A Network-Based Approach to Earthquake Pattern Analysis

By

Mirela Suteanu

A Thesis Submitted to Saint Mary's University, Halifax, Nova Scotia,
in Partial Fulfillment of the Requirements for the
Degree of Masters of Science in Applied Science.

May 23, 2014, Halifax, Nova Scotia

Approved:	Dr. Pawan Lingras	Supervisor Department of Mathematics and Computing Science
Approved:	Dr. Robert Shcherbakov	External Examiner Department of Earth Sciences Western University
Approved:	Dr. Sageev Oore	Supervisory Committee Member Department of Mathematics and Computing Science
Approved:	Dr. Danika van Proosdij	Supervisory Committee Member Department of Geography
Approved:	Dr. Diane Crocker	Chair Associate Dean of Graduate Studies

Date: May 23, 2014

Acknowledgments

I would like to thank Dr. Pawan Lingras for his continuous and vigorous support over the years, for his patience, for always finding the best approaches, for giving me the freedom to explore, yet firmly guiding me towards the completion of the program. I would like to thank Dr. Danika van Proosdij, for her strong support during this extended period of time; not only did she find practical solutions when I needed them, but she also found ways to sympathetically show me that the final accomplishment was within reach when the struggle was hard. I would also like to thank Dr. Sageev Oore, for his support and understanding over time.

Furthermore, I would like to thank my mother, without whose help I could not have dedicated the required time to this project. Last, but not least, I would like to thank my husband, Cristian, and my daughter, Maria Cristina for all their support and encouragement.

A Network-Based Approach to
Earthquake Pattern Analysis

By Mirela Suteanu

Abstract

This research proposes a new network-based method for the assessment of earthquake relationships in space-time-magnitude patterns. The method is applied to the study of volcanic seismicity in Hawaii, over a time period from January 1st, 1989 to December 31st, 2012. It is shown that networks with high values of the minimum edge weight W_{min} enjoy strong scaling properties, as opposed to networks with low values for W_{min}, which exhibit poor or no such properties. The scaling behaviour along the spectrum of W_{min}, in conjunction with the robustness regarding parameter variations, endorse the idea of a relationship between fundamental properties of seismicity and the scaling properties of the earthquake networks, and can be used to discern the interrelated earthquakes from the rest of the dataset. The scale free behaviour of the connectivity distribution along the spectrum of the minimum weight values is mirrored by a similar behaviour of the distribution of the number of nodes' linked neighbours. The patterns found in the distributions of temporal and spatial intervals between earthquakes are similar in various networks, from large to small networks. Notable similarities are found between the variation of the network clustering coefficient, C, and the variation of the exponents of the connectivity distribution, β, and of the weight distribution, γ. Results of this method are further applied for the study of temporal changes in volcanic seismicity patterns. It is shown that β, γ, and C manifest a generally synchronous variation over successive temporal windows, which can be related to changes in seismicity and in the life of the volcanic system. A Zipf distribution is found for the ranked sets of magnitude values of successive network nodes. The distribution of differences between the magnitude values of successive nodes is also governed by a power law.

Date: May 23, 2014

Table of Contents

Chapter 1

Introduction

1.1. Background

Understanding the fundamental laws that govern seismicity is an important, but challenging task. Extensive research dedicated to this topic shows various aspects of the correlations found in earthquake patterns. Correlations have been found in magnitude (Gutenberg and Richter, 1954; Lippiello et al., 2012b), time (Omori, 1894; Shcherbakov et al., 2004; Shcherbakov et al., 2006), and space (Turcotte, 1977; Felzer and Brodsky, 2006; Lippiello et al., 2009). Integrated approaches have been developed to find space-time-magnitude patterns (Bak et al., 2002). Over the past decade, approaches based on complex networks have shown not only that networks of correlated earthquakes can be created, but also that these networks enjoy scaling properties (Baiesi and Paczuski, 2004; Baiesi and Paczuski, 2005; Davidsen et al., 2008).

1.2. Objectives

The purpose of this research is to use instruments from complex networks to study relationships between earthquakes. The earthquakes are seen as sets of space-time-magnitude events that can be related with each other, while the features of the interactions among earthquakes can vary over time. This approach is applied to seismicity associated to hotspot volcanism in Hawaii. In order to assess earthquakes interactions and their change in time, the following main objectives are pursued:

1. The development of an integrative method that maps seismic information to directed weighted networks.

2. The analysis of the earthquake networks, in search for possible patterns related to seismicity.

3. The analysis of earthquake networks in temporal windows, with the purpose of finding new instruments able to reflect the changes in the real-world system.

1.3.   Organization of the Thesis

This thesis consists of six chapters and four appendices.

Chapter 1, the current chapter, is meant to present the main objectives of the research.

Chapter 2 reviews important results in earthquake pattern analysis and complex networks that support the studies presented in the thesis.

Chapter 3 discusses aspects regarding the data and the experimental setup.

Chapter 4 shows results presented in the paper *"Scale Free Properties in a Network-Based Integrated Approach to Earthquake Pattern Analysis"*, which was published on March 24, 2014 in the journal *Nonlinear Processes in Geophysics*. A network-based method for the assessment of earthquake relationships in space-time-magnitude patterns is proposed. It is shown that networks with high values for the minimum edge weight W_{min} enjoy strong scaling properties, as opposed to networks with low values for W_{min}, which exhibit no such properties. The scaling behaviour along the spectrum of W_{min} values, in conjunction with the robustness regarding parameter variations, endorse the idea of a relationship between fundamental properties of

seismicity and the scaling properties of the earthquake networks. Results of this method are further applied for the study of temporal changes in volcanic seismicity patterns.

Chapter 5 presents results discussed in the paper *"Aspects of Structure in Earthquake Networks"*, which was submitted for publication to the journal *Pure and Applied Geophysics* on March 30, 2014. The paper discusses various aspects revealed by the study of earthquake networks introduced in chapter 4, such as: patterns found in the distributions of temporal and spatial intervals between earthquakes; similarities found between the variation of the network clustering coefficient, *C,* and the variation of the exponents of the connectivity distribution, β, and of the weight distribution, γ; the Zipf distribution found for the ranked sets of magnitude values of successive network nodes, a.s.o.

Chapter 6 summarizes the results of the thesis research and presents aspects of the study which can be developed in future research.

The appendices refer to the code of the main programs that have been developed in the framework of this research.

In order to avoid confusion involving the two articles in Chapter 4 and Chapter 5, the numbers of tables and figures start from number 1 for each chapter. Also, where references are required, they are added at the end of each chapter.

References

Bak, P., Christensen, K., Danon, L., and Scanlon, T.: Unified scaling law for earthquakes, Phys. Rev. Lett., 88, doi 10.1103/PhysRev Lett. 88.178501, 2002.

Baiesi, M., and Paczuski, M.: Scale-free networks for earthquakes and aftershocks, Phys. Rev. E, 69, 066106-1–8, 2004.

Baiesi, M., and Paczuski, M.: Complex networks of earthquakes and aftershocks, Nonlinear Proc. Geoph. 12, 1-11, 2005.

Davidsen, J., Grassberger, P., and Paczuski, M.: Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity, Phys. Rev. E, 77, 066104, 2008.

Felzer, K. R., and Brodsky, E. E.: Decay of aftershock density with distance indicates triggering by dynamic stress, Nature, 441, 735-738, 2006.

Gutenberg, B., and Richter, C. F.: Seismicity of the Earth. Princeton University Press, Princeton, 1954.

Lippiello, E., de Arcangelis, L., and Godano, C.: The role of static stress diffusion in the spatio-temporal organization of aftershocks, Phys. Rev. Lett., 103, 038501, doi:10.1103/PhysRevLett.103.038501, 2009.

Lippiello, E., Godano, C., and de Arcangelis, L.: The earthquake magnitude is influenced by previous seismicity, Geophysical Research Letters, 39, 5, doi: 10.1029/2012GL051083, 2012b.

Omori, F.: On the aftershocks of earthquakes, J. College of Science, Imperial University of Tokyo, 7: 111–200, 1894.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B.: A generalized Omori's law for earthquake aftershock decay, Geophys. Res. Lett., 31, L11613, doi 10.1029/2004GL019808, 2004.

Shcherbakov, R., Turcotte, D. L., and Rundle, J. B.: Scaling properties of the Parkfield aftershock sequence, B. Seismol. Soc. Am., 96, 4B, 376–S384, doi: 10.1785/0120050815, 2006.

Turcotte, D.: Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge, 1977.

Chapter 2

Related Work

An earthquake is an oscillatory movement of the ground as a result of a sudden release of energy inside the Earth (Hyndman and Hyndman, 2009). The magnitude of an earthquake is a measure of the size of the earthquake, and is proportional to the logarithm of the energy (Hanks and Kanamori, 1979). As Bak and Tang (1989), and Kagan (1994) pointed out years ago, there is still no complete understanding of the underlying mechanisms of seismicity, and even the most well-established relations that come from empirical observations cannot be logically derived from the fundamental laws of physics (Davidsen et al., 2008; van Stiphout et al., 2012).

Although significant progress has been done in recent years, seismic forecasting is still an aspiration (Shcherbakov et al., 2010; Tiampo and Shcherbakov, 2012; Varotsos et al., 2012; Varotsos et al., 2013). Among many hypotheses and models, a fact that is generally acknowledged by the whole scientific community is the ubiquity of power law forms in earthquake occurrence.

Omori (1895) was the first scientist to note that the aftershock rate decays with time according to a power law:

$$n(t) \approx t^{-1}, \tag{1}$$

where $n(t)$ is the frequency of aftershocks per time unit, and K and c are constants.

What is known today as the modified Omori law was the work of Utsu (1961), who showed that the aftershocks occurrence rate n fits the formula:

$$n(t) = K(t + c)^{-p}, \tag{2}$$

where p is a constant with values between 0.7 and 1.8.

Newer studies reveal more complex aspects of this formula (Shcherbakov et al., 2004; Shcherbakov et al., 2006).

It was also discovered by Gutenberg and Richter (1954) that the magnitude-frequency distribution of the cumulative number of earthquakes N of magnitude greater than m has the form:

$$N = 10^{a-bm} , \tag{3}$$

where a and b are constants, and in most cases $b \approx 1$. Although eq. (3) represents an exponential with respect to m, since the magnitude is proportional to the logarithm of the seismic moment, it becomes a power law when considered with respect to the seismic moment.

Felzer and Brodsky (2006) found that the decay of aftershocks as a function of distance also fits a power law over distances of 0.2-50 km:

$$\rho(r) = cr^{-n} , \tag{4}$$

where ρ is the linear density of aftershocks, r is the distance of the aftershock from the mainshock, c is a constant, and $n \approx 1.3$.

A power law was also found by Lippiello et al. (2009), who showed that the distribution $\rho\,(r)$ of the epicentral distance r between each aftershock and its mainshock exhibits a maximum followed by a power law decay.

The concept of mainshock refers to earthquakes that are independent of other earthquakes (also called background seismicity), while earthquakes that depend on other earthquakes are called aftershocks, foreshocks, or triggered earthquakes. The process of discrimination of earthquakes into these two categories is called seismicity declustering

(van Stiphout et al., 2012). This process is important for practical and scientific reasons, such as hazard management, seismicity modelling or prediction research. Different declustering methods have been developed, and, as van Stiphout et al. (2012) point out, each of them rely on their own model for seismicity.

In the window declustering methods, aftershocks are identified if they fall in a space-time window related to a mainshock, which is a function of the mainshock's magnitude (Knopoff and Gardner,1972; Gardner and Knopof, 1974). If d is the spatial distance of influence of a mainshock of magnitude M, and t is its temporal interval of influence, then all earthquakes situated at a distance d in space and a distance t in time are aftershocks or foreshocks of that mainshock. As an example, in Gardner and Knopoff (1974) d and t developed for southern California are defined as:

$$d \text{ (in km)} = 10^{0.1238*M+0.983} \tag{5}$$

$$t \text{ (in days)} = \begin{cases} 10^{0.032*M+2.7389}, \text{ if } M \geq 6.5 \\ 10^{0.5409*M-0.547}, \text{ if } M < 6.5 \end{cases} \tag{6}$$

Various choices of parameter values in such formulas lead to significant variations in aftershock identification.

Reasenberg's algorithm (Reasenberg, 1985) identifies foreshocks and aftershocks within a cluster based on Omori's law for the cluster's time extension and on a window-type function for the cluster's spatial extension; also in this case, different choices of fixed parameter values may lead to substantially different estimates of the correlations between earthquakes.

The ETAS model (abbreviation from Epidemic Type Aftershock Sequence) is a statistical model, in which seismicity is considered the sum between a factor that

represents independent (background) earthquakes and a factor that represents the dependent (triggered) events, which is a superposition of modified Omori functions shifted in time (Ogata, 1988; Ogata and Zhuang, 2006):

$$\lambda_\theta(t) = \mu + \sum_{\{ j:t_j<t \}} e^{\alpha\{M_j-M_c\}} \nu(t-t_j) , \qquad (7)$$

where λ_θ represents the total seismicity rate, μ represents the rate of the background seismicity, $\nu(t-t_j)$ are modified Omori functions (Eq. 2) of the aftershocks that occurred before time t (expressed in days), the exponential is a weight factor for the size of the aftershock j and is a function of its magnitude M_j, and M_c is the cut-off magnitude of the fitted data. Eq. (7) expands the modified Omori formula to cover situations of cascading aftershocks that may result in very intricate structures.

The simple ETAS model described by the above equation is based on the modified Omori relation and does not include any spatial component. Based on ETAS, more complex models have been developed, for example space-time extensions of ETAS generated by multiplying $\nu(t-t_j)$ with a function of space under the summation in Eq. (7):

$$\lambda_\theta(t) = \mu + \sum_{\{ j:t_j<t \}} \nu(t-t_j) \times g(x-x_j, y-y_j; M_j - M_c) \qquad (8)$$

More elaborate models use probabilistic approaches, such as the stochastic declustering method of Zhuang et al. (2002); they start from the space-time extension of the ETAS model, and replace the functions for background and aftershock seismicity from Eq. 8 with probability functions.

A seismicity declustering method was also proposed by Baiesi and Paczuski (2004). They suggest that the identification of scaling properties in their earthquake networks endorse the idea of a relationship between these properties and the governing

12

features of real-life seismicity (Baiesi and Paczuski, 2004; Baiesi and Paczuski, 2005). In this context, the focus of their research is on the identification of correlations between earthquakes, and not on the precise classification of earthquakes in mainshocks and aftershocks; as they point out, an exact classification mainshock-aftershocks may be an intrinsically impossible task. Since their model was an insightful introduction of complex networks in the study of seismicity, their work will be discussed in more detail. They showed not only that networks of correlated earthquakes can be created, but also that these networks exhibit scaling properties. Albert and Barabási (2004) had already proven that the distributions of real networks enjoy power law properties: they showed that the Internet, science collaboration networks, biological networks, and many other real networks have connectivity distributions with scaling properties.

Baiesi and Paczuski (2004) built a directed weighted network in which the nodes were earthquakes, the direction of edges was given by the temporal succession of earthquakes, and the edge weight was given by a metric that included components in time, distance, and magnitude. Their method works as follows: if j is a given earthquake, the relationship between j and any of the earthquakes i that occurred before j and had a magnitude within an interval Δm of m_i is described by the expected number of events between i and j, n_{ij}, or by the correlation between the two events i and j, c_{ij}, as follows:

$n_{ij} = C\ t\ l^{df}\ \Delta m\ 10^{-bmi}$,

or

$c_{ij} = 1/n_{ij} = C^{-1}\ t^{-1}\ l^{-df}\ \Delta m^{-1}\ 10^{bmi}$,

where t is the time interval between i and j measured in seconds, l is the distance in space between i and j measured in meters, d_f is the fractal dimension of earthquake epicenters in the studied area, and b is the exponent of the magnitude-frequency distribution.

Each new earthquake j attaches with a single link to the previous earthquake i that minimizes n_{ij} (or maximises c_{ij}), which means that an earthquake can have only one related predecessor and the network has a tree structure. Only the links that are strong enough are retained for the network; the strength of a link is assessed against a threshold value n_c , with $(n_{ij} < n_c)$, or $c_<$, with $(c_{ij} > c_<)$. In 2005, they extended the network model to accept more than one predecessor, as long as $n_{ij} < n_c$ or $c_{ij} > c_<$, i.e. the link is strong enough. The authors give fixed values to the thresholds:

- $c_< = 10^2$ ($n_c = 10^{-2}$) for the tree structure (Baiesi and Paczuski, 2004);

- $c_< = 10^4$ ($n_c = 10^4$) for the extended network that allows more than one predecessor for a node.

Singularities are eliminated by taking a cut-off value for spatial distance $l_{min} = 100$ m, and a cut-off value for the time interval $t_{min} = 180$ s (Baiesi and Paczuski, 2004) or $t_{min} = 60$ s (Baiesi and Paczuski, 2005). The method is probably robust with respect to variations in the values of these parameters, but a study of the effect of choosing different values for them is not shown.

Also, when constructing the networks, Baiesi and Paczuski choose to discard all magnitudes lower than a magnitude threshold $m_<$. In their first paper (Baiesi and Paczuski, 2004) the magnitude threshold is $m_< = 2.5$, while in their second paper (Baiesi and Paczuski, 2005) the magnitude threshold is $m_< = 3$. Although the properties of the resulting networks are robust with respect to the actual value of $m_<$, as Baiesi and

Paczuski (2005) point out, the selection of a magnitude threshold may result in the loss of valid correlations between earthquakes.

The criterion for the selection of earthquakes into the network is the maximization of the correlation function c_{ij} with respect to preceding events i as long as c_{ij} does not drop under the threshold value $c_<$. The event pairs are checked one by one against the threshold value $c_<$, and the events with the highest c_{ij} above the threshold $c_<$ are considered nodes with a mainshock-aftershock relationship.

The components of c_{ij} are functions with large discrepancies in their values. For example, for the network with tree structure (Baiesi and Paczuski, 2004), the maximum possible value of the component in distance is $100^{-1} = 0.01$, the maximum possible value of the component in time is $180^{-1} \approx 0.006$, while the minimum possible value of the component in magnitude is $10^{2.5} \approx 316.23$ (a cut-off value of 2.5 was applied to the magnitude values). A comparison between the contribution of each of the three components should thus consider 0.01, 0.006 and 316.23, while this discrepancy can only grow with time and distance. This choice seems to effectively address the identification of event clusters around the largest shocks.

Various distributions are assessed, such as the distributions of node connectivity, of the correlation c between event pairs, of the link weights, and it is shown that they exhibit strong scaling properties.

The tree structure analysis with the measure n_{ij} introduced by Baiesi and Paczuski (Baiesi and Paczuski, 2004) is expanded by Zaliapin et al. (2008) in a new statistical methodology aimed at the identification of clusters of earthquakes, and particularly of aftershocks. Zaliapin and Ben-Zion apply this method to southern California data and

show that the method is accurate and robust with respect to a number of parameters, as well as to catalogue incompleteness and location errors, and identifies clusters comprised of foreshocks, mainshocks and aftershocks (Zaliapin and Ben-Zion, 2013a). Furthermore, they classify the detected clusters into three major types, which correspond largely to singles, burst-like and swarm-like sequences, and study correlations between different cluster types and geographic locations (Zaliapin and Ben-Zion, 2013b).

Overall, the work of Bayesi and Paczuski opened an elegant path for the future study of seismicity. One might recognize a trend in some areas of today's scientific research that resorts to network-based approaches when complicated systems, containing many interacting subsystems, are involved, as for example the evolving climate networks (Hlinka et al., 2014), or the networks of recurrent events constructed by Davidsen et al. (2008).

References

Albert, R., and Barabási, A.-L.: Statistical mechanics of complex networks, Rev. Mod. Phys., 74, 47-97, 2002.

Bak, P., and Tang, C.: Earthquakes as a Self-Organized Critical Phenomenon, Journal of Geophysical Research, 94, B 1L, 15635-15637, 1989.

Baiesi, M., and Paczuski, M.: Scale-free networks for earthquakes and aftershocks, Phys. Rev. E, 69, 066106-1–8, 2004.

Baiesi, M., and Paczuski, M.: Complex networks of earthquakes and aftershocks, Nonlinear Proc. Geoph. 12, 1-11, 2005.

Davidsen, J., Grassberger, P., and Paczuski, M.: Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity, Phys. Rev. E, 77, 066104, 2008.

Felzer, K. R., and Brodsky, E. E.: Decay of aftershock density with distance indicates triggering by dynamic stress, Nature, 441, 735-738, 2006.

Gardner, J. K., and Knopoff, L.: Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, Bull. Seis. Soc. Am., 64, 5, 1363–1367, 1974.

Gutenberg, B., and Richter, C. F.: Seismicity of the Earth. Princeton University Press, Princeton, 1954.

Hanks, T.C., Kanamori, H.: Moment magnitude scale, Journal of Geophysical Research 84, B5, 2348–2350, 1979.

Hlinka, J., Hartman, D., Jajcay, N., Vejmelka, M., Donner, R., Marwan, N., Kurths, J., and Paluš, M.: Regional and inter-regional effects in evolving climate networks, Nonlin. Processes Geophys., 21, 451–462, 2014.

Hyndman, D., Hyndman, D.: Natural Hazards and Disasters, Brooks/Cole: Cengage Learning, Belmont, 2009.

Kagan, Y.Y.: Observational evidence for earthquakes as a nonlinear dynamic process, Physica D, 77, 160-192, 1994.

Knopoff, L., and Gardner, J. K.: Higher seismic activity during local night on the raw worldwide earthquake catalogue, Geophys. J. R. Astr. Soc., 28, 311–313, 1972.

Lippiello, E., de Arcangelis, L., and Godano, C.: The role of static stress diffusion in the spatio-temporal organization of aftershocks, Phys. Rev. Lett., 103, 038501, doi:10.1103/PhysRevLett.103.038501, 2009.

Ogata, Y., and Zhuang, J.: Space–time ETAS models and an improved extension, Tectonophysics, 413, 13–23, 2006.

Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes, J. Am. Stat. Assoc. 83, 9 –27, 1988.

Omori, F.: On the aftershocks of earthquakes, J. College of Science, Imperial University of Tokyo, 7: 111–200, 1894.

Reasenberg, P.: Second-order moment of central California seismicity, 1969-1982, J. Geophys. Res., 90, 5479– 5495, 1985.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B.: A generalized Omori's law for earthquake aftershock decay, Geophys. Res. Lett., 31, L11613, doi 10.1029/2004GL019808, 2004.

Shcherbakov, R., Turcotte, D. L., and Rundle, J. B.: Scaling properties of the Parkfield aftershock sequence, B. Seismol. Soc. Am., 96, 4B, 376–S384, doi: 10.1785/0120050815, 2006.

Shcherbakov, R., Turcotte, D. L., Rundle, J. B., Tiampo, K.F., and Holliday, J.R.: Forecasting the Locations of Future Large Earthquakes: An Analysis and Verification, Pure and Applied Geophysics, 167, 743–749, 2010.

Tiampo, K.F., and Shcherbakov, R.: Seismicity-based earthquake forecasting techniques: Ten years of progress, Tectonophysics, 522-523, 89-121, 2012.

Utsu, T. (1961). A statistical study of the occurrence of aftershocks, Geophysical Magazine, 30, 521–605.

van Stiphout, T., Zhuang, J., and Marsan, D.: Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-52382934, http://www.corssa.org, 2012.

Varotsos, P.A., Sarlis, N.V., and Skordas, E.S.: Order parameter fluctuations in natural time and b-value variation before large earthquakes, Natural Hazards and Earth System Sciences, doi:10.5194/nhess-12-3473-2012, 3473-3481, 2012.

Varotsos, P.A., Sarlis, N.V., and Skordas, E.S., Lazaridou, M.S.: Seismic Electric Signals: An additional fact showing their physical interconnection with seismicity, Tectonophysics, 589, 116–125, 2013.

Zaliapin, I., Gabrielov, A., Keilis-Borok, V, and Wong, H.: Clustering Analysis of Seismicity and Aftershock Identification, Phys. Rev. Lett., 101, 018501, 2008.

Zaliapin, I., and Ben-Zion, Y.: Earthquake clusters in southern California I: Identification and stability, Journal of Geophysical Research, 118, 2847-2864, 2013.

Zaliapin, I., and Ben-Zion, Y.: Earthquake clusters in southern California I: Identification and stability, Journal of Geophysical Research, 118, 2865-2877, 2013a.

Zhuang, J., Ogata, Y., and Vere-Jones, D.: Stochastic declustering of space-time earthquake occurrences, J.Am. Stat. Assoc., 97, 369-380, 2002b.

Chapter 3

Study Data and Experimental Setup

3.1. Preliminaries: Data Preprocessing

The data source for this study is the Advanced National Seismic System (ANSS) catalogue for Hawaii (18.5° ÷ 20.5° N and 154.5° ÷ 156.5° W), with events ranging from January 1st, 1989 to December 31st, 2012. The conversion to Cartesian coordinates was performed by using the software GEOTRANS from the National Geospatial-Intelligence Agency (http://gcmd.nasa.gov/records/GEOTRANS.html). Given the flexibility of Visual FoxPro 9.0 in raw data manipulation, this software was used for the preliminary pre-processing. Table 1 presents the description of the main fields of the table *Data.DBF* that was created from the downloaded text file, and Fig. 1 shows a data sample.

Field	Field Name	Type	Width	Description
1	PID	Character	7	Node (earthquake) identifier (unique)
2	YEAR	Character	4	Year of the earthquake
3	MONTH	Character	2	Month of the earthquake
4	DAY	Character	2	Day of the first earthquake
5	HOUR	Character	2	Hour of the earthquake
6	MINUT	Character	2	Minute of the earthquake
7	SEC	Character	5	Second of the earthquake
8	LAT	Character	10	Latitude of the earthquake
9	LONG	Character	10	Longitude of the earthquake
10	DEPTH	Character	10	Depth of the earthquake
11	MAGN	Character	5	Magnitude of the earthquake
12	CHAR_X	Character	10	x rectangular coordinate of the earthquake
13	CHAR_Y	Character	10	y rectangular coordinate of the earthquake

Table 1. Table Data.DBF that provides the data for this thesis research.

Fig. 1. Data sample for table Data.DBF.

The number of events in the catalogue is 64,392. Scaling properties in the Gutenberg-Richter magnitude-frequency distribution start approximately at magnitude 1.6, as shown by the regression line in Fig. 2, and therefore, for catalogue completeness, only events with magnitude $m >= 1.6$ are selected for the analysis (37,451 earthquakes).
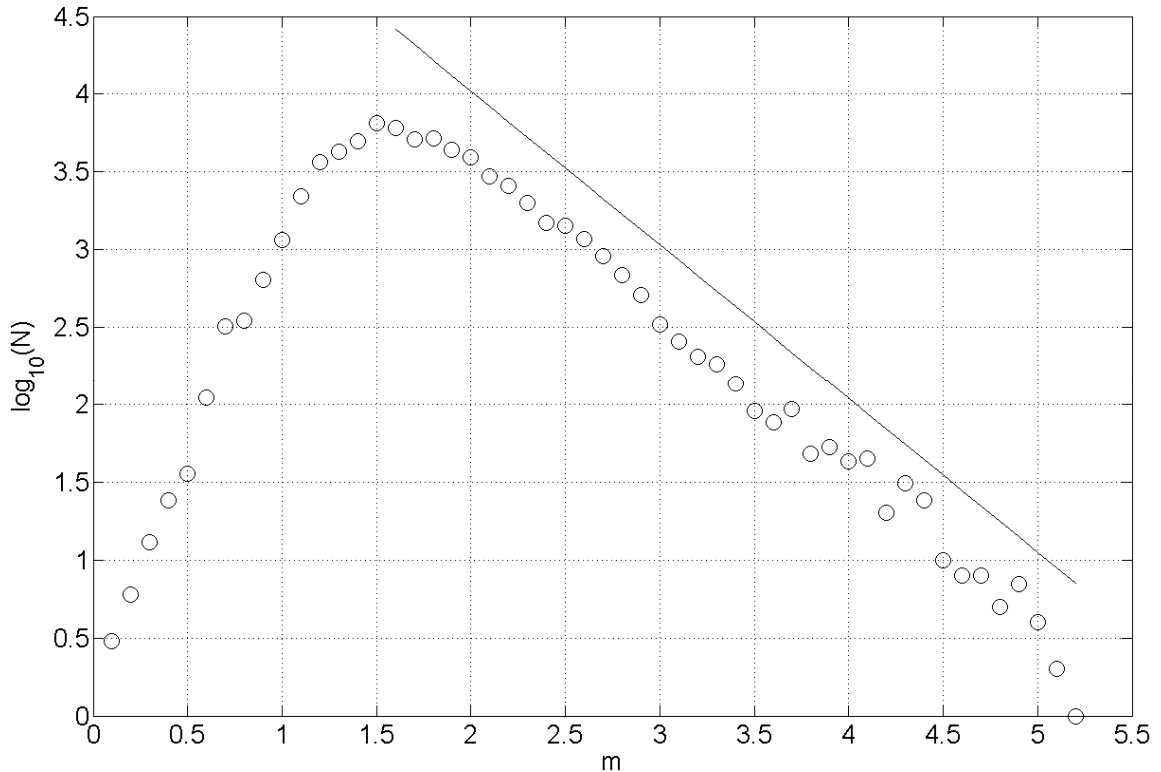

Fig. 2. Gutenberg-Richter magnitude-frequency distribution for the Hawaii data set ranging from January 1st, 1989 to December 31st, 2012. N is the cumulative number of earthquakes of magnitude greater than m, and the slope of the regression line (coefficient b of the distribution) is 0.998.

