Noname manuscript No.
(will be inserted by the editor)

Robust PageRank: Stationary Distribution on a Growing Network Structure

Anna Timonina-Farkas

Received: date / Accepted: date

Abstract PageRank (PR) is a challenging and important network ranking algorithm, which plays a crucial role in information technologies and numerical analysis due to its huge dimension and wide range of possible applications. The traditional approach to PR goes back to the pioneering paper of S. Brin and L. Page [5], who developed the initial method in order to rank websites in the search engine results.
Recently, A. Juditsky and B. Polyak in the work [13] proposed a robust formulation of the PageRank model for the case, when links in the network structure may vary, i.e. some links may appear or disappear influencing the transportation matrix defined by the network structure. In this article, we make a further step forward, allowing the network to vary not only in links, but also in the number of nodes. We focus on growing network structures (e.g. Internet) and we propose a new robust formulation of the PageRank problem for uncertain networks with fixed growth rate (i.e. the expected number of pages which appear in the future is fixed). Further, we compare our results with the ranks estimated by the method of A. Juditsky and B. Polyak [13], as well as with the true ranks of tested network structures.
We formulate the robust PageRank in terms of non-convex optimization problems and we bound these formulations from above by convex but non-smooth optimization problems. In the numerical part of the article, we propose some smoothening techniques, which allow to obtain the solution accurately and efficiently in case of middle-size networks by the use of the well-known subgradient algorithm avoiding all non-smooth points. Furthermore, we address high-dimensional algorithms by modelling PageRank via the use of the multinomial distribution.

Keywords World Wide Web · PageRank · robust optimization · growing networks

The research is conducted in line with the Schrödinger Fellowship of the Austrian Science Fund (FWF)

Anna Timonina-Farkas, PhD
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne (RAO, EPFL)
EPFL-CDM-MTEI-RAO, Station 5, CH-1015 Lausanne, Switzerland
T: +41 (0) 21 693 00 36; F: +41 (0) 21 693 24 89
E-mail: anna.farkas@epfl.ch
Homepage: http://homepage.univie.ac.at/anna.timonina/

1 Introduction

PageRank (PR) is an automated information retrieval algorithm, which is designed by S. Brin and L. Page [5,21] during the rise of the World Wide Web in order to improve the quality of at the moment existing search engines. The algorithm relies not only on keyword matching but also on additional structure present in the hypertext providing higher quality search results.
Mathematically speaking, the rank of some page i, $\forall i = 1,...,N$ is the probability for a random user to be on this page. We denote the rank of the page i by x_i, $\forall i = 1,...,N$. Let $\mathscr{L}_i$ be the set of all web-pages, which refer to the page i. The probability, that a random user sitting on a page $j \in \mathscr{L}_i$ clicks on the link to the page i is equal to P_{ij} (Figure 1).

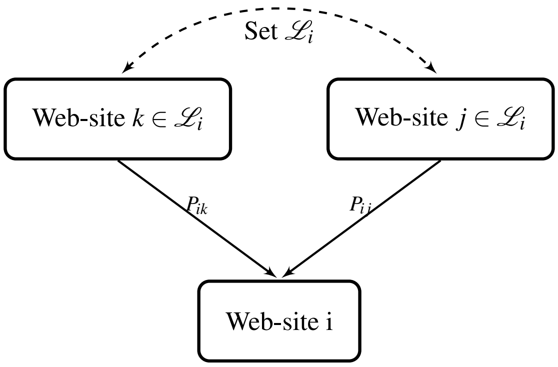


Fig. 1: Links outgoing from the web-site j.

The idea of the PageRank algorithm is, therefore, based on the interchange of ranks between pages, which is formulated in line with probability theory rules:

$$x_i = \sum_{j \in \mathscr{L}_i} P_{ij} x_j, \ \forall i = 1,...,N. \tag{1}$$

Intuitively, the equation (1) can be seen in the following way: the page j gives a part of its rank (or score) to the page i, if there is a direct link from the page j to the page i. The amount of the transferred score is proportional to the probability for a random user to click on the link to the page i sitting on the page j.
In the initial PageRank formulation [6,8,15,21], the probability P_{ij} is assumed to be an inverse of the total number of links outgoing from the page j denoted by n_j, i.e. $P_{ij} = \frac{1}{n_j}$. This assumption lacks realism, as it assigns equal probabilities for a random user to move to the page i from the page j. In reality, these probabilities are rank-dependent: the user is more likely to move to the page with a higher rank, than to the page with a lower rank.
If one denotes by P the *transportation (or transition) matrix* with entries $\{P_{ij}\}$, $\forall i, j =$

$1,\ldots,N$, the PageRank problem can be formulated as the problem of finding the *principal* eigenvector $\bar{x} = (\bar{x}_1,\ldots,\bar{x}_N)^T$ of this matrix, which exists due to the well-known Perron-Frobenius theorem [10] for non-negative matrices:

$$P\bar{x} = \bar{x}.$$

As the principal eigenvector corresponding the eigenvalue $\lambda = 1$ is not necessarily unique for non-negative stochastic matrices [10], the original PageRank problem is replaced with finding the principal eigenvector of the following modified matrix [5, 8,21] (known as *the Google matrix*):

$$G = \alpha P + (1-\alpha)S,$$

where $\alpha \in (0,\ 1)$ is the *damping factor* [5] and S is the doubly stochastic matrix of the form: $S = \left\{\frac{1}{N}\right\}_{ij},\ \forall i,j = 1,\ldots,N$.
In the pioneering paper of S. Brin and L. Page [5], the following intuitive justification is given to the matrix G: "We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. And, the α damping factor is the probability at each page the "random surfer" will get bored and request another random page."
The principal eigenvector $\bar{y} = (\bar{y}_1,\ldots,\bar{y}_N)$ of the matrix G is unique according to the Perron-Frobenius theorem for positive matrices [10]. Moreover, the well-known *power method* $y^{(k+1)} = Gy^{(k)}$ [22] converges to $\bar{y}$ for each starting value $y^{(0)}$ satisfying simplex constraints, which allows to work with high-dimensional cases. However, in the work of B. Polyak and A. Timonina (see [25]), it is shown that the solution $\bar{y}$ of the modified problem $G\bar{y} = \bar{y}$ can be far enough from the original $\bar{x}$ of the system $P\bar{x} = \bar{x}$. Moreover, for specific network structures it may happen that the eigenvector of the matrix G stops distinguishing between high-ranked pages [25], that build the core of search engine results: if a large amount of high-ranked pages obtains the same score, the ranking becomes ineffective, making the difference between "high-ranked" and "low-ranked" pages to be almost binary. Therefore, the approach with the perturbed matrix G works well only in the presence of a small number of high-ranked pages, which has been true when the initial algorithm [5] has been developed but which currently needs reconsideration as the amount of information on the web continues to grow and the number of experienced users in the art of web research is increasing.
In the work of B. Polyak and A. Timonina (see [25]), the authors propose an l_1-regularization method which avoids low-ranked pages:

$$\min_{x\in\Sigma_N} \left\{ ||Px - x||_2^2 + \varepsilon ||x||_1 \right\}, \tag{2}$$

where parameter ε regulates the number of high-ranked pages one would like to consider and where Σ_N denotes the standard simplex on $\mathbb{R}^N$, i.e. $\Sigma_N = \{v \in \mathbb{R}^N : \sum_{i=1}^N v_i,\ v_i \geq 0\}$. This optimization problem can be solved via the coordinate descent method analogous to the Gauss-Seidel algorithm for solving linear equations [6,15, 22,25,27], which allows to enhance efficiency of numerical ranking and to improve accuracy by differentiating between high-importance pages.

However, the optimization problem (2) has a drawback: it is not robust to variations in the Internet structure, while the World Wide Web is, clearly, changing over time: the amount of information on the web is growing, new pages appear, some web-sites (old or spam ones) disappear. This is particularly important due to the fact, that PageRank computation takes such a significant amount of time (Google re-evaluation of PageRank takes a month), that around 1-3% of new web-pages opens during this time with new links influencing the transportation matrix P.
In the work of A. Juditsky and B. Polyak [13], the robust reformulation of the PageRank model is proposed for the case, when links in the network structure may vary, i.e. some links may appear or disappear influencing the transition matrix P:

$$\min_{x\in\Sigma_N}\left\{\max_{\xi\in\mathscr{P}}||(P+\xi)x-x||_{(*)}\right\}, \tag{3}$$

where $||\cdot||_{(*)}$ is l_1- or l_2- norm, $\mathscr{P}$ is a set of allowed perturbations, such that the matrix $P+\xi$ stays column-stochastic, Σ_N is the standard simplex on $\mathbb{R}^N$. The optimization problem (3) can be bounded from above by a convex non-smooth optimization problem of the following type (see [2,3,13]):

$$\min_{x\in\Sigma_N}\left\{||Px-x||_{(*)}+\varepsilon||x||_{(*)}\right\}, \tag{4}$$

where ε is a parameter dependent on the perturbation set $\mathscr{P}$.
However, the formulation (3) lacks some realism: though the matrix $P+\xi$ is uncertain in this case, the dimension of the matrix stays the same, meaning that no pages may appear or disappear. In reality though, (i) the number of web-pages in the Internet grows and (ii) one does not explicitly know, how the structure of the Internet changes. In this article, we allow the network to vary not only in links, but also in the number of nodes, considering, therefore, growing network structures with transition matrix $Q(P)$:

$$\min_{x\in\Sigma_N}\left\{\max_{Q(P)\in\Xi}||Q(P)x-x||_{(*)}\right\}, \tag{5}$$

where $Q(P)$ depends on the matrix P. Further, Ξ is the perturbation set adjusted for the network growth so that the matrix Q stays column-stochastic.
In Section 2 we propose l_1-, l_2- and Frobenius-norm robust reformulations for PageRank of the uncertain network with fixed growth rate (i.e. the expected number of pages which appear in the future is fixed) and we bound these reformulations from above by convex but non-smooth optimization problems. In Section 3 we demonstrate that formultations robust to network growth impose upper bounds on the formulations robust to perturbations in links, i.e. on the formulations proposed by A. Juditsky and B. Polyak in the work [13]. In Section 4 we study properties of the chosen perturbation set, which helps to shed some light on the parameter ε of problems (2) and (4). In Section 5 we consider smoothening techniques for the subgradient method in order to solve formulated optimization problems in middle-dimensional cases. Further in the article, we address high-dimensional algorithms by modelling PageRank via the use of the multinomial distribution and, afterwards, we conclude, as well as we state directions for future research.

2 Problem Formulation

Consider a column-stochastic non-negative transportation matrix $P \in \mathbb{R}^{N\times N}$, which satisfies $P_{ij} \geq 0,\ \forall i,j = 1,...,N$ and $\sum_{i=1}^{N} P_{ij} = 1,\ \forall j = 1,...,N$ by definition. The well-known Perron-Frobenius theorem [10] states that there exists *dominant* (or *principal*) eigenvector $\bar{x} \in \Sigma_N$, where we denote by $\Sigma_N = \{v \in \mathbb{R}^N :\ \sum_{i=1}^{N} v_i = 1,\ v_i \geq 0\}$ the standard simplex on $\mathbb{R}^N$:

$$P\bar{x} = \bar{x}. \tag{6}$$

For each column j of the matrix P, the entry $P_{ij}, \forall i = 1,...,N$ corresponds to the transition probability from the node j to the node i of the network (i.e. the probability to move from the web-page j to the web-page i in the Internet). In general, matrix P is huge-dimensional and very sparse, as the number of links outgoing from every node is much smaller than the total number of nodes in the network: the average number of links outgoing from a web-page in the Internet is equal to 20 for the whole web (see [13]), while the total number of pages in the Internet is ca. 10^9. The dominant vector $\bar{x}$ describes the stationary distribution on the network structure: each element of the vector $\bar{x}$ denotes the rank of the corresponding node. For the Internet, the ranks can be seen as the time which an average user spends on the particular web-site.
The dominant vector $\bar{x}$ is not robust and may be highly vulnerable to small changes in matrix P. A. Juditsky and B. Polyak in their work [13] reformulated the problem in the robust optimization form, allowing matrix P to vary according to the law $P+\xi$ under some conditions on ξ. For this matrix, they found the stationary distribution $\tilde{x}$ robust to variations in links and, therefore, stable with respect to small changes of P. However, in their work, the size of matrix ξ was assumed to be the same as of matrix P, i.e. $N \times N$, meaning that growing in the number of nodes was not considered in the network (further, this formulation is referred to as *a fixed-size model*). In reality, though, changes of matrix P happen not only in links, but also in the number of nodes. The number of domains being registered per minute corresponds to the 1-3% growth of the Internet per month (`http://www.internetlivestats.com/total-number-of-websites/`). That is why, in this article we consider the following changes of matrix P:

$$Q = \begin{pmatrix} P+\xi & \zeta \\ \psi & \chi \end{pmatrix}, \tag{7}$$

where P is the column-stochastic transportation matrix describing the current state of the network with N pages; ξ is the matrix describing variations in links of the initial network; ζ, ψ and χ are matrices describing links to and from M new pages, which may appear in the future (ξ is of the size $N\times N$, ψ is of the size $M\times N$, ζ is of the size $N\times M$ and χ is of the size $M\times M$). In reality, $M \approx 0.03N$ per month.
As matrices P and Q must be column-stochastic, ξ, ζ, ψ and χ must satisfy the following properties:

$$\begin{cases} \xi_{ij} \geq -P_{ij},\ \forall i,j = 1,...,N; \\ \psi_{ij} \geq 0,\ \forall i = 1,...,M,\ j = 1,...,N; \\ \zeta_{ij} \geq 0,\ \forall i = 1,...,N,\ j = 1,...,M; \\ \chi_{ij} \geq 0,\ \forall i,j = 1,...,M, \end{cases} \tag{8}$$

saying that all elements of the matrix Q are non-negative, as well as the following properties must hold:

$$\begin{cases} \sum_{i=1}^{N}(P_{ij}+\xi_{ij})+\sum_{i=1}^{M}\psi_{ij}=1,\ \forall j=1,...,N \\ \sum_{i=1}^{N}\zeta_{ij}+\sum_{i=1}^{M}\chi_{ij}=1,\ \forall j=1,...,M \end{cases} \Leftrightarrow$$
$$\Leftrightarrow \begin{cases} \sum_{i=1}^{N}\xi_{ij}+\sum_{i=1}^{M}\psi_{ij}=0,\ \forall j=1,...,N \\ \sum_{i=1}^{N}\zeta_{ij}+\sum_{i=1}^{M}\chi_{ij}=1,\ \forall j=1,...,M, \end{cases} \tag{9}$$

saying that every column of the matrix Q sums up to 1 (notice, that here we use the fact that P is also column-stochastic and $\sum_{i=1}^{N}P_{ij}=1,\ \forall j=1,...,N$).
Similar to the work of A. Juditsky and B. Polyak [13], the function $\max_{Q\in\Xi}||Qx-x||_{(*)}$, where $\|\cdot\|_{(*)}$ is some norm, can be seen as a measure of "goodness" of a vector x as a common dominant eigenvector of the family Ξ, where Ξ stands for the set of perturbed stochastic matrices of the form (7) under conditions (8) and (9).
Further, let us denote by $x=\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}$ a feasible point, which is a candidate for the common dominant eigenvector of the family Ξ. Let $x^{(1)}$ be of the size $N\times 1$ and $x^{(2)}$ be of the size $M\times 1$. Hence, x is of the size $(N+M)\times 1$. Notice, that the vector x must belong to the standard simplex $\Sigma_{N+M}=\{v\in\mathbb{R}^{N+M}:\sum_{i=1}^{N+M}v_i=1,\ v_i\geq 0\}$, that means $x_i^{(1)}\geq 0,\ \forall i=1,...,N$, $x_j^{(2)}\geq 0,\ \forall j=1,...,M$ and $\sum_{i=1}^{N}x_i^{(1)}+\sum_{j=1}^{M}x_j^{(2)}=1$ (i.e. $||x^{(1)}||_1+||x^{(2)}||_1=1$).
We say that the vector $\hat{x}$ is a robust solution of the eigenvector problem on Ξ if

$$\hat{x}\in\underset{x\in\Sigma_{N+M}}{\text{Argmin}}\left\{\max_{Q\in\Xi}||Qx-x||_{(*)}\right\}, \tag{10}$$

where $\|\cdot\|_{(*)}$ is some norm (further, we consider l_1-, l_2- and Frobenius-norm robust formulations).
The reasonable choice of the uncertainty set Ξ would impose some bounds on the column-wise norms of matrices ξ, ζ, ψ and χ, meaning that the perturbation in links of the current and future states of the network would be bounded: i.e. $||[\xi]_j||\leq\varepsilon_j^{(\xi)}$, $||[\psi]_j||\leq\varepsilon_j^{(\psi)}$, $||[\zeta]_j||\leq\varepsilon_j^{(\zeta)}$ and $||[\chi]_j||\leq\varepsilon_j^{(\chi)}$, where $[\cdot]_j$ denotes the j-th column of a matrix. Moreover, the *total uncertainty budget* for matrices ξ, ζ, ψ and χ could be fixed (see [13]): this would imply constraints on the overall possible perturbations of the transportation matrix P.
By solving the optimization problem (10), one would protect the rank vector $\hat{x}$ against high fluctuation in case of link or node perturbations influencing the transportation matrix P. Currently, the Google scores vector is being updated once per month without accounting for the 1-3% growth rate during this month. By solving the optimization problem of the type (10) one could decrease the number of updates of the score vector and, therefore, reduce the underlying personnel and machinery costs.
In the following sections, we formulate robust optimization problems for finding the stationary distribution of the extended network Q for the case of l_1-, l_2- and Frobenius-norms, and we bound these problems from above by convex but non-smooth optimization problems.

