

Hyperbolic map unravels eight regions in temperature volatility regionalization of Mainland China

Yuxuan Song(宋雨轩), Changgui Gu(顾长贵), Muhua Zheng(郑木华), Aixia Feng(冯爱霞), Yufei Xi(席雨菲), Haiying Wang(王海英), and Huijie Yang(杨会杰)
Citation: Chin. Phys. B, 2024, 33 (**12**): 128902. DOI: 10.1088/1674-1056/ad84c2
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Identify information sources with different start times in complex networks based on sparse observers

Yuan-Zhang Deng(邓元璋), Zhao-Long Hu(胡兆龙), Feilong Lin(林飞龙), Chang-Bing Tang(唐长兵), Hui Wang(王晖), and Yi-Zhen Huang(黄宜真)
Chin. Phys. B, 2024, 33 (**11**): 118901. DOI: 10.1088/1674-1056/ad7af4

Detecting the core of a network by the centralities of the nodes

Peijie Ma(马佩杰), Xuezao Ren(任学藻), Junfang Zhu(朱军芳), and Yanqun Jiang(蒋艳群)
Chin. Phys. B, 2024, 33 (**8**): 088903. DOI: 10.1088/1674-1056/ad4cd4

Identifying influential spreaders in complex networks based on density entropy and community structure

Zhan Su(苏湛), Lei Chen(陈磊), Jun Ai(艾均), Yu-Yu Zheng(郑雨语), and Na Bie(别娜)
Chin. Phys. B, 2024, 33 (**5**): 058901. DOI: 10.1088/1674-1056/ad20d6

A multilayer network diffusion-based model for reviewer recommendation

Yiwei Huang(黄羿炜), Shuqi Xu(徐舒琪), Shimin Cai(蔡世民), and Linyuan Lü(吕琳媛)
Chin. Phys. B, 2024, 33 (**3**): 038901. DOI: 10.1088/1674-1056/ad181d

Source localization in signed networks with effective distance

Zhi-Wei Ma(马志伟), Lei Sun(孙蕾), Zhi-Guo Ding(丁智国), Yi-Zhen Huang(黄宜真), and Zhao-Long Hu(胡兆龙)
Chin. Phys. B, 2024, 33 (**2**): 028902. DOI: 10.1088/1674-1056/ad1482

Hyperbolic map unravels eight regions in temperature volatility regionalization of Mainland China

Yuxuan Song(宋雨轩)[1], Changgui Gu(顾长贵)[1,†], Muhua Zheng(郑木华)[2], Aixia Feng(冯爱霞)[3], Yufei Xi(席雨菲)[1], Haiying Wang(王海英)[1], and Huijie Yang(杨会杰)[1]

[1] *School of Management, University of Shanghai for Science and Technology, Shanghai* 200093, *China*
[2] *School of Physics and Electronic Engineering, Jiangsu University, Zhenjiang, Jiangsu* 212013, *China*
[3] *National Meteorological Information Center, China Meteorological Administration, Beijing* 100081, *China*

(Received 3 July 2024; revised manuscript received 5 September 2024; accepted manuscript online 9 October 2024)

Abrupt temperature volatility has detrimental effects on daily activities, macroeconomic growth, and human health. Predicting abrupt temperature volatility and thus diminishing its negative impacts can be achieved by exploring homogeneous regions of temperature volatility and analyzing the driving factors. To investigate the regionalization of temperature volatility in Mainland China, a network constructed by the cosine similarity of temperature volatility series from Mainland China was embedded in hyperbolic space. Subsequently, we partitioned the network on the hyperbolic map using the critical gap method and then found eight regions in all. Ultimately, a network of communities was constructed while the interaction among communities was quantified. This yields a perspective of temperature volatility regionalization that can accurately reflect factors including altitude, climate type, and the geographic location of mountains. Further analysis demonstrates that the regionalization in the hyperbolic map is distinct from provinces and has a realistic basis: communities in southwest China show strong correlations due to the temperature sensitivity to altitude, and communities in northern China show a convergence in the area of Dingxi, Gansu, mainly owing to the strong temperature sensitivity to climate types. As a consequence, node distributions and community divisions in the hyperbolic map can offer new insights into the regionalization of temperature volatility in Mainland China. The results demonstrate the potential of hyperbolic embedding of complex networks in forecasting future node associations in real-world data.

Keywords: complex network, hyperbolic embedding, community detection, regionalization

PACS: 89.75.–k, 89.60.–k, 92.60.Ry, 92.70.Aa **DOI:** 10.1088/1674-1056/ad84c2

1. Introduction

Abrupt temperature fluctuations can negatively impact human activities and health.[1] Researchers across nations, including China, Australia, and the United States, believe in a strong connection between mortality and abrupt changes in temperature.[2–4] Additionally, abrupt temperature fluctuations have the potential to seriously impair people's mental health by causing psychological discomfort, anxiety, and sadness.[5] Sudden changes in day-to-day temperature can lead to higher mortality rates in weather-suicide associations and can increase the incidence of hospitalizations together with the risk of influenza outbreaks.[6] Moreover, the perturbations in temperature patterns impede macroeconomic expansion and produce negative effects, especially on regions characterized by lower latitudes and income levels.[7] It has been demonstrated that daily maximum temperatures have a greater impact on human health than average daily temperatures, thus the term 'volatility' is employed to denote fluctuations in the highest temperature between consecutive days.[8]