The fact that the earthquake catalogue is incomplete for low magnitude values can be noticed in Fig. 2. To find the catalogue completeness threshold, successive slopes from the *log(N)* vs. *m* graph have been determined and represented as a function of event magnitude: Fig. 3 shows that for low magnitudes the slopes have positive values, which decrease quickly with increasing magnitude; the decreasing tendency is followed by a plateau, and it is the *m*-value for the beginning of this plateau that indicate where the completeness of the catalogue

should begin. Therefore, the threshold value of *m* = 1.6 has been selected for this catalog, as illustrated in Fig.3. This value has been used throughout the whole study. For larger magnitudes the point pattern is increasingly scattered, as expected from the distribution presented in Fig. 2.
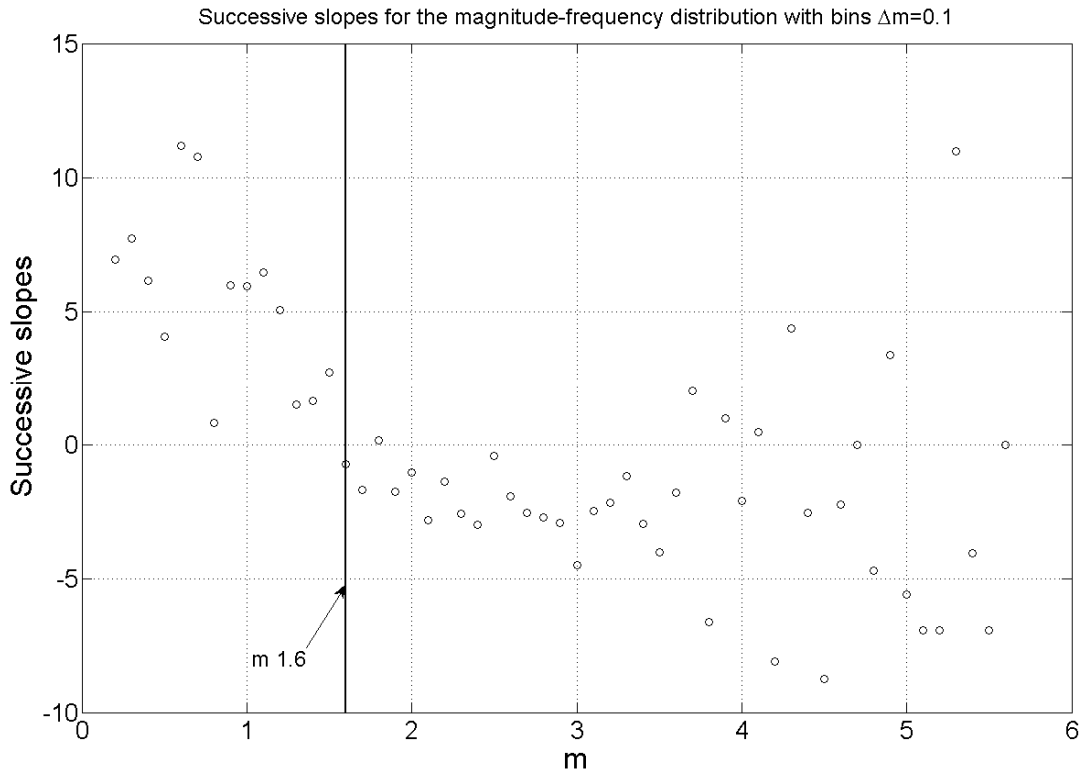
Fig. 3. Successive slopes determined for the magnitude-frequency distribution with a magnitude bin size Δm = 0.1. The selected threshold is marked by a thick vertical line.

The programming tools used for the research presented in this thesis are:

- Microsoft Visual FoxPro 9.0: for data preprocessing.

- Microsoft SQL Server 2008: for network generation and analysis.

- MATLAB R2007b-R2012b: for information visualisation.

3.2. Construction of the Earthquake Networks

The first experimental objective is to build the earthquake network. The conceptual definition is presented in the papers in Chapters 4 and 5. The networks are generated in the database *Network,* which was created on a Microsoft SQL Server 2008. A diagram of the main tables of this database, including their primary keys and foreign keys, is shown in Fig. 4. The table *Data.DBF* is imported on the server through the Integration Services into the table *Data*. The user defined procedure *usp_Create_Network*, which is presented in Appendix A, uses the information in the input data *Data* to build the actual network in the table *Network_Edges* (metadata are presented in *Table 2*).

Examples of images from a network of earthquakes are shown in Fig. 5a. Details are also illustrated in Fig.5b, where, by zooming in, the actual nodes and edges can be seen. The software used to create these images was ArcGIS 9.2. Fig. 6 presents data samples from a network table.

By running the procedure *Network.dbo.usp_Create_Network* with different values of input parameters, different classes of networks are built as described in Chapters 4 and 5. Inside each class, series of networks (i.e. *Network_Edges* tables) are created by setting different threshold values for the minimum edge weight W_{min}. The newly created networks of each class are then copied into separate databases (*Net_M, Net_O, Net_P*,…) for future analysis and storage purposes.
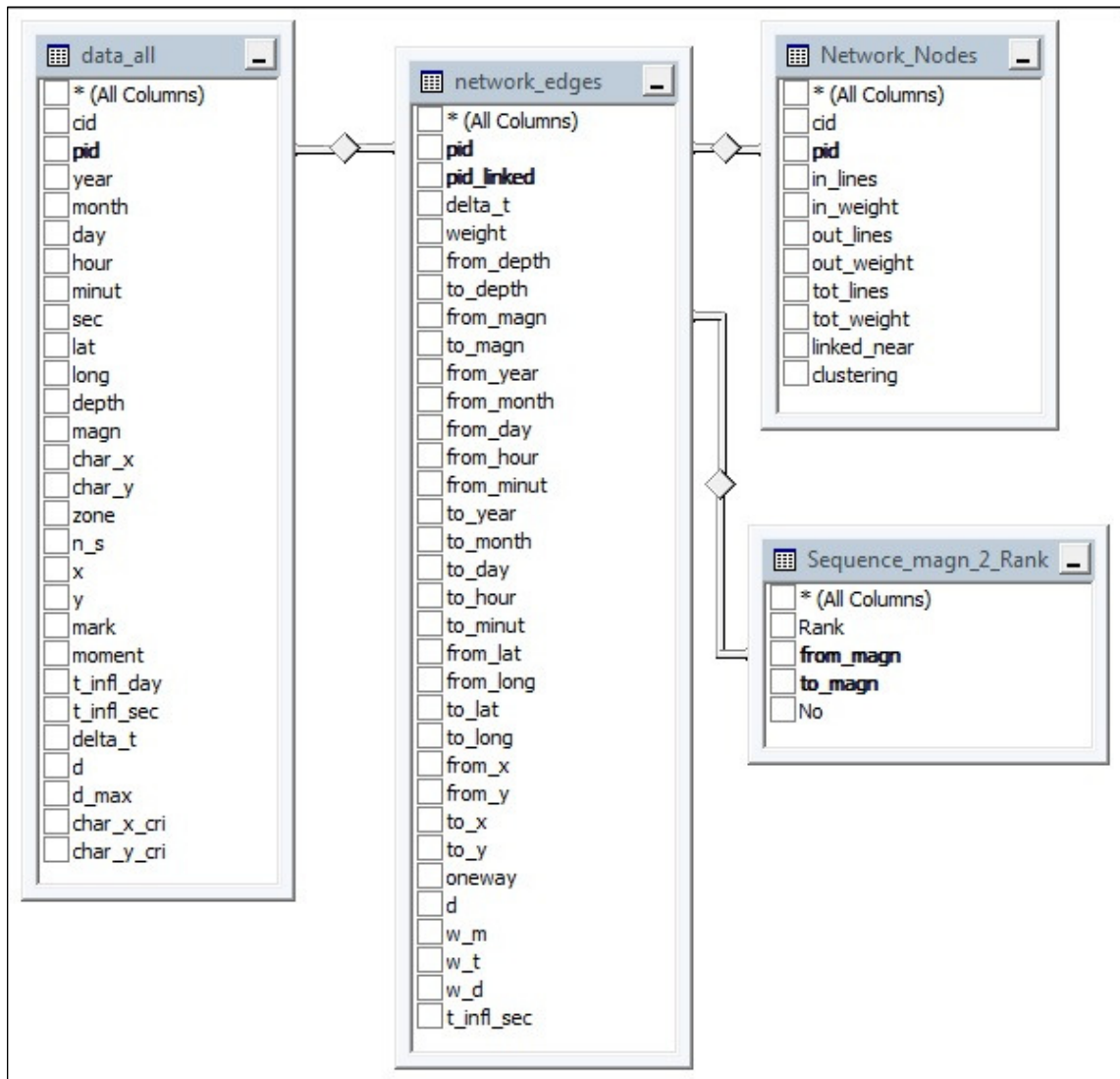
Fig. 4.Diagram of the Network database. The primary keys and the foreign keys of the tables are presented in bold and the relationships are represented by thin bars.
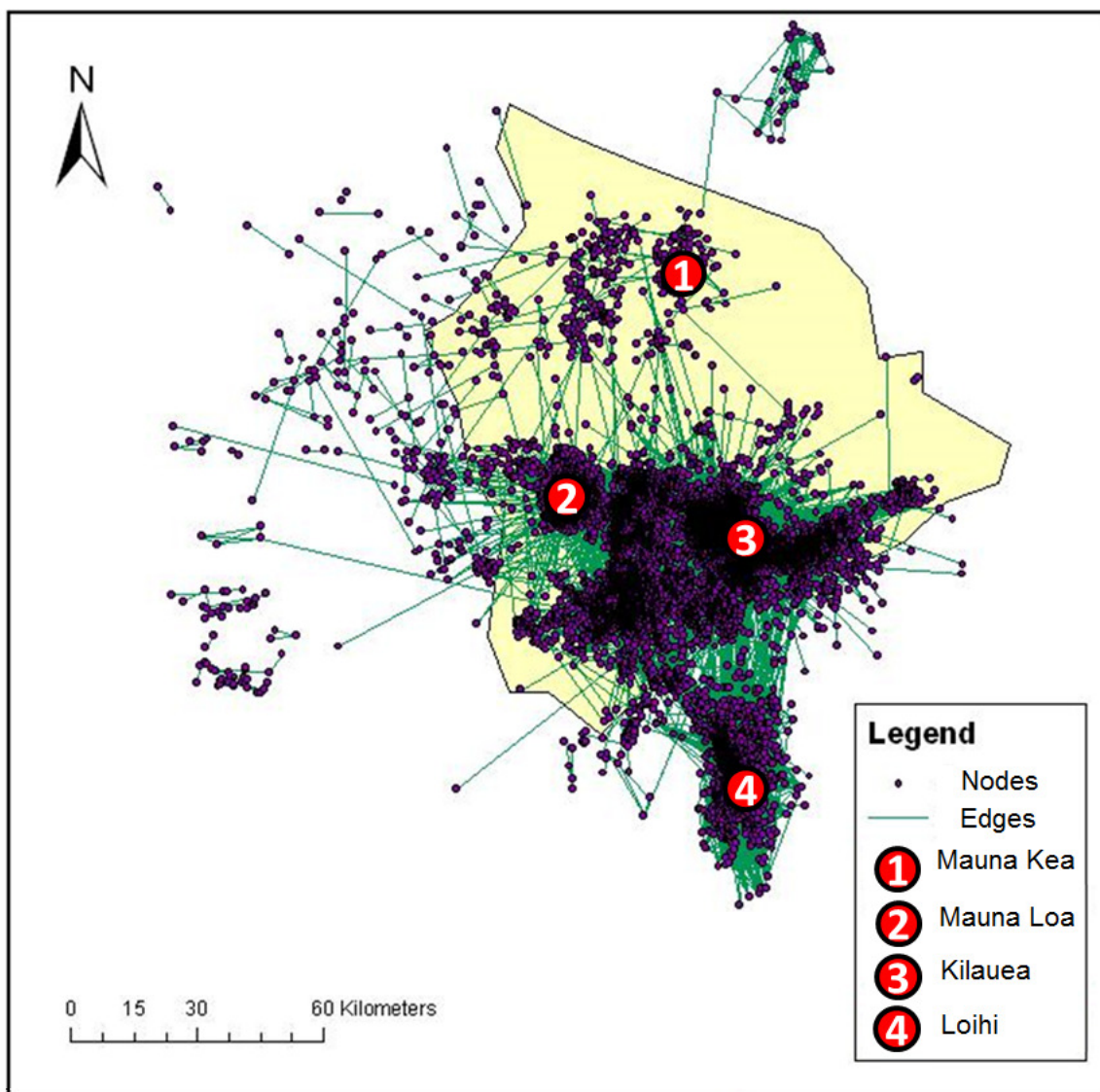
Fig. 5a. Example of an earthquake network with locations of the main volcanoes on the Big Island of Hawaii.
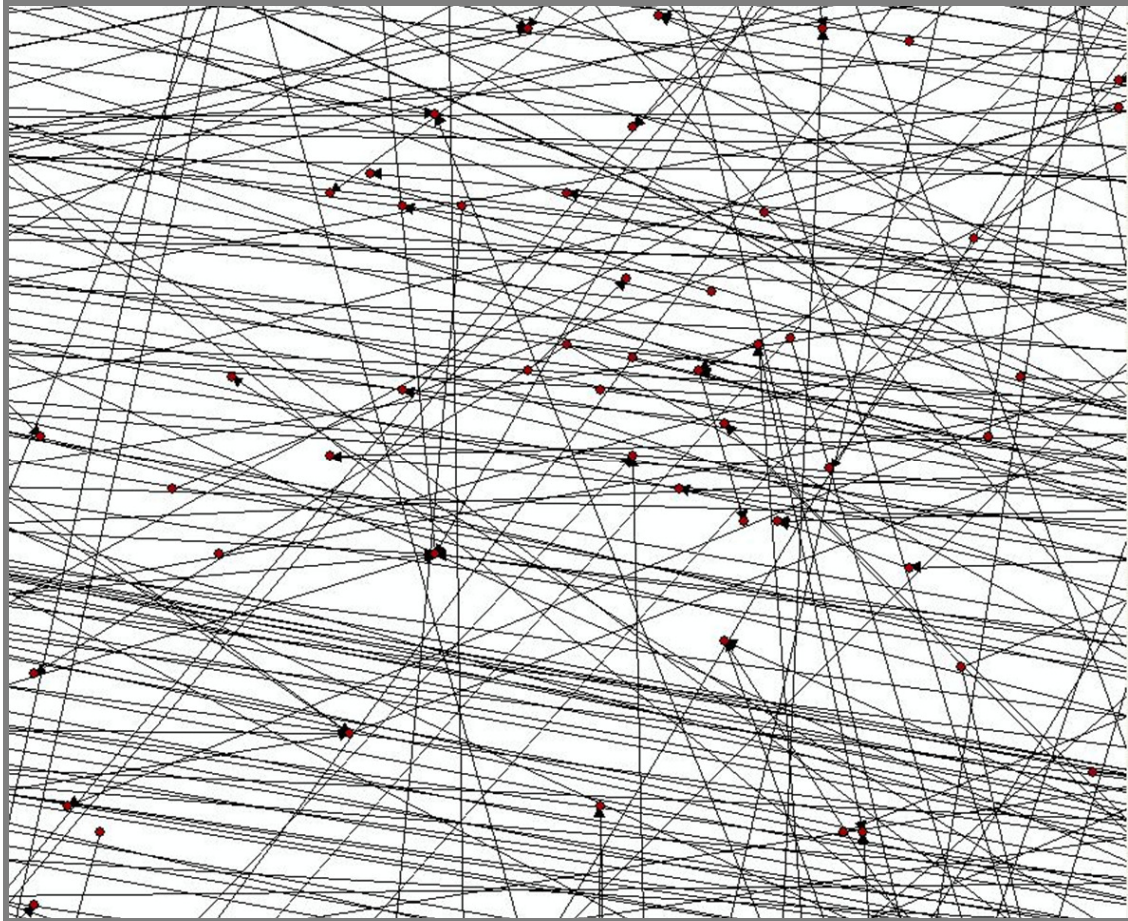
Fig. 5b. Nodes and edges of an earthquake network.

Network_edges

Pid	Pid_linked	Delta_t	Weight	From_depth	To_depth	D	From_magn	To_magn	From_year	From_month	From_day	From_hour	From_minut	To_year	To_month	To_day
4	8	4723	0.0000002811153242000	9.80	16.07	2.042	1.83	1.87	1989	12	10	16	59	1989	12	10
4	10	10208	0.0000000473287970000	9.80	5.71	23.768	1.83	2.03	1989	12	10	16	59	1989	12	10
4	12	36333	0.000000041047997000	9.80	4.52	10.313	1.83	1.93	1989	12	10	16	59	1989	12	11
4	15	57516	0.0000000063645723000	9.80	10.00	5.303	1.83	1.81	1989	12	10	16	59	1989	12	11
4	17	64464	0.0000000053964678000	9.80	12.45	5.507	1.83	1.75	1989	12	10	16	59	1989	12	11
4	18	86557	0.0000000006339599000	9.80	7.28	21.629	1.83	1.98	1989	12	10	16	59	1989	12	11
4	21	98629	0.0000000170239659000	9.80	7.79	1.716	1.83	1.70	1989	12	10	16	59	1989	12	11
4	22	100163	0.000000285067972000	9.80	9.88	1.158	1.83	1.83	1989	12	10	16	59	1989	12	11
4	24	107095	0.0000000046484216000	9.80	6.04	4.223	1.83	1.82	1989	12	10	16	59	1989	12	11
4	25	109477	0.0000000019837464000	9.80	10.83	7.807	1.83	2.09	1989	12	10	16	59	1989	12	11
4	26	110376	0.0000000049923040000	9.80	12.50	3.917	1.83	2.18	1989	12	10	16	59	1989	12	11
4	27	126913	0.0000000003287484000	9.80	9.67	26.496	1.83	2.14	1989	12	10	16	59	1989	12	12
4	34	162012	0.0000000071163383000	9.80	10.30	2.267	1.83	1.85	1989	12	10	16	59	1989	12	12
4	37	162882	0.0000000021665176000	9.80	6.49	5.449	1.83	1.80	1989	12	10	16	59	1989	12	12
4	38	163775	0.0000000042737248000	9.80	11.15	3.281	1.83	1.86	1989	12	10	16	59	1989	12	12

Network_edges

To_day	To_hour	To_minut	From_lat	From_long	To_lat	To_long	From_x	From_y	To_x	To_y	Oneway	W_m	W_t	W_d	T_infl_sec
10	18	18	19.42220	-155.2857	19.41580	-155.3040	127.323	102.530	125.409	101.819	FT	0.000169824365246000	0.381126610259928000	0.043432541994440000	864000
10	19	50	19.42220	-155.2857	19.33970	-155.0773	127.323	102.530	149.250	93.358	FT	0.000169824365246000	0.176329524615602000	0.001580521586834000	864000
11	3	5	19.42220	-155.2857	19.35430	-155.3532	127.323	102.530	120.296	94.981	FT	0.000169824365246000	0.049541738915036000	0.004878886556680000	864000
11	8	58	19.42220	-155.2857	19.43030	-155.2358	127.323	102.530	132.549	103.431	FT	0.000169824365246000	0.031295574179899000	0.011975297121654000	864000
11	10	54	19.42220	-155.2857	19.42950	-155.2337	127.323	102.530	132.770	103.342	FT	0.000169824365246000	0.027922509451769000	0.011380337476069000	864000
11	17	2	19.42220	-155.2857	19.36420	-155.0893	127.323	102.530	147.969	96.082	FT	0.000169824365246000	0.020795554742218000	0.001795110963495000	864000
11	20	23	19.42220	-155.2857	19.41130	-155.2742	127.323	102.530	128.538	101.318	FT	0.000169824365246000	0.018250239990656000	0.054927794524263000	864000
11	20	49	19.42220	-155.2857	19.41970	-155.2750	127.323	102.530	128.447	102.252	FT	0.000169824365246000	0.017907679040078000	0.093407971118464000	864000
11	22	44	19.42220	-155.2857	19.38480	-155.2930	127.323	102.530	126.587	98.372	FT	0.000169824365246000	0.016807513630894000	0.016285535151227000	864000
11	23	24	19.42220	-155.2857	19.35420	-155.3048	127.323	102.530	125.373	94.970	FT	0.000169824365246000	0.016441797680391000	0.007104554531474000	864000
11	23	39	19.42220	-155.2857	19.42980	-155.3163	127.323	102.530	124.130	100.262	FT	0.000169824365246000	0.016307893020222000	0.018026157969206000	864000
12	4	15	19.42220	-155.2857	19.29780	-155.5020	127.323	102.530	104.723	88.639	FT	0.000169824365246000	0.014182948691765000	0.001364888167561000	864000
12	14	0	19.42220	-155.2857	19.40320	-155.2780	127.323	102.530	128.146	100.418	FT	0.000169824365246000	0.011110296356045000	0.037716461577450000	864000
12	14	14	19.42220	-155.2857	19.42980	-155.2343	127.323	102.530	132.706	103.375	FT	0.000169824365246000	0.011050944855785000	0.011544172391241000	864000
12	14	29	19.42220	-155.2857	19.39270	-155.2850	127.323	102.530	127.420	99.250	FT	0.000169824365246000	0.010990677707169000	0.022897184478309000	864000

Fig. 6. Data sample from the earthquake network table.

25

Table 2. Metadata for table Network_Edges.

#	Field Name	Type	Prec	Scale	Description
1	pid	character	7		Identifier of the first occuring earthquake
2	pid_linked	character	7		Linked node identifier
3	delta_t	numeric	10	0	Time interval between the 2 nodes (seconds)
4	weight	numeric	19	18	Edge weight
5	from_depth	numeric	4	2	Depth of the first occuring earthquake
6	to_depth	numeric	4	2	Depth of the linked earthquake
7	from_magn	numeric	3	2	Magnitude of the first occuring earthquake
8	to_magn	numeric	3	2	Magnitude of the linked earthquake
9	from_year	smallint	4		Year of the first occuring earthquake
10	from_month	tinyint	2		Month of the first occuring earthquake
11	from_day	tinyint	2		Day of the first occuring earthquake
12	from_hour	tinyint	2		Hour of the first occuring earthquake
13	from_minut	tinyint	2		Minute of the first occuring earthquake
14	from_sec	numeric	4	2	Second of the first occuring earthquake
15	to_year	smallint	4		Year of the linked earthquake
16	to_month	tinyint	2		Month of the linked earthquake
17	to_day	tinyint	2		Day of the linked earthquake
18	to_hour	tinyint	2		Hour of the linked earthquake
19	to_minut	tinyint	2		Minute of the linked earthquake
20	to_sec	numeric	4	2	Second of the linked earthquake
21	from_lat	character	8		Latitude of the first occuring earthquake
22	from_long	character	9		Longitude of the first occuring earthquake
23	to_lat	character	8		Latitude of the linked earthquake
24	to_long	character	9		Longitude of the linked earthquake
25	from_x	numeric	9	3	x rectangular coordinate of the first earthquake
26	from_y	numeric	9	3	y rectangular coordinate of the first earthquake
27	to_x	numeric	9	3	x rectangular coordinate of the linked earthquake
28	to_y	numeric	9	3	y rectangular coordinate of the linked earthquake
29	oneway	character	2		Sense of the edge: From-To (FT/TF)(for ArcGIS)
30	d	numeric	7	3	Distance between the 2 nodes (kilometers)
31	w_m	numeric	19	18	Weight due to the magnitude of the first event
32	w_t	numeric	19	18	Weight due to the time interval between events
33	w_d	numeric	19	18	Weight due to the distance between events
34	t_infl_sec	numeric	10	0	Max. time interval of influence between events

3.3. Network Analysis

The experimental analysis of the earthquake networks starts on the SQL Server and is finalized in MATLAB, where the network distributions are assessed and their image is represented. Starting from the data in table *Network_Edges*, the node connectivity, the node weight, the number of nodes' linked neighbours, and the node clustering coefficient are calculated in table *Network_Nodes* by the procedure *usp_Compute_Network_Nodes* (Appendix B). Fig. 7 presents a chart of the system with the functional modules of the *Network* database and the main tables they contain.
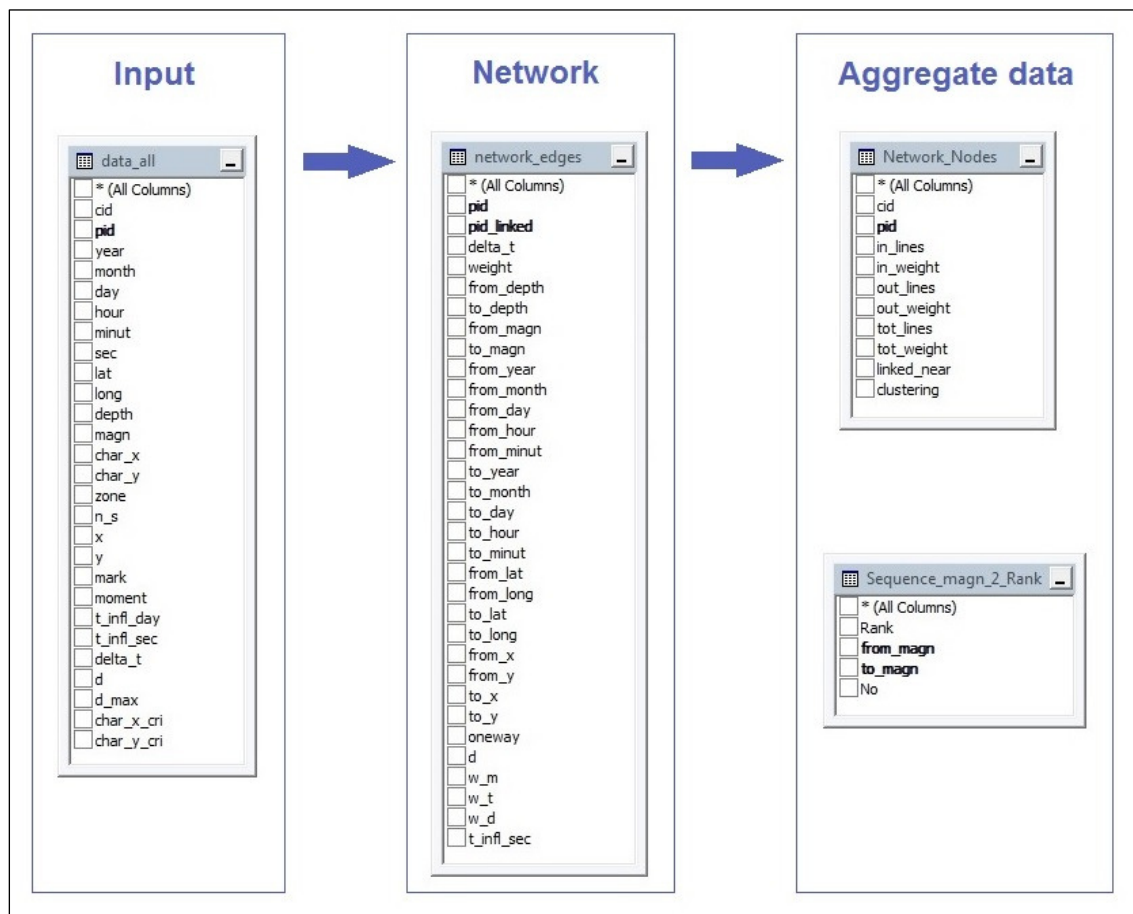

Fig. 7. Functional chart of the system.

The ranked sequences of two and three magnitude values of successive network nodes are calculated in the table *Sequence_Magn_2_Rank* and *Sequence_Magn_3_Rank* (not shown).

A description of the *Network_ Nodes* table is presented in Table 3.

Table 3. Metadata for table Network_Edges.

#	Field Name	Type	Prec.	Scale	Description
1	cid	int			Node identifier (unique, Identity) - in computing
2	pid	int			Node (earthquake) identifier (unique)
3	in_lines	int			Number of edges that enter into the node
4	in_weight	numeric	21	18	Total weight of edges that enter into the node
5	out_lines	int			Number of edges that leave the node
6	out_weight	numeric	21	18	Total weight of edges that leave the node
7	tot_lines	int			Total number of edges linked to the node
8	tot_weight	numeric	21	18	Total weight of all the edges in the node
9	linked_near	int			Number of linked neighbours of the node
10	clustering	numeric	4	3	Clustering coefficient of the node

The network tables *Network_Edges* and *Network_ Nodes* are then exported to MATLAB for analysis of the distributions of node connectivity, node weight, linked neighbours, spatial and temporal distances, etc., and for generation of the graphs.

