Physica A 481 (2017) 224–234

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Network similarity and statistical analysis of earthquake seismic data

Krishanu Deyasi [a], Abhijit Chakraborty [b,*], Anirban Banerjee [a,c]

[a] Department of Mathematics and Statistics, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, India
[b] Graduate School of Simulation Studies, University of Hyogo, Kobe 650-0047, Japan
[c] Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, India

HIGHLIGHTS

- Complex network and statistical analysis of distinct earthquake catalogs.
- Peakedness of graph spectra of different earthquake regions varies significantly.
- Hierarchical clustering of earthquake networks is explained by tectonic plates.
- We have calculated the conditional probabilities for forthcoming earthquake events.

ARTICLE INFO

Article history:
Received 25 November 2016
Received in revised form 15 March 2017
Available online 19 April 2017

Keywords:
Earthquake network
Seismic catalogs
Hierarchical clustering
Jensen–Shannon divergence
Graph spectra
Conditional probability

ABSTRACT

We study the structural similarity of earthquake networks constructed from seismic catalogs of different geographical regions. A hierarchical clustering of underlying undirected earthquake networks is shown using Jensen–Shannon divergence in graph spectra. The directed nature of links indicates that each earthquake network is strongly connected, which motivates us to study the directed version statistically. Our statistical analysis of each earthquake region identifies the hub regions. We calculate the conditional probability of the forthcoming occurrences of earthquakes in each region. The conditional probability of each event has been compared with their stationary distribution.

1. Introduction

Earthquakes are one of the most devastating natural calamities that can shatter human civilization in large extent. Naturally, scientists are investigating this phenomena in great detail. The most robust empirically established facts in the phenomenology of earthquakes are the Gutenberg–Richter law [1] and the Omori law [2]. While the Gutenberg–Richter law expresses the relationship between frequency and magnitude of tremors in a region, the Omori law describes the temporal rate of decay of aftershocks. Different models have been proposed to study the phenomena of earthquakes. One well-known example is the Burridge–Knopoff model that demonstrates the statistical properties of earthquakes using friction on a fault surface as a stickslip process [3]. It is considered that large faults in the Earth's crust are formed due to the action of plate tectonic forces.

Plate tectonic theory, which is based on Alfred Wegener's continental drift theory [4], is considered to be the fundamental theoretical framework in the field of Earth Science and plays the pivotal role to explain tremors. This theory states that Earth's

* Corresponding author.
E-mail address: abhiphyiitg@gmail.com (A. Chakraborty).

http://dx.doi.org/10.1016/j.physa.2017.04.050
0378-4371/

outer shell, the lithosphere is divided into several rigid pieces, called plates. There are mainly eight major plates: African, Antarctic, Eurasian, North American, South American, Pacific, and Indo-Australian. When a pair of plates move with respect to each other they do not deform internally rather they deform along their edges and create earthquakes and volcanoes along the edges of the plates.

Recently a complex network approach has been applied to observe the universal features of the earthquake phenomena from seismic catalogs. Based on the work of Bak et al. [5], Baiesi and Paczuski, using a correlation metric between a pair of events, have constructed a network in which tremors are the nodes and a pair of nodes are linked if the correlation between them is higher than a certain threshold [6,7]. It is shown that the earthquake network exhibits a scale-free nature with highly heterogeneous degree distribution characterized by a power law. In another study, Abe and Suzuki have constructed the network using a grid, covering the entire earthquake events over a region [8–15]. A cell of the grid is considered to be a node if at least one epicenter occurs within the cell, and the cell size is a tunable parameter for the model. A pair of nodes is connected by a link if two successive events occur on those two nodes. Using this network model, they found various robust features of earthquake networks. They have shown that earthquake networks exhibit a power-law degree distribution [8] small-world phenomena [9], assortativity [10] and scaling in local clustering [15]. Subsequently, different structural properties of weighted earthquake network have also been studied extensively in [16].

Although universal features of earthquake networks of different regions have been studied extensively [6,8–15], but no study has been devoted to capture the similarity and dissimilarity between earthquake networks of different regions. A study of hierarchical clustering [17] of earthquake networks will be very useful in this regard. Hierarchical clustering is usually represented by a dendrogram that pictorially shows the similar entities are clustered together in groups. Quantitatively, the similarity between networks could be measured based on different properties, viz., Euclidean distance, structural properties and dynamical behavior.

Primarily, earthquake networks are directed sequence of consecutive events, so analyzing the directed structure of earthquake networks can give more information than the underlying structure of earthquake networks [18]. Not only the directed structure of earthquake network, studying the directed sequence of earthquake events can also give us important insight, such as, prediction of consecutive earthquakes.

In this paper, we divide our study of earthquake data into two parts. In the first part, we follow the method of Abe and Suzuki to construct earthquake networks as it is connected to a universal law [19]. The similarity between each pair of spectral probability functions of eleven earthquake networks is measured using a probabilistic measure viz., Jensen–Shannon divergence. Using the similarity distances, the hierarchical clustering between earthquake networks is shown as a dendrogram. The hierarchical clustering of earthquake networks reveals the similar or dissimilar nature of different earthquake regions based on their positions on different tectonic plates. We also measure the frequency of earthquake events on a node and identify the earthquake prone locations for different regions. In the second part, we find the pair of regions where consecutive earthquake events have occurred with relatively higher frequency. For this, we calculate the conditional probability between two successive events. We also compare the conditional probability of two consecutive earthquakes with their stationary conditional probabilities to predict the occurrence of an earthquake event at a node consecutively after one occurs at a certain node.

2. Data

We have analyzed eleven distinct earthquake catalogs for different parts of the world, namely the Southern California Earthquake Data Center catalog (SC), Northern California Earthquake Catalog (NC), Japan University Network Earthquake catalog (JAP), Canada's National Earthquake Database catalog (CAN), International Institute of Earthquake Engineering and Seismology—Iran catalog (IRAN), Institute of Geodynamics—Greece catalog (GR), Center for Earthquake Research and Information—New Madrid catalog (NM) and British Geological Survey Earthquake Database around the British Isles catalog (BI), Geoscience Australia catalog(AUS), Swiss Seismological Service catalog (SZ) and GeoNet—the official source of geological hazard information for New Zealand catalog (NZ). Each catalog contains the geographical positions of the epicenters, specified by their latitudes (θ) and longitudes (ϕ) and the exact occurrence times of tremors. The positions of epicenters of the different earthquake regions are shown in Fig. 1. All the related parameters are mentioned in Table 1. The minimum and maximum values of the latitude–longitude coordinates, i.e., $(\theta_{\min}, \theta_{\max})$ and $(\phi_{\min}, \phi_{\max})$ characterize the extent of a earthquake region. The entire earthquake region is divided into a large number of square cells, following the approach in [8]. In this method, the cell size L is the parameter of the model. We use the definition [14] of the dimensionless cell size parameter $\ell = L/(L_{lat}L_{lon})^{1/2}$. Here, the total extent along the north–south and the east–west directions of the entire earthquake region are L_{lat} and L_{lon}, respectively. The North–South distance between (θ_i, ϕ_i) and $(\theta_{\min}, \phi_{\min})$ is $d_{NS} = R(\theta_i - \theta_{\min})$ and the East–West distance is $d_{EW} = R(\phi_i - \phi_{\min})\cos\theta_{av}$, where the radius of the earth is $R = 6370$ km and θ_{av} is $(\theta_{\min} + \theta_{\max})/2$.

