more nuanced perspective on the spatial clustering patterns of the global drought network. The findings further validate the substantial added value of multi-feature fusion in graph-based learning and establish a foundation for uncovering the networked mechanisms underlying drought propagation.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W14-2025-395-2025 | © Author(s) 2025. CC BY 4.0 License.

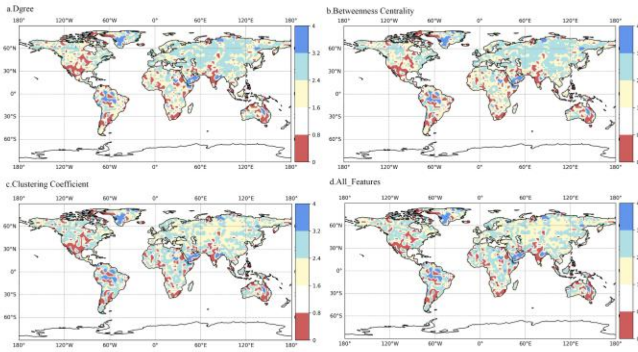

Figure 3 Clustering results based on different structural feature inputs
(a. Degree centrality input; b. Betweenness centrality input; c. Clustering coefficient input; d. Combined input of all three features)

4.3 Contribution Analysis of Network Features in Drought Clustering

To further interpret the structural-functional roles of each cluster, the average values of the three structural indicators were calculated for each category, and a normalized radar chart was drawn (Figure 4). The results show that Cluster 1 is characterized by high degree and betweenness centrality but low clustering coefficient, indicating a dual role as a network hub and information transmission bridge. This suggests that nodes in Cluster 1 are key connectors that facilitate the rapid spread of drought signals across otherwise distant regions. Cluster 2 displays moderate structural features and may represent intermediate connector regions within the network, potentially serving as transition zones that mediate interactions between core hubs and peripheral areas. Cluster 3 exhibits low values across all three indicators, reflecting peripheral or isolated characteristics, possibly corresponding to areas with delayed drought response or limited information flow. Additionally, the radar chart highlights the clear differentiation among clusters in their structural profiles, underscoring the importance of integrating multiple topological attributes to accurately capture the functional heterogeneity of the drought network. Such distinctions are essential for identifying priority regions in drought monitoring and for designing targeted intervention strategies to improve early warning systems and resilience planning.

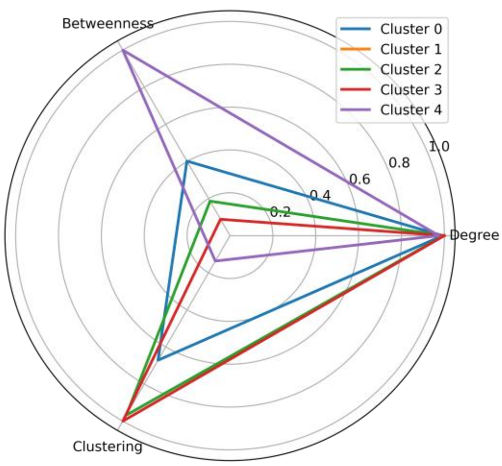

Figure 4 Radar Chart of Structural Feature Distributions for Different Drought Clusters

5. Discussions

The findings of this study provide compelling evidence that combining complex network theory with graph learning methods enables a more nuanced characterization of drought spatial structures compared to traditional approaches. In particular, the extraction of node degree, betweenness centrality, and clustering coefficient allowed the identification of regions that serve as hubs or bridges in the global drought propagation network, which are not readily captured by index-based analyses alone.

The performance comparison of different feature input schemes further illustrates that multi-feature integration can significantly enhance the discriminative power of node embeddings. The resulting clusters display clearer spatial boundaries and more coherent regional patterns, which are critical for operational drought monitoring and intervention planning. This supports the hypothesis that structural heterogeneity is an essential aspect of drought dynamics and must be accounted for when designing early warning systems and resilience strategies.

Nevertheless, this study has several limitations. First, the temporal resolution of the SPEI data (monthly) may obscure short-term drought episodes or rapid transitions. Second, while the chosen topological indicators are representative, they do not exhaustively capture all possible structural characteristics of drought networks. Incorporating other graph metrics (e.g., eigenvector centrality, modularity) and temporal graph models could yield additional insights. Finally, this framework has not yet integrated socioeconomic and land use factors that may modulate the actual impacts of drought events. Future research should explore multilayer graph representations that combine climatic, environmental, and socioeconomic networks to support holistic risk assessment and management.

Overall, this study demonstrates the value of graph-based learning approaches for uncovering hidden spatial structures in drought systems, offering an innovative foundation for more targeted and proactive drought governance.

References

Agarwal A., Caesar L., Marwan N., Maheswaran R., Merz B., Kurths J. 2019. Network-based identification and characterization of teleconnections on different scales. *Scientific Reports*, 9: 8808.

Chen J., Li M., Wang W. 2012. Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast. *Mathematical Problems in Engineering*, 915053.

Chen Y., Li P., Hao Z., Feng S., Liao Z., An N. 2023. The 2022 Sichuan-Chongqing spatio-temporally compound extremes: a bitter taste of novel hazards. *Science Bulletin*, 68(13): 1337-1339.

Feng P., Wang B., Liu D. L., Yu Q. 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*, 173: 303-316.

Hao Z., Chen Y., Feng S., Liao Z., An N., Li P. 2023. The 2022 Sichuan-Chongqing spatio-temporally compound extremes: a bitter taste of novel hazards. *Science Bulletin*, 68(13): 1337-1339.

Lu Z. H., Dong W. J., Lu B., Yuan N. M., Ma Z. G., Bogachev M. I., Kurths J. 2022. Early warning of the Indian Ocean Dipole using climate network analysis. *Proceedings of The National Academy of Sciences of The United States of America*, 119(11): 2109089119.

Mondal S., Mishar A. K., Leung R., et al. 2023. Global droughts connected by linkages between drought hubs. *Nature Communications*, 14: 144.

Murphy C., Laurence, Allard A. 2021. Deep learning of contagion dynamics on complex networks. *Nature Communications*, 12: 4720.

Scarselli F., Gori M., Tsoi A. C., et al. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20: 61-80.

Singh J., Ashfaq M., Skinner C. B., et al. 2021. Amplified risk of spatially compounding droughts during co-occurrences of modes of natural ocean variability. *Climate and Atmospheric Science*, 4: 7.

Yuan X., Wang Y., Ji P., Wu P., Sheffield J., Otkin J. A. 2023. A global transition to flash droughts under climate change. *Science*, 380(6641): 187-191.

Zscheischler J., Lehner F. 2022. Attributing Compound Events to Anthropogenic Climate Change. *Bulletin of the American Meteorological Society,* 103(3): 936-953.