PERTURBATION AND INVERSE PROBLEMS OF STOCHASTIC MATRICES

JOOST BERKHOUT , BERND HEIDERGOTT , AND PAUL VAN DOOREN

Abstract. Perturbation analysis of stochastic matrices is a classical area of research concerned with finding norm bounds on the effect of a perturbation matrix Δ of a stochastic matrix G on its stationary distribution, i.e., the unique normalized left Perron eigenvector. A common assumption is to consider Δ to be given and to find bounds on its impact, but in this paper, we rather focus on an inverse optimization problem called Target Stationary Distribution Problem (TSDP). The starting point is a target stationary distribution, and we search for a perturbation Δ of minimum norm such that $G + \Delta$ remains stochastic and has the desired target stationary distribution. It is shown that TSDP has relevant applications in the design of, for example, road networks, social networks, hyperlink networks, and queuing systems. The key to our approach is that we work with rank-1 perturbations. Building on those results for rank-1 perturbations, we provide a methodology to construct arbitrary rank perturbations as sums of appropriately constructed rank-1 perturbations.

Key words. Markov Chains, Perturbation Analysis, Inverse Problems, Target Stationary Distribution Problem

AMS subject classifications. 60J10, 93C73, 65F15

1. Introduction.

In this paper, we analyze perturbations of finite-dimensional Markov chains. We are given an irreducible stochastic matrix G with stationary distribution $\mu > 0$, which is the unique normalized left Perron eigenvector with $\mu^\top \mathbf{1} = 1$, where $\mathbf{1}$ is a vector of ones. Throughout the paper, we will use the terms "stochastic matrix" and "Markov chain" as synonyms. We study the *Target Stationary Distribution Problem* (TSDP) of finding the smallest-norm perturbation Δ so that the perturbed stochastic matrix $G+\Delta$ has a given target stationary distribution $\hat{\mu}$ ($\neq \mu$). More specifically, for G and $\hat{\mu}$ given, the TSDP is given by

$$\begin{aligned} \min_{\Delta} \quad & \|\Delta\| \\ \text{s.t.} \quad & \hat{\mu}^\top (G + \Delta) = \hat{\mu}^\top, \\ & \Delta \mathbf{1} = 0, \\ & G + \Delta \geq 0, \end{aligned} \tag{1.1}$$

for some specific norm $\|\cdot\|$ that is relevant for the considered application. The feasible set of (1.1) can be characterized using [14] as $\Delta = \hat{G} - G$ for all $\hat{G}$ in the convex polytope of stochastic matrices with stationary distribution $\hat{\mu}$. This feasible set always contains $\mathbf{1}\hat{\mu}^\top - G$, where the rank-1 matrix $\mathbf{1}\hat{\mu}^\top$ is the Riesz projector associated with the Perron root 1, also known as the ergodic projector in Markov chain theory. We study the TSDP for the 1-norm, the 2-norm, the v-norm, and the ∞-norm (see Section 4 for definitions). As shown later on, for some of these norms, the problem can be cast into a linear programming (LP) problem that can be solved in polynomial time. However, it is shown that solving a corresponding LP is computationally infeasible for realistically-sized instances.

We are considering applications where the stochastic matrices G and $G+\Delta$ model, for example, hyperlink networks, social networks, or queuing networks. Their stationary distributions contain important information on the nodes in the network, such as their centrality or other types of rankings. The target stationary distribution $\hat{\mu}$ then captures some desired state of the system. In practice, one is interested in reaching that desired state with minimum effort, i.e., we are interested in finding minimal norm perturbations. For example, a social agent may want to obtain a certain influence level within a social network with minimum effort.

TSDP deviates from problems in the literature on perturbation analysis of Markov chains (see Section 3 for details), where G and Δ are considered given, and bounds on the impact of the perturbation Δ on the stationary distribution of $G+\Delta$ compared to that of G are established. In this paper, we address the inverse problem and ask: "what kind of perturbations can attain a given stationary distribution?". The focus of this paper is on gaining a deeper understanding of perturbation analysis and of the structure of the solutions to our problem, but also to provide algorithms to compute (approximate) solutions.

For convenience, we define the feasible set of the TSDP as $\boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}$, where

$$\boldsymbol{\Delta}^{\hat{\mu}} := \boldsymbol{\Delta}^{\hat{\mu}}(G,\hat{\mu}) := \{\Delta : \hat{\mu}^\top (G+\Delta) = \hat{\mu}^\top,\ \Delta\mathbf{1} = 0\}$$

and

$$\boldsymbol{\Delta}^{\geq 0} := \boldsymbol{\Delta}^{\geq 0}(G) := \{\Delta : G + \Delta \geq 0\},$$

in which the arguments in brackets are omitted for simplicity when appropriate. These definitions allow us to write our TSDP as $\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$.

The key step in our analysis is to look at rank-1 perturbations. We justify this by showing that in relevant settings, explicit rank-1 perturbations can be found that solve subproblem $\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|$, which sometimes also solves the TSDP and in any case, provides bounds on the TSDP's solution(s). In particular, defining rank-1 perturbations as

$$\boldsymbol{\Delta}^{\text{rank-1}} = \{\Delta : \text{rank}(\Delta) = 1\},$$

we will provide problem instances where $\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\text{rank-1}}} \|\Delta\| = \min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$. Similarly, we study the TSDP when only rank-1 perturbations are allowed. To that end, define

$$\boldsymbol{\Delta} := \boldsymbol{\Delta}(G,\hat{\mu}) := \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0} \cap \boldsymbol{\Delta}^{\text{rank-1}}$$

where again, the arguments in brackets are omitted for simplicity when appropriate. We thus also study the problem $\min_{\Delta\in\boldsymbol{\Delta}} \|\Delta\|$ and present explicit constructions to find a solution. Note that if solutions exist, it holds that

$$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\| \leq \min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\geq 0}} \|\Delta\| \leq \min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|,$$

and we will show (in Section 5) that $\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\| = \min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\text{rank-1}}} \|\Delta\|$.

We call a perturbation *non-structural* if G and $G+\Delta$ have the *same support*, and call it *structural* if G and $G+\Delta$ have *different support*, where the support of a matrix A is defined as the set of indices (i,j) for which $A_{i,j} \neq 0$. The distinction between structural and non-structural perturbations is motivated by the fact that removing or adding links in a network is of a different nature than adjusting the weight of an established link. We will provide results that show "how far" μ can be moved towards $\hat{\mu}$ without having to change the support of $G+\Delta$. The feasible set of non-structural perturbations can be characterized using the results from [6].

The price we have to pay for the analytical elegance and simplicity of our explicit rank-1 solutions is that they may not solve the TSDP. Fortunately, as we show in this paper, in such cases, an approximate (i.e., not achieving minimal norm) solution Δ can often be obtained via a sequence of rank-1 perturbations. We develop heuristics for finding a sequence of rank-1 perturbation steps so that the accumulated perturbation is of higher rank and does allow to reach the target stationary distribution. Numerical experiments will show the efficiency of our approach for dense random matrices and for specific sparse matrices.

The paper is organized as follows. Motivating applications are presented in Section 2, and a literature survey is given in Section 3. Section 4 is devoted to technical preliminaries, and Section 5 focuses on $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|$. Section 6 presents the analysis of $\min_{\Delta \in \boldsymbol{\Delta}} \|\Delta\|$, and Section 7 analyses the same problem when perturbations can only affect one row (which is often the case in practice). Finally, a heuristic for approximately solving $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$ as sums of rank-1 perturbations is presented in Section 8. The paper concludes with suggestions for further research. Proofs, additional examples, and detailed numerical results are given in the appendix.

2. Applications.

In this section, we mention a series of applications as motivation for this research. Illustrating examples from these applications will also be given at a later stage.

2.1. Road Networks.

Consider a road network consisting of n roads and represented by a graph in the following manner. The nodes of the graph represent the road segments, and a directed link between road segments i and j means that there is a junction that allows going from road i to road j. We assume that information is available on the traffic flow and that this is modeled as a discrete-time random walk on the road network: the probability $G_{i,j}$ thus models the probability that a car on road segment i turns into road segment j at a particular time instant. As detailed in [11], self-links $G_{i,i}$ can be chosen in such a way that they mimic the travel times along the road elements. The value of μ_i of the stationary distribution of G then represents the long-range time average with which a typical car will be found to drive on road segment i. In other words, the vector μ represents the (relative) road congestion.

In perturbation analysis, we have a desired traffic distribution given by $\hat{\mu}$, and Δ is the adjustment in traffic that will achieve a transition from the distribution μ towards the $\hat{\mu}$ regime. The condition that Δ should be minimal follows naturally from the fact that influencing the traffic by, e.g., signaling or changes to the road infrastructure, is costly. Moreover, a 1-norm minimal Δ reflects an adjustment that is easier for travelers on the network to adjust to. Finally, in this setting, it is preferable that Δ is non-structural so that no road segments have to be built or closed down since this would lead to substantial costs.

2.2. Social Networks.

Social network analysis investigates the social structures of relationships between agents [44]. A social network can be modeled as a finite set of nodes, and the edges connecting them represent the social relationship between the nodes [36]. Social networks are typically represented by weighted graphs, where the nodes set is the set of social agents, and a directed link (i, j) between agents i and j means either that i follows j (i.e., i puts trust in j) or that i influences j (i.e., i sends information to j). The relative strength of the link is expressed via a weight function $W_{i,j}$. Through normalizing the weights, a Markov chain G can be constructed of which the stationary distribution expresses the influence or centrality of the social agents. For example, if the weights reflect trust, then the stationary distribution expresses the relative trust the agents receive in the network.

In perturbation analysis of social networks, one is interested in perturbing the stationary distribution. For example, agent i can influence his or her outgoing nodes, and the question arises which perturbation of the i-th row will maximize the importance of i. In the same vein, agent i may be interested in decreasing the importance of some other node $j \neq i$ by adjusting its outgoing links. Finally, coalition games can be considered where a group of agents S either wants to maximize its importance or tries to diminish the importance of another group Z, with $S \cap Z = \emptyset$. As social networks are typically obtained from collecting data based on observations and questionnaires, it is of interest to identify the maximal Δ such that $\hat{\mu}_i$ changes no more than some pre-specified value, which amounts to a robustness analysis. Finally, it is of interest to identify the minimal Δ such that $\hat{\mu}_i$ does change no more than some pre-specified precision value, which provides a safety margin against organized attacks on the network.

2.3. Hyperlink Networks. Consider an unweighted directed graph with n nodes modeling a hyperlink network such as the world-wide-web. Perturbation analysis of hyperlink networks is well studied and we refer to [20]. The PageRank algorithm originally introduced for unweighted graphs has been extended to weighted graphs, where the weight of a link (i, j) is the number of outlinks of page j divided by the total number of outlinks of all webpages i is directly linked to (see [43, 42]). The input data on link visits defines, after appropriate rescaling, a Markov chain G. Assuming, for ease of presentation, that the resulting Markov chain is irreducible, we may set the damping factor to 1 in the PageRank algorithm. The PageRank vector is then equal to the stationary distribution of G, and the owner of page i is then interested in boosting the PageRank of page i. Due to the dynamics of the web, the inflow to page i is variable. Indeed, the owner of page i may choose to invest in a better, more prominent placement of the link (j, i) of some webpage j, thereby increasing the weight of link (j, i). This leads to the problem of finding the smallest Δ that maximizes the value of $\hat{\mu}_i$ by focusing on the i-th column.

2.4. Queuing Systems. Markov models are prominent in the analysis of queuing systems. Under appropriate conditions, such as exponentially distributed inter-arrival times and service times of customers, the discrete-time queue-length process becomes a Markov chain on an at most denumerable state space. These Markov chains typically have a so-called *birth-and-death* structure, i.e., the corresponding matrix is tridiagonal and irreducible. For illustration and later use, we present a typical Markov model in the following example.

Consider the queue-length process of an M/M/s/K queue, where $s \geq 1$ denotes the number of service places and $K \geq 0$ is the number of buffer places. Let λ denote the arrival rate and ν the service rate. For $\eta = \lambda + s\nu$, where $\lambda, \nu > 0$, the subordinated chain of the queue-length process of an M/M/s/K queue has the following $(1 + s + K) \times (1 + s + K)$ transition probability matrix G in which only the non-zeros are indicated:

$$\begin{bmatrix}
1-\lambda/\eta & \lambda/\eta & & & & & & \\
\nu/\eta & 1-(\lambda+\nu)/\eta & \lambda/\eta & & & & & \\
 & 2\nu/\eta & 1-(\lambda+2\nu)/\eta & \lambda/\eta & & & & \\
 & & \ddots & \ddots & \ddots & & & \\
 & & s\nu/\eta & 1-(\lambda+s\nu)/\eta & \lambda/\eta & & & \\
 & & & s\nu/\eta & 1-(\lambda+s\nu)/\eta & \lambda/\eta & & \\
 & & & \ddots & \ddots & \ddots & & \\
 & & & & s\nu/\eta & 1-(\lambda+s\nu)/\eta & \lambda/\eta \\
 & & & & & s\nu/\eta & 1-s\nu/\eta
\end{bmatrix}.$$

In robustness analysis of queuing networks, it is of importance to relate a change in stationary distribution to a perturbation Δ of an analytically tractable model such as the one above.

