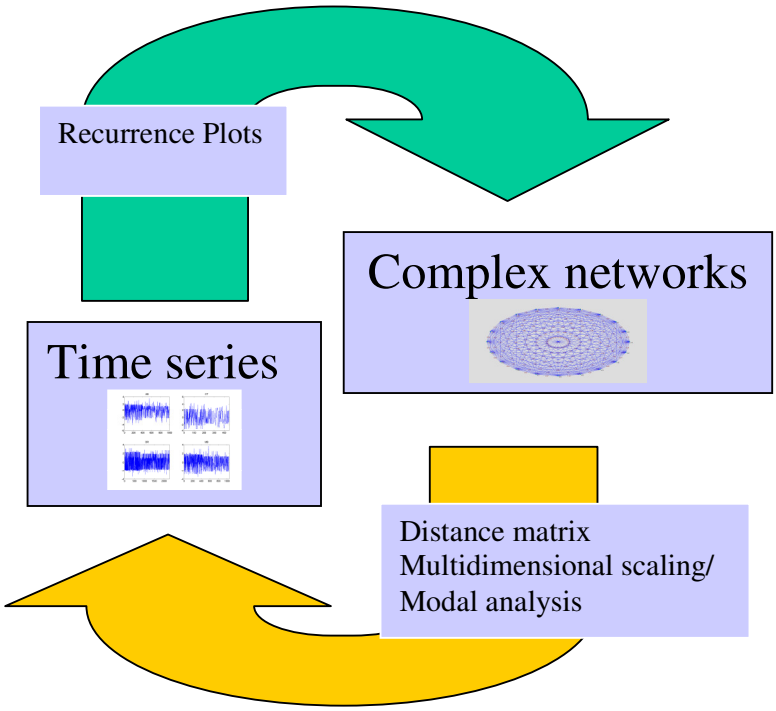
JRC Scientific and Technical Reports

FROM COMPLEX NETWORKS TO TIME SERIES ANALYSIS AND *VICEVERSA*: Application to metabolic networks

F. Strozzi[1], J.M. Zaldívar[2], K. Poljansek[3], F. Bono[3] and E. Gutiérrez[3]

[1]CarloCattaneo University, Engineering Faculty, Castellanza, Italy
European Commission, Joint Research Centre, Ispra, Italy:
[2]Institute for Health and Consumer Protection
[3]Institute for the Protection and Security of the Citizen



EUR 23947 EN - 2009

JRC
EUROPEAN COMMISSION

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

European Commission
Joint Research Centre

Contact information
Address: Via E. Fermi 2749, TP 202
E-mail: jose.zaldivar-comenges@jrc.it
Tel.: +39-0332-789202
Fax: +39-0332-785807

http://www.jrc.ec.europa.eu/

JRC 52892

EUR 23947 EN
ISBN 978-92-79-12955-1
ISSN 1018-5593
DOI 10.2788/25588

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Quality control insert

	Name	*Signature*	*Date*
Report Prepared by:	José-Manuel Zaldívar		17/06/09
Reviewed by: **(Scientific level)**	Andrew Worth		28/06/09
Approved by: **(Head of Unit)**	Elke Anklam (ff)		03/07/09
Final approval: **(IHCP Director)**	Elke Anklam		03/07/09

CONTENTS

EXECUTIVE SUMMARY

One way to describe complex interactions between different elements in a system is through complex network analysis. This approach has been applied to many real networks, such as WWW, protein-protein interaction, metabolic and social networks. In addition, it has been established that complex networks have certain common characteristics with non-linear time series such as fractality and self similarity, i.e. invariance under change of scale. In this work, we have developed a general framework to transform complex networks into non-linear time series (and vice-versa). This has allowed us to explore new research avenues that this dual vision of complex phenomena may provide. As examples we have chosen standard case studies from non-linear dynamical systems, which we have complemented with a more applied case from metabolic networks. Preliminary results are the development of a new type of complex network: chaotic networks and the analysis of the statistical properties of time series generated by several network structures. This work is a preliminary step to focus the research and to assess the most important avenues to explore more in depth.

1. INTRODUCTION

One of the main objectives of Systems Toxicology is to be able to describe the response of a functioning organism to toxicants at all levels of biological organization and complexity by combining the information from several sources to gain a deeper understanding of the mechanisms of toxic action and by developing approaches allowing a more sensitive and earlier detection of adverse effects in toxicity studies. This constitutes a big challenge since it requires to study interactions between genes, proteins and metabolites, and to integrate the results from different theoretical and experimental set-ups.

One way to describe complex interactions between different elements that has gained considerable attention during the last decade is complex network analysis. Complex networks describe a considerable amount of natural and social systems with large, irregular, and time-dependent structures, built up of thousands of nodes and connections between them. In recent years a number of new techniques for their analysis have been developed and implemented (for a review see Boccaletti et al., 2006). In addition, it has been established that complex networks have certain common characteristics with nonlinear time series such as fractality and self similarity, i.e. invariance under change of scale (Barabasi and Albert, 1999; Albert and Barabasi, 2002). For example, Song et al. (2005) have shown that many real complex networks, such as WWW, social, protein-protein interaction and cellular networks as self-similar under length-scale transformation. Strogaz (2005) argued that the self-similarity that it is not present in some standard models of complex networks (random or scale-free) may be a necessary step that appears to optimize some aspects of their performance or, as pointed out by Song et al. (2006), a necessity for robustness against targeted attacks.

The main idea of this work is to develop a general framework to transform complex networks into nonlinear time series (and vice-versa), and to explore the new research avenues that this dual vision of complex phenomena can provide us. As examples we chose standard case studies from non-linear dynamical systems, which we complement with a more applied case from metabolic networks. This work is a preliminary step to focus the research and to assess the most important avenues to explore more in depth.

2. METHODS AND APPROACH

2.1. COMPLEX NETWORK ANALYSIS

A network is a set of items called vertices or nodes with connections between them called edges (Newman, 2003). Whereas the study of networks, developed by the mathematical theory of graphs,

was mainly focused on the study and properties of small networks, recent years have witnessed the emergence of research in the field of complex networks. Complex networks describe a considerable amount of natural and social systems with large, irregular, and time-dependent structures, built up of thousands of nodes and connections between them. Certainly the continuous increase of computational power has supported the analysis of available databases of real networks, and, contrary to graph theory, has created the basis for the development of statistical properties and the identification of the unifying laws common to most of these networks (Albert and Barábasi, 2002; Newman, 2003).

Examples of real complex networks are: social networks formed by group of people with some pattern of interactions between them (e.g. friendships, genealogies, several types of collaboration, etc.); information networks formed around the flow of information (examples are the network of Web pages - the World Wide Web- linked together by hyperlinks from one page to another, the network of citations between academic papers: citation networks); technological networks which are man-made networks developed for several function such as supply chains, electric power grids (high-voltage transmission lines), networks of airlines routes, of roads, or railways, internet: the physical network of the computers connected; and biological systems networks like metabolic networks, food web networks, neural networks, blood vessel networks, etc.

Even though at the beginning it was supposed that complex networks should be treated as random graphs, i.e. undirected networks with edges placed randomly between a fixed number of vertices, it was rapidly recognised that this was not always the case and that other types of structures existed, such as small-world (Watts and Strogatz, 1998) and scale free networks (Barabási and Albert, 1999). This is important, since the properties of complex networks are related and/or encoded in their topology. For this reason, tools and measurements to capture in quantitative terms these aspects have been developed. Boccaletti et al. (2006) have recently reviewed the indices and structural properties (e.g. size, density, degree, clustering, diameter, etc.) used normally to define network topology, whereas Melián and Bascompte (2004) have shown how the network structure correlates with its robustness and its response to external perturbations, showing that the cohesive organization of a network in dense sub webs increases the resistance against fragmentation. In addition, models to understand the evolution over time (the appearance or disappearance of vertices and edges) of complex networks have been proposed (Newman, 2003) as well as methods to analyze the dynamics on the networks given a certain structure and certain statistical properties i.e. how the topology can influence the dynamics.

Real complex networks have several characteristics that make them different from random networks; for example, connection patterns that have both random and deterministic characteristics. Between them, they posses a heavy tail (i.e. vertices exist having usually high degree) in the distribution degree; a high clustering coefficient (a parameter that quantifies how it compares in connectivity to a completely connected graph); a community structure (i.e. a group of nodes connected densely between

them; assortativity or dissasortattivity among vertices, which refers to preferential attachment between similar or dissimilar nodes; small world effect, which refers to the fact that most pairs of vertices are connected between them by short paths (contrary to the longer distances typical in random networks).

Concerning biological systems, there has been in the last years a considerable development in their representation and analysis as networks. This is mainly due to the generation of a large number of datasets at different levels of biological organization and complexity. Between these systems, we can highlight the following:

- Food-web networks: A food-web constitutes a special case of a biological network, which describes a biological community focusing on trophic interactions between consumers and resources (de Ruiter et al., 2005). Therefore, food-webs are deeply interrelated with ecosystem processes and functioning, since the trophic interactions represent the transfer rates of energy and matter within the ecosystem. It is well-known that trophic webs are not randomly assembled, but are the result of the history of the community organization, ecosystem size, availability of the resources, interactions between prey and predators and disturbance (Post, 2002). Networks resulting from trophic structures show some common patterns still only partially understood, such as: small world effect, scale free topology (Newman, 2003), hierarchical organization (Clauset et al., 2008; Redner, 2008), and the presence of compartments (Krause et al., 2003). For this reason a considerable effort in ecosystem theory has been devoted to understand how food-webs are structured and how this structure influences ecosystem processes (Dunne, 2006). The development of static food-web models starts with the "cascade" (Cohen and Newman, 1985) and "niche" models (Williams and Martinez, 2000), followed by the introduction of dynamics, through bioenergetic-based models in food-webs, which has allowed the development of explicit dynamic network models of shared nutrient consumption; including competition among producers for multiple resources, as well as effects of anthropogenic pressures. This has also allowed including population dynamics within this framework (Martinez et al., 2006) and to extend the type of network parameters and analysis one is able to calculate.

- Metabolic networks: In these networks the nodes correspond to metabolites whereas the connections between these nodes correspond to reactions. These are directed networks and their statistical properties have been studied by several authors: Jeong et al. (2000); Fell and Wagner (2000); Podani et al. (2001); Wagner and Fell (2001); Stelling et al. (2002); Gulmerà and Nunes Amaral (2005), amongst others. Jeong et al. (2000) showed that metabolic networks were small world networks, however in their analysis they included connections through current metabolites, e.g. ATP, ADP, NADH, NAD^+, etc. In a subsequent analysis by Ma and Zeng (2003), they showed that even after the removal of these connections, the networks still showed a power law distribution, but with longest pathways and diameters than previously calculated by Jeong et al. (2000). In addition, they identified several hub metabolites that were common in all organisms.

- Protein interaction networks: In these undirected networks the nodes correspond to proteins whereas the connections correspond to physical interaction between them. Since many biological functions involve interactions between proteins at different levels, these networks have been analyzed by several authors: Schwikowski et al. (2000), Ito et al. (2001); Jeong et al. (2001); Maslov and Sneppen (2002); Solé and Pastor-Satorras (2003), amongst others. In addition, these networks have also been studied with drug design applications in mind (Drews, 2000).
- Genetic regulatory networks: A cell expresses only a small fraction of genes, their transcription being regulated by DNA-binding proteins. In these switching directed networks the vertices represent the proteins produced by the expression of a gene, whereas the edges represent the dependence of protein production on the proteins at other vertices. These networks have been analyzed by several authors: Guelzim et al. (2002); Shen-Orr et al. (2002); Farkas et al. (2003). In addition, transcription factors have been analyzed from the perspective of control of disease conditions and drug design applications (Yang et al., 2009).
- Neural networks: Measuring the topology of real neural networks is a challenging task (White et al. 1986; Sporns 2002). However, the development of new experimental *in vitro* approaches with simple geometries has allowed starting to gain understanding of the conceptual modelling behind neuronal networks (Novellino and Zaldívar, 2009). From these *in vitro* experiments some kind of network self-organization arises by randomly organized connectivity (Eckmann et al., 2007). However, it is not clear if this corresponds to in vivo neural networks.

In addition, in the last years there has been a continuous development of application of networks theory to biological problems, for example: conformational networks during protein folding, vascular networks, blood vessels networks, etc.

2.2. STRUCTURAL MEASURES ON COMPLEX NETWORKS

During the last years a considerable number of indices and structural properties to define network topology and behaviour have been developed. Here we refer only to the definitions and properties that will be used in the next Sections; however, a detailed review can be found in Boccaletti et al. (2006).

- Average path length (characteristic length) and diameter

The distance between two vertices is the number of edges in the shortest path connecting them. This is also known as the geodesic distance. The diameter of a graph is the maximal distance between any pair of its nodes. Milgram in the sixties examined the average path length (or characteristic length) for social networks (Milgram, 1967) and he found that human society is a small world type network characterized by short path lengths, in particular "six path length". It seems that many natural network show small world property that mathematically is expressed as a slow increase of the average path length ($\bar{l}$) when the number of nodes (N) increases i.e. $\bar{l} \approx \ln(N)$.

- Giant component

The giant component of a network is the subnetwork with all the nodes connected to each other directly or indirectly and that consists of the majority of the network.

- Clustering coefficient

The clustering coefficient [0-1] is the property of a node that tells us how well a node is connected with its neighbourhood. A value of 1 means that a node neighbourhood is fully connected. The clustering coefficient of a node is the quotient between the number of connections between its neighbourhood and the maximum possible number of connections between them. If they are *n* and they are fully connected, it will be: *n*(n-1)/2*.

The clustering coefficient of a graph is the average clustering coefficient of all nodes in the graph.

Slightly different definitions of clustering exist in the literature (such as the one based on the density of triangles in the network), but, irrespective of the precise nature of these definitions, it always results that their values are always higher for real networks in comparison to random graphs with a similar number of vertices and edges.

- Centrality measures

The centrality measures of a vertex in a graph determine the relative importance of the vertex itself. There are different centrality measures we will define and use two of them:

Degree centrality

The degree centrality of a vertex in an undirected network is the number of ties a node has. This measure represents the risk of a node to catch what is flowing in the network.

For a graph G: = (V,E) with n vertices, the degree centrality $C_D(v)$ for vertex v is:

$$C_D(v) = \frac{\deg(v)}{n-1} \tag{1}$$

The degree centrality of the graph G is defined as:

$$C_D(G) = \frac{\sum_{i=1}^{|V|}[C_D(v_m) - C_D(v_i)]}{n-2} \tag{2}$$

where v_m is the node with the maximum centrality in G. n-2 is the correspondent value of the numerator of Eq. (2) when the n-graph has a star structure. Then this quantity measures the distance between a given graph and a *star structure*.

Betweenness centrality

This can be defined on a vertex or an edge, and it measures how many times a vertex or an edge occur on the shortest paths (geodesic) between other vertices in comparison with the total number of shortest paths. For a graph G: = (V,E) with n vertices, the betweenness $C_B(v)$ for vertex v is:

$$C_B(v) = \sum_{\substack{s,t \in V \\ s,t \neq v}} \frac{\vartheta_{st}(v)}{\vartheta_{st}} \tag{3}$$

where $\vartheta_{st}(v)$ is the number of shortest paths passing across *v* and ϑ_{st} is the total number of shortest paths. This may be normalised by dividing through the number of pairs of vertices not including *v*, which is $(n - 1)(n - 2)$. Similarly it can be defined the betweenness centrality of an edge.

- **Network resilience**

The property of resilience of a network to node-removal has been studied in depth in the literature and is related to the degree distribution property. If vertices are removed from a network, the typical length of the connection between two vertices will increase and ultimately become disconnected. The level of network resilience varies in accordance with node removal, which can be performed randomly or following some special ranking, for example nodes with highest degree or with the highest "betweenness" (Holme et al. 2002). Typically, the resilience of a network is measured calculating the mean distance on the network or the size of the giant component as the fraction of vertex removed increases or the change in connectivity properties that is related to the network ability to diffuse information, illness, etc. Other measurements exist depending on the scope of the network. Looking to the size of the giant component it seems that scale free networks are more robust than random networks to random attacks but more vulnerable (less resilient) to selective attacks for example based on nodes degree.

-Cumulative Degree Distribution

Poisson distribution

The simplest network model is the Bernoulli random graph called often just random graph (Erdos and Reny, 1960). The model is developed taking a fixed number *n* of nodes and connecting every couple of nodes independently with a probability *p* (*G*(*n,p*) model). The *G*(*n,p*) model has various interesting properties. The distribution of the degree of any particular vertex is *binomial*:

$$B(n, p) = P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{4}$$

The binomial distribution converges towards the *Poisson distribution* for large *n* and small *p* to maintain $n \cdot p$ constant:

$$B(n, p) = P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \cong \frac{z^k e^{-z}}{k!} \tag{5}$$

where $z=(n-1)p$ is the mean degree.

If *n* is large enough but *p* remains fixed and the skew of the distribution is not too great, a continuous approximation of *B*(*n*, *p*) is given by the *normal distribution*

The tails of the binomial, normal and Poisson distributions approach zero *exponentially fast* (Poisson distribution decays rapidly for a large values of z) this means that the probability of large values shrinks quickly and becomes nearly zero. Nevertheless, in the literature the degree distributions of many real networks seems to be highly non-Poisson and often with a fat tail, i.e. vertices exist having unusually high degree (Albert and Barabasi, 2002). Moreover the high degree nodes in the tail can have an important effect on the behaviour of the networked system.

Power law distribution

Power law distributions have, on the other hand, a fat or heavy tail. This means that the probability to have to find a large value is small but it does happen. Mathematically one can say that

$$1 - P(x < k) = P(x \geq k) \cong k^{-\alpha} \qquad (6)$$

i.e the cumulative probability that the degree of a node (x) is higher than k is proportional to $k^{-\alpha}$ for some exponent α. Usually this function is represented in a logarithmic scale in order to measure the slope α of the straight line.

Sometimes the power law scaling is evident only in the tail e.g. scientific paper citations, or earthquake magnitude (Barabasi and Albert, 1999). A network is often called *scale-free* if its degree distribution follows a power law.

