COMSTA: 6392 Model 3Gsc pp. 1–17 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Q1 # Correlation between graphs with an application to brain network analysis

Q2 André Fujita[a,*], Daniel Yasumasa Takahashi[b], Joana Bisol Balardin[c], Maciel Calebe Vidal[a], João Ricardo Sato[d]

[a] *Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Brazil*
[b] *Department of Psychology and Neuroscience Institute, Green Hall, Princeton University, USA*
[c] *Hospital Israelita Albert Einstein, Brazil*
[d] *Center of Mathematics, Computation, and Cognition, Federal University of ABC, Brazil*

ARTICLE INFO

Article history:
Received 6 June 2016
Received in revised form 25 November 2016
Accepted 26 November 2016
Available online xxxx

Keywords:
Network
Correlation
fMRI
Functional brain network
Autism

ABSTRACT

The global functional brain network (graph) is more suitable for characterizing brain states than local analysis of the connectivity of brain regions. Therefore, graph-theoretic approaches are natural methods to use for studying the brain. However, conventional graph theoretical analyses are limited due to the lack of formal statistical methods of estimation and inference. For example, the concept of correlation between two vectors of graphs has not yet been defined. Thus, the introduction of a notion of correlation between graphs becomes necessary to better understand how brain sub-networks interact. To develop a framework to infer correlation between graphs, one may assume that they are generated by models and that the parameters of the models are the random variables. Then, it is possible to define that two graphs are independent when the random variables representing their parameters are independent. In the real world, however, the model is rarely known, and consequently, the parameters cannot be estimated. By analyzing the graph spectrum, it is shown that the spectral radius is highly associated with the parameters of the graph model. Based on this, a framework for correlation inference between graphs is constructed and the approach illustrated on functional magnetic resonance imaging data on 814 subjects comprising 529 controls and 285 individuals diagnosed with autism spectrum disorder (ASD). Results show that correlations between the default-mode and control, default-mode and somatomotor, and default-mode and visual sub-networks are higher in individuals with ASD than in the controls.

1. Introduction

There is an increasing number of evidences suggesting that complex cognitive processes arise from orchestrated communication of brain areas namely, functional connectivity networks (Cassidy et al., 2016). Functional connectivity between two brain areas can be ascertained with correlation measures, such as Pearson's and Spearman's correlation coefficients (Garcés et al., 2016; Richiardi et al., 2011) between functional magnetic resonance imaging (fMRI) signals. The analysis of functional connectivities is beginning to elucidate cognition at a systems level, yet it remains largely unknown

* Correspondence to: Rua do Matão, 1010 – Cidade Universitária, São Paulo- SP, 05508-090, Brazil.
E-mail address: fujita@ime.usp.br (A. Fujita).

http://dx.doi.org/10.1016/j.csda.2016.11.016

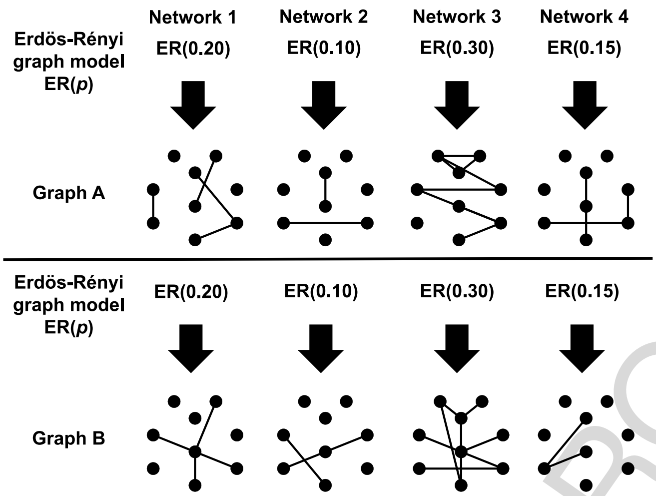

Fig. 1. Two vectors of random graphs **A** and **B** generated by Erdös–Rényi (ER) random graph models with the parameter p between brackets. Notice that the parameters of the ER models are perfectly correlated, however, this correlation cannot be easily identified by directly analyzing the structure of graphs **A** and **B**. Notice that although they are generated by the same graph model and parameters, they are structurally different. Thus, one solution for identifying correlation between graphs consists of identifying correlation between the parameters of the random graph models.

how they are associated with mental disorders (Cassidy et al., 2016; Di Martino et al., 2009; Nomi and Uddin, 2015; Kana et al., 2014; Uddin et al., 2013). Some studies indicate that dysregulations in functional connectivity are associated with neuropsychiatric disorders such as Alzheimer's disease, multiple sclerosis, and epilepsy (Stam, 2014).

More recently, special attention is given to the idea that brain regions are organized into interconnected communities or sub-networks (Sporns, 2013). Thus, the identification of connectivity between sub-networks could provide insights into the basis of inter-group differences (controls versus patients) (Bassett et al., 2011, 2015). Therefore, the development of methods to identify correlation between sub-networks becomes essential.

Mathematically, networks can be represented as graphs, i.e., a set of vertices and a set of edges, where the edges indicate which pairs of vertices are associated. For our purposes, these vertices represent ROIs and the edges represent the functional connectivity (correlation) between two ROIs.

Networks or graphs are difficult to be manipulated from a statistical viewpoint because they are not numbers. For an illustrative example, see Fig. 1. By analyzing Fig. 1, it is very difficult to identify a correlation between the two vectors of graphs **A** and **B** by only analyzing their structures. Thus, to construct a framework to infer correlation, one natural idea would be to imagine that a graph is generated by a mathematical model with a set of parameters which are the random variables. Intuitively, two vectors of graphs are correlated when the parameters (random variables) of the graph model are correlated (Fig. 1). However, given two vectors of graphs, the model that generates them is rarely known, and consequently, the parameters cannot be estimated.

Thus, it is necessary to identify a feature of the graph that is highly associated with the parameters of the graph model. To identify the feature that contains information on the parameters, we investigated the spectral properties of random graphs (set of eigenvalues of the adjacency matrix). It is known that some structural properties, such as the number of walks, diameter, and cliques can be described by the spectrum of the graph (Van Mieghem, 2010). Here, we propose to estimate the correlation between graphs by using the spectral radius (largest eigenvalue) of the graphs. Our results show that the spectral radius is highly associated with the parameters that generate the graph, and thus, it can be a good feature for calculating correlation between two graphs.

By simulations, we show both the statistical power and the effective control of type I error. Then we illustrate the usefulness of our method by analyzing a large fMRI dataset (ABIDE – The Autism Brain Imaging Data Exchange – Consortium website – http://fcon_1000.projects.nitrc.org/indi/abide/) composed of 814 participants comprising 529 controls and 285 individuals with ASD.

2. Description of the method

2.1. Graph

A graph is an ordered pair $G = (V, E)$, where V is a set of n vertices $(v_1, v_2, \ldots, v_n)$ and E is a set of m edges that connect two vertices of V. In this study we will consider solely the case of graphs with non-empty set of nodes and edges.

Any undirected graph G with n vertices can be represented by its adjacency matrix $\mathbf{A}^G$ with $n \times n$ elements $\mathbf{A}^G_{ij}$ $(i, j = 1, \ldots, n)$; its value is $\mathbf{A}^G_{ij} = \mathbf{A}^G_{ji} = 1$ if vertices v_i and v_j are connected and 0 otherwise. The spectrum of graph G is the set

Please cite this article in press as: Fujita, A., et al., Correlation between graphs with an application to brain network analysis. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.11.016