3. Construction of the earthquake network

We use the method of Abe and Suzuki [8] to construct an earthquake network. Although the Abe–Suzuki method is not limited to two dimensional cells, but also to three dimensional cells. We consider a two dimension version of the model for our study. An entire earthquake region has been discretized into a two dimensional rectangular grid with the dimensionless

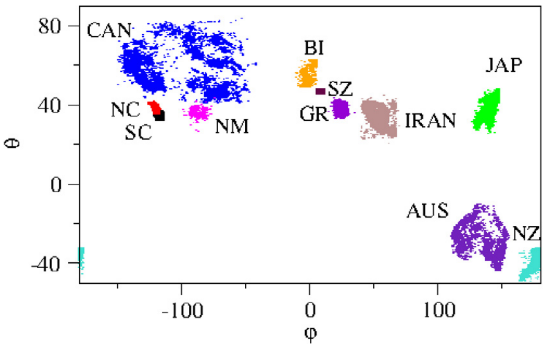

Fig. 1. (Color online) Data points representing the positions of the epicenters of the earthquakes events. Different regions are indicated with different colors.

Table 1
Specification of the parameters of different earthquake catalogs. The minimum and maximum values of the latitude–longitude coordinates, i.e., $(\theta_{\min}, \theta_{\max})$ and $(\phi_{\min}, \phi_{\max})$ characterize the extent of a earthquake region. n is the number of earthquake in the catalogs. All angles are measured in degree and $(L_{lat} L_{lon})^{1/2}$ are measured in kilometer where L_{lat} and L_{lon} is the total extent along the north–south and the east–west directions, respectively.

Region	Period	$\theta_{\min}$	$\theta_{\max}$	$\phi_{\min}$	$\phi_{\max}$	n	$(L_{lat} L_{lon})^{1/2}$
SC	1973–2011	32	37	−122	−114	572 601	638.33
NC	1984–1989	35.301	42.223	−126.145	−116.643	99 998	796.21
JAP	1985–1998	25.73	47.964	126.43	148.0	200 910	2178.01
CAN	1985–2015	41.01	83.54	−149.89	−41.83	83 404	5140.65
IRAN	1900–2014	20.650	44.490	40.000	69.590	21 057	2710.74
GR	1970–2012	33.100	43.680	14.810	35.030	139 126	1439.64
NM	1974–2014	26.716	43.780	−98.880	−74.540	10 425	2047.57
BI	1970–2014	49.009	63.000	−10.904	5.000	9 754	1240.08
AUS	1955–2015	−43.78	−10.16	110.50	154.89	24 126	4054.84
SZ	1951–2008	45.40	48.30	5.60	11.10	15 767	367.20
NZ	2000–2015	−49.18	−32.28	−179.99	180.00	319 353	7547.32

cell size ℓ as tunable parameter. Here, a cell is considered to be a node if at least one earthquake has its epicenter within this cell. A pair of nodes is connected by a link if and only if at least one pair of successive events occur whose epicenters are located within these two cells. The weight w_{ij} of a link connecting two distinct nodes i and j, is the total number of consecutive events between them [16]. The strength of the ith node is defined as $s_i = \sum_j w_{ij}$. In general, there can be self loops of nodes in the network. We do not consider self loops for our study as the ratio of self loops and total number of links are very small. Moreover we are interested in the relationships between distinct locations.

4. Results

We study the spectral plots of the normalized graph Laplacian operator (Δ) of earthquake networks. The normalized graph Laplacian operator (Δ) [20] of an unweighted and undirected graph Γ of size N is defined as: $\Delta = (\Delta)_{ij}, i, j = 1, \ldots, N$ where

$$(\Delta)_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\dfrac{1}{k_i} & \text{if } i \text{ and } j \text{ are neighbors}, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where k_i is the degree of the ith node. Note that this operator is similar to the operator studied in [21] but it is different than the operator widely studied in algebraic graph Laplacian (see [22]). Now the eigenvalue equation can be written as $\Delta u - \lambda u = 0$ where $u(\neq 0)$ is called the eigenfunction corresponding to the eigenvalue λ. Divers eigenvalues among N eigenvalues of Δ can occur with higher multiplicity. The eigenvalues of Δ are bounded between $[0, 2]$. The lowest eigenvalue is always $\lambda_1 = 0$ which happens for a constant u. The multiplicity λ_1 reflects the number of connected components in the graph. The lowest non-zero eigenvalue λ_2 signifies how easily the graph can be cut into two disjoint components. A graph is bipartite if and only if the highest eigenvalue λ_N is 2. Moreover, the distribution of eigenvalues about 1 is symmetric when the graph is bipartite. All the non-zero eigenvalues are equal to $N/(N-1)$ for complete graph (for more see [21]).

The eigenvalues of Δ also encode the local properties produced by certain graph operation like motif joining or duplication [23], e.g., an eigenvalue 1 is produced by the duplication of a single vertex (the simplest motif). The eigenvalue 1 is extensively observed with higher multiplicity in many biological networks. The duplication of an edge (i, j) (motif of size two) generates the eigenvalues $\lambda_{\pm} = 1 \pm \frac{1}{\sqrt{k_i k_j}}$ and the duplication of a chain $(i - j - l)$ of length 3 produces the

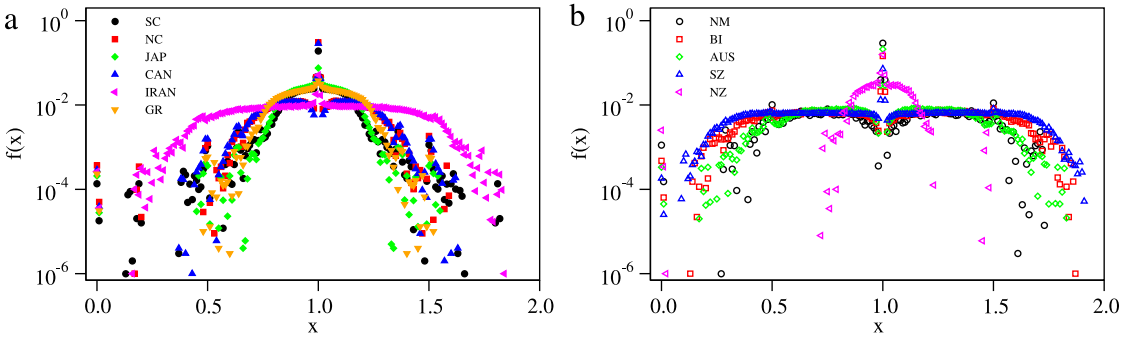

Fig. 2. (Color online) Spectral plot of earthquake networks using $\ell = 0.01$. The spectral probability function $f(x)$ has been plotted as a collection of the eigenvalues λ_i by convolving with a Gaussian kernel with $\sigma = 0.005$.

