

RANKING OF SEISMIC ZONES IN GREECE USING MEASURES OF NETWORKS FORMED FROM EARTHQUAKE HISTORICAL DATA

Chorozoglou D.¹, Kugiumtzis D.², Papadimitriou E.¹ and Tsaklidis G.³

¹Department of Geophysics, Aristotle University of Thessaloniki, chorozod@geo.auth.gr,
ritsa@geo.auth.gr

²Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki,
dkugiu@auth.gr

³Department of Mathematics, Aristotle University of Thessaloniki, tsaklidi@math.auth.gr

Abstract

Interdependencies in earthquakes with $M \geq 5.0$ that occurred in the Greek territory during 1911-2014 are investigated by means of network analysis. The nodes of the seismic network represent seismic zones and the connections between them are given by the time succession of two earthquakes of any seismic zones. Five different approaches for determining the weighted connections are studied based on different normalizations of the frequencies of successions for each pair of nodes (areas). In particular, two approaches transform the transition matrix to transition probability matrix of a Markov chain or Gibbs sequence. The network structure is quantified by five selected network measures, and we assess how changes in seismicity (number of earthquakes or seismic moment release) in sliding time windows (e.g. decades) agree with changes in the network structure. It turns out that the best matching of seismicity and network measures is succeeded with the two approaches normalizing the maximum frequency of a succession pair, over the current time window or over all time windows, and with the approach that is based on probability distribution for the pairs of nodes. For the approaches transforming the transition matrix to transition probability matrix, a test is developed for the null hypothesis according to which the observed transition in a time window is equal to the predicted one on the basis on the information given from the previous time window. The developed parametric test is based on Chi-square null distribution and shows that the null hypothesis could not be rejected in some time windows, providing a tool for future seismic hazard assessment.
Keywords: weighted matrix, transition matrix, Markov chain, Gibbs sequence, network measure.

Περίληψη

Επιχειρείται διερεύνηση των αλληλεξαρτήσεων μεταξύ σεισμών με $M \geq 5.0$ που έγιναν στον ελληνικό χώρο κατά την περίοδο 1911-2014 με τη βοήθεια της ανάλυσης δικτύου. Οι κόμβοι του δικτύου αντιπροσωπεύουν σεισμικές ζώνες και οι συνδέσεις μεταξύ τους δίνονται από την χρονική διαδοχή δύο σεισμών που πραγματοποιούνται σε οποιαδήποτε σεισμική ζώνη. Εξετάζονται πέντε διαφορετικές προσεγγίσεις για τον καθορισμό της σταθμισμένης σύνδεσης, με βάση τη διαφορετική εξομάλυνση των συχνοτήτων των διαδοχικών σεισμών για κάθε ζεύγος κόμβων (περιοχές). Ειδικότερα, δύο προσεγγίσεις

μετατρέπουν τον πίνακα μετάβασης σε πίνακα πιθανοτήτων μετάβασης μιας αλυσίδας Markov ή αλληλουχίας Gibbs. Η δομή του δικτύου ποσοτικοποιείται με πέντε επιλεγμένα μέτρα δικτύου, και εκτιμώνται αν οι μεταβολές της σεισμικότητας (πλήθος σεισμών ή απελευθέρωση σεισμικής ενέργειας) σε όλα τα συρόμενα παράθυρα του χρόνου (π.χ. δεκαετίες) συμφωνούν με τις αλλαγές στη δομή του δικτύου. Αποδεικνύεται ότι η καλύτερη αντιστοίχιση της σεισμικότητας και των μέτρων δικτύου επιτυγχάνεται με τις δύο προσεγγίσεις στις οποίες η ομαλοποίηση γίνεται με τη μέγιστη συχνότητα ενός ζεύγους διαδοχής, πάνω από το τρέχον χρονικό παράθυρο ή πάνω από όλα τα παράθυρα του χρόνου, και με την προσέγγιση που βασίζεται στην κατανομή πιθανοτήτων για τα ζεύγη των κόμβων. Για τις προσεγγίσεις που μετατρέπουν τον πίνακα μετάβασης σε πίνακα πιθανοτήτων μετάβασης, αναπτύχθηκε έλεγχος με μηδενική υπόθεση ότι η παρατηρούμενη μετάβαση σε ένα χρονικό παράθυρο είναι ίση με την προβλεπόμενη με βάση την πληροφορία που δίνεται από το προηγούμενο χρονικό παράθυρο. Η ανάπτυξη παραμετρικής δοκιμασίας που βασίζεται στην κατανομή χ-τετράγωνο έδειξε ότι η μηδενική υπόθεση δεν μπορεί να απορριφθεί σε κάποια χρονικά παράθυρα, παρέχοντας ένα χρήσιμο εργαλείο για εκτίμηση της μελλοντικής σεισμικής επικινδυνότητας.