- Fractal structure

Song et al. (2005) showed that many real complex networks, such as WWW, social, protein-protein interaction networks and cellular networks are self-similar under length-scale transformation. They tiled networks with boxes and identify a power-law relation between boxes needed to cover the network and the size of boxes just as if they were of fractal shape. Strogatz (2005) recognized the same self-similarity in Romanesque Broccoli and he assumed that the fractal shape may be the indication of a new architectural law of complex systems. Moreover he observed that some real networks and some standard models such as random graphs or scale-free networks lack of self-similarity and such scaling is not automatic but could be a necessary step to optimize some aspect of their performance. Song et al. (2006) suggested the possibility of a unique growth mechanism of real complex networks. They observed that many real networks are self-similar and they showed that a robust network such as a cellular network necessitates a fractal topology. Moreover they suggested growth rules that ensure the fractal geometry together with other well known properties of real complex networks i.e. small-world effect (Watts and Strogatz, 1998) and scale-freeness. They showed that the fractal property of networks significantly increases the robustness against targeted attacks because hubs are more dispersed. They demonstrated that, in general, scale free networks are not fractal and showed that the rule of growth they proposed (the inverse of renormalization procedure they use to measure the fractal dimension) allows the emergence of self-similarity thanks to the imposition of strong repulsion (disassortativity) between the hubs at all length scales i.e. the hubs prefer to grow by connections to

less-connected nodes rather than to other hubs i.e. there is a hub repulsion. They modified "the rich get richer" at the basis of preferential attachment principle with "the rich still get richer although at the expense of the poor".

Goth et al (2006) recognized too that fractal scaling is one important characteristic of real networks together with small-world and scale-free properties. They found that the fractal scaling in the scale-free networks originates from a special type of spanning tree based on edge betweenness centrality and they called this tree: fractal skeleton. Kim et al (2007) improved the algorithm of Song et al (2005) to study the fractal scaling of some real networks and propose another fractal network model.

- Evolving networks

Real networks are often evolving networks, which imply that new edges and nodes appear and others disappear. In the literature the process of network growth has been studied to explain the characteristic features of the networks (Newman, 2003). The most studied class of network growth models by far are those that aim at explaining the origin of the highly skewed degree distributions. The most popular is the one of Barabasi and Albert (1999) which is based on the principle of growth and of preferential attachment. For example, a paper gets new citations proportionally to the number of citations that it already has. When a new node appears it will be connected to the others proportionally to their number of links. It seems that this model justifies the power law distribution. However, this model implies a correlation between the age of vertices and their degree which not always exists in real networks.

Different ways to study network evolution applied to real networks were introduced by Latapy and Magnien (2007, 2008) and Borgnat et al (2008). First the evolving network can be considered as a sequence of snapshots and each of these can be studied separately using the properties of static networks; secondly the properties can be defined as a function of the evolution itself, such as the duration of contacts in a network or the evolution of communities in time or the rate of appearance of nodes and edges; finally, specific users or phenomena such as diffusion of information can be studied in an evolving network.

-Reproductive number R_0 /connectedness

The basic reproduction number (or rate or ratio) R_0 of an infection is the mean number of secondary cases infected by a single case in a population without immunity or control abilities. Another context in which it appears is the spread of information (idea, rumours or others) in a network. Let us consider the diffusion of an idea: if a node with k degree communicates an idea with a probability r to reach the friends and we exclude the ones which communicated to it the idea, then the expected number of people to whom he will pass the idea will be $r(k\text{-}1)$. Taking the weighted average over the vertices we obtain R_0:

$$R_0 = r\frac{\sum_i k_i(k_i-1)}{\sum_i k_i} = r\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} \qquad (7)$$

If R_0>1 then the number of people hearing the idea grows exponentially; if R_0<1 the idea will die. The dependence by mean square degree means that a person with high degree has more possibility to hear and to spread the idea. The spread of the idea occurs if and only if

$$r > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{8}$$

When $\langle k^2 \rangle$ is very large in comparison with $\langle k \rangle$, r need not to be large to spread the idea.

Instead to check if R_0>1 one can check for $\kappa = \langle k^2 \rangle / \langle k \rangle > 2$. κ is called connectivity parameter and the value of 2 is a percolation threshold when the giant component disappears and the network becomes a set of small clusters.

2.3. NONLINEAR TIME SERIES ANALYSIS

The mathematical basis of continuous dynamical modelling is formed by differential equations of the following type:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \alpha) \tag{9}$$

where the real variable t denotes time, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ represents the state variables of the system, depending on time t and on the initial conditions, and α_j are parameters of the system, while $\boldsymbol{F} = (F_1, F_2, \ldots, F_n)$ is a nonlinear function of these variables and parameters. Actual states of these systems are described by the vector variable $\boldsymbol{x}$ consisting of n independent components. Each state of the system corresponds to a definite point in phase space, which is called the phase point. The time variation of the state of the system is represented as a motion along some curve called phase trajectory. Experimentally, it is not always possible to measure the complete state of a system and, normally, when analysing a dynamical system, we have access to few observable quantities which, in the absence of noise, are related to the state space coordinates by:

$$s(t) = \mathbf{h}(\mathbf{x}(t)) \tag{10}$$

where $\boldsymbol{h}$ is normally an unknown nonlinear function called the measurement function. The theory of embedding is a way to map a temporal time series of measurements to a state space "similar" -in a topological sense- to that of the underlying dynamical system we are interested in analysing. Techniques of state space reconstruction were introduced by Packard *et al.* (1981) and Takens (1981), who showed that it is possible to address this problem using measurements of a sufficient long time series, $s(t)$, of the dynamical system of interest. Takens proved that, under certain conditions, the dynamics on the attractor of the underlying original system has a one-to-one correspondence with measurements of a limited number of variables. This observation opened a new field of research. In fact, if the equations defining the underlying dynamical system are not known, and we are not able to

measure all the state space variables, the state space of the original system is not directly accessible to us. However, if by measuring a fewer variables we are able to reconstruct a one-to-one correspondence between the reconstructed state space and the original, this means that it is possible to identify unambiguously the original state space from measurements. Embedding theory has opened a new field of research: nonlinear time series analysis (Tong, 1990; Abarbanel, 1996; Kantz and Schreiber, 1997; Diks, 1999, amongst others).

Nonlinear analysis of experimental time series has, among its goals, the separation of high-dimensional and stochastic dynamics from low-dimensional deterministic signals, estimation of system parameters or invariants (characterisation), and, finally, prediction, modelling and control.

Unfortunately, often it is very difficult to tell whether a series is stochastic or deterministically chaotic or some combination of these categories (Zaldívar et al., 2000; Strozzi and Zaldívar, 2002). More generally, the extent to which a non-linear deterministic process retains its properties when corrupted by noise is also unclear (Takens, 1996). The noise can affect a system in different ways, either in an additive sense or as a measurement error, even though the equations of the system remain deterministic.

Since a single reliable statistical test for chaoticity is not available, combining multiple tests is a crucial aspect; especially when one is dealing with limited and noisy data sets.

There are different aspects that should be carefully studied before attempting to go further using nonlinear time series analysis methods. A long and exhaustive discussion can be found in Schreiber (1998) and the basic methodologies will be presented during the analysis part. Here, we are briefly going to indicate the main problems one should be aware of. These can be summarized as follows:

- has the phenomenon been sufficiently sampled?;
- is the data set stationary or can one remove the nonstationary part?;
- is the level of noise sufficiently low so that one can obtain useful information using nonlinear time series techniques?

Some tests to study these questions have been recently implemented in the TISEAN software package (Hegger *et al*., 1999), which has incorporated a substantial quantity of algorithms developed for nonlinear time series analysis.

2.2.1. Preliminary R/S analysis

A tool for studying long-term memory and fractality of a time series is the Rescaled Range analysis (R/S analysis) first introduced by Hurst (1951) in hydrology. Mandelbrot (1983) argued that R/S analysis is a more powerful tool in detecting long range dependence when compared to more conventional analysis like autocorrelation analysis, variance ratios and spectral analysis. In this method, one measures how the range of cumulative deviations from the mean of the series is changing

with the time. It has been found that, for some time series, the dependence of R/S on the number of data points (or time) follows an empirical power law described as $(R/S)_n=(R/S)_0 n^H$, where $(R/S)_0$ is a constant, n is the time index for periods of different length, and H is the Hurst exponent. $(R/S)_n$ is defined as

$$\left(\frac{R}{S}\right)_n = \frac{\max_{1\le t\le n} A(t,n) - \min_{1\le t\le n} A(t,n)}{\sqrt{\frac{1}{n}\sum_{t=1}^{\tau}(s(t)-\langle s\rangle_n)^2}} \tag{11}$$

where $A(t,n)$ is the accumulated departure of the time series $s(t)$ from the time average over the time interval n: $\langle s\rangle_n \quad A(t,n) = \sum_{i=t}^{t+n}(s(i)-\langle s\rangle_n)$.

The Hurst exponent, $0\le H\le 1$, is equal to 0.5 for random walk time series, <0.5 for anticorrelated series, and >0.5 for positively correlated series. The Hurst exponent is directly related to the "fractal dimension", which gives a measure of the roughness of a surface. The relationship between the fractal dimension, D, and the Hurst exponent, H, is:

$$D = 2 - H \tag{12}$$

Hurst exponents quantify the correlation of a fractional Brownian motion. A fractional Brownian motion (fBm) is a random walk with a Hurst exponent different from 0.5 and then with a memory. The decaying of spectral density of a fBm has a relationship with the Hurst exponent as follow:

$$spectral\ density \propto \frac{1}{f^{\alpha}} \tag{13}$$

where $\alpha = 2H + 1$.

A long memory process is a process with a random component, where a past event has a decaying effect on future events. The process has some memory of past events, which is "forgotten" as time moves forward. The mathematical definition of long memory processes is given in terms of autocorrelation. When a data set exhibits autocorrelation, a value x_i at time t_i is correlated with a value x_{i+d} at time t_{i+d}, where d is some time increment in the future. In a long memory process autocorrelation decays over time and the decay follows a power law, i.e

$$p(k) = Ck^{-\beta} \tag{14}$$

where, C is a constant and $p(k)$ is the autocorrelation function with lag k. The Hurst exponent is related to the exponent β by

$$H = 1 - \frac{\beta}{2} \tag{15}$$

In this work we have used the standard scaled windowed variance method (Cannon *et al.*, 1997) to estimate H by linear regression of *log(R/S)* versus *log(Windowsize)*.

2.2.2. State space reconstruction: Finding the time delay and embedding dimension

Let *s(t)* be the measure of some variable of our system, see Eq. (10), Takens (1981) shown that instead of derivatives, $\{s(t), \dot{s}(t), \ddot{s}(t), ...\}$, one can use delay coordinates, $\{s(t), s(t+\Delta t), s(t+2\Delta t), ...\}$, where Δt is a suitably chosen time delay. In fact, looking at the following approximation of the derivative of *s(t)*:

$$\frac{ds(t)}{dt} \cong \frac{s(t+\Delta t) - s(t)}{\Delta t} \tag{16}$$

$$\frac{d^2 s(t)}{dt^2} \cong \frac{s(t+2\Delta t) - 2s(t+\Delta t) + s(t)}{2\Delta t^2} \tag{17}$$

It is clear that the new information brought from every new derivative is contained in the series of the delay coordinates. The advantage of using delay coordinates instead of derivatives is that in case of high dimensions higher order derivatives will tend to amplify considerably the noise in the measurements. The embedding theorem is important because it gives a rigorous justification for the state space reconstruction. However, Takens' theorem is true for the unrealistic case of an infinite, noise-free, number of points. Takens showed that, in this case, the choice of the time delay is not relevant, and gave indications only on the choice of the embedding dimension.

Nevertheless, in real applications, the proper choice of the time delay τ and the calculation of an embedding dimension, d_E, are fundamental for starting to analyse the data. As a matter of fact, a lot of research on state space reconstruction has centred on the problems of choosing the time delay and the embedding dimension which we can call the parameters of the reconstruction for delay coordinates.

- Time delay

The first step in phase space reconstruction is to choose an optimum delay parameter τ. Different prescriptions have appeared in the literature to choose τ, but they are all empirical in nature and do not necessarily provide appropriate estimates. In this work, we have used the first minimum of the Average mutual information (AMI) function (Fraser and Swinney, 1986). *I(τ)*, as a kind of nonlinear correlation function to determine when the values of *s(n)* and *s(n+ τ)* are independent enough of each other to be useful as coordinates in a time delay vector but not so independent as to have no connection which each other at all. For a discrete time series, *I(τ)* can be calculated as,

$$I(\tau) = \sum_{n,n+T} P(s(n), s(n+\tau)) \log_2 \left[\frac{P(s(n), s(n+\tau)}{P(s(n))P(s(n+\tau))} \right] \tag{18}$$

where *P(s(n))* refers to individual probability and *P(s(n),s(n+ τ))* is the joint probability density. Following the method developed by Abarbanel (1996), to determine *P(s(n))* we simply project the values taken form *s(n)* versus *n* back onto the *s(n)* axis and form an histogram of the values. Once normalised, this gives us *P(s(n))*. For the join distribution of *s(n)* and *s(n+ τ)* we form the two-dimensional histogram in the same way.

- Embedding dimension

The dimension, where a time delay reconstruction of the system phase space provides a necessary number of coordinates to unfold the dynamics from overlaps on itself caused by projection, is called the embedding dimension, d_E. This is a global dimension, which can be different from the real dimension. Furthermore, this dimension depends on the time series measurement, and hence, if we measure two different variables of the system, there is no guarantee that the d_E from time delay reconstruction will be the same from each of them.

The usual method for choosing the minimum embedding dimension is to compute some invariant of the attractor. By increasing the embedding dimension used for the computations, one notes when the value of the invariant stops changing. Since these invariants are geometric properties of the dynamics, they become independent of d for $d \geq d_E$, i.e. after the geometry is unfolded.

In this work, we have used three methods the method of False Nearest Neighbours (FNN) developed by Kennel *et. al* (1992). In this case, the condition of no self-intersection states that if the dynamics is to be reconstructed successfully in R^d, then all the neighbour points in R^d should be also neighbours in R^{d+1}. The method checks the neighbours in successively higher embedding dimensions until it finds only a negligible number of false neighbours when increasing dimension from d to $d+1$. This d is chosen as the embedding dimension.

It was found by Kennel *et al.* (1992) that if the data set is noise-free, the percentage of false nearest neighbours will drop from nearly 100% in dimension one to strictly zero when d_E is reached. Further, it will remain zero from then on since the dynamics is unfolded. If the signal is contaminated with noise (infinite dimension signal) we may not see the percentage of false nearest neighbours drop to near zero in any dimension. In this case, depending on the signal-to-noise ratio the determination of d_E will degrade.

A problem in embedding theory is that Takens' theorem has been proven for noise-free systems. Unfortunately, there is always a certain amount of noise, $\sigma(t)$, in real data. Such noise can appear in both the measurements and the dynamics (Diks, 1999). Observational noise, i.e. $s(t)=h(\mathbf{x}(t))+\sigma(t)$, does not affect the evolution of the dynamical system, whereas dynamical noise acts directly on the state of the dynamical system influencing its evolution, for example: $d\mathbf{x}/dt=\mathbf{F}(\mathbf{x},\alpha)+\sigma(t)$.

2.4. RECURRENCE PLOTS AND RECURRENCE QUANTIFICATION ANALYSIS

Eckmann et al. (1987) introduced a new graphical tool, which they called a recurrence plot (RP). The recurrence plot is based on the computation of the distance matrix between the reconstructed points in the phase space:

$$d_{ij} = \left\| S_i - S_j \right\| \tag{19}$$

This produces an array of distances in a *nxn* square matrix, **D**, n being the number of points under study. If this distance is lower that a predetermined cut-off, ε, the pixel located at specific (*i,j*) coordinates is darkened. These points highlight the recurrences of the dynamical systems and the recurrent plot provides insight into periodic structures and clustering properties that are not apparent in the original time series.

To extend the original concept and make it more quantitative, Zbilut and Webber (1992) developed a methodology called Recurrence Quantification Analysis (RQA). As a result, they defined several measures of complexity to quantify the small scale structures in RP (http://homepages.luc.edu/~cwebber). These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP (Webber and Zbilut, 1994). A computation of these measures in small windows (sub-matrices) of the RP moving along the main diagonal yields the time dependent behaviour of these variables (Weber and Zbilut, 1994). Some studies based on RQA measures show that they are able to identify bifurcation points, especially chaos-order transitions (Trulla *et al*., 1996). The vertical structures in the RP are related to intermittency and laminar states: those measures quantifying the vertical structures enable to detect chaos-chaos transitions (Marwan *et al*., 2002). The measures to quantify complexity of RPs are the following:

2.3.1. Measures based on recurrence density

Here we define *%recurrence* (RR) as the percentage of darkened pixels in recurrence plot:

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}(\varepsilon) \tag{20}$$

where $\mathbf{R}_{i,j}(\varepsilon)$ is one if the state of the system at time I and the one at time j have a distance less than ε and zero otherwise. *RR* is a measure of the density of recurrence points in RP.

2.3.2 Measures based on diagonal lines

Let P(ε,l) be the histogram of diagonal lines of length l. If we assume we have obtained the right value of ε then we can consider $P(\varepsilon, l) = P(l)$. Processes with uncorrelated or weakly correlated behaviour cause none or very short diagonals, whereas deterministic processes cause longer diagonals. In analogues manner *%determinism* (*DET*) is the ratio of recurrence points that form diagonal structures (of at least length l_{min}) to all recurrence points

$$DET = \frac{\sum_{l=l_{\min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{21}$$

DET is then the percentage of recurrent points forming diagonal line structures. If l_{min} =1 the determinism is one. For the choice of l_{min} we have to take into account that the histogram *P(l)* can become sparse if l_{min} is too large, and, thus, the reliability of *DET* decreases.

Another RQA measure considers the length $L_{\max}$ of the longest diagonal line found in the RP, or its inverse, the *divergence (DIV)*

$$L_{\max} = \max\left(\{l_i\}_{i=1}^{N_l}\right), \text{ respectively } DIV = \frac{1}{L_{\max}} \tag{22}$$

where $N_l = \sum_{l \geq l_{\min}} P(l)$ is the total number of diagonal lines.

These measures are related to the exponential divergence of the phase space trajectory. The faster the trajectory segments diverge, the shorter are the diagonal lines and the higher is *DIV*.