2.1 l_1-norm formulation

Let us consider $\|\cdot\|_{(*)} = \|\cdot\|_1$. We say that the vector $\hat{x}^{(l_1)}$ is the l_1-robust solution of the eigenvector problem on the set of perturbed matrices $\Xi^{(l_1)}$, if

$$\begin{aligned}\hat{x}^{(l_1)} \in \underset{x\in\Sigma_{N+M}}{\text{Argmin}} \; \Big\{ & \max_{Q} ||Qx-x||_1 : \\ & \text{Q is column-stochastic}, \\ & ||[\xi]_j||_1 \le \varepsilon_j^{(\xi)}, \;\; \forall j = 1,...,N, \text{ and } \sum_{i,j} |\xi_{ij}| \le \varepsilon^{(\xi)}, \\ & ||[\psi]_j||_1 \le \varepsilon_j^{(\psi)}, \; \forall j = 1,...,N, \text{ and } \sum_{i,j} |\psi_{ij}| \le \varepsilon^{(\psi)}, \\ & ||[\zeta]_j||_1 \le \varepsilon_j^{(\zeta)}, \;\; \forall j = 1,...,M, \text{ and } \sum_{i,j} |\zeta_{ij}| \le \varepsilon^{(\zeta)}, \\ & ||[\chi]_j||_1 \le \varepsilon_j^{(\chi)}, \; \forall j = 1,...,M, \text{ and } \sum_{i,j} |\chi_{ij}| \le \varepsilon^{(\chi)} \Big\},\end{aligned} \tag{11}$$

where we have l_1-norm constraints on each column $[\cdot]_j$, $\forall j$ of matrices ξ, ζ, ψ and χ, as well as we bound the sum of absolute values of all elements of these matrices. These constraints bound the perturbation in links of the current and future states of the network. At the same time, the total uncertainty budget for matrices ξ, ζ, ψ and χ is fixed. Notice, that the problem (11) is not convex-concave, meaning that the function

$$\phi_1(x) = \max_{Q\in\Xi^{(l_1)}} \|Qx-x\|_1$$

cannot be computed efficiently.

Importantly, the formulation (11) does not discourage sparsity in transition probabilities. Consider, for example, the uncertainty matrix ψ, which describes links from N existing pages to M new ones. All elements of the matrix ψ are non-negative. If we reduce one positive element from j-th column of this matrix by a small enough δ, the norm $||[\psi]_j||_1$ and the sum $\sum_{i,j}|\psi_{ij}|$ decrease by this δ, regardless of the value of the element we decrease. This means, that l_1-norm formulation (11) does not make a preference which transition probabilities to decrease in order to satisfy the constraints. By this, a lot of transition probabilities of the matrix Q can result in being zeros. In contrast, for l_2- and Frobenius-norm formulations, which we consider in the next sections, the reduction of larger terms of the matrix ψ by δ results in a much greater reduction in norms than doing so with smaller terms. Therefore, l_2- and Frobenius-norm formulations discourage sparsity by yielding diminishing reductions for elements closer to zero.

Proposition 1 *Optimal value $\phi_1(\hat{x}^{(l_1)})$ of the non-convex optimization problem (11) can be bounded from above by the optimal value of the following convex optimization problem with $\varepsilon_1 = \varepsilon^{(\xi)} + \varepsilon^{(\psi)}$ and $\varepsilon_2^{(l_1)} = \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M$:*

$$\phi_1(\hat{x}^{(l_1)}) \le \min_{x\in\Sigma_{N+M}} \Big\{ ||Px^{(1)} - x^{(1)}||_1 + \varepsilon_1 \|x^{(1)}\|_{(a)} + \varepsilon_2^{(l_1)} \|x^{(2)}||_{(b)} \Big\}, \tag{12}$$

where $\|x^{(1)}\|_{(a)} = \min\limits_{\lambda+\mu=x^{(1)}} \left\{ \|\lambda\|_\infty + \sum_{j=1}^{N} \frac{\varepsilon_j^{(\xi)}+\varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)}+\varepsilon^{(\psi)}} |\mu_j| \right\}$,

$\|x^{(2)}\|_{(b)} = \min\limits_{\lambda+\mu=x^{(2)}} \left\{ \|\lambda\|_\infty + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} |\mu_j| \right\}$.

Proof See the Appendix 9.1 for the proof.

Notice, that $\|x^{(2)}\|_{(b)} = 0$ if there are no new pages in the network (i.e. if $M = 0$). In this case, the optimization problem (12) completely coincides with the l_1-reformulation proposed by A. Juditsky and B. Polyak in the work [13].

2.2 l_2-norm formulation

Let us consider $\|\cdot\|_{(*)} = \|\cdot\|_2$. We say that the vector $\hat{x}^{(l_2)}$ is the l_2-robust solution of the eigenvector problem on the set of perturbed matrices $\Xi^{(l_2)}$, if

$$
\begin{aligned}
\hat{x}^{(l_2)} \in \operatorname*{Argmin}_{x\in\Sigma_{N+M}} \Big\{ & \max_Q ||Qx-x||_2 : \\
& \text{Q is column-stochastic,} \\
& ||[\xi]_j||_1 \le \varepsilon_j^{(\xi)},\ \forall j=1,...,N, \text{ and } ||\xi||_F \le \varepsilon^{(\xi)}, \\
& ||[\psi]_j||_1 \le \varepsilon_j^{(\psi)},\ \forall j=1,...,N, \text{ and } ||\psi||_F \le \varepsilon^{(\psi)}, \\
& ||[\zeta]_j||_1 \le \varepsilon_j^{(\zeta)},\ \forall j=1,...,M, \text{ and } ||\zeta||_F \le \varepsilon^{(\zeta)}, \\
& ||[\chi]_j||_1 \le \varepsilon_j^{(\chi)},\ \forall j=1,...,M, \text{ and } ||\chi||_F \le \varepsilon^{(\chi)} \Big\},
\end{aligned} \tag{13}
$$

where we have constraints on j-th column $[\cdot]_j$ of matrices ξ, ζ, ψ and χ, as well as we have second-order constraints on matrices themselves. Notice, that the problem (13) is not convex-concave, meaning that the function

$$\phi_2(x) = \max_{Q\in\Xi^{(l_2)}} \|Qx-x\|_2$$

cannot be computed efficiently.

Proposition 2 *Optimal value* $\phi_2(\hat{x}^{(l_2)})$ *of the non-convex optimization problem (13) can be bounded from above by the optimal value of the following convex optimization problem with* $\varepsilon_1 = \varepsilon^{(\xi)} + \varepsilon^{(\psi)}$ *and* $\varepsilon_2^{(l_2)} = \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1$*:*

$$\phi_2(\hat{x}^{(l_2)}) \le \min_{x\in\Sigma_{N+M}} \left\{ ||Px^{(1)}-x^{(1)}||_2 + \varepsilon_1 \|x^{(1)}\|_{(c)} + \varepsilon_2^{(l_2)} ||x^{(2)}||_{(d)} \right\}, \tag{14}$$

where $\|x^{(1)}\|_{(c)} = \min\limits_{\lambda+\mu=x^{(1)}} \left\{ \|\lambda\|_2 + \sum_{j=1}^{N} \frac{\varepsilon_j^{(\xi)}+\varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)}+\varepsilon^{(\psi)}} |\mu_j| \right\}$,

$\|x^{(2)}\|_{(d)} = \min\limits_{\lambda+\mu=x^{(2)}} \left\{ \|\lambda\|_2 + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} |\mu_j| \right\}$.

Proof See the Appendix 9.2 for the proof.

2.3 Frobenius-norm formulation

Let us further consider $\|\cdot\|_{(*)} = \|\cdot\|_2$. We say that the vector $\hat{x}^{(F)}$ is the robust solution of the eigenvector problem on the set of perturbed matrices $\Xi^{(F)}$, if

$$\hat{x}^{(F)} \in \operatorname*{Argmin}_{x\in\Sigma_{N+M}} \Big\{ \max_{Q} ||Qx-x||_2 : \\ \text{Q is column-stochastic,} \\ ||\xi||_F \le \varepsilon^{(\xi)},\ ||\psi||_F \le \varepsilon^{(\psi)}, \\ ||\zeta||_F \le \varepsilon^{(\zeta)},\ ||\chi||_F \le \varepsilon^{(\chi)} \Big\}, \tag{15}$$

where $\|\cdot\|_F$ is the Frobenius norm. Notice, that the problem (15) is not convex-concave, meaning that the function

$$\phi_3(x) = \max_{Q\in\Xi^{(F)}} \|Qx-x\|_2$$

cannot be computed efficiently. Notice also, that the formulation (15) is an upper bound for the l_2-formulation (13).

Proposition 3 *Optimal value $\phi_3(\hat{x}^{(F)})$ of the non-convex optimization problem (15) can be bounded from above by the optimal value of the following convex optimization problem:*

$$\phi_3(\hat{x}^{(F)}) \le \min_{x\in\Sigma_{N+M}} \Big\{ \|Px^{(1)}-x^{(1)}\|_2 + \varepsilon_1\|x^{(1)}\|_2 + \varepsilon_2^{(F)}\|x^{(2)}\|_2 \Big\}, \tag{16}$$

where $\varepsilon_1 = \varepsilon^{(\xi)} + \varepsilon^{(\psi)}$ and $\varepsilon_2^{(F)} = \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1$.

Proof See the Appendix 9.3 for the proof.

Notice, that the parameter ε_1 is equal to $(\varepsilon^{(\xi)} + \varepsilon^{(\psi)})$ and does not depend on the problem formulation. This parameter describes the total uncertainty in links of current N pages, implied by the change in already existing links (i.e. $P+\xi$) and by the uncertainty ψ corresponding to newly appeared links from N existing to M new pages. However, parameters $\varepsilon_2^{(l_1)}$ and $\varepsilon_2^{(l_2)} = \varepsilon_2^{(F)}$ depend on the problem formulation. For the l_1-norm formulation the parameter $\varepsilon_2^{(l_1)}$ is equal to $(\varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M)$ and denotes the total uncertainty in links between M new pages, as well as from them. This parameter is clearly dependent on the number of new pages M, giving more weight to their ranks as the number grows. Differently, for the l_2- and Frobenius-norm formulations, parameters $\varepsilon_2^{(l_2)} = \varepsilon_2^{(F)}$ are equal to $(\varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1)$ and do not show explicit dependency on the number of new pages M.
Let us denote by ε_2 the following parameter:

$$\varepsilon_2 = \begin{bmatrix} \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M, \text{ for } l_1 - \text{norm formulation;} \\ \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1, \text{ for } l_2 - \text{and Frobenius-norm formulations.} \end{bmatrix}$$

In this case, optimization problems (12), (14) and (16) can be written in the following general form:

$$\tilde{x} \in \underset{x \in \Sigma_{N+M}}{\text{Argmin}} \left\{ \|Px^{(1)} - x^{(1)}\|_{(*)} + \varepsilon_1 \|x^{(1)}\|_{(1)} + \varepsilon_2 \|x^{(2)}\|_{(2)} \right\}, \tag{17}$$

where $\|\cdot\|_{(*)}$, $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ correspond to the norms from the formulations above. For the l_1-norm formulation (12), $\|\cdot\|_{(*)}$ defines the l_1-norm, $\|\cdot\|_{(1)}$ is equal to $\|\cdot\|_{(a)}$ and $\|\cdot\|_{(2)}$ is equal to $\|\cdot\|_{(b)}$. For the l_2-norm formulation (14), $\|\cdot\|_{(*)}$ defines the l_2-norm, $\|\cdot\|_{(1)}$ is equal to $\|\cdot\|_{(c)}$ and $\|\cdot\|_{(2)}$ is equal to $\|\cdot\|_{(d)}$. For the Frobenius-norm formulation (16), $\|\cdot\|_{(*)}$, $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ denote l_2-norms.
We refer to the vector $\tilde{x}$ as to the (computable) *robust dominant eigenvector* of the corresponding family of perturbed matrices $\Xi^{(l_1)}$, $\Xi^{(l_2)}$ or $\Xi^{(F)}$.
Further in the article and before we proceed with the numerical solution of the problem (17), we show that the formulation (17) provides the upper bound on the formulation of A. Juditsky and B. Polyak in the work [13]. Moreover, we discuss the choice of the perturbation set defined by parameters $\varepsilon^{(\xi)}$, $\varepsilon^{(\psi)}$, $\varepsilon^{(\zeta)}$ and $\varepsilon^{(\chi)}$.

3 Comparison to the model with fixed-size network

Optimization problems (11), (13) and (15) account for uncertainties in links between current and future (not yet existing) pages: these uncertainties are incorporated via matrices ξ, ψ, ζ and χ, which worst-case realization gives us the opportunity to compute the robust PageRank. These optimization problems differ from robust formulations corresponding to the fixed-size network model proposed by A. Juditsky and B. Polyak in the work [13], who studied uncertainties implied by the matrix $P+\xi$, describing variations in links of existing pages with constant network size. In this section, we study the relationship between the fixed-size and the growing network model. For this, we compute the lower bound of the norm $\max_{Q \in \Xi} ||Qx - x||$ for the case of l_1-, l_2- and Frobenius-norm and we prove that the growing network model imposes the upper bound for the fixed-size network model.

Theorem 1 (Upper Bounds) *Optimization problems (11), (13) and (15) under conditions (8) and (9) impose upper bounds on the fixed-size network model in the following sense:*

$$\phi_1(x) = \max_{Q \in \Xi^{(l_1)}} ||Qx - x||_1 \geq \max_{\substack{\mathbb{1}_N^T [\xi]_j = 0 \\ \|[\xi]_j\|_1 \leq \varepsilon_j^{(\xi)} \\ \sum_{i,j} |\xi_{ij}| \leq \varepsilon^{(\xi)}}} \|(P+\xi)x^{(1)} - x^{(1)}\|_1, \tag{18}$$

$$\phi_2(x) = \max_{Q \in \Xi^{(l_2)}} ||Qx - x||_2 \geq \max_{\substack{\mathbb{1}_N^T [\xi]_j = 0 \\ \|[\xi]_j\|_1 \leq \varepsilon_j^{(\xi)} \\ \|\xi\|_F \leq \varepsilon^{(\xi)}}} \|(P+\xi)x^{(1)} - x^{(1)}\|_2, \tag{19}$$

$$\phi_3(x) = \max_{Q \in \Xi^{(F)}} ||Qx - x||_2 \geq \max_{\substack{\mathbb{1}_N^T [\xi]_j = 0 \\ \|\xi\|_F \leq \varepsilon^{(\xi)}}} \|(P+\xi)x^{(1)} - x^{(1)}\|_2. \tag{20}$$

Proof Let $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, where u_1 is a vector of the length N and u_2 is a vector of the length M.
For the case of l_1-norm, the following equality holds

$$||Qx - x||_1 = \left\| \begin{pmatrix} P+\xi & \zeta \\ \psi & \chi \end{pmatrix} \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} - \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \right\|_1 = \left\| \begin{pmatrix} (P+\xi - I_N)x^{(1)} + \zeta x^{(2)} \\ \psi x^{(1)} + (\chi - I_M)x^{(2)} \end{pmatrix} \right\|_1 =$$
$$= \|(P+\xi - I_N)x^{(1)} + \zeta x^{(2)}\|_1 + \|\psi x^{(1)} + (\chi - I_M)x^{(2)}\|_1,$$

where I_N and I_M are identity matrices of the size $N \times N$ and $M \times M$ correspondingly.
Using the norm duality, i.e.

$$\|(P+\xi - I_N)x^{(1)} + \zeta x^{(2)}\|_1 = \max_{\substack{u_1 \in \mathbb{R}^N \\ \|u_1\|_\infty \le 1}} u_1^T \left((P+\xi - I_N)x^{(1)} + \zeta x^{(2)}\right)$$
$$\|\psi x^{(1)} + (\chi - I_M)x^{(2)}\|_1 = \max_{\substack{u_2 \in \mathbb{R}^M \\ \|u_2\|_\infty \le 1}} u_2^T \left(\psi x^{(1)} + (\chi - I_M)x^{(2)}\right),$$

we choose such feasible $u_1 = u_1^*$, that $\|(P+\xi)x^{(1)} - x^{(1)}\|_1 = (u_1^*)^T (P+\xi - I_N)x^{(1)}$, $u_1^* \in \mathbb{R}^N$ and $\|u_1^*\|_\infty \le 1$ and we fix $u_2 = \mathbb{1}_M$, where $\mathbb{1}_M$ is the $M \times 1$ vector of all-ones.
By this, we compute the lower bound for the norm $||Qx - x||_1$:

$$||Qx - x||_1 \ge (u_1^*)^T \left((P+\xi - I_N)x^{(1)} + \zeta x^{(2)}\right) + \mathbb{1}_M^T \left(\psi x^{(1)} + (\chi - I_M)x^{(2)}\right) =$$
$$= \|(P+\xi)x^{(1)} - x^{(1)}\|_1 + (u_1^*)^T \zeta x^{(2)} + \mathbb{1}_M^T \psi x^{(1)} + \mathbb{1}_M^T \chi x^{(2)} - \mathbb{1}_M^T x^{(2)} =$$
$$= \|(P+\xi)x^{(1)} - x^{(1)}\|_1 + \mathbb{1}_M^T \psi x^{(1)} + (u_1^* - \mathbb{1}_N)^T \zeta x^{(2)},$$

where the final equation holds due to equalities (21) and (22) with $\mathbb{1}_M^T \psi x^{(1)} \ge 0$ and $(u_1^* - \mathbb{1}_N)^T \zeta x^{(2)} \le 0$:

$$\mathbb{1}_M^T \psi = -\mathbb{1}_N^T \xi; \tag{21}$$
$$\mathbb{1}_M^T \chi = \mathbb{1}_M^T - \mathbb{1}_N^T \zeta. \tag{22}$$

Notice, that equalities (21) and (22) hold due to the column-stochasticity of the matrix Q (i.e. due to conditions (8) and (9)).
Therefore, we compute the lower bound for the function $\phi_1(x)$, i.e.

$$\phi_1(x) \ge \|(P+\xi)x^{(1)} - x^{(1)}\|_1 + \mathbb{1}_M^T \psi x^{(1)} + (u_1^* - \mathbb{1}_N)^T \zeta x^{(2)}, \ \forall \xi, \psi, \zeta \in \Xi^{(l_1)}$$

As soon as this bound holds for all ξ, ψ, ζ in the perturbation set, we can set $\zeta = 0$ and $\psi = 0$ without any loss of generality. Moreover, by setting $\psi = 0$ we guarantee that conditions of A. Juditsky and B. Polyak in the work [13] are satisfied, i.e. that the matrix $P + \xi$ is column-stochastic.
Therefore, we guarantee that the l_1-norm formulation of A. Juditsky and B. Polyak in the work [13] is a lower bound for our optimization problem, i.e. the bound (18) follows.
Analogically, one can show, that the same type of bound holds for the case of l_2- and

Frobenius-norm robust formulations.
Notice, that the following equality holds due to the duality of the second norm:

$$||Qx-x||_2 = \left\| \begin{pmatrix} P+\xi & \zeta \\ \psi & \chi \end{pmatrix} \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} - \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} (P+\xi-I)x^{(1)}+\zeta x^{(2)} \\ \psi x^{(1)}+(\chi-I)x^{(2)} \end{pmatrix} \right\|_2 =$$

$$= \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_2\le 1}} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^T \begin{pmatrix} (P+\xi-I_N)x^{(1)}+\zeta x^{(2)} \\ \psi x^{(1)}+(\chi-I_M)x^{(2)} \end{pmatrix}.$$

To compute the lower bound for the norm $||Qx-x||_2$, we choose such feasible $u_1 = u_1^*$, that $\|(P+\xi)x^{(1)}-x^{(1)}\|_2 = (u_1^*)^T(P+\xi-I)x^{(1)}$, $u_1^* \in \mathbb{R}^N$, $\|u_1^*\|_2 \le 1$ (which exists due to the duality of the norm) and we fix $u_2 = \mathbb{0}_M$ (i.e. zero-vector of the length M). In this case, we can write the following:

$$||Qx-x||_2 \ge (u_1^*)^T \Big((P+\xi-I_N)x^{(1)}+\zeta x^{(2)} \Big) = \|(P+\xi)x^{(1)}-x^{(1)}\|_2 + (u_1^*)^T \zeta x^{(2)},$$

which leads to the following bounds for l_2- and Frobenius-norm formulations correspondingly:

$$\max_{Q\in\Xi^{(l_2)}} ||Qx-x||_2 \ge \|(P+\xi)x^{(1)}-x^{(1)}\|_2 + (u_1^*)^T\zeta x^{(2)},\ \forall\, \xi \text{ and } \zeta \in \Xi^{(l_2)},$$

$$\max_{Q\in\Xi^{(F)}} ||Qx-x||_2 \ge \|(P+\xi)x^{(1)}-x^{(1)}\|_2 + (u_1^*)^T\zeta x^{(2)},\ \forall\, \xi \text{ and } \zeta \in \Xi^{(F)}.$$

Analogically to the l_1-norm formulation, we can set $\zeta = 0$ without any loss of generality, as the bounds hold for all ξ, ζ in the perturbation set. Therefore, one can guarantee that the l_2- and the Frobenius-norm formulations of A. Juditsky and B. Polyak in the work [13] impose lower bounds for our optimization problems, i.e. bounds (19) and (20) hold.

