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

Argimiro Arratia[a,2], Carlos Marijuán[b,3]

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

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.

Key words: Centrality, Eigenvector, PageRank, Topology, Network
2010 MSC: 15B51, 68R10, 05C99, 05C40, 05C05

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,

Email addresses: argimiro@cs.upc.edu (Argimiro Arratia), marijuan@mat.uva.es (Carlos Marijuán)

[1]Published in: Linear Algebra and Applications, Vol. 504, 2016, pp. 325–353

[2]Supported by MINECO project APCOM (TIN2014-57226-P), Gen. Cat. SGR2014-890 (MACDA)

[3]Supported by MICINN project SINGACOM (MTM2012-36917-C03-01)

which was first studied in [12], 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 [3], [10]).

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, affects 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* [8], *Katz or alpha-centrality* [9] and PageRank [4, 5]. 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 PageRank centrality measure by dividing the centrality scores inherited from neighbor vertices by their out-degree. We will provide mathematical formulations of all these eigenvector-based centrality measures so that the reader can see how each generalizes the other in a formal way. For a more in-depth exposition of these and other centrality measures see [16].

The paper is organized as follows: In Section 2 we fix the notation to be used for digraphs, present the linear algebraic formulation of each of the eigenvector based directed network centrality measures, and discuss these measures from a perspective of power series. In Section 3 we specify the PageRank formula for directed networks organized as rooted trees. This formula depends solely on the number of vertices at each level of the tree structure, and provides us with a full mathematical justification of the fact that erasing the vertices farthest away from the root improves the PageRank. We then give some rules to optimize the link structure of a web site that stem from our results. Over rooted trees α–centrality coincides with PageRank so all results in this section apply to α–centrality as well. Section 4 presents a through analysis of how the basic combinatorial results of previous section adapt to the α-centrality, as well as PageRank measure, in the more general context of acyclic digraphs. In Section 5, we extend the PageRank and the α–centrality formula obtained in Section 3 to vertices of trees with bidirectional arcs, and cyclical trees (obtained by closing cycles on unidirectional rooted trees), modeling these more complex hierarchical structures by means of infinite unidirectional rooted trees. We also give a vectorial version of these formulas for bidirectional trees. In Section 6, we analyze the behavior of PageRank and α–centrality on bidirectional and cyclical trees when their topology is modified. We give qualitative and quantitative justifications on the consequences of these actions. We close with Section 7 looking at the directed network through its condensation digraph as the acyclic digraph of its strongly connected components, where PageRank can be calculated independently, thus justifying its computation in parallel as suggested by some authors.

2. Preliminaries on eigenvector based centrality measures

By a **digraph** $\mathcal{D}$ we mean a pair $\mathcal{D} = (V, A)$ where V is a finite nonempty set and $A \subset V \times V \setminus \{(v, v) : v \in V\}$. Elements in V and A are called **vertices** and **arcs** respectively. For an arc (u, v) we will say that u is **adjacent to** v, and we also use uv to denote an arc (u, v). The **in-degree** $id(v)$ (**out-degree** $od(v)$) of a vertex v is the number of arcs uv (vu) in A.

A sequence of vertices $v_1 v_2 \dots v_q, q \geq 2$, such that $v_i v_{i+1}$ is an arc for $i = 1, 2, \dots, q-1$ is a **walk** of **length** $q-1$ **joining** v_1 **with** v_q or more simply a v_1–v_q **walk**. If the vertices of $v_1 v_2 \dots v_q$ are distinct the walk is called a **path**. A **cycle of length** q or a q-**cycle** is a path $v_1 v_2 \dots v_q$ closed by the arc $v_q v_1$. A digraph is **acyclic** if it has no cycle.

By a **subdigraph** of the digraph (V, A) we mean a digraph (W, B) such that $W \subset V$ and $B \subset A$. The subdigraph is called a **partial digraph** when $W = V$. The **induced subdigraph** by the digraph (V, A) on $W \subset V$ is the digraph $(W, A/W)$ where $A/W = A \cap (W \times W)$.

Let $\mathbf{M} = (m_{ij})$ be the $N \times N$ adjacency matrix of the digraph $\mathcal{D} = (V, A)$. If $\mathcal{D}$ represents a network (social or informational as WWW) $m_{ij} > 0$ stands for the contribution of vertex v_i to v_j's status (and $v_i v_j \in A$), and so we let $\mathbf{x} = (x_1, \dots, x_N)^t$ be a vector of centrality scores for the elements in V. The **eigenvector centrality** measure assigns to each vertex v_i a proportion of the weighted sum of the centrality of the vertices connected to it:

$$\lambda x_i = m_{1i} x_1 + m_{2i} x_2 + \ldots + m_{Ni} x_N, \quad \text{for } i = 1, \dots, N, \tag{1}$$

or in matrix form

$$\mathbf{M}^t \mathbf{x} = \lambda \mathbf{x} \tag{2}$$

Then, the eigenvector centrality of the network $\mathcal{D}$ is given by the unique nonnegative eigenvector $\mathbf{x}$ associated to the spectral radius $\rho(\mathbf{M}^t)$ of the nonnegative matrix $\mathbf{M}^t$ (by Perron–Frobenius theory). Furthermore, such eigenvector is non null if the corresponding spectral radius $\rho(\mathbf{M}^t)$ is non null, which is equivalent to the existence of a cyclic structure in the digraph $\mathcal{D}$. Consequently, eigenvector centrality is useless in acyclic digraphs.

A more appropriate measure of centrality was introduced by Katz [11] in 1953. Katz considers for each vertex v_i the influence of all the vertices connected by a walk to v_i. Arc connections are penalized by an attenuation factor α, and the contribution of each v_j–v_i walk of length k to the score of vertex v_i is α^k. Taking into account that each element $(\mathbf{M}^k)_{ji}$ of the matrix $\mathbf{M}^k$ gives the number of v_j–v_i walks of length k, Katz centrality assigns to each vertex v_i the score

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (\mathbf{M}^t)_{ji}, \quad \text{for} \quad i = 1, \dots, N \tag{3}$$

Then the Katz centrality vector is given by the column sums of the matrix $\sum_{k=1}^{\infty} \alpha^k (\mathbf{M}^t)^k$. If we set $\mathbf{e} = (1, \ldots, 1)^t$ to be the N-vector of ones, we have the following matrix form for Katz centrality,

$$\mathbf{x} = \left(\sum_{k=1}^{\infty} \alpha^k (\mathbf{M}^t)^k \right) \mathbf{e} = \left(-\mathbf{I} + \sum_{k=0}^{\infty} \alpha^k (\mathbf{M}^t)^k \right) \mathbf{e} \tag{4}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. If α is smaller than the inverse of the spectral radius of $\mathbf{M}^t$, then the series $\sum_{k=0}^{\infty} \alpha^k (\mathbf{M}^t)^k$ converges to $(\mathbf{I} - \alpha \mathbf{M}^t)^{-1}$. In this case Eq. (4) can be expressed as

$$\mathbf{x} = (-\mathbf{I} + (\mathbf{I} - \alpha \mathbf{M}^t)^{-1}) \mathbf{e} \tag{5}$$

Another similar measure of centrality that also resolves the problems encountered by the eigenvector centrality is α*–centrality* (see [9]). This measure consists in adding to the Katz score of each vertex a constant term independent of the connective structure of the network, so that every vertex has a non zero centrality value. In this way,

$$x_i = \alpha \sum_{j=1}^{N} m_{ji} x_j + \beta_i \tag{6}$$

with α, $\beta_i > 0$ for each $i = 1, \ldots, N$, where α is a parameter reflecting the relative importance of endogenous versus exogenous factors in the determination of centrality [9]. In matrix form

$$\mathbf{x} = \alpha \mathbf{M}^t \mathbf{x} + \beta \mathbf{e}$$

and rearranging for $\mathbf{x}$, we have

$$\mathbf{x} = \beta \left(\mathbf{I} - \alpha \mathbf{M}^t \right)^{-1} \mathbf{e} \tag{7}$$

Since we are interested in the relative values of the scores x_i, the factor β is irrelevant, and for convenience we set $\beta = 1$. So we have the α**–centrality** measure given by

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{M}^t)^{-1} \mathbf{e} \tag{8}$$

Thus, α-centrality is a simple translation of Katz centrality. In the remainder of this paper we will refer to both measures as α-centrality.

The next step to extend Eq. (8) to a more fair centrality measure that distributes the centrality of a node among its neighbors in proportion to their number, is to consider the contribution of centrality from each node divided by its out-degree. In mathematical terms, the centrality x_i of vertex v_i is formalized as

$$x_i = \beta + \alpha \sum_{j=1}^{N} m_{ji} \frac{x_j}{od(v_j)} \tag{9}$$

where $\beta > 0$ is some constant. This is the general form of the PageRank scoring function which originally sets $\beta = (1-\alpha)/N$ and α is a constant in the real interval $(0,1)$ which is usually set to 0.85 [4, 5]. Additionally it is assumed that all nodes have at least out-degree 1 in order to avoid indeterminate terms in the sum. We use $\mathcal{P}(v_i)$ to denote the original PageRank centrality measure for a vertex v_i. Thus, Eq. (9) becomes

$$\mathcal{P}(v_i) = \frac{1-\alpha}{N} + \alpha \sum_{j=1}^{N} m_{ji} \frac{\mathcal{P}(v_j)}{od(v_j)} \tag{10}$$

where $m_{ji} = 1$ iff $v_i v_j \in A$ or 0 otherwise. In matrix form

$$\mathbf{p} = \frac{(1-\alpha)}{N} (I - \alpha \mathbf{M}^t \mathbf{D}^{-1})^{-1} \mathbf{e} \tag{11}$$

where $\mathbf{D}$ is the diagonal matrix with $D_{ii} = \max[od(v_i), 1]$, and $\mathbf{p} = (\mathcal{P}(v_1), \ldots, \mathcal{P}(v_N))^t$ is the **PageRank vector** of $\mathcal{D}$. This $\mathbf{p}$ is a probability vector (the sum over j on the right side of Eq. (10) is one, for all $i = 1, \ldots, N$) and, in fact, is the positive dominant eigenvector of the *transition* matrix $T = \frac{(1-\alpha)}{N} J + \alpha \mathbf{M}^t \mathbf{D}^{-1}$ (where J is the $N \times N$ matrix of 1's), associated to the greatest positive eigenvalue λ and is the solution to the system of equations given by (10) for all possible vertices of $\mathcal{D}$. Since the digraph $\mathcal{D}$ has no **sinks** (i.e. vertices with out-degree 0), the matrix T is stochastic by columns with dominant eigenvalue $\lambda = 1$, and the PageRank of a node v can be interpreted as the probability of a user reaching v directly (with probability $(1-\alpha)/N$) or after following all appropriate links, each with probability α. On the other hand, if there are some sinks, then T is not stochastic, and so the proposed method to make it stochastic is to connect every sink with all the vertices of the digraph, including the sink itself, which amounts to normalizing to 1 the vector $\mathbf{p}$ by simply dividing all its components by their overall sum. This 1-normalization is necessary to compare the PageRank of vertices in different websites under the same metric conditions. The PageRank vector $\mathbf{p}$ is computed by iterative methods based on the power method where fast convergence is guaranteed by the domination of the spectral radius of T, and the convergence speed is given by the second eigenvalue of T. For an in-depth exposition of PageRank and the related linear algebra methods see [13] and references therein.