Λέξεις κλειδιά: Πίνακας βαρών, πίνακας μετάβασης, αλυσίδα Markov, ακολουθία Gibbs, μέτρο δικτύου.

1. Introduction

The network approach is a powerful tool for analyzing dynamic structures of complex systems. Complex network analysis is an emerging field that was introduced recently also in seismology by Abe and Suzuki (2004) in order to study seismicity as a spatiotemporal complex system. The seismic network has a number of interesting properties, some of which are common with many other natural and artificial complex systems, such as the metabolic network and the network of the web (Albert and Barabasi, 2002), which enable the study of seismicity. Global physical properties of seismicity can be explored by examining the geometrical (topological) and dynamic characteristics. The changes in the structure of the network (dynamic characteristics) can reveal a main shock few days before it occurs and the topological characteristics show the interdependencies between the variables of the complex network (Abe and Suzuki, 2009).

The purpose of this study is the construction of seismic networks, with directed and weighted connections, where the nodes of networks represent seismic zones and the connections between them are given by the time succession of two earthquakes of any seismic zone. The weights of connections are formed by different normalizations of the frequencies of successions. Six approaches are shown, one for simple connections and five for weighted connections, to forming the adjacency or weighted matrix for introducing the connections of seismic network. Two of the five approaches for weighted connections transform the transition matrix to transition probability matrix of a Markov chain or Gibbs sequence to purpose the estimation of future seismic hazard. In addition, we examine which of the six approaches can depict better the seismic activity observed in the study area during 1911-2014, using two ranking lists of seismic zones that come from the real seismic activity and the network measures for each time window. Finally, we make a statistical test which examines whether the a priori known transition probability matrix differs from the estimated one. The results encourage research in the next step for evaluation of future seismic hazard.

2. Data and Methodology

The observation data are obtained from the earthquake catalog compiled in the Geophysics Department of the Aristotle University of Thessaloniki (<http://geophysics.geo.auth.gr/ss/>), and comprise crustal earthquakes (focal depth less than 40 Km) of magnitude $M \geq 5.0$ (2,018 events) that occurred in 1911-2014 (Fig. 1). In this study the seismic measures which are used to construct

the seismic network are indicated by the sum of the number of successive earthquakes or by the sum of the seismic moment M_0 released for each pair of nodes. The seismic moment is calculated from the empirical relation $\log M_0 = 1.5M + 16.01$, where M the earthquake magnitude (Kanamori and Anderson, 1975). The analysis is performed in sliding non-overlapping time windows of 2, 3, 5 and 10 years duration. The approaches which are based on Markov chain and Gibbs sequence are applied only for 10 years duration because they require at least one transition from each seismic zone.

2.1 Approaches for determining the connections of networks

Generally with the term of network we mean the graph $G = (N, E)$ which is defined by the nodes and the connections between them, where N is the set of nodes and E the set of connections. The $|E|$ network connections of the $|N|$ nodes can be directed or undirected, weighted or binary such that the network of N nodes is fully described by a square matrix of size $N \times N$ and the value at each position $(i, j) \in E$ of this matrix indicates the connection between nodes i and j . For any two nodes i and j the distance $d(i, j)$ between them is defined as the length of the shortest path from i to j , if the nodes are connected and called neighbors, whereas $d(i, j) = \infty$ otherwise. The adjacency matrix A introduces simple connections, and in case of directed connections of network is a square non-symmetric matrix $A = \{a_{ij}\}$, $i, j \in N$ of size $|N| \times |N|$, having entries (i, j) . The adjacency matrix A takes a value of one (1) if there is a connection between nodes i and j and zero (0) otherwise. On a network with weighted and directed connections the weighted matrix W is also a square non-symmetric matrix $W = \{w_{ij}\}$, $i, j \in N$ of size $|N| * |N|$, where w_{ij} is the weight which characterizes the weighted connection.