eigenvalues $\lambda = 1, 1 \pm \sqrt{\frac{1}{k_j}(\frac{1}{k_i} + \frac{1}{k_l})}$. With the specific values of vertices' degrees, the above two motif duplication produce the eigenvalues 1 ± 0.5 and $1 \pm \sqrt{0.5}$ which are also observed in many real networks. By adding a small graph G (with an eigenvalue μ and corresponding eigenfunction that vanishes at a vertex $i \in G$) after identifying the vertex i with any vertex of a graph Γ produces a new graph with the same eigenvalue μ. e.g., adding a triangle (which itself has an eigenvalue 1.5) to a graph generates the same eigenvalue 1.5 to the new graph produced by the joining process (for more details see [23]).

The spectrum distribution of a network has been convolved with a kernel $g(x, \lambda)$. After convolution, we get the function

$$f(x) = \int g(x, \lambda) \sum_k \delta(\lambda, \lambda_k) d\lambda = \sum_k g(x, \lambda_k). \tag{2}$$

Evidently,

$$0 < \int f(x) dx < \infty. \tag{3}$$

Here, we use a Gaussian kernel $1/\sqrt{2\pi\sigma^2}\exp(-(x - m_x)^2/2\sigma^2)$ with $\sigma = 0.005$. It is observed that by selecting different kernels does not alter the result remarkably. However the value of the parameter is important. Many random fluctuations observe and obscures the global features of the network structure in the spectral density plot for small value of σ but, for large values the details become unclear [23]. Hence $\sigma = 0.005$ is chosen as an optimum value from a range. Now $f(x)$ can be written as $f(x) = \sum_k 1/0.005\sqrt{2\pi}\exp(-(x - \lambda_k)^2/0.00005)$.

The spectral probability functions $f(x)$ of earthquake networks for different earthquake regions are shown in Fig. 2 for $\ell = 0.01$. It is observed that each function has a peak at $x = 1$ and the degree of the peakedness of different distributions varies significantly.

4.1. Part A: structural similarity in earthquake networks

We now measure the pairwise similarity of the spectral probability functions using the Jensen–Shannon (JS) divergence [24] for a pair of probability distribution $p_1(x)$ and $p_2(x)$ of a discrete random variable X namely

$$JS(p_1, p_2) = \frac{1}{2}KL(p_1, p) + \frac{1}{2}KL(p_2, p),$$

where $p = (p_1 + p_2)/2$. The above JS divergence is defined in terms of Kullback–Leibler (KL) divergence:

$$KL(p_1, p_2) = \sum_{x \in X} p_1(x) \log \frac{p_1(x)}{p_2(x)}.$$

The JS divergence has many advantages over the KL divergence. For instance, it is symmetric and it is also defined when any one of the probability distribution (p_1 or p_2) is zero. It is known that the square root of JS divergence is a metric [25].

The structural distance $D(\Gamma_1, \Gamma_2)$ between a pair of networks Γ_1 and Γ_2 is defined as [26]:

$$D(\Gamma_1, \Gamma_2) = \sqrt{JS(f_1, f_2)}, \tag{4}$$

where f_1 and f_2 are the spectral probability functions of normalized graph Laplacian for the networks Γ_1 and Γ_2, respectively.

The structural distance $D(\Gamma_1, \Gamma_2)$ between a pair of earthquake networks Γ_1 and Γ_2 has been calculated using the above Eq. (4). The distance matrix $D(\Gamma_i, \Gamma_j)$ is used for hierarchical clustering of earthquake networks of different regions. The hierarchical clustering method [17] consists of four steps: (i) Each individual network is treated as a cluster. (ii) A pair of

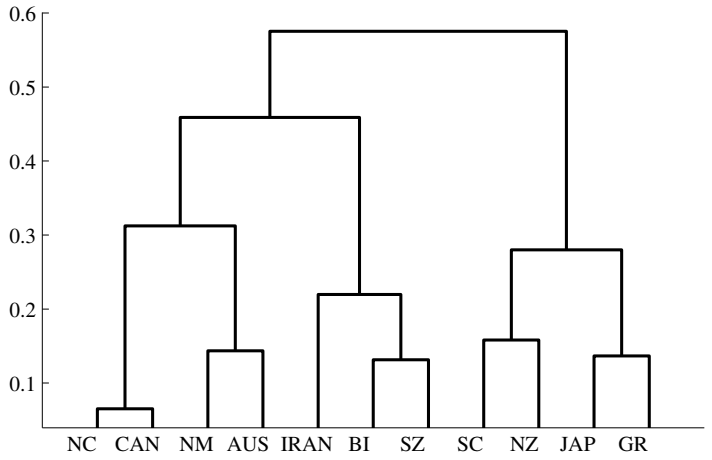

Fig. 3. Dendrogram representing the hierarchical clustering of the different earthquake networks for different regions, based on the similarity of their spectral distributions. The distance metric D is computed from the Jensen–Shannon divergence between the corresponding spectral distributions of a pair of earthquake networks. Different spectral distributions are obtained by convolving with a Gaussian kernel for $\sigma = 0.005$. Earthquake networks are clustered using complete linkage algorithm. The height of the branch represents the distance between clusters, known as the linkage function d. The clusters are mainly formed on the basis of their positions within different tectonic plates.

clusters separated by the shortest distance are merged and form a single cluster together. (iii) The distances between all pair of clusters are computed. (iv) Steps (ii) and (iii) are repeated until all the individual networks form a single cluster.

In step (iii) the distance between clusters can be calculated in different ways. For instance, when the maximum distance between the pair of clusters is considered, it is known as complete linkage clustering. Similarly, it is called the single linkage clustering (average linkage clustering) if the minimum distance (average distance) is considered between the pair of clusters. Note that the hierarchical clustering formed using the complete linkage algorithm is equivalent to a network in which a pair of nodes are linked if the distance between them exceeds certain threshold. In case of the single linkage algorithm it is identical to the minimal spanning tree [27].