□

Now, consider robust reformulations (12), (14) and (16). In case $M = 0$ (i.e. if there is no growth in the network), these reformulations fully coincide with upper bounds proposed by A. Juditsky and B. Polyak in the work [13]. In general, for $M > 0$, robust reformulations (12), (14) and (16) differ from the bounds imposed by the fixed-size network. However, if there are no links from old *to* new web-pages, the current links are perturbed only by ξ, as $\psi = 0$. Therefore, Q is *reducible* and the difference between the fixed-size and the growing network models is fully imposed by the links from new pages.
If a random user starting from some new page (among $i = N+1,...,N+M$) keeps clicking on links, he eventually results in one of pages $i = 1,...,N$ as $\zeta \neq 0$. However, he is not able to get back to the starting page by clicking links as $\psi = 0$. Therefore, ranks $x^{(2)}$ should a priori be lower than $x^{(1)}$. Moreover, one cannot say which of pages $i = N+1,...,N+M$ have higher ranks and which have lower ones, as links are actually unknown to the search engine before the Internet structure is updated. Therefore, without loss of generality, the ranks $x^{(2)}$ could be assumed to be zeros and the problem could be solved numerically using algorithms proposed in [13].
Therefore, further we focus on the case of *irreducible* transition matrices Q, which imply $\varepsilon^{(\psi)} > 0$, $\varepsilon^{(\zeta)} > 0$, $\varepsilon^{(\chi)} > 0$.

4 Bounds on the perturbation set Ξ

Consider optimization problems (12), (14) and (16) in the general form (17):

$$\tilde{x} \in \underset{x \in \Sigma_{N+M}}{\text{Argmin}} \left\{ \|Px^{(1)} - x^{(1)}\|_{(*)} + \varepsilon_1 \|x^{(1)}\|_{(1)} + \varepsilon_2 \|x^{(2)}\|_{(2)} \right\}.$$

Notice, that $(\varepsilon^{(\xi)} + \varepsilon^{(\psi)})$ is denoted by ε_1, while ε_2 is equal to $(\varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M)$ for the l_1-norm robust formulation (12) and to $(\varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1)$ for both l_2- and Frobenius-norm formulations (14) and (16).
The size of the optimization problem (17) can be reduced, as the problem can be subdivided into two separate smaller-size optimization problems.

Lemma 1 *Let $\tilde{y}^{(1)}$ and $\tilde{y}^{(2)}$ be solutions of optimization problems (23) and (24):*

$$\tilde{y}^{(1)} \in \underset{y^{(1)} \in \Sigma_N}{\text{Argmin}} \left\{ \|Py^{(1)} - y^{(1)}\|_{(*)} + \varepsilon_1 \|y^{(1)}\|_{(1)} \right\}, \tag{23}$$

$$\tilde{y}^{(2)} \in \underset{y^{(2)} \in \Sigma_M}{\text{Argmin}} \left\{ \varepsilon_2 \|y^{(2)}\|_{(2)} \right\}, \tag{24}$$

where $\Sigma_N = \{\nu \in \mathbb{R}^N : \sum_{i=1}^N \nu_i,\ \nu_i \geq 0\}$ and $\Sigma_M = \{\nu \in \mathbb{R}^M : \sum_{i=1}^M \nu_i,\ \nu_i \geq 0\}$.
Then the following holds:

1. *If $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_{(*)} + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} < \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$, then the optimal solution to the problem (17) is $\tilde{x} = \begin{pmatrix} \tilde{x}^{(1)} \\ \tilde{x}^{(2)} \end{pmatrix}$ with $\tilde{x}^{(1)} = \tilde{y}^{(1)}$ and $\tilde{x}^{(2)} = 0$;*
2. *If $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_{(*)} + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} > \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$, then the optimal solution to the problem (17) is $\tilde{x} = \begin{pmatrix} \tilde{x}^{(1)} \\ \tilde{x}^{(2)} \end{pmatrix}$ with $\tilde{x}^{(1)} = 0$ and $\tilde{x}^{(2)} = \tilde{y}^{(2)}$;*
3. *If $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_{(*)} + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} = \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$, then there are infinitely many optimal solutions to the problem (17).*

Proof Consider the optimization problem (17). Based on the fact, that $\|x^{(1)}\|_1 + \|x^{(2)}\|_1 = 1$ and $x^{(1)} \geq 0$, $x^{(2)} \geq 0$, let us make the following change of variables with $s \in [0,\ 1]$:

$$\begin{aligned} x^{(1)} &= sy^{(1)}, \\ x^{(2)} &= (1-s)y^{(2)}, \end{aligned} \tag{25}$$

where $\|y^{(1)}\|_1 = 1$, $\|y^{(2)}\|_1 = 1$ and $y^{(1)} \geq 0, y^{(2)} \geq 0$. Further, $\|x^{(1)}\|_1 = s$ and $\|x^{(2)}\|_1 = 1 - s$, $\forall s \in [0,\ 1]$. Hence, simplex constraints on $x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}$ are satisfied.
Therefore, the optimization problem (17) can be rewritten as:

$$\tilde{y} \in \underset{\substack{y^{(1)} \in \Sigma_N \\ y^{(2)} \in \Sigma_M \\ s \in [0,\ 1]}}{\text{Argmin}} \left\{ s\left(\|Py^{(1)} - y^{(1)}\|_{(*)} + \varepsilon_1 \|y^{(1)}\|_{(1)} \right) + (1-s)\left(\varepsilon_2 \|y^{(2)}\|_{(2)} \right) \right\}, \tag{26}$$

where $\tilde{y} = \begin{pmatrix} \tilde{y}^{(1)} \\ \tilde{y}^{(2)} \end{pmatrix}$ and s^* denote the optimal solution of the optimization problem (26). Furthermore, $\tilde{y}^{(1)}$ and $\tilde{y}^{(2)}$ are independent of each other.
At optimality of the problem (26), $s^* = 1$ if $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_* + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} < \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$ and $s^* = 0$ if $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_* + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} > \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$. In case $\|P\tilde{y}^{(1)} - \tilde{y}^{(1)}\|_* + \varepsilon_1 \|\tilde{y}^{(1)}\|_{(1)} = \varepsilon_2 \|\tilde{y}^{(2)}\|_{(2)}$, s^* can take any value in the interval [0, 1].
Hence, the statement of the Lemma 1 follows. By comparing optimal values of problems (23) and (24), one can conclude the optimal solution s^* and, therefore, one can get the optimal solution of the problem (17) by the system (25).

□

In general, one would like to avoid the optimal solution $\tilde{x}^{(1)} = 0$, $\tilde{x}^{(2)} > 0$, as it would mean that the uncertainty about current pages is larger than the uncertainty about future (not yet existent) pages. For this, one needs to guarantee that the optimal value of the problem (23) is not greater than the optimal value of the problem (24) (i.e. one would like to avoid point 2 of the Lemma 1). Further, we consider l_1-, l_2- and Frobenius-norm formulations (12), (14) and (16) and we explicitly solve the optimization problem (24) for each of these norms. Moreover, we state conditions on parameters $\varepsilon^{(\xi)}$, $\varepsilon^{(\psi)}$, $\varepsilon^{(\zeta)}$ and $\varepsilon^{(\chi)}$, which guarantee that points 1 or 3 of the Lemma 1 are sufficiently satisfied.

Statement 1 *Consider the robust reformulation (12) and apply the proposed change of variables (25). In this case,* $\|y^{(2)}\|_{(2)} = \min_{\lambda+\mu=y^{(2)}} \left\{ \|\lambda\|_\infty + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} |\mu_j| \right\}$ *and the following holds:*

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(2)} = \begin{bmatrix} \frac{1}{M}, \text{ if } \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} \geq \frac{1}{M}, \ \forall j = 1, ..., M \\ \min_j \left\{ \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} \right\}, \text{ if } \min_j \left\{ \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} \right\} < \frac{1}{M}. \end{bmatrix} \tag{27}$$

Proof See the Appendix 9.4 for the proof.

Statement 2 *Consider the robust reformulation (14) and apply the proposed change of variables (25). In this case,* $\|y^{(2)}\|_{(2)} = \min_{\lambda+\mu=y^{(2)}} \left\{ \|\lambda\|_2 + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} |\mu_j| \right\}$ *and the following holds:*

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(2)} = \begin{bmatrix} \frac{1}{\sqrt{M}}, \text{ if } \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} \geq \frac{1}{\sqrt{M}}, \ \forall j = 1, ..., M \\ \min_j \left\{ \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} \right\}, \text{ if } \min_j \left\{ \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} \right\} < \frac{1}{\sqrt{M}}. \end{bmatrix} \tag{28}$$

Proof See the Appendix 9.5 for the proof.

Statement 3 *Consider the robust reformulation (16) and apply the proposed change of variables (25). In this case,* $\|y^{(2)}\|_{(2)} = \|y^{(2)}\|_2$ *and the following holds:*

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(2)} = \frac{1}{\sqrt{M}}. \tag{29}$$

Proof The proof obviously follows from the direct minimization of the l_2-norm.

Statements 1, 2 and 3 impose conditions on the optimal value of the problem (23), under which points 1 or 3 of the Lemma 1 are satisfied. These conditions vary for l_1-, l_2- and Frobenius-norm formulations and can be stated in the following way:

For the l_1-case:

$$\begin{cases} \min\limits_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_1+\varepsilon_1\|y^{(1)}\|_{(a)}\right\}\leq \frac{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M}{M}, \\ \min\limits_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_1+\varepsilon_1\|y^{(1)}\|_{(a)}\right\}\leq \min_j\left\{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1\right\}. \end{cases} \tag{30}$$

For the l_2-case:

$$\begin{cases} \min\limits_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_2+\varepsilon_1\|y^{(1)}\|_{(c)}\right\}\leq \frac{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1}{\sqrt{M}}, \\ \min\limits_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_2+\varepsilon_1\|y^{(1)}\|_{(c)}\right\}\leq \min_j\left\{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1\right\}. \end{cases} \tag{31}$$

For the Frobenius-norm:

$$\min_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_2+\varepsilon_1\|y^{(1)}\|_2\right\}\leq \frac{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1}{\sqrt{M}}. \tag{32}$$

Notice, that $\varepsilon_1=\varepsilon^{(\xi)}+\varepsilon^{(\psi)}$.

Theorem 2 (Sufficient Conditions) *Consider statements (30), (31) and (32).*

1. *Condition (30) is sufficiently satisfied, if* $\varepsilon^{(\xi)}+\varepsilon^{(\psi)}\leq 1$.
2. *Condition (31) is sufficiently satisfied, if* $\begin{cases}\varepsilon^{(\xi)}+\varepsilon^{(\psi)}\leq 1,\\ \varepsilon^{(\zeta)}+\varepsilon^{(\chi)}\geq\sqrt{M}-1.\end{cases}$
3. *Condition (32) is sufficiently satisfied, if* $\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}\geq(\varepsilon^{(\xi)}+\varepsilon^{(\psi)})\sqrt{M}-1$.

Proof Consider l_1-norm and let $\bar{y}^{(1)}$ be such a vector, that $\bar{y}^{(1)}=P\bar{y}^{(1)}$ and $\bar{y}^{(1)}\in\Sigma_N$ (notice, that it exists due to the Perron-Frobenius theorem). In this case,

$$\min_{y^{(1)}\in\Sigma_N}\left\{\|Py^{(1)}-y^{(1)}\|_1+\varepsilon_1\|y^{(1)}\|_{(a)}\right\}\leq\varepsilon_1\|\bar{y}^{(1)}\|_{(a)}\leq\varepsilon_1=\varepsilon^{(\xi)}+\varepsilon^{(\psi)},$$

where the last inequality holds due to the definition of the norm $\|\cdot\|_{(a)}$.
Therefore, condition (30) is satisfied, if

$$\begin{cases}\varepsilon^{(\xi)}+\varepsilon^{(\psi)}\leq 1+\frac{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}}{M},\\ \varepsilon^{(\xi)}+\varepsilon^{(\psi)}\leq 1+\min_j\left\{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}\right\}.\end{cases}$$

It is, therefore, sufficient that $\varepsilon^{(\xi)}+\varepsilon^{(\psi)}\leq 1$ in the absence of information about new pages ambiguity parameters (i.e. about $\varepsilon_j^{(\zeta)}$ and $\varepsilon_j^{(\chi)}$). Hence, the statement 1 of the Theorem 2 follows.
Analogically, we obtain the sufficient conditions for l_2- and Frobenius-norm reformulations, i.e. statements 2 and 3 of the Theorem 2.

□

Condition 1 of the Theorem 2 depends neither on the number M of new pages in the network nor on the uncertainty levels $\varepsilon^{(\zeta)}$ and $\varepsilon^{(\chi)}$, which makes the l_1- formulation convenient for the analysis. For the case of l_2- robust reformulation, condition 2 of the Theorem 2 states that the uncertainty about future pages $\varepsilon^{(\zeta)} + \varepsilon^{(\chi)}$ should grow faster than $(\sqrt{M} - 1)$ as the number M of new pages increases. For the case $M = 0$ and $M = 1$, this condition is automatically satisfied. For higher dimensions (i.e. $M > 1$), this condition can be statistically tested from real-world data. Similarly, for the Frobenius-norm formulation (16), condition 3 of the Theorem 2 requires that the uncertainty about future pages $\varepsilon^{(\zeta)} + \varepsilon^{(\chi)}$ grows faster than $(\varepsilon^{(\xi)} + \varepsilon^{(\psi)})\sqrt{M} - 1$ as the number M of new pages increases.
Notice, that in the work of A. Juditsky and B. Polyak [13] small perturbations $\varepsilon_1 = 0.01$ were considered in order to avoid complete "equalization" of the scores. Moreover, $M = 0$ in the work [13]. Thus, conditions 1, 2 and 3 of the Theorem 2 were satisfied in the work of A. Juditsky and B. Polyak [13].
Further, let us assume that conditions of the Theorem 2 are satisfied. In this case, we focus on the numerical solution of the problem (23), which we formulate in the following general form:

$$\min_{x \in \Sigma_N} \left\{ \|Px - x\|_{(*)} + \varepsilon \|x\|_{(1)} \right\}, \tag{33}$$

where we, with some abuse of notations, denote $\tilde{y}^{(1)}$ by x and ε_1 by ε and where norms $\|\cdot\|_{(*)}$ and $\|\cdot\|_{(1)}$ correspond to the norms in l_1-, l_2- and Frobenius-norm formulations (12), (14) and (16).
In the following section, we study numerical techniques for the solution of the optimization problem (33).

5 Numerical algorithms

Consider optimization problems (12), (14) and (16). As it is demonstrated in the previous section for each of these problems, it is sufficient to solve the optimization problem of the type (33) if conditions (30), (31) or (32) are correspondingly satisfied. In medium- to large-dimensional cases (i.e. N = 1.e3 – 1.e6), convex non-smooth optimization problem (33) can be solved numerically using available optimization techniques, including interior-point methods, mirror-descent algorithms [14, 19] and randomized techniques [11, 12, 18]. In huge-dimensional cases (i.e. N = 1.e6 – 1.e9), one can employ randomized subgradient algorithms, which are specifically designed for sparse matrices by Y. Nesterov [20]. Moreover, one can use less accurate but extremely fast numerical methods proposed in [13, 17, 25].
In this section, we do not consider these algorithms, but we propose some techniques, which allow to smoothen optimization problems (12), (14), (16) and to solve them numerically via approximation.
Further, we consider the optimization problem (33) with $\|\cdot\|_{(*)} = \|\cdot\|_2$ and $\|\cdot\|_{(1)} = \|\cdot\|_2$, i.e. we focus on the Frobenius-norm formulation (16) under conditions of the Theorem 2. We choose this formulation, as it imposes the upper bound on the l_2- formulation (14) and as it is a non-smooth non-separable optimization problem, which

implies additional difficulty.
Differently, one could use the following upper bound for the approximate solution of the optimization problem (14):

$$\min_{\substack{x\in\Sigma_N\\ \lambda+\mu=x}}\left\{\|Px-x\|_2+\varepsilon\|\lambda\|_2+\sum_{j=1}^{N}\varepsilon_j|\mu_j|\right\}\le\min_{x\in\Sigma_N}\left\{\|Px-x\|_2+\sum_{j=1}^{N}\varepsilon_j x_j\right\},$$

which we do not consider in this article but which can be approached analogically to the Frobenius-norm formulation.
We also do not consider numerical algorithms for the solution of l_1- norm formulation (12), as this problem can be bounded by the optimization problem with a separable non-smooth part and, therefore, it can be solved approximately via the well-known projected coordinate descent algorithm (see, for example, [25]).
Therefore, let us consider the optimization problem (16). Its reformulation (23) under condition (32) implies the following optimization problem:

$$\tilde{x}\in\operatorname*{Argmin}_{x\in\Sigma_N}\left\{\|Px-x\|_2+\varepsilon\|x\|_2\right\}, \tag{34}$$

where $\varepsilon=\varepsilon^{(\xi)}+\varepsilon^{(\psi)}$.
We solve the optimization problem (34) using the following steps:

Step 1: Apply a *nonstandard normalization* instead of simplex constraints $x\in\Sigma_N$. This would allow us to simplify the constraints of the optimization problem (34);
Step 2: Bound the feasible set of the problem (34) so, that it does not include any of non-smooth points;
Step 3: Solve the optimization problem (34) via *projected subgradient method* on the feasible set.