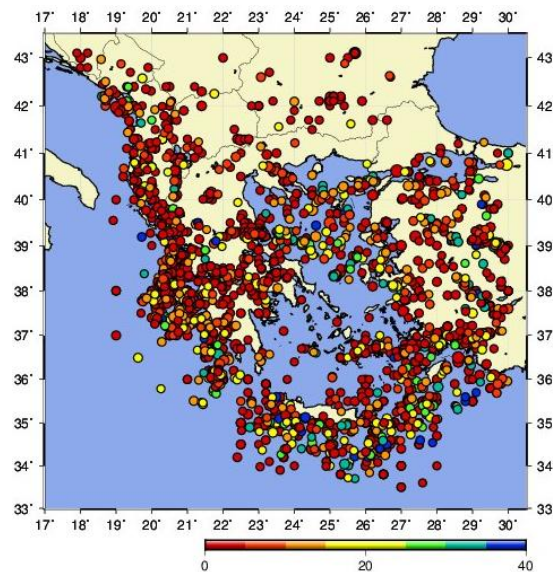


Figure 1 - Epicentral distribution of $M \geq 5.0$ that occurred in 1911-2014 in the broader area of Greece. The diameter of the circles is proportional to the earthquake magnitude and the color proportional to the focal depth as given by a color scale.

For the construction of networks firstly we divide the study area into subareas, the seismic zones, that are homogenous from the seismotectonic point of view (faulting type, seismic moment release) and taking into account previous publications (Papazachos *et al.*, 1998; Papaioannou and Papazachos, 2000). The N seismic zones represent the N nodes of the seismic network. Two successive earthquakes define a connection between the respective nodes. If two successive earthquakes occur inside the same zone, they form a loop. The connections and loops represent the correlations between two successive earthquakes.

The first approach, for introducing connections, constructs the seismic network with simple connections, i.e. the adjacency matrix A contains the value one (1) or zero (0) for each pair (i, j) of nodes. This approach, called Binary, has the disadvantage that full exploitation of data is not achieved, because the multiple successions of earthquakes are not recognized but it is considered as a simple one.

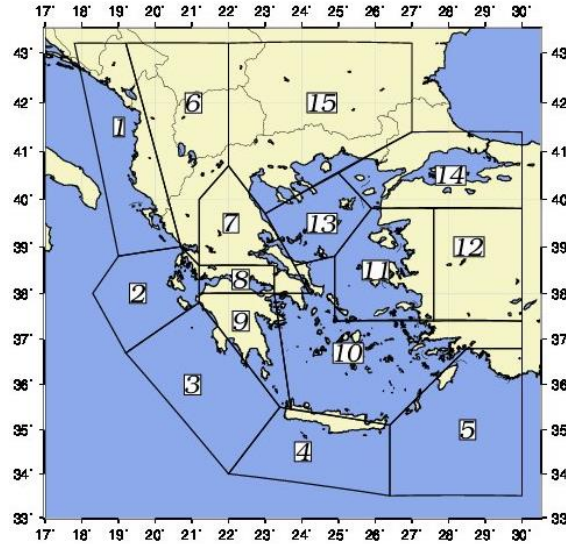


Figure 2 - The 15 seismic zones in which the study area is divided and which represent the network nodes.

Five approaches are now introduced to constructing weighted connections in space $[0,1]$. The construction process of the weighted matrix W is as follows: For each time window a square non-symmetric transition matrix $S = \{s_{ij}\}$ is created, where s_{ij} indicates how many times the succession of earthquakes appears for each pair (i, j) of nodes. Alternatively it is a positive real number that indicates the value of seismic moment M_0 which is released between nodes i and j . Having created such square matrices, which are equal in number with the time windows, we define the weighted matrix W for each time window as $W = S/\max\{s_{ij}\}$, where $\max\{s_{ij}\}$ is the maximum value of all matrices S . This approach, called Wallwin, has as variation the definition of the weighted matrix W at each time window as $W = S/\max\{s_{ij}\}$, where now $\max\{s_{ij}\}$ is the maximum value of matrix S for each time window. This approach is called Weachwin. The fourth approach, the third for weighted connections, is based on probability frequency distribution for the transition matrix S , and is called Wprob. A probability frequency distribution is a way to show the eventual occurrence frequency at each pair (i, j) of nodes. A probability distribution assigns a probability to each pair (i, j) of matrix as $W = \{s_{ij}\} / \sum_{i,j=1}^N \{s_{ij}\}$. Thus, the sum of elements of weighted matrix W is equal to the unity.