We use the complete linkage clustering algorithm to cluster the earthquake networks hierarchically as shown in the dendrogram in Fig. 3. We observe that different pairs of earthquake networks form clusters at the lower level. This pair-wise clustering of earthquake networks reflects their relative positions in tectonic plates. The hierarchical clustering indicates that the strongest structural similarity is observed between the earthquake networks of NC and CAN. A possible reason for this is that NC and CAN belong to the North American tectonic plate. Similarly, the clusters SZ–BI and JAP–GR belong to Eurasian tectonic plate. The earthquake network for SC is over the region close to the border of the Asia pacific and North American tectonic plate and NZ is over the Asia pacific plate. So, here we observe that SC–NZ together form a cluster.

Clusters are formed following the principle of homophily. Community formation based on homophily is also observed in social, economic and biological networks [28,29]. All the community detection algorithms including hierarchical clustering have limitations in detecting perfect clustering of nodes. It can only capture clustering of nodes with limited accuracy. It is not only the fact that geographically nearby regions as more closely connected, but also other features (nature fault zone) might be involved in forming communities. However, clusters between different earthquake networks show interesting similarities in some aspects in their tectonic plates.

Next, we study the highly earthquake prone locations of different regions as it has been reported earlier [8,16] that the degree distributions of unweighted as well as the strength distribution of weighted earthquake networks are highly heterogeneous with power-law distributions. It indicates that few nodes of the network are having very high degree or strength known as hub and large number of nodes are having small degree or strength. We have found the location of the node (the midpoint of a square cell where maximum number of earthquakes occurred) having maximum strength $s_{\max}$ in different earthquake regions. The latitude–longitude coordinates of nodes having maximum strength are (34.36, 116.45), (37.56, −118.43), (34.97, 139.16) (47.39, −70.08), (38.40, 46.73), (38.39, 21.89), (36.56, −89.64), (55.83, −3.09), (−30.82, 117.24), (46.32, 7.36), (−39.03, 175.15) with $s_{\max} = 7815, 3935, 2026, 3412, 173, 2774, 1343, 655, 712, 184, 13\,417$ for SC, NC, JAP, CAN, IRAN, GR, NM, BI, AUS, SZ and NZ, respectively. Top ten earthquake prone locations of each earthquake region are tabulated in the Table 2. We show the position of all nodes for SC, NC and JAP regions on maps in Fig. 4. As can be seen from the Fig. 4, the strength of nodes are highly heterogeneous with a wide range of values for the nodal strength spanning over a several orders of magnitude. The groups of nodes having high strength are indicated by red color for each region are showing different fault zones. These hub regions not only reveal the occurrence of very large number of earthquake events, but also represents the locations of main shocks. In fact, there is empirical evidence that the aftershocks have a tendency to return to the location of associated main shock [8].

In the case of the underlying undirected structure of earthquake networks, we observe that their similarities between their topologies are dependent on that between their underlying tectonic plates. However, one can consider a directed nature

Table 2
Ten major earthquake prone locations are identified for different earthquake regions. Here, θ and ϕ represents the latitude and longitude in degree of the midpoint of a square cell with side $\ell = 0.01$. The total strength of a node is denoted by s.

Region	θ, ϕ, s				
SC	34.36, −116.45, 7815 33.67, −116.73, 5259	35.05, −117.70, 6123 36.09, −117.84, 4865	35.74, −117.63, 5467 36.03, −117.84, 4689	36.03, −117.77, 5419 35.80, −117.63, 4440	33.50, −116.45, 5291 33.50, −116.52, 4432
NC	37.56, −118.43, 3935 37.56, −118.80, 2332	38.78, −122.75, 3598 37.56, −118.89, 2290	38.78, −122.84, 3444 37.28, −121.65, 2140	38.85, −122.84, 3001 38.85, −122.75, 2085	37.63, −118.89, 2484 37.49, −118.43, 1933
JAP	34.97, 139.16, 2026 36.73, 139.41, 1133	34.18, 135.25, 1316 33.99, 135.25, 1080	42.02, 139.16, 1254 35.55, 140.14, 1035	36.14, 139.90, 1207 32.03, 130.35, 1023	35.75, 137.45, 1158 36.14, 140.14, 997
CAN	47.39, −70.08, 3412 49.71, −127.73, 1366	52.48, −131.70, 2834 50.17, −127.73, 1172	52.94, −132.70, 2520 47.86, −70.08, 1121	50.63, −130.71, 1481 48.78, −128.72, 1116	48.78, −122.76, 1385 48.32, −122.76, 1087
IRAN	38.40, 46.73, 173 35.72, 49.04, 81	30.84, 56.85, 105 32.55, 48.75, 72	27.67, 56.56, 85 28.41, 57.14, 70	28.41, 51.64, 84 33.77, 49.04, 69	33.77, 48.75, 83 27.43, 55.98, 68
GR	38.39, 21.89, 2774 38.26, 21.72, 838	38.39, 22.06, 2086 37.62, 20.90, 655	37.23, 22.06, 1437 38.26, 22.22, 607	38.26, 22.06, 1222 38.39, 22.22, 595	38.39, 21.72, 890 38.01, 21.56, 582
NM	36.56, −89.64, 1343 36.01, −89.87, 242	36.19, −89.42, 848 35.27, −92.35, 239	36.37, −89.42, 615 36.56, −89.42, 172	36.37, −89.64, 452 35.82, −90.09, 169	36.19, −89.64, 244 35.45, −92.35, 154
BI	55.83, −3.09, 655 50.14, −5.09, 112	56.16, −3.69, 319 53.04, −2.29, 107	55.94, −3.09, 288 53.15, −1.10, 99	52.93, −4.49, 157 53.04, −2.09, 88	56.27, −3.69, 136 52.93, −4.29, 81
AUS	−30.82, 117.24, 712 −30.45, 117.65, 275	−34.83, 149.16, 667 −33.37, 138.52, 260	−19.88, 134.02, 512 −30.45, 116.83, 256	−30.45, 117.24, 376 −31.55, 117.24, 256	−31.55, 116.83, 293 −32.28, 117.24, 233
SZ	46.32, 7.36, 184 46.32, 7.46, 71	46.32, 7.41, 181 47.68, 7.46, 68	47.58, 7.61, 99 46.55, 10.31, 68	46.32, 7.32, 83 46.29, 7.22, 66	45.99, 7.90, 73 46.35, 7.41, 62
NZ	−39.03, 175.15, 13 417 −37.68, 176.94, 11 327	−45.14, 167.08, 13 127 −41.07, 175.15, 10 664	−38.36, 176.04, 11 752 −39.03, 176.04, 9916	−40.39, 176.04, 11 642 −41.75, 174.25, 9861	−39.71, 176.94, 11 385 −43.79, 172.46, 8070

