Linear Algebra and its Applications 504 (2016) 325–353

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

On graph combinatorics to improve eigenvector-based measures of centrality in directed networks

Argimiro Arratia [a,1], Carlos Marijuán [b,*,2]

[a] *Department of Computer Science/BGSMath, Universitat Politècnica de Catalunya, Barcelona, Spain*
[b] *Departament of Applied Mathematics/IMUVA, Universidad de Valladolid, Valladolid, Spain*

ARTICLE INFO

Article history:
Received 19 October 2014
Accepted 8 April 2016
Available online 18 April 2016
Submitted by R. Brualdi

MSC:
15B51
68R10
05C99
05C40
05C05

Keywords:
Centrality
Eigenvector
PageRank
Topology
Network

ABSTRACT

We present a combinatorial study on the rearrangement of links in the structure of directed networks for the purpose of improving the valuation of a vertex or group of vertices as established by an eigenvector-based centrality measure. We build our topological classification starting from unidirectional rooted trees and up to more complex hierarchical structures such as acyclic digraphs, bidirectional and cyclical rooted trees (obtained by closing cycles on unidirectional trees). We analyze different modifications on the structure of these networks and study their effect on the valuation given by the eigenvector-based scoring functions, with particular focus on α-centrality and PageRank.

* Corresponding author.
E-mail addresses: argimiro@cs.upc.edu (A. Arratia), marijuan@mat.uva.es (C. Marijuán).

[1] Supported by MINECO project APCOM (TIN2014-57226-P), Gen. Cat. SGR2014-890 (MACDA).
[2] Supported by MICINN project SINGACOM (MTM2015-65764-C03-1-P).

http://dx.doi.org/10.1016/j.laa.2016.04.011
0024-3795/

1. Introduction

In the analysis of propagation of ideas and influence through social networks, a much studied optimization problem is the selection of the most authoritative nodes. This is the so-called influence maximization problem, which was first studied in [10], where it is shown NP-hard for several models of social networks and approximation guarantees for efficient solutions are given. A related problem is to model the dynamics of social networks that change in time by modifications on the topology of the network. These topological modifications can significantly alter the hierarchy of influence previously existing in a social network. For example, the situation arises in academic networks, such as *Academia* or *Research Gate*, where participants are often enticed by the administrator of the network to link (or *follow*) others, in order to raise their social presence and consequently their network score, which is computed by a form of centrality measure. Also, in the World Wide Web, the role played by the topology of the internet has been widely recognized as a key factor in the computation and improvement of the scores given by the most used ranking measure, namely PageRank (see [1,8]).

In this paper we address the problem of how the modifications in the link structure of a directed network, whose nodes are ranked by a measure of centrality based on eigenvectors, affect the distribution of values given by this type of scoring function. We propose to do this analysis progressively with respect to the topological complexity of the network. Hence, we present here the case of unidirectional rooted trees, acyclic digraphs through their rooted subtrees, trees with bidirectional arcs and trees extended with cycles, and for all these trees we set as our objective to improve the eigenvector-based centrality score of the root.

We focus our analysis on α-centrality and PageRank scoring functions. PageRank is arguably the most general form of eigenvector-based centrality measure, producing more meaningful scores in directed networks than other centrality measures in its class. As a matter of fact, measures of centrality based on the eigenvectors of the adjacency matrix of directed networks are basically three: *eigenvector centrality* [6], *Katz or alpha-centrality* [7] and PageRank [2,3]. Eigenvector centrality is useless in acyclic digraphs because it assigns a null score to all vertices. In general, a vertex having arcs coming from source nodes (vertices with in-degree zero) obtains a score of zero. More precisely, only vertices in, or connected from, a strongly connected component have positive score. Katz and α-centrality fix the eigenvector scoring limitations by aggregating a term to the scoring function independent of the link structure. This additional term accounts for exogenous sources of information and in this way every vertex gets some non-zero score that can transmit to its neighbors. However, the Katz (and α) centrality score is transmitted uniformly, so that any number of vertices receiving a link from one vertex with high centrality score becomes equally highly central too. This poses an unfair gain of relevance by many individuals in social networks, or pages in the World Wide Web, since it is enough for them to have a highly reputed "sponsor", regardless of their level of popularity quantified by the number of links received. This anomaly is corrected by the