A completely different approach for the construction of the weighted matrix W is based on the Markov chain. Consider $V = \{1, 2, \dots, N\}$ the state space of a Markov chain, i.e. the space V consists of all nodes. Let us define X_t ($t = 1, 2, \dots, n$) as the Markov chain which was formed by the time succession of two earthquakes of any seismic zones with values from the state space V . Firstly, the transition matrix $S = \{s_{ij}\}$ is formed as in the previous approaches. We define the probabilities $\{p_{ij}\}$, that form the square non-symmetric transition probability matrix P because we have directed connections, as follows: $P = \{p_{ij}\} = \frac{\{s_{ij}\}}{\sum_{j=1}^N \{s_{ij}\}}$ with $i, j = 1, 2, \dots, N$. The main characteristic of the transition probability matrix P is that the sum of each row is unity, i.e. $\sum_{j=1}^N p_{ij} = 1$. The matrix P is the

weighted matrix and this approach is called Markov. The last approach is based on Gibbs sequence (Casella and George, 1992). Starting with random variables $X = \{X'_0, X'_1, \dots, X'_N\}$ and $Y = \{Y'_0, Y'_1, \dots, Y'_N\}$, the matrices $P_{X|Y}$ and $P_{Y|X}$ thought of as transition matrices the probabilities of getting to X states from Y states and vice versa are given. If we are only interested in generating the marginal distribution of X , we are mainly concerned with the X' sequence. To go from $X'_0 \rightarrow X'_1$ we have to go through Y'_1 , so the iteration sequence is $X'_0 \rightarrow Y'_1 \rightarrow X'_1$ and $X'_0 \rightarrow X'_1$ forms a Markov chain with transition probability $P(X'_1 \rightarrow x_1 | X'_0 \rightarrow x_0) = \sum_y P(X'_1 = x_1 | Y'_1 = y) * P(Y'_1 = y | X'_0 = x_0)$. Thus, the transition probability matrix of the X' sequence is given by $P_{X|X} = P_{Y|X} * P_{X|Y}$, and is a weighted matrix. This approach is called Gibbs.

The value of each network measure gives different information on the network structure (Table 1). The network properties are quantified with a number of characteristics (indices) computed on A and W or P .

Table 1 - Information provided from each network measure.

Network measures (for weighted connections)	Information
Degree/Strength: $K^W = \frac{1}{N} \sum_{i \in N} k_i^W = \sum_{j \in N} \{w_{ij}\}$	One of the most basic statistics that characterize the graphs, measuring the connections in each node.
Clustering coefficient: $C^W = \frac{1}{N} \sum_{i \in N} \frac{\sum_{j,h \in N} w_{ij} w_{ih} w_{jh}}{k_i(k_i - 1)}$	Probability that two nodes connected to a common node to be interconnected. It shows the tendency of the nodes to group by triads.
Global efficiency: $E^W = \frac{1}{N} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} (d_{ij}^W)^{-1}}{N - 1}$	The small distances between nodes give higher efficiency of the information flow on the network.
Eigenvector centrality: $X_i = \lambda_i^{-1} \sum_{j \in N} w_{ij} x_j,$ λ the largest eigenvalue of the solution of equation $WX = \lambda X$.	The nodes with many neighbors have high value and contribute more to the flow of information in network, i.e. the connections linking high degree nodes contribute to the high value of the node.
Pagerank: $P^W(n_i) = \frac{1-d}{N} + d \sum_{n_j \in N(n_i)} \frac{P^W(n_j)}{outdeg(n_j)}$	Calculate the rank value that indicates the importance of a node. A node that is connected with many nodes with high Pagerank receives a high rank itself. *outdeg, in type, is the number of outgoing connections.

3. Results and Discussion

The study area, i.e. the broader area of Greece, is divided into 15 seismic zones which represent the nodes of the seismic network and the connections between them are given by the time succession of two earthquakes of any seismic zones. After forming the seismic network, with the six approaches for each of the sliding time window, the changes in the structure of networks are examined per approach as for example during the time window 1971-1980 when the width of the sliding window is 10 years (Fig. 3). The different structure of network, is shown by the different values of network measures for each approach (Table 2).