The measure *entropy* (*ENTR*) refers to the Shannon entropy of the probability $p(l) = P(l) / N_l$ to find a diagonal line of length l in RP.

$$ENTR = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l) \tag{23}$$

ENTR reflects the complexity of the RP in respect of the diagonal lines. For uncorrelated noise the value of ENTR is rather small, indicating its low complexity.

Trend is a measure of the paling recurrence points away from the central diagonal. It is a linear regression coefficient over recurrence point density of the diagonals parallel to main diagonal as a function of the time distance between these diagonals and the main diagonal. It provides information about non-stationarity in the process, especially if a drift is present in the trajectory. *Trend* will depend strongly on the size of the window and may yield ambiguous results for different window sizes.

2.3.3. Measures based on vertical lines

We can find vertical lines in presence of laminar states in intermittence regimes. Let the total number of vertical lines of length ν in RP is given by the histogram $P(v)$ and, in analogy to the definition of the determinism, the ratio between the recurrence points forming the vertical structures and the entire set of recurrence points can be computed:

$$LAM = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)} \tag{24}$$

which is referred to as *laminarity*. The computation of *LAM* is realised for those ν that exceed a minimal length $\nu_{\min}$. *LAM* represents the occurrence of laminar stares in the system without describing the length of these laminar phases. *LAM* will decrease if the RP consists of more single recurrence points than vertical structures.

The average length of vertical structures is given by

$$TT = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=v_{\min}}^{N} P(v)} \tag{25}$$

and is called *Trapping Time*. *TT* estimates the mean time that the system will abide at a specific state or how long the state will be trapped.

In contrast to the RQA measures based on diagonal lines, these measures are able to find chaos-chaos transitions. Since periodic dynamics the measures quantifying vertical structures are zero, chaos-order transition can be identified (Marwan et al., 2002).

For a recent overview of the quantifying techniques and their applications, the reader is referred to Marwan et al. (2007) and to http://tocsy.agnld.uni-postdam.de.

2.5. FROM TIME SERIES TO NETWORKS AND *VICEVERSA*

From time series to networks

The idea to build up a bridge between complex networks and time series is quite recent. Zhang and Small (2006) constructed complex networks from pseudoperiodic time series. They considered a noisy periodic signal and a chaotic time series generated by the Rössler system and showed that such series exhibit small-world effect (Watts and Strogatz, 1998) and scale free features. The procedure developed consists in associating to each cycle of the time series a node of a graph and then to determine if two nodes are connected to each other by measuring if their distance in the state space is lower than a predetermined value. The generalization of their approach was to consider a fully connected graph with weighted connections between nodes depending on their distance in the reconstructed state space.

Another method to convert time series into a graph, termed the visibility graph, was proposed by Lacasa et al. (2008). Their algorithm considers the time series as a landscape. Each point is a node and a point is connected with all those that can be seen from the top of it preserving the temporal order. The algorithm transforms periodic time series into regular graphs, random time series into random graphs and fractal time series into scale-free networks. The advantage of this approach with respect to Zhang and Small (2006) is that their algorithm does not require pseudoperiodicity in the time series. This method can also be generalized to produce weighted networks or directed graphs.

Xu et al. (2008) developed another approach to mode from time series to networks. They embedded the time series in an appropriate state space with a given time delay and embedding dimension, then they selected a fixed number of nearest neighbours (four, in their example) to each point to connect the nodes. They observed that subgraph patterns were related to the recurrence properties of the original time series. Furthermore, they noticed the strong similarity of all these methods with Recurrence plots,

suggesting that it is possible to view the recurrence plot as an adjacency matrix of a complex network but the typical network properties have no correspondence with the time ordering of RPs introduced by the time series.

Hirata et al. (2008), with a different idea in mind, proposed a method to reproduce distance matrices and original time series from recurrence plots. They proposed a procedure to convert a time series in a weighted graph using RPs and then from the graph reproduce the original time series (Fig. 2.1). However, they did not use the RP to build up the adjacency matrix but rather for constructing a new distance matrix.

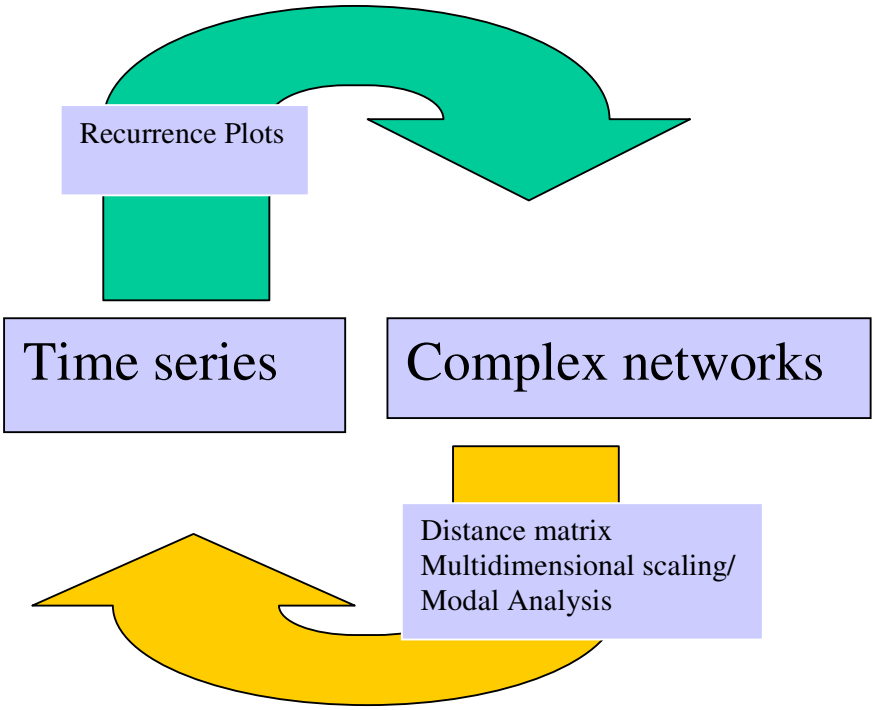


Figure 2.1. Schematic representation of how to move from time series to networks and *viceversa*.

The algorithm presented in this work consists in the calculation of recurrence plot (RP) of the time series. This RP is then considered as an adjacency matrix of an undirected graph as Xu et al. (2008) suggested. Then complex network properties can be calculated using this adjacency matrix.

From network to time series: multidimensional scaling and modal analysis

Hirata et al. (2008) transformed the recurrence plot obtained from a time series into a distance matrix considering this as a weighted network and then they rebuilt the time series using a multidimensional scaling.

Metric multidimensional scaling transforms a distance matrix into a set of coordinates; for example, in an Euclidean space. A distance matrix cannot be analyzed directly because it is not positive semi-definite, but it can be transformed into an equivalent cross-product matrix. The multidimensional scaling consists in transforming the distance matrix in a cross-product matrix, if the variables are nor centred nor normalized, or correlation matrix, if the variables are centred and normalized, or

covariance, if they are centred but not normalized, and then in applying a principal component analysis by performing the eigen-decomposition.

The algorithm proposed by Hirata et al. (2008) is the following:

1) Construct the recurrence plot. They construct a recurrence plot starting form a time series: if $z(i)$ is the state at time index i ($1 \le i \le n$), and let $c(i,j)$ the distance between the states at time i and time j in respect to some metric and in some state space. $c(i,j)$ tells us if the two states are close or not then c returning 1 or 0 respectively. The RP can be defined as $\{(i,j): c(z_i, z_j) = 1, i, j = 1,2,...n\}$

2) Construct a graph from RP. In this graph each node corresponds to a time index and an edge between nodes i and j exists if and only if $c(z_i, z_j) = 1$.

3) For each edge they assigned a weight: let $G_i = \{j : c(z_i, z_j) = 1\}$ i.e. the indices of the states close to state $z(i)$. For each existing edged between i and j it is defined a weight w:

$$w(i,j) = 1 - \frac{|G_i \cap G_j|}{|G_i \cup G_j|} \tag{26}$$

where $|B|$ is the number of elements of set B.

4) Calculated the shortest distance between each pair of nodes of the weighted graph $w(i,j)$ using Dijkstra method (Dijkstra,1959).

5) Apply a method for multidimensional scaling (Grower, 1966) which redistributed a set of points in order to preserve the distance matrix. Hirata et al. (2008) plotted only the component with the highest eigenvalues, and it works, but in general we have to check the explained variance along the eigen directions to decide. Hirata et al. (2008) applied the classical multidimensional scaling implemented in Matlab© (Grower, 1966).

In this work we called the reconstruction method introduced by Hirata et al. 2008 “method 1” where multidimensional scaling is applied using the Matlab function *cmdscale*. In addition, we propose another reconstruction method (“method 2”; Gutiérrez and Zaldívar, 2000) in which we reconstructed directly the time series starting from the adjacency matrix without using the transformation function proposed in point 3) of the Hirata et al. (2008) algorithm. Method 2 consists in the following steps with the corresponding Matlab code:

1) Let A the adjacency matrix of given undirected networks

2) Calculate the covariance matrix: *CD=cov(A)*

3) Calculate the eigenvalues and the eigenvectors of covariance matrix: *[Evec,Eval]=eig(CD)*

4) Project the covariance matrix along the eigenvector basis: *P=CD2'*Evec*

5) Sum the modes: *sr=sum(P,2)*

As Xu et al. (2008) underlined, a problem of this approach is that if we change the order in which the network nodes are visited, the locations of 1’s in the adjacency matrix change and then the

reconstructed time series can be different even if the network is the same, but once we have established an order in which the nodes are visited, the reconstruction of the time series is unique. The order can be given by the appearance of a new node, and than using the time series we can compare different evolutions rules or, for example, given by some ranking procedure such as the ones based on centrality measures (degree, betweenness, etc).

A complete loop: Time series-Network-Time series

Let us consider as an example the case already illustrated in Hirata et al. (2008) for the chaotic Hénon map. Figure 2.2 illustrates the results from the algorithm (in this case method 1, but similar results are obtained using method 2). Even though there is always some los of information, the reconstructed time series preserves most of the information and characteristics of the original time series, as it was demonstrated by Hirata et al. (2008). Better results are obtained for more periodic time series, e.g. periodic functions.

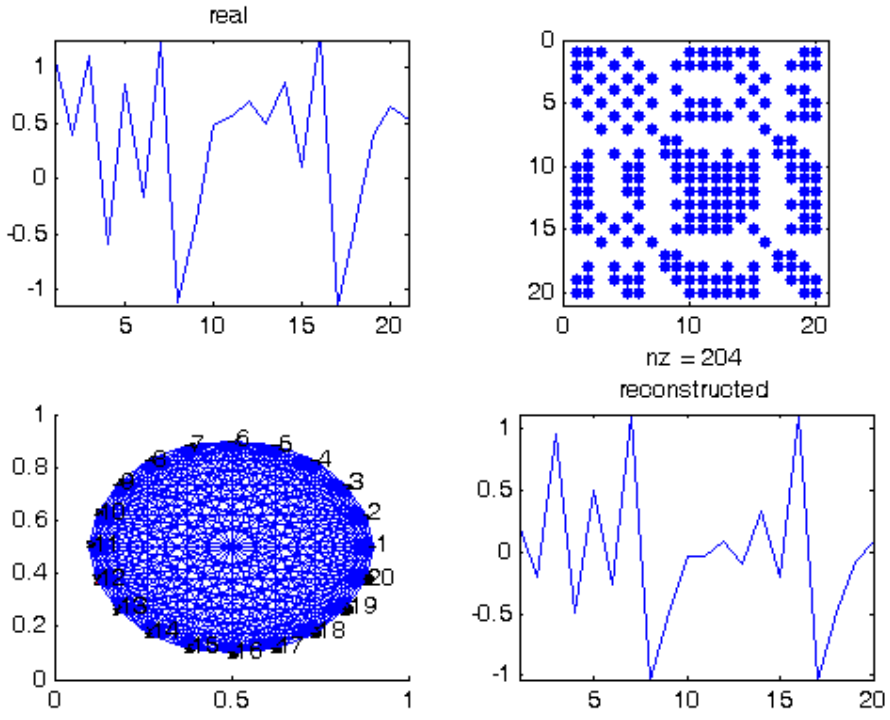


Figure 2.2. From time series to recurrence plot, to network and to time series again.

3. CASE STUDIES

3.1. FROM TIME SERIES TO NETWORKS

- Time series considered:

We have considered the time series generated by three dynamical systems:

a) Random numbers, generated using the Matlab function *rand* as follows:

y=rand(1,500)

i.e a the states have an uniform distribution between 0 and 1.

b) Sinusoidal (two periods) generated as follows:

*x=[0:5 *pi/500:5*pi]; y=sin(x);*

c) Lorenz oscillator (Lorenz, 1963) is given by the system equations:

$$\begin{cases} \dfrac{dx(t)}{dt} = \sigma\big(y(t) - x(t)\big) \\ \dfrac{dy(t)}{dt} = -y(t) + (r - z(t)) \cdot x(t) \\ \dfrac{dz(t)}{dt} = -b \cdot z(t) + x(t) \cdot y(t) \end{cases} \tag{27}$$

The Lorenz oscillator is non-linear, deterministic and three-dimensional. Tucker (2002) proved that for a certain set of parameters the system exhibits chaotic behaviour. To ensure chaotic behaviour we use the values *r=28*, *b=8/3* and $\sigma = 10$, where r and σ are the Rayleigh and Prandtl numbers, respectively. When the behaviour is chaotic the system shows a "strange" attractor and Grasberger (1983) estimated an Housdorff dimension between 2 and 3 i.e. 2.06 ± 0.01.

We have integrated the system Eq. (27) using Matlab function *ode45* with initial conditions:

$x(0) = 1; y(0) = 1; z(0) = 1;$ with a fixed time step *h=0.1*. from *t=0* until $t = 500 \cdot h$ in order to obtain 500 state points. Figure 3.1 shows the three time series.

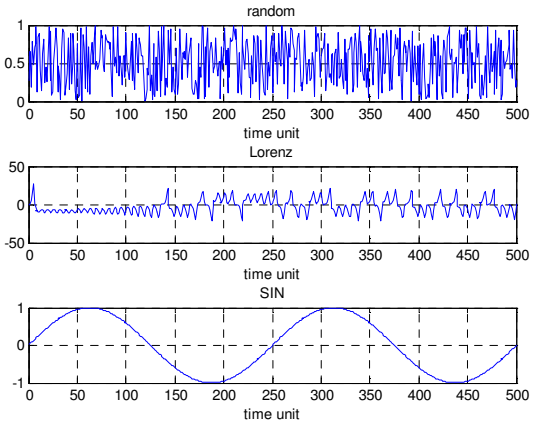


Figure 3.1. Time series considered.

The time series have then been reconstructed using a time delay $\tau = 3$ and an embedding dimension, d_E =3. These parameters ensure the unfolding of Lorenz and sin trajectories and are used even for random in order to compare the results. The reconstructed attractors are shown in Fig. 3.2. As can be observed, the random series covers the phase space, the Lorenz attractor shows its typical butterfly structure, whereas the sine function is represented by a limit cycle.

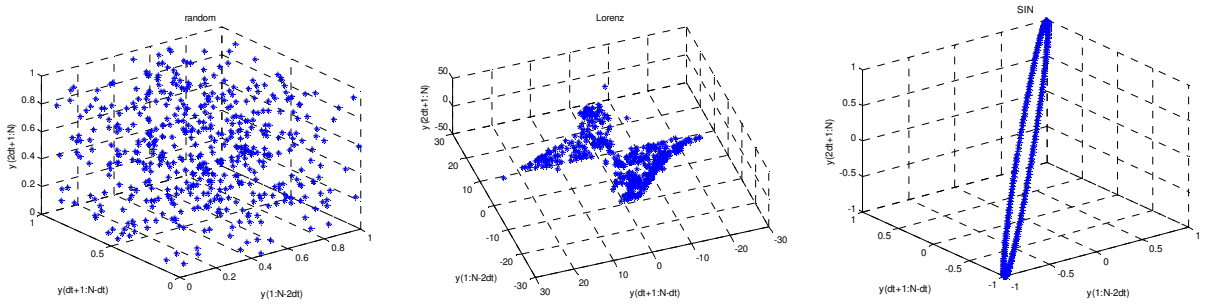


Figure 3.2. Reconstructed state spaces for the three time series.

In addition, the corresponding RP have been constructed, see Fig. 3.3. Following Marwan et al. (2008), we have taken ε as 1/10 of the maximum distance of the normalized time series, Eq. (28):

$$p = \frac{1}{10}\max\left(\frac{y(t) - \min_t(y(t))}{\max_t(y(t)) - \min_t(y(t))}\right) \tag{28}$$

where *y(t)* is the original time series.

As explained before, these RPs will constitute the adjacency matrices of some networks. In this way, we have created a series of evolving networks, with an increasing number of nodes, by considering time series of increasing lengths. Other evolving networks can be built starting from a given time series of a fixed length, modifying the cut-off radius, ε, in the RP; in this case we are increasing the number of connections. In the limit case, $\varepsilon >$ *max. distance*, then we obtain full connectivity.

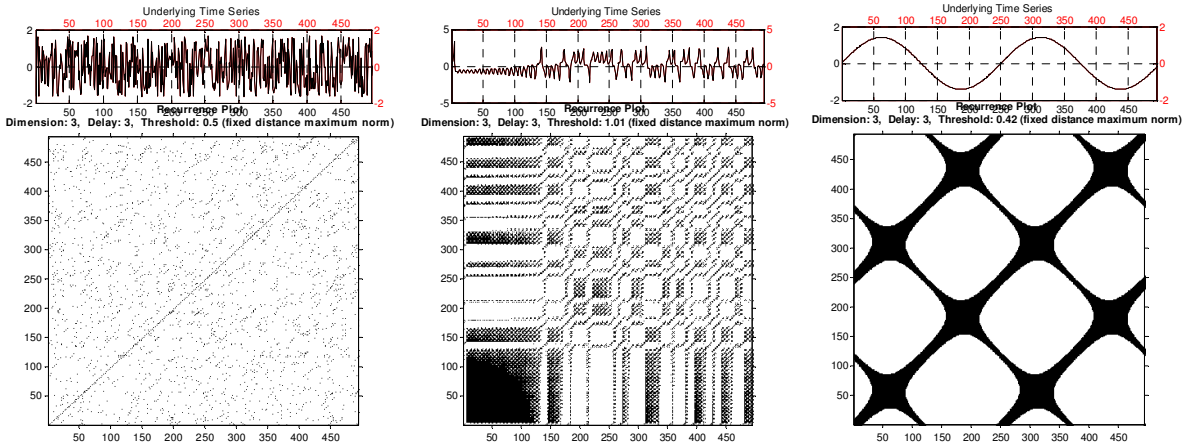


Figure 3.3. Corresponding RPs for the three selected time series.