3. Literature Survey. Perturbation analysis of stochastic matrices studies the effect a perturbation Δ of a stochastic matrix G has on the stationary distribution of G, where $G + \Delta$ is stochastic. More formally, in perturbation analysis, one looks for establishing a bound

$$\|\hat{\mu} - \mu\|_\alpha \leq D(\Delta, G), \tag{3.1}$$

where $\|\cdot\|_\alpha$ denotes a suitable vector norm, and $D(\Delta, G)$ is a scalar function of Δ and G. This type of perturbation analysis dates back to Schweitzer's pioneering paper [35]. To the best of our knowledge, the first paper putting this perturbation question into the framework of (3.1) is [27]. This paper proposed bounds of the form

$$D(\Delta, G) = \kappa \|\Delta\|_\beta, \tag{3.2}$$

for some appropriate matrix norm $\|\cdot\|_\beta$, where κ is the so-called *condition number* of the Markov chain G for the $(\|\cdot\|_\alpha, \|\cdot\|_\beta)$-norm pair. Finding bounds of the type (3.1) is a field of active research [4, 21, 32, 33, 37, 30, 8, 18, 5, 26, 1] and various condition number bounds have been proposed in the literature [10, 18]. Perturbation bounds like (3.1) are of interest in a wide range of application areas, such as mathematical physics [41], climate modeling [9], Bayesian statistics [3, 2], and bio-informatics [29, 34]. Conditions numbers for quantum Markov chains in mathematical physics can be found in [40].

In our paper, we address the inverse problem: we take G and $\hat{\mu}$ as starting point, and we search for Δ such that (i) $G + \Delta$ is stochastic, (ii) $\hat{\mu}$ is the normalized left Perron vector of $G + \Delta$, and (iii) Δ has minimum norm.

4. Technical Preliminaries. In this paper, we consider square $n \times n$ non-negative matrices A, i.e., matrices with non-negative elements, which we denote by $A \geq 0$. If in the matrix A all elements are strictly larger than 0, we call A a positive matrix and denote this by $A > 0$. The positive semi-definite matrices, on the other hand, will be denoted by $A \succeq 0$. The support of a general matrix A, denoted by $\text{supp}(A)$, is the set of indices (i, j) for which $A_{i,j} \neq 0$. It is well known that non-negative matrices have an eigenvalue that is equal to its spectral radius $\rho := \rho(A)$ and hence is real and non-negative. Moreover, if A is irreducible, then this so-called *Perron-root* ρ is simple and positive. Therefore the matrix $(A - \rho I)$ has rank $n - 1$, where I denotes the $n \times n$ identity matrix. Moreover, the corresponding left and right eigenvectors $\mathbf{v}$ and $\mathbf{u}$ are also positive, i.e., $\mathbf{v}^\top A = \rho \mathbf{v}^\top > 0$ and $A\mathbf{u} = \rho \mathbf{u} > 0$. The Perron vectors are typically normalized using $\mathbf{v}^\top \mathbf{1} = 1$ and $\mathbf{1}^\top \mathbf{u} = 1$, where $\mathbf{1}$ is the n-vector of all ones. The non-negative matrix A is said to be stochastic if $A\mathbf{1} = \mathbf{1}$. For such a matrix, the spectral radius $\rho(A) = 1$. We will denote the i-th canonical basis vector of $\mathbb{R}^n$ by e_i.

The dual norm of a vector $y \in \mathbb{R}^n$ for a vector norm $\|\cdot\|$, is defined as

$$\|y\|_* := \sup_{z \neq 0} \frac{|z^\top y|}{\|z\|} = \sup_{z \neq 0} \frac{|y^\top z|}{\|z\|}. \tag{4.1}$$

Vector norms are extended to matrix norms by using the subordinate norm defined via

$$\|A\| := \sup_{z \neq 0} \frac{\|Az\|}{\|z\|}. \tag{4.2}$$

For $x \in \mathbb{R}^n$, we denote by $\|x\|_\infty$ the maximum absolute value (a.k.a. the infinity norm or ∞-norm), by $\|x\|_2$ the square root of the sum of the squared entries of x (a.k.a. the 2-norm or L_2-norm), and by $\|x\|_1$ the sum of absolute values (a.k.a. the L_1 norm or 1-norm). Furthermore, for $v \geq \mathbf{1}$ and $v_1 = 1$, we define

$$\|x\|_v := \sup_{1\leq i \leq n} \frac{|x_i|}{v_i} = \|D_v^{-1}x\|_\infty \quad \text{where} \quad D_v = \operatorname{diag}(v_1, \dots, v_n) \tag{4.3}$$

for $x \in \mathbb{R}^n$, which is called the *v-norm*. In the following, we choose

$$v_i = \alpha^i, \quad 1 \leq i \leq n,$$

with $\alpha \in [1, \infty)$ some specified constant. The v-norm is frequently used in the analysis of denumerable Markov chains that exhibit a drift towards a small finite set; think, for example, of a queuing model where stability implies the queue has the tendency to return to the empty state; see [28]. The v-norm, as defined above, was restricted to the finite-dimensional case. In the following, we will omit the subscript α whenever the results stated hold for general $\alpha \geq 1$. Following (4.1) the dual norm of the v-norm is given by $\|y\|_{v,*} = \sum_i v_i |y_i|$, and following (4.2) the subordinate matrix norm for the v-norm satisfies $\|A\|_v = \|D_v^{-1} A D_v\|_\infty$.

5. General Rank-1 Perturbations.

In this section, we show that rank-1 perturbations can be used to try to solve the TSDP. We will drop the constraint that Δ has to belong to $\mathbf{\Delta}^{\geq 0}$ and impose instead that Δ is rank-1, that is, we consider

$$\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}}} ||\Delta||. \tag{5.1}$$

While the following theorem is fairly standard, we provide, for the sake of completeness, a proof in Appendix A.1.

Theorem 5.1. *Any matrix $\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}}$ can be written as $\Delta = \frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}(I - G)$, for some x such that $\hat{\mu}^\top x \neq 0$, i.e.,*

$$\mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}} = \{\Delta : \Delta = \frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}(I - G) \textit{ for all } x \textit{ with } \hat{\mu}^\top x \neq 0\},$$

where the rank-1 matrix $\frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}$ is the skew projector onto the range of x and parallel to $\hat{\mu}$. For any subordinate matrix norm, a minimum norm choice of $\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}}$ is obtained by any x such that

$$\frac{|x^\top \hat{\mu}|}{\|x\|} = \|\hat{\mu}\|_* \tag{5.2}$$

and the corresponding minimum norm Δ has the norm

$$\|\Delta\| = \|(I - G)^\top \hat{\mu}\|_* / \|\hat{\mu}\|_*. \tag{5.3}$$

Moreover, these are also minimizers of arbitrary rank in $\mathbf{\Delta}^{\hat{\mu}}$, i.e.

$$\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}}} \|\Delta\| = \min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}}} \|\Delta\|.$$

Remark 5.2. It follows from (5.3) and $(I-G)^\top \mu = 0$ that the minimum norm rank-1 matrices Δ in the set $\mathbf{\Delta}^{\hat{\mu}}$ satisfy

$$\|\Delta\| = \|(I-G)^\top(\hat{\mu}-\mu)\|_*/\|\hat{\mu}\|_* \leq (1+\|G\|)\|\hat{\mu}-\mu\|_*/\|\hat{\mu}\|_*$$

which bounds those minimum norm Δ's in terms of the requested perturbation $\hat{\mu}-\mu$. This can be viewed as a converse perturbation theorem to the classical results described in Section 3.

Note that the projector does not depend on the scaling factor of x, but only on its direction. The following corollary provides explicit expressions for minimal rank-1 norm perturbations.

COROLLARY 5.3. *The solutions to (5.1) for the 1-, 2-, v- and ∞-norms are given by the vectors cx (with scale factor $c \neq 0$), where x is defined as follows:*
- *for the 1-norm: $x = e_i$ where i is any maximizing index of the vector $\hat{\mu}$, and $\|\Delta\|_1 = \|(I-G)^\top\hat{\mu}\|_\infty/\|\hat{\mu}\|_\infty$*
- *for the 2-norm: $x = \hat{\mu}$, and $\|\Delta\|_2 = \|(I-G)^\top\hat{\mu}\|_2/\|\hat{\mu}\|_2$*
- *for the v-norm: $x = D_v\mathbf{1}$, and $\|\Delta\|_v = \|D_v(I-G)^\top\hat{\mu}\|_1/\|D_v\hat{\mu}\|_1$*
- *for the ∞-norm: $x = \mathbf{1}$, and $\|\Delta\|_\infty = \|(I-G)^\top\hat{\mu}\|_1/\|\hat{\mu}\|_1$.*

We illustrate Corollary 5.3 with the following example.

EXAMPLE 1. *Let*

$$G = \begin{bmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{bmatrix}$$

with stationary distribution $\mu^\top = (9/17, 8/17)$. Following Corollary 5.3 for the ∞-norm, the smallest rank-1 perturbation to achieve a uniform distribution $\hat{\mu}^\top = (1/2, 1/2)$ is

$$\Delta = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2/3 & -2/3 \\ -3/4 & 3/4 \end{bmatrix} = \frac{1}{24}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

with $\|\Delta\|_\infty = 1/12$. Indeed,

$$G + \Delta = \frac{1}{24}\left(\begin{bmatrix} 8 & 16 \\ 18 & 6 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\right) = \frac{1}{24}\begin{bmatrix} 7 & 17 \\ 17 & 7 \end{bmatrix}$$

which has the stationary distribution $\hat{\mu}^\top$. So Δ perturbs elements $G_{1,2}$ and $G_{2,1}$ to their average $(G_{1,2}+G_{2,1})/2$.

Clearly, $G+\Delta$ is non-negative in the above example, and thus, $G+\Delta$ is stochastic as well, which means Δ is a solution to $\min_{\Delta\in\mathbf{\Delta}^{\hat{\mu}}\cap\mathbf{\Delta}^{\geq 0}}\|\Delta\|$. That $G+\Delta$ is again stochastic is a mere coincidence and does not hold in general, as we illustrate in Appendix B.1.

6. Rank-1 Perturbations Preserving Stochasticity.

In this section, we focus on solving

$$\min_{\Delta\in\mathbf{\Delta}} ||\Delta|| \tag{6.1}$$

that compared to Problem 5.1 also forces $G+\Delta$ to be non-negative (and thus stochastic). Therefore, a solution to (6.1) provides a candidate to Problem (1.1), i.e., $\min_{\Delta\in\mathbf{\Delta}^{\hat{\mu}}\cap\mathbf{\Delta}^{\geq 0}}\|\Delta\|$. We will characterize $\mathbf{\Delta}$ and provide explicit solutions to (6.1).

The following theorem characterizes $\mathbf{\Delta}$. For a proof, please refer to Appendix A.2.

THEOREM 6.1. *Let G be a given irreducible stochastic matrix with stationary distribution μ. Let $\hat{\mu}$ $(\neq \mu)$ be the target stationary distribution. Define $z := z(G, \hat{\mu})$ as*

$$z^\top := (\hat{\mu} - \mu)^\top (I - G) = z_+^\top + z_-^\top,$$

with $z_+ \geq 0$, $z_- \leq 0$ and $\operatorname{supp}(z_+) \cap \operatorname{supp}(z_-) = \emptyset$. Introduce vectors $\ell := \ell(G, \hat{\mu})$ and $u := u(G, \hat{\mu})$ defined through

$$\ell_i := \max_{j \in \operatorname{supp}(z_+)} \frac{-G_{i,j}}{z_j} \leq 0 \;\; \forall i, \quad u_i := \min_{j \in \operatorname{supp}(z_-)} \frac{-G_{i,j}}{z_j} \geq 0 \;\; \forall i. \tag{6.2}$$

Then the set of stochasticity-preserving rank-1 perturbations can be characterized by

$$\boldsymbol{\Delta}(G, \hat{\mu}) = \Big\{ \Delta = x z^\top : \forall x \in \mathbb{R}^n \text{ with } \hat{\mu}^\top x = 1 \text{ and } \ell \leq x \leq u \Big\}. \tag{6.3}$$

It further holds:
(i) $\boldsymbol{\Delta}(G, \hat{\mu}) \neq \emptyset$ if and only if $\hat{\mu}^\top u \geq 1$.
(ii) If $\hat{\mu}^\top u = 1$, then

$$\boldsymbol{\Delta}(G, \hat{\mu}) = \left\{ \Delta^* := \frac{u}{\hat{\mu}^\top u} \right\}, \tag{6.4}$$

i.e., Δ^* is the only candidate in $\boldsymbol{\Delta}(G, \hat{\mu})$.

The relation between Theorem 5.1 and Theorem 6.1 is that Theorem 5.1 provides the generic form of a minimal-norm rank-1 perturbation that possibly violates the stochasticity of the perturbed matrix $G + \Delta$, while Theorem 6.1 provides the generic form of a rank-1 perturbation that does not violate the stochasticity of the perturbed matrix $G + \Delta$ but is possibly not a solution to the TSDP.

How far can we go in the *direction* of $\hat{\mu}$ with a rank-1 perturbation? To answer this question, we introduce a scaling factor $\alpha > 0$ and consider as target stationary distribution $\hat{\mu}_\alpha := \mu + \alpha d$, where $d := \hat{\mu} - \mu$. It follows that $u(G, \hat{\mu}_\alpha) = u(G, \hat{\mu})/\alpha$. To simplify notation, let $u = u(G, \hat{\mu})$. Then, condition $\hat{\mu}_\alpha^\top u/\alpha = 1$ in Theorem 6.1 is satisfied by letting $\alpha \leq \alpha^* := \mu^\top u/(1 - d^\top u)$ when $d^\top u < 1$, otherwise α^* is effectively 0. In the following example, we consider the maximal feasible step-size α^*.