(a)

(b)

(c)

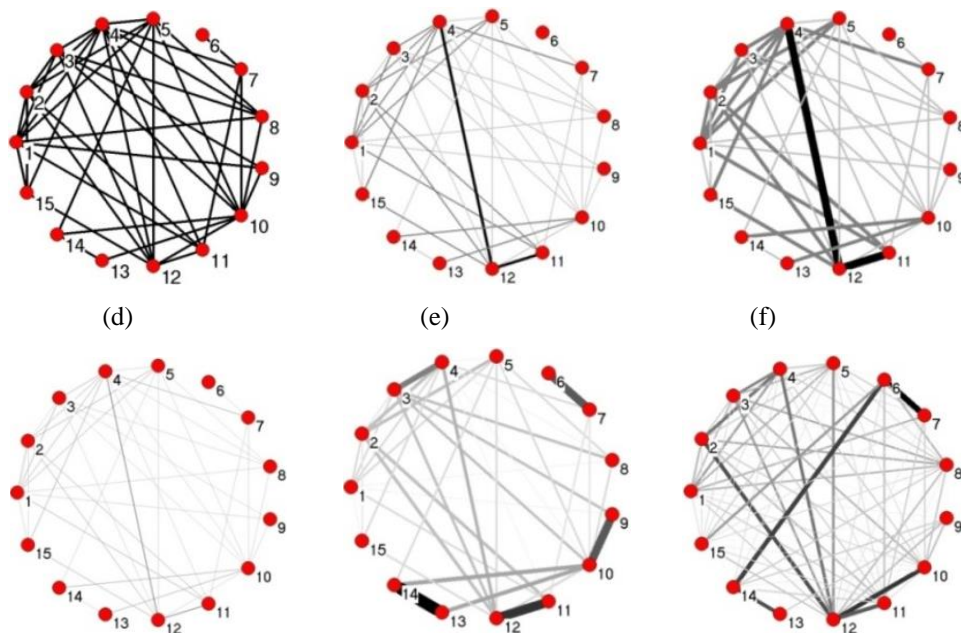


Figure 3 - The structure of the seismic network when as a seismic measure the number of earthquakes is considered during the time window 1971-1980 with the six approaches a) Binary, b) Wallwin, c) Weachwin, d) Wprob, e) Markov and f) Gibbs.

After the investigation of the structure of networks we study the seismic activity which is observed for each seismic zone during 1911-2014 using sliding non-overlapping time windows of 2, 3, 5 and 10 years duration as for example for the period 1971-1980 which is shown in Figure 4.

Table 2 - The values of the network measures, when as a seismic measure the number of earthquakes is considered, for all approaches during the time window 1971-1980.

Approach	Binary	Wallwin	Weachwin	Wprob	Markov	Gibbs
Degree/Strength	6.4	0.39	1.01	0.06	1	1
Clustering coefficient	0.73	0.04	0.10	0.01	0.11	0.06
Global efficiency	0.69	0.03	0.08	0.04	0.09	0.06
Eigenvector centrality	0.23	0.20	0.20	0.20	0.25	0.25
Pagerank	0.15	0.07	0.07	0.03	0.10	0.09

To perform the ranking, we are interested in the first position for each time window when the number of earthquakes and the seismic moment release in each seismic zone is taken into account (Fig. 5). The ranking of the seismic zones in many cases, based on the above seismic measures, is different because the number of earthquakes as a seismic measure does not express the amount of energy release unlike the measure when the seismic moment is considered.

