Université catholique de Louvain
École polytechnique de Louvain
Département d'ingénierie mathématique

Dominant Vectors of Nonnegative Matrices

Application to Information Extraction in Large Graphs

Laure Ninove

Thèse présentée en séance publique le 21 février 2008
en vue de l'obtention du grade de
Docteur en Sciences appliquées.

DOMINANT VECTORS OF NONNEGATIVE MATRICES
APPLICATION TO INFORMATION EXTRACTION IN LARGE GRAPHS

Laure NINOVE

Thèse présentée en séance publique, le 21 février 2008, au Département d'Ingénierie mathématique de l'École polytechnique de Louvain à l'Université catholique de Louvain, à Louvain-la-Neuve, en vue de l'obtention du grade de Docteur en Sciences appliquées.

PROMOTEURS : Professeurs Vincent BLONDEL et Paul VAN DOOREN

JURY :

Professeur Georges BASTIN, président, Université catholique de Louvain
Professeur Vincent BLONDEL, Université catholique de Louvain
Professeur Lars ELDÉN, Linköping Universitet, Suède
Professeur Stéphane GAUBERT, INRIA Rocquencourt, France
Professeur Ulrich KRAUSE, Universität Bremen, Allemagne
Professeur Paul VAN DOOREN, Université catholique de Louvain

Acknowledgements

Summary

Objects such as documents, people, words or utilities, that are related in some way, for instance by citations, friendship, appearance in definitions or physical connections, may be conveniently represented using graphs or networks. An increasing number of such relational databases, as for instance the World Wide Web, digital libraries, social networking web sites or phone calls logs, are available. Relevant information may be hidden in these networks. A user may for instance need to get authority web pages on a particular topic or a list of similar documents from a digital library, or to determine communities of friends from a social networking site or a phone calls log. Unfortunately, extracting this information may not be easy.

This thesis is devoted to the study of problems related to information extraction in large graphs with the help of dominant vectors of nonnegative matrices. The graph structure is indeed very useful to retrieve information from a relational database. The correspondence between nonnegative matrices and graphs makes Perron–Frobenius methods a powerful tool for the analysis of networks.

In a first part, we analyze the fixed points of a normalized affine iteration used by a database matching algorithm. Then, we consider questions related to PageRank, a ranking method of the web pages based on a random surfer model and used by the well known web search engine Google. In a second part, we study optimal linkage strategies for a web master who wants to maximize the average PageRank score of a web site. Finally, the third part is devoted to the study of a nonlinear variant of PageRank. The simple model that we propose takes into account the mutual influence between web ranking and web surfing.

Contents

List of Figures

List of notations

$\mathbf{1}$	vector of all ones	
0	zeros scalar, vector or matrix	
$A, B, \ldots$	matrices, $A = [A_{ij}]_{i,j=1}^{m,n}$	
A^T	transpose of the matrix A	
A^D	Drazin inverse of the matrix A	33
$A_{\mathcal{I},\mathcal{J}}$	submatrix of A determined by the index sets $\mathcal{I}, \mathcal{J}$	
$A_{\mathcal{I}}$	principal submatrix of A determined by $\mathcal{I}$, i.e. $A_{\mathcal{I}\mathcal{I}}$	
$A \geq B$	entrywise inequality for A and B, i.e. $A_{ij} \geq B_{ij}$	
$A > B$	entrywise inequality for A and B, i.e. $A_{ij} > B_{ij}$	
$A \gneqq B$	$A \geq B$ and $A \neq B$	
$A \otimes B$	Kronecker product of A and B	
α^+	positive part of the scalar, i.e. $\max\{0, \alpha\}$	
$\mathcal{A}(r)$	set of spanning arborescences of $\mathcal{G}$ rooted at r	133
δ_{ij}	Kronecker delta, i.e. $\delta_{ij} = 1$ if $i = j$ and 0 otherwise	
$\boldsymbol{d}_i$	outdegree of the node i, i.e. $\lvert\{k \in \mathcal{N} : (i,k) \in \mathcal{E}\}\rvert$	
$d_H(\boldsymbol{x}, \boldsymbol{y})$	Hilbert's projective metric between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$	37
$d_H(A, B)$	induced projective metric between matrices A and B	138
$\operatorname{diag}(\boldsymbol{x})$	diagonal matrix determined by the vector $\boldsymbol{x}$	
$\boldsymbol{e}_i$	i^{th} standard basis vector in $\mathbb{R}^n$	
$\boldsymbol{e}_{\mathcal{I}}$	vector with a 1 in the entries of $\mathcal{I}$ and a 0 elsewhere	
$\mathcal{E}_{\mathcal{I}}$	set of internal links, i.e. $\{(i,j) \in \mathcal{E} : i, j \in \mathcal{I}\}$	
$\mathcal{E}_{\text{out}(\mathcal{I})}$	set of external outlinks, i.e. $\{(i,j) \in \mathcal{E} : i \in \mathcal{I}, j \notin \mathcal{I}\}$	
$\mathcal{E}_{\text{in}(\mathcal{I})}$	set of external inlinks, i.e. $\{(i,j) \in \mathcal{E} : i \notin \mathcal{I}, j \in \mathcal{I}\}$	
$\mathcal{E}_{\bar{\mathcal{I}}}$	set of external links, i.e. $\{(i,j) \in \mathcal{E} : i, j \notin \mathcal{I}\}$	

$\mathcal{G} = (\mathcal{N}, \mathcal{E})$	graph defined by a set of nodes $\mathcal{N}$ and of edges $\mathcal{E}$	27
$\mathcal{G} \times \widetilde{\mathcal{G}}$	graph product of the graphs $\mathcal{G}$ and $\widetilde{\mathcal{G}}$	49
$\mathcal{G}(A)$	graph of the nonnegative matrix A	31
$\mathcal{G}_{\mathcal{I}}$	subgraph of $\mathcal{G}$ determined by the set $\mathcal{I}$, i.e. $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$	
I	identity matrix	
(i, j)	edge from node i to node j	
$i \to j$	the node i belongs to the set of in-neighbors of j	28
$\mathrm{ind}(A)$	index of the matrix A	33
$\mathrm{int}(\Sigma)$	relative interior of Σ, i.e. $\Sigma \cap \mathbb{R}^n_{>0}$	
$\langle i_0, \ldots, i_\ell \rangle$	path from node i_0 to node i_ℓ	28
$j \leftarrow i$	the node j belongs to the set of out-neighbors of i	28
$\mathbb{N}$	set of nonnegative integer numbers	
$\lVert\cdot\rVert^\star$	dual vector norm of a vector norm $\lVert\cdot\rVert$	68
$\lvert\lvert\lvert\cdot\rvert\rvert\rvert$	matrix norm induced by a vector norm $\lVert\cdot\rVert$	
$\lVert\cdot\rVert_1$	ℓ_1 vector norm, i.e. sum vector norm	
$\lVert\cdot\rVert_2$	ℓ_2 vector norm, i.e. Euclidean vector norm	
$\lvert\lvert\lvert\cdot\rvert\rvert\rvert_2$	matrix norm induced by $\lVert\cdot\rVert_2$, i.e. spectral norm	
$\lVert\cdot\rVert_\infty$	ℓ_∞ vector norm, i.e. max vector norm	
$\lVert\cdot\rVert_F$	Frobenius matrix norm	
$\mathbb{P}(X \mid Y)$	probability of event X given Y	
$\mathbb{R}$	set of real numbers	
$\mathbb{R}_{\geq 0}$	set of nonnegative real numbers	
$\mathbb{R}_{>0}$	set of positive real numbers	
$\mathrm{rank}(A)$	rank of the matrix A	
$\rho(A)$	spectral radius of the matrix A	
Σ	standard simplex in $\mathbb{R}^n$, i.e. $\{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \colon \sum_i \boldsymbol{x}_i = 1\}$	
$\mathrm{sign}(\alpha)$	sign of the scalar α	
$\mathrm{supp}(\boldsymbol{x})$	support of the vector $\boldsymbol{x}$, i.e. $\{i \colon \boldsymbol{x}_i > 0\}$	59
$\tau_B(A)$	Birkhoff's coefficient of ergodicity of the matrix A	138
$\mathrm{vec}(A)$	vector of stacked columns of A	49
$\boldsymbol{x}, \boldsymbol{y}, \ldots$	vectors, $\boldsymbol{x} = [\boldsymbol{x}_i]_{i=1}^n$	
$\boldsymbol{x}^T$	transpose of the vector $\boldsymbol{x}$	
$\boldsymbol{x}_{\mathcal{I}}$	subvector of $\boldsymbol{x}$ determined by the index set $\mathcal{I}$	

Chapter 1

Introduction

1.1 World Wide Web and search engines

The concept of hypertext was introduced in order to make the retrieval of information easier. Its starting point can be traced back to the forties [31, 84], soon after World War II, when Vannevar Bush described the Memex, a futuristic device that was intended to be used as a supplement of memory for the user [28] (see Figure 1.1). The user could store various documents and resources in his Memex, which would store them on microfilms. These resources would of course be indexed by codes as in libraries, in order to allow the user to quickly find a document knowing its code. But the Memex was not to be limited to usual hierarchical indexing but also provided *associative indexing*. As Bush explained, humans think by association. When the user thinks that two documents in his Memex should be related, he can order the Memex to link them permanently. The Memex would then add to one document the code of the other. So each time the user would consult one of these two documents, he would be able to recall instantly the other one.

The term *hypertext* was coined by Ted Nelson twenty years later to refer to systems where documents and other resources are linked together. So from some document, the user may be able to reach on demand some other related information (text, document or other resources). In 1960, Ted Nelson launched the Project Xanadu, the first project of a hyper-

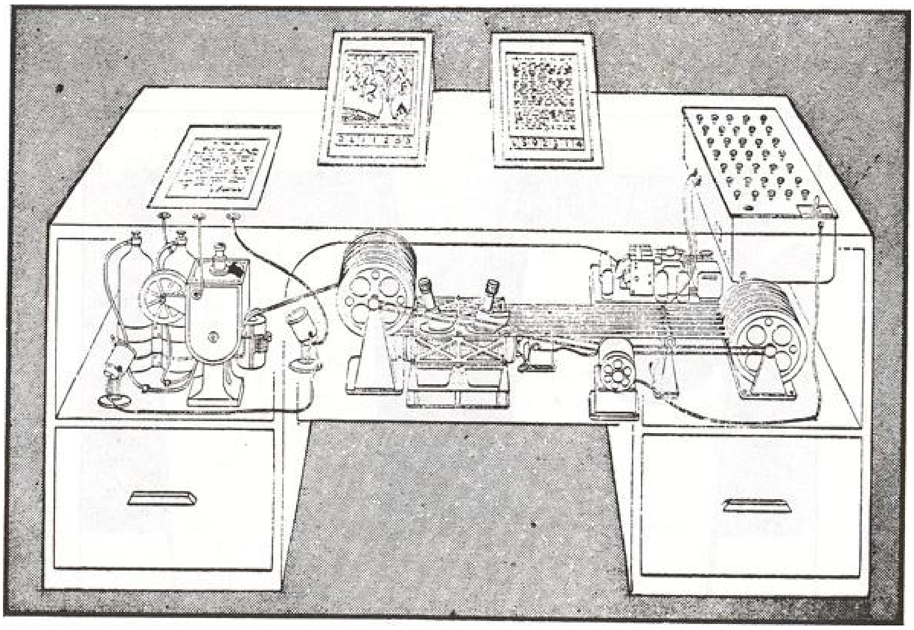


Figure 1.1: A drawing of Bush's Memex appeared in Life [29].

text system. This was an ambitious project, much more than today's World Wide Web: it proposed a system with e.g. automatic version management, side-by-side inter comparison of documents, valid copyright system and no breaking links [98]. But the first hypertext system to be operational was Doug Engelbart's NLS (oNLine System). In 1968, Doug Engelbart made the first public demonstration of an hypertext system together with several other innovations, as the computer mouse. He showed for instance how by clicking on some graphic, he could jump to other resources linked to the graphic.

In 1989, Tim Berners-Lee assessed that a lot of information was lost at CERN because of the high turnover of people and made a proposal for information management [15]. The system he described should store information without restraining it to a hierarchical system. It could be viewed as a web of interconnected nodes. Nodes could represent pieces of information, like documents, projects or people. Directed links between nodes could for instance represent relations as "refers to", "is part of" or "is an example of". Berners-Lee was soon joined by Robert Cailliau. Combining hypertext with the Internet, the network of computers networks, they created the World Wide Web. In 1990, Berners-Lee had developed the first web browser, web server and web pages.

Hypertext was intended as a way to organize resources. Paradoxically, the success of the World Wide Web and the huge size it achieved in a few years urged to develop tools to make the search of information on the Web easier.

The first web search engine was created in 1993: Wandex consisted in an index of web pages generated by a web crawler called the World Wide Web Wanderer. Early search engines then provided full-text search, that is, the possibility for the user to search any word in any indexed web page. But the results obtained were listed arbitrarily, without any relevance based ranking. In 1994, relevance retrieval was introduced, providing ranked results to a query. Relevance algorithms were at this time based on statistical analysis of the content of the web pages. Later, meta-search engines appeared. The idea of meta-search engines was to send the user's query to several other search engines and then to display results from each of them. Meta-search engines were useful when the different search engines' robots did not crawl a large enough part of the Web and had very different databases of indexed web pages. Let us also note that, from the very start of the Web, people used web directories, that are not strictly speaking web search engines. Yahoo! for instance, created in 1994, was at that time a hierarchical directory of web sites, that was manually compiled and maintained.

Progressively, limitations of content based relevance retrieval started to be felt. Indeed, the number of occurrences of terms of the query, or other similar statistics, is not always a good indicator of the relevance of the web page. Kleinberg [76], for instance, noticed that, for the search query *Harvard*, the home page of Harvard University was not the page where this term appeared most often. About 1998, the idea was in the air: the hyperlink structure itself could be used to define relevance scores for web pages.

On the one hand, Brin and Page proposed the PageRank algorithm that became the heart of their Google search engine [25, 103]. PageRank assigns to each indexed web page an importance score, depending only on the link structure of the Web. To compute these scores, PageRank considers a hyperlink from a web page to another as a vote of confidence from the first page to the second. Web pages divide their vote between the pages they link to, that is, if a web page links to four other pages, each of these pointed pages gets a quarter of its vote. Moreover, PageRank gives more weight to votes from important pages, that is, web pages that receive themselves many votes. So the importance scores of the web pages are computed iteratively, the score of each page being reinforced by the score the web pages linking to it. These scores are called PageRank scores of the web pages. Basically, for a particular request, the search engine selects related web pages by text matching based search and ranks them according to the PageRank scores. Note that PageRank scores of the web pages are computed independently from the user's query.

On the other hand, Kleinberg proposed a two levels algorithm in order to use the link structure of the web graph [76]. He assigns to each web page two scores: an authority score and a hub score. These scores are mutually reinforced: a web page has a good authority score if it is pointed to by many web pages with a good hub score. Similarly, a good hub is a web page that links to many good authorities. Unlike PageRank, the hub and authority scores are not computed independently from the user query. Here, for a particular query, the algorithm first constructs a focused subgraph of the Web. Then it computes iteratively authority and hub scores for this subgraph. Kleinberg's hubs and authorities algorithm is also called HITS (Hypertext Induced Topic Search). Even if it did not have the same success as Google's PageRank, it has been integrated in a search engine like Teoma, now merged with Ask search engine.

PageRank and HITS both use the link structure of the Web to rank web pages by computing iteratively scores. We will see that the reinforcement relations considered by these algorithms can be expressed as a simple linear iterative process. So PageRank scores, as well as Kleinberg's hub and authority scores can be represented by dominant eigenvectors of matrices constructed from the adjacency matrix of the web graph. Search engines based on these algorithms are called *eigen-*

vector based search engines.

1.2 Context of the thesis

The present work fits in the context of information extraction in large graphs by computing dominant vectors of nonnegative matrices. Various sets of data can be represented using graphs, as soon as the considered objects, such as documents, people, words or utilities, are related in some way, for instance by citations, friendship, appearance in definition or physical connections. By information extraction in large graphs, we mean the use of the structure of the graph in order to get information that is relevant to the user's need and that is not easy to find a priori. This can be for instance a list of authority documents, the determination of communities of friends, of synonyms or of central items in physical networks.

PageRank [25] and HITS [76] are two examples of information extraction methods in large graphs based on the computation of dominant vectors of nonnegative matrices. Another example is the computation of similarity scores between nodes of two graphs. Blondel et al. [17] define a similarity measure that they use [109] in order to automatically extract synonyms from a monolingual dictionary. Melnik et al. [93] consider the logical schemas of two relational databases, i.e., the descriptions of the objects of the databases and the relationships between them, and try to provide a matching of the corresponding fields. For this, they first construct a graph for each database schema, then they compute similarity scores between every pair of fields from both schemas and they finally select a subset of pairs with high scores in order to make a matching. Both these similarity algorithms will be described with more details in the next chapter.

There are examples of such methods in many other contexts, including non science applications. Fowler et al. [47], for instance, construct a graph from the reports of majority opinions of the U.S. Supreme Court. Each case is represented by a node and the links represent the citations made by the Court to preceding cases. Their analysis of the graph allows them to automatically find authority cases that closely correspond to those chosen by legal experts for their relevance. They also characterize the evolution of the use of precedents in the judgments. Dominant vectors of nonnegative matrices may also be useful in the context of voting problems [106], where they allow to get out of the Condorcet paradox. A nonnegative matrix is constructed from pairwise scores of dominances for the item to be ranked and the dominant eigenvector of the matrix gives a ranking of the items. Note that such voting problems may occur for instance in the context of ranking sport teams [73, 51] or graduate programs [107].

We choose in this work to study the problem of ranking the nodes of a graph, with PageRank [25] or HITS [76], and the problem of allocating similarity scores to pairs of nodes of two directed graphs, as considered by Blondel et al. [17] and Melnik et al. [93]. These four algorithms use similar link analysis methods: the scores are computed iteratively, by considering that the score of each node (or pair of nodes) is reinforced by the score of its neighbors. These mutual reinforcements of the scores are expected to converge to an equilibrium. So the authority or similarity scores are the equilibrium of a kind of flow propagating in the graphs.

Representing the graphs by their adjacency matrix, these algorithms may be seen as simple linear iterative processes. More precisely, a nonnegative matrix can be constructed from the adjacency matrix of the considered graph(s). At each step, the scores are represented by a nonnegative vector. Then, the mutual reinforcement relation can simply be expressed as the iterated application of the matrix on the successive vectors of scores. If the process converges to an equilibrium, the vector of scores at the equilibrium is simply a dominant nonnegative eigenvector of the considered matrix.

The Perron–Frobenius theory deals with dominant eigenvalues and eigenvectors of nonnegative matrices. For instance, existence, uniqueness and convergence to the vector of scores of the PageRank algorithm is a direct application of the Perron–Frobenius theory. This theory also allows one to understand why the mutual reinforcement relations considered in the HITS algorithm and the similarity measure of Blondel et al. do not necessarily lead directly to a unique nonnegative score distribution.

Nonlinear generalizations of the Perron–Frobenius theory can also be useful in the study of nonlinear variants of such eigenvector based algorithms. For instance, for computing their similarity scores, Melnik et al. slightly modify the linear iterative process in order to cope with the non uniqueness of the score vector at equilibrium. The nonlinear iteration resulting from this can be studied using results of the nonlinear Perron–Frobenius theory.

Let us finally emphasize that the literature about large graphs and networks is very diverse. As we have already mentioned, there is a wide variety of situations leading to the study of graphs [99]: besides information networks like the Web or graphs constructed from a dictionary [17, 109], one may analyze social networks (e.g. graphs of friendship relations or of phone calls [102]), technological networks (e.g. the Internet [45] or the electric power grid) or biological networks (e.g. neural networks). The variety comes also from the methods used to analyze these networks. We are here concerned by information retrieval methods based on the computation of dominant eigenvectors of nonnegative matrices and describing an equilibrium. One may also use spectral methods to project the nodes of the graphs in a vector space of lower dimension in order for instance to compare different graphs for matching them approximately [30, 48]. Other methods come for instance from statistical physics. Statistics as the distribution of node degrees or clustering coefficients, based for instance on the proportion of triangles in the network, may be useful to compare the network with classes of random networks and therefore highlight some of its distinctive features [3, 99].

1.3 A foretaste of the thesis

We now present the main parts of this thesis and our main contributions. In the preliminary Chapter 2, we first review classical notions about graphs and nonnegative matrices that will be useful throughout the thesis, e.g. irreducibility of a nonnegative matrix, classes of a graph, the Perron–Frobenius Theorem, random walks. We then present Brin and Page's PageRank [25], Kleinberg's hub and authority scores [76], and definitions of similarity between nodes of graphs proposed by Blondel et al. [17] and Melnik et al. [93].

An affine eigenvalue problem Chapter 3 is devoted to the study of the fixed points of the iteration representing the reinforcement of the similarity scores as defined by Melnik et al. [93]. As we already said, they modify the normalized linear iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k)}{\|A\boldsymbol{x}(k)\|}$$

in a normalized affine iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k)+\boldsymbol{b}}{\|A\boldsymbol{x}(k)+\boldsymbol{b}\|},$$

with A a nonnegative $n \times n$ matrix, $\boldsymbol{b}$ a nonnegative vector and $\|\cdot\|$ a monotone vector norm. In their paper, A, $\boldsymbol{b}$ and $\|\cdot\|$ are, respectively, a matrix constructed from the adjacency matrix of the products of graphs considered, the vector of all ones and the ℓ_∞ norm. The fixed points of this iteration can be studied by considering an equivalent conditional affine eigenvalue problem on the nonnegative orthant,

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \in \mathbb{R}^n_{\geq 0}, \quad \|\boldsymbol{x}\| = 1.$$

We study the solutions of this eigenvalue problem from two points of view. The first one follows the study of M-matrix equations as initiated by Carlson [32] and Victory [116]. This analysis is based on accessibility relations in the graph corresponding to the matrix A for nodes in the support of $\boldsymbol{b}$. In Theorem 3.7, given a solution satisfying particular graph-theoretic conditions, we study the existence and properties of other possible solutions of the considered eigenvalue problem.

In the second approach, we show the link between the solution of the conditional affine eigenvalue problem and the solution of a maximizing problem of the spectral radius of a rank-one perturbation of the matrix A. More precisely, let $\boldsymbol{c}_*$ be a maximizer of $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ on the unit sphere of the dual norm. If $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) > \rho(A)$, then the solution of the conditional affine eigenvalue problem is unique. Moreover this solution $(\lambda_*, \boldsymbol{x}_*)$ is the Perron pair of the matrix $A + \boldsymbol{b}\boldsymbol{c}_*^T$, that is, $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_* = (A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_*$. We have also that $(\boldsymbol{x}_*, \boldsymbol{c}_*)$ is a dual pair with respect to the considered norm. These results are stated in Theorem 3.19.

We then unify both approaches. This leads us to a characterization of the solutions of the conditional affine eigenvalue problem on the nonnegative orthant, which we summarize in Theorem 3.1.

Melnik et al. introduced the normalized affine iteration as a small perturbation of the simple linear iteration. In their case, since the vector $\boldsymbol{b}$ has only positive entries, the fixed point of the normalized affine iteration is indeed unique. A natural question then arises: which fixed point of the linear iteration (or equivalently which eigenvector of A) will be approached by the fixed point of the normalized affine iteration when the norm of the vector $\boldsymbol{b}$ is sufficiently small? In Proposition 3.51, we prove that, when $\boldsymbol{b}$ is sufficiently small in norm and when A is symmetric, the eigenvector of A approached by the fixed point of the normalized affine iteration is the orthogonal projection of $\boldsymbol{b}$ on the invariant subspace of A associated with $\rho(A)$.

Maximizing PageRank via outlinks In Chapter 4, we move to the study of Brin and Page's PageRank algorithm for ranking web pages. PageRank is behind the success of the most famous web search engine at this time, Google. What Google exactly does is a well kept secret. But Google's team claims that PageRank ranking remains at the heart of the search engine. So it is quite natural that some web masters try to increase the PageRank of their web pages in order to be better referenced on Google. Remember that PageRank is a story of links. So, the only way users have to alter PageRank scores is to modify the link structure of the web graph.

PageRank scores are known to be robust with respect to small perturbations of the web graph, at least when these perturbations concern web pages with a low PageRank score [16, 75, 100]. In other words, the addition or deletion of a few hyperlinks does not change a lot the PageRank scores of the web pages. But! The value of PageRank scores does not matter. What is crucial for a web page to be well referenced on Google is its ranking. Indeed, even a small change in the value of PageRank scores can lead to important modifications in the ranking of the web pages. So PageRank is not rank-stable with respect to small modifications in the link structure of the web graph [86].

A new link from somebody's web page to your page, without any other change in the web graph, will increase the PageRank of your page [64]. Therefore, some web masters buy hyperlinks to their web pages from link farms or make an alliance with other web masters by trading a link for a link [10, 53].

The impact on your PageRank of links leaving your web page is not so evident. However, these outgoing links are the only control you directly have on the web graph. That is why we were interested in the maximization of the PageRank via outlinks. In Theorem 4.21, we characterize optimal link structures for a set of web pages for which we want to maximize the sum of the PageRank scores. We show that, in order to maximize its PageRanks sum, the web site must be organized as follows. The internal link structure, i.e., links between nodes of the web site, must consist in a forward chain of links together with all possible backward links. The external outlink structure, i.e., links from the nodes of the web site to the outside, must consist of a unique outlink, starting from the last node of the chain. In Theorems 4.17 and 4.19, we also give results in the case where for instance the internal link structure is given and we want to optimize the external outlink structure. In all cases, we make the assumption that every node of the web site must keep an access to the rest of the Web. We of course explain and justify this assumption.

We also look at some related questions. For instance, the addition of external inlinks, i.e., links from the rest of the graph to pages of the web site, is not always profitable for the sum of the PageRanks of this web site, in contrast to the case of an inlink to a single page.

Self-validating web rankings A web page has a high PageRank if it is pointed to by many pages with a high PageRank. PageRank has also a stochastic interpretation: it measures how often a given web page would be visited by a random web surfer. The PageRank model considers a random web surfer that follows randomly hyperlinks of the web graph and sometimes gets bored and zaps to a web page taken at random in the graph. Several authors proposed variants of this model in order to have a more realistic model of the behavior of web surfers. For instance, the possible use by web surfers of the back button of the web browser can be taken in account [22, 89, 112, 43], as well as the bookmarks or the query of the random web surfer [56, 57, 66, 105].

In Chapter 5, we start from the idea that the web ranking itself may influence the behavior of web surfers. We then propose a simple model that takes into account the mutual influence between web ranking and web surfing.

Consider a web surfer that regularly consults the web search engine. Maybe he has noticed that for a wide variety of scientific or cultural queries on Google, the free encyclopedia Wikipedia arrives in the top of the results' list. Similarly, when he searches some LaTeX package with the web search engine, he is often directed towards some page of the CTAN web site. Suppose this web surfer is now surfing on the Web and is currently visiting a web page pointing to two other pages. The first one is from the Wikipedia web site and the second one from a web site he has never heard before. It seems realistic to think he will more likely click to the hyperlink to the web page from the reputed Wikipedia instead of to the other unknown web page.

The PageRank iteration is a simple linear iteration $\boldsymbol{\pi}(k+1)^T = \boldsymbol{\pi}(k)^T G$, where G is a stochastic matrix obtained from the adjacency matrix of the web graph. These iterates converge to a vector describing the stationary distribution of the random walk of the web surfer. In our model, we suppose that the current web ranking, say $\boldsymbol{\pi}(k)$ induces a particular random walk on the web graph. This walk has a stationary distribution, represented by a vector that we denote $\boldsymbol{u}_T(\boldsymbol{\pi}(k))$. So $\boldsymbol{u}_T(\boldsymbol{\pi}(k))$ is the left dominant eigenvector of some stochastic matrix $G_T(\boldsymbol{\pi}(k))$. This stationary distribution is then used to update the web pages ranking. This leads us to iterations such as $\boldsymbol{\pi}(k+1) = \boldsymbol{u}_T(\boldsymbol{\pi}(k))$. The parameter $T > 0$, called temperature, is fixed and represents the confidence of the web surfer in the web ranking. We call T-PageRank the limit of $\boldsymbol{\pi}(k)$ when k tends to infinity, if it exists. We also consider the iteration defined by $\tilde{\boldsymbol{\pi}}(k+1)^T = \tilde{\boldsymbol{\pi}}(k)^T G_T(\tilde{\boldsymbol{\pi}}(k))$, which is similar from a computational point of view to the power method.

The study of both iterations uses nonlinear Perron–Frobenius theory [80, 101]. We prove in Theorem 5.14 and 5.15 the existence and uniqueness of the T-PageRank when the temperature is large enough. Under some assumptions, we also show that this T-PageRank can be computed with a method analogous to the power method. On the other hand, when considering small values of the temperature, i.e., when the web surfer has a strong confidence in the web ranking, several T-PageRanks exist, depending on the choice of the initial ranking. In some cases, the T-PageRank only strongly reinforces the initial belief about the importance of the web pages given by the initial web ranking. We prove in Theorem 5.19 that such a self-validating effect is a general feature.

We also study the following variant of our model. When surfing on the Web by following hyperlinks, the web surfer chooses uniformly the next page to visit, as in the standard PageRank model. But, when he gets bored of following links, instead of zapping to a random web page as in the standard PageRank, we suppose that he jumps to the search engine's web page. From this search engine's page, he is more likely to move to some well ranked page rather than to a badly ranked one. So, in this variant, our web surfer takes the web ranking into account only when visiting the search engine's web page. For such variants of the model, we prove similar results about uniqueness or multiplicity of T-PageRanks, depending on the temperature.

1.4 Related publications

Chapter 3 about the affine eigenvalue problem, is in a large part based on a joint work with Vincent Blondel and Paul Van Dooren published in *Linear Algebra and its Applications* [20]. This part corresponds more or less to Sections 3.3 and 3.4, that is, the approach considering the maximization of the spectral radius. The first results of Section 3.5 also appear in this paper. But we did not at this time see precisely the relation between our "Path Condition" [20] and the literature about M-matrix equations [32, 116]. Subsequent research led us to the results of Section 3.2 as well as most of Section 3.5. Results of Section 3.6 about the approximation by normalized affine iterations where presented at the SIAM Meeting 2005, New Orleans. Finally, some previous work about normalized affine iterations was presented at the Benelux Meeting 2004, Helvoirt [19] and at MTNS 2004, Leuven [18]. In this last reference, we also analyzed a variant of the normalized affine iteration.

Chapter 4 about the maximization of the PageRank, is based on a joint work with Cristobald de Kerchove and Paul Van Dooren. Preliminary results, together with results about collusion of two web pages, were announced at ILAS 2006, Amsterdam, and the Benelux Meeting 2007, Lommel [40]. A full paper has been accepted for publication in *Linear Algebra and its Applications* [39] for a special issue devoted to ILAS 2006.

Chapter 5 about self-validating web rankings, is based on a joint work with Marianne Akian and Stéphane Gaubert that has been submitted to the *SIAM Journal of Matrix Analysis and Application* [1]. Part of the results were announced at POSTA 2006, Grenoble, and appeared in its proceedings in the *Lecture Notes in Control and Information Sciences* [2].

Chapter 2

Graphs, matrices, ranking and similarity scores

In this preliminary chapter, we first present classical notions about graphs and nonnegative matrices (see [14, 61, 74, 110] for a detailed exposition), as well as some results about particular nonlinear maps for which important properties of nonnegative matrices can be extended [80, 81, 101]. We then present the PageRank and HITS approaches to rank web pages, based on the link structure of the Web [25, 76], and compare two definitions of similarity between the nodes of two graphs [17, 93].

2.1 Graphs and nonnegative matrices

2.1.1 Graphs

A *(directed) graph* $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is defined by a finite set of *nodes* $\mathcal{N}$ and a set of ordered pairs $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, called *(directed) edges* or *links*[1]. Typically, we will consider $\mathcal{N} = \{1, \ldots, n\}$.

A link (i, j) is said to be an *outlink* for node i and an *inlink* for node j.

[1] In this work, we will only consider directed graphs, so we will often simply write graph, edge or link to refer to a directed graph, edge or link. In general, when defining mathematical objects, we will use parenthesis for terms that may be omitted when the context is sufficiently clear.

By

$$j \leftarrow i,$$

we mean that j belongs to the set of *children* or *out-neighbors* of i, that is, $j \in \{k \in \mathcal{N} \colon (i,k) \in \mathcal{E}\}$. Similarly, $i \rightarrow j$ means that i belongs to the set of *parents* or *in-neighbors* of j, that is, $i \in \{k \in \mathcal{N} \colon (k,j) \in \mathcal{E}\}$. The *outdegree* $\boldsymbol{d}_i$ of a node i is its number of children, that is,

$$\boldsymbol{d}_i = |\{k \in \mathcal{N} \colon (i,k) \in \mathcal{E}\}|.$$

A *(directed) path* of *length* ℓ from a node i_0 to a node i_ℓ is a sequence of nodes $\langle i_0, i_1, \ldots, i_\ell \rangle$ such that $(i_k, i_{k+1}) \in \mathcal{E}$ for every $k = 0, 1, \ldots, \ell - 1$.

Accessibility relations and classes A node i *has an access to* a node j if there exists a directed path from i to j. By convention we say that a node i always has an access to itself (even if $(i,i) \notin \mathcal{E}$). Two nodes i and j *communicate* if they have an access to each other. The communication relation defines equivalence classes: two nodes belong to the same *class* if they communicate. A class $\mathcal{I}$ *has an access to* a class $\mathcal{J}$ if some node $i \in \mathcal{I}$ has an access to some node $j \in \mathcal{J}$. A class is a *final class* if it has access to no other class in the graph. A class is an *initial class* if no other class has an access to it. These notions are illustrated in Figure 2.1. Note that a graph always has at least one final and one initial class.

In this thesis, we will also say that a node i *has an access to a set of nodes* $\mathcal{J}$ if i has an access to at least one node $j \in \mathcal{J}$. Similarly, a class $\mathcal{I}$ has an access to a set of nodes $\mathcal{J}$ if any $i \in \mathcal{I}$ has an access to at least one node $j \in \mathcal{J}$.

A graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is a *strongly connected graph* if it consists of a unique class, that is, every two nodes $i, j \in \mathcal{N}$ communicate. A strongly connected graph is moreover *aperiodic* if, for every node $i \in \mathcal{N}$, the greatest common divisor of the length of all paths from i to i is equal to 1.

Subgraphs For any subset of nodes $\mathcal{I} \subseteq \mathcal{N}$, the *(induced) subgraph* is $\mathcal{G}_\mathcal{I} = (\mathcal{I}, \mathcal{E}_\mathcal{I})$, where

$$\mathcal{E}_\mathcal{I} = \{(i,j) \in \mathcal{E} \colon i, j \in \mathcal{I}\}$$

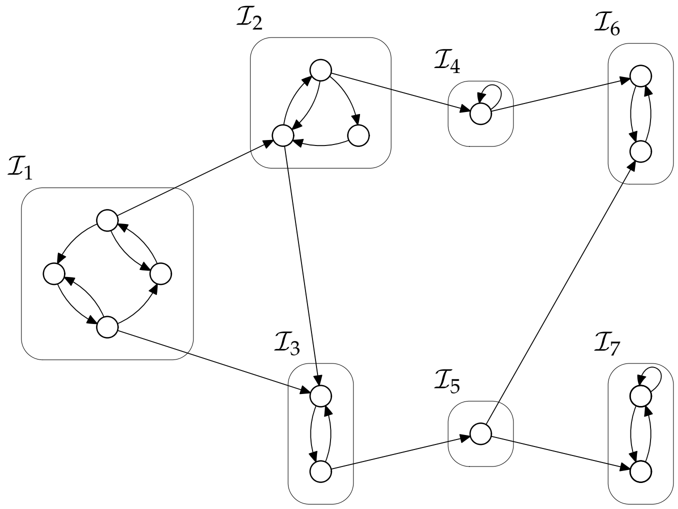


Figure 2.1: Graph with seven classes. Class $\mathcal{I}_1$ is initial and classes $\mathcal{I}_6$ and $\mathcal{I}_7$ are final. Class $\mathcal{I}_3$ has an access to classes $\mathcal{I}_5$, $\mathcal{I}_6$ and $\mathcal{I}_7$ but has no access to classes $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_4$.

is the set of *internal links*.

Of course, all the notions presented above for graphs also apply to subgraphs. For instance, the subgraph $\mathcal{G}_\mathcal{I} = (\mathcal{I}, \mathcal{E}_\mathcal{I})$ is strongly connected if every two nodes $i, j \in \mathcal{I}$ communicates by paths contained entirely in $\mathcal{I}$.

2.1.2 Nonnegative matrices and Perron–Frobenius theory

Let us first begin with some notations. By $\boldsymbol{e}_i$ we denote the i^{th} column of the identity matrix I. For a set of nodes $\mathcal{I} \subseteq \{1, \ldots, n\}$, the vector $\boldsymbol{e}_\mathcal{I}$ is the vector of $\mathbb{R}^n$ with a 1 in the entries of $\mathcal{I}$ and a 0 elsewhere. The vector of all ones is denoted by $\mathbf{1}$. In general, for a subset $\mathcal{I} \subseteq \{1, \ldots, n\}$, we denote by $\boldsymbol{x}_\mathcal{I}$ the corresponding subvector of a vector $\boldsymbol{x} = [\boldsymbol{x}_i]_{i=1}^n$ and by $M_\mathcal{I}$ the corresponding principal submatrix of a square matrix $M = [M_{ij}]_{i,j=1}^n$. For two different sets $\mathcal{I}, \mathcal{J} \subseteq \{1, \ldots, n\}$, we also denote by $M_{\mathcal{I}\mathcal{J}}$ the corresponding submatrix, possibly rectangular.

A matrix $A = [A_{ij}]_{i,j=1}^{m,n}$ is a *nonnegative matrix* if all its entries A_{ij}

are nonnegative. In this case we write $A \geq 0$ or $A \in \mathbb{R}^{m\times n}_{\geq 0}$. Similarly, a matrix A is a *positive matrix*, denoted by $A > 0$ or $A \in \mathbb{R}^{m\times n}_{>0}$, if all its entries are positive. Finally, we denote by $A \gneqq 0$ a nonzero nonnegative matrix. The same notations apply to vectors. A nonnegative matrix will be assumed to be square, unless it is specified otherwise.

A nonnegative matrix is said to be *irreducible* if there exists *no* permutation matrix P such that $P^T MP$ is block upper triangular with at least two diagonal blocks. By convention, one-by-one zero matrices are considered to be irreducible. The number of eigenvalues of largest magnitude of an irreducible matrix is called its *index of cyclicity*. A *primitive matrix* is an irreducible matrix with an index of cyclicity equal to one, that is, an irreducible matrix with only one eigenvalue of largest magnitude.

Perron–Frobenius theory Perron–Frobenius theory deals with nice properties of nonnegative matrices. In particular, the spectral radius

$$\rho(A) = \max\{|\lambda| \colon Ax = \lambda x \text{ for some } x \neq 0\}$$

is always an eigenvalue of the nonnegative matrix A and is called the *Perron root* of A. Moreover, there always exists at least one nonnegative eigenvector associated to the Perron root, i.e., a vector $\boldsymbol{x} \gneqq 0$ such that $A\boldsymbol{x} = \rho(A)\boldsymbol{x}$. Such a vector is called a *(right) Perron vector*. Also, a nonnegative matrix always has at least one *left Perron vector*, i.e., a vector $\boldsymbol{y} \gneqq 0$ such that $\boldsymbol{y}^T A = \rho(A)\boldsymbol{y}^T$.

Theorem 2.1 (Perron–Frobenius, see [14, 61])**.** *Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be a nonnegative matrix. Then*

- (a) *$\rho(A)$ is an eigenvalue of A;*
- (b) *A has at least one Perron vector;*
- (c) *every positive eigenvector of A corresponds to the eigenvalue $\rho(A)$.*

Moreover, if A is irreducible, then

- (d) *$\rho(A) > 0$;*
- (e) *A has exactly one Perron vector and this vector is positive;*
- (f) *every eigenvalue of maximum magnitude is an algebraically simple eigenvalue of A;*

(g) *the eigenvalues of maximum magnitude are equally spaced on a circle centered in* 0 *and of radius* $\rho(A)$ *in the complex plane.*

Moreover, if A *is primitive, then*

(h) $\rho(A)$ *is the unique eigenvalue of maximum modulus;*

(i) $\lim_{k\to\infty}(\rho(A)^{-1}A)^k = \boldsymbol{x}\boldsymbol{y}^T$, *where* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are respectively the right and left Perron vectors of* A, *scaled such that* $\boldsymbol{y}^T\boldsymbol{x} = 1$.

Graphical representation As Perron–Frobenius Theorem shows, properties like positivity or uniqueness of the Perron vector of a nonnegative matrix essentially depend on its irreducibility or primitivity properties. These notions have nice graphical characterizations. The *(directed) graph* $\mathcal{G}(A) = (\mathcal{N}, \mathcal{E})$ of a matrix $A \in \mathbb{R}^{n\times n}_{\geq 0}$ is defined by $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{E} = \{(i,j) \colon A_{ij} \neq 0\}$. A nonnegative matrix is irreducible if and only if its graph is strongly connected. Its index of cyclicity is equal, for every node $i \in \mathcal{N}$, to the greatest common divisor of all paths in $\mathcal{G}(A)$ from i to i. In particular, a matrix is primitive if and only if its graph is aperiodic. Moreover, a positive matrix is always primitive.

Frobenius normal form The parallelism between nonnegative matrices and graphs goes further. Up to a symmetric permutation of its rows and columns, a nonnegative matrix can always be written in a block upper triangular form whose diagonal blocks are irreducible. This canonical form is the *Frobenius normal form* and is not necessarily unique. It is strongly related to the notions of accessibility and classes for a graph. Indeed, each irreducible diagonal block of the Frobenius normal form of a matrix A corresponds to a class of the graph $\mathcal{G}(A)$.

Let $\mathcal{I}_1, \ldots, \mathcal{I}_s$ be the classes of $\mathcal{G}(A)$ and let them be ordered according to the triangular structure of the Frobenius normal form of A, i.e., such that $A_{\mathcal{I}_k\mathcal{I}_\ell} = 0$ if $k > \ell$. It is easily proved that a node $i \in \mathcal{N}$ has an access to $j \in \mathcal{N}$ in $\mathcal{G}(A)$ if and only if there exists some $k \in \mathbb{N}$ such that $(A^k)_{ij} \neq 0$. So, clearly, a class $\mathcal{I}_k$ of the graph $\mathcal{G}(A)$ can not have an access to a class $\mathcal{I}_\ell$ with $\ell < k$. So the class $\mathcal{I}_1$ must be an initial class and the class $\mathcal{I}_s$ must be a final class. An example of nonnegative matrix written in a Frobenius normal form is given in Figure 2.2.

$$
A = \left(\begin{array}{cccc|ccc|cc|c|c|cc|cc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}\right)
$$

Figure 2.2: Nonnegative matrix under its Frobenius normal form. Its graph $\mathcal{G}(A)$ is the graph represented in Figure 2.1. The matrix A has seven irreducible diagonal blocks $A_{\mathcal{I}_1}, \dots, A_{\mathcal{I}_7}$, corresponding to the classes $\mathcal{I}_1, \dots, \mathcal{I}_7$ of $\mathcal{G}(A)$. As seen from the block triangular form, a class $\mathcal{I}_k$ does not have an access to any class $\mathcal{I}_\ell$ with $\ell < k$.

By extension, we talk about *classes of a matrix* A to refer to the classes of the corresponding graph $\mathcal{G}(A)$. This allows one to work for instance with the spectral radius of the corresponding diagonal blocks. A class $\mathcal{I}$ of A is a *basic class* if $\rho(A_{\mathcal{I}}) = \rho(A)$. Note that a nonnegative matrix always has at least one basic class. Indeed, its eigenvalues are these of the diagonal blocks of the block triangular structure of the Frobenius normal form. So at least one of these diagonal blocks must have $\rho(A)$ as spectral radius.

Several extensions of the Perron–Frobenius Theorem are based on the Frobenius normal form of a matrix and its graphical representation with classes and accessibility relations between them. We refer the interested reader to Schneider's survey [108].

***M*-matrices** When working with a nonnegative matrix A, it is often useful to consider matrices of the form $\lambda I - A$ for some scalar λ. An M-

matrix is a matrix of the form $\lambda I - A$ for which $\lambda \geq \rho(A)$. Nonsingular M-matrices have the nice property to have a nonnegative inverse, as stated by the following lemma.

Lemma 2.2 (see [14, 62])**.** *Let A be a nonnegative matrix and let $\lambda \geq \rho(A)$. Then, $(\lambda I - A)$ is nonsingular and $(\lambda I - A)^{-1} \geq 0$ if and only if $\lambda > \rho(A)$.*

It is sometimes useful to characterize matrices $(\lambda I - A)^{-1}$ when λ approaches $\rho(A)$. The *index* of a square matrix $M \in \mathbb{R}^{n\times n}$, denoted $\operatorname{ind}(M)$, is the smallest nonnegative integer k such that $\operatorname{rank}(M^{k+1}) = \operatorname{rank}(M^k)$. Clearly, the index of a matrix is zero if and only if this matrix is nonsingular. The *Drazin inverse* of a square matrix M, denoted M^D, is the unique solution X to the equations $M^{k+1}X = M^k$, $XMX = X$, $MX = XM$, where k is the index of M. The following lemma particularizes results of Meyer [95, Theo. 3.1 and 3.2].

Lemma 2.3 (Meyer [95])**.** *Let $M \in \mathbb{R}^{n\times n}$. Then, for any $m \in \mathbb{N}$, the limit $\lim_{\varepsilon\to 0} \varepsilon^m (M + \varepsilon I)^{-1}$ exists and is non zero if and only if $m = \operatorname{ind}(M)$. In particular, if $\operatorname{ind}(M) > 0$,*

$$\lim_{\varepsilon\to 0} \varepsilon^{\operatorname{ind}(M)} (M + \varepsilon I)^{-1} = (-1)^{\operatorname{ind}(M)-1} (I - MM^D) M^{\operatorname{ind}(M)-1}.$$

Moreover, let $\boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{b} \neq 0$ and let $r \leq \operatorname{ind}(M)$ the smallest nonnegative integer such that $M^r\boldsymbol{b}$ is in the range of $M^{\operatorname{ind}(M)}$. Then, for any $m \in \mathbb{N}$, $\lim_{\varepsilon\to 0} \varepsilon^m (M + \varepsilon I)^{-1}\boldsymbol{b}$ exists and is nonzero if and only if $m = r$. In particular, if $r > 0$,

$$\lim_{\varepsilon\to 0} \varepsilon^r (M + \varepsilon I)^{-1}\boldsymbol{b} = (-1)^{r-1} (I - MM^D) M^{r-1}\boldsymbol{b}.$$

For instance, if A is an irreducible nonnegative matrix, a stochastic matrix or a symmetric nonnegative matrix, then the index of $M = \rho(A)I - A$ is equal to 1 (see Lemma 1 in [96] and Theorem 8.4.2 in [14]). Moreover, if A is an irreducible nonnegative matrix and $M = \rho(A)I - A$, then $(I - MM^D) = \boldsymbol{x}\boldsymbol{y}^T$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are respectively the right and left Perron vectors of A such that $\boldsymbol{y}^T\boldsymbol{x} = 1$ [96].

Adjacency matrix We have seen how a graph can be associated to a matrix. It is also often useful to represent a graph by its adjacency matrix. The *adjacency matrix* $A = [A_{ij}]_{i,j \in \mathcal{N}}$ of a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with $\mathcal{N} = \{1, \dots, n\}$, is given by

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that sometimes, one has to consider a *weighted graph*, that is, a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with a *weight function* $w \colon \mathcal{E} \to \mathbb{R}$ associated. The *weighted adjacency matrix* $A = [A_{ij}]_{i,j \in \mathcal{N}}$ of such a graph is then defined as $A_{ij} = w(i,j)$ if $(i,j) \in \mathcal{E}$ and $A_{ij} = 0$ otherwise.

2.1.3 Random walks and stochastic matrices

Let us now consider a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with $\mathcal{N} = \{1, \dots, n\}$ and choose some initial node. Suppose that at each time step, we choose at random a child of the current node and move to this node. The resulting sequence of nodes is called a *random walk*. In its simplest form, the probability of moving from a node i to a node $j \leftarrow i$ is equal to $\boldsymbol{d}_i^{-1}$, where $\boldsymbol{d}_i$ is the outdegree of i. One can also consider a random walk on a graph with weighted edges, where we move from i to j with a probability proportional to the weight of the edge (i,j).

The concept of random walk with weighted edges is equivalent to the concept of finite Markov chain. Let us consider a random process defined by a finite set of *states*, $\mathcal{N} = \{1, \dots, n\}$, and a sequence $X_0, X_1, X_2, \dots$ of *random variables* taking values in $\mathcal{N}$. This process is a *(finite) Markov chain* if the *transition probabilities* at step k depend only on the state at step k, that is

$$\mathbb{P}(X_{k+1} = j \mid X_k = i_k, \dots, X_0 = i_0) = \mathbb{P}(X_{k+1} = j \mid X_k = i_k).$$

A Markov chain can therefore be described by its directed graph, possibly weighted. This graphical representation is useful to determine for instance if the Markov chain is *aperiodic*, which is the case if the graph is strongly connected and aperiodic.

A Markov chain is also often described by its *transition matrix* $P = [P_{ij}]_{i,j \in \mathcal{N}}$ defined by $P_{ij} = \mathbb{P}(X_{k+1} = j \mid X_k = i)$ for all $i, j \in \mathcal{N}$. The transition matrix is a *(row) stochastic matrix*, that is, a square nonnegative matrix with all row sums equal to 1. Note that every stochastic matrix can represent a Markov chain. For all $k \in \mathbb{N}$, let $\boldsymbol{x}(k)^T = [\boldsymbol{x}(k)_i]_{i\in\mathcal{N}}$ be the *probability distribution vector* at step k, that is, $\boldsymbol{x}(k)_i = \mathbb{P}(X_k = i)$. The probability distribution vector is a *stochastic vector*, that is, a nonnegative vector with all row sums equal to 1.

The transition matrix of a Markov chain is useful to characterize the time evolution of probability distributions. Indeed, $\boldsymbol{x}(k+1)^T = \boldsymbol{x}(k)^T P$. A probability distribution vector $\boldsymbol{x}^T$ is then called a *stationary distribution* of the Markov chain if $\boldsymbol{x}^T = \boldsymbol{x}^T P$. Equivalently, $\boldsymbol{x}^T$ is called an *invariant measure* of the stochastic matrix P.