Further, we consider each of these steps in more details.

Nonstandard normalization: Assume that one page (page N) is known to have the highest rank (see [25]) and, therefore, let us use the nonstandard normalization $x_N=1$ instead of $\sum_{i=1}^N x_i=1$. By this, we introduce the following optimization problem:

$$z\in\operatorname*{Argmin}_{x\in\mathbb{X}} f(x),\ f(x)=\|Px-x\|_2+\varepsilon\|x\|_2, \tag{35}$$

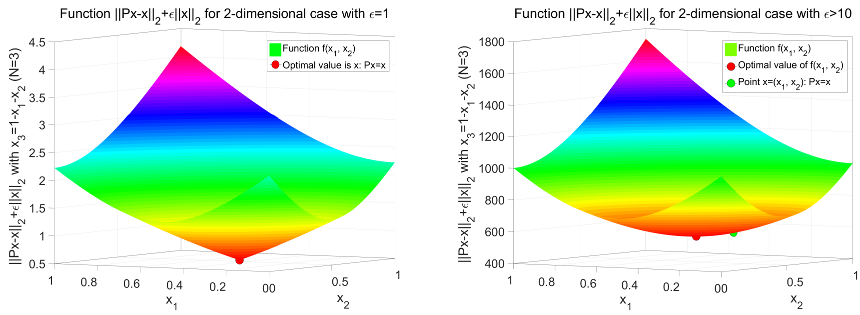
where $\mathbb{X}=\{x\in\mathbb{R}^N,\ x_N=1,\ x\ge 0\}$.
Notice, that the optimal value and the optimal solution of the problem (34) are related to those of the problem (35) in the following way:

$$\begin{cases} f(\tilde{x})=\tilde{x}_N f(z),\\ \tilde{x}=\tilde{x}_N z,\end{cases}$$

where $z=\begin{pmatrix}\frac{\tilde{x}_1}{\tilde{x}_N}\\ \frac{\tilde{x}_2}{\tilde{x}_N}\\ \vdots\\ 1\end{pmatrix}$. This leads to the statement $\tilde{x}_N=\frac{1}{\sum_{i=1}^N z_i}$.

Non-smooth points: The optimization problem (35) is non-smooth at least at one point $P\bar{x} = \bar{x}$, which is a possible optimal solution corresponding to the dominant (or principal) eigenvector of the non-perturbed matrix P. Therefore, we first of all check if the optimal solution of the problem (35) is always the point $\bar{x}$: $P\bar{x} = \bar{x}$. For this, we solve the optimization problem (35) in small-dimensional cases with randomly chosen non-negative column-stochastic matrix P. We can see that the following two cases are possible:

1. Optimal solution of the problem (35) is the point $\bar{x}: \; P\bar{x} = \bar{x}$ (Figure 2 (a)), in which case the function $f(x)$ is non-smooth at optimality;
2. Optimal solution of the problem (35) is a point $x: \; Px \neq x$ (Figure 2 (b)), in which case the function $f(x)$ is smooth at optimality.



(a) Non-smooth optimality. (b) Smooth at optimality.

Fig. 2: Function $f(x)$ for randomly chosen P in small-dimensional cases.

The optimal solution of the problem (35) is not necessarily the principal eigenvector of the matrix P. Therefore, there may exist at least one not optimal point, where the function is non-smooth.

Further, notice that the optimization problem (35) is non-smooth $\forall \bar{x}$ such that $P\bar{x} = \bar{x}$. As soon as the matrix P is non-negative and column-stochastic, its maximal eigenvalue ($\lambda = 1$) is not necessarily a simple root of the characteristic polynomial according to the Perron-Frobenius theorem [10]. Therefore, there may exist multiple number of points $\bar{x}: \; P\bar{x} = \bar{x}$. If the matrix P would be irreducible, the maximal eigenvalue ($\lambda = 1$) would be the simple root of the characteristic polynomial. However, there still could exist multiple complex eigenvalues with absolute value 1, which would strongly influence convergence of numerical algorithms (for example, the well-known *power method* would not converge).

We propose a technique, which guarantees that at each iteration k the objective function is smooth at the point $x^{(k)}$.

Notice, that the subgradient of the function $f(x)$ at any point $x: \; Px \neq x$ is unique and is equal to:

$$\partial f(x) = \frac{(P-I)^T(P-I)x}{||Px-x||_2} + \varepsilon \frac{x}{\|x\|_2}, \text{ where } Px \neq x. \tag{36}$$

Notice also, that $||x||_2 > 0$ for all points in the feasible set of the problem (35), as $x_N = 1$.
Now, consider a perturbed Google matrix $G = \alpha P + (1-\alpha)S$, where S is the doubly stochastic matrix with entries equal to $\frac{1}{N}$ and where $\alpha = 0.85$ is the damping factor [9,16]. This matrix was initially proposed by S. Brin and L. Page [5] and was used by Google in the well-known *PageRank algorithm* $y^{(k+1)} = Gy^{(k)}$. For the matrix G one can guarantee the uniqueness of the maximal (in absolute value) eigenvalue (i.e. $|\lambda| = 1$) and, therefore, the uniqueness of the eigenvector $\bar{y}$ corresponding to it [25]. Moreover, one can guarantee the convergence of the power method $y^{(k+1)} = Gy^{(k)}$ to $\bar{y}: \; G\bar{y} = \bar{y}$ due to the Perron-Frobenius theorem for positive matrices.
Further, consider the following convex function and its corresponding subgradient:

$$g(x) = ||Gx - x||_2 + \varepsilon ||x||_2,$$
$$\partial g(x) = \frac{(G-I)^T(G-I)x}{||Gx - x||_2} + \varepsilon \frac{x}{||x||_2}.$$

Notice, that the function $g(x)$ has a unique non-smooth point $\bar{y}$, which corresponds to $G\bar{y} = \bar{y}$ and which, in general, may differ from $\bar{x}: \; P\bar{x} = \bar{x}$.
The Perron-Frobenius theorem for positive matrices implies that the principal eigenvector of the matrix G has only positive entries. Hence, one can guarantee that $Gx \neq x$ for each chosen x by setting ranks of some spam pages to zero (i.e. $x_i = 0$ for some i). By this, one guarantees the uniqueness of the subgradient at each feasible point.
Therefore, we solve the following optimization problem numerically:

$$\min_{x \in \bar{\mathbb{X}}} g(x), \; g(x) = \|Gx - x\|_2 + \varepsilon \|x\|_2, \tag{37}$$

where $\bar{\mathbb{X}} = \{x \in \mathbb{R}^N, \, x_1 = 0, \, x_N = 1, \, x \geq 0\}$, where x_1 corresponds to the spam page.

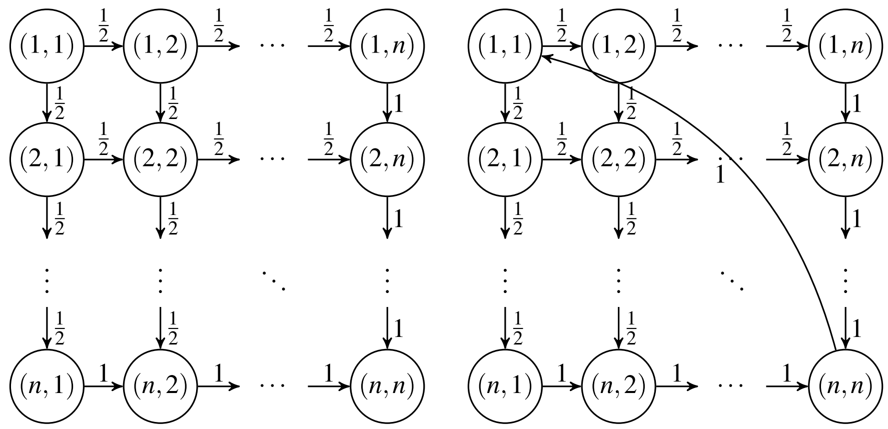
Subgradient method: We solve the optimization problem (37) by the projected subgradient algorithm, where we do not iterate over elements x_1 and x_N:

$$x_i^{(k+1)} = \max\left\{ x_i^{(k)} - t_k \left(\frac{([G]_i - e_i, Gx - x)}{||Gx - x||_2} + \varepsilon \frac{x_i}{||x||_2} \right), \, 0 \right\}, \; \forall i = \overline{2, N-1}, \tag{38}$$

where $[G]_i$ is the i-th column of the matrix G, e_i is the i-th column of the matrix I_N and t_k is the chosen step size. In the method (38), the subgradient is unique at each feasible point of the problem (37).
In general, subgradient methods are not necessarily the descent methods. Therefore, one keeps track of the best function value at every iteration k. i.e. $g^{best} = \min_{i \in \{1,2,\ldots,k\}} g(x^{(i)})$. In this article, we test the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$ instead of keeping track of the best function value, as well as we do not iterate over pages with zero-ranks. This allows to enhance efficiency of the algorithm.

5.1 Numerical results

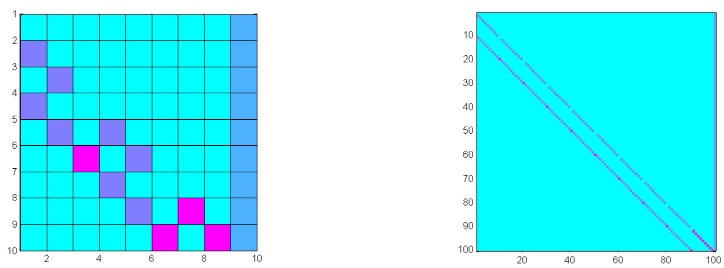
Consider the following two ranking models (Figure 3 (a) and (b)), which are also discussed in the works [13,25,28]. For both models, nodes (e.g. web-pages) are shown as circles, while references (e.g. web-links) are denoted by arrows.



(a) Model 1. (b) Model 2.

Fig. 3: Network models for testing purposes.

Model 1: Let us start by considering Model 1 (Figure 3 (a)). In this model, node (n,n) represents a *dangling* vertex, which makes the transition matrix P reducible, leading to non-uniqueness of the eigenvector corresponding to the eigenvalue $\lambda = 1$. To avoid reducibility of the matrix P, we need to guarantee that the number of outgoing links is non-zero for each page. For this, we assume the ability of equally probable transitions from the node (n,n) to any node in the Model 1 similarly to the approach used by search engines. The transition matrix P corresponding to the Model 1 with equally probable transitions from the vertex (n,n) becomes *irreducible* and *aperiodic*, which guarantees the uniqueness of the eigenvector corresponding to the eigenvalue $\lambda = 1$, as well as the convergence of the well-known *power method* $x^{(k+1)} = Px^{(k)}$ [13,25, 28].
Figure 4 demonstrates the sparsity of the matrix P corresponding to the Model 1 for networks with $N = n^2 = 9$ and $N = n^2 = 100$ nodes: zero elements of the matrix P are shown with light blue color. The matrix P is, clearly, highly sparse.



(a) N=9. (b) N=100.

Fig. 4: Non-zero elements of the transition matrix P corresponding to the Model 1.

We are interested in finding ranks $x_{i,j}$ of each node of the network.
Taking arbitrary value of $x_{n,n}$, say $x_{n,n} = n^2$, we obtain the system of equations describing ranks of the Model 1 [25]:

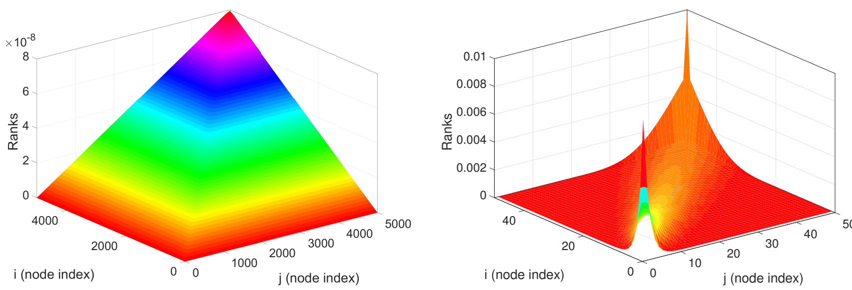
$$\begin{cases} x_{1,1} = \frac{1}{n^2} x_{n,n} = 1, \\ x_{i,1} = x_{1,i} = \frac{1}{2} x_{i-1,1} + 1, \quad i = 2, \ldots, n, \\ x_{i,j} = \frac{1}{2} x_{i-1,j} + \frac{1}{2} x_{i,j-1} + 1, \quad j, i = 2, \ldots, n-1, \\ x_{n,j} = x_{j,n} = x_{n,j-1} + \frac{1}{2} x_{n-1,j} + 1, \quad j = 2, \ldots, n-1, \\ x_{n,n} = x_{n-1,n} + x_{n,n-1} + 1 = n^2. \end{cases} \tag{39}$$

The system of equations (39) has a closed form solution and, therefore, it can be solved explicitly for each $x_{i,j}$ [25,28]. This allows us to test the performance of the subgradient algorithm (38) in comparison with the true ranks of the Model 1.

Model 2: The second model (i.e. Model 2 in the Figure 3 (b)) differs from the Model 1, as it has only one link from the node (n,n). As the transition matrix corresponding to the Model 2 is *periodic*, the power method $x^{(k+1)} = Px^{(k)}$ does not converge for it [25,28].
Taking arbitrary value of $x_{1,1}$, we obtain the system of equations describing ranks of the Model 2 [25]:

$$\begin{cases} x_{i,1} = x_{1,i} = \frac{1}{2} x_{i-1,1}, \quad i = 2, \ldots, n, \quad x_{1,1} = 1, \\ x_{i,j} = \frac{1}{2} x_{i-1,j} + \frac{1}{2} x_{i,j-1}, \quad j, i = 2, \ldots, n-1, \\ x_{n,j} = x_{jn} = x_{n,j-1} + \frac{1}{2} x_{n-1,j}, \quad j = 2, \ldots, n-1, \\ x_{n,n} = x_{n-1,n} + x_{n,n-1}. \end{cases} \tag{40}$$

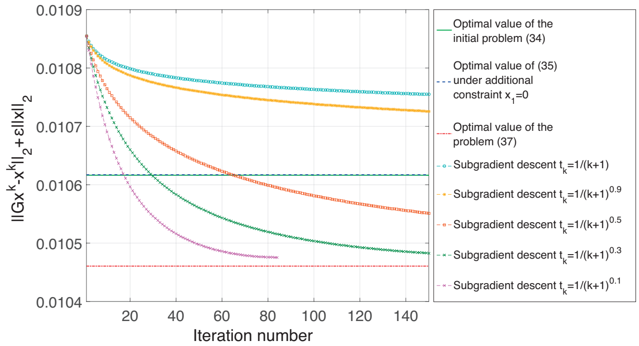
Analogically to the Model 1, the system of equations (40) can be solved explicitly.
Further, we proceed to standard normalization and get the normalized ranks as $x^*_{i,j} = \frac{x_{i,j}}{\sum_{i,j} x_{i,j}}$ for both Model 1 and Model 2 (see Figure 5). Notice, that these ranks correspond to the dominant eigenvector of the matrix P. However, they are not necessarily the optimal solution of optimization problems (34) or (37).



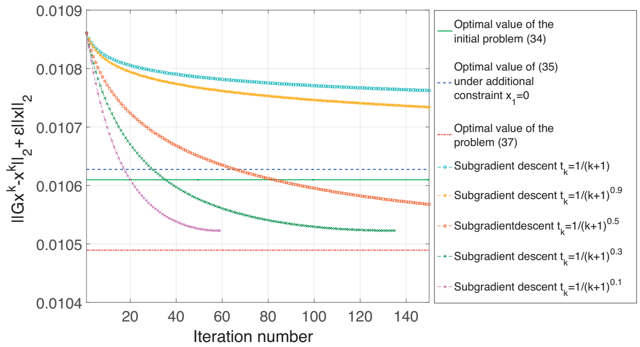
(a) Model 1. (b) Model 2.

Fig. 5: Explicit ranks of (a) Model 1 and (b) Model 2 ($N = 5000^2$ and $N = 50^2$ correspondingly).

For $N = 10.000$ we (i) compute exact ranks of the described models (i.e. Model 1 and Model 2) via systems (39) and (40). Afterwards, we (ii) solve the reformulation (37) by the algorithm (38) described in the Section 5. Further, we test different possible step sizes t_k in the algorithm (38).
For Model 1 and Model 2 correspondingly, Figure 6 demonstrates the optimal value convergence of the problem (37) obtained by the algorithm (38) with the *diminishing step size* (i.e. $\lim_{k\to\infty} t_k = 0$, $\sum_{k=1}^{\infty} t_k = \infty$) and with the starting point $x = (0,\ 1, \ldots, 1)^T$.



(a) Model 1.



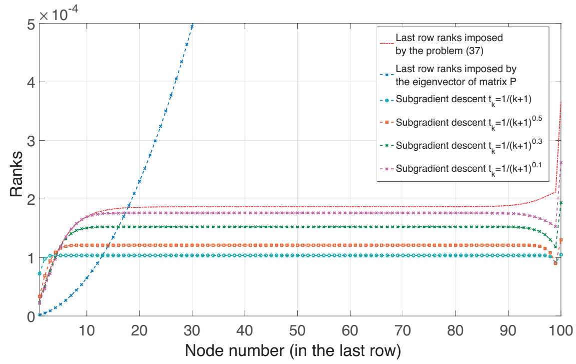
(b) Model 2.

Fig. 6: Optimal value convergence for the problem (37) obtained by the algorithm (38) with diminishing step size $t_k = \frac{1}{(k+1)^d}$, $\forall k$ with $\varepsilon = 1$ and $d \in (0,1)$.

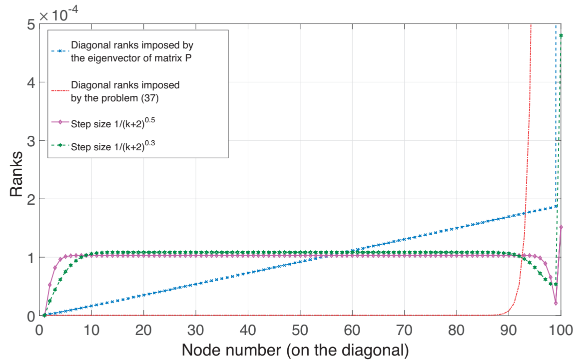
As the subgradient method is not necessarily the descent method, one should, in general, keep track of the best function value at every iteration k: $g^{best} = \min_{i\in\{1,2,\ldots,k\}} g(x^{(i)})$.
We, however, test the performance of the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$ instead of keeping track of the best function value. This allows to enhance efficiency of the algorithm (Figure 6). In the Figure 6, one can see that larger step sizes lead to the earlier

break of iterations under the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$.