Remark 1. *Obviously, by the definitions, α–centrality and PageRank are equivalent measures of centrality on digraphs without vertices of out-degree greater than one. For this class of digraphs, formula (11) with $D = I$ coincides with formula (7) for $\beta = (1-\alpha)/N$.*

Centrality measures as power series. Yet another view of PageRank is the analytical formulation given by Brinkmeier (see [6]), who conceived the PageRank function as a power series. In this setting, a formula is given that highlights the fact that the ranking of a vertex v, as assigned by PageRank, depends on the weighted contributions of each vertex in every walk that leads into v, being these contributions higher in value for vertices that are nearer in distance from v.

For a given walk $\rho = v_1 v_2 \dots v_n$ in the graph (V, A), define the **branching factor** of ρ by the formula

$$D(\rho) = \frac{1}{od(v_1)od(v_2)\cdots od(v_{n-1})} \tag{12}$$

Then, for any vertex $a \in V$, we have

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{w \in V} \sum_{\rho : w \longrightarrow a} \alpha^{l(\rho)} D(\rho) \tag{13}$$

where $\rho : w \longrightarrow a$ denotes a walk ρ joining a vertex w with a, and $l(\rho)$ is the length of ρ. We should remark that for effectively computing $\mathcal{P}(a)$ by the power series in (13), Brinkmeier implements a breadth-first search strategy where the inner sum is taken over all walks ending in a of a fixed length, and this is done for all possible lengths; that is,

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{l \geq 0} \sum_{\rho : w \xrightarrow{l} a} \alpha^{l} D(\rho) \tag{14}$$

where $\rho : w \xrightarrow{l} a$ denotes a walk ρ from any w to a of length l.

For α-centrality $D(\rho) = 1$ for all walks ρ, and formulas (13) and (14) become for each vertex v_i

$$x_i = \beta \sum_{w \in V} \sum_{\rho : w \longrightarrow v_i} \alpha^{l(\rho)} = \beta \sum_{l \geq 0} \sum_{\rho : w \xrightarrow{l} v_i} \alpha^{l}$$

with $\beta = 1$ as in Eq. (8) or $\beta = (1-\alpha)/N$. From now on we use $\beta = (1-\alpha)/N$ for a precise comparison of α–centrality with PageRank.

3. Rearranging the structure of rooted trees

Our starting case study is the set of *rooted trees*, where a **tree with root** r is an acyclic digraph with a maximal vertex r, such that for every vertex $v \neq r$ there is a unique v–r path. Vertices with in–degree 0 are called **leaves**. The root is the targeted page for improving its α-centrality or PageRank valuation. The **height of a vertex** in a rooted tree is the length of the path from the vertex to the root. The **level** N_k is the set of vertices with height k; the root is at level N_0. The **height of a rooted tree** is the length of the longest path from a leaf to the root.

Rooted trees belong to the class of digraphs without vertices of out-degree greater than one and so all the results of this Section expressed in terms of PageRank are also valid in the same way for α–centrality (cf. Remark 1).

Remark 2. *Since we are interested in studying the behavior of PageRank when localized in certain subdigraphs of the Web digraph, we think, in particular, of our trees as local closed web sites. This means that the value of N in formula (10) is the number of vertices in the tree.* □

To compute the PageRank of the root r of a tree all we need to do is count the number of vertices at each level of the tree. For each vertex w there is a unique w–r path of length k if $w \in N_k$, then by Eq. (13) we have:

Theorem 3. *If a rooted tree has N vertices and height h, then the PageRank of its root r is given by the formula*

$$\mathcal{P}(r) = \frac{1-\alpha}{N} \sum_{k=0}^{h} \alpha^k n_k \tag{15}$$

where $n_k := |N_k|$ is the number of vertices of the kth–level, N_k, of the tree. □

Remark 4. *Theorem 3 shows that we can do any rearrangements of links between two consecutive levels of a web set up as a rooted tree, and the PageRank of the root will be the same.* □

Remark 5. *Due to Theorem 3, we will from now on describe a rooted tree $\mathcal{T}^r$, with root r and $h \geq 0$ levels, each of cardinality $n_0 = 1, n_1, \ldots, n_h$, as the string $\mathcal{T}^r = 1n_1 \ldots n_h$. Also the PageRank for the root r of $\mathcal{T}^r$, or for any other vertex seemed as the root of a subtree in $\mathcal{T}^r$, will only depend on the height and the number of vertices at each level of $\mathcal{T}^r$.* □

The following result shows that erasing vertices farthest away from the root improves the PageRank. This corroborates the known fact that the optimal configuration is a **star**, i.e. a rooted tree of height 1 (see e.g. [13]).

Theorem 6. *If in a tree $\mathcal{T}^r = 1n_1 \dots n_h$ we have that p vertices, $1 \leq p \leq n_h$, of the last level N_h are erased, then the PageRank of its root r, $\mathcal{P}(r)$, increases its value.*

PROOF. After passing from the tree $\mathcal{T}^r = 1n_1 \dots n_h$, with $N = 1 + n_1 + \dots + n_h$ vertices and PageRank $\mathcal{P}(r)$, to the tree $\mathcal{T}'^r = 1n_1 \dots n_{h-1}(n_h - p)$ with $N - p$ vertices and PageRank $\mathcal{P}'(r)$, we get

$$\begin{aligned}
\mathcal{P}'(r) \quad - \quad \mathcal{P}(r) \; &= \frac{(1-\alpha)p}{(N-p)N}(1 + n_1\alpha + \dots + n_{h-1}\alpha^{h-1} - (N - n_h)\alpha^h) \\
&= \frac{(1-\alpha)p}{(N-p)N}(1 + n_1\alpha + \dots + n_{h-1}\alpha^{h-1} - (1 + n_1 + \dots + n_{h-1})\alpha^h) \\
&= \frac{(1-\alpha)p}{(N-p)N}((1-\alpha^h) + n_1(\alpha - \alpha^h) + \dots + n_{h-1}(\alpha^{h-1} - \alpha^h)) > 0
\end{aligned}$$

because $0 < \alpha < 1$ and $h \geq 1$. $\square$

Remark 7. *Thus, in order to improve the PageRank of the root of a tree one can delete as many vertices from highest level to lowest, as the context permits. Conversely, if a new level of vertices is added to a tree, then the PageRank of its root decreases.* $\square$

Remark 8. *The previous result holds in absolute terms, i.e., not disregarding the existence of a sink in the tree. Since in practice one needs to normalize the PageRank vector to guarantee the stochastic properties of the transition matrix ruling the system, and to compare the PageRanks of the pages in different trees under the same metric conditions, we should establish the truth of Theorem 6 for the 1-normalized version of PageRank.* $\square$

Proof of normalized version of Theorem 6. The key observation is that the *PageRank at level N_k*, understood as the sum of all PageRank of vertices at level N_k and which we denote as $\mathcal{P}(N_k)$, only depends of the quantities n_k, n_{k+1}, ..., n_h (see Remark 5). We have

$$\begin{aligned}
\mathcal{P}(N_1) &= \frac{1-\alpha}{N}(n_1 + n_2\alpha + n_3\alpha^2 + \ldots + n_{h-1}\alpha^{h-2} + n_h\alpha^{h-1}) \\
\mathcal{P}(N_2) &= \frac{1-\alpha}{N}(n_2 + n_3\alpha + \ldots + n_{h-1}\alpha^{h-3} + n_h\alpha^{h-2}) \\
\vdots \quad &\quad \vdots \\
\mathcal{P}(N_{h-1}) &= \frac{1-\alpha}{N}(n_{h-1} + n_h\alpha) \\
\mathcal{P}(N_h) &= \frac{1-\alpha}{N}n_h
\end{aligned}$$

The PageRank at level N_0 is exactly $\mathcal{P}(r)$. The sum of all levels' PageRank is then

$$\begin{aligned}
\mathcal{P}(\mathcal{T}^r) \;:=\; & \frac{1-\alpha}{N}(N + (N-1)\alpha + (N-1-n_1)\alpha^2 \; + \\
& (N-1-n_1-n_2)\alpha^3 + \ldots + (n_{h-1}+n_h)\alpha^{h-1} + n_h\alpha^h)
\end{aligned}$$

and the normalization of $\mathcal{P}(r)$ is obtained by the quotient

$$\frac{\mathcal{P}(r)}{\mathcal{P}(\mathcal{T}^r)} = \frac{1 + n_1\alpha + n_2\alpha^2 + \ldots + n_{h-1}\alpha^{h-1} + n_h\alpha^h}{N + (N-1)\alpha + (N-1-n_1)\alpha^2 + \ldots + (n_{h-1}+n_h)\alpha^{h-1} + n_h\alpha^h}$$

If p vertices, $1 \leq p \leq n_h$, are removed from the last level N_h of $\mathcal{T}^r$, then the normalized PageRank of r in the pruned tree $\mathcal{T}'^r$ is

$$\frac{\mathcal{P}'(r)}{\mathcal{P}(\mathcal{T}'^r)} = \frac{1 + n_1\alpha + n_2\alpha^2 + \ldots + n_{h-1}\alpha^{h-1} + (n_h-p)\alpha^h}{N' + (N'-1)\alpha + (N'-1-n_1)\alpha^2 + \ldots + (n_h-p)\alpha^h}$$

where $N' = N - p$. Therefore,

$$\frac{\mathcal{P}'(r)}{\mathcal{P}(\mathcal{T}'^r)} \geq \frac{\mathcal{P}(r)}{\mathcal{P}(\mathcal{T}^r)} \iff \mathcal{P}'(r)\mathcal{P}(\mathcal{T}^r) \geq \mathcal{P}(r)\mathcal{P}(\mathcal{T}'^r)$$

Note that both terms in the last inequality are polynomials in α of degree $2h$. Then the inequality holds because the coefficients accompanying α^k, for $k < h$, are greater in $\mathcal{P}'(r)\mathcal{P}(\mathcal{T}^r)$ than in $\mathcal{P}(r)\mathcal{P}(\mathcal{T}'^r)$, and for $k \geq h$ the corresponding coefficients are equal. □

Erasing p leaves from any other level N_k distinct from the last level N_h can either increase or decrease the PageRank of the root. Hence, doing an unorderly pruning has mixed consequences to PageRank, as the following example shows.