To assess the influence of the reconstruction parameters (time delay, embedding dimension and cut-off radius) on the properties of the networks, we have performed a series of numerical experiments considering time series on windows of increasing width, that give rise to evolving networks with an increasing number of nodes, and we measured how some networks properties change when the RP parameters change.

We have considered a data window that increases by 25 points every time unit, i.e., we started with networks of 25 nodes and we stopped with a window of 19*25 nodes. The reconstruction parameters are time delay (τ) and embedding dimension (d_E). Another important parameter of recurrence plot is the cut-off radius (ε) that determines the density of recurrence plot and then of the adjacency matrix. We have considered some values of those parameters around ε =3, d_E =3, τ =3 i.ee. values that ensure the unfolding of Lorenz and sinusoidal dynamics. The values considered are: [3,7,11,15] for τ, [1,2,3,4] for d_E and [2*p, 3*p, 4*p, 5*p] for ε where p is given by Eq. (28).

In Figures 3.4-3.12 we have plotted different measures of the networks with respect to the increase of the numbers of nodes in accordance with the three different dynamics i.e. Random, Lorenz or Sine.

Fig. 3.4 shows the cumulative degree distributions i.e. the probability that the degree of a node is higher than a fixed value k, P($x \geq k$) where x is the node degree. It seems that Lorenz network has an exponential cumulative degree distribution until a certain degree and the nodes have a higher probability of higher degree in comparison with Random and Sin networks independently from the RQA parameters.

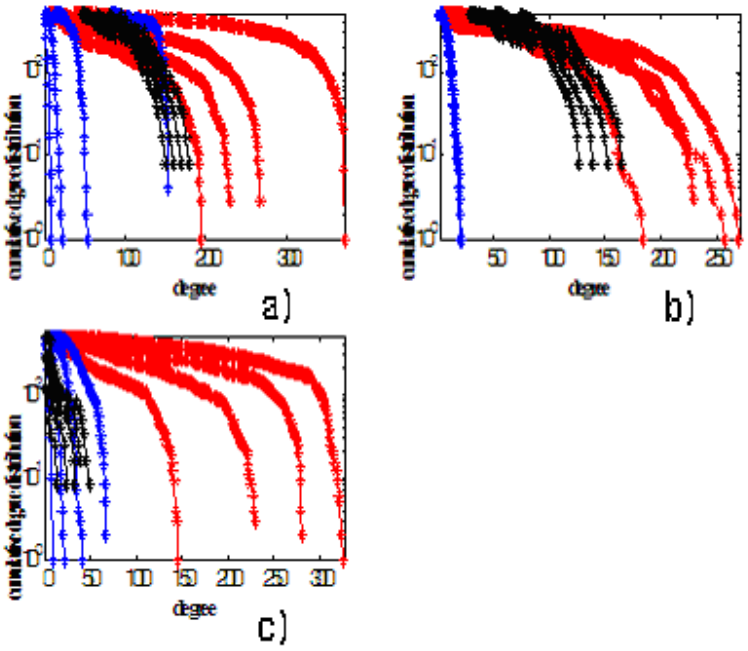
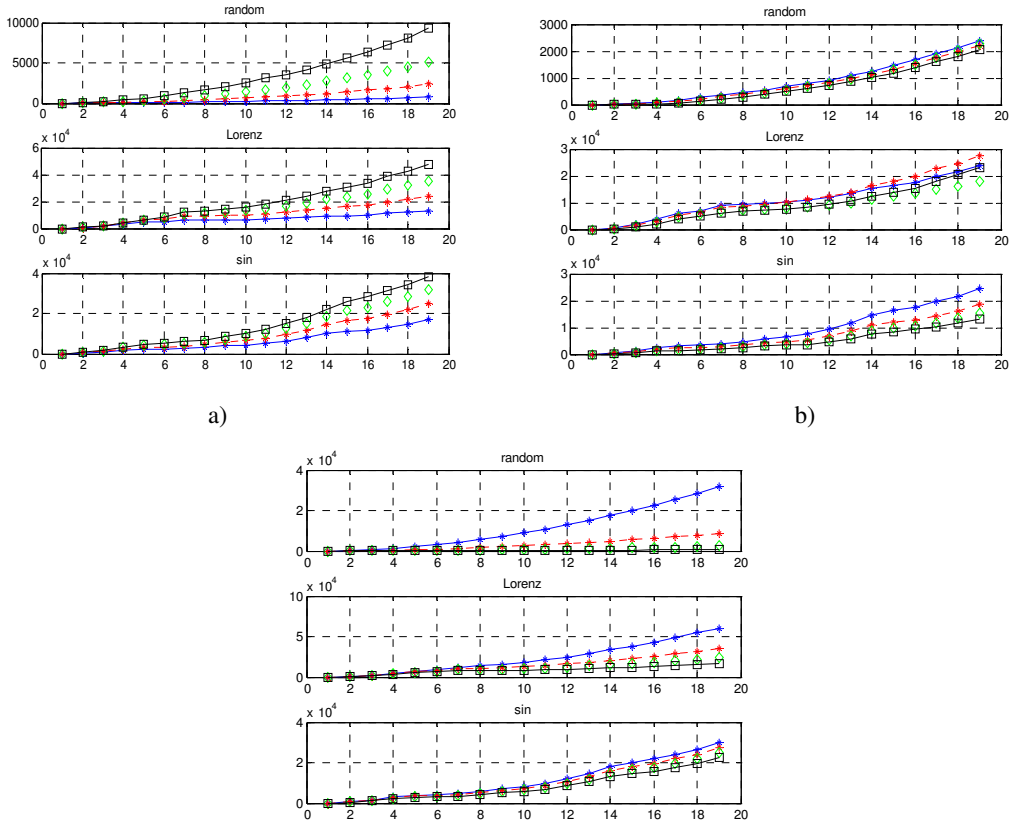


Figure 3.4. Cumulative degree distribution: sin (black), Lorenz (red), random (blue). a) $\varepsilon \in$ [2p, 3p, 4p, 5p], τ =3, d_E =3. b) $\tau \in$ [3,7,11,15], ε =3, d_E =3. c) $d_E \in$ [1,2,3,4], τ =3, ε =3p. N=500.

In Figure 3.5 we can see that the number of edges increases with ε and it decreases when either d_E or τ increase; moreover Lorenz network has more edges than Sin or Random. This property could be related to the folding property of Lorenz dynamics on the attractor for which a state comes near to the others many times, and this corresponds to a new connection in the adjacency matrix.

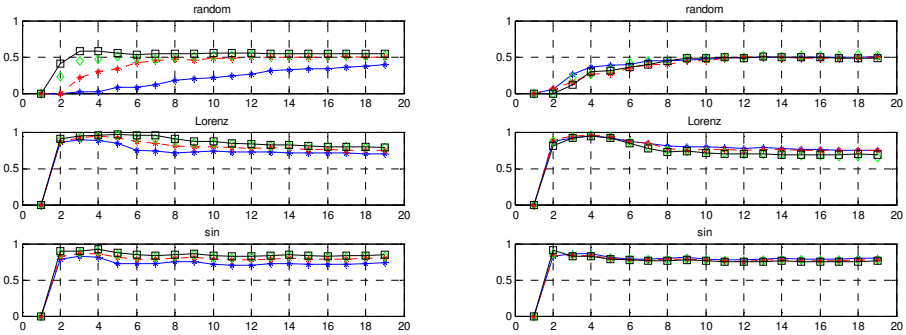


a) b)

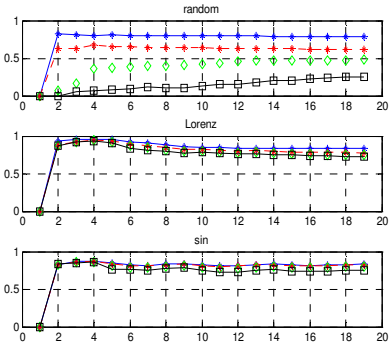
c)

Figure 3.5. Increase in the number of edges. a) $\varepsilon \in [2p, 3p, 4p, 5p]$, τ =3, d_E =3. b) $\tau \in$ [3,7,11,15], ε =3, d_E =3. c) $d_E \in$ [1,2,3,4], τ =3, ε =3p. The order of the graphics in respect to the parameter variation is blue, red, green , black.

In Figure 3.6 it is possible to observe that Lorenz networks have a mean clustering coefficient higher than Random and very similar to Sine networks and it converges to a value between 0.7 and 0.8. Since the mean clustering coefficient does not describe completely the internal structure of the networks, we have also measured the maximum and minimum clustering. The maximum clustering coefficient (not shown) is always one, whereas the minimum values are shown in Figure 3.7. Lorenz and Sine networks have a minimum clustering coefficient higher than Random and this correspond to the fact that more nodes have a higher clustering value and then the nodes connected with them are better connected the ones with the others.

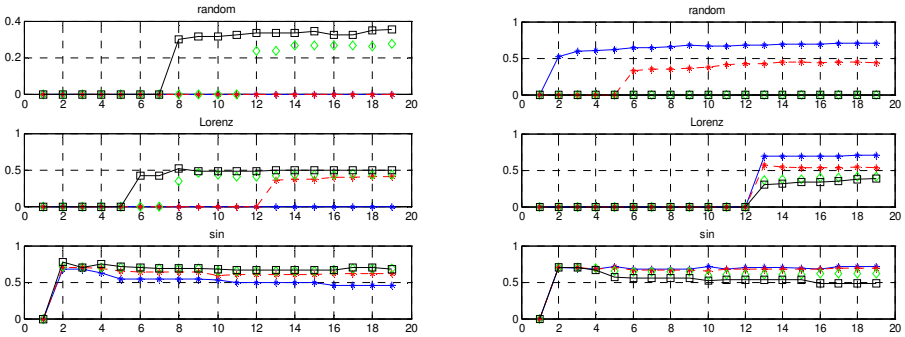


a) b)

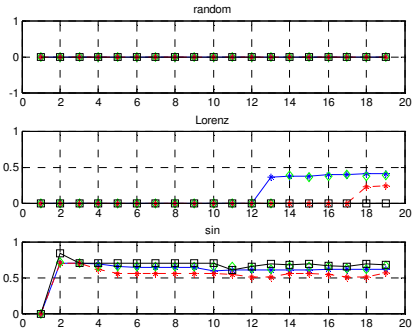


c)

Figure 3.6. Mean clustering: a) $\varepsilon \in [2p, 3p, 4p, 5p]$, τ=3, d_E =3; b) $\tau \in$ [3,7,11,15], ε =3, d_E=3; c) $d_E \in$ [1,2,3,4], τ=3, ε =3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.



a) b)



c)

Figure 3.7. Min clustering. A) $\varepsilon \in$ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) $\tau \in$ [3,7,11,15], ε =3, d_E =3. c) $d_E \in$ [1,2,3,4], τ=3, ε =3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.

The degree centrality measures how much the network structure is similar to a star structure, i.e., the most centralized one. In Figure 3.8 the degree centrality is represented. We can see that Lorenz converges to a value between 0.2 and 0.3, higher than random (0.1 or smaller) and it is more stable than Sine that shows oscillations in correspondence to the periodicity.

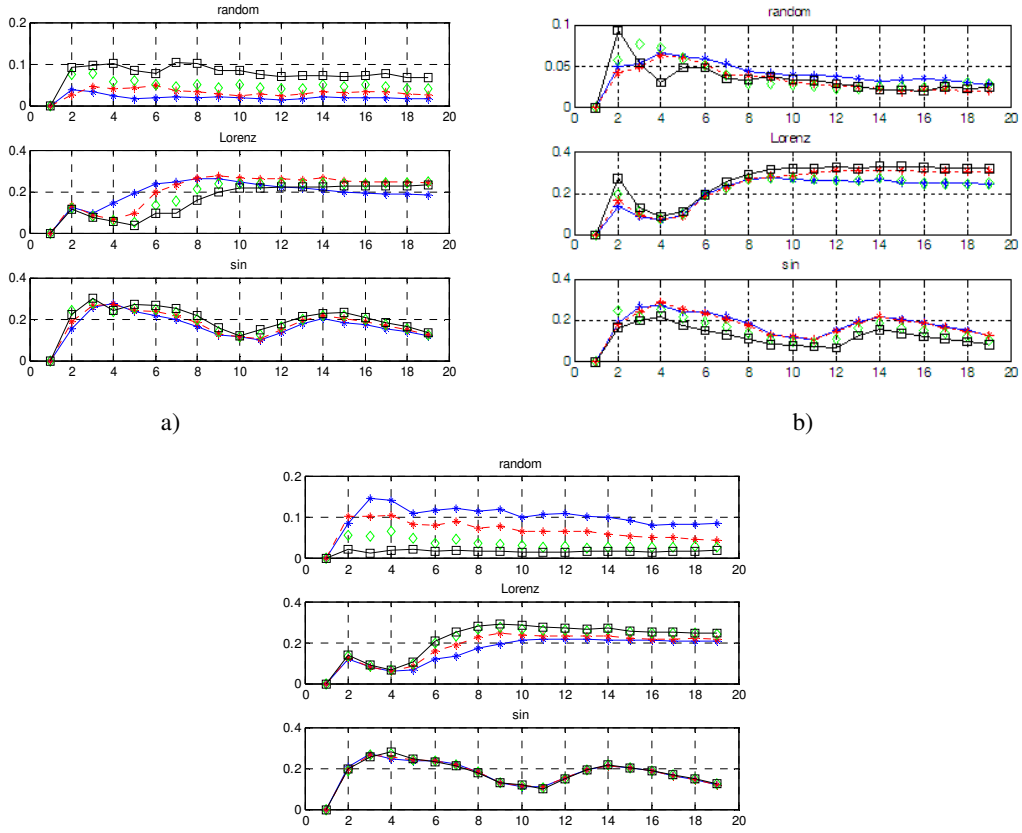
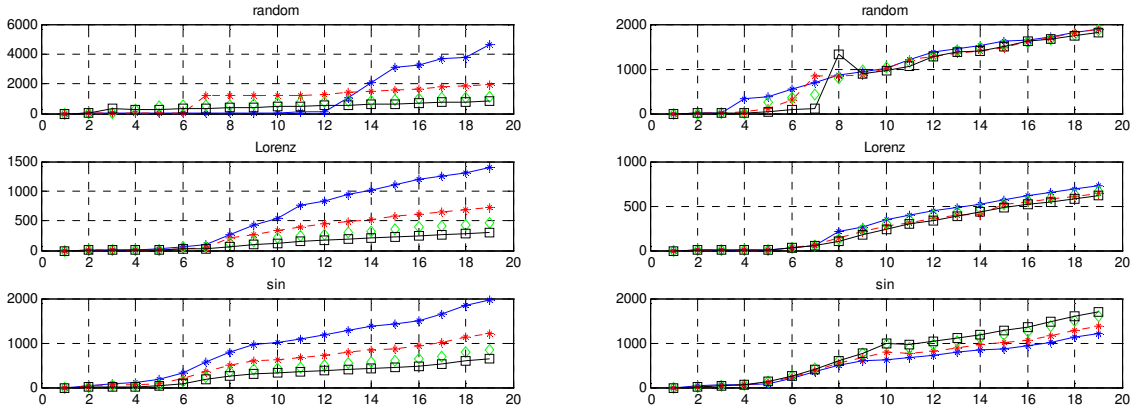
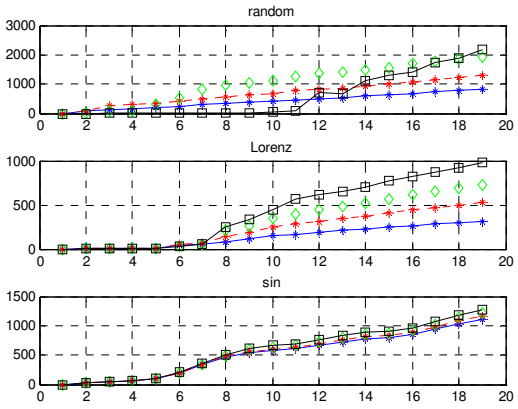


Figure 3.8. Degree Centrality. a) ε∈ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) τ∈ [3,7,11,15], ε=3, d_E =3. c) d_E ∈ [1,2,3,4], τ=3, ε=3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.

In Figure 3.9 it is possible to observe that the mean betweenness centrality in Lorenz networks converges to values lower than Random or Sine independently from the RP parameters. The mean value of betweenness is not very significant but since the min values (not shown) are always zero, it tells us that even the maximum betweenness of the Lorenz network is lower than the Sine or Random networks during their evolution: this can be interpreted as a sign of low vulnerability to fragmentation.



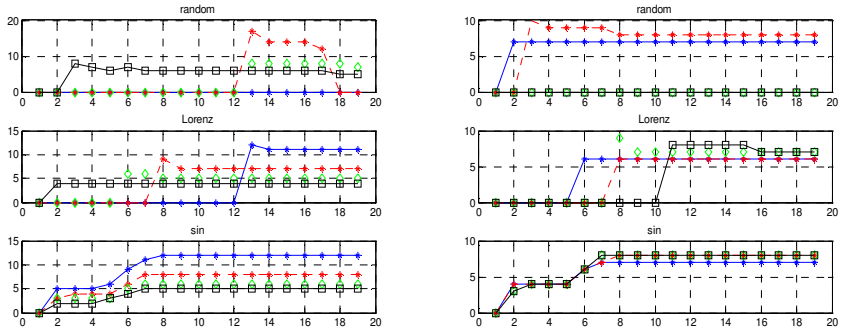
a) b)



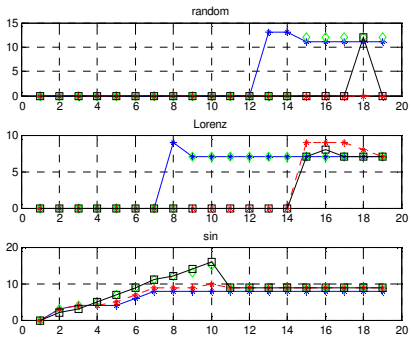
c)

Figure 3.9. Mean Betweenness centrality. a) $\varepsilon \in$ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) τ∈ [3,7,11,15], ε =3, d_E =3. c) $d_E \in$ [1,2,3,4], τ=3, ε =3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.