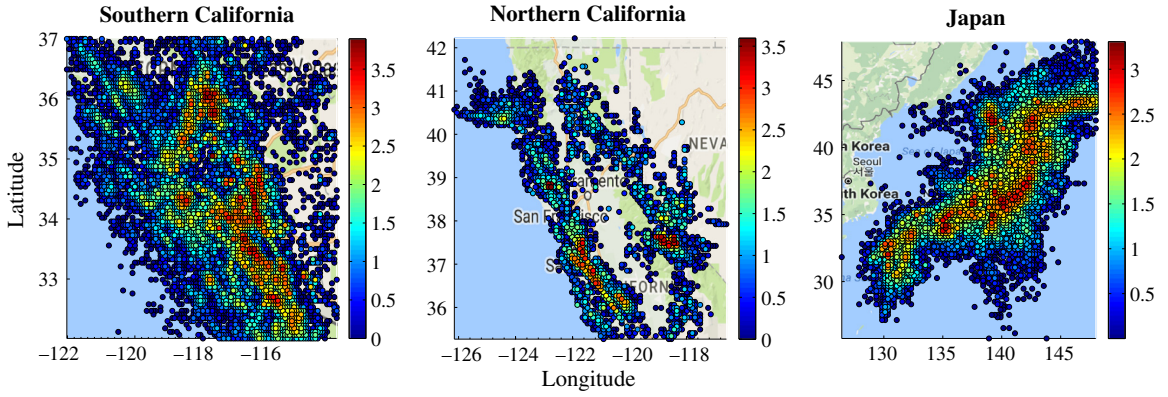

Fig. 4. Graphical representation for the position of all nodes in Southern California, Northern California and Japan earthquake regions on maps. Color bars indicate the strength (s) of nodes in logarithmic scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of links in earthquake network. We further extend our study to investigate the directed version of the earthquake network. A directed link between nodes A and B ($A \rightarrow B$) indicates that two successive events have occurred on A and B respectively. A strongly connected component analysis shows that each of these earthquake networks are strongly connected, i.e., there exists at least one directed path between every pair of nodes. In our subsequent analysis, we focus on the directed version of earthquake networks to capture the statistical relations between distinct locations for all earthquake regions.

4.2. Part B: statistical analysis of earthquake regions

Now, we analyze the consecutive occurrence of earthquake events by observing the trend of the consecutive directed sequences. We also do the statistical analysis for probabilistic predictions of consecutive events.

4.2.1. Number of consecutive earthquakes

First, we quantify the pair of nodes having the link with maximum weight $w_{\max}$ for different earthquake regions. Here, the maximum weight indicates highest number of consecutive earthquake events between the pair of nodes in that earthquake region. The latitude–longitude coordinates for the top ten pairs of nodes AB with maximum weight w_{AB} are shown in the

Table 3
Statistical analysis of earthquake regions, SC, NC and JAP. w_{AB} represents the total number of successive earthquake events between node A and B. $P(B^*|A^*)$ is the conditional probability that there will be an earthquake at B (i.e., B^*) immediately after one occurs at A (i.e., A^*). $Z_{B^*|A^*}$ is the z-score for the successive events A^*B^*.

A	B	w_{AB}	P(B^* | A^*)	$Z_{B^*\|A^*}$
SC				
35.74, −117.63	35.80, −117.63	1111	0.20	154.06
35.80, −117.63	35.74, −117.63	1066	0.24	147.69
36.03, −117.91	36.03, −117.84	552	0.15	87.58
36.03, −117.84	36.03, −117.91	545	0.12	86.32
34.02, −116.31	33.96, −116.31	422	0.11	92.21
33.96, −116.31	34.02, −116.31	401	0.16	87.51
33.16, −115.61	33.21, −115.61	387	0.11	96.99
33.04, −114.99	35.05, −117.70	380	0.21	76.06
33.21, −115.61	33.16, −115.61	380	0.18	95.31
33.04, −114.99	35.05, −117.70	380	0.21	76.06
NC				
37.63, −118.43	37.56, −118.43	530	0.32	52.80
37.56, −118.43	37.63, −118.43	509	0.13	49.66
37.49, −118.43	37.56, −118.43	450	0.23	39.02
37.56, −118.43	37.49, −118.34	448	0.11	39.19
37.56, −118.43	37.49, −118.43	442	0.11	37.70
37.49, −118.34	37.56, −118.43	406	0.22	35.07
37.49, −118.62	37.42, −118.62	319	0.21	54.96
37.42, −118.62	37.49, −118.62	296	0.19	50.61
38.78, −122.84	38.78, −122.75	247	0.07	8.57
37.49, −118.43	37.63, −118.43	239	0.12	33.00
JAP				
34.97, 139.16	34.97, 139.41	206	0.10	92.04
34.97, 139.41	34.97, 139.16	182	0.41	81.40
34.57, 135.00	34.77, 135.25	128	0.15	68.33
34.77, 135.25	34.57, 135.00	126	0.17	67.26
42.02, 139.16	41.82, 139.16	70	0.06	27.95
34.77, 139.41	34.97, 139.16	66	0.19	32.50
34.97, 139.16	34.77, 139.41	59	0.03	28.72
41.82, 139.16	42.02, 139.16	54	0.07	21.06
42.02, 139.41	42.02, 139.16	48	0.06	18.73
34.57, 139.41	34.77, 139.41	47	0.21	72.62

third column of the Tables 3–6. We observe a wide range of variation for the values of link weights in each earthquake regions which is also reported in [16]. This fact prompted us to study the conditional probability of consecutive events.

4.2.2. Conditional probability in earthquake networks

Now we study the conditional probability of the occurrence of an earthquake event at node B, right after one takes place at node A.

Let A^* denote an earthquake at epicenter A and B^* is the same for B. The conditional probability of B^* given A^* is:

$$P(B^* \mid A^*) = \frac{P(A^* \cap B^*)}{P(A^*)}, \tag{5}$$

where $P(A^* \cap B^*) = w_{AB}/(n-1)$ represents probability that the earthquake happened at A and B consequently and $P(A^*) = s_A/n$ is the probability of earthquake events at A^*. Here we compute the conditional probability for all pairs of nodes in each earthquake region.

The latitude–longitude coordinates of the top 10 two consecutive nodes are given in the fourth column of Tables 3–6. As can be seen from Fig. 5 the average conditional probability $\langle P(B^*|A^*)\rangle$ decreases with the haversine distance between A and B for three different earthquake regions. Similar trend is also observed for other earthquake regions. We further show the statistically significant negative correlation between the $\langle P(B^*|A^*)\rangle$ and distance in Table 7 for each earthquake region.