EXAMPLE 2. *Consider the queuing system with $s = 2$, $K = 1$, $\lambda = 1$ and $\nu = 1.8$, which has the stationary distribution $\mu = (0.5705, 0.317, 0.088, 0.0245)$. We then try to perturb this queuing system in order to achieve different stationary distributions $\hat{\mu}$ of the same queuing system with the same arrival rate $\hat{\lambda} = 1$ but with different service rates $\hat{\nu}$. For varying $\hat{\nu}$, Table 1 gives the corresponding $\hat{\mu}$, the extremal value $\alpha^\star$, the value of $\|\Delta\|_\infty$ for Δ from (6.4) with $\alpha = 1$ in case $\alpha^\star \geq 1$, and the minimum value of $\|\Delta\|_\infty$ for $\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}$ and $\Delta \in \boldsymbol{\Delta}(G, \hat{\mu})$.*

It follows from Table 1 that the $\hat{\mu}$'s for $\hat{\nu} = 0.2$ and $\hat{\nu} = 2$ are too different from μ to allow for a rank-1 perturbation. For $\hat{\nu} = 0.2$, however, $(1 - \alpha)\mu + \alpha\hat{\mu}$ can be reached for $\alpha \leq 0.223$. For the other $\hat{\nu}$'s, we can reach $\hat{\mu}$ and in fact we can even go beyond $\hat{\mu}$, for example, for $\hat{\nu} = 1.6$, the μ and $\hat{\mu}$ are close enough that we can reach $(1 - \alpha)\mu + \alpha\hat{\mu}$ for $1 < \alpha \leq 14.629$.

An interesting observation is that the rank-1 perturbation in Theorem 6.1 may lead to structural breaks. Recall that we call a perturbation *non-structural,* if G and $G + \Delta$ have the same support, and we call this perturbation *structural*, otherwise. Before we illustrate this with the following example, we point out that based upon

$\hat{\nu}$	$\hat{\mu}$	$\alpha^\star$	$\|\Delta\|_\infty$ of (6.4)	$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$
0.2	(0.02, 0.101, 0.251, 0.628)	0.223	no candidate	0.874	no solution
1.2	(0.43, 0.358, 0.149, 0.062)	4.062	0.153	0.093	0.119
1.4	(0.485, 0.347, 0.124, 0.044)	6.695	0.092	0.06	0.072
1.6	(0.532, 0.332, 0.104, 0.032)	14.629	0.041	0.029	0.033
2	(0.604, 0.302, 0.075, 0.019)	0	no candidate	0.026	no solution

Table 1: Perturbing a queuing system with $s = 2$, $K = 1$, $\lambda = 1$ and $\nu = 1.8$, with $\mu = (0.5705, 0.317, 0.088, 0.0245)$, to the same system with different service rates $\hat{\nu}$

the construction of the vectors z, u and x we can identify the subset of $\boldsymbol{\Delta}(G, \hat{\mu})$ such that $G + \Delta$, for $\Delta \in \boldsymbol{\Delta}(G, \hat{\mu})$, has the same support as G. This is discussed in Appendix C.

EXAMPLE 3. *The following two examples provide some instances of structural perturbations. Consider the $n \times n$* **ring network***, introduced below.*

$$G_r(b) = \begin{bmatrix} 1-2b & b & 0 & 0 & \dots & b \\ b & 1-2\,b & b & 0 & \dots & 0 \\ 0 & b & 1-2\,b & b & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & b & 1-2\,b & b \\ b & 0 & \dots & 0 & b & 1-2b \end{bmatrix},$$

with $b \in (0, 1/2]$. It has the stationary distribution $\mu = \frac{1}{n}\mathbf{1}$. Consider also the following $n \times n$ **star network**

$$G_s(\beta,\gamma) = \begin{bmatrix} 1-\beta & \frac{\beta}{n-1} & \frac{\beta}{n-1} & \frac{\beta}{n-1} & \cdots & \frac{\beta}{n-1} \\ 1-\gamma & \gamma & 0 & 0 & \dots & 0 \\ 1-\gamma & 0 & \gamma & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1-\gamma & 0 & \dots & 0 & \gamma & 0 \\ 1-\gamma & 0 & \dots & 0 & 0 & \gamma \end{bmatrix},$$

with $\beta \in (0, 1]$ and $\gamma \in [0, 1)$. Its stationary distribution is given by

$$\mu_1 = \frac{1-\gamma}{1-\gamma+\beta} \quad \textit{and} \quad \mu_i = \frac{\beta}{(n-1)(1-\gamma+\beta)}, \quad \textit{for } i > 1.$$

We now construct two small examples with these general structures. Consider $G_r(b) \in \mathbb{R}^{4\times 4}$ with $b = 0.3$. Its stationary probability is $\mu^\top = (1/4, 1/4, 1/4, 1/4)$. For $\hat{\mu}$ we take the stationary distribution of the star network $G_s(\beta, \gamma)$ with $\beta = \gamma = 0.9$, which is $\hat{\mu}^\top = (0.1, 0.3, 0.3, 0.3)$. The Δ obtained by Theorem 6.1 (without stepsize α) is

$$\Delta = \begin{bmatrix} -0.2182 & 0.1091 & 0 & 0.1091 \\ -0.1636 & 0.0818 & 0 & 0.0818 \\ 0 & 0 & 0 & 0 \\ -0.1636 & 0.0818 & 0 & 0.0818 \end{bmatrix},$$

with $\|\Delta\|_1 = 0.5455$. The perturbation is structural since $(G_{4,2} + \Delta_{4,2}) > 0$ while $G_{4,2} = 0$.

For a second example, we consider $G_s(\beta,\gamma)$ with $\beta = 0.2$, $\gamma = 0.9$, and we take for $\hat\mu$ the stationary distribution of the star network $G_s(\beta,\gamma)$ with $\beta = 0.3$, $\gamma = 0.3$. This gives

$$\Delta = \begin{bmatrix} 0.1571 & -0.0524 & -0.0524 & -0.0524 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

with $\|\Delta\|_1 = 0.1571$. This perturbation is non-structural as G and $G+\Delta$ have the same support. This happens to be also the minimal 1-norm of a rank-1 perturbation without the constraint $G + \Delta \geq 0$.

We now look at the matrices $\Delta = xz^\top$ of minimum norm in the feasible set $\Delta \in \boldsymbol{\Delta}(G,\hat\mu)$ introduced in (6.3) of Theorem 6.1. When looking for matrices of minimum norm, we can tighten the following conditions for x

$$\hat\mu^\top x = 1, \quad \ell \leq x \leq u, \tag{6.5}$$

to

$$\hat\mu^\top x = 1, \quad 0 \leq x \leq u, \tag{6.6}$$

because if x has a negative component, then replacing it by 0 maintains the feasibility of the candidate and can only reduce the norm of $\Delta = xz^\top$. If we want to minimize the subordinate norm of Δ, we need to minimize the corresponding vector norm of x, as indicated above.

Using the results of Theorem 6.1, the optimization problem $\min_{\Delta\in\boldsymbol{\Delta}} ||\Delta||$ can therefore be tightened to the following problem

$$\begin{aligned} \min_x \quad & \|xz^\top\| \\ \text{s.t.} \quad & \hat\mu^\top(G + xz^\top) = \hat\mu^\top, \\ & \hat\mu^\top x = 1, \\ & G + xz^\top \geq 0, \\ & 0 \leq x \leq u \end{aligned} \tag{6.7}$$

since the additional constraints do not affect the feasibility of the candidates $\Delta = xz^\top$. Sufficient conditions for a candidate of (6.7) for specific norms are provided below. For the proof see Appendix A.3.

THEOREM 6.2. *Let G be an irreducible stochastic matrix with stationary distribution μ, and let $\hat\mu$ ($\neq \mu$) be a target stationary distribution. Define the vectors z, u and x as in Theorem 6.1, then Problem (6.7) has a solution if and only if $\hat\mu^\top u \geq 1$. This solution is unique if $\hat\mu^\top u = 1$ and is given by $\Delta = uz^\top$. If $\hat\mu^\top u > 1$, then $\Delta = xz^\top$ is a solution for every x solving the following convex optimization problems for the 1-, 2-, v- and ∞-norms, respectively:*

- $\min \gamma$, *s.t.* $x^\top \mathbf{1} \leq \gamma, \quad 0 \leq x \leq u, \quad \hat\mu^\top x = 1$ *for the 1-norm*
- $\min \gamma$, *s.t.* $\begin{bmatrix} \gamma & x^\top \\ x & \gamma I \end{bmatrix} \succeq 0, \quad 0 \leq x \leq u, \quad \hat\mu^\top x = 1$ *for the 2-norm*
- $\min \gamma$, *s.t.* $x \leq \gamma D_v \mathbf{1}, \quad 0 \leq x \leq u, \quad \hat\mu^\top x = 1$ *for the v-norm*
- $\min \gamma$, *s.t.* $x \leq \gamma \mathbf{1}, \quad 0 \leq x \leq u, \quad \hat\mu^\top x = 1$ *for the ∞-norm*

In Appendix D, we provide explicit constructions for solving the convex optimization problems given in Theorem 6.2.

7. Stochastic Rank-1 Perturbations That Only Affect One Row. When solving $\min_{\Delta \in \boldsymbol{\Delta}} ||\Delta||$, a natural condition is to impose that Δ has only non-zero elements in the i-th row, namely the row that is "controlled" by user i in the case that G represents the "influence" each node has on the other nodes, see the social network and the hyperlink network example in Section 2. In that case, $x = e_i/\hat{\mu}_i$ in the generic form of Δ from (6.3), i.e.,

$$\Delta = e_i\, z^\top / \hat{\mu}_i,$$

where we need

$$G_{i,j} + \frac{z_j}{\hat{\mu}_i} \geq 0, \ \forall j,$$

for $G + \Delta$ to be stochastic. As the above inequality does not hold in general, we study in this setting how far we can go in the direction of $\hat{\mu}$ while ensuring that $G + \Delta \geq 0$. To that end, the direction is $d^\top = \hat{\mu}^\top - \mu^\top$ and the relaxed target stationary distribution is $\hat{\mu}_\alpha = \mu + \alpha d$, where $\alpha \geq 0$ denotes a relaxation parameter. Note that $d^\top \mathbf{1} = 0$. The parameter α^* then gives the maximal value of α for which relaxed target $\hat{\mu}_\alpha$ can be reached while ensuring $G + \Delta \geq 0$ and where Δ is zero except for the i-th row. Below, we discuss two specific cases for this setting.

7.1. Increasing Only One Stationary Distribution Element. Let us assume that direction d is chosen such that the only positive value is d_i (corresponding to the row we perturb in G) and that the other values are negative or zero. A particular choice that is useful is $d = e_i - \mu$ which gives for the elements in $\hat{\mu} := \hat{\mu}_\alpha = \mu + \alpha d$:

$$\hat{\mu}_i := (\hat{\mu}_\alpha)_i = \mu_i + \alpha(1 - \mu_i) \quad \text{and} \quad \hat{\mu}_j := (\hat{\mu}_\alpha)_j = (1-\alpha)\mu_j, \ \ \forall j \neq i, \tag{7.1}$$

and the corresponding z and matrix $G + \Delta$ satisfy

$$z^\top = d^\top (I - G) = \alpha e_i^\top (I - G), \quad G + \Delta = G + \frac{\alpha}{\mu_i + \alpha(1-\mu_i)} e_i e_i^\top (I - G). \tag{7.2}$$

In this specific case, the lower bound $l(G, \hat{\mu})$ and upper bound $u(G, \hat{\mu})$ in (6.2) can be calculated explicitly. By (7.2), the perturbed matrix $G + \Delta$ is non-negative in its i-th row since

$$\frac{G_{i,j}\mu_i(1-\alpha)}{\mu_i + \alpha(1-\mu_i)} \geq 0 \ \ \forall j \neq i, \quad \frac{G_{i,j}\mu_i(1-\alpha) + \alpha}{\mu_i + \alpha(1-\mu_i)} \geq 0$$

holds for $0 < \alpha \leq 1$. This implies that $\alpha^* = 1$. For this extremal value, the i-th row of G becomes the vector $e_i^\top$, and the left eigenvector $\hat{\mu}^\top$ becomes the vector $e_i^\top$. In terms of ranking, this is also the best deal for node i since its so-called "reputation" is maximal. But, of course, eliminating all elements of the i-th row (except the diagonal element) is hardly achievable in practice. So a relaxation to a smaller value than the extremal $\alpha^* = 1$, ought to be recommended. The above perturbation results allow for a robustness analysis of G as detailed in the following example.

EXAMPLE 4. *Consider a traffic network G, see Section 2.1, where i represents a road segment that is of key importance for traffic congestion control. We tolerate deviations from the traffic network as long as they do not increase the congestion at i above a pre-specified fraction $\beta > 0$, and we compute the minimal perturbation of the given traffic network that reaches this tolerance bound. This minimal perturbation gives robustness insights on, for example, the maximal measurement errors we can* *accept to ensure that the current congestion at i does not exceed the pre-specified fraction, or on which road segments are crucial to be accurately measured. We are thus looking for a minimal perturbation Δ of the i-th row such that $\hat{\mu}_i \leq (1+\beta)\mu_i$. For solving Δ, we assume for ease of presentation that the mass that is shifted to i is taken uniformly from the other nodes so that*

$$\hat{\mu}_i \leq (1+\beta)\mu_i \quad \Leftrightarrow \quad \mu_i + \alpha(1-\mu_i) \leq (1+\beta)\mu_i \quad \Leftrightarrow \quad \alpha \leq \frac{\beta\mu_i}{1-\mu_i}.$$

Choosing $\alpha = \frac{\beta\mu_i}{1-\mu_i}$, we get via (7.2) the following maximum allowable perturbation

$$\Delta = \frac{\beta}{1+\beta}\frac{1}{1-\mu_i}e_i e_i^\top (I-G).$$

7.2. Maximal Weight Shift Between Two Elements of μ. For a given fixed i, let us consider direction $d = e_i - e_j$ for some $j \neq i$, i.e., d has only two non-zero elements, d_i and d_j. We then have for the elements in $\hat{\mu} := \hat{\mu}_\alpha = \mu + \alpha d$:

$$\hat{\mu}_i := (\hat{\mu}_\alpha)_i = \mu_i + \alpha \quad \text{and} \quad \hat{\mu}_j := (\hat{\mu}_\alpha)_j = \mu_j - \alpha, ,$$

and hence $z^\top = \alpha(e_i - e_j)^\top (I-G)$ yielding

$$G + \Delta = G + \frac{\alpha}{\mu_i + \alpha}(e_i e_i^\top - e_i e_j^\top)(I-G). \tag{7.3}$$

In order to check the non-negativity of this matrix, we only have to verify that the elements in row i are non-negative, which implies

$$\frac{1}{\mu_i+\alpha}\Big(\mu_i G_{i,i} + \alpha(1+G_{j,i})\Big) \geq 0, \quad \frac{1}{\mu_i+\alpha}\Big(\mu_i G_{i,j} - \alpha(1-G_{j,j})\Big) \geq 0,$$

and

$$\frac{1}{\mu_i+\alpha}\Big(\mu_i G_{i,k} + \alpha G_{j,k}\Big) \geq 0 \quad \forall k \neq i, j.$$

The first and last of these inequalities hold for every $\alpha \geq 0$, but the second inequality holds only for $\alpha \leq \mu_i G_{i,j}/(1-G_{j,j})$. In order to maximize the increase of $\mu_i + \alpha$, the best choice for the index j is therefore to choose a maximal solution of

$$\alpha^* = \max_{j \neq i} \mu_i G_{i,j}/(1-G_{j,j}). \tag{7.4}$$

That will increase μ_i to $\hat{\mu}_i = \mu_i + \alpha^*$ and decrease μ_j to $\hat{\mu}_j = \mu_j - \alpha^*$, while all the other entries of the vector μ remain unchanged. For examples illustrating this we refer to Appendix E.