3.4. Study of Temporal Windows

The temporal windows are sub-sets of a regular network that are created and stored in the database *Windows*. An edge table *Window_i_Edges* and a node table *Window_i_Nodes* are created for each of these sub-sets, where *i* is the rank of the window. The windows are generated by splitting up the main table of the network nodes in smaller tables *Window_i_Nodes* of equal size, which can have an overlap factor. The

temporal succession of events is accomplished by ensuring that the main table *Network_Nodes* is ordered by the node ID, *pid*, which implicitly gives the chronological order of the events. The corresponding table *Window_i_Edges* is generated by collecting all edges adjacent to the nodes in the table *Window_i_Node* from the main table *Network_Edges*.
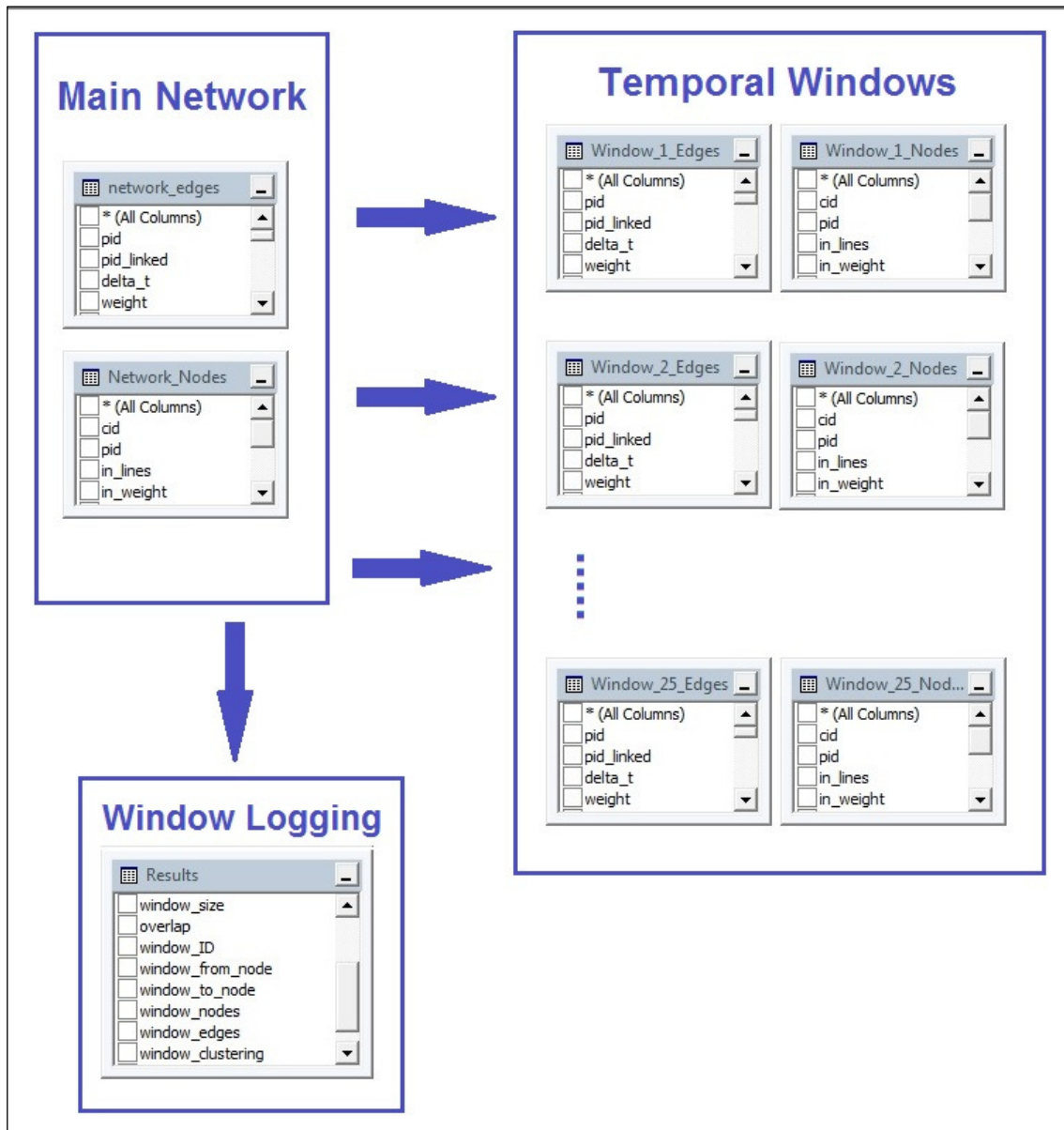
Fig. 8. Diagram of the tables in the database Windows (the database of the successive temporal windows).

During this process, the characteristics of each window are logged in the table *Results*, including the network average clustering coefficient for each window *i*. A diagram of the tables in the database *Windows* is shown in Fig 8.

The programs that generate the temporal windows are *usp_Compute_Network_Nodes_in_tempdb* (Appendix   C) and *usp_Create_Windows* (Apendix D).

The tables are then exported to MATLAB for further analysis.

Chapter 4

Scale Free Properties in a Network-Based Integrated Approach to Earthquake Pattern Analysis

Abstract

This paper proposes a network-based method for the assessment of earthquake relationships in space-time-magnitude patterns. It is shown that networks with high values for the minimum edge weight W_{min} enjoy strong scaling properties, as opposed to networks with low values for W_{min}, which exhibit no such properties. The scaling behaviour along the spectrum of W_{min} values, in conjunction with the robustness regarding parameter variations, endorse the idea of a relationship between fundamental properties of seismicity and the scaling properties of the earthquake networks. Results of this method are further applied for the study of temporal changes in volcanic seismicity patterns.

1 Introduction

Extensive research is dedicated to earthquake pattern analysis in an on-going effort to understand the laws that govern seismicity. Correlations in earthquake patterns have been found in magnitude (Gutenberg and Richter, 1954; Lippiello et al., 2012b), time (Omori, 1894; Shcherbakov et al., 2004; Shcherbakov et al., 2006), and space (Turcotte, 1977; Felzer and Brodsky, 2006; Lippiello et al., 2009). Integrated approaches have been developed to find space-time-magnitude patterns (Bak et al., 2002). Network-based approaches have shown not only that networks of correlated earthquakes can be created, but also that these networks enjoy scaling properties (Baiesi and Paczuski, 2004; Baiesi and Paczuski, 2005; Davidsen et al., 2008; Suteanu and Suteanu, 2011).

A space-time-magnitude metric defined for a directed network of earthquakes was proposed by Baiesi and Paczuski with their seismicity declustering method (Baiesi and Paczuski, 2004; Baiesi and Paczuski, 2005). The study in our paper also creates directed weighted networks of earthquakes with the purpose of assessing relationships between them. Given the ubiquity of power laws governing earthquake distributions (Nanjo and Nagahama, 2000; Lapenna et al., 2000; Shcherbakov et al., 2004; Carbone et al., 2005; Felzer and Brodsky, 2006; Shcherbakov et al., 2006; Bunde and Lennartz, 2012; Lippiello et al., 2012a; Lippiello et al., 2012b), power law forms are used to estimate quantitatively the relationships between events in a space-time-size perspective. However, there are major differences between Baiesi and Paczuski's model and the work presented in this paper. Not only do the two methods use different metrics, but, most importantly, they use different criteria for the discrimination of interrelated earthquakes from the rest of the set: in the method of Baiesi and Paczuski (2004, 2005) the criterion is

32

the maximization of a correlation function, while in this study series of networks are created, assessed and searched for scale free properties. Since the main component in Baiesi and Paczuski's metric is a function that is exponential in magnitude, their method effectively addresses the identification of event clusters around the largest shocks, while our method addresses earthquakes of all sizes that are considered close enough in space-time-magnitude to be interrelated.

Although a new quantitative metric is defined in our study and a new type of networks is built, the results show power law properties that are consistent with previous work of Baiesi and Paczuski (2005) and with their interpretation that the underlying correlations of the seismicity structure are unambiguous, sufficiently strong to survive the approximation of the metric, and can be reliably detected.

Our method is applied to seismicity associated to hotspot volcanism in Hawaii. The earthquakes are seen as sets of space-time-magnitude events that can be related with each other, while the quality of the interactions among earthquakes can vary over time. In order to assess these interactions and their change in time, an integrative approach that maps seismic information to directed weighted networks is developed. Different classes of networks of earthquakes are studied, and results show scale free properties that are robust with respect to certain variations in the definition of the networks. Networks with values of the minimum edge weight W_{min} in the middle to upper range of the spectrum of edge weight values enjoy strong scaling properties, as opposed to networks with W_{min} in the lower range, which exhibit poor or no such properties. It is shown that network parameters studied for successive event windows are able to reflect the way the

relationships between earthquakes are changing over time, and that patterns of change can be related to important events in the life of the volcanic system.

2   Construction of the earthquake networks

The epicenters of earthquakes that could be related to other earthquakes are seen as network nodes that are connected through directed edges. The edge direction is given by the temporal succession of the events. Ideally, only interrelated earthquakes can be nodes of this network, and the edges that link them to other nodes carry a space-time-magnitude weight. Therefore, a combination of three factors is evaluated before deciding whether or not any two earthquakes belong to the network: the size (magnitude) of the first occurring event, and its proximity to future events in space and in time. There are many possible combinations of these three factors. Even small earthquakes may be related to subsequent events if the latter were close enough in space and time.

In order to quantitatively assess the relationship between earthquakes, three weight variables are defined: the weight in distance w_d, the weight in time w_t, and the weight in magnitude w_m. A total weight W characterizes every edge as a combination of the previous three variables. Considering the Gutenberg-Richter law, the Omori law, and other scaling relationships regarding the distributions of earthquakes in space, time, and magnitude (e.g. Lei and Kusunose, 1999; Richards-Dinger et al, 2010; Felzer and Brodsky, 2006; Shcherbakov et al., 2006; Lennartz et al., 2008; Lippiello et al., 2009; Lippiello et al., 2012a; Sanchez and Shcherbakov, 2012), the following forms for the node weights of any one edge have been chosen:

a) Distance weight:

$$w_d = cd^r, r < 0,$$
(1)

where d is the spatial distance between the two nodes of the edge measured in km, and c is a positive constant.

b) Time weight:

$$w_t = st^p, p < 0,$$
(2)

where t is the time interval between the two nodes of the edge measured in hours, and s is a positive constant.

c) Magnitude weight:

$$w_m = \frac{m}{m_{max}},$$
(3)

where m is the magnitude of the first occurring node of the edge, and m_{max} is the maximum magnitude value in the data set.

The total weight of an edge is calculated as the product of the weights in space, time, and magnitude. Only the nodes that carry enough total weight belong to the network, which means that only edges that have a value of the total weight W higher than a minimum threshold W_{min} are selected for the network:

$$W = \begin{cases} w_d \cdot w_t \cdot w_m, & W \geq W_{min} \\ 0, & W < W_{min} \end{cases}$$
(4)

The generality of this definition allows various combinations of space-time-magnitude correlations between any two events and includes the possibility of multiple interactions for any given event: any node can have any number of edges that enter the

node and any number of edges that leave the node, as long as these edges carry enough total weight.

For practical reasons and with the purpose of avoiding singularities, a small cutoff value is used for the weights in space and time (it is also reasonable to assume that all earthquakes that are very close in space or in time could be related to each other). Therefore, modified forms of Eqs. (1) and (2) are used in the actual construction of the networks:

$$
w_d = \begin{cases} 1, & d \le d_{\min} \\ cd^r, & d \ge d_{\min}, \quad r < 0 \end{cases} \tag{1'}
$$

$$
w_t = \begin{cases} 1, & t \le t_{\min} \\ st^p, & t \ge t_{\min}, \quad p < 0 \end{cases} \tag{2'}
$$

Various values for the exponents r and p, and for the cutoff values $d_{\min}$ and $t_{\min}$ are explored. The constants c and s are calculated using the boundary conditions:

$$
w_d = cd^r_{\min} = 1 \tag{1''}
$$

and

$$
w_t = st^p_{\min} = 1 \tag{2''}
$$

An essential difference between the total edge weight defined in this paper and Baiesi and Paczuski's (2004, 2005) metric consists in the contribution of each of the three factors (time interval, space interval, and magnitude). In contrast with their approach, in this paper the three components (time interval, spatial distance, and magnitude) are seen independently, as separate components with useful statistical properties, and each of them

36

can have comparable contributions to the total edge weight W. This is accomplished by limiting the upper value of each of the three components to 1. The definition of a magnitude weight proportional to m, and not exponential in m is therefore meant to support a balance of factors in the total edge weight formula. The resulting networks and network distributions are governed by statistical contributions of each of the three components. This choice is especially important in the study of volcano-tectonic seismicity, where the seismic sources are associated not only with tectonic stress, but also with thermodynamic processes and the dynamics of gas, fluid and solid**.**

To simplify the computation, a maximum interval of influence in time T_{max}, and a maximum interval of influence in space D_{max} are assigned. Although fixed values of T_{max}, and D_{max} are chosen for creating the initial network, making this choice is different from making a choice of parameters involving a subjective factor, such as in window declustering methods or in Reasenberg's cluster method (Reasenberg, 1985). For example, in order to identify aftershocks, Knopoff and Gardner define space-time windows that are functions of the mainshock magnitude (Knopoff and Gardner, 1972; Gardner and Knopoff, 1974); various choices of parameter values lead to significant variations in the aftershock identification. Reasenberg's algorithm (Reasenberg, 1985) identifies foreshocks and aftershocks within a cluster based on Omori's law for the cluster's time extension and on a window-type function for the cluster's spatial extension; also in this case, different choices of fixed parameter values may lead to substantially different estimates of the correlations between earthquakes. In this paper, T_{max} and D_{max} receive fixed values only with the purpose of simplifying the computation. In principle, T_{max} and D_{max} could cover the whole extent of the catalogue in time and space. The study

shows that the final outcome is not affected by the initial choice of T_{max} and D_{max}, since large distances and long time intervals between events result in very small values of the edge weights w_d and w_t, and therefore lead to small values of the total edge weight W. The links carrying small weights are eliminated from the network in the next step anyway, when the network definition (4) is applied.

Different values for the maximum interval of influence in time T_{max}, and for the maximum interval of influence in space D_{max} are explored, as well as various values for r, p, d_{min} and t_{min}. This way, a series of network classes are generated with the purpose of creating a structured framework for the analysis: all networks that belong to a certain class B, C…, N share the same initial choice of D_{max}, r, d_{min}, T_{max}, p, t_{min}. Letters are used to name the classes of earthquake networks, and the choice of every letter has only a classification purpose. A description of the classes that have been studied is shown in Table 1.

In each class, an initial network is created when assigning the specific values to parameters. For clarity, an index $_0$ is used to describe these initial networks: B_0, C_0, etc. The highest value of the total edge weight in each class, H, is the highest value of the total edge weight in the initial network $(NETWORK\ CLASS)_0$, while the lowest value of the total edge weight in the class, L, is the lowest value of the total edge weight in the initial networks $(NETWORK\ CLASS)_0$. For example, the highest value in the network B_0 is $H=1$, and the lowest value in the network B_0 is $L=1.01*10^{-5}$. In general, the first networks B_0, C_0,…, N_0 are simply a collection of earthquakes, and not networks of interrelated events, and they serve for the operational initiation of the method.

Inside each class, specific values for the threshold W_{min} define distinct networks.

3 Data

The data source for this study is the Advanced National Seismic System (ANSS) catalogue for the Big Island of Hawaii, with events ranging from January 1st, 1989 to December 31st, 2012. Fig. 1 shows an example of a network of earthquakes. By zooming in, the actual nodes and edges can be seen (Fig. 2).
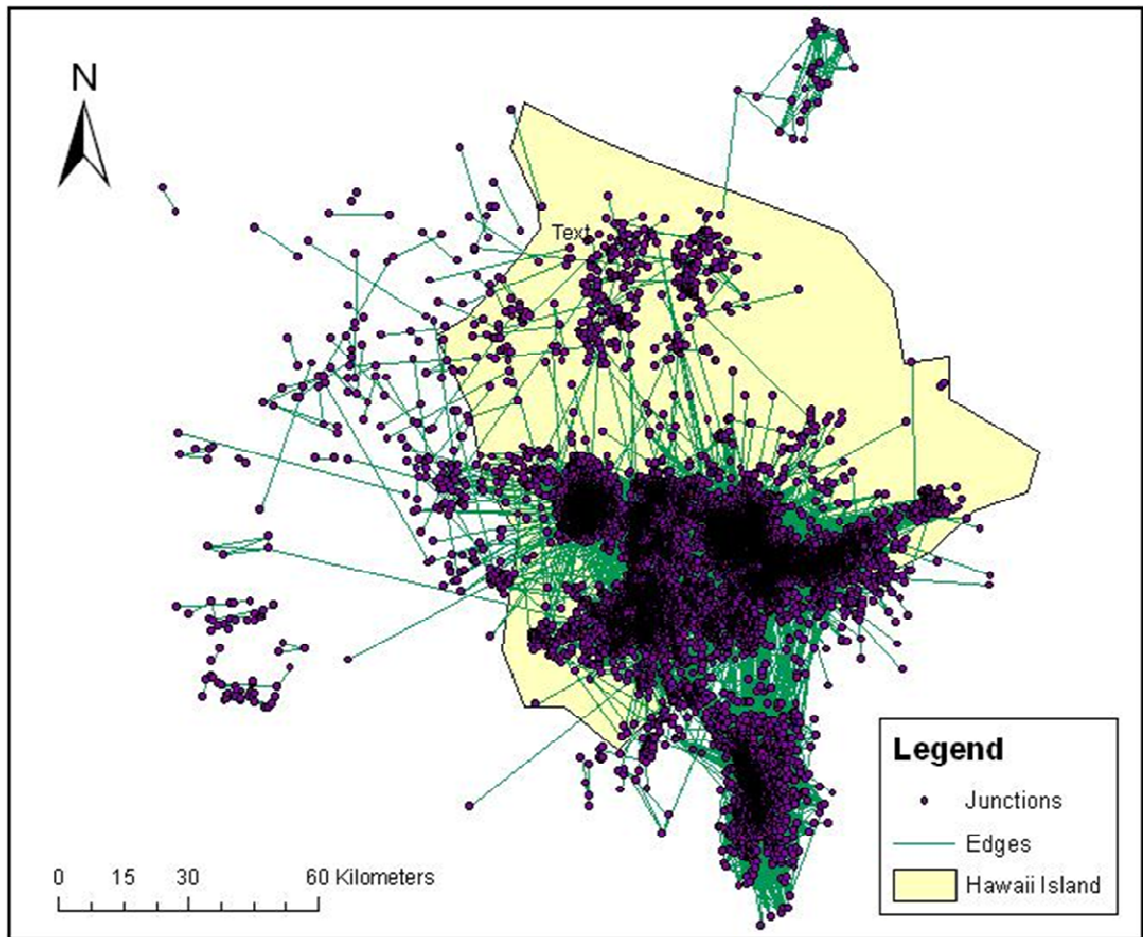
Figure 1. The earthquake network: an example for the Big Island of Hawaii.
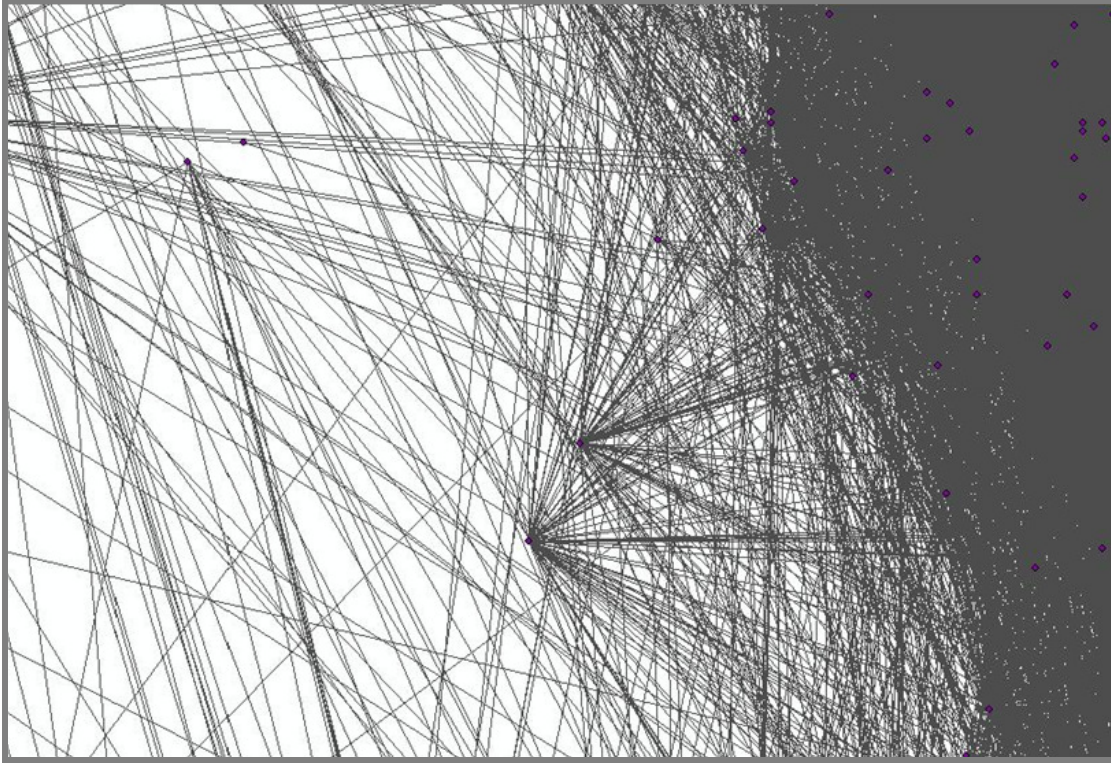
Figure 2. Zoomed-in example of nodes and edges of the earthquake network presented in Fig.1.

The number of events in the catalogue is 64,392. For catalogue completeness, only events with magnitude $m >= 1.6$ are used in the analysis (37,451 earthquakes); the b value in the Gutenberg-Richter magnitude-frequency distribution for this data set is b $\approx$ 0.99 (Fig. 3).

Different sets of networks in various network classes have been analyzed. The following values of the parameters have been studied:

- D_{max}: 10 km; 30 km; 50 km.

- T_{max}: 7 days; 8 days; 10 days; 30 days; 40 days.

- d_{min}: 0.025km; 0.1 km; 0.2 km; 1 km; 2 km.

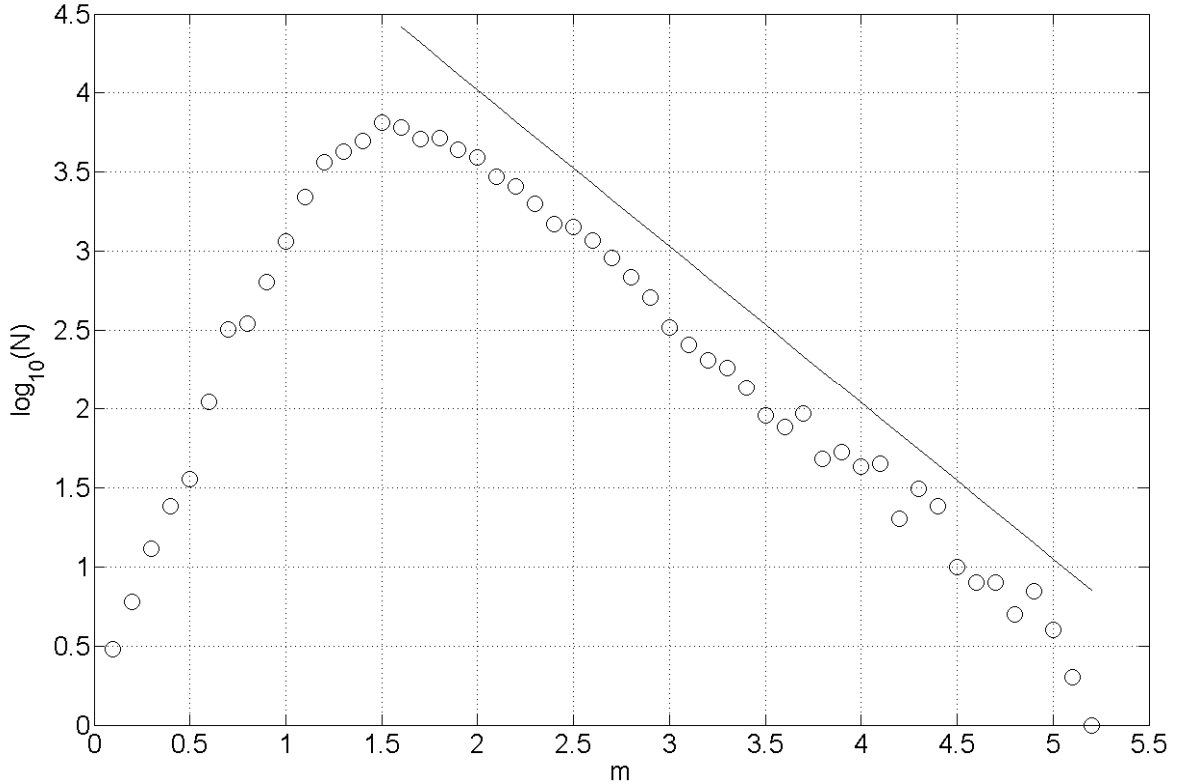- t_{min}: 3 min; 30 min; 1 h; 14h.

- r: -1; -1.35.

- p: -0.5; -1.

Figure 3. Gutenberg-Richter magnitude-frequency distribution for the Hawaii data set ranging from January 1st, 1989 to December 31st, 2012.

Depending on the lowest edge weight value *L* and the highest edge weight value *H* in each class, specific values for the threshold W_{min} are chosen in order to create and analyze distinct networks inside each class.

4   Results

4.1   Network parameters and analysis

Network parameters such as node connectivity and node weight distribution (Boccaletti et al., 2006) are assessed for sets of earthquake networks in different classes. The node connectivity (degree) represents the total number of edges in the node. The number of edges that enter the node (in-connectivity), and the number of edges that go out of the node (out-connectivity) are also studied. Similarly, the node weight, in-weight,

and out-weight are calculated and analyzed. The node weight is given by the sum of the weights of all edges in the node, the in-weight is the weight of all edges that enter the node, and the out-weight is the weight of all edges that go out of the node.

Table 1. All classes. T_{max} is the maximum time interval between events, D_{max} is the maximum distance between events, r is the exponent of the distance weight w_d (Eq. 1), p is the exponent of the time weight w_t (Eq. 2), d_{min} and t_{min} represent cutoff values, H is the highest value of the total edge weight in the class, and L is the lowest value of the total edge weight in the class.

Class	T_{max} (days)	D_{max} (km)	r	p	t_{min} (h)	d_{min} (km)	H	L
B	10	30	-1.35	-1	1	1	1.00	$1.01*10^{-5}$
C	10	30	-1.35	-1	0.5	0.2	0.65	$5.76*10^{-7}$
D	30	30	-1.35	-1	1	1	1.00	$3.39*10^{-6}$
E	40	50	-1.35	-1	0.05	0.2	6.07	$7.24*10^{-9}$
F	7	10	-1.35	-1	0.05	0.1	2.72	$1.44*10^{-7}$
G	7	10	-1.35	-1	0.05	0.025	2.72	$2.22*10^{-8}$
H	8	10	-1.35	-1	0.5	0.2	0.65	$3.19*10^{-6}$
I	8	11	-1.35	-1	0.05	0.1	2.72	$1.11*10^{-7}$
J	8	10	-1.35	-1	1	1	1.00	$5.60*10^{-5}$
L	7	10	-1	-0.5	14	2	1.00	$1.39*10^{-2}$
M	7	10	-1	-0.5	1	1	1.00	$1.86*10^{-3}$
N	40	50	-1	-0.5	0.5	0.2	0.66	$2.19*10^{-5}$

The results of the studies that were performed on all classes in Table 1 show that, for each class of networks, the connectivity distribution enjoys power law properties for all networks that have W_{min} in the upper range of the interval between the lowest edge weight value L and the highest edge weight value H in the class, while for networks that have values of W_{min} in the lower range of the interval between L and H, the connectivity

distribution is irregular and scattered. In many of the networks with irregular shapes, constants with a power law tail are present.