The intricate structures of hydro-climatic systems can be analyzed by complex networks, such as the investigation of summer extreme precipitation,[9] drought events in the United States,[10] vapor topology,[11] and the spread of heatwaves.[12] The construction of a climate network (CN) often relies on spatiotemporal datasets that describe the dynamics of the climate system.[13] Within these networks, geographical locations serve as nodes, and edges between nodes are determined by the similarity of corresponding climate time series.[14] This similarity reflects statistical associations between climate subsystems across different geographical locations. Many studies have successfully constructed networks based on the similarity of climate time series, employing measures such as event synchronization (ES),[15–17] Pearson correlation coefficient,[18] cross-correlation (CC),[19] mutual information (MI),[20] and mean distance.[21]

"Rationalization" means a process of recognizing the homogeneous region.[22] Given the regional disparities in climate variables, regionalization provides a straightforward and effective approach to address spatial heterogeneity.[23] It facilitates the identification of homogeneous areas and is commonly utilized in exploring underlying driving factors that may influence spatial variations.[24] Considering homogeneous regions as a whole can also reduce modeling complexity.[25] There have been several methods to achieve regionalization (e.g., multiple regression, principal component analysis, clustering algorithms), but they can only quantify homogeneous regions,

†Corresponding author. E-mail: gu_changgui@163.com
© 2024 Chinese Physical Society and IOP Publishing Ltd

whereas community detection in complex network theory can investigate the interactions between communities as well as achieve regionalization.[26] Community detection of CNs is effective in identifying homogeneous regions and quantifying associations between heterogeneous regions.[27–29]

Although the current studies have all been relatively successful, some aspects have not been considered. Current efforts in constructing CNs tend to concentrate on extreme precipitation events and heatwaves, with scant exploration into the analysis of important day-to-day temperature volatility.[6,8] Most research on climate regionalization worldwide focuses on the regionalization of climate states rather than climate change, which cannot fulfill the requirements of adaptation to climate change.[30] Studies investigating regional heterogeneity in temperature volatility by community detection and exploring associations between heterogeneous regions are even more scarce. Meanwhile, prevailing methodologies for network construction of hydro-climatic systems and distance calculation primarily rely on Euclidean space. However, hyperbolic space, different from Euclidean space, has been shown to reflect the structure of networks more accurately.[31,32] The space suitable for embedding complex networks is also shown to be hyperbolic, rather than Euclidean.[33] The critical gap method (CGM) based on hyperbolic space has also been shown to outperform traditional clustering methods and community detection methods. For example, k-mean clustering suffers from instability of the final clusters, extreme value theory (EVT)-based community detection performs worse in modularity, and Louvain detects a smaller number of communities with less consistency between their partitions and reality than that obtained from the CGM.[34,35] Hyperbolic embedding of temperature volatility networks may help to elucidate the global structure of temperature volatility, revealing hidden similarities between distant regions.

Hyperbolic space is a geometric space characterized by constant negative curvature.[36] Theories in network geometry and topology propose a hyperbolic relationship between nodes in complex networks, whereby connected nodes in hyperbolic space are closer in an underlying hidden metric space.[32,37,38] Beyond the conventional Euclidean space, real-world complex networks can also be embedded into hyperbolic space, yielding hyperbolic maps that provide more accurate interpretations of network connectivity structures.[31,39] Many characteristics observed in most real networks, including small-world properties, heterogeneous degree distributions, clustering coefficients, community, navigability, and self-similarity, can be explained by hyperbolic models.[40] Mercator provides a reliable approach for embedding real-world networks into hidden metric spaces. By combining machine learning techniques with maximum likelihood (ML) methods, Mercator can handle arbitrary degree distributions and replicate the properties of real networks.[41] Currently, hyperbolic embeddings of complex networks perform well in a variety of domains.[34,40,42] Notably, the successful application of analyzing long-term evolution in the international trade system using spatial networks embedded in hyperbolic space demonstrates that hyperbolic embedding performs well in analyzing correlations between subsystems across different geographical locations.[34] That is consistent with the purpose of constructing climate networks based on similarity between climate time series and can provide a more accurate explanation of the network structure, but there are scarcely any studies that embed temperature volatility networks in hyperbolic space.

In this paper, we explore the regionalization of temperature volatility in Mainland China and analyzed the underlying causes. We first construct a temperature volatility similarity network and embedded it in hyperbolic space to obtain additional information. The embedded network is then subjected to community detection based on the critical gap method, and the additional information obtained is used to find communities that are accurately related to the underlying structure. Finally, a community network is constructed in which each node represents a community. The associations between communities are quantified by potential energy and realized energy. A detailed description of the dataset and methodology is provided in Section 2. The quality of hyperbolic embedding is presented in Section 3. The community detection results and analysis of eight regions of Mainland China are shown in Section 4. Conclusions and discussion are given in Section 5.

2. Data description and methods

The process of our work can be shown in Fig. 1. There are three main steps: network construction, network embedding, and regionalization. The methods involved will be described in this section, where the methods for network construction are in Subsection 2.2, network embedding in Subsection 2.3, and regionalization in Subsections 2.4 and 2.5.

2.1. Data description

The data in this work are obtained from the Chinese daily surface air temperature $0.5^\circ \times 0.5^\circ$ grid point dataset (V2.0), with 3830 grid points belonging to Mainland China, which contains latitude 18°N–54°N and longitude 72°E–136°E. The National Meteorological Information Center of China (NMIC) generated and provided this dataset by reprocessing the daily temperature data from national stations (basic, reference, and general stations) from 1961 to the present. We have selected daily surface maximum air temperature data from 1 January 2001 to 8 May 2023 for 3830 grid points in Mainland China from the dataset. The length of the time series for each grid point is 8528 days.