Here we see that $P(B^*|A^*)$, in all the earthquake regions, for some A and B are not negligible. The higher values of $P(B^*|A^*)$ may depend on the high frequency of B^*. So to predict the occurrence of B^* right after A^*, with high statistically significance, we compute the z-score for $P(B^*|A^*)$.

Table 4
Statistical analysis of earthquake regions, CAN, IRAN and GR. w_{AB} represents the total number of successive earthquake events between node A and B. $P(B^*|A^*)$ is the conditional probability that there will be an earthquake at B (i.e., B^*) immediately after one occurs at A (i.e., A^*). $Z_{B^*|A^*}$ is the z-score for the successive events A^*B^*.

A	B	w_{AB}	$P(B^* \mid A^*)$	$Z_{B^*\|A^*}$
CAN				
52.48, −131.70	52.94, −132.70	284	0.10	19.21
52.94, −132.70	52.48, −131.70	262	0.10	16.98
52.48, −132.70	52.48, −131.70	184	0.18	23.04
52.48, −131.70	52.48, −132.70	173	0.06	21.01
52.94, −132.70	52.48, −132.70	158	0.06	20.51
52.48, −132.70	52.94, −132.70	145	0.14	18.52
52.48, −131.70	47.39, −70.08	107	0.04	−2.23
52.94, −132.70	47.39, −70.08	100	0.04	−1.65
47.39, −70.08	52.94, −132.70	92	0.03	−2.39
52.02, −131.70	52.48, −131.70	90	0.09	7.84
IRAN				
38.40, 46.73	38.65, 46.73	18	0.10	31.02
38.65, 46.73	38.40, 46.73	17	0.47	29.37
28.41, 59.17	28.16, 59.17	15	0.42	45.17
28.16, 59.17	28.41, 59.17	15	0.26	45.14
28.41, 59.17	28.16, 59.17	15	0.42	45.17
28.16, 59.17	28.41, 59.17	15	0.26	45.14
36.45, 51.64	36.45, 51.35	14	0.29	43.94
36.45, 51.35	36.45, 51.64	12	0.30	37.63
30.84, 56.56	30.84, 56.85	12	0.33	26.64
36.45, 51.35	36.45, 51.64	12	0.30	37.63
GR				
38.39, 22.06	38.39, 21.89	192	0.09	22.19
38.39, 21.89	38.39, 22.06	187	0.07	21.38
38.26, 22.06	38.39, 22.06	69	0.06	11.16
38.13, 26.52	38.13, 26.68	65	0.20	71.54
38.39, 21.89	38.39, 21.72	62	0.02	9.82
38.39, 22.06	38.26, 22.06	62	0.03	9.54
38.39, 21.89	38.39, 21.72	62	0.02	9.82
38.39, 22.06	38.26, 22.06	62	0.03	9.54
37.62, 20.90	37.75, 20.90	59	0.09	37.56
38.39, 21.72	38.39, 21.89	57	0.06	8.74

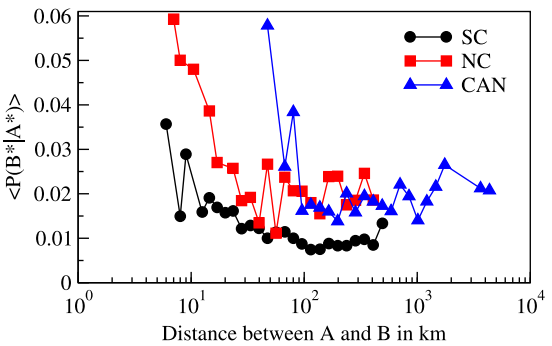

Fig. 5. Average conditional probability $\langle P(B^*|A^*)\rangle$ is plotted with the haversine distance between A and B. We have used logarithmic binning of the horizontal axis.

4.2.3. *Statistical significance factor for consecutive events*

The statistical significance can be measured by computing the z-score for the successive events at A and B:

$$Z_{B^*|A^*} = \frac{P(B^*|A^*) - \pi(A^*, B^*)}{\sqrt{\frac{\pi(A^*,B^*)(1-\pi(A^*,B^*))}{n(A^*)}}}, \tag{6}$$

where π is the stationary probability matrix (for more details see Ref. [30]) (in our case, stationary matrix π exists, since the underlying graph is strongly connected and non-bipartite), calculated from matrix M. The rows and columns of M are

Table 5
Statistical analysis of earthquake regions, NM, BI and AUS. w_{AB} represents the total number of successive earthquake events between node A and B. $P(B^*|A^*)$ is the conditional probability that there will be an earthquake at B (i.e., B^*) immediately after one occurs at A (i.e., A^*). $Z_{B^*|A^*}$ is the z-score for the successive events A^*B^*.

A	B	w_{AB}	$P(B^* \mid A^*)$	$Z_{B^*\|A^*}$
NM				
36.56, −89.64	36.19, −89.42	213	0.16	5.61
36.19, −89.42	36.56, −89.64	197	0.23	4.37
36.37, −89.42	36.56, −89.64	165	0.27	6.07
36.56, −89.64	36.37, −89.42	148	0.11	4.04
36.56, −89.64	36.37, −89.64	121	0.09	4.84
36.37, −89.64	36.56, −89.64	118	0.26	4.80
36.19, −89.42	36.37, −89.42	106	0.12	4.80
36.37, −89.42	36.19, −89.42	83	0.13	1.93
36.56, −89.64	36.01, −89.87	68	0.05	4.00
36.37, −89.64	36.19, −89.42	62	0.14	1.80
BI				
55.94, −3.09	55.83, −3.09	115	0.40	19.32
55.83, −3.09	55.94, −3.09	105	0.16	16.77
55.83, −3.09	56.16, −3.69	30	0.05	0.65
53.04, −2.29	53.04, −2.09	30	0.28	26.37
55.83, −3.09	56.16, −3.69	30	0.05	0.65
53.04, −2.29	53.04, −2.09	30	0.28	26.37
55.83, −3.09	52.93, −4.49	27	0.04	3.85
53.04, −2.09	53.04, −2.29	26	0.30	22.72
56.16, −3.69	55.83, −3.09	23	0.07	−0.76
53.48, −2.09	53.48, −2.29	23	0.77	64.37
AUS				
−30.45, 117.24	−30.45, 116.83	77	0.20	32.18
−30.45, 116.83	−30.45, 117.24	74	0.29	30.94
−30.09, 117.65	−30.45, 117.65	51	0.50	41.28
−30.45, 117.65	−30.09, 117.65	45	0.16	36.11
−31.91, 116.83	−31.55, 116.83	33	0.21	19.67
−31.55, 116.83	−31.91, 116.83	32	0.11	18.97
−34.83, 149.16	−30.82, 117.24	30	0.04	1.02
−30.45, 117.24	−30.82, 117.24	29	0.08	4.05
−30.82, 117.24	−30.45, 117.24	28	0.04	3.75
−19.88, 134.02	−34.83, 149.16	23	0.04	1.21

different events and (A^*, B^*)th element of M is defined as:

$$M_{A^*B^*} = P(B^*|A^*),$$

and $\sqrt{\frac{\pi(A^*,B^*)(1-\pi(A^*,B^*))}{n(A^*)}}$ is the corresponding standard deviation. Here, $n(A^*)$ denotes the cardinality of A^*, i.e., s_A.