For example, a study on the large class E of networks is shown in Fig. 4. The class definition, the initial network E_0 and the networks characteristics are presented in Table 2. The large size of the class (8,488,767 edges) originates in the assumption that, for any earthquake of the network, the interval of influence may go up to 40 days in time, and up to 50 km in space. This is a broad supposition for the active volcanic system of Hawaii; for the majority of these volcanic earthquakes, which do not have large magnitudes (Fig.3), correlations with earthquakes so far away in space and time are quite unlikely. This situation is suggestively illustrated in Fig. 4. The set of six images in Fig. 4 shows the change in shape of the connectivity distribution when the minimum value of the total weight changes from low values, such as in networks E1, E2, E3, towards higher values, as in networks E10, E11, from highly irregular and scattered shapes to well-organized shapes that exhibit significant power law properties. This behaviour is characteristic of networks in all the other classes. When weak links are included (low values of W_{min}), most of the nodes in the emerging networks have little or no relationship with each other, and results show that this choice for network nodes translates in irregular and scattered shapes of the node connectivity. When only strong links are retained (high values of W_{min}), the events selected to participate in the network are primarily earthquakes that are related to each other, and the results show that the underlying properties of seismicity manifest themselves in the well-organized, scale free appearance of the node connectivity.
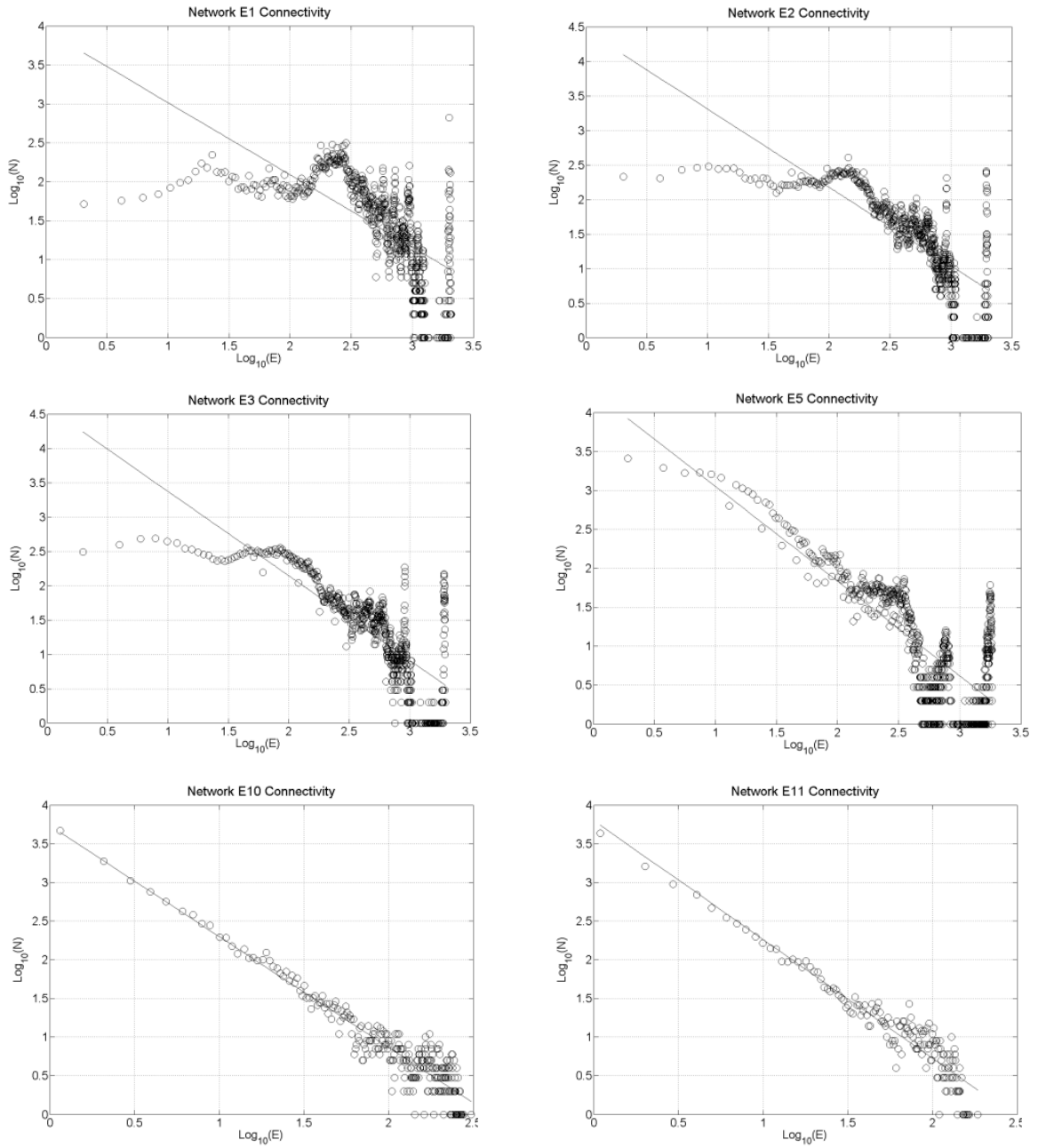
Figure 4. Connectivity distribution for networks E1, E2, E3, E5, E10, and E11. N (on the Y axis) is the number of nodes that have a number E of edges(on the X axis).

Table 2. Class E Networks. E_0 is the initial network that was generated using the parameter values shown in the first column. See Table 1 for the meaning of T_{max}, D_{max}, r, p, d_{min} , t_{min}, H, and L.

Class E definition	Characteristics of the initial network E_0	Network		Number of nodes	Number of edges
		Name	W_{min}		
- T_{max}= 40 days	- 37,441 nodes	E1	10^{-8}	37,441	8,461,301
- D_{max}= 50 km	- 8,488,767 edges	E2	$5*10^{-8}$	37,419	6,628,575
- r = - 1.35	- H = 6.07	E3	10^{-7}	37,357	5,660,485
- p = - 1	- L = $7.24*10^{-9}$	E4	$5*10^{-7}$	36,808	3,782,245
- d_{min} = 0.2 km		E5	10^{-6}	36,172	3,102,765
- t_{min} = 3 min		E6	$5*10^{-6}$	32,886	1,784,667
		E7	10^{-5}	30,377	1,330,446
		E8	$5*10^{-5}$	23,032	584,548
		E9	10^{-4}	20,063	376,852
		E10	$5*10^{-4}$	14,091	117,835
		E11	10^{-3}	11,955	68,408

The threshold values W_{min} are evaluated in the context of the whole set of earthquakes in the class, and not individually. W_{min} values that identify networks of interrelated events emerge from the global assessment of network properties: they are found as those values for which scale free properties appear and become stronger when subsequent networks are created using increasing values of W_{min}. In this sense, the

threshold W_{min} is a global parameter; it belongs to a range of values in the middle to upper zone of the interval between the lowest edge weight value L and the highest edge weight value H in any given class of networks. This represents a major difference between this method and the method of Baiesi and Paczuski (2004, 2005).

Moreover, the generality of this method allows a variety of correlations between earthquakes: any event of the network can have any number of predecessors and any number of successors if the corresponding edges carry enough space-time-magnitude weight, with no arbitrary limitation on magnitude, time, or distance.

Power law properties can also be found in the distributions of time intervals and distances between nodes. Fig. 5 shows the distribution of time intervals in the initial network E_0, i.e. the distribution of all the time intervals between any two earthquakes within a space-time window that is quite large in the Hawaii volcano-tectonic context.
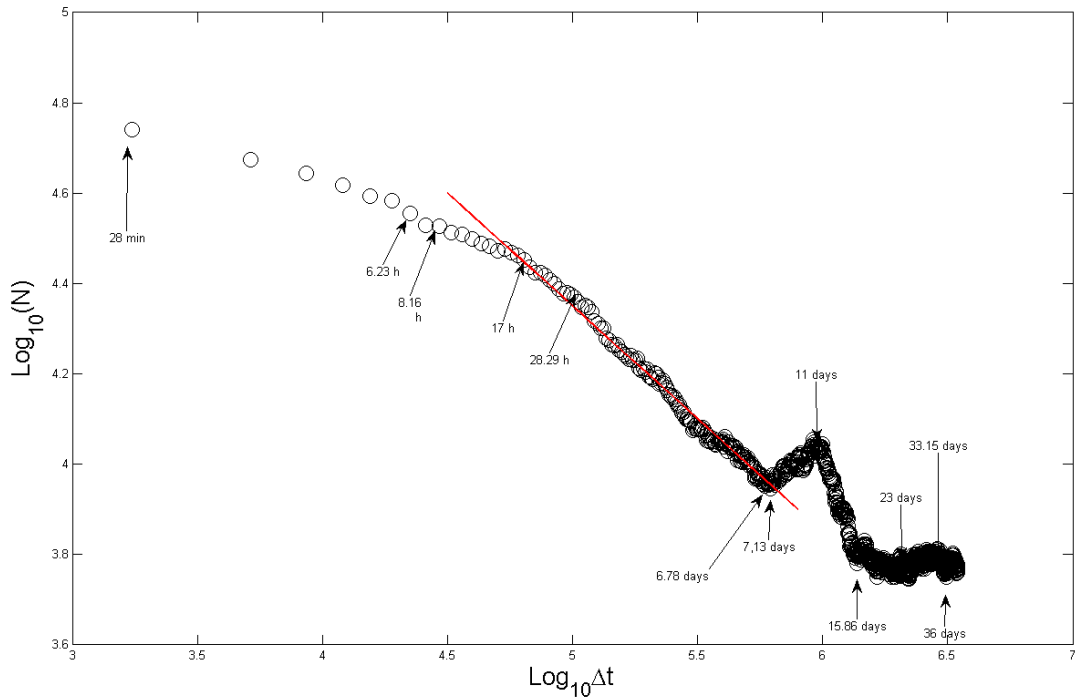
Figure 5. Distribution of time intervals (in seconds) between any two nodes in the initial network E_0. N is the number of time intervals of Δt seconds between any two earthquakes in a space-time window of 50 km and 40 days. The red line represents a reference line with the slope -0.5.

As shown in Fig. 5, there is a distinct scale free zone that goes up to 7 days with a power law exponent of -0.5. The peak between 7 and 15 days with a maximum at 11 days is consistent with studies that show that the precursory sequences in Hawaii follow a power law acceleration with 10-15 days before eruption (Chastin and Main, 2003).
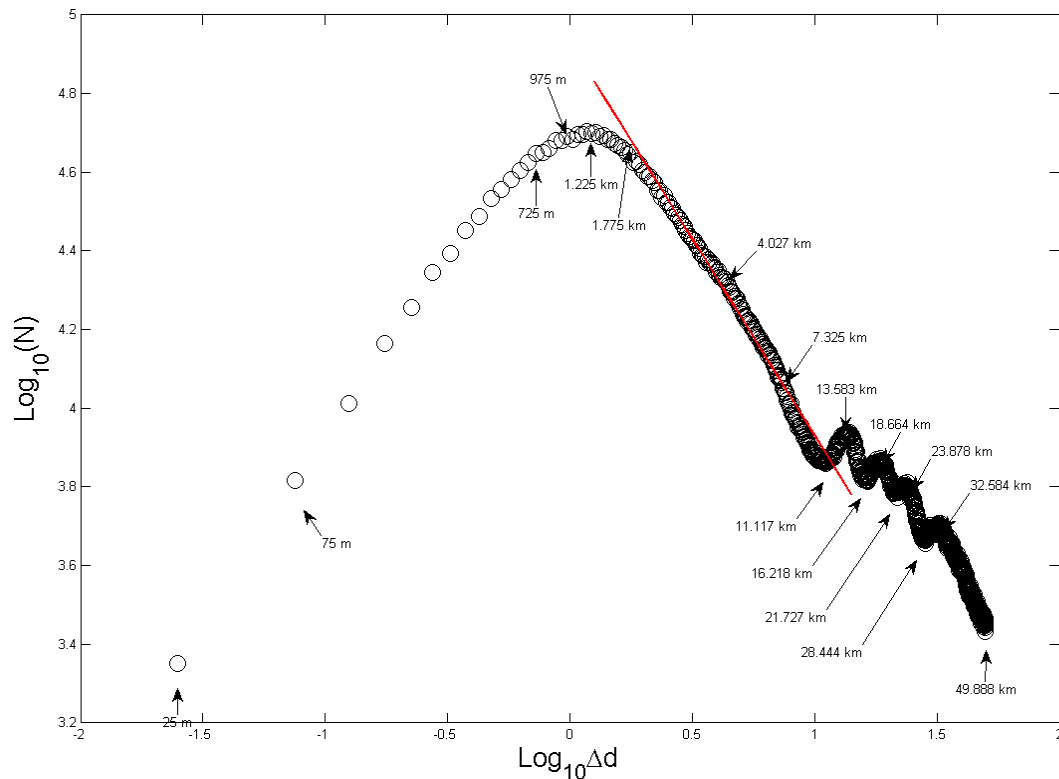
Figure 6. Distribution of distances (in km) between any two nodes in the initial network E_0. N is the number of distances of Δd km between any two earthquakes in a space-time window of 50 km and 40 days. The red line represents a reference line with the slope -1.

The distribution of the distances between events also has scaling properties. Fig. 6 shows the distribution of space intervals in the initial network E_0, which exhibits an exponent of -1 for the power law interval between 1 km and 10 km. Although the context is volcano-tectonic and not all earthquakes in this space-time window are interrelated, the overall shape of the distribution shows remarkable similarities with the distribution of distances of aftershocks from the mainshock in (Lippiello et al., 2009): an increase up to

a maximum value (1 km in this case), followed by a power law decrease. We believe that the peaks with maximums at 13, 18… km refer to events that are spatially clustered around the neighbouring volcanoes, distinct vents and fracture zones, and reflect the spatial characteristics of the Hawaii volcanic system: as shown in Fig.1, the network exhibits intense clusters spatially centered on the volcanoes, their vents and fracture zones, which are situated at distances compatible with the peaks in Fig. 6.

As another example, one of the network classes, class M, was created with characteristics drawn from the two distributions shown in Fig. 5 and Fig. 6: T_{max} = 7 days, p = - 0.5, D_{max} = 10 km, r = -1. The full description of the class M is summarized in Table 3.

Table 3. Class M Networks. M_0 is the initial network that was generated using the parameter values shown in the first column. See Table 1 for the meaning of T_{max}, D_{max}, r, p, d_{min} , t_{min}, H, and L.

Class M definition	Characteristics of the initial network M_0	Network		Number of nodes	Number of edges
		Name	W_{min}		
- T_{max}= 7 days - D_{max}= 10 km - r = - 1 - p = - 0.5 - d_{min} = 1 km - t_{min} = 1 h	- 33,065 nodes - 1,913,280 edges - H = 1 - L = $1.86*10^{-3}$	M1	$1.5*10^{-2}$	26,919	1,270,458
		M2	$2 * 10^{-2}$	25,396	1,088,015
		M3	$3 * 10^{-2}$	22,682	782,659
		M4	$4 * 10^{-2}$	20,440	584,686
		M5	$5 * 10^{-2}$	18,825	443,533
		M6	10^{-1}	14,006	149,235

The choice of the parameter values in the definition of class M was largely based on specific statistical characteristics of the dataset; however, the analysis reveals the same behaviour and the same qualitative patterns as those found in all the other classes. For example, Fig. 7a shows the exponential character of the dependency of the number of edges and number of nodes on the minimum weight in the network, and Fig. 7b shows the power law dependency of the number of edges on the number of nodes for class M of networks.

a.   b.
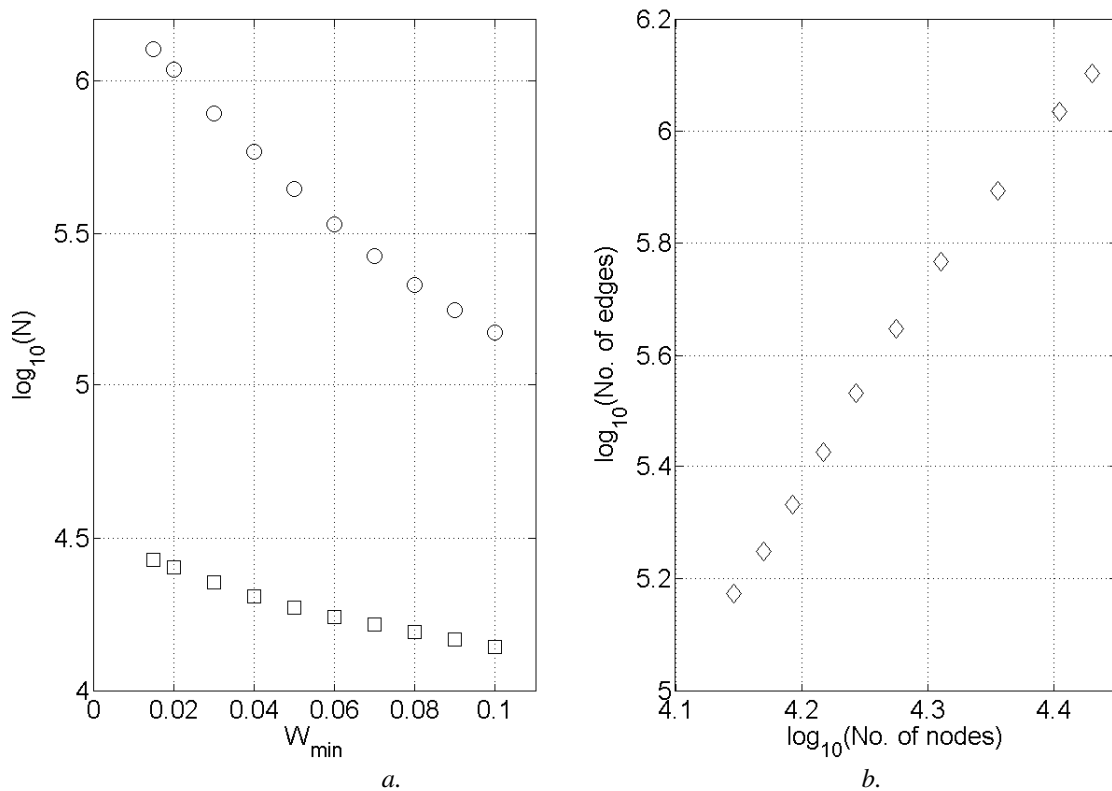
Figure 7. Class M networks. a. The dependency of the number N of edges (circles) and the number N of nodes (squares) on the minimum weight W_{min}. b. The dependency of the number of edges on the number of nodes.

In the example of class M, the connectivity, in-connectivity and out-connectivity distributions are assessed for networks M1 to M6. Since the maximum intervals in time and space T_{max} and D_{max} that define class M have rather low values, inside class M even networks with relatively low values of W_{min} display power law properties, as illustrated in Fig. 8a. The connectivity distributions for the other five networks M2 to M6 are quite similar; another example is shown for network M5 in Fig. 9a. The exponent β for each type of connectivity distribution (all, in, out) is calculated and a graph with all β values is shown in Fig. 10. Similarly, power law properties are detected for the weight, in-weight and out-weight distributions in networks M1 to M6, as shown in Fig. 8b and Fig. 9b, and the exponent γ (Albert and Barabasi, 2002) for each weight distribution is illustrated in Fig. 11.

In general, the analysis shows that the scale free behaviour observed in all networks with superior values of W_{min} inside their class is remarkably robust with respect to variations of parameter values r, p, d_{min} and t_{min}. Moreover, networks of the same class that have lower values of W_{min} exhibit poor scaling characteristics or even no such characteristics at all. The scaling behaviour along the spectrum of W_{min} values, in conjunction with the robustness regarding parameter variations, endorse the idea of a relationship between fundamental properties of seismicity and the scaling characteristics found in earthquake networks.
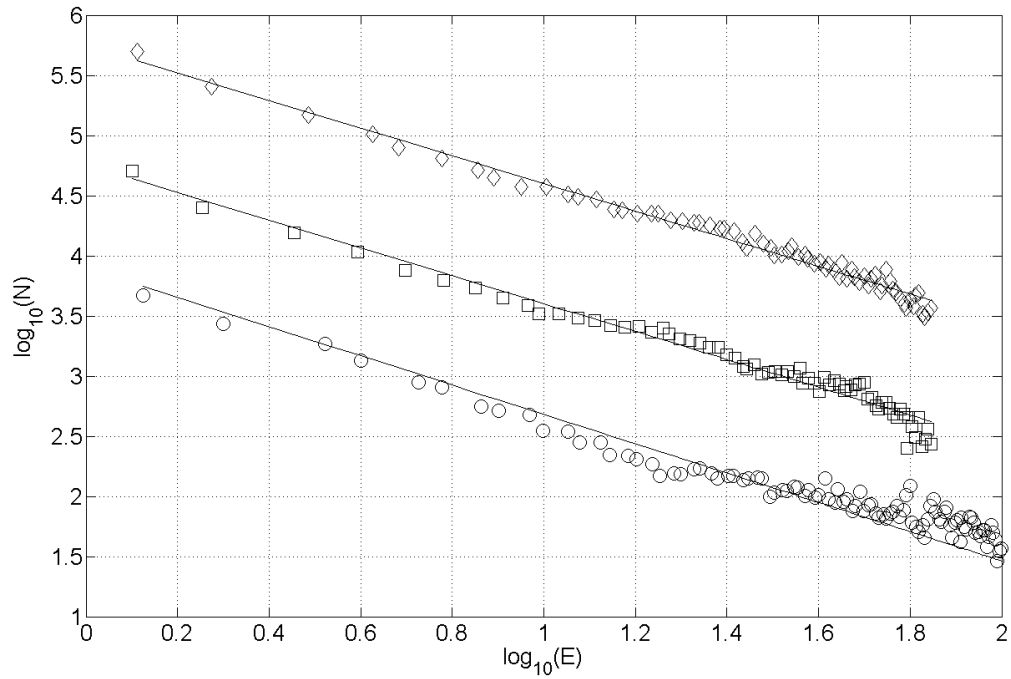
Figure 8 a.
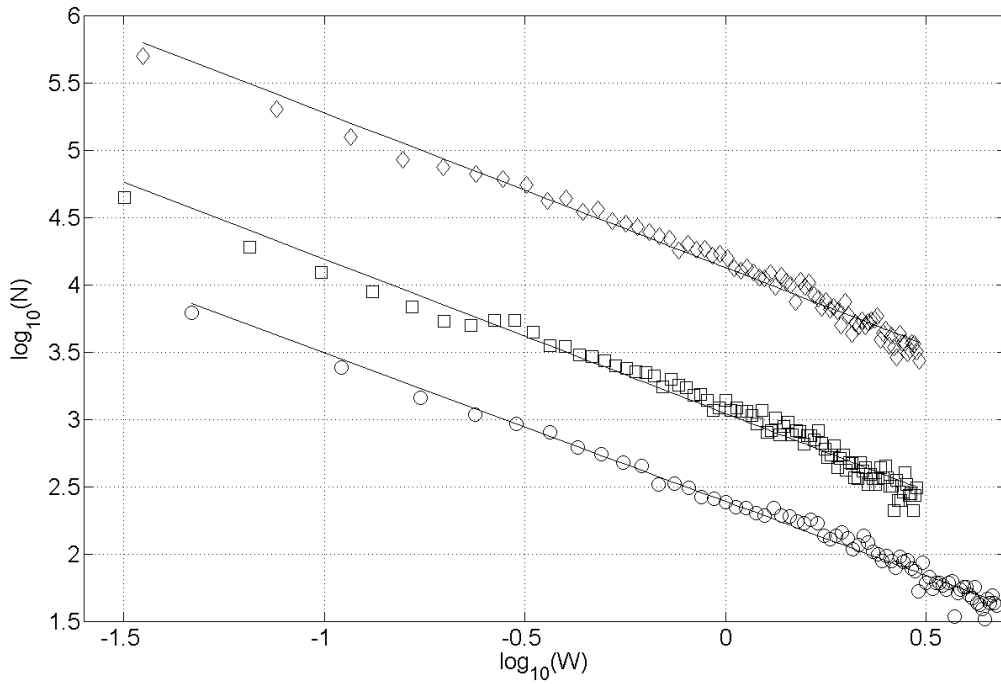
Figure 8 b.

Figure 8. Connectivity and weight distributions for network M1. The curves are shifted by 10 units each along the Y axis. a. Total connectivity (circles), in-connectivity (squares), out-connectivity (diamonds). N is the number of nodes that have a number of edges E (on the X axis). b. Total weight (circles), in-weight (squares), out-weight (diamonds). N is the number of nodes that have a weight of W (on the X axis).
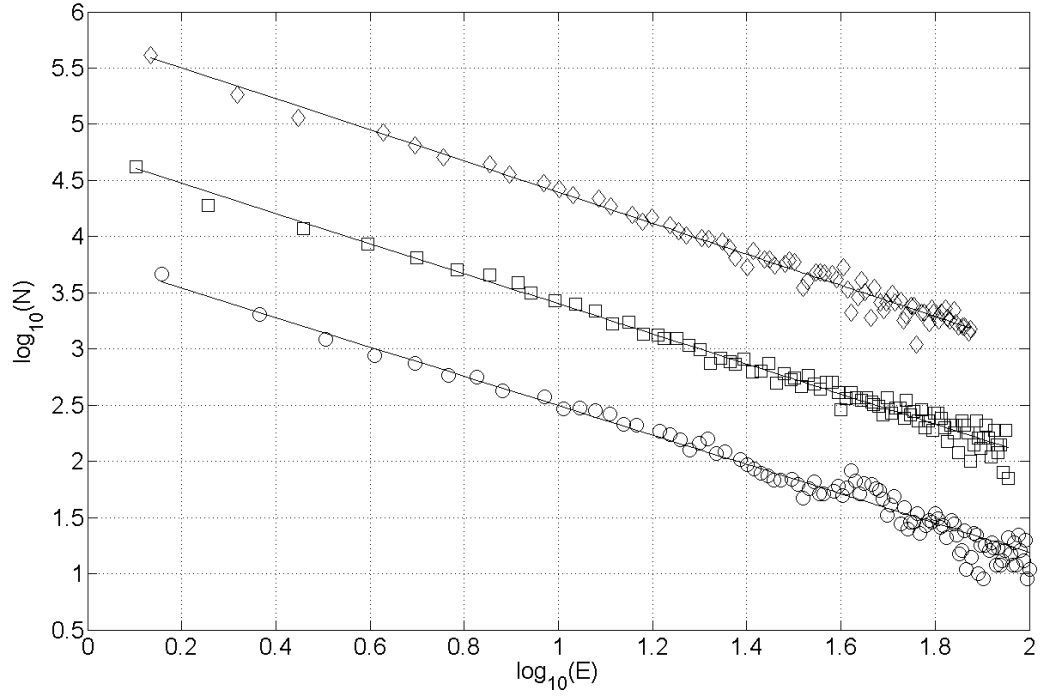
Figure 9 a.
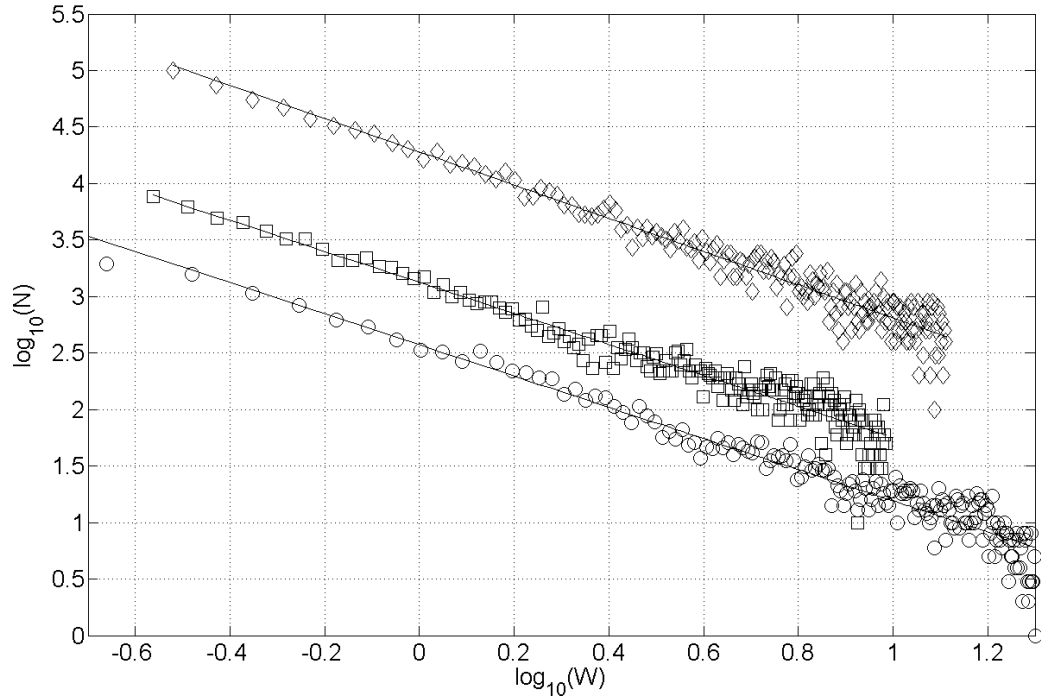
Figure 9 b.