Example 9. *Let the tree $\mathcal{T}^r = 1n_1n_2$ with PageRank $\mathcal{P}(r)$. Then consider removing p leaves from the level N_1, with $1 \le p \le n_1 - 1$. The resulting pruned tree $\mathcal{T}'^r = 1(n_1 - p)n_2$ has PageRank $\mathcal{P}'(r)$ and we have that*

$$\mathcal{P}'(r) - \mathcal{P}(r) = \frac{(1-\alpha)^2 p}{(N-p)N}(1 - n_2\alpha)$$

which is positive for $n_2 = 1$ and $\alpha \in (0, 1)$, and negative for any $\alpha > 1/n_2$.

This reduction of the PageRank of the root of this tree also holds in relative terms. From the normalized version of Theorem 6 we have

$$\frac{\mathcal{P}(r)}{\mathcal{P}(\mathcal{T}^r)} = \frac{1 + n_1\alpha + n_2\alpha^2}{1 + n_1 + n_2 + (n_1 + n_2)\alpha + n_2\alpha^2}$$

and

$$\frac{\mathcal{P}'(r)}{\mathcal{P}(\mathcal{T}'^r)} = \frac{1 + (n_1 - p)\alpha + n_2\alpha^2}{1 + n_1 - p + n_2 + (n_1 - p + n_2)\alpha + n_2\alpha^2}$$

so that

$$\mathcal{P}'(r)\mathcal{P}(\mathcal{T}^r) - \mathcal{P}(r)\mathcal{P}(\mathcal{T}'^r) > 0 \iff p(1 - n_2\alpha) > 0$$

Hence, for this deletion of leaves at intermediate level, the relative variation of PageRank of the root is equivalent to its absolute variation. □

If it were the case that for practical, or any other reason, we were obliged to keep certain height, then a natural question is how much can we prune the tree to improve on PageRank. The extreme situation is to prune all but one arc at each level, so we take that structure as benchmark.

Theorem 10. *The PageRank of the root of the tree $\mathcal{T}^r = 1n_1 \dots n_h$ is smaller than the PageRank of the root of the tree*

$$\mathcal{T}_q^r := 1n_1 \dots n_{\lfloor \frac{h-1}{2} \rfloor} \underbrace{1 \dots 1}_{\lfloor h/2 \rfloor + 1}$$

The tree $\mathcal{T}_q^r$ is called **queue tree**.

Proof. We proceed recursively from the last level down to $\lfloor (h-1)/2 \rfloor$.
(a) The PageRank $\mathcal{P}(r)$ of the root r of $\mathcal{T}^r = 1n_1 \dots n_{h-1}n_h$ is smaller than the PageRank $\mathcal{P}'(r)$ of $\mathcal{T}'^r = 1n_1 \dots n_{h-1}1$. Indeed, let $N = 1+n_1+\dots+n_h$,

then

$$\begin{aligned}
\mathcal{P}'(r) &- \mathcal{P}(r) = \frac{(n_h-1)(1-\alpha)}{(N-(n_h-1))N}\left(\sum_{k=0}^{h-1} n_k\alpha^k - (N-n_h)\alpha^h\right) \\
&= \frac{(n_h-1)(1-\alpha)}{(N-(n_h-1))N}\sum_{k=0}^{h-1} n_k(\alpha^k-\alpha^h) > 0
\end{aligned}$$

Apply the same methodology for $\mathcal{T}^r = 1n_1\ \dots n_{h-2}n_{h-1}1$ and $\mathcal{T}'^r = 1n_1 \dots n_{h-2}11$, and so on, up to $\lfloor h/2 \rfloor$. At this last step we have
(*b*) $\mathcal{T}^r = 1n_1 \dots n_{\lfloor \frac{h-1}{2} \rfloor} n_{\lfloor \frac{h+1}{2} \rfloor} \underbrace{1\dots1}_{\lfloor h/2 \rfloor}$, and we shall see that its PageRank is less than that of the queue tree $\mathcal{T}'^r = 1n_1 \dots n_{\lfloor \frac{h-1}{2} \rfloor} \underbrace{1\dots1}_{\lfloor h/2 \rfloor + 1}$. We work separately the cases of h even or h odd.
(*b.i*) If $h = 2p-1$ then $\mathcal{T}^r = 1n_1 \dots n_{p-1}n_p \underbrace{1\dots1}_{p-1}$, $\mathcal{T}'^r = 1n_1 \dots n_{p-1} \underbrace{1\dots1}_{p}$ and $N = n_1 + \dots + n_p + p$. Let $M = \frac{(n_p-1)(1-\alpha)}{(N-(n_p-1))N}$. Then

$$\begin{aligned}
\mathcal{P}'(r) &- \mathcal{P}(r) = M\left(1 + \sum_{k=1}^{p-1} n_k\alpha^k - (N-n_p)\alpha^p + \sum_{k=p+1}^{2p-1}\alpha^k\right) \\
&= M\left((1-\alpha^p) + \sum_{k=1}^{p-1} n_k(\alpha^k-\alpha^p) + \sum_{k=p+1}^{2p-1}(\alpha^k-\alpha^p)\right) \\
&= M\left((1-\alpha^p) + \sum_{k=1}^{p-1}(n_k-\alpha^{p-k})(\alpha^k-\alpha^p)\right) > 0
\end{aligned}$$

(*b.ii*) If $h = 2p$ then $\mathcal{T}^r = 1n_1 \dots n_{p-1}n_p \underbrace{1\dots1}_{p}$, $\mathcal{T}'^r = 1n_1 \dots n_{p-1} \underbrace{1\dots1}_{p+1}$ and $N = n_1 + \dots + n_p + p + 1$. One then shows $\mathcal{P}'(r) - \mathcal{P}(r) > 0$ by a similar argument as in (*b.i*). $\square$

Remark 11. *Theorem 10 can not be improved, in the sense that deleting further vertices (but keeping the height) in a queue tree may or may not improve the PageRank of the root. For small values of h, the queue tree is the optimal pruning of a tree for increasing PageRank. For example, if $h = 4$ the corresponding queue tree is $\mathcal{T}_q^r = 1n_1111$ with PageRank $\mathcal{P}(r)$, and if $n_1 > 1$ and we remove a vertex from level N_1, we get the tree $\mathcal{T}'^r = 1(n_1-1)111$*

with PageRank $\mathcal{P}'(h)$, and their difference is

$$\mathcal{P}'(r) - \mathcal{P}(r) = \frac{1-\alpha}{(n_1+3)(n_1+4)}(1 - 4\alpha + \alpha^2 + \alpha^3 + \alpha^4) < 0$$

for any α such that $0.27568 < \alpha < 1$. For larger values of h, an improvement of PageRank will depend on α and on the cardinalities of the levels N_1, ..., $N_{\lfloor \frac{h-1}{2} \rfloor}$. □

A theoretically as well as commercially important problem is to find a scheme for modifying the link structure of a local web in order to improve its ranking, as set by PageRank or any other ranking function. In this section we have presented the case of a network with a tree–like structure, where the PageRank of the main page, located at the root of the tree, should have the highest possible value, but at the same time the overall structure of the web should satisfy certain conditions given by the context. We shall not make precise the details of the context, which are surely determined by the general conditions imposed by design. Let us refer to the context as Π. By virtue of Theorem 3 this translates into the following optimization problem. *Main Objective*: Given a certain context Π, to maximize the function

$$\mathcal{P}(h, n_1, \ldots, n_h) = \frac{1-\alpha}{1 + n_1 + \ldots + n_h} \sum_{k=0}^{h} \alpha^k n_k$$

for fixed α, such that $0 < \alpha < 1$, and all trees $\mathcal{T}^r = 1n_1 \ldots n_h$ with integer values $h, n_i \geq 1$, $1 \leq i \leq h$. If the total number N of vertices is bounded then we can assure that the maximum exists. The complexity of the problem depends mostly on the conditions imposed by the context Π. This justifies approaching the solution through heuristics. Here we give an ad hoc list of rules that clearly stem from our theorems.

Rule 1: Due to Theorem 6, the first action to take is to reduce the height as much as the context allows.

Rule 2: Keep in mind that while applying Rule 1 (and deleting levels), links between consecutive levels can be rearrange in any way you like, as long as the context is kept consistent, and this has no effect on the root's PageRank value (by Theorem 3).

Rule 3: Once the optimal height $h > 1$ is attained[4], we delete (as much as possible) vertices from levels in the upper half of the tree, trying to

[4]Optimality here again depends on maintaining the context consistent. This height could mean the minimal levels of a hierarchy that we need to reflect in the web site; say, for example, of a corporation or a hypertext.

get it close to its underlying queue tree (Theorem 10), and those vertices that cannot be deleted should be moved as closer to level 1 as possible (by Theorem 3).

4. Acyclic digraphs

We continue in this section with an analysis of the extend to which the previous results hold for the different centrality measures in a general acyclic digraph.

For an acyclic digraph (V, A) there exists at least one vertex v with $od(v) = 0$. Such vertices are called **maximals** or sinks in the digraph. The set of maximal vertices of (V, A) will be denoted by M, and a path $v_1 v_2 \dots v_q$ with $v_q \in M$ will be called **path with maximal end**.

Moreover, the vertices in the acyclic digraph (V, A) can be distributed by **levels** $N_0, N_1, \dots$, where $N_0 = M$ and, recursively for $p > 0$,

$$N_p = \{v \in V \backslash \bigcup_{i=0}^{p-1} N_i : v \text{ is maximal in the induced subdigraph on } V \backslash \bigcup_{i=0}^{p-1} N_i\}$$

Thus one has a partition of V, $V = N_0 \cup N_1 \cup \dots \cup N_h$, h being the **height of the acyclic digraph**, i.e. the last index such that $N_h \neq \emptyset$.