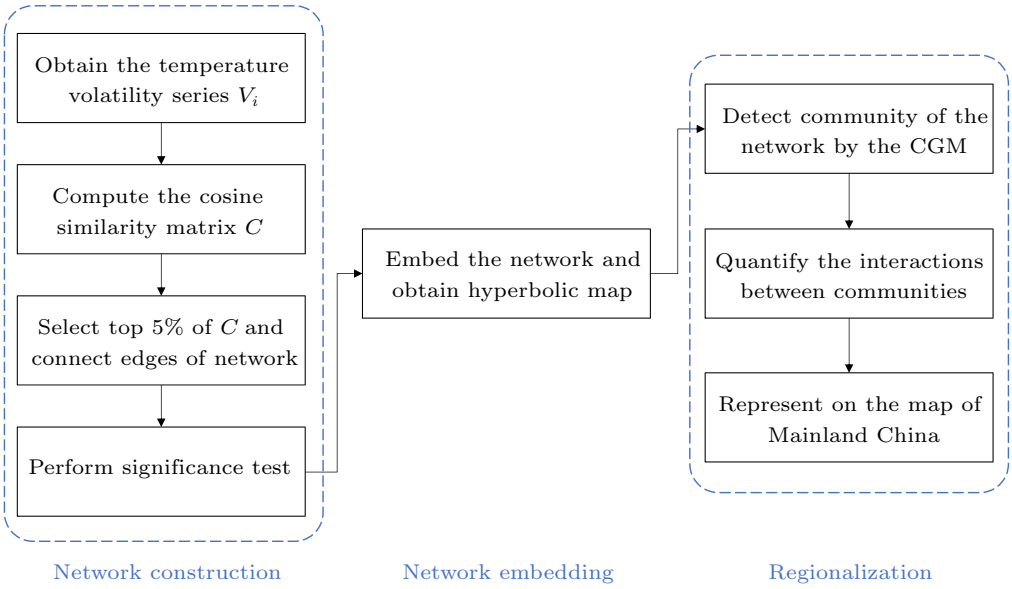


Fig. 1. Process of regionalization of Mainland China. This shows the process of our work, which is divided into three main steps: network construction, network embedding, and regionalization.

2.2. Construction of temperature volatility network

As shown in Fig. 1, the construction of the temperature volatility network can be divided into four steps.

Step 1, obtain the temperature volatility series from the raw data. For any particular grid point i, the daily maximum temperature time series data can be processed according to the following equation to obtain the maximum temperature volatility series V_i of the grid point:[8]

$$V_i(t) = H_i(t+1) - H_i(t), \quad (i = 1, \ldots, 3830,\ t = 1, \ldots, 8327), \tag{1}$$

where $H_i(t)$ is the maximum surface air temperature data of the grid point i on the t^{th} day.

Step 2, compute the cosine similarity matrix. The cosine similarity between temperature volatility series can be computed using

$$C(i,j) = \frac{V_i \cdot V_j}{\|V_i\| \times \|V_j\|}, \quad (i = 1, \ldots 3830,\ j = 1, \ldots 3830), \tag{2}$$

yielding a cosine similarity matrix $\boldsymbol{C}$.

In step 3, we determined the connected edges of the network by setting the thresholds. In numerous prior studies of establishing networks with similarity matrices, only the top 2% or 5% of node pairs with the highest similarity are utilized to establish networks, which ensures the statistical significance of the network by excluding weakly correlated edges.[16,43,44] However, in this study, compared to 5%, 2% is not a satisfactory choice. Given the small-world properties, typically only 5% of the strongest links among all edges are preserved. In this study, the absolute value of the cosine similarity matrix is employed to iterate through varying percentages of concatenated edges, ranging from 1% to 10%. Subsequently, the largest clusters are retained, and the average path length is computed. Differential calculations for the average path length are conducted, and the result shows a trend towards stabilization in average path length when the percentage of connected edges reaches 5%. Subsequently, considering the quality of hyperbolic embedding, only the top 5% of the absolute values of cosine similarity are kept in network construction in this work. The threshold C_θ is determined according to

$$C_\theta = \theta^{\text{th}} \text{ of } \{C(i,j) \,|\, i = 1, \ldots, 3830, j = i+1, \ldots, 3830\}, \tag{3}$$

with $\theta^{\text{th}} = 95^{\text{th}}$. The adjacency matrix is obtained by establishing connections between pairs of nodes whose absolute cosine similarity is larger than the threshold defined by

$$A_{ij} = \begin{cases} 1, & C(i,j) \geq C_\theta, \\ 0, & C(i,j) < C_\theta, \end{cases} \quad (i, j = 1, \ldots, 3830). \tag{4}$$

In step 4, Monte Carlo simulations and randomization were used to calculate p-values and test for significance, and the results show that all the edges are statistically significant.

2.3. Embedding in hyperbolic space

The $\mathbb{S}^1$ model has been validated to reflect the topology of realistic networks in one-dimensional circle.[31,41] Each node i in the model is characterized by two variables, denoted as (κ_i, θ_i), representing the hidden degree and angular position of node i. The hidden degree κ_i exhibits a positive correlation with the degree of node i, indicating its popularity.[32] The angular position θ_i serves as a descriptor of similarity of node i, where a smaller angular distance between nodes i and j, denoted as $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$, implies a greater resemblance. In the $\mathbb{S}^1$ model, N points are uniformly distributed on a circle, with a radius of $R = N/2\pi$ to maintain the density of 1. The theoretical analysis indicates that the probability of connectivity between nodes i and j, p_{ij}, should increase with

$\kappa_i \cdot \kappa_j$ and decrease with $\Delta\theta_{ij}$. The connection probability p_{ij} can be expressed as

$$p_{ij} = \frac{1}{1+\left(\frac{d_{ij}}{\mu\kappa_i\kappa_j}\right)^{\beta}}, \tag{5}$$

where $d_{ij} = R\Delta\theta_{ij}$. μ and β are global parameters that respectively adjust the average degree and clustering coefficient of the network.