Figure 9. Connectivity and weight distributions for network M5. The curves are shifted by 10 units each along the Y axis. a. Total connectivity (circles), in-connectivity (squares), out-connectivity (diamonds). N is the number of nodes that have a number of edges E (on the X axis). b. Total weight (circles), in-weight (squares), out-weight (diamonds). N is the number of nodes that have a weight of W (on the X axis).
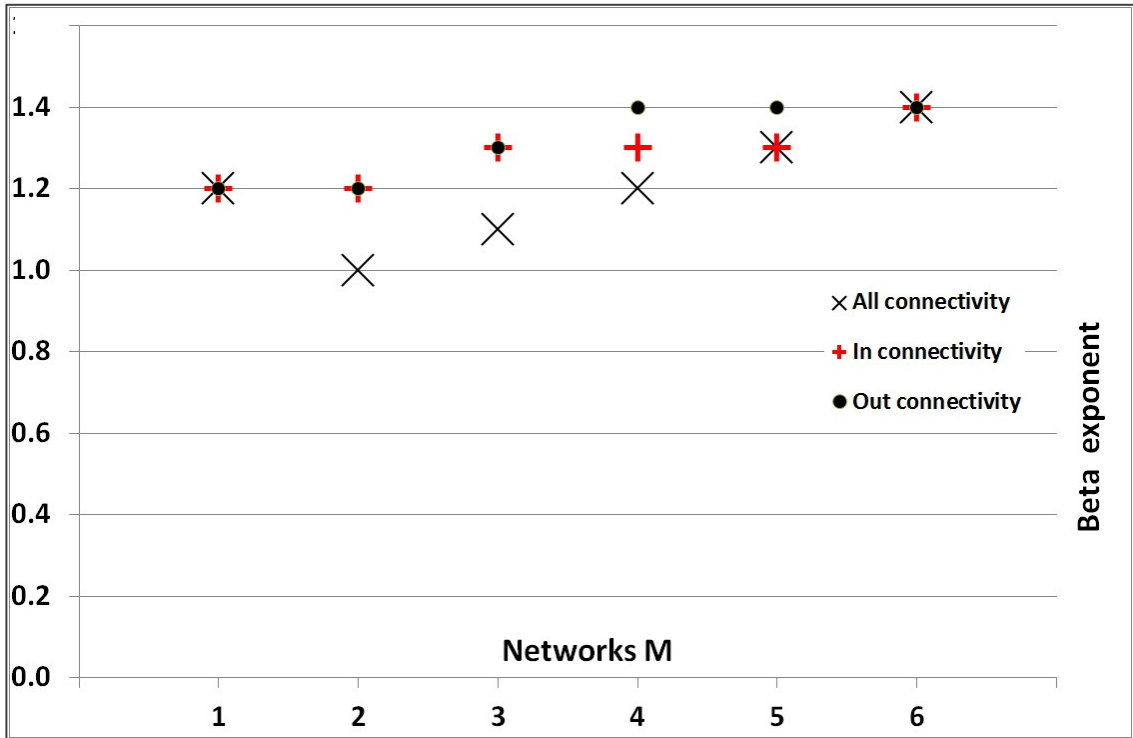
Figure 10. Exponent of connectivity distributions for networks M1 to M6.
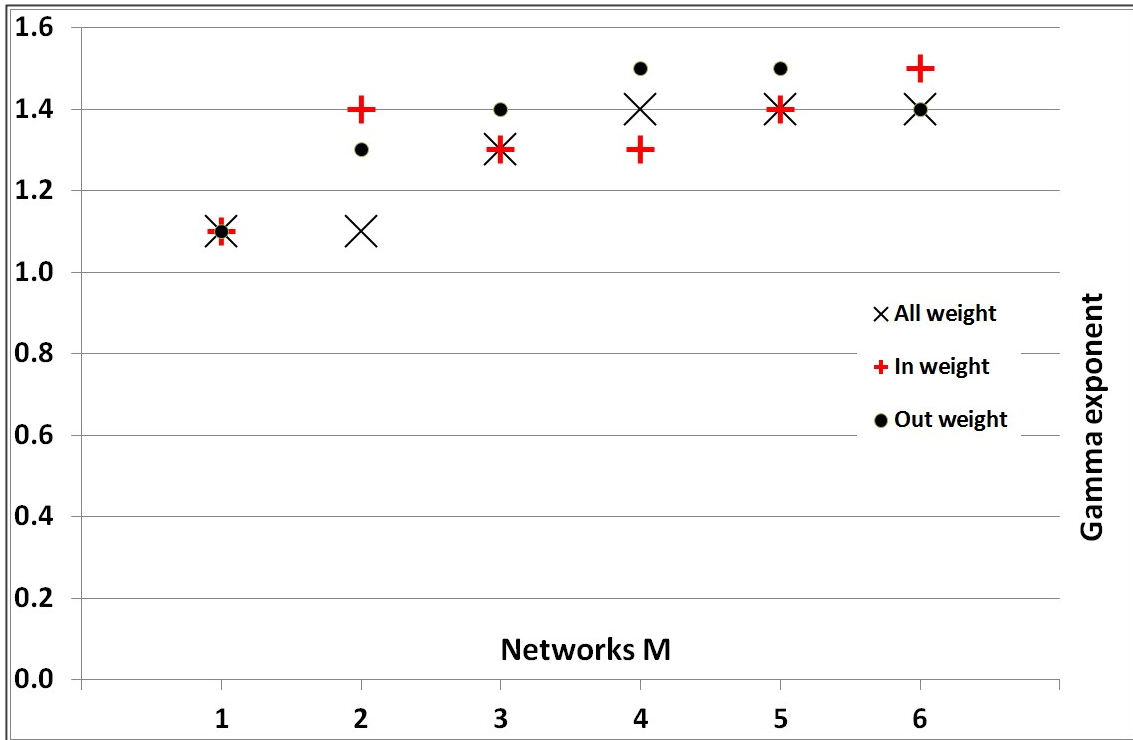
Figure 11. Exponent of the weight distribution for networks M1 to M6.

This observation suggests that a way of testing the reliability of the method is to question the identity of the earthquakes selected in networks that possess scaling properties. If the method is reliable, networks with strong scaling properties should retain only the nodes that correspond to earthquakes that are truly related to each other, regardless of the choice in the parameter values. Since a series of parameter values were explored in this study, a discussion around a few samples of results would be relevant. As shown in Table 1, each specific set of parameter values corresponds to the definition of a class. Table 4 compares four networks with strong scale free properties (D9, J6, M6 and E10) from four different combinations of initial parameters, i.e. from four different classes, with the purpose of showing that all four of them identify the same correlated earthquakes. D9 is the most selective network, having 11,966 nodes (the lowest number of nodes), the next selective is J6, with 13,523 nodes, while M6 and E10 have almost the same number of nodes, 14,006, and 14,091 respectively. Although they come from different classes, with different characteristics, all these networks possess a high value of W_{min} inside their class, and their connectivity distributions enjoy significant power law properties. The question is how many earthquakes selected in the smallest network, D6, have also been selected in the slightly larger networks J6, M6, and E10, then how many earthquakes chosen for the network J6 have also been accepted in the networks M6, and E10, and how many earthquakes included in the network M6 have also been included in the network E10.

Table 4. Robustness of the method: regardless of various choices of parameter values shown under column "Class definition", the method identifies the same statistical population of events as being the earthquakes that are interconnected. In this example, the overlap between the nodes selected by networks D9, J6, M6, and E10 ranges between 91.34% and 100%.

Network	Network characteristics	Comparative node selection	Class	Class definition	Class characteristics
D9	11,966 nodes $W \geq 7.5*10^{-3}$	11,966 nodes in J6 (100%) 11,966 nodes in M6 (100%) 11,813 nodes in E10 (98.72%)	**D**	T_{max}= 30 days D_{max}= 30 km r = - 1.35 p = - 1 d_{min} = 1 km t_{min} = 1 h	37,325 nodes 5,801,083 edges H = 1.00 L = $3.39*10^{-6}$
J6	13,523 nodes $W \geq 5*10^{-2}$	13,154 nodes in M6 (97.27%) 12,702 nodes in E10 (93.93%)	**J**	T_{max}= 8 days D_{max}= 10 km r = - 1.35 p = - 1 d_{min} = 1 km t_{min} = 1 h	33,447 nodes 2,035,257 edges H = 1.00 L = $5.60*10^{-5}$
M6	14,006 nodes $W \geq 10^{-1}$	12,793 nodes in E10 (91.34%)	**M**	T_{max}= 7 days D_{max}= 10 km r = - 1 p = - 0.5 d_{min} = 1 km t_{min} = 1 h	33,065 nodes 1,913,280 edges H = 1.00 L = $1.86*10^{-3}$
E10	14,091 nodes $W \geq 5*10^{-4}$		**E**	T_{max}= 40 days D_{max}= 50 km r = - 1.35 p = - 1 d_{min} = 0.2 km t_{min} = 3 min	37,441 nodes 8,488,767 edges H = 6.07 L = $7.24*10^{-9}$

The results in Table 4 show that all the earthquakes selected in the network D9 have also been selected in the networks J6 and M6, and 98.72% have also been included in E10, regardless of the variations in the choice of the parameter values. In network J6, 97.27% of the earthquakes have also been selected by the slightly larger network M6, and 93.93% of the events have also been chosen for the network E10. Finally, 91.34% of the events included in the network M6 have also been included in the network E10. We are actually looking, in each of these cases, at the same statistical population of earthquakes. These are the earthquakes that, being close enough in space-time-magnitude, are most likely to be related to each other.

This is an interesting result, indicating that the method is reliable, robust with respect to variations in parameter values, and reflects fundamental properties of seismicity. Consequently, the process of identification of correlations between earthquakes can start with a certain choice of parameter values (class definition), and end when networks with scale free properties are found.

4.2 Evolution of network properties over time

Results of this method are further applied for the study of the way the relationships between earthquakes change over time. The network is split up in successive event windows, each window having the same number of successive events. The first objective of the analysis is to determine whether scaling properties can be identified in the temporal windows. If that is the case, the next objective is to study whether changes of scaling properties in successive temporal windows can be related to real-life changes in the volcanic system.

Numerous networks from different classes have been studied. The networks chosen for the analysis were those with strong scaling properties. They were split in event windows of 1,000 successive events and sub-networks of 1,000 nodes were generated accordingly. In each case, the node connectivity distribution and the node weight distribution were assessed.

The results (not shown) confirm that also these distributions manifest power law characteristics; for each scaling regime, the corresponding exponents β and γ were calculated. As an example, the study of successive event windows in the network M2 is presented in Fig. 12 (the variation of the connectivity distribution exponent β) and Fig. 13 (the variation of the weight distribution γ).
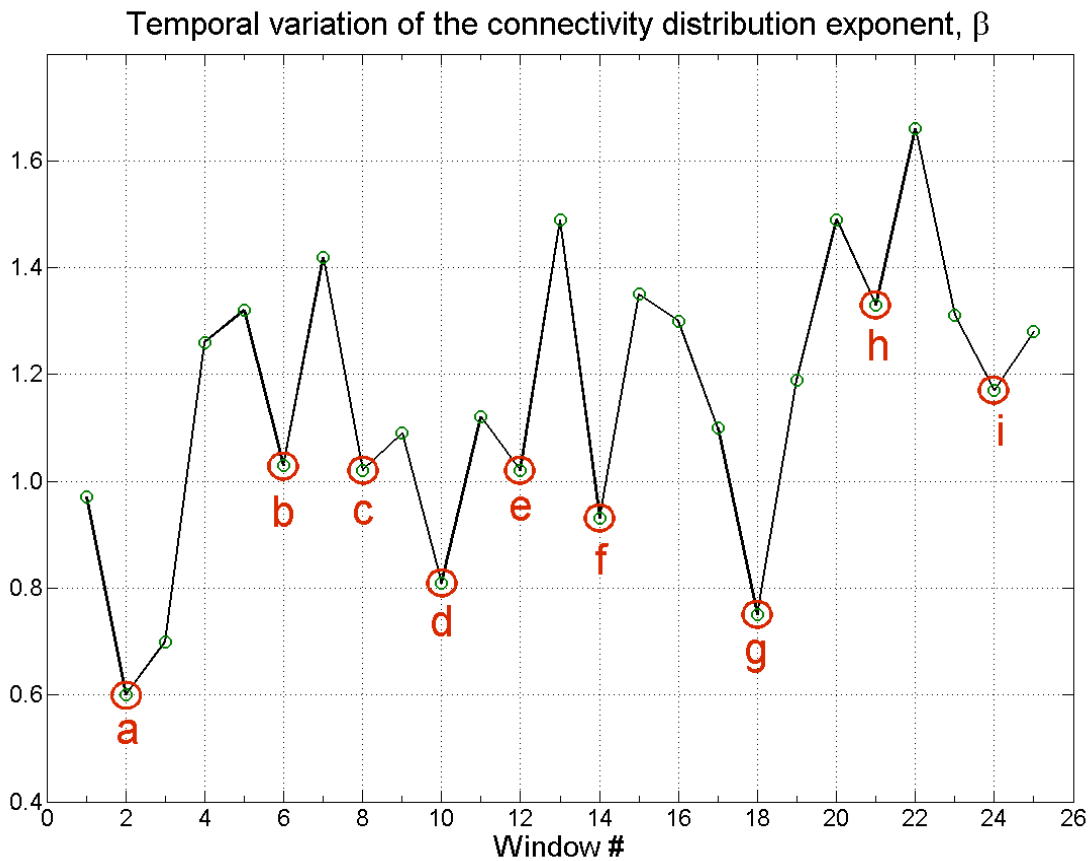
Figure 12. Temporal variation of the connectivity distribution exponent in successive temporal windows of network M2.
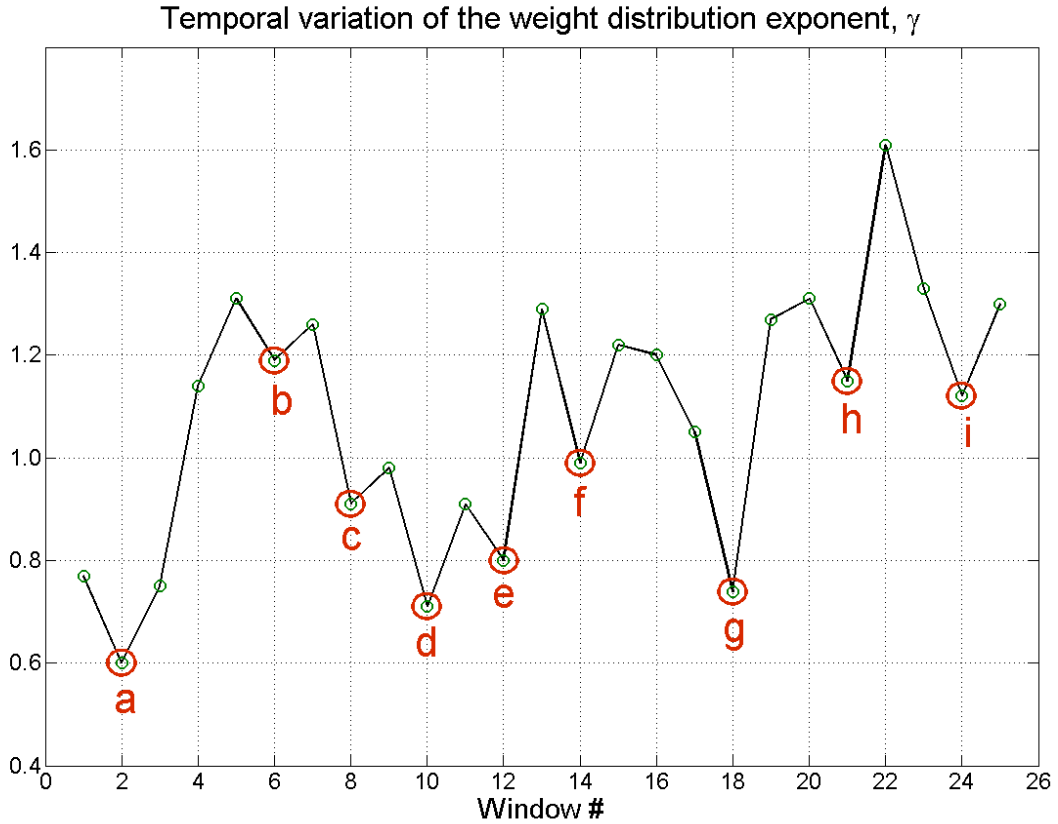
Figure 13. Temporal variation of the weight distribution exponent in successive temporal windows of network M2.

Although the graphs of the two distributions are not identical, they show the same trend in their evolution. In this example, the minimum values in the variation of β correspond to the windows where γ also has minimum values; therefore, the same lower case letters from "a" to "i" were used on both graphs to tag the corresponding minimum values of the two exponents.

Fig. 14 presents the graph of the cumulative number of earthquakes from January 1989 to December 2012. On this graph, the areas corresponding to the temporal windows tagged with letters from "a" to "i" in Fig. 13 and Fig. 14, were tagged with the same letters. Each of the labeled areas in Fig. 14 is therefore associated with the minimum values of β and γ in successive temporal windows. A steep increase in the number of

earthquakes can be noticed in Fig. 14 for the tagged areas. The history of the volcano shows that sudden events, with important discharges of energy, such as rapid openings of new fissures, violent massive eruptions or explosions, occurred in the volcano in the corresponding time intervals.
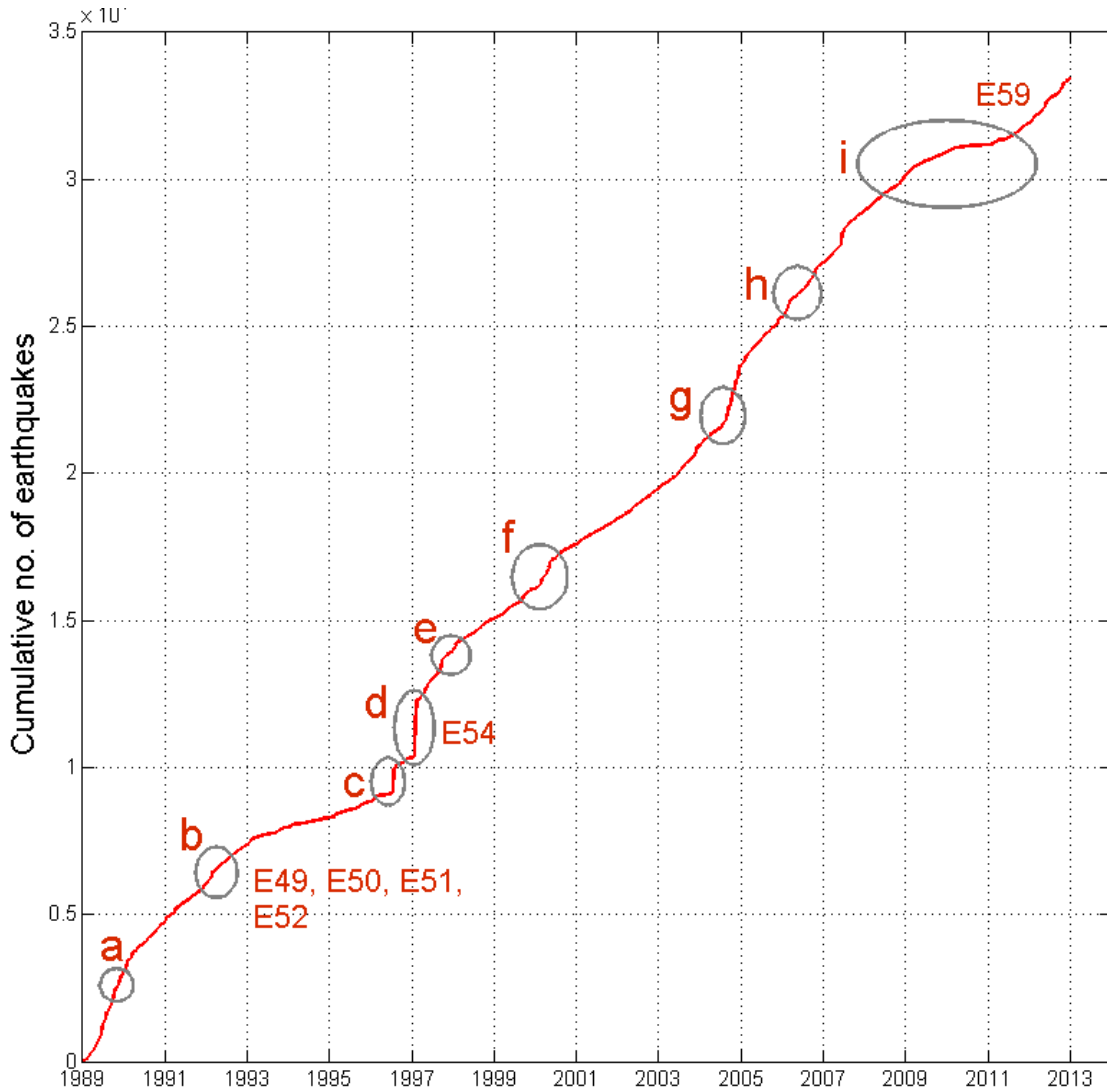
Figure 14. Cumulative number of earthquakes between January 1989 and December 2012. The small letters tag the areas corresponding to the minimum values of the exponents β and γ in successive temporal windows of network M2.

For example, minimum "a", corresponding to window number 2, can be related to the braking of the Kupaianaha tube system in 1989 that caused massive surface lava

flows; these surface lava flows spectacularly invaded new territory, overran the Waha`ula Visitor Center and residences in Hawaii Volcanoes National Park. Minimum "b", corresponding to window number 6, can be related to the four episodes that occurred between the end of 1991 and the end of 1992: episodes 49, 50, 51, and 52. New fissures developed with these four episodes, and were accompanied by important seismic phenomena, as for example the 4.5 magnitude earthquake that preceded the episode 52 from October 1992. As another example, minimum "d", corresponding to window number 10 can be related to episode 54 from 1997, when another new fissure developed, lava fountains reached tens of meters in height, and a period with 2,000 to 4,000 earthquakes per day followed.

Each of these minimums can be related to such sudden events, with surges in the activity of the volcano, important discharges of energy, and changes in seismicity. However, the steep slope situated between "h" and "i" in Fig. 14, which is probably due to episode 58 from July 2007, cannot be related to a minimum in the values of β and γ. A possible explanation for this exception is the process of artificially breaking down the network M2 in sub-networks with an equal number of nodes. A study on the optimization of the temporal windows selection should address this issue and is subject to further research.

The meaning of the minima in the exponents β and γ is an increased connectivity in the corresponding networks; the proportion of nodes that have high connectivity is larger. In the studied context, energy dissipates through various processes such as magma flows, lava effusion, explosions, heat emission, tectonic phenomena, degassing, etc. (Wright and Pilger, 2008). Although the increased connectivity in minima of the

exponents β and γ is not consistently related to higher dissipation in tectonic energy, it could be related to peaks in the overall energy emitted by the volcanic system during eruptions or large outpourings of lava. An analysis of the clustering coefficient and of associations with energy dissipation in the system is subject of future research.

Overall, the study shows that variations in the values of the exponents β and γ are able to reflect the way the relationships between earthquakes are changing over time. Minimum values of β and γ in successive temporal windows can be related to important events in the life of the volcanic system and the associated seismicity.

5 Conclusions

A new type of directed networks has been proposed for the assessment of relationships between earthquakes. The method was applied to volcanic seismicity in Hawaii. The nodes of the networks are epicenters of earthquakes; the edges that link the nodes carry space-time-magnitude weights, and have a direction given by the temporal succession of the events. The generality of the definition of the edge weight, W, as a combination of a factor in time, a factor in space, and a factor in magnitude is comprehensive and permits various combinations of space-time-magnitude correlations between earthquakes. Since any node can have any number of edges that enter the node and any number of edges that leave the node, any given event may have multiple predecessors, and any given event can contribute to multiple future events, as long as its edges carry enough weight.

Parameters and formulas used in the calculation of the weights take into consideration well-established properties of seismicity. High values of W are associated

with strong relationships between earthquakes, while low values of W are associated with either weak relationships or no relationships at all. Various classes of networks can be generated based on distinct values of the parameters. Inside each class, different networks can be created by setting different thresholds for the minimum edge weight W_{min}.

It is shown that networks that have W_{min} in the middle to upper range of the interval between the lowest edge weight value L and the highest edge weight value H in their class manifest significant scaling properties of node connectivity distributions, as opposed to networks with low values of W_{min}, which exhibit poor or no scaling characteristics. Since high values of weight describe the strong links, the events selected in the networks with high values of W_{min} are primarily the earthquakes that are most likely to be related to each other. Therefore, it is reasonable to see a relationship between the fundamental characteristics of seismicity and the well-organized, scale free distributions of node connectivity. In networks with low values of W_{min}, most of the nodes have little or no relationship with each other. In this context, the irregular and scattered shapes of their connectivity distributions are not a surprise.

It is also shown that the scale free behaviour observed in networks with superior values of W_{min} is robust with respect to variations in parameter values. Tests performed on networks that manifest strong power law properties, but originating in different choices of parameter values, confirm the reliability of the method. They show that the same statistical population of earthquakes is chosen to participate in these networks, i.e. the earthquakes most likely to be interrelated. The results indicate that the method is reliable, robust with respect to variations of parameter values, and reflects fundamental properties of seismicity.

The threshold values W_{min} that identify networks of interrelated events are assessed in the context of all the earthquakes in the class: they are found as those values for which scale free properties appear and become stronger when subsequent networks are created using increasing values of W_{min}. It can be said that the threshold W_{min} is a global parameter that characterizes the set of earthquakes and its values are meaningful only inside that set.

There are also other significant scaling properties that are detected in the analysis of the classes of networks. Node weight distributions also enjoy scaling properties. For each class, the dependency of the number of edges on the number of nodes is a power law. The distribution of the distances between events exhibits distinct regimes with scale free properties. Similarly, the distribution of the time intervals between events is characterized by different domains with scaling properties as well.

It is also shown that the evolution of the relationships among earthquakes over time can be studied by splitting up the network in successive event windows with an equal number of nodes. The distributions of node connectivity and node weight in the emerging sub-networks manifest scaling properties that can be used to follow the evolution of seismicity over time. The exponents β and γ of these distributions have a similar trend in their evolution over the temporal windows. The increased connectivity in minima of β and γ can be associated with sudden, important discharges of energy in the life of the volcanic system and accompanying earthquakes. It is shown that the exponents of connectivity and weight distributions for successive event windows are able to reflect the way the relationships between earthquakes are changing over time. Aspects regarding

the clustering coefficient, energy dissipation and optimization of the selection of the temporal windows are subject to further research.

References

Albert, R., and Barabási, A.-L.: Statistical mechanics of complex networks, Rev. Mod. Phys., 74, 47-97, 2002.

Bak, P., Christensen, K., Danon, L., and Scanlon, T.: Unified scaling law for earthquakes, Phys. Rev. Lett., 88, doi 10.1103/PhysRev Lett. 88.178501, 2002.

Baiesi, M., and Paczuski, M.: Scale-free networks for earthquakes and aftershocks, Phys. Rev. E, 69, 066106-1–8, 2004.

Baiesi, M., and Paczuski, M.: Complex networks of earthquakes and aftershocks, Nonlinear Proc. Geoph. 12, 1-11, 2005.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U.: Complex networks: structure and dynamics, Phys. Reports, 424, 175-308, 2006.

Bunde, A., and Lennartz, S.: Long-term correlations in earth sciences, Acta Geophys., 60, 3, 562-588, 2012.

Carbone, V., Sorriso-Valvo, L., Harabaglia, P., Guerra, I.: Unified scaling law for waiting times between seismic events, Europhys. Lett., 71, 6, 1036–1042, 2005.

Chastin, S.F.M., and Main, I.G.: Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions, Geophys. Res. Lett., 30, 13, 1671. doi:10.1029/2003GL016900, 2003.

Davidsen, J., Grassberger, P., and Paczuski, M.: Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity, Phys. Rev. E, 77, 066104, 2008.

Felzer, K. R., and Brodsky, E. E.: Decay of aftershock density with distance indicates triggering by dynamic stress, Nature, 441, 735-738, 2006.

Gardner, J. K., and Knopoff, L.: Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, Bull. Seis. Soc. Am., 64, 5, 1363–1367, 1974.

Gutenberg, B., and Richter, C. F.: Seismicity of the Earth. Princeton University Press, Princeton, 1954.

Knopoff, L., and Gardner, J. K.: Higher seismic activity during local night on the raw worldwide earthquake catalogue, Geophys. J. R. Astr. Soc., 28, 311–313, 1972.

Lapenna, V., Macchiato, M., Piscitelli, S., and Telesca L.: Scale invariance properties in seismicity of Southern Apennine Chain (Italy), Pure Appl. Geophys., 157, 4, 589–602, 2000.

Lei, X., and Kusunose, K.: Fractal structure and characteristic scale in the distributions of earthquake epicentres, active faults and rivers in Japan, Geophys. J. Int., 139, 3, 754–762, 1999.

Lennartz, S., Livina, V. N., Bunde, A., and Havlin S.: Long-term memory in earthquakes and the distribution of interoccurrence times, Europhys. Lett., 89, 69001, doi: 10.1209/0295-5075/81/69001, 2008.

Lippiello, E., Corral, A., Bottiglieri, M., Godano, C., and de Arcangelis, L.: Scaling behavior of the earthquake intertime distribution: Influence of large shocks and time scales in the Omori law, Phys. Rev. E, 86, 6-2, 066119, doi:10.1103/PhysRevE.86.066119, 2012a.

Lippiello, E., de Arcangelis, L., and Godano, C.: The role of static stress diffusion in the spatio-temporal organization of aftershocks, Phys. Rev. Lett., 103, 038501, doi:10.1103/PhysRevLett.103.038501, 2009.

Lippiello, E., Godano, C., and de Arcangelis, L.: The earthquake magnitude is influenced by previous seismicity, Geophysical Research Letters, 39, 5, doi: 10.1029/2012GL051083, 2012b.

Nanjo, K., and Nagahama, H.: Spatial distribution of aftershocks and the fractal structure of active fault systems, Pure Appl. Geophys., 157, 4, 575–588, 2000.