The **closure**[5] of a vertex v in an acyclic digraph (V, A) is the set of vertices

$$\bar{v} = \{u : \text{there is a path from } u \text{ to } v\} \cup \{v\}$$

Clearly, the union of the closures of the maximal vertices of the acyclic digraph (V, A) covers the set of vertices V, and the arcs of its induced subdigraphs cover the set of arcs A:

$$\bigcup_{m \in M} \bar{m} = V, \qquad \bigcup_{m \in M} A/\bar{m} = A.$$

A **labeling** of a digraph (V, A) by the label set E is a bijective map $\underline{v} : E \longrightarrow V$, $\underline{v}(e)$ being denoted by v_e for any $e \in E$. If $(E, \leq)$ is a totally ordered set, the digraph (V, A) is said to be E-**ordered** if it is labeled by E in such a way that if $(v_i, v_j) \in A$ then $i < j$. We have the following characterization of acyclic digraphs [14]: A digraph (V, A), with $card(V) = N$, is acyclic if and only if (V, A) is E-ordered by the set $E = \{1, \dots, N\}$.

[5]The set $\{\bar{v} : v \in V\}$ can be taken as a sub-basis of closed sets for a topology over the set of vertices V (see [15]).

Definition 12. *We call* **forest of paths** *associated to an acyclic digraph* (V, A) *to the digraph* $(\tilde{V}, \tilde{A})$ *where* $\tilde{V}$ *is the set of paths with maximal end of* (V, A) *together with the set of maximal points* M *and* $\tilde{A} = \{(p, q) \in \tilde{V} \times \tilde{V} : p \notin M$ *and* q *is the path obtained from* p *by deleting the first element*$\}$.

If the acyclic digraph (V, A) is labeled by means of $E = \{1, \ldots, N\}$ by the bijection $\underline{v} : E \longrightarrow V$, then the forest of paths digraph $(\tilde{V}, \tilde{A})$ will be considered labeled by $\tilde{E} = \{K \subset E : \underline{v}_K \in \tilde{V}\}$ by means of the bijection $\tilde{v} : \tilde{E} \longrightarrow \tilde{V}$ given by $\tilde{v}(K) = v_{i_1} \ldots v_{i_q} \in \tilde{V}$, where $\underline{v}_K$ denotes the restriction of $\underline{v}$ to the naturally ordered set $K = \{i_1 < \ldots < i_q\}$. Moreover, if the acyclic digraph (V, A) is E-ordered by $E = \{1, \ldots, N\}$, then $\tilde{E}$ is totally ordered by the lexicographic ordering and therefore, the forest of paths digraph $(\tilde{V}, \tilde{A})$ is an $\tilde{E}$-ordered digraph.

Proposition 13 (cf. [14]). *The forest of paths digraph* $(\tilde{V}, \tilde{A})$ *of an acyclic digraph* (V, A) *is a forest with* m *rooted trees, where* $m = Card(M)$. □

Remark 14. *Each non maximal vertex of the acyclic digraph* (V, A) *gives rise to a new vertex in the forest of paths* $(\tilde{V}, \tilde{A})$ *for each one of the paths with maximal end in* (V, A) *starting from it. And each vertex in the forest of paths is in the level* N_k *of this forest if and only if* k *is the length of the corresponding path.*

Proposition 15 (cf. [14]). *The labels of the leaves in the forest of paths* $(\tilde{V}, \tilde{A})$ *describe the digraph from the start* (V, A).

Proof. It is clear that $v_i \in V$ if and only if v_i is in the label of some leaf of $(\tilde{V}, \tilde{A})$ and $(v_i, v_j) \in A$ if, and only if, v_i and v_j are consecutive (in that order) in the label of some leaf in $(\tilde{V}, \tilde{A})$. If we denote the set of leaves of $(\tilde{V}, \tilde{A})$ by L, then we have:

$$V = \bigcup_{v \in L} \{v_i : v_i \text{ being in the label of } v \in L \subset \tilde{V}\} \text{ and}$$

$$A = \{(v_i, v_j) : \exists v = v_{j_i} ... v_{j_r} v_{j_{r+1}} \ldots v_{j_q} \in L \text{ in } (\tilde{V}, \tilde{A}), i = j_r, j = j_{r+1}\} \quad \square$$

Now we show how a labeling in an acyclic digraph (V, A) induces a "prelabeling" in its forest of paths $(\tilde{V}, \tilde{A})$ which can be done by using the same label set and in such a way as to enable the recovery of the original structure of (V, A).

Definition 16. *A* **prelabeling** *on a forest* (Z, H) *by the prelabel set* E *is a surjective map* $p : Z \longrightarrow E$. *For every* $x \in Z$, $p(x)$ *is the* **prelabel** *of* x.

Proposition 17. *Let (V, A) be an acyclic digraph labeled by $E = \{1, \dots, N\}$, with $\underline{x}(i) = x_i$, and let $(\tilde{V}, \tilde{A})$ be its forest of paths labeled by $\tilde{E}$. The following properties hold:*

1. *The map $p : \tilde{V} \longrightarrow E$ such that $p(x_{i_1} \dots x_{i_q}) = i_1$ for any $x_{i_1} \dots x_{i_q} \in \tilde{V}$, is a prelabeling on $(\tilde{V}, \tilde{A})$ by E.*
2. *If K is the arc set given by:*

   $$K = \{(i, j) \in E \times E : \text{ there exists } (x, y) \in \tilde{A} \text{ with } p(x) = i \text{ and } p(y) = j\}$$

   then the labeling bijection $p : E \longrightarrow V$ is a digraph isomorphism between (E, K) and (V, A).
3. *The acyclic digraph (V, A) labeled by E can be recovered from its forest of paths $(\tilde{V}, \tilde{A})$ prelabeled by E, being V the set of prelabels of $\tilde{V}$ and $(u, v) \in A$ if and only if $(u, v) \in \tilde{A}$.*

Remark 18. *The vertices of $\tilde{V}$ with the same prelabel i have the same closure $\bar{i}$, and the induced subdigraph by the forest $(\tilde{V}, \tilde{A})$ in $\bar{i}$ is a rooted subtree with root i.*

Figure 1 includes, on the left, an E-ordered acyclic digraph (V, A) by $E = \{1, \dots, 7\}$, with two maximal vertices; in the middle, its $\tilde{E}$-ordered forest of paths $(\tilde{V}, \tilde{A})$ by $\tilde{E}$, the set of paths with maximal end; and, on the right, the forest of paths $(\tilde{V}, \tilde{A})$ prelabeled by E.

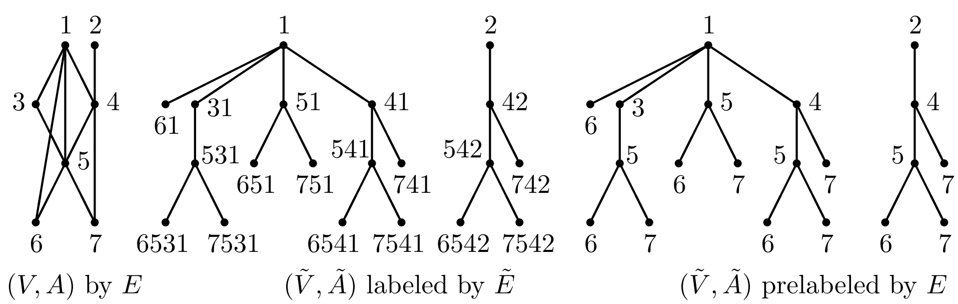

Figure 1: An acyclic digraph and its labeled and prelabeled forest of paths

The α–centrality and PageRank vectors of an acyclic digraph can be obtained from its associated forest of paths. We then have the following extension of Theorem 3 to acyclic digraphs.

Theorem 19. *Let (V, A) be an acyclic digraph E-ordered by $E = \{1, \dots, N\}$ and let $(\tilde{V}, \tilde{A})$ be its forest of paths prelabeled by E. Then*

1. *The α–centrality vector of (V, A) is*

$$\mathbf{x} = (x_1, \ldots, x_N) = \frac{1-\alpha}{N}(y_1, \ldots, y_N), \text{ with } y_i = \sum_{k=0}^{i} n_{i_k} \alpha^k \tag{16}$$

   where n_{i_k} is the number of vertices of the level N_{i_k} in the rooted subtree of the forest $(\tilde{V}, \tilde{A})$ with root i (the induced sub-digraph in $\bar{i}$).

2. *The normalized α-centrality vector of (V, A) is*

$$\mathbf{x}^1 = \frac{1}{\sum_{i=1}^{N} y_i}(y_1, \ldots, y_N)$$

3. *The PageRank vector of (V, A) is*

$$\mathbf{p} = (\mathcal{P}_1, \ldots, \mathcal{P}_N) = \frac{1-\alpha}{N}(q_1, \ldots, q_N), \text{ with } q_i = \sum_{k=0}^{i} b_{i_k} \alpha^k \tag{17}$$

   where b_{i_k} is the sum of the vertex's branching of the level N_{i_k} in the rooted subtree of $(\tilde{V}, \tilde{A})$ with root prelabeled by i.

4. *The normalized PageRank vector of (V, A) is*

$$\mathbf{p}^1 = \frac{1}{\sum_{i=1}^{N} q_i}(q_1, \ldots, q_N)$$

Proof. The α-centrality measures the contribution of the paths to each vertex, and this is what has been modelled with the forest of paths. This shows 1. For each vertex labeled by i we consider the induced subdigraph by the forest of paths on the closure of i, and where each vertex is weighted by the branching of the corresponding subpath. Then statement 3 follows by Remark 14 and Proposition 15. Statements 2 and 4 are the 1-normalization of the corresponding formulas. $\square$

Theorem 19 gives us another way of computing the α-centrality and PageRank measures for acyclic digraphs.