The $\mathbb{H}^2$ model is an isomorphic purely geometric formulation of $\mathbb{S}^1$ model and is a representation on two-dimensional circle of the $\mathbb{S}^1$ model.[32] In the $\mathbb{H}^2$ model, the angular position θ_i of each node i is the same as it is in the $\mathbb{S}^1$ model, while the radial coordinate r_i is transformed from the hidden degree κ_i in the $\mathbb{S}^1$ model using

$$r_i = R_{\mathbb{H}^2} - 2\ln\frac{\kappa_i}{\kappa_0}, \tag{6}$$

where $R_{\mathbb{H}^2} = 2\ln\frac{N}{\pi\mu\kappa_0^2}$ denotes the radius of the hyperbolic circle, and $\kappa_0 = \min(\{\kappa\})$ represents the minimum hidden degree. According to Eq. (6), the greater the hidden degree κ_i, the greater the node importance and the closer the node is to the center of the hyperbolic circle.

In the $\mathbb{H}^2$ model, the connection probability p_{ij} between nodes i and j can be expressed as

$$p_{ij} = \frac{1}{1+\mathrm{e}^{\frac{\beta}{2}(H_{ij}-R_{\mathbb{H}^2})}}, \tag{7}$$

where H_{ij} is the hyperbolic distance, includes the popularity and similarity information of nodes i and j, and can be expressed as[31,32]

$$H_{ij} = \cosh^{-1}(\cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos\Delta\theta_{ij}). \tag{8}$$

The optimal embedding is obtained by finding the hidden degree and angular coordinates of each node, $\{\kappa_i\}$ and $\{\theta_i\}$, maximizing the likelihood of reproducing the real network structure, L, which can be expressed as

$$L = \prod_{i<j} p_{ij}^{A_{ij}}(1-p_{ij})^{(1-A_{ij})}, \tag{9}$$

where p_{ij} is the node connection probability of the $\mathbb{S}^1$ model in Eq. (5) and A_{ij} is the adjacency matrix of the network expressed in Eq. (4). Embedding by Mercator yields inferred hidden degree κ_i, angular coordinates θ_i, radial coordinates r_i of each node, and global parameters β and μ.

2.4. Critical gap method

Community detection in complex networks is an important branch of network topology, aiming to detect clusters characterized by strong internal connections and sparse inter-cluster connections. For networks embedded in hyperbolic spaces, the critical gap method (CGM) is an effective community detection method.[35] A critical gap $\Delta\theta_{\max}$ is defined by the CGM. Two consecutive nodes i and j belong to different communities if the angular distance between them is beyond the critical gap ($\Delta\theta_{ij} > \Delta\theta_{\max}$), otherwise, they belong to the same community. We modified $\Delta\theta_{\max}$ to different values then computed the communities and modularity Q separately according to

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\cdot\delta(C_i, C_j), \tag{10}$$

where $k_i = \sum_j A_{ij}$ is the degree of node i, and k_j can be obtained. The variable m means the number of edges. The variables C_i and C_j correspond to the communities i and j, while δ is the delta-function. Finally, we chose $\Delta\theta_{\max}$ that maximizes the modularity Q.

2.5. Network of communities

The CGM is used for community detection. Then a community network can be constructed, with each node representing a community. In this network, the size of the nodes is proportional to the number of nodes contained in the community on a logarithmic scale, and the geometric centers of the nodes in each community serve as the node location.[35] The attraction potential energy between nodes can be defined in terms of the logarithm of the probability that no link exists between nodes, which can be expressed by

$$E(i,j) = \ln(1-p_{ij}), \tag{11}$$

where p_{ij} is the connection probability between nodes i and j. The theoretical mutual attraction energy between two communities can be defined as the potential energy between two communities. By summing up the attraction energies of all node pairs between the two communities, the potential energy between the two communities can be obtained as

$$E(C_i, C_j) = \ln\left[\prod_{i\in C_i, j\in C_j}(1-p_{ij})\right] = \sum_{i\in C_i, j\in C_j}\ln(1-p_{ij}). \tag{12}$$

Considering the structure of the real network, the realized energy between two communities can be obtained. This means the realized attraction energy between two communities in the real network, which can be expressed as

$$R(C_i, C_j) = \sum_{i\in C_i, j\in C_j} A_{ij}\ln(1-p_{ij}). \tag{13}$$

3. Network embedding of temperature volatility of Mainland China

After constructing the temperature volatility network based on the grid points' similarity, embedding it into the hyperbolic space with Mercator yields additional information

such as the hidden degree κ_i, radial coordinates r_i, and angular position θ_i of each node, as well as the global parameters $\beta = 4.6318$ and $\mu = 0.0024$. To validate the embeddings, 1000 synthetic networks have been generated according to Eq. (5) as an ensemble with the global parameters β and μ, as well as the inferred coordinates $\{\kappa_i,\ \theta_i\}$.

The embedding quality can be regarded from aspects of topological properties and local properties of each node (in Fig. 2). The comparisons of the topological properties of the original temperature volatility network and synthetic ensembles include the aspect of complementary cumulative degree distribution $P_c(k)$, the degree-dependent clustering coefficient $\bar{c}(k)$, and the average degree of nearest neighbors $\bar{k}_{nn,n}(k)$, respectively (see Figs. 2(a)–2(c)). The black hollow dots indicate the real temperature volatility network. The blue line indicates the empirical temperature volatility network, which is the average of the ensemble of 1000 synthetic networks. The red dashed line indicates the 2σ confidence interval around the average. It is observed that the vast majority of points in the original network are distributed within the confidence intervals.