The Perron–Frobenius Theorem allows one to analyze stationary distributions of a Markov chain. Indeed, by Theorem 2.1(c), the spectral radius $\rho(P) = 1$, since $P\mathbf{1} = \mathbf{1}$, that is, $\mathbf{1}$ is a positive eigenvector of the stochastic matrix P. Therefore, invariant measures of a stochastic matrix correspond exactly to its left Perron vectors. In particular, if P is irreducible, then the Markov chain has a unique stationary distribution $\boldsymbol{x}^T$. If P is moreover primitive, then, for any initial probability distribution $\boldsymbol{x}(0)^T$, we converge to this stationary distribution, since

$$\lim_{k\to\infty} \boldsymbol{x}(k)^T = \boldsymbol{x}(0)^T \lim_{k\to\infty} P^k = \boldsymbol{x}(0)^T \mathbf{1}\boldsymbol{x}^T = \boldsymbol{x}^T,$$

by Theorem 2.1(i). So, let us particularize some of the results of Perron–Frobenius Theorem for stochastic matrices.

Corollary 2.4 (Perron–Frobenius for stochastic matrices). *Let $P \in \mathbb{R}^{n\times n}_{\geq 0}$ be a stochastic matrix. Then*

(a) $\rho(P) = 1$ and is an eigenvalue of P;
(b) P has at least one invariant measure;

Moreover, if P is irreducible, then

(e) P has exactly one invariant measure $\boldsymbol{x}^T$, and $\boldsymbol{x}^T$ is positive;

Moreover, if P is primitive, then

(i) $\lim_{k\to\infty} P^k = \mathbf{1}\boldsymbol{x}^T$.

Let us note that the irreducibility of the stochastic matrix P is not necessary in order to have a unique invariant measure. With the concept of classes introduced in Section 2.1.1, we can consider the following extension of Perron–Frobenius Theorem in the context of stochastic matrices. It is based on the fact that the basic classes of a stochastic matrix, that is, the classes with a spectral radius equal to one, are exactly its final classes. Although it is a particular case of Theorem 3.1 in [108], we give here a simple proof in this case.

Proposition 2.5 (see [108])**.** *Let P be a stochastic matrix. Then P has a unique invariant measure if and only if it has a unique final class.*

Proof. Suppose P has t final classes. Let P be written under a Frobenius normal form, with irreducible diagonal blocks $P_{\mathcal{I}_1}, \ldots, P_{\mathcal{I}_s}$. We can assume without loss of generality that the final classes of P are la belled $\mathcal{I}_{s-t+1}, \ldots, \mathcal{I}_s$. Let $\boldsymbol{x}^T$ be an invariant measure of P. So, for every $k = 1, \ldots, s$,

$$\boldsymbol{x}_{\mathcal{I}_k}^T = \boldsymbol{x}_{\mathcal{I}_k}^T P_{\mathcal{I}_k} + \sum_{i=1}^{k-1} \boldsymbol{x}_{\mathcal{I}_i}^T P_{\mathcal{I}_i \mathcal{I}_k}$$

Since $\rho(P_{\mathcal{I}_1}) < 1$, we first have $\boldsymbol{x}_{\mathcal{I}_1} = 0$. By induction, $\boldsymbol{x}_{\mathcal{I}_k} = 0$ for every $k = 1, \ldots, s-t$, since these classes $\mathcal{I}_k$ are not final and therefore not basic classes. Now, for every $k = s-t+1, \ldots, s$, we have $\boldsymbol{x}_{\mathcal{I}_k}^T = \boldsymbol{x}_{\mathcal{I}_k}^T P_{\mathcal{I}_k}$, since these classes are final. For each $k = s-t+1, \ldots, s$, the matrix $P_{\mathcal{I}_k}$ is irreducible and stochastic, so $\boldsymbol{x}_{\mathcal{I}_k}^T$ must be a scalar multiple of the unique invariant measure of $P_{\mathcal{I}_k}$. Therefore, $\boldsymbol{x}^T$ is an invariant measure of P if and only if $\boldsymbol{x}_{\mathcal{I}_k} = 0$ for every $k = 1, \ldots, s-t$ and if, $\boldsymbol{x}_{\mathcal{I}_{s-t+1}}^T, \ldots, \boldsymbol{x}_{\mathcal{I}_s}^T$ are respectively scalar multiples of the invariant measures of $P_{\mathcal{I}_{s-t+1}}, \ldots, P_{\mathcal{I}_s}$ such that $\boldsymbol{x}^T\mathbf{1} = 1$. The conclusion follows then readily. □

2.1.4 Some results of nonlinear Perron–Frobenius theory

Several applications in fields like economics [97] or biology [33] lead to the study of iterated nonlinear maps on cones. Questions arise there such as the existence and the uniqueness of a fixed point in the considered cone or in its interior, or the convergence of the iterates. Many results of the Perron–Frobenius theory can be extended to certain classes of nonlinear maps. See for instance [49, 94] for recent references. Here, we briefly present some nonlinear Perron–Frobenius theorems that we will use in the sequel. Note that we state here only the results we need for this thesis. Original theorems are stated in much more generality by their authors [80, 101]. In particular, we will restrict ourselves to $\mathbb{R}^n_{\geq 0}$.

Given a map $f\colon \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ on the nonnegative orthant and some vector norm $\|\cdot\|$ on $\mathbb{R}^n$, we are interested in the behavior of the iterates

$$x(k+1) = \frac{f(x(k))}{\|f(x(k))\|}, \tag{2.1}$$

for some initial vector $x \in S$, where $S = \{x \in \mathbb{R}^n_{\geq 0} \colon \|x\| = 1\}$. Obviously, a vector $x_* \in S$ is a fixed point of iteration (2.1) if and only if $(\|f(x_*)\|, x_*)$ is a solution of the following nonlinear conditional eigenvalue problem

$$\lambda x = f(x), \quad \lambda \in \mathbb{R}, \quad x \in S. \tag{2.2}$$

A vector x satisfying (2.2) for some $\lambda \in \mathbb{R}$ is called an *eigenvector* of the map f.

It is easily proved by Brouwer's Fixed Point Theorem that the eigenvalue problem (2.2) has at least one solution when the map f is continuous [97].

Hilbert's projective metric is a usual tool for proving further results about the fixed points and the convergence of iteration (2.1) [77, 80, 101]. The convergence to a unique and positive fixed point is generally proved with Banach's Fixed Point Theorem, by showing that the normalized map $f/\|f\|$ is a contraction for Hilbert's projective metric. Moreover, the norm $\|\cdot\|$ is usually supposed to be *monotone*, i.e., for all $x, y \in \mathbb{R}^n$, $|x| \geq |y|$ implies $\|x\| \geq \|y\|$.

Hilbert's (projective) metric d_H is defined as

$$d_H\colon \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{\geq 0}\colon (x, y) \mapsto \max_{i,j} \ln \frac{x_i y_j}{y_i x_j}.$$

For any vector norm $\|\cdot\|$, this metric defines a distance on $\operatorname{int}(S)$, where $S = \{x \in \mathbb{R}^n_{\geq 0} \colon \|x\| = 1\}$ and $\operatorname{int}(S) = S \cap \mathbb{R}^n_{>0}$ is its *relative interior*. Moreover, $(\operatorname{int}(\Sigma), d_H)$ is a complete metric space.

The following theorem characterizes a class of iterated maps over the nonnegative orthant that have a unique fixed point, that is moreover positive, and possesses a global convergence to it on the cone. The result is due to Krause [80] who states it in a more general framework.

Theorem 2.6 (Krause [80]). *Let $\|\cdot\|$ be a monotone vector norm and $S = \{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} : \|\boldsymbol{x}\| = 1\}$. Let $\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ be a map on the nonnegative orthant. Suppose there exist $\alpha, \beta > 0$ and $\boldsymbol{v} \in \mathbb{R}^n_{>0}$ such that $\alpha\boldsymbol{v} \leq \boldsymbol{f}(\boldsymbol{x}) \leq \beta\boldsymbol{v}$ for all $\boldsymbol{x} \in S$. Suppose also that $\lambda\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\lambda \in [0, 1]$ such that $\lambda\boldsymbol{x} \leq \boldsymbol{y}$; and that $\lambda\boldsymbol{f}(\boldsymbol{x}) < \boldsymbol{f}(\boldsymbol{y})$ if moreover $\lambda < 1$ and $\lambda\boldsymbol{x} \neq \boldsymbol{y}$.*

Then $\boldsymbol{f}$ has a unique eigenvector $\boldsymbol{x}_*$. As a consequence, $\boldsymbol{x}_*$ is the unique fixed point of the normalized map $\boldsymbol{u} : S \to \operatorname{int}(S) : \boldsymbol{x} \mapsto \boldsymbol{f}(\boldsymbol{x})/\|\boldsymbol{f}(\boldsymbol{x})\|$. Moreover, all the orbits of $\boldsymbol{u}$ converge to its unique fixed point $\boldsymbol{x}_*$.

Note that in Theorem 2.6, the map $\boldsymbol{f}$ is not required to be continuous. More precisely, the assumption implies that $\boldsymbol{f}$ is continuous on $\operatorname{int}(S)$, but not necessary on S. For instance, the map

$$\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0} : x \mapsto \begin{cases} 2(\boldsymbol{x} + \mathbf{1}) & \text{if } \boldsymbol{x} \in \mathbb{R}^n_{>0}, \\ \boldsymbol{x} + \mathbf{1} & \text{if } \boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \setminus \mathbb{R}^n_{>0}, \end{cases}$$

is not continuous but satisfies the assumptions of Theorem 2.6.

Krause also particularizes his result for the case of concave maps. The map $\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ is *concave* if $\boldsymbol{f}(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq \lambda\boldsymbol{f}(\boldsymbol{x}) + (1-\lambda)\boldsymbol{f}(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_{\geq 0}$ and $\lambda \in [0, 1]$.

Theorem 2.7 (Krause [80, 81]). *Let $\|\cdot\|$ be a monotone vector norm and $S = \{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} : \|\boldsymbol{x}\| = 1\}$. Let $\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ be a concave map such that $\boldsymbol{f}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \gneqq 0$.*

Then $\boldsymbol{f}$ has a unique eigenvector $\boldsymbol{x}_*$. As a consequence, $\boldsymbol{x}_*$ is the unique fixed point of the normalized map $\boldsymbol{u} : S \to \operatorname{int}(S) : \boldsymbol{x} \mapsto \boldsymbol{f}(\boldsymbol{x})/\|\boldsymbol{f}(\boldsymbol{x})\|$. Moreover, all the orbits of $\boldsymbol{u}$ converge to its unique fixed point $\boldsymbol{x}_*$.

Before presenting another result, due to Nussbaum, let us present some definitions useful when working with iterated maps on cones [101, 110]. A function $\boldsymbol{h}$ is *subhomogeneous* on a set U if $\lambda\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{h}(\lambda\boldsymbol{x})$ for all $\boldsymbol{x} \in U$ and every $\lambda \in [0, 1]$. It is *order-preserving* on U if $\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{h}(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in U$ such that $\boldsymbol{x} \leq \boldsymbol{y}$.

The following theorem is a simple formulation of the general Theorems 2.5 and 2.7 (or more precisely of Corollaries 2.2 and 2.5) of Nussbaum in [101].

Theorem 2.8 (Nussbaum [101]). *Let $\|\cdot\|$ be a monotone vector norm and $S = \{x \in \mathbb{R}^n_{\geq 0} \colon \|x\| = 1\}$. Let $f \colon \mathbb{R}^n_{>0} \to \mathbb{R}^n_{>0}$ be a continuous, order-preserving map which is subhomogeneous on $\operatorname{int}(S)$. Suppose moreover that if $x \in \operatorname{int}(S)$ is an eigenvector of f, then f is continuously differentiable on an open neighborhood of x and the matrix $f'(x)$ is nonnegative and irreducible.*

Then f has at most one eigenvector $x_* \in \operatorname{int}(S)$. As a consequence, the normalized map $u \colon \operatorname{int}(S) \to \operatorname{int}(S) \colon x \mapsto f(x)/\|f(x)\|$ has at most one fixed point.

Moreover, if f has an eigenvector $x_* \in \operatorname{int}(S)$, and if $f'(x_*)$ is a primitive matrix, then all the orbits of u converge to its unique fixed point x_*.

2.2 Pertinence and similarity scores

We are interested in this thesis by questions related to the problem of ranking web pages of the Web, that is, giving to each web page a score of pertinence. The other problem motivating this work is about measuring similarity between objects like words in a dictionary. In both cases, the relationships between the objects are used to assign them relevance or similarity scores. As we have seen, the modeling by graphs is quite natural for these problems: the objects are represented by nodes and links correspond to the relationships between the objects.

2.2.1 Ranking the Web with PageRank

PageRank's measure of web pages' relevance has been introduced in 1998 by Brin and Page [25, 103]. We here present its definition and well known interpretations as well as its basic calculation. For a more detailed exposition of PageRank and related topics, we refer to the surveys of Bianchini et al. [16] and of Langville and Meyer [83] and to the recent book of Langville and Meyer [84].

Votes of confidence The basic idea of PageRank is simple and intuitive: a hyperlink to a web page can be viewed as a *vote of confidence* to the pointed page (see Figure 2.3). In other words, *a web page is relevant if it is*

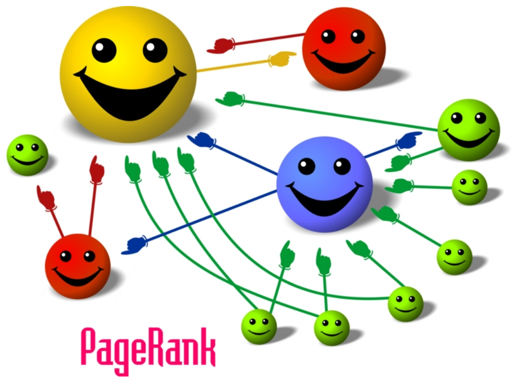


Figure 2.3: PageRank and votes of confidence among web pages as illustrated by Felipe Micaroni Lalli [82]. Reproduced with the permission of its author.

pointed to by many pages and these pages are themselves relevant and do not point to many other web pages.

So PageRank uses the graph structure of the Web in order to assign a pertinence score to each web page. Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be the graph representing the Web. Web pages are represented by the set of nodes $\mathcal{N} = \{1, \ldots, n\}$ and hyperlinks by the set of directed links $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, where there is a link $(i, j) \in \mathcal{E}$ if and only if there exists a hyperlink in page i pointing to page j.

Iterative computation of the scores The pertinence scores of the web pages are computed iteratively, according to the idea of mutual reinforcement of web pages' scores. Let $\boldsymbol{p}(k)_i$ be the pertinence score of page i, supposed to be nonnegative, at iteration step $k \in \mathbb{N}$. Let $\boldsymbol{d}_i = |\{j \in \mathcal{N} : (i, j) \in \mathcal{E}\}|$ be its outdegree. The mutual reinforcement of web pages' scores could be naively modeled as

$$\boldsymbol{p}(k+1)_j = \sum_{i \to j} \frac{\boldsymbol{p}(k)_i}{\boldsymbol{d}_i} \quad \text{for all } j \in \mathcal{N}. \tag{2.3}$$

In other words, the relevance score of page j is the sum of the scores of its parents in the graph, weighted by their outdegree. If these iterations converge to an equilibrium, these equilibrium scores are taken as pertinence scores.

For the sake of simplicity, we will make the assumption that *each node has at least one outlink*, i.e., $\boldsymbol{d}_i \neq 0$ for every $i \in \mathcal{N}$. Indeed, this can be done without loss of generality by making a preprocessing on the web graph [16, 84]. With this assumption, the system (2.3) has at least one equilibrium. Indeed, let $\boldsymbol{p}(k) = [\boldsymbol{p}(k)_i]_{i \in \mathcal{N}}$ be the vector of scores at iteration $k \in \mathbb{N}$ and $P = [P_{ij}]_{i,j \in \mathcal{N}}$ be a stochastic matrix defined by

$$P_{ij} = \begin{cases} \boldsymbol{d}_i{}^{-1} & \text{if } (i,j) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that P is a stochastic scaling of the adjacency matrix of $\mathcal{G}$, that is, $P = D^{-1}A$, where $D = \operatorname{diag}(\boldsymbol{d})$ is the diagonal *outdegree matrix*, with $\boldsymbol{d} = [\boldsymbol{d}_i]_{i \in \mathcal{N}}$. Then the system (2.3) can be written as

$$\boldsymbol{p}(k+1)^T = \boldsymbol{p}(k)^T P,$$

and by Corollary 2.4, P has at least one invariant measure, which is an equilibrium of the system.

Damping on the scores Some difficulties remain nevertheless with this definition of the equilibrium vector of pertinence scores. Indeed, without any additional assumption, the vector of pertinence scores may be not uniquely defined. This lack of uniqueness does not just occur in theory. Indeed, the graph of the Web is *not* strongly connected in practice [26]. There are for instance several web sites which do not have links to the rest of the Web. As these sets of nodes form final classes, the invariant measure of the matrix P is not unique, as stated in Proposition 2.5.

Brin and Page propose a simple solution to this problem: to "damp" the matrix P with a positive matrix. This allows one to work with a matrix for which the invariant measure is unique and for which the iterates converge for any initial vector of scores. With a *damping factor* $0 < c < 1$ and a stochastic *personalization vector* $\boldsymbol{z} \in \mathbb{R}^n_{>0}$, the positive and stochastic *Google matrix* G is defined as

$$G = cP + (1-c)\mathbf{1}\boldsymbol{z}^T.$$

So by Corollary 2.4, the matrix G has exactly one invariant measure $\boldsymbol{\pi}^T$ and moreover, for any initial stochastic vector $\boldsymbol{\pi}(0)^T$, the iterates defined by

$$\boldsymbol{\pi}(k+1)^T = \boldsymbol{\pi}(k)^T G \tag{2.4}$$

converge to $\boldsymbol{\pi}^T$. So the *PageRank vector* $\boldsymbol{\pi}$, the entries of which give the relevance score of the web pages, is defined by

$$\begin{aligned} \boldsymbol{\pi}^T &= \boldsymbol{\pi}^T G, \\ \boldsymbol{\pi}^T \mathbf{1} &= 1. \end{aligned} \tag{2.5}$$

The *PageRank of a node* i is the i^{th} entry $\boldsymbol{\pi}_i = \boldsymbol{\pi}^T \boldsymbol{e}_i$ of the PageRank vector.

In this context of votes of confidence, the use of a Google matrix G obtained by a rank-one correction of the matrix P can be interpreted as a damping of these votes of confidence by an initial score fixed by the web search engine. This appears clearly by considering the following reformulation of system (2.5),

$$\boldsymbol{\pi}^T = c\boldsymbol{\pi}^T P + (1-c)\boldsymbol{z}^T.$$

Random surfers on the Web PageRank has also a famous *stochastic interpretation*: it is an attempt to measure *how often a given web page would be visited by a web surfer* with the following behavior. Suppose indeed that the behavior of a web surfer can be modeled by the following Markov chain: the web surfer moves randomly on the web graph, using hyperlinks between pages with a probability c and otherwise *zapping* to some new page according to the personalization vector $\boldsymbol{z}$. Precisely, when visiting a page i, with probability c, the web surfer chooses randomly the next web page he will visit, among the pages referenced by page i, with the uniform distribution. And with probability $(1-c)$, he moves to any page j of the web graph, with a probability proportional to $\boldsymbol{z}_j$. Then the Google matrix G and the PageRank vector are respectively the transition matrix and the stationary distribution of this random walk. Note that, in this stochastic interpretation, the PageRank of a node is equal to the inverse of its *mean return time*, that is, $\boldsymbol{\pi}_i^{-1}$ is the mean number of steps a random surfer starting in node i will take for coming back to i (see [36, 74]).

Computation of PageRank Algorithms to compute the PageRank vector are essentially based on the power method. But the matrix G is a huge and dense matrix. In order to take advantage of the sparse structure of the matrix P, due to the sparse structure of the Web, the following reformulation of iteration (2.4)

$$\boldsymbol{\pi}(k+1)^T = c\boldsymbol{\pi}(k)^T P + (1-c)\boldsymbol{z}^T \tag{2.6}$$

is used to compute the successive iterations, with $\boldsymbol{\pi}(0)^T$ an arbitrary stochastic vector.

The rate of convergence of the power method depends on the ratio between the second and the first larger eigenvalues in magnitude. For PageRank, it is easily shown [42, 58, 83] that if $\{1, \lambda_2, \ldots, \lambda_n\}$ is the set of eigenvalues of the stochastic matrix P, then the eigenvalues of its rank-one correction G are $\{1, c\lambda_2, \ldots, c\lambda_n\}$. Moreover, we have already said that since the web graph has several final classes, the matrix P has several invariant measures so it has several eigenvalues equal to one. So the ratio characterizing the rate of convergence of iteration (2.4) and its equivalent formulation (2.6) is then equal to c. That means that the approximation error is like $\|\boldsymbol{\pi}(k) - \boldsymbol{\pi}\| = O(c^k)$. Brin and Page propose to take a damping factor equal to $c = 0.85$.

Nevertheless, with the huge dimensions of the matrix P, a few dozen iterations may take days and days of intensive computation. So algorithms are needed to compute PageRank more efficiently than with the power method. Several methods for accelerating the PageRank computation or its updating have been proposed, for instance extrapolation methods [23, 24, 72, 111], adaptive methods [70], aggregation methods [26, 63, 71, 85], and others [7]. See Berkhin's survey [13] and Langville and Meyer's book [84] for a general explanation of most of these methods.

2.2.2 Ranking the Web with hubs and authorities

About the same time as Brin and Page, Kleinberg was also thinking about a way to exploit the graph structure of the Web in order to rank web pages. But, instead of giving one score to each web page, he proposed to give *two* scores to each web page: an authority score and a hub score [76]. His algorithms is called HITS, for Hypertext Induced Topic Search.

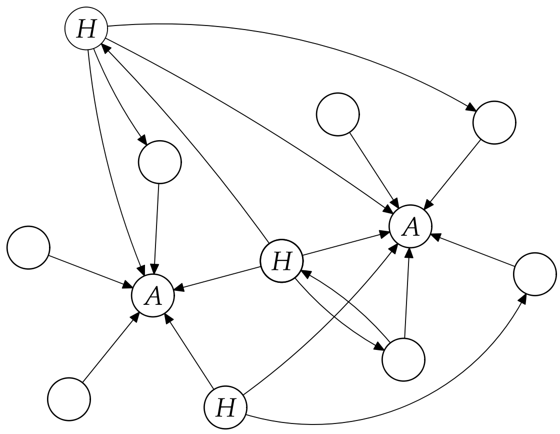


Figure 2.4: HITS basic idea: good authorities are pointed to by many good hubs and good hubs point to many good authorities. Here nodes that are good hubs are labeled H and good authorities are labeled A.

Hubs and authorities These two scores are mutually reinforced (see Figure 2.4): *a web page has a good authority score if it is pointed to by many pages having a good hub score. Conversely, a web page with a good hub score is a page that points to many good authority pages.*

Construction of a focused subgraph The Web is represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ as in the previous Section 2.2.1. But, unlike PageRank, Kleinberg proposed to construct a focused subgraph of the Web before computing the scores of web pages. For a particular query, a root set of web pages is first constructed with the help of a text-matching based search engine. Then, roughly, this set is extended with pages that link to or are linked to by pages of the set. This provides a base set of web pages $\mathcal{I} \subseteq \mathcal{N}$. The corresponding *focused subgraph* is $\mathcal{G}_{\mathcal{I}} = (\mathcal{I}, \mathcal{E}_{\mathcal{I}})$.

Iterative computation of the scores Let $\boldsymbol{a}(k)_j$ be the *authority score* of page j and $\boldsymbol{h}(k)_i$ be the *hub score* of page i at iteration's step $k \in \mathbb{N}$. These scores are nonnegative and normalized in order to have $\sum_{i \in \mathcal{N}} \boldsymbol{a}(k)_i^2 = 1$ and $\sum_{i \in \mathcal{N}} \boldsymbol{h}(k)_i^2 = 1$. Hub scores and authority scores are computed iteratively, according to the idea of mutual reinforcement between these two scores,

$$\begin{aligned} \tilde{\boldsymbol{a}}(k+1)_j &= \sum_{\substack{i \to j \\ i \in \mathcal{I}}} \boldsymbol{h}(k)_i \quad \text{for all } j \in \mathcal{I}, \\ \tilde{\boldsymbol{h}}(k+1)_i &= \sum_{\substack{j \leftarrow i \\ j \in \mathcal{I}}} \boldsymbol{a}(k)_j \quad \text{for all } i \in \mathcal{I}, \end{aligned} \tag{2.7a}$$

and then normalized

$$\begin{aligned} \boldsymbol{a}(k+1)_j &= \frac{\tilde{\boldsymbol{a}}(k+1)_j}{\left(\sum_{\ell \in \mathcal{I}} \tilde{\boldsymbol{a}}(k+1)_\ell^2\right)^{1/2}} \quad \text{for all } j \in \mathcal{I}, \\ \boldsymbol{h}(k+1)_i &= \frac{\tilde{\boldsymbol{h}}(k+1)_j}{\left(\sum_{\ell \in \mathcal{I}} \tilde{\boldsymbol{h}}(k+1)_\ell^2\right)^{1/2}} \quad \text{for all } i \in \mathcal{I}. \end{aligned} \tag{2.7b}$$

Let $\boldsymbol{a}(k) = [\boldsymbol{a}(k)_i]_{i \in \mathcal{I}}$ and $\boldsymbol{h}(k) = [\boldsymbol{h}(k)_i]_{i \in \mathcal{I}}$ be the vectors of authority and hub scores at iteration $k \in \mathbb{N}$ and let $A \in \mathbb{R}_{\geq 0}^{n \times n}$ be the adjacency matrix of graph $\mathcal{G}_\mathcal{I}$. Then the system (2.7) can be written as

$$\begin{aligned} \boldsymbol{a}(k+1) &= \frac{A^T \boldsymbol{h}(k)}{\|A^T \boldsymbol{h}(k)\|_2}, \\ \boldsymbol{h}(k+1) &= \frac{A \boldsymbol{a}(k)}{\|A \boldsymbol{a}(k)\|_2}. \end{aligned}$$

With initial vectors $\boldsymbol{a}(0)$ and $\boldsymbol{h}(0)$, the iterates at step $2k$ for $k \in \mathbb{N}$ can be expressed independently as

$$\begin{aligned} \boldsymbol{a}(2k) &= \frac{(A^T A)^k \boldsymbol{a}(0)}{\|(A^T A)^k \boldsymbol{a}(0)\|_2}, \\ \boldsymbol{h}(2k) &= \frac{(A A^T)^k \boldsymbol{h}(0)}{\|(A A^T)^k \boldsymbol{h}(0)\|_2}. \end{aligned} \tag{2.8}$$

The matrices $A^T A$ and $A A^T$ are called respectively the *authority* and *hub matrices* [41, 84]. Let us note that, when the adjacency matrix A is not weighted, i.e., has only 0 and 1 entries, the entry $(A^T A)_{ij}$ of the authority matrix is the number of nodes that point to both i and j, i.e., the number of parents that nodes i and j have in common. Similarly, the entry $(A A^T)_{ij}$ of the hub matrix is the number of children that nodes i and j have in common.

Convergence concerns It is not difficult to show that, for any vectors $\boldsymbol{a}(0)$ and $\boldsymbol{h}(0)$, the iterates (2.8) converge to limit vectors that in general depend on the chosen $\boldsymbol{a}(0)$ and $\boldsymbol{h}(0)$ [17, 76, 84]. Indeed, let $A = U\Sigma V^T$ be the singular value decomposition of A. Then $A^T A = V\Sigma^2 V^T$ and this matrix has only nonnegative eigenvalues. Moreover, since $A^T A$ is nonnegative, by Perron–Frobenius Theorem 2.1, $\rho(A^T A)$ is its only eigenvalue of maximum modulus (possibly multiple). Decomposing $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix}$, where V_1 corresponds to the eigenvectors of $A^T A$ associated to $\rho(A^T A)$ and V_2 to the rest of the spectrum Σ_2^2, we have $A^T A = \rho(A^T A) V_1 V_1^T + V_2 \Sigma_2^2 V_2^T$. Note that this decomposition is not always unique. Moreover, the projector $V_1 V_1^T$ is nonnegative even if V_1 is not necessarily nonnegative itself. The same development can be done for the hub matrix AA^T. Finally, one gets

$$\lim_{k\to\infty} \boldsymbol{a}(k) = \frac{V_1 V_1^T \boldsymbol{a}(0)}{\|V_1 V_1^T \boldsymbol{a}(0)\|_2} \quad \text{and} \quad \lim_{k\to\infty} \boldsymbol{h}(k) = \frac{U_1 U_1^T \boldsymbol{h}(0)}{\|U_1 U_1^T \boldsymbol{h}(0)\|_2}.$$

The *authority* and *hub score vectors* are then defined as these limit vectors $\boldsymbol{a} = \lim_{k\to\infty} \boldsymbol{a}(k)$ and $\boldsymbol{h} = \lim_{k\to\infty} \boldsymbol{h}(k)$. One sees that they do in general depend on the initial vectors $\boldsymbol{a}(0)$ and $\boldsymbol{h}(0)$. In fact, the authority score vector $\boldsymbol{a}$ is independent from $\boldsymbol{a}(0)$ if and only if $\rho(A^T A)$ is a simple eigenvalue of the nonnegative matrix $A^T A$. Since $A^T A$ is symmetric, its Frobenius normal form is block diagonal. So $\rho(A^T A)$ is a simple eigenvalue of $A^T A$ if and only if $A^T A$ has a unique basic class. Note finally that $\rho(A^T A)$ is a simple eigenvalue of $A^T A$ if and only if $\rho(AA^T) = \rho(A^T A)$ is a simple eigenvalue of AA^T.

Example 2.9. Irreducibility of matrices A and $A^T A$ are not directly linked. Consider for instance the cyclic matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

whose graph $\mathcal{G}(A)$ is a directed ring of three nodes. The matrix A is irreducible, but $A^T A = AA^T$ is the identity matrix, that has three basic classes. In fact, this is true for such cyclic matrices of any dimensions. Primitivity of A is also not sufficient to have irreducibility of $A^T A$, as can be seen by considering for instance

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad A A^T = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, let us note that for the reducible matrix $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, the authority and hub matrices $A^T A$ and $A A^T$ are positive. ◇

Choice of the initial score vectors Kleinberg proposed to take the normalized vector of all ones as initial authority and hub vectors: $\boldsymbol{a}(0) = \boldsymbol{h}(0) = \mathbf{1}/\|\mathbf{1}\|_2$. But in fact, in his original paper [76], the computation of $\boldsymbol{h}(k+1)$ is made with $\boldsymbol{a}(k+1)$ and not $\boldsymbol{a}(k)$. So the authority and hub score vectors obtained are

$$\boldsymbol{a} = \frac{V_1 V_1^T A^T \mathbf{1}}{\|V_1 V_1^T A^T \mathbf{1}\|_2} \quad \text{and} \quad \boldsymbol{h} = \frac{U_1 U_1^T \mathbf{1}}{\|U_1 U_1^T \mathbf{1}\|_2}.$$

Nevertheless, several subsequent papers [17, 41, 100] consider the iterates (2.8) with initial vectors $\boldsymbol{a}(0) = \boldsymbol{h}(0) = \mathbf{1}/\|\mathbf{1}\|_2$. This leads to the following authority and hub score vectors

$$\boldsymbol{a} = \frac{V_1 V_1^T \mathbf{1}}{\|V_1 V_1^T \mathbf{1}\|_2} \quad \text{and} \quad \boldsymbol{h} = \frac{U_1 U_1^T \mathbf{1}}{\|U_1 U_1^T \mathbf{1}\|_2}.$$

Extremal property of the score vectors Let us finally note that, since the authority matrix $A^T A$ is symmetric, the possible authority score vectors satisfy an extremal property. Indeed, the sets

$$\{\boldsymbol{a} = \lim_{k\to\infty} \boldsymbol{a}(k) \colon \boldsymbol{a}(0) \in \mathbb{R}^n_{\geq 0}\} \quad \text{and} \quad \operatorname*{argmax}_{\|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}^T A^T A \boldsymbol{x}$$

are equal. A similar property can be expressed for the hub score vectors.

2.2.3 Measuring similarity between nodes of two graphs

Now, let us leave web rankings a little and talk about the similarity between nodes of two graphs. Blondel et al. [17] introduced a measure of similarity between the nodes of two graphs that generalizes Kleinberg's hubs and authorities [76]. They defined what they call the *similarity matrix*, whose entries give the similarity scores between the pair of nodes. Independently, Melnik et al. [92, 93] have proposed a definition of similarity between the nodes of two graphs based on similar ideas. They called their method the *similarity flooding* algorithm. We here present both definitions and compare them.

Similar neighbors As for PageRank and HITS, the basic idea of both similarity measures is based on a mutual reinforcement relation. Here, *two nodes are similar if they have similar parents and similar children.*

Iterative computation of the scores Let two graphs $(\mathcal{N}, \mathcal{E})$ and $(\mathcal{M}, \mathcal{F})$ that do not have necessarily the same number of nodes $n = |\mathcal{N}|$ and $m = |\mathcal{M}|$. Let respectively A and B be their adjacency matrices. For convenience, we write $\mathcal{G}(A) = (\mathcal{N}, \mathcal{E})$ and $\mathcal{G}(B) = (\mathcal{M}, \mathcal{F})$. Let $S(k)_{ij}$ be the similarity score of the pair of nodes (i, j), with $i \in \mathcal{M}$ and $j \in \mathcal{N}$, at the iteration step $k \in \mathbb{N}$. The reinforcement relation is modeled by Blondel et al. [17] as

$$\widetilde{S}(k+1)_{ij} = \sum_{\substack{b:\ (i,b)\in\mathcal{F} \\ a:\ (j,a)\in\mathcal{E}}} S(k)_{ba} + \sum_{\substack{b:\ (b,i)\in\mathcal{F} \\ a:\ (a,j)\in\mathcal{E}}} S(k)_{ba} \quad \text{for all } i \in \mathcal{M}, j \in \mathcal{N}, \tag{2.9a}$$

with a normalization

$$S(k+1)_{ij} = \frac{\widetilde{S}(k+1)_{ij}}{\left(\sum_{\substack{b\in\mathcal{M} \\ a\in\mathcal{N}}} \widetilde{S}(k+1)_{ba}^2\right)^{1/2}} \quad \text{for all } i \in \mathcal{M}, j \in \mathcal{N}. \tag{2.9b}$$

This is illustrated in Figure 2.5.

Melnik et al. [93] consider roughly the same iteration scheme (2.9). In addition, they propose to add propagation coefficients, that is, to make a weighted sum instead of (2.9a). Moreover their algorithm allows one to work with graphs having labeled edges. Finally, they normalize differently the scores: all the values $\widetilde{S}(k+1)$ are divided by the maximum one. For the sake of clarity, we will not consider weighted or labeled edges nor distinguish between different norms.

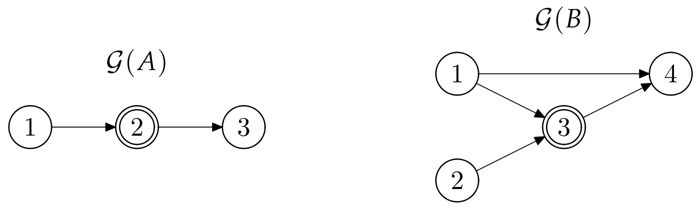


Figure 2.5: Basic idea of similarity between nodes: two nodes are similar if they have similar parents and children. Here, node 2 *of* $\mathcal{G}(A)$ *has parents and children similar to those of node* 3 *of* $\mathcal{G}(B)$.

Let $S(k) = [S(k)_{ij}]_{i\in\mathcal{M},j\in\mathcal{N}} \in \mathbb{R}^{m\times n}$ be the matrix of similarity scores at iteration k. Then the system (2.9) can be expressed in a matrix form, as

$$S(k+1) = \frac{BS(k)A^T + B^T S(k)A}{\|BS(k)A^T + B^T S(k)A\|}.$$

Product graph The reinforcement relations (2.9) of similarity scores for nodes of $\mathcal{G}(A)$ and $\mathcal{G}(B)$ can be interpreted using the product graph $\mathcal{G}(A) \times \mathcal{G}(B)$. The nodes of this graph are the pairs with one node of $\mathcal{G}(A)$ and one node for $\mathcal{G}(B)$. There is a link in the product graph from the pair of nodes (a_1, b_1) to the pair of nodes (a_2, b_2) if $(a_1, a_2) \in \mathcal{E}$ and $(b_1, b_2) \in \mathcal{F}$, i.e., if a_1 is linked to a_2 and b_1 to b_2. An example of product of two graphs is represented in Figure 2.6. So the score $S(k+1)_{ba}$ of the pair of nodes (a, b) is proportional to the sum of the scores of the pairs of nodes which are parents and children of (a, b) in the product graph $\mathcal{G}(A) \times \mathcal{G}(B)$. It is easy to see [91, 118] that the adjacency matrix of the product graph $\mathcal{G}(A) \times \mathcal{G}(B)$ is simply the Kronecker product $A \otimes B$ of the two original adjacency matrices. Let $\boldsymbol{s}(k) = \mathrm{vec}(S(k)) \in \mathbb{R}^{mn}$ for all $k \in \mathbb{N}$, where vec denotes the operator that transforms a matrix in a vector by stacking its columns (see Chapter 4 in [62]). Then the system (2.9) can also be expressed equivalently [17] as

$$\boldsymbol{s}(k+1) = \frac{(A \otimes B + A^T \otimes B^T)\boldsymbol{s}(k)}{\|(A \otimes B + A^T \otimes B^T)\boldsymbol{s}(k)\|}. \tag{2.10}$$

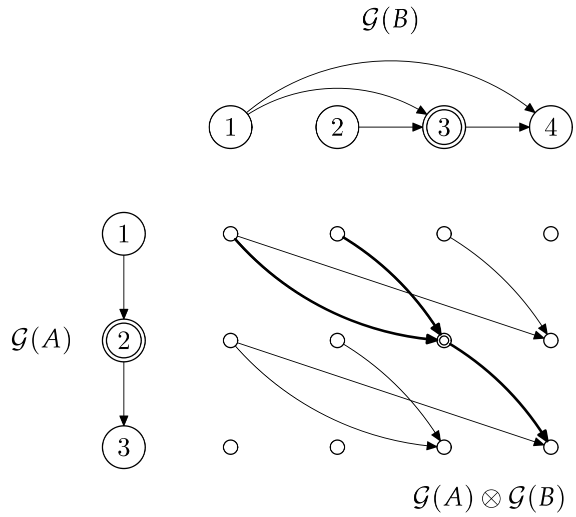


Figure 2.6: The iterative computation of the similarity scores of the pairs of nodes can be interpreted as the propagation of scores in the product graph, by following the links in both directions. Here, for the graphs $\mathcal{G}(A)$ and $\mathcal{G}(B)$ considered in Figure 2.5, the similarity score of the pair of nodes 3 of $\mathcal{G}(B)$ and 2 of $\mathcal{G}(A)$ is reinforced by the scores of its neighbors in the product graph, via the links represented with bold arrows.

Convergence concerns The authors of both papers observe that the iterates (2.9) do not converge in general. Blondel et al. use the equivalent expression (2.10) to analyze the convergence. They prove that both even and odd sequences $(\boldsymbol{s}(2k))_{k\in\mathbb{N}}$ and $(\boldsymbol{s}(2k+1))_{k\in\mathbb{N}}$ converge to limit vectors that depend in general on the initial vector $\boldsymbol{s}(0)$. Let $M = A \otimes B + A^T \otimes B^T$. Since M is symmetric, it can be diagonalized as $M = U\Lambda U^T$ with U unitary and Λ a diagonal matrix with the eigenvalues of M, that are all real. Moreover, since M is nonnegative, Perron–Frobenius Theorem ensures that $\rho(M)$ is an eigenvalue of M. They then consider the decomposition $U = \begin{pmatrix} U_1 & U_2 & U_3 \end{pmatrix}$, where U_1 corresponds to the block $\Lambda_1 = \rho(M)I$, the columns U_2 to the block $\Lambda_2 = -\rho(M)I$, and U_3

to the block Λ_3 with $\rho(\Lambda_3) < \rho(M)$. From this, they get

$$\lim_{k\to\infty} \boldsymbol{s}(2k) = \frac{(U_1U_1^T + U_2U_2^T)\boldsymbol{s}(0)}{\|(U_1U_1^T + U_2U_2^T)\boldsymbol{s}(0)\|_2},$$
$$\lim_{k\to\infty} \boldsymbol{s}(2k+1) = \frac{(U_1U_1^T - U_2U_2^T)\boldsymbol{s}(0)}{\|(U_1U_1^T - U_2U_2^T)\boldsymbol{s}(0)\|_2}.$$

They then prove that the even limit vector $\lim_{k\to\infty} \boldsymbol{s}(2k)$ with $\boldsymbol{s}(0) = \mathbf{1}/\|\mathbf{1}\|_2$ is the unique vector of largest ℓ_1 norm in the set of all even and odd limit vectors for every possible initial vector. So they define the *similarity matrix* as $S = \lim_{k\to\infty} S(2k)$ where $S(0)$ is the matrix whose entries are all identical. And the corresponding vector $\boldsymbol{s} = \mathrm{vec}(S)$ is therefore

$$\boldsymbol{s} = \frac{(U_1U_1^T + U_2U_2^T)\mathbf{1}}{\|(U_1U_1^T + U_2U_2^T)\mathbf{1}\|}.$$

Melnik et al. on their side, propose to change the iteration formula (2.10) to

$$\boldsymbol{z}(k+1) = \frac{M\boldsymbol{z}(k) + \boldsymbol{d}}{\|M\boldsymbol{z}(k) + \boldsymbol{d}\|}, \tag{2.11}$$

where M is constructed from the adjacency matrix of the product graph and its transpose, as above, or possibly a weighted form of these matrices, and $\boldsymbol{d}$ is a positive vector, for instance taken as $\mathbf{1}$. They observe experimentally the convergence of these modified iterates [93]. They also try to justify theoretically the convergence when M is taken stochastic [92]. They argue that the damping with the positive vector $\boldsymbol{d}$ is like modifying the product graph to a graph with positive weighted adjacency matrix to which a result such as Corollary 2.4 could be applied. The problem is that with such construction, the weights on the edges of the modified graph change for each iteration, that is, the matrix is positive but changes at each step. In fact the convergence of iterates (2.11) is a direct consequence of nonlinear Perron–Frobenius results we have presented in Section 2.1.4 . Indeed, with $\boldsymbol{d} > 0$, the affine map $\mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{>0} \colon \boldsymbol{x} \mapsto M\boldsymbol{x} + \boldsymbol{d}$ is a concave map that satisfies the hypothesis of Krause's Theorem 2.7.

Let $\boldsymbol{z} = \lim_{k\to\infty} \boldsymbol{z}(k)$ be the vector giving the similarity scores for the similarity flooding algorithm of Melnik et al. We were interested in comparing $\boldsymbol{z}$ with $\boldsymbol{s}$ corresponding to the similarity matrix of Blondel et al. So, suppose $M = A \otimes B + A^T \otimes B^T$ and $M = U \Lambda U^T$, where U and Λ are decomposed in three parts according to the eigenvalues $\rho(M)$, $-\rho(M)$ and the rest of the spectrum, as above. In the next chapter, we prove in Proposition 3.51 that, if $\boldsymbol{d} = \varepsilon \mathbf{1}$ with $\varepsilon > 0$ but ε small in comparison with $\rho(M)$, then

$$\boldsymbol{z} \approx \frac{U_1 U_1^T \mathbf{1}}{\|U_1 U_1^T \mathbf{1}\|}.$$

Note then that the similarity scores $\boldsymbol{z}$ of Melnik et al. are approximately proportional to the average between the similarity scores $\boldsymbol{s}$ and the odd limit $\lim_{k\to\infty} \boldsymbol{s}(2k+1)$ for the same initial vector $\boldsymbol{s}(0) = \mathbf{1}/\|\mathbf{1}\|$, considered by Blondel et al.

Extremal property of the similarity matrix Let us finally note that, since the matrix matrix $A \otimes B + A^T \otimes B^T$ is symmetric, the possible similarity matrices satisfy an extremal property. Indeed, the sets

$$\{\lim_{k\to\infty} S(k) \colon S(0) \in \mathbb{R}^n_{\geq 0}\} \quad \text{and} \quad \operatorname*{argmax}_{\|X\|_F = 1} \langle X, BXA^T + B^T XA \rangle$$

are equal, where the inner product $\langle X, Y \rangle = \sum_{i,j} X_{ij} Yij$ for two matrices X and Y of same dimensions.

Similarity scores and hubs and authorities As shown by Blondel et al. [17], Kleinberg's hub and authority scores can be seen as similarity scores between the nodes of the focused web graph, say $\mathcal{G}(B)$, and the simple graph $\mathcal{G}(A) = (\{1,2\}, \{(1,2)\})$, with adjacency matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, see Figure 2.7. Indeed, in this case,

$$(A \otimes B + A^T \otimes B^T)^2 = \begin{pmatrix} BB^T & 0 \\ 0 & B^T B \end{pmatrix},$$

which gives exactly the hub and authority matrices (see Section 2.2.2).

Other applications Blondel et al. [17] as well as Senellart and Blondel [109] show how the similarity matrix can be used in order to automatically extract synonyms from a monolingual dictionary. They first

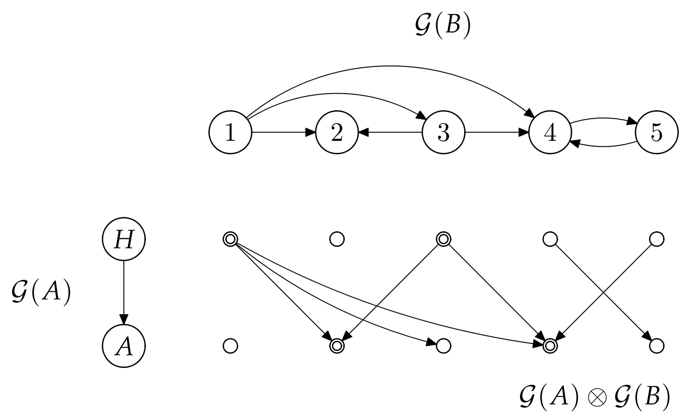


Figure 2.7: Kleinberg's hub and authority scores can be seen as similarity scores of the nodes of the focused graph with the two nodes of the simple path graph. Here, the product graph illustrates quite clearly that nodes 1 *and* 3 *of* $\mathcal{G}(B)$ *are rather hubs and nodes* 2 *and* 4 *rather authorities.*

construct a graph $\mathcal{G}$ representing the dictionary: there is a node for each word of the dictionary and a directed link between two nodes if the second one appears in the definition of the first one. Then, for a particular query word, they construct a neighborhood graph, that is, the subgraph induced by the corresponding node, its parents and its children in the graph. They then compute the similarity between the nodes of the neighborhood graph and the central node of the directed path graph of length three.

The principal application that Melnik et al. [93] have in mind is the automatic matching of diverse data structures. They present an example of two relational databases with information about employees, such as the name, the birth date or the department, of two different companies. Of course, both companies do not have in general the same names for corresponding fields nor exactly the same data fields. Melnik et al. want to make a matching between corresponding fields of the logical schemas of both databases. For each database schema, they construct a directed graph, with labeled nodes and links (see Figure 2.8). Then they proceed to an initialization of similarity scores with a simple string matching algorithm which compares prefixes and suffixes of the nodes' labels.

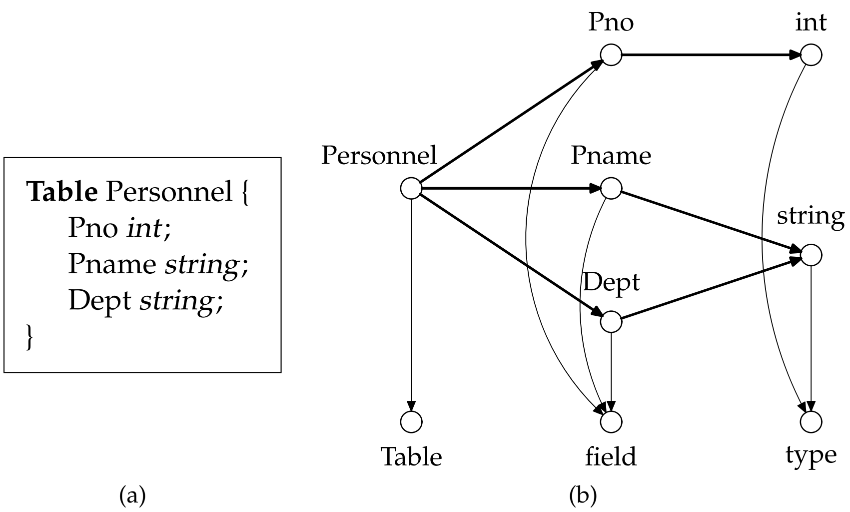


Figure 2.8: Melnik et al. represent a database's schema by a graph with labeled edges. (a) An example of logical schema of a database. (b) A simplified version of the graph corresponding to the schema given in (a), with two kinds of edges.

As a third step, they compute the similarity scores with their similarity flooding algorithm that we have presented in this section. That gives them a similarity score for each pair of elements. Finally, they use a filter which automatically selects the most plausible pairs of elements in order to match them: to each element of the first graph must correspond at most one element of the other graph, and conversely.

Chapter 3

An affine eigenvalue problem on the nonnegative orthant

This chapter is devoted to the analysis of the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad x \in \mathbb{R}^n_{\geq 0}, \quad \|\boldsymbol{x}\| = 1,$$

where A is an $n \times n$ nonnegative matrix, $\boldsymbol{b}$ a nonnegative vector, and $\|\cdot\|$ a monotone vector norm. We approach this problem from two complementary points of view. The first considers *graph-theoretic properties* of A and b. The second approach characterizes, under suitable assumptions, the unique solution $(\lambda, \boldsymbol{x})$ as the Perron pair of a rank-one modification of the matrix A satisfying a *maximizing property*.

3.1 Introduction

In Section 2.2.3, we have seen how Melnik et al. [93] proposed to consider the normalized affine iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k) + \boldsymbol{b}}{\|A\boldsymbol{x}(k) + \boldsymbol{b}\|}, \tag{3.1}$$

in order to tackle with a lack of convergence of the iterates

$$\boldsymbol{x}(k) = \frac{A\boldsymbol{x}(k)}{\|A\boldsymbol{x}(k)\|},$$

for some nonnegative matrix A. They propose to take a positive vector $\boldsymbol{b}$, and indeed in this case, the normalized affine iteration converges to a unique fixed point. We have noticed in Section 2.2.3 that, if $A \in \mathbb{R}^{n\times n}_{\geq 0}$ and $\boldsymbol{b} \in \mathbb{R}^{n}_{\geq 0}$ are such that $A\boldsymbol{x} + b > 0$ for all $\boldsymbol{x} \gneqq 0$, the convergence of iteration (3.1) is a particular case of Krause's Theorem 2.7 about concave maps.