Omori, F.: On the aftershocks of earthquakes, J. College of Science, Imperial University of Tokyo, 7: 111–200, 1894.

Reasenberg, P.: Second-order moment of central California seismicity, 1969-1982, J. Geophys. Res., 90, 5479– 5495, 1985.

Richards-Dinger, K., Stein, R., and Toda, S.: Decay of aftershock density with distance does not indicate triggering by dynamic stress, Nature, 467, 583-586, 2010.

Sanchez, L., and Shcherbakov, R: Temporal scaling of volcanic eruptions, J. Volcanol. Geoth. Res., 247-248, 115-121, 2012.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B.: A generalized Omori's law for earthquake aftershock decay, Geophys. Res. Lett., 31, L11613, doi 10.1029/2004GL019808, 2004.

Shcherbakov, R., Turcotte, D. L., and Rundle, J. B.: Scaling properties of the Parkfield aftershock sequence, B. Seismol. Soc. Am., 96, 4B, 376–S384, doi: 10.1785/0120050815, 2006.

Suteanu, C., and Suteanu, M.: Identification of change in spatio-temporal patterns: two multiscale approaches to volcanic seismicity, Rom. Geophys. J., 55, 7-16, 2011.

Turcotte, D.: Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge, 1977.

Wright, R., and Pilger, E.: Radiant flux from Earth's subaerially erupting volcanoes, International Journal of Remote Sensing, 29, 22, 6443-6466, 2008.

Chapter 5

Aspects of Structure in Earthquake Networks

Submitted for publication:

Abstract

Analysis performed on multiple sets of earthquake networks created for the Hawaii volcanic system reveals characteristics that can be associated with fundamental properties of seismicity. The scale free behaviour of the connectivity distribution along the spectrum of the minimum weight values, which can be used to discern the interrelated earthquakes from the rest of the data set, is mirrored by a similar behaviour of the distribution of the number of linked neighbours. The patterns found in the distributions of temporal and spatial intervals between earthquakes are similar from large to small networks. Similarities are found between the variation of the network clustering coefficient, C, and the variation of the exponents of the connectivity distribution, β, and of the weight distribution, γ; their synchronous variation over successive temporal windows can be related to changes in seismicity and in the life of the volcanic system. A Zipf distribution is found for the ranked sets of magnitude values of successive network nodes. The distribution of differences between the magnitude values of successive nodes is also governed by a power law.

Keywords: seismicity, earthquakes, networks, Hawaii volcanoes, nonlinear systems, scaling properties.

1 Introduction

Current research dedicated to understanding seismicity and its ruling phenomena illustrates various aspects of the correlations found in earthquake patterns. Over the past decade, approaches based on complex networks revealed that networks built on earthquake data enjoy scaling properties (Baiesi and Paczuski, 2004; Baiesi and Paczuski, 2005; Davidsen et al., 2008; Suteanu, 2014). This paper presents an analysis of earthquake networks introduced by Suteanu (2014) and used for the study of volcanic seismicity in Hawaii. An objective of this research is to determine whether the distributions of the spatial and temporal distances between connected nodes have a similar pattern in different networks, and, if that is the case, whether the pattern enjoys scaling properties. Since the network average clustering coefficient (C), the exponent of the connectivity distribution (β), and the exponent of the weight distribution (γ) are global parameters, which characterize the network as a whole, we investigate if any coherent relationship can be found between them. We also inquire whether distributions of magnitude values for successive nodes of a network enjoy specific properties, and, if that is the case, whether those properties are characteristic of all networks.

2 The earthquake networks

In this analysis we generate directed weighted networks of earthquakes to identify interrelated events following the method in Suteanu (2014). The source of data is the Advanced National Seismic System (ANSS) catalogue for the Big Island of Hawaii between January 1st, 1989 and December 31st, 2012. For catalogue completeness, only the earthquakes with a magnitude $m \geq 1.6$ are used in the analysis (37,451 events). The b

value in the Gutenberg-Richter magnitude-frequency distribution is b ≈ 0.99. The nodes of the networks are epicenters of earthquakes and are linked by edges that carry space-time-magnitude weights, with a direction given by the temporal succession of the events. The edge weight, *W*, is a combination of a variable in space, a variable in time, and a variable in magnitude. All three weight components are seen independently, as separate ingredients with comparable contributions to the total edge weight and a maximum value limited to 1. This approach is particularly important in the case of volcano-tectonic seismicity, where the seismic sources are diverse (tectonic stress, thermodynamic processes, dynamics of gas, fluid and solid).

Given the omnipresence of scaling relationships in earthquake distributions (Omori, 1894; Utsu., 1961; Kagan, 1994; Nanjo and Nagahama, 2000; Lapenna et al., 2000; Shcherbakov et al., 2004; Carbone et al., 2005; Felzer and Brodsky, 2006; Shcherbakov et al., 2006; Lennartz et al., 2008; Bunde and Lennartz, 2012; Lippiello et al., 2012a; Lippiello et al., 2012; Varotsos et al., 2012), power laws are used for the quantitative estimation of the relationships between earthquakes in space and time:

a) Distance weight:

$$w_d = cd^r, r < 0, \tag{1}$$

where *d* is the spatial distance between the two nodes of an edge measured in km, and *c* is a positive constant.

b) Time weight:

$$w_t = st^p, p < 0, \tag{2}$$

where *t* is the time interval between the two nodes of an edge measured in hours, and *s* is a positive constant.

Assuming that all earthquakes that are very close in space or in time could be related to each other, and in order to avoid singularities, a small cutoff value is used for the weights in space and time. Therefore, the method uses modified forms of Equations (1) and (2):

$$w_d = \begin{cases} 1, & d \le d_{min} \\ \\ cd^r, & d \ge d_{min}, \quad r < 0 \end{cases} \tag{1'}$$

$$w_t = \begin{cases} 1, & t \le t_{min} \\ \\ st^p, & t \ge t_{min}, \quad p < 0 \end{cases} \tag{2'}$$

The constants c and s are calculated using the boundary conditions:

$$w_d = cd_{min}^r = 1 \tag{1''}$$

and

$$w_t = st_{min}^p = 1 \tag{2''}$$

c) The magnitude weight is proportional to the magnitude of the first occurring event of the edge:

$$w_m = \frac{m}{m_{max}}, \tag{3}$$

where m is the magnitude of the first occurring earthquake associated with the edge, and m_{max} is the maximum magnitude value in the data set.

The total weight of an edge is given by the product of the weights in space, time, and magnitude, but only the nodes that carry enough weight are selected in the network

of correlated events, therefore only edges with a total weight W higher than a minimum threshold W_{min} are chosen for this network:

$$W = \begin{cases} w_d \cdot w_t \cdot w_m, & W \geq W_{min} \\ 0, & W < W_{min} \end{cases} \qquad (4)$$

This definition allows various combinations of space-time-magnitude correlations between any two events: any node can have any number of predecessors and any number of successors, as long as its edges carry enough total weight. An example of an earthquake network is illustrated in Fig. 1.

A maximum interval of influence in time T_{max}, and a maximum interval of influence in space D_{max} are assigned in order to simplify the computation, but, hypothetically, T_{max} and D_{max} may cover the whole extent of the catalogue: large spatial and temporal intervals between events produce very small values of the edge weights w_d and w_t, and, implicitly, of the total edge weight W; the resulting weak links are eventually eliminated by the network definition (4).

The network classes are sets of earthquake networks that share the same values of the parameters D_{max}, r, d_{min}, T_{max}, p, t_{min}. A summary of all classes that have been studied is presented in Table 1. An initial network is created in every class when specific values are assigned to the above parameters, but, in general, this first network does not hold predominantly correlated earthquakes, it represents mainly a collection of events that serves for the initiation of the method. An index $_0$ is used to differentiate the initial networks B_0, C_0, …, P_0 from the rest of the networks. Inside each class, a series of networks is generated by setting various threshold values for the minimum edge weight W_{min}.
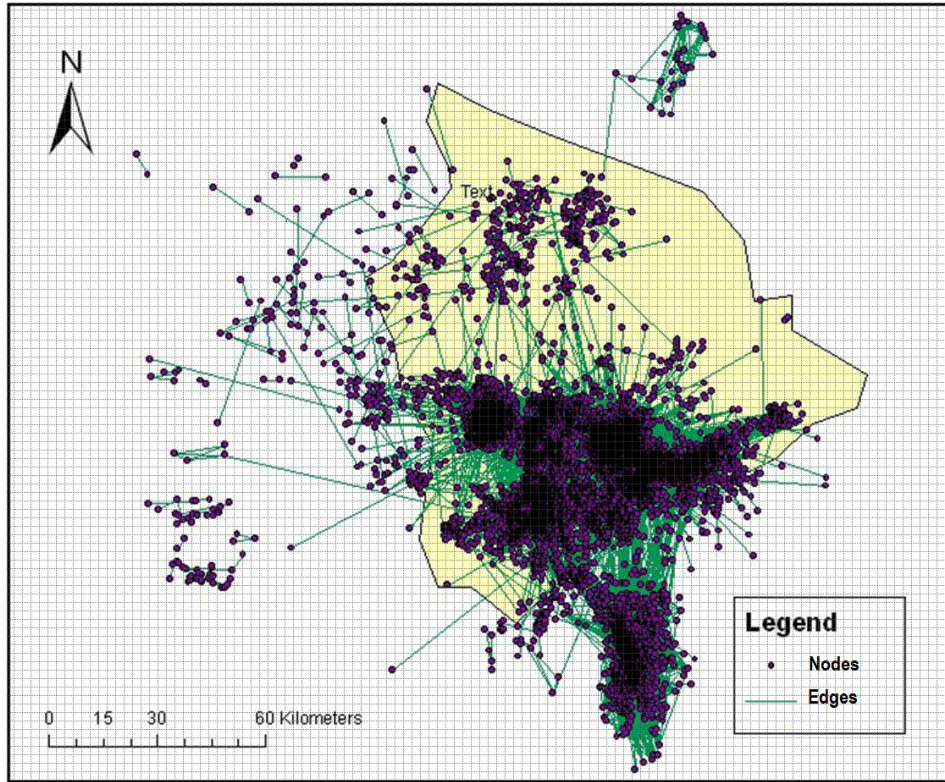
Figure 1a.
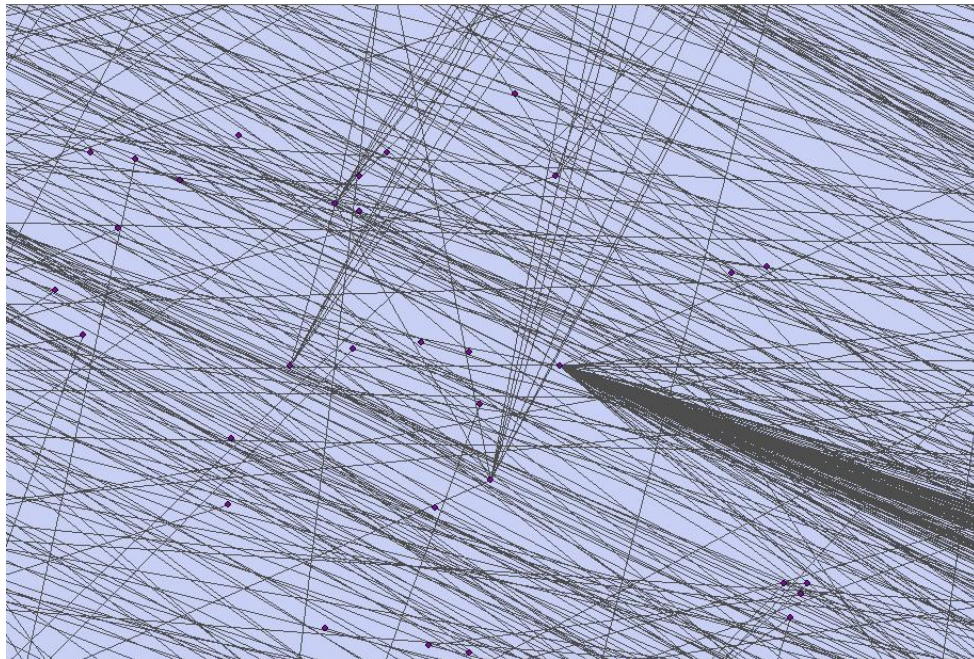
Figure 1b.

Figure 1. The earthquake network. (a) An example for the Big Island of Hawaii. (b) Zoomed-in example of nodes and edges.
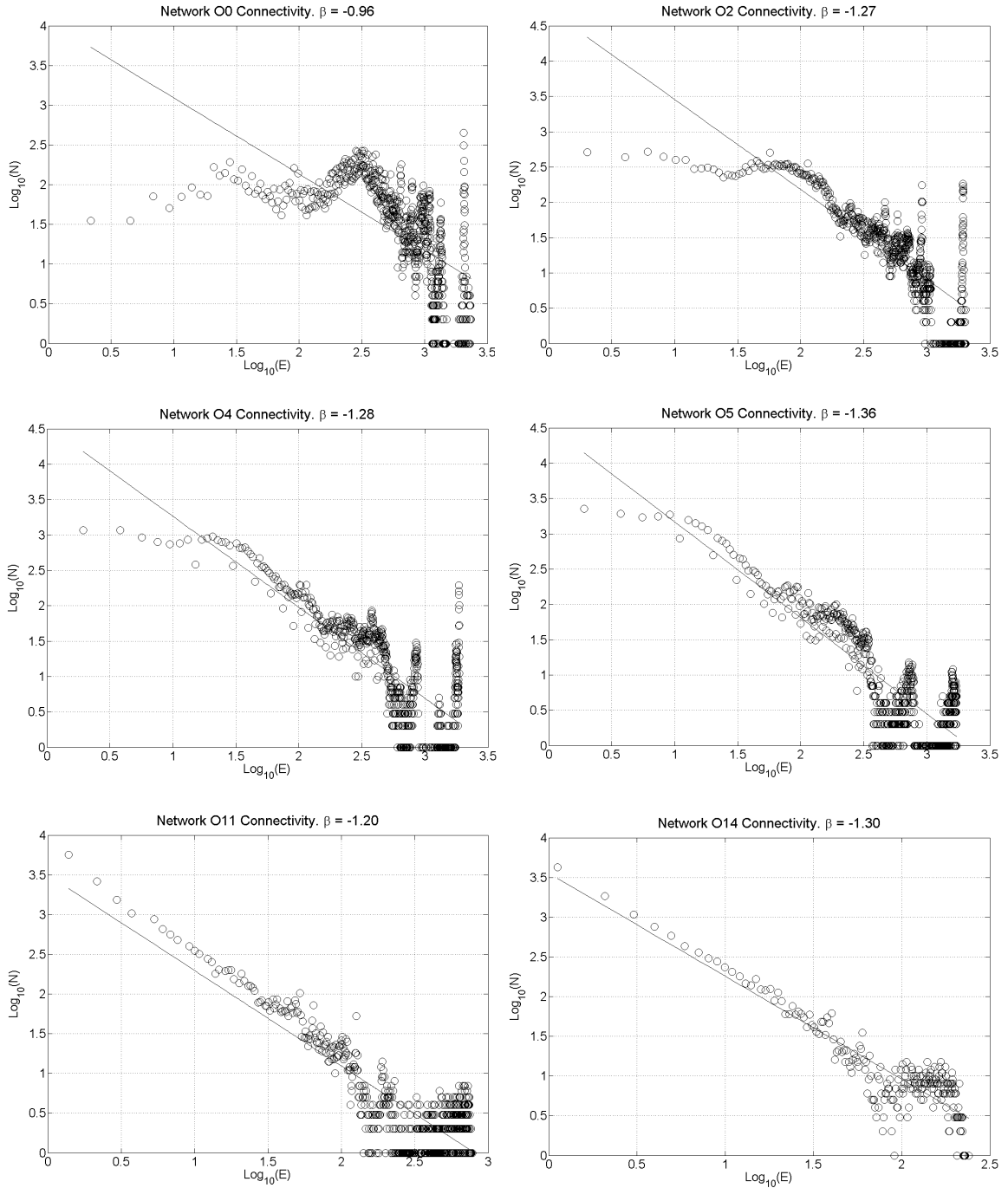
Figure 2. Connectivity distribution for networks O_0, $O2$, $O4$, $O5$, $O11$, and $O14$. N (on the Y axis) is the number of nodes that have E edges (on the X axis).

In Suteanu (2014) network parameters such as node connectivity distribution, node weight distribution (Boccaletti et al., 2006), and the exponents of these distributions, β and respectively γ (Albert and Barabasi, 2002), are assessed for sets of

earthquake networks in different classes. Since the weight distribution can be related to the specific choices of parameter values and the study was focused on finding patterns that do not depend on variations in parameter values, only the behaviour of node connectivity was assessed along the spectrum of W_{min} over different classes. The study shows that networks having W_{min} in the middle to the upper range of the interval between the lowest edge weight value L and the highest edge weight value H in their class exhibit significant scaling properties of the node connectivity distributions, while networks with low values of W_{min} exhibit poor or no such characteristics (Fig. 2). Since the higher values of weight in the strong links are associated with the earthquakes that are most likely to be related with each other, a relationship can be discerned between the major characteristics of seismicity and the structured, power law characteristics in the distributions of node connectivity.

This behavior along the spectrum of W_{min} is robust with respect to variations of parameters that cover the range of values found in numerous studies on temporal and spatial distributions of earthquakes (Omori, 1894; Utsu., 1961; Utsu et al., 1995; Felzer and Brodsky, 2006; Shcherbakov, 2005; Shcherbakov, 2006; Davidsen et al., 2008; Lippiello et al., 2009; Lennartz et al., 2011; Lippiello et al., 2012a;). For example, since p-values usually range between 0.7 and 1.8 (Utsu., 1961), in this study p-values were meant to cover this interval and therefore they were chosen from the slightly larger interval [0.5,2]. Also, since r-values of 1.35-1.37 have been observed for distance distributions (Felzer and Brodsky, 2006), or equal to the earthquake's fractal distribution in the studied area (Bayesi and Paczuski, 2004), in this paper r-values were chosen from the larger interval [0.5,2] (Table 1).

Table 1. All classes. T_{max} is the maximum time interval between events, D_{max} is the maximum distance between events, r is the exponent of the distance weight w_d (Eq. 1), p is the exponent of the time weight w_t (Eq. 2), d_{min} and t_{min} represent cutoff values, H is the highest value of the total edge weight in the class, and L is the lowest value of the total edge weight in the class.

Class	T_{max} (days)	D_{max} (km)	r	p	t_{min} (h)	d_{min} (km)	H	L
B	10	30	-1.35	-1	1	1	1.00	$1.01*10^{-5}$
C	10	30	-1.35	-1	0.5	0.2	0.65	$5.76*10^{-7}$
D	30	30	-1.35	-1	1	1	1.00	$3.39*10^{-6}$
E	40	50	-1.35	-1	0.05	0.2	6.07	$7.24*10^{-9}$
F	7	10	-1.35	-1	0.05	0.1	2.72	$1.44*10^{-7}$
G	7	10	-1.35	-1	0.05	0.025	2.72	$2.22*10^{-8}$
H	8	10	-1.35	-1	0.5	0.2	0.65	$3.19*10^{-6}$
I	8	11	-1.35	-1	0.05	0.1	2.72	$1.11*10^{-7}$
J	8	10	-1.35	-1	1	1	1.00	$5.60*10^{-5}$
L	7	10	-1	-0.5	14	2	1.00	$1.39*10^{-2}$
M	7	10	-1	-0.5	1	1	1.00	$1.86*10^{-3}$
N	40	50	-1	-0.5	0.5	0.2	0.66	$2.19*10^{-5}$
O	50	50	-2	-2	1	1	1.00	$6.75*10^{-11}$
P	30	30	-0.5	-1.5	1	1	1.00	$2.27*10^{-6}$

The identification of correlated earthquakes starts with a certain choice of parameter values (the initial network) from the range of values exemplified previously,

and, after setting increasing thresholds for the minimum weight W_{min}, ends when networks with scale free properties are found. An important aspect of this method is that, although different parameter values may be chosen initially, the same statistical population of interrelated earthquakes is identified in the end (Suteanu, 2014).

Our present analysis shows that the distributions of the spatial and temporal intervals between connected nodes have a similar pattern and scaling properties in all networks, from small to large networks, from networks inside the same class to networks from different classes.

Table 2. Class E Networks. E_0 is the initial network that was generated using the parameter values shown in the first column. See Table 1 for the meaning of T_{max}, D_{max}, r, p, d_{min} , t_{min}, H, and L.

Class E definition	Range of total weight W values	Network Name	W_{min}	Number of nodes	Number of edges
T_{max}= 40 days	H = 6.07	E_0	$7.24*10^{-9}$	37,441	8,488,767
D_{max}= 50 km	L = $7.24*10^{-9}$				
r = - 1.35		E8	$5*10^{-5}$	23,032	584,548
p = - 1					
d_{min} = 0.2 km		E10	$5*10^{-4}$	14,091	117,835
t_{min} = 3 min					

An example is shown for three networks in class E (Table 2): Fig. 3 presents the connectivity distribution in the three networks, from the highly irregular distribution of the large initial network E_0, in which most of the earthquakes are not necessarily

correlated, to the scaling distributions of the networks E8 and E10. The left column in Fig. 4 shows how the distribution of distances between events changes from E_0 to E8 and E10. The three distributions exhibit similarities with each other and with distributions found in other studies (Davidsen et al., 2008; Lippiello et al., 2009): an increase up to a maximum value, followed by a power law decrease. While the maximum value is ~ 1.2 km in E_0, it shifts towards lower values in E8 (~ 0.7 km) and E10 (~ 0.3 km). The peaks at 13 km, 18 km, 23 km… in the distribution of E_0 may be attributed to the spatial relations among events that are clustered around neighbouring volcanoes, distinct vents and fracture zones (see Fig. 1a); these peaks are not present in the distributions of E8 and E10, suggesting that these networks have fewer nodes correlated over long distances.

The right column of Fig. 4 shows the change in time interval distributions from E_0 to E8 and E10. Although the overall shape of the distribution is similar in the three networks and power law properties are present in all three of them, the time interval between events decreases significantly when the scaling properties of the connectivity distribution become stronger in networks E8 and E10. The increase between 7 and 15 days with a peak at 11 days from E_0, which can be attributed to precursory sequences in Hawaii (Chastin and Main, 2003), diminishes in E8, and completely disappears from E10, since E10 retains only the strongest links and the earthquakes that are extremely likely to be interrelated. These differences between the three networks illustrate the fact that in networks that exhibit increasingly strong scaling properties of the connectivity distribution, i.e. networks with nodes that are more and more likely to be interrelated, the earthquakes selected for the networks are increasingly close in time and space.
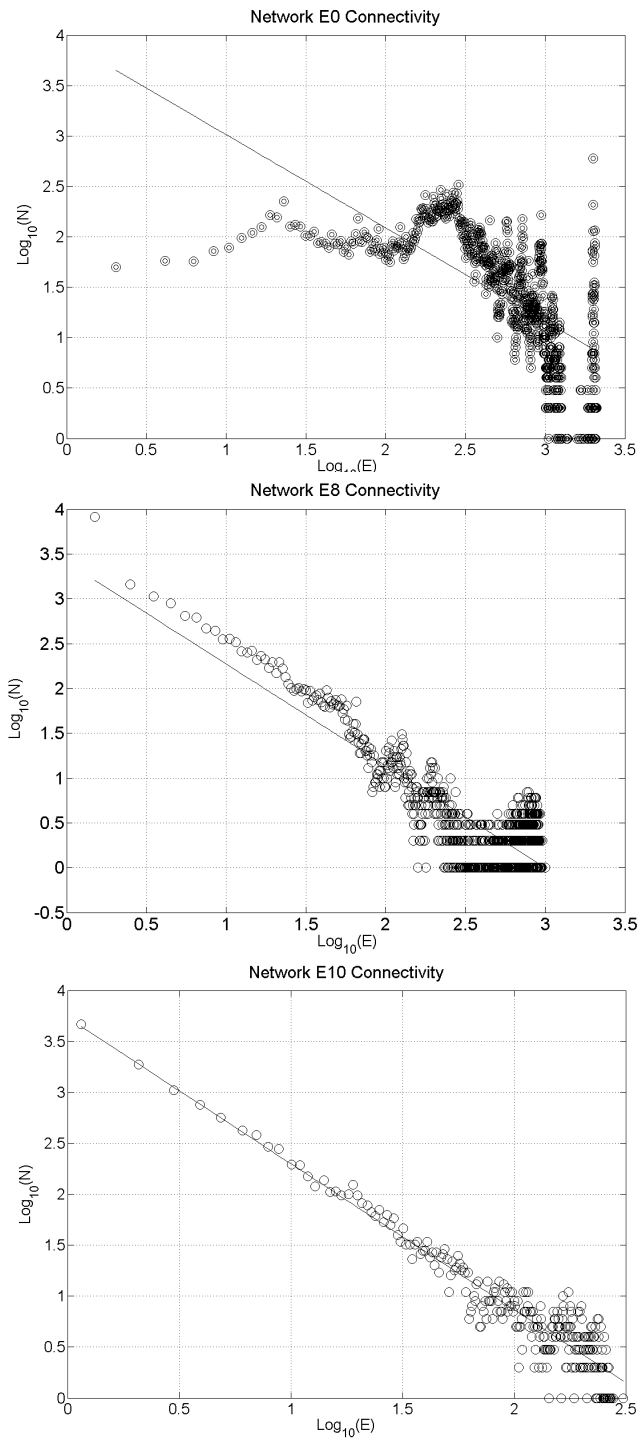
Figure 3. Connectivity distribution for networks E_0, E8, and E10. N (on the Y axis) is the number of nodes that have E edges (on the X axis).
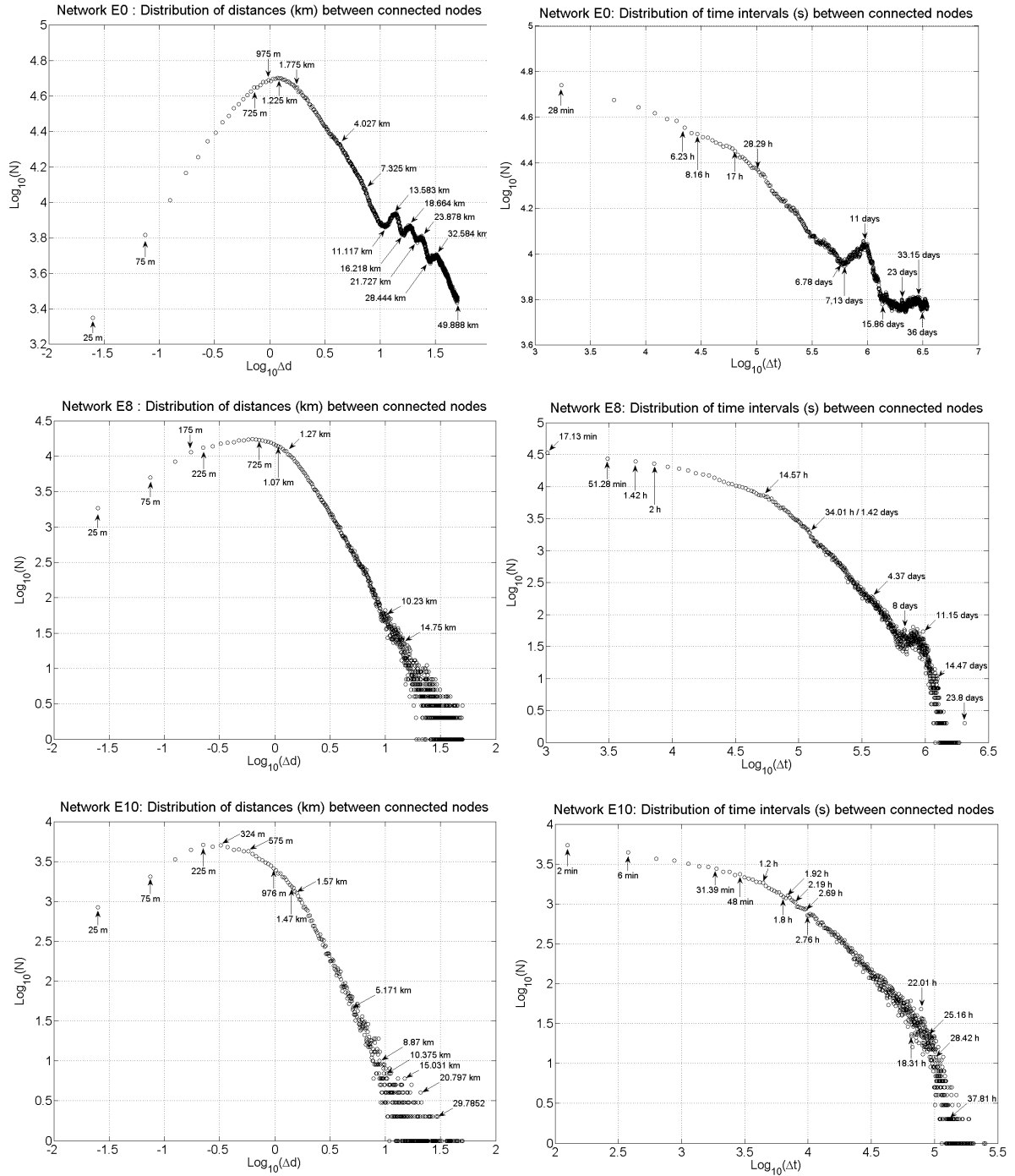
Figure 4. Left column: distribution of distances (km) between any two nodes in networks E_0, E8 and E10; N is the number of distances of Δd km between any two earthquakes in each network. Right column: distribution of time intervals (s) between any two nodes in the networks E_0, E8 and E10; N is the number of time intervals of Δt seconds between any two earthquakes in each network.

