Physica A 428 (2015) 302–313

Contents lists available at ScienceDirect

Physica A

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

Multifractal characterization of protein contact networks

Enrico Maiorino [a,*], Lorenzo Livi [b], Alessandro Giuliani [c], Alireza Sadeghian [b], Antonello Rizzi [a]

[a] Department of Information Engineering, Electronics, and Telecommunications, SAPIENZA University of Rome, Via Eudossiana 18, 00184 Rome, Italy

[b] Department of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

[c] Department of Environment and Health, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

HIGHLIGHTS

- We study protein contact networks by techniques proper for time series analysis.
- Time series generated from protein contact networks show persistence.
- Such time series present features in-between typical mono and multifractal signals.
- The multifractal spectra are embedded in a principal components space.
- Fluctuations with large magnitude are more characterizing than small fluctuations.

ARTICLE INFO

Article history:
Received 9 October 2014
Available online 11 February 2015

Keywords:
Multifractal analysis
Complex biological networks
Time series analysis
Random walk

ABSTRACT

The multifractal detrended fluctuation analysis of time series is able to reveal the presence of long-range correlations and, at the same time, to characterize the self-similarity of the series. The rich information derivable from the characteristic exponents and the multifractal spectrum can be further analyzed to discover important insights into the underlying dynamical process. In this paper, we employ multifractal analysis techniques in the study of protein contact networks. To this end, initially a network is mapped to three different time series, each of which is generated by a stationary unbiased random walk. To capture the peculiarities of the networks at different levels, we accordingly consider three observables at each vertex: the degree, the clustering coefficient, and the closeness centrality. To compare the results with suitable references, we consider also instances of three well-known network models and two typical time series with pure monofractal and multifractal properties. The first result of notable interest is that time series associated to protein contact networks exhibit long-range correlations (strong persistence), which are consistent with signals in-between the typical monofractal and multifractal behavior. Successively, a suitable embedding of the multifractal spectra allows to focus on ensemble properties, which in turn gives us the possibility to make further observations regarding the considered networks. In particular, we highlight the different role that small and large fluctuations of the considered observables play in the characterization of the network topology.

* Corresponding author.
E-mail addresses: enrico.maiorino@uniroma1.it (E. Maiorino), llivi@scs.ryerson.ca (L. Livi), alessandro.giuliani@iss.it (A. Giuliani), asadeghi@ryerson.ca (A. Sadeghian), antonello.rizzi@uniroma1.it (A. Rizzi).
URLs: https://sites.google.com/site/lorenzlivi/ (L. Livi), http://www.scs.ryerson.ca/~asadeghi/ (A. Sadeghian), http://infocom.uniroma1.it/~rizzi/ (A. Rizzi).

http://dx.doi.org/10.1016/j.physa.2015.02.026
0378-4371/

1. Introduction

Beside the prophecy made by Weaver [1], identifying in the study of "organized complexity" (i.e., systems made by many interconnected elements) as the new theoretical frontier of Sciences, the Dutch electrical engineer Bernard Tellegen, already in 1952 demonstrated the existence of regularities and laws dependent only on the wiring architecture of the studied systems [2,3]. These cause systems made by elements obeying different physical laws (e.g. electrical circuits, molecules, metabolic networks) to display a very similar mesoscopic behavior. These "network laws" are still largely unknown and there is a huge room of investigation for catching these principles. This is the main focus of the emerging research field of complex networks analysis [4–10].