We find z-score values among top 10 two consecutive nodes for each of the region. This is shown in the fifth column of the Tables 3–6. A statistically significant negative correlation is also found between haversine distance and z-score shown in Table 7, implying a decaying statistical significance level of the conditional probability $\langle P(B^*|A^*)\rangle$ with the distance.

5. Conclusion

We have analyzed eleven earthquake catalogs from the different parts of the world using a complex network framework and statistical techniques. The spectral probability functions of various regions have distinct peakedness. The dissimilarity between the spectral probability functions are shown as a distance matrix, which is calculated using a probabilistic metric, viz., Jensen–Shannon divergence. We have hierarchically clustered the earthquake regions from the distance matrix and linked it with their proximity in tectonic plates. The locations of highly earthquake prone regions have also been identified. To understand the chain of earthquake events, we have delved deeper by considering directed version of these networks, and it is revealed that the entire network is strongly connected for all regions. Furthermore, we have calculated the conditional probability of a forthcoming earthquake event on a node and the statistical significance of those probabilities is also estimated by comparing with their stationary probability.

The structural similarities between networks can be computed by the spectrum of non-normalized Laplacian matrix or adjacency matrix. However, a correlation is observed between the spectral densities of these matrices with the degree distribution of the network [31,32]. It is difficult to compare different real networks because of their irregular structure and heterogeneous sizes. Networks with same sizes can be aligned to each other to find the structural difference. As the spectrum of normalized Laplacian is bounded between 0 and 2, we have the advantage for comparing the spectral plots of

Table 6
Statistical analysis of earthquake regions, SZ and NZ. w_{AB} represents the total number of successive earthquake events between node A and B. $P(B^*|A^*)$ is the conditional probability that there will be an earthquake at B (i.e., B^*) immediately after one occurs at A (i.e., A^*). $Z_{B^*|A^*}$ is the z-score for the successive events A^*B^*.

A	B	w_{AB}	$P(B^* \mid A^*)$	$Z_{B^*\|A^*}$
SZ				
46.29, 7.17	46.29, 7.22	22	0.37	40.92
46.29, 7.22	46.29, 7.17	17	0.26	31.48
47.35, 11.03	47.35, 10.99	12	0.63	70.96
47.58, 7.61	47.64, 7.56	11	0.11	32.35
47.35, 10.99	47.35, 11.03	11	0.52	65.03
47.58, 7.61	47.64, 7.56	11	0.11	32.35
47.35, 10.99	47.35, 11.03	11	0.52	65.03
46.32, 7.41	46.32, 7.36	10	0.06	4.97
47.68, 7.51	47.68, 7.46	10	0.33	25.88
47.58, 7.61	47.61, 7.56	10	0.10	31.46
NZ				
−45.14, 167.08	−39.03, 175.15	725	0.06	3.39
−37.68, 176.94	−45.14, 167.08	692	0.06	6.64
−39.03, 175.15	−45.14, 167.08	681	0.05	1.61
−39.71, 176.94	−40.39, 176.04	679	0.06	9.14
−39.03, 175.15	−38.36, 176.04	642	0.05	2.90
−40.39, 176.04	−39.03, 175.15	637	0.05	2.94
−45.14, 167.08	−37.68, 176.94	633	0.05	4.02
−40.39, 176.04	−39.71, 176.94	629	0.05	6.81
−41.07, 175.15	−39.03, 175.15	616	0.06	4.27
−38.36, 176.04	−39.03, 175.15	615	0.05	1.75

Table 7
The Pearson correlation coefficient of haversine distance between A and B with conditional probability $\langle P(B^*|A^*)\rangle$ (in third column) and with z-score (in fourth column) for each earthquake region. Different threshold of W_{AB} are chosen to get statistically significant result for different region.

Region	W_{AB} threshold	Correlation coefficient between $\langle P(B^*\|A^*)\rangle$ and distance (p value)	Correlation coefficient between z-score and distance (p value)
SC	10	$-0.22\ (<10^{-10})$	$-0.31\ (<10^{-10})$
NC	10	$-0.19\ (<10^{-10})$	$-0.38\ (<10^{-10})$
JAP	10	$-0.26\ (<10^{-10})$	$-0.44\ (<10^{-10})$
CAN	10	$-0.02\ (<10^{-3})$	$-0.13\ (<10^{-10})$
IRAN	5	$-0.16\ (<10^{-4})$	$-0.26\ (<10^{-10})$
GR	10	$-0.26\ (<10^{-10})$	$-0.37\ (<10^{-10})$
NM	5	$-0.06\ (<10^{-2})$	$-0.04\ (<10^{-2})$
BI	10	$-0.35\ (<10^{-2})$	$-0.35\ (<10^{-2})$
AUS	10	$-0.26\ (<10^{-6})$	$-0.31\ (<10^{-10})$
SZ	2	$-0.05\ (<10^{-10})$	$-0.16\ (<10^{-10})$
NZ	2	$-0.01\ (<10^{-10})$	$-0.06\ (<10^{-10})$

networks with different sizes. Our study can be extended for further investigations of earthquake prediction and prevention for different earthquake prone regions.

6. Data sources

SC: Network data for SC region was downloaded from Southern California Earthquake Data Center, http://www.data.scec.org/ earthquake catalog for this study were accessed through the SCEDC: SCEDC (2013):Southern California Earthquake Center Caltech Dataset. doi:10.7909/C3WD3xH1.

NC: NC network data was downloaded from USGS NCSN catalog, http://quake.geo.berkeley.edu/ncedc/catalog-search.html earthquake catalog for this study were accessed through the NCEDC (2014), Northern California Earthquake Data Center. UC Berkeley Seismological Laboratory. Dataset. doi:10.7932/NCEDC.

JAP: Data of JAP region was downloaded from Japan University Network Earthquake Catalog, http://wwweic.eri.u-tokyo.ac.jp/db/junec/ [33].

CAN: Network data of CAN region was downloaded from Canada's National Earthquake Database, http://earthquake.usgs.gov/earthquakes/eqarchives/. We gratefully acknowledge US Geological Survey for keeping the data publicly available.

IRAN: IRAN network data was downloaded from International Institute of Earthquake Engineering and Seismology, http://www.iiees.ac.ir and we duly acknowledge International Institute of Earthquake Engineering and Seismology for keeping the data publicly available.

