J. Math. Anal. Appl. 449 (2017) 1581–1600

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Accounting for the role of long walks on networks via a new matrix function

Ernesto Estrada *, Grant Silver

Department of Mathematics & Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XQ, UK

ARTICLE INFO

Article history:
Received 23 May 2016
Available online 3 January 2017
Submitted by J.A. Filar

Keywords:
Matrix error functions
Matrix tanh function
Communicability functions
Double-factorial
Chordless cycles
Complex networks

ABSTRACT

We introduce a new matrix function for studying graphs and real-world networks based on a double-factorial penalization of walks between nodes in a graph. This new matrix function is based on the matrix error function. We find a very good approximation of this function using a matrix hyperbolic tangent function. We derive a communicability function, a subgraph centrality and a double-factorial Estrada index based on this new matrix function. We obtain upper and lower bounds for the double-factorial Estrada index of graphs, showing that they are similar to those of the single-factorial Estrada index. We then compare these indices with the single-factorial one for simple graphs and real-world networks. We conclude that for networks containing chordless cycles—holes—the two penalization schemes produce significantly different results. In particular, we study two series of real-world networks representing urban street networks, and protein residue networks. We observe that the subgraph centrality based on both indices produces a significantly different ranking of the nodes. The use of the double-factorial penalization of walks opens new possibilities for studying important structural properties of real-world networks where long-walks play a fundamental role, such as the cases of networks containing chordless cycles.

1. Introduction

The study of large graphs and networks has become an important topic in applied mathematics, computer sciences and beyond [35,23]. The role played by such large graphs in representing the structural skeleton of complex systems—ranging from social to ecological and infrastructural ones—has triggered the production of many indices that try to quantify the different structural characteristics of these networks [23,8]. Among those mathematical approaches used nowadays for studying networks, matrix functions [33] of adjacency matrices of graphs have received an increasing visibility due to their involvement in the so-called *communicability functions* [27,26,16,28,29,34,36,11,32,40,39,5,4,2,14,15]. These functions characterize how much

* Corresponding author.
E-mail address: ernesto.estrada@strath.ac.uk (E. Estrada).

http://dx.doi.org/10.1016/j.jmaa.2016.12.062
0022-247X/

information flows between two different nodes of a graph by accounting for a weighted sum of all the routes connecting them. Here, a route is synonymous with a walk connecting two nodes, which is a sequence of (not necessarily distinct) consecutive vertices and edges in the graph. Then, the communicability function between the nodes p and q is defined by the p, q-entry of the following function of the adjacency matrix (see [27,29] and references therein)

$$G = \sum_{k=0}^{\infty} c_k A^k, \tag{1.1}$$

where the coefficients c_k are responsible of giving more weight to shorter walks. The most popular of these communicability functions is the one derived from the scaling of $c_k = \frac{1}{k!}$, which gives rise to the exponential of the adjacency matrix (see further for definitions). This function, and the graph-theoretic invariants derived from it, have been widely applied in practical problems covering a wide range of areas. Just to mention a few, the communicability function is used for studying real-world brain networks and the effects of diseases on the normal functioning of the human brain [9,10]. On the other hand, the so-called subgraph centrality [28]—a sort of self-communicability of a node in a graph—has been used to detect essential proteins in protein–protein interaction networks [19,17]. The network bipartivity—a measure derived from the use of the self-communicability—has found applications ranging from detection of cracks in granular material [38], to the stability of fullerenes [13], and transportation efficiency of airline networks [24].

A typical question when studying the structural indices derived from (1.1) when using $c_k = \frac{1}{k!}$ is whether or not we are penalizing the longer routes in the graph too heavily (see Preliminaries for formal definitions) [21]. To understand this problem let us consider the communicability function between the nodes p and q in the graph:

$$G_{pq} = \sum_{k=0}^{\infty} c_k \left(A^k\right)_{pq}, \tag{1.2}$$

where $\left(A^k\right)_{pq}$ gives the number of routes of length k between these two nodes. Then, when we use $c_k = \frac{1}{k!}$, a route of length 2 is penalized by $1/2$ and a walk of length 3 is penalized by $1/6$. However, a walk of length 5 is already penalized by $1/120 \approx 0.008$, which could be seen as a very heavy penalization for a relatively short walk between these two nodes. This means that the longer walks connecting two nodes make a little contribution to the communicability function. If we consider the function accounting for the self-returning walks starting (and ending) at a given node G_{pp}, a heavy penalization of longer walks means that this index is mainly dependent on the degree of the corresponding node, i.e., the number of edges incident to it. That is,

$$G_{pp} = 1 + c_2 k_p + \sum_{k=3}^{\infty} c_k \left(A^k\right)_{pp}, \tag{1.3}$$

where k_p is the degree of the node p. Then, the main question here is to study whether using coefficients c_k that do not penalize the longer walks so heavily will reveal some structural information of networks which is important in practical applications of these indices.

Here we consider the use of a double-factorial penalization $1/k!!$ [31] of walks as a way to increase the contribution of longer walks in communicability-based functions for graphs and real-world networks. The goal of this paper is two-fold. First, we want to investigate whether this new penalization of walks produces structural indices that are significantly different from the ones derived from the single-factorial penalization. The other goal is to investigate whether the information contained in longer walks is of significant relevance for describing the structure of graphs and real-world networks. While in the first case we can obtain analytical