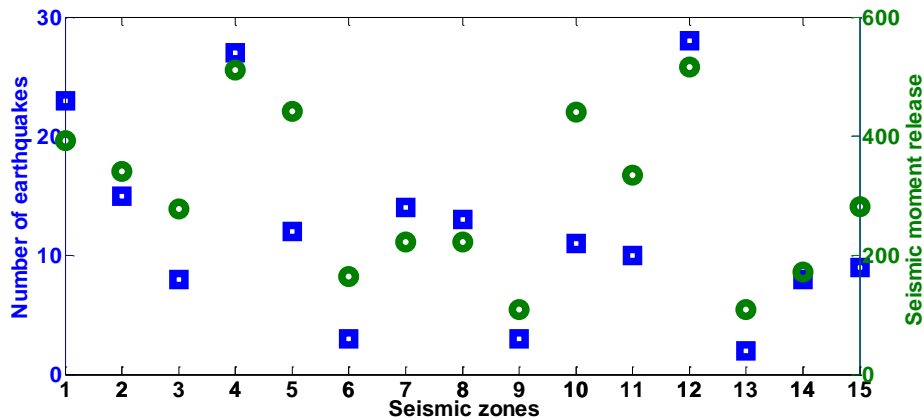


Figure 4 - The seismic activity of the seismic zones (x – axis) during the time window 1971-1980, with seismic measure the number of earthquakes (y – axis, squares) and the seismic moment (y – axis in a logarithmic scale, cycles).

Using five network measures, for weighted or simple connections the values of which were calculated for each seismic zone in each approach, a second ranking is generated as regards the first position for each time window of the seismic zones on the network measures.

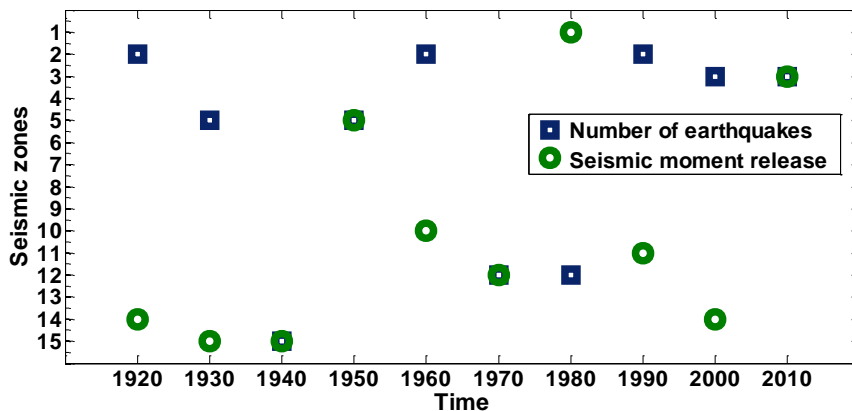


Figure 5 - The first position for each time window (x – axis) when the sliding window lasts 10 years, for the larger earthquake number and seismic moment release (y – axis).

Having the two ranking lists for the first position for each time window, when the first list results from the observations and the second one from the network measures, the agreement of the two ranking lists is investigated (Fig. 6). The results of the Binary approach (Fig. 6, the five first network measures) show the ineffective way to construct the seismic network because the success rates do not exceed 60% in any of the seismic and network measure. The Wallwin, Weachwin and Wprob approaches show the best results, if we exclude the almost expected 100% success rate with network measure the degree and seismic measure the seismic moment. This is due to the method of introducing connections of the network, because by definition the seismic zone with the higher seismic activity has the most connections. The rates of success are over 80% with network measure the eigenvector centrality and pagerank when the seismic measure is the number of events. Particularly low rates are derived from the approach of Markov and Gibbs for all network measures when as seismic measure the seismic moment release is considered. When the number of events is considered the success rates for the Gibbs approach is very high particularly when the network measure is the degree.