In this chapter, we focus on the *fixed points* of the normalized affine iteration. As we already noticed in Section 2.1.4, a vector $\boldsymbol{x}_*$ is a fixed point of (3.1) if and only if $(\|A\boldsymbol{x}_* + \boldsymbol{b}\|, \boldsymbol{x}_*)$ is a solution of the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1. \tag{3.2}$$

We are interested in the existence and possible uniqueness of the solutions of this eigenvalue problem, for a matrix $A \in \mathbb{R}^{n\times n}_{\geq 0}$, a vector $\boldsymbol{b} \in \mathbb{R}^{n}_{\geq 0}$ and a monotone norm $\|\cdot\|$.

We take two different approaches on the conditional affine eigenvalue problem. We first analyze (3.2) from the point of view of accessibility relations in the graph of A to nodes such that the corresponding entry of the vector $\boldsymbol{b}$ is positive. This approach uses M-matrix equations as analyzed by Carlson [32] and subsequent authors. In Theorem 3.7, we show how a solution of problem (3.2) satisfying particular graph properties gives information about the existence of other solutions.

Our second point of view on the conditional affine eigenvalue problem (3.2) is to look at its solution as a Perron pair of a rank-one perturbation of the matrix A. Our main result using this approach is Theorem 3.19: this rank-one perturbation of A satisfies a *maximizing property*. More precisely, the conditional affine eigenvalue problem (3.2) has a unique solution $(\lambda_*, \boldsymbol{x}_*)$ with $\lambda > \rho(A)$, if and only if $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) > \rho(A)$, where $\boldsymbol{c}_*$ is a maximizer of $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ on the unit sphere of the dual norm. Moreover, in this case, λ_* is the spectral radius and $\boldsymbol{x}_*$ the unique normalized Perron vector of the matrix $A + \boldsymbol{b}\boldsymbol{c}_*^T$.

These graph-theoretic and optimization approaches can be related to give a quite complete characterization of the solutions of the conditional affine eigenvalue problem (3.2). This characterization uses rank-one perturbation matrices $A + \boldsymbol{b}\boldsymbol{c}^T$ for vectors $\boldsymbol{c}$ in the unit sphere for the dual norm, as well as graph-theoretic properties of A and $\boldsymbol{b}$. The results

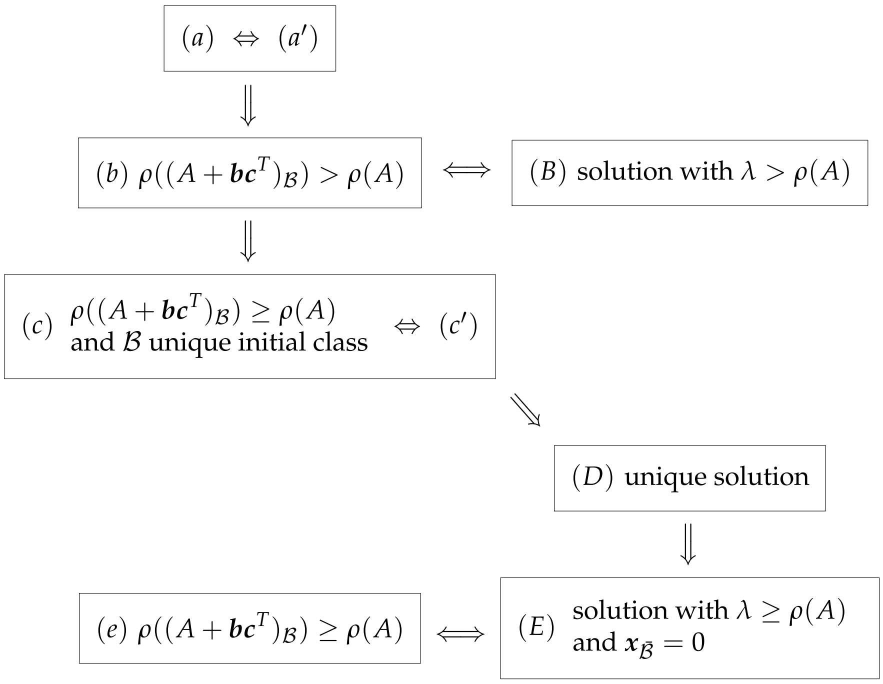


Figure 3.1: Diagram of implications of Theorem 3.1.

obtained in this chapter can be summarized in the following Theorem 3.1. Statements about the conditional affine eigenvalue problem and its solution are denoted by capital letters while statements about properties of rank-one perturbations of the matrix A are denoted by small letters. Figure 3.1 explains the implication relations between the statements. Notions as the dual norm and weak Sraffa matrices will be defined in the sequel, when needed.

Theorem 3.1. *Let A be a nonnegative matrix and $\boldsymbol{b}$ be a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. Let $\mathcal{B}$ be the set of nodes which have an access in the graph $\mathcal{G}(A)$ to the nodes i such that $\boldsymbol{b}_i > 0$. Consider the following conditions.*

(a) *For all $\varepsilon > 0$, there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = \varepsilon$ such that $\rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) > \rho(A)$.*

(a') $\rho(A_{\mathcal{B}}) = \rho(A)$.

(b) *There exists* $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^{\star} = 1$ *such that* $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) > \rho(A)$.

(B) *The conditional affine eigenvalue problem* (3.2) *has a solution* $(\lambda, \boldsymbol{x})$ *such that* $\lambda > \rho(A)$.

(c) *There exists* $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^{\star} = 1$ *such that* $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ *and* $\mathcal{B}$ *is the unique initial class of* $A + \boldsymbol{b}\boldsymbol{c}^T$.

(c') *There exists* $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^{\star} = 1$ *such that* $A + \boldsymbol{b}\boldsymbol{c}^T$ *is a weak Sraffa matrix.*

(D) *The conditional affine eigenvalue problem* (3.2) *has a unique solution.*

(e) *There exists* $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^{\star} = 1$ *such that* $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$.

(E) *The conditional affine eigenvalue problem* (3.2) *has a solution* $(\lambda, \boldsymbol{x})$ *such that* $\lambda \geq \rho(A)$ *and* $\boldsymbol{x}_{\mathcal{B}} = 0$.

Then $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (D) \Rightarrow (e)$ *while none of the converse implications holds. Moreover* $(a) \Leftrightarrow (a')$, $(b) \Leftrightarrow (B)$, $(c) \Leftrightarrow (c')$ *and* $(e) \Leftrightarrow (E)$. *For* (a) *and* (b), *the vector* $\boldsymbol{c}$ *can be chosen such that* $\mathcal{B}$ *is the unique initial class of* $A + \boldsymbol{b}\boldsymbol{c}^T$.

This theorem is a collection of several results that we prove in this chapter. The equivalences $(a) \Leftrightarrow (a')$, $(b) \Leftrightarrow (B)$, $(c) \Leftrightarrow (c')$ and $(e) \Leftrightarrow (E)$ follow respectively from Proposition 3.28, Theorem 3.19, Proposition 3.45 and Proposition 3.42. The implication $(b) \Rightarrow (c)$ follows from the fact that the vector $\boldsymbol{c}$ of condition (b) can be chosen such that $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$, which is proved in Proposition 3.34. The implication $(c) \Rightarrow (D)$ is proved in Proposition 3.37 and $(D) \Rightarrow (e)$ is proved in Proposition 3.39. Counterexamples are given for each converse implication.

It would have been nice to find an equivalent condition to the uniqueness of the solution (condition (D)) similar to $(b) \Leftrightarrow (B)$ and $(e) \Leftrightarrow (E)$. This is unfortunately not possible in general, as we show in Example 3.41. But, in the case where the norm $\|\cdot\|$ is *strictly monotone*, that is, $\|\cdot\|$ is monotone and satisfies moreover $\|\boldsymbol{x}\| > \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $|\boldsymbol{x}| \gneqq |\boldsymbol{y}|$, we prove in Theorem 3.43 that the conditional affine eigenvalue problem (3.2) has a unique solution if and only if there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_\mathcal{B}) \geq \rho(A)$. That is, in this particular case of strictly monotone norm $\|\cdot\|$, the conditions (D), (e) and (E) of Theorem 3.1 are equivalent. Note that in general $(D) \not\Rightarrow (c)$, even with strictly monotone norms, as shown with Example 3.41 for the case of the ℓ_2 norm. However, for a particular class of norms, we prove in Proposition 3.44 that the conditions (c), (D) and (e) of Theorem 3.1 are equivalent.

This chapter is organized as follows. We first analyze the affine eigenvalue problem (3.2) in Section 3.2. Then, in Section 3.3 we show how, under suitable assumptions, the unique solution of problem (3.2) can be expressed as the eigenpair of a rank-one perturbation of A satisfying a maximizing property. We particularize these results for the ℓ_1, ℓ_∞ and ℓ_2 norms in Section 3.4. We then relate both graph-theoretic and optimization approaches in Section 3.5. In Section 3.6, we see how, for a small $\varepsilon > 0$ and under suitable assumptions, the solution of the affine eigenvalue problem $\lambda \boldsymbol{x} = A\boldsymbol{x} + \varepsilon \boldsymbol{b}$, $\lambda \in \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^n_{\geq 0}$, $\|\boldsymbol{x}\| = 1$ can be a good approximation of a Perron pair of the matrix A. Finally, in Section 3.7,we present our experiments on a subgraph of the Web.

3.2 A graph-theoretic condition

Let us first analyze the solutions of the conditional affine eigenvalue problem (3.2) from the point of view of accessibility relations in the graph $\mathcal{G}(A)$ to nodes i such that $\boldsymbol{b}_i > 0$.

Let $\operatorname{supp}(\boldsymbol{d}) = \{j \colon \boldsymbol{d}_j > 0\}$ be the *support* of a vector $\boldsymbol{d}$. In Section 2.1.2, we have seen that the matrix A can be written in a Frobenius normal form. In particular, let $\mathcal{B}$ be the set of nodes having an access to $\operatorname{supp}(\boldsymbol{b})$ in the graph of A (or equivalently the union of classes having an access so $\operatorname{supp}(\boldsymbol{b})$). Then, up to a permutation of the indices, the matrix A and the vector $\boldsymbol{b}$ can be rewritten as

$$A = \begin{pmatrix} A_\mathcal{B} & A_{\mathcal{B}\bar{\mathcal{B}}} \\ 0 & A_{\bar{\mathcal{B}}} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_\mathcal{B} \\ 0 \end{pmatrix}.$$

Decomposing $\boldsymbol{x}$ according to $\mathcal{B}$, the affine eigenvalue problem (3.2) can be rewritten as

$$\begin{aligned}
&\lambda \boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}} \boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}} + A_{\mathcal{B}\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}}, \\
&\lambda \boldsymbol{x}_{\bar{\mathcal{B}}} = A_{\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}}, \\
&\lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1.
\end{aligned} \tag{3.3}$$

Before going further in the study of the solutions of the conditional affine eigenvalue problem, let us first present the following classical result about nonnegative matrices (see for instance [14, Chap. 2] and [61, Chap. 8]).

Lemma 3.2. *Let M be a nonnegative matrix and α, β two nonnegative scalars. Then*

(a) $\alpha \boldsymbol{x} \leq M\boldsymbol{x}$ *with* $\boldsymbol{x} \gneqq 0$ *implies* $\alpha \leq \rho(M)$,
$\alpha \boldsymbol{x} < M\boldsymbol{x}$ *with* $\boldsymbol{x} \gneqq 0$ *implies* $\alpha < \rho(M)$;

(b) $M\boldsymbol{x} \leq \beta \boldsymbol{x}$ *with* $\boldsymbol{x} > 0$ *implies* $\rho(M) \leq \beta$,
$M\boldsymbol{x} < \beta \boldsymbol{x}$ *with* $\boldsymbol{x} > 0$ *implies* $\rho(M) < \beta$.

Moreover, if M is irreducible, then

(c) $\alpha \boldsymbol{x} \lneqq M\boldsymbol{x} \lneqq \beta \boldsymbol{x}$ *with* $\boldsymbol{x} \gneqq 0$ *implies* $\alpha < \rho(A) < \beta$ *and* $\boldsymbol{x} > 0$.

The following Theorem 3.4 is due to Carlson [32] and Victory [116] (see also Schneider's survey [108] as well as [59, 114] and the references therein). It gives a graph-theoretic condition that allows to verify if, *for a given λ*, the linear equation $\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ has a nonnegative solution $\boldsymbol{x}$. We restate and reprove it here with our notations, decomposing matrices and vectors according to $\mathcal{B}$, the *set of nodes having an access to* $\mathrm{supp}(\boldsymbol{b})$. See Figure 3.2 for an illustration. We begin with a simple lemma.

Lemma 3.3. *Let M be a nonnegative matrix and $\boldsymbol{d}$ a nonnegative vector such that every node in the graph $\mathcal{G}(M)$ has an access to* $\mathrm{supp}(\boldsymbol{d})$. *If $M\boldsymbol{x} + \boldsymbol{d} \leq \lambda \boldsymbol{x}$ with $\boldsymbol{x} \geq 0$, then $\rho(M) < \lambda$ and $\boldsymbol{x} > 0$.*

Proof. Clearly, if every node has an access to $\mathrm{supp}(\boldsymbol{d})$ in the graph of M, the matrix $M + \boldsymbol{d}\boldsymbol{c}^T$ is irreducible for any positive vector $\boldsymbol{c}$. Let $\boldsymbol{c} > 0$ be such that $\boldsymbol{c}^T\boldsymbol{x} < 1$. Then $(M + \boldsymbol{d}\boldsymbol{c}^T)\boldsymbol{x} \lneqq M\boldsymbol{x} + \boldsymbol{d} \leq \lambda \boldsymbol{x}$ and since $M + \boldsymbol{d}\boldsymbol{c}^T$ is irreducible, $\rho(M) \leq \rho(M + \boldsymbol{d}\boldsymbol{c}^T) < \lambda$ and $\boldsymbol{x} > 0$ by Lemma 3.2(c). $\square$

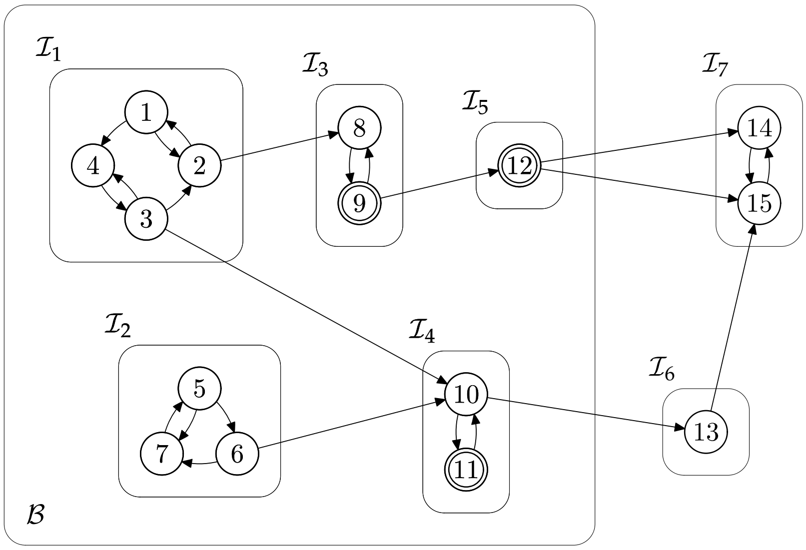


Figure 3.2: Nodes $9 \in \mathcal{I}_3$, $11 \in \mathcal{I}_4$ *and* $12 \in \mathcal{I}_5$ *correspond to positive entries of* $\boldsymbol{b}$, *that is,* $\operatorname{supp}(\boldsymbol{b}) = \{9, 11, 12\}$. *The set of nodes having an access to* $\operatorname{supp}(\boldsymbol{b})$ *in* $\mathcal{G}(A)$ *is* $\mathcal{B} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4 \cup \mathcal{I}_5$.

Theorem 3.4 (Carlson–Victory, see [108]). *Let A be a nonnegative matrix, $\boldsymbol{b}$ a nonnegative vector and $\lambda \in \mathbb{R}$. Let $\mathcal{B}$ be the set of nodes having an access to* $\operatorname{supp}(\boldsymbol{b})$ *in $\mathcal{G}(A)$. Then $\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ has a nonnegative solution $\boldsymbol{x} \geq 0$ if and only if $\lambda > \rho(A_{\mathcal{B}})$. Moreover, in this case, $\boldsymbol{x}_{\mathcal{B}} > 0$.*

Proof. Suppose first that there exists $\boldsymbol{x} \geq 0$ such that $\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$. Decomposing $\boldsymbol{x}$ according $\mathcal{B}$, we have

$$\lambda \boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}} \boldsymbol{x}_{\mathcal{B}} + A_{\mathcal{B}\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}} + \boldsymbol{b}_{\mathcal{B}} \geq A_{\mathcal{B}} \boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}.$$

By Lemma 3.3, since every node of $\mathcal{B}$ has an access to $\operatorname{supp}(\boldsymbol{b}_{\mathcal{B}})$ in $\mathcal{G}(A_{\mathcal{B}})$, it follows that $\lambda > \rho(A_{\mathcal{B}})$.

Conversely, suppose that $\lambda > \rho(A_{\mathcal{B}})$. Then by Lemma 2.2, $(\lambda I - A_{\mathcal{B}})^{-1} \geq 0$. Therefore, $\boldsymbol{x}$ is a nonnegative solution of $\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ if and only if

$$\begin{aligned} \boldsymbol{x}_{\mathcal{B}} &= (\lambda I - A_{\mathcal{B}})^{-1} \boldsymbol{b}_{\mathcal{B}} + (\lambda I - A_{\mathcal{B}})^{-1} A_{\mathcal{B}\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}}, \\ \lambda \boldsymbol{x}_{\bar{\mathcal{B}}} &= A_{\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}}, \quad \text{with } \boldsymbol{x}_{\bar{\mathcal{B}}} \geq 0. \end{aligned} \tag{3.4}$$

By Lemma 3.3, $\boldsymbol{x}_{\mathcal{B}} \geq (\lambda I - A_{\mathcal{B}})^{-1}\boldsymbol{b}_{\mathcal{B}} > 0$. Note also that $\lambda\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ has a unique nonnegative solution if and only if there does not exist a nonnegative eigenvector of $A_{\bar{\mathcal{B}}}$ corresponding to λ. If such an eigenvector exists (for instance if $\lambda = \rho(A_{\bar{\mathcal{B}}}) = \rho(A)$), then the linear equation has infinitely many nonnegative solutions. □

Example 3.5. Let us illustrate that, for a given $\lambda > 0$, the equation $\lambda\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ can have one, infinitely many or no solution, as stated by Theorem 3.4. Let

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

The set of nodes having an access to $\operatorname{supp}(\boldsymbol{b})$ in $\mathcal{G}(A)$ is $\mathcal{B} = \{1, 2\}$, as seen in Figure 3.3. By Theorem 3.4, the equation $\lambda\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ has at least one nonnegative solution, satisfying equation (3.4), if and only if $\lambda > \rho(A_{\mathcal{B}}) = 2$. If $\lambda = 3$, then equation (3.4) has infinitely many solutions, given by

$$\begin{pmatrix} 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 2 \\ 2 \\ 0 \end{pmatrix} \qquad \text{for any } \alpha \geq 0.$$

Also for $\lambda = 4$, equation (3.4) has infinitely many solutions, given by

$$\begin{pmatrix} 1/6 \\ 1/2 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 3 \\ 3 \\ 3 \end{pmatrix} \qquad \text{for any } \alpha \geq 0.$$

Finally, if $\lambda > 2$, $\lambda \neq 3, 4$, then

$$\boldsymbol{x} = \frac{1}{(\lambda - 1)(\lambda - 2)} \begin{pmatrix} 1 \\ \lambda - 1 \\ 0 \\ 0 \end{pmatrix}$$

is the unique solution to the system. ◇

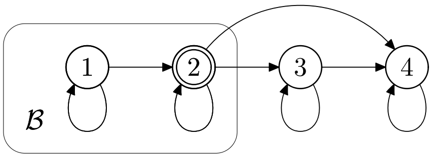


Figure 3.3: The set of nodes having an access to $\text{supp}(\boldsymbol{b})$ *is* $\mathcal{B} = \{1, 2\}$. *So, by Carlson–Victory's Theorem 3.4, the equation* $\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$ *has a nonnegative solution* $\boldsymbol{x}$ *if and only if* $\lambda > \rho(A_{\mathcal{B}}) = 2$.

Now, let us come back to the conditional affine eigenvalue problem (3.2),

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1.$$

By Carlson–Victory Theorem, if $(\lambda, \boldsymbol{x})$ is a solution of this conditional affine eigenvalue problem then $\lambda > \rho(A_{\mathcal{B}})$ and $\boldsymbol{x}_{\bar{\mathcal{B}}}$ is a nonnegative eigenvector of $A_{\bar{\mathcal{B}}}$. In fact, a particular solution of the eigenvalue problem (3.2) with $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$ informs us about the existence of other solutions.

Lemma 3.6. *Let M be a nonnegative matrix, $\boldsymbol{d}$ a nonnegative vector such that every node has an access to* $\text{supp}(\boldsymbol{d})$ *in the graph $\mathcal{G}(M)$. Then, for every $\lambda_1 > \lambda_2 > \rho(M)$,*

$$0 < (\lambda_1 I - M)^{-1}\boldsymbol{d} < (\lambda_2 I - M)^{-1}\boldsymbol{d}.$$

Proof. Let $\lambda_1 > \lambda_2 > \rho(M)$. Then

$$(\lambda_1 I - M)^{-1}\boldsymbol{d} = \frac{1}{\lambda_1} \sum_{k \in \mathbb{N}} \left(\frac{M}{\lambda_1}\right)^k \boldsymbol{d} \leq \frac{1}{\lambda_2} \sum_{k \in \mathbb{N}} \left(\frac{M}{\lambda_2}\right)^k \boldsymbol{d} = (\lambda_2 I - M)^{-1}\boldsymbol{d},$$

and moreover this inequality is strict since $\lambda_1 \neq \lambda_2$ and $(\lambda_1 I - M)^{-1}\boldsymbol{d} > 0$ by Lemma 3.3. □

Theorem 3.7. *Let A be a nonnegative matrix, $\boldsymbol{b}$ a nonnegative vector and $\|\cdot\|$ a monotone norm. The conditional affine eigenvalue problem* (3.2) *has exactly one solution $(\lambda^*, \boldsymbol{x}^*)$ such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$. Moreover, for that solution, $\lambda^* > \rho(A_{\mathcal{B}})$, $\boldsymbol{x}^*_{\mathcal{B}} > 0$ and the following holds*

(*a*) *if $\lambda^* > \rho(A)$, then that solution $(\lambda^*, \boldsymbol{x}^*)$ of* (3.2) *is unique;*

(*b*) *if $\lambda^* < \rho(A)$, then* (3.2) *has at least one additional solution $(\lambda, \boldsymbol{x})$ such that $\lambda = \rho(A)$ and $\boldsymbol{x}_{\bar{\mathcal{B}}} \neq 0$;*

(c) *if* $\lambda^* = \rho(A)$*, then any other solution* $(\lambda, \boldsymbol{x})$ *of* (3.2)*, if it exists, must be such that* $\lambda = \rho(A)$ *and* $\boldsymbol{x}_{\bar{\mathcal{B}}} \neq 0$.

Proof. Let $\lambda \in \mathbb{R}$, $\boldsymbol{x} \geq 0$ and suppose that $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$. Then, by equation (3.4), $(\lambda, \boldsymbol{x})$ is a solution of the affine eigenvalue problem (3.2) if and only if $\lambda \boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}} \boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}$ and $\|\boldsymbol{x}\| = 1$. By Theorem 3.4, we must have $\lambda > \rho(A_{\mathcal{B}})$ and by Lemma 3.6, $\boldsymbol{x}_{\mathcal{B}} = (\lambda I - A_{\mathcal{B}})^{-1} \boldsymbol{b}_{\mathcal{B}} > 0$, since by definition every node of $\mathcal{B}$ has an access to supp $(\boldsymbol{b})$ in the graph $\mathcal{G}(A_{\mathcal{B}})$. Let

$$f\colon]\rho(A_{\mathcal{B}}), \infty[\;\to \mathbb{R}_{>0} \colon \lambda \mapsto \left\| \begin{pmatrix} (\lambda I - A_{\mathcal{B}})^{-1} \boldsymbol{b}_{\mathcal{B}} \\ 0 \end{pmatrix} \right\|. \tag{3.5}$$

By Lemma 3.6 and the fact that $\|\cdot\|$ is monotone, the continuous map f is strictly decreasing in λ, with $\lim_{\lambda \to \rho(A)} f(\lambda) = \infty$ and $\lim_{\lambda \to \infty} f(\lambda) = 0$. Therefore, there exists a unique $\lambda^* > \rho(A)$ such that $f(\lambda^*) = 1$. Hence there exists a unique solution $(\lambda^*, \boldsymbol{x}^*)$ of the eigenvalue problem (3.2) such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$. This solution satisfies $\boldsymbol{x}^*_{\mathcal{B}} = (\lambda^* I - A_{\mathcal{B}})^{-1} \boldsymbol{b}_{\mathcal{B}} > 0$.

It follows also that any other solution $(\lambda, \boldsymbol{x})$ of the eigenvalue problem (3.2) must satisfy $\lambda \geq \lambda^*$ and $\lambda \boldsymbol{x}_{\bar{\mathcal{B}}} = A_{\bar{\mathcal{B}}} \boldsymbol{x}_{\bar{\mathcal{B}}}$ with $\boldsymbol{x}_{\bar{\mathcal{B}}} \gneqq 0$. This is impossible if $\lambda^* > \rho(A)$, so assertion (a) follows. If $\lambda^* < \rho(A)$, there must at least exist a solution $(\lambda, \boldsymbol{x})$ with $\lambda = \rho(A) = \rho(A_{\bar{\mathcal{B}}})$ and $\boldsymbol{x}_{\bar{\mathcal{B}}}$ a nonzero Perron vector of $A_{\bar{\mathcal{B}}}$, as stated in (b). Finally, if $\lambda = \rho(A)$, then, depending on the norm and on $A_{\mathcal{B}\bar{\mathcal{B}}}$, there may exist another solution $(\lambda, \boldsymbol{x})$. This solution must be such that $\lambda = \rho(A) = \rho(A_{\bar{\mathcal{B}}})$ and $\boldsymbol{x}_{\bar{\mathcal{B}}}$ is a nonzero Perron vector of $A_{\bar{\mathcal{B}}}$, so (c) follows. □

Example 3.8. Let us first illustrate statement (a) of Theorem 3.7. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right)$ and let $\|\cdot\|$ be the ℓ_1 norm. The set of nodes having an access to supp$(\boldsymbol{b})$ is $\mathcal{B} = \{1\}$. The solution $(\lambda^*, \boldsymbol{x}^*)$ of (3.2) such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$ is $\lambda^* = 3 > \rho(A)$ and $\boldsymbol{x}^* = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. This is the unique solution of (3.2). ◇

Example 3.9. Now let us give an example for statement (b). Let $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 4 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and let $\|\cdot\|$ be the ℓ_1 norm. The set of nodes having an access to supp$(\boldsymbol{b})$ is $\mathcal{B} = \{1\}$. So the solution $(\lambda^*, \boldsymbol{x}^*)$ of (3.2) such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$ is $\lambda^* = 3 < \rho(A)$ and $\boldsymbol{x}^* = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. The only other solution $(\lambda, \boldsymbol{x})$ of (3.2) is $\lambda = 4 = \rho(A)$ and $\boldsymbol{x} = \left(\begin{smallmatrix} 0.5 \\ 0.5 \end{smallmatrix}\right)$. ◇

Example 3.10. Let us finally consider statement (c). Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ and $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. The set of nodes having an access to supp$(\boldsymbol{b})$ is $\mathcal{B} = \{1\}$.

Let us consider the cases where $\|\cdot\|$ is the ℓ_1 norm or the ℓ_∞ norm. In both case, the solution $(\lambda^*, \boldsymbol{x}^*)$ of (3.2) such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$ is $\lambda^* = 2 = \rho(A)$ and $\boldsymbol{x}^* = \binom{1}{0}$. But, when $\|\cdot\|$ is the ℓ_1 norm, this solution is the unique solution of (3.2). On the other hand, when $\|\cdot\|$ is the ℓ_∞ norm, the problem (3.2) has infinitely many solutions $(\lambda, \boldsymbol{x})$ that can all be expressed as $\lambda = 2 = \rho(A)$ and $\boldsymbol{x} = \binom{1}{0} + \binom{0}{\alpha}$ for any $\alpha \in [0, 1]$. $\diamond$

Theorem 3.7 can be simplified in the case where the norm $\|\cdot\|$ is *strictly monotone*, that is, $\|\cdot\|$ is monotone and $\|\boldsymbol{x}\| > \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $|x| \gneqq |y|$. For instance, the ℓ_1 and ℓ_2 norms are strictly monotone while the ℓ_∞ norm is not.

Corollary 3.11. *Let A be a nonnegative matrix, $\boldsymbol{b}$ a nonnegative vector and $\|\cdot\|$ a monotone and strictly monotone norm. Let $(\lambda^*, \boldsymbol{x}^*)$ be the only solution to problem* (3.2) *such that $\boldsymbol{x}^*_{\bar{\mathcal{B}}} = 0$. Then this solution is the unique solution of the conditional affine eigenvalue problem* (3.2) *if and only if $\lambda^* \geq \rho(A)$.*

Proof. This follows directly from expressions (3.4) and (3.5). $\square$

Note that Corollary 3.11 shows the equivalence $(D) \Leftrightarrow (E)$ of Theorem 3.1 in the case of a strictly monotone norm.

3.3 A maximal property of the spectral radius

Let us now take another point of view on the conditional affine eigenvalue problem (3.2). We will derive a maximizing condition that ensures the uniqueness of the solution and give its expression as the spectral radius and normalized Perron vector of a particular rank-one perturbation of the matrix A.

Let $(\lambda, \boldsymbol{x})$ be a solution of the problem (3.2), and let $\boldsymbol{c} \geq 0$ be any nonnegative vector such that $\boldsymbol{c}^T\boldsymbol{x} = 1$. Note that we then have

$$\lambda \boldsymbol{x} = (A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{x},$$

so λ is an eigenvalue of $A + \boldsymbol{b}\boldsymbol{c}^T$ and $\boldsymbol{x}$ is a corresponding nonnegative eigenvector. In this section, we are interested in characterizing the solution $(\lambda, \boldsymbol{x})$ of the conditional affine eigenvalue problem (3.2) when

$\lambda > \rho(A)$. Remember indeed that, in this case, this solution is unique by Theorem 3.7.

In a first step we prove the uniqueness of the Perron vector corresponding to a spectral radius $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, for an arbitrary nonnegative vector $\boldsymbol{c}$.

Lemma 3.12. *Let A be a nonnegative matrix and $\boldsymbol{b}, \boldsymbol{c}$ two nonnegative vectors. If $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, then the matrix $A + \boldsymbol{b}\boldsymbol{c}^T$ has only one Perron vector. Moreover, for any nonnegative vector $\boldsymbol{d}$, if the matrices $A + \boldsymbol{b}\boldsymbol{c}^T$ and $A + \boldsymbol{b}\boldsymbol{d}^T$ have the same spectral radius $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) = \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, then their normalized Perron vectors are equal.*

Proof. Suppose $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. Let $\boldsymbol{u} \gneqq 0$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = (A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u}$. We must have $\boldsymbol{c}^T\boldsymbol{u} > 0$, since otherwise $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = A\boldsymbol{u}$ with $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. So, from $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = A\boldsymbol{u} + (\boldsymbol{c}^T\boldsymbol{u})\boldsymbol{b}$, it follows

$$\frac{\boldsymbol{u}}{\boldsymbol{c}^T\boldsymbol{u}} = (\rho(A + \boldsymbol{b}\boldsymbol{c}^T)I - A)^{-1}\boldsymbol{b},$$

which shows that the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}^T$ is unique.

Similarly, if $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) = \rho(A + \boldsymbol{b}\boldsymbol{c}^T)$, then, for any Perron vector $\boldsymbol{v}$ of $A + \boldsymbol{b}\boldsymbol{d}^T$,

$$\frac{\boldsymbol{u}}{\boldsymbol{c}^T\boldsymbol{u}} = (\rho(A + \boldsymbol{b}\boldsymbol{c}^T)I - A)^{-1}\boldsymbol{b} = (\rho(A + \boldsymbol{b}\boldsymbol{d}^T)I - A)^{-1}\boldsymbol{b} = \frac{\boldsymbol{v}}{\boldsymbol{d}^T\boldsymbol{v}},$$

and hence $\boldsymbol{u}$ and $\boldsymbol{v}$ are equal, up to a scalar factor. $\square$

Example 3.13. Let us illustrate that two matrices $A + \boldsymbol{b}\boldsymbol{c}^T$ and $A + \boldsymbol{b}\boldsymbol{d}^T$ that have the same spectral radius larger than $\rho(A)$, also have the same Perron vector. Let

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

Then $\rho(A) = 1$ and, for instance, $\rho(A + \boldsymbol{b}\boldsymbol{e}_1^T) = \rho(A + \boldsymbol{b}\boldsymbol{e}_2^T) = 3$. Therefore, by Lemma 3.12, the corresponding normalized Perron vectors of $A + \boldsymbol{b}\boldsymbol{e}_1^T$ and $A + \boldsymbol{b}\boldsymbol{e}_2^T$ are equal:

$$\boldsymbol{u}_1 = \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Furthermore, it is easily proved that if $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho(A + \boldsymbol{b}\boldsymbol{d}^T) > \rho(A)$, then the Perron vector of these matrices is also the Perron vector of any matrix $A + \boldsymbol{b}(\alpha \boldsymbol{c}^T + (1-\alpha)\boldsymbol{d}^T)$, with $0 \le \alpha \le 1$, which has moreover the same spectral radius. ◇

Example 3.14. Let us show by an example that the converse is not true in general. Let $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, $\boldsymbol{c} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and $\boldsymbol{d} = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right)$. Then $A + \boldsymbol{b}\boldsymbol{c}^T$ and $A + \boldsymbol{b}\boldsymbol{d}^T$ have the same Perron vector $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$. But $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = 1 \neq \rho(A + \boldsymbol{b}\boldsymbol{d}^T) = 4$. ◇

The next stage is to show that, for a nonnegative vector $\boldsymbol{c}$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, we can compare $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ with $\rho(A + \boldsymbol{b}\boldsymbol{d}^T)$ for any $\boldsymbol{d} \ge 0$ by comparing the scalar product of the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}^T$ with $\boldsymbol{c}$ or $\boldsymbol{d}$, and reciprocally. In the following lemma, the sign of a scalar $\alpha \in \mathbb{R}$ is denoted by $\operatorname{sign}(\alpha)$.

Lemma 3.15. *Let A be a nonnegative matrix and $\boldsymbol{b}, \boldsymbol{c}$ nonnegative vectors. If $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$ then, for any nonnegative vector $\boldsymbol{d}$,*

$$\operatorname{sign}(\rho(A + \boldsymbol{b}\boldsymbol{c}^T) - \rho(A + \boldsymbol{b}\boldsymbol{d}^T)) = \operatorname{sign}(\boldsymbol{c}^T\boldsymbol{u} - \boldsymbol{d}^T\boldsymbol{u}),$$

where $\boldsymbol{u}$ is the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}^T$.

Proof. Let $\mathcal{J} = \{j \colon \boldsymbol{u}_j > 0\}$ be the support of $\boldsymbol{u}$, and let $\bar{\mathcal{J}} = \{j \colon \boldsymbol{u}_j = 0\}$ be the complement of $\mathcal{J}$. From $(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = \rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u}$ with $\boldsymbol{u}_{\mathcal{J}} > 0$ and from $\boldsymbol{u}_{\bar{\mathcal{J}}} = 0$, it follows that, up to a permutation, A is block upper triangular with diagonal blocks $A_{\mathcal{J}}$ and $A_{\bar{\mathcal{J}}}$. Moreover, since $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{J}}) = \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A) \ge \rho(A_{\mathcal{J}})$, it follows that $\boldsymbol{c}_{\mathcal{J}} \neq 0$ and hence $\boldsymbol{b}_{\bar{\mathcal{J}}} = 0$.

Suppose first that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A + \boldsymbol{b}\boldsymbol{d}^T)$. If we had $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} \le (A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u}$, we would have $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) \le \rho(A + \boldsymbol{b}\boldsymbol{d}^T)$ by Lemma 3.2(a). Therefore there must exist an index i such that

$$\boldsymbol{e}_i^T(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = \boldsymbol{e}_i^T\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} > \boldsymbol{e}_i^T(A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u},$$

and hence $\boldsymbol{c}^T\boldsymbol{u} > d^T\boldsymbol{u}$.

Suppose now that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) < \rho(A + \boldsymbol{b}\boldsymbol{d}^T)$. Then $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) > \rho(A) \ge \rho(A_{\bar{\mathcal{J}}})$ and hence $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) = \rho((A + \boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{J}})$, since, up to a permutation, $A + \boldsymbol{b}\boldsymbol{d}^T$ is block upper triangular with diagonal blocks $(A + \boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{J}}$ and $A_{\bar{\mathcal{J}}}$. If we had $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} \geq (A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u}$, then we would have $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u}_{\mathcal{J}} \geq (A + \boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{J}}u_{\mathcal{J}}$ with $\boldsymbol{u}_{\mathcal{J}} > 0$ and hence $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) \geq \rho(A + \boldsymbol{b}\boldsymbol{d}^T)$ by Lemma 3.2(b). Therefore, there must exist an index i such that

$$\boldsymbol{e}_i^T(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = \boldsymbol{e}_i^T\rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} < \boldsymbol{e}_i^T(A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u},$$

and hence $\boldsymbol{c}^T\boldsymbol{u} < d^T\boldsymbol{u}$.

Finally, if $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho(A + \boldsymbol{b}\boldsymbol{d}^T)$, then $\boldsymbol{u}$ is also a Perron vector of $A + \boldsymbol{b}\boldsymbol{d}^T$ by Lemma 3.12. Therefore

$$(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = \rho(A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{u} = \rho(A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u} = (A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{u},$$

and $\boldsymbol{c}^T\boldsymbol{u} = d^T\boldsymbol{u}$. □

Example 3.16. Let us illustrate that two spectral radii $\rho(A + \boldsymbol{b}\boldsymbol{d}^T)$ and $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$ can be compared by comparing the scalar products $\boldsymbol{d}^T\boldsymbol{u}$ and $\boldsymbol{c}^T\boldsymbol{u}$, where $\boldsymbol{u}$ is the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}^T$. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$ and $\boldsymbol{c} = \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right)$. Then $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = 5 > \rho(A)$ and $\boldsymbol{u} = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$ is the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}^T$. If $\boldsymbol{d} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, since $\boldsymbol{d}^T\boldsymbol{u} = 3 < \boldsymbol{c}^T\boldsymbol{u} = 4$, we know by Lemma 3.15 that $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) < \rho(A + \boldsymbol{b}\boldsymbol{c}^T)$. Indeed, $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) = 4$. ◇

Now, given a nonnegative vector $\boldsymbol{c}$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, we would like to have a nonnegative eigenvector $\boldsymbol{x}$ of the matrix $A + \boldsymbol{b}\boldsymbol{c}^T$ such that $\boldsymbol{c}^T\boldsymbol{x} = 1$ and $\|\boldsymbol{x}\| = 1$, since this would give us a solution of the conditional affine eigenvalue problem (3.2). That leads us naturally to the notions of dual norm and dual pair.

The *dual norm* $\|\cdot\|^\star$ of a vector norm $\|\cdot\|$ is defined by

$$\|\boldsymbol{y}\|^\star = \max_{\|\boldsymbol{x}\|=1} |\boldsymbol{y}^T\boldsymbol{x}|.$$

For a fixed $\boldsymbol{x} \in \mathbb{R}^n$, the nonempty set

$$\{\boldsymbol{y} \in \mathbb{R}^n \colon \|\boldsymbol{y}\|^\star\|\boldsymbol{x}\| = \boldsymbol{y}^T\boldsymbol{x} = 1\}$$

is the *dual of $\boldsymbol{x}$ with respect to* $\|\cdot\|$. A pair $(\boldsymbol{x}, \boldsymbol{y})$ of vectors of $\mathbb{R}^n$ is said to be a *dual pair with respect to* $\|\cdot\|$ if $\|\boldsymbol{y}\|^\star\|\boldsymbol{x}\| = \boldsymbol{y}^T\boldsymbol{x} = 1$. It can be shown that if $\|\cdot\|^\star$ is the dual norm of $\|\cdot\|$, then $\|\cdot\|$ is the dual norm of $\|\cdot\|^\star$ (see [61, Sec. 5.4 and 5.5]). We also have that a norm $\|\cdot\|$ is monotone if and only if its dual norm $\|\cdot\|^\star$ is monotone. Moreover, a norm $\|\cdot\|$ is monotone if and only if $\|\boldsymbol{x}\| = \||\boldsymbol{x}|\|$ for every $\boldsymbol{x} \in \mathbb{R}^n$ [12].

Therefore, for a nonnegative matrix A, a nonnegative vector $\boldsymbol{b}$ and a monotone norm $\|\cdot\|$, we have

$$\begin{aligned}\max_{\substack{\|\boldsymbol{c}\|^\star=1\\ c\geq 0}} \rho(A+\boldsymbol{b}\boldsymbol{c}^T) &\leq \max_{\|\boldsymbol{c}\|^\star=1} \rho(A+\boldsymbol{b}\boldsymbol{c}^T)\\ &\leq \max_{\|\boldsymbol{c}\|^\star=1} \rho(A+\boldsymbol{b}|\boldsymbol{c}|^T) = \max_{\substack{\|\boldsymbol{c}\|^\star=1\\ c\geq 0}} \rho(A+\boldsymbol{b}\boldsymbol{c}^T).\end{aligned}$$

Similarly, for a nonnegative vector $\boldsymbol{u}$ and a monotone norm $\|\cdot\|$,

$$\max_{\|\boldsymbol{c}\|^\star=1} \boldsymbol{c}^T\boldsymbol{u} = \max_{\substack{\|\boldsymbol{c}\|^\star=1\\ c\geq 0}} \boldsymbol{c}^T\boldsymbol{u}.$$

Example 3.17. Let us show by an example that if the norm $\|\cdot\|$ is not monotone, then we do not have $\max_{\|\boldsymbol{c}\|^\star=1, c\geq 0} \rho(A+\boldsymbol{b}\boldsymbol{c}^T) = \max_{\|\boldsymbol{c}\|^\star=1} \rho(A+\boldsymbol{b}\boldsymbol{c}^T)$ in general. Let $A = \left(\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix}2\\1\end{smallmatrix}\right)$ and a non monotone norm $\|\cdot\|$ and its dual norm $\|\cdot\|^\star$ be given by

$$\begin{aligned}&\|\cdot\|\colon \mathbb{R}^2 \to \mathbb{R}_{\geq 0}\colon \boldsymbol{x} \mapsto |x_1+x_2| + 2|x_1-x_2|,\\ &\|\cdot\|^\star\colon \mathbb{R}^2 \to \mathbb{R}_{\geq 0}\colon \boldsymbol{x} \mapsto \frac{1}{8}|3x_1+x_2| + \frac{1}{8}|x_1+3x_2|.\end{aligned}$$

Then, $\max_{\|\boldsymbol{c}\|^\star=1, c\geq 0} \rho(A+\boldsymbol{b}\boldsymbol{c}^T) = 4$ while $\max_{\|\boldsymbol{c}\|^\star=1} \rho(A+\boldsymbol{b}\boldsymbol{c}^T) = 5$. Maximizers of these expressions are for instance given respectively by $\boldsymbol{c} = \left(\begin{smallmatrix}2\\0\end{smallmatrix}\right)$ and $\boldsymbol{c} = \left(\begin{smallmatrix}3\\-1\end{smallmatrix}\right)$. ◇

The following result now follows directly from Lemma 3.15.

Proposition 3.18. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. If there exists a nonnegative vector $\boldsymbol{d}$, with $\|\boldsymbol{d}\|^\star = 1$ such that $\rho(A+\boldsymbol{b}\boldsymbol{d}^T) > \rho(A)$, then*

$$\rho(A+\boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|^\star=1} \rho(A+\boldsymbol{b}\boldsymbol{c}^T),$$

with $\boldsymbol{c}_* \in \mathbb{R}^n_{\geq 0}$, $\|\boldsymbol{c}_*\|^\star = 1$, *if and only if*

$$\boldsymbol{c}_*^T \boldsymbol{u}_* = \max_{\|\boldsymbol{c}\|^\star = 1} \boldsymbol{c}^T \boldsymbol{u}_*,$$

with $\boldsymbol{c}_* \in \mathbb{R}^n_{\geq 0}$, $\|\boldsymbol{c}_*\|^\star = 1$ *and where* $\boldsymbol{u}_*$ *is the Perron vector of* $A + \boldsymbol{b}\boldsymbol{c}_*^T$.

In other words, Proposition 3.18 says that $\boldsymbol{c}_*$ is a maximizer of the spectral radius $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ among all $\boldsymbol{c}$ of dual norm $\|\boldsymbol{c}\|^\star = 1$ if and only if $(\boldsymbol{u}_*, \boldsymbol{c}_*)$ is a dual pair with respect to $\|\cdot\|$, where $\boldsymbol{u}_*$ is the normalized Perron vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$.

Now we are ready to prove the result announced in the introduction for this approach. Under some assumptions, the solution of a conditional affine eigenvalue problem can be expressed as the Perron root and vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$ where $\boldsymbol{c}_*$ is a maximizer of $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ among all $\boldsymbol{c}$ with a dual norm equal to 1. This result corresponds to the equivalence $(b) \Leftrightarrow (B)$ of Theorem 3.1.

Theorem 3.19. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. Let $\boldsymbol{c}_*$ be a nonnegative vector, with $\|\boldsymbol{c}_*\|^\star = 1$, such that*

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|^\star = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T).$$

Then the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1,$$

has a unique solution $(\lambda_*, \boldsymbol{x}_*)$ which moreover satisfies $\lambda_* > \rho(A)$, if and only if $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) > \rho(A)$. Moreover, in this case, $\lambda_* = \rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$, the vector $\boldsymbol{x}_*$ is the unique normalized Perron vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$, and $(\boldsymbol{x}_*, \boldsymbol{c}_*)$ is a dual pair with respect to $\|\cdot\|$.

Proof. Suppose $\boldsymbol{c}_*$ is a nonnegative vector, with $\|\boldsymbol{c}_*\|^\star = 1$ such that

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|^\star = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A),$$

and let $\boldsymbol{x}_* \in \mathbb{R}^n_{\geq 0}$, $\|\boldsymbol{x}_*\| = 1$ be the unique normalized Perron vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$, by Lemma 3.12. Then, by Proposition 3.18 and by the property of dual norms,

$$\boldsymbol{c}_*^T \boldsymbol{x}_* = \max_{\|\boldsymbol{c}\|^\star = 1} \boldsymbol{c}^T \boldsymbol{x}_* = \|\boldsymbol{x}_*\| = 1.$$

From $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_* = (A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_*$, we have $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_* = A\boldsymbol{x}_* + \boldsymbol{b}$ with $\boldsymbol{x}_* \geq 0$, $\|\boldsymbol{x}_*\| = 1$ and therefore $(\lambda_*, \boldsymbol{x}_*)$, with $\lambda_* = \rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$, is a solution of the conditional affine eigenvalue problem (3.2). Moreover, $(\boldsymbol{x}_*, \boldsymbol{c}_*)$ is a dual pair with respect to $\|\cdot\|$. And since $\lambda_* > \rho(A)$, by Theorem 3.7, $(\lambda_*, \boldsymbol{x}_*)$ is the only solution to problem (3.2).

On the other hand, suppose that $(\lambda_*, \boldsymbol{x}_*)$ is a solution to problem (3.2) with $\lambda_* > \rho(A)$. Let $\boldsymbol{d}$ be a nonnegative vector in the dual of $\boldsymbol{x}_*$. Then $\lambda_*\boldsymbol{x}_* = (A + \boldsymbol{b}\boldsymbol{d}^T)\boldsymbol{x}_*$, so $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) \geq \lambda_* > \rho(A)$ and therefore $\max_{\|\boldsymbol{c}\|^\star=1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. □

Theorem 3.19 shows how a problem of maximizing the spectral radius of particular *rank-one perturbations* of nonnegative matrices is related to a conditional affine eigenvalue problem. The problem of maximizing the spectral radius of *diagonal perturbations* of a nonnegative matrices has been studied by several authors [44, 54, 67]. Han et al. [54] have also studied the problem of maximizing the spectral radius of *fixed Frobenius norm perturbations* of nonnegative matrices. In particular, they prove the following result.

Proposition 3.20 (Han et al. [54]). *Let A be a nonnegative matrix. Then*

$$\lambda_* = \max_{\|X\|_F=1} \rho(A + X) \quad \textit{if and only if} \quad |||(\lambda^* I - A)^{-1}|||_2 = 1.$$

A similar result can be stated in our case as a direct consequence of Theorem 3.19.

Proposition 3.21. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. Suppose that $\max_{\|\boldsymbol{c}\|^\star=1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. Then*

$$\lambda_* = \max_{\|\boldsymbol{c}\|^\star=1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T) \quad \textit{if and only if} \quad \|(\lambda^* I - A)^{-1}\boldsymbol{b}\| = 1.$$

3.4 Particular norms

In this section, we see how Theorem 3.19 can be specialized for the ℓ_1, ℓ_∞ and ℓ_2 norms, denoted respectively by $\|\cdot\|_1$, $\|\cdot\|_\infty$ and $\|\cdot\|_2$.

Let us first consider the ℓ_1 norm. Let $\boldsymbol{u} \gneqq 0$, $\|\boldsymbol{u}\|_1 = 1$ be a non-negative normalized vector. The dual of $\boldsymbol{u}$ with respect to $\|\cdot\|_1$ is given by

$$\{\boldsymbol{c} \in \mathbb{R}^n \colon \|\boldsymbol{c}\|_\infty = \boldsymbol{c}^T\boldsymbol{u} = 1\},$$

since $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are dual to each other. Clearly, the vector $\mathbf{1} \in \mathbb{R}^n$ of all ones belongs to the dual of $\boldsymbol{u}$. Moreover $\mathbf{1} \geq \boldsymbol{c}$ for any vector $\boldsymbol{c} \in \mathbb{R}^n$ such that $\|\boldsymbol{c}\|_\infty = 1$, and hence, for a nonnegative matrix A and a nonnegative vector $\boldsymbol{b}$,

$$\rho(A + \boldsymbol{b}\mathbf{1}^T) = \max_{\|\boldsymbol{c}\|_\infty = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T).$$

Therefore, Theorem 3.19 can be specialized as follows.

Corollary 3.22. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. If $\rho(A + \boldsymbol{b}\mathbf{1}^T) > \rho(A)$, then the conditional eigenvalue problem*

$$\lambda\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\|_1 = 1,$$

has a unique solution $(\lambda_*, \boldsymbol{x}_*)$, where $\lambda_* = \rho(A + \boldsymbol{b}\mathbf{1}^T)$ and $\boldsymbol{x}_*$ is the unique normalized Perron vector of $A + \boldsymbol{b}\mathbf{1}^T$.