3   The clustering coefficient

A relevant measure of the extent to which the nodes in the network tend to group together is the clustering coefficient. If a node i has k_i neighbours, the local clustering coefficient C_i of the node i is defined as the actual number of links L_i between its neighbours divided by the total number of possible links between them (Watts and Strogatz, 1998):

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \qquad\qquad (5)$$

In this study, C_i is 0 for nodes with neighbours that don't have any link between them and for nodes with a node degree equal to 1.

In order to calculate the clustering coefficient, the number of links L_i between the neighbours of each node i were first calculated and assessed. Results show significant similarities between the change in the connectivity distributions towards a scaling structure and the change in the neighbours' links distributions, as illustrated in the examples from class O in Fig. 2 and Fig. 5. A description of these networks is presented in Table 3.

Fig. 5 shows that the distribution of the number of links between nodes' neighbours L_i observed in the initial network O_0 has a similar appearance with the connectivity distribution in O_0 (Fig. 2): an approximately constant interval followed by an approximate power law tail and scattered dot patterns. Moreover, corresponding to the increasingly strong scaling properties of the connectivity distribution that appear in the networks with W_{min} in the upper range of the weight spectrum, a more prominent scaling

structure seems to emerge in the L_i links distributions with the increase of the threshold W_{min}.

Table 3. Class O Networks. O_0 is the initial network that was generated using the parameter values shown in the first column. See Table 1 for the meaning of T_{max}, D_{max}, r, p, d_{min} , t_{min}, H, and L.

Class O definition	Range of total weight W values	Network Name	W_{min}	Number of nodes	Number of edges
T_{max}= 50 days	H = 1.00	O_0	$6.75*10^{-11}$	37,443	9,966,177
D_{max}= 50 km	$L = 6.75*10^{-11}$	O2	10^{-8}	37,354	5,737,672
r = - 2		O4	10^{-7}	37,019	3,826,589
p = - 2		O5	$5*10^{-7}$	36,366	2,816,955
d_{min} = 1 km		O11	$5*10^{-4}$	21,996	444,071
t_{min} = 1 h		O14	10^{-2}	14,158	130,196

The network average clustering coefficient C was calculated as the average of all C_i:

$$C = \frac{1}{N_{nodes}} \sum_{i=1}^{N_{nodes}} C_i \ , \tag{6}$$

where N_{nodes} is the total number of nodes, and C_i is the clustering coefficient of node i (Watts and Strogatz, 1998).
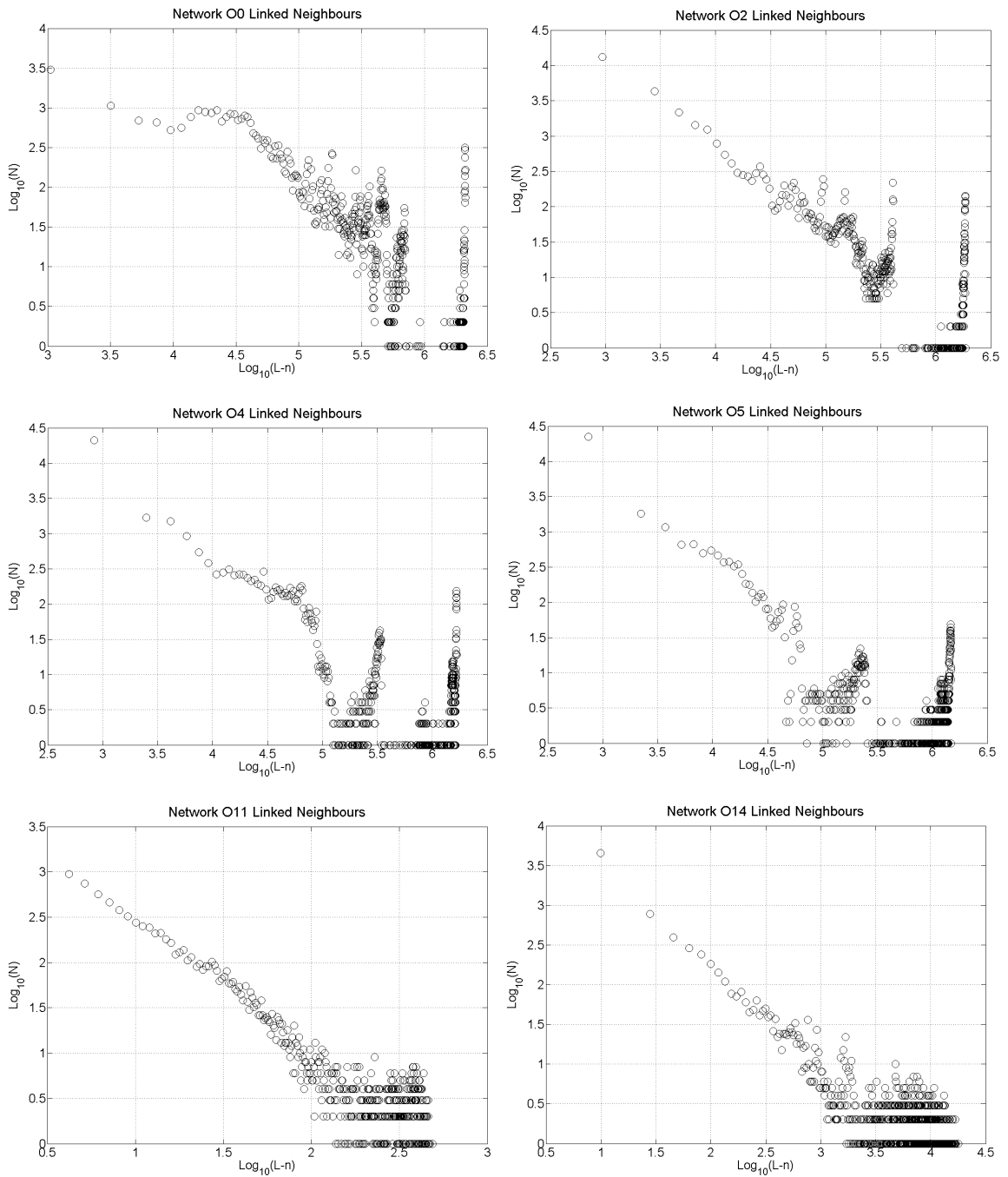
Figure 5. Distribution of actual links between the neighbours of each node for networks O_0, $O2$, $O4$, $O5$, O11, and O14. N (on the Y axis) is the number of nodes that have L-n ("Linked- neighbours") links between their neighbours (on the X axis).

Table 4. Class M Networks. M_0 is the initial network that was generated using the parameter values shown in the first column. See Table 1 for the meaning of T_{max}, D_{max}, r, p, d_{min}, t_{min}, H, and L.

Class M definition	Range of total weight W values	Name	Network W_{min}	Number of nodes	Number of edges
T_{max}= 7 days	H = 1	M_0	$1.86*10^{-3}$	33,065	1,913,280
D_{max}= 10 km	L = $1.86*10^{-3}$	M1	$1.50*10^{-2}$	26,919	1,270,458
r = - 1		M2	$2*10^{-2}$	25,396	1,088,015
p = - 0.5		M3	$3*10^{-2}$	22,682	782,659
d_{min} = 1 km		M4	$4*10^{-2}$	20,440	584,686
t_{min} = 1 h		M5	$5*10^{-2}$	18,825	443,533
		M6	10^{-1}	14,006	149,235

Compared with many other types of networks (Newman, 2003), rather high values of the network clustering coefficient (> 0.4) are found in the earthquake networks studied in this paper. High values of the network clustering coefficient have also been found by Baiesi and Paczuski (2005). Fig. 6a shows an example of network clustering coefficient values for the networks in class M, which are described in Table 4. The corresponding values of the exponents of the connectivity distribution, β, and weight distribution, γ, are presented in Fig. 6b. A comparison between Fig. 6a and 6b shows that the variation of the network average clustering coefficient C and the variation of the exponents β and γ in networks M1 to M6 display a similar overall behaviour.
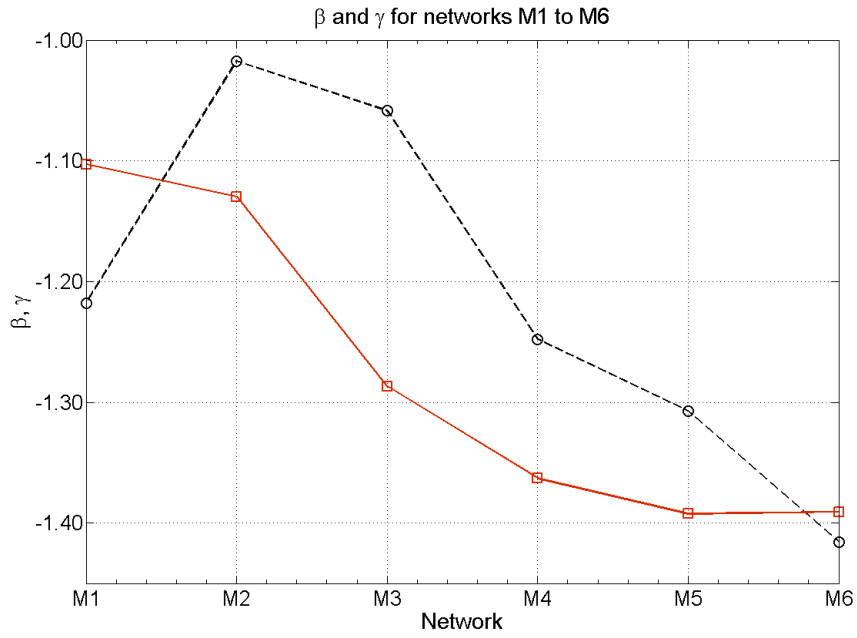
Figure 6 a.
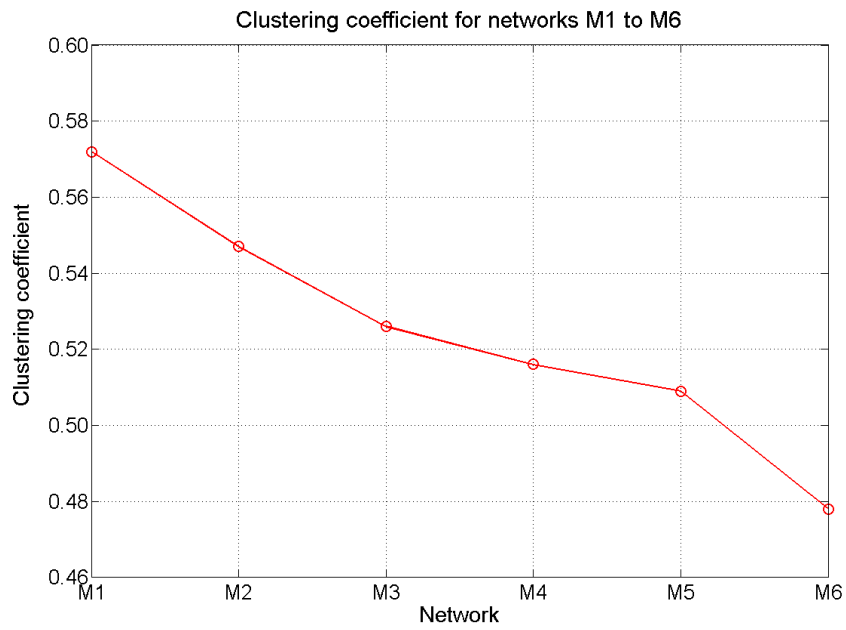
Figure 6 b.

*Figure 6. Comparison between the clustering coefficient (**a**), and the signed values of β, the exponent of the connectivity distribution (black dashed line), and γ, the exponent of the weight distribution (red continuous line) for networks M1 to M6 (**b**).*
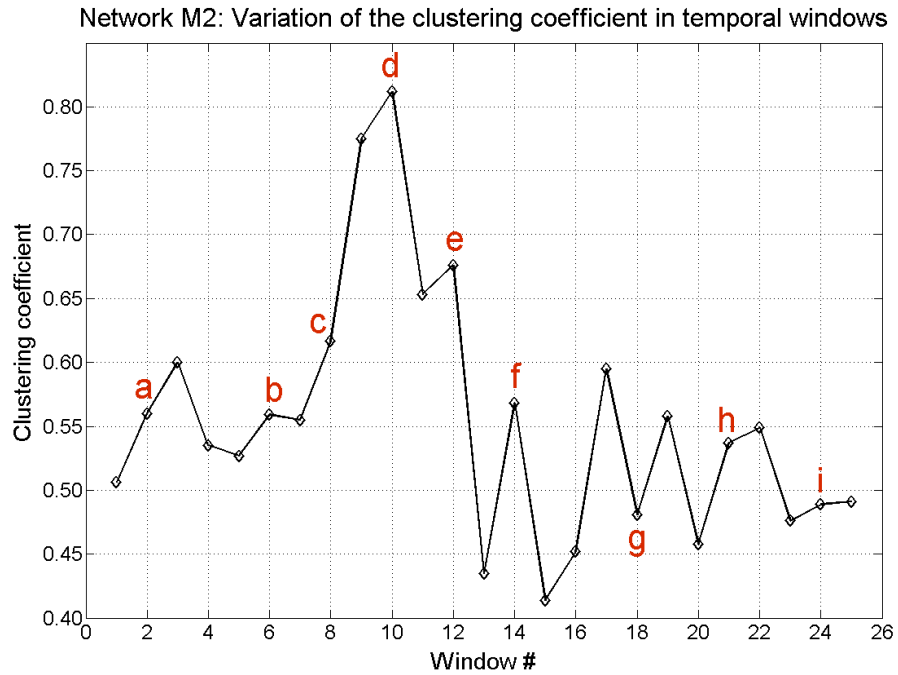
Figure 7 a.
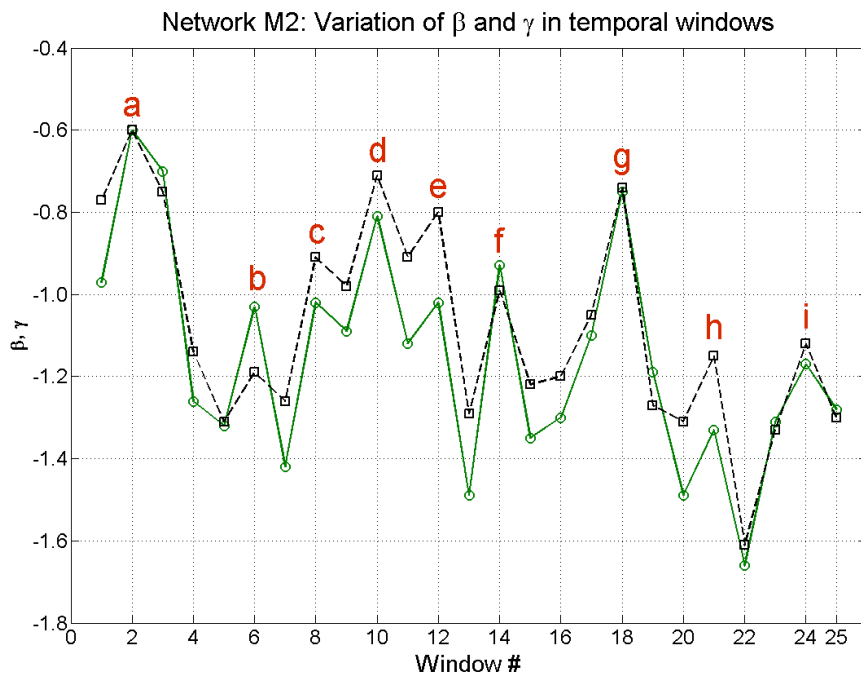
Figure 7 b.

Figure 7. *Successive temporal windows of network M2. (**a**) Variation of the network average clustering coefficient, C. (**b**) Variation of the connectivity distribution exponent, β (green continuous line), and of the weight distribution exponent, γ(black dashed line).*

The variation of β and γ over successive event windows can be used to study the way in which relationships between earthquakes change over time (Suteanu, 2014). The network is divided in successive event windows of various lengths. In this study, equal sized windows of 1,000 events and their corresponding networks were generated. The values of β and γ were estimated using linear regression with uncertainties ranging in most cases between 0.08 and 0.2 at a confidence level of 95%. Fig. 7 presents a comparison between the variation of β and γ in temporal windows of network M2 (Fig. 7b), and the variation of the network clustering coefficient C in the same temporal windows (Fig. 7a). Since in this paper we consider the signed values of β and γ, and not their absolute values, the maxima in Fig. 7b correspond to the minima discussed by Suteanu (2014). In a similar way, the lower case letters from "a" to "i" are used to tag the maximum values of the two exponents β and γ (Fig. 7b), the values of the network clustering coefficient C in the corresponding windows (Fig. 7a), and the corresponding areas on the cumulative number of earthquakes graph in Fig. 8.

Each of the maxima of the exponents β and γ in Fig. 7b was associated in Suteanu (2014) with enhanced volcanic activity, major discharges of energy, changes in earthquake patterns, etc. Fig. 7 shows that the network clustering coefficient in successive temporal windows has an overall variation that closely resembles the variation of β and γ. Although one important discrepancy can be observed for window 18 (g), which may be related to the abrupt cut-off in window definition, the similarity in the variation of the three network parameters is remarkable. The synchronous maxima of the three parameters show that, in the corresponding windows, there is an increase in the number of nodes that are highly connected and grouped together.
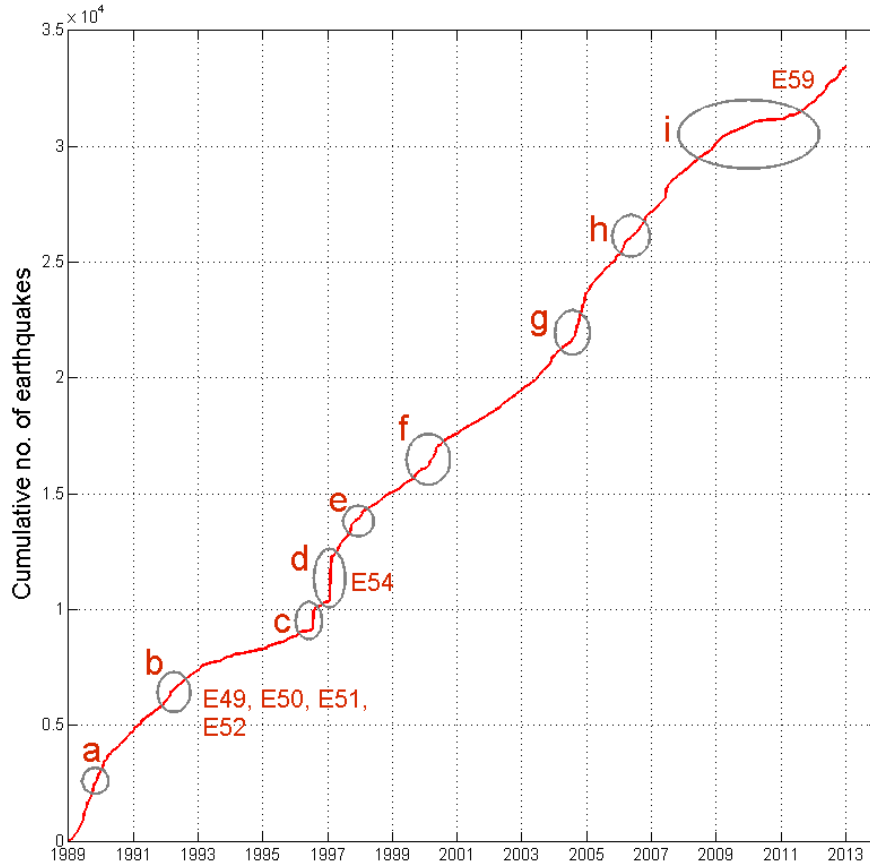
Figure 8. Cumulative number of earthquakes from January 1989 to December 2012. The small letters tag the areas corresponding to the maximum values of the exponents β and γ and the network clustering coefficient C in successive temporal windows of network M2.

We believe that this is another confirmation that the analysis we perform in successive temporal windows is able to reflect the change in the relationships between earthquakes over time.

4 Magnitude of successive nodes

All the sequences of values (m_i, m_j) – where i and j are nodes of magnitude m_i and m_j, respectively, and node i has a directed edge to node j – are ranked and assessed. This is accomplished by taking magnitude values of every two linked nodes i and j to create

ordered pairs (m_i, m_j). The number of occurrences of every instance (m_i, m_j) is calculated and the set of all occurrences is arranged in a decreasing order. The distribution of the resulting ordered vector is ultimately assessed. Moreover, if the node j has a directed edge to the node k of magnitude m_k, then the sequences of values (m_i, m_j, m_k) are similarly ranked and assessed. Results show that a Zipf distribution is found in both cases. Examples are presented in Fig. 9: Fig 9a shows the example of a network with strong scaling properties in the connectivity distribution, network D5, while Fig. 9b shows the example of a large initial network, network E_0, which has an irregular distribution of node connectivity.

This property can probably also be found for sequences of four, five, or more successive magnitude values. Since it is unlikely that the majority of nodes in E_0 represent earthquakes that are correlated with each other (Suteanu, 2014), we believe that these results are not characteristics of interconnected earthquakes, but rather root in the Gutenberg-Richter magnitude frequency distribution (Gutenberg and Richter, 1954).

5   Conclusions

The study of earthquake networks is able to reveal aspects of structure that can be reasonably associated with fundamental properties of seismicity. The scale free behaviour of the connectivity distribution along the spectrum of W_{min}, which can be used to discern the interrelated earthquakes from the rest of the data set, is mirrored by a similar behaviour of the distribution of the number of nodes' linked neighbours. This behaviour is robust with respect to variations in parameters within the range of values found in temporal and spatial distributions of earthquakes.
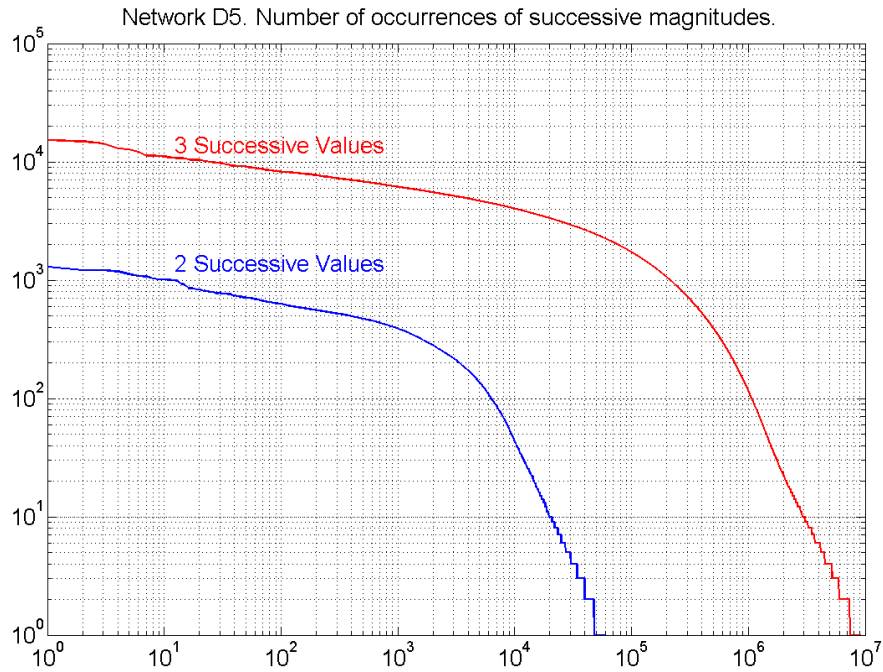
Figure 9a.
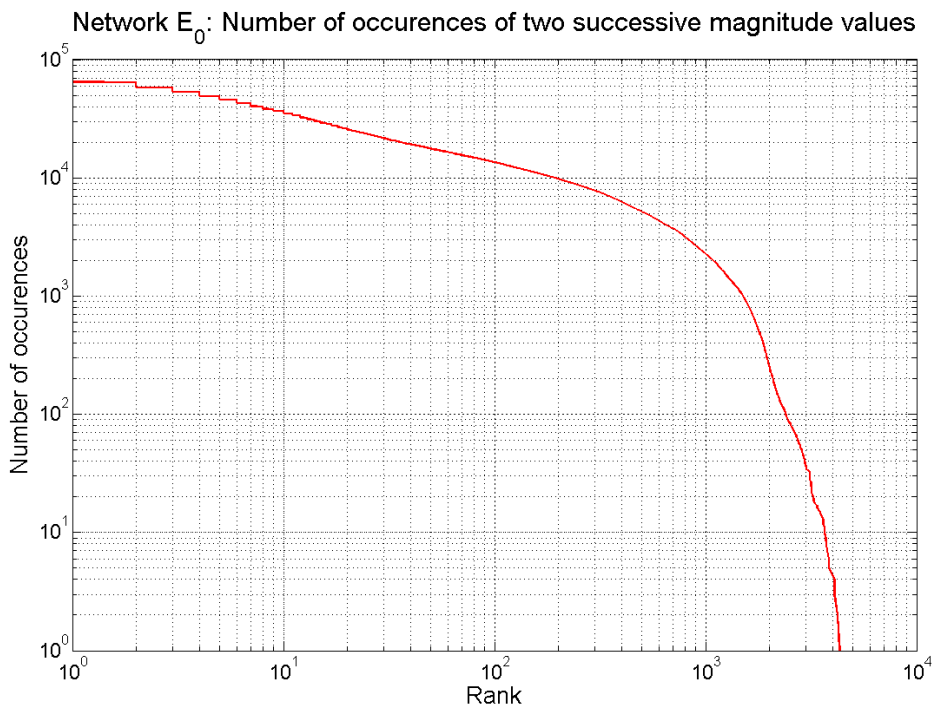
Figure 9b.

Figure 9. Figure 9. Magnitude values of successive nodes. (a) Magnitude of two and three successive nodes in a network that enjoys scaling properties of the connectivity distribution (network D5). (b) Pairs of magnitude values of successive nodes in a large initial network (network E_0).

A similar pattern of increase up to a maximum followed by a power law decrease was found in the distribution of distances between earthquakes in a variety of networks: large networks, small networks, networks in the same class, or networks that belong to different classes. Similarly, an overall pattern of slow decrease followed by a power law decrease was found in the distribution of time intervals between events in various networks: networks inside each class, networks across the classes, as well as in the selective networks of interrelated events. In this latter case, the above distributions reflect the increased proximity in space and time of the correlated events.

In comparison to most of the biological, technological, or social networks, high values have been found for the network average clustering coefficient, which ranges generally between 0.4 and 0.6. A notable similarity was found between the variation of the network clustering coefficient, C, along the series of networks in a class, and the variation of the exponents of the connectivity distribution, β, and of the weight distribution, γ, especially in networks that exhibit significant scaling properties of these distributions. Moreover, the synchronous variation of β, γ, and C in successive temporal windows can be related to changes in the characteristics of seismicity and in the life of the volcanic system.

The intrinsic scale free structure of seismicity is also revealed in the Zipf distribution found for the ranked number of occurrences of groups of magnitude values of successive network nodes.

References

Albert, R., and Barabási, A.-L. (2002), Statistical mechanics of complex networks, Rev. Mod. Phys., 74, 47-97.

Baiesi, M., and Paczuski, M. (2004), Scale-free networks for earthquakes and aftershocks, Phys. Rev. E, 69, 066106-1–8.

Baiesi, M., and Paczuski, M. (2005), Complex networks of earthquakes and aftershocks, Nonlinear Processes in Geophysics, 12, 1-11.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006), Complex networks: structure and dynamics, Phys. Reports, 424, 175-308.

Bunde, A., and Lennartz, S. (2012), Long-term correlations in earth sciences, Acta Geophys., 60, 3, 562-588.

Carbone, V., Sorriso-Valvo, L., Harabaglia, P., Guerra, I. (2005), Unified scaling law for waiting times between seismic events, Europhys. Lett., 71, 6, 1036–1042.

Chastin, S.F.M., and Main, I.G. (2003), Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions, Geophys. Res. Lett., 30, 13, 1671. doi,10.1029/2003GL016900.

Davidsen, J., Grassberger, P., and Paczuski, M. (2008), Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity, Phys. Rev. E, 77, 066104.

Felzer, K. R., and Brodsky, E. E. (2006), Decay of aftershock density with distance indicates triggering by dynamic stress, Nature, 441, 735-738.

Gutenberg, B., and Richter, C. F., Seismicity of the Earth (Princeton University Press, Princeton 1954).

Kagan, Y.Y. (1994), Observational evidence for earthquakes as a nonlinear dynamic process, Elsevier Physica D, Nonlinear Phenomena, 77, 1-3, 160-192.