GR: Data of GR region was downloaded from Institute of Geodynamics, http://www.gein.noa.gr/services/cat.html. We acknowledge National Observatory of Athens for keeping the data publicly available.

NM: Network data of NM region was downloaded from Centre for Earthquake Research and Information, Univ. of Memphis, http://www.ceri.memphis.edu/seismic/catalogs/. We acknowledge University of Memphis of keeping the data publicly available.

BI: Network data of BI region was downloaded from BGS Earthquake Database, http://www.earthquakes.bgs.ac.uk/earthquakes/dataSearch.html and access through Reproduced with the permission of the British Geological Survey NERC. All rights Reserved.

AUS: AUS network data was downloaded from Geoscience Australia—Earthquake Database, http://www.ga.gov.au/earthquakes/searchQuake.do and access through the Australian Governments Open Access and Licensing Framework (AusGOAL).

SZ: Data of SZ was downloaded from Swiss Seismological Service, http://hitseddb.ethz.ch:8080/ecos09/query_sum [34].

NZ: Network data of NZ region data was downloaded from GeoNet the official source of geological hazard information for New Zealand, http://quakesearch.geonet.org.nz/. We acknowledge the New Zealand GeoNet project and its sponsors EQC, GNS Science and LINZ, for keeping data publicly available.

Acknowledgments

We thank Shakti N. Menon for helpful suggestions. We gratefully acknowledge the assistance of Buddhananda Banerjee in the statistical part of this work. KD gratefully acknowledges the financial support from CSIR (file number 09/921 (0070)/2012-EMR-I), India.

References

[1] B. Gutenberg, C.F. Richter, Frequency of earthquakes in California, Bull. Seismol. Soc. Amer. 34 (1944) 185–188.
[2] F. Omori, On the after-shocks of earthquakes, J. Colloid Sci. 7 (1894) 111–200.
[3] R. Burridge, L. Knopoff, Model and theoretical seismicity, Bull. Seismol. Soc. Am. 57 (1967) 341–371.
[4] A. Cox, R.B. Hart, Plate Tectonics: How it Works, John Wiley & Sons, 2009.
[5] P. Bak, K. Christensen, L. Danon, T. Scanlon, Unified scaling law for earthquakes, Phys. Rev. Lett. 88 (2002) 178501.
[6] M. Baiesi, M. Paczuski, Scale-free networks of earthquakes and aftershocks, Phys. Rev. E 69 (2004) 066106.
[7] J. Davidsen, M. Paczuski, Analysis of the spatial distribution between successive earthquakes, Phys. Rev. Lett. 94 (2005) 048501.
[8] S. Abe, N. Suzuki, Scale-free network of earthquakes, Europhys. Lett. 65 (2004) 581.
[9] S. Abe, N. Suzuki, Small-world structure of earthquake network, Physica A 337 (2004) 357–362.
[10] S. Abe, N. Suzuki, Complex earthquake networks: Hierarchical organization and assortative mixing, Phys. Rev. E 74 (2006) 026113.
[11] S. Abe, N. Suzuki, Dynamical evolution of clustering in complex network of earthquakes, Eur. Phys. J. B 59 (2007) 93–97.
[12] S. Abe, N. Suzuki, Violation of the scaling relation and non-markovian nature of earthquake aftershocks, Physica A 388 (2009) 1917–1920.
[13] S. Abe, N. Suzuki, Scaling relation for earthquake networks, Physica A 388 (2009) 2511–2514.
[14] S. Abe, N. Suzuki, Determination of the scale of coarse graining in earthquake networks, Europhys. Lett. 87 (2009) 48008.
[15] S. Abe, D. Pastén, N. Suzuki, Finite data-size scaling of clustering in earthquake networks, Physica A 390 (2011) 1343–1349.
[16] A. Chakraborty, G. Mukherjee, S. Manna, Weighted network analysis of earthquake seismic data, Physica A 433 (2015) 336–343.
[17] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (1967) 241–254.
[18] S. Abe, N. Suzuki, Scale-invariant statistics of period in directed earthquake network, Eur. Phys. J. B 44 (2005) 115–117.
[19] S. Abe, N. Suzuki, Universal law for waiting internal time in seismicity and its implication to earthquake network, Europhys. Lett. 97 (2012) 49002.
[20] A. Banerjee, J. Jost, On the spectrum of the normalized graph Laplacian, Linear Algebra Appl. 11 (2008) 3015–3022.
[21] F.R. Chung, Spectral Graph Theory, Vol. 92, American Mathematical Society, 1997.
[22] B. Mohar, Y. Alavi, G. Chartrand, O.R. Oellermann, The Laplacian spectrum of graphs, Graph Theory Combin. Appl. 2 (1991) 871–898.
[23] A. Banerjee, J. Jost, Graph spectra as a systematic tool in computational biology, Discrete Appl. Math. 157 (2009) 2425–2431.
[24] J. Lin, Divergence measures based on the shannon entropy, IEEE Trans. Inform. Theory 37 (1991) 145–151.
[25] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inform. Theory 49 (2003) 1858–1860.
[26] A. Banerjee, Structural distance and evolutionary relationship of networks, Biosystems 107 (2012) 186–196.
[27] J.C. Gower, G. Ross, Minimum spanning trees and single linkage cluster analysis, Appl. Stat. 18 (1969) 54–64.
[28] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. 99 (2002) 7821–7826.
[29] S. Currarini, M.O. Jackson, P. Pin, An economic model of friendship: Homophily, minorities, and segregation, Econometrica 77 (2009) 1003–1045.
[30] L. Lovász, Random walks on graphs, Combinatorics 2 (1993) 1–46. Paul erdos is eighty.
[31] S. Dorogovtsev, A. Goltsev, J. Mendes, A. Samukhin, Random networks: eigenvalue spectra, Physica A 338 (2004) 76–83.
[32] C. Zhan, G. Chen, L.F. Yeung, On the distributions of Laplacian eigenvalues versus node degrees in complex networks, Physica A 389 (2010) 1779–1788.
[33] S. Tsuboi, K. Koketsu, K. Takano, T. Miyatake, K. Abe, Y. Hagiwara, Hypocenter determination procedure of Japan university network earthquake catalog, J. Seismol. Soc. Japan 42 (1989) 277–284.
[34] D. Fäh, D. Giardini, P. Kästli, N. Deichmann, M. Gisler, G. Schwarz-Zanetti, S. Alvarez-Rubio, S. Sellami, B. Edwards, B. Allmann, F. Bethmann, J. Wössner, G. Gassner-Stamm, S. Fritsche, D. Eberhard, Ecos-09 earthquake catalogue of switzerland release 2011 report and database. public catalogue, 17. 4. 2011. swiss seismological service eth zurich, RISK (2011).