Example 20. *Let us compute the α-centrality and PageRank of the acyclic digraph shown in Figure 1. Using formula (7) with adjacency matrix $\mathbf{M} = (m_{ij})_{1\le i,j\le 7}$, with non null elements $m_{31} = m_{41} = m_{42} = m_{51} = m_{53} = m_{54} = m_{61} = m_{65} = m_{74} = m_{75} = 1$, we obtain the α-centrality*

$$\begin{aligned}\mathbf{x} &= \frac{1-\alpha}{7}\big(1 + 4\alpha + 5\alpha^2 + 4\alpha^3, 1 + \alpha + 2\alpha^2 + 2\alpha^3, \\ &\qquad 1 + \alpha + 2\alpha^2, 1 + 2\alpha + 2\alpha^2, 1 + 2\alpha, 1, 1\big)\end{aligned}$$

One can see that this expression coincides with that given by formula (16).
Now compute the PageRank using formula (11) to get

$$\begin{aligned}\mathbf{p} &= \frac{1-\alpha}{7}\Bigg(1+\frac{7}{3}\alpha+\frac{13}{12}\alpha^2+\frac{1}{2}\alpha^3, 1+\frac{1}{2}\alpha+\frac{5}{12}\alpha^2+\frac{1}{6}\alpha^3,\\ &\qquad 1+\frac{1}{3}\alpha+\frac{1}{3}\alpha^2, 1+\frac{5}{6}\alpha+\frac{1}{3}\alpha^2, 1+\alpha, 1, 1\Bigg)\end{aligned}$$

Once again this expression can be easily obtained using formula (17). For instance, to compute the PageRank of the vertex 1 in the acyclic digraph (first term in the vector above) we look in the induced subgraph by the forest of paths on the closure of 1, and obtain the value

$$\mathcal{P}_1 = \frac{1-\alpha}{7}\left(1+\frac{7}{3}\alpha+\frac{13}{12}\alpha^2+\frac{1}{2}\alpha^3\right)$$

where $\frac{7}{3}$, $\frac{13}{12}$ *and* $\frac{1}{2}$ *are each equal to the sum of the branching of the vertices of its corresponding level, namely,* N_1, N_2 *and* N_3, *in the forest of paths* $(\tilde{V}, \tilde{A})$. *Similarly the PageRank of vertex 4 is*

$$\mathcal{P}_4 = \frac{1-\alpha}{7}\left(1+\frac{5}{6}\alpha+\frac{1}{3}\alpha^2\right)$$

where the coefficients are obtained from the branching of the levels in the induced subgraph by the forest of paths on the closure of 4. Figure 2 shows the induced subdigraph in the closure of the vertices 1, 2, 3, 4 and 5, of the prelabeled forest of paths in Figure 1.

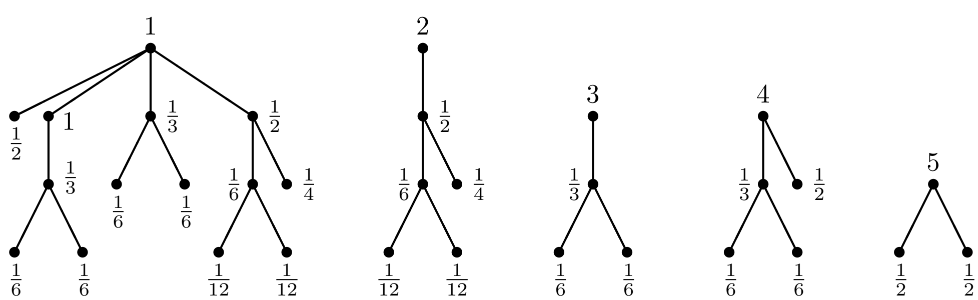

Figure 2: Rooted subtrees of $(\tilde{V}, \tilde{A})$ with branching

The previous Theorem shows that Remark 4 holds for α–centrality as well as PageRank (provided one considers the sum b_k of all branchings of vertices at level k) on acyclic digraphs; that is, one can do any rearrangement of links between consecutive levels of the network (being acyclic digraph) and the α–centrality or PageRank measure will be the same. Remark 5 can also be updated in this context of acyclic digraphs considering b_k instead of n_k.

However, Theorem 6 does not hold for either α-centrality or PageRank applied to measuring centrality of maximal vertices in acyclic digraphs. We give an example below where it occurs that removing vertices with out-degree greater than one from the last level can either increase or decrease the values of the α-centrality and/or PageRank measures of the root of the acyclic digraph. In Figure 3 we have, on the left, an acyclic digraph with $2k+3$ vertices, k vertices from the last level (level N_3) connected to the vertex 2. In the middle, we have the forest of paths with vertices weighted by their branching. The right-most figure shows the reduced digraph obtained by removing the k vertices of degree 2 at level N_3.

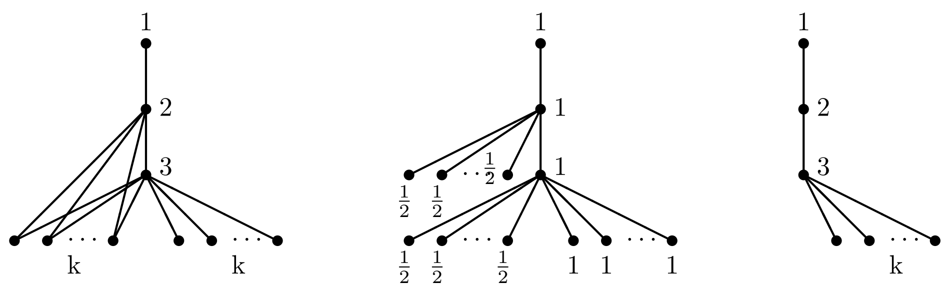

Figure 3: Acyclic digraph, its forest of paths and reduced digraph

With the notations of Theorem 19, for the α–centrality of the acyclic digraph in Figure 3 we have

$$\begin{aligned} y_1 &= 1+\alpha+(k+1)\alpha^2+2k\alpha^3, \;\; y_2 = 1+(k+1)\alpha+2k\alpha^2, \\ y_3 &= 1+2k\alpha, \;\; y_i = 1, \text{ for } i = 4, \ldots, 2k+3, \quad \text{and} \end{aligned}$$

$$\sum_{i=1}^{2k+3} y_i = 2k+3+(3k+2)\alpha+(3k+1)\alpha^2+2k\alpha^3$$

and for the acyclic digraph without the k vertices of outdegree 2 we have

$$\begin{aligned} y'_1 &= 1+\alpha+\alpha^2+k\alpha^3, \;\; y'_2 = 1+\alpha+k\alpha^2, \\ y'_3 &= 1+k\alpha, \;\; y'_i = 1, \text{ for } i = 4, \ldots, k+3, \quad \text{and} \end{aligned}$$

$$\sum_{i=1}^{k+3} y'_i = k+3+(k+2)\alpha+(k+1)\alpha^2+k\alpha^3$$

Then, in relative terms, we have

$$\frac{x'_1}{\sum_{i=1}^{k+3} x'_i} \geq \frac{x_1}{\sum_{i=1}^{2k+3} x_i} \Longleftrightarrow y'_1 \sum_{i=1}^{2k+3} y_i \geq y_1 \sum_{i=1}^{k+3} y'_i$$

$$\Longleftrightarrow (\alpha+1)\left(\alpha - \frac{1-\sqrt{k+1}}{k}\right)\left(\alpha - \frac{1+\sqrt{k+1}}{k}\right) \leq 0$$

Hence,

$$\begin{cases} \dfrac{x'_1}{\sum_{i=1}^{k+3} x'_i} \geq \dfrac{x_1}{\sum_{i=1}^{2k+3} x_i} & \text{for } \ k \leq 3, \ \alpha \in (0,1), \quad \text{and} \\ \\ \dfrac{x'_1}{\sum_{i=1}^{k+3} x'_i} \leq \dfrac{x_1}{\sum_{i=1}^{2k+3} x_i} & \text{for } \ k > 3, \ \alpha \in \left(\frac{1+\sqrt{k+1}}{k}, 1\right) \end{cases}$$

For the PageRank of the acyclic digraph in Figure 3 we have $q_1 = 1+\alpha+(\frac{k}{2}+1)\alpha^2+\frac{3k}{2}\alpha^3$, $q_2 = 1+(\frac{k}{2}+1)\alpha+\frac{3k}{2}\alpha^2$, $q_3 = 1+\frac{3k}{2}\alpha$, $q_i = 1$, for $i = 4, \ldots, 2k+3$, and $\displaystyle\sum_{i=1}^{2k+3} q_i = 2k+3+(2k+2)\alpha+(2k+1)\alpha^2+\frac{3k}{2}\alpha^3$, and for the acyclic digraph without the k vertices of outdegree 2 we have $q'_i = y'_i, i = 1, \ldots, k+3$. Then, in relative terms, we have

$$\frac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \geq \frac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} \Longleftrightarrow q'_1 \sum_{i=1}^{2k+3} q_i \geq q_1 \sum_{i=1}^{k+3} q'_i \Longleftrightarrow (m-3)\alpha^2 - 4\alpha - 2 \leq 0$$

Hence,

$$\begin{cases} \dfrac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \geq \dfrac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} & \text{for } \ k \leq 9, \ \alpha \in (0,1), \quad \text{and} \\ \\ \dfrac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \leq \dfrac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} & \text{for } \ k > 9, \ \alpha \in \left(\frac{2+\sqrt{2k-2}}{k-3}, 1\right) \end{cases}$$

As a consequence, the notion of queue tree has no analogue in the context of acyclic digraphs.

In the remainder of this work we shall concentrate on the analysis of the behavior of PageRank in more general structures. Observing that α-centrality can be seen as the particular case of PageRank where the branching of all paths is 1, from each result about PageRank we will obtain a similar result for α-centrality as corollary.

5. The bidirectional case

We turn now to trees with bidirectional as well as unidirectional arcs. A digraph $\mathcal{B}^r = (V, A)$ is a **bidirectional tree with root** r if its set of arcs A can be partitioned in two disjoint sets A_1 and A_2 such that:

- (V, A_1) is a partial tree with root r (the underlying tree of $\mathcal{B}^r$), and
- if $uv \in A_2$ then $vu \in A_1$, and in this case we have the **bidirectional arc** (or 2-cycle) vuv.

Observe that for each arc $uv \in A_2$ the corresponding bidirectional arc vuv defines an infinite number of walks ending at the root r (just as would do any cycle within a tree). Henceforth, to the effect of computing the PageRank of r with formula (13), we can view each arc $uv \in A_2$ as a path of infinite length hanging from the vertex v, and containing alternatively copies of vertices u and v, where at each v hangs a copy of the sub–tree rooted at v, $\mathcal{T}^v$, and at each u hangs a copy of the remainder of the sub–tree rooted at u after removing from it the sub–tree $\mathcal{T}^v$, that is, $\mathcal{T}^u \setminus \mathcal{T}^v$. Note that $\mathcal{T}^u$ (and $\mathcal{T}^v$) may contain bidirectional arcs. Extending this idea through all bidirectional arcs, we can view the bidirectional tree $\mathcal{B}^r$ as its associated infinite tree. Figure 4 shows a bidirectional tree $\mathcal{B}^r$ with two disjoint bidirectional arcs, vuv and $v'u'v'$ (leftmost tree); next to it the bidirectional tree with an infinite branch corresponding to vuv; and the rightmost tree is the full infinite tree associated to $\mathcal{B}^r$.