8. Heuristics for General-Rank Perturbations Preserving Stochasticity.

In this section, we develop heuristics for the TSDP, i.e.,

$$\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|,$$

by making use of the developed theory from the previous sections. In words, for a given stochastic matrix G, we are looking for a minimum-norm perturbation Δ of general rank such that $G + \Delta$ is stochastic and has stationary distribution $\hat{\mu}$. The heuristics can find approximate solutions to $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$ in cases when existing (commercial) convex problem solvers fail to find a solution in a reasonable time.

The developed theory so far concerns rank-1 perturbations. Example 5 illustrates that no feasible rank-1 perturbation may exist for $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$ even if the target $\hat{\mu}$ is arbitrarily close to the original μ (i.e., $\mathbf{\Delta}(G, \hat{\mu}) = \emptyset$).

EXAMPLE 5. *Consider a ring network from Example 3 of size $n \geq 3$. Suppose, with $a \in (0, 1/n]$, we aim for*

$$\hat{\mu}_i = \begin{cases} 1/n + a & i = 1 \\ 1/n - a & i = 3 \\ 1/n & i \notin \{1, 3\} \end{cases} .$$

For this $\hat{\mu}$, it holds that $z^\top = (2ab, 0, -2ab, ab, -ab, 0, \ldots, 0)$, and because there is no row in G for which $G(i, 3)$ and $G(i, n)$ are both > 0 it follows that $u = 0$. This means that there is no rank-1 perturbation for all $a \in (0, 1/n]$.

While there may be no rank-1 perturbation that allows to reach $\hat{\mu}$ from μ, the accumulation of a sequence of rank-1 perturbations can lead to perturbations of general rank and thus to candidates for $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$. This key idea will be used in scalable heuristics that can find approximate solutions to $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$.

In the remainder of this section, Section 8.1 presents mathematical programming problem formulations that can be solved using (commercial) solvers. In Section 8.2 we then use the rank-1 perturbation theory developed in this paper to develop heuristics for $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|$. Numerical experiments of both approaches are presented in Section 8.3.

8.1. Mathematical Programming Problems. We reformulate our TSDP so that existing algorithms from literature can be used and which are implemented in (commercial) solvers. In the cases of the 1-norm, the v-norm, and the ∞-norm, we can cast the TSDP into a linear programming (LP) problem, and for the 2-norm, we can reformulate it as a linear matrix inequality (LMI) problem. Let $Z := |\Delta|$, then the TSDP can be written as the following LP problem in the variables Δ, Z and γ, for the 1-, ∞- and $v-$norms, respectively:

$$(8.1) \quad \begin{aligned} \min_{\Delta, Z, \gamma} \quad & \gamma \\ \text{s.t.} \quad & \hat{\mu}^\top (G + \Delta) = \hat{\mu}^\top \\ & G + \Delta \geq 0 \\ & Z \geq \Delta \\ & Z \geq -\Delta \\ & \begin{cases} \text{if } \gamma \geq \sum_i Z_{i,j}, \ \forall j & \text{for the 1-norm} \\ \text{if } \gamma \geq \sum_j Z_{i,j}, \ \forall i & \text{for the } \infty\text{-norm} \\ \text{if } v_i \gamma \geq \sum_j Z_{i,j} v_j, \ \forall i & \text{for the } v\text{-norm,} \end{cases} \end{aligned}$$

and it can be written as the following LMI problem for the 2-norm:

$$(8.2) \quad \begin{aligned} \min_{\Delta, \gamma} \quad & \gamma \\ \text{s.t.} \quad & \hat{\mu}^\top (G + \Delta) = \hat{\mu}^\top \\ & G + \Delta \geq 0 \\ & \begin{bmatrix} \gamma I & \Delta^\top \\ \Delta & \gamma I \end{bmatrix} \succeq 0 \, . \end{aligned}$$

Remark 8.1. The LP and LMI problem formulations can also be used to find rank-1 solutions by adding the constraint $\Delta = xz^\top$, where $x \in \mathbb{R}^n$ are extra decision variables. This particular rank-1 structure follows from Theorem 5.1.

Problems (8.1) and (8.2) always have a non-empty feasible set which can be characterized using [14] as $\Delta = \hat{G} - G$ for all $\hat{G}$ in the convex polytope of stochastic matrices with stationary distribution $\hat{\mu}$. Note that the feasible set includes $\Delta = \mathbf{1}\hat{\mu}^\top - G$. Finding a general-rank solution is much more complex than finding a rank-1 solution since the number of decision variables is quadratic in n, rather than linear. Therefore, these programs have a worst-case time complexity of $\mathcal{O}(n^6)$. As a result, (commercial) solvers are not recommended for large-scale problems.

8.2. Rank-1 Steps Heuristics. In this section, so-called rank-1 steps heuristics are developed that compute approximate solutions of the TSDP. Starting from μ, the idea is to iteratively reach intermediate stationary distributions that are getting "closer and closer" to the target $\hat{\mu}$ as illustrated in Figure 1. The i-th intermediate stationary distribution after i perturbations/steps is denoted by $\mu^{(i)}$. The $\mu^{(i)}$'s need to be determined upfront or dynamically along the way. To make the heuristic computationally efficient, $\mu^{(i)}$ should be reachable from $\mu^{(i-1)}$ via a rank-1 perturbation that preserves stochasticity as analyzed in previous sections. The heuristic later on prescribes how possible $\mu^{(i)}$'s can be determined (for example, by fixing its elements to that of $\hat{\mu}$, respectively, giving $\mu^{(n-1)} = \hat{\mu}$). Although numerical experiments show that it often works, there is no guarantee that the heuristic leads to a sequence of $\mu^{(i)}$'s leading to $\hat{\mu}$. In case it cannot, one can fall back to the candidate $\Delta = \mathbf{1}\hat{\mu}^\top - G$.

To further formalize the rank-1 steps heuristics, let us introduce some notation. Define the i-th perturbation, or step, by $\Delta^{(i)}$. Then the accumulated perturbation after $i-1$ steps is given by $\widetilde{\Delta}^{(i)} := \sum_{j=1}^{i-1} \Delta^{(j)}$ (for which $\widetilde{\Delta}^{(1)} = 0$). At each step i, $\mu^{(i)}$ is chosen such that $\mathbf{\Delta}(G + \widetilde{\Delta}^{(i)}, \mu^{(i)}) \neq \emptyset$. Then, the perturbation from this set with smallest norm is chosen, i.e., $\Delta^{(i)} = \arg\min_{\Delta \in \mathbf{\Delta}(G+\widetilde{\Delta}^{(i)}, \mu^{(i)})} \|\Delta\|$. Consequently, $G + \widetilde{\Delta}^{(i+1)}$ is a stochastic matrix with stationary distribution $\mu^{(i)}$. In case no appropriate $\mu^{(i)}$ can be found in reasonable time, $\Delta = \mathbf{1}\hat{\mu}^\top - G$ can be returned. While there are uncountably many accumulated sequences of rank-1 perturbations leading to $\hat{\mu}$, finding one is challenging. Eventually, we hope to reach $\hat{\mu}$ at, say, the $(n-1)$-th step, which gives us the approximate solution $\widetilde{\Delta}^{(n)}$ for $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$; this is illustrated in Figure 1.

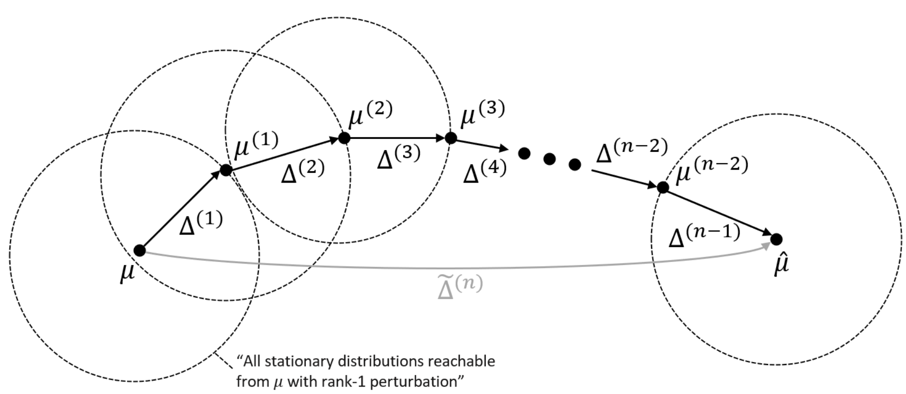

Fig. 1: Illustration of a rank-1 steps heuristic that takes $n-1$ rank-1 perturbations (or steps) towards $\hat{\mu}$ to approximately solve $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$.

There are two problems with this approach that need to be addressed. The first problem is that it is unknown what a good feasible sequence of $\mu^{(i)}$'s are. The second problem is that at every step the rank-1 perturbation does not take previous perturbations into account which unnecessarily increases the norm of the accumulated perturbation. Instead, ensuring that each perturbation "connects efficiently" to the previous accumulated perturbation can significantly reduce the norm of the final accumulated perturbation.

To ensure that a new perturbation connects efficiently to previous perturbations, we will straightforwardly extend Theorem 6.1 so that it also considers previous perturbations. In particular, in step i, our goal is to minimally perturb $G + \widetilde{\Delta}^{(i)}$, while considering previous perturbations, so that $\mu^{(i)}$ is reached:

$$\min_{\Delta \in \mathbf{\Delta}(G+\widetilde{\Delta}^{(i)}, \mu^{(i)})} \|\widetilde{\Delta}^{(i)} + \Delta\|. \tag{8.3}$$

Problem (8.3) is a generalization of problem $\min_{\Delta \in \mathbf{\Delta}(G,\mu^{(i)})} \|\Delta\|$ that was formalized as convex optimization problems in Theorem 6.2 and for which the solutions are explicitly constructed in Section D. Problem (8.3) can be solved similarly to obtain its solution. In particular, Problem (8.3) can be rewritten, using Theorem 6.1, as

$$\min_{x} \|\widetilde{\Delta}^{(i)} + xz^\top\| \quad \text{s.t.} \quad l \le x \le u, \quad \mu^{(i)\top} x = 1, \tag{8.4}$$

where we did not decorate z, l and u with superscript (i) for simplicity but do note that they depend on $G + \widetilde{\Delta}^{(i)}$ and $\mu^{(i)}$ (instead of G and $\hat{\mu}$, respectively). Indeed, it now can happen that it is beneficial to take $x < 0$ to reduce the objective value, it even may happen that $\|\widetilde{\Delta}^{(i)} + xz^\top\| < \|\widetilde{\Delta}^{(i)}\|$, i.e., the previous objective value can be reduced by a new rank-1 perturbation. Substituting x by $x - l$ in (8.4) gives

$$\min_{x} \|(\widetilde{\Delta}^{(i)} + lz^\top) + xz^\top\| \quad \text{s.t.} \quad 0 \le x \le u - l, \quad \mu^{(i)\top} x = 1 - \mu^{(i)\top} l.$$

In comparison to Theorem 6.1, there are three changes in this minimization problem: (i) the upperbound of x has changed, (ii) $\mu^{(i)\top} x$ should equal $1 - \mu^{(i)\top} l$ instead of 1, and (iii) the objective now contains $\widetilde{\Delta}^{(i)} + lz^\top (\neq 0)$. The first two differences are not fundamental and the same algorithmic procedures from Section D apply. The third difference demands a change in the algorithm: When we now start with $x = 0$ and start increasing x to ensure $\mu^{(i)\top} x = 1 - \mu^{(i)\top} l$, we have to take into account that different x_i-increases have different effects on the objective due to $\widetilde{\Delta}^{(i)} + lz^\top$. As a result, the algorithm should first focus on decreasing the objective as much as possible, then increase x as much as possible without affecting the objective, and then lastly, increase x proportionally to their effect on the objective until $\mu^{(i)\top} x = 1 - \mu^{(i)\top} l$ is reached. While doing this, one has to take the upperbounds $u - l$ of x into account and check throughout whether $\mu^{(i)\top} x = 1 - \mu^{(i)\top} l$ is met. Once this restriction is met, one can return the solution $(l+x)z^\top$ (reversing the substitution) for $\min_{\Delta \in \mathbf{\Delta}(G+\widetilde{\Delta}^{(i)}, \mu^{(i)})} \|\widetilde{\Delta}^{(i)} + \Delta\|$. For notational convenience, we denote the solution of this procedure by

$$P(G, \mu^{(i)}, \widetilde{\Delta}^{(i)}) := \operatorname*{arg\,min}_{\Delta \in \mathbf{\Delta}(G+\widetilde{\Delta}^{(i)}, \mu^{(i)})} \|\widetilde{\Delta}^{(i)} + \Delta\|.$$

We implemented this procedure using a binary search with a tolerance of ξ and therefore it has a time complexity of $\mathcal{O}(\log_2(\|\widetilde{\Delta}^{(i)} + uz^\top\|/\xi)n^2)$.