This means that, for the ℓ_1 norm, the solution of the conditional affine eigenvalue problem (3.2) is explicitly defined. Actually, under the hypotheses of Theorem 3.19, the iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k) + \boldsymbol{b}}{\|A\boldsymbol{x}(k) + \boldsymbol{b}\|_1},$$

for $\boldsymbol{x}_0 \geq 0$, is equivalent to the iteration

$$\boldsymbol{x}(k+1) = \frac{(A + \boldsymbol{b}\mathbf{1}^T)\boldsymbol{x}(k)}{\|(A + \boldsymbol{b}\mathbf{1}^T)\boldsymbol{x}(k)\|_1},$$

that is, the power method applied to the matrix $A + \boldsymbol{b}\mathbf{1}^T$.

Remark 3.23. PageRank iteration $\boldsymbol{\pi}(k+1)^T = c\boldsymbol{\pi}(k)^T P + (1-c)\boldsymbol{z}^T$ (see Section 2.2.1) can be seen as a particular case of the normalized affine iteration with the ℓ_1 norm. Taking $A = P^T$, $\boldsymbol{b} = (1-c)\boldsymbol{z}/c$ and $\boldsymbol{x}(0) = \boldsymbol{\pi}(0)$, the iterates $\boldsymbol{x}(k)$ and $\boldsymbol{\pi}(k)$ are equal for every $k \in \mathbb{N}$. $\diamond$

Consider now the ℓ_∞ norm. The dual of a vector $\boldsymbol{u} \gneqq 0$, $\|\boldsymbol{u}\|_\infty = 1$ with respect to $\|\cdot\|_\infty$ is

$$\{\boldsymbol{c} \in \mathbb{R}^n \colon \|\boldsymbol{c}\|_1 = \boldsymbol{c}^T\boldsymbol{u} = 1\}.$$

Clearly, there exists at least a basis vector $\boldsymbol{e}_k$ in the dual of $\boldsymbol{u}$, which satisfies $\boldsymbol{u}_k = \max_i \boldsymbol{u}_i = 1$. As it was noticed in Example 3.13, if $\rho(A + \boldsymbol{b}\boldsymbol{e}_i^T) = \rho(A + \boldsymbol{b}\boldsymbol{e}_j^T) > \rho(A)$, then the convex combination $A + \boldsymbol{b}(\alpha\boldsymbol{e}_i^T + (1-\alpha)\boldsymbol{e}_j^T)$, $0 \le \alpha \le 1$, has also the same spectral radius and Perron vector. Therefore, Theorem 3.19 can be specialized in the following way.

Corollary 3.24. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. If $\rho(A + \boldsymbol{b}\boldsymbol{e}_\ell^T) = \max_i \rho(A + \boldsymbol{b}\boldsymbol{e}_i^T) > \rho(A)$, then the conditional eigenvalue problem*

$$\lambda\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \ge 0, \quad \|\boldsymbol{x}\|_\infty = 1,$$

has a unique solution $(\lambda_*, \boldsymbol{x}_*)$, where $\lambda_* = \rho(A + \boldsymbol{b}\boldsymbol{e}_\ell^T)$ and $\boldsymbol{x}_*$ is the unique normalized Perron vector of $A + \boldsymbol{b}\boldsymbol{e}_\ell^T$.

Let us notice that in this case, in contrast to the case of the ℓ_1 norm, it can not be said *a priori* which matrix $A + \boldsymbol{b}\boldsymbol{e}_i^T$ will give the solution, but there are potentially n choices.

It is known that the ℓ_2 norm is its own dual norm, and that the dual of a vector $\boldsymbol{u} \ge 0$, $\|\boldsymbol{u}\|_2 = 1$, with respect to $\|\cdot\|_2$ is the singleton $\{\boldsymbol{u}\}$. Therefore, Proposition 3.18 and Theorem 3.19 can be particularized as follows.

Corollary 3.25. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. If there exists a nonnegative vector $\boldsymbol{d}$, with $\|\boldsymbol{d}\|_2 = 1$, such that $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) > \rho(A)$, then*

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|_2 = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T),$$

with $\boldsymbol{c}_* \in \mathbb{R}^n_{\ge 0}$, $\|\boldsymbol{c}_*\|_2 = 1$, if and only if $\boldsymbol{c}_*$ is the Perron vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$ and $\boldsymbol{c}_* \in \mathbb{R}^n_{\ge 0}$, $\|\boldsymbol{c}_*\|_2 = 1$.

Corollary 3.26. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\boldsymbol{c}_*$ be a nonnegative vector, with $\|\boldsymbol{c}_*\|_2 = 1$, such that*

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|_2 = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T).$$

If $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) > \rho(A)$, then the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\|_2 = 1,$$

has a unique solution $(\lambda_*, \boldsymbol{x}_*)$, where $\lambda_* = \rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$ and $\boldsymbol{x}_* = \boldsymbol{c}_*$.

3.5 Relation between graph-theoretic and optimization approaches

In this section, we show how the graph-theoretic and the spectral radius maximizing approaches can be related. In particular, we are interested in the existence of nonnegative vectors $\boldsymbol{c}$ on the unit sphere of the dual norm such that the principal submatrix $(A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}$ has particular properties, where $\mathcal{B}$ is the set of nodes having an access to $\mathrm{supp}(\boldsymbol{b})$ in $\mathcal{G}(A)$. Remember that in Theorem 3.19 we have already proved the equivalence $(b) \Leftrightarrow (B)$ of Theorem 3.1. In this section, we prove all the remaining implications of Theorem 3.1.

Lemma 3.27. *Let M be a nonnegative matrix and $\boldsymbol{d}$ a nonnegative vector such that every node has an access to $\mathrm{supp}(\boldsymbol{d})$ in the graph $\mathcal{G}(M)$. Then $\rho(M + \boldsymbol{d}\boldsymbol{c}^T) > \rho(M)$ for every vector $\boldsymbol{c} > 0$.*

Proof. Let $\boldsymbol{c} > 0$ and let $\boldsymbol{x} \geq 0$ be a Perron vector of the matrix $M + \boldsymbol{d}\boldsymbol{c}^T$ such that $\boldsymbol{c}^T\boldsymbol{x} = 1$. Then $M\boldsymbol{x} + \boldsymbol{d} = (M + \boldsymbol{d}\boldsymbol{c}^T)\boldsymbol{x} = \rho(M + \boldsymbol{d}\boldsymbol{c}^T)\boldsymbol{x}$, and therefore $\rho(M) < \rho(M + \boldsymbol{d}\boldsymbol{c}^T)$ by Lemma 3.3. □

The following proposition shows that having $\rho(A_{\mathcal{B}}) = \rho(A)$ ensures the existence of a nonnegative vector $\boldsymbol{c}$ satisfying $\|\boldsymbol{c}\|^\star = 1$ and $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, while if $\rho(A_{\mathcal{B}}) < \rho(A)$ then the existence of such normalized nonnegative vector $\boldsymbol{c}$ depends on the norm of $\boldsymbol{b}$ and the gap between the spectral radii $\rho(A_{\mathcal{B}})$ and $\rho(A)$. This result corresponds to the equivalence $(a) \Leftrightarrow (a')$ of Theorem 3.1.

Proposition 3.28. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Then, $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$ for every $\boldsymbol{c} \geq 0$ such that $\boldsymbol{c}_{\mathcal{B}} > 0$ if and only if $\rho(A_{\mathcal{B}}) = \rho(A)$. Moreover if $\rho(A_{\mathcal{B}}) < \rho(A)$, there exists $\varepsilon > 0$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho(A)$ for every $\boldsymbol{c} \geq 0$ with $\|\boldsymbol{c}\|^\star \leq \varepsilon$.*

Proof. Suppose $\rho(A_{\mathcal{B}}) = \rho(A)$ and let $\boldsymbol{c} \geq 0$ such that $\boldsymbol{c}_{\mathcal{B}} > 0$. Then by Lemma 3.27,

$$\rho(A + \boldsymbol{b}\boldsymbol{c}^T) \geq \rho(A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T) > \rho(A_{\mathcal{B}}) = \rho(A).$$

On the other hand, suppose $\rho(A_{\mathcal{B}}) < \rho(A)$. Then, for any nonnegative vector $\boldsymbol{c}$,

$$\begin{aligned}\rho(A + \boldsymbol{b}\boldsymbol{c}^T) &= \max\{\rho(A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T), \rho(A_{\bar{\mathcal{B}}})\} \\ &= \max\{\rho(A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T), \rho(A)\},\end{aligned}$$

and since the spectral radius of a matrix is a continuous function of its entries, there exists an $\varepsilon > 0$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho(A)$ for every nonnegative vector $\boldsymbol{c}$ with $\|\boldsymbol{c}\|^\star \leq \varepsilon$. □

Example 3.29. Let us emphasize that $\rho(A_{\mathcal{B}}) = \rho(A)$ is a sufficient but not necessary condition in order to have a unique solution $(\lambda, \boldsymbol{x})$ with $\lambda > \rho(A)$ of the conditional affine eigenvalue problem (3.2). Consider for instance the affine eigenvalue problem given in Example 3.8. Clearly, $\rho(A_{\mathcal{B}}) < \rho(A)$ but the problem has a unique solution $(\lambda, \boldsymbol{x})$ with $\lambda > \rho(A)$. ◇

Since in Theorem 3.19 we proved the equivalence $(b) \Leftrightarrow (B)$ of Theorem 3.1, the implication $(a) \Rightarrow (B)$ of Theorem 3.1 now follows readily.

Corollary 3.30. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. If $\rho(A_{\mathcal{B}}) = \rho(A)$, then the conditional affine eigenvalue problem*

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1,$$

has a unique solution $(\lambda_*, \boldsymbol{x}_*)$, where $\lambda_* = \rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$, with $\boldsymbol{c}_* \geq 0$, $\|\boldsymbol{c}_*\|^\star = 1$, such that

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \max_{\|\boldsymbol{c}\|^\star = 1} \rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A),$$

and $\boldsymbol{x}_*$ is the unique normalized Perron vector of $A + \boldsymbol{b}\boldsymbol{c}_*^T$. Moreover $(\boldsymbol{x}_*, \boldsymbol{c}_*)$ is a dual pair with respect to $\|\cdot\|$.

Remark 3.31. The Path Condition, that we defined in our paper [20, p. 78], is equivalent to the condition $\rho(A_{\mathcal{B}}) = \rho(A)$. $\diamond$

Remark 3.32. Remember that the condition $A\boldsymbol{x} + \boldsymbol{b} > 0$ for all $\boldsymbol{x} \gneqq 0$ ensures the convergence of the normalized affine iterations (3.1), by Krause's Theorem 2.7. In fact, supposing that $\boldsymbol{b} \neq 0$, this condition implies that $\mathcal{B} = \{1, \dots, n\}$, that is, every node has an access to supp($\boldsymbol{b}$) in the graph $\mathcal{G}(A)$. Henceforth, the condition $\rho(A_{\mathcal{B}}) = \rho(A)$ is obviously satisfied and the uniqueness of the solution follows. We have already shown in Example 3.29 that $\rho(A_{\mathcal{B}}) = \rho(A)$ is not necessary to have a unique solution. So neither is the condition $A\boldsymbol{x} + \boldsymbol{b} > 0$ for all $\boldsymbol{x} \gneqq 0$. Let us note finally that obviously, $A\boldsymbol{x} + \boldsymbol{b} > 0$ for all $\boldsymbol{x} \gneqq 0$ is not a necessary condition for $\rho(A_{\mathcal{B}}) = \rho(A)$. $\diamond$

Remark 3.33. Let us also note that existence and uniqueness of the fixed point of the normalized affine iteration (3.1) in the case $A\boldsymbol{x} + \boldsymbol{b} > 0$ for all $\boldsymbol{x} \gneqq 0$ could also be easily proved directly with an argument similar to the proof of Theorem 3.7 (see [20, Appendix A]). $\diamond$

The implication $(b) \Rightarrow (c)$ of Theorem 3.1 follows directly from the following result.

Proposition 3.34. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. If there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$, then this vector $\boldsymbol{c}$ can be chosen so that $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$.*

Proof. Let $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. For all $0 < \varepsilon < 1$, let $\boldsymbol{c}_\varepsilon = (1-\varepsilon)\boldsymbol{c} + \delta_\varepsilon \mathbf{1} > 0$, where $\delta_\varepsilon > 0$ is such that $\|\boldsymbol{c}_\varepsilon\|^\star = 1$. Since $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$ and the spectral radius is continuous, there exists $\varepsilon > 0$ such that $\rho(A + \boldsymbol{b}\boldsymbol{c}_\varepsilon) \geq \rho(A + (1-\varepsilon)\boldsymbol{b}\boldsymbol{c}^T) > \rho(A)$. Since $\boldsymbol{c}_\varepsilon > 0$, the unique initial class of the matrix $A + \boldsymbol{b}\boldsymbol{c}_\varepsilon$ is $\mathcal{B}$, the set of nodes having an access to supp($\boldsymbol{b}$) in $\mathcal{G}(A)$. $\square$

Example 3.35. Let us show by an example that the converse implication of Theorem 3.1 does not hold in general, that is, $(c) \not\Rightarrow (b)$. Consider for instance $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and let $\|\cdot\|$ be the ℓ_1 norm so $\|\cdot\|^\star$ be the ℓ_∞ norm. Then $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho(A)$ for every nonnegative vector $\boldsymbol{c}$ with $\|\boldsymbol{c}\|^\star = 1$, so condition (b) is not satisfied. On the other hand, condition (c) is satisfied for $\boldsymbol{c} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, since the matrix $A + \boldsymbol{b}\boldsymbol{c}^T$ has a unique initial class $\mathcal{B} = \{1\}$ and moreover $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) = \rho(A)$. $\diamond$

Now, let us prove implication $(c) \Rightarrow (D)$ of Theorem 3.1 with the help of an elementary lemma.

Lemma 3.36. *Let M be a nonnegative matrix and let $\boldsymbol{x}$ be a nonnegative eigenvector of M associated to a nonzero eigenvalue. If $\boldsymbol{x}_j > 0$ then $\boldsymbol{x}_i > 0$ for every i which has an access to j in the graph $\mathcal{G}(M)$.*

Proof. Suppose $\lambda \boldsymbol{x} = M\boldsymbol{x}$ with $\lambda > 0$. If $\boldsymbol{x}_j > 0$ and there is a link (i, j) in the graph $\mathcal{G}(M)$, i.e. $M_{ij} > 0$ then clearly $\lambda \boldsymbol{x}_i = \sum_k M_{ik}\boldsymbol{x}_k \geq M_{ij}\boldsymbol{x}_j > 0$. The conclusion then follows by induction. □

Proposition 3.37. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. If there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$, then the conditional affine eigenvalue problem* (3.2) *has a unique solution.*

Proof. Let $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$. Suppose by contradiction that the problem (3.2) has several solutions. Then, by Theorems 3.4 and 3.7, it must have a solution $(\rho(A), \boldsymbol{x})$ with $\boldsymbol{x}_{\mathcal{B}} > 0$ and $\boldsymbol{x}_{\bar{\mathcal{B}}} \neq 0$ such that $\rho(A)\boldsymbol{x}_{\bar{\mathcal{B}}} = A_{\bar{\mathcal{B}}}\boldsymbol{x}_{\bar{\mathcal{B}}}$. By the property of the dual norms, $\boldsymbol{c}^T\boldsymbol{x} \leq 1$, so we have $\rho(A)\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b} \geq (A + \boldsymbol{b}\boldsymbol{c}^T)\boldsymbol{x}$ and therefore

$$\rho(A)\boldsymbol{x}_{\mathcal{B}} \geq (A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}} + (A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}\bar{\mathcal{B}}}\boldsymbol{x}_{\bar{\mathcal{B}}}.$$

Since $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$, it has an access to every node of $\bar{\mathcal{B}}$ in $\mathcal{G}(A + \boldsymbol{b}\boldsymbol{c}^T)$. So $(A_{\mathcal{B}\bar{\mathcal{B}}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\bar{\mathcal{B}}}^T)\boldsymbol{x}_{\bar{\mathcal{B}}} \neq 0$ by Lemma 3.36. Therefore $\rho(A)\boldsymbol{x}_{\mathcal{B}} \gneqq (A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T)\boldsymbol{x}_{\mathcal{B}}$, and since $A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T$ is irreducible, it follows by Lemma 3.2(c) that $\rho(A) > \rho(A_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}}\boldsymbol{c}_{\mathcal{B}}^T)$, which contradicts the hypotheses. □

Example 3.38. The converse of implication $(c) \Rightarrow (D)$ of Theorem 3.1 is not true. Consider for instance $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and let $\|\cdot\|$ and $\|\cdot\|^\star$ be the ℓ_2 norm. The affine eigenvalue problem has a unique solution $(\lambda, \boldsymbol{x})$ with $\lambda = 2$ and $\boldsymbol{x} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, but there does not exist $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$. ◇

Let us now look at the implication $(D) \Rightarrow (e)$ of Theorem 3.1.

Proposition 3.39. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. If the conditional affine eigenvalue problem* (3.2) *has a unique solution then there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$.*

Proof. If the conditional affine eigenvalue problem (3.2) has a unique solution $(\lambda, \boldsymbol{x})$ then by Theorem 3.7, we must have $\lambda \geq \rho(A)$ and $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$. In the case where $\lambda > \rho(A)$, the conclusion follows from Theorem 3.19 and Proposition 3.34. Let us therefore consider the case where $\lambda = \rho(A)$. Let $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ in the dual of $\boldsymbol{x}$. Since $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$, we have

$$\rho(A)\boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}} = (A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}}.$$

It follows that $\rho(A)$ is an eigenvalue of $(A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}$ and hence $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$. □

Example 3.40. Let us show by an example that the converse of implication $(D) \Rightarrow (e)$ of Theorem 3.1 does not hold in general. Consider for instance the conditional affine eigenvalue problem given in Example 3.10 with $\|\cdot\|$ the ℓ_∞ norm. Then with $\boldsymbol{c} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, we have $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) = \rho(A)$. But there are infinitely many solutions to the conditional affine eigenvalue problem. ◇

Example 3.41. Let us show by an example that it is not possible in general to characterize pairs A and $\boldsymbol{b}$ leading to a unique solution for the eigenvalue problem (3.2) with equivalences like $(b) \Leftrightarrow (B)$ or $(e) \Leftrightarrow (E)$ of Theorem 3.1. Indeed, consider again $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ and $\boldsymbol{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ of Example 3.10. Then the set

$$\begin{aligned} &\{A + \boldsymbol{b}\boldsymbol{c}^T : \boldsymbol{c} \geq 0, \|\boldsymbol{c}\|^\star = 1 \text{ and } \rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)\} \\ &\qquad = \begin{cases} \left\{\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\right\} & \text{if } \|\cdot\| \text{ is the } \ell_2 \text{ or } \ell_\infty \text{ norm,} \\ \left\{\begin{pmatrix} 2 & \alpha \\ 0 & 2 \end{pmatrix} : \alpha \in [0,1]\right\} & \text{if } \|\cdot\| \text{ is the } \ell_1 \text{ norm.} \end{cases} \end{aligned}$$

On the other hand, the conditional affine eigenvalue problem has a unique solution for $\|\cdot\|$ being the ℓ_2 or ℓ_1 norm and infinitely many solutions for the ℓ_∞ norm. ◇

We now prove the equivalence $(e) \Leftrightarrow (E)$ of Theorem 3.1.

Proposition 3.42. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. Then the conditional affine eigenvalue problem* (3.2) *has a solution $(\lambda, \boldsymbol{x})$ with $\lambda \geq \rho(A)$ and $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$ if and only if there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$.*

Proof. By Theorem 3.7, let $(\lambda, \boldsymbol{x})$ be the solution of problem (3.2) such that $\boldsymbol{x}_{\mathcal{B}} > 0$ and $\boldsymbol{x}_{\bar{\mathcal{B}}} = 0$. Suppose first that $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ is such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$. Since $\boldsymbol{c}^T\boldsymbol{x} \leq 1$ by property of the dual norm, we have $\lambda \boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}} \geq (A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}}$. Therefore, by Lemma 3.2(b), $\lambda \geq \rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$.

On the other hand, suppose that $\lambda \geq \rho(A)$. Let $\boldsymbol{c} \geq 0$, $\|c\|^\star = 1$ be in the dual of $\boldsymbol{x}$. Then $\lambda \boldsymbol{x}_{\mathcal{B}} = A_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}} + \boldsymbol{b}_{\mathcal{B}} = (A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}\boldsymbol{x}_{\mathcal{B}}$. Therefore $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \lambda \geq \rho(A)$. □

We showed in Example 3.41 that it is not possible in general to characterize pairs A and $\boldsymbol{b}$ leading to a unique solution for the eigenvalue problem (3.2) with equivalences like these considered in Theorem 3.1. Now, in the particular case *when the norm $\|\cdot\|$ is strictly monotone*, we have the following characterization of conditional affine eigenvalue problems having a unique solution.

Theorem 3.43. *Let A be a nonnegative matrix and $\boldsymbol{b}$ be a nonnegative vector. Let $\|\cdot\|$ be a strictly monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. Let $\mathcal{B}$ be the set of nodes which have an access in the graph $\mathcal{G}(A)$ to the nodes i such that $\boldsymbol{b}_i > 0$. Then the conditional affine eigenvalue problem* (3.2) *has a unique solution if and only if there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$.*

Proof. This follows directly from Corollary 3.11 and Proposition 3.42. □

We now consider the particular case of a norm defined by

$$\|\cdot\| \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0} \colon \boldsymbol{x} \mapsto \boldsymbol{d}^T|\boldsymbol{x}|,$$

for some given positive vector $\boldsymbol{d} > 0$, as for instance the ℓ_1 norm. Such a norm is strictly monotone, and its dual norm is defined by $\|\cdot\|^\star \colon \mathbb{R}^n \to$

$\mathbb{R}_{\geq 0}\colon \boldsymbol{y} \mapsto \max_i |\boldsymbol{y}_i|/\boldsymbol{d}_i$. For such norms, the conditions (c), (D) and (e) of Theorem 3.1 are equivalent, as it follows from Theorem 3.43 and the following proposition.

Proposition 3.44. *Let A be a nonnegative matrix, $\boldsymbol{b}$ be a nonnegative vector and $\boldsymbol{d}$ be a positive vector. Let $\|\cdot\|$ be the vector norm defined by $\|\boldsymbol{x}\| = \boldsymbol{d}^T|\boldsymbol{x}|$ and let $\|\cdot\|^\star$ be its dual norm. If there exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ then $\rho((A+\boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A+\boldsymbol{b}\boldsymbol{d}^T$.*

Proof. Clearly, $\|\boldsymbol{d}\|^\star = 1$ and $\boldsymbol{d} \geq \boldsymbol{c}$ for every vector $\boldsymbol{c}$ such that $\|\boldsymbol{c}\|^\star = 1$. So, if $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|_\infty = 1$ is such that $\rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$, then we have $\rho((A+\boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{B}}) \geq \rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$. Moreover, since $\boldsymbol{d}$ is positive, $\mathcal{B}$ must be the unique initial class of $A+\boldsymbol{b}\boldsymbol{d}^T$. □

Let us close this section by a description of the connections with weak Sraffa matrices, a concept introduced by Krause [78]. By definition, a *weak Sraffa matrix* is a nonnegative matrix M that has a unique initial class $\mathcal{I}$ which is moreover a basic class, i.e., $\rho(M_{\mathcal{I}}) = \rho(M)$.

The following proposition shows the equivalence of conditions (c) and (c') of Theorem 3.1.

Proposition 3.45. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector. Let $\|\cdot\|$ be a monotone vector norm and let $\|\cdot\|^\star$ be its dual norm. There exists $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A+\boldsymbol{b}\boldsymbol{c}^T$ if and only if there exists $\widetilde{\boldsymbol{c}} \geq 0$, $\|\widetilde{\boldsymbol{c}}\|^\star = 1$ such that $A+\boldsymbol{b}\widetilde{\boldsymbol{c}}^T$ is a weak Sraffa matrix.*

Proof. Clearly, if $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $\rho((A+\boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{B}}) \geq \rho(A)$ and $\mathcal{B}$ is the unique initial class of $A+\boldsymbol{b}\boldsymbol{c}^T$ then $A+\boldsymbol{b}\boldsymbol{c}^T$ is a weak Sraffa matrix.

On the other hand, suppose that $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ is such that $A+\boldsymbol{b}\boldsymbol{c}^T$ has a unique initial class which is moreover basic. We want to prove that this vector $\boldsymbol{c}$ can be chosen such that this initial class is exactly the set $\mathcal{B}$ of nodes having an access to $\operatorname{supp}(\boldsymbol{b})$ in the graph $\mathcal{G}(A)$.

Let $\mathcal{I}_1, \ldots \mathcal{I}_r$ be the initial classes of A which have an access to $\operatorname{supp}(\boldsymbol{b})$, so $\mathcal{I}_i \subseteq \mathcal{B}$ for every $i \in \{1, \ldots, r\}$. Let $\boldsymbol{c} \geq 0$, $\|\boldsymbol{c}\|^\star = 1$ such that $A + \boldsymbol{b}\boldsymbol{c}^T$ has a unique initial class and that this class is moreover basic. Every node in $\mathcal{B}$ must have an access in the graph $\mathcal{G}(A + \boldsymbol{b}\boldsymbol{c}^T)$ to every node i such that $\boldsymbol{c}_i \neq 0$. Moreover, no node in $\bar{\mathcal{B}}$ has an access to a node in $\mathcal{B}$ in $\mathcal{G}(A + \boldsymbol{b}\boldsymbol{c}^T)$. So two cases can occur. Either $\boldsymbol{c}_{\mathcal{I}_i} \neq 0$ for every $\mathcal{I}_i$ with $i \in \{1, \dots, r\}$ and in this case the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$ is exactly $\mathcal{B}$. Either there is exactly one class $\mathcal{I}_i$ with $i \in \{1, \dots, r\}$ such that $\boldsymbol{c}_{\mathcal{I}_i} = 0$. So $\mathcal{I}_i$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$ and therefore it is basic. Hence $\rho(A + \boldsymbol{b}\boldsymbol{c}^T) = \rho((A + \boldsymbol{b}\boldsymbol{c}^T)_{\mathcal{I}_i}) = \rho(A_{\mathcal{I}_i})$, and it follows that $\rho(A_{\mathcal{I}_i}) = \rho(A) \geq \rho(A_{\bar{\mathcal{B}}})$. But then, we can choose for instance $\widetilde{\boldsymbol{c}} = \mathbf{1}/\|\mathbf{1}\|^\star$ instead of $\boldsymbol{c}$. Indeed, in this case, $\mathcal{B}$ is the unique initial class of $A + \boldsymbol{b}\boldsymbol{c}^T$. Moreover

$$\rho((A + \boldsymbol{b}\widetilde{\boldsymbol{c}}^T)_{\mathcal{B}}) \geq \rho((A + \boldsymbol{b}\widetilde{\boldsymbol{c}}^T)_{\mathcal{I}_i}) \geq \rho(A_{\mathcal{I}_i}) = \rho(A)$$
$$\rho((A + \boldsymbol{b}\widetilde{\boldsymbol{c}}^T)_{\bar{\mathcal{B}}}) = \rho(A_{\bar{\mathcal{B}}}) \leq \rho(A),$$

so $\mathcal{B}$ is also a basic class of $A + \boldsymbol{b}\boldsymbol{c}^T$. □

Remark 3.46. Weak Sraffa matrices are exactly the matrices that have a unique nonnegative eigenvector (up to a scalar multiple), as noted and proved in [78]. This nice property can be also derived from Theorems 3.1 and 3.7 of [108]. ◇

Remark 3.47. Weak Sraffa matrices are a weak version of what Krause calls Sraffa matrices in [79]. A Sraffa matrix is a nonnegative matrix whose unique initial class is also its unique basic class. Sraffa matrices are exactly the nonnegative matrices that have only one positive left eigenvector [79, 108]. ◇

3.6 Approximation of a Perron pair by the solution of a conditional affine eigenvalue problem

In this section, we see how a solution $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ of

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \varepsilon \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1, \tag{3.6}$$

approaches a Perron vector of A for small $\varepsilon > 0$. Since we are interested in the case where ε goes to 0, we only consider nonnegative matrix and vector A and $\boldsymbol{b}$ such that, for each $\varepsilon > 0$, there is a unique solution $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ of (3.6).

Lemma 3.48. *Let A be a nonnegative matrix, $\boldsymbol{b}$ a nonnegative vector and $\|\cdot\|$ a monotone vector norm. Then the conditional affine eigenvalue problem* (3.6) *has exactly one solution $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ for every $\varepsilon > 0$ if and only if $\rho(A_{\mathcal{B}}) = \rho(A)$, where $\mathcal{B}$ is the set of nodes having an access to* $\operatorname{supp}(\boldsymbol{b})$ *in the graph $\mathcal{G}(A)$.*

Proof. If $\rho(A_{\mathcal{B}}) = \rho(A)$, then clearly the conditional affine eigenvalue problem (3.6) has exactly one solution $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ for every $\varepsilon > 0$ by implications $(a') \Rightarrow (B) \Rightarrow (c) \Rightarrow (D)$ of Theorem 3.1. On the other hand, suppose the problem (3.6) has exactly one solution $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ for every $\varepsilon > 0$. For each $\varepsilon > 0$, let $\boldsymbol{c}_\varepsilon \geq 0$, $\|\boldsymbol{c}_\varepsilon\|^\star = 1$ be in the dual of $\boldsymbol{x}_\varepsilon$. Then, by Theorem 3.7, we must have $\rho(A_{\mathcal{B}}) < \lambda_\varepsilon \leq \rho((A + \varepsilon \boldsymbol{b}\boldsymbol{c}_\varepsilon^T)_{\mathcal{B}})$. By Proposition 3.28, this can possible for every $\varepsilon > 0$ only if $\rho(A_{\mathcal{B}}) = \rho(A)$. □

Proposition 3.49. *Let A be a nonnegative matrix and $\boldsymbol{b}$ a nonnegative vector such that $\rho(A_{\mathcal{B}}) = \rho(A)$. Let $\|\cdot\|$ a monotone vector norm. For each $\varepsilon > 0$, let $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ be the unique solution of the conditional affine eigenvalue problem* (3.6). *Then*

- (a) $\lim_{\varepsilon \to 0} \lambda_\varepsilon = \rho(A)$;
- (b) $\boldsymbol{x}_0 \doteq \lim_{\varepsilon \to 0} x_\varepsilon$ *exists and is a Perron vector of A;*
- (c) *the sets $\{\alpha \boldsymbol{x}_\varepsilon : \varepsilon > 0, \alpha \in \mathbb{R}\}$ and $\{\alpha \boldsymbol{x}_0 : \alpha \in \mathbb{R}\}$ do not depend on the norm $\|\cdot\|$.*

Proof. By Proposition 3.28 and Theorem 3.19, $\lambda_\varepsilon = \max_{\|\boldsymbol{c}\|^\star = 1} \rho(A + \varepsilon \boldsymbol{b}\boldsymbol{c}^T)$. The first statement then follows from the continuity of the spectral radius of a matrix on its entries. Moreover, by Lemma 3.27, $\lambda_\varepsilon > \rho(A)$ and λ_ε is strictly increasing in ε since $\rho(A_{\mathcal{B}}) = \rho(A)$. Therefore, for all $\delta > 0$ we can define without ambiguity $\varepsilon_\delta > 0$ such that

$$\lambda_{\varepsilon_\delta} = \rho(A) + \delta.$$

Hence, for all $\delta > 0$, we can write

$$\boldsymbol{x}_{\varepsilon_\delta} = \varepsilon_\delta((\rho(A) + \delta)I - A)^{-1}\boldsymbol{b}. \tag{3.7}$$

For every $\delta > 0$, $\|\boldsymbol{x}_{\varepsilon_\delta}\| = 1$, so the limit $\lim_{\varepsilon \to 0} \boldsymbol{x}_\varepsilon = \lim_{\delta \to 0} \boldsymbol{x}_{\varepsilon_\delta}$ exists by Lemma 2.3. Its expression is given in terms of the Drazin inverse of $\rho(A)I - A$. So we can define

$$\boldsymbol{x}_0 \doteq \lim_{\varepsilon \to 0} \boldsymbol{x}_\varepsilon.$$

Clearly, $\boldsymbol{x}_0$ is a Perron vector of A, since

$$(\rho(A)I - A)\boldsymbol{x}_0 = \lim_{\varepsilon\to 0}(\lambda_\varepsilon I - A)\lim_{\varepsilon\to 0}\boldsymbol{x}_\varepsilon = \lim_{\varepsilon\to 0}(\lambda_\varepsilon I - A)\boldsymbol{x}_\varepsilon = \lim_{\varepsilon\to 0}\varepsilon\boldsymbol{b} = 0.$$

Finally, from the equation (3.7), it is clear that none of the sets $\{\alpha\boldsymbol{x}_\varepsilon\colon \varepsilon > 0, \alpha \in \mathbb{R}\}$ or $\{\alpha\boldsymbol{x}_0\colon \alpha \in \mathbb{R}\}$ depends on the norm $\|\cdot\|$. □

Example 3.50. Let us illustrate Proposition 3.49. Consider for instance

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}.$$

Basis vectors for the eigenspace of A associated to its Perron root $\rho(A)$ are

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Clearly, $\mathcal{B} = \{1, 2, 3\}$ and $\rho(A_{\mathcal{B}}) = \rho(A)$. For every $\varepsilon > 0$, let $(\lambda, \boldsymbol{x}_\varepsilon)$ be the unique solution of problem (3.6) with $\|\cdot\|$ the ℓ_1 norm and let $(\widetilde{\lambda}, \widetilde{\boldsymbol{x}}_\varepsilon)$ be the unique solution of problem (3.6) with $\|\cdot\|$ the ℓ_∞ norm. Then

$$\lim_{\varepsilon\to 0}\boldsymbol{x}_\varepsilon = \begin{pmatrix} 1/2 \\ 1/3 \\ 1/6 \\ 0 \end{pmatrix} \quad \text{and} \quad \lim_{\varepsilon\to 0}\widetilde{\boldsymbol{x}}_\varepsilon = \begin{pmatrix} 1 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix},$$

which are parallel to $2\boldsymbol{v}_1 + \boldsymbol{v}_2$. Moreover, as Proposition 3.49 shows, the orbits $\{\boldsymbol{x}_\varepsilon\colon \varepsilon > 0\}$ and $\{\widetilde{\boldsymbol{x}}_\varepsilon/\|\widetilde{\boldsymbol{x}}_\varepsilon\|_1\colon \varepsilon > 0\}$ are equal. However, for some $\varepsilon > 0$ the vectors $\boldsymbol{x}_\varepsilon$ and $\widetilde{\boldsymbol{x}}_\varepsilon/\|\widetilde{\boldsymbol{x}}_\varepsilon\|_1$ are not equal in general. This example is illustrated in Figure 3.4 ◇

In the next proposition, we show that, when the matrix A is symmetric, the Perron vector that is approached is the orthogonal projection of the vector $\boldsymbol{b}$ on the invariant subspace of A corresponding to its Perron root $\rho(A)$.

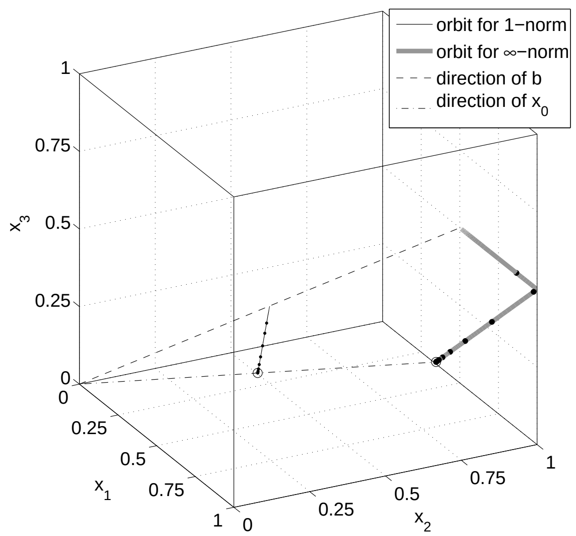


Figure 3.4: The orbits $\{\boldsymbol{x}_\varepsilon \colon \varepsilon > 0\}$ and $\{\tilde{\boldsymbol{x}}_\varepsilon \colon \varepsilon > 0\}$ defined in Example 3.50 converge to a Perron vector of A. For a few values of $\varepsilon > 0$, $\boldsymbol{x}_\varepsilon$ and $\tilde{\boldsymbol{x}}_\varepsilon$ are represented by a dot.

Proposition 3.51. *Let A be a nonnegative and symmetric matrix and $\boldsymbol{b}$ a nonnegative vector such that $\rho(A_{\mathcal{B}}) = \rho(A)$. Let $\|\cdot\|$ be a monotone vector norm. For each $\varepsilon > 0$, let $(\lambda_\varepsilon, \boldsymbol{x}_\varepsilon)$ be the unique solution of the conditional affine eigenvalue problem* (3.6) *and let $\boldsymbol{x}_0 = \lim_{\varepsilon \to 0} \boldsymbol{x}_\varepsilon$. Then*

$$\boldsymbol{x}_0 = \frac{U_1 U_1^T \boldsymbol{b}}{\|U_1 U_1^T \boldsymbol{b}\|},$$

where U_1 is an orthonormal basis of the invariant subspace of A corresponding to $\rho(A)$.

Proof. Since the matrix A is symmetric, it is diagonalizable by unitary transformation as

$$U^T A U = \begin{pmatrix} \rho(A) I & 0 \\ 0 & D_2 \end{pmatrix},$$

where $U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$ is a unitary matrix, and where $\rho(A)$ is not an eigenvalue of the diagonal block D_2. So, for every $\varepsilon > 0$,

$$\begin{aligned} \boldsymbol{x}_\varepsilon &= \varepsilon(\lambda_\varepsilon I - A)^{-1}\boldsymbol{b} \\ &= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \varepsilon(\lambda_\varepsilon - \rho(A))^{-1} I & 0 \\ 0 & \varepsilon(\lambda_\varepsilon I - D_2)^{-1} \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} \boldsymbol{b} \\ &= \frac{\varepsilon}{\lambda_\varepsilon - \rho(A)} U_1 U_1^T \boldsymbol{b} + \varepsilon U_2 (\lambda_\varepsilon I - D_2)^{-1} U_2^T \boldsymbol{b}. \end{aligned}$$

Let $\zeta = \min\{\rho(A) - \lambda \colon \lambda \neq \rho(A) \text{ is an eigenvalue of } A\}$ be the eigengap of A. Then $\|\varepsilon U_2(\lambda_\varepsilon I - D_2)^{-1} U_2^T \boldsymbol{b}\| < \varepsilon \zeta^{-1} \|U_2 U_2^T \boldsymbol{b}\|$ for every $\varepsilon > 0$. Therefore, since $\|\boldsymbol{x}_\varepsilon\| = 1$ for every $\varepsilon > 0$, we have $\lim_{\varepsilon \to 0} \boldsymbol{x}_\varepsilon = U_1 U_1^T \boldsymbol{b} / \|U_1 U_1^T \boldsymbol{b}\|$. □

Remark 3.52. Note that it follows from the proof of Proposition 3.51 that λ_ε behaves like $\rho(A) + \varepsilon$ when ε goes to zero. This could also have been noted by Lemma 2.3. Indeed, since $\rho(A_{\mathcal{B}}) = \rho(A)$, there does not exist an $\boldsymbol{x} \geq 0$ such that $\rho(A)\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}$, by Carlson–Victory Theorem 3.4. In other words, $\boldsymbol{b}$ is not in the range of $\rho(A)I - A$. On the other hand, the index of $\rho(A)I - A$ is equal to one, since A is a symmetric nonnegative matrix. Therefore by Lemma 2.3, the limit $\lim_{\varepsilon \to 0} \varepsilon((\rho(A) + \varepsilon)I - A)^{-1}\boldsymbol{b}$ exists and is nonzero. ◇

Remark 3.53. As we noted in Section 2.2.3, p. 52, the vector

$$\boldsymbol{x}_0 = \frac{U_1 U_1^T \boldsymbol{b}}{\|U_1 U_1^T \boldsymbol{b}\|}$$

is proportional to the average between the vectors

$$\lim_{k \to \infty} \frac{A^{2k}\boldsymbol{b}}{\|A^{2k}\boldsymbol{b}\|} \quad \text{and} \quad \lim_{k \to \infty} \frac{A^{2k+1}\boldsymbol{b}}{\|A^{2k+1}\boldsymbol{b}\|},$$

that is, the limit vectors of the even and odd subsequences resulting of the power method on the matrix A with $\boldsymbol{b}$ taken as initial vector. ◇

3.7 Experiments on a subgraph of the Web

In this section, we briefly present our experiments on a large-scale example. As we noticed in Remark 3.23, the PageRank iteration can

	# iterations	$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$	$\lvert\lambda_2(A + \boldsymbol{b}\boldsymbol{c}_*^T)\rvert$	$\lvert\lambda_2/\lambda_1\rvert$
ℓ_1 norm	118	1.1765	1	0.85
ℓ_2 norm	4,562	1.0049	1.0000	0.9951
ℓ_∞ norm	7,920	1.0029	1.0000	0.9971

Figure 3.5: Number of iterations needed for convergence and estimation of the speed of convergence near the fixed point for the ℓ_1, ℓ_2 and ℓ_∞ norms on a large scale-example.

be seen as a particular case of the normalized affine iteration with the ℓ_1 norm, by taking $A = P^T$ and $\boldsymbol{b} = (1-c)\boldsymbol{z}/c$. For these A and $\boldsymbol{b}$, we compare the convergence of the iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k) + \boldsymbol{b}}{\|A\boldsymbol{x}(k) + \boldsymbol{b}\|},$$

for the ℓ_1, ℓ_2 and ℓ_∞ norm on a large scale example.

We consider a subgraph of the Web with about 280,000 nodes which has been obtained by S. Kamvar from a crawl on the Stanford web [69]. The damping factor is taken as $c = 1$ and the personalization vector as $\boldsymbol{z} = \frac{1}{n}\mathbf{1}$. The scaled adjacency matrix P is preprocessed in order to be stochastic. We take $A = P^T$, $\boldsymbol{b} = (1-c)\boldsymbol{z}/c$ and $\boldsymbol{x}(0) = \mathbf{1}/\|\mathbf{1}\|$. For each norm ℓ_1, ℓ_2 and ℓ_∞, we compute iteratively the fixed point of the normalized affine iteration. The number of iterations needed to have $\|\boldsymbol{x}(k) - \boldsymbol{x}(k-1)\|_1 < 10^{-10}$ strongly depends on the chosen norm. Few iterations are needed for the ℓ_1 norm in comparison with the cases of the ℓ_2 and ℓ_∞ norms, as shown in Figure 3.5. At the equilibrium, we know that the fixed point $\boldsymbol{x}_*$ satisfies

$$\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_* = (A + \boldsymbol{b}\boldsymbol{c}_*^T)\boldsymbol{x}_*,$$

where $\boldsymbol{c}_*$ is a dual vector of $\boldsymbol{x}_*$. In order to estimate the speed of convergence near the fixed point, we compute the ratio $|\lambda_2(A + \boldsymbol{b}\boldsymbol{c}_*^T)|/\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$. The spectral radius $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T)$ at equilibrium can simply be computed as $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) = \|A\boldsymbol{x}_* + \boldsymbol{b}\|$. We compute the magnitude of the subdominant eigenvalue, $|\lambda_2(A + \boldsymbol{b}\boldsymbol{c}_*^T)|$, with the `eig` function of Matlab. As expected, the ratio $|\lambda_2/\lambda_1|$ is much better for the ℓ_1 norm than for the ℓ_2 and ℓ_∞ norms. The results are given in Figure 3.5.

3.8 Conclusions

In this chapter, we have analyzed the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1,$$

for a nonnegative matrix A, a nonnegative vector $\boldsymbol{b}$, and a monotone norm $\|\cdot\|$.

In Theorem 3.1, we have characterized the pairs A and $\boldsymbol{b}$ such that the conditional affine eigenvalue problem has a unique solution, possibly with $\lambda > \rho(A)$. This characterization uses graph-theoretic properties as well as maximizing properties of the spectral radius. We also gave an example showing that the characterization with such properties could not be finer, unless imposing more assumptions on the norm. For the case where $\|\cdot\|$ is strictly monotone, we characterize in Theorem 3.43 the pairs of matrices A and $\boldsymbol{b}$ leading to a unique solution.

One of the main results of this chapter, stated in Theorem 3.19, is that the solution $(\lambda_*, \boldsymbol{x}_*)$ of the conditional affine eigenvalue problem

$$\lambda \boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{b}, \quad \lambda \in \mathbb{R}, \quad \boldsymbol{x} \geq 0, \quad \|\boldsymbol{x}\| = 1,$$

can be expressed as the spectral radius and normalized Perron vector of a matrix $A + \boldsymbol{b}\boldsymbol{c}_*^T$, where $\boldsymbol{c}_*$ is a maximizer of the spectral radius $\rho(A + \boldsymbol{b}\boldsymbol{c}^T)$ among all $\boldsymbol{c} \geq 0$ such that $\|\boldsymbol{c}\|^\star = 1$. The assumption required is that $\rho(A + \boldsymbol{b}\boldsymbol{c}_*^T) > \rho(A)$.

We noticed in Section 2.2.3 that, if $A \in \mathbb{R}^{n\times n}_{\geq 0}$ and $\boldsymbol{b} \in \mathbb{R}^n_{\geq 0}$ are such that $A\boldsymbol{x} + b > 0$ for all $\boldsymbol{x} \gneqq 0$, the convergence of the normalized affine iteration

$$\boldsymbol{x}(k+1) = \frac{A\boldsymbol{x}(k) + \boldsymbol{b}}{\|A\boldsymbol{x}(k) + \boldsymbol{b}\|},$$

is a particular case of Krause's Theorem 2.7 about concave maps.

We have seen in this chapter that this assumption is not necessary in order to ensure the uniqueness of the fixed point of this iteration. It could maybe be interesting to see if this assumption could be weakened to some of the conditions of Theorem 3.1. For instance, for the particular case of a norm defined by $\|\boldsymbol{x}\| = \boldsymbol{d}^T|\boldsymbol{x}|$ for some given positive vector $\boldsymbol{d}$, it can easily be proved that the normalized affine iteration converges to its unique fixed point if $\rho(A + \boldsymbol{b}\boldsymbol{d}^T) > \rho(A)$, since $(A + \boldsymbol{b}\boldsymbol{d}^T)_{\mathcal{B}}$ is primitive. That is, condition (b) of Theorem 3.1 is sufficient in this case to have global convergence on the nonnegative orthant. We wonder if accessibility properties could enable to weaken this condition to the conditions (c), (D) and (e) of Theorem 3.1, which we proved equivalent in this case.

Chapter 4

Maximizing PageRank via outlinks

In this chapter we see how a web master can design the link structure of a web site in order to obtain a PageRank as large as possible. The web master can only choose the hyperlinks *starting* from her pages but has no control on the hyperlinks from other web pages. We provide an optimal linkage strategy under some reasonable assumptions.

4.1 Introduction

Google is probably the most popular web search engine at this time. It is therefore not surprising that some web masters want to increase the PageRank of their web pages in order to get more visits from web surfers. Since PageRank is based on the link structure of the Web, it is thus useful to understand how addition or deletion of hyperlinks influence it.

Mathematical analysis of PageRank's sensitivity with respect to perturbations of the matrix describing the web graph is a topical subject of interest (see for instance [9, 16, 75, 83, 84, 86] and the references therein). Normwise and componentwise conditioning bounds [75] as well as the derivative [83, 84] are used to measure the sensitivity of the PageRank vector. It appears that the PageRank vector is relatively insensitive to small changes in the graph structure, at least when these changes concern web pages with a low PageRank score [16, 83]. One could therefore think that trying to modify its PageRank via changes in the link structure of the Web is a waste of time. However, what is important for web masters is not the values of the PageRank vector but the *ranking* that results from it. Lempel and Moran [86] show that PageRank is not rank-stable, i.e., small modifications in the link structure of the web graph may cause dramatic changes in the ranking of the web pages. Therefore, the question of how the PageRank of a particular page or set of pages can be increased–even slightly–by adding or removing links to the web graph remains of interest.

If a hyperlink from a page i to a page j is added, without no other modification in the Web, then the PageRank of j will increase [8, 64]. But in general, a web master does not have control on the *inlinks* of her web pages unless she pays another web master to add a hyperlink from his page to her or she makes an *alliance* with him by trading a link for a link [10, 53]. But it is natural to ask how a web master could modify her PageRank by herself. This leads to analyze how the choice of the *outlinks* of a page can influence its own PageRank. Sydow [113] showed via numerical simulations that adding well chosen outlinks to a web page may increase significantly its PageRank ranking. Avrachenkov and Litvak [9] analyzed theoretically the possible effect of new outlinks on the PageRank of a page and its neighbors. Supposing that a web page has control only on its outlinks, they gave the optimal linkage strategy for this single page. Bianchini et al. [16] as well as Avrachenkov and Litvak in [8] consider the impact of links between web communities (web sites or sets of related web pages), respectively on the sum of the PageRanks and on the individual PageRank scores of the pages of some community. They give general rules in order to have a PageRank as high as possible but they do not provide an optimal link structure for a web site.

Our aim in this chapter is to find a *generalization* of Avrachenkov–Litvak's optimal linkage strategy [9] *to the case of a web site with several pages*. We consider a given set of pages and suppose we have control on the *outlinks* of these pages. We are interested in the problem of *maximizing the sum of the PageRanks* of these pages.

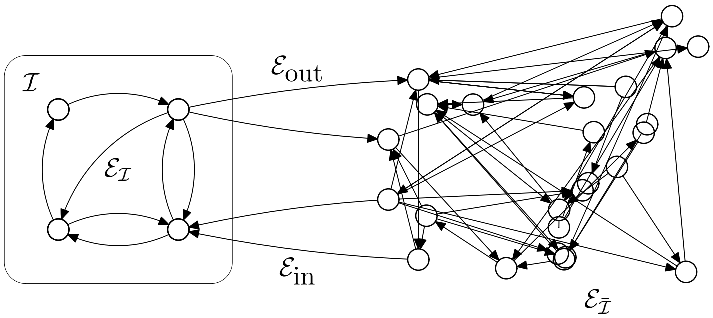


Figure 4.1: Sets of internal links, external links and external inlinks and outlinks.