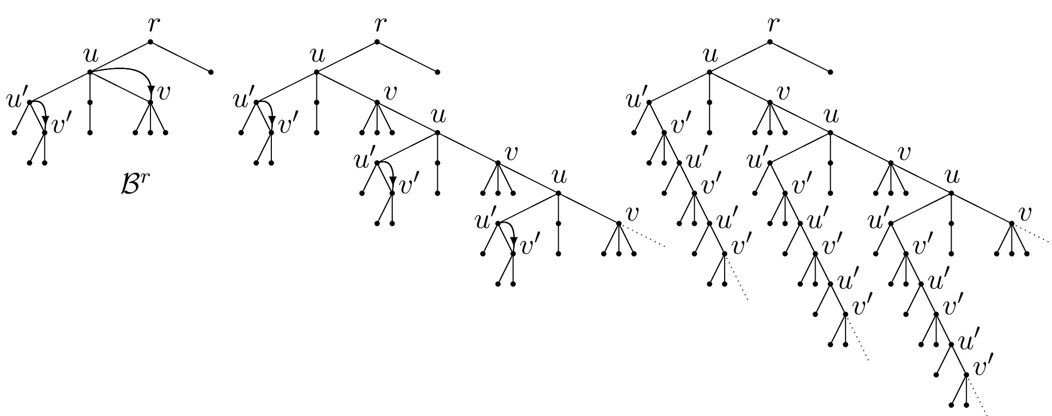

Figure 4: Bidirectional tree $\mathcal{B}^r$ and its associated infinite tree in two stages.

This view of $\mathcal{B}^r$ as an infinite tree makes it easier to understand the interpretations we do below of formula (13) adapted to our trees. In formula

(13), the sum is taken over all vertices w connected through a walk to a. In the associated infinite tree this walk is a unique path ρ connecting w with a. This path could have various incidence of bidirectional arcs. On the other hand, each bidirectional arc vuv, with $u \neq r$ and $od(u) = 2$, produces an infinite number of walks: u, uvu, $uvuvu$, $\ldots$, with branching factors $D(u) = 1$, $D(uvu) = 1/2$, $D(uvuvu) = 1/2^2$, $\ldots$; hence, summing over all these walks we get

$$\sum_{\rho\,:\,u \longrightarrow u} \alpha^{l(\rho)} D(\rho) = 1 + \frac{\alpha^2}{2} + \frac{\alpha^4}{2^2} + \cdots = \frac{1}{1-\alpha^2/2}$$

Therefore, if the path $\rho : w \longrightarrow a$ in formula (13) contains q vertices, each meeting a bidirectional arc, the contribution to $\mathcal{P}(a)$ of the possible walks produced on ρ is $1/(1-\alpha^2/2)^q$. If the bidirectional arc is vrv, with $od(r) = 1$, and hence $D(rvr \ldots vr) = 1$ for any walk on this arc, we get that the contribution to $\mathcal{P}(a)$ is $1/(1-\alpha^2)$. All the above observations lead to the following result on computing the PageRank on bidirectional trees.

Theorem 21. *Let $\mathcal{B}^r = (V, A)$ be a bidirectional tree rooted at r.*
(1) If $od(r) = 0$, then the PageRank of any $a \in V$ is given by

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho : w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n (1-\alpha^2/2)^q} \tag{18}$$

(2) If $od(r) = 1$ with bidirectional arc rvr, then

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho : w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n (1-\alpha^2/2)^q}, \quad \textit{for} \;\; a \notin \{r, v\} \tag{19}$$

and

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho : w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n (1-\alpha^2/2)^{q-1}(1-\alpha^2)}, \quad \textit{for} \;\; a \in \{r, v\} \tag{20}$$

where in all cases, $\rho : w \longrightarrow a$ is the unique path from the vertex w to a, and $l(\rho)$ is the length of this path; n is the number of bidirectional vertices (i.e. with $od(u) = 2$) not being an end-vertex in ρ; q is the number of bidirectional arcs meeting ρ. $\square$

In particular, if $od(r) = 0$, then $n = q$ and so

$$\mathcal{P}(r) = \frac{1-\alpha}{N} \sum_{\substack{w \in V \\ \rho: w \longrightarrow a}} \frac{\alpha^{l(\rho)}}{(2-\alpha^2)^q} \tag{21}$$

And if $od(r) = 1$, then $n = q - 1$ and

$$\mathcal{P}(r) = \frac{1-\alpha}{N} \sum_{\substack{w \in V \\ \rho: w \longrightarrow a}} \frac{\alpha^{l(\rho)}}{(2-\alpha^2)^{q-1}(1-\alpha^2)} \tag{22}$$

For α-centrality formulas (18) to (22) coincide.

Corollary 22. *Let $\mathcal{B}^r = (V, A)$ be a bidirectional tree rooted at r. Then the α-centrality of a vertex $v_i \in V$, $i = 1, \ldots, N$, is given by*

$$x_i = \frac{1-\alpha}{N} \sum_{\substack{w \in V \\ \rho: w \longrightarrow v_i}} \frac{\alpha^{l(\rho)}}{(1-\alpha^2)^q} \tag{23}$$

where $\rho : w \longrightarrow v_i$ is the unique path from the vertex w to v_i, $l(\rho)$ is the length of this path, and q is the number of bidirectional arcs meeting ρ. □

Our proposed formula for computing the PageRank of the root in the case of unidirectional trees (Eq. (15)) is founded on Brinkmeier's breadth-first search implementation of his analytical formulation (Eqs. (14) and (13)). We would like to have a result on the same spirit of counting by levels for bidirectional trees.

For a breadth-first search type of computation of PageRank on a bidirectional tree, we must classify somehow the vertices by levels of the tree. For each $k > 0$, the vertices at level $N_k = \{v_{k1}, \ldots, v_{kn_k}\}$ are characterize by the number of bidirectional arcs met by their paths which ends in the root, $v_{ki} \ldots r$. Hence, $n_k = n_k^0 + \cdots + n_k^{k+1}$, where n_k^q denotes the number of vertices at level N_k having q bidirectional arcs meeting their paths to r. Some of these n_k^q could be null. The non-null n_k^q many vertices contributes to the summation in equations (21) and (22) the quantities $n_k^q \alpha^k/(2-\alpha^2)^q$ and $n_k^q \alpha^k/(2-\alpha^2)^{q-1}(1-\alpha^2)$ according to either case of $od(r) = 0$ or $od(r) = 1$. Thus, we have the following result.

Theorem 23. *Let $\mathcal{B}^r$ be a bidirectional tree rooted at r, with N vertices and height $h > 0$.*

(1) *If* $od(r) = 0$, $\qquad \mathcal{P}(r) = \dfrac{1-\alpha}{N}\displaystyle\sum_{k=0}^{h}\sum_{q=0}^{k}\dfrac{n_k^q \alpha^k}{(2-\alpha^2)^q}$

(2) *If* $od(r) = 1$, $\qquad \mathcal{P}(r) = \dfrac{1-\alpha}{N}\displaystyle\sum_{k=0}^{h}\sum_{q=0}^{k}\dfrac{n_k^{q+1} \alpha^k}{(2-\alpha^2)^q(1-\alpha^2)}$

where q *is the number of bidirectional arcs met by the path ending in* r*, but distinct from the bidirectional arc incidence with* r*, if such bidirectional arc exists.* □

We can give a more succinct vectorial formulation of the previous result, if we develop the sums "by rows" (outmost sum) and group column terms in a vector.

Theorem 24. *Let* $\mathcal{B}^r$ *be a bidirectional tree rooted at* r*, with* N *vertices and height* $h > 0$*. Then*

$$\mathcal{P}(r) = \begin{cases} \dfrac{1-\alpha}{N}\displaystyle\sum_{q=0}^{h}\dfrac{\Delta_q \cdot \Lambda_q}{(2-\alpha^2)^q} & \text{if } od(r) = 0 \\[2ex] \dfrac{1-\alpha}{N}\displaystyle\sum_{q=0}^{h}\dfrac{\Delta'_q \cdot \Lambda_q}{(2-\alpha^2)^q(1-\alpha^2)} & \text{if } od(r) = 1 \end{cases}$$

where $\Delta_q = (n_q^q, n_{q+1}^q, \ldots, n_h^q)$*,* $\Delta'_q = (n_q^{q+1}, n_{q+1}^{q+1}, \ldots, n_h^{q+1})$ *and* $\Lambda_q = (\alpha^q, \alpha^{q+1}, \ldots, \alpha^h)$*.* □

Corollary 25. *Let* $\mathcal{B}^r$ *be a bidirectional tree rooted at* r*, with* N *vertices and height* $h > 0$*. Then the* α*–centrality of the root* r *is given by*

$$x_r = \frac{1-\alpha}{N}\sum_{k=0}^{h}\sum_{q=0}^{k}\frac{n_k^q\alpha^k}{(1-\alpha^2)^q} = \frac{1-\alpha}{N}\sum_{q=0}^{h}\frac{\Delta_q\cdot\Lambda_q}{(1-\alpha^2)^q}$$

where q *is as in Theorem 23,* Δ_q *and* Λ_q *are as in Theorem 24.* □

5.1. Case of s-cycles

In this section we generalize the computation of PageRank to bidirectional trees of height $h > 1$ on which we close permissible cycles of any length obtained by joining vertices from level N_j with vertices from level N_k, for $0 \le j < k \le h$. In this way we can transform bidirectional arcs vuv into cycles $vuv_{s-1}\ldots v_2 v$ of longer length, where the arc uv close the new

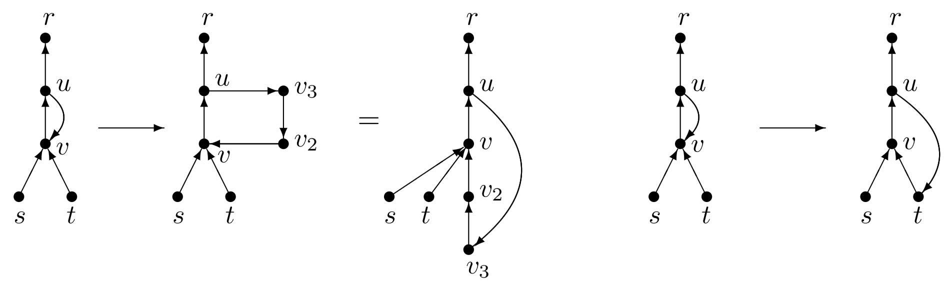

Figure 5: Examples of cyclical trees.

cycle inserted in the rooted tree. Also the arc uv of the bidirectional arc vuv can be substituted by a new arc ut closing a larger path $t \dots vu$ in the tree. In Figure 5 we exhibit some examples of these transformations.