The success of the rank-1 step heuristic depends on the chosen sequence of $\mu^{(i)}$'s. A straightforward sequence, that will also be used below, is to iteratively set the elements of μ to the corresponding elements in $\hat{\mu}$ (and keep those elements fixed in consecutive steps). Specifically, $\mu^{(i)}$ has i elements fixed to those of $\hat{\mu}$, of which $i-1$ elements were already fixed to $\hat{\mu}$ in $\mu^{(i-1)}$, and the remaining elements of $\mu^{(i)}$ divide the remaining mass proportionally to μ. The elements can be set in a random order, but experiments show that it is better to consider the elements of $|\hat{\mu}-\mu|$ in decreasing order (the "preparation" in the rank-1 steps heuristic below). The reasoning is that at the beginning you have the most flexibility to overcome the largest differences. After $n-1$ steps, $\hat{\mu}$ will possibly be reached, but there is no guarantee that this sequence indeed reaches $\hat{\mu}$, i.e., it is not guaranteed that there is always a rank-1 perturbation from $\mu^{(i-1)}$ to $\mu^{(i)}$. However, numerical experiments showed that it often finds a "path" to $\hat{\mu}$, and if it does not, one can fall back to candidate $\mathbf{1}\hat{\mu}^\top - G$. In the following, we elaborated the rank-1 steps heuristic for this $\mu^{(i)}$ sequence, named R1SH, that converges after $n-1$ steps if the sequence is feasible.

Rank-1 steps heuristic (R1SH): (approximately solving $\min_{\Delta\in\mathbf{\Delta}^{\hat{\mu}}\cap\mathbf{\Delta}^{\geq 0}} \|\Delta\|$)
Given G, μ and $\hat{\mu}$ perform:

1. **Preparation**: Relabel the indices of G, μ and $\hat{\mu}$ so that

$$|\hat{\mu}_1-\mu_1| \geq |\hat{\mu}_2-\mu_2| \geq \cdots \geq |\hat{\mu}_n-\mu_n|,$$

   i.e., $|\hat{\mu}-\mu|$ is sorted from large to small without loss of generality.
2. **Initialization**: Set $\mu^{(0)}=\mu$.
3. **For** $i\in\{1,2,\ldots,n-1\}$, respectively, **do**:
   (a) Determine $\mu_j^{(i)}$ for $j=1,2,\ldots,n$ as follows:

$$\mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } 1\leq j\leq i \\ \frac{\mu_j}{\sum_{k=i+1}^n \mu_k}\left(1-\sum_{k=1}^i \hat{\mu}_k\right), & \text{if } i+1\leq j\leq n \end{cases}$$

   i.e., we fix $\mu_j^{(i)}$ to $\hat{\mu}_j$, for $j=1,2,\ldots,i$, and the remaining mass of $1-\sum_{k=1}^i \hat{\mu}_k$ is distributed over $\mu_j^{(i)}$, for $j=i+1,\ldots,n$, in proportion to the corresponding values in μ.
   (b) Calculate u (see Theorem 6.1) for stochastic matrix $G+\widetilde{\Delta}^{(i)}$ and new stationary vector $\mu^{(i)}$.
   (c) **If** $u^\top\mu^{(i)}\geq 1$:
   Calculate $\Delta^{(i)}=P(G,\mu^{(i)},\widetilde{\Delta}^{(i)})$.
   Else:
   Return $\Delta=\mathbf{1}\hat{\mu}^\top-G$ (intended sequence is infeasible).
4. **Return** $\Delta=\widetilde{\Delta}^{(n)}$ as approximate solution to $\min_{\Delta\in\mathbf{\Delta}^{\hat{\mu}}\cap\mathbf{\Delta}^{\geq 0}} \|\Delta\|$.

The time complexity of R1SH is $\mathcal{O}(\log_2(\mathcal{C}/\xi)n^3)$, where constant $\mathcal{C} := \max_i \|\widetilde{\Delta}^{(i)} + u^{(i)}z^{(i)\top}\|$. R1SH can be generalised by fixing more elements at once in each step. This procedure, indicated as R1SH(K) and introduced in the following, allows for a trade-off between computation time and quality of the approximate solution .

R1SH(K): To reduce the complexity of R1SH at the expense of the quality of the approximate solution, subsets of elements can be fixed at each step, rather than one at a time. More specifically, after the first preparation step in R1SH, we partition the set of indices $1,\ldots,n-1$ into K (almost) equally sized subsets $P_1,P_2,\ldots,P_K$. Then the for-loop of step 3 in R1SH loops over $i\in\{1,\ldots,K\}$, and step 3a in R1SH becomes

$$\mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } j \in \cup_{k=1}^{i} P_k \\ \frac{\mu_j}{\sum_{k=i+1}^{n} \mu_k} \left(1 - \sum_{k=1}^{i} \hat{\mu}_k\right), & \text{if } j \notin \cup_{k=1}^{i} P_k \end{cases}.$$

This version of R1SH is denoted as R1SH(K). Note that R1SH = R1SH($n-1$). Its time complexity is $\mathcal{O}(\log_2(\mathcal{C}/\xi)n^2K)$.

As an alternative to the $\mu^{(i)}$ sequence in R1SH, one can use more than $n-1$ steps to reach $\hat{\mu}$. This allows one to do finer steps. Also, intended steps that are infeasible can be skipped and retried later. This is exploited in the so-called finer-R1SH that is introduced in the following.

FR1SH(ϕ) (finer-R1SH)**:** To increase the quality of the approximate solution at the expense of the computing time, one can choose smaller $\mu^{(i)}$ increments and repeat the for-loop more than $n-1$ times till the $\mu^{(i)}$'s converge. In particular, one can set $\mu_j^{(i)}$ in step 3a of R1SH for $j = 1, 2, \ldots, n$ as follows (where mod represents a modulo operation):

$$\mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } j = i \bmod n \\ \frac{\mu_j^{(i-1)}}{\sum_{k \neq i} \mu_k} (1 - \hat{\mu}_j), & \text{if } j \neq i \bmod n \end{cases},$$

i.e., in the i-th for-loop force $\mu_{i \bmod n}^{(i)} = \hat{\mu}_{i \bmod n}$ and divide the remaining mass of $1 - \hat{\mu}_{i \bmod n}$ proportionally over the other elements of $\mu^{(i)}$. One can repeat the for-loops until $\|\mu^{(i)} - \mu^{(i-1)}\| < \phi$, where $\phi > 0$ is a given precision. Then, we hope to reach $\hat{\mu}$ from $\mu^{(i)}$ with a single rank-1 perturbation. Note that in contrast to R1SH only $\mu_{i \bmod n}^{(i)}$ is fixed to $\hat{\mu}_{i \bmod n}$ in the i-th for-loop of FR1SH(ϕ). The time complexity of FR1SH(ϕ) is $\mathcal{O}(\log_2(\mathcal{C}/\xi)n^3/\phi)$.

There are different ways to increase the chance of finding better approximate solutions with R1SH, R1SH(K) or FR1SH(ϕ) at the expense of larger computing time. For example, at each iteration in the for-loop one can try to jump directly from $\mu^{(i)}$ to $\hat{\mu}$ via a rank-1 perturbation. While doing so, one can keep track of the best candidate solution of $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$ and return the best candidate at the end. A less computationally intensive way, that we will always use when applying R1SH and R1SH(K) later on, is to compare the final candidate solution with $\arg\min_{\Delta \in \mathbf{\Delta}(G,\hat{\mu})} \|\Delta\|$ (if it exists) and return the best. In Section B.2, we provide a numerical example on applying R1SH, R1SH(K) and FR1SH(ϕ) to the queuing example. Also the Riesz projector $\mathbf{1}\hat{\mu}^\top - G$ (referred to as "Riesz" in short) is applied to that example for comparison.

8.3. Numerical Experiments. We present in this section experiments for larger numerical instances. In particular, the tests in Section 8.3.1 make use of randomly generated dense matrices, whereas Section 8.3.2 performs tests on real-life sparse matrices. Throughout this section, the ∞-norm is considered.

8.3.1. Dense Random Matrices. To test R1SH and its variants for larger examples, we generated random problem instances of $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$. A random G is generated by drawing a $n \times n$ matrix with random values in $(0, 1)$ and scaling the rows such that row sums are all one. Similarly, to generate $\hat{\mu}$, a random $n \times 1$ vector v of random values in $(0, 1)$ is first generated and scaled so that it sums up to one, and then a 0.1-fraction of this random vector v is then mixed with μ (of G)

to generate the random $\hat{\mu} = 0.1v + 0.9\mu$ vector. The pair $(G, \hat{\mu})$ is then a problem instance of $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$.

We sample 25 random problem instances of $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|$, for different sizes n. Each problem instance is solved with the methods from the previous sections. In particular, $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$ is found with Theorem 5.1, $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|_\infty$ is found by solving the LP from Section 8.1 using Gurobi 9.1.2, and $\min_{\Delta \in \mathbf{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$ is found by applying the algorithm from Section D.3. The results for $n = 100$, $n = 500$ and $n = 1000$ can be found in Table 4 in Section F.

The following can be observed from the numerical experiments:

- R1SH and the LP solution method do not scale well. For $n = 100$, the LP solution method is faster than R1SH and finds the solution instead of an approximate solution. But for larger instances with $n = 500$, the LP solution method takes significantly more time than R1SH, which is in line with the complexity analysis. As a result, within 10 minutes R1SH could solve 96% of the $n = 500$ instances, whereas the LP solution method could solve only 56% of the instances with $n = 500$ nodes.
- The approximate solution quality found by R1SH(K) increases with K, just as the computation time (which increases linearly in K). In particular, if $\hat{\mu}$ is not too far away from μ, good approximate solutions are found by R1SH(K) for relatively small K. Also for the $n = 1000$ instances, R1SH(16) finds near optimal approximate solutions, as can be seen from a comparison with the lower bound $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$ for $\min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|_\infty$.

8.3.2. Sparse Matrices. The applicability of our heuristics for sparse G is hindered because $\mathbf{\Delta}(G, \hat{\mu})$ is empty for many $\hat{\mu}$. As a result, there is limited flexibility in jumping to intermediate stationary distributions. More specifically, in a rank-1 perturbation, the same vector (such as $z^\top = \hat{\mu}(I - G)$) is used to modify every row. Since the perturbation "transfers" mass within a row, some elements will be positive and some negative. In a sparse setting, there are many zeros, meaning that a single vector can often only be used for perturbing a single row (or a few rows at most). Being able to perturb only one row in a stochastic matrix, it is not hard to imagine that the number of reachable stochastic matrices is limited. In other words, finding a rank-1 perturbation towards a *specific* stationary distribution (the main focus of this paper) is often infeasible. Example 5 demonstrates this for the (sparse) ring network.

Nevertheless, the rank-1 steps heuristics do apply to specific cases of sparse matrix instances where $\hat{\mu}$ changes most significantly for a subset of nodes that constitutes a dense subgraph. Intuitively, rank-1 perturbations will have more flexibility to adjust connections between nodes from a dense subgraph. To create test instances for sparse matrices, we will find large cliques in the undirected graph constituted by $G + G^\top$ (ignoring self-loops) and will increase or decrease the share of the cliques in μ to obtain $\hat{\mu}$.

To illustrate the applicability and verify the quality of the approximate solution for sparse matrices, we consider Barabási–Albert preferential attachment social networks. In particular, for our experiments, a graph of $n = 100$ nodes is grown by attaching new nodes each with 5 edges that are preferentially attached to existing nodes with high degrees. When applying the rank-1 steps heuristics in this sparse matrix setting, but also later on in other sparse matrix experiments, we look after each step whether we can reach $\hat{\mu}$ with a rank-1 perturbation and we keep track of the Δ with smallest norm. Furthermore, we consider FR1SH in the current and following sparse matrix experiments, as this increases the change of finding (better) candidates and allows the comparison of the two approaches.

In the first experiment, $\hat{\mu}$ is based on making the largest clique as uniform as possible while keeping their total mass fixed. A practical meaning of this objective is to make the "network leaders" more cooperative. The average results of 25 random social networks can be found in Table 5a. It indeed shows that candidates can be reached with R1SH and FR1SH. More specifically, in contrast to R1SH, FR1SH is able to find candidates for all instances. The quality is significantly better than the Riesz projector, but relatively far away from the solution found by solving the LP.

In the second experiment, $\hat{\mu}$ is determined by reducing the total mass of the largest clique by 10% while keeping the relative weights inside the clique, as well as outside, respectively, equal. The results in Table 5b show that FR1SH is again able to solve all instances while obtaining results close to the optimum.

To further explore the applicability of rank-1 steps heuristics in sparse networks, we will consider the following three real-life networks from three different domains with different objectives regarding $\hat{\mu}$ (see also the applications overview in Section 2):

Social network: A high-school network of student relationships where we aim to increase the popularity of a clique of students by 10%. This could potentially enhance the group's cohesion.

Road network: Road network between the largest cities in Europe where we aim to decrease the traffic congestion of a chosen clique by 10% (assuming the traffic flows uniformly through the network as described by a random walk).

Organizational network: An email-conversation network of university employees where we aim to decrease the organizational importance of a chosen clique by 10%. This could potentially lower the hierarchical nature of an organization.