In Figures 3.10 and 3.11 the diameter and the mean distance between nodes are represented. The mean distance in Lorenz networks is generally smaller than for Sin or Random ones. In Fig 3.12 the maximum eigenvalue of Lorenz networks increases faster.

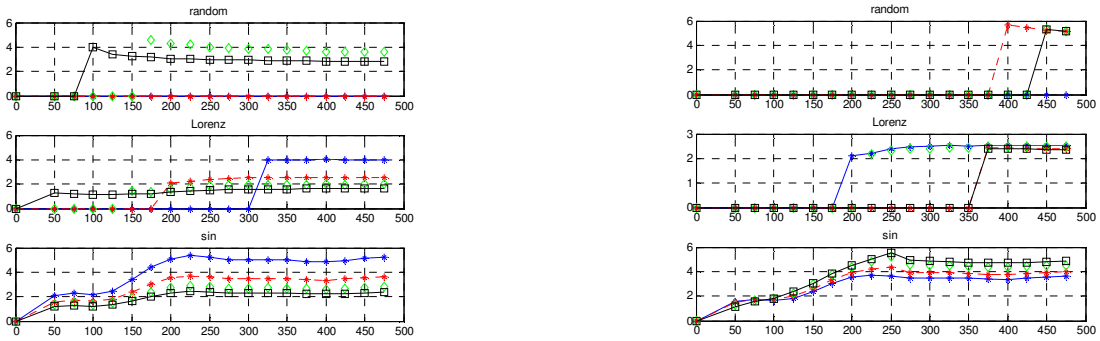


a) b)

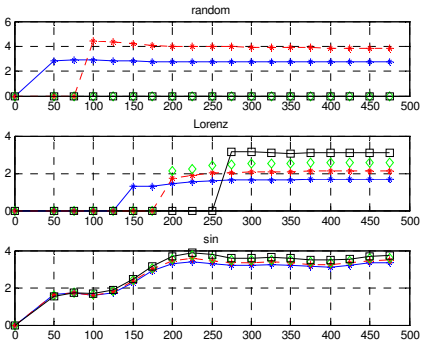


c)

Figure 3.10. Diameter. a) ε∈ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) τ∈ [3,7,11,15], ε=3, d_E =3. c) d_E ∈ [1,2,3,4], τ=3, ε=3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.



a) b)



c)

Figure 3.11. Mean Distance between nodes. a) ε∈ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) τ∈ [3,7,11,15], ε=3, d_E =3. c) d_E ∈ [1,2,3,4], τ=3, ε=3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green , black.

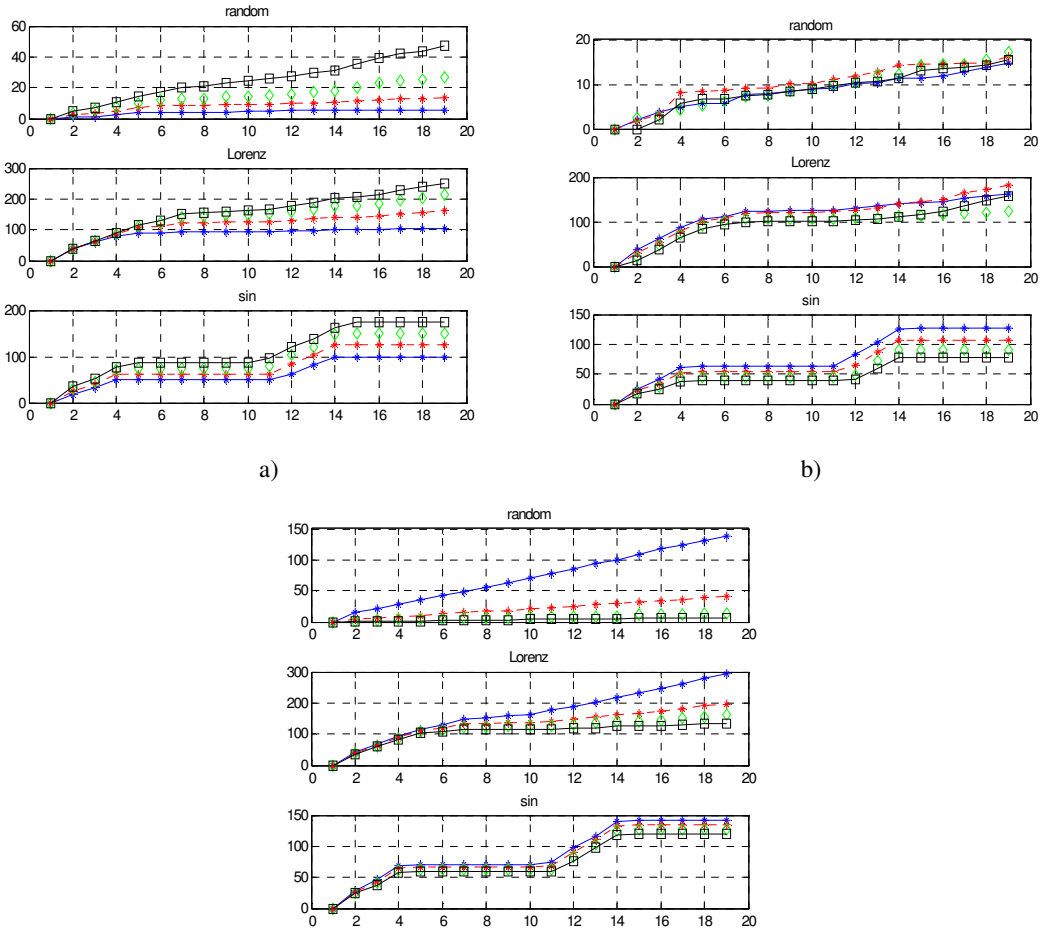


Figure 3.12. Max eigenvalue of adjacency matrix. a) ε∈ [2p, 3p, 4p, 5p], τ=3, d_E =3. b) τ∈ [3,7,11,15], ε=3, d_E =3. c) d_E ∈ [1,2,3,4], τ=3, ε=3p. The order of the graphics in respect to the parameter variation in the figures is blue, red, green, black.

In conclusion Lorenz networks have a faster edger number increase in comparison with Random networks and they have a higher degree centrality (similarity with star structure), but with lower mean betweenness centrality that it could make them less vulnerable (disconnectable) to selective attacks. Perhaps due to the higher edges number, the mean path is lower than in a Random network (or Sine network). The maximum eigenvalue is always higher in comparison with Sine and Random, therefore Lorenz networks seem to ensure more diffusion (stretching and folding properties ensure good mixing).

- Comparison between Lorenz, Sin, Random and Scale Free networks

To compare the evolution of Lorenz, Sin and Random networks with τ=3, d_E =3 and ε=3p. with the one of scale free (SF) networks, we have generated a set of SF networks, using Pajek (de Nooy et al., 2005) in order to have a strong component with approximately 200, 300, 400 and 500 nodes, a high probability of connections (0.9) and the maximum number of edges parameter (the same of Lorenz),

but higher in comparison with Sine and Random networks. The parameters used to generate the SF networks are listed in Table 3.1.

Table 3.1. Scale free parameters of the generated networks

Initial n° of nodes	Max n° of lines	Average degree of vertices	Initial E-R n° of nodes	Initial probability of lines	Resulting n° nodes of the strong component
210	9339	50	10	0.9	201
305	12105	75	10	0.9	299
410	17696	75	10	0.9	403
503	26410	100	10	0.9	499

The adjacency matrices of the strong components are plotted in Fig. 3.13.

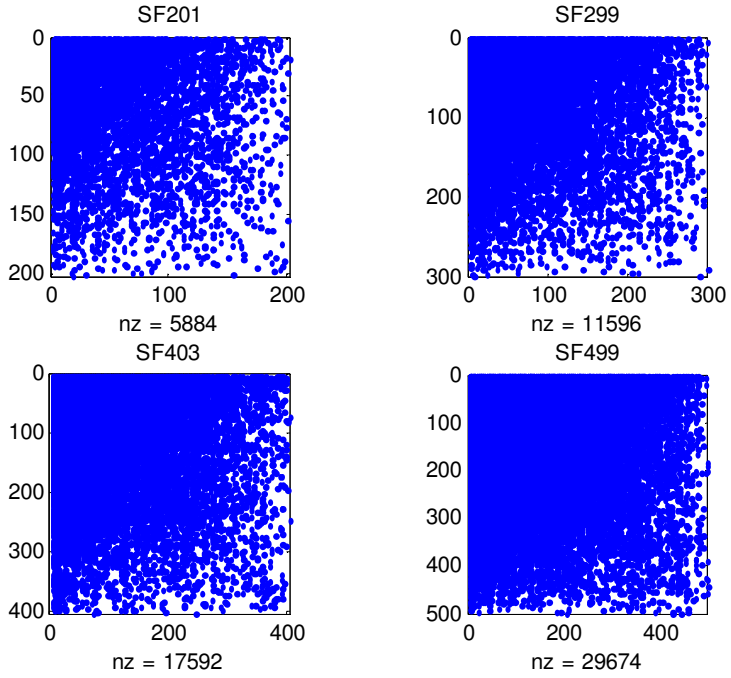


Figure 3.13. Adjacency matrices of the strong components of SF networks listed in Table 3.1.

In Figure 3.14, we have plotted some measurements of SF, Lorenz, Random and Sine networks when the dimension of the networks is increasing from nearly 200 until 500 nodes. Mean node degree and edges number of Lorenz and Sine network increases faster than for scale free networks, but they have a lower degree centrality. Moreover Lorenz has a higher max and mean betweenness centrality than scale free networks. The Lorenz diameter and mean distance are higher than Scale Free but smaller than Sine or Random ones.

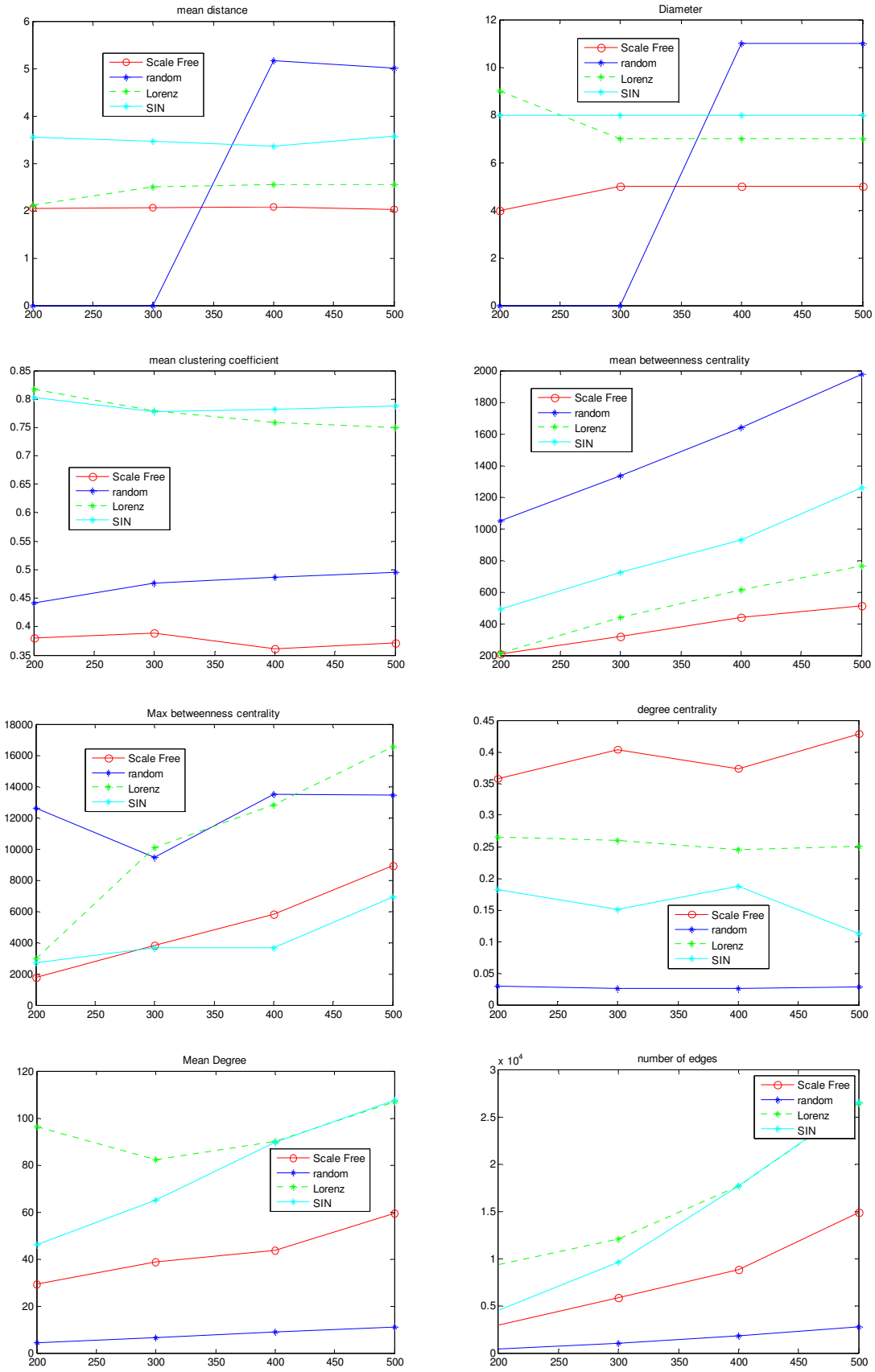
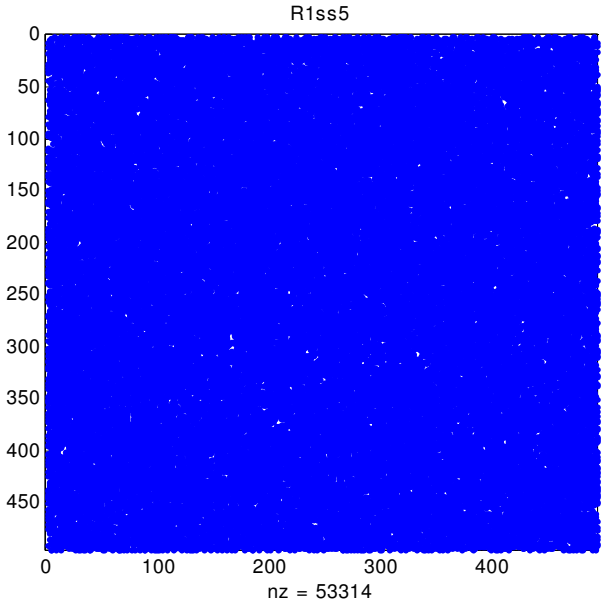


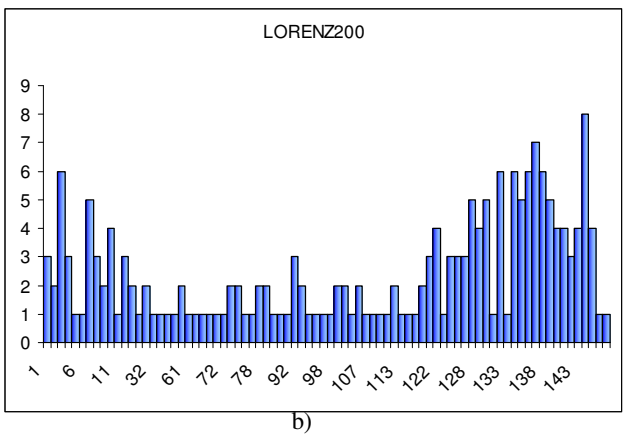
Figure 3.14. Comparison between Lorenz, Sin and Random networks with different number of nodes (200, 300, 400, 500) in respect to different network measures.

- Chaotic Networks as a paradigm for spreading information

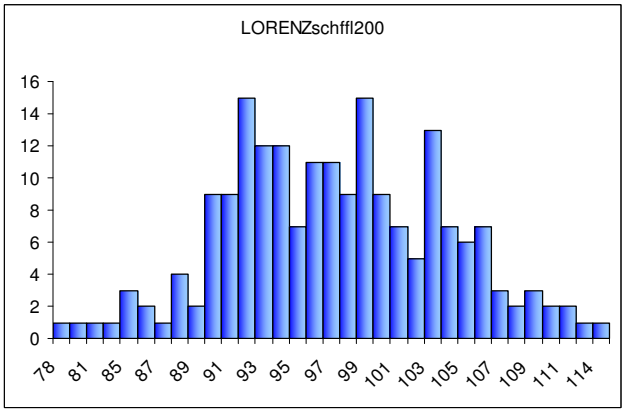
To understand if the differences of Lorenz network structures could be useful for some real applications, we have studied two types of network vulnerabilities: the first when the network nodes are attacked randomly and the second when they are subject to selective attack based on node degree. It seems that Lorenz networks are more robust in both cases, but this could be due to the huge number of connections of Lorenz adjacency matrix. To check if this is the reason we have randomly shuffled the connections to disrupt the structure of Lorenz adjacency matrix and we have obtained a nearly regular network (Fig 3.15a), in which all the nodes have the same degree. The degree distribution of the 200 node network Lorenz shuffled is plotted in Fig 3.15 together with the one of Lorenz 200 itself.



a) Lorenz Shuffled 500



b)



c)

Figure 3.15. a)Adjacency matrix of Lorenz shuffled with 500 of nodes; b) Node degree distribution of Lorenz and c) of Lorenz shuffled.

We have also checked the connectivity value for Lorenz, Shuffled Lorenz, Sine, Random and scale free networks, the results are shown in Figs 3.16-3.17.