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be the web graph, with a set of nodes $\mathcal{N} = \{1, \ldots, n\}$ and a set of links $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. As in Section 2.2.1, $(i, j) \in \mathcal{E}$ if and only if there exists a hyperlink linking page i to page j. For a subset of nodes $\mathcal{I} \subseteq \mathcal{N}$, we define

$$\begin{aligned}
\mathcal{E}_{\mathcal{I}} &= \{(i,j) \in \mathcal{E} \colon i, j \in \mathcal{I}\} \text{ the set of internal links,} \\
\mathcal{E}_{\text{out}(\mathcal{I})} &= \{(i,j) \in \mathcal{E} \colon i \in \mathcal{I}, j \notin \mathcal{I}\} \text{ the set of external outlinks,} \\
\mathcal{E}_{\text{in}(\mathcal{I})} &= \{(i,j) \in \mathcal{E} \colon i \notin \mathcal{I}, j \in \mathcal{I}\} \text{ the set of external inlinks,} \\
\mathcal{E}_{\bar{\mathcal{I}}} &= \{(i,j) \in \mathcal{E} \colon i, j \notin \mathcal{I}\} \text{ the set of external links.}
\end{aligned}$$

If we do not impose any condition on $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$, the problem of maximizing the sum of the PageRanks of pages of $\mathcal{I}$, i.e., $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$, is quite trivial and does not have much interest (see the discussion in Section 4.3). Therefore, when characterizing optimal link structures, we make the following *accessibility assumption*: every page of the web site must have an access to the rest of the Web.

Our first main result concerns the *optimal external outlink structure* for a given web site. In the case where the subgraph corresponding to the web site is strongly connected, Theorem 4.17 can be specialized as follows.

Theorem 4.1. *Let $\mathcal{E}_{\mathcal{I}}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose that the subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$ is strongly connected and $\mathcal{E}_{\mathcal{I}} \neq \emptyset$. Then every optimal outlink structure $\mathcal{E}_{\text{out}(\mathcal{I})}$ consists of a unique outlink to a particular page outside of $\mathcal{I}$.*

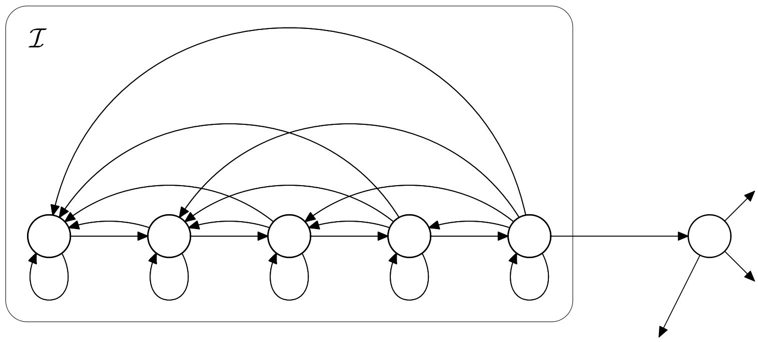


Figure 4.2: Every optimal linkage strategy for a set $\mathcal{I}$ of five pages must have this structure.

We are also interested in the optimal *internal* link structure for a web site. In the case where there is a unique leaking node in the web site, that is, only one node linking to the rest of the web, Theorem 4.19 can be specialized as follows.

Theorem 4.2. *Let $\mathcal{E}_{\text{out}(\mathcal{I})}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose that there is only one leaking node in $\mathcal{I}$. Then every optimal internal link structure $\mathcal{E}_{\mathcal{I}}$ consists of a forward chain of links together with every possible backward link.*

Putting together Theorems 4.17 and 4.19, we get in Theorem 4.21 the *optimal link structure* for a web site. This optimal structure is illustrated in Figure 4.2.

Theorem 4.3. *Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Then, for every optimal link structure, $\mathcal{E}_{\mathcal{I}}$ consists of a forward chain of links together with every possible backward link, and $\mathcal{E}_{\text{out}(\mathcal{I})}$ consists of a unique outlink, starting from the last node of the chain.*

Before going ahead in this chapter, let us remind that the algorithm actually used by Google is not publicized. We only know that Google's team claims they use PageRank in one way or another. One has therefore to limit oneself to a mathematical study of PageRank, as it is defined in [103]. In this chapter, we characterize the optimal structures for the maximization of the PageRank but we do not pretend to provide a practical way to improve one's ranking on Google.

This chapter is organized as follows. We first develop tools for analyzing the PageRank of a set of pages $\mathcal{I}$ in Section 4.2. Then in Section 4.3 we provide the *optimal linkage strategy for a set of nodes*. In Section 4.4, we give some extensions and variants of the main theorems. Finally, in Section 4.5,we present our experiments on a subgraph of the Web.

4.2 PageRank of a web site

We have seen in Section 2.2.1 that the PageRank vector $\boldsymbol{\pi}$ is defined by

$$\begin{aligned} \boldsymbol{\pi}^T &= \boldsymbol{\pi}^T G, \\ \boldsymbol{\pi}^T \mathbf{1} &= 1, \end{aligned} \tag{4.1}$$

where $G = cP + (1-c)\mathbf{1}\boldsymbol{z}^T$ is the Google matrix and $P = [P_{ij}]_{i,j\in\mathcal{N}}$ is a stochastic scaling of the adjacency matrix of $\mathcal{G}$ defined by $P_{ij} = \boldsymbol{d}_i^{-1}$ if $(i,j) \in \mathcal{E}$ and $P_{ij} = 0$ otherwise. Remember that it is supposed that *each node has at least one outlink*, i.e., the outdegree $\boldsymbol{d}_i \neq 0$ for every $i \in \mathcal{N}$.

We are interested in characterizing the *PageRank of a set* $\mathcal{I}$. We define this as the sum

$$\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = \sum_{i\in\mathcal{I}} \boldsymbol{\pi}_i,$$

where $\boldsymbol{e}_{\mathcal{I}}$ denotes the vector with a 1 in the entries of $\mathcal{I}$ and 0 elsewhere. Note that the PageRank of a set corresponds to the notion of energy of a community in [16].

Let $\mathcal{I} \subseteq \mathcal{N}$ be a subset of the nodes of the graph. The PageRank of $\mathcal{I}$ can be expressed as $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = (1-c)\boldsymbol{z}^T(I-cP)^{-1}\boldsymbol{e}_{\mathcal{I}}$ from PageRank equations (4.1). Let us then define the vector

$$\boldsymbol{v} = (I - cP)^{-1}\boldsymbol{e}_{\mathcal{I}}. \tag{4.2}$$

With this, we have the following expression for the PageRank of the set $\mathcal{I}$:

$$\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = (1-c)\boldsymbol{z}^T \boldsymbol{v}. \tag{4.3}$$

The vector $\boldsymbol{v}$ plays a crucial role throughout this chapter. In this section, we first present a probabilistic interpretation for this vector and prove some of its properties. We then show how it can be used in order to analyze the influence of some page $i \in \mathcal{I}$ on the PageRank of the set $\mathcal{I}$. We end this section by briefly introducing the concept of basic absorbing graph, which is useful in order to analyze optimal linkage strategies.

4.2.1 Mean number of visits before zapping

Let us first see how the entries of the vector $\boldsymbol{v} = (I - cP)^{-1}\boldsymbol{e}_{\mathcal{I}}$ can be interpreted. Consider a random surfer on the web graph $\mathcal{G}$ that, as described in Section 2.2.1, follows the hyperlinks of the web graph with a probability c. Assume now that, instead of zapping to some page of $\mathcal{G}$ with a probability $(1-c)$, he *stops* his walk with probability $(1-c)$ at each step of time. This is equivalent to consider a random walk on the extended graph $\mathcal{G}_e = (\mathcal{N} \cup \{n+1\}, \mathcal{E} \cup \{(i, n+1) \colon i \in \mathcal{N}\})$ with a transition probability matrix

$$P_e = \begin{pmatrix} cP & (1-c)\mathbf{1} \\ 0 & 1 \end{pmatrix}.$$

At each step of time, with probability $1-c$, the random surfer can *disappear* from the original graph, that is, he can reach the absorbing node $n+1$.

The nonnegative matrix $(I - cP)^{-1}$ is commonly called the fundamental matrix of the absorbing Markov chain defined by P_e (see for instance [74, 110]). In the extended graph $\mathcal{G}_e$, the entry $[(I-cP)^{-1}]_{ij}$ is the expected number of visits to node j before reaching the absorbing node $n+1$ when starting from node i. From the point of view of the standard random surfer described in Section 2.2.1, the entry $[(I-cP)^{-1}]_{ij}$ is the expected number of visits to node j before zapping for the first time when starting from node i.

Therefore, the vector $\boldsymbol{v}$ defined in equation (4.2) has the following probabilistic interpretation. The entry $\boldsymbol{v}_i$ is the *expected number of visits to the set $\mathcal{I}$ before zapping* for the first time when the random surfer starts his walk in node i.

Let us first prove some simple properties about this vector.

Lemma 4.4. *Let $\boldsymbol{v} \in \mathbb{R}^n_{\geq 0}$ be defined by $\boldsymbol{v} = cP\boldsymbol{v} + \boldsymbol{e}_{\mathcal{I}}$. Then,*

(a) $\max_{i \notin \mathcal{I}} \boldsymbol{v}_i \le c \max_{i \in \mathcal{I}} \boldsymbol{v}_i$,

(b) $\boldsymbol{v}_i \le 1 + c\,\boldsymbol{v}_i$ *for all* $i \in \mathcal{N}$; *with equality if and only if the node* i *does not have an access to* $\bar{\mathcal{I}}$,

(c) $\boldsymbol{v}_i \ge \min_{j \leftarrow i} \boldsymbol{v}_j$ *for all* $i \in \mathcal{I}$; *with equality if and only if the node* i *does not have an access to* $\bar{\mathcal{I}}$;

Proof. (a) Since $c < 1$, for all $i \notin \mathcal{I}$,

$$\max_{i \notin \mathcal{I}} \boldsymbol{v}_i = \max_{i \notin \mathcal{I}} \left(c \sum_{j \leftarrow i} \frac{\boldsymbol{v}_j}{\boldsymbol{d}_i} \right) \le c \max_j \boldsymbol{v}_j.$$

Since $c < 1$, it then follows that $\max_j \boldsymbol{v}_j = \max_{i \in \mathcal{I}} \boldsymbol{v}_i$.

(b) The inequality $\boldsymbol{v}_i \le \frac{1}{1-c}$ follows directly from

$$\max_i \boldsymbol{v}_i \le \max_i \left(1 + c \sum_{j \leftarrow i} \frac{\boldsymbol{v}_j}{\boldsymbol{d}_i} \right) \le 1 + c \max_j \boldsymbol{v}_j.$$

From (a) it then also follows that $\boldsymbol{v}_i \le \frac{c}{1-c}$ for all $i \notin \mathcal{I}$. Now, let $i \in \mathcal{N}$ such that $\boldsymbol{v}_i = \frac{1}{1-c}$. Then $i \in \mathcal{I}$. Moreover,

$$1 + c\,\boldsymbol{v}_i = \boldsymbol{v}_i = 1 + c \sum_{j \leftarrow i} \frac{\boldsymbol{v}_j}{\boldsymbol{d}_i},$$

that is, $\boldsymbol{v}_j = \frac{1}{1-c}$ for every $j \leftarrow i$. Hence node j must also belong to $\mathcal{I}$. By induction, every node k such that i has an access to k must belong to $\mathcal{I}$.

(c) Let $i \in \mathcal{I}$. Then, by (b)

$$1 + c\,\boldsymbol{v}_i \ge \boldsymbol{v}_i = 1 + c \sum_{j \leftarrow i} \frac{\boldsymbol{v}_j}{\boldsymbol{d}_i} \ge 1 + c \min_{j \leftarrow i} \boldsymbol{v}_j,$$

so $\boldsymbol{v}_i \ge \min_{j \leftarrow i} \boldsymbol{v}_j$ for all $i \in \mathcal{I}$. If $\boldsymbol{v}_i = \min_{j \leftarrow i} \boldsymbol{v}_j$ then also $1 + c\,\boldsymbol{v}_i = \boldsymbol{v}_i$ and hence, by (b), the node i does not have an access to $\bar{\mathcal{I}}$. $\square$

Let us denote the set of nodes of $\bar{\mathcal{I}}$ which on average give the most visits to $\mathcal{I}$ before zapping by

$$\mathcal{V} = \operatorname*{argmax}_{j \in \bar{\mathcal{I}}} \boldsymbol{v}_j.$$

Then the following lemma is quite intuitive. It says that, among the nodes of $\bar{\mathcal{I}}$, those that provide the highest mean number of visits to $\mathcal{I}$ are parents of $\mathcal{I}$, i.e., parents of some node of $\mathcal{I}$.

Lemma 4.5 (Parents of $\mathcal{I}$). *If $\mathcal{E}_{\mathrm{in}(\mathcal{I})} \neq \emptyset$, then*

$$\mathcal{V} \subseteq \{j \in \bar{\mathcal{I}} \colon \text{ there exists } \ell \in \mathcal{I} \text{ such that } (j, \ell) \in \mathcal{E}_{\mathrm{in}(\mathcal{I})}\}.$$

If $\mathcal{E}_{\mathrm{in}(\mathcal{I})} = \emptyset$, then $\boldsymbol{v}_j = 0$ for every $j \in \bar{\mathcal{I}}$.

Proof. Suppose first that $\mathcal{E}_{\mathrm{in}(\mathcal{I})} \neq \emptyset$. Let $k \in \mathcal{V}$ with $\boldsymbol{v} = (I - cP)^{-1}\boldsymbol{e}_{\mathcal{I}}$. If we supposed that there does not exist $\ell \in \mathcal{I}$ such that $(k, \ell) \in \mathcal{E}_{\mathrm{in}(\mathcal{I})}$, then we would have, since $\boldsymbol{v}_k > 0$,

$$\boldsymbol{v}_k = c \sum_{j \leftarrow k} \frac{\boldsymbol{v}_j}{\boldsymbol{d}_k} \leq c \max_{j \notin \mathcal{I}} \boldsymbol{v}_j = c\boldsymbol{v}_k < \boldsymbol{v}_k,$$

which is a contradiction. Now, if $\mathcal{E}_{\mathrm{in}(\mathcal{I})} = \emptyset$, then there is no access to $\mathcal{I}$ from $\bar{\mathcal{I}}$, so clearly $\boldsymbol{v}_j = 0$ for every $j \in \bar{\mathcal{I}}$. $\square$

Lemma 4.5 shows that the nodes $j \in \bar{\mathcal{I}}$ that provide the highest value of $\boldsymbol{v}_j$ must belong to the set of parents of $\mathcal{I}$. The converse is not true, as we see in the following example: some parents of $\mathcal{I}$ can provide a lower mean number of visits to $\mathcal{I}$ than other nodes which are not parents of $\mathcal{I}$. In other words, Lemma 4.5 gives a necessary but not sufficient condition in order to maximize the entry $\boldsymbol{v}_j$ for some $j \in \bar{\mathcal{I}}$.

Example 4.6. Let us see with an example that having $(j, i) \in \mathcal{E}_{\mathrm{in}(\mathcal{I})}$ for some $i \in \mathcal{I}$ is not sufficient to have $j \in \mathcal{V}$. Consider the graph in Figure 4.3. Let $\mathcal{I} = \{1\}$ and take a damping factor $c = 0.85$. For $\boldsymbol{v} = (I - cP)^{-1}\boldsymbol{e}_1$, we have

$$\boldsymbol{v}_2 = \boldsymbol{v}_3 = \boldsymbol{v}_4 = 4.359 > \boldsymbol{v}_5 = 3.521 > \boldsymbol{v}_6 = 3.492 > \boldsymbol{v}_7 > \cdots > \boldsymbol{v}_{11},$$

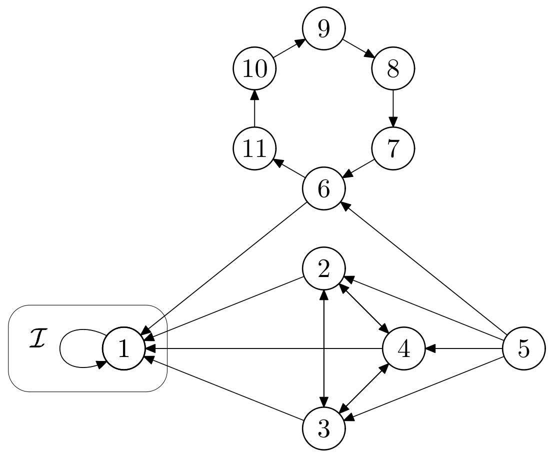


Figure 4.3: The node $6 \notin \mathcal{V}$ and yet it is a parent of $\mathcal{I} = \{1\}$ (see Example 4.6).

so $\mathcal{V} = \{2, 3, 4\}$. As ensured by Lemma 4.5, every node of the set $\mathcal{V}$ is a parent of node 1. But here, $\mathcal{V}$ does not contain all parents of node 1. Indeed, the node $6 \notin \mathcal{V}$ while it is a parent of 1 and is moreover its parent with the lowest outdegree. Moreover, we see in this example that node 5, which is a not a parent of node 1 but a parent of node 6, gives a higher value of the expected number of visits to $\mathcal{I}$ before zapping, than node 6, parent of 1. Let us try to get some intuition about that. When starting from node 6, a random surfer has probability 0.5 to reach node 1 in only one step. But he has also a probability 0.5 to move to node 11 and to be send far away from node 1. On the other hand, when starting from node 5, the random surfer can not reach node 1 in only one step. But with probability 0.75 he will reach one of the nodes 2, 3 or 4 in one step. And from these nodes, the web surfer stays very near to node 1 and can not be sent far away from it. $\diamond$

In the next lemma, we show that from some node $i \in \mathcal{I}$ which has an access to $\bar{\mathcal{I}}$, there always exists what we call a *decreasing path* to $\bar{\mathcal{I}}$. That is, we can find a path such that the mean number of visits to $\mathcal{I}$ is higher when starting from some node of the path than when starting from the successor of this node in the path.

Lemma 4.7 (Decreasing paths to $\bar{\mathcal{I}}$)**.** *For every $i_0 \in \mathcal{I}$ which has an access* *to $\overline{\mathcal{I}}$, there exists a path $\langle i_0, i_1, \dots, i_s\rangle$ with $i_1, \dots, i_{s-1} \in \mathcal{I}$ and $i_s \in \bar{\mathcal{I}}$ such that*

$$v_{i_0} > v_{i_1} > \dots > v_{i_s}.$$

Proof. Let us simply construct a decreasing path recursively by

$$i_{k+1} \in \operatorname*{argmin}_{j \leftarrow i_k} v_j,$$

as long as $i_k \in \mathcal{I}$. If i_k has an access to $\bar{\mathcal{I}}$, then $v_{i_{k+1}} < v_{i_k} < \frac{1}{1-c}$ by Lemma 4.4(b) and (c), so the node i_{k+1} has also an access to $\bar{\mathcal{I}}$. By assumption, i_0 has an access to $\bar{\mathcal{I}}$. Moreover, the set $\mathcal{I}$ has a finite number of elements, so there must exist an s such that $i_s \in \bar{\mathcal{I}}$. □

4.2.2 Influence of the outlinks of a node

We are now interested in how a modification of the outlinks of some node $i \in \mathcal{N}$ can change the PageRank of a subset of nodes $\mathcal{I} \subseteq \mathcal{N}$. So we compare two graphs on $\mathcal{N}$ defined by their set of links, $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ respectively.

Every item corresponding to the graph defined by the set of links $\widetilde{\mathcal{E}}$ will be written with a tilde symbol. So $\widetilde{P}$ denotes its scaled adjacency matrix, $\tilde{\pi}$ the corresponding PageRank vector, $\widetilde{d}_i = |\{j\colon (i,j) \in \widetilde{\mathcal{E}}\}|$ the outdegree of some node i in this graph, $\tilde{v} = (I - c\widetilde{P})^{-1} e_{\mathcal{I}}$ and $\widetilde{\mathcal{V}} = \operatorname{argmax}_{j \in \bar{\mathcal{I}}} \tilde{v}_j$. Finally, by $j \widetilde{\leftarrow} i$ we mean $j \in \{k\colon (i,k) \in \widetilde{\mathcal{E}}\}$.

Let us consider two graphs defined respectively by their set of links $\mathcal{E}$ and $\widetilde{\mathcal{E}}$. Suppose that they differ only in the links starting from some given node i, that is, $\{j\colon (k,j) \in \mathcal{E}\} = \{j\colon (k,j) \in \widetilde{\mathcal{E}}\}$ for all $k \neq i$. Then their scaled adjacency matrices P and $\widetilde{P}$ are linked by a rank-one correction. Let us then define the vector

$$\delta = \sum_{j \widetilde{\leftarrow} i} \frac{e_j}{\widetilde{d}_i} - \sum_{j \leftarrow i} \frac{e_j}{d_i},$$

which gives the correction to apply to the line i of the matrix P in order to get $\widetilde{P}$.

Now let us first express the difference between the PageRank of $\mathcal{I}$ for two configurations differing only in the links starting from some node i. Note that in the following lemma the personalization vector $\boldsymbol{z}$ does not appear explicitly in the expression of $\widetilde{\boldsymbol{\pi}}$.

Lemma 4.8. *Let two graphs defined respectively by $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ and let $i \in \mathcal{N}$ such that for all $k \neq i$, $\{j\colon (k,j) \in \mathcal{E}\} = \{j\colon (k,j) \in \widetilde{\mathcal{E}}\}$. Then*

$$\widetilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} = \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} + c\, \boldsymbol{\pi}_i \,\frac{\boldsymbol{\delta}^T \boldsymbol{v}}{1 - c\,\boldsymbol{\delta}^T (I - cP)^{-1} \boldsymbol{e}_i}.$$

Proof. Clearly, the scaled adjacency matrices are linked by $\widetilde{P} = P + \boldsymbol{e}_i\, \boldsymbol{\delta}^T$. Since $c < 1$, the matrix $(I - cP)^{-1}$ exists and the PageRank vectors can be expressed as

$$\begin{aligned} \boldsymbol{\pi}^T &= (1-c)\boldsymbol{z}^T (I - cP)^{-1}, \\ \widetilde{\boldsymbol{\pi}}^T &= (1-c)\boldsymbol{z}^T (I - c\,(P + \boldsymbol{e}_i \boldsymbol{\delta}^T))^{-1}. \end{aligned}$$

Applying the Sherman–Morrison formula to $((I - cP) - c\boldsymbol{e}_i \boldsymbol{\delta}^T)^{-1}$, we get

$$\widetilde{\boldsymbol{\pi}}^T = (1-c)\boldsymbol{z}^T (I - cP)^{-1} + (1-c)\boldsymbol{z}^T (I - cP)^{-1} \boldsymbol{e}_i \,\frac{c\boldsymbol{\delta}^T (I - cP)^{-1}}{1 - c\boldsymbol{\delta}^T (I - cP)^{-1} \boldsymbol{e}_i},$$

and the result follows immediately. □

Let us now give an equivalent condition in order to increase the PageRank of $\mathcal{I}$ by changing outlinks of some node i. The PageRank of $\mathcal{I}$ increases when the new set of links favors nodes giving a higher mean number of visits to $\mathcal{I}$ before zapping.

Theorem 4.9 (PageRank and mean number of visits before zapping)**.** *Let two graphs defined respectively by $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ and let $i \in \mathcal{N}$ such that for all $k \neq i$, $\{j\colon (k,j) \in \mathcal{E}\} = \{j\colon (k,j) \in \widetilde{\mathcal{E}}\}$. Then*

$$\widetilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} > \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} \quad \text{if and only if} \quad \boldsymbol{\delta}^T \boldsymbol{v} > 0$$

and $\widetilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} = \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ if and only if $\boldsymbol{\delta}^T \boldsymbol{v} = 0$.

Proof. Let us first show that $\delta^T(I - cP)^{-1}\boldsymbol{e}_i \leq 1$ is always verified. Let $\boldsymbol{u} = (I - cP)^{-1}\boldsymbol{e}_i$. Then $\boldsymbol{u} - cP\boldsymbol{u} = \boldsymbol{e}_i$ and, by Lemma 4.4(a), $\boldsymbol{u}_j \leq \boldsymbol{u}_i$ for all j. So

$$\delta^T \boldsymbol{u} = \sum_{j \widetilde{\leftarrow} i} \frac{\boldsymbol{u}_j}{\widetilde{\boldsymbol{d}}_i} - \sum_{j \leftarrow i} \frac{\boldsymbol{u}_j}{\boldsymbol{d}_i} \leq \boldsymbol{u}_i - \sum_{j \leftarrow i} \frac{\boldsymbol{u}_j}{\boldsymbol{d}_i} \leq \boldsymbol{u}_i - c \sum_{j \leftarrow i} \frac{\boldsymbol{u}_j}{\boldsymbol{d}_i} = 1.$$

Now, since $c < 1$ and $\boldsymbol{\pi} > 0$, the conclusion follows by Lemma 4.8. $\square$

The next proposition shows how to add a new link (i, j) starting from a given node i in order to increase the PageRank of the set $\mathcal{I}$. The PageRank of $\mathcal{I}$ increases as soon as a node $i \in \mathcal{I}$ adds a link to a node j with a larger or equal expected number of visits to $\mathcal{I}$ before zapping.

Proposition 4.10 (Adding a link)**.** *Let $i \in \mathcal{I}$ and let $j \in \mathcal{N}$ be such that $(i, j) \notin \mathcal{E}$ and $\boldsymbol{v}_i \leq \boldsymbol{v}_j$. Let $\widetilde{\mathcal{E}} = \mathcal{E} \cup \{(i, j)\}$. Then*

$$\widetilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} \geq \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$$

with equality if and only if the node i does not have an access to $\bar{\mathcal{I}}$.

Proof. Let $i \in \mathcal{I}$ and let $j \in \mathcal{N}$ be such that $(i, j) \notin \mathcal{E}$ and $\boldsymbol{v}_i \leq \boldsymbol{v}_j$. Then

$$1 + c \sum_{k \leftarrow i} \frac{\boldsymbol{v}_k}{\boldsymbol{d}_i} = \boldsymbol{v}_i \leq 1 + c\boldsymbol{v}_i \leq 1 + c\boldsymbol{v}_j,$$

with equality if and only if i does not have an access to $\bar{\mathcal{I}}$ by Lemma 4.4(b). Let $\widetilde{\mathcal{E}} = \mathcal{E} \cup \{(i, j)\}$. Then

$$\delta^T \boldsymbol{v} = \frac{1}{\boldsymbol{d}_i + 1} \left(\boldsymbol{v}_j - \sum_{k \leftarrow i} \frac{\boldsymbol{v}_k}{\boldsymbol{d}_i} \right) \geq 0,$$

with equality if and only if i does not have an access to $\bar{\mathcal{I}}$. The conclusion follows from Theorem 4.9. $\square$

Now let us see how to remove a link (i, j) starting from a given node i in order to increase the PageRank of the set $\mathcal{I}$. If a node $i \in \mathcal{N}$ removes a link to its worst child from the point of view of the expected number of visits to $\mathcal{I}$ before zapping, then the PageRank of $\mathcal{I}$ increases.

Proposition 4.11 (Removing a link). *Let $i \in \mathcal{N}$ and let $j \in \operatorname{argmin}_{k \leftarrow i} v_k$. Let $\widetilde{\mathcal{E}} = \mathcal{E} \setminus \{(i,j)\}$. Then*

$$\widetilde{\pi}^T e_{\mathcal{I}} \geq \pi^T e_{\mathcal{I}}$$

with equality if and only if $v_k = v_j$ for every k such that $(i,k) \in \mathcal{E}$.

Proof. Let $i \in \mathcal{N}$ and let $j \in \operatorname{argmin}_{k \leftarrow i} v_k$. Let $\widetilde{\mathcal{E}} = \mathcal{E} \setminus \{(i,j)\}$. Then

$$\delta^T v = \sum_{k \leftarrow i} \frac{v_k - v_j}{d_i(d_i - 1)} \geq 0,$$

with equality if and only if $v_k = v_j$ for all $k \leftarrow i$. The conclusion follows by Theorem 4.9. $\square$

In order to increase the PageRank of $\mathcal{I}$ with a new link (i,j), Proposition 4.10 only requires that $v_j \leq v_i$. On the other hand, Proposition 4.11 requires that $v_j = \min_{k \leftarrow i} v_k$ in order to increase the PageRank of $\mathcal{I}$ by deleting link (i,j). One could wonder whether this condition could be weakened to $v_j < v_i$, so as to have symmetric conditions for the addition or deletion of links. In fact, this can not be done as shown in the following example.

Example 4.12. Let us see that the condition $j \in \operatorname{argmin}_{k \leftarrow i} v_k$ in Proposition 4.11 can not be weakened to $v_j < v_i$. Consider the graph in Figure 4.4 and take a damping factor $c = 0.85$. Let $\mathcal{I} = \{1,2,3\}$. We have

$$v_1 = 2.63 > v_2 = 2.303 > v_3 = 1.533.$$

As ensured by Proposition 4.11, if we remove the link $(1,3)$, the PageRank of $\mathcal{I}$ increases (from 0.199 to 0.22 with a uniform personalization vector $z = \frac{1}{n}\mathbf{1}$), since $3 \in \operatorname{argmin}_{k \leftarrow 1} v_k$. But, if we remove instead the link $(1,2)$, the PageRank of $\mathcal{I}$ decreases (from 0.199 to 0.179 with z uniform) even if $v_2 < v_1$. $\diamond$

Remark 4.13. Let us note that, if the node i does not have an access to the set $\bar{\mathcal{I}}$, then for every *deletion* of a link starting from i, the PageRank of $\mathcal{I}$ will not be modified. Indeed, in this case $\delta^T v = 0$ since by Lemma 4.4(b), $v_j = \frac{1}{1-c}$ for every $j \leftarrow i$. $\diamond$

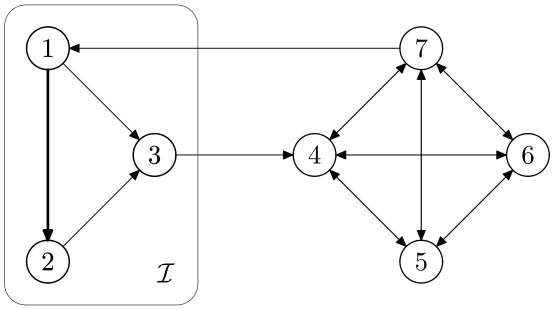


Figure 4.4: For $\mathcal{I} = \{1, 2, 3\}$, removing link $(1, 2)$ gives $\tilde{\pi}^T e_{\mathcal{I}} < \pi^T e_{\mathcal{I}}$, even if $v_1 > v_2$ (see Example 4.12).

4.2.3 Basic absorbing graph

Let us introduce briefly the notion of basic absorbing graph (see Kemeny and Snell's book [74, Chapt. III]).

For a given graph $(\mathcal{N}, \mathcal{E})$ and a specified subset of nodes $\mathcal{I} \subseteq \mathcal{N}$, the *basic absorbing graph* is the graph $(\mathcal{N}, \mathcal{E}^0)$ defined by $\mathcal{E}^0_{\text{out}(\mathcal{I})} = \emptyset$, $\mathcal{E}^0_{\mathcal{I}} = \{(i,i) \colon i \in \mathcal{I}\}$, $\mathcal{E}^0_{\text{in}(\mathcal{I})} = \mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}^0_{\bar{\mathcal{I}}} = \mathcal{E}_{\bar{\mathcal{I}}}$. In other words, the basic absorbing graph $(\mathcal{N}, \mathcal{E}^0)$ is a graph constructed from $(\mathcal{N}, \mathcal{E})$, keeping the same sets of external inlinks and external links $\mathcal{E}_{\text{in}(\mathcal{I})}, \mathcal{E}_{\bar{\mathcal{I}}}$, removing the external outlinks $\mathcal{E}_{\text{out}(\mathcal{I})}$ and changing the internal link structure $\mathcal{E}_{\mathcal{I}}$ in order to have only self-links for nodes of $\mathcal{I}$.

As in the previous subsection, every item corresponding to the basic absorbing graph will be written with a zero symbol. For instance, we write π_0 for the PageRank vector corresponding to the basic absorbing graph and $\mathcal{V}_0 = \text{argmax}_{j \in \bar{\mathcal{I}}} [(I - cP_0)^{-1} e_{\mathcal{I}}]_j$.

Proposition 4.14 (PageRank for a basic absorbing graph)**.** *Let a graph defined by a set of links $\mathcal{E}$ and let $\mathcal{I} \subseteq \mathcal{N}$. Then*

$$\pi^T e_{\mathcal{I}} \leq \pi_0^T e_{\mathcal{I}},$$

with equality if and only if $\mathcal{E}_{\text{out}(\mathcal{I})} = \emptyset$.

Proof. Up to a permutation of the indices, equation (4.2) can be written as

$$\begin{pmatrix} I - cP_{\mathcal{I}} & -cP_{\mathrm{out}(\mathcal{I})} \\ -cP_{\mathrm{in}(\mathcal{I})} & I - cP_{\bar{\mathcal{I}}} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_{\mathcal{I}} \\ \boldsymbol{v}_{\bar{\mathcal{I}}} \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix},$$

so we get

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_{\mathcal{I}} \\ c(I - cP_{\bar{\mathcal{I}}})^{-1} P_{\mathrm{in}(\mathcal{I})} \boldsymbol{v}_{\mathcal{I}} \end{pmatrix}. \tag{4.4}$$

By Lemma 4.4(c) and since $(I - cP_{\bar{\mathcal{I}}})^{-1}$ is a nonnegative matrix by Lemma 2.2, we then have

$$\boldsymbol{v} \leq \begin{pmatrix} \frac{1}{1-c} \mathbf{1} \\ \frac{c}{1-c} (I - cP_{\bar{\mathcal{I}}})^{-1} P_{\mathrm{in}(\mathcal{I})} \mathbf{1} \end{pmatrix} = \boldsymbol{v}_0,$$

with equality if and only if no node of $\mathcal{I}$ has an access to $\bar{\mathcal{I}}$, that is, $\mathcal{E}_{\mathrm{out}(\mathcal{I})} = \emptyset$. The conclusion now follows from equation (4.3) and $\boldsymbol{z} > 0$. □

Let us finally prove a nice property of the set $\mathcal{V}$ when $\mathcal{I} = \{i\}$ is a singleton: it is independent of the outlinks of i. In particular, $\mathcal{V}$ can be found from the basic absorbing graph.

Lemma 4.15. *Let a graph defined by a set of links $\mathcal{E}$ and let $\mathcal{I} = \{i\}$ Then there exists an $\alpha \neq 0$ such that $(I - cP)^{-1}\boldsymbol{e}_i = \alpha(I - cP_0)^{-1}\boldsymbol{e}_i$. As a consequence,*

$$\mathcal{V} = \mathcal{V}_0.$$

Proof. Let $\mathcal{I} = \{i\}$. Since $\boldsymbol{v}_{\mathcal{I}} = \boldsymbol{v}_i$ is a scalar, it follows from equation (4.4) that the direction of the vector $\boldsymbol{v}$ does not depend on $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ but only on $\mathcal{E}_{\mathrm{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$. □

4.3 Optimal linkage strategy for a web site

In this section, we consider a set of nodes $\mathcal{I}$. For this set, we want to choose the sets of internal links $\mathcal{E}_{\mathcal{I}} \subseteq \mathcal{I} \times \mathcal{I}$ and external outlinks $\mathcal{E}_{\mathrm{out}(\mathcal{I})} \subseteq \mathcal{I} \times \bar{\mathcal{I}}$ that maximize the total PageRank score of nodes in $\mathcal{I}$, that is, $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$.

Let us first discuss the constraints on $\mathcal{E}$ that we will consider. If we do not impose any condition on $\mathcal{E}$, the problem of maximizing $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is quite trivial. As shown by Proposition 4.14, the web master should take in this case $\mathcal{E}_{\mathrm{out}(\mathcal{I})} = \emptyset$ and $\mathcal{E}_{\mathcal{I}}$ an arbitrary subset of $\mathcal{I} \times \mathcal{I}$ such that each node has at least one outlink. She just tries to lure the random walker to her pages, not allowing him to leave $\mathcal{I}$ except by zapping according to the preference vector. Therefore, it seems sensible to impose that $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ *must be nonempty*.

Now, in order to avoid trivial solutions to our maximization problem, it is not enough to assume that $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ must be nonempty. Indeed, with this single constraint, in order to lose as few as possible visits from the random walker, the web master should take a unique leaking node $k \in \mathcal{I}$ (i.e., $\mathcal{E}_{\mathrm{out}(\mathcal{I})} = \{(k, \ell)\}$ for some $\ell \in \bar{\mathcal{I}}$) and isolate it from the rest of the set $\mathcal{I}$ (i.e., $\{i \in \mathcal{I} \colon (i,k) \in \mathcal{E}_{\mathcal{I}}\} = \emptyset$).

Moreover, it seems reasonable to imagine that Google penalizes (or at least tries to penalize) such behavior in the context of spam alliances [53].

All this discussion leads us to make the following assumption.

Assumption 4.16 (Accessibility)**.** Every node of $\mathcal{I}$ has an access to at least one node of $\bar{\mathcal{I}}$.

Let us explain the basic ideas we use in order to determine an optimal linkage strategy for a set of web pages $\mathcal{I}$. We determine some forbidden patterns for an optimal linkage strategy and deduce the only possible structure an optimal strategy can have. In other words, we assume that we have a configuration which gives an optimal PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$. Then, we prove that if some particular pattern appears in this optimal structure, then we can construct another graph for which the PageRank $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}}$ is strictly larger than $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$.

We firstly determine the shape of an optimal external outlink structure $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$, when the internal link structure $\mathcal{E}_{\mathcal{I}}$ is given, in Theorem 4.17. Then, given the external outlink structure $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$, we determine the possible optimal internal link structure $\mathcal{E}_{\mathcal{I}}$ in Theorem 4.19. Finally, we put both results together in Theorem 4.21 in order to get the general shape of an optimal linkage strategy for a set $\mathcal{I}$ when $\mathcal{E}_{\mathrm{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ are given.

Proofs of this section are illustrated by several figures for which we take the following drawing convention. When nodes are drawn from left to right on the same horizontal line, they are arranged by decreasing value of v_j. Links are represented by continuous arrows and paths by dashed arrows.

4.3.1 Optimal outlink structure

We begin with the determination the optimal *outlink* structure $\mathcal{E}_{\text{out}(\mathcal{I})}$ for the set $\mathcal{I}$, while its internal structure $\mathcal{E}_{\mathcal{I}}$ is given. An example of optimal outlink structure is given after the theorem.

Theorem 4.17 (Optimal outlink structure). *Let $\mathcal{E}_{\mathcal{I}}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{F}_1, \dots, \mathcal{F}_r$ be the final classes of the subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$. Let $\mathcal{E}_{\text{out}(\mathcal{I})}$ be such that the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. Then $\mathcal{E}_{\text{out}(\mathcal{I})}$ has the following structure:*

$$\mathcal{E}_{\text{out}(\mathcal{I})} = \mathcal{E}_{\text{out}(\mathcal{F}_1)} \cup \cdots \cup \mathcal{E}_{\text{out}(\mathcal{F}_r)},$$

where for every $s = 1, \dots, r$,

$$\mathcal{E}_{\text{out}(\mathcal{F}_s)} \subseteq \{(i,j) \colon i \in \operatorname*{argmin}_{k \in \mathcal{F}_s} \boldsymbol{v}_k \text{ and } j \in \mathcal{V}\}.$$

Moreover, for every $s = 1, \dots, r$, if $\mathcal{E}_{\mathcal{F}_s} \neq \emptyset$, then $|\mathcal{E}_{\text{out}(\mathcal{F}_s)}| = 1$.

Proof. Let $\mathcal{E}_{\mathcal{I}}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose $\mathcal{E}_{\text{out}(\mathcal{I})}$ is such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. We will determine the possible leaking nodes of $\mathcal{I}$ by analyzing three different cases.

Firstly, let us consider some node $i \in \mathcal{I}$ such that i does not have children in $\mathcal{I}$, i.e., $\{k \in \mathcal{I} \colon (i,k) \in \mathcal{E}_{\mathcal{I}}\} = \emptyset$. Then clearly we have $\{i\} = \mathcal{F}_s$ for some $s = 1, \dots, r$, with $i \in \operatorname{argmin}_{k \in \mathcal{F}_s} \boldsymbol{v}_k$ and $\mathcal{E}_{\mathcal{F}_s} = \emptyset$. From Assumption 4.16, we have $\mathcal{E}_{\text{out}(\mathcal{F}_s)} \neq \emptyset$, and from Theorem 4.9 and the optimality assumption, we have $\mathcal{E}_{\text{out}(\mathcal{F}_s)} \subseteq \{(i,j) \colon j \in \mathcal{V}\}$ (see Figure 4.5).

Secondly, let us consider some $i \in \mathcal{I}$ such that i has children in $\mathcal{I}$, i.e., $\{k \in \mathcal{I} \colon (i,k) \in \mathcal{E}_{\mathcal{I}}\} \neq \emptyset$, and

$$\boldsymbol{v}_i \leq \min_{\substack{k \leftarrow i \\ k \in \mathcal{I}}} \boldsymbol{v}_k.$$

Let $j \in \operatorname{argmin}_{k \leftarrow i} \boldsymbol{v}_k$. Then $j \in \bar{\mathcal{I}}$ and $\boldsymbol{v}_j < \boldsymbol{v}_i$ by Lemma 4.4(c). Suppose by contradiction that the node i would keep an access to $\bar{\mathcal{I}}$ if we took

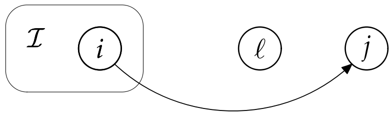


Figure 4.5: If $v_j < v_\ell$, then $\tilde{\pi}^T e_{\mathcal{I}} > \pi^T e_{\mathcal{I}}$ with $\widetilde{\mathcal{E}}_{\text{out}(\mathcal{I})} = \mathcal{E}_{\text{out}(\mathcal{I})} \cup \{(i,\ell)\} \setminus \{(i,j)\}$.

$\widetilde{\mathcal{E}}_{\text{out}(\mathcal{I})} = \mathcal{E}_{\text{out}(\mathcal{I})} \setminus \{(i,j)\}$ instead of $\mathcal{E}_{\text{out}(\mathcal{I})}$. Then, by Proposition 4.11, considering $\widetilde{\mathcal{E}}_{\text{out}(\mathcal{I})}$ instead of $\mathcal{E}_{\text{out}(\mathcal{I})}$ would increase strictly the PageRank of $\mathcal{I}$ while Assumption 4.16 remains satisfied (see Figure 4.6). This

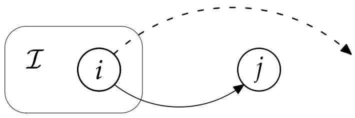


Figure 4.6: If $v_j = \min_{k \leftarrow i} v_k$ and i has another access to $\bar{\mathcal{I}}$, then $\tilde{\pi}^T e_{\mathcal{I}} > \pi^T e_{\mathcal{I}}$ with $\widetilde{\mathcal{E}}_{\text{out}(\mathcal{I})} = \mathcal{E}_{\text{out}(\mathcal{I})} \setminus \{(i,j)\}$.

would contradict the optimality assumption for $\mathcal{E}_{\text{out}(\mathcal{I})}$. From this, we conclude that

- the node i belongs to a final class $\mathcal{F}_s$ of the subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$, with $\mathcal{E}_{\mathcal{F}_s} \neq \emptyset$ for some $s = 1, \ldots, r$;
- there does not exist another $\ell \in \bar{\mathcal{I}}$, $\ell \neq j$ such that $(i,\ell) \in \mathcal{E}_{\text{out}(\mathcal{I})}$;
- there does not exist another k in the same final class $\mathcal{F}_s$, $k \neq i$ such that such that $(k,\ell) \in \mathcal{E}_{\text{out}(\mathcal{I})}$ for some $\ell \in \bar{\mathcal{I}}$.

Again, by Theorem 4.9 and the optimality assumption, we have $j \in \mathcal{V}$. Let us now notice that

$$\max_{k \in \bar{\mathcal{I}}} v_k < \min_{k \in \mathcal{I}} v_k. \tag{4.5}$$

Indeed, with $i \in \operatorname{argmin}_{k \in \mathcal{I}} v_k$, we are in one of the two cases analyzed above for which we have seen that $v_i > v_j = \max_{k \in \bar{\mathcal{I}}} v_k$.

Finally, consider a node $i \in \mathcal{I}$ that does not belong to any of the final classes of the subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$. Suppose by contradiction that there exists $j \in \bar{\mathcal{I}}$ such that $(i,j) \in \mathcal{E}_{\text{out}(\mathcal{I})}$. Let $\ell \in \operatorname{argmin}_{k \leftarrow i} v_k$. Then it follows from inequality (4.5) that $\ell \in \bar{\mathcal{I}}$. But the same argument as

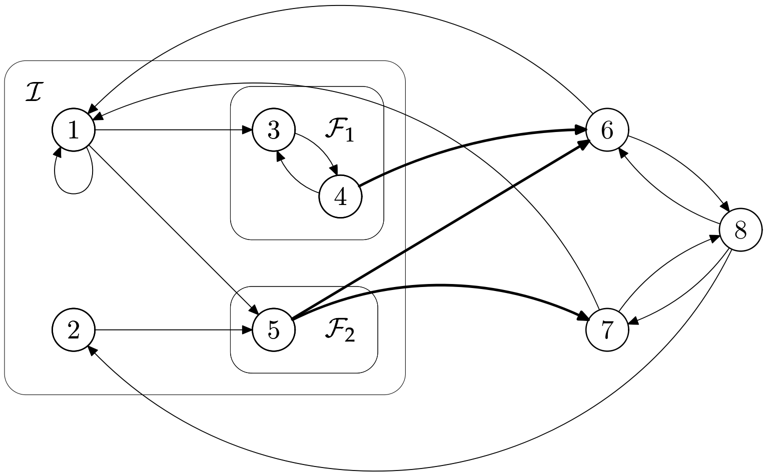


Figure 4.7: Bold arrows represent one of the six optimal outlink *structures for this configuration with two final classes (see Example 4.18).*

above shows that the link $(i, \ell) \in \mathcal{E}_{\text{out}(\mathcal{I})}$ must be removed since $\mathcal{E}_{\text{out}(\mathcal{I})}$ is supposed to be optimal (see Figure 4.6 again). So, there does not exist $j \in \bar{\mathcal{I}}$ such that $(i, j) \in \mathcal{E}_{\text{out}(\mathcal{I})}$ for a node $i \in \mathcal{I}$ which does not belong to any of the final classes $\mathcal{F}_1, \ldots, \mathcal{F}_r$. $\square$

Example 4.18. Consider the graph given in Figure 4.7. The internal link structure $\mathcal{E}_{\mathcal{I}}$, as well as $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ are given. The subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$ has two final classes $\mathcal{F}_1$ and $\mathcal{F}_2$. With $c = 0.85$ and $\boldsymbol{z}$ the uniform probability vector, this configuration has six optimal outlink structures (one of these solutions is represented by bold arrows in Figure 4.7). Each one can be written as $\mathcal{E}_{\text{out}(\mathcal{I})} = \mathcal{E}_{\text{out}(\mathcal{F}_1)} \cup \mathcal{E}_{\text{out}(\mathcal{F}_2)}$, with $\mathcal{E}_{\text{out}(\mathcal{F}_1)} = \{(4, 6)\}$ or $\mathcal{E}_{\text{out}(\mathcal{F}_1)} = \{(4, 7)\}$ and $\emptyset \neq \mathcal{E}_{\text{out}(\mathcal{F}_2)} \subseteq \{(5, 6), (5, 7)\}$. Indeed, since $\mathcal{E}_{\mathcal{F}_1} \neq \emptyset$, as stated by Theorem 4.17, the final class $\mathcal{F}_1$ has exactly one external outlink in every optimal outlink structure. On the other hand, the final class $\mathcal{F}_2$ may have several external outlinks, since it is composed of a unique node and this node does not have a self-link. Note that $\mathcal{V} = \{6, 7\}$ in each of these six optimal configurations, but this set $\mathcal{V}$ can not be determined a priori since it depends on the chosen outlink structure. $\diamond$

4.3.2 Optimal internal link structure

Now we determine the optimal *internal* link structure $\mathcal{E}_{\mathcal{I}}$ for the set $\mathcal{I}$, while its outlink structure $\mathcal{E}_{\text{out}(\mathcal{I})}$ is given. Examples of optimal internal structure are given after the proof of the theorem.

Theorem 4.19 (Optimal internal link structure). *Let $\mathcal{E}_{\text{out}(\mathcal{I})}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{L} = \{i \in \mathcal{I} \colon (i,j) \in \mathcal{E}_{\text{out}(\mathcal{I})}$ for some $j \in \bar{\mathcal{I}}\}$ be the set of leaking nodes of $\mathcal{I}$ and let $n_{\mathcal{L}} = |\mathcal{L}|$ be the number of leaking nodes. Let $\mathcal{E}_{\mathcal{I}}$ be such that the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. Then there exists a permutation of the indices such that $\mathcal{I} = \{1, 2, \dots, n_{\mathcal{I}}\}$, $\mathcal{L} = \{n_{\mathcal{I}} - n_{\mathcal{L}} + 1, \dots, n_{\mathcal{I}}\}$,*

$$\boldsymbol{v}_1 > \cdots > \boldsymbol{v}_{n_{\mathcal{I}} - n_{\mathcal{L}}} > \boldsymbol{v}_{n_{\mathcal{I}} - n_{\mathcal{L}} + 1} \geq \cdots \geq \boldsymbol{v}_{n_{\mathcal{I}}},$$

and $\mathcal{E}_{\mathcal{I}}$ has the following structure:

$$\mathcal{E}_{\mathcal{I}}^L \subseteq \mathcal{E}_{\mathcal{I}} \subseteq \mathcal{E}_{\mathcal{I}}^U,$$

where

$$\begin{aligned} \mathcal{E}_{\mathcal{I}}^L &= \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon j \leq i\} \cup \{(i,j) \in (\mathcal{I} \setminus \mathcal{L}) \times \mathcal{I} \colon j = i+1\}, \\ \mathcal{E}_{\mathcal{I}}^U &= \mathcal{E}_{\mathcal{I}}^L \cup \{(i,j) \in \mathcal{L} \times \mathcal{L} \colon i < j\}. \end{aligned}$$

Proof. Let $\mathcal{E}_{\text{out}(\mathcal{I})}$, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose $\mathcal{E}_{\mathcal{I}}$ is such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16.