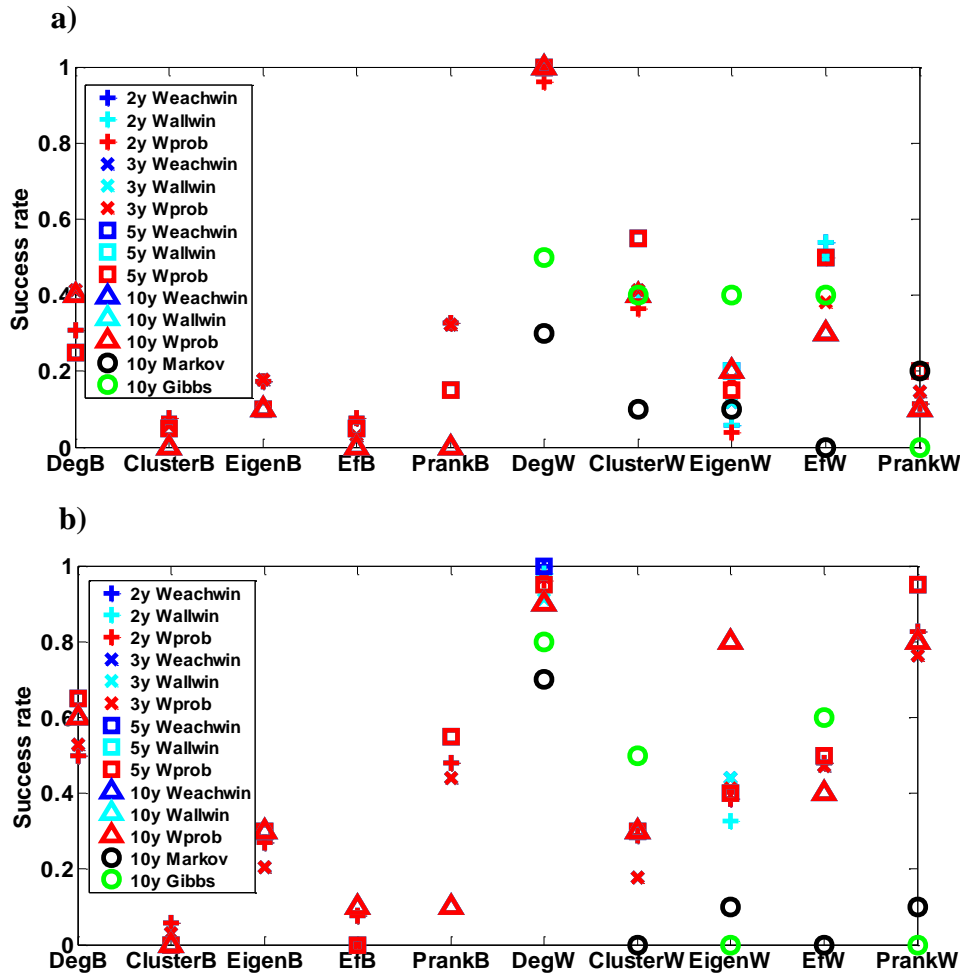


Figure 6 - Success rate (y – axis, agreement between the two ranking lists for the first position) with seismic measure a) the seismic moment release and b) the number of earthquakes based on the network measures (x – axis, the first five measures with the suffix B next to the abbreviation of the name of the measures for simple connections and the next five for weighted connections with the suffix W).

The satisfactory results from Gibbs and Markov approaches, when the measure is the number of events and the network measure is the degree, encourage the effort to estimate the predicted transition probability matrix for the next time window in order to assess the future seismic hazard. The estimation to predicted transition probability matrix is performed with Gibbs and Markov approaches because they use the stochastic matrix P as weighted matrix for introducing the connections of network.

3.1 Statistical test

The elements of the a priori known transition propability matrix P are obtained as described in subsection 2.1 about the approach of Markov and Gibbs, respectively. We test, for every time window, if the elements of the transition probability matrix P which is constructed with the approach of Gibbs and Markov, respectively, agree with the values of the elements of the predicted transition probability matrix P^* . To estimate the predicted matrix P^* we create the square non-symmetric matrix P_S which is formed from the transition matrix S . The transition matrix S is calculated from

all historical data without dividing them in time windows and following the known process of transformation to the stochastic matrix for each approach (described in subsection 2.1), and the transition probability matrix P_S is generated. According to Markov process, the predicted transition probability matrix P^* in different periods is formed by $P_{t+1}^* = P_t * P_S$, where $t=1, 2, \dots, n$ (Zhang and Zhang, 2009). Suppose that the null hypothesis $H_0: P = P^*$, in case of testing all transitions together then a statistical criterion for testing the hypothesis can be based on the fact that the test statistic $\sum_{i=1}^N \sum_{j=1}^N \frac{n_i(P_{ij}^* - P_{ij})^2}{P_{ij}}$ follows the χ^2 distribution with $N(N - 1) - a$ degrees of freedom, where N is the number of states (nodes), a the number of zero elements in matrix P , n_i the number of transitions of the row i in transition matrix S and the sums are for those values of i and j for which $P_{ij} > 0$. The statistical test shows the acceptance of the null hypothesis H_0 in many cases (Fig. 7, when dashed line is above the continuous line for the same time window) only for the Gibbs approach. This happened because the Gibbs approach provides higher success rates, concerning the agreement between of the two ranking lists, than the Markov approach. The results of the statistical test show that network approach is a useful tool for future seismic hazard assessment.