Lapenna, V., Macchiato, M., Piscitelli, S., and Telesca L. (2000), Scale invariance properties in seismicity of Southern Apennine Chain (Italy), Pure Appl. Geophys., 157, 4, 589–602.

Lennartz, S., Livina, V. N., Bunde, A., and Havlin S. (2008), Long-term memory in earthquakes and the distribution of interoccurrence times, Europhys. Lett., 89, 69001, doi: 10.1209/0295-5075/81/69001.

Lennartz, S., Bunde, A., and Turcotte, D.L. (2011), Modelling seismic catalogues by cascade models: Do we need long-term magnitude correlations?, Geophysical Journal International, 184, 1214-1222.

Lippiello, E., Corral, A., Bottiglieri, M., Godano, C., and de Arcangelis, L. (2012a), Scaling behavior of the earthquake intertime distribution: Influence of large shocks and time scales in the Omori law, Phys. Rev. E, 86, 6-2, 066119, doi:10.1103/PhysRevE.86.066119.

Lippiello, E., de Arcangelis, L., and Godano, C. (2009), The role of static stress diffusion in the spatio-temporal organization of aftershocks, Phys. Rev. Lett., 103, 038501, doi:10.1103/PhysRevLett.103.038501.

Lippiello, E., Godano, C., and de Arcangelis, L. (2012b), The earthquake magnitude is influenced by previous seismicity, Geophysical Research Letters, 39, 5, doi: 10.1029/2012GL051083.

Nanjo, K., and Nagahama, H. (2000), Spatial distribution of aftershocks and the fractal structure of active fault systems, Pure Appl. Geophys., 157, 4, 575–588.

Newman, M.E.J. (2003), The Structure and Function of Complex Networks, SIAM REVIEW, 45, 2, 167-256.

Omori, F. (1894), On the aftershocks of earthquakes, J. College of Science, Imperial University of Tokyo, 7, 111–200.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B. (2004), A generalized Omori's law for earthquake aftershock decay, Geophys. Res. Lett., 31, L11613, doi 10.1029/2004GL019808.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B. (2005), Aftershock statistics, Pure and Applied Geophysics, 162, 1051-1076.

Shcherbakov, R., Turcotte, D.L., and Rundle, J. B. (2006), Scaling properties of the Parkfield aftershock sequence, B. Seismol. Soc. Am., 96, 4B, 376–S384, doi: 10.1785/0120050815.

Suteanu, M. (2014), Scale free properties in a network-based integrated approach to earthquake pattern analysis, Nonlinear Processes in Geophysics, 21, 427-438.

Utsu, T. (1961). A statistical study of the occurrence of aftershocks, Geophysical Magazine, 30, 521–605.

Utsu, T., Ogata, Y., and Matsu'ura, R.S. (1995), The Centenary of the Omori Formula for a Decay Law of Aftershock Activity, J. Phys. Earth, 43, 1-33.

Varotsos, P.A., Sarlis, N.V., and Skordas, E.S. (2012), Order parameter fluctuations in natural time and b-value variation before large earthquakes, Natural Hazards and Earth System Sciences, doi:10.5194/nhess-12-3473-2012, 3473-3481.

Watts, D.J., and Strogatz, S. (1998), Collective dynamics of 'small-world' networks, Nature, 393, 440–442.

Chapter 6

Conclusions

In this research, a new type of directed networks is proposed for the assessment of relationships between earthquakes. The method was applied to volcanic seismicity in Hawaii. The nodes of the networks are epicenters of earthquakes; the edges that link the nodes carry space-time-magnitude weights, and have a direction given by the temporal succession of the events. The generality of the definition of the edge weight, W, as a combination of a factor in time, a factor in space, and a factor in magnitude is comprehensive and permits various combinations of space-time-magnitude correlations between earthquakes. Since any node can have any number of edges that enter the node and any number of edges that leave the node, any given event may have multiple predecessors, and any given event can contribute to multiple future events, as long as its edges carry enough weight.

Parameters and formulas used in the calculation of the weights take into consideration well-established properties of seismicity. High values of W are associated with strong relationships between earthquakes, while low values of W are associated with either weak relationships or situations for which no relationships can be established. Various classes of networks can be generated based on distinct values of the parameters. Inside each class, different networks can be created by setting different thresholds for the minimum edge weight W_{min}.

It is shown that networks that have W_{min} in the middle to upper range of the interval between the lowest edge weight value L and the highest edge weight value H in

their class manifest significant scaling properties of node connectivity distributions, as opposed to networks with low values of W_{min}, which exhibit poor or no scaling characteristics. Since high values of weight describe the strong links, the events selected in the networks with high values of W_{min} are primarily the earthquakes that are most likely to be related to each other. Therefore, a relationship can be discerned between the fundamental characteristics of seismicity and the well-organized, scale free distributions of node connectivity. In networks with low values of W_{min}, most of the nodes have little or no relationship with each other. In this context, the irregular and scattered shapes of their connectivity distributions are not a surprise.

It is also shown that the scale free behaviour in the connectivity distributions is robust with respect to variations of parameters within the range of values found by numerous studies in temporal and spatial distributions of earthquakes. Tests performed on networks that manifest strong power law properties, but originating in different choices of parameter values, confirm the reliability of the method. They show that the same statistical population of earthquakes is chosen to participate in these networks, i.e. the earthquakes most likely to be interrelated. The results indicate that the method is reliable, robust with respect to variations of parameter values, and reflects fundamental properties of seismicity.

The threshold values W_{min} that identify networks of interrelated events are assessed in the context of all the earthquakes in the class: they are found as those values for which scale free properties appear and become stronger when subsequent networks are created using increasing values of W_{min}. It can be said that the threshold W_{min} is a

global parameter that characterizes the set of earthquakes in a class, and its values are meaningful only inside that set.

There are also other significant scaling properties that are detected in the analysis of the classes of networks. Node weight distributions also enjoy scaling properties. Also, for each class, the dependency of the number of edges on the number of nodes is a power law.

A similar pattern of increase up to a maximum followed by a power law decrease was found in the distribution of distances between earthquakes in a variety of networks: large networks, small networks, networks in the same class, or networks that belong to different classes. Also, an overall pattern of slow decrease followed by a power law decrease was found in the distribution of time intervals between events in different networks: networks belonging to the same class, networks across classes, as well as in the selective networks of interrelated events. In this latter case, the distributions of spatial and temporal intervals between nodes reflect the increased proximity in space and time of the correlated events.

In comparison to most of the biological, technological, or social networks, high values have been found for the network average clustering coefficient, which ranges generally between 0.4 and 0.6. A notable similarity was found between the variation of the network clustering coefficient, C, along the series of networks in a class, and the variation of the exponents of the connectivity distribution, β, and of the weight distribution, γ, especially in networks that exhibit significant scaling properties of these distributions.

The evolution of the relationships between earthquakes over time can be studied by splitting up the network in successive event windows of various lengths. The distributions of node connectivity and node weight in the emerging sub-networks manifest scaling properties that can be used to follow the evolution of seismicity over time. The exponents β and γ of these distributions have a similar evolution over the temporal windows. The increased connectivity in the maxima of β and γ can be associated with sudden, important discharges of energy in the life of the volcanic system; the exponents of connectivity and weight distributions for successive event windows can be used to study the way the relationships between earthquakes are changing over time.

Moreover, the network clustering coefficient, C, has an overall variation over the same temporal windows that closely resembles the variation of β and γ. Although a few differences can be observed, which may be related to the abrupt cut-off in window definition, the similarity in the variation of the three network parameters is remarkable. The presence of synchronous maxima of the three parameters confirms that the analysis performed in successive temporal windows is able to reflect the change in the relationships between earthquakes over time. The maxima show that, in the corresponding windows, there is an increase in the number of nodes that are highly connected and grouped together.

The intrinsic scale free structure of seismicity is also revealed in the Zipf distribution found for the ranked magnitude values of successive network nodes.

Aspects regarding the energy dissipation and the optimization of the selection of the temporal windows are the subject of further research.

Appendix A

```
CREATE PROCEDURE dbo.usp_Create_Network
       @nTMAx int, @nDMAx int,@nTmin dec(3,2), @nDmin dec(3,2),
       @nc dec (19,18), @ns dec (19,18),
       @nr dec(3,1), @np dec(3,1)
AS
DECLARE @nMAx int, @i int, @nPid int,
       @nFrom_depth dec(4,2),@nFrom_magn dec(3,2),
       @nFrom_year  smallint,@nFrom_month  tinyint,@nFrom_day  tinyint,
       @nFrom_hour tinyint,@nFrom_minut tinyint,@nFrom_sec dec(4,2),
       @cFrom_lat char(8),@cFrom_long char(9),@nFrom_x dec(9,3),
       @nFrom_y dec(9,3),@dFrom_moment datetime, @nMagnMAx dec(3,2)
SET @i = 1
SET @nMAx = (SELECT COUNT(*) FROM dbo.data_all)
SET @nMagnMAx = (SELECT MAX(CAST(magn AS dec(3,2)))
                   FROM dbo.data_all)
WHILE @i < @nMAx
       BEGIN
              SET @nPid = (SELECT CAST(pid AS int) FROM dbo.data_all
                                    WHERE cid = @i)
              SET @nFrom_depth = (SELECT depth FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_magn = (SELECT magn FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_year = (SELECT year FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_month = (SELECT month FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_day = (SELECT day FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_hour = (SELECT hour FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_minut = (SELECT minut FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_sec = (SELECT sec FROM dbo.data_all
                                               WHERE cid = @i)
              SET @cFrom_lat = (SELECT lat FROM dbo.data_all
                                               WHERE cid = @i)
              SET @cFrom_long = (SELECT long FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_x = (SELECT x FROM dbo.data_all
                                               WHERE cid = @i)
              SET @nFrom_y = (SELECT y FROM dbo.data_all
                                               WHERE cid = @i)
              SET @dFrom_moment = (SELECT moment FROM dbo.data_all
                                               WHERE cid = @i)
              INSERT  dbo.Network_Edges
                   (pid,     pid_linked,delta_t,from_depth,    to_depth,
                   from_magn,  to_magn,  from_year,from_month,  from_day,
                   from_hour,  from_minut,from_sec,  to_year,   to_month,
                   to_day,    to_hour,to_minut,    to_sec,     from_lat,
                   from_long,  to_lat,  to_long,  from_x,  from_y,  to_x,
                   to_y, d)
```

```sql
                SELECT    @nPid,pid,    DATEDIFF(SECOND,@dFrom_moment,
                CAST(year+'-'+month+'-'+day+'    '+hour+':'+minut+':'
                +sec                AS                datetime)),
                @nFrom_depth,depth,@nFrom_magn,magn,@nFrom_year,
                @nFrom_month,          @nFrom_day,          @nFrom_hour,
                @nFrom_minut,@nFrom_sec,year,  month,day,  hour,minut,
                sec,@cFrom_lat,@cFrom_long,lat,long,@nFrom_x,@nFrom_y
                ,x,y,   SQRT((x-@nFrom_x)*(x-@nFrom_x)+(y-@nFrom_y)*(y-
                @nFrom_y))
                FROM dbo.data_all
                WHERE CAST(pid AS int) > @nPid
                        AND
                        DATEDIFF(SECOND,@dFrom_moment,CAST(year+'-'+
                        month  +'-'+day+'   '+hour+':'+minut+':'+sec  AS
                        datetime))<= @nTMAx * 3600 *24
                        AND
                        SQRT((x-@nFrom_x)*(x-@nFrom_x)
                        + (y-@nFrom_y)*(y-@nFrom_y)) <= @nDMAx
                ORDER BY CAST(pid AS int)
                SET @i = @i +1
        END
UPDATE dbo.Network_Edges
        SET w_m = from_magn*1.000000000000000000/@nMagnMAx,
        w_d = CASE
            WHEN d <= @nDmin THEN 1
            WHEN d > @nDmin THEN @nc * POWER(1.000000000000000000*d,(-
1)*@nr)
            END,
        w_t = CASE
            WHEN delta_t/3600 <= @nTmin THEN 1
            WHEN delta_t/3600 > @nTmin
                THEN   @ns  *  POWER(1.000000000000000000*delta_t/3600,(-
            1)*@np)
            END
UPDATE dbo.Network_Edges
        SET weight = w_m * w_d * w_t
GO
```

Appendix B

CREATE PROCEDURE [dbo].[usp_Compute_Network_Nodes]

```sql
AS
TRUNCATE TABLE dbo.Network_Nodes
SELECT DISTINCT pid INTO dbo.#tmp_pid
       FROM dbo.network_edges
INSERT dbo.#tmp_pid
       SELECT DISTINCT pid_Linked
       FROM dbo.network_edges
INSERT dbo.Network_Nodes
       (pid)
       SELECT DISTINCT CAST(pid AS int) AS pid
       FROM dbo.#tmp_pid
       ORDER BY CAST(pid AS int)
--------- OUT edges:
SELECT CAST(pid AS int) AS pid,COUNT(*) AS out_lines,SUM(weight) AS
out_weight INTO #Out
       FROM dbo.network_edges
       GROUP BY pid
       ORDER BY CAST(pid AS int)
UPDATE dbo.Network_Nodes
       SET out_lines = tt.out_lines,
             out_weight = tt.out_weight
       FROM dbo.Network_Nodes nn
       INNER JOIN dbo.#Out tt
       ON nn.pid = tt.pid
UPDATE dbo.Network_Nodes
       SET out_lines = -1,
             out_weight = -1
       WHERE out_lines IS NULL
--------- IN edges:
SELECT CAST(pid_linked AS int) AS pid,COUNT(*) AS in_lines,SUM(weight)
AS in_weight INTO #IN
       FROM dbo.network_edges
       GROUP BY pid_linked
       ORDER BY CAST(pid_linked AS int)
UPDATE dbo.Network_Nodes
       SET in_lines = tt.in_lines,
             in_weight = tt.in_weight
       FROM dbo.Network_Nodes nn
       INNER JOIN dbo.#IN tt
       ON nn.pid = tt.pid
UPDATE dbo.Network_Nodes
       SET in_lines = -1,
             in_weight = -1
       WHERE in_lines IS NULL
--------- Total edge:
UPDATE dbo.Network_Nodes
       SET tot_lines = CASE
                           WHEN in_lines>0 AND out_lines>0
                                 THEN in_lines+out_lines
                           WHEN in_lines>0 AND out_lines<0 THEN in_lines
                           WHEN in_lines<0 AND out_lines>0 THEN out_lines
```

```sql
                        END,
        tot_weight= CASE
                WHEN in_weight>0 AND out_weight>0
                        THEN in_weight+out_weight
                WHEN in_weight>0 AND out_weight<0 THEN in_weight
                WHEN in_weight<0 AND out_weight>0 THEN out_weight
                    END
--------- Linked neighbour:
CREATE TABLE dbo.#tmp (pid char(7) NOT NULL)
CREATE TABLE dbo.#tmp_1 (pid char(7) NOT NULL)
DECLARE @nNodes int, @i int, @cPid char(7), @nLinks int
SET @nNodes = (SELECT COUNT(*) FROM dbo.Network_Nodes)
SET @i = 1
WHILE @i <= @nNodes
    BEGIN
        IF (SELECT tot_lines FROM dbo.Network_Nodes WHERE cid = @i)
> 1
            BEGIN
                SET @cPid = (SELECT pid FROM dbo.Network_Nodes
                                    WHERE cid = @i)
                INSERT dbo.#tmp
                    SELECT pid FROM dbo.network_edges
                    WHERE pid_linked = @cPid     -- 98
                    ORDER BY CAST(pid AS INT)
                INSERT dbo.#tmp
                    SELECT pid_linked FROM dbo.network_edges
                    WHERE pid = @cPid -- 98
                    ORDER BY CAST(pid AS INT)
                INSERT dbo.#tmp_1
                    SELECT * FROM dbo.#tmp
                SET @nLinks=(SELECT COUNT(*)
                                    FROM network_edges nn
                                    INNER JOIN dbo.#tmp tt
                                    ON nn.pid=tt.pid
                                    INNER  JOIN  dbo.#tmp_1
t1
                                    ON
nn.pid_linked=t1.pid)
                UPDATE dbo.Network_Nodes
                    SET linked_near = @nLinks
                    WHERE cid = @i
                TRUNCATE TABLE dbo.#tmp
                TRUNCATE TABLE dbo.#tmp_1
            END
        SET @i = @i + 1
    END
UPDATE dbo.Network_Nodes
    SET clustering= CASE
                WHEN tot_lines >1
                  THEN (CAST (linked_near AS  numeric(13,3)) /
                  (CAST((tot_lines * (tot_lines - 1))/2 AS
                    numeric(13,3))))
                WHEN tot_lines = 1 THEN 0
                END
GO
```

Appendix C

CREATE PROCEDURE [dbo].[usp_Compute_Network_Nodes_in_tempdb]

```sql
AS
TRUNCATE TABLE tempdb.dbo.Network_Nodes
SELECT DISTINCT pid INTO dbo.#tmp_pid
      FROM tempdb.dbo.network_edges
INSERT dbo.#tmp_pid
      SELECT DISTINCT pid_Linked
      FROM tempdb.dbo.network_edges
INSERT tempdb.dbo.Network_Nodes
      (pid)
      SELECT DISTINCT CAST(pid AS int) AS pid
      FROM dbo.#tmp_pid
      ORDER BY CAST(pid AS int)
--------- OUT edges:
SELECT  CAST(pid AS int)  AS  pid,COUNT(*)  AS  out_lines,SUM(weight)  AS
out_weight  INTO #Out
      FROM tempdb.dbo.network_edges
      GROUP BY pid
      ORDER BY CAST(pid AS int)
UPDATE tempdb.dbo.Network_Nodes
      SET out_lines = tt.out_lines,
            out_weight = tt.out_weight
      FROM tempdb.dbo.Network_Nodes nn
      INNER JOIN dbo.#Out tt
      ON nn.pid = tt.pid
UPDATE tempdb.dbo.Network_Nodes
      SET out_lines = -1,
            out_weight = -1
      WHERE out_lines IS NULL
--------- IN edges:
SELECT  CAST(pid_linked AS int)  AS pid,COUNT(*)  AS in_lines,SUM(weight)
AS in_weight  INTO #IN
      FROM tempdb.dbo.network_edges
      GROUP BY pid_linked
      ORDER BY CAST(pid_linked AS int)
UPDATE tempdb.dbo.Network_Nodes
      SET in_lines = tt.in_lines,
            in_weight = tt.in_weight
      FROM tempdb.dbo.Network_Nodes nn
      INNER JOIN dbo.#IN tt
      ON nn.pid = tt.pid
UPDATE tempdb.dbo.Network_Nodes
      SET in_lines = -1, in_weight = -1
      WHERE in_lines IS NULL
--------- Total edge:
UPDATE tempdb.dbo.Network_Nodes
      SET tot_lines = CASE
                        WHEN in_lines>0 AND out_lines>0
                          THEN in_lines+out_lines
                        WHEN in_lines>0 AND out_lines<0 THEN in_lines
                        WHEN in_lines<0 AND out_lines>0 THEN out_lines
                      END,
```

```sql
                tot_weight= CASE
                              WHEN in_weight>0 AND out_weight>0
                                THEN in_weight+out_weight
                              WHEN in_weight>0 AND out_weight<0
                                THEN in_weight
                              WHEN in_weight<0 AND out_weight>0
                                THEN out_weight
                            END
---------- Linked neighbour:
CREATE TABLE dbo.#tmp (pid char(7) NOT NULL)
CREATE TABLE dbo.#tmp_1 (pid char(7) NOT NULL)
DECLARE @nNodes int, @i int, @cPid char(7), @nLinks int
SET @nNodes = (SELECT COUNT(*) FROM tempdb.dbo.Network_Nodes)
SET @i = 1
WHILE @i <= @nNodes
      BEGIN
            IF (SELECT tot_lines
                  FROM tempdb.dbo.Network_Nodes WHERE cid = @i) > 1
                  BEGIN
                        SET @cPid = (SELECT pid
                                          FROM tempdb.dbo.Network_Nodes
                                          WHERE cid = @i)
                        INSERT dbo.#tmp
                              SELECT pid FROM tempdb.dbo.network_edges
                              WHERE pid_linked = @cPid
                              ORDER BY CAST(pid AS INT)
                        INSERT dbo.#tmp
                              SELECT pid_linked
                              FROM tempdb.dbo.network_edges
                              WHERE pid = @cPid -- 98
                              ORDER BY CAST(pid AS INT)
                        INSERT dbo.#tmp_1
                              SELECT * FROM dbo.#tmp
                        SET @nLinks = (SELECT COUNT(*)
                                          FROM  tempdb.dbo.network_edges
                                nn
                                          INNER JOIN dbo.#tmp tt
                                          ON nn.pid=tt.pid
                                          INNER JOIN dbo.#tmp_1 t1
                                          ON nn.pid_linked=t1.pid)
                        UPDATE tempdb.dbo.Network_Nodes
                              SET linked_near = @nLinks
                              WHERE cid = @i
                        TRUNCATE TABLE dbo.#tmp
                        TRUNCATE TABLE dbo.#tmp_1
                  END
            SET @i = @i + 1
      END
UPDATE tempdb.dbo.Network_Nodes
      SET clustering= CASE
                        WHEN tot_lines >1
                          THEN (CAST (linked_near AS
                                numeric(13,3))/(CAST((tot_lines *
                                (tot_lines - 1))/2 AS numeric(13,3))))
                        WHEN tot_lines = 1 THEN 0
                      END
GO
```

Appendix D

CREATE PROCEDURE [dbo].[usp_Create_Windows]

```sql
                @cNetworkName varchar(10), @nSize int, @nOverlap int
AS
/*
Input:
------
       - Tables:
            - Network_Edges
            - Result
       - Parameters:
            - @nSize = Window size
            - @nOverlap = Overlap with previous window

Network_Nodes is populated with all nodes (distinct) in Network_Edges.

Output:
-------
       - Creates windows based on nSize & nOverlap
       - Generates tables:
            - Window_(No)_Edges
            - Window_(No)_Nodes (with node characteristics)
       - Fills up table Result with summary of computation
*/
DECLARE @nEdges int,@nNodes int, @i int, @nMinPid int, @nMaxPid int,
@nMinCid int, @nMaxCid int,
       @nWindow_Edges              int,@nWindow_Nodes              int,@nCluster
dec(4,3),@nCluster_NA dec(4,3), @c varchar(10)
TRUNCATE TABLE dbo.Network_Nodes
SELECT DISTINCT pid INTO dbo.#tmp_pid
       FROM dbo.Network_Edges
INSERT dbo.#tmp_pid
       SELECT DISTINCT pid_Linked
       FROM dbo.network_edges
INSERT dbo.Network_Nodes
       (pid)
       SELECT DISTINCT CAST(pid AS int) AS pid
       FROM dbo.#tmp_pid
       ORDER BY CAST(pid AS int)
SET @nEdges = (SELECT COUNT(*) FROM dbo.Network_Edges)
SET @nNodes = (SELECT COUNT(*) FROM dbo.Network_Nodes)
IF @nSize>=@nNodes
       BEGIN
            PRINT 'Error:  Parameter  Window_Size  >  Total  number  of
nodes.'
            RETURN
       END
IF @nOverlap>=@nSize
       BEGIN
            PRINT 'Error: Parameter Overlap > Window_Size.'
            RETURN
       END
---------
```

```sql
IF OBJECT_ID('tempdb.dbo.Network_Edges') IS NOT NULL
      TRUNCATE TABLE tempdb.dbo.Network_Edges
ELSE
      CREATE TABLE tempdb.dbo.Network_Edges (
            [pid] [char](7) NOT NULL,
            [pid_linked] [char](7) NOT NULL,
            [delta_t] [decimal](10, 0) NULL,
            [weight] [decimal](19, 18) NULL,
            [from_depth] [decimal](4, 2) NULL,
            [to_depth] [decimal](4, 2) NULL,
            [from_magn] [decimal](3, 2) NULL,
            [to_magn] [decimal](3, 2) NULL,
            [from_year] [smallint] NULL,
            [from_month] [tinyint] NULL,
            [from_day] [tinyint] NULL,
            [from_hour] [tinyint] NULL,
            [from_minut] [tinyint] NULL,
            [to_year] [smallint] NULL,
            [to_month] [tinyint] NULL,
            [to_day] [tinyint] NULL,
            [to_hour] [tinyint] NULL,
            [to_minut] [decimal](2, 0) NULL,
            [from_lat] [char](8) NULL,
            [from_long] [char](9) NULL,
            [to_lat] [char](8) NULL,
            [to_long] [char](9) NULL,
            [from_x] [decimal](9, 3) NULL,
            [from_y] [decimal](9, 3) NULL,
            [to_x] [decimal](9, 3) NULL,
            [to_y] [decimal](9, 3) NULL,
            [oneway] [char](2) NULL,
            [d] [decimal](5, 3) NULL,
            [w_m] [decimal](19, 18) NULL,
            [w_t] [decimal](19, 18) NULL,
            [w_d] [decimal](19, 18) NULL,
            [t_infl_sec] [decimal](10, 0) NULL)
IF OBJECT_ID('tempdb.dbo.Network_Nodes') IS NOT NULL
      TRUNCATE TABLE tempdb.dbo.Network_Nodes
ELSE
      CREATE TABLE tempdb.dbo.Network_Nodes (
            [cid] [int] IDENTITY(1,1) NOT NULL,
            [pid] [int] NOT NULL,
            [in_lines] [int] NULL,
            [in_weight] [numeric](21, 18) NULL,
            [out_lines] [int] NULL,
            [out_weight] [numeric](21, 18) NULL,
            [tot_lines] [int] NULL,
            [tot_weight] [numeric](22, 18) NULL,
            [linked_near] [int] NULL,
            [clustering] [numeric](4, 3) NULL)
--------------------------------------------
SET @i = 1
SET @nMinCid = 1
SET @nMaxCid = @nMinCid + (@nSize-1)
```

```sql
WHILE @nMaxCid <= @nNodes
      BEGIN
            SET @c = (SELECT CAST(@i AS varchar(10)))
            SET   @nMinPid   =   (SELECT    CAST(pid   AS    int)   FROM
dbo.Network_Nodes
                                        WHERE cid = @nMinCid)
            SET   @nMaxPid   =   (SELECT    CAST(pid   AS    int)   FROM
dbo.Network_Nodes
                                        WHERE cid = @nMaxCid)
            TRUNCATE TABLE tempdb.dbo.Network_Edges
            INSERT tempdb.dbo.Network_Edges
                  SELECT * FROM dbo.Network_Edges
                        WHERE CAST(pid AS int) >= @nMinPid AND
                              CAST(pid AS int)<= @nMaxPid
            EXECUTE ('
            IF OBJECT_ID(''dbo.Window_'+@c+'_Edges'') IS NOT NULL
                  BEGIN
                        TRUNCATE TABLE dbo.Window_'+@c+'_Edges
                        INSERT dbo.Window_'+@c+'_Edges
                              SELECT * FROM tempdb.dbo.Network_Edges
                  END
            ELSE
                  SELECT * INTO dbo.Window_'+@c+'_Edges
                        FROM tempdb.dbo.Network_Edges
                  ')
            EXECUTE dbo.usp_Compute_Network_Nodes_in_tempdb
            EXECUTE ('
            IF OBJECT_ID(''dbo.Window_'+@c+'_Nodes'') IS NOT NULL
                  TRUNCATE TABLE dbo.Window_'+@c+'_Nodes
            ELSE
                  CREATE TABLE dbo.Window_'+@c+'_Nodes(
                        [cid] [int] NOT NULL,
                        [pid] [int] NOT NULL,
                        [in_lines] [int] NULL,
                        [in_weight] [numeric](21, 18) NULL,
                        [out_lines] [int] NULL,
                        [out_weight] [numeric](21,18) NULL,
                        [tot_lines] [int] NULL,
                        [tot_weight] [numeric](22,18) NULL,
                        [linked_near] [int] NULL,
                        [clustering] [numeric](4, 3) NULL
                        )
            INSERT dbo.Window_'+@c+'_Nodes
                  SELECT * FROM tempdb.dbo.Network_Nodes')
            SET @nCluster = (SELECT AVG(clustering)
                              FROM tempdb.dbo.Network_Nodes)
            SET @nCluster_NA = (SELECT AVG(clustering)
                                    FROM tempdb.dbo.Network_Nodes
                                    WHERE tot_lines > 1)
      SET @nWindow_Nodes = (SELECT COUNT(*)
                                    FROM tempdb.dbo.Network_Nodes)
            SET @nWindow_Edges = (SELECT COUNT(*)
                                    FROM tempdb.dbo.Network_Edges)
            INSERT dbo.Results
                  VALUES (@cNetworkName,@nNodes,@nEdges,@nSize,
                        @nOverlap,@i,@nMinPid,@nMaxPid,@nWindow_Nodes,
                        @nWindow_Edges,@nCluster,@nCluster_NA)
```

```
                SET @nMinCid = (@nMaxCid + 1) - @nOverlap
                SET @nMaxCid = @nMinCid + (@nSize-1)
                SET @i = @i +1
        END
PRINT 'Total number of windows: ' + CAST((@i-1) AS varchar(10))
GO
```