Firstly, by Proposition 4.10 and since every node of $\mathcal{I}$ has an access to $\bar{\mathcal{I}}$, every node $i \in \mathcal{I}$ links to every node $j \in \mathcal{I}$ such that $\boldsymbol{v}_j \geq \boldsymbol{v}_i$ (see Figure 4.8), that is

$$\{(i,j) \in \mathcal{E}_{\mathcal{I}} \colon \boldsymbol{v}_i \leq \boldsymbol{v}_j\} = \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon \boldsymbol{v}_i \leq \boldsymbol{v}_j\}. \tag{4.6}$$

Secondly, let $(k,i) \in \mathcal{E}_{\mathcal{I}}$ such that $k \neq i$ and $k \in \mathcal{I} \setminus \mathcal{L}$. Let us prove that, if the node i has an access to $\bar{\mathcal{I}}$ by a path $\langle i, i_1, \dots, i_s \rangle$ such that $i_j \neq k$ for all $j = 1, \dots, s$ and $i_s \in \bar{\mathcal{I}}$, then $\boldsymbol{v}_i < \boldsymbol{v}_k$ (see Figure 4.9). Indeed, if we had $\boldsymbol{v}_k \leq \boldsymbol{v}_i$ then, by Lemma 4.4(c), there would exists $\ell \in \mathcal{I}$ such that $(k,\ell) \in \mathcal{E}_{\mathcal{I}}$ and $\boldsymbol{v}_\ell = \min_{j \leftarrow k} \boldsymbol{v}_j < \boldsymbol{v}_i \leq \boldsymbol{v}_k$. But, with $\widetilde{\mathcal{E}}_{\mathcal{I}} =$

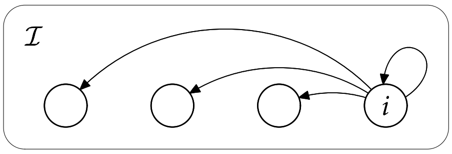


Figure 4.8: Every $i \in \mathcal{I}$ must link to every $j \in \mathcal{I}$ with $\boldsymbol{v}_j \geq \boldsymbol{v}_i$.

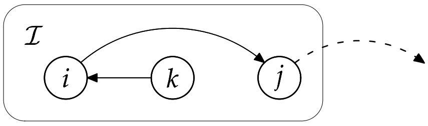


Figure 4.9: The node i can not have an access to $\bar{\mathcal{I}}$ without crossing k since in this case we should then have $\boldsymbol{v}_i < \boldsymbol{v}_k$.

$\mathcal{E}_\mathcal{I} \setminus \{(k,\ell)\}$, we would have $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_\mathcal{I} > \boldsymbol{\pi}^T \boldsymbol{e}_\mathcal{I}$ by Proposition 4.11 while Assumption 4.16 remains satisfied since the node k would keep access to $\bar{\mathcal{I}}$ via the node i (see Figure 4.10). That contradicts the optimality assumption. This leads us to the conclusion that $\boldsymbol{v}_k > \boldsymbol{v}_i$ for every $k \in \mathcal{I} \setminus \mathcal{L}$ and $i \in \mathcal{L}$. Moreover $\boldsymbol{v}_i \neq \boldsymbol{v}_k$ for every $i, k \in \mathcal{I} \setminus \mathcal{L}$, $i \neq k$. Indeed, if we had $\boldsymbol{v}_i = \boldsymbol{v}_k$, then $(k,i) \in \mathcal{E}_\mathcal{I}$ by (4.6) while by Lemma 4.7, the node i would have an access to $\bar{\mathcal{I}}$ by a path independent from k. So we should have $\boldsymbol{v}_i < \boldsymbol{v}_k$.

We conclude from this that we can relabel the nodes of $\mathcal{N}$ such that $\mathcal{I} = \{1, 2, \ldots n_\mathcal{I}\}$, $\mathcal{L} = \{n_\mathcal{I} - n_\mathcal{L} + 1, \ldots, n_\mathcal{I}\}$ and

$$\boldsymbol{v}_1 > \boldsymbol{v}_2 > \cdots > \boldsymbol{v}_{n_\mathcal{I}-n_\mathcal{L}} > \boldsymbol{v}_{n_\mathcal{I}-n_\mathcal{L}+1} \geq \cdots \geq \boldsymbol{v}_{n_\mathcal{I}}. \tag{4.7}$$

It follows also that, for $i \in \mathcal{I} \setminus \mathcal{L}$ and $j > i$, $(i,j) \in \mathcal{E}_\mathcal{I}$ if and only if $j = i+1$. Indeed, suppose first $i < n_\mathcal{I} - n_\mathcal{L}$. Then, we cannot have $(i,j) \in \mathcal{E}_\mathcal{I}$ with $j > i+1$ since in this case we would contradict the ordering of the

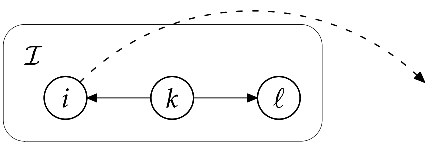


Figure 4.10: If $\boldsymbol{v}_\ell = \min_{j \leftarrow k} \boldsymbol{v}_j$, then $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_\mathcal{I} > \boldsymbol{\pi}^T \boldsymbol{e}_\mathcal{I}$ with $\widetilde{\mathcal{E}}_{\mathrm{out}(\mathcal{I})} = \mathcal{E}_{\mathrm{out}(\mathcal{I})} \setminus \{(k,\ell)\}$.

nodes given by equation (4.7) (see Figure 4.9 again with $k = i + 1$ and remember that by Lemma 4.7, node j has an access to $\bar{\mathcal{I}}$ by a decreasing path). Moreover, node i must link to some node $j > i$ in order to satisfy Assumption 4.16, so $(i, i+1)$ must belong to $\mathcal{E}_\mathcal{I}$. Now, consider the case $i = n_\mathcal{I} - n_\mathcal{L}$. Suppose we had $(i, j) \in \mathcal{E}_\mathcal{I}$ with $j > i + 1$. Let us first note that there can not exist two or more different links (i, ℓ) with $\ell \in \mathcal{L}$ since in this case we could remove one of these links and increase strictly the PageRank of the set $\mathcal{I}$. If $\boldsymbol{v}_j = \boldsymbol{v}_{i+1}$, we could relabel the nodes by permuting these two indices. If $\boldsymbol{v}_j < \boldsymbol{v}_{i+1}$, then with $\widetilde{\mathcal{E}}_\mathcal{I} = \mathcal{E}_\mathcal{I} \cup \{(i, i+1)\} \setminus \{(i, j)\}$, we would have $\widetilde{\boldsymbol{\pi}}^T \boldsymbol{e}_\mathcal{I} > \boldsymbol{\pi}^T \boldsymbol{e}_\mathcal{I}$ by Theorem 4.9 while Assumption 4.16 remains satisfied since the node i would keep access to $\bar{\mathcal{I}}$ via node $i + 1$. That contradicts the optimality assumption. So we have proved that

$$\{(i, j) \in \mathcal{E}_\mathcal{I} \colon i < j \text{ and } i \in \mathcal{I} \setminus \mathcal{L}\} = \{(i, i+1) \colon i \in \mathcal{I} \setminus \mathcal{L}\}. \tag{4.8}$$

Thirdly, it is obvious that

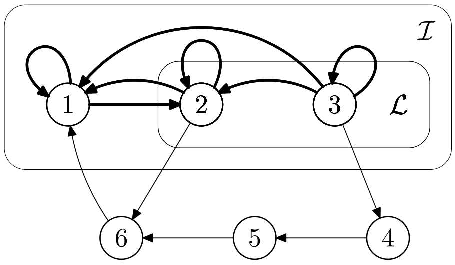
$$\{(i, j) \in \mathcal{E}_\mathcal{I} \colon i < j \text{ and } i \in \mathcal{L}\} \subseteq \{(i, j) \in \mathcal{L} \times \mathcal{L} \colon i < j\}. \tag{4.9}$$

The announced structure for a set $\mathcal{E}_\mathcal{I}$ giving a maximal PageRank score $\boldsymbol{\pi}^T \boldsymbol{e}_\mathcal{I}$ under Assumption 4.16 now follows directly from equations (4.6), (4.8) and (4.9). □

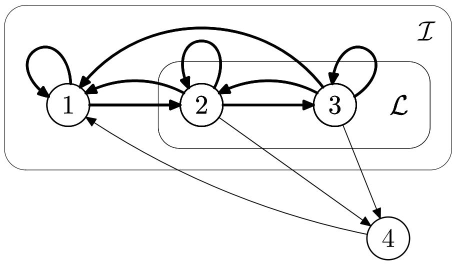
Example 4.20. Consider the graphs given in Figure 4.11. For both cases, the external outlink structure $\mathcal{E}_{\text{out}(\mathcal{I})}$ with two leaking nodes, as well as $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ are given. With $c = 0.85$ and $\boldsymbol{z}$ the uniform probability vector, the optimal internal link structure for configuration (a) is given by $\mathcal{E}_\mathcal{I} = \mathcal{E}_\mathcal{I}^L$, while in configuration (b) we have $\mathcal{E}_\mathcal{I} = \mathcal{E}_\mathcal{I}^U$, with $\mathcal{E}_\mathcal{I}^L$ and $\mathcal{E}_\mathcal{I}^U$ defined in Theorem 4.19. ◇

4.3.3 Optimal link structure

Combining the optimal outlink structure and the optimal internal link structure described in Theorems 4.17 and 4.19, we now find the *optimal linkage strategy* for a set of web pages. Let us note that, since we have here control on both $\mathcal{E}_\mathcal{I}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$, there are no more cases of several final classes or several leaking nodes to consider. For an example of optimal link structure, see Figure 4.2.



(a)



(b)

Figure 4.11: Bold arrows represent optimal internal *link structures. In (a) we have $\mathcal{E}_{\mathcal{I}} = \mathcal{E}^L_{\mathcal{I}}$, while $\mathcal{E}_{\mathcal{I}} = \mathcal{E}^U_{\mathcal{I}}$ in (b).*

Theorem 4.21 (Optimal link structure)**.** *Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ be such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. Then there exists a permutation of the indices such that $\mathcal{I} = \{1, 2, \ldots, n_{\mathcal{I}}\}$,*

$$\boldsymbol{v}_1 > \cdots > \boldsymbol{v}_{n_{\mathcal{I}}} > \boldsymbol{v}_{n_{\mathcal{I}}+1} \geq \cdots \geq \boldsymbol{v}_n,$$

and $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ have the following structure:

$$\begin{aligned} \mathcal{E}_{\mathcal{I}} &= \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon j \leq i \text{ or } j = i+1\}, \\ \mathcal{E}_{\text{out}(\mathcal{I})} &= \{(n_{\mathcal{I}}, n_{\mathcal{I}}+1)\}. \end{aligned}$$

Proof. Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given and suppose $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ are such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. Let us relabel the nodes of $\mathcal{N}$ such that $\mathcal{I} = \{1, 2, \ldots, n_{\mathcal{I}}\}$ and $\boldsymbol{v}_1 \geq \cdots \geq \boldsymbol{v}_{n_{\mathcal{I}}} > \boldsymbol{v}_{n_{\mathcal{I}}+1} = \max_{j \in \bar{\mathcal{I}}} \boldsymbol{v}_j$. By Theorem 4.19, $(i,j) \in \mathcal{E}_{\mathcal{I}}$ for every nodes $i, j \in \mathcal{I}$ such that $j \leq i$. In particular, every node of $\mathcal{I}$ has an access to node 1. Therefore, there is a unique final class $\mathcal{F}_1 \subseteq \mathcal{I}$ in the subgraph $(\mathcal{I}, \mathcal{E}_{\mathcal{I}})$. So, by

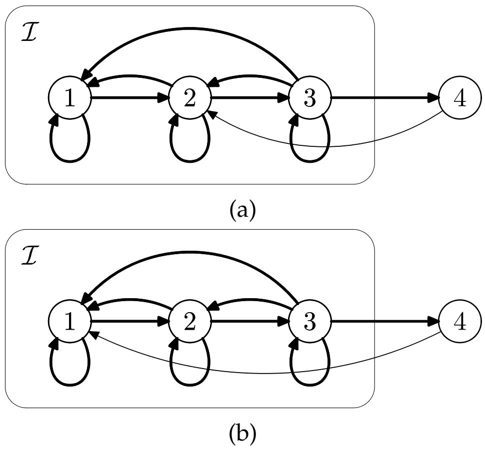


Figure 4.12: For $\mathcal{I} = \{1, 2, 3\}$, $c = 0.85$ and z uniform, the link structure in (a) is not optimal and yet it satisfies the necessary conditions of Theorem 4.21 (see Example 4.22).

Theorem 4.17, $\mathcal{E}_{\mathrm{out}(\mathcal{I})} = \{(k, \ell)\}$ for some $k \in \mathcal{F}_1$ and $\ell \in \bar{\mathcal{I}}$. Without loss of generality, we can suppose that $\ell = n_{\mathcal{I}} + 1$. By Theorem 4.19 again, the leaking node k must be equal to $n_{\mathcal{I}}$. Therefore $(i, i+1) \in \mathcal{E}_{\mathcal{I}}$ for every node $i \in \{1, \ldots, n_{\mathcal{I}} - 1\}$. $\square$

The link structure described in Theorem 4.21 is a necessary *but not sufficient* condition in order to have a maximal PageRank, as shown by the following example.

Example 4.22. Consider the graphs in Figure 4.12. Let $c = 0.85$ and a uniform personalization vector $z = \frac{1}{n}\mathbf{1}$. Both graphs have the link structure required Theorem 4.21 in order to have a maximal PageRank, with $v_{(a)} = \begin{pmatrix} 6.484 & 6.42 & 6.224 & 5.457 \end{pmatrix}^T$ and $v_{(b)} = \begin{pmatrix} 6.432 & 6.494 & 6.247 & 5.52 \end{pmatrix}^T$. The configuration (a) is not optimal since in this case, the PageRank $\pi_{(a)}^T e_{\mathcal{I}} = 0.922$ is strictly less than the PageRank $\pi_{(b)}^T e_{\mathcal{I}} = 0.926$ obtained by the configuration (b). Let us nevertheless note that, with a non uniform personalization vector $z = \begin{pmatrix} 0.7 & 0.1 & 0.1 & 0.1 \end{pmatrix}^T$, the link structure (a) is optimal. $\diamond$

4.4 Extensions and variants

We now present some extensions and variants of the results of the previous section. We first emphasize the role of parents of $\mathcal{I}$. Secondly, we briefly talk about Avrachenkov–Litvak's optimal link structure for the case where $\mathcal{I}$ is a singleton. Then we give variants of Theorem 4.21 when self-links are forbidden or when a minimal number of external outlinks is required. Finally, we make some comments of the influence of external *inlinks* on the PageRank of $\mathcal{I}$.

4.4.1 Linking to parents

If some node of $\mathcal{I}$ has at least one parent in $\bar{\mathcal{I}}$ then the optimal linkage strategy for $\mathcal{I}$ is to have the internal link structure described in Theorem 4.21 together with a single link to one of the parents of $\mathcal{I}$.

Corollary 4.23 (Necessity of linking to parents)**.** *Let $\mathcal{E}_{\text{in}(\mathcal{I})} \neq \emptyset$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. Then $\mathcal{E}_{\text{out}(\mathcal{I})} = \{(i,j)\}$, for some $i \in \mathcal{I}$ and $j \in \bar{\mathcal{I}}$ such that $(j,k) \in \mathcal{E}_{\text{in}(\mathcal{I})}$ for some $k \in \mathcal{I}$.*

Proof. This is a direct consequence of Lemma 4.5 and Theorem 4.21. □

Let us nevertheless remember that not every parent of nodes of $\mathcal{I}$ will give an optimal link structure, as we have already discussed in Example 4.6 and we develop now.

Example 4.24. We continue Example 4.6 and consider the graph in Figure 4.3 as basic absorbing graph for $\mathcal{I} = \{1\}$, that is, $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ are given. We take $c = 0.85$ as damping factor and a uniform personalization vector $\boldsymbol{z} = \frac{1}{n}\mathbf{1}$. We have seen in Example 4.6 than $\mathcal{V}_0 = \{2,3,4\}$. Let us consider the value of the PageRank $\boldsymbol{\pi}_1$ for different sets $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$:

	$\mathcal{E}_{\text{out}(\mathcal{I})}$: $\emptyset$	$\{(1,2)\}$	$\{(1,5)\}$	$\{(1,6)\}$	$\{(1,2),(1,3)\}$
$\mathcal{E}_{\mathcal{I}} = \emptyset$	/	0.1739	0.1402	0.1392	0.1739
$\mathcal{E}_{\mathcal{I}} = \{(1,1)\}$	0.5150	0.2600	0.2204	0.2192	0.2231

As expected from Corollary 4.26, the optimal linkage strategy for $\mathcal{I} = \{1\}$ is to have a self-link and a link to one of the nodes 2, 3 or 4. We note also that a link to node 6, which is a parent of node 1 provides a lower PageRank than a link to node 5, which is not parent of 1. Finally, if we forbid self-links (see below), then the optimal linkage strategy is to link to one *or more* of the nodes $2, 3, 4$. $\diamond$

In the case where no node of $\mathcal{I}$ has a parent in $\bar{\mathcal{I}}$, then *every* structure like described in Theorem 4.21 will give an optimal link structure.

Proposition 4.25 (No external parent)**.** *Let $\mathcal{E}_{\mathrm{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose that $\mathcal{E}_{\mathrm{in}(\mathcal{I})} = \emptyset$. Then the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16 if and only if*

$$\begin{aligned} \mathcal{E}_{\mathcal{I}} &= \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon j \leq i \text{ or } j = i+1\}, \\ \mathcal{E}_{\mathrm{out}(\mathcal{I})} &= \{(n_{\mathcal{I}}, n_{\mathcal{I}}+1)\}. \end{aligned}$$

for some permutation of the indices such that $\mathcal{I} = \{1, 2, \ldots, n_{\mathcal{I}}\}$.

Proof. This follows directly from $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = (1-c)\boldsymbol{z}^T \boldsymbol{v}$ and the fact that, if $\mathcal{E}_{\mathrm{in}(\mathcal{I})} = \emptyset$,

$$\boldsymbol{v} = (I - cP)^{-1} \boldsymbol{e}_{\mathcal{I}} = \begin{pmatrix} (I - cP_{\mathcal{I}})^{-1} \mathbf{1} \\ 0 \end{pmatrix},$$

up to a permutation of the indices. $\square$

4.4.2 Optimal linkage strategy for a singleton

The optimal outlink structure for a single web page has already been given by Avrachenkov and Litvak in [9]. Their result becomes a particular case of Theorem 4.21. Note that in the case of a single node, the possible choices for $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ can be found a priori by considering the basic absorbing graph, since $\mathcal{V} = \mathcal{V}_0$ in this case.

Corollary 4.26 (Optimal link structure for a single node)**.** *Let $\mathcal{I} = \{i\}$ and let $\mathcal{E}_{\mathrm{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Then the PageRank $\boldsymbol{\pi}_i$ is maximal under Assumption 4.16 if and only if $\mathcal{E}_{\mathcal{I}} = \{(i,i)\}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})} = \{(i,j)\}$ for some $j \in \mathcal{V}_0$.*

Proof. This follows directly from Lemma 4.15 and Theorem 4.21. $\square$

4.4.3 Optimal linkage strategy under additional assumptions

Let us consider the problem of maximizing the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ when *self-links are forbidden*. Indeed, it seems to be often supposed that Google's PageRank algorithm does not take self-links into account [90]. In this case, Theorem 4.21 can be adapted readily for the case where $|\mathcal{I}| \geq 2$. When $\mathcal{I}$ is a singleton, we must have $\mathcal{E}_{\mathcal{I}} = \emptyset$, so $\mathcal{E}_{\text{out}(\mathcal{I})}$ can contain *several* links, as stated in Theorem 4.17.

Corollary 4.27 (Optimal link structure with no self-links)**.** *Suppose $|\mathcal{I}| \geq 2$. Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ be such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16 and under the assumption that there does not exist $i \in \mathcal{I}$ such that $\{(i,i)\} \in \mathcal{E}_{\mathcal{I}}$. Then there exists a permutation of the indices such that $\mathcal{I} = \{1, 2, \ldots, n_{\mathcal{I}}\}$, $\boldsymbol{v}_1 > \cdots > \boldsymbol{v}_{n_{\mathcal{I}}} > \boldsymbol{v}_{n_{\mathcal{I}}+1} \geq \cdots \geq \boldsymbol{v}_n$, and $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ have the following structure:*

$$\begin{aligned} \mathcal{E}_{\mathcal{I}} &= \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon j < i \text{ or } j = i+1\}, \\ \mathcal{E}_{\text{out}(\mathcal{I})} &= \{(n_{\mathcal{I}}, n_{\mathcal{I}}+1)\}. \end{aligned}$$

Corollary 4.28 (Optimal link structure for a single node with no self-link)**.** *Suppose $\mathcal{I} = \{i\}$. Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Suppose $\mathcal{E}_{\mathcal{I}} = \emptyset$. Then the PageRank $\boldsymbol{\pi}_i$ is maximal under Assumption 4.16 if and only if $\emptyset \neq \mathcal{E}_{\text{out}(\mathcal{I})} \subseteq \mathcal{V}_0$.*

We now consider the problem of maximizing the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ when *several external outlinks are required*. Then the proof of Theorem 4.17 can be modified in order to obtain the following variant of Theorem 4.21.

Corollary 4.29 (Optimal link structure with several external outlinks)**.** *Let $\mathcal{E}_{\text{in}(\mathcal{I})}$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16 and assumption that $|\mathcal{E}_{\text{out}(\mathcal{I})}| \geq r$. Then there exists a permutation of the indices such that $\mathcal{I} = \{1, 2, \ldots, n_{\mathcal{I}}\}$, $\boldsymbol{v}_1 > \cdots > \boldsymbol{v}_{n_{\mathcal{I}}} > \boldsymbol{v}_{n_{\mathcal{I}}+1} \geq \cdots \geq \boldsymbol{v}_n$, and $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\text{out}(\mathcal{I})}$ have the following structure:*

$$\begin{aligned} \mathcal{E}_{\mathcal{I}} &= \{(i,j) \in \mathcal{I} \times \mathcal{I} \colon j < i \text{ or } j = i+1\}, \\ \mathcal{E}_{\text{out}(\mathcal{I})} &= \{(n_{\mathcal{I}}, j_k) \colon j_k \in \mathcal{V} \text{ for } k = 1, \ldots, r\}. \end{aligned}$$

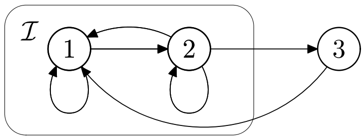


Figure 4.13: For $\mathcal{I} = \{1, 2\}$, adding the external inlink $(3, 2)$ gives $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} < \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ (see Example 4.31).

4.4.4 External inlinks

Finally, let us make some comments about the addition of external inlinks to the set $\mathcal{I}$. It is well known that adding an inlink to a particular page always increases the PageRank of this page [8, 64]. This can be viewed as a direct consequence of Theorem 4.9 and Lemma 4.4. The case of a set of several pages $\mathcal{I}$ is not as simple. The following theorem shows that, if the set $\mathcal{I}$ has a link structure as described in Theorem 4.21 then adding an inlink to a page of $\mathcal{I}$ from a page $j \in \bar{\mathcal{I}}$ which is *not* a parent of some node of $\mathcal{I}$ will increase the PageRank of $\mathcal{I}$. But in general, adding an inlink to some page of $\mathcal{I}$ from $\bar{\mathcal{I}}$ may *decrease* the PageRank of the set $\mathcal{I}$, as shown in Examples 4.31 and 4.32.

Theorem 4.30 (External inlinks)**.** *Let $\mathcal{I} \subseteq \mathcal{N}$ and a graph defined by a set of links $\mathcal{E}$. If*

$$\min_{i \in \mathcal{I}} \boldsymbol{v}_i > \max_{j \notin \mathcal{I}} \boldsymbol{v}_j,$$

then, for every $j \in \bar{\mathcal{I}}$ which is not a parent of $\mathcal{I}$, and for every $i \in \mathcal{I}$, the graph defined by $\widetilde{\mathcal{E}} = \mathcal{E} \cup \{(j, i)\}$ gives $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} > \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$.

Proof. This follows directly from Theorem 4.9. □

Example 4.31. Let us show with an example that a new external inlink is not always profitable for a set $\mathcal{I}$ in order to improve its PageRank, even if $\mathcal{I}$ has an optimal linkage strategy. Consider for instance the graph in Figure 4.13. With $c = 0.85$ and $\boldsymbol{z}$ uniform, we have $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = 0.8481$. But if we consider the graph defined by $\widetilde{\mathcal{E}}_{\text{in}(\mathcal{I})} = \mathcal{E}_{\text{in}(\mathcal{I})} \cup \{(3, 2)\}$, then we have $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} = 0.8321 < \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$. ◇

Example 4.32. A new external inlink does not not always increase the PageRank of a set $\mathcal{I}$ even if this new inlink comes from a page which is

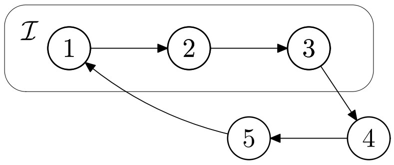


Figure 4.14: For $\mathcal{I} = \{1, 2, 3\}$, adding the external inlink $(4, 3)$ gives $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} < \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ (see Example 4.32).

not already a parent of some node of $\mathcal{I}$. Consider for instance the graph in Figure 4.14. With $c = 0.85$ and $\boldsymbol{z}$ uniform, we have $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}} = 0.6$. But if we consider the graph defined by $\tilde{\mathcal{E}}_{\text{in}(\mathcal{I})} = \mathcal{E}_{\text{in}(\mathcal{I})} \cup \{(4, 3)\}$, then we have $\tilde{\boldsymbol{\pi}}^T \boldsymbol{e}_{\mathcal{I}} = 0.5897 < \boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$. ◇

4.5 Experiments on a subgraph of the Web

In this section, we briefly present our experiments on a large-scale example. Our aim is to show how the PageRank and the ranking of a web site may be improved by modifying its internal and external outlink structure. As Theorem 4.21 does not give us a constructive way to find the optimal link structure of a web site, we use a simple heuristic to try to approach it.

We first select a small community $\mathcal{I}$ of web pages with many internal and relatively few external links. We start with the original graph, that we call Structure **A**:

$$\mathcal{E}_{\mathcal{I}}^A = \mathcal{E}_{\mathcal{I}} \quad \text{and} \quad \mathcal{E}_{\text{out}(\mathcal{I})}^A = \mathcal{E}_{\text{out}(\mathcal{I})}.$$

We first add internal links to $\mathcal{I}$ in order to transform it in a clique, that is a subgraph with every possible link (we do not consider self-links). This is what we call Structure **B**:

$$\mathcal{E}_{\mathcal{I}}^B = \{(i, j) \in \mathcal{I} \times \mathcal{I} : i \neq j\} \quad \text{and} \quad \mathcal{E}_{\text{out}(\mathcal{I})}^B = \mathcal{E}_{\text{out}(\mathcal{I})}.$$

Then, we try to modify the link structure of $\mathcal{I}$ in a way similar to the optimal link structure given in Theorem 4.21, with an internal structure given by a forward chain of links and every possible backward link and with a unique external outlink. For the Structure **B**, we compute the vector $\boldsymbol{v}^B$ of expected numbers of visits in $\mathcal{I}$ before zapping, by iterating $\boldsymbol{v}^B(k+1) = cP^B\boldsymbol{v}^B(k) + \boldsymbol{e}_{\mathcal{I}}$. Then, we sort the nodes of $\mathcal{I}$ by decreasing value of $\boldsymbol{v}^B$: the mapping $\sigma^B\colon \{1,\ldots,n_{\mathcal{I}}\} \to \{1,\ldots,n_{\mathcal{I}}\}$ is such that $\boldsymbol{v}_{\sigma^B(1)} \geq \cdots \geq \boldsymbol{v}_{\sigma^B(n_{\mathcal{I}})}$, with $n_{\mathcal{I}} = |\mathcal{I}|$. We then successively delete all external outlinks for the first node of $\mathcal{I}$, then for the second node and so on. For the last node of $\mathcal{I}$, we delete every external outlink and add an outlink to a node of $\bar{\mathcal{I}}$ which has the higher expected number of visits in $\mathcal{I}$ before zapping. This gives us the Structure **C**:

$$\begin{aligned} \mathcal{E}_{\mathcal{I}}^C &= \mathcal{E}_{\mathcal{I}}^B \\ \mathcal{E}_{\mathrm{out}(\mathcal{I})}^C &= \{(\sigma^B(n_{\mathcal{I}}),\ell) \in \mathcal{I}\times\bar{\mathcal{I}}\colon \ell \in \operatorname*{argmax}_{j\in\bar{\mathcal{I}}} \boldsymbol{v}_j\}. \end{aligned}$$

Then, we successively delete internal links in order to have a forward chain of links together with every possible backward link. The Structure **D** so obtained is hoped to be a nearly optimal linkage strategy.

$$\begin{aligned} \mathcal{E}_{\mathcal{I}}^D &= \{(i,j) \in \mathcal{I}\times\mathcal{I}\colon \sigma^B(j) < \sigma^B(i) \text{ or } \sigma^B(j) = \sigma^B(i+1)\} \\ \mathcal{E}_{\mathrm{out}(\mathcal{I})}^D &= \mathcal{E}_{\mathrm{out}(\mathcal{I})}^C. \end{aligned}$$

Finally, we consider, in Structure **E**, a set $\mathcal{I}$ without any external outlink and with an internal link structure consisting only of links from every node of the set to the first one. By Proposition 4.14, this will give us an upper bound on the PageRank of the set $\mathcal{I}$. Moreover, with no external outlinks, this internal structure is the best for giving a maximal ranking to the first page [53]:

$$\mathcal{E}_{\mathcal{I}}^E = \{(i,\sigma^B(1)) \in \mathcal{I}\times\mathcal{I}\colon i \in \mathcal{I}\} \quad \text{and} \quad \mathcal{E}_{\mathrm{out}(\mathcal{I})}^E = \emptyset.$$

The algorithm can be outlined as following.

1: Find a community $\mathcal{I}$ of well connected pages with at least one external outlink.
2: Let $n_{\mathcal{I}} \leftarrow |I|$.
3: (**A**) Compute the PageRank and the best and worst ranks of $\mathcal{I}$.
4: Transform $\mathcal{I}$ into a clique, i.e., $\mathcal{E}_{\mathcal{I}} \leftarrow \{(i,j) \in \mathcal{I}\times\mathcal{I}\colon i \neq j\}$.
5: (**B**) Compute the PageRank and the best and worst ranks of $\mathcal{I}$.
6: Compute iteratively $\boldsymbol{v} = (I - cP)^{-1}\boldsymbol{e}_{\mathcal{I}}$.
7: Let $\sigma\colon \{1,\ldots,n_{\mathcal{I}}\} \to \{1,\ldots,n_{\mathcal{I}}\}$ such that $\boldsymbol{v}_{\sigma(1)} \geq \cdots \geq \boldsymbol{v}_{\sigma(n_{\mathcal{I}})}$.

if $\{(\sigma(n_{\mathcal{I}}), j) \in \mathcal{E}_{\mathrm{out}(\mathcal{I})}\} = \emptyset$ **then**
    let $\mathcal{E}_{\mathrm{out}(\mathcal{I})} \leftarrow \mathcal{E}_{\mathrm{out}(\mathcal{I})} \cup \{(\sigma(n_{\mathcal{I}}), \ell)\}$ for some $\ell \notin \mathcal{I}$.
end if
for $k = 1$ to $n_{\mathcal{I}} - 1$ **do**
    let $\{(\sigma(k), j) \in \mathcal{E}_{\mathrm{out}(\mathcal{I})}\} \leftarrow \emptyset$;
    compute the PageRank of $\mathcal{I}$.
end for
Let $\{(\sigma(n_{\mathcal{I}}), j) \in \mathcal{E}_{\mathrm{out}(\mathcal{I})}\} \leftarrow \{(\sigma(n_{\mathcal{I}}), \ell)\}$ for some $\ell \in \operatorname{argmax}_{j \in \bar{\mathcal{I}}} v_j$.
(**C**) Compute the PageRank and the best and worst ranks of $\mathcal{I}$.
for $k = 1$ to $n_{\mathcal{I}} - 2$ **do**
    let $\{(\sigma(k), j) \in \mathcal{E}_{\mathcal{I}}\} \leftarrow \{(\sigma(k), \sigma(k+1))\} \cup \{(\sigma(k), \sigma(j)) : j < k\}$;
    compute the PageRank of $\mathcal{I}$.
end for
(**D**) Compute the PageRank and the best and worst ranks of $\mathcal{I}$.
Let $\mathcal{E}_{\mathrm{out}(\mathcal{I})} \leftarrow \emptyset$ and $\mathcal{E}_{\mathcal{I}} \leftarrow \{(i, \sigma(1)) : i \in \mathcal{I}\}$.
(**E**) Compute the PageRank and the best and worst ranks of $\mathcal{I}$.

We experiment this heuristic on a subgraph of the Web with about 280,000 nodes which has been obtained by S. Kamvar from a crawl on the Stanford web [69]. We take a damping factor c equal to 0.85 and a personalization vector $z = \frac{1}{n}\mathbf{1}$. We choose three sets of web pages, with many internal links and relatively few external inlinks and outlinks. The first considered set has 14 pages, with originally 80 internal links, 22 external outlinks and 18 external inlinks. The second one is a set of 15 pages, with originally 125 internal links, 282 external outlinks and 133 external inlinks. And the third set has 24 pages, with originally 552 internal links, 912 external outlinks and 3,196 external inlinks. For each of them, independently, we modify the external outlink structure and the internal link structure with the algorithm described above.

In Figure 4.15, we give the PageRank sum of the set of pages, as well as the rank of its best and worst pages, for the structures **A**, **B**, **C**, **D** and **E**. In the three cases, the PageRank of the set of pages increases significantly when the external outlink structure is modified from **A** to **C** and approaches quite well the theoretical upper bound, given by the PageRank of the isolated structure **E**. Then, when modifying the internal structure, from **C** to **D**, the value of the PageRank does not change a lot. But note that the restructuring of the internal structure leads to a

	structure	PageRank	min rank	max rank
A	original	$5.02 \cdot 10^{-5}$	12,158	261,589
B	clique	$6.01 \cdot 10^{-5}$	20,749	30,509
C	external	$6.56 \cdot 10^{-5}$	21,727	23,903
D	internal	$6.59 \cdot 10^{-5}$	5,199	179,286
E	isolated	$6.59 \cdot 10^{-5}$	1,573	261,589

(a)

	structure	PageRank	min rank	max rank
A	original	$2.028 \cdot 10^{-5}$	62,569	230,044
B	clique	$2.356 \cdot 10^{-5}$	61,833	137,223
C	external	$9.343 \cdot 10^{-5}$	14,182	18,261
D	internal	$9.458 \cdot 10^{-5}$	3,501	162,446
E	isolated	$9.470 \cdot 10^{-5}$	999	258,862

(b)

	structure	PageRank	min rank	max rank
A	original	$1.224 \cdot 10^{-3}$	1,419	1,821
B	clique	$1.224 \cdot 10^{-3}$	1,419	1,821
C	external	$5.117 \cdot 10^{-3}$	322	359
D	internal	$5.136 \cdot 10^{-3}$	23	2,931
E	isolated	$5.139 \cdot 10^{-3}$	4	3,091

(c)

Figure 4.15: Evolution of the PageRank, the minimum rank and the maximum rank of three sets of web pages for which we modify the external outlink and the internal link structure. (a) Set of 14 *pages. (b) Set of* 15 *pages. (c) Set of* 24 *pages.*

considerable improvement of the rank of the best page within the set.

In Figure 4.16, we represent graphically the evolution of the PageRank for one of the three web sites. Here, there are as many steps between **B** and **C** than the number of pages of the considered web site: they correspond to the removal of the external outlinks of the first page, then of the second page, and so on. Similarly, the steps between **C** and **D** correspond to the gradual choice of the internal outlinks for each page of the set.

Remark 4.33. If, for some reason, the computation of the vector v for the Structure **B** is not possible, it can be avoided by the following variant of the heuristic. The nodes of $\mathcal{I}$ may be reordered according to their

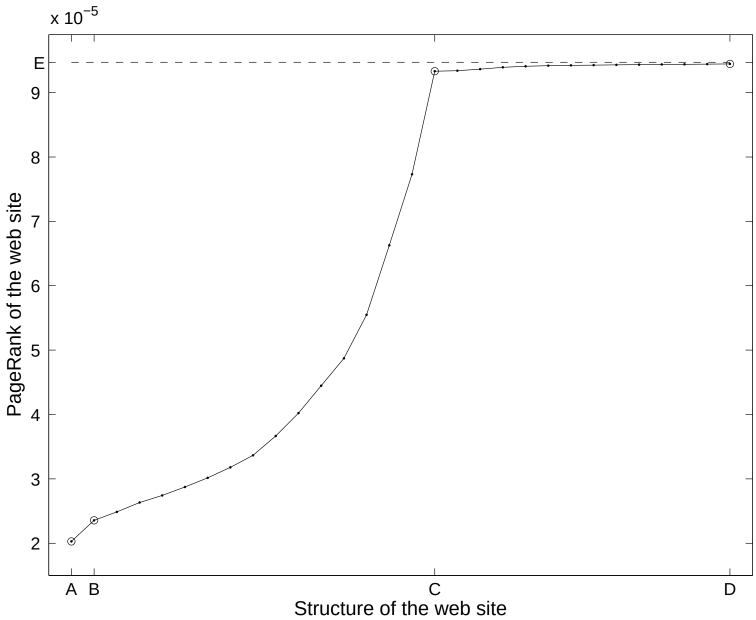


Figure 4.16: Evolution of the PageRank of a set of 15 *web pages whose structure is progressively modified to the structure* ***D****.*

PageRank instead of the vector $\boldsymbol{v}$ for the Structure **B**. Also, the external outlink may be chosen to a node $j \in \bar{\mathcal{I}}$ which maximizes $(P\boldsymbol{e}_{\mathcal{I}})_j$. On our three examples, the results where not much affected by this variant. Note that $\boldsymbol{v} = \boldsymbol{e}_{\mathcal{I}} + cP\boldsymbol{e}_{\mathcal{I}} + c^2P^2\boldsymbol{e}_{\mathcal{I}} + \cdots$, so $\operatorname{argmax}_{j\in\bar{\mathcal{I}}}(P\boldsymbol{e}_{\mathcal{I}})_j$ gives a first order estimation of $\operatorname{argmax}_{j\in\bar{\mathcal{I}}} v_j$. $\diamond$

4.6 Conclusions

In this chapter we describe the optimal link structure for a web site in order to maximize its PageRank. This optimal structure with a forward chain and every possible backward link may be not intuitive. To our knowledge, it has never been mentioned, while topologies such as cliques, rings or stars are considered in the literature on collusion and alliance between pages [10, 53]. Note that, in these papers, the interaction between the considered web pages and the rest of the Web is not studied properly. They suppose in particular that changing an external outlink of a web page would not change the PageRank of this page at all. They in some way study the influence of internal links on the PageRank of a set of pages by isolating it from the rest of the Web.

Moreover, the optimal structure we provide gives new insight into the claim of Bianchini et al. [16] that, in order to maximize the PageRank of a web site, hyperlinks to the rest of the web graph "should be in pages with a small PageRank and that have many internal hyperlinks". More precisely, we have seen that the leaking pages must be chosen with respect to the mean number of visits before zapping that they give to the web site, rather than with respect to their PageRank.

Let us now present some possible directions for future work. We have noticed in Example 4.22 that the first node of $\mathcal{I}$ in the forward chain of an optimal link structure is not necessarily *a child* of some node of $\bar{\mathcal{I}}$. In the example we gave, the personalization vector was not uniform. We wonder if this could occur with a uniform personalization vector and make the following conjecture.

Conjecture 4.34. Let $\mathcal{E}_{\mathrm{in}(\mathcal{I})} \neq \emptyset$ and $\mathcal{E}_{\bar{\mathcal{I}}}$ be given. Let $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ such that $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{I}}$ is maximal under Assumption 4.16. If $\boldsymbol{z} = \frac{1}{n}\mathbf{1}$, then there exists $j \in \bar{\mathcal{I}}$ such that $(j,i) \in \mathcal{E}_{\mathrm{in}(\mathcal{I})}$, where $i \in \operatorname{argmax}_k \boldsymbol{v}_k$.

If this conjecture is correct we can also ask if the node $j \in \bar{\mathcal{I}}$ such that $(j,i) \in \mathcal{E}_{\mathrm{in}(\mathcal{I})}$ where $i \in \operatorname{argmax}_k \boldsymbol{v}_k$ belongs to $\mathcal{V}$.

Another question concerns the optimal linkage strategy in order to maximize an arbitrary linear combination of the PageRanks of the nodes of $\mathcal{I}$. In particular, we may want to maximize the PageRank $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{S}}$ of *a target subset* $\mathcal{S} \subseteq \mathcal{I}$ by choosing $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ as usual. A general shape for an optimal link structure seems difficult to find, as shown in the following example.

Example 4.35. Consider the graphs in Figure 4.17. In both cases, let $c = 0.85$ and $\boldsymbol{z} = \frac{1}{n}\mathbf{1}$. Let $\mathcal{I} = \{1,2,3\}$ and let $\mathcal{S} = \{1,2\}$ be the target set. In the configuration (a), the optimal sets of links $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ for maximizing $\boldsymbol{\pi}^T \boldsymbol{e}_{\mathcal{S}}$ have the link structure described in Theorem 4.21. But in (a), the optimal $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathrm{out}(\mathcal{I})}$ do not have this structure. Let us

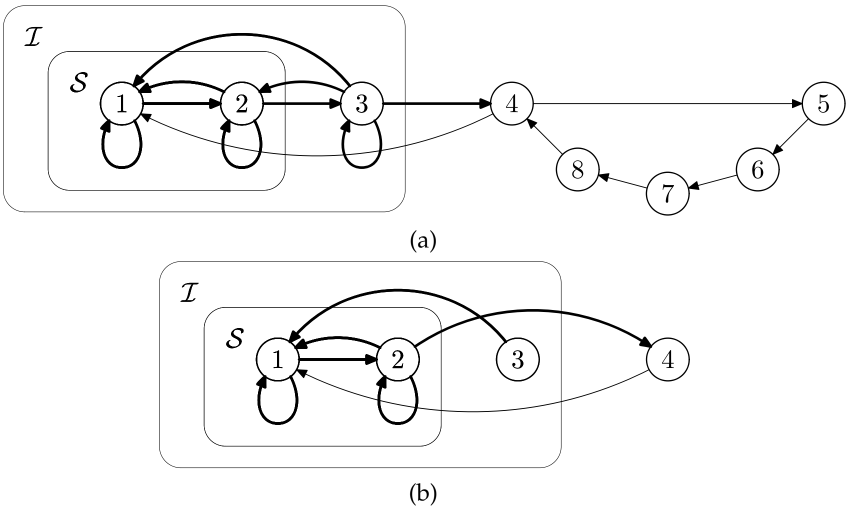


Figure 4.17: In (a) and (b), bold arrows represent optimal link structures for $\mathcal{I} = \{1, 2, 3\}$ with respect to a target set $\mathcal{S} = \{1, 2\}$ (see Example 4.35).

note nevertheless that, by Theorem 4.21, the subsets $\mathcal{E}_\mathcal{S}$ and $\mathcal{E}_{\text{out}(\mathcal{S})}$ must have the link structure described in Theorem 4.21. $\diamond$

It should also be interesting to consider the problem of maximizing the PageRank sum of a set of pages under other assumptions that our Assumption 4.16 of accessibility. This assumption is in particular not scalable. Indeed, having exactly one external outlink does not mean the same for a set of two pages and a set of a thousand pages. One could for instance impose a lower bound on the ratio between the number of external outlinks and internal links, or on the probability for a random surfer to leave the set of pages at some time. It may also be interesting to see how to extend our results to the case where we allow to consider any stochastic matrix P and not only diagonal scalings of adjacency matrices.

One could also want to consider the problem of maximizing the PageRank sum of a set of web pages when the damping factor c approaches 1 in the case the matrix P is irreducible. Unfortunately, the vector $\boldsymbol{v}$ will not be useful anymore in this case. Indeed, since P is irreducible, $\lim_{c\to\infty}(1-c)(I-cP)^{-1} = \mathbf{1}\boldsymbol{u}^T$, where $\boldsymbol{u}^T$ is the invariant measure of P (see p. 33). So, the vector $\boldsymbol{v} \approx (\boldsymbol{u}^T\boldsymbol{e}_\mathcal{I})\mathbf{1}$ for a damping factor $c \approx 1$. The difference between the invariant measures of a stochastic matrix P and a perturbation of it can be expressed by using the group inverse $P^\sharp$ [36]. With this expression, the techniques used in this chapter could perhaps be adapted in order to find the optimal structure in the case where $c = 1$.

Chapter 5

Multiple equilibria of nonhomogeneous Markov chains and self-validating web rankings

Google's PageRank measures how often a given web page is visited by a random surfer on the web graph. It seems realistic that PageRank may also have an influence on the behavior of web surfers. In this chapter, we propose a simple model taking into account the mutual influence between web ranking and web surfing. Our ranking, the T-PageRank, is a nonlinear generalization of the PageRank. A positive parameter T, the temperature, measures the confidence of the web surfer in the web ranking. When the temperature is large, the web surfer is not much influenced by the web ranking and the T-PageRank is then well defined. But when the temperature is small, that is, when the web surfer rely too much on the web ranking, the T-PageRank may depend strongly on the initial ranking and self-validating effects may appear. This suggest that in order to keep a good quality ranking, Google should not use information about the trajectory actually followed by its users for the computation of its scores.

5.1 Introduction

In Section 2.2.1, we have seen that PageRank measures *how often a given web page would be visited by a random walker on the web graph* for a particular model of web surfing. In the most basic definition of the PageRank, the web graph is assumed to be strongly connected and the random web surfer moves from pages to pages by following the hyperlinks at random. If $A = [A_{ij}]_{i,j=1}^n$ is the irreducible adjacency matrix of the web graph then the stochastic matrix $P = [A_{ij}/\sum_k A_{ik}]_{i,j=1}^n$ is the transition matrix of the random walk describing the trajectory of the random web surfer. The vector of scores $\boldsymbol{p}$ is then defined as the unique invariant measure of P. In other words, $\boldsymbol{p}$ is the unique stochastic vector that is such that

$$\boldsymbol{p}^T = \boldsymbol{p}^T P.$$

However, the assumption that a web surfer makes uniform draws on the web graph may seem unrealistic: a web surfer could have an a priori idea of the value of web pages, therefore favoring pages from reputed sites. The web rank may influence the reputation of the web sites, and hence it may influence the behavior of the web surfers, which ultimately may influence the web rank. In this chapter, we propose a simple model taking into account the mutual influence between web ranking and web surfing.

The T-PageRank We consider a sequence of stochastic vectors, representing successive web ranks, $\boldsymbol{p}(0), \boldsymbol{p}(1), \ldots$. The sequence is defined as follows. The current ranking $\boldsymbol{p}(k)$ induces a random walk on the web graph. We assume that the web surfer moves from page i to page j with probability proportional to $A_{ij}e^{E(\boldsymbol{p}(k)_j)/T}$, where E is an increasing function, the *energy*, and $T > 0$ is a fixed positive parameter, the *temperature*. The web surfer's trajectory is therefore a Markov chain with transition matrix $P(\boldsymbol{p}(k))$, where $P(\boldsymbol{x})$ is defined for any nonnegative vector $\boldsymbol{x}$ by

$$P(\boldsymbol{x})_{ij} = \frac{A_{ij}e^{E(\boldsymbol{x}_j)/T}}{\sum_k A_{ik}e^{E(\boldsymbol{x}_k)/T}}.$$

The unique stationary distribution of this Markov chain, i.e., the invariant measure of the matrix $P(\boldsymbol{p}(k))$, is then used to update the web rank.

Thus,

$$\boldsymbol{p}(k+1) = \boldsymbol{u}(\boldsymbol{p}(k)), \tag{5.1a}$$

where, for any nonnegative vector $\boldsymbol{x}$, $\boldsymbol{u}(\boldsymbol{x})$ is the unique stochastic vector such that

$$\boldsymbol{u}(\boldsymbol{x})^T = \boldsymbol{u}(\boldsymbol{x})^T P(\boldsymbol{x}). \tag{5.1b}$$

We call the *T-PageRank* the limit of $\boldsymbol{p}(k)$ when k tends to infinity, if it exists, or a fixed point of the map $\boldsymbol{u}$.

Note that if $\boldsymbol{p}(0)$ is the uniform distribution, then $\boldsymbol{p}(1)$ is the classical PageRank. Note also that the temperature T measures the selectivity of the process. If T is small, with overwhelming probability, the web surfer moves from page i to one of the pages j referenced by page i of best rank, i.e., maximizing $\boldsymbol{p}(k)_j$, whereas if $T = \infty$, the web surfer draws the next page among the pages j referenced by the page i, with the uniform distribution, as in the standard web rank definition. For $T = \infty$, the T-PageRank coincides with the classical PageRank, because in this case $P(\boldsymbol{p}(k)) = P$.

We also consider the simple iteration defined by

$$\widetilde{\boldsymbol{p}}(k+1)^T = \boldsymbol{f}(\widetilde{\boldsymbol{p}}(k))^T, \quad \text{with } \boldsymbol{f}(\boldsymbol{x})^T = \boldsymbol{x}^T P(\boldsymbol{x}), \tag{5.2}$$

where $\widetilde{\boldsymbol{p}}(0)$ is an arbitrary stochastic vector. From a computational point of view, this is similar to the standard power method.

Main results Our first main result shows that, if the temperature T is sufficiently large, the T-PageRank exists, is unique and does not depend on the initial ranking. Moreover, if the matrix A is primitive, the generalized power algorithm (5.2) can be used to compute the T-PageRank.

Theorem 5.1. *Assume that A is irreducible. If $T \geq n\,\mathrm{Lip}(E)$, where $\mathrm{Lip}(E)$ is the Lipschitz constant of the function E, then the map $\boldsymbol{u}$ given by* (5.1b) *has a unique fixed point and the iterates* (5.1a) *converge to it for every initial ranking. Moreover, if A is primitive and if T is large enough, the iterates* (5.2) *converge to this unique fixed point for every initial ranking.*

For small values of T, several T-PageRanks exist, depending on the choice of the initial ranking. In some cases, the T-PageRank does

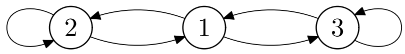


Figure 5.1: For this graph, self-validating effects appear for small temperatures.

nothing but validating the initial "belief" in the interest of pages given by the initial ranking.

Example 5.2. Consider for instance the graph given in Figure 5.1 with adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Let $E(x) = x$ for all $x \in \mathbb{R}$ and let $T = \frac{1}{4}$. Let $\boldsymbol{p}(0) = \widetilde{\boldsymbol{p}}(0) = \left(\frac{1}{3} \quad \frac{1}{3}+\varepsilon \quad \frac{1}{3}-\varepsilon\right)^T$ for an arbitrary small $\varepsilon > 0$. Then the iterates (5.1) and (5.2) converge to a T-PageRank close to $\left(0.021 \quad 0.978 \quad 0.001\right)^T$, so the initial belief that the node 2 is more interesting than node 3 has strongly increased. $\diamond$

Our second main result shows that the existence of multiple T-PageRanks is in fact a general feature, when T is small enough.

Theorem 5.3. *If A is irreducible and has at least two positive diagonal entries, then multiple T-PageRanks exist for T small enough.*

These two theorems follow respectively from Theorems 5.14 and 5.15, and from Theorem 5.19 and Remark 5.20 , which are stated in a more general framework.

We also study variants, bringing more realistic models of the behavior of the web surfer. This includes the presence of a "damping factor", as in the standard definition of Google's PageRank. We also consider the situation where the web surfer may take the web rank into account only when visiting a special page (the search engine's web page). Similar conclusions apply to such variants. These variants also have the advantage of allowing one to work with non strongly connected web graphs.