The weighted adjacency matrices of all networks are normalized so that they are stochastic and we only considered the largest strongly connected component (so that the stationary distribution exists). More details about the considered datasets can be found in Table 2. For each real-life network, we search for the 25 largest cliques (its computation time turns out to be negligible in our examples, probably due to sparsity), and for each clique we apply the rank-1 steps heuristics. We did not solve the LP with Gurobi for these networks because of the scalability issues of that approach (after computing for a relatively long time, it still did not find a candidate solution). Table 6 and Table 7 (in the appendix) present the average results over the 25 cliques for the different real-life networks with a time limit of 60 and 600 seconds, respectively. Note that the time limit does not necessarily have to be reached because the rank-1 steps heuristics are terminated once a full loop over the nodes did not lead to an improvement.

All instances could be solved using (F)R1SH within 60 seconds. To get an indication of the quality of the candidate solutions found, we can again compare it with lower bound $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$ for $\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|_\infty$. From this it follows that, especially for the road and organizational networks, (F)R1SH is able to find candidate solutions that lie relatively close to the lower bound as compared to the Riesz projector. Furthermore, a comparison between Table 6 and Table 7 shows that the performance of candidates found by (F)R1SH with a time limit of 60 seconds are often close to the candidates found with a time limit of 600 seconds. Only in the organizational network, the average of the norms decreased from 0.0264 to 0.0263 on average. Moreover, it follows from Table 7 (in the appendix) that on average the time limit of 600 seconds is often not reached and the performance of FR1SH is only slightly better on average than R1SH.

To conclude, the numerical experiments demonstrate that the rank-1 steps heuris-

Name	Description	Goal $\hat{\mu}$	Ref.
Social network	A directed network based on a survey from 1994/1995 on a high school. Each student was asked to list his/her 5 best female friends and 5 best male friends. A node represents a student and an edge (i, j) between two students shows that student i chose student j as a friend. Higher edge weights indicate more interactions. The network consists of 2155 nodes and 11467 edges (0.25% of all possible connections).	Increase popularity of a chosen clique (when assuming that edges are undirected) with 10%.	[23, 22, 25]
Road network	This is the international E-road network that lies mostly in Europe. The network is undirected where nodes represent cities and an edge between two nodes means that they are connected by an E-road. The network consists of 1039 nodes and 2834 edges (0.24% of all possible connections).	Decrease traffic intensity of a chosen clique with 10%.	[23, 39, 25]
Org. network	Email communication network at the University Rovira i Virgili in Spain. Nodes are employees and each undirected edge represents that at least one email was sent between the employees. The network consists of 1133 nodes and 10902 edges (0.85% of all possible connections).	Decrease the organizational importance of a chosen clique with 10%.	[24, 15, 25]

Table 2: Overview of the real-life (sparse) networks used to test rank-1 steps heuristics.

tics provide a scalable alternative for solving the LP that leads to significantly better candidate solutions than the Riesz projector. It particularly works well for dense random matrices and specific sparse matrix instances in case $\hat{\mu}$ is not too far away from μ.

9. Conclusion and Further Research. In this paper we established an inverse theory of perturbation analysis of Markov chains to solve the Target Stationary Distribution Problem (TSDP). The key ingredient of our approach was to work with rank-1 perturbations only, and we established closed-form solutions for rank-1 perturbations achieving a given target stationary distribution. To overcome the limitation to rank-1 perturbations, we developed rank-1 steps heuristics for finding a sequence of rank-1 perturbations/steps so that the accumulated perturbation is of higher rank and does allow to reach the target stationary distribution. Different applications are discussed and numerical experiments show the efficiency of our approach for artificial dense random instances and for specific sparse matrices issued from real-life data.

There are still open questions regarding the rank-1 steps heuristics for solving the TSDP. In particular, one can look for other $\mu^{(i)}$ sequences that improve the performance of our iterative procedure. Also, a rigorous convergence analysis would valuable, as well as performance guarantees and approximation error estimates for the approximate solutions. Also it remains open whether the structural knowledge about the feasible set from [14, 6] can be exploited for other scalable (approximate) solution methods for the TSDP.

Acknowledgement. The authors want to express their gratitude to the anonymous reviewers for their valuable and constructive suggestions.

REFERENCES

[1] K. Abbas, J. Berkhout and B. Heidergott (2016) A critical account of perturbation analysis of Markov chains. *Markov Processes and Related Fields* **22** 227–266.
[2] P. Alquier, N. Friel, R. Everitt and A. Boland (2016) Noisy Monte Carlo: convergence of Markov chains with approximate transition kernels. *Statistics and Computing* **26** 29–47.
[3] C. Andrieu and G. Roberts (2009) The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistic* **37** 697–725.
[4] V. Anisimov (1988) Estimates for the deviations of the transition characteristics of nonhomogeneous Markov processes. *Ukrainian Math. J.* **40** 588–592.
[5] L. Bortolussi and R. Hayden (2013) Bounds on the deviation of discrete-time Markov chains from their mean-field model. *Performance Evaluation* **70** 736–749.
[6] J. Breen and S. Kirkland (2015) Stationary vectors of stochastic matrices subject to combinatorial constraints. *Electronic Journal of Linear Algebra* **28**: 25-53.
[7] P. Carrington, J. Scott and S. Wasserman (2005) *Models and methods in social network analysis.* Cambridge University Press.
[8] H. Caswell (2013) Sensitivity analysis of discrete Markov chains via matrix calculus. *Linear Algebra and its Applications* **438** 1727–1745.
[9] M. Chekroun, J. Neelin, D. Kondrashov, McWilliams and M. Ghil (2014) Rough parameter dependence in climate models and the role of Ruelle-Pollicott resonances. *Proceedings of the National Academy of Sciences of the United States of America* **111** 1684–1690.
[10] G. Cho and C. Meyer (2001) Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications* **335** 137–150.
[11] E. Crisostomi, S. Kirkland and R. Stephen (2011) A Google-like model of road network dynamics and its application to regulation and control. *International Journal of Control* **84** 633–651.
[12] B. Golub and M. Jackson (2010) Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* **2** 112–49.
[13] A. Greenbaum, R.C. Li and M. Overton (2020) First-order perturbation theory for eigenvalues and eigenvectors. *SIAM Review* **62(2)** 463–482.
[14] D. Gregory, S. Kirkland and N. Pullman (1992) Row-stochastic matrices with a common left fixed vector. *Linear Algebra and its Applications* **169** 131–149.
[15] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas (2003) Self-similar community structure in a network of human interactions. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics* **68(6)** 065103.
[16] B. Heidergott, A. Hordijk and M. Van Uitert (2007) Series expansions for finite-state Markov chains. *Probability in the Engineering and Informational Sciences* **21(3)** 381–400.
[17] B. Heidergott and A. Hordijk (2003) Taylor series expansions for stationary Markov chains. *Advances in Applied Probability* **35(4)** 1046–1070.
[18] I. Ipsen and T. Selee (2011) Ergodicity coefficients defined by vector norms. *SIAM Journal on Matrix Analysis and Applications* **32** 153–200.
[19] M. Jackson (2010) *Social and economic networks.* Princeton University Press.
[20] C. de Kerchove, L. Ninove and P. Van Dooren (2008) Maximizing PageRank via outlinks. *Linear Algebra and its Applications* **429** 1254–1276.
[21] S. Kirkland (2002) On a question concerning condition numbers for Markov chain. *SIAM Journal on Matrix Analysis and Applications* **23** 1109–1119.
[22] KONECT (2016) Adolescent health network dataset – KONECT. http://konect.uni-koblenz.de/networks/moreno_health.
[23] KONECT (2017) Euroroads network dataset – KONECT. http://konect.cc/networks/subelj_euroroad.
[24] KONECT (2017) U. Rovira i Virgili network dataset – KONECT. http://konect.uni-koblenz.de/networks/moreno_health.
[25] J. Kunegis (2013) KONECT – The Koblenz Network Collection. *WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web*, 1343–1350.
[26] W. Li, L. Jiang, W.-K. Ching and L.-B. Cui (2013) On perturbation bounds for the joint stationary distribution of multivariate Markov chain models. *East Asian Journal on Applied Mathematics* **3** 1–17.
[27] C.D. Meyer (1980) The condition of a finite Markov chain and perturbation bounds for the limiting probabilities. *SIAM Journal of Algebraic Discrete Methods* **1** 273–283.
[28] S. Meyn and R. Tweedie (2012) *Markov chains and stochastic stability.* Springer Science & Business Media.
[29] A. Mitrophanov, A. Lomsadze and M. Borodovsky (2005) Sensitivity of hidden Markov models. *Journal of Applied Probability* **42** 632–642.
[30] A. Mitrophanov (2005) Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability* **42** 1003–1014.

[31] J. Moody (2001) Peer influence groups: Identifying dense clusters in large networks. *Social Networks*, **23(4)** 261–283.

[32] Z. Mouhoubi and D. Aïssani (2010) New perturbation bounds for denumerable Markov chains. *Linear Algebra and its Applications* **432** 1627–1649.

[33] M. Neumann and J. Xu (2003) Improved bounds for a condition number of Markov chains. *Linear Algebra and its Applications* **386** 225–241.

[34] R. Pal, A. Datta and E.R. Dougherty (2008) Robust intervention in probabilistic Boolean networks. *IEEE Transactions on Signal Processing* **56** 1280–1294.

[35] P. Schweitzer (1968) Perturbation theory and finite Markov chains. *Journal of Applied Probability* **5** 410–413.

[36] J. Scott(1988) Social network analysis. *Sociology* **22** 109–127.

[37] E. Seneta (2001) Sensitivity analysis, ergodicity coefficients, and rank-one updates of finite Markov chains. In: Stewart, W. J. (ed.) *Numerical Solutions of Markov Chains*, Marcel Dekker, New York.

[38] G. Stewart (1973) Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review* **15(4)** 727–764.

[39] L. Šubelj and M. Bajec (2011) Robust network community detection using balanced propagation. *The European Physical Journal B*, **81(3)** 353–362.

[40] O. Szehr (2015) Eigenvalue estimates for the resolvent of a non-normal matrix. *Journal of Spectral Theory* **4** 783–813.

[41] O. Szehr and M. Wolf (2013) Perturbation bounds for quantum Markov processes and their fixed points. *Journal of Mathematical Physics* **54** 032203.

[42] S. Tuteja (2013) Enhancement in weighted PageRank algorithm using VOL. *Journal of Computer Engineering*, **14** 135–141.

[43] N. Tyagi and S. Sharma (2012) Weighted Page Rank algorithm based on number of visits of links of web page. *International Journal of Soft Computing and Engineering (IJSCE)* **2(3)** 441–446.

[44] S. Wasserman and K. Faust (1994) *Social network analysis: Methods and applications.* Cambridge University Press.

Appendix A. Proofs.

A.1. Proof of Theorem 5.1. By definition, it holds for all $\Delta \in \mathbf{\Delta}$ that (i) $\hat{\mu}^\top(G+\Delta) = \hat{\mu}^\top \Leftrightarrow \hat{\mu}^\top\Delta = \hat{\mu}^\top(I-G)$, and (ii) $\Delta = xy^\top$ for some vectors x and y. Inserting (ii) into (i) yields

$$\hat{\mu}^\top xy^\top = \hat{\mu}^\top(I-G).$$

Since G is irreducible and $\hat{\mu} \neq \mu$ by assumption, the right-hand side is nonzero and so is the scalar $c := \hat{\mu}^\top x$. This implies that

$$y^\top = \frac{1}{c}\hat{\mu}^\top(I-G) \quad \text{and thus} \quad \Delta = \frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}(I-G).$$

For a rank-1 matrix $\Delta := xy^\top$ the subordinate norm can be computed as follows

$$\|xy^\top\| = \sup_{z\neq 0}\frac{\|(xy^\top)z\|}{\|z\|} = \|x\|\sup_{z\neq 0}\frac{|y^\top z|}{\|z\|} = \|x\|\|y\|_*.$$

If we now minimize the norm of the rank-1 matrix $\Delta = x\hat{\mu}^\top(I-G)/(\hat{\mu}^\top x)$ over x, we obtain

$$\inf_{x\neq 0}\|\Delta\| = \inf_{x\neq 0}\frac{\|x\|}{|\hat{\mu}^\top x|}\|(I-G)^\top\hat{\mu}\|_* = \|(I-G)^\top\hat{\mu}\|_*\left(\sup_{x\neq 0}\frac{|\hat{\mu}^\top x|}{\|x\|}\right)^{-1}$$

$$= \frac{\|(I-G)^\top\hat{\mu}\|_*}{\|\hat{\mu}\|_*}.$$

This is also the minimum norm solution for arbitrary matrices Δ satisfying constraints of Problem (1.1) since $(I-G)^\top\hat{\mu} = \Delta^\top\hat{\mu}$ implies

$$\|(I-G)^\top\hat{\mu}\|_* = \|\Delta^\top\hat{\mu}\|_* \leq \|\Delta^\top\|_*\|\hat{\mu}\|_*.$$

But we also have $\|\Delta^\top\|_* = \|\Delta\|$ since

$$\|\Delta\| = \sup_{x\neq 0}\frac{\|\Delta x\|}{\|x\|} = \sup_{x\neq 0}\sup_{y\neq 0}\frac{|y^\top\Delta x|}{\|y\|_*\|x\|}$$

$$= \sup_{y\neq 0}\sup_{x\neq 0}\frac{|x^\top\Delta^\top y|}{\|x\|\|y\|_*} = \sup_{y\neq 0}\frac{\|\Delta^\top y\|_*}{\|y\|_*} = \|\Delta^\top\|_*,$$

which completes the proof.