Besides, we also investigated the local properties of each node in the real temperature volatility network and synthetic ensemble. The local properties of each node include degree, number of triangles connected to each node, and sum of degrees of neighbors (see Figs. 2(d)–2(f)). The dashed line shows the case where the true and inferred values are equal, and the nodes are distributed around the dashed line. The error bars show the 2σ confidence interval. Subsequently, we validated the embeddings by some statistical tests, such as the Pearson correlation coefficient ρ, the χ^2 test normalized by the number of nodes N, and the score ζ that quantifies the proportion of nodes with values outside the 2σ confidence interval around the expected value. It can be shown that the embeddings perform well and the $\mathbb{S}^1$ model can accurately replicate the topological properties of the temperature volatility network.

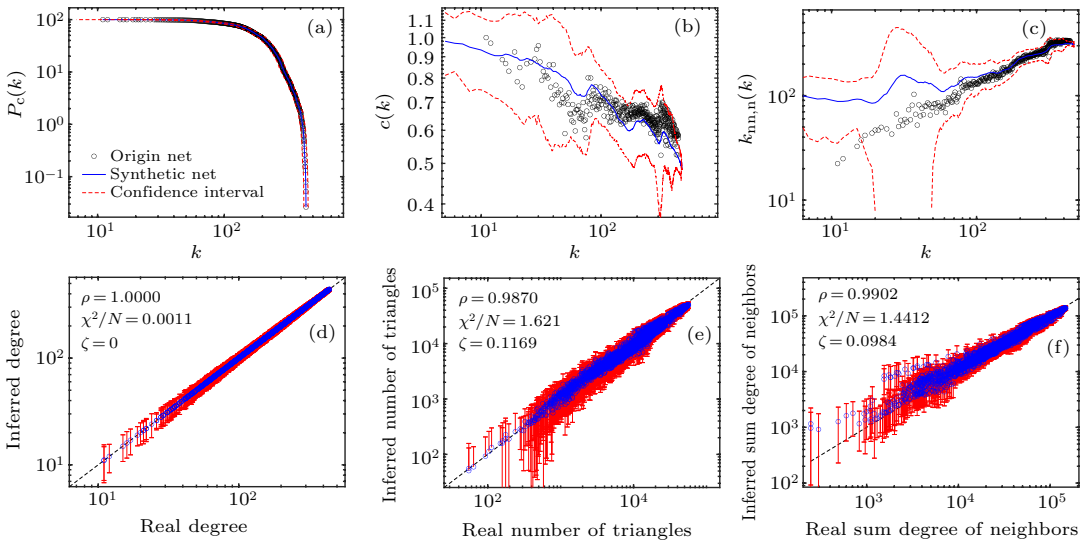


Fig. 2. Embedding quality. (a)–(c) Network properties of origin compared to the model predictions, (a) complementary cumulative degree distribution, (b) degree-dependent clustering coefficient, (c) average degree of nearest neighbors. The red dashed line marks the 2σ confidence interval around the average. (d)–(f) Comparison of the local properties of each node in the synthetic ensemble with the real network, (d) degrees, (e) number of triangles, and (h) sum of degrees of neighbors. Error bars show the 2σ confidence interval around the average. Statistical tests with ρ, the χ^2 test and the score ζ are reported in each subfigure.

4. Community detection of the network of temperature volatility of Mainland China

4.1. Independence between community and province

Because the embedding quality has been verified as satisfactory, regionalization of the temperature volatility network in Mainland China can be implemented by community detection. Community detection by the CGM reveals eight communities, as depicted in Fig. 3(a), with a corresponding modularity $Q = 0.6291$. According to the definition of the angular position θ in the $\mathbb{S}^1$ and $\mathbb{H}^2$ models, the more similar the two nodes are, the smaller the angular difference $\Delta\theta_{ij}$ is between these two nodes. Consequently, nodes within the same community tend to cluster within a specific angular region. The results of community detection for the temperature volatility network can be visually presented in Fig. 3, where the same color represents nodes within the same community. Starting from $\theta = 0°$, the communities are sequentially labeled from 1 to 8 in counterclockwise order. Notably, communities 4, 6, and 7 are larger communities containing more nodes. Communities 1, 5, and 8 are smaller communities containing fewer nodes. The results support the theory that nodes in the same community have close angular coordinates θ and cluster in particular angular regions in hyperbolic space. It indicates that the CGM performs well in detecting communities of the temperature volatility network.

Subsequently, we visualized a hyperbolic map of the province to investigate if there is a correlation between community and province. In Fig. 3(b), the coordinates of the nodes $\{\theta_i, r_i\}$ are the same as those in Fig. 3(a), and the nodes in the same province are represented by the same color, with a total of 31 provinces. As can be observed, there is a chaotic pattern among the color distribution, and the nodes in the same province are not all close to each other. This implies that in the temperature volatility network, the community is not related to the province, which means that pairs of nodes with high similarity and more likely to be connected are not necessarily distributed in the same province.

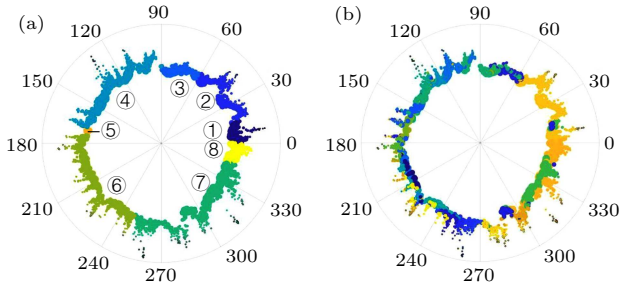


Fig. 3. Visualization of communities and provinces. (a) Community detection, nodes in the same community are represented by the same color. (b) Province visualization, nodes in the same province are represented by the same color. The coordinates of the nodes in both subgraphs are the same. It shows the independence of the community division from the province division.

