The Small World of Seismic Events

Douglas S.R. Ferreira, Andrés R.R. Papa, and Ronaldo Menezes

Abstract. The understanding of long-distance relations between seismic activities has for long been of interest to seismologists and geologists. In this paper we have used data from the world-wide earthquake catalog for the period between 1972 and 2011, to generate a network of sites around the world for earthquakes with magnitude $m \geq 4.5$ on the Richter scale. After the network construction, we have analyzed the results under two viewpoints. Firstly, in contrast to previous works, which have considered just small areas, we showed that the best fitting for networks of seismic events is not a pure power law, but a power law with an exponential cutoff. We also have found that the global network presents small-world properties. The implications of our results are discussed.

1 Introduction

The general belief in seismic theory is that the relationship between events that are located far apart is hard to be understood. However we live today in a world where data is being collected on most aspects of our lives and better yet, computer power is cheaply available for analyzing the data. The work on seismic data analysis is no

Douglas S.R. Ferreira
Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro, Brazil
Observatório Nacional, Rio de Janeiro, Brazil
e-mail: douglas.ferreira@ifrj.edu.br

Andrés R.R. Papa
Geophysics Department, Observatório Nacional, Rio de Janeiro, Brazil
Instituto de Física, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: papa@on.br

Ronaldo Menezes
BioComplex Laboratory, Department of Computer Sciences,
Florida Institute of Technology, Melbourne, Florida, USA
e-mail: rmenezes@cs.fit.edu

P. Contucci et al. (eds.), *Complex Networks V*,
Studies in Computational Intelligence 549,
DOI: 10.1007/978-3-319-05401-8_10, © Springer International Publishing Switzerland 2014

different; we have now large collections of millions of seismic events from around the world which deserves analysis. In this paper we have found some evidence that point to small-world characteristics in the existent data on seismic events. An event in a particular geographical site appears to be related to many other sites around the world and not only to other events at nearby sites.

Since the work from Barabási and Albert [5] researchers have turned their attention not on mining the data itself but rather organizing the data in a network which captures relationships between pieces of data and only then mining the network structure and hence the relations between pieces of data. The network can review information hard to see from mining the raw pieces of data. The use of networks as a framework for the understanding of natural phenomena is nowadays called *Network Science*.

Through the analysis of a model using successive earthquakes, recent studies [1, 2] have applied concepts of complex networks to study the relationship between seismic events. In these studies, networks of geographical sites are constructed by choosing a region of the world (e.g. Iran, California) and its respective earthquake catalog. The region is then divided into small cubic cells, where a cell will become a node of the network if an earthquake occurred therein. Two different cells will be connected by a directed edge when two successive earthquakes occurred in these respective cells. If two earthquakes occur in the same cell we have a loop, i.e., the cell is connected to itself. This method of describing the complexity of seismic phenomena has found that, at least for some regions, the common features of complex networks (e.g. scale-free, small-world) are present. However, in spite of the importance of the results which show that seismic networks for some specific regions present small-world effects, these results are somewhat expected since it makes sense for areas located geographically close to each other to be correlated.

In this paper we have used data from the world-wide earthquake catalog for the period between 1972 and 2011, to generate a network of sites around the world. Since only seismic events with $m \geq 4.5$ are recorded for all locations around the world, we then consider them *significant events* and used this set in our analysis.

2 Theoretical Background

2.1 Complex Networks Features

Scale-free networks are defined as those in which the degree distribution of nodes (or vertices) follows a power law, that is, the probability that a network will have nodes of degree k, denoted by $P(k)$ is given by $P(k) \sim k^{-\gamma}$, where γ is a positive constant. This equation states that scale-free networks have a very small number of highly-connected nodes (called hubs) and a large number of nodes with low connectivity. These networks exist in contrast with general random networks with a very large number of nodes in which the probability distribution follows a Poisson distribution. Random networks have nice properties but the truth of the matter is that most real networks are not random.