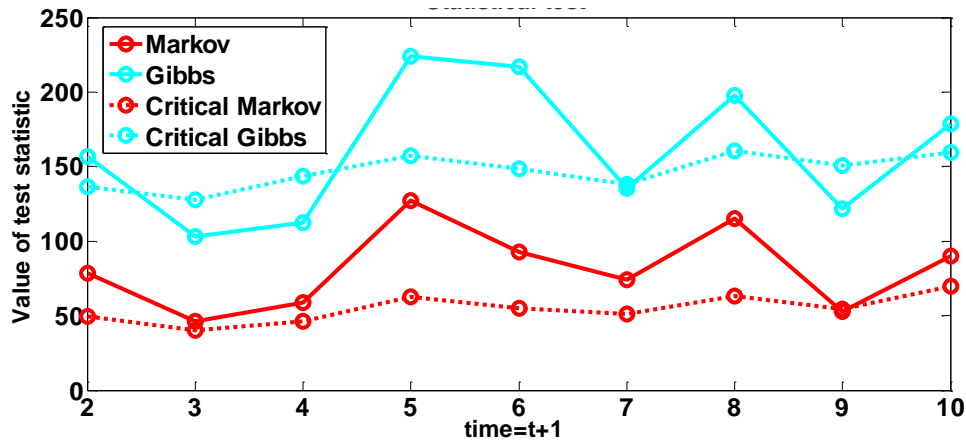


Figure 7 - The value of test statistic (y – axis) as a function of time for 10-years sliding windows (x – axis). The dashed (critical) lines are defined from statistical tables for significance level $\alpha = 0.05$.

4. Concluding Remarks

These first results suggest the contribution of network analysis in deriving from seismicity properties of a seismically active area which is divided in seismic zones. The approach of Gibbs estimates reliably, in some cases, the future seismic activity as evidenced by the statistical test unlike the Markov approach where we have rejections of null hypothesis H_0 . The Binary approach is the most unsuitable for introducing connections unlike with the other three approaches, Wallwin, Weachwin and Wprob. These approaches show high success rates with network measure the eigenvector centrality and pagerank, and excluding the degree, when the seismic measure is the number of earthquakes. Generally, the results are better when the seismic measure is the number of earthquakes than the seismic moment. In addition, the network index of degree can be regarded as an index of the level of seismicity and may be a useful tool in the study of earthquake networks.

The network approach has been found to be a powerful tool that contributes significantly to investigate properties of complex phenomena such as seismic activity. The construction of the seismic network with new seismic measures, the use of approach through the time series for introducing the network connections, the introduction of new network measures for examining the network structure can be significantly contribute to seismic hazard assessment. An open issue arising

from this work is the application of the same analysis, which is presented in this study, with the above-mentioned settings.

5. References

- Abe, S. and Suzuki, N., 2009. Scaling relation for earthquake networks, *Physica A*, 388, 2511-2514.
- Abe, S. and Suzuki, N., 2004. Scale-free network of earthquakes, *Europhys Letters*, 65, 581-586.
- Albert, R. and Barabasi, A.L., 2002. Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74, 47-97.
- Casella, G. and George, E., 1992. Explaining the Gibbs Sampler, *The American Statistician*, 46(3), 167-174.
- Kanamori, H. and Anderson, L., 1975. Theoretical basis of some empirical relations in seismology, *Bulletin of the Seismological Society of America*, 65(5), 1073-1095.
- Papaoannou, Ch.A. and Papazachos, B.C., 2000. Time-independent and time - dependent seismic hazard in Greece based on seismogenic sources, *Bulletin Seismology Society of America*, 90, 22-33.
- Papazachos, B.C., Papadimitriou, E.E., Kiratzi, A.A., Papazachos, C.B. and Louvari, E.K., 1998. Fault plane solutions in the Aegean Sea and the surrounding area and their tectonic implication, *Bollettino di geofisica teorica e applicata*, 39, 199-218.
- Zhang, D. and Zhang, X., 2009. Study on Forecasting the Stock Market Trend Based on Stochastic Analysis Method, *International Journal of Business and Management*, 4(6), 163-170.