Figure 7 compares ranks estimated via algorithm (38) with the dominant eigenvector of the matrix P.



(a) Last row elements.

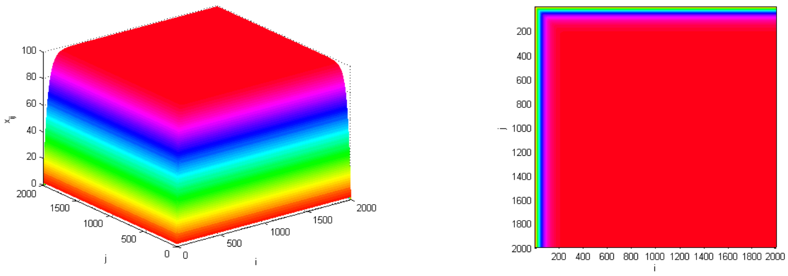


(b) Diagonal elements.

Fig. 7: **Model 1**: Estimated ranks v.s the dominant eigenvector.

In the example of Figure 7 we use $\varepsilon = 1$ as the parameter of the optimization problem (37). For larger values of ε, the robust eigenvector stops distinguishing ranks of high-importance pages. This is in line with conditions (30), (31) and (32) and the Theorem 2, which claim that the uncertainty about future pages becomes dominant if the value of $\varepsilon_1 = \varepsilon$ gets too high, which makes ranks of current pages indistinguishable from each other. Therefore, increasing the value of ε even further would lead to the solution, where one cannot differ between ranks of pages $x_2,...,x_{N-1}$ at all (notice, that $x_1 = 0$ and $x_N = 1$ are fixed in the optimization problem (37)).
For small values of ε (e.g. $\varepsilon << 1$), the optimal solution of the problem (34) should approach the dominant eigenvector of the unperturbed transition matrix P. As we solve the optimization problem (37) instead of the problem (34) for numerical pur-

poses, our optimal solution approaches the unique dominant eigenvector of the Google matrix $G = \alpha P + (1-\alpha)S$, with S being the doubly stochastic matrix with elements $\frac{1}{N}$ and the damping factor $\alpha = 0.85$.
Figure 8 (a) demonstrates values $x_{i,j}$, $\forall i, j = 1, ..., n$ of the dominant eigenvector of the matrix G with $\alpha = 0.99$ for Model 1. Large amount of high-importance pages have the same rank already for $\alpha = 0.99$, which can be also observed in the Figure 8 (b) (view from above) [25,28]. Even more pages become indistinguishable if we decrease the damping factor α.



(a) Ranks $x_{i,j}$, $\forall i, j = 1, ..., n$. (b) View from above.

Fig. 8: Eigenvector of the Google matrix $G = \alpha P + (1-\alpha)S$ with $\alpha = 0.99$ corresponding to the Model 1.

By this, one could claim that the dominant eigenvector of the matrix G should be viewed as the robust eigenvector of the perturbed family of matrices Q defined by (7), (8) and (9). This, however, would not provide the methodology to rank high-importance pages, which are the core of all search engine results. For small- and medium-size problems, the optimal solution of the optimization problem (37) with $\varepsilon \leq 1$ provides better results in terms of ranking of high-importance pages than the direct use of the transition matrix G.
Further, we provide the results for the subgradient method (38) for the following step sizes for Model 2 with $N = 10.000$:

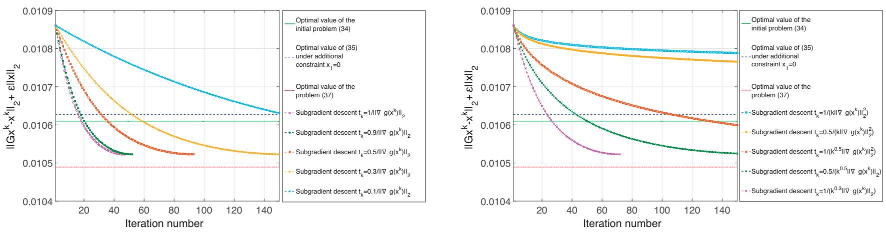
Constant step length: $t_k = \frac{h}{||\partial g(x^{(k)})||_2}$, $\forall k$;

Polyak's step size: $t_k = \frac{g(x^{(k)}) - g^*}{||\partial g(x^{(k)})||_2^2}$, where g^* is the (unknown) optimal value for the problem (34):

$$t_k = \frac{g(x^{(k)}) - g^*}{||\partial g(x^{(k)})||_2^2} \approx \frac{g(x^{(k)}) - \min_{i \in \{1,2,...,k\}} g(x^{(i)}) + \frac{1}{k}}{||\partial g(x^{(k)})||_2^2}, \tag{41}$$

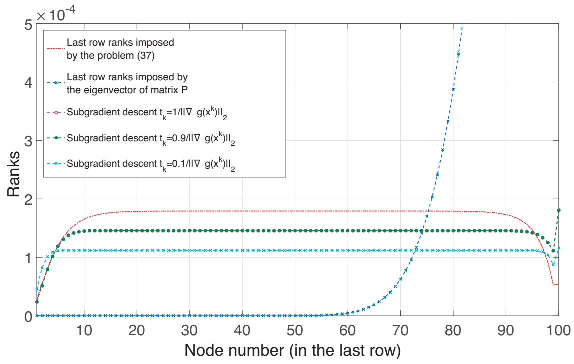
$$t_k = \frac{g(x^{(k)}) - g^*}{||\partial g(x^{(k)})||_2^2} \approx \frac{g(x^{(k)}) - \min_{i \in \{1,2,...,k\}} g(x^{(i)}) + \frac{1}{\sqrt{k}}}{||\partial g(x^{(k)})||_2^2}. \tag{42}$$

Notice, that one can use $t_k = \frac{1}{k||\partial g(x^{(k)})||_2^2}$ and $t_k = \frac{1}{\sqrt{k}||\partial g(x^{(k)})||_2^2}$ instead of corresponding step sizes (41) and (42) in case one applies the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$. This follows from the fact, that $g(x^{(k)}) = g^{best} = \min_{i\in\{1,2,\dots,k\}} g(x^{(i)})$ if the stopping rule is implemented. Otherwise, one needs to use step sizes (41) and (42) directly.

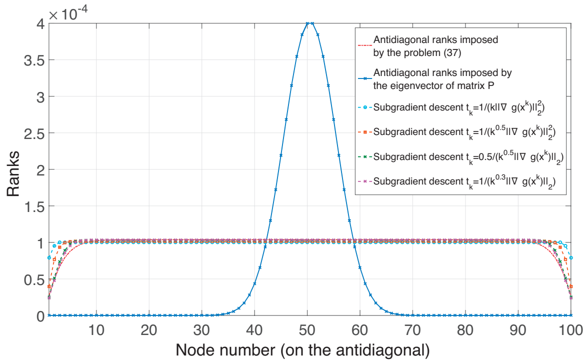


(a) Constant step length. (b) Polyak's step size.

Fig. 9: **Model 2**: Optimal solution of the problem (37) obtained by the algorithm (38) with $\varepsilon = 1$ and the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$.



(a) Last row elements.



(b) Antidiagonal elements.

Fig. 10: **Model 2**: Estimated ranks v.s the dominant eigenvector.

Similarly to the Figure 7, in Figures 9 and 10 we use $\varepsilon = 1$ as the parameter of the optimization problem (37). For larger values of ε, the robust eigenvector does not distinguish ranks of high-importance pages according to conditions (30), (31) and (32) and the Theorem 2.
Finally, we compare results obtained by the use of the stopping rule $g(x^{(k+1)}) > g(x^{(k)})$ and the results implied by the standard for subgradient algorithms choice of the function value: $g^{best} = \min_{i \in \{1,2,\dots,k\}} g(x^{(i)})$. From Figures 11 and 12, one can see that there is no sufficient gain in estimation accuracy in case one keeps track of the best function value till the iteration k: the function value decreases monotonically till the iteration number $k \approx 100$ (see Figures 11 and 12), after which the step size is being reduced but no additional accuracy is being achieved.

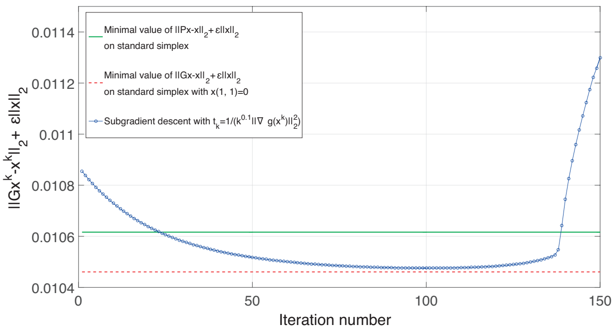


Fig. 11: **Model 1**: Optimal solution of the problem (37) obtained by the algorithm (38) without stopping rule.

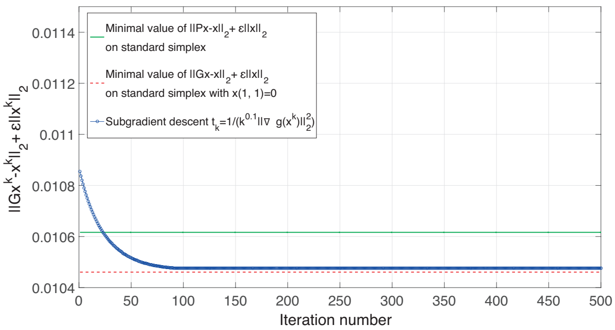


Fig. 12: **Model 1**: Optimal solution of the problem (35) obtained by the algorithm (38) with the standard for subgradient algorithms choice of the function value g^{best}.

In the next section, we discuss the Google matrix G and corresponding power methods designed for high-dimensional cases.

6 Meaning of α in the Google matrix $G = \alpha P + (1-\alpha)S$

Let L be the number of users currently sitting on the web-site j. Suppose, there are r links from this site. The users can follow the links independently from each other. Suppose also, that there are two additional possibilities for the users: (i) they can leave the site j and go to some random web-site known to the search engine and (ii) they can leave the Internet completely (this possibility includes leaving to some site which is not yet ranked by the search engine, i.e. to a very recently appeared web-site). Notice, that one can consider possibilities (i) and (ii) as additional links from the site.
Let us denote by E_i the event when a user sitting on the web-site j chooses a link i, $\forall i = 1,...,r+2$, where "links" $r+1$ and $r+2$ correspond to the observed possibilities (i) and (ii). If the random variable L_i indicates the number of times link number i is observed over L trials, the vector $(L_1,...,L_{r+2})$ follows a multinomial distribution with parameters L and p, where $p = (p_1,...,p_{r+2})$ is the vector of probabilities corresponding to events E_i, $\forall i = 1,...,r+2$ with $\sum_{i=1}^{r+2} p_i = 1$. In our setting, the users make their decisions independently of each other and links are supposed to be available at all times. Hence, we can now state the probability mass function $\mathbb{P}(L_1 = l_1,...,L_{r+2} = l_{r+2})$ that the decision E_i occurs exactly l_i times $\forall i = 1,...,r+2$ with $\sum_{i=1}^{r+2} l_i = L$:

$$\mathbb{P}(L_1 = l_1,...,L_{r+2} = l_{r+2}) = \frac{L!}{l_1!l_2!...l_{r+2}!} {p_1}^{l_1} {p_2}^{l_2} ... (p_{r+2})^{l_{r+2}}. \quad (43)$$

For the fixed observation $(l_1,...,l_{r+2})$, the maximum likelihood estimator of the probability, that a random user chooses the link i is equal to $\hat{p}_i = \frac{l_i}{L}$ (Figure 13).

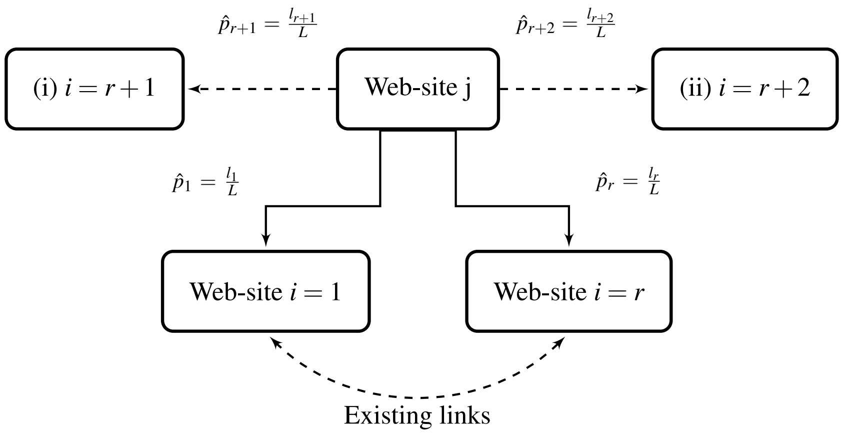


Fig. 13: Links outgoing from the web-site j.

By the Law of Large Numbers, $\forall i = 1,...,r+2$ and $\varepsilon > 0$, $\mathbb{P}\big(|\hat{p}_i - p_i| > \varepsilon\big) \to 0$, as

$L \to \infty$. Therefore, the distribution of the fraction $\hat{p}_i = \frac{l_i}{L}$ is increasingly concentrated near the expected value of the fraction, which we denote by $p_i = \frac{\bar{l}_i}{L}$ with a small abuse of notations. This means, that the probability $\mathbb{P}(L_1 = \bar{l}_1, ..., L_{r+2} = \bar{l}_{r+2})$ converges to 1 as $L \to \infty$.
Asymptotically, we can use the Stirling's approximation $\bar{l}_i! \approx \sqrt{2\pi \bar{l}_i}\ (\bar{k}_i)^{\bar{l}_i} e^{-\bar{l}_i}$ in order to receive the statement

$$\bar{l}_1 \bar{l}_2 ... \bar{l}_{r+2} \approx \frac{(L!)^2 e^{2L}}{(2\pi)^{r+2} L^{2L}},$$

where we set $\mathbb{P}(L_1 = \bar{l}_1, ..., L_{r+2} = \bar{l}_{r+2}) \approx 1$ for large enough L.
This especially holds true for high-ranked web-sites with high probabilities to enter. These web-sites are the core of all search engine results and are, therefore, of primal interest to us.
Further, it is not difficult to obtain the following system of equations:

$$\begin{cases} \prod_{i=1}^{r+2} p_i \approx \frac{1}{(2\pi L)^{r+1}} \\ \sum_{i=1}^{r+2} p_i = 1, \end{cases}$$

which can be written as

$$\begin{cases} p_{r+1} p_{r+2} \approx \frac{1}{(2\pi L)^{r+1} \prod_{i=1}^{r} p_i} \\ p_{r+1} + p_{r+2} = 1 - \sum_{i=1}^{r} p_i. \end{cases} \tag{44}$$

Notice, that the probability $\sum_{i=1}^{r} p_i$ describes the average ratio of users who follow the existing links on a web-site j, i.e. links i, $\forall i = 1, ..., r$. This probability, in general, can be estimated for any web-site by setting a counter of users who follow the links. Let us suppose for a moment, that this probability is known for the site j and let us denote it by $\bar{p}$, i.e. $\bar{p} = \sum_{i=1}^{r} p_i$. Further, let us notice that $\prod_{i=1}^{r} p_i > 0$ and let us denote by $\bar{q}$ the function $\frac{1}{(2\pi L)^{r+1} \prod_{i=1}^{r} p_i}$, which depends on the number of users on the web-site and on the product of probabilities $\prod_{i=1}^{r} p_i$.
We can now solve the system (44) with respect to p_{r+1} and p_{r+2} in order to estimate the probabilities of options (i) and (ii):

$$\begin{cases} p_{r+1} \approx 0.5(1-\bar{p}) \left(1 \pm \sqrt{1 - \frac{4\bar{q}}{(1-\bar{p})^2}}\right) \\ p_{r+2} = 1 - \bar{p} - p_{r+1}, \end{cases} \tag{45}$$

where $\bar{p} = \sum_{i=1}^{r} p_i$ and $\bar{q} = \frac{1}{(2\pi L)^{r+1} \prod_{i=1}^{r} p_i}$.
Importantly, one can always set the number of users L on the web-site to be so high, that $1 - \frac{4\bar{q}}{(1-\bar{p})^2} > 0$ is satisfied.
Further, we assume that the probability p_{r+2} to leave the Internet or to go to some recently appeared web-site to be small enough. This leads to the probability choice $p_{r+1} = 0.5(1-\bar{p}) \left(1 + \sqrt{1 - \frac{4\bar{q}}{(1-\bar{p})^2}}\right)$. Otherwise, the assumption would be opposite: high probability to leave the Internet and small probability to choose randomly among

ranked web-sites.
Using Taylor approximation in the system (45) w.r.t. $\bar{q} \sim \frac{1}{L^{r+1}}$, we claim

$$\begin{cases} p_{r+1} \approx (1-\bar{p})\left(1-\frac{\bar{q}}{(1-\bar{p})^2}\right) \\ p_{r+2} \approx (1-\bar{p})\frac{\bar{q}}{(1-\bar{p})^2}. \end{cases}$$

Further, if we suppose that
Assumption 1: the probability p_{r+1} is equally distributed through all N web-sites in the network;
Assumption 2: probabilities p_i are equal $\forall i = 1,...,r$;
Assumption 3: the term $\frac{\bar{q}}{(1-\bar{p})^2}$ is small enough to be neglected,
we can write the following transiting matrix $\bar{P}$ for current N pages:

$$\bar{P} \approx \bar{p}P + (1-\bar{p})S, \tag{46}$$

where S is the doubly stochastic matrix with entries $\frac{1}{N}$ and P is the initial transition matrix with r non-zero entries $\frac{1}{r}$ in the column j, $\forall j = 1,...,N$ (notice, that number r of ougoing links is web-site dependent).
Therefore, the matrix (46) coincides with the Google matrix $G = \alpha P + (1-\alpha)S$, where $\alpha = \bar{p}$ is the probability to follow links provided on a web-site. Google uses $\alpha = 0.85$ for the computations. It means, that 15% of users do not follow the links provided on web-sites.
Differently, one could change the Assumption 1 and claim the following:
Assumption 1 (new): the probability p_{r+1} is equally distributed among all web-sites except those which are provided as links on the current web-site j.
In this case, one would use another stochastic matrix $\tilde{S}$ with zero entry (i, j), if there is a direct link from j to i, $\forall i, j = 1,...,N$. The transition matrix $\tilde{P}$ would, therefore, become

$$\tilde{P} \approx \bar{p}P + (1-\bar{p})\tilde{S}, \tag{47}$$

where $\tilde{S}$ is a stochastic matrix with the entry $\tilde{S}_{ij} = \frac{1}{N-r}$ if there is no direct link from j to i, $\forall i, j = 1,...,N$.
Importantly, transition matrices $\bar{P}$ and $\tilde{P}$ have only positive entries. Therefore, their eigenvectors, corresponding to the eigenvalue $\lambda = 1$ are unique and, moreover, power methods $x^{(k+1)} = \bar{P}x^{(k)}$ and $x^{(k+1)} = \tilde{P}x^{(k)}$ converge correspondingly to their dominant eigenvectors according to the Perron-Frobenius theorem for positive matrices for all starting points $x^{(0)}$ satisfying simplex constraints.