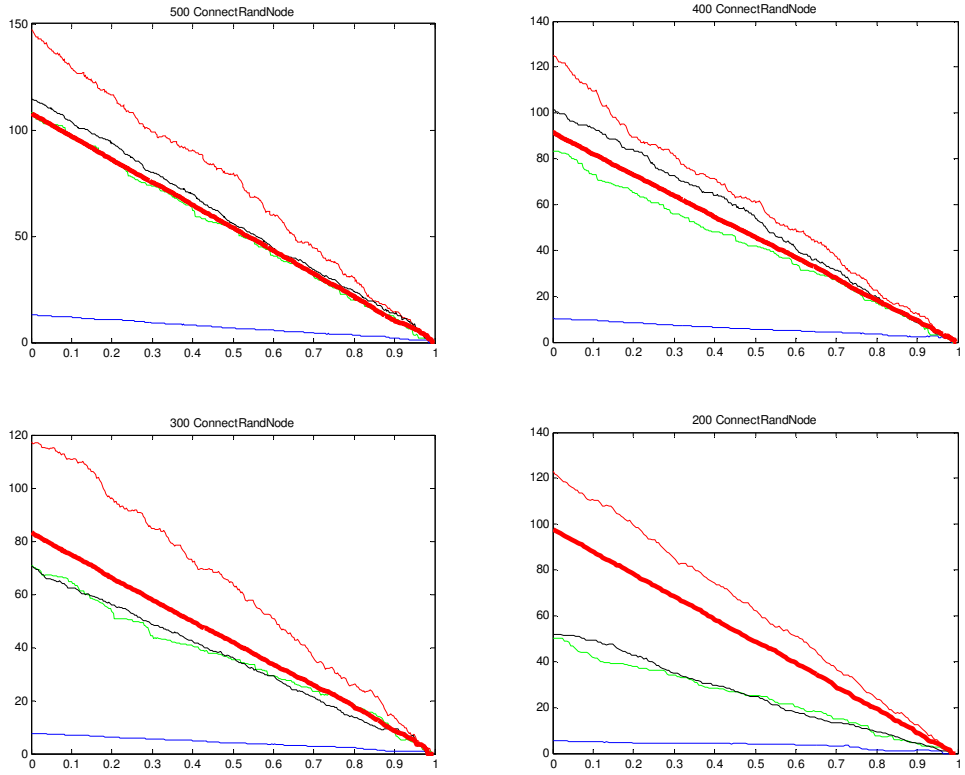


Figure 3.16..Connectivity of Lorenz Shuffled (red bold), Lorenz (red), Sin (black), Scale Free (green), Random (blue) networks subject to random attacks.

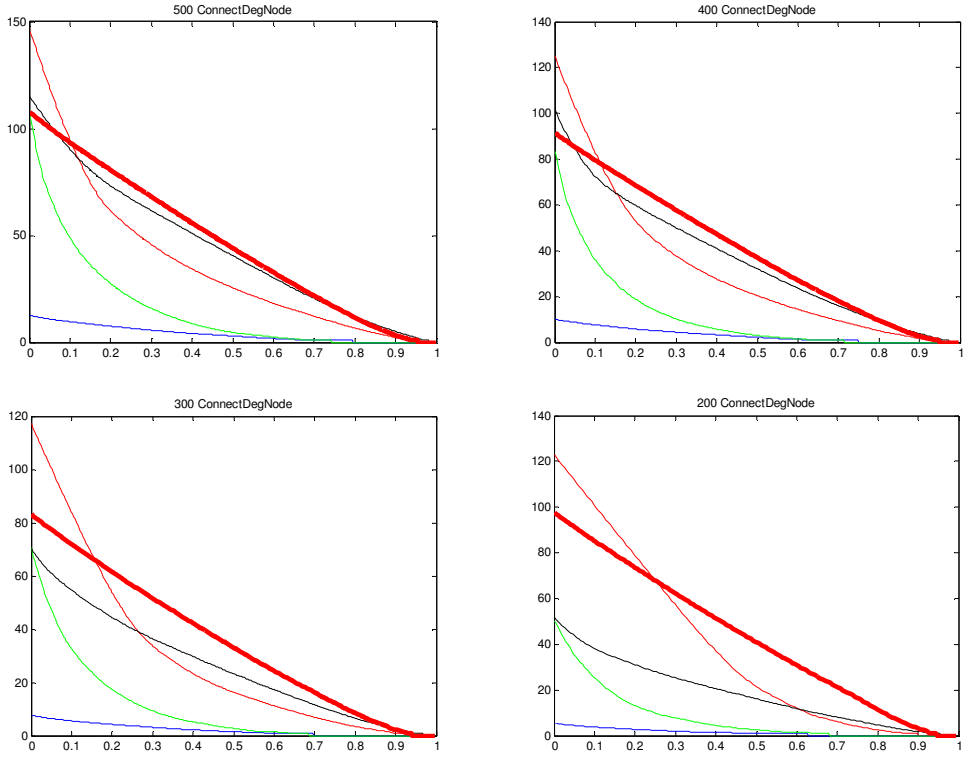


Figure 3.17. Connectivity of Lorenz Shuffled (red bold), Lorenz (red), Sin (black), Scale Free (green), Random (blue) networks subject to degree attacks.

Applying random attacks, the connectivity of Lorenz is always higher than that of the shuffled Lorenz or that of the other networks. In the case of selective attacks Lorenz has higher values of connectivity than Lorenz shuffled, but only if the attacks are not too strong (0.2%-0.3%); this limit increases if the number of network nodes decreases.

High value of connectivity means $\langle k^2 \rangle >> \langle k \rangle$ and then from Eq. (8) a lower probability r is required to transmit the idea/illness. In conclusion, it seems that a Lorenz structure allows a better diffusion in comparison to a shuffled Lorenz shuffled since the connectivity is higher and this is due to the structure of connection and not to the number of edges and nodes

3.2. FROM NETWORKS TO TIME SERIES

3.2.1. Simple networks

Let us analyze the case of simple periodic and random networks. Figure 3.18. shows the results applying the reconstruction methods described in Section 2.5, when a circular network is considered. As it can be observed the output is a periodic function as one may have expected.

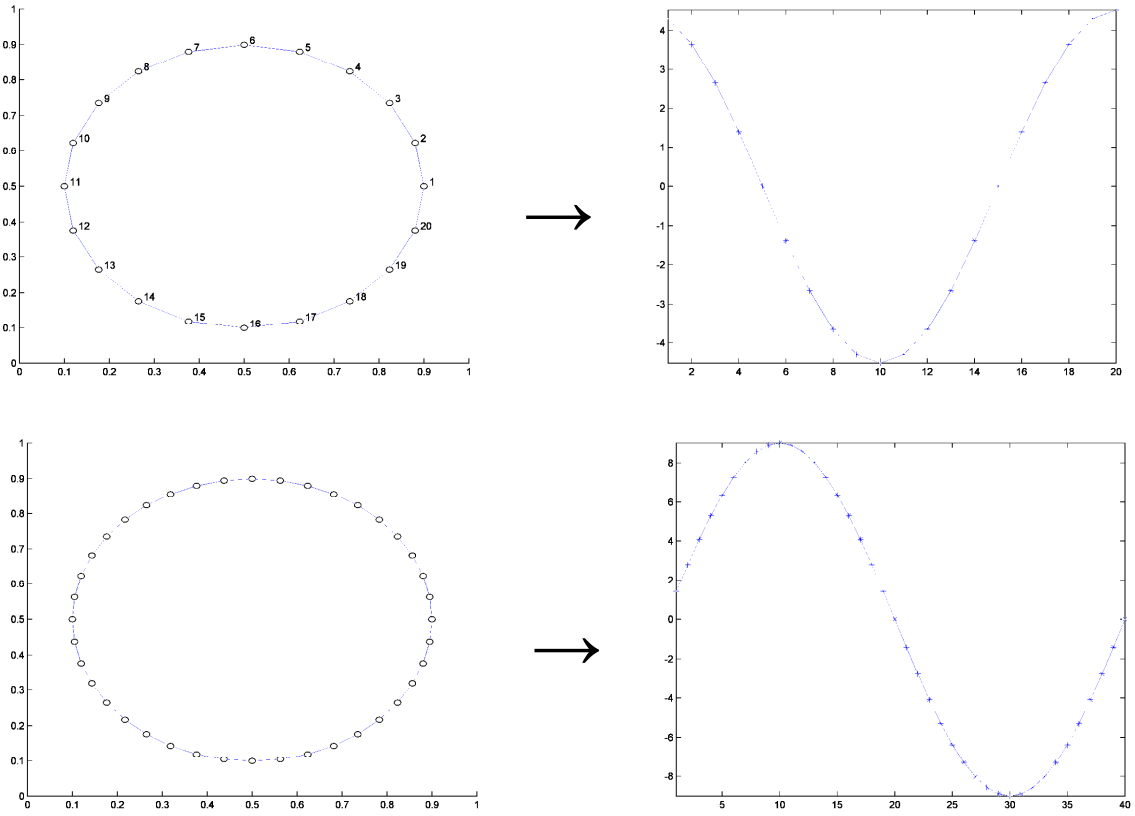


Figure 3.18. Results of the conversion of two circular networks of 20 and 40 nodes into time series.

In a similar way, the results of a random network are also a random time series. We have used the Pajek software (de Nooy et al., 2005) to generate random Erdos-Reny (ER) network with a giant component of 1000 nodes and then we have applied the two algorithms described in Section 2.5 (Method 1 and Method 2) to generate the corresponding time series (Figure 3.19). The application of nonlinear time series techniques show that there is no clear low dimensional deterministic component in both series and that they can be treated as stochastic ones.

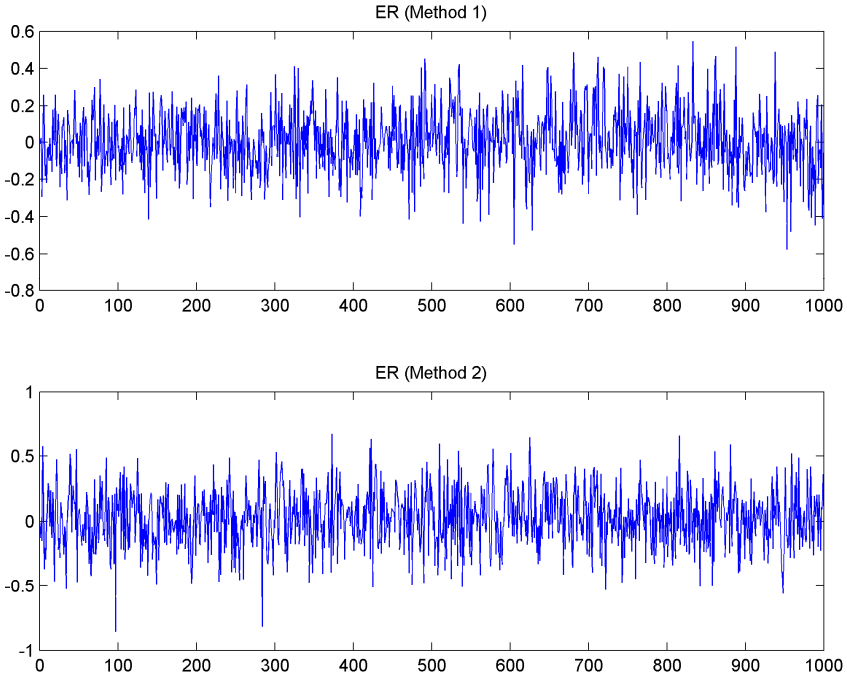


Figure 3.19. Time series generated from the ER network.

To compare both series, we have fitted the statistical distribution of their first differences, *y(t+1)-y(t)*, using the STABLE software (http://www.american.edu/cas/faculty/jpnolan.cfm). Stable distributions (Nolan, 1999) are a class of distributions that have the property of stability: if a number of independent and identically distributed random variables have a stable distribution, then a linear combination of these variables will have the same distribution, except for possibly different shift and scale parameters. They are described by four parameters, X~S(α, β, γ, δ; 0), the first two are $\alpha \in (0,2]$, an index of stability and $\beta \in [-1,1]$, a skewness parameter. α and β determine the shape of the distribution. The last parameters are a scale parameter, $\gamma \in [0,\infty)$, and a location parameter, $\delta \in (-\infty,\infty)$. Special cases of stable distributions include Gaussian (α = 2), Cauchy (α=1, β=0), Levy (α=1/2, β=1) and Dirac (α→0 or γ→ 0) distributions. In this case both distributions are Gaussian (α =2.0). Figure 3.20 shows the fitting results.

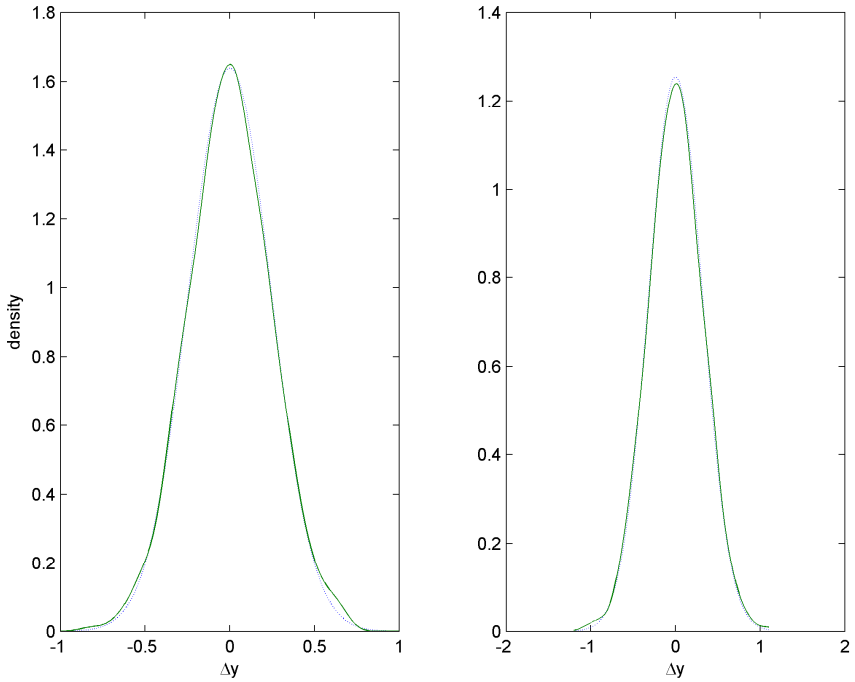


Figure 3.20. Fitted density plots of the first difference of both time series. Continuous line: Gaussian distribution, discontinuous: data.

Therefore, the time series generated from an ER network, independently of using Method 1 or Method 2, correspond to random time series with Gaussian probability distribution.

3.2.2. Scale free networks

We have used Pajek software (de Nooy et al., 2005) to generate one scale free (SF) network with a giant component of 1178 nodes. In this case, the ordering of the nodes follows the growth of the network. Then we have applied the two algorithms described in Section 2.5. Figures 3.21. shows the resulting time series. Also in this case, the application of nonlinear time series techniques shows that there is no clear low dimensional deterministic component in both series and that they can be treated as stochastic ones.

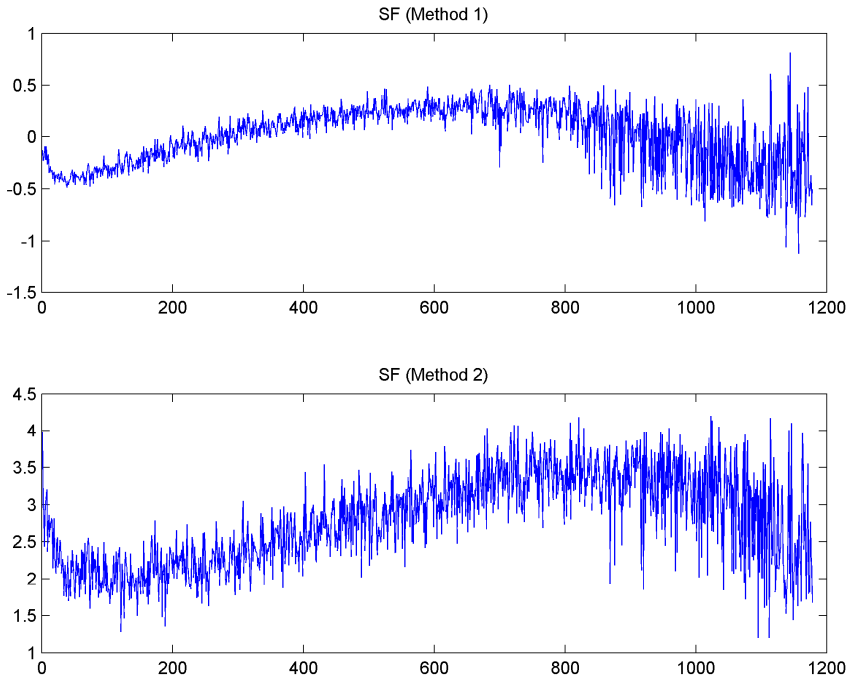


Figure 3.21. Time series generated from the SF network.

We have carried out a similar analysis to the one described in Section 3.2.1. The obtained parameters are summarized in Table 3.2, whereas fig. 3.22 shows the fitting results.

Table 3.2. Scale Free (SF) time series (*Method1* and *Method2*) fitted parameters using STABLE (Nolan, 1999).

Data set	**α**	**β**	**γ x10^1**	**δ x10^2**
SF1	1.22	-0.039	0.99	0.23
SF2	1.70	0.026	3.22	-0.42

As it can be observed, and contrary to the ER time series, that both methods produce different parameters in terms of the fitting of the probability distribution function; even though, in both cases heavy tail distributions are obtained the values differ.