One of the best approaches for defining Small-world Networks is based on the work of Watts [13] which states that in small-world networks, every node is "close" to every other node in the network. It is generally agreed that "close" refers to the average path length in the network, ℓ, which has the same order of magnitude as the logarithm of the number of nodes, i.e., $\ell \sim \ln N$. In addition, and what makes small-world networks even more interesting, is the fact that these networks have a high degree of clustering representing a transitivity in the relation of nodes; if a node i has two connections, the theory argues that the two connections are also likely to "know" each other. More formally, the clustering coefficient, C_i, of that node is given by:

$$C_i = \frac{\triangle(i)}{\triangle_{all}(i)} \tag{1}$$

where, $\triangle(i)$ is the number of the directed triangles formed by i with its neighbors and $\triangle_{all}(i)$ is the number of all possible triangles that i could form with its neighbors; the clustering coefficient of the entire network, C, is just the average of all C_i over the number of nodes in the network, N. In random networks the clustering coefficient can be estimated using the closed form $C_{\text{rand}} = \langle k \rangle / N$, where $\langle k \rangle$ is the average degree in the random network.

In the context of networks of seismic events, if it contains hubs, one can argue that the distribution of earthquakes should also follow a power law. On the other hand, if the network has small-world properties one can argue that there is some indication of long-range relations between far-apart earthquake sites.

3 A Geographical Network from Seismic Events

The use of networks to understand phenomena associated with geographical locations has been used in many instances in science including diseases [9], scientific collaborations [8, 10], and organ transplantation [12] to mention just a few. Seismic activity is intrinsically linked to geography because todays instruments can pinpoint with great accuracy the location in the globe where each seismic event take place.

It is important to precisely locate the geographical location of a seismic event but if we want to understand relations between events we should concentrate on creating a network in which locations are linked based on an acceptable criteria. In this paper we use the same procedure employed by [1] in their studies of earthquakes in specific regions of the world. The construction of the network is as follows. We first have to decide on what should represent the nodes. Obviously our first choice are the sites where the earthquake took place. The problem of doing this is that an earthquake epicenter is rarely located exactly in the same location and given the accuracy of today's instruments we would have an infinitely large number of possible sites. We decided instead to define nodes representing a larger region of the world we here call it cell. A cell will become a node of the network if an earthquake has its epicenter therein. The creation of edges follows a temporal order of seismic events. For instance, if an earthquake occurs in a cell C_1 and the next earthquake in a cell C_2, we assume a relation between C_1 and C_2 and we represent the event by a directed edge in the network. The process continues linking cells according to the temporal order.

The degree of each node (the total number of its connections) is not affected by the direction of the network. The nature of the way the network is constructed means that for each node in the network, its in-degree is equal to its out-degree (the exceptions are only the first and last sites in the sequence of seismic events but for all practical purposes we can disregard this small difference).

Although the use of temporal ordering of events is not new in our paper, there are two main differences between our study and others. Firstly and most importantly, the region considered in our investigation is the entire globe, instead of just some specific geographical subarea of the globe; this is, to the best our knowledge, the first worldwide study of seismic events using networks and consequently the first one to investigate the possibility of long-range links between seismic events. Secondly, we have used a two dimensional model in which the depth dimension of the earthquake epicenter is not considered, since we are interested in looking for spatial connections between different regions around the world and besides 82% of the earthquakes, in our dataset, have their hypocenters in a depth less than or equal to 100 km.

Before we divide the globe into cells, we need to choose the size of such cells particularly because we are dealing with the entire globe; if the cells are too small we will not have any useful information in the network, if the cells are large we lose information due to the grouping of events into a single network node. There are no rules to define the sizes. Therefore we have taken three different sizes, the same sizes used in previous studies [1, 11], where the authors conducted studies about earthquake networks using data from California, Chile and Japan. The quadratic cells have, 5 km $\times$ 5 km, 10 km $\times$ 10 km and 20 km $\times$ 20 km. To set up cells around the globe, we have used the latitude and longitude coordinates of each epicenter in relation to the origin of the coordinates, i.e., where both latitude and longitude are equal to zero (we have chosen the referential at the origin for simplicity). So, if a seismic event occurs with epicenter E with location (θ_E, ϕ_E), where θ_E and ϕ_E are the values of latitude and longitude in radians of the epicenter, we are able to find the distances north–south and east–west between this point and the origin. These distances can be calculated, considering the spherical approximation for the Earth, by:

$$\begin{aligned} S_E^{ns} &= R.\theta_E \\ S_E^{ew} &= R.\phi_E.\cos\theta_E, \end{aligned} \tag{2}$$

where S_E^{ns} and S_E^{ew} are, the north-south and east-west distances for the earthquake E, respectively, and R is the Earth radius, considered equal to 6.371×10^3 km. With this computation we can identify the cell in the lattice for each event using the values of S_E^{ns} and S_E^{ew}.

Note that the distances between different cells are irrelevant for the present part of our study. For now we are just interested in the connectivity of nodes. However, from the sequence there are important consequences to be obtained which we present below.

The seismic data used to build our network was taken from the Global Earthquake Catalog, provided by Advanced National Seismic System[1], which records events from the entire globe. The data spans all seismic events between the period from January 1, 1972 to December 31, 2011. This catalog has a limitation because it is not consistent in all regions of the world; it includes events of all magnitudes for the United States of America but only events with $m \geq 4.5$ (in the Richter scale) for the rest of the world (unless they received specific information that the event was felt or caused damage). Therefore, in order to obtain a more homogeneous distribution of data through the world, we have analyzed only events with $m \geq 4.5$ and we also excluded data that represent artificial seismic events ("quarry blasts"). In the end, we were left with 185 747 events, where 82% of them happen near the surface of the world (depth $\leq 100\,\mathrm{km}$).

4 Results

Given the network build as described in the previous section, we have performed a few experiments to understand this structure. Following the procedure explained earlier, the 185 747 events yielded three different networks depending on the size used for the cells: $20\,\mathrm{km} \times 20\,\mathrm{km}$ with 65 355 nodes, $10\,\mathrm{km} \times 10\,\mathrm{km}$ with 104 516 nodes, and $5\,\mathrm{km} \times 5\,\mathrm{km}$ with 144 974 nodes.

4.1 Scale-Free Property of the Seismic Network

It has been shown recently [3, 11] that seismic networks for specific regions of the globe (e.g. California) appear to have scale-free properties, or in other words that the construction of the network employs preferential attachment as described by [5] insofar that a node added to the network has a higher probability to be connected to an existing node that already has a large number of connections. This is somewhat trivial to understand because active sites in the world will tend to appear in the temporal sequence of seismic events many times. The preferential attachment states that the probability P that a new node i will be linked to an existing node j, depends on the degree $\deg(j)$ of the node j, that is, $P(i \to j) = \deg(j)/\sum_u \deg(u)$. This rule generates a scale-free behavior whose connectivity distribution follows a power-law with a negative exponent as shown in Section 2.1.

In [3, 11], earthquake networks were built for some specific regions (California, Chile and Japan), and their connectivity distributions were found to follow power-laws. However, if we look carefully to the connectivity distribution and plot its cumulative probability, instead of its probability density, we can observe that the power-law distributions that emerge from these network are truncated. According to [4], there are at least two classes of factors that may affect the preferential attachment and consequently the scale-free degree distribution: the aging of the nodes and the cost of adding links to the nodes (or the limited capacity of a node). The

[1] `http://quake.geo.berkeley.edu/anss`

aging effect means that even a highly connected node may, eventually, stop receiving new links. The presence of an aging-like effect in our work could be expected from the fact that relaxation times for tectonics are much longer that the time interval under study. Some cells can stop of receiving new connections during a period of time comparable to our own time window by a temporal quiet period due to a transitory stress accumulation. The second factor that affects the preferential attachment occurs when the number of possible links attaching to a node is limited by physical factors or when this node has, for any reason, a limited capacity to receive connections, like in a network of world airports. We have not found a suitable parallel to this factor in the case of earthquakes. When any of these factors is present, the distribution is better represented with a power law with an exponential cutoff, $P(k) \sim k^{-\alpha} e^{-k/k_c}$, where α and k_c are constants.