A.2. Proof of Theorem 6.1. It follows from Theorem 5.1 that any $\Delta \in \mathbf{\Delta}$ is of the form $\Delta = \frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}(I-G)$ for some x such that $\hat{\mu}^\top x \neq 0$ (note that $\mathbf{\Delta} \subseteq \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\text{rank-1}}$). Since Δ of this form does not depend on the scaling factor of x but only on its direction, we can force the scaling such that $\hat{\mu}^\top x = 1$ and thus Δ simplifies to

$$\Delta = x\hat{\mu}^\top(I-G) = xz^\top$$

with $z := (\mu - \mu)^\top(I-G)$. The condition $G + \Delta \geq 0$ can then be rewritten as

$$G_{i,j} + x_i z_j \geq 0 \quad \forall i,j.$$

Since $z^\top \mathbf{1} = 0$ and $z^\top \neq 0$ the non-negative vectors z_+ and z_- of the decomposition $z = z_+ + z_-$, $\operatorname{supp}(z_+) \cap \operatorname{supp}(z_-) = \emptyset$ both have non-empty support. The above equations then yield the intervals

$$\ell \le x \le u, \text{ where } \ell_i := \max_{j \in \operatorname{supp}(z_+)} \frac{-G_{i,j}}{z_j} \le 0, \quad u_i := \min_{j \in \operatorname{supp}(z_-)} \frac{-G_{i,j}}{z_j} \ge 0.$$

We point out that possibly $\ell_i = u_i = 0$ for some i, but then $x_i = 0$ as well. It follows from the inequalities $x \le u$ and $\hat{\mu} > 0$, that the condition $\hat{\mu}^\top x = 1$ can be achieved if and only if

$$\hat{\mu}^\top u \ge 1, \tag{A.1}$$

and this will guarantee that the matrix $G + \Delta$ is non-negative. If (A.1) is satisfied, then a candidate is given by

$$y = z, \quad x = \frac{u}{\hat{\mu}^\top u} \quad \Longrightarrow \quad \Delta = \frac{u}{\hat{\mu}^\top u} z^\top. \tag{A.2}$$

But also any other x for which $\ell \le x \le u$ and $\hat{\mu}^\top x = 1$, yields a candidate $\Delta = xz^\top$ that satisfies all conditions. Note that it is recommended to avoid negative components in x since they would make the inequality (A.1) harder to reach.

A.3. Proof of Theorem 6.2. It follows from Theorem 6.1 that if $\hat{\mu}^\top u < 1$, the feasible set is empty, and that if $\hat{\mu}^\top u = 1$, the feasible set is a single point $x = u$. If $\hat{\mu}^\top u > 1$, then the optimization problems merely express that one should minimize the norm γ of the vector x over the set of constraints. This is formulated as a convex optimization problem that is feasible, as was pointed out in Theorem 6.1. The problems listed above can be solved using a descent method, and details are provided in Section D.

Appendix B. Queuing Networks.

B.1. General Rank-1 Perturbations. We consider the queuing system in Section 2.4, where we set $s = 2$, $K = 1$ and $\lambda = 1$, $\nu = 1.8$. For $\hat{\mu}$ we choose the uniform distribution over the states $\{0, s+K\}$. Note that this cannot be achieved by a queuing system since its stationary distribution is known to be of a power-law structure. By Theorem 5.1 and Corollary 5.3, the minimal ∞-norm rank-1 perturbation matrix Δ is given by

$$\Delta = \begin{bmatrix} -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \end{bmatrix},$$

with $\|\Delta\|_\infty = 0.2826$, which gives

$$G + \Delta = \begin{bmatrix} 0.7391 & 0.1196 & 0 & 0.1413 \\ 0.3478 & 0.2935 & 0.2174 & 0.1413 \\ -0.0435 & 0.6848 & 0 & 0.3587 \\ -0.0435 & -0.0978 & 0.7826 & 0.3587 \end{bmatrix}.$$

While the left-eigenvector of $G + \Delta$ is indeed the uniform distribution over the state-space, $G + \Delta$ contains negative values and thus fails to be a stochastic matrix.

B.2. Rank-1 Heuristics. Consider again the experiment from Section B.1 Table 3 shows the minimum ∞-norms for different optimization problems and the results for the R1SH, R1SH(2), FR1SH(10^{-3}) and the Riesz projector candidate $\mathbf{1}\hat{\mu}^\top - G$ (in short "Riesz"). For comparison, we also show the minimum norms of the optimization problem with different feasible sets: $P_1 := \min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$ is found with Theorem 5.1, $P_2 := \min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|_\infty$ is found by solving the LP from Section 8.1 using Gurobi 9.1.2 and $P_3 := \min_{\Delta \in \mathbf{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$ is found by applying the algorithm from Subsection D.3. It follows from the results that R1SH is able to solve all instances, also those for which $\mathbf{\Delta}(G, \hat{\mu}) = \emptyset$. This means that it successfully finds a sequence of rank-1 perturbations leading to $\hat{\mu}$. In contrast, R1SH(2) did not find a candidate for $\hat{\nu} = 0.2$. Though better than the Riesz projector, R1SH is not that successful in finding a candidate near $P_2 = \min_{\Delta \in \mathbf{\Delta}^{\hat{\mu}} \cap \mathbf{\Delta}^{\geq 0}} \|\Delta\|_\infty$ for $\hat{\nu} \in \{0.2, 2\}$. For $\hat{\nu} \in \{0.2, 2\}$, FR1SH(10^{-3}) finds candidates that are significantly better than the candidates found by R1SH. Also R1SH(2) finds a better candidate for $\hat{\nu} = 2$ compared to R1SH, which shows that more subsets (R1SH = R1SH(3)) does not necessarily lead to better candidates.

$\hat{\nu}$	P_1	P_2	P_3	R1SH	FR1SH(10^{-3})	R1SH(2)	Riesz
0.2	0.874	0.94	no candidate	1.458	1.028	no candidate	1.759
1.2	0.093	0.119	0.119	0.119	0.119	0.119	1.577
1.4	0.06	0.072	0.072	0.072	0.072	0.072	1.664
1.6	0.029	0.033	0.033	0.033	0.033	0.033	1.727
2	0.026	0.035	no candidate	0.111	0.058	0.087	1.811

Table 3: Perturbing a queuing system with $s = 2$, $K = 1$, $\lambda = 1$ and $\nu = 1.8$, with $\mu = (0.5705, 0.317, 0.088, 0.0245)$, to the same system with different service rates $\hat{\nu}$

Appendix C. Structural vs. Non-Structural Perturbations. Based upon the construction of the vectors z, u and x in Theorem 6.1 we can identify the subset of $\mathbf{\Delta}(G, \hat{\mu})$ such that $G + \Delta$, for $\Delta \in \mathbf{\Delta}(G, \hat{\mu})$, has the same support as G. To that end, let, for arbitrary matrix B, zeros(B) denote the set of indices for which $B_{i,j} = 0$. There are two ways for $\Delta = xz^\top$ to be structural (i.e., $\text{supp}(G) \neq \text{supp}(G + \Delta)$):

1. A non-existing edge appears ($0 = G_{i,j} < (G_{i,j} + \Delta_{i,j})$). The set of row indices for which this may happen is

$$S_0^{\text{supp}} := \{i : \exists (i,j) \in \text{zeros}(G) \text{ with } (j \in \text{supp}(z_+) \text{ and } u_i > 0)$$
$$\text{or } (j \in \text{supp}(z_-) \text{ and } l_i < 0)\}.$$

   In particular, when $x_i > 0$ for $i \in S_0^{\text{supp}}$ a non-existing edge appears.

2. An existing edge disappears ($G_{i,j} > (G_{i,j} + \Delta_{i,j}) = 0$). The set of row indices for which this may happen is

$$S_{\neq 0}^{\text{supp}} := \{i : l_i < 0 \text{ or } u_i > 0\}.$$

   In particular, when $x_i = l_i$ or $x_i = u_i$ for $i \in S_{\neq 0}^{\text{supp}}$ an existing edge disappears.

Therefore, the set of candidates such that $G + \Delta$ for $\Delta \in \mathbf{\Delta}(G, \hat{\mu})$ has the same support as G is

$$\mathbf{\Delta}^{\text{supp}}(G, \hat{\mu}) := \Big\{ \Delta = xz^\top \in \mathbf{\Delta}(G, \hat{\mu}) : \; x_i = 0 \text{ for } i \in S_0^{\text{supp}} \text{ and} \\ l_i < x_i < u_i \text{ for } i \in S_{\neq 0}^{\text{supp}} \Big\}.$$

Note that the set of non-structural candidates $\mathbf{\Delta}^{\text{supp}}(G, \hat{\mu})$ is not closed and an infimum is sought. Let the vector $\bar{x}$ be defined as

$$\bar{x}_i = \begin{cases} 0, & \text{if } i \in S_0^{\text{supp}} \\ u_i, & \text{if } i \notin S_0^{\text{supp}} \end{cases}.$$

If $\hat{\mu}^\top \bar{x} > 1$, then the vector $\bar{x}/(\hat{\mu}^\top \bar{x})$ is strictly smaller than u in its nonzero components, and the implied candidate $G + \Delta$ will have the same support as G and will therefore be irreducible if G was irreducible.

Practically, $\mathbf{\Delta}(G, \hat{\mu})$ equals $\mathbf{\Delta}^{\text{supp}}(G, \hat{\mu})$ if one uses a (non-zero) precision $\phi > 0$ and sets: i) $l_i = l_i + \phi$ when $l_i < 0$, ii) $u_i = u_i - \phi$ when $u_i > 0$, and iii) $l_i = u_i = 0$ for $i \in S_0^{\text{supp}}$. This means that the results concerning $\mathbf{\Delta}(G, \hat{\mu})$ also generalize to $\mathbf{\Delta}^{\text{supp}}(G, \hat{\mu})$.

Appendix D. Rank-1 Perturbations that preserve Stochasticity.

D.1. The Minimal 1-Norm Rank-1 Perturbation. The minimum 1-norm problem is given by

$$\min \gamma, \quad x^\top \mathbf{1} \leq \gamma, \quad 0 \leq x \leq u, \quad \hat{\mu}^\top x = 1,$$

where we assumed $\hat{\mu}^\top u > 1$, which implies that the feasible set is non-empty. Notice that the problem is essentially the same if we permute all elements in the vectors u, x, and $\hat{\mu}$ simultaneously. Therefore we can assume, without loss of generality, that the elements of the non-negative vector u are ordered in a non-increasing fashion:

$$u_1 \geq u_2 \geq \ldots \geq u_k > u_{k+1} = \ldots = u_n = 0,$$

where u_k is the last non-zero element of u. It follows from $0 \leq x \leq u$ that the last $n - k$ components of x must also be zero and that we only must consider the first k components of x in the minimization problem. Let us start with a tentative candidate $x = u$. In order to decrease the 1-norm of the nonzero part of x as much as possible with respect to the upper bound, we choose a uniform perturbation $x_i = u_i - \delta$, for $1 \leq i \leq k$, yielding $\hat{\mu}^\top x = \hat{\mu}^\top u - k\delta$. But in order to maintain $0 \leq x$, δ must be bounded by u_k. Therefore, if

$$\hat{\mu}^\top x = \hat{\mu}^\top u - \sum_{i=1}^{k} \hat{\mu}_i u_k \leq 1 < \hat{\mu}^\top u,$$

then the minimum norm solution is given by setting $\delta = (\hat{\mu}^\top u - 1)/\sum_{i=1}^{k} \hat{\mu}_i$ and $x_i = u_i - \delta$ for $1 \leq i \leq k$. If, on the other hand,

$$1 < \hat{\mu}^\top u - \sum_{i=1}^{k} \hat{\mu}_i u_k,$$

then we modify the nonzero upper bounds u_i, for $1 \le i \le k$, by $\hat{u}_i = u_i - u_k$ and keep the zero ones $\hat{u}_i = u_i,\ k+1 \le i \le n$, yielding

$$\hat{u}_1 = u_1 - u_k \ge \hat{u}_2 = u_2 - u_k \ge \ldots \ge \hat{u}_k = u_k - u_k = \hat{u}_{k+1} = \ldots = \hat{u}_n = 0,$$

and the quantity $\hat{\mu}^\top u$ by $\hat{\mu}^\top \hat{u}$. This implies $\hat{u}_k = 0$ and we can then repeat the above procedure with a shorter vector of nonzero upper bounds. It is clear that we achieve a maximum decrease of γ at each step, and that the computed solution is unique.

D.2. The Minimal 2-Norm Rank-1 Perturbation. The minimum 2-norm problem stated in Theorem 6.2 is equivalent to

$$\min \gamma, \quad x^\top x \le \gamma^2, \quad 0 \le x \le u, \quad \hat{\mu}^\top x = 1,$$

where we assume $\hat{\mu}^\top u > 1$. Even though this is a convex problem that can be solved via LMI techniques, the quadratic inequality makes it harder to characterize the solution in analytic form. But the solution is unique since the level sets of the 2-norm form a strictly convex set. Also, if $x = \hat{\mu}/(\hat{\mu}^\top \hat{\mu})$ satisfies the constraints $0 \le x \le u$, then it is the minimum norm solution of our problem since it is already the minimum 2-norm solution without those constraints (see Theorem 5.1 and Corollary 5.3).