4.2. Quantification of interactions between communities with potential and realized energy

Following the previously detailed methods in Subsection 2.5, a network is constructed from communities we obtained from the result of community detection. Then, interactions between communities are quantified by calculating the potential and realized energy between them according to Eqs. (12) and (13).

The potential energy and realized energy between communities of the temperature volatility network are demonstrated by the network of community in Fig. 4. Each node of the community network represents a community in the temperature volatility network and the node color corresponds to the community color depicted in Fig. 3(a). The node size is proportional to the size of the community on a logarithmic scale, and the nodes are located at the geometric centers of the nodes of the community. The width of the edge represents the energy exchanged between the two communities.

It can be observed that the interactions between community 1 and community 8, and between community 7 and community 8 are strong correlations. Also, the realized influences between community 1 and community 2, community 3 and community 4, community 4 and community 5, and community 5 and community 6 are weak, even though they are close in the hyperbolic map. In addition, there exist energy exchange edges that are similar in width in the potential energy network but are only partially represented in the realized network. For example, the edges between community 5 and community 7 as well as the edges between community 5 and community 8 are of about the same width in the potential energy network, but in the realized energy network, only the edge between community 5 and community 7 exists. This suggests a potential for hyperbolic embedding of complex networks in predicting future node associations for real-world data.

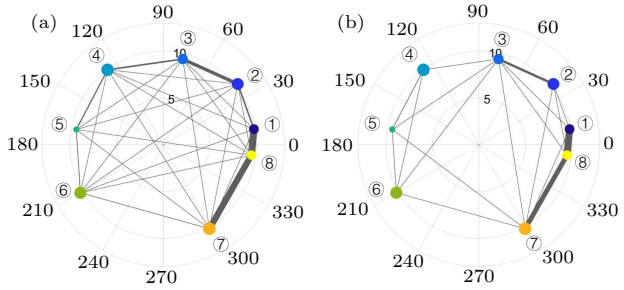


Fig. 4. Network of communities. (a) Potential energy of community network. (b) Realized energy of community network. Nodes correspond to the communities. The size of the nodes is proportional to the size of the community on the logarithmic scale. The width of the edges between two nodes means the interaction between the communities.

4.3. Regionalization of Mainland China

We represented the corresponding grid points of the communities in the China map to visualize the communities better and identify possible driving factors. The grid point's color matches that of its affiliated community in Fig. 3(a), the numbers on the map correspond to the community numbers, and the bottom blue curves represent the provincial boundaries of China. Notably, the regionalization results of climate volatility in Mainland China obtained by community detection for hyperbolic embedded networks are based on a combination of factors such as altitude, climate, and location of mountains.

As shown in Fig. 4, two strong associations exceed all other associations between communities: one between community 7 and community 8, and the other between community 1 and community 8. Community 1 contains grid points located in Xinjiang Uygur Autonomous Region (XUAR), Qinghai Province (QPR), Tibet Autonomous Region (TAR), and Gansu Province (GSM), and has grid points count of 107, 41, 34, and 11, respectively. Community 8 contains grid points located in the TAR, XUAR, and Qinghai Province, counting 163, 4, and 3. Community 7 contains grid points located in the TAR, Qinghai Province, Yunnan Province, Sichuan Province, Guangxi Zhuang Autonomous Region, Guangdong Province, Gansu Province, XUAR, Fujian Province, Hainan Province, and Jiangxi Province, counting 259, 227, 132, 123, 69, 59, 33, 31, 14, 10, 9, and 5, respectively. The strong correlation can be attributed to the greater sensitivity of temperature to altitude in southern China compared to northern China, and the greater independence of the climate in southwestern China due to the

influence of the Tibetan Plateau.[45] Altitude has a strong influence on temperature volatility in southern China, but it is also an influence factor in the north. The regions corresponding to community 1 and community 2 are geographically adjacent, but they have a large difference in altitude, and their boundary coincides with one of the boundaries of the three steps of China's terrain. The difference in altitude precisely explains why these two communities are located close in the hyperbolic map but exchange very little energy. Similarly, the regions corresponding to communities 2 and 3 are within the same steps of China's terrain, so the energy exchanged between them is more than that between communities 1 and 2.

The boundaries of communities 3, 4, 5, and 6 converge around Dingxi, Gansu in Fig. 5. There are a variety of climates distributed near the convergence point, including the plateau climate zone, mesothermal zone, southern temperate zone, northern subtropical zone, and central subtropical zone. In terms of China's eco-geographical subdivisions, there are plateau temperate humid/semi-humid regions, warm temperate humid regions, and meso-temperate humid regions distributed near the convergence point. The community divisions in the vicinity of this convergence point exhibit a strong correlation with the climate type, which is consistent with Zhai and Ren's theory that temperature is more sensitive to climate generally at high latitudes (particularly in regions north of 35°N).[46] This convergence point is also located in the border area between the first and second terrains of the three steps of China's terrain, as well as on the boundary of the Loess Plateau and the Inner Mongolian Plateau.



Fig. 5. Community representation in Chinese map. Different colors correspond to different communities, representing eight regions in the temperature volatility regionalization of Mainland China. Grid points without colors (including Taiwan, China) are lacking in data, so they are not taken into account.

In addition, the boundary between communities 1 and 2 passes through the Kunlun Mountains, that between communities 3 and 7 passes through the Qilian Mountains, and that between communities 6 and 7 passes through the Nanling and the Hengduan Mountains.