In Fig. 1, we plot the cumulative probability distribution for the earthquake network built for the Southern California (32°N – 37°N and 114°W – 122°W), using the data catalog provided by Advanced National Seismic System, where we considered all seisms with magnitude $m > 0$ for the period between January 1, 2002 and December 31, 2011. The total number of events is 147 435. It is possible to observe in this plot that the data is better fitted to a power-law with exponential cutoff than a pure power law which is a good fit only for small values of k with an exponent $\gamma - 1 = 0.513$, which is consistent with the value $\gamma = 1.5$ reported in [3] for the probability density function. These results apply for a network built using cell sizes of 5 km × 5 km. It is noteworthy that in a probability density plot, the cutoff does not seem to exist, because the fluctuations are higher than in a cumulative probability plot.

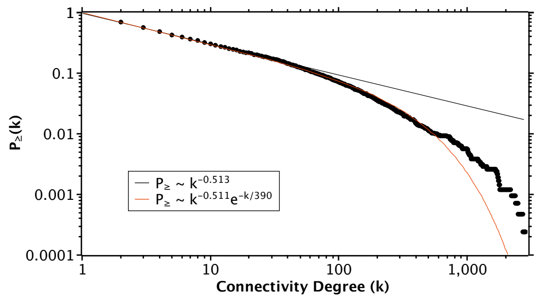

Fig. 1 Cumulative probability distribution of connectivity for the earthquake network in California using cell size 5 km × 5 km. The solid lines represent two possible fittings: a power-law (black) and a power-law with exponential cutoff (red). There are 4 187 nodes in this network.

Looking at the world earthquake network constructed using the data from the Global Earthquake Catalog, we note that the aging-cost effect are visibly stronger in the connectivity distribution; the exponential cutoff is clearly visible in both the degree distribution and the cumulative degree distribution presented in Fig. 2.

Fig. 2(left) represents the connectivity distribution for the global networks using the three different cell sizes for the global lattice. It is interesting to note that, comparing these plots, we observe that the behavior is the same in all three cases (in the sense that they present a power law with an exponential cutoff), which indicates that the cell size does not change the complex features behind the global seismic phenomena.

In Fig. 2(right), we have the same plot of Fig. 2(left), but using the cumulative probability only for cell size 20 km × 20 km. Note that the cumulative probability plot for the global network shows the same exponential cutoff behavior than for local network, as shown in Fig. 1.

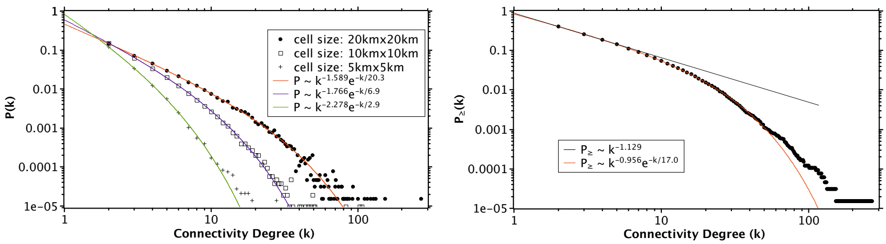

Fig. 2 Connectivity distributions in the global earthquake network. Plot for the cell sizes 20 km × 20 km (solid circles), 10 km × 10 km (squares) and 5 km × 5 km (cross), where the solid lines represent the best fit using power-law with exponential cutoff (on the left). Cumulative probability for the cell size 20 km × 20 km. The solid lines represent a standard power-law (black) and a power-law with exponential cutoff (red) (on the right).

4.2 *Small-World Property of the Global Seismic Network*

Small-world networks [13] have the general characteristic that they contain groups of near-cliques (dense areas of connectivity) but long jumps between these areas. These two properties lead to a network in which the *average shortest path* is very small and the *clustering coefficient* very high. It is important to note that the term *average shortest path* does not refer to a spatial distance but the number of "steps" on the network to move from a node to another.