A well-known approach to characterize complex networks consists in analyzing the fractal properties of their topology (i.e., scaling properties), which is usually implemented by means of the box counting method [11]. Recently, new methods [12–14] have been proposed to generalize this analysis to the multifractal setting [15]. A network (or equivalently, a graph) is a very general mathematical construct having the power of unifying apparently different objects like correlation matrices, recurrence plots, adjacency matrices. The only necessary property to call something a graph is the existence of a wiring scheme that, for any pair of constituting elements, returns a binary information (or in certain cases, the strength) about the existence of a link between the elements. The definition of this wiring scheme allows to transfer virtually any problem to the graph domain, granting the possibility to take advantage of all the tools coming from classical graph theory and the more recently developed complex networks analysis. In particular, when studying time series it is possible to generate a network (in the form of recurrence plots, correlation matrices, or other mapping methods) that preserves their temporal correlation structure in the form of topological invariants [16–18]. Nonetheless, the time series/network "isomorphism" works in both directions. The exploration of large networks by means of a random walker, taking at each vertex different directions along the graph topology according to some probabilistic criteria [19], has the potential to offer a dynamical perspective of the studied network. However, while the former approach has been deeply investigated [16], less attention has been devoted to the latter in the case of (multi)fractal characterization, apart from few recent exceptions [20–22].

In this paper we exploit the Multifractal Detrended Fluctuation Analysis (MFDFA) [23–25], a generalization of the Detrended Fluctuation Analysis (DFA) [26], to study time series obtained from complex networks via stationary unbiased random walks (RW). The MFDFA builds upon a generalization of the so-called Hurst exponent as a detector of long-range correlations [27,28]. On the basis of Hurst exponent is the idea of characterizing time series in terms of their degree of *persistence*: roughly speaking, a series is long-range correlated (persistent) if the underlying process has memory of the past states, a property that is firstly noticeable as a heavy-tail in the corresponding autocorrelation function. Brownian motion corresponds to Hurst exponent equal to 0.5 and it is considered as the baseline uncorrelated process. Series with Hurst exponent greater than 0.5 are considered as persistent; series with Hurst exponent smaller than 0.5 are anti-persistent (consecutive values tend to be very different). Additionally, if the value of this exponent does not vary significantly with the magnitude of fluctuations, then the time series is considered monofractal and it can be consistently analyzed via DFA; in the opposite case, it is multifractal and the MFDFA is a more suitable choice. If the studied time series corresponds to a sequence of discrete observables attached to the vertices of a network and the ordering is determined by the subsequent encounters of a random walker exploring the graph, then its persistence/antipersistence property can be translated into the assortative/disassortative character of the graph with respect to said observables. An assortative graph [29] is a graph in which vertices with similar properties (typically the degree is used, but in theory any property of the vertex can be taken into account) tend to be in contact more frequently than what expected by chance, while a disassortative graph has the opposite feature. Studying a complex network by the action of a random walker producing a collection of time series of encounters with vertices has an advantage with respect to the simple computation of the static assortative indexes of the graph. Indeed, the walker trajectories offer also a sampling of the paths distribution in the graph. This distribution is affected by the whole set of mutual relations of vertices at different scales, which are not fully appreciable by a single static snapshot of the network by means of classical network invariants. In the same manner, we are able to gain an insight into the different scaling of the autocorrelation function and hence on the distribution of the corresponding observable across different locations and scales of the network.

In this study, we primarily focus on the protein contact networks (PCN) elaborated from the E. coli [30]. The networks consist of amino acid residues put in contact according to the 3D protein structure [31]. We compare the properties of PCN with those of different known network and time series models. In particular, we bias the study on their analogies and differences with a model of synthetic protein contact networks (PCN-S) as theorized by Bartoli et al. [32]; PCN-S consist in coiled coils of polymers in which the probability of contact is a decreasing function of the distance between residues along the chain. From our study, PCN clearly emerge as a particular class of networks whose time series exhibit properties in between mono and multi fractal signals. The synthetic (simulated coiled coils) proteins, i.e., PCN-S, preserve sufficiently well the multifractality (thus the presence of different scaling relations at different scales or positions), while they seem to lose much of the strong persistence typical of PCN time series. Comparing the PCN with their synthetic counterparts allowed us to highlight a different level of order of the former not shown in the latter due to the pure synthetic polymer folding in space (i.e. no excluded volume constraints). Indeed, the additional freedom given by pure folding causes the loss of many peculiar properties of proteins, like allosteric effect (i.e. the global configurational change of the system upon ligand binding at a specific site) and rapid folding [33,34]. Notably, the particular wiring of protein contact networks encompasses two