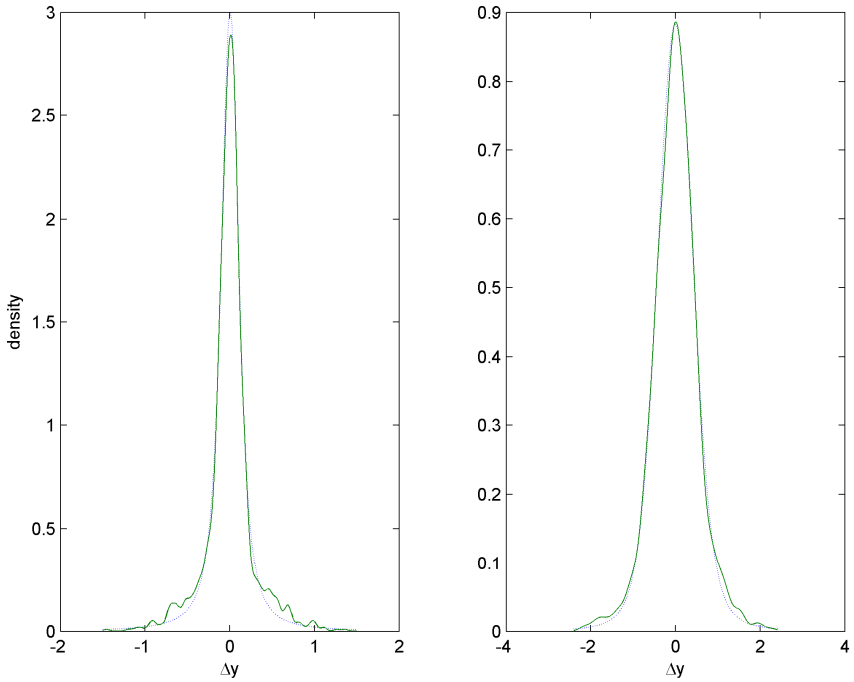


Figure 3.22. Fitted density plots of the first difference of the time series in Fig. 3.21. Continuous line: stable distribution, discontinuous: data.

4. APPLICATION TO METABOLIC NETWORKS

In metabolic networks the nodes correspond to metabolites, whereas the connections between these nodes correspond to reactions. In this work, we have analyzed the metabolic networks corresponding to 43 different organisms provided by Jeong et al. (2000) based on data deposited in the WIT database (http://igweb.integratedgenomics.com/IGwit/).

4.1. APPLICATION OF RQA ANALYSIS

By considering the network adjacency matrix as the recurrence plot (RP) of some time series, it is possible to calculate the properties of the metabolic networks analyzed by Jeong et al. (2000) using the RQA parameters. Table 4.1 summarizes the results obtained, whereas Fig. 4.1 shows the scaling behaviour of *RR*, *Trend* and *LAM* ads a function of the number of substrates.

Table 4.1. RQA parameters of the metabolic networks from Jeong et al. (2000). The color code corresponds to Archae, Bacterium and Eukaryote (pink, green and blue), respectively. N° nodes is the number of substrate whereas *RR*, *DET*, L_{max}, *ENTR*, *Trend*, *LAM* and *TT* correspond to several RQA parameters (Section 2.3): percentage of recurrence and determinism, maximum line, entropy, trend, percentage of laminarity and trapping time, respectively.

N	Code	Name	N° nodes	RR	DET	L_{max}	ENTR	Trend	LAM	TT
1	AP	*A. pernix*	204	0.467	8.348	4	0.26	3.16	17.57	2.02
2	AG	*A. fulgidus*	496	0.182	10.445	5	0.38	0.506	20.28	2.23
3	TH	*M. thermoautotro picum*	430	0.211	9.992	4	0.48	0.644	20.21	2.223
4	MJ	*M. jannaschii*	424	0.214	10.063	4	0.5	0.681	20.6	2.202
5	PF	*P. furiosus*	316	0.3	10.842	4	0.53	1.4	18.57	2.3
6	PH	*P. horikoshii*	323	0.293	11.905	4	0.47	1.31	17.91	2.324
7	AA	*A. aeolicus*	419	0.202	9.367	4	0.35	0.639	19.62	2.112
8	CQ	*C. pneumoniae*	194	0.41	9.463	3	0.67	3.694	22.51	2.146
9	CT	*C. trachomatis*	215	0.379	11.688	4	0.89	3.002	21.43	2.106
10	CY	*Synechocystis sp.*	546	0.157	11.054	5	0.43	0.354	21.99	2.207
11	PG	*P. gingivalis*	424	0.203	9.948	4	0.26	0.686	20.93	2.241
12	MB	*M. bovis*	429	0.204	9.5	4	0.45	0.717	19.49	2.183
13	ML	*M. leprae*	422	0.205	9.727	4	0.34	0.673	20.26	2.154
14	MT	*M. tuberculosis*	587	0.147	10.642	4	0.3	0.335	21.28	2.282
15	BS	*B. subtilis*	785	0.105	10.653	4	0.25	0.149	22.18	2.338
16	EF	*E. faecalis*	386	0.217	10.837	4	0.2	0.726	20.53	2.066
17	CA	*C. acetobutylicum*	494	0.172	9.632	4	0.25	0.451	22.05	2.231
18	MG	*M. genitalium*	209	0.436	8.381	2	0	2.885	28.38	3.311
19	MP	*M. pneumoniae*	178	0.512	11.159	2	0	3.407	22.32	2.476
20	PN	*S. pneumoniae*	416	0.203	9.707	4	0.12	0.662	19.95	2.106
21	ST	*S. pyogenes*	403	0.209	10.337	4	0.12	0.676	19.73	2.136
22	CL	*C. tepidum*	389	0.229	8.38	4	0.32	0.883	19.87	2.221
23	RC	*R. capsulatus*	670	0.126	9.095	4	0.29	0.235	22.1	2.223
24	RP	*R, prowazekii*	214	0.394	12.302	3	0.35	2.764	24.01	2.086
25	NG	*N. gonorrhoeae*	406	0.206	9.843	4	0.33	0.636	20.71	2.087
26	NM	*N. meningitidis*	381	0.221	9.738	4	0.26	0.725	20.58	2.077
27	CJ	*C. jejuni*	380	0.225	10.314	4	0.26	0.803	19.46	2.127
28	HP	*H. pylori*	375	0.23	10.14	4	0.35	0.765	20.19	2.221
29	EC	*E. coli*	778	0.103	10.423	4	0.36	0.137	22.95	2.27

30	TY	*S. typhi*	819	0.099	10.064	4	0.32	0.133	23.15	2.331
31	YP	*Y. pestis*	568	0.148	9.504	4	0.41	0.336	23.27	2.229
32	AB	*A. actinomycetem comitans*	395	0.213	10.377	4	0.42	0.754	20.07	2.127
33	HI	*H. influenzae*	526	0.155	10.309	4	0.43	0.337	21.65	2.237
34	PA	*P. aeruginosa*	734	0.115	9.758	4	0.4	0.187	23.44	2.23
35	TP	*T. pallidum*	207	0.447	12.613	2	0	2.928	19.82	2.115
36	BB	*B. burgdorfei*	187	0.486	11.872	2	0	4.106	19.86	2.122
37	TM	*T. maritima*	338	0.255	9.734	4	0.3	0.973	20.29	2.129
38	DR	*D. radiourans*	815	0.101	9.89	4	0.3	0.14	22.8	2.447
39	EN	*E. nidulans*	383	0.236	10.638	5	0.49	0.836	22.11	2.173
40	SC	*S. cerevisiae*	561	0.156	11.374	5	0.46	0.309	20.64	2.166
41	CE	*C. elegans*	462	0.19	11.989	5	0.61	0.487	20.45	2.214
42	OS	*O. sativa*	292	0.333	10.519	5	0.67	1.604	21.17	2.239
43	AT	*A. thaliana*	302	0.322	10.266	5	0.66	1.51	19.9	2.211

A preliminary analysis shows that *RR* and *Trend* follow a power law ($RR \propto N^{-\gamma}$; $Trend \propto N^{-\gamma}$) as a function of the number of substrates as well as of the number of links, number of individual reactions or temporary substrate-enzyme complexes and enzymes -values of *L*, *R* and *E* in Jeong et al. (2000), Table 1 supplementary material-. Table 4.2 summarizes the coefficients and Fig. 4.1 shows the fitting results for the case of *RR* and *Trend* as a function of the number of substrates (*N° nodes*, *N*).

Table 4.2. Exponential coefficients (γ) of the power low data fitting.

	N° nodes	L_{in}	L_{out}	***R***	***E***
RR	1.032	0.765	0.767	0.782	0.717
Trend	2.078	1.542	1.604	1.604	1.351

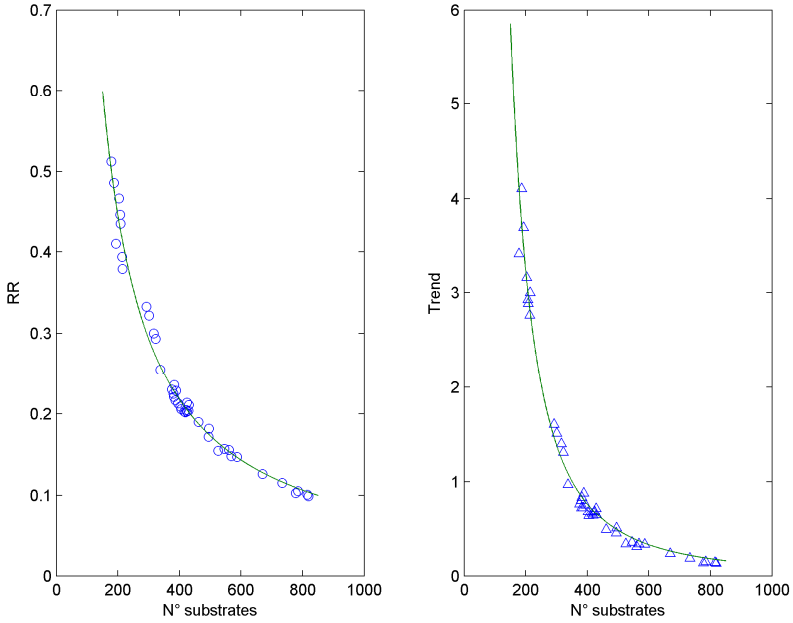


Figure 4.1. Power scaling of *RR* and *Trend* as a function of the number of substrates (see Table 4.2 for the coefficients).

4.2. METABOLIC TIME SERIES ANALYSIS

The application of the algorithms described in Section 2.5 to the giant component of the 43 metabolic networks produces the corresponding time series. Figure 4.2 shows the resulting time series for four examples.

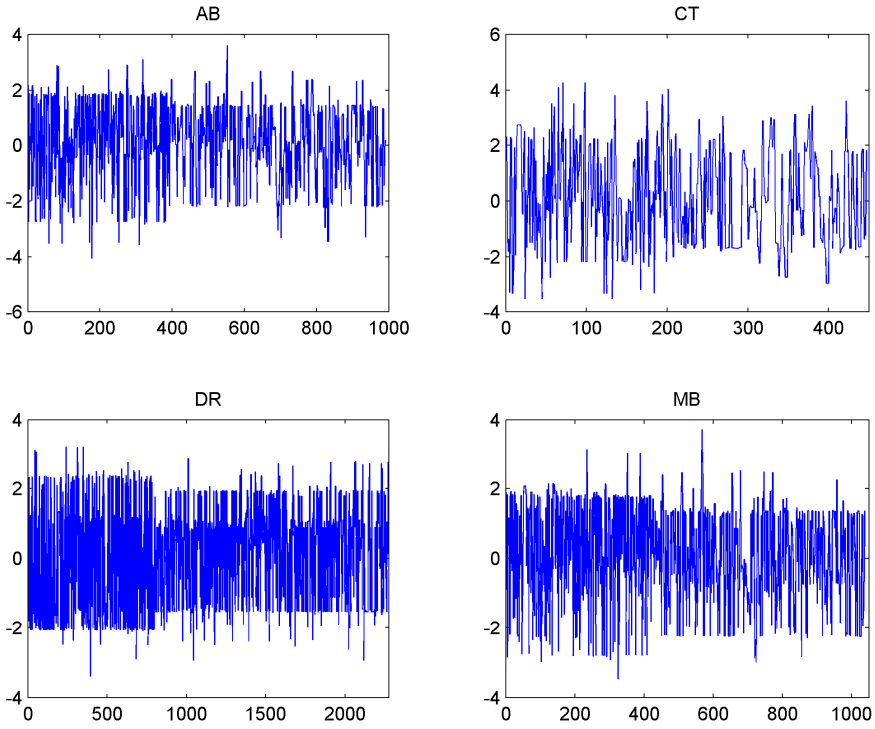


Figure 4.2. Examples of time series obtained from metabolic networks (Method 1).

The non-linear time series analysis for the metabolic series show that there is no low dimension non-linear structure, i.e. the series can be considered as stochastic time series. For this reason, we have carried out a rescaled range analysis (R/S). The results for the generated time series are shown in Table 4.2. As it can be seen the series are not random noise ($H \approx 0.5$), but they have a fractal structure. In addition, all time series show strong antipersistence, i.e. $H<0.5$.

However, there is arbitrariness in the selection of the initial node or initial time and in the way the nodes are visited -once transformed in time series. To check the consistency of the results in view of the selection of the initial node –time series point-, we have re-ordered the metabolic network as a function of the number of connections and carried out the same approach. Figure 4.3 shows the generated time series using the same examples as in Fig. 4.2. In addition, we have also calculated the Hurst exponent, see Table 4.3. Even though the antipersistence ($H<0.5$) in the new time series is maintained, the results are different without any clear correlation between both Hurst exponents sets. Therefore, we can conclude that this value is not an invariant against network re-ordering, but depends

on the numbering of the network which implies that the fractal structure is somehow changed with this transformation.

Table 4.2. Hurst exponent of the corresponding metabolic time series.

Code	Name	Nodes GC	H	Code	Name	Nodes GC	H
AP	*A. pernix*	490	0.0701	RC	*R. capsulatus*	1760	0.2688
AG	*A. fulgidus*	1268	0.1990	RP	*R, prowazekii*	451	0.1416
TH	*M. thermoautotro picum*	1112	0.2868	NG	*N. gonorrhoeae*	1042	0.2657
MJ	*M. jannaschii*	1082	0.3490	NM	*N. meningitidis*	970	0.4120
PF	*P. furiosus*	746	0.1885	CJ	*C. jejuni*	939	0.3192
PH	*P. horikoshii*	764	0.2721	HP	*H. pylori*	940	0.2665
AA	*A. aeolicus*	1052	0.1614	EC	*E. coli*	2268	0.2300
CQ	*C. pneumoniae*	387	0.0830	TY	*S. typhi*	2354	0.2880
CT	*C. trachomatis*	447	0.1940	YP	*Y. pestis*	1444	0.2988
CY	*Synechocystis sp.*	1427	0.3649	AB	*A. actinomycetem comitans*	991	0.4024
PG	*P. gingivalis*	1001	0.2534	HI	*H. influenzae*	1413	0.2650
MB	*M. bovis*	1038	0.3991	PA	*P. aeruginosa*	1958	0.2625
ML	*M. leprae*	1050	0.2500	TP	*T. pallidum*	467	0.3288
MT	*M. tuberculosis*	1515	0.3434	BB	*B. burgdorfei*	392	0.0212
BS	*B. subtilis*	2207	0.2350	TM	*T. maritima*	821	0.3163
EF	*E. faecalis*	1000	0.3622	DR	*D. radiourans*	2273	0.2486
CA	*C. acetobutylicum*	1294	0.4770	EN	*E. nidulans*	911	0.1833
MG	*M. genitalium*	467	0.0521	SC	*S. cerevisiae*	1511	0.3136
MP	*M. pneumoniae*	409	0.2450	CE	*C. elegans*	1173	0.2020
PN	*S. pneumoniae*	1069	0.2584	OS	*O. sativa*	658	0.3496
ST	*S. pyogenes*	1043	0.3957	AT	*A. thaliana*	689	0.2564
CL	*C. tepidum*	913	0.3339				

In this case, contrary to the visibility graph proposed by Lacasa et al. (2008), there is no functional relation between the Hurst exponent and the degree distribution of the graph since this last value does not change by the re-ordering on the nodes in the network.

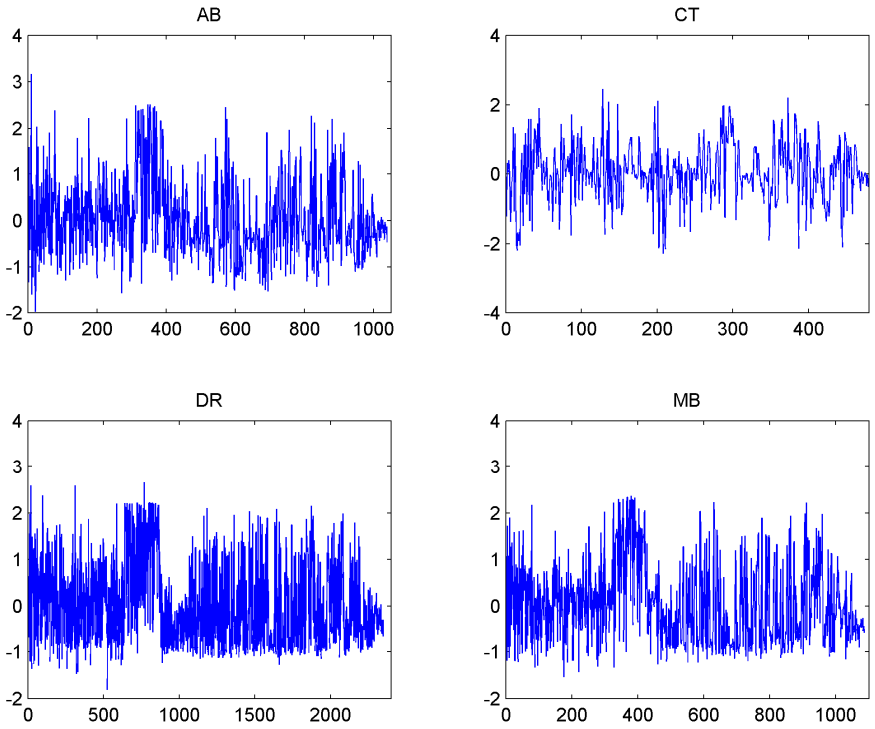


Figure 4.3. Examples of time series obtained from metabolic networks after re-ordering by number of connections.

Table 4.3. Hurst exponent of the corresponding metabolic time series after re-ordering by number of connections.