Method In order to analyze the map (5.1b), we first use Tutte's Matrix Tree Theorem [115] to express explicitly the invariant measure $\boldsymbol{u}(\boldsymbol{x})$ in terms of the entries of $P(\boldsymbol{x})$. Then we study the convergence and the fixed points of (5.1) by using results of nonlinear Perron–Frobenius theory due to Nussbaum [101] and Krause [80], that we already presented in Section 2.1.4. To analyze the iteration (5.2), we use two different approaches, that give convergence results under distinct technical assumptions. Our first approach is to show that, if T is sufficiently large, the map $\boldsymbol{f}$ from the simplex to itself is a contraction for some norm. Our second approach uses Hilbert's projective metric and Birkhoff's coefficient of ergodicity.

Related work Several variants of the PageRank are considered in the literature in order to have a more realistic model of the behavior of the web surfers. For instance, some authors propose to introduce the browser's back button in the model [43, 22, 112]. Others try to develop a topic-sensitive ranking by considering a surfer whose trajectory on the web graph depends on his query or bookmarks [56, 57, 66, 105].

Our use of transition probabilities proportional to $A_{ij}e^{E(\boldsymbol{p}(k)_j)/T}$, where E is an energy function and T the temperature, is reminiscent of simulated annealing algorithms. For a reference, see Catoni [34]. In the context of opinion formation, Holyst et al. [60] study a social impact model where the probability that an individual changes his opinion depends on a "social temperature" T, which measures the randomness of the process.

The iteration (5.1) can be studied in the settings of nonlinear Perron–Frobenius theory. We have seen in Section 2.1.4 that many works exist in the literature in order to generalize the classical Perron–Frobenius theorems to nonlinear maps on cones satisfying some hypotheses such as primitivity, positivity or homogeneity.

The iteration (5.2) has been studied by several authors in an abstract setting. Artzrouni and Gavart [6] analyze its dynamics when $\boldsymbol{x}(k)$ behaves asymptotically like $\lambda^k \boldsymbol{x}_*$ for some $\lambda \neq 1$ and $\boldsymbol{x}_*$. For $\lambda = 1$, it can be useful to look at the stability of a linearization of the system near one of its fixed points [33]. When $P(\boldsymbol{x})$ is stochastic and satisfies certain monotonicity conditions, Conlisk [38] proves the convergence of the iterates (5.2) to a stable limit. Lorenz [87] proves their convergence for *column*-stochastic matrices satisfying classic properties of opinion dynamics models. In [88], he experimentally studies a reformulation of these models with stochastic matrices $P(\boldsymbol{x})$, where $\boldsymbol{x}$ is an opinion distribution vector. Iterations like (5.2) could also be studied in the setting of nonhomogeneous products of matrices. In this case, iterations like $\boldsymbol{x}(k+1)^T = \boldsymbol{x}(k)^T P(k)$ are considered, where the matrices do not depend explicitly on $\boldsymbol{x}$. Two classical approaches to study their dynamics and convergence are the use of ergodicity coefficients [5, 55, 110, 119] or of the joint spectral radius [52, 55, 65]. However, the main results of this chapter can not be deduced from these works.

Finally, several authors try to understand how web rankings may have an impact on the evolution of the link structure of the Web, with experimental measures on real data or models of preferential attachment [11, 35, 37, 46]. Most of these studies suggest that search engines introduce an unfortunate bias on the evolution of the Web: the well established pages become more and more popular, while recently created pages are penalized. Note however that Fortunato et al. [46] claims that the bias introduced by search engines is much weaker than what is found by other authors. Their model takes into account the topical interests of the web search engines' users.

Outline of the chapter This chapter is organized as follows. In Section 5.2, we analyze the existence, uniqueness or multiplicity of the T-PageRank, and the convergence of the iterates (5.1) and (5.2). Then, in Section 5.3, we introduce a refinement of our model, inspired by the damping factor of the classical PageRank algorithm. Section 5.4 is devoted to the estimation of the *critical temperature*, that is, the temperature corresponding to the loss of the uniqueness of the T-PageRank, for some particular cases. We shall see that, even for very small or regular web graphs, the T-PageRank can have a complex behavior. We end this chapter by an experiment of the T-PageRank algorithm on a large-scale example in Section 5.5.

5.2 Existence, uniqueness, and approximation of the T-PageRank

5.2.1 Hypotheses

We work with *row* vectors throughout this chapter. We denote by $\Sigma = \{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \colon \sum_i \boldsymbol{x}_i = 1\}$ the *simplex*, i.e., the set of stochastic row vectors.

Let A be an $n \times n$ irreducible nonnegative matrix. For all *temperature* $T > 0$, and all $\boldsymbol{x} \in \Sigma$, let $P_T(\boldsymbol{x})$ be the irreducible stochastic matrix such that

$$P_T(\boldsymbol{x})_{ij} = \frac{A_{ij}\, g_T(\boldsymbol{x}_j)}{\sum_k A_{ik}\, g_T(\boldsymbol{x}_k)},$$

where $g_T \colon [0,1] \to \mathbb{R}_{>0}$ is a continuously differentiable map with $g_T(0) = 1$. We suppose moreover that g_T is increasing with $g'_T \colon [0,1] \to \mathbb{R}_{>0}$ and we make the following assumptions on the asymptotic behavior of g_T

$$\lim_{T \to 0} L_{\min}(g_T) = \infty, \tag{A_0}$$

$$\lim_{T \to \infty} L_{\max}(g_T) = 0, \tag{A_∞}$$

where $L_{\min}(g_T)$ and $L_{\max}(g_T)$ are defined as

$$L_{\min}(g_T) = \min_{x \in [0,1]} \frac{g'_T(x)}{g_T(x)} \quad \text{and} \quad L_{\max}(g_T) = \max_{x \in [0,1]} \frac{g'_T(x)}{g_T(x)}.$$

Remark 5.4. If we consider $G_T \colon [0,1] \to \mathbb{R}_{\geq 0} \colon x \mapsto \ln g_T(x)$, with $G_T(0) = 0$ and $G'_T(x) > 0$ for all $x \in [0,1]$, then $L_{\min}(g_T)$ and $L_{\max}(g_T)$ may simply be rewritten as $L_{\min}(g_T) = \min_{x \in [0,1]} G'_T(x)$ and $L_{\max}(g_T) = \max_{x \in [0,1]} G'_T(x)$. $\diamond$

Example 5.5. The hypotheses (A_0) and (A_∞) are satisfied in particular for $G_T(x) = E(x)\theta(T)$, where the energy function $E \colon [0,1] \to \mathbb{R}_{\geq 0}$ is continuously differentiable, independent of T and such that $E(0) = 0$ and $E'(x) > 0$ for all $x \in [0,1]$, and where $\theta \colon \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ is a map independent of x such that $\lim_{T \to 0} \theta(T) = \infty$ and $\lim_{T \to \infty} \theta(T) = 0$. This covers in particular the case, reminiscent of simulated annealing algorithms, where $g_T(x) = e^{E(x)/T}$. It also covers the case where $g_T(x) = (1 + \psi(x))^{\theta(T)}$, for a continuously differentiable map ψ with $\psi(0) = 0$ and $\psi'(x) > 0$ for all $x \in [0,1]$. $\diamond$

5.2.2 Preliminary results

The following elementary lemmas will be useful in the sequel.

Lemma 5.6. *For all $x, y \in [0,1]$,*

$$L_{\min}(g_T)\,(x-y)^+ \le \left(\ln \frac{g_T(x)}{g_T(y)}\right)^+ \le L_{\max}(g_T)\,(x-y)^+,$$

where for all $x \in \mathbb{R}$, $x^+ = \max\{0, x\}$. Moreover, if $x, y \neq 0$,

$$\left(\ln \frac{g_T(x)}{g_T(y)}\right)^+ \le L_{\max}(g_T)\,\left(\ln \frac{x}{y}\right)^+ .$$

Proof. Let $x, y \in [0,1]$. Then, using the fact that g_T is increasing and the logarithm is monotone,

$$(\ln g_T(x) - \ln g_T(y))^+ \ge \min_{a\in[0,1]} \frac{g_T'(a)}{g_T(a)}\,(x-y)^+ = L_{\min}(g_T)\,(x-y)^+,$$
$$(\ln g_T(x) - \ln g_T(y))^+ \le \max_{a\in[0,1]} \frac{g_T'(a)}{g_T(a)}\,(x-y)^+ = L_{\max}(g_T)\,(x-y)^+.$$

Moreover, $(x-y)^+ \le (\ln x - \ln y)^+$ if $x, y \in]0,1]$. □

Lemma 5.7. *Let $x, y \in [0,1]$. If $x > y$, then $\lim_{T\to 0} g_T(x)/g_T(y) = \infty$.*

Proof. This result follows directly from Lemma 5.6 and Assumption (A_0). □

Lemma 5.8. *The map g_T tends to the constant function equal to 1, uniformly in $[0,1]$, when T tends to infinity.*

Proof. For all $x \in [0,1]$, by Lemma 5.6,

$$\ln g_T(x) = \ln \frac{g_T(x)}{g_T(0)} \le L_{\max}(g_T)\,(x-0) \le L_{\max}(g_T).$$

Therefore, by Assumption (A_∞),

$$\lim_{T\to\infty} \sup_{x\in[0,1]} \ln g_T(x) \le \lim_{T\to\infty} L_{\max}(g_T) = 0.$$

□

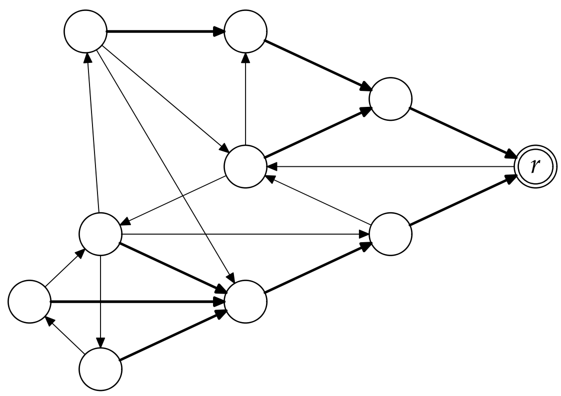


Figure 5.2: This graph has several spanning arborescences rooted at node r. Here, one of these spanning arborescence is represented with bold arrows.

From Lemma 5.8, the following corollary follows directly.

Corollary 5.9. *The limit* $\lim_{T\to\infty} P_T(\boldsymbol{x}) = \operatorname{diag}(A\mathbf{1})^{-1}A$ *uniformly for* $\boldsymbol{x} \in \Sigma$.

5.2.3 Fixed points and convergence of $\boldsymbol{u}_T$

When A is irreducible, for all $\boldsymbol{x} \in \Sigma$, the matrix $P_T(\boldsymbol{x})$ is irreducible, and we can define the map

$$\boldsymbol{u}_T \colon \Sigma \to \Sigma \colon \boldsymbol{x} \mapsto \boldsymbol{u}_T(\boldsymbol{x}),$$

that sends $\boldsymbol{x}$ to the unique invariant measure $\boldsymbol{u}_T(\boldsymbol{x})$ of $P_T(\boldsymbol{x})$. We use Tutte's Matrix Tree Theorem [115] in order to give an explicit expression for $\boldsymbol{u}_T(\boldsymbol{x})$ (for a proof of this theorem, see [27, Sec. 9.6]).

Let $M \in \mathbb{R}^{n\times n}_{\geq 0}$ be a nonnegative matrix and let $\mathcal{G}(M)$ be its directed graph. Let r be a node of $\mathcal{G}(M)$. A directed subgraph R of $\mathcal{G}(M)$ which contains no directed cycles and such that, for each node $i \neq r$, there is exactly one edge leaving i in R, is called a *spanning arborescence* rooted at r. The set of spanning arborescences of $\mathcal{G}(M)$ rooted at r is denoted by $\mathcal{A}(r)$. An example of spanning arborescence of a graph is given in Figure 5.2.

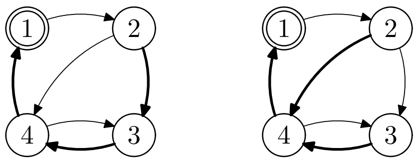


Figure 5.3: The graph considered in Example 5.11 has two spanning arborescences rooted at node 1*. These are represented here with bold arrows.*

Theorem 5.10 (Matrix Tree Theorem, Tutte [115])**.** *Let* $M \in \mathbb{R}^{n\times n}_{\geq 0}$ *be an irreducible stochastic matrix, and let* $\boldsymbol{u}$ *be its invariant measure. Then* $\boldsymbol{u} = \boldsymbol{v}/\sum_i \boldsymbol{v}_i$*, where for all index* r

$$\boldsymbol{v}_r = \sum_{R\in\mathcal{A}(r)} \prod_{(i,j)\in R} M_{ij}. \tag{5.3}$$

Example 5.11. Let us illustrate Tutte's Matrix Tree Theorem by an example. Consider the irreducible stochastic matrix

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}.$$

There exist two spanning arborescences rooted at node 1, one rooted at node 2, three rooted at node 3 and two rooted at node 4. By Matrix Tree Theorem, the invariant measure $\boldsymbol{u}$ of M is equal to $\boldsymbol{v}/\|\boldsymbol{v}\|_1$, where $\boldsymbol{v}$ can be computed as follows:

$$\begin{aligned}
\boldsymbol{v}_1 &= M_{23}M_{34}M_{41} + M_{24}M_{34}M_{41} = 0.5, \\
\boldsymbol{v}_2 &= M_{34}M_{41}M_{12} = 0.5, \\
\boldsymbol{v}_3 &= M_{41}M_{12}M_{23} + M_{12}M_{24}M_{43} + M_{12}M_{23}M_{43} = 0.65, \\
\boldsymbol{v}_4 &= M_{12}M_{24}M_{34} + M_{12}M_{23}M_{34} = 1.
\end{aligned}$$

For instance, the two spanning arborescence rooted at node 1 are represented in Figure 5.3. ◇

Lemma 5.12. *Assume that* A *is irreducible. Then, the invariant measure of* $P_T(\boldsymbol{x})$ *is given by*

$$\boldsymbol{u}_T(\boldsymbol{x}) = \frac{\boldsymbol{h}_T(\boldsymbol{x})}{\sum_k \boldsymbol{h}_T(\boldsymbol{x})_k},$$

where

$$\boldsymbol{h}_T(\boldsymbol{x})_r = \Big(\sum_k A_{rk}\, g_T(\boldsymbol{x}_k)\Big)\Big(\sum_{R\in\mathcal{A}(r)} \prod_{(i,j)\in R} A_{ij}\, g_T(\boldsymbol{x}_j)\Big). \tag{5.4}$$

Proof. Apply Theorem 5.10 to $P_T(\boldsymbol{x})$, and take $\boldsymbol{h}_T(\boldsymbol{x}) = \mu\, \boldsymbol{v}$, where $\boldsymbol{v}$ is given by (5.3) and $\mu = \prod_i \sum_k A_{ik} g_T(\boldsymbol{x}_k)$. □

The existence of fixed points for $\boldsymbol{u}_T$ is then proved using Brouwer's Fixed Point Theorem.

Proposition 5.13. *Assume that A is irreducible. The map $\boldsymbol{u}_T$ has at least one fixed point in* $\operatorname{int}(\Sigma)$*. Moreover, every fixed point of $\boldsymbol{u}_T$ is in* $\operatorname{int}(\Sigma)$*.*

Proof. By Lemma 5.12, the map $\boldsymbol{u}_T\colon \Sigma \to \Sigma$ is continuous, and therefore Brouwer's Fixed Point Theorem ensures the existence of at least one fixed point for $\boldsymbol{u}_T$. Moreover, since the invariant measure of an irreducible matrix is positive, and since $P_T(\boldsymbol{x})$ is irreducible, $\boldsymbol{u}_T$ maps Σ to $\operatorname{int}(\Sigma)$, and therefore every fixed point of $\boldsymbol{u}_T$ is in $\operatorname{int}(\Sigma)$. □

The following result concerns the uniqueness of the fixed point and the convergence of the orbits of $\boldsymbol{u}_T$. With Assumption (A_∞) satisfied, it shows that the map $\boldsymbol{u}_T$ has a unique fixed point and that all its orbits converge to this fixed point, for a sufficiently large temperature T. It can be proved using Nussbaum's Theorem 2.8, as we do it here, or, under the same hypotheses, using Krause's Theorem 2.6 (take $\|\boldsymbol{x}\| = \sum_i \boldsymbol{x}_i$ on $\mathbb{R}^n_{\geq 0}$).

Theorem 5.14. *Assume that A is irreducible. If $nL_{\max}(g_T) \leq 1$, the map $\boldsymbol{u}_T$ has a unique fixed point $\boldsymbol{x}_T$, which belongs to* $\operatorname{int}(\Sigma)$*. Moreover, all the orbits of $\boldsymbol{u}_T$ converge to this fixed point.*

Proof. Since g_T is increasing, $\boldsymbol{h}_T$ is an order-preserving map from $\mathbb{R}^n_{>0}$ to itself: $\boldsymbol{x} \leq \boldsymbol{y}$ implies $\boldsymbol{h}_T(\boldsymbol{x}) \leq \boldsymbol{h}_T(\boldsymbol{y})$. Now, let us show that $\boldsymbol{h}_T$ is *subhomogeneous* on $\operatorname{int}(\Sigma)$. Let $\boldsymbol{x} \in \operatorname{int}(\Sigma)$ and $0 < \lambda \leq 1$. Any entry of $\boldsymbol{h}_T(\boldsymbol{x})$ is a sum of positively weighted terms like

$$\prod_k g_T(\boldsymbol{x}_k)^{\gamma_k},$$

with $\sum_k \gamma_k = n$. By Lemma 5.6, for each $k \in \{1, \dots, n\}$,

$$\ln \frac{g_T(\boldsymbol{x}_k)}{g_T(\lambda \boldsymbol{x}_k)} \leq L_{\max}(g_T) \ln \frac{1}{\lambda}.$$

Therefore, if $nL_{\max}(g_T) \leq 1$, then $\lambda^{1/n} g_T(\boldsymbol{x}_k) \leq g_T(\lambda \boldsymbol{x}_k)$, and it follows that $\lambda \boldsymbol{h}_T(\boldsymbol{x}) \leq \boldsymbol{h}_T(\lambda \boldsymbol{x})$. Since $0 \leq \boldsymbol{h}_T(0)$, this shows that $\boldsymbol{h}_T$ is subhomogeneous on $\operatorname{int}(\Sigma)$.

Finally, let $\boldsymbol{x}_T$ be a fixed point of $\boldsymbol{u}_T$, by Proposition 5.13. The derivative $\boldsymbol{h}'_T(\boldsymbol{x})$ is a nonnegative continuous function of $\boldsymbol{x}$:

$$\begin{aligned} \frac{\partial \boldsymbol{h}_T(\boldsymbol{x})_r}{\partial \boldsymbol{x}_\ell} &= A_{r\ell}\, g'_T(\boldsymbol{x}_\ell) \Big(\sum_{R \in \mathcal{A}(r)} \prod_{(i,j) \in R} A_{ij}\, g_T(\boldsymbol{x}_j) \Big) \\ &\quad + \Big(\sum_k A_{rk}\, g_T(\boldsymbol{x}_k) \Big) \Big(\sum_{R \in \mathcal{A}(r)} m_{\ell,R} \frac{g'_T(\boldsymbol{x}_\ell)}{g_T(\boldsymbol{x}_\ell)} \prod_{(i,j) \in R} A_{ij}\, g_T(\boldsymbol{x}_j) \Big), \end{aligned}$$

where $m_{\ell,R} = |\{i \colon (i, \ell) \in R\}|$. Moreover, since g'_T takes positive values, $\partial \boldsymbol{h}_T(\boldsymbol{x})_r / \partial \boldsymbol{x}_\ell > 0$ as soon as $A_{r\ell} > 0$ or $m_{\ell,R} > 0$ for some $R \in \mathcal{A}(r)$. In particular, $m_{r,R} > 0$ for all $R \in \mathcal{A}(r)$. Since an irreducible matrix with positive diagonal is primitive, $\boldsymbol{h}'_T(\boldsymbol{x})$ is also primitive (see for instance Corollary 2.2.28 in [14]). Therefore, by Theorem 2.8, $\boldsymbol{x}_T$ is the unique fixed point of $\boldsymbol{u}_T$, and all the orbits of $\boldsymbol{u}_T$ converge to $\boldsymbol{x}_T$. □

5.2.4 Fixed points and convergence of $\boldsymbol{f}_T$

We now consider the map

$$\boldsymbol{f}_T \colon \Sigma \to \Sigma \colon \boldsymbol{x} \mapsto \boldsymbol{x} P_T(\boldsymbol{x}).$$

The following result shows that if T is sufficiently large, the fixed point of the map $\boldsymbol{u}_T$ can be computed by iterating $\boldsymbol{f}_T$.

Theorem 5.15. *The fixed points of $\boldsymbol{u}_T$ and $\boldsymbol{f}_T$ are the same. If A is primitive, then, for T sufficiently large, all the orbits of $\boldsymbol{f}_T$ converge to the fixed point $\boldsymbol{x}_T$ of $\boldsymbol{u}_T$.*

Proof. Clearly, $\boldsymbol{f}_T$ and $\boldsymbol{u}_T$ have the same fixed points. Suppose now that A is primitive, and let us show then that, for T is sufficiently large, $\boldsymbol{f}_T$ is a contraction for some particular norm. For every $\boldsymbol{x}, \boldsymbol{y} \in \Sigma$ and for any norm $\|\cdot\|$,

$$\|\boldsymbol{f}_T(\boldsymbol{x}) - \boldsymbol{f}_T(\boldsymbol{y})\| \leq \sup_{\boldsymbol{v}\in\Sigma} \|(\boldsymbol{x}-\boldsymbol{y})\, \boldsymbol{f}'_T(\boldsymbol{v})\| .$$

The derivative of $\boldsymbol{f}_T$ satisfies

$$\frac{\partial \boldsymbol{f}_T(\boldsymbol{v})_j}{\partial \boldsymbol{v}_\ell} = P_T(\boldsymbol{v})_{\ell j} + \sum_i \left(\delta_{\ell j} - P_T(\boldsymbol{v})_{ij}\right) \frac{A_{i\ell}\, \boldsymbol{v}_i\, g'_T(\boldsymbol{v}_\ell)}{\sum_k A_{ik}\, g_T(\boldsymbol{v}_k)},$$

where $\delta_{\ell j}$ denotes the Kronecker delta. It follows from Assumption (A_∞) and Corollary 5.9 that

$$\lim_{T\to\infty} \boldsymbol{f}'_T(\boldsymbol{v}) = \lim_{T\to\infty} P_T(\boldsymbol{v}) = \mathrm{diag}(A\mathbf{1})^{-1}A,$$

uniformly for $\boldsymbol{v} \in \Sigma$. Let $P = \mathrm{diag}(A\mathbf{1})^{-1}A$, and let $\mathcal{S} = \{\boldsymbol{z} \in \mathbb{R}^n \colon \sum_k \boldsymbol{z}_k = 0\}$ be the space of row vectors orthogonal to the vector $\mathbf{1}$. The map $\boldsymbol{x} \mapsto \boldsymbol{x}P$ preserves the space $\mathcal{S}$, because $P\mathbf{1} = \mathbf{1}$. Let $P_{\mathcal{S}}$ denote the restriction of the map $\boldsymbol{x} \mapsto \boldsymbol{x}P$ to $\mathcal{S}$. Since the matrix P is primitive, its Perron root, $\rho(P)$, is a simple eigenvalue and all the other eigenvalues of P have a strictly smaller modulus by Perron–Frobenius Theorem 2.1. Moreover, the left Perron vector $\boldsymbol{u}$ of P does not belong to the space $\mathcal{S}$ because it must have positive entries, contradicting $\sum_k \boldsymbol{u}_k = 0$. We deduce that $\rho(P_{\mathcal{S}}) < \rho(P) = 1$. It follows that there exists a norm $\|\cdot\|$ such that $|||P_{\mathcal{S}}||| < 1$, where $|||\cdot|||$ is the matrix norm induced by $\|\cdot\|$ (see for instance Lemma 5.6.10 in [61]). Therefore, since $\boldsymbol{f}'_T(\boldsymbol{v})$ tends uniformly to P for $\boldsymbol{v} \in \Sigma$ when T tends to ∞,

$$\lim_{T\to\infty} \sup_{\boldsymbol{v}\in\Sigma} |||\boldsymbol{f}'_T(\boldsymbol{v})_{\mathcal{S}}||| = |||P_{\mathcal{S}}||| < 1,$$

where $\boldsymbol{f}'_T(\boldsymbol{v})_{\mathcal{S}}$ denotes the restriction of $\boldsymbol{f}'_T(\boldsymbol{v})$ on $\mathcal{S}$. It follows that, for all $\alpha \in \,]|||P_{\mathcal{S}}|||, 1[$, there exists T_α such that for all $T > T_\alpha$,

$$\|\boldsymbol{f}_T(\boldsymbol{x}) - \boldsymbol{f}_T(\boldsymbol{y})\| \leq \|\boldsymbol{x}-\boldsymbol{y}\| \sup_{\boldsymbol{v}\in\Sigma} |||\boldsymbol{f}'_T(\boldsymbol{v})_{\mathcal{S}}||| \leq \alpha \|\boldsymbol{x}-\boldsymbol{y}\| .$$

Hence, for such temperature T, by Banach's Fixed Point Theorem, $\boldsymbol{f}_T$ has a unique fixed point and every orbit of $\boldsymbol{f}_T$ converges to this fixed point. $\square$

Remark 5.16. Note that the maps $\boldsymbol{f}_T$ and $\boldsymbol{u}_T$ have the same fixed points but their iterates do not converge under the same conditions. In particular, for the convergence of the orbits of $\boldsymbol{f}_T$, the primitivity of A cannot be dispensed with. Let for instance $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T > 0$. Then $P_T(\boldsymbol{x}) = A$ for all $\boldsymbol{x} \in \Sigma$. The only fixed point for $\boldsymbol{f}_T$ and $\boldsymbol{u}_T$ is $\left(\frac{1}{2} \quad \frac{1}{2}\right)$. Moreover, for every initial vector, the iterates of $\boldsymbol{u}_T$ converge in one step, since $\boldsymbol{u}_T(\boldsymbol{x}) = \left(\frac{1}{2} \quad \frac{1}{2}\right)$ for every $\boldsymbol{x} \in \Sigma$. On the other hand, the iterates of $\boldsymbol{f}_T$ do not converge in general, since $\boldsymbol{f}_T^k(\boldsymbol{x})$ oscillates when k tends to infinity, unless $\boldsymbol{x} = \left(\frac{1}{2} \quad \frac{1}{2}\right)$. ◇

For positive matrices $M \in \mathbb{R}^{n\times n}_{>0}$, let us now derive another convergence criterion, depending on Birkhoff's coefficient of ergodicity. Recall that Hilbert's projective metric d_H is defined as

$$d_H\colon \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{\geq 0}\colon (\boldsymbol{x}, \boldsymbol{y}) \mapsto \max_{i,j} \ln \frac{\boldsymbol{x}_i \boldsymbol{y}_j}{\boldsymbol{y}_i \boldsymbol{x}_j}.$$

The *coefficient of ergodicity* τ_B, also known as Birkhoff's contraction coefficient, is defined for a nonnegative matrix M having no zero column as

$$\tau_B(M) = \sup_{\substack{\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n_{>0} \\ \boldsymbol{x} \neq \lambda \boldsymbol{y}}} \frac{d_H(\boldsymbol{x}M, \boldsymbol{y}M)}{d_H(\boldsymbol{x}, \boldsymbol{y})}.$$

We always have $0 \leq \tau_B(M) \leq 1$. Moreover, $\tau_B(M) < 1$ if and only if M is positive [110]. It can also be easily proved that $\tau_B(M) = \tau_B(DM)$ for every diagonal matrix D with positive diagonal elements. We also define an *induced projective metric* between two positive matrices as

$$d_H\colon \mathbb{R}^{n\times n}_{>0} \times \mathbb{R}^{n\times n}_{>0} \to \mathbb{R}_{\geq 0}\colon (M, N) \mapsto \sup_{z \in \mathbb{R}^n_{>0}} d_H(zM, zN).$$

Lemma 5.17. *Assume that A is positive. Then, for any $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{int}(\Sigma)$,*

$$d_H(P_T(\boldsymbol{x}), P_T(\boldsymbol{y})) \leq 2L_{\max}(g_T)\, d_H(\boldsymbol{x}, \boldsymbol{y}).$$

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{int}(\Sigma)$ be fixed. Let us define $\alpha = \max_{i,j} \frac{P_T(\boldsymbol{x})_{ij}}{P_T(\boldsymbol{y})_{ij}}$ and

$\beta = \min_{i,j} \frac{P_T(\boldsymbol{x})_{ij}}{P_T(\boldsymbol{y})_{ij}}$. By definition,

$$\begin{aligned} d_H(P_T(\boldsymbol{x}), P_T(\boldsymbol{y})) &= \sup_{\boldsymbol{z}\in\mathbb{R}^n_{>0}} \max_{i,j} \ln \frac{(\boldsymbol{z}P_T(\boldsymbol{x}))_i}{(\boldsymbol{z}P_T(\boldsymbol{y}))_i} \frac{(\boldsymbol{z}P_T(\boldsymbol{y}))_j}{(\boldsymbol{z}P_T(\boldsymbol{x}))_j} \\ &\leq \sup_{\boldsymbol{z}\in\mathbb{R}^n_{>0}} \max_{i,j} \ln \frac{(\alpha\boldsymbol{z}P_T(\boldsymbol{y}))_i}{(\boldsymbol{z}P_T(\boldsymbol{y}))_i} \frac{(\boldsymbol{z}P_T(\boldsymbol{y}))_j}{(\beta\boldsymbol{z}P_T(\boldsymbol{y}))_j} = \ln \frac{\alpha}{\beta}. \end{aligned}$$

Moreover, by Lemma 5.6,

$$\begin{aligned} \ln \alpha &= \max_{i,j} \ln \left(\frac{A_{ij}\, g_T(\boldsymbol{x}_j)}{\sum_k A_{ik}\, g_T(\boldsymbol{x}_k)} \frac{\sum_k A_{ik}\, g_T(\boldsymbol{y}_k)}{A_{ij}\, g_T(\boldsymbol{y}_j)} \right) \\ &\leq \max_{j,k} \left(\ln \frac{g_T(\boldsymbol{x}_j)}{g_T(\boldsymbol{y}_j)} + \ln \frac{g_T(\boldsymbol{y}_k)}{g_T(\boldsymbol{x}_k)} \right) \\ &\leq L_{\max}(g_T) \left(\max_j \left(\ln \frac{\boldsymbol{x}_j}{\boldsymbol{y}_j} \right)^+ + \max_k \left(\ln \frac{\boldsymbol{y}_k}{\boldsymbol{x}_k} \right)^+ \right) \\ &= L_{\max}(g_T) \left(\left(\max_j \ln \frac{\boldsymbol{x}_j}{\boldsymbol{y}_j} \right)^+ + \left(\max_k \ln \frac{\boldsymbol{y}_k}{\boldsymbol{x}_k} \right)^+ \right) \\ &= L_{\max}(g_T)\, d_H(\boldsymbol{x}, \boldsymbol{y}), \end{aligned}$$

since $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{int}(\Sigma)$ implies $\max_j \ln \frac{\boldsymbol{x}_j}{\boldsymbol{y}_j} \geq 0$ and $\max_k \ln \frac{\boldsymbol{y}_k}{\boldsymbol{x}_k} \geq 0$. We get similarly $-\ln \beta \leq L_{\max}(g_T)\, d_H(\boldsymbol{x}, \boldsymbol{y})$. □

Proposition 5.18. *Assume that A is positive. If $2L_{\max}(g_T) < 1 - \tau_B(A)$, then $\boldsymbol{f}_T$ has a unique fixed point $\boldsymbol{x}_T \in \operatorname{int}(\Sigma)$ and all the orbits of $\boldsymbol{f}_T$ converge to this fixed point.*

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{int}(\Sigma)$. By Lemma 5.17, and since $P_T(\boldsymbol{x})$ is a diagonal scaling of A,

$$\begin{aligned} d_H(\boldsymbol{f}_T(\boldsymbol{x}), \boldsymbol{f}_T(\boldsymbol{y})) &\leq d_H(\boldsymbol{x}P_T(\boldsymbol{x}), \boldsymbol{y}P_T(\boldsymbol{x})) + d_H(\boldsymbol{y}P_T(\boldsymbol{x}), \boldsymbol{y}P_T(\boldsymbol{y})) \\ &\leq \tau_B(P_T(\boldsymbol{x}))\, d_H(\boldsymbol{x}, \boldsymbol{y}) + d_H(P_T(\boldsymbol{x}), P_T(\boldsymbol{y})) \\ &\leq \big(\tau_B(A) + 2L_{\max}(g_T)\big)\, d_H(\boldsymbol{x}, \boldsymbol{y}). \end{aligned}$$

Therefore, if $2L_{\max}(g_T) < 1 - \tau_B(A)$, then $\boldsymbol{f}_T$ is a contraction on $\operatorname{int}(\Sigma)$ with respect to the distance d_H. Since $(\operatorname{int}(\Sigma), d_H)$ is a complete metric space, by Banach's Fixed Point Theorem, $\boldsymbol{f}_T$ has a unique fixed point $\boldsymbol{x}_T \in \operatorname{int}(\Sigma)$ and all the orbits of $\boldsymbol{f}_T$ converge to this fixed point. $\square$

5.2.5 Existence of multiple fixed points of $\boldsymbol{u}_T$ and $\boldsymbol{f}_T$

Theorems 5.14 and 5.15 show that for a sufficiently large temperature T, the maps $\boldsymbol{u}_T$ and $\boldsymbol{f}_T$ have a unique fixed point. We can naturally wonder about the uniqueness of the fixed point of these maps for small T: we show that, at least when A is positive, multiple fixed points always exist.

Theorem 5.19. *Assume that A is irreducible and that the first column of A is positive. Then, for all $0 < \varepsilon < \frac{1}{2}$, there exists T_ε such that for $T \leq T_\varepsilon$, the map $\boldsymbol{u}_T$ has a fixed point in $\Sigma_\varepsilon = \{\boldsymbol{x} \in \Sigma, \boldsymbol{x}_1 \geq 1 - \varepsilon\}$.*

Assume now that A is irreducible with $A_{11} > 0$ only, and that there exists $\varepsilon_n > 0$, independent of T, such that $g_T(\varepsilon_n)^{n-1} \leq g_T(1 - \varepsilon_n)$ for all $T > 0$. Then, for all $0 < \varepsilon < \varepsilon_n$, there exists T_ε such that for $T \leq T_\varepsilon$, the map $\boldsymbol{u}_T$ has a fixed point in $\Sigma_\varepsilon = \{\boldsymbol{x} \in \Sigma, \boldsymbol{x}_1 \geq 1 - \varepsilon\}$.

Proof. Let k be the number of indices $i \neq 1$ such that $A_{i1} > 0$. Let $0 < \varepsilon < \frac{1}{2}$, and let $\boldsymbol{x} \in \Sigma_\varepsilon$. By the irreducibility of A, there exists a spanning arborescence R rooted at 1, containing all the k arcs $(i, 1)$ with $i \neq 1$ and $A_{i1} > 0$. Hence

$$\boldsymbol{h}_T(\boldsymbol{x})_1 \geq A_{11} g_T(\boldsymbol{x}_1) \prod_{(i,j) \in R} A_{ij} g_T(\boldsymbol{x}_j) \geq \alpha g_T(\boldsymbol{x}_1)^{k+1} \geq \alpha g_T(\boldsymbol{x}_1)^k g_T(1 - \varepsilon),$$

where $\alpha = A_{11} \prod_{(i,j) \in R} A_{ij} > 0$. Let $r \neq 1$. If $A_{r1} \neq 0$, then a spanning arborescence rooted at r can have at most $k - 1$ arcs $(i, 1)$ with $A_{i1} > 0$, whereas it can have at most k arcs $(i, 1)$ with $A_{i1} > 0$ in general. Hence, in all cases, $\boldsymbol{h}_T(\boldsymbol{x})_r$ is a sum of positively weighted terms like $\prod_\ell g_T(\boldsymbol{x}_\ell)^{\gamma_\ell}$ with $\sum_\ell \gamma_\ell = n$ and $\gamma_1 \leq k$. This implies that, for $r \neq 1$,

$$\boldsymbol{h}_T(\boldsymbol{x})_r \leq \beta g_T(\boldsymbol{x}_1)^k g_T(\varepsilon)^{n-k},$$

for some positive constant β. Therefore,

$$\boldsymbol{u}_T(\boldsymbol{x})_1 = \frac{1}{1 + \sum_{r \neq 1} \frac{\boldsymbol{h}_T(\boldsymbol{x})_r}{\boldsymbol{h}_T(\boldsymbol{x})_1}} \geq \frac{1}{1 + (n-1) \frac{\beta}{\alpha} \frac{g_T(\varepsilon)^{n-k}}{g_T(1-\varepsilon)}}.$$

If the first column of A is positive, then $k = n-1$. By Lemma 5.6,

$$\ln \frac{g_T(1-\varepsilon)}{g_T(\varepsilon)} \geq L_{\min}(g_T)\,(1-2\varepsilon).$$

Therefore, if $L_{\min}(g_T)(1-2\varepsilon) \geq \ln \frac{(n-1)\beta}{\alpha}\frac{1-\varepsilon}{\varepsilon}$, we get $\boldsymbol{u}_T(\boldsymbol{x})_1 \geq 1-\varepsilon$. This shows that $\boldsymbol{u}_T(\Sigma_\varepsilon) \subset \Sigma_\varepsilon$. By Brouwer's Fixed Point Theorem, the continuous map $\boldsymbol{u}_T$ has therefore at least one fixed point in Σ_ε.

Now, suppose we know only that $A_{11} > 0$, but there exists $\varepsilon_n > 0$ such that $g_T(\varepsilon_n)^{n-1} \leq g_T(1-\varepsilon_n)$ for all $T > 0$. The map $\varphi_T \colon \varepsilon \mapsto (n-1)\ln g_T(\varepsilon) - \ln g_T(1-\varepsilon)$ is increasing and its derivative satisfies $\varphi'_T(\varepsilon) \geq nL_{\min}(g_T)$. Let $0 < \varepsilon < \varepsilon_n$. We have $\varphi_T(\varepsilon_n) - \varphi_T(\varepsilon) \geq nL_{\min}(g_T)(\varepsilon_n - \varepsilon)$. Moreover, $k \geq 1$ by irreducibility of A and $\varphi_T(\varepsilon_n) \leq 0$, hence

$$\ln \frac{g_T(1-\varepsilon)}{g_T(\varepsilon)^{n-k}} \geq \ln \frac{g_T(1-\varepsilon)}{g_T(\varepsilon)^{n-1}} = -\varphi_T(\varepsilon) \geq nL_{\min}(g_T)(\varepsilon_n - \varepsilon).$$

Therefore, if $nL_{\min}(g_T)(\varepsilon_n - \varepsilon) \geq \ln \frac{(n-1)\beta}{\alpha}\frac{1-\varepsilon}{\varepsilon}$, we get $\boldsymbol{u}_T(\boldsymbol{x})_1 \geq 1-\varepsilon$. The result follows by the same argument as above. □

Remark 5.20. When $g_T(x) = e^{E(x)/T}$ for some increasing energy E, then ε_n satisfies the condition $g_T(\varepsilon_n)^{n-1} \leq g_T(1-\varepsilon_n)$ for all $T > 0$ if and only if $(n-1)E(\varepsilon_n) \leq E(1-\varepsilon_n)$, which holds for some $0 < \varepsilon_n < 1$, since $E(0) = 0$ and $E(1) > 0$. ◇

Corollary 5.21. *If A is positive, then, for $T > 0$ sufficiently small, the map $\boldsymbol{u}_T$ has several fixed points in Σ.*

Example 5.22. If A is not positive, the existence of several fixed points for small T is not insured. Indeed, we shall see in Remark 5.35 that for $A = \left(\begin{smallmatrix} 1 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ and $g_T(x) = e^{x/T}$, the fixed point of $\boldsymbol{u}_T$ and $\boldsymbol{f}_T$ is unique for each $T > 0$. ◇

5.3 Refinement of the model

In the present section, we study a more general model, which includes a *damping factor* $0 < c < 1$, as in the standard definition of Google's PageRank (see Section 2.2.1, in which the web surfer either jumps to the search engine with probability $1-c$ or moves to a neighboring page with probability c. The presence of a damping factor yields a more realistic model of the web surfer's walk. Moreover, it allows one to deal with reducible matrices, and it improves the convergence speed of iterative methods.

Let A be a $n \times n$ nonnegative matrix with no zero row and let $\boldsymbol{z} \in \mathbb{R}^n_{>0}$ be a *personalization vector*. For all temperature $0 < T < \infty$, let us define as previously $g_T \colon [0,1] \to \mathbb{R}_{>0}$ as a continuously differentiable and increasing map, with $g_T(0) = 1$ and $g'_T \colon [0,1] \to \mathbb{R}_{>0}$. Suppose that Assumptions (A_0) and (A_∞) are satisfied. For a temperature $T = \infty$, let us also define $g_\infty(x) = 1$ for all $x \in [0,1]$.

For any two temperatures $0 < T_1, T_2 \leq \infty$, and for all $\boldsymbol{x} \in \Sigma$, we can consider the positive transition matrix $P_{T_1,T_2,c}(\boldsymbol{x})$ defined as

$$P_{T_1,T_2,c}(\boldsymbol{x})_{ij} = c\,\frac{A_{ij}\, g_{T_1}(\boldsymbol{x}_j)}{\sum_k A_{ik}\, g_{T_1}(\boldsymbol{x}_k)} + (1-c)\,\frac{\boldsymbol{z}_j\, g_{T_2}(\boldsymbol{x}_j)}{\sum_k \boldsymbol{z}_k\, g_{T_2}(\boldsymbol{x}_k)}. \tag{5.5}$$

Remark 5.23. For simplicity, we consider the same family of weight functions g_T for the first and the second term of $P_{T_1,T_2,c}(\boldsymbol{x})$. Note however that the results of this section remain true if two families g_{T_1} and $\tilde{g}_{T_2}$ are considered. $\diamond$

Remark 5.24. Suppose $T_1 = \infty$, $T_2 < \infty$ and $0 < c < 1$, and let $\boldsymbol{x}$ be the current ranking vector. Then $P_{T_1,T_2,c}(\boldsymbol{x})$ is the transition matrix of the following random walk on the graph. At each step of his walk, either, with probability c, the web surfer draws the next page uniformly among the pages referenced by his current page. Or, with probability $1-c$, he refers to the Web search engine, and therefore preferentially chooses for the next page a web page with a good ranking. $\diamond$

The maps $\boldsymbol{u}_{T_1,T_2,c}$ and $\boldsymbol{f}_{T_1,T_2,c}$ are defined as previously: $\boldsymbol{u}_{T_1,T_2,c}(\boldsymbol{x})$ is the unique invariant measure of $P_{T_1,T_2,c}(\boldsymbol{x})$ and $\boldsymbol{f}_{T_1,T_2,c}(\boldsymbol{x}) = \boldsymbol{x}P_{T_1,T_2,c}(\boldsymbol{x})$. Theorem 5.14 about the uniqueness of the fixed point of $\boldsymbol{u}_T$ can be adapted in the following way.

Proposition 5.25. *If* $nL_{\max}(g_{T_1}) + (n-1)L_{\max}(g_{T_2}) \leq 1$*, the map* $\boldsymbol{u}_{T_1,T_2,c}$ *has a unique fixed point* $\boldsymbol{x}_{T_1,T_2,c}$ *in* Σ*. Moreover, all the orbits of* $\boldsymbol{u}_{T_1,T_2,c}$ *converge to the fixed point* $\boldsymbol{x}_{T_1,T_2,c}$*.*

Proof. If $T_1 = T_2 = \infty$, the result follows directly from Perron–Frobenius theory. Let us therefore suppose that $T_1 < \infty$ or $T_2 < \infty$. For every $\boldsymbol{x} \in \Sigma$, by Theorem 5.10, $\boldsymbol{u}_{T_1,T_2,c} = \boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x}) / \sum_k \boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})_k$, where

$$\boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})_r = \Big(\sum_k A_{rk}\, g_{T_1}(\boldsymbol{x}_k)\Big)\Big(\sum_{R\in\mathcal{A}(r)} \prod_{(i,j)\in R} W(\boldsymbol{x})_{ij}\Big),$$

with $W(\boldsymbol{x})_{ij} = \sum_k \big(c\, A_{ij}\, g_{T_1}(\boldsymbol{x}_j)\, \boldsymbol{z}_k\, g_{T_2}(\boldsymbol{x}_k) + (1-c)\, A_{ik}\, g_{T_1}(\boldsymbol{x}_k)\, \boldsymbol{z}_j\, g_{T_2}(\boldsymbol{x}_j)\big)$.

Since g_{T_1} and g_{T_2} are nondecreasing, $\boldsymbol{h}_{T_1,T_2,c}$ is an order-preserving map. Moreover, assume that $nL_{\max}(g_{T_1}) + (n-1)L_{\max}(g_{T_1}) \le 1$. Then, as in the proof of Theorem 5.14, $\boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})$ is shown to be subhomogeneous on $\mathrm{int}(\Sigma)$.

Finally, the derivative $\boldsymbol{h}'_{T_1,T_2,c}(\boldsymbol{x})$ is a nonnegative continuous function:

$$\begin{aligned}\frac{\partial \boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})_r}{\partial \boldsymbol{x}_\ell} &= A_{r\ell}\, g'_{T_1}(\boldsymbol{x}_\ell)\Big(\sum_{R\in\mathcal{A}(r)} \prod_{(i,j)\in R} W(\boldsymbol{x})_{ij}\Big)\\ &+ \Big(\sum_k A_{rk}\, g_{T_1}(\boldsymbol{x}_k)\Big)\Big(\sum_{R\in\mathcal{A}(r)} \Big(\prod_{(i,j)\in R} W(\boldsymbol{x})_{ij}\Big)\Big(\sum_{(i,j)\in R} \frac{\partial W(\boldsymbol{x})_{ij}/\partial \boldsymbol{x}_\ell}{W(\boldsymbol{x})_{ij}}\Big)\Big),\end{aligned}$$

where

$$\begin{aligned}\frac{\partial W(\boldsymbol{x})_{ij}}{\partial \boldsymbol{x}_\ell} &= c\, A_{ij}\, g_{T_1}(\boldsymbol{x}_j)\, \boldsymbol{z}_\ell\, g'_{T_2}(\boldsymbol{x}_\ell) + (1-c)\, A_{i\ell}\, g'_{T_1}(\boldsymbol{x}_\ell)\, \boldsymbol{z}_j\, g_{T_2}(\boldsymbol{x}_j)\\ &+ \delta_{\ell j} \sum_k \big(c\, A_{i\ell}\, g'_{T_1}(\boldsymbol{x}_\ell)\, \boldsymbol{z}_k\, g_{T_2}(\boldsymbol{x}_k) + (1-c)\, A_{ik}\, g_{T_1}(\boldsymbol{x}_k)\, \boldsymbol{z}_\ell\, g'_{T_2}(\boldsymbol{x}_\ell)\big).\end{aligned}$$

Let us now prove that $\boldsymbol{h}'_{T_1,T_2,c}(\boldsymbol{x})$ is a positive matrix for every $\boldsymbol{x}$. Suppose first that $T_2 < \infty$. Then $g'_{T_2}(\boldsymbol{x}_\ell) > 0$, and there exists a spanning arborescence $R \in \mathcal{A}(r)$ and a node i such that $(i,\ell) \in R$, since $\mathcal{G}(P_{T_1,T_2,c}(\boldsymbol{x}))$ is the complete graph. It follows that, for this R,

$$\begin{aligned}\sum_{(i,j)\in R} \frac{\partial W(\boldsymbol{x})_{ij}/\partial \boldsymbol{x}_\ell}{W(\boldsymbol{x})_{ij}} &\ge \frac{\partial W(\boldsymbol{x})_{i\ell}/\partial \boldsymbol{x}_\ell}{W(\boldsymbol{x})_{i\ell}}\\ &\ge \frac{\sum_k (1-c)\, A_{ik}\, g_{T_1}(\boldsymbol{x}_k)\, \boldsymbol{z}_\ell\, g'_{T_2}(\boldsymbol{x}_\ell)}{W(\boldsymbol{x})_{i\ell}} > 0,\end{aligned}$$

and hence $\partial \boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})_r / \partial \boldsymbol{x}_\ell > 0$. Now, suppose that $T_2 = \infty$ and $T_1 < \infty$. Then we can suppose without loss of generality that A has no zero column (see Remark 5.26 below). Either $A_{r\ell} > 0$, and therefore $\partial \boldsymbol{h}_{T_1}(\boldsymbol{x})_r / \partial \boldsymbol{x}_\ell > 0$. Or there exists $i \neq r$ such that $A_{i\ell} > 0$, and for all $R \in \mathcal{A}(r)$, there exists j such that $(i,j) \in R$, that is

$$\frac{\partial W(\boldsymbol{x})_{ij}}{\partial \boldsymbol{x}_\ell} \geq (1-c)\, A_{i\ell}\, \boldsymbol{z}_j\, g'_{T_1}(\boldsymbol{x}_\ell) > 0,$$

and hence $\partial \boldsymbol{h}_{T_1,T_2,c}(\boldsymbol{x})_r / \partial \boldsymbol{x}_\ell > 0$.

Since Brouwer's Fixed Point Theorem ensures the existence of at least one fixed point $\boldsymbol{x}_{T_1,T_2,c} \in \operatorname{int}(\Sigma)$ for the continuous map $\boldsymbol{u}_{T_1,T_2,c}$ which sends Σ to $\operatorname{int}(\Sigma)$, by Theorem 2.8, this fixed point $\boldsymbol{x}_{T_1,T_2,c}$ is the unique fixed point of $\boldsymbol{u}_{T_1,T_2,c}$, and all the orbits of $\boldsymbol{u}_{T_1,T_2,c}$ converge to $\boldsymbol{x}_{T_1,T_2,c}$. □

Remark 5.26. If $T_2 = \infty$ and the matrix A has a zero column, the problem can be reduced to a problem of smaller dimension with a matrix with no zero column. Indeed, suppose the n^{th} column of A is zero. Then

$$\begin{aligned}(\boldsymbol{u}_{T_1,\infty,c}(\boldsymbol{x})_1 \;\cdots\; \boldsymbol{u}_{T_1,\infty,c}(\boldsymbol{x})_{n-1}) &= \frac{\sum_k \boldsymbol{z}_k - (1-c)\boldsymbol{z}_n}{\sum_k \boldsymbol{z}_k}\, \tilde{\boldsymbol{u}}_{T_1,\infty,c}(\tilde{\boldsymbol{x}}),\\ \boldsymbol{u}_{T_1,\infty,c}(\boldsymbol{x})_n &= (1-c)\frac{\boldsymbol{z}_n}{\sum_k \boldsymbol{z}_k},\end{aligned}$$

where $\tilde{\boldsymbol{x}} = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_{n-1})$ and $\tilde{\boldsymbol{u}}_{T_1,\infty,c}(\tilde{\boldsymbol{x}})$ is the invariant measure of a matrix $\tilde{P}_{T_1,\infty,c}(\tilde{\boldsymbol{x}})$, with $\tilde{A}$ the principal submatrix of A corresponding to the indices $1, \ldots, n-1$, and $\tilde{\boldsymbol{d}}$ some positive vector of length $n-1$. ◇

The following adaptations of Theorem 5.15 and Proposition 5.18 about the uniqueness of the fixed point of $\boldsymbol{f}_T$ are quite direct.

