Physica A 387 (2008) 6436–6442

Contents lists available at ScienceDirect

Physica A

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

Spectral reconstruction of complex networks☆

Francesc Comellas *, Jordi Diaz-Lopez

Departament de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya, Avda. Canal Olímpic s/n, 08860, Castelldefels, Catalonia, Spain

ARTICLE INFO

Article history:
Received 2 May 2008
Received in revised form 18 July 2008
Available online 8 August 2008

Keywords:
Complex networks
Eigenvalues
Laplacian spectrum
Tabu search

ABSTRACT

In this paper we study the reconstruction of a network topology from the eigenvalues of its Laplacian matrix. We introduce a simple cost function and consider the tabu search combinatorial optimization method, while comparing its performance when reconstructing different categories of networks – random, regular, small-world, scale-free and clustered – from their eigenvalues. We show that this combinatorial optimization method, together with the information contained in the Laplacian spectrum, allows an exact reconstruction of small networks and leads to good approximations in the case of networks with larger orders. We also show that the method can be used to generate a quasi-optimal topology for a network associated to a dynamic process (like in the case of metabolic or protein–protein interaction networks of organisms).

1. Introduction

In recent years there has been a growing interest in the study of complex networks, related to transportation and communication systems – WWW, Internet, power grid, etc. – including social and biological networks (like the metabolic or the protein–protein interaction networks of organisms). Many of these networks are large, with a number of nodes, very often in the thousands. To store the topological details of the network requires knowing the list of adjacencies and, although usually the networks are sparse, this means the use of a large amount of memory. In contrast, the spectrum of the network (eigenvalues of the associated adjacency or Laplacian matrix) contains important information with significantly less memory use. Also, the spectrum provides information on the behavior of dynamic processes supported by the network (see Ref. [1].) Therefore, it is of interest to reconstruct or generate a network from its spectrum.

In Ref. [2], Ipsen and Mikhailov use simulated annealing with an elaborated cost function based on the spectral density to perform such a reconstruction. The spectral density of a network has also been considered by different authors for classification purposes, see for example Refs. [3–5]. Here, we propose a simple cost function which, together with the information provided by the knowledge of the spectrum, drives the tabu search method towards a good network reconstruction. The method is probabilistic, i.e. it has a random component, and as a consequence we can not guarantee that the algorithm will find an optimal reconstruction, but we show that the final networks match the originals in their main topological properties.

We also show that the method can be used, without modifications, to generate a quasi-optimal topology for a network associated to a dynamic process. As an example, we have considered the Laplacian spectrum of the largest connected components from the metabolic network of *Saccharomyces cerevisiae* and the protein–protein interaction network of *Helicobacter pylori*, to generate new networks with a similar topology. This can be used to produce network models when the spectrum of a network is known or it can be inferred, even partially, from its dynamical properties.

☆ Research supported by the Ministerio de Educación y Ciencia, Spain, and the European Regional Development Fund under project TEC2005-03575 and by the Catalan Research Council under project 2005SGR00256.

* Corresponding author. Tel.: +34 934 137 072; fax: +34 934 137 007.
E-mail address: comellas@ma4.upc.edu (F. Comellas).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.07.032

In the next section, we introduce the mathematical notation and concepts necessary for our study. Section 3 contains a description of tabu search, the combinatorial optimization algorithm considered, with details on its use with information provided by the spectrum. Our main results are presented in Section 4.

2. The Laplacian spectrum of a graph and its reconstruction

Let us consider the Laplacian of the graph $G = G(V, E)$, with vertex set V (order $n = |V|$) and edge set E, associated with a network. The Laplacian is a symmetric matrix with zero row-sums that accounts for the topology of the network, defined to be $L_{ij} = -1$ if nodes i and j are connected, $L_{ii} = \delta_i$ if node i has degree δ_i (i.e. is connected to δ_i other nodes), and $L_{ij} = 0$ otherwise. The Laplacian matrix can be related to the adjacency matrix A of G by $L = D - A$, where D is the diagonal matrix of vertex degrees of G.