Code	Name	Nodes GC	H	Code	Name	Nodes GC	H
AP	*A. pernix*	490	0.3223	RC	*R. capsulatus*	1760	0.2411
AG	*A. fulgidus*	1268	0.1673	RP	*R, prowazekii*	451	0.1191
TH	*M. thermoautotro picum*	1112	0.1344	NG	*N. gonorrhoeae*	1042	0.2766
MJ	*M. jannaschii*	1082	0.1484	NM	*N. meningitidis*	970	0.2857
PF	*P. furiosus*	746	0.1930	CJ	*C. jejuni*	939	0.2788
PH	*P. horikoshii*	764	0.1855	HP	*H. pylori*	940	0.2154
AA	*A. aeolicus*	1052	0.3265	EC	*E. coli*	2268	0.1027
CQ	*C. pneumoniae*	387	0.1931	TY	*S. typhi*	2354	0.1374
CT	*C. trachomatis*	447	0.4230	YP	*Y. pestis*	1444	0.3050
CY	*Synechocystis sp.*	1427	0.1641	AB	*A. actinomycetem comitans*	991	0.2226
PG	*P. gingivalis*	1001	0.3893	HI	*H. influenzae*	1413	0.0979
MB	*M. bovis*	1038	0.3418	PA	*P. aeruginosa*	1958	0.2544
ML	*M. leprae*	1050	0.1976	TP	*T. pallidum*	467	0.0743
MT	*M. tuberculosis*	1515	0.2666	BB	*B. burgdorfei*	392	0.0429
BS	*B. subtilis*	2207	0.1957	TM	*T. maritima*	821	0.3841
EF	*E. faecalis*	1000	0.2878	DR	*D. radiourans*	2273	0.1942
CA	*C. acetobutylicum*	1294	0.1613	EN	*E. nidulans*	911	0.1329
MG	*M. genitalium*	467	0.2080	SC	*S. cerevisiae*	1511	0.1040
MP	*M. pneumoniae*	409	0.1298	CE	*C. elegans*	1173	0.2293
PN	*S. pneumoniae*	1069	0.2275	OS	*O. sativa*	658	0.4552
ST	*S. pyogenes*	1043	0.2924	AT	*A. thaliana*	689	0.1058
CL	*C. tepidum*	913	0.4530				

5. CONCLUSIONS

In this work we present a simple and fast approach to generate network structures based on time series recurrence plots and *viceversa*. In addition, we discuss the application of the different analysis techniques developed in both fields, i.e. complex networks and time series analysis.

In the last decade a considerable number of models of network growth have been developed with the main objective of explaining the properties of real world networks (Newman, 2003). These models provide a set of rules on how the addition of vertices and edges should occur in a network. A characteristic of these models is their stochastic nature in the sense that there is a certain probability of the appearance of new vertices or new attachments, e.g. the preferential attachment growth model proposed by Barabasi and Albert (1999) to explain the WWW growth.

Concerning the transformation from time series to networks, here we propose a deterministic growth procedure which produces new types of complex network structures that have some interesting features. This simple and fast approach is able to generate deterministic network structures based on time series recurrence plots. The generated networks contain several properties of the original time series. In this case, networks generated from chaotic attractors display interesting features from the point of view of robustness which could help in designing systems with high tolerance against errors and transfer of information. Chaotic networks based on the Lorenz attractor show that they are highly tolerant against attacks and they have a high ability for the transfer of information or, on the contrary, are able to transmit infections faster. It is still necessary to investigate if such chaotic networks already exist in natural or man-made systems or, if possible, to construct such networks and test their properties.

On the other hand, the transformation from networks to time series presents some problems concerning the selection of the initial time or in our case the initial node and the way in which the nodes are visited. If a network has been generated following a certain growth law it seems logical to choose the first node as the origin and then proceed following the network growth pattern. However, the situation is not so clear, for example, with metabolic networks, where it is difficult to select which is the first metabolite. Similar concerns would apply to other types of biological networks. In this case several alternatives could be considered, e.g. ordering using the number of connections. However, we have still to find if there are some invariant/preserved properties in the generated time series from the same network. We have found that rescaled range analysis does not preserve the fractal structure in the time series. In any case, if time series parameters would be invariant against the initial node selection, then they could be used to analyze the networks that have generated said time series. Our future work will continue along these lines concerning the generation of time series from biological networks.

6. REFERENCES

Abarbanel, H.D.I., *Analysis of Observed Chaotic Data*, 1996, Springer-Verlag, New-York.

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

Barabási, A. L. and Albert R., 1999. Energence of scaling in random networks. Science 286, 509-512.

Boccaletti, S., Latora, V., Moreno, Y., Chavez , M., Hwanga D.-U., 2006. Complex networks: Structure and dynamics. Physics Reports 424, 175-308.

Borgnat, P., Fleury, E., Guillaume, J.-L., Magnien, C., Robardet, C., Scherrer, A., 2008. Evolving Networks," in Proceedings of *NATO ASI 'Mining Massive Data Sets for Security'*, Françoise Fogelman-Soulié and Domenico Perrotta and Jakub Piskorski and Ralf Steinberg editors, *NATO Science for Peace and Security Series D: Information and Communication Security*, vol. 19, pp. 198-204, IOS Press.

Cannon, M. J. Percival, D. B., Caccia, D. C., Raymond, G. M. and Bassingthwaighte, J. B., 1997, Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A* **241**, 606-626.

Clauset, A., Moore, C., Newman, M.E.J., 2008. Hierarchical structure and the prediction of missing links in networks. Nature 453, 98-101.

Cohen, J. E. and Newman, C. M., 1985b. A Stochastic Theory of Community Foodwebs: I. Models and Aggregated Data. Proceedings of the Royal Society of London (Editor), Series B, Biological Sciences, 224 (1237), 421-448.

De Ruiter, P. C., Wolters, V. and Moore, J. C. (Eds.), 2005a. Dynamic foodwebs: Multispecies assemblages, ecosystem development and environmental change. Academic Press, Amsterdam, 608 pp.

Dijkstra, E.,W., Numer. Math. 1, 269 (1959)

Diks, C., 1999, *Nonlinear Time Series Analysis: Methods and applications*.World Scientific, Singapore.

Drews, J. 2000 Drug discovery: A historical perspective. Science 297, 1960-1964.

Dunne, J.A., 2006. The network structure of foodwebs. Pages 27-86 in Ecological Networks: Linking Structure to Dynamics in Foodwebs. M. Pascual and J.A. Dunne (Editors), Oxford University Press, UK.

Eckmann, J.P., Feinerman, O., Gruendlinger, L., Moses, E., Soriano, J., Tlusty, T. 2007. The physics of living neural networks. Phys. Rep. 449, 54-76.

Eckmann, J.P., Kamphorst J.O., Ruelle, D., 1987. Recurrence plots of dynamical systems, Europhys. Lett. 4, 973-977.

Erdos, P. Reny, A., 1960. On the evolution of random graphs. Publications of the mathematical

Institute of the Hungarian Academy of Sciences 5, 17-61.

Farkas, I.J., Jeong, H., Vicsek, T., Barabasi, A.L., Oltvai, Z.N., 2003. The topology of the transcription regulatory network in yeast, *Saccharomyces cerevisiae*. Physica A 381, 601-612.

Fell, D.A., Wagner, A. 2000. The small world of metabolism. Nature Biotechnology 18, 1121-1122.

Fraser, A. and Swinney, H., 1986, Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134-1140.

Goh, K.I., Salvi, G., Kahng, B., and Kim D., 2006. Skeleton and fractal scaling in complex networks. Physical Review E 96, 1-4.

Grassberger, P., 1983, Generalized dimension of strange attractors. Phys. Lett A 97, 227-230.

Grower, J.C., Biometrika 53, 325 (1966)

Guelzim, N., Bottani, S., Bourgine, P., Kepes, F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. Nature Genetics 31, 60-63.

Gulmerà, R and Nunes Amaral, L.A.2005. Functional cartography of complex metabolic networks. Nature 433, 895-900.

Gutiérrez E, and Zaldívar, J. M., 2000, The application of Karhunen-Loéve, or Principal Component Analysis Method, to study the non-linear seismic response of structures. Earthquake Engng Struct. Dyn. 29,1261-1286.

Hegger, R., Kantz, H., Schreiber, T., 1999, Practical implementation of nonlinear time series methods: The TISEAN package. CHAOS 9, 413-. The software package is publicly available at http://www.mpipks-dresden.mpg.de/~tisean .

Hirata, Y., Horai, S., Kzuyuki,A., 2008. Reproduction of distance matrices and original time series from recurrence plots and their applications. European Physical Journal Special Topics 164, 13-22.

Holme, P., Kim, B.J., Yoon, C.N., Han, S. K. 2002. Attack vulnerability of complex networks. Phys. Rev. E 65, 056109.

Hurst, H. E., 1951, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng*. **116**, 770-779.

Ito, T., Chiba, T., Owaza, R., Yoshida, M., Hattori, M., Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 4569-4574.

Jeong, H., Mason, S., Barabási, A.L., Oltvai, Z.N. 2001. Lethality and centrality in protein networks. Nature 411, 41-42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic netwoks. Nature 407, 651-654.

Kantz H., Shreiber T., 1997. *Nonlinear Time Series Analysis*, Cambridge University Press.

Kennel, M. B., R. Brown and H. D. I. Abarbanel, 1992, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403-3411.

Kim, J.S., Goh, K.I., Salvi, G., Oh, E., Kahng, B. and Kim, D., 2007. Fractality in complex networks: Critical and supercritical skeletons. Physical Review E 75, 1-14.

Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E., and Taylor, W. W., 2003. Compartments revealed in food-web structure. Nature 426, 282-285.

Lacasa, L., Luque, B., Ballestros, F., Luque, J., Nuno, J.C.. 2008. From time series to complex networks: the visibility graph. Proc. Nat. Acad. Sci. USA. 105, 4972-497.

Latapy, M., Magnien, C. 2007. Measuring Fundamental Properties of Real-World Complex Networks, eprint arXiv:cs/0609115.

Latapy, M., Magnien, C. 2008. Complex Network Measurements: Estimating the Relevance of Observed Properties. Infocom'08, 2008, Phoenix, USA. pp. 1660-1668.

Lorenz, E.N. 1963. Deterministic nonperiodic flow. J.Atmos. Sci., 20, 130-141.

Ma, H. and Zeng, A.-P. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19, 270-277.

Mandelbrot, B. B., *The Fractal Geometry of Nature*, 1983, W. H. Freeman. New York.

Martinez, N.D., Williams, R.J., Dunne, J.A., 2006. Diversity, complexity, and persistence in large model ecosystems. Pages 163-185 in Ecological Networks: Linking Structure to Dynamics in Foodwebs. M. Pascual and J.A. Dunne (Editors), Oxford University Press, UK.

Marwan, N., Romano, M.C., Thiel, M. Kurths, J., 2007. Recurrence plots for the analysis of complex systems. Physical Reports 438:237-329.

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., Kurths, J. 2002. Recurrence plot based measures of complexity and its application to heart rate variability data. Phys. Rev. E 66(2), 026702.

Maslov S., Sneppen, K. 2002. Specificity and stability in topology of protein networks. Science 296, 910-913.

Melián and Bascompte (2004)

Milgram, S. 1967. The small world problem. Phys. Today 2, 60-67.

Newman, M.E.J. 2003. The structure and function of complex networks. SIAM Review 45, 167-256.

Nolan, J.P. 1999. Fitting data and assessing goodness of fit with stable distributions, in: Proceedings of the Conference on Applications of Heavy Tailed distributions in Economics, Engineering and Statistics. American University, Washington DC.

Nooy W. de, A. Mrvar, V. Batagelj, 2005. Exploratory Social Network Analysis with Pajek, Cambridge University Press.

Novellino, A. Zaldívar J.M., 2009. Recurrence quantification analysis and spontaneous electrophysiological activity characterization in *in vitro* neuronal networks coupled to multi electrode array (MEA) chips. Biological Cybernetics (submitted).

Packard, N., Crutchfield, J., Farmer, D. and Shaw, R., 1981, Geometry from a time series. Phys. Rev. Lett. 45, 712-715.

Podani, J., Oltvai, Z.N., Jeong, H., Tombor, B., Barabasi, A.L., Szathmary, E. 2001. Comparable system-level organization of *Archaea* and *Eurokaryotes*. Nature Genetics 29, 54-56.

Post, D.M., 2002. The long and short of food chain length. Trends in Ecology and Evolution, 17 (6), 269-277.

Redner S. 2008. Networks: Teasing out the missing links. Nature 453, 47-48.

Schreiber, T., 1998, Interdisciplinary application of nonlinear time series methods. *Physics Reports*.

Schwikowski, B., Uetz, P., Fields, S. 2000. A network of protein-protein interactions in yeast. Nature Biotechnology 18, 1257-1261.

Shen-Orr, S., Milo, R., Mangan, S, Alon, U. 2002. Network motifs in the transcriptional regulation netwok of *Escherichia coli*. Nature Genetics 31, 64-68.

Solé, R.V. and Pastor-Satorras, R. 2003. Complex networks in genomics and proteomics, in Handbook of Graphs and Networks, Bornholdt, S. and Schuster, H.H. eds. Wiley-VCH, Berlin. pp. 145-167.

Song, C., Havlin, S., Makse, H.A., 2005. Self-similarity of complex networks. Nature 433, 392-395.

Song, C., Havlin, S., Makse, H.A., 2006. Origins of fractality in the growth of complex networks. Nature Physics 2, 275-281.

Sporns, O. 2002. Network analysis, complexity, and brain function. Complexity 8, 56-60.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D. 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420, 190-193.

Strogatz, H.S., 2005. Romanesque networks. Nature 433, 365-366

Strozzi, F. and Zaldívar, J. M., 2002, Embedding theory: Introduction and applications to time series analysis, in *Modelling and forecasting financial data: Techniques of Nonlinear Dynamics*. A. Soofi and L. Cao (Eds.), Kluwer Academic Publishers, Boston.

Takens, F., 1981, in Dynamical Systems and Turbulence, Warwick 1980, vol. 898 of Lecture Notes in Mathematics, edited by A. Rand and L.S Young, Springer, Berlin, pp. 366-381.

Takens, F., 1996, The effect of small noise on systems with chaotic dynamics. In *Stochastic and Spatial Structures of Dynamical Systems*, S. J. van Strien and S. M. Verduyn Lunel, Verhandelingen KNAW, Afd. Natuurkunde, vol. 45, pp. 3-15. North-Holland, Amsterdam.

Tong, H., *Nonlinear Time Series: a Dynamical System Approach*. 1990, Oxford University Press. Oxford.

Trulla, L.L, Giuliani, A., Zbilut, J.P., Webber, C.L. 1996. Recurrence quantification analysis of the logistic equation with transients. Phys. Lett. A 223, 255-26.

Tucker, W., 2002. A rigorous ODE solver and Smale's 14th Problem. Found Comp. Math., 2, 53-117

Wagner, A., Fell, D. 2001. The small world inside large metabolic networks. Proc. Roy. Soc. London

Ser. B 268, 1803-1810.

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

Webber Jr. C. L. and Zbilut, J. P., 1994. Dynamical assessment of physiological systems and states using recurrence plot strategies. J. Appl. Physiol. 76, 965-973.

White, J.G., Southgate, E., Thompson, J.N. and Brenner, S. 1986. The structure of the nervous system of the nematode *C. elegans*. Philos. Trans. Roy. Soc. London 314, 1-340.

Williams, R.J. and Martinez, N.D., 2000. Simple rules yield complex foodwebs. Nature 404, 180-183.

Xu, X., Zhang, J., Small, M., Superfamily phenomena and motifs of networks induced from time series. Proc. Nat. Acad. Sci. USA 105, 19601-19605.

Yang, E., Misra, A., Maguire, T.J., Androulakis, I.P. 2009. Analysis of regulatory and interaction networks from clusters of co-expressed genes. In Clustering challenges in biological networks, S. Butenko, W.A. Chaovalitwongse and P.M. Pardalos (Eds.) Scientific World, Singapore. pp 53-82.

Zaldívar, J.M., Galván, I. M., Strozzi, F., Gutiérrez, E., and Tomasin, A., 2000, Forecasting high waters at Venice Lagoon using chaotic time series analysis and non-linear neural networks. J. of Hydroinformatics 2, 61-84.

Zbilut J.P., Webber, C.L., 1992, Embeddings and delays as derived from quantification of recurrence plots. Phys. Lett. A 171, 199-203.

Zhang, J., Small, M., 2006. Complex Network from pseudoperiodic time series: topology versus Dynamics. Phys. Rev. Lett. 96, 1-4.

European Commission

EUR 23947 EN – Joint Research Centre
Title: From complex networks to time series analysis and *viceversa*: Application to metabolic networks
Author(s): F. Strozzi, J. M. Zaldívar, K. Pioljansek, F. Bono and E. Gutiérrez
Luxembourg: Office for Official Publications of the European Communities
2009 – 56 pp. – 21 x 29,7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-12955-1
DOI 10.2788/25588

Abstract. In this work we present a simple and fast approach to generate network structures based on time series recurrence plots and *viceversa*. In addition, we discuss the application of the different analysis techniques developed in both fields, i.e. complex networks and time series analysis. Concerning the transformation from time series to networks, we propose a deterministic growth procedure which produces a new types of complex network structures that have some interesting features. This simple and fast approach is able to generate deterministic network structures based on time series recurrence plots. The generated networks contain several properties of the original time series. In this case, networks generated from chaotic attractors display interesting features from the point of view of robustness which could help in designing systems with high tolerance against errors and transfer of information. Chaotic networks based on the Lorenz attractor show that they are highly tolerant against attacks and they have a high ability for the transfer of information or on the contrary they are able to transmit infections faster. It is still necessary to investigate if such chaotic networks exist already in natural or man-made systems or, if possible, to construct such networks and test their properties. On the other hand, the transformation from networks to time series presents some problems concerning the selection of the initial time or in our case the initial node and the way in which the nodes are visited. If a network has been generated following a certain growth law it seems logical to choose the first node as the origin and then proceed following the network growth pattern. However, the situation is not so clear for example with metabolic networks, where it is difficult to select which is the first metabolite. Similar concerns would apply to other types of biological networks. In this case several alternatives could be considered, e.g. ordering using the number of connections. However, we have still to find if there are some invariant/preserved properties in the generated time series from the same network. We have found that rescaled range analysis does not preserve the fractal structure in the time series. In any case, if time series parameters would be invariant against the initial node selection, then they could be used to analyze the networks that have generated said time series. Our future work will continue along these lines.
.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-23947-EN-C

ISBN 978-92-79-12955-1