High-dimensional algorithms for robust PageRank: Numerically, it is convenient to use the power method $x^{(k+1)} = \bar{P}x^{(k)}$ used by Google, as it can be written as

$$x^{(k+1)} = \bar{p}Px^{(k)} + \frac{1-\bar{p}}{N}\mathbb{1}_N, \tag{48}$$

where $\mathbb{1}_N$ is the vector of all-ones of the length N arising from $\frac{1}{N}\mathbb{1}_N = Sx^{(k)}$.
However, under the Assumption 1 (new), the following power method is to be used:

$$x^{(k+1)} = \bar{p}Px^{(k)} + (1-\bar{p})\tilde{S}x^{(k)}, \tag{49}$$

where $\tilde{S}$ is the column-stochastic matrix with (i, j) being zero if there is a direct link from the site j to the site i.
Power iterations (48) and (49) both work in case of huge dimensionality and converge linearly with the rate $\frac{\lambda_1}{\lambda_2} = \frac{1}{\lambda_2}$, where λ_2 is the second eigenvalue of the corresponding transition matrix (either $\bar{P}$ or $\tilde{P}$). They converge to PageRank estimates, which are biased with respect to the dominant eigenvector of the matrix P.
We compare dominant eigenvectors of matrices $\bar{P}$ and $\tilde{P}$ with the dominant eigenvector of the matrix P for Model 2 described in the Section 5.1. The number of nodes in the network is $N = n^2 = 40.000$ (Figures 14, 15, 16, 17).

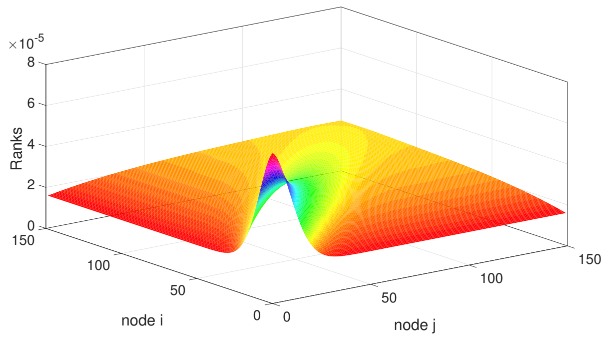


Fig. 14: **Model 2**: Part of $x_{i,j}$ elements computed via algorithms (48) and (49).

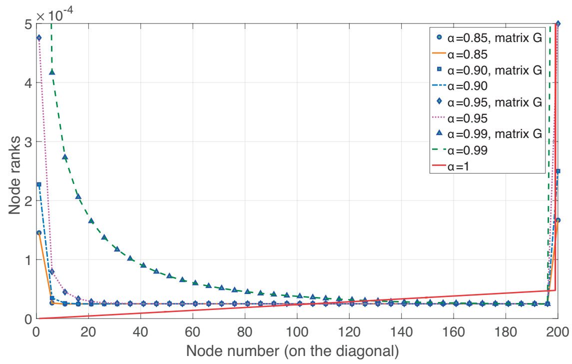


Fig. 15: **Model 2**: Diagonal elements computed via algorithms (48) and (49).

In general, the dominant eigenvector of the matrix $\bar{P}$ may differ from the dominant eigenvector of the matrix $\tilde{P}$. Both of these vectors can approximate PageRank, which is robust in terms of perturbations in links. However, as the probability p_{r+2} to leave the Internet or to go to a newly appeared web-site has been neglected (as it is proportional to $\frac{1}{L^{r+1}}$), dominant eigenvectors of $\bar{P}$ and $\tilde{P}$ cannot be considered as approximations for the PageRank robust to the long-term perturbations in the number of nodes. In order to take these perturbations into account, one would need to account for the probability p_{r+2} avoiding neglect of terms in the system (45).

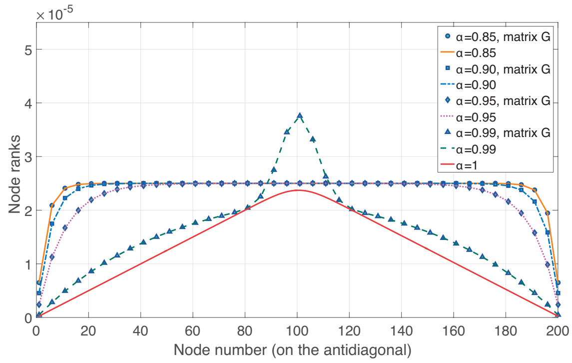


Fig. 16: **Model 2**: Antidiagonal elements computed via algorithms (48) and (49).

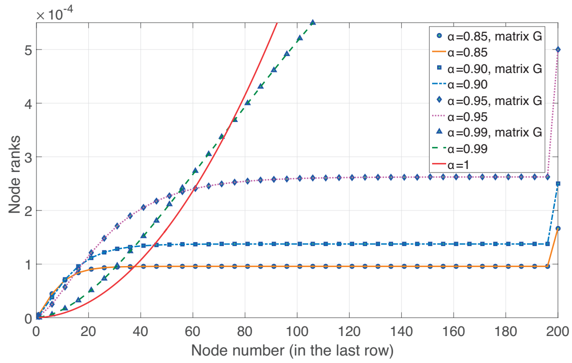


Fig. 17: **Model 2**: Last row elements computed via algorithms (48) and (49).

In the Figures 14, 15, 16, 17 one can see, that the difference between principal eigenvectors of matrices $\bar{P}$ and $\tilde{P}$ is negligible for Model 2. Algorithm (48) is, however, more efficient numerically.

In order to recover the dominant eigenvector of the unperturbed transition matrix P in huge dimensional cases, one could implement fast iterative algorithms of the type $x^{(k+1)} = \bar{P}_k x^{(k)}$ with transition matrix $\bar{P}_k$ adapted iteratively so, that it converges to the matrix P. Such algorithms are discussed in details in the works of A. Juditsky and B. Polyak [13], B. Polyak and A. Timonina [25], A. Timonina [28]. Furthermore, these iterative regularization schemes remind methods for solving variational inequalities discussed in the work of A. Bakushinskij and B. Polyak [1].

7 Direction for future research

The PageRank problem is one of the most challenging problems in information technologies and numerical analysis due to its huge dimension and wide range of possible applications.
First realizable application of the PageRank problem lies in the field of scientometrics, i.e. the study for analyzing science and innovation by measuring impact of articles, journals and institutes via their scientific citations. The main difference between the PageRank formulation (1) and the formulation suitable for scientometrics is incorporated in transition probabilities P_{ij}: journals may refer to each other multiple times via one publication, while in the Internet multiple references from the web-page i to the web-page j are considered as a single reference.
Furthermore, there is a potential application of the research to the area of finance: the study on robust PageRank can be extended to the robust ranking measurement technique of systemic risk in the financial sector. To realize this application, one could consider the financial system as a complex network of financial institutions, where financial dependencies represent existing links between these institutions, and one would define the systemic risk as the probability of default of a large portion of financial institutions in the network [26]. The robust approach is especially useful for the systemic risk application, as the dependence structure of financial institutions varies very fast, being subject to changes in loans, book and market values of accredited firms, etc.
In our future research, we plan to study statistical methods to estimate the probability for a random user in the Internet to leave to some web-page which is not yet ranked by the search engine. In this article, this probability is considered to be small enough (e.g. the probability p_{r+2} in the system (45)). However, explicit incorporation of this probability in the analysis would lead to a better estimate for the robust PageRank.
Further, we plan to focus on different types of structured perturbations [13] and randomized techniques for the computation of the robust stationary distribution in high-dimensional cases [18,23,24]. We would also like to test the proposed approach on a large-scale real-life data available for the World Wide Web [4].

8 Acknowledgements

The author would like to express her vast gratitude to Prof. Dr. Boris T. Polyak for inspirational and enlightening discussions on Robust PageRank, as well as for continuous and irreplaceable support in the development of the article. Simultaneously, the author would like to acknowledge Prof. Dr. Daniel Kuhn for beneficial assessment during the evolution of the article.

References

1. Bakushinskij, A.B., Polyak, B.T. *On the Solution of Variational Inequalities.* Sov. Math. Doklady, Volume 14, pp. 1705–1710 (1974).
2. Ben–Tal, A., Nemirovski, A. *Robust convex optimization.* Mathematics of Operations Research, Volume 23(4), pp. 769–805 (1998).
3. Boyd, S., Vandenberghe, L. *Convex Optimization.* Cambridge University Press, Cambridge (2004).
4. Borodin, A., Roberts, G.O., Rosenthal, J.S., Tsaparas, P. *Finding Authorities and Hubs from Link Structures on the World Wide Web.* 10th International World Wide Web Conference (2000).
5. Brin, S., Page, L. *The Anatomy of a Large-Scale Hypertextual Web Search Engine.* Computer Networks and ISDN Systems, Volume 30(1-7), pp. 107–117 (1998).
6. Bryan, K., Leise, T. *The $25,000,000,000 Eigenvector: The Linear Algebra behind Google.* SIAM Review, Volume 48(3), pp. 569–581 (2006).
7. El Ghaoui, L., Lebret, H. *Robust Solutions to Least-Squares Problems with Uncertain Data.* SIAM Journal on Matrix Analysis and Applications, Volume 18(4), pp. 1035–1064 (1997).
8. Franceschet, M. *PageRank: Standing on the Shoulders of Giants.* Communications of the ACM, Volume 54(6), pp. 92–101 (2011).
9. Haveliwala, T.H., Kamvar, S.D. *The Second Eigenvalue of the Google Matrix.* Technical Report, Computer Science Department, Stanford University (2003).
10. Horn, R.A., Johnson, C.R. *Matrix Analysis.* 575pp., Cambridge University Press, Cambridge (1990).
11. Ishii, H., Tempo, R. *Distributed Randomized Algorithms for the PageRank Computation.* IEEE Transactions on Automatic Control, Volume 55(9), pp. 1987–2002 (2010).
12. Ishii, H., Basar, T. and Tempo, R. *Randomized Algorithms for Synthesis of Switching Rules for Multimodal Systems.* IEEE Transactions on Automatic Control, Volume 50(6), pp. 754-767 (2005).
13. Juditsky, A., Polyak, B. *Robust Eigenvector of a Stochastic Matrix with Application to PageRank.* 51st IEEE Conference on Decision and Control, Maui, Hawaii, USA (2012).
14. Lan, G., Nemirovskij, A.S., Shapiro, A. *Validation Analysis of Mirror Descent Stochastic Approximation Method.* Mathematical Programming, Volume 134(2), pp. 425–458 (2012).
15. Langville, A.N., Meyer, C.D. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press (2006).
16. Langville, A.N., Meyer, C.D. *Deeper Inside PageRank.* Internet Mathematics, Volume 1(3), pp. 335–380 (2004).
17. Lei, J. *Distributed Randomized PageRank Algorithm Based on Stochastic Approximation.* IEEE Transactions on Automatic Control, Volume 60(6), pp. 1641–1646 (2014).
18. Nazin, A.V., Polyak, B.T. *Adaptive Randomized Algorithm for Finding Eigenvector of Stochastic Matrix with Application to PageRank.* Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai (2009).
19. Nemirovskij, A.S., Yudin, D.B. *Problem Complexity and Method Efficiency in Optimization.* John Wiley, New York (1983).
20. Nesterov, Y. *Subgradient Methods for Huge-Scale Optimization Problems.* Mathematical Programming, Volume 146(1-2), pp. 275–297 (2014).
21. Page, L., Brin, S., Motwani, R. and Winograd, T. *The Pagerank Citation Ranking: Bringing Order to the Web.* Technical Report, Stanford Digital Library Technologies Project, USA (1998).
22. Polyak, B.T. *Introduction to Optimization.* Optimization Software, 464 pp. (1987).
23. Polyak, B.T. *Random Algorithms for Solving Convex Inequalities.* Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (eds. Butnariu, D., Censor, Y. and Reich, S.), Elsevier, pp. 409–422 (2001).
24. Polyak, B.T., Tempo, R. *Probabilistic Robust Design with Linear Quadratic Regulators.* Systems and Control Letters, Volume 43(5), pp. 343–353 (2001).
25. Polyak, B.T., Timonina, A.V. *PageRank: New Regularizations and Simulation Models.* 18th IFAC World Congress, Volume 18(1), pp. 11202–11207 (2011).
26. Sadoghi, A. Measuring Systemic Risk: Robust Ranking Techniques Approach. 7th Financial Risks International Forum (2014).
27. Tibshirani, R. *Regression Shrinkage and Selection via the Lasso.* Journal of the Royal Statistical Society, Series B (Methodological), Volume 58(1), pp. 267-288 (1996).
28. Timonina, A.V. *The Rank-Model and Its Investigations.* (in Russian) Stochastic Optimization in Informatics, Volume 5, (O.N. Granichin ed., ISSN 1992–2922), St.-Petersburg University, pp. 139–156 (2009).

9 Appendix

Lemma 2 *Conic optimization problem*

$$f(x) = \max_{\substack{z\in\mathbb{R}^N\\ ||z||_1\leq\varepsilon\\ |z_j|\leq\varepsilon_j}} z^T x \tag{50}$$

is equivalent to the following minimization problem:

$$f(x) = \min_{\lambda+\mu=x} \left\{ \varepsilon\|\lambda\|_\infty + \sum_{j=1}^{N} \varepsilon_j |\mu_j| \right\}. \tag{51}$$

Proof Let us dualize the optimization problem (50). First of all, notice that

$$\max_{\substack{z\in\mathbb{R}^N\\ ||z||_1\leq\varepsilon\\ |z_j|\leq\varepsilon_j}} z^T x \iff \max_{\substack{z\in\mathbb{R}^N,\, t\in\mathbb{R}^N_{\{0,+\}}\\ \sum_{i=1}^N t_i\leq\varepsilon\\ z_j\leq t_j\\ -z_j\leq t_j\\ z_j\leq\varepsilon_j\\ -z_j\leq\varepsilon_j}} z^T x.$$

Therefore, the Lagrangian $\mathscr{L}$ can be written in the following form for dual variables $\alpha,\beta,\gamma,\eta,\nu$, where $\alpha\in\mathbb{R}_{\{0,+\}}$ and $\beta,\gamma,\eta,\nu\in\mathbb{R}^N_{\{0,+\}}$:

$$\begin{aligned}\mathscr{L} = {} & z^T x - \alpha\left(\sum_{i=1}^N t_i - \varepsilon\right) - \sum_{j=1}^N \Big(\beta_j(z_j - t_j) + \gamma_j(-z_j - t_j)\Big) - \\ & - \sum_{j=1}^N \Big(\eta_j(z_j-\varepsilon_j) + \nu_j(-z_j-\varepsilon_j)\Big) = \alpha\varepsilon + \sum_{j=1}^N (\eta_j+\nu_j)\varepsilon_j + \\ & + \sum_{j=1}^N z_j\big(x_j - \beta_j + \gamma_j - \eta_j + \nu_j\big) + \sum_{j=1}^N t_j\big(\beta_j+\gamma_j-\alpha\big).\end{aligned}$$

By strong duality, the following holds

$$\max_{\substack{z\in\mathbb{R}^N\\ ||z||_1\leq\varepsilon\\ |z_j|\leq\varepsilon_j}} z^T x = \max_{\substack{z\in\mathbb{R}^N\\ t\in\mathbb{R}^N_{\{0,+\}}}} \min_{\substack{\beta,\gamma,\eta,\nu\in\mathbb{R}^N_{\{0,+\}}\\ \alpha\in\mathbb{R}_{\{0,+\}}}} \mathscr{L} = \min_{\substack{\beta,\gamma,\eta,\nu\in\mathbb{R}^N_{\{0,+\}}\\ \alpha\in\mathbb{R}_{\{0,+\}}}} \max_{\substack{z\in\mathbb{R}^N\\ t\in\mathbb{R}^N_{\{0,+\}}}} \mathscr{L},$$

where the following is true at the point of maximum over z,t:

$$\begin{aligned}& x_j - \beta_j + \gamma_j - \eta_j + \nu_j = 0,\ \forall j = 1,...,N.\\ & \beta_j + \gamma_j - \alpha \leq 0,\ \forall j = 1,...,N.\end{aligned}$$

Substituting these equations into the Lagrangian and maximizing over z and t, we get

$$\mathscr{L} = \alpha\varepsilon + \sum_{j=1}^N (\eta_j+\nu_j)\varepsilon_j. \tag{52}$$

Now, let us make the following change of variables:

$$\lambda_j = \beta_j - \gamma_j, \ \forall j = 1, ..., N,$$
$$\mu_j = \eta_j - \nu_j \ \forall j = 1, ..., N.$$

Notice, that $x_j = \lambda_j + \mu_j, \ \forall j = 1, ..., N$. At the point of minimum over $\alpha, \ \eta, \ \nu$ the term $\eta_j + \nu_j$ behaves as $|\mu_j|$. This happens, because at optimality $\mu_j = \eta_j, \ \nu_j = 0$ if $\mu_j \geq 0$ and $\mu_j = -\nu_j, \ \eta_j = 0$ if $\mu_j \leq 0$. Similarly, $\beta_j + \gamma_j, \ \forall j = 1, ..., N$ behaves as $|\lambda_j|, \ \forall j = 1, ..., N$ at optimality, which leads to $\alpha = \|\lambda\|_\infty$.
Hence, equation (52) under the proposed change of variables applies the statement of the Lemma 2.