Although the result of the regionalization based on hyperbolic embedding can be explained by a combination of altitude, climate type, and other realistic bases, it still differs from all existing regional divisions in Mainland China, and is not obtained by a simple combination of several existing divisions. For example, it is not common for the grid points in the rightmost part of community 7 and the lower part of community 6 to be divided into two communities, and for the grid points contained in community 5 to be divided into a single community. This is a new insight with more system thinking obtained through hyperbolic embedding, which has significant implications for the development of climate adaptation strategies, resource allocation, and policy responses to climate change.

In summary, community detection in the hyperbolic map of temperature volatility can lead to a new perspective of comprehensive regionalization with a realistic basis. Besides, community detection in hyperbolic maps is universal in environmental science research and can be extended to other types of ecological data, such as precipitation, providing a new perspective for identifying regions with similar precipitation patterns.

5. Discussion and conclusion

Abrupt temperature changes have negative impacts on human health and activities, macroeconomic growth, and other areas. Despite its importance, daily temperature volatility has received little attention compared to mean annual temperatures, long-term temperature changes, and even daily extreme temperatures. Delineation of heterogeneous regions and analysis of driving factors facilitate the prediction of temperature volatility, thereby mitigating negative impacts. Community detection in complex networks is an effective tool for regionalization, which not only explores heterogeneous regions but also investigates the associations between them. As an effective method to investigate the macroscopic properties of complex systems, complex networks have been sparsely studied for temperature volatility and are all based on Euclidean space. However, hyperbolic embedding of complex networks has been shown to more accurately reflect the underlying metric space of the topology.

To explore the temperature volatility regionalization of Mainland China, we first established a similarity network with the temperature volatility series of 3830 grid points in Mainland China. Then, hyperbolic embedding was applied according to the connectivity structure of the network. We validated the quality of the hyperbolic embedding from both the topology of the network and the local properties of each node, and the results show that the quality of embedding of the temperature volatility network is satisfactory. Subsequently, community detection was performed on the hyperbolic map using the CGM. Eight communities were obtained with modularity $Q = 0.6291$. By comparing with the province visualization map, it can be observed that communities are independent of

provinces in the temperature volatility network. A network of community is established by treating each community as a node, and then the interactions between the communities are quantified by potential energy and realized energy. Ultimately, the regionalization of temperature volatility can be obtained, which can be explained by a combination of factors such as elevation, climate type, and location of mountains. The result shows that hyperbolic embedding of complex networks has the potential to predict future node associations in real-world data and can provide new insight into the regionalization of temperature volatility in Mainland China.

We also tried the k-means clustering and the Louvain method. The results show that both methods are inferior to the CGM method (in Appendix A). The modularity $Q = 0.5827$ obtained by the k-means clustering method is lower than modularity $Q = 0.6291$ of the CGM method, and it suffers from the instability of the final clusters. The Louvain method obtains a higher modularity $Q = 0.8327$, but only divides 2 communities, ignoring much local information. The comparison results are consistent with the conclusion researchers have mentioned before.[34,35]

This work is replicable. The approach to regionalizing Mainland China in our work essentially requires embedding complex networks in the hyperbolic space and then detecting communities on the hyperbolic map, which has been validated to be effective when dealing with spatial networks. Within most of the CNs, geographical locations serve as nodes, and edges between nodes are determined by the similarity of corresponding climate time series.[14] This means that theoretically most of the regionalized studies conducted through climate networks could implement our proposed methodology. However, limitations such as embedding quality need to be considered when judging whether it can be used specifically.

This study still has some limitations. The similarity measure we adopted is cosine similarity, and different choices may lead to different results. Cosine similarity was adopted because it has the best embedding quality among cosine similarity, Spearman rank correlation coefficient, and Pearson correlation coefficient, and embedding quality is taken as a prerequisite for the subsequent analyses in this paper. In the future, we can take more methods of network construction for analysis and discussion. In addition, the time scale selected in this study may have neglected the short-term trend, so a more detailed time scale can be considered for analysis in future studies.

Appendix A: Comparison of different community detection methods

We tried the k-means clustering and the Louvain method. The following figure shows the community detection results using these two methods. The Louvain method obtains a higher modularity $Q = 0.8327$, but only divides 2 communities, ignoring much local information. The k-means clustering method obtains the modularity $Q = 0.5827$, which is lower than modularity $Q = 0.6291$ of the CGM method. Also, it suffers from the instability of the final clusters. In addition, we can see that the community represented by purple in the upper right corner has been separated by other communities, which also indicates that this method does not perform well. It is worth noting that when using k-means clustering, we set the final number of communities $k = 8$, which is consistent with the community detection results obtained by the CGM. It shows that both methods are inferior to the CGM method.

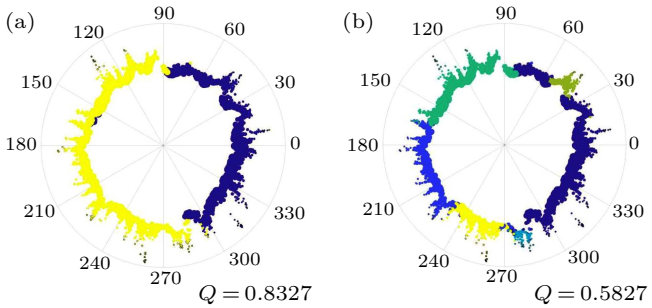


Fig. A1. Community detection results of (a) Louvain and (b) k-means. Q is the modularity value.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 12275179, 12005079, and 41975100), the Shanghai Natural Science Foundation of China (Grant No. 21ZR1443900), Natural Science Foundation of Jiangsu Province (Grant No. BK20220511), the funding for Scientific Research Startup of Jiangsu University (Grant No. 4111710001), and the Joint Research Project for Meteorological Capacity Improvement (Grant No. 22NLTSZ004).