Formally we define a digraph $\mathcal{C}^r = (V, A)$ as a **cyclical tree with root** r, if its set of arcs A can be partitioned in two disjoint sets A_1 and A_2 such that:

- (V, A_1) is a partial tree with root r (the underlying tree of $\mathcal{C}^r$), and
- if $uv \in A_2$ then there is a path $v_s v_{s-1} \dots v_1$, beginning at $v_s = u$, ending at $v_1 = v$ and with intermediate vertices and arcs $v_{i+1}v_i$ in A_1, and in this case we have the s-cycle $vuv_{s-1} \dots v_2 v$.

We proceed to compute the PageRank of these cyclical trees. Similarly to the bidirectional case, we have that each cycle $vu \dots v$ of length $l \geq 2$ and $od(u) = 2$ produces an infinite number of walks: u, $uv \dots u$, $uv \dots uv \dots u$, $\dots$, with branching factors $D(u) = 1$, $D(uv \dots u) = 1/2$, $D(uv \ \dots uv \dots u) = 1/2^2$, $\dots$; hence, summing over all these walks we get

$$\sum_{\rho : u \longrightarrow u} \alpha^{l(\rho)} D(\rho) = 1 + \frac{\alpha^l}{2} + \frac{\alpha^{2l}}{2^2} + \dots = \frac{1}{1 - \alpha^l/2}$$

Therefore, if the path $\rho : w \longrightarrow a$ contains q vertices, meeting q cycles of length l_1, l_2, $\dots$, l_q, respectively, then the contribution to $\mathcal{P}(a)$ of the possible walks produced on ρ is

$$\frac{1}{1 - \alpha^{l_1}/2} \cdot \frac{1}{1 - \alpha^{l_2}/2} \cdots \frac{1}{1 - \alpha^{l_q}/2}$$

If the cycle is $vr \overset{l}{\dots} v$, with $od(r) = 1$, and hence $D(rv \dots r) = 1$, we get that the contribution to $\mathcal{P}(a)$ is $1/(1 - \alpha^l)$.

Theorem 26. *Let $\mathcal{C}^r = (V, A)$ be a cyclical tree rooted at r.*
(1) *If $od(r) = 0$, then PageRank for a vertex $a \in V$ is given by*

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n(1-\alpha^{l_1}/2)\cdots(1-\alpha^{l_q}/2)}$$

(2) *If $od(r) = 1$ in the cycle $rv_1 \dots v_{l_{q-1}} r$, then*

$$\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n(1-\alpha^{l_1}/2)\cdots(1-\alpha^{l_q}/2)}, \quad \text{for } a \notin \{r, v_1, \dots, v_{l_{q-1}}\}$$

and $\mathcal{P}(a) = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{2^n(1-\alpha^{l_1}/2)\cdots(1-\alpha^{l_{q-1}}/2)(1-\alpha^{l_q})}$, *for $a \in \{r, v_1, \dots, v_{l_{q-1}}\}$, where in all cases $\rho : w \longrightarrow a$ is the unique path from w to a, and $l(\rho)$ is the length of this path; n is the number of bidirectional vertices (i.e. with $od(u) = 2$) not being an end-vertex in ρ; q is the number of cycles meeting ρ and of lengths $l_1, l_2, \dots, l_q$.* □

In particular, if $od(r) = 0$, $n = q$, and

$$\mathcal{P}(r) = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{(2-\alpha^{l_1})\dots(2-\alpha^{l_q})} \tag{24}$$

And if $od(r) = 1$, $n = q-1$, and

$$\mathcal{P}(r) = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} a} \frac{\alpha^{l(\rho)}}{(2-\alpha^{l_1})\dots(2-\alpha^{l_{q-1}})(1-\alpha^{l_q})} \tag{25}$$

Corollary 27. *Let $\mathcal{C}^r = (V, A)$ be a cyclical tree rooted at r. Then the α-centrality for a vertex $v_i \in V$, $i = 1, \dots, N$, is given by*

$$x_i = \frac{1-\alpha}{N} \sum_{\rho: w \xrightarrow{w \in V} v_i} \frac{\alpha^{l(\rho)}}{(1-\alpha^{l_1})\cdots(1-\alpha^{l_q})}$$

where $\rho : w \longrightarrow v_i$ is the unique path from the vertex w to v_i, $l(\rho)$ is the length of this path, and q is the number of cycles meeting ρ and of lengths $l_1, l_2, \dots, l_q$. □

6. Rearrangements in rooted bidirectional and cyclical trees

Analogously to the case of unidirectional trees we shall analyze in this section the behavior of PageRank on bidirectional, and more general, cyclical trees when their topology is modified. Our first result shows that on a unidirectional tree changing unidirectional arcs to bidirectional enhance the PageRank value of the end-vertices of the transformed arc, but reduces the PageRank of the root of the tree.

Theorem 28. *If in a unidirectional tree $\mathcal{T}^r$ an arc vu, with $u \neq r$, is changed to a bidirectional arc vuv, then $\mathcal{P}(u)$ and $\mathcal{P}(v)$ both increase, but $\mathcal{P}(r)$ decreases. The same holds for α-centrality.*

Proof. We introduce some notation first. The term $\mathcal{P}_x(\mathcal{T}^y)$ denotes the PageRank of vertex x in the tree $\mathcal{T}^y$ with root y and $n_p(\mathcal{T}^y)$ denotes the number of vertices at level N_p in the tree $\mathcal{T}^y$. Now, assume that u is at level N_k in the tree $\mathcal{T}^r$ and, hence, $v \in N_{k+1}$ (see Figure 6).

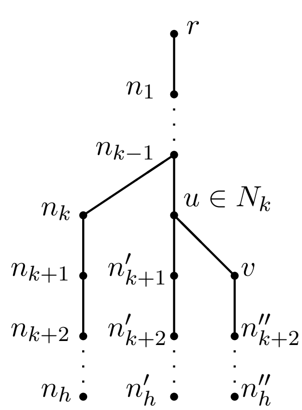

Figure 6: Number of vertices by levels

Then, we have that

$$\begin{aligned}
\mathcal{P}_r(\mathcal{T}^r) &= \frac{1-\alpha}{N} \sum_{p=0}^{h} n_p(\mathcal{T}^r)\alpha^p \\
&= \frac{1-\alpha}{N} \left(\sum_{p=0}^{h} n_p(\mathcal{T}^r - \mathcal{T}^u)\alpha^p + \sum_{p=k}^{h} n_p(\mathcal{T}^u)\alpha^p \right)
\end{aligned}$$

Therefore, if $\mathcal{B}^r$ is the bidirectional tree obtained from $\mathcal{T}^r$ by just changing the arc vu to bidirectional arc vuv, using the results of Section 5, we have

$$\mathcal{P}_r(\mathcal{B}^r) = \frac{1-\alpha}{N}\left(\sum_{p=0}^{h} n_p(\mathcal{T}^r - \mathcal{T}^u)\alpha^p + \frac{1}{2(1-\alpha^2/2)}\sum_{p=k}^{h} n_p(\mathcal{T}^u)\alpha^p\right) < \mathcal{P}_r(\mathcal{T}^r)$$

which shows that the PageRank of the root r decreases. On the other hand, the PageRanks of u and v are given by the equations:

$$\mathcal{P}_u(\mathcal{B}^u) = \frac{1-\alpha}{N(1-\alpha^2/2)}\sum_{p=k}^{h} n_p(\mathcal{T}^u)\alpha^{p-k} = \frac{1}{1-\alpha^2/2}\mathcal{P}_u(\mathcal{T}^r) > \mathcal{P}_u(\mathcal{T}^r)$$

$$\begin{aligned}\mathcal{P}_v(\mathcal{B}^v) &= \frac{1-\alpha}{N(1-\alpha^2/2)}\left(\frac{\alpha}{2}\sum_{p=k}^{h} n_p(\mathcal{T}^u - \mathcal{T}^v)\alpha^{p-k} + \sum_{p=k+1}^{h} n_p(\mathcal{T}^v)\alpha^{p-(k+1)}\right)\\ &> \mathcal{P}_v(\mathcal{T}^v)\end{aligned}$$

For α-centrality, the corresponding expressions can be obtained replacing $\frac{1}{2\left(1-\frac{\alpha^2}{2}\right)}$ by $\frac{1}{1-\alpha^2}$ in $\mathcal{P}_r(B^r)$; $\frac{1}{1-\frac{\alpha^2}{2}}$ by $\frac{1}{1-\alpha^2}$ in $\mathcal{P}_u(B^u)$; $\frac{1}{1-\frac{\alpha^2}{2}}$ by $\frac{1}{1-\alpha^2}$ and $\alpha/2$ by α in $\mathcal{P}_v(B^v)$. $\square$

Using same arguments as given for the previous theorem, we can generalized the result to the case where the original tree is bidirectional, and some of its unidirectional arcs (if any) is promoted to being bidirectional.