Lemma 3 *Conic optimization problem*

$$f(x) = \max_{\substack{z \in \mathbb{R}^N \\ ||z||_2 \leq \varepsilon \\ |z_j| \leq \varepsilon_j}} z^T x \tag{53}$$

is equivalent to the following minimization problem:

$$f(x) = \min_{\lambda + \mu = x} \left\{ \varepsilon \|\lambda\|_2 + \sum_{j=1}^{N} \varepsilon_j |\mu_j| \right\}. \tag{54}$$

Proof Let us dualize the optimization problem (53). First of all, notice that

$$\max_{\substack{z \in \mathbb{R}^N \\ ||z||_2 \leq \varepsilon \\ |z_j| \leq \varepsilon_j}} z^T x \iff \max_{\substack{z \in \mathbb{R}^N \\ \sqrt{\sum_{i=1}^N z_i^2} \leq \varepsilon \\ z_j \leq \varepsilon_j \\ -z_j \leq \varepsilon_j}} z^T x.$$

Therefore, the Lagrangian $\mathscr{L}$ can be written in the following form for dual variables α, β, γ, where $\alpha \in \mathbb{R}_{\{0,+\}}$ and $\beta, \gamma \in \mathbb{R}^N_{\{0,+\}}$:

$$\mathscr{L} = z^T x - \alpha \left(\sqrt{\sum_{i=1}^{N} z_i^2} - \varepsilon \right) - \sum_{j=1}^{N} \beta_j (z_j - \varepsilon_j) - \sum_{j=1}^{N} \gamma_j (-z_j - \varepsilon_j).$$

By strong duality, the following holds

$$\max_{\substack{z \in \mathbb{R}^N \\ ||z||_2 \leq \varepsilon \\ |z_j| \leq \varepsilon_j}} z^T x = \max_{z \in \mathbb{R}^N} \min_{\substack{\alpha \in \mathbb{R}_{\{0,+\}} \\ \beta, \gamma \in \mathbb{R}^N_{\{0,+\}}}} \mathscr{L} = \min_{\substack{\alpha \in \mathbb{R}_{\{0,+\}} \\ \beta, \gamma \in \mathbb{R}^N_{\{0,+\}}}} \max_{z \in \mathbb{R}^N} \mathscr{L},$$

where the following must be true at the point of maximum over z:

$$\frac{\partial L(z, \alpha, \beta, \gamma)}{\partial z_j} = x_j - \beta_j + \gamma_j - \alpha \frac{z_j}{\|z\|_2} = 0, \ \forall j = 1, ..., N.$$

Substituting this equation into the Lagrangian, we get

$$\mathscr{L}(z, \alpha, \beta, \gamma) = \alpha \varepsilon + \sum_{j=1}^{N} (\beta_j + \gamma_j) \varepsilon_j. \tag{55}$$

Now, let us make the following change of variables:

$$\lambda_j = \alpha \frac{z_j}{\|z\|_2},\ \forall j = 1,...,N,$$
$$\mu_j = \beta_j - \gamma_j,\ \forall j = 1,...,N.$$

Notice, that $\alpha = \|\lambda\|_2$ and that at the point of minimum over $\alpha,\ \beta,\ \gamma$ the term $\beta_j + \gamma_j$ behaves as $|\mu_j|$. This happens, because at optimality $\beta_j = \mu_j,\ \gamma_j = 0$ if $\mu_j \geq 0$ and $\beta_j = 0,\ \gamma_j = -\mu_j$ if $\mu_j \leq 0$.
Hence, equation (55) applies the statement of the Lemma 3 under the proposed change of variables.

Lemma 4 *(see Theorem 3.1 in [7]) For $a_i \in \mathbb{R}^{n_i},\ \forall i = 0,...,N,\ \xi_j \in \mathbb{R}^{n_0 \times n_j},\ j = 1,...,N$ the following holds:*

$$\max_{\substack{\|\xi_1\|_F \leq \varepsilon^{(\xi_1)} \\ \|\xi_2\|_F \leq \varepsilon^{(\xi_2)} \\ \cdots \\ \|\xi_N\|_F \leq \varepsilon^{(\xi_N)}}} \left\| a_0 + \sum_{i=1}^{N} \xi_i a_i \right\|_2 = \|a_0\|_2 + \sum_{i=1}^{N} \varepsilon^{(\xi_i)} \|a_i\|_2.$$

Proof

$$\begin{aligned}
&\|a_0 + \sum_{i=1}^{N} \xi_i a_i\|_2^2 = \left(a_0 + \sum_{i=1}^{N} \xi_i a_i\right)^T \left(a_0 + \sum_{i=1}^{N} \xi_i a_i\right) = \\
&= \|a_0\|_2^2 + \sum_{i=1}^{N} \|\xi_i a_i\|_2^2 + 2\sum_{j=1}^{N} \|a_0^T \xi_j a_j\|_2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \|a_i^T \xi_i^T \xi_j a_j\|_2 \leq \\
&\leq \|a_0\|_2^2 + \sum_{i=1}^{N} \left(\varepsilon^{(\xi_i)}\right)^2 \|a_i\|_2^2 + 2\|a_0\|_2 \sum_{j=1}^{N} \varepsilon^{(\xi_j)} \|a_j\|_2 + \\
&+2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \varepsilon^{(\xi_i)} \varepsilon^{(\xi_j)} \|a_i\|_2 \|a_j\|_2 = \left(\|a_0\|_2 + \sum_{i=1}^{N} \varepsilon^{(\xi_i)} \|a_i\|_2\right)^2.
\end{aligned}$$

Hence,

$$\|a_0 + \sum_{i=1}^{N} \xi_i a_i\|_2 \leq \|a_0\|_2 + \sum_{i=1}^{N} \varepsilon^{(\xi_i)} \|a_i\|_2.$$

Equality holds if $\xi_i = \xi_i^* = \frac{\varepsilon^{(\xi_i)} a_0 a_i^T}{\|a_0\|_2 \|a_i\|_2}$ for $a_0 \neq 0$ (for $a_0 = 0$ one can take arbitrary $a_0 :\ \|a_0\|_2 = 1$).

9.1 Proof of the Proposition 1

Proof Let $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, where u_1 is a vector of the length N and u_2 is a vector of the length M. Notice, that the following equality holds due to the duality of the l_1-norm:

$$
\begin{aligned}
||Qx-x||_1 &= \left\| \begin{pmatrix} P+\xi & \zeta \\ \psi & \chi \end{pmatrix} \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} - \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \right\|_1 = \left\| \begin{pmatrix} (P+\xi-I_N)x^{(1)}+\zeta x^{(2)} \\ \psi x^{(1)}+(\chi-I_M)x^{(2)} \end{pmatrix} \right\|_1 = \\
&= \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^T \begin{pmatrix} (P+\xi-I_N)x^{(1)}+\zeta x^{(2)} \\ \psi x^{(1)}+(\chi-I_M)x^{(2)} \end{pmatrix} = \\
&= \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T(P-I_N)x^{(1)} + \left(u_1^T\xi + u_2^T\psi\right)x^{(1)} + \left(u_1^T\zeta + u_2^T(\chi-I_M)\right)x^{(2)} \right),
\end{aligned}
$$

where I_N and I_M are identity matrices of sizes $N\times N$ and $M\times M$ correspondingly. Further, the function $||Qx-x||_1$ can be bounded from above in line with the triangle inequality and norm duality:

$$
\begin{aligned}
||Qx-x||_1 \le\ & \|Px^{(1)}-x^{(1)}\|_1 + \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T\xi + u_2^T\psi\right)x^{(1)} + \\
& + \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T\zeta + u_2^T(\chi-I_M)\right)x^{(2)}.
\end{aligned}
$$

This leads to the statement:

$$
\begin{aligned}
\max_{Q\in\Xi^{(l_1)}} ||Qx-x||_1 \le\ & \|Px^{(1)}-x^{(1)}\|_1 + \max_{\substack{\|[\xi]_j\|_1\le\varepsilon_j^{(\xi)} \\ \sum_{i,j}|\xi_{ij}|\le\varepsilon^{(\xi)} \\ \|[\psi]_j\|_1\le\varepsilon_j^{(\psi)} \\ \sum_{i,j}|\psi_{ij}|\le\varepsilon^{(\psi)}}} \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T\xi + u_2^T\psi\right)x^{(1)} + \\
& + \max_{\substack{\|[\zeta]_j\|_1\le\varepsilon_j^{(\zeta)} \\ \sum_{i,j}|\zeta_{ij}|\le\varepsilon^{(\zeta)} \\ \|[\chi]_j\|_1\le\varepsilon_j^{(\chi)} \\ \sum_{i,j}|\chi_{ij}|\le\varepsilon^{(\chi)}}} \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T\zeta + u_2^T(\chi-I_M)\right)x^{(2)}. \qquad (56)
\end{aligned}
$$

Now, consider the following subproblem:

$$
g_1(x) = \max_{\substack{\|[\xi]_j\|_1\le\varepsilon_j^{(\xi)} \\ \sum_{i,j}|\xi_{ij}|\le\varepsilon^{(\xi)} \\ \|[\psi]_j\|_1\le\varepsilon_j^{(\psi)} \\ \sum_{i,j}|\psi_{ij}|\le\varepsilon^{(\psi)}}} \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty\le 1}} \left(u_1^T\xi + u_2^T\psi\right)x^{(1)}.
$$

Let $z = \xi^T u_1 + \psi^T u_2$. Based on the conditions of the problem (12), we can write

$$|z_j| = \left|\sum_{i=1}^{N} \xi_{ij} u_{1i} + \sum_{i=1}^{M} \psi_{ij} u_{2i}\right| \le \sum_{i=1}^{N} |\xi_{ij}| + \sum_{i=1}^{M} |\psi_{ij}| \le \varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)},$$
$$\|z\|_1 = \sum_{j=1}^{N} \left|\sum_{i=1}^{N} \xi_{ij} u_{1i} + \sum_{i=1}^{M} \psi_{ij} u_{2i}\right| \le \sum_{j=1}^{N}\sum_{i=1}^{N} |\xi_{ij}| + \sum_{j=1}^{N}\sum_{i=1}^{M} |\psi_{ij}| \le \varepsilon^{(\xi)} + \varepsilon^{(\psi)},$$

where u_{1i}, $\forall i = 1,...,N$ and u_{2i}, $\forall i = 1,...,M$ are i-th coordinates of vectors u_1 and u_2 correspondingly.
Hence,

$$g_1(x) \le \max_{\substack{z\in\mathbb{R}^N \\ ||z||_1 \le \varepsilon^{(\xi)}+\varepsilon^{(\psi)} \\ |z_j| \le \varepsilon_j^{(\xi)}+\varepsilon_j^{(\psi)}}} z^T x^{(1)} = (\varepsilon^{(\xi)} + \varepsilon^{(\psi)}) \max_{\substack{z\in\mathbb{R}^N \\ ||z||_1 \le 1 \\ |z_j| \le \frac{\varepsilon_j^{(\xi)}+\varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)}+\varepsilon^{(\psi)}}}} z^T x^{(1)},$$

which is a strictly feasible conic optimization problem. Dualizing the constraints we come to

$$g_1(x) \le \left(\varepsilon^{(\xi)} + \varepsilon^{(\psi)}\right) \min_{\lambda+\mu=x^{(1)}} \left\{ \|\lambda\|_\infty + \sum_{j=1}^{N} \frac{\varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)} + \varepsilon^{(\psi)}} |\mu_j| \right\} =$$
$$= \left(\varepsilon^{(\xi)} + \varepsilon^{(\psi)}\right) \|x^{(1)}\|_{(a)}, \tag{57}$$

where we denote the underlying norm by $\|x^{(1)}\|_{(a)}$.
Analogically, consider the other subproblem:

$$g_2(x) = \max_{\substack{\|[\zeta]_j\|_1 \le \varepsilon_j^{(\zeta)} \\ \sum_{i,j} |\zeta_{ij}| \le \varepsilon^{(\zeta)} \\ \|[\chi]_j\|_1 \le \varepsilon_j^{(\chi)} \\ \sum_{i,j} |\chi_{ij}| \le \varepsilon^{(\chi)}}} \max_{\substack{u\in\mathbb{R}^{N+M} \\ ||u||_\infty \le 1}} \left(u_1^T \zeta + u_2^T (\chi - I_M)\right) x^{(2)}.$$

Let $z = \zeta^T u_1 + (\chi^T - I_M) u_2$. Based on the conditions of the problem (12), we can write

$$|z_j| = \left|\sum_{i=1}^{N} \zeta_{ij} u_{1i} + \sum_{i=1}^{M} \chi_{ij} u_{2i} - u_{2j}\right| \le \sum_{i=1}^{N} |\zeta_{ij}| + \sum_{i=1}^{M} |\chi_{ij}| + 1 \le \varepsilon_j^{(\zeta)} + \varepsilon_j^{(\chi)} + 1,$$
$$\|z\|_1 = \sum_{j=1}^{M} \left|\sum_{i=1}^{N} \zeta_{ij} u_{1i} + \sum_{i=1}^{M} \chi_{ij} u_{2i} - u_{2j}\right| \le \sum_{j=1}^{M}\sum_{i=1}^{N} |\zeta_{ij}| + \sum_{j=1}^{M}\sum_{i=1}^{M} |\chi_{ij}| + M \le$$
$$\le \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M.$$

Hence,

$$g_2(x) \le \max_{\substack{z\in\mathbb{R}^M \\ ||z||_1 \le \varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M \\ |z_j| \le \varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}} z^T x^{(2)} = (\varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + M) \max_{\substack{z\in\mathbb{R}^M \\ ||z||_1 \le 1 \\ |z_j| \le \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M}}} z^T x^{(2)},$$

which is a strictly feasible conic optimization problem. Dualizing the constraints we come to

$$
\begin{aligned}
g_2(x) &\le \left(\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M\right)\min_{\lambda+\mu=x^{(2)}}\left\{\|\lambda\|_\infty+\sum_{j=1}^{N}\frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M}|\mu_j|\right\}\\
&= \left(\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M\right)\|x^{(2)}\|_{(b)},
\end{aligned}
\tag{58}
$$

where we denote the underlying norm by $\|x^{(2)}\|_{(b)}$.
Therefore, the statement of the Proposition 1 follows from equations (56), (57) and (58).

9.2 Proof of the Proposition 2

Proof Let $u=\begin{pmatrix}u_1\\u_2\end{pmatrix}$, where u_1 is a vector of the length N and u_2 is a vector of the length M. Notice, that the following equality holds due to the duality of the second norm:

$$
\begin{aligned}
&||Qx-x||_2=\left\|\begin{pmatrix}P+\xi & \zeta\\ \psi & \chi\end{pmatrix}\begin{pmatrix}x^{(1)}\\x^{(2)}\end{pmatrix}-\begin{pmatrix}x^{(1)}\\x^{(2)}\end{pmatrix}\right\|_2=\left\|\begin{pmatrix}(P+\xi-I_N)x^{(1)}+\zeta x^{(2)}\\ \psi x^{(1)}+(\chi-I_M)x^{(2)}\end{pmatrix}\right\|_2=\\
&=\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\begin{pmatrix}u_1\\u_2\end{pmatrix}^T\begin{pmatrix}(P+\xi-I_N)x^{(1)}+\zeta x^{(2)}\\ \psi x^{(1)}+(\chi-I_M)x^{(2)}\end{pmatrix}=\\
&=\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\left(u_1^T(P-I_N)x^{(1)}+\left(u_1^T\xi+u_2^T\psi\right)x^{(1)}+\left(u_1^T\zeta+u_2^T(\chi-I_M)\right)x^{(2)}\right),
\end{aligned}
$$

where I_N and I_M are identity matrices of sizes $N\times N$ and $M\times M$ correspondingly.
Furthermore, one can bound the norm $||Qx-x||_2$ from above based on the triangle inequality and norm duality:

$$
\begin{aligned}
&||Qx-x||_2\le\\
&\le\|Px^{(1)}-x^{(1)}\|_2+\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\left(u_1^T\xi+u_2^T\psi\right)x^{(1)}+\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\left(u_1^T\zeta+u_2^T(\chi-I_M)\right)x^{(2)},
\end{aligned}
$$

which would lead to the upper bound for the value $\max_{Q\in\Xi^{(l_2)}}\|Qx-x\|_2$ (analogous to the case of l_1-norm).

$$
\begin{aligned}
&\max_{Q\in\Xi^{(l_2)}}||Qx-x||_2\le\|Px^{(1)}-x^{(1)}\|_2+\\
&+\max_{\substack{\|[\xi]_j\|_1\le\varepsilon_j^{(\xi)}\\ \|\xi\|_F\le\varepsilon^{(\xi)}\\ \|[\psi]_j\|_1\le\varepsilon_j^{(\psi)}\\ \|\psi\|_F\le\varepsilon^{(\psi)}}}\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\left(u_1^T\xi+u_2^T\psi\right)x^{(1)}+\max_{\substack{\|[\zeta]_j\|_1\le\varepsilon_j^{(\zeta)}\\ \|\zeta\|_F\le\varepsilon^{(\zeta)}\\ \|[\chi]_j\|_1\le\varepsilon_j^{(\chi)}\\ \|\chi\|_F\le\varepsilon^{(\chi)}}}\max_{\substack{u\in\mathbb{R}^{N+M}\\||u||_2\le 1}}\left(u_1^T\zeta+u_2^T(\chi-I_M)\right)x^{(2)}.
\end{aligned}
$$

Considering the following subproblem

$$g_3(x) = \max_{\substack{\|[\xi]_j\|_1 \le \varepsilon_j^{(\xi)} \\ \|\xi\|_F \le \varepsilon^{(\xi)} \\ \|[\psi]_j\|_1 \le \varepsilon_j^{(\psi)} \\ \|\psi\|_F \le \varepsilon^{(\psi)}}} \max_{\substack{u \in \mathbb{R}^{N+M} \\ ||u||_2 \le 1}} \left(u_1^T \xi + u_2^T \psi\right) x^{(1)}$$

and denoting $z = \xi^T u_1 + \psi^T u_2$, we reformulate it based on the conditions of the problem (14):

$$|z_j| = \big|\sum_{i=1}^{N} \xi_{ij} u_{1i} + \sum_{i=1}^{M} \psi_{ij} u_{2i}\big| \le \sum_{i=1}^{N} \big|\xi_{ij}\big| + \sum_{i=1}^{M} \big|\psi_{ij}\big| \le \varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)},$$
$$\|z\|_2 \le \|\xi\|_F \|u_1\|_2 + \|\psi\|_F \|u_2\|_2 \le \varepsilon^{(\xi)} + \varepsilon^{(\psi)},$$

where u_{1i}, $\forall i = 1,...,N$ and u_{2i}, $\forall i = 1,...,M$ are i-th coordinates of vectors u_1 and u_2 correspondingly.
Therefore,

$$g_3(x) \le \max_{\substack{z \in \mathbb{R}^N \\ ||z||_2 \le \varepsilon^{(\xi)} + \varepsilon^{(\psi)} \\ |z_j| \le \varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)}}} z^T x^{(1)} = \left(\varepsilon^{(\xi)} + \varepsilon^{(\psi)}\right) \max_{\substack{z \in \mathbb{R}^N \\ ||z||_2 \le 1 \\ |z_j| \le \frac{\varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)} + \varepsilon^{(\psi)}}}} z^T x^{(1)},$$

which is a strictly feasible conic optimization problem. Dualizing the constraints (see Lemma 3 in the Appendix for details) we obtain:

$$g_3(x) \le \left(\varepsilon^{(\xi)} + \varepsilon^{(\psi)}\right) \min_{\lambda + \mu = x^{(1)}} \left\{ \|\lambda\|_2 + \sum_{j=1}^{N} \frac{\varepsilon_j^{(\xi)} + \varepsilon_j^{(\psi)}}{\varepsilon^{(\xi)} + \varepsilon^{(\psi)}} |\mu_j| \right\}. \tag{59}$$