The (Laplacian) spectrum of G consists of the n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the Laplacian matrix and they satisfy $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. We have also considered the reconstruction of a network from the spectra of the sign-less Laplacian matrix, defined as $|L| = D + A$, and the normalized Laplacian matrix, defined to be $\bar{L}_{ij} = -1/\sqrt{\delta_i \delta_j}$ if nodes i and j are connected, $\bar{L}_{ii} = 1$ if node i has degree $\delta_i > 0$, and $\bar{L}_{ij} = 0$ otherwise. For this last matrix, the eigenvalues satisfy $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ and $1 < \lambda_2 \leq 2$. In this paper, when we refer to the Laplacian matrix we mean the first definition unless stated otherwise. The spectrum of a graph is important as it provides bounds on its diameter, maximum and minimum degrees, and gives information about possible partitions etc. It can also be used to count the number of paths of a given length in the network, number of triangles, total number of links, etc., see Refs. [6,7]. Dynamic properties of a network, like its synchronizability, can also be determined from the eigenvalues, see Refs. [8,1] We note that two isomorphic graphs have the same spectrum, independent of the labeling of the vertices, but there also exist non-isomorphic graphs (topologically different) with the same spectrum, known as cospectral graphs. For $n < 6$ there are no connected cospectral graphs with respect to the usual Laplacian matrix. For $n = 6$ there exist four pairs, there are 130 pairs for $n = 7$, 1767 pairs for $n = 8$, etc. The number of cospectral graphs increases rapidly with the order of the graph, but the fraction is very small (e.g. it is 0.09 for $n = 11$) and approaches zero as the order of the graph increases, see Refs. [9,10]. Hence, two graphs with the same spectrum would indeed be isomorphic with a high probability.

In this paper, we study the reconstruction of graphs from their Laplacian spectra. Note that the number of different graphs of a given order n is large even for relatively small order. For example, for $n = 40$ there are roughly 10^{186} graphs. It makes no sense to check all of these graphs to find one matching spectrum, even in an approximate way. We are in the classical situation where combinatorial optimization algorithms are useful.

The generic process is as follows: we will reconstruct a given reference graph G_0 from its spectrum $\{\lambda_i^0, 0 \leq i \leq n\}$. In the reconstruction process, we generate an initial random graph G_{ini} using the information provided by the spectrum. The graph should have n vertices, $(\sum_{i=1}^{n} \lambda_i)/2$ edges and fulfill the following contraints on the degrees maximum and minimum: $\frac{4}{n\lambda_2} \leq \Delta \leq \frac{n-1}{n}\lambda_n$ and $\delta \geq \frac{n-1}{n}\lambda_2$, see [11]. Some bounds on the diameter provided by Mohar in Ref. [12] could also be considered: $\frac{4}{n\lambda_2} \leq D \leq 2\lceil \frac{\Delta+\lambda_2}{4\lambda_2} \ln(n-1) \rceil$.

Next we apply a process of change and selection. Change is done by random modifications of the pattern of connections, whereas the selection process is based on the spectral distance between two graphs, a concept which we discuss below.

A typical modification of a graph consists in reconnecting one edge, while keeping again the constraints deduced from the spectrum.

To decide if the changes should be accepted (that is, to perform *selection*), we need a measure (cost function) of the "distance" of a given graph G_t with eigenvalues $\{\lambda_i^t, 0 \leq i \leq n\}$ from the reference graph G_0. This measure is given by a spectral distance ϵ. In this paper, we introduce a simple spectral distance based on the quadratic difference of the eigenvalues $\epsilon = \sum_{i=0}^{n} (\lambda_i^0 - \lambda_i^t)^2$. We have tested other spectral distances which give different weights to the eigenvalues, but they are more complex and their efficiency is similar. Also, our new distance is simpler than the distance considered in the related study by Ipsen and Mikhailov [2]. These authors use the distance function $\epsilon = \sqrt{\int_0^\infty [\rho(\omega) - \rho_0(\omega)]^2 \mathrm{d}\omega}$, where $\rho(\omega)$ is the spectral density defined as $\rho(\omega) = K \sum_{k=1}^{n-1} \gamma/((\omega - \omega_k)^2 + \gamma^2)$ with $\omega_k^2 = -\lambda_k$, K is a normalization constant and γ is the width of the Lorentz distribution. Note that the spectral distance, which has to be evaluated many times by the two algorithms, involves in both cases the computation of all the eigenvalues of the graph. After this, in our case, the distance is calculated with simple operations: the sum of the squares of the eigenvalue differences. The distance proposed in [2], on top of these last operations, requires an integration.

The main problem with the reconstruction of a graph is to relate the generated graph to the reference graph. The graphs can be isomorphic but with permuted vertices or non-isomorphic with topological similarity that might not be manifest. As in Ref. [2], we check the similarity between two graphs in terms of the singular value decomposition [13] of their adjacency matrices (the details of which are in Section 4). Recently, another method, which measures and visualizes the similarity between networks, has been published in Ref. [14]. It would be of interest to check if it can be also applied in this context.

3. Combinatorial optimization algorithms

When exact methods are not possible, sometimes it is sufficient to obtain an approximate solution with a fast easily