In general, a simple approximate solution is obtained as follows (and could be used as starting point for an optimization scheme). Clearly $x_u := u/(\hat{\mu}^\top u)$ is a candidate of our problem, and $x_\mu := \hat{\mu}/(\hat{\mu}^\top \hat{\mu})$ is a solution of the unconstrained problem, i.e., without $0 \le x \le u$. Moreover, the convex combinations

$$x_c := (1-c)x_u + c x_\mu, \quad 0 \le c \le 1,$$

all satisfy $\hat{\mu}^\top x_c = 1$. Therefore the largest value of c for which $0 \le x_c \le u$, implies a candidate that minimizes the norm of x on this line interval. This maximum value of c is given by

$$c = \min_{x_{\mu_i} > u_i} \frac{u_i - x_{u_i}}{x_{\mu_i} - x_{u_i}}.$$

D.3. The Minimal ∞-Norm Rank-1 Perturbation. The minimum ∞-norm problem is given by

$$\min \gamma, \quad x \le \gamma \mathbf{1}, \quad 0 \le x \le u, \quad \hat{\mu}^\top x = 1,$$

where we assume $\hat{\mu}^\top u > 1$. Again, we can assume without loss of generality that the elements of u are ordered in a nonincreasing manner:

$$u_1 = \ldots = u_\ell > u_{\ell+1} \ge \ldots \ge u_n,$$

where there are ℓ elements of maximal size. In order to decrease the ∞-norm of x as much as possible with respect to the upper bound, we choose a perturbation of all largest elements $x_i = u_i - \delta$, for $1 \le i \le \ell$, and bound δ by $u_1 - u_{\ell+1}$ so that x_i for $1 \le i \le \ell$ are still the largest elements in x. As a result, $\hat{\mu}^\top x = \hat{\mu}^\top u - \sum_{i=1}^{\ell} \hat{\mu}_i \delta$. If

$$\hat{\mu}^\top u - \sum_{i=1}^{\ell} \hat{\mu}_i (u_1 - u_{\ell+1}) \le 1 < \hat{\mu}^\top u,$$

then the minimum norm solution is given by setting

$$\delta = (\hat{\mu}^\top u - 1) \left(\sum_{i=1}^{\ell} \hat{\mu}_i \right)^{-1}.$$

If, on the other hand,

$$1 < \hat{\mu}^\top u - \sum_{i=1}^{\ell} \hat{\mu}_i (u_1 - u_{\ell+1}),$$

then we set the new maximal upper bounds to $\hat{u}_i = u_{\ell+1}$, $1 \le i \le \ell$, keep the other ones unchanged, i.e., $\hat{u}_i = u_i$ for $\ell + 1 \le i \le n$, and change the quantity $\hat{\mu}^\top u$ to $\hat{\mu}^\top \hat{u}$. This yields

$$\hat{u}_1 = \cdots = \hat{u}_\ell = \hat{u}_{\ell+1} \ge \ldots \ge \hat{u}_n.$$

implying that the number of equal largest elements has increased. We can then repeat the above procedure with the updated vector of upper bounds. It is clear that we achieve a maximum decrease of γ at each step, and that the computed solution is unique. Example 6 demonstrates the results of this procedure.

EXAMPLE 6. *Reconsider Example B.1. Using the algorithm as described above, we find the following rank-1 Δ of minimal ∞-norm that preserves stochasticity of $G + \Delta$:*

$$\Delta = \begin{bmatrix} -0.087 & -0.1957 & 0 & 0.2826 \\ -0.087 & -0.1957 & 0 & 0.2826 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which gives

$$G + \Delta = \begin{bmatrix} 0.6957 & 0.0217 & 0 & 0.2826 \\ 0.3043 & 0.1957 & 0.2174 & 0.2826 \\ 0 & 0.7826 & 0 & 0.2174 \\ 0 & 0 & 0.7826 & 0.2174 \end{bmatrix}$$

Compared to Example B.1, $G + \Delta$ is now a stochastic matrix which is achieved by a larger perturbation: $\|\Delta\|_\infty = 0.565$ instead of $\|\Delta\|_\infty = 0.2826$.

Now reconsider Example 2. The last column in Table 1 presents the minimal norms found by the algorithm from this section when $\alpha^\star \ge 1$. As expected, the norms are smaller than the norms of (6.4) but larger than the norms for $\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}$.

It was pointed out earlier that the subordinate v-norm $\|\Delta\|_v$ is essentially the ∞-norm of the scaled matrix $\|D_v^{-1} \Delta D_v\|_\infty$. The minimization of $\|\Delta\|_v$ can therefore also be performed using the procedure just described for the ∞-norm.

Appendix E. Application of Equation (7.4). We illustrate the solution proposed in (7.4) with two examples that are motivated from the theory of the wisdom of crowds in social network analysis, see [12, 19].

EXAMPLE 7. *Consider the ring network described in Section C. Suppose we want to maximize the weight of node 1 by changing the weight of a link from node 1 to one other node. By (7.4), we have*

$$\alpha^* = \frac{1}{n} \frac{b}{1 - (1 - 2b)} = \frac{1}{2n},$$

where we can choose either node 2 or node n to shift the mass from. Suppose we shift mass from the link of node 1 to node 2. This then gives a new stationary weight $3/(2n)$ for node 1, a weight of $1/(2n)$ for node 2, and the weight of the rest of the nodes remains $1/n$.

From (7.3), the corresponding perturbation matrix Δ can be found as follows:

$$\frac{\alpha}{\mu_i + \alpha} = \frac{\alpha^*}{\mu_1 + \alpha^*} = \frac{1}{3},$$

$$\Delta = \frac{\alpha^*}{\mu_1 + \alpha^*}(e_1 e_1^\top - e_1 e_2^\top)(I - G) = \frac{1}{3}\begin{bmatrix} 3b & -3b & b & \cdots & -b \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

which gives

$$G + \Delta = \begin{bmatrix} 1-b & 0 & \frac{1}{3}b & 0 & \dots & \frac{2}{3}b \\ b & 1-2\,b & b & 0 & \dots & 0 \\ 0 & b & 1-2\,b & b & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & b & 1-2\,b & b \\ b & 0 & \dots & 0 & b & 1-2b \end{bmatrix}.$$

Note that Δ is structural.

EXAMPLE 8. *Consider the star network given in Section C. Let $i = 2$, then $j = 1$ is the only possibility to shift mass from and (7.4) becomes*

$$\alpha^* = \frac{\beta}{(n-1)(1-\gamma+\beta)}.$$

Node 1 is the leader if $(n-1)(1-\gamma) > \beta$. Moreover, node i, for $i > 1$, can achieve a weight higher than the leader has if

$$\beta < (n-1)(1-\gamma) < 2\beta.$$

This can be seen as follows

$$\frac{1-\gamma}{1-\gamma+\beta} > \frac{\beta}{(n-1)(1-\gamma+\beta)}$$

being equivalent to

$$(n-1)(1-\gamma) > \beta.$$

The highest increase node i can realize is α^*, which gives the new stationary value

$$\hat{\mu}_i = \mu_i + \alpha^* = \frac{2\beta}{(n-1)(1-\gamma+\beta)}.$$

This value exceeds μ_1 if

$$\frac{2\beta}{(n-1)(1-\gamma+\beta)} > \frac{1-\gamma}{1-\gamma+\beta},$$

which proves the claim.

Appendix F. Numerical Results. All numerical results are obtained on a Windows laptop with an Intel i7 processor with 16.0 GB RAM. While solving the instances, we keep track of the norm found, whether the (approximate) solution is feasible, the running time in seconds, and the rank of the (approximate) solution. The results of the (approximate) solutions are then averaged over the different random instances.

The results for dense random matrices of dimensions $n = 100$, $n = 500$ and $n = 1000$ can be found in Table 4. We have imposed a time limit of 10 minutes on each method. To prevent excessive running times for $n = 1000$, we did not use the LP and R1SH and we mix a fraction of 0.01 of the random vector to generate random $\hat{\mu} = 0.01v + 0.99\mu$. In the results, "nan" stands for "not a number" which means that no candidate was found for any of the generated instances, i.e., each instance could either not be solved by the particular method or the method exceeded the time limit. The notation "ff" indicates which fraction of the problems yielded a candidate.

The numerical results in Table 5, Table 6 and Table 7 show that rank-1 steps heuristics are applicable to specific cases of sparse matrix instances where $\hat{\mu}$ changes most significantly on a subset of nodes that constitutes a dense subgraph.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0506	0	0.0002	1
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\geq 0}} \|\Delta\|_\infty$	0.0506	1	6.632	98.4
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.001	nan
R1SH(2)	0.1202	0.56	0.1335	2
R1SH(4)	0.0751	0.96	0.4162	4
R1SH(8)	0.0573	1	1.0055	8
R1SH(16)	0.0526	1	2.2879	16
R1SH	0.0515	1	22.2669	80.36
Riesz projector	0.6084	1	0.0001	99

(a) $n = 100$ nodes.

Method	mean($\|\Delta\|_\infty$)	ff.	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.049871	0	0.006601	1
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}\cap\boldsymbol{\Delta}^{\geq 0}} \|\Delta\|_\infty$	0.04974	0.56	611.633	497.929
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.010003	nan
R1SH(2)	nan	0	0.006959	nan
R1SH(4)	nan	0	0.007039	nan
R1SH(8)	nan	0	0.007119	nan
R1SH(16)	nan	0	0.300243	nan
R1SH	0.050379	0.96	294.73	498.875
Riesz projector	0.557979	1	0.001922	499

(b) $n = 500$ nodes.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.005014	0	0.049828	1
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.049937	nan
R1SH(2)	0.013155	0.8	5.06878	2
R1SH(4)	0.006194	1	11.9585	4
R1SH(8)	0.005166	1	23.3796	8
R1SH(16)	0.005034	1	47.3646	16
Riesz projector	0.543417	1	0.008795	999

(c) $n = 1000$ nodes.

Table 4: Mean results for 25 dense random matrices of different sizes.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0587	0	0.0001	1
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|_\infty$	0.0683	1	3.7114	78.68
$\min_{\Delta \in \boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0001	nan
R1SH(K) for $K = 2, 4, 8, 16$	nan	0	≈0.0001	nan
R1SH	0.7777	0.72	1.2242	4.9444
FR1SH(1e-08)	0.7172	1	1.9234	9.04
Riesz projector	1.9158	1	0	99

(a) $\hat{\mu}$ where the largest clique is made uniform.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0287	0	0.0002	1
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}} \cap \boldsymbol{\Delta}^{\geq 0}} \|\Delta\|_\infty$	0.0347	1	4.5041	85
$\min_{\Delta \in \boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0018	nan
R1SH(K) for $K = 2, 4, 8, 16$	nan	0	≈0.0004	nan
R1SH	0.15	0.84	0.4457	5.2381
FR1SH(1e-08)	0.0384	1	1.4358	10.88
Riesz projector	1.9135	1	0	99

(b) $\hat{\mu}$ where the importance of the largest clique is increased by 10%.

Table 5: Mean results of 25 Barabási–Albert preferential attachment social networks of $n = 100$ nodes for different goals.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0008	0	0.3725	1
$\min_{\Delta \in \boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0534	nan
R1SH(2)	nan	0	0.1055	nan
R1SH(4)	nan	0	0.1057	nan
R1SH(8)	nan	0	0.1064	nan
R1SH(16)	nan	0	0.1035	nan
R1SH	0.1867	1	60.1135	6.32
FR1SH(1e-08)	0.1866	1	53.3658	6.4
Riesz projector	2	1	0.0538	2119

(a) Results social network.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0005	0	0.0661	1
$\min_{\Delta \in \boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	0.558	0.2	0.0214	1
R1SH(2)	0.558	0.2	0.3396	1
R1SH(4)	0.558	0.2	0.3456	1
R1SH(8)	0.558	0.2	0.3724	1
R1SH(16)	0.558	0.2	0.3626	1
R1SH	0.0584	1	43.9586	2.88
FR1SH(1e-08)	0.0584	1	59.9862	2.84
Riesz projector	1.9985	1	0.0159	1004.56

(b) Results road network.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta \in \boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0026	0	0.0857	1
$\min_{\Delta \in \boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0228	nan
R1SH(2)	nan	0	0.0348	nan
R1SH(4)	nan	0	0.0341	nan
R1SH(8)	nan	0	0.037	nan
R1SH(16)	nan	0	0.0397	nan
R1SH	0.0266	1	59.7658	7.48
FR1SH(1e-08)	0.0264	1	62.1489	8.52
Riesz projector	1.9996	1	0.0151	1090

(c) Results organizational network.

Table 6: Mean results for real-life sparse networks (descriptions can be found in Table 2). For each network, 25 largest cliques are considered and the results are averaged. The methods have a time limit of 60 seconds.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0008	0	0.3194	1
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0399	nan
R1SH(2)	nan	0	0.0865	nan
R1SH(4)	nan	0	0.0871	nan
R1SH(8)	nan	0	0.0873	nan
R1SH(16)	nan	0	0.0822	nan
R1SH	0.1867	1	227.599	6.32
FR1SH(1e-08)	0.1866	1	228.049	6.4
Riesz projector	2	1	0.0456	2119

(a) Results social network.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0005	0	0.071	1
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	0.558	0.2	0.0238	1
R1SH(2)	0.558	0.2	0.36	1
R1SH(4)	0.558	0.2	0.3471	1
R1SH(8)	0.558	0.2	0.3664	1
R1SH(16)	0.558	0.2	0.3678	1
R1SH	0.0584	1	49.6037	2.88
FR1SH(1e-08)	0.0584	1	77.0852	2.84
Riesz projector	1.9985	1	0.0171	1004.56

(b) Results road network.

Method	mean($\|\Delta\|_\infty$)	ff	mean run time	mean rank
$\min_{\Delta\in\boldsymbol{\Delta}^{\hat{\mu}}} \|\Delta\|_\infty$	0.0026	0	0.0856	1
$\min_{\Delta\in\boldsymbol{\Delta}(G,\hat{\mu})} \|\Delta\|_\infty$	nan	0	0.0198	nan
R1SH(2)	nan	0	0.0368	nan
R1SH(4)	nan	0	0.0388	nan
R1SH(8)	nan	0	0.0399	nan
R1SH(16)	nan	0	0.0342	nan
R1SH	0.0266	1	74.93	7.48
FR1SH(1e-08)	0.0263	1	148.367	9.96
Riesz projector	1.9996	1	0.0163	1090

(c) Results organizational network.

Table 7: Mean results for real-life sparse networks (descriptions can be found in Table 2). For each network, 25 largest cliques are considered and the results are averaged. In contrast to the results from Table 6, the methods here got a time limit of 10 minutes instead of 60 seconds.