Proposition 5.27. *The fixed points of $\boldsymbol{u}_{T_1,T_2,c}$ and $\boldsymbol{f}_{T_1,T_2,c}$ are the same. Moreover, for T_1 and T_2 sufficiently large, all the orbits of $\boldsymbol{f}_{T_1,T_2,c}$ converge to the fixed point $\boldsymbol{x}_{T_1,T_2,c}$ of $\boldsymbol{u}_{T_1,T_2,c}$.*

Proposition 5.28. *Assume that A is positive. If $2(L_{\max}(g_{T_1}) + L_{\max}(g_{T_2})) < 1 - \tau_B(A)$, then $\boldsymbol{f}_{T_1,T_2,c}$ has a unique fixed point $\boldsymbol{x}_{T_1,T_2,c} \in \operatorname{int}(\Sigma)$ and all the orbits of $\boldsymbol{f}_{T_1,T_2,c}$ converge to the fixed point $\boldsymbol{x}_{T_1,T_2,c}$.*

We next show that the map $f_{T_1,T_2,c}$ has multiple fixed points if either T_1 or T_2 is sufficiently small. Of course, this requires the damping factor to give enough weight to the terms corresponding to the small temperature in equation (5.5).

Proposition 5.29. *For all $\frac{1}{2} < \alpha < 1$, the map $f_{T_1,T_2,c}$ has a fixed point in $\Sigma_\alpha = \{x \in \Sigma\colon x_1 \geq \alpha\}$ if one of the two following conditions hold:*

(a) *T_1 is sufficiently small, $c > \alpha$, and the first column of A is positive,*

(b) *T_2 is sufficiently small and $1 - c > \alpha$.*

Proof. Let $\frac{1}{2} < \alpha < 1$ and let $x \in \Sigma_\alpha$. For all $k \neq 1$, since $x_k \leq 1 - x_1 < \frac{1}{2}$, we have by Lemma 5.6

$$\frac{g_T(x_k)}{g_T(x_1)} \leq e^{-L_{\min}(g_T)\,(x_1 - x_k)} \leq e^{-L_{\min}(g_T)\,(2\alpha - 1)},$$

for any $0 < T \leq \infty$. It follows that

$$\begin{aligned} f_{T_1,T_2,c}(x)_1 &= c \sum_i \frac{A_{i1} x_i}{A_{i1} + \sum_{k\neq 1} A_{ik} \frac{g_{T_1}(x_k)}{g_{T_1}(x_1)}} + (1-c) \frac{z_1}{z_1 + \sum_{k\neq 1} z_k \frac{g_{T_2}(x_k)}{g_{T_2}(x_1)}} \\ &\geq c \sum_i \frac{A_{i1} x_i}{A_{i1} + \sum_{k\neq 1} A_{ik} e^{-L_{\min}(g_{T_1})\,(2\alpha-1)}} \\ &\quad + (1-c) \frac{z_1}{z_1 + \sum_{k\neq 1} z_k e^{-L_{\min}(g_{T_2})\,(2\alpha-1)}}. \end{aligned}$$

In the first case, suppose that the first column of A is positive and that $c > \alpha$, and let $\mu = \max_i \sum_{k\neq 1} \frac{A_{ik}}{A_{i1}}$. If T_1 is small enough to have $L_{\min}(g_{T_1})\,(2\alpha - 1) \geq \ln \frac{\mu\alpha}{c-\alpha}$, then

$$f_{T_1,T_2,c}(x)_1 \geq \frac{c}{1 + \mu e^{-L_{\min}(g_{T_1})\,(2\alpha-1)}} \geq \alpha.$$

In the second case, suppose that $1 - c > \alpha$ and let $\nu = \sum_{k\neq 1} \frac{z_k}{z_1}$. If T_2 is small enough to have $L_{\min}(g_{T_2})\,(2\alpha - 1) \geq \ln \frac{\nu\alpha}{1-c-\alpha}$, then

$$f_{T_1,T_2,c}(x)_1 \geq \frac{1-c}{1 + \nu e^{-L_{\min}(g_{T_2})\,(2\alpha-1)}} \geq \alpha.$$

In both cases, $f_{T_1,T_2,c}(x) \in \Sigma_\alpha$. Therefore, by Brouwer's Fixed Point Theorem, the continuous map $f_{T_1,T_2,c}$ has at least one fixed point in Σ_α. $\square$

The conclusion of Proposition 5.29 is weaker than that of Theorem 5.19. The latter shows that for a sufficiently small temperature, we can find a fixed point of f_T arbitrarily close to a vertex of the simplex, whereas the former shows that for a sufficiently small temperature, we can find a fixed point of $f_{T_1,T_2,c}$ in a region whose size depends on the damping factor. In fact, such a fixed point may not approach a vertex of the simplex as one of the temperature tends to 0.

Remark 5.30. If the first column of A is not positive, the existence of a fixed point such that $x_1 > x_2, \ldots, x_n$ for small T_1 is not guaranteed. Consider for instance $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, $g_T(x) = e^{x/T}$, $T_2 = \infty$, $z = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $0 < c < 1$. Then, for each each $T_1 > 0$, any fixed point $x_{T_1,\infty,c}$ of $f_{T_1,\infty,c}$ belongs to $[0, \frac{1}{2}[\times]\frac{1}{2}, 1]$. $\diamond$

We now consider the case where the damping factor c approaches 1 and $T_1 = \infty$. Then, the corresponding value of the generalized PageRank converges to an invariant measure of the matrix $\operatorname{diag}(A\mathbf{1})^{-1}A$, independently of the choice of T_2.

Proposition 5.31. *Assume $T_1 = \infty$. For every vector norm $\|\cdot\|$ and every $\varepsilon > 0$, there exists $c_\varepsilon < 1$ such that for every fixed point x of $f_{\infty,T_2,c}$, with $c_\varepsilon < c < 1$, there exists an invariant measure u of $\operatorname{diag}(A\mathbf{1})^{-1}A$ such that $\|x - u\| < \varepsilon$.*

Proof. Let $P = \operatorname{diag}(A\mathbf{1})^{-1}A$, and let $M = I - P$. Since P is stochastic, the index of M is $\operatorname{ind}(M) = 1$ (see p. 33). Therefore, by Lemma 2.3, with $c = (1+\varepsilon)^{-1}$,

$$\lim_{c \to 1} (1-c)(I - cP)^{-1} = \lim_{\varepsilon \to 0} \varepsilon (M + \varepsilon I)^{-1} = I - MM^D.$$

Let $\|\cdot\|$ be a vector norm and $|||\cdot|||$ its induced matrix norm, and let $\nu > 0$ such that $\|v\| \leq \nu$ for all stochastic vector v. Let $\varepsilon > 0$. There exists $c_\varepsilon < 1$ such that if $c_\varepsilon < c < 1$,

$$|||(1-c)(I - cP)^{-1} - (I - MM^D)||| < \nu^{-1}\varepsilon.$$

Let $c \in]c_\varepsilon, 1[$, and let $\boldsymbol{x}$ be a fixed point of $\boldsymbol{f}_{\infty,T_2,c}$, that is, $\boldsymbol{x} = \boldsymbol{v}(\boldsymbol{x})(1 - c)(I - cP)^{-1}$, where $\boldsymbol{v}(\boldsymbol{x})_i = \boldsymbol{z}_i\, g_{T_2}(\boldsymbol{x}_i)/\sum_k \boldsymbol{z}_k\, g_{T_2}(\boldsymbol{x}_k)$ for all i. Then,

$$\|\boldsymbol{x} - \boldsymbol{v}(\boldsymbol{x})(I - MM^D)\| < \varepsilon.$$

But $\boldsymbol{v}(\boldsymbol{x})(I - MM^D)$ is an invariant measure of the matrix P. Indeed, $I - MM^D$ is stochastic, and $(I - MM^D)(I - P) = M - MM^DM = M - M^2M^D = 0$, by definition of the Drazin inverse. □

5.4 Estimating the critical temperature

We call *critical temperature* the largest temperature for which the number of fixed points of $\boldsymbol{u}_T$ changes. It corresponds to the loss of the uniqueness of the fixed point. In this section, we are interested in estimating the critical temperature for some particular cases. We study in detail the case of $n \times n$ matrices of all ones with the particular weight function $g_T(x) = e^{x/T}$.

We suppose that $g_T(x) = e^{x/T}$ and first consider the particular case where the graph is *complete* with

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

For this matrix, the point $\boldsymbol{x} = \frac{1}{n}\mathbf{1}$ is a fixed point of $\boldsymbol{u}_T$ for all T. We are interested in the existence of other fixed points, depending on the temperature T.

Lemma 5.32. *Assume that A is the $n \times n$ matrix of all ones. The point $\boldsymbol{x}$ is a fixed point of $\boldsymbol{u}_T$ if and only if $\boldsymbol{x} \in \Sigma$ and there exists $\lambda \in \mathbb{R}$ such that $\lambda = \boldsymbol{x}_i e^{-\boldsymbol{x}_i/T}$ for every $i = 1, \dots, n$.*

Proof. This follows directly from $\boldsymbol{x} = \boldsymbol{f}_T(\boldsymbol{x})$. □

Lemma 5.33. *Assume that A is the $n \times n$ matrix of all ones. The point $\boldsymbol{x}$ is a fixed point of $\boldsymbol{u}_T$ if and only if $\boldsymbol{x} \in \Sigma$ and there exists $\mathcal{K} \subseteq \{1, \dots, n\}$, $y \in [0, T]$ and $z \geq T$ such that $ye^{-y/T} = ze^{-z/T}$, $|\mathcal{K}|y + (n - |\mathcal{K}|)z = 1$, $\boldsymbol{x}_i = y$ for all $i \in \mathcal{K}$ and $\boldsymbol{x}_i = z$ for all $i \notin \mathcal{K}$.*

Proof. Since the map $x \mapsto xe^{-x/T}$ is increasing for $0 \le x < T$ and decreasing for $x > T$, there can be at most two values $y \neq z$ such that $ye^{-y/T} = ze^{-z/T} = \lambda$ for a given $\lambda \in \mathbb{R}$. The result hence follows from Lemma 5.32 and $\boldsymbol{x} \in \Sigma$. □

Since in our case, $L_{\max}(g_T) = T^{-1}$, we know from Theorem 5.14 that the critical temperature is at most n. Proposition 5.34 shows that this critical temperature is in fact roughly $(\ln n)^{-1}$ when n tends to infinity.

Proposition 5.34. *Assume that A is the $n \times n$ matrix of all ones. If $n > 2$, the map $\boldsymbol{u}_T$ has a unique fixed point $\boldsymbol{x} = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$ if and only if $T > T^*(n)$, where*

$$\frac{1 - \frac{1}{\ln n}}{\ln((\ln n - 1)n + 1)} \le T^*(n) = \sup_{\alpha > 1} \frac{1 - \frac{1}{\alpha}}{\ln((\alpha - 1)n + 1)} < \frac{1}{\ln (n-1)},$$

thus $T^*(n) \sim \frac{1}{\ln n}$ when n tends to ∞. If $n = 2$, the map $\boldsymbol{u}_T$ has a unique fixed point $\boldsymbol{x} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ if and only if $T \ge T^*(2) = \frac{1}{2}$.

Proof. From Lemma 5.33, $\boldsymbol{x} \in \operatorname{int}(\Sigma)$ is a fixed point of $\boldsymbol{u}_T$, with $\boldsymbol{x} \neq \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$, if and only if there exists $\mathcal{K} \subset \{1, \ldots, n\}$, $y, z \in \mathbb{R}$, such that $\boldsymbol{x}_i = y$ for $i \in \mathcal{K}$, $\boldsymbol{x}_i = z$ for $i \notin \mathcal{K}$, $0 < k = |\mathcal{K}| < n$, $ky + (n-k)z = 1$, $ye^{-y/T} = ze^{-z/T}$, and $y < z$. Denote $\alpha = \frac{1}{ny}$. Since $y < \frac{1}{n}$, we get necessary that $\alpha > 1$. From $ye^{-y/T} = ze^{-z/T}$, we get $T = T_{\alpha,k}$, where

$$T_{\alpha,k} = \frac{1 - \frac{1}{\alpha}}{(n-k) \ln\left(\frac{(\alpha-1)n}{n-k} + 1\right)}.$$

This implies that $\boldsymbol{u}_T$ has a fixed point $\boldsymbol{x} \in \operatorname{int}(\Sigma)$, $\boldsymbol{x} \neq \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$ if and only if $T \in \mathcal{T} = \{T_{\alpha,k}, \alpha > 1, k \in \{1, \ldots, n-1\}\}$. Let

$$T^*(n) = \sup_{\alpha > 1} T_{\alpha,n-1} = \sup_{\alpha > 1} \frac{1 - \frac{1}{\alpha}}{\ln((\alpha - 1)n + 1)}.$$

We shall show that $\mathcal{T} =]0, T^*(n)]$ when $n > 2$ and $\mathcal{T} =]0, T^*(2)[$ when $n = 2$.

First, a study of $T_{\alpha,k}$ as a function of k shows that it is increasing. It is therefore sufficient to show that $\{T_{\alpha,n-1}, \alpha > 1\} =]0, T^*(n)]$ when

$n > 2$, and $\{T_{\alpha,1}, \alpha > 1\} =]0, T^*(2)[$ when $n = 2$. Second, a study of $T_{\alpha,n-1}$ as a function of $\alpha > 1$ shows that, when $n > 2$, $T_{\alpha,n-1}$ is increasing, then decreasing, tends to 0 when α goes to infinity, and its maximum is attained for $\alpha = \alpha_n$, where $\alpha_n > \frac{2(n-1)}{n}$. Hence $\mathcal{T} =]0, T^*(n)]$. When $n = 2$, $T_{\alpha,1}$ is decreasing, tends to 0 when α goes to infinity, and to $\frac{1}{2}$ when α goes to 1. Hence $T^*(2) = \frac{1}{2}$, and $\mathcal{T} =]0, T^*(2)[$.

Moreover, for $n \geq 3$,

$$T^*(n) = T_{\alpha_n,n-1} = \frac{1 - \frac{1}{\alpha_n}}{\ln((\alpha_n - 1)n + 1)} < \frac{1}{\ln((\alpha_n - 1)n + 1)},$$

and since $\alpha_n > \frac{2(n-1)}{n}$, we get $T^*(n) < \frac{1}{\ln(n-1)}$. For the lower bound, we get $T_{\ln n,n-1} \leq T^*(n)$, since $\ln n > 1$. □

Proposition 5.34 deals with the very special case of a complete graph. In more general circumstances, the exact computation of the critical temperature seems out of range. However, we can obtain numerically a lower bound of the critical temperature, which seems to be an accurate estimate, using the following homotopy-type method. We first choose two random initial vectors on the simplex. Then, we iterate the map f_T from each of these vectors. For small values of T, this yields with an overwhelming probability two different web ranks. Then, we increase the temperature T, and keep iterating the map f_T on each of these web ranks, until the two web ranks coincide. This yields a lower bound of the critical temperature. Then, we repeat this procedure, with new random initial vectors, until the lower bound of the critical temperature is not improved any more. Note that the simpler method consisting in keeping T fixed and iterating f_T from various initial conditions (random vectors or Dirac distributions on a vertex of the simplex) experimentally yields an under estimate of the critical temperature.

Using the previously described homotopy-type method, we computed numerically the critical temperature for two families of graphs. These experiments reveal that the $1/\ln(n)$ asymptotic obtained for the complete graph gives a good general estimate. We first considered the ring graph, with n nodes, in which node i is connected to its two neighbors and to itself. The critical temperature, for $n = 51, 201, 501$ and 1001 is shown by stars in Figure 5.4. The exact value of the critical

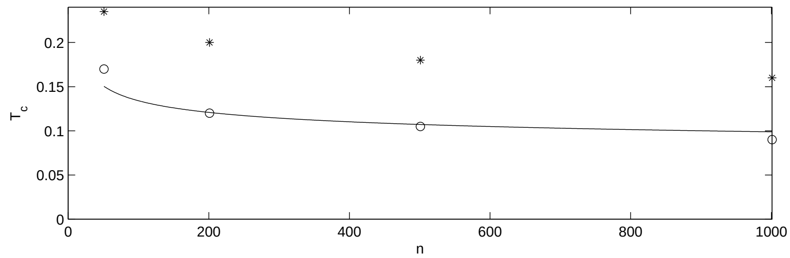


Figure 5.4: Estimation of the critical temperature T_c as a function of the number of nodes n for the complete graph (continuous curve), the ring graph (stars) and random graphs (circles).

temperature of the complete graph with n nodes, $T^*(n)$, is drawn as a continuous curve. We see that the critical temperatures of the ring and complete graphs are essentially proportional. We also computed numerically the critical temperature for a standard model of random directed graph, in which the presence of the different arcs are independent random variables, and for every (i, j), the probability of presence of the arc (i, j) is given by the same number p. We took $p = 10/n$, so that every node is connected to an average number of 10 nodes. The corresponding critical temperatures are represented by circles. The values of these critical temperatures do not seem to change significantly with the realization of the random graph, hence, each of the values which are represented correspond to a unique realization.

We have noted in our numerical experiments that the convergence of the iterates of $\boldsymbol{f}_T$ may be very slow for some temperatures T. In some cases, the iteration of $\boldsymbol{u}_T$ appeared to be more efficient. Note that the implementation of $\boldsymbol{u}_T$ is much more difficult, and that a single iteration of $\boldsymbol{u}_T$ needs more time to be computed, since it needs the resolution of an implicit system.

Remark 5.35. Let us briefly discuss the case of an arbitrary 2×2 irreducible matrix A with the weight function $g_T = e^{x/T}$. In this case, some elementary calculations give information about the critical temperature [2]. Firstly, the critical temperature for a graph of only two nodes is always less than 1, since it can be shown that $\boldsymbol{u}_T$ has a unique fixed point if $T \geq 1$. This is the best general upper bound that can be given

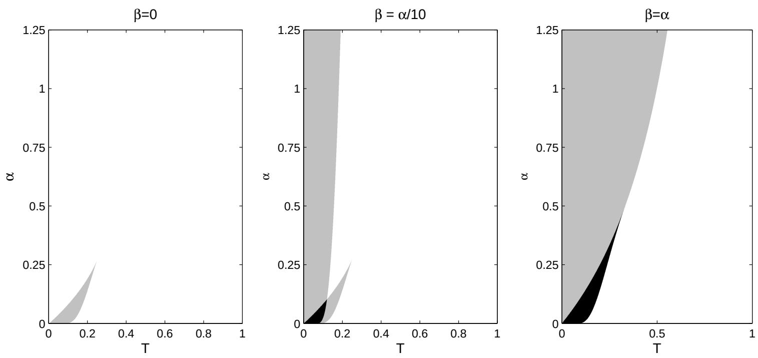


Figure 5.5: For 2×2 matrices, the map $\boldsymbol{u}_T$ has 5 fixed points for (T, α) in the black region, 3 fixed points for (T, α) in the gray region, and 1 fixed point otherwise.

for problems of this dimension since for every $T < 1$, we can construct a 2×2 matrix such that $\boldsymbol{u}_T$ has at least two fixed points. Moreover, one can show that, for every $T > 0$, the map $\boldsymbol{u}_T$ has at most 5 fixed points and does not have any orbit of period greater than 1. Numerical experiments show that for a 2×2 irreducible matrix, the number of fixed points of the map $\boldsymbol{u}_T$ can change $0, 1, 2$ or even 3 times when decreasing the temperature T.

This can be seen in Figures 5.5, which were obtained experimentally. Let $\alpha = A_{11}/A_{12}$ and $\beta = A_{22}/A_{21}$. For a specified β, $\boldsymbol{u}_T$ has 5 fixed point if (T, α) belongs to the black region, 3 fixed points in the gray region and 1 fixed point in the white region. $\diamond$

5.5 Experiments on a subgraph of the Web

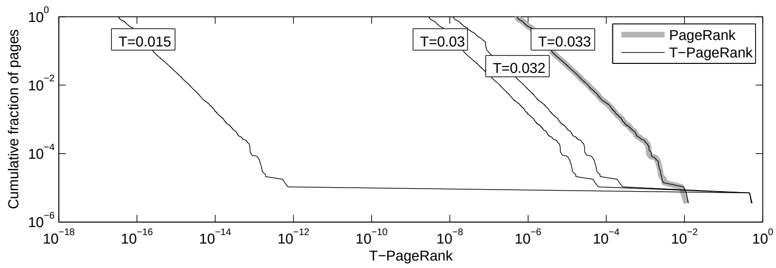
In this section, we briefly present our experiments of the T-PageRank on a large-scale example. We consider a subgraph of the Web with about 280,000 nodes which has been obtained by S. Kamvar from a crawl on the Stanford web [69]. We use the variant of our model presented in Section 5.3, with a transition matrix given by

$$P(\boldsymbol{x})_{ij} = c\,\frac{A_{ij}\,e^{x_j/T}}{\sum_k A_{ik}\,e^{x_k/T}} + (1-c)\,\frac{e^{x_j/T}}{\sum_k e^{x_k/T}},$$

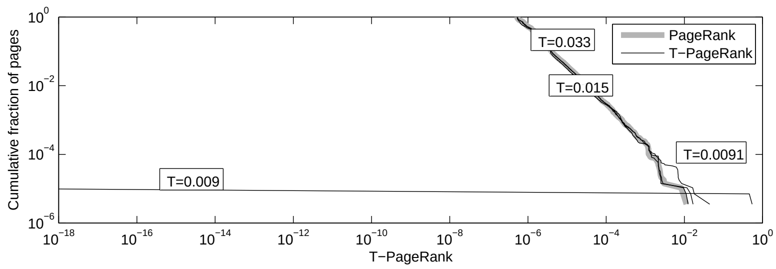
where we suppose that for each dangling node i (i.e., a node corresponding to a web page without hyperlink), the i^{th} row of the matrix A is a row of all ones. The chosen damping factor is $c = 0.85$. We have computed the T-PageRank from the recurrence (5.2) for various temperatures T and initial rankings. As expected, when the temperature T is large, the T-PageRank is very close to the classical PageRank, and when T approaches zero, arbitrary close initial rankings can induce totally different T-PageRanks. The critical temperature experimentally seems to be about $T = 0.033$. It has the same order of magnitude as the $T^*(n) = 0.06148$ estimate discussed in Section 5.4.

As in [117], we represent in Figure 5.6, in a log-log scale, the cumulative distribution function of the PageRank, i.e the proportion of pages for which the T-PageRank is larger than a given value, as a function of this value. In Figure 5.6(a), we show the successive T-PageRanks obtained for *increasing* temperatures from $T = 0.015$ to the critical temperature $T = 0.033$ by the following variant of the previously described homotopy method: for $T = 0.015$, we iterate the map $\boldsymbol{f}_T$, with a Dirac mass on a vertex of the simplex as initial ranking, until a fixed point is reached. Then, for each new value of T, we iterate $\boldsymbol{f}_T$ until a fixed point is reached, starting from the previous fixed point. For $T \leq 0.032$ the distribution of the T-PageRank is quite different from that of the PageRank and it comes closer suddenly for $T = 0.033$. In Figure 5.6(b), we show the successive T-PageRanks obtained by a similar method for *decreasing* temperatures from $T = 0.033$ to 0.009, with the classical PageRank as an initial ranking. The latter procedure may be compared with simulated annealing schemes, in which the temperature is gradually decreased. Until $T = 0.0091$, the distribution of the T-PageRank is quite similar to that of the classical PageRank (see a zoom in Figure 5.6(c)). With $T = 0.009$, the T-PageRank moves suddenly away from the PageRank. These figures suggest that the gap between web pages considered as "good" and "bad" is more pronounced with the T-PageRank than with the classical PageRank.

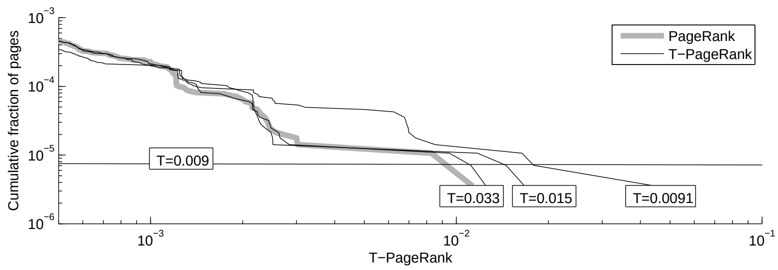
We have also compared the five best nodes for the classical PageRank and for the T-PageRank with decreasing temperatures $T = 0.033$, 0.015 and 0.0091. As we see in Figure 5.7, for $T = 0.033$, the PageRank and T-PageRank give a similar ranking for the top-five. But for smaller temperatures as $T = 0.015$ or $T = 0.0091$, even the two best nodes are



(a)



(b)



(c)

Figure 5.6: Fraction of pages having a PageRank larger than a particular value. (a) T-PageRanks computed with increasing temperatures from $T = 0.015$ to 0.033, with a vertex of the simplex as initial ranking. (b) T-PageRanks computed with decreasing temperatures from $T = 0.033$ to 0.009, with the classical PageRank vector as initial ranking. (c) Zoom of Figure (b).

$T = 0.033$	$T = 0.015$	$T = 0.0091$
1	2	2
2	1	1
3	3	3
4	6	46
5	7	33

Figure 5.7: The five best nodes of the T-PageRank for several values of T: the numbers refer to the rankings according to the classical PageRank.

exchanged.

Since for this special set of data, the correspondence between the page numbers and the urls is not available, one cannot interpret the discrepancies between the PageRank and the T-PageRank. In [104], J.-P. Poveda made similar experiments on the larger matrix obtained by S. Kamvar for a crawl of the union of the Stanford and Berkeley webs [68], with about $685,000$ nodes, for which, this time, the correspondence between some pages and the main urls is given. These experiments suggest that the T-PageRank obtained by the latter scheme, in which the temperature is gradually decreased, as illustrated in Figure 5.6(b), might be of practical interest.

5.6 Conclusions

Google informs the users of services such as the Google Toolbar, that information about web pages they visit may be collected and used in order to "improve Google technologies and services" [50].

Within the limits of our model, our results show that Google should *not* use this information in order to update the PageRank scores. Indeed, we have seen that, if the web surfers excessively rely on the web ranking, this could lead to pathological phenomena, like getting non unique and even meaningless rankings.

Our results might be considered as an argument in favor of the claim that PageRank type measures should not be used to assess *quality* of web pages but only *popularity* of them. Indeed, the validity of the classical PageRank relies on an ideal view of the web, in which the web masters are thought of as experts, creating hyperlinks only to pages they carefully examined, and judged by themselves to be of interest. In the real world, however, the web masters may be influenced by factors like reputation of web pages, that may be heavily influenced by the web ranking.

The literature propose several preferential attachment models of the impact of web ranking on the evolution of the web graph [11, 35, 37, 46]. We have here proposed a model of the mutual influence between web ranking and *web surfing*. Note however that, the basic version of our T-PageRank model has also the following naive interpretation in the context of the mutual influence of the web ranking and the evolution of the web graph. Web masters do not delete nor add any hyperlink but only weight their hyperlinks according to the web ranking, for instance by organizing them on their web pages in decreasing order with respect to the web ranking.

Chapter 6

Conclusions

In this thesis, we have studied problems related to information extraction in large graphs, with the help of dominant vectors of nonnegative matrices.

Data mining in large networks is a hot topic. Technological advances such as the Internet and the Web, computer-based databases or mobile phone networks have created the need for efficient methods of information extraction in large graphs.

Nonnegative matrices and nonlinear maps on the nonnegative orthant have applications in various fields. Economical or population growth models, chemical processes or problems of flows in networks, for instance, lead naturally to consider nonnegative quantities. In this context, dominant eigenvectors of nonnegative matrices may describe for instance an equilibrium, a probability distribution or an optimal network property. Moreover, the correspondence between nonnegative matrices and graphs makes Perron–Frobenius methods a powerful tool for the analysis of networks.

Many more problems could have been described in a thesis with a title as general as "Dominant vectors of nonnegative matrices. Application to information extraction in large graphs". We have focused in this work on three particular problems that motivated us.

Although our results are not intended to be directly usable in real applications, this research was motivated from real information extraction problems.

Our first topic started from a comparison between two measures of similarity in graphs and their use for databases matching and automatic extraction of synonyms in a dictionary. Our goal to provide a consistent mathematical framework for the matching algorithm of Melnik et al. has led us to the study of the conditional affine eigenvalue problem on the nonnegative orthant. This analysis gives insight into the properties of the fixed point of the normalized affine iteration. This iteration may possibly be useful in some cases as a variant of the power method, for example when this latter does not converge.

The two others problems we consider in the thesis are related to the PageRank. One may perhaps object that studying PageRank is useless since Google does not use PageRank as such. Anchor text is used [25], as well as personalized search [56, 57, 84], for instance. And Google has most probably developed even more efficient and more complicated tools for ranking web pages. I am nevertheless convinced that studying PageRank is meaningful. Indeed, the PageRank model is now used in other contexts, such as in bibliometrics [21] or in attempts to rank graduate programs [107] or sport teams [51]. Moreover, PageRank is a very simple model that could be used as a toy model in order to understand more complicated situations. It is also interesting to note a certain universality in ranking methods like PageRank. Altman and Tennenholtz [4] prove that, for a strongly connected graph, every ranking of the nodes satisfying a few set of intuitive axioms must coincide with the ranking induced by the PageRank for a damping factor taken as $c = 1$, i.e., with no zapping.

The results we obtain about optimal linkage strategies for maximizing the PageRank of a web site, or about self-validating web rankings, could probably not be easy to adapt to other kinds of networks. Indeed, a feature of popularity measures such as PageRank, is that high scores, corresponding to high probabilities of presence in some nodes, are expected. On the contrary, for problems of flow on roads networks or on the network of routers on the Internet for instance, a high probability of presence may lead to an overload of some nodes of the network and should therefore be avoided. So, optimal link structures for subset of nodes should rather be structures minimizing a probability of presence in these nodes. And random walkers on these networks should rather use information about nodes that are often visited in order to avoid them. It may nevertheless be possible that the kind of arguments we used could be adapted in order to find link structures minimizing the probability of presence in some nodes, under some accessibility assumptions.

Before concluding, I would like to present two questions which I think are both challenging and interesting. The first one concerns *the possible impact of PageRank on the evolution of the web graph in the context of preferential attachment models*. Search engines may have an influence on the discovery of web pages by web masters. So, whenever a new page is created by a web master, this page may be more likely to link to some well ranked web pages. And similarly, new links from existing pages may be preferentially added to well ranked pages while other links to non popular pages may be deleted. In Section 5.1, we mentioned studies about the bias that search engines may introduce, that is, how popular, well ranked web pages may become more and more popular. One could also ask how the connectivity of the web graph may evolve if web masters are influenced by the web ranking. Will there be more and more separated communities focusing on specific topics? Or on the contrary will the web ranking have a unifying role? It may be interesting to see if the conclusions differ from those obtained for classical preferential attachment models. Note that it could also be interesting to look at such questions from the point of view of opinion dynamics.

The other question is related to *PageRank in the context of game theory*. What if several or even all web masters try to maximize the PageRank of their web site by choosing their hyperlinks? Or if they try to have a ranking as good as possible? What would be optimal strategies in the case it does not much matter for web masters to be ranked in third or fourth position but well to be listed on the first page of results returned by the search engine? I would expect that for some objective functions, some alliance strategies may be optimal, maybe with interesting link structures.

I am convinced that information extraction in large graphs have very good prospects. Graphs are an easy and convenient way to store many kinds of data for which there exist physical, social or information connections. There exist more and more of such relational databases, as for instance digital libraries (arXiv), online auction (eBay) or social networking web sites (Facebook). But, at present, there is a lack of techniques for exploring these data. Methods based on dominant vectors of nonnegative matrices, on statistic distributions or on SVD computations for instance have a future for data mining in large networks.

Bibliography

[1] M. Akian, S. Gaubert, and L. Ninove, *Multiple equilibria of nonhomogeneous Markov chains and self-validating Web rankings*, submitted to SIAM J. Matrix Anal. Appl.

[2] ———, *The T-PageRank: A model of self-validating effects of web surfing*, Positive systems, Lecture Notes in Control and Inform. Sci., vol. 341, Springer, 2006, pp. 239–246.

[3] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Modern Phys. **74** (2002), no. 1, 47–97.

[4] A. Altman and M. Tennenholtz, *Ranking systems: the PageRank axioms*, Proc. 6th ACM Conference on Electronic Commerce (EC'05), ACM, 2005, pp. 1–8.

[5] M. Artzrouni, *On the dynamics of the linear process* $Y(k) = A(k)Y(k-1)$ *with irreducible matrices* $A(k)$, SIAM J. Matrix Anal. Appl. **17** (1996), no. 4, 822–833.

[6] M. Artzrouni and O. Gavart, *Nonlinear matrix iterative processes and generalized coefficients of ergodicity*, SIAM J. Matrix Anal. Appl. **21** (2000), no. 4, 1343–1353.

[7] K. Avrachenkov, N. Litvak, D. Nemirovsky, and N. Osipova, *Monte Carlo methods in PageRank computation: when one iteration is sufficient*, SIAM J. Numer. Anal. **45** (2007), no. 2, 890–904.

[8] K. Avrachenkov and N. Litvak, *Decomposition of the Google PageRank and Optimal Linking Strategy*, Tech. report, INRIA, 2004.

[9] ———, *The effect of new links on Google PageRank*, Stoch. Models **22** (2006), no. 2, 319–331.

[10] R. Baeza-Yates, C. Castillo, and V. López, *PageRank Increase under Different Collusion Topologies*, Proc. 1st International Workshop on Adversarial Information Retrieval on the Web (AIRWeb'05), 2005.

[11] R. Baeza-Yates, C. Castillo, and F. Saint-Jean, *Web Dynamics, Structure and Page Quality*, Web Dynamics: Adapting to Change in Content, Size, Topology and Use (M. Levene and A. Poulovassilis, eds.), Springer, 2004, pp. 93–109.

[12] F. L. Bauer, J. Stoer, and C. Witzgall, *Absolute and monotonic norms*, Numer. Math. **3** (1961), 257–264.

[13] P. Berkhin, *A survey on PageRank computing*, Internet Math. **2** (2005), no. 1, 73–120.

[14] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Classics in Applied Mathematics, vol. 9, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[15] T. Berners-Lee, *Information Management: A Proposal*, Tech. report, CERN, 1989.

[16] M. Bianchini, M. Gori, and F. Scarselli, *Inside PageRank*, ACM Trans. Inter. Tech. **5** (2005), no. 1, 92–128.

[17] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, *A measure of similarity between graph vertices: Applications to synonym extraction and web searching*, SIAM Rev. **46** (2004), no. 4, 647–666.

[18] V. D. Blondel, L. Ninove, and P. Van Dooren, *Affine iterations on nonnegative vectors*, Proc. 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS'04), 2004.

[19] ———, *Convergence of graph similarity algorithms*, Book of Abstracts 23rd Benelux Meeting on Systems and Control, 2004, p. 121.

[20] ———, *An affine eigenvalue problem on the nonnegative orthant*, Linear Algebra Appl. **404** (2005), 69–84.

[21] J. Bollen, M. A. Rodriguez, and H. Van de Sompel, *Journal Status*, Scientometrics **69** (2006), no. 3, 669–687.

[22] M. Bouklit and F. Mathieu, *BackRank: An alternative for PageRank?*, Proc. 14th International World Wide Web Conference on Special Interest Tracks & Posters (WWW'05), ACM Press, 2005, pp. 1122–1123.

[23] C. Brezinski, M. Redivo-Zaglia, and S. Serra-Capizzano, *Extrapolation methods for PageRank computations*, C. R. Math. Acad. Sci. Paris **340** (2005), no. 5, 393–397.

[24] C. Brezinski and M. Redivo-Zaglia, *The PageRank vector: Properties, computation, approximation, and acceleration*, SIAM J. Matrix Anal. Appl. **28** (2006), no. 2, 551–575.

[25] S. Brin and L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems **30** (1998), no. 1–7, 107–117.

[26] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, *Graph structure in the Web*, Comput. Networks **33** (2000), no. 1-6, 309–320.

[27] R. A. Brualdi and H. J. Ryser, *Combinatorial matrix theory*, Encyclopedia of Mathematics and its Applications, vol. 39, Cambridge University Press, 1991.

[28] V. Bush, *As we may think*, Atlantic Monthly **176** (1945), 101–108.

[29] ———, *As we may think: A top U.S. scientist foresees a possible future world in which man-made machines will start to think*, Life Magazine **19** (1945), no. 11, 112–124.

[30] T. Caelli and S. Kosinov, *An Eigenspace Projection Clustering Method for Inexact Graph Matching*, IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004), no. 4, 515–519.

[31] R. Cailliau and J. Gillies, *How the Web Was Born: The Story of the World Wide Web*, Oxford University Press, 2000.

[32] D. Carlson, *A note on M-matrix equations*, J. Soc. Indust. Appl. Math. **11** (1963), 1027–1033.

[33] H. Caswell, *Matrix Population Models: Construction, Analysis, and Interpretation*, second ed., Sinauer, 2001.

[34] O. Catoni, *Simulated annealing algorithms and Markov chains with rare transitions*, Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, Springer, 1999, pp. 69–119.

[35] S. Chakrabarti, A. Frieze, and J. Vera, *The influence of search engines on preferential attachment*, Proc. 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05), Society for Industrial and Applied Mathematics, 2005, pp. 293–300.

[36] G. E. Cho and C. D. Meyer, *Markov chain sensitivity measured by mean first passage times*, Linear Algebra Appl. **316** (2000), no. 1-3, 21–28.

[37] J. Cho and S. Roy, *Impact of search engines on page popularity*, Proc. 13th International Conference on World Wide Web (WWW'04), ACM, 2004, pp. 20–29.

[38] J. Conlisk, *Stability and monotonicity for interactive Markov chains*, J. Math. Sociol. **17** (1992), no. 2-3, 127–143.

[39] C. de Kerchove, L. Ninove, and P. Van Dooren, *Maximizing PageRank via outlinks*, accepted for publication in Linear Algebra Appl.

[40] ———, *Maximizing PageRank via outlinks*, Book of Abstracts 26rd Benelux Meeting on Systems and Control, 2004, p. 121.

[41] C. H. Q. Ding, H. Zha, X. He, P. Husbands, and H. D. Simon, *Link analysis: Hubs and Authorities on the World Wide Web*, SIAM Rev. **46** (2004), no. 2, 256–268.

[42] L. Eldén, *A Note on the Eigenvalues of the Google Matrix*, Tech. Report LiTH-MAT-R-04-01, Department of Mathematics, Linköping University, 2004.

[43] R. Fagin, A. R. Karlin, J. Kleinberg, P. Raghavan, S. Rajagopalan, R. Rubinfeld, M. Sudan, and A. Tomkins, *Random walks with "back buttons"*, Ann. Appl. Probab. **11** (2001), no. 3, 810–862.

[44] S. Fallat, D. D. Olesky, and P. van den Driessche, *Graph-theoretic aspects of maximizing the spectral radius of nonnegative matrices*, Linear Algebra Appl. **253** (1997), 61–77.

[45] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the Internet topology*, Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM'99), ACM, 1999, pp. 251–262.

[46] S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, *Topical interests and the mitigation of search engine bias*, Proc. Natl. Acad. Sci. **103** (2006), no. 34, 12684–12689.

[47] J. H. Fowler and S. Jeon, *The authority of Supreme Court precedent*, Social Networks (2007), to appear.

[48] C. Fraikin and P. Van Dooren, *Graph matching with type constraints on nodes and edges*, Web Information Retrieval and Linear Algebra Algorithms, Dagstuhl Seminar Proceedings, no. 07071, 2007.

[49] S. Gaubert and J. Gunawardena, *The Perron-Frobenius theorem for homogeneous, monotone functions*, Trans. Amer. Math. Soc. **356** (2004), no. 12, 4931–4950.

[50] Google, *Google Privacy Policy Highlights*, `http://www.google.com/privacy.html`, October 2005.

[51] A. Govan, C. D. Meyer, and R. Albright, *Generalizing Google's PageRank to Rank Sports Teams*, preprint.

[52] L. Gurvits, *Stability of discrete linear inclusion*, Linear Algebra Appl. **231** (1995), 47–85.

[53] Z. Gyöngyi and H. Garcia-Molina, *Link spam alliances*, Proc. 31st International Conference on Very Large Data Bases (VLDB'05), VLDB Endowment, 2005, pp. 517–528.

[54] L. Han, M. Neumann, and M. Tsatsomeros, *Spectral radii of fixed Frobenius norm perturbations of nonnegative matrices*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 1, 79–92.

[55] D. J. Hartfiel, *Nonhomogeneous matrix products*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

[56] T. Haveliwala, S. Kamvar, and G. Jeh, *An Analytical Comparison of Approaches to Personalizing PageRank*, Tech. report, Stanford University, 2003.

[57] T. H. Haveliwala, *Topic-sensitive PageRank*, Proc. 11th International Conference on World Wide Web (WWW'02), ACM Press, 2002, pp. 517–526.

[58] T. H. Haveliwala and S. D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Tech. report, Stanford University, 2003.

[59] D. Hershkowitz and H. Schneider, *Solutions of Z-matrix equations*, Linear Algebra Appl. **106** (1988), 25–38.

[60] J. A. Holyst, K. Kacperski, and F. Schweitzer, *Social Impact Models of Opinion Dynamics*, Annual Reviews of Computational Physics (D. Stauffer, ed.), vol. 9, World Scientific, Singapore, 2001, pp. 253–273.

[61] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1990.

[62] ______, *Topics in matrix analysis*, Cambridge University Press, 1994.

[63] I. C. F. Ipsen and S. Kirkland, *Convergence analysis of a PageRank updating algorithm by Langville and Meyer*, SIAM J. Matrix Anal. Appl. **27** (2006), no. 4, 952–967.

[64] I. C. F. Ipsen and R. S. Wills, *Mathematical properties and analysis of Google's PageRank*, Bol. Soc. Esp. Mat. Apl. (2006), no. 34, 191–196.

[65] A. Jadbabaie, J. Lin, and A. S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control **48** (2003), no. 6, 988–1001.

[66] G. Jeh and J. Widom, *Scaling personalized web search*, Proc. 12th International Conference on World Wide Web (WWW'03), ACM Press, 2003, pp. 271–279.

[67] C. R. Johnson, R. Loewy, D. D. Olesky, and P. van den Driessche, *Maximizing the spectral radius of fixed trace diagonal perturbations of nonnegative matrices*, Linear Algebra Appl. **241–243** (1996), 635–654.

[68] S. D. Kamvar, *Stanford–Berkeley Web Matrix*, `http://www.stanford.edu/~sdkamvar/research.html`, from a December 2002 crawl.

[69] ———, *Stanford Web Matrix*, `http://www.stanford.edu/~sdkamvar/research.html`, from a September 2002 crawl.

[70] S. D. Kamvar, T. H. Haveliwala, and G. H. Golub, *Adaptive methods for the computation of PageRank*, Linear Algebra Appl. **386** (2004), 51–65.

[71] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, *Exploiting the Block Structure of the Web for Computing PageRank*, Tech. report, Stanford University, 2003.

[72] ———, *Extrapolation methods for accelerating PageRank computations*, Proc. 12th International Conference on World Wide Web (WWW'03), ACM Press, 2003, pp. 261–270.

[73] J. P. Keener, *The Perron-Frobenius theorem and the ranking of football teams*, SIAM Rev. **35** (1993), no. 1, 80–93.

[74] J. G. Kemeny and J. L. Snell, *Finite Markov chains*, The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.

[75] S. Kirkland, *Conditioning of the entries in the stationary vector of a Google-type matrix*, Linear Algebra Appl. **418** (2006), no. 2-3, 665–681.

[76] J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM **46** (1999), no. 5, 604–632.

[77] E. Kohlberg and J. W. Pratt, *The contraction mapping approach to the Perron-Frobenius theory: Why Hilbert's metric?*, Math. Oper. Res. **7** (1982), no. 2, 198–210.

[78] U. Krause, *On minimal and maximal eigenvalues*, Unpublished note.

[79] ______, *Heterogeneous labour and the fundamental Marxian theorem*, Rev. Econom. Stud. **48** (1981), no. 1, 173–178.

[80] ______, *Perron's stability theorem for nonlinear mappings*, J. Math. Econom. **15** (1986), no. 3, 275–282.

[81] ______, *Concave Perron-Frobenius theory and applications*, Nonlinear Anal. **47** (2001), no. 3, 1457–1466.

[82] F. M. Lalli, *PageRank*, Art draw, `http://commons.wikimedia.org/wiki/Image:PageRank-hi-res.png`.

[83] A. N. Langville and C. D. Meyer, *Deeper inside PageRank*, Internet Math. **1** (2004), no. 3, 335–380.

[84] ______, *Google's PageRank and beyond: The science of search engine rankings*, Princeton University Press, 2006.

[85] ______, *Updating Markov chains with an eye on Google's PageRank*, SIAM J. Matrix Anal. Appl. **27** (2006), no. 4, 968–987.

[86] R. Lempel and S. Moran, *Rank-Stability and Rank-Similarity of Link-Based Web Ranking Algorithms in Authority-Connected Graphs.*, Inf. Retr. **8** (2005), no. 2, 245–264.

[87] J. Lorenz, *A stabilization theorem for dynamics of continuous opinions*, Phys. A **355** (2005), no. 1, 217–223.

[88] ______, *Consensus Strikes Back in the Hegselmann-Krause Model of Continuous Opinion Dynamics Under Bounded Confidence*, Journal of Artificial Societies and Social Simulation **9** (2006), no. 1, 8.

[89] F. Mathieu and M. Bouklit, *The effect of the back button in a random walk: Application for PageRank*, Proc. 13th International World Wide Web Conference on Alternate track papers & posters (WWW Alt.'04), ACM Press, 2004, pp. 370–371.

[90] F. Mathieu, *Graphes du Web: Mesures d'importance à la PageRank*, Ph.D. thesis, Université Montpellier II, 2004.

[91] M. H. McAndrew, *On the product of directed graphs*, Proc. Amer. Math. Soc. **14** (1963), 600–606.

[92] S. Melnik, H. Garcia-Molina, and E. Rahm, *Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching*, Extended tech. report, Stanford University, 2001.

[93] ———, *Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching*, Proc. 18th International Conference on Data Engineering (ICDE'02), IEEE Computer Society, 2002, pp. 117–128.

[94] V. Metz, *Nonlinear Perron-Frobenius theory in finite dimensions*, Nonlinear Anal. **62** (2005), no. 2, 225–244.

[95] C. D. Meyer, *Limits and the index of a square matrix*, SIAM J. Appl. Math. **26** (1974), 469–478.

[96] C. D. Meyer and M. W. Stadelmaier, *Singular M-matrices and inverse positivity*, Linear Algebra Appl. **22** (1978), 139–156.

[97] M. Morishima, *Equilibrium, stability, and growth: A multi-sectoral analysis*, Clarendon Press, Oxford, 1964.

[98] T. H. Nelson, *Xanalogical structure, needed now more than ever: Parallel documents, deep links to content, deep versioning, and deep re-use*, ACM Comput. Surv. **31** (1999), no. 4es, 33.

[99] M. E. J. Newman, *The structure and function of complex networks*, SIAM Rev. **45** (2003), no. 2, 167–256.

[100] A. Y. Ng, A. X. Zheng, and M. I. Jordan, *Link analysis, eigenvectors and stability*, Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI'01), 2001.

[101] R. D. Nussbaum, *Hilbert's projective metric and iterated nonlinear maps*, Mem. Amer. Math. Soc. **75** (1988), no. 391.

[102] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási, *Structure and tie strengths in mobile communication networks*, Proc. Natl. Acad. Sci. **104** (2007), no. 18, 7332–7336.

[103] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*, Tech. report, Computer Science Department, Stanford University, 1998.

[104] J. P. Poveda, *Calcul du T-PageRank d'un graphe du Web*, PPL internship report, École Nationale Supérieure des Techniques Avancées, Août 2007.

[105] M. Richardson and P. Domingos, *The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank*, Advances in Neural Information Processing Systems 14 (NIPS'01), MIT Press, 2001, pp. 1441–1448.

[106] T. L. Saaty, *Rank According to Perron: A New Insight*, Mathematics Magazine **60** (1987), no. 4, 211–213.

[107] B. M. Schmidt and M. M. Chingos, *Ranking Doctoral Programs by Placement: A New Method*, PS: Political Science and Politics **40** (2007), 523–529.

[108] H. Schneider, *The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and on related properties: A survey*, Linear Algebra Appl. **84** (1986), 161–189.

[109] P. P. Senellart and V. D. Blondel, *Automatic discovery of similar words*, Survey of Text Mining: Clustering, classification, and retrieval (M. Berry, ed.), Springer-Verlag, 2003, pp. 25–44.

[110] E. Seneta, *Nonnegative matrices and Markov chains*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 1981.

[111] S. Serra-Capizzano, *Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 2, 305–312.

[112] M. Sydow, *Random surfer with back step*, Proc. 13th International World Wide Web Conference on Alternate track papers & posters (WWW Alt.'04), ACM Press, 2004, pp. 352–353.

[113] ———, *Can One Out-Link Change Your PageRank?*, Advances in Web Intelligence (AWIC'05), Lecture Notes in Computer Science, vol. 3528, Springer, 2005, pp. 408–414.

[114] B.-S. Tam and H. Schneider, *Linear equations over cones and Collatz-Wielandt numbers*, Linear Algebra Appl. **363** (2003), 295–332.

[115] W. T. Tutte, *Graph theory*, Encyclopedia of Mathematics and its Applications, vol. 21, Cambridge University Press, 2001.

[116] H. D. Victory, Jr., *On nonnegative solutions of matrix equations*, SIAM J. Algebraic Discrete Methods **6** (1985), no. 3, 406–412.

[117] Y. V. Volkovich, N. Litvak, and D. Donato, *Determining factors behind the PageRank log-log plot*, Memorandum, University of Twente, February 2007.

[118] P. M. Weichsel, *The Kronecker product of graphs*, Proc. Amer. Math. Soc. **13** (1962), 47–52.

[119] J. Wolfowitz, *Products of indecomposable, aperiodic, stochastic matrices*, Proc. Amer. Math. Soc. **14** (1963), 733–737.