Physica A 415 (2014) 73–86

Contents lists available at ScienceDirect

Physica A

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

Variability analysis of complex networks measures based on stochastic distances

Raquel S. Cabral [a,b,*], Alejandro C. Frery [b], Jaime A. Ramírez [a]

[a] Graduate Program in Electrical Engineering, Federal University of Minas Gerais, Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil

[b] Laboratório de Computação Científica e Análise Numérica – LaCCAN, Universidade Federal de Alagoas – Ufal, Av. Lourival Melo Mota, s/n, 57072-900, Maceió, AL, Brazil

HIGHLIGHTS

- We evaluate the variability of complex network measures face to perturbations.
- We analyze theoretical models and real networks with stochastic distances.
- Clustering coefficient is not sensitive to perturbations in theoretical models.
- Hypothesis tests verified the perturbations that break the distribution of degrees.
- The Hellinger distance is the most sensitive distance to the perturbations.

ARTICLE INFO

Article history:
Received 8 September 2013
Received in revised form 30 May 2014
Available online 8 August 2014

Keywords:
Complex network
Measures
Stochastic distance
Variability

ABSTRACT

Complex networks can model the structure and dynamics of different types of systems. It has been shown that they are characterized by a set of measures. In this work, we evaluate the variability of complex network measures face to perturbations and, for this purpose, we impose controlled perturbations and quantify their effect. We analyze theoretical models (*random*, *small-world* and *scale-free*) and real networks (*a collaboration network* and *a metabolic networks*) along with the *shortest path length*, *vertex degree*, *local cluster coefficient* and *betweenness centrality* measures.

In such an analysis, we propose the use of three stochastic quantifiers: the Kullback–Leibler divergence and the Jensen–Shannon and Hellinger distances. The sensitivity of these measures was analyzed with respect to the following perturbations: edge addition, edge removal, edge rewiring and node removal, all of them applied at different intensities. The results reveal that the evaluated measures are influenced by these perturbations. Additionally, hypotheses tests were performed to verify the behavior of the degree distribution to identify the intensity of the perturbations that leads to break this property.

1. Introduction

Complex networks are systems whose structure is irregular, complex and dynamically evolving in time [1]. In recent years, a number of measures have been developed to quantify the structure and behavior of such systems, which provide

* Corresponding author at: Graduate Program in Electrical Engineering, Federal University of Minas Gerais, Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil. Tel.: +55 8232141882.

E-mail addresses: raquelcabral@gmail.com (R.S. Cabral), acfrery@gmail.com (A.C. Frery), jramirez@ufmg.br (J.A. Ramírez).

http://dx.doi.org/10.1016/j.physa.2014.07.079
0378-4371/

a framework that allows its characterization, analysis and modeling, reflecting different features of the network such as, connectivity, centrality, cycles, distances, among others.

The choice of an appropriate measure for the characterization of a network is performed by evaluating its behavior, and depends mainly on three factors: (i) data availability; (ii) storage capacity and processing; and (iii) interest in characterizing the behavior of the measures. In this procedure, the network is mapped into a feature vector [2]; however, in many cases the mapping is not complete and does not accurately describe the network's real structure. In such cases, it is important to evaluate the performance of the measures when unexpected changes occur in the networks. For instance, *what is the behavior of the measures if the network loses links or nodes?* or, *do these changes break the properties used to describe the network structure?* To address such problems, it is necessary to compare different states of the network. In this work, we investigate the use of methods from the Information Theory, in particular the concept of Stochastic Quantifiers, as means to quantify the changes.

There are a few works that explore the use of stochastic measures to analyze the behavior of complex networks. Wang et al. [3] employed the concept of Entropy of the degree distribution to provide an amount of the network's heterogeneity, since that measure quantifies the diversity of the link distribution. They also studied the robustness of scale-free networks using the Entropy.

A common practice to address the questions raised previously is to use samples of the network, instead of considering it entirely. This saves memory and processing time. Boas et al. [4] employed the Kullback–Leibler divergence to compare different states of the network and assessed the appropriateness of using network samples. However, this divergence cannot be considered a distance since it is not commutative. We extend those results in two manners, namely, (i) we analyze the behavior of four measures applied to different theoretical network models and real networks; and (ii) we compare the sensitivity of three stochastic quantifiers with respect to several perturbations of the networks, including node removal.

The study of the evolution of complex networks is an important issue due to their dynamic nature. Among the changes a network may be subjected to, we will consider node removal, edge addition, edge removal, and edge rewiring. Such perturbations describe common changes in practical situations as, for instance, the death of a node, the creation and the deletion of a connection, and both operations at once.

Carpi et al. [5] proposed a new quantifier based on Information Theory for the analysis of the evolution of small-world dynamic networks. The quantifier, a statistical complexity measure, is used to compute changes in the topological randomness for degree distribution of the network. It is obtained by the product of the normalized Shannon entropy and the normalized Jensen–Shannon distance. This quantifier requires the use of a probability distribution as a reference to compute the Jensen–Shannon distance. The authors used three reference distributions: Poisson, uniform and the distribution corresponding to the regular lattice.

The analysis of perturbations in networks has also been studied with two practical purposes in mind: their vulnerability to attacks and the identification of elements whose failure leads to a breakdown. The vulnerability is associated with the decrease of network performance when structural changes occur; these can be caused by the random or directed removal of vertices, termed failure and attack, respectively. Measures related to this property are commonly defined in terms of the shortest path length and the size of the connected components of the graph. The main idea is to intentionally apply a sequence of failures (or attacks) to the network, and to observe its behavior [6–10].

In this context, many real networks have been studied. Pu et al. [11] studied the behavior of network controllability under vulnerability for networks of different topologies under two different kinds of attacks, including the power grid networks. Jeong et al. [12] evaluated the diameter behavior in protein networks by removal of the most connected protein, and Kaiser and Hilgetag [13] evaluated the robustness toward edge elimination of metabolic networks showing that intercluster connections represent the most vulnerable edges in these networks.

The breakdown phenomenon in networks refers to a type of cascade process, where the failure of a single or a few nodes may result in the collapse of their functionality. In networks with power distribution, for instance, the failure of a node requires that its load is redistributed to other nodes, causing a network overload and possibly other faults [14,15].

Cabral et al. [16] presented the analysis of communication strategies in wireless sensor networks by means of analyzing the variation of the shortest path length measure in flooding, random, small-world, and scale-free networks. The variation of this measure was analyzed with respect to the insertion and removal of nodes in flooding; and with respect to insertion, removal and rewiring of links in the strategy based in complex networks. Stochastic quantifiers, namely the normalized Kullback–Leibler divergence and Hellinger distance were used to quantify the variation of the shortest paths.

The goal of this work is to analyze the behavior of measures face to network perturbations. We propose the use of quantifiers based on information theoretic tools to perform this analysis. The complex network measures evaluated were shortest paths length, vertex degree, local cluster coefficient and betweenness centrality. The first one is calculated for each pair of nodes and the others are calculated for each network node. We compare the results, which are vector-valued measures, using of stochastic divergences between discrete probability distributions. In addition, we propose the use of three quantifiers: the Kullback–Leibler divergence, the Jensen–Shannon and Hellinger distances.

The proposed methodology allows the comparison of the behavior of different quantifiers and the identification of the intensity of the perturbations that leads to significant changes of their properties. The results reveal that all the evaluated measures are influenced by the perturbations considered, but to different extent.