Linear Algebra and its Applications 489 (2016) 30–49

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Spectral classes of regular, random, and empirical graphs

Jiao Gu [a,b,c,*], Jürgen Jost [a,d], Shiping Liu [e], Peter F. Stadler [b,a,d,f]

[a] *Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany*
[b] *Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*
[c] *School of Science, Jiangnan University, Wuxi, 214122, China*
[d] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, United States*
[e] *Department of Mathematical Sciences, Durham University, DH1 3LE Durham, United Kingdom*
[f] *Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*

ARTICLE INFO

Article history:
Received 13 May 2015
Accepted 28 August 2015
Submitted by R. Brualdi

MSC:
05C50
05C75
05C82

Keywords:
Quasimetric
Laplacian spectrum
Radon measure
Graph families

ABSTRACT

We define a (pseudo-)distance between graphs based on the spectrum of the normalized Laplacian. Since this quantity can be computed easily, or at numerically estimated, it is suitable for comparing in particular large graphs. Numerical experiments demonstrate that the spectral distance provides a practically useful measure of graph dissimilarity. The asymptotic behavior of the Laplacian spectrum furthermore yields a tool for classifying families of graphs in such a way that the distance of two graphs from the same family is bounded by $\mathcal{O}(1/n)$ in terms of size n of their vertex sets.

* Corresponding author.
E-mail addresses: jiao@bioinf.uni-leipzig.de (J. Gu), jost@mis.mpg.de (J. Jost), shiping.liu@durham.ac.uk (S. Liu), studla@bioinf.uni-leipzig.de (P.F. Stadler).

http://dx.doi.org/10.1016/j.laa.2015.08.038
0024-3795/

1. Introduction

Structural comparison of graphs has important applications in biology and pattern recognition, see e.g. [1,2]. The problem comes in two distinct flavors: it is comparably easy when correspondences between nodes are known. This is the case e.g. for the comparison of metabolic networks or protein–protein interaction networks [3]. The problem becomes much more difficult when node correspondences are unknown, as is the case e.g. in the atom-mapping problem reviewed in [4]. A classical combinatorial formulation of the latter problem is to find the largest graph G that is isomorphic to a subgraph of each of two given input graphs G_1 and G_2. A natural metric distance is given by $d_{MCSI}(G_1, G_2) := \|G_1 \setminus G\| + \|G_2 \setminus G\|$, were $\|\cdot\|$ is measure of graph size, e.g. the sum of edges and vertices. The main difficulty for practical applications is that "maximum common subgraph isomorphism problem" is NP-complete [5,6] and even APX-hard [7].

For large graphs, thus, more computationally efficient distance measures are required. Graph kernels [8] describe graphs as vectors of features, usually the occurrence data of small subgraphs have increasingly been used in bioinformatics [9] and chemoinformatics [10]. A related approach computed the earth movement (Wasserstein) distance between the distributions of graph features [11]. A practical difficulty is the fact that a very large number of features is required to achieve sufficient resolution for very large graphs.

Here we pursue a different approach that makes use of the representation of graphs by its adjacency or its Laplacian matrix. Spectral properties of these matrix representation are closely related to the graph structure [12,13]. Spectral graph theory in turn has received much inspiration from eigenvalue estimates in Riemannian geometry, see e.g. [12,14,15]. Many of the estimates and bounds for graph properties involve only particular eigenvalues, such as the smallest or largest.

Here we compare the entire spectra of two graphs [16]. This approach is attractive because there are very efficient and numerically stable algorithms for computing the eigenvalues of a large $(N \times N)$-matrix, in fact with an effort of only $\mathcal{O}(N^2)$ in practice [17]. For the set of graphs of the same size, a spectral distance based on the adjacency matrix was suggested by Richard Brualdi ([18], Problem AWGS.4) as a cospectral measure and further studied in [19]. In a related approach, the rank ordered list of eigenvalues is padded by 0 entries if the graphs have different size [20].

The spectrum does not uniquely determine a graph, i.e., there are pairs of non-isomorphic graphs with the same Laplacian or adjacency spectrum [21]. Although it remains an open problem what fraction of graphs is uniquely determined by its spectrum [22,17], we shall see that the comparison of graph spectra nevertheless provides a sensitive and computationally attractive graph distance. We propose here a spectral distance associated with the normalized Laplacian instead of the adjacency matrix, without any constraint on the graph sizes. The reason is that the normalized Laplacian, with its natural interpretation in terms of random walks or diffusion, seems to capture some geometric properties better than the adjacency matrix.