References

[1] Donadelli M, Jüppner M, Paradiso A and Schlag C 2019 *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series* **05/WP/2019**
[2] Zhan Z, Zhao Y, Pang S, Zhong X, Wu C and Ding Z 2017 *Science of The Total Environment* **584–585** 1152
[3] Guo Y, Barnett AG, Yu W, Pan X, Ye X, Huang C and Tong S 2011 *PLoS One* **6** e16511
[4] Cheng J, Zhu R, Xu Z, Xu X, Wang X, Li K and Su H 2014 *International Journal of Public Health* **59** 923
[5] Trombley J, Chalupka S and Anderko L 2017 *AJN, American Journal of Nursing* **117** 44
[6] Pińskwar I, Choryński A and Graczyk D 2023 *International Journal of Biometeorology* **68** 317
[7] Kotz M, Wenz L, Stechemesser A, Kalkuhl M and Levermann A 2021 *Nature Climate Change* **11** 319
[8] Ombadi M and Risser M D 2022 *Weather and Climate Extremes* **38** 100515
[9] Mondal S, Mishra A K and Leung L R 2020 *Geophysical Research Letters* **47** e2020GL088185
[10] Konapala G and Mishra A K 2017 *Journal of Hydrology* **555** 600
[11] Feng A X, *et al.* 2022 *Chin. Phys. B* **31** 049201

[12] Mondal S and Mishra A K 2021 *Geophysical Research Letters* **48** e2020GL090411
[13] Tsonis A A, Swanson K L and Roebber P J 2006 *Bulletin of the American Meteorological Society* **87** 585
[14] Tsonis A A and Roebber P J 2004 *Physica A* **333** 497
[15] Gao M, Wang Z, Ning J and Wang Y 2023 *Chaos, Solitons & Fractals* **173** 113650
[16] Boers N, Bookhagen B, Barbosa H M, Marwan N, Kurths J and Marengo J A 2014 *Nat. Commun.* **5** 5199
[17] He S H, Wang J P, Li X M and Zhang Y X 2014 *Chin. Phys. B* **23** 059202
[18] Li K, Wang M and Liu K 2021 *International Journal of Climatology* **42** 4445
[19] Meng J, Fan J, Ashkenazy Y, Bunde A and Havlin S 2018 *New J. Phys.* **20** 043036
[20] Deza JI, Barreiro M and Masoller C 2015 *Chaos* **25** 033105
[21] Tsonis AA 2007 *Geophysical Research Letters* **34**
[22] Mallick J, Talukdar S, Almesfer M K, *et al.* 2022 *Environmental Science and Pollution Research* **29** 25112
[23] Yang T, Xu C Y, Shao Q X and Chen X 2009 *Stochastic Environmental Research and Risk Assessment* **24** 165
[24] El Kenawy A, López-Moreno JI and Vicente-Serrano S M 2013 *Theoretical and Applied Climatology* **113** 387
[25] Yu Y, Shao Q and Lin Z 2018 *Journal of Hydrology* **564** 149
[26] Agarwal A, Marwan N, Maheswaran R, Merz B and Kurths J 2018 *Journal of Hydrology* **563** 802
[27] Agarwal A, Marwan N, Ozturk U and Maheswaran R 2019 *Water Resources and Environmental Engineering II* **17** 179
[28] Gao C, Liu L, Zhang S, Xu Y P, Wang X and Tang X 2023 *Atmospheric Research* **292** 106874
[29] Wang G L and Tsonis A A 2009 *Chin. Phys. B* **18** 5149
[30] Shi P, Sun S, Wang M, Li N, Wang J, Jin Y, Gu X and Yin W 2014 *Science China Earth Sciences* **57** 2676
[31] Krioukov D, Papadopoulos F, Kitsak M, Vahdat A and Boguñá M 2010 *Phys. Rev. E* **82** 036106
[32] Krioukov D, Papadopoulos F, Vahdat A and Boguñá M 2009 *Phys. Rev. E* **80** 035101
[33] Chamberlain B P, Clough J and Deisenroth M P 2017 arXiv **cs.LG** 1705.10359
[34] García-Pérez G, Boguñá M, Allard A and Serrano M A 2016 *Scientific Reports* **6** 33441
[35] Bruno M, Sousa Sandro F, Gursoy F, Serafino M, Vianello F V, Vranic A and Boguna M 2019 arXiv:1906.09082
[36] Wang X, Zhang Y and Shi C 2019 *Proceedings of the AAAI Conference on Artificial Intelligence* **33** 5337
[37] Boguná M, Papadopoulos F and Dmitri K 2010 *Nat. Commun.* **1** 62
[38] Papadopoulos F, Kitsak M, Serrano M Á, Marián B and Dmitri K 2012 *Nature* **489** 537
[39] Serrano M A, Krioukov D and Boguñá M 2008 *Phys. Rev. Lett.* **100** 078701
[40] Wu J, Zhang X, Gu C, Bi H, Xu K and Zheng M 2023 *Nonlinear Dyn.* **111** 18493
[41] García-Pérez G, Allard A, Serrano M A and Boguñá M 2019 *New J. Phys.* **21** 123033
[42] Ortiz E and Serrano M A 2022 *Chaos, Solitons & Fractals* **165** 112847
[43] Yang W, Yang H and Yang D 2019 *Journal of Hydrology* **576** 726
[44] Sun A Y, Xia Y, Caldwell T G and Hao Z 2018 *Advances in Water Resources* **112** 203
[45] Dong D, Huang G, Qu X, Tao W and Fan G 2014 *Theoretical and Applied Climatology* **122** 285
[46] Zhai P and Ren F 1997 *Acta Meteorologica Sinica* **4** 418