Note, that $g_3(x)$ upper bound is a norm.
Analogically, we can bound the following subproblem by its dualization:

$$g_4(x) = \max_{\substack{\|[\zeta]_j\|_1 \le \varepsilon_j^{(\zeta)} \\ \|\zeta\|_F \le \varepsilon^{(\zeta)} \\ \|[\chi]_j\|_1 \le \varepsilon_j^{(\chi)} \\ \|\chi\|_F \le \varepsilon^{(\chi)}}} \max_{\substack{u \in \mathbb{R}^{N+M} \\ ||u||_2 \le 1}} \left(u_1^T \zeta + u_2^T (\chi - I_M)\right) x^{(2)}. \tag{60}$$

Let $z = \zeta^T u_1 + (\chi^T - I_M) u_2$. Based on the conditions of the problem (14), we can write

$$|z_j| = \left|\sum_{i=1}^{N} \zeta_{ij} u_{1i} + \sum_{i=1}^{M} \chi_{ij} u_{2i} - u_{2j}\right| \le \sum_{i=1}^{N} \big|\zeta_{ij}\big| + \sum_{i=1}^{M} \big|\chi_{ij}\big| + 1 \le \varepsilon_j^{(\zeta)} + \varepsilon_j^{(\chi)} + 1,$$
$$\|z\|_2 \le \|\zeta\|_F \|u_1\|_2 + \|\chi\|_F \|u_2\|_2 + \|u_2\|_2 \le \varepsilon^{(\zeta)} + \varepsilon^{(\chi)} + 1.$$

Therefore,

$$g_4(x) \leq \max_{\substack{z\in\mathbb{R}^M \\ ||z||_2\leq \varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1 \\ |z_j|\leq \varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}} z^T x^{(2)} = \left(\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1\right) \max_{\substack{z\in\mathbb{R}^M \\ ||z||_2\leq 1 \\ |z_j|\leq \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1}}} z^T x^{(2)},$$

which is a strictly feasible conic optimization problem. Dualizing the constraints (see Lemma 3 in the Appendix for details) we obtain:

$$g_4(x) \leq \left(\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1\right) \min_{\lambda+\mu=x^{(2)}} \left\{ \|\lambda\|_2 + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} |\mu_j| \right\}. \quad (61)$$

Equations (59), (59) and (61) imply the statement of the Proposition 2.

9.3 Proof of the Proposition 3

Proof The statement of the Proposition 3 directly follows from the Lemma 4, implying the following upper bound:

$$\begin{aligned}
&\max_{Q\in\Xi^{(F)}} ||Qx-x||_2 \leq \max_{\substack{||\xi||_F\leq\varepsilon^{(\xi)} \\ ||\psi||_F\leq\varepsilon^{(\psi)} \\ ||\zeta||_F\leq\varepsilon^{(\zeta)} \\ ||\chi||_F\leq\varepsilon^{(\chi)}}} \left\| \begin{pmatrix} (P+\xi-I_N)x^{(1)}+\zeta x^{(2)} \\ \psi x^{(1)}+(\chi-I_M)x^{(2)} \end{pmatrix} \right\|_2 \leq \\
&\leq \max_{\substack{||\xi||_F\leq\varepsilon^{(\xi)} \\ ||\zeta||_F\leq\varepsilon^{(\zeta)}}} \|(P+\xi-I_N)x^{(1)}+\zeta x^{(2)}\|_2 + \max_{\substack{||\psi||_F\leq\varepsilon^{(\psi)} \\ ||\chi||_F\leq\varepsilon^{(\chi)}}} \|\psi x^{(1)}+(\chi-I_M)x^{(2)}\|_2 = \\
&\leq \left(\|Px^{(1)}-x^{(1)}\|_2+\varepsilon^{(\xi)}\|x^{(1)}\|_2+\varepsilon^{(\zeta)}\|x^{(2)}\|_2\right)+ \\
&+\left(\|x^{(2)}\|_2+\varepsilon^{(\psi)}\|x^{(1)}\|_2+\varepsilon^{(\chi)}\|x^{(2)}\|_2\right) = \|Px^{(1)}-x^{(1)}\|_2+ \\
&+(\varepsilon^{(\xi)}+\varepsilon^{(\psi)})\|x^{(1)}\|_2+(1+\varepsilon^{(\zeta)}+\varepsilon^{(\chi)})\|x^{(2)}\|_2,
\end{aligned}$$

where I_N and I_M are identity matrices of sizes $N\times N$ and $M\times M$ correspondingly.

9.4 Proof of the Statement 1

Proof Consider the optimization problem

$$\min_{y^{(2)}\in\Sigma_M} \|y^{(2)}\|_{(b)} = \min_{\substack{y^{(2)}\in\Sigma_M \\ \lambda+\mu=y^{(2)}}} \left\{ \|\lambda\|_\infty + \sum_{j=1}^{M} \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} |\mu_j| \right\}.$$

One can rewrite it as the following linear optimization problem:

$$\min_{y^{(2)}\in\Sigma_M} \|y^{(2)}\|_{(b)} = \min_{\substack{y^{(2)}\geq 0\\ \sum_{j=1}^M y_j^{(2)}=1\\ \lambda+\mu=y^{(2)}\\ \lambda_j\leq u,\,\forall j\\ -\lambda_j\leq u,\,\forall j\\ \mu_j\leq v_j,\,\forall j\\ -\mu_j\leq v_j,\,\forall j}} \left\{ u + \sum_{j=1}^M \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M} v_j \right\}. \tag{62}$$

Let us denote $c_j = \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+M}$ and solve the optimization problem (62) by its dualization.
The Lagrangian $\mathscr{L}$ for the problem (62) is

$$\begin{aligned}\mathscr{L} &= u + \sum_{j=1}^M c_j v_j + \alpha\left(\sum_{j=1}^M y_j^{(2)} - 1\right) + \\ &+ \sum_{j=1}^M \left(\beta_j\big(\lambda_j+\mu_j-y_j^{(2)}\big) + \gamma_j\big(\mu_j - v_j\big) + \eta_j\big(-\mu_j - v_j\big)\right) + \\ &+ \sum_{j=1}^M \left(\kappa_j\big(\lambda_j - u\big) + \nu_j\big(-\lambda_j - u\big)\right),\end{aligned}$$

where dual variables are $\alpha\in\mathbb{R}$, $\beta\in\mathbb{R}^M$, $\gamma,\eta,\kappa,\nu\in\mathbb{R}^M_{\{0,+\}}$.
Differently, one can rewrite the Lagrangian in the following form:

$$\begin{aligned}\mathscr{L} &= u\left(1-\sum_{j=1}^M \kappa_j - \sum_{j=1}^M \nu_j\right) + \sum_{j=1}^M v_j\left(c_j - \gamma_j - \eta_j\right) + \sum_{j=1}^M y_j^{(2)}\left(\alpha - \beta_j\right) + \\ &+ \sum_{j=1}^M \lambda_j\left(\beta_j + \kappa_j - \nu_j\right) + \sum_{j=1}^M \mu_j\left(\beta_j + \gamma_j - \eta_j\right) - \alpha.\end{aligned}$$

By strong duality, the following holds

$$\min_{y^{(2)}\in\Sigma_M} \|y^{(2)}\|_{(b)} = \min_{\substack{y^{(2)}\in\mathbb{R}^M_{\{0,+\}}\\ u\in\mathbb{R}_{\{0,+\}}\\ v\in\mathbb{R}^M_{\{0,+\}}\\ \lambda,\mu\in\mathbb{R}^M}} \max_{\substack{\alpha\in\mathbb{R}\\ \beta\in\mathbb{R}^M\\ \gamma,\eta,\kappa,\nu\in\mathbb{R}^M_{\{0,+\}}}} \mathscr{L} = \max_{\substack{\alpha\in\mathbb{R}\\ \beta\in\mathbb{R}^M\\ \gamma,\eta,\kappa,\nu\in\mathbb{R}^M_{\{0,+\}}}} \min_{\substack{y^{(2)}\in\mathbb{R}^M_{\{0,+\}}\\ u\in\mathbb{R}_{\{0,+\}}\\ v\in\mathbb{R}^M_{\{0,+\}}\\ \lambda,\mu\in\mathbb{R}^M}} \mathscr{L},$$

where the following must hold true:

$$\begin{cases} 1-\sum_{j=1}^M \kappa_j - \sum_{j=1}^M \nu_j \geq 0, \\ c_j - \gamma_j - \eta_j \geq 0,\ \forall j = 1,...,N, \\ \alpha - \beta_j \geq 0,\ \forall j = 1,...,N, \\ \beta_j + \kappa_j - \nu_j = 0,\ \forall j = 1,...,N, \\ \beta_j + \gamma_j - \eta_j = 0,\ \forall j = 1,...,N. \end{cases}$$

Therefore,

$$\min_{y^{(2)}\in\Sigma_M}\|y^{(2)}\|_{(b)} = \max_{\substack{\tilde{a}\in\mathbb{R}\\ \tilde{b}\in\mathbb{R}^M\\ \gamma,\eta,\kappa,\nu\in\mathbb{R}^M_{\{0,+\}}}} \left\{\tilde{a}, \text{ subject to}\right.$$

$$\sum_{j=1}^{M}(\kappa_j+\nu_j)\leq 1,$$
$$c_j\geq \eta_j+\gamma_j,\ \forall j=1,...,N,$$
$$\tilde{a}\leq \tilde{b}_j,\ \forall j=1,...,N,$$
$$\tilde{b}_j=\kappa_j-\nu_j,\ \forall j=1,...,N,$$
$$\left.\tilde{b}_j=\gamma_j-\eta_j,\ \forall j=1,...,N\right\},$$

where we denote $\tilde{a}=-\alpha$ and $\tilde{b}_j=-\beta_j$. At optimality, $\nu_j=\eta_j=0$ and, therefore, one can rewrite the optimization problem in the following equivalent form:

$$\min_{y^{(2)}\in\Sigma_M}\|y^{(2)}\|_{(b)} = \max_{\substack{\tilde{a}\in\mathbb{R}\\ \tilde{b}\in\mathbb{R}^M}} \left\{\tilde{a}, \text{ subject to}\right.$$

$$\sum_{j=1}^{M}\tilde{b}_j\leq 1,$$
$$\tilde{a}\leq \tilde{b}_j,\ \forall j=1,...,N,$$
$$\left.c_j\geq \tilde{b}_j,\ \forall j=1,...,N\right\}.$$

According to this optimization problem, if $\exists\, \tilde{b}_j>\frac{1}{M}$, then $\tilde{a}<\frac{1}{M}$ $\forall j$. Hence, the best possible choice for $\tilde{b}_j$ would be $\frac{1}{M}$. However, there is an additional constraint $c_j\geq\tilde{b}_j,\ \forall j=1,...,M$. Therefore, if $\min_j\{c_j\}<\frac{1}{M}$, then $\tilde{a}=\min_j\{c_j\}<\frac{1}{M}$.
One can summarize it as follows:

$$\min_{y^{(2)}\in\Sigma_M}\|y^{(2)}\|_{(b)} = \begin{bmatrix}\frac{1}{M},\ \text{if } c_j\geq\frac{1}{M},\ \forall j=1,...,M\\ \min_j\{c_j\},\ \text{if } \min_j\{c_j\}<\frac{1}{M},\end{bmatrix}$$

which implies the Statement (1).

9.5 Proof of the Statement 2

Proof Consider the optimization problem

$$\min_{y^{(2)}\in\Sigma_M}\|y^{(2)}\|_{(d)} = \min_{\substack{y^{(2)}\in\Sigma_M\\ \lambda+\mu=y^{(2)}}}\left\{\|\lambda\|_2+\sum_{j=1}^{M}\frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1}|\mu_j|\right\},$$

which is equivalent to the following convex optimization problem:

$$\min_{y^{(2)}\in\Sigma_M} \|y^{(2)}\|_{(d)} = \min_{\substack{y^{(2)}\geq 0\\ \sum_{j=1}^M y_j^{(2)}=1\\ \lambda+\mu=y^{(2)}\\ \mu_j\leq v_j,\forall j\\ -\mu_j\leq v_j,\forall j}} \left\{\|\lambda\|_2 + \sum_{j=1}^M \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1} v_j\right\}. \tag{63}$$

Let us denote $c_j = \frac{\varepsilon_j^{(\zeta)}+\varepsilon_j^{(\chi)}+1}{\varepsilon^{(\zeta)}+\varepsilon^{(\chi)}+1}$ and solve the optimization problem (63) by its dualization.

The Lagrangian $\mathscr{L}$ for the problem (63) is

$$\begin{aligned}\mathscr{L} &= \|\lambda\|_2 + \sum_{j=1}^M c_j v_j + \alpha\Big(\sum_{j=1}^M y_j^{(2)} - 1\Big) + \\ &+ \sum_{j=1}^M \Big(\beta_j\big(\lambda_j+\mu_j-y_j^{(2)}\big) + \gamma_j\big(\mu_j - v_j\big) + \eta_j\big(-\mu_j - v_j\big)\Big),\end{aligned}$$

where dual variables are $\alpha\in\mathbb{R}$, $\beta\in\mathbb{R}^M$, $\gamma,\eta\in\mathbb{R}^M_{\{0,+\}}$.

Differently, one can rewrite the Lagrangian in the following form:

$$\begin{aligned}\mathscr{L} &= \Big(\|\lambda\|_2 + \sum_{j=1}^M \beta_j\lambda_j\Big) + \sum_{j=1}^M v_j\Big(c_j - \gamma_j - \eta_j\Big) + \\ &+ \sum_{j=1}^M y_j^{(2)}\Big(\alpha - \beta_j\Big) + \sum_{j=1}^M \mu_j\Big(\beta_j + \gamma_j - \eta_j\Big) - \alpha.\end{aligned}$$

By strong duality (guaranteed by the Slater's conditions), the primal and the dual problems are equivalent:

$$\min_{y^{(2)}\in\Sigma_M} \|y^{(2)}\|_{(d)} = \min_{\substack{y^{(2)}\in\mathbb{R}^M_{\{0,+\}}\\ v\in\mathbb{R}^M_{\{0,+\}}\\ \lambda,\mu\in\mathbb{R}^M}} \max_{\substack{\alpha\in\mathbb{R}\\ \beta\in\mathbb{R}^M\\ \gamma,\eta\in\mathbb{R}^M_{\{0,+\}}}} \mathscr{L} = \max_{\substack{\alpha\in\mathbb{R}\\ \beta\in\mathbb{R}^M\\ \gamma,\eta\in\mathbb{R}^M_{\{0,+\}}}} \min_{\substack{y^{(2)}\in\mathbb{R}^M_{\{0,+\}}\\ v\in\mathbb{R}^M_{\{0,+\}}\\ \lambda,\mu\in\mathbb{R}^M}} \mathscr{L}.$$

For all j such that $\beta_j = 0$, we can conclude, that $\lambda_j = 0$ at optimality. For all j such that $\beta_j \neq 0$, optimal λ_j must follow the equation:

$$\frac{\partial\mathscr{L}}{\partial\lambda_j} = \frac{\lambda_j}{\|\lambda\|_2} + \beta_j = 0,\ \forall j = 1,...,N.$$

From these follows

$$\sum_{j:\,\beta_j\neq 0} \frac{\lambda_j^2}{\|\lambda\|_2^2} = \sum_{j:\,\beta_j\neq 0} \beta_j^2 \leq 1,$$

where the last inequality holds due to the fact that $\sum_{j=1}^M \frac{\lambda_j^2}{\|\lambda\|_2^2} = 1$ (i.e. the equality holds if $\beta_j \neq 0$, $\forall j = 1,...,M$).

Hence, the following conditions must be satisfied in order to guarantee feasibility of the optimization problem:

$$\begin{cases} \sum_{j=1}^{M} \beta_j^2 \leq 1, \\ \beta_j + \gamma_j - \eta_j = 0,\ \forall j = 1, \ldots, N, \\ c_j - \gamma_j - \eta_j \geq 0,\ \forall j = 1, \ldots, N, \\ \alpha - \beta_j \geq 0,\ \forall j = 1, \ldots, N, \end{cases}$$

Therefore,

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(d)} = \max_{\substack{\tilde{a} \in \mathbb{R} \\ \tilde{b} \in \mathbb{R}^M \\ \gamma, \eta \in \mathbb{R}^M_{\{0,+\}}}} \Bigg\{ \tilde{a}, \text{ subject to} \\ \sum_{j=1}^{M} \tilde{b}_j^2 \leq 1, \\ c_j \geq \eta_j + \gamma_j,\ \forall j = 1, \ldots, N, \\ \tilde{a} \leq \tilde{b}_j,\ \forall j = 1, \ldots, N, \\ \tilde{b}_j = \gamma_j - \eta_j,\ \forall j = 1, \ldots, N \Bigg\},$$

where we denoted $\tilde{a} = -\alpha$ and $\tilde{b}_j = -\beta_j$.
At optimality, $\eta_j = 0$. Hence, one can rewrite the optimization problem in the following equivalent form:

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(d)} = \max_{\substack{\tilde{a} \in \mathbb{R} \\ \tilde{b} \in \mathbb{R}^M}} \Bigg\{ \tilde{a}, \text{ subject to} \\ \sum_{j=1}^{M} \tilde{b}_j^2 \leq 1, \\ \tilde{b}_j \leq c_j,\ \forall j = 1, \ldots, N, \\ \tilde{a} \leq \tilde{b}_j,\ \forall j = 1, \ldots, N \Bigg\},$$

which solution can be summarized as follows:

$$\min_{y^{(2)} \in \Sigma_M} \|y^{(2)}\|_{(d)} = \begin{bmatrix} \frac{1}{\sqrt{M}}, \text{ if } c_j \geq \frac{1}{\sqrt{M}},\ \forall j = 1, \ldots, M \\ \min_j\{c_j\}, \text{ if } \min_j\{c_j\} < \frac{1}{\sqrt{M}}, \end{bmatrix}$$

which leads to the Statement (2).