Theorem 29. *Let $\mathcal{B}^r$ be a bidirectional tree, and let $\mathcal{B}'^r$ be the tree resulting from $\mathcal{B}^r$ when a unidirectional arc vu, with $u \neq r$, is changed to a bidirectional arc vuv (see Figure 7). Then*

1. $\mathcal{P}_u(\mathcal{B}'^u) = \dfrac{1}{1-\alpha^2/2}\mathcal{P}_u(\mathcal{B}^u) > \mathcal{P}_u(\mathcal{B}^u)$.
2. $\mathcal{P}_v(\mathcal{B}'^u) > \mathcal{P}_v(\mathcal{B}^u)$.
3. *If $v'u'v'$ is a bidirectional arc intersecting the path $uv_1 \dots v_{k-1}r$, then $\mathcal{P}_{u'}(\mathcal{B}'^r) < \mathcal{P}_{u'}(\mathcal{B}^r)$ and $\mathcal{P}_{v'}(\mathcal{B}'^r) < \mathcal{P}_{v'}(\mathcal{B}^r)$.*
4. *$\mathcal{P}_x(\mathcal{B}'^r) < \mathcal{P}_x(\mathcal{B}^r)$ for all vertex x in the path $v_1 \dots v_{k-1}r$.*
5. *In particular, $\mathcal{P}_r(\mathcal{B}'^r) < \mathcal{P}_r(\mathcal{B}^r)$.*
6. *The vertices which are neither contained in the path $uv_1 \dots v_{k-1}r$ nor in the bidirectional arcs intersecting this path preserve their original PageRank.*

Similar inequalities hold for α-centrality changing $(1-\alpha^2/2)$ by $(1-\alpha^2)$. □

Theorems 28 and 29 suggest that in order to increase the PageRank of the root r of a tree we have to directly promote to bidirectional the arcs incidence to r. The consequences of this manipulation is summarized in the following theorem, which is a direct consequence of the two previous results.

Theorem 30. *Let $\mathcal{B}^r$ be a bidirectional tree, with $od(r) = 0$, and let $\mathcal{B}'^r$ be the tree resulting from $\mathcal{B}^r$ when one of its arcs vr is changed to a bidirectional arc vrv. Then*

1. $\mathcal{P}_r(\mathcal{B}'^r) = \dfrac{\mathcal{P}_r(\mathcal{B}^r)}{1-\alpha^2}$.
2. $\mathcal{P}_v(\mathcal{B}'^r) = \mathcal{P}_v(\mathcal{B}^r) + \dfrac{\alpha\mathcal{P}_r(\mathcal{B}^r)}{1-\alpha^2}$.
3. $\mathcal{P}_r(\mathcal{B}'^r) \geq \mathcal{P}_v(\mathcal{B}'^r) \Longleftrightarrow \mathcal{P}_r(\mathcal{B}^r) \geq (1+\alpha)\mathcal{P}_v(\mathcal{B}^r)$.
4. *All other vertices (apart from r and v) preserve their PageRank.* □

Note that, for $\alpha = 0.85$, we have that $\mathcal{P}_r(\mathcal{B}'^r) \approx 3.6\mathcal{P}_r(\mathcal{B}^r)$, $\mathcal{P}_v(\mathcal{B}'^r) - \mathcal{P}_v(\mathcal{B}^r) \approx 3.06\mathcal{P}_r(\mathcal{B}^r)$ and that the PageRank of the root r is kept greater than the PageRank of the vertex v if and only if the original PageRank of r is greater than 1.85 times the PageRank of v.

Remark 31. *For cyclical trees we have results similar to Theorems 28–30 but replacing $1/(1-\alpha^2/2)$ by $1/(1-\alpha^l/2)$ for PageRank, or replacing $1/(1-\alpha^2)$ by $1/(1-\alpha^l)$ for α-centrality.*

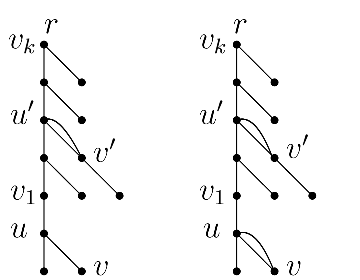

Figure 7: Theorem 29

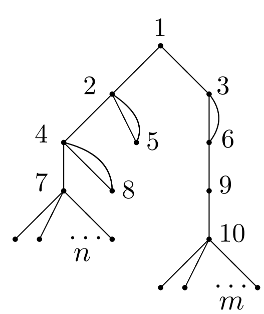

Figure 8: Exam. 32, case $od(r) = 0$

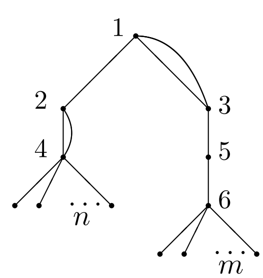

Figure 9: Exam. 32, case $od(r) = 1$

Now, the pruning of the lower levels of a bidirectional tree has mix consequences for the α-centrality and the PageRank of the root, in the same way as it happened for acyclic digraphs. The following example illustrates the possible outcomes of pruning lower levels of a bidirectional tree.

Example 32. *Consider the tree shown in Figure 8, with root labelled 1 and out-degree 0. We should compute the PageRank of vertex labeled by* 1 *before and after removing the* m *vertices of the last level. Applying Equation (21) we get*

$$\mathcal{P}(1) = \frac{1-\alpha}{N}\left(1 + \frac{2\alpha + 2\alpha^2 + \alpha^3 + \alpha^4 + m\alpha^5}{2-\alpha^2} + \frac{\alpha^2 + 2\alpha^3 + n\alpha^4}{(2-\alpha^2)^2}\right)$$

where $N = n + m + 10$*. Now, pruning the* m *vertices of the last level, we get that the new PageRank of 1 in the pruned tree is*

$$\mathcal{P}'(1) = \frac{1-\alpha}{N'}\left(1 + \frac{2\alpha + 2\alpha^2 + \alpha^3 + \alpha^4}{2-\alpha^2} + \frac{\alpha^2 + 2\alpha^3 + n\alpha^4}{(2-\alpha^2)^2}\right)$$

and $N' = n + 10$*. Then, for* $\alpha = 0.85$*, we have that*

$$\mathcal{P}'(1) > \mathcal{P}(1) \iff m(1443654850 - 19126309n) > 0$$

which holds for $n \leq 75$*, and independently of the positive value of* m*. Thus, for* $n \leq 75$ $(n \geq 76)$ *and for all* $m \geq 1$*, successive removal of the* m *vertices of the last level increments (decrements) the PageRank of the root,* $\mathcal{P}(1)$*. By similar arguments and using equation (22), in the tree shown in Figure 9, which is an example of a tree with root having out-degree 1, we have that for* $n \leq 31$ $(n \geq 32)$ *and for all* $m \geq 1$*, successive removal of the* m *vertices of the last level increments (decrements)* $\mathcal{P}(1)$*.*

The previous results give us some clues on ways of optimizing the α–centrality or PageRank of tree-like organized networks. Obviously these rules for rearrangement should apply insofar as the context allows.

Rule 1 To augment either eigenvector-based centrality value of the root, transform incoming arcs into bidirectional ones. Furthermore, link the root with vertices below in the tree (so that cycles passing by the root are built).

Rule 2 To augment either eigenvector-based centrality value of a vertex u different from the root, link u with a bidirectional arc to each one of the vertices on the subtree with root u (hence obtaining a cyclical tree).

Keep in mind that this enhances the α–centrality or PageRank of u but reduces the corresponding score of the root. One may interpret this action as linking an individual with all its subordinates in a hierarchical organization.

7. A note on fast computation of PageRank

There are several approaches in the literature to the task of speeding up the calculation of PageRank, based upon the following general scheme (see, for example, [10, 3, 7]):

> Partition the directed network into local sub-nets; then compute some independent ranking for each local sub-net, which will apply to the whole sub-net treated as a unit; and then compute the ranking of the digraph of sub-nets.

In [3] and [7] the local splitting of the directed network is done in strongly connected components (SCC), and further in [7, Thm 2.1], it is shown that the PageRank can be calculated independently on each SCC, provided we know the PageRank of all vertices outside the SCC, but directly linking to vertices in the SCC.

We observe that if a directed network $\mathcal{D} = (V, A)$ is a cyclical tree with root r then the set of arcs A can be partitioned in two disjoint sets A_1 and A_2 such that:

- (V, A_1) is a partial digraph whose condensation digraph is a tree of SCCs with distinguished roots, where each pair of adjacent SCCs are linked by a unique arc and the maximal SCC contains the root r (the underlying digraph of $\mathcal{D}$); and
- if $uv \in A_2$ then there is a path $v_s v_{s-1} \dots v_1$, beginning at $v_s = v$, ending at $v_1 = u$ and with intermediate vertices and arcs $v_{i+1}v_i$ in A_1.

Therefore, cyclical trees with root give a simple splitting of a directed network in the way of [3] and [7], namely as a tree of SCCs, with the additional strongest condition of having a single link between components, which by the previously mentioned result of [7], can have PageRank computed independently in each SCC, and on a very simple way, provided we know the PageRank of their descendants in the topological structure of the tree. This suggests computing PageRank in parallel and through *layers*, as is proposed in [7, §3], following an iterated process on the tree from a top level N_h down to the root at N_0. The cyclical tree is a suitable structure for the application of this process.

References

[1] C. C. Aggarwal. *Social Network Data Analytics*, Springer, 2011.

[2] C. C. Aggarwal, A. Khan, and X. Yan. On Flow Authority Discovery in Social Networks. *Proc. SIAM Conf. on Data Mining* (2011) 522–533.

[3] A. Arasu, J. Novak, A. Tomkins, J. Tomlin. PageRank computation and the structure of the Web (2002) *11th Int. WWW Conf.*

[4] S. Brin, L. Page. The anatomy of a large scale hypertextual web search engine. *Comp. Net. and ISDN Syst.*, 33 (1998) 107–117.

[5] S. Brin, R. Motwami, L. Page, T. Winograd. The PageRank citation ranking: Bringing order to the web. *Tech. Rep.*, CS, Stanford, 1998.

[6] M. Brinkmeier. PageRank revisited, *ACM Trans. on Internet Technologies*, 6, 3 (2006) 282–301.

[7] M. Brinkmeier. Distributed calculation of PageRank using strongly connected components, *Proc. of the I2CS'05 in Paris* (LNCS), 2005.

[8] P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology* (1987), 1170–1182.

[9] P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23 (3), (2001) 191–201.

[10] S. D. Kamvar, T. H. Haveliwala, C. D. Manning and G. H. Golub. Exploiting the block structure of the Web for computing PageRank. *Tech. Rep.*, Stanford, 2003.

[11] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18 (1953) 39–43.

[12] D. Kempe, J. Kleinberg and E. Tardos. Maximizing the spread of influence through a social network. In: *Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (2003) 137–146.

[13] A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, 2006.

[14] C. Marijuán, Birational equivalence of reduced graphs, *Journal of Pure and Applied Algebra* 124 (1998) 173-199.

[15] C. Marijuán, Finite topologies and digraphs, *Proyecciones Journal of Mathematics* 29, 3 (2010) 295-311.

[16] M. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[17] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In: *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (2009) 807–816.