Here we would like to test if the global seismic network has small-world properties. The consequence of such a finding would be an indicative that seismic events around the world are correlated and not independent. To study these properties we need to introduce slight changes to our original network. The first is that the loops have to be removed, since we are looking for correlations between nodes and it only makes sense when these nodes are different. The second change is, that we move from a network with multi-graph characteristics to a weighted network. That is, if two nodes are linked by w edges in the original network, they will be linked by a single edge with weight w in the new version of the network.

We have analyzed the seismic network for the entire world under two viewpoints: directed and undirected. The cell size used in this construction was 20 km × 20 km. The data were the same used to construct the Fig. 2. Table 1 shows the results obtained for the clustering coefficient (C) [6] and the average path length (ℓ) [7].

Table 1 Results for the clustering coefficient (C) compared to the clustering coefficient of a random network of the same size (C_{rand}) and the average path length (ℓ) compared to the $\ln N$, where N is 65 355 in network with cell size 20 km × 20 km

Network	C	C_{rand}	ℓ	$\ln N$
Directed	7.0×10^{-3}	4.2×10^{-5}	17.19	11.08
Undirected	4.2×10^{-2}	4.2×10^{-5}	12.24	11.08

From Table 1, we note that both versions of the earthquake network has small-world properties; the clustering coefficient is much higher than an equivalent for a random network, and the average path length has the same order of magnitude as the logarithm of the number of nodes. It is worth noticing that the regional earthquake networks built for California, Japan and Chile also are small-world [3, 11] although the significance of small world at the global level is higher because with these world-wide results we have an indicative of long-range relations between different places around the world.

5 Conclusions

The use of networks to model and study relationships between seismic events has been used in the past for small areas of the globe. Here we demonstrate that similar techniques could also be used at the global level. More importantly, many of the techniques used in complex network analysis were used here to show that there seem to exist long-distance relations between seismic events.

We argued in favor of the long-distance relation hypothesis by showing that the network has small-world characteristics. Given the small-world characteristics of high clustering and low average path length, we were able to argue that seisms around the world appear not to be independent of each other.

Another interesting approach we intend to do in the future relates to using community analysis or community detection to understand how seismic locations are grouped.

Acknowledgements. Douglas S. R. Ferreira thanks the Capes Foundation, Ministry of Education of Brazil, for the scholarship under process BEX 13748/12-2. Andrés R. R. Papa thanks CNPq (Brazilian Science Foundation) for his productivity fellowship.

References

1. Abe, S., Suzuki, N.: Scale-free network of earthquakes. Europhys. Lett. 65(4), 581–586 (2004)
2. Abe, S., Suzuki, N.: Small-world structure of earthquake network. Physica A 337(1), 357–362 (2004)

3. Abe, S., Suzuki, N.: Complex-network description of seismicity. Nonlin. Processes Geophys. 13(2), 145–150 (2006)
4. Amaral, L.A., Scala, A., Barthelemy, M., Stanley, H.E.: Classes of small-world networks. Proc. Natl. Acad. Sci. 97(21), 11149–11152 (2000)
5. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286(5439), 509–512 (1999)
6. Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101(11), 3747–3752 (2004)
7. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25(2), 163–177 (2001)
8. Divakarmurthy, P., Biswas, P., Menezes, R.: A temporal analysis of geographical distances in computer science collaborations. In: IEEE Third International Conference on Privacy, Security, Risk and Trust (Passat) and 2011 IEEE Third International Conference on Social Computing (SocialCom), pp. 657–660. IEEE (2011)
9. Newman, M.E.J.: Spread of epidemic disease on networks. Physical Review E 66(1), 016128 (2002)
10. Pan, R.K., Kaski, K., Fortunato, S.: World citation and collaboration networks: uncovering the role of geography in science. Scientific Reports 2 (2012)
11. Pasten, D., Abe, S., Munoz, V., Suzuki, N.: Scale-free and small-world properties of earthquake network in chile. arXiv preprint arXiv:1005.5548 (2010)
12. Venugopal, S., Stoner, E., Cadeiras, M., Menezes, R.: Understanding organ transplantation in the usa using geographical social networks. Social Network Analysis and Mining, 1–17 (2012)
13. Watts, D.J.: Networks, dynamics, and the small-world phenomenon. American Journal of Sociology 105(2), 493–527 (1999)