

1     **PERTURBATION AND INVERSE PROBLEMS OF STOCHASTIC**  
2                                     **MATRICES**

3             JOOST BERKHOUT , BERND HEIDERGOTT , AND PAUL VAN DOOREN

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

15     **Key words.** Markov Chains, Perturbation Analysis, Inverse Problems, Target Stationary Dis-  
16 tribution Problem

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

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

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

26 (1.1)

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

37     We are considering applications where the stochastic matrices  $G$  and  $G + \Delta$  model,  
38 for example, hyperlink networks, social networks, or queuing networks. Their station-  
39 ary distributions contain important information on the nodes in the network, such as  
40 their centrality or other types of rankings. The target stationary distribution  $\hat{\mu}$  then  
41 captures some desired state of the system. In practice, one is interested in reaching  
42 that desired state with minimum effort, i.e., we are interested in finding minimal norm

43 perturbations. For example, a social agent may want to obtain a certain influence  
44 level within a social network with minimum effort.

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

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

$$54 \quad \Delta^{\hat{\mu}} := \Delta^{\hat{\mu}}(G, \hat{\mu}) := \{\Delta : \hat{\mu}^\top(G + \Delta) = \hat{\mu}^\top, \Delta \mathbf{1} = 0\}$$

55 and

$$56 \quad \Delta^{\geq 0} := \Delta^{\geq 0}(G) := \{\Delta : G + \Delta \geq 0\},$$

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

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

$$64 \quad \Delta^{\text{rank-1}} = \{\Delta : \text{rank}(\Delta) = 1\},$$

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

$$68 \quad \Delta := \Delta(G, \hat{\mu}) := \Delta^{\hat{\mu}} \cap \Delta^{\geq 0} \cap \Delta^{\text{rank-1}}$$

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

$$72 \quad \min_{\Delta \in \Delta^{\hat{\mu}}} \|\Delta\| \leq \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta^{\geq 0}} \|\Delta\| \leq \min_{\Delta \in \Delta(G, \hat{\mu})} \|\Delta\|,$$

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

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

82 The price we have to pay for the analytical elegance and simplicity of our explicit  
83 rank-1 solutions is that they may not solve the TSDP. Fortunately, as we show in this  
84 paper, in such cases, an approximate (i.e., not achieving minimal norm) solution  $\Delta$  can  
85 often be obtained via a sequence of rank-1 perturbations. We develop heuristics for  
86 finding a sequence of rank-1 perturbation steps so that the accumulated perturbation  
87 is of higher rank and does allow to reach the target stationary distribution. Numerical

88 experiments will show the efficiency of our approach for dense random matrices and  
 89 for specific sparse matrices.

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

98 **2. Applications.** In this section, we mention a series of applications as motiva-  
 99 tion for this research. Illustrating examples from these applications will also be given  
 100 at a later stage.

101 **2.1. Road Networks.** Consider a road network consisting of  $n$  roads and rep-  
 102 resented by a graph in the following manner. The nodes of the graph represent the  
 103 road segments, and a directed link between road segments  $i$  and  $j$  means that there  
 104 is a junction that allows going from road  $i$  to road  $j$ . We assume that information is  
 105 available on the traffic flow and that this is modeled as a discrete-time random walk  
 106 on the road network: the probability  $G_{i,j}$  thus models the probability that a car on  
 107 road segment  $i$  turns into road segment  $j$  at a particular time instant. As detailed in  
 108 [11], self-links  $G_{i,i}$  can be chosen in such a way that they mimic the travel times along  
 109 the road elements. The value of  $\mu_i$  of the stationary distribution of  $G$  then represents  
 110 the long-range time average with which a typical car will be found to drive on road  
 111 segment  $i$ . In other words, the vector  $\mu$  represents the (relative) road congestion.

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

120 **2.2. Social Networks.** Social network analysis investigates the social structures  
 121 of relationships between agents [44]. A social network can be modeled as a finite set  
 122 of nodes, and the edges connecting them represent the social relationship between  
 123 the nodes [36]. Social networks are typically represented by weighted graphs, where  
 124 the nodes set is the set of social agents, and a directed link  $(i, j)$  between agents  
 125  $i$  and  $j$  means either that  $i$  follows  $j$  (i.e.,  $i$  puts trust in  $j$ ) or that  $i$  influences  $j$   
 126 (i.e.,  $i$  sends information to  $j$ ). The relative strength of the link is expressed via a  
 127 weight function  $W_{i,j}$ . Through normalizing the weights, a Markov chain  $G$  can be  
 128 constructed of which the stationary distribution expresses the influence or centrality  
 129 of the social agents. For example, if the weights reflect trust, then the stationary  
 130 distribution expresses the relative trust the agents receive in the network.

131 In perturbation analysis of social networks, one is interested in perturbing the  
 132 stationary distribution. For example, agent  $i$  can influence his or her outgoing nodes,  
 133 and the question arises which perturbation of the  $i$ -th row will maximize the impor-  
 134 tance of  $i$ . In the same vein, agent  $i$  may be interested in decreasing the importance  
 135 of some other node  $j \neq i$  by adjusting its outgoing links. Finally, coalition games



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

$$180 \quad (3.1) \quad \|\hat{\mu} - \mu\|_{\alpha} \leq D(\Delta, G),$$

181 where  $\|\cdot\|_{\alpha}$  denotes a suitable vector norm, and  $D(\Delta, G)$  is a scalar function of  $\Delta$  and  
 182  $G$ . This type of perturbation analysis dates back to Schweitzer's pioneering paper  
 183 [35]. To the best of our knowledge, the first paper putting this perturbation question  
 184 into the framework of (3.1) is [27]. This paper proposed bounds of the form

$$185 \quad (3.2) \quad D(\Delta, G) = \kappa \|\Delta\|_{\beta},$$

186 for some appropriate matrix norm  $\|\cdot\|_{\beta}$ , where  $\kappa$  is the so-called *condition number* of  
 187 the Markov chain  $G$  for the  $(\|\cdot\|_{\alpha}, \|\cdot\|_{\beta})$ -norm pair. Finding bounds of the type (3.1)  
 188 is a field of active research [4, 21, 32, 33, 37, 30, 8, 18, 5, 26, 1] and various condition  
 189 number bounds have been proposed in the literature [10, 18]. Perturbation bounds  
 190 like (3.1) are of interest in a wide range of application areas, such as mathematical  
 191 physics [41], climate modeling [9], Bayesian statistics [3, 2], and bio-informatics [29,  
 192 34]. Conditions numbers for quantum Markov chains in mathematical physics can be  
 193 found in [40].

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

197 **4. Technical Preliminaries.** In this paper, we consider square  $n \times n$  non-  
 198 negative matrices  $A$ , i.e., matrices with non-negative elements, which we denote by  
 199  $A \geq 0$ . If in the matrix  $A$  all elements are strictly larger than 0, we call  $A$  a positive  
 200 matrix and denote this by  $A > 0$ . The positive semi-definite matrices, on the other  
 201 hand, will be denoted by  $A \succeq 0$ . The support of a general matrix  $A$ , denoted by  
 202  $\text{supp}(A)$ , is the set of indices  $(i, j)$  for which  $A_{i,j} \neq 0$ . It is well known that non-  
 203 negative matrices have an eigenvalue that is equal to its spectral radius  $\rho := \rho(A)$   
 204 and hence is real and non-negative. Moreover, if  $A$  is irreducible, then this so-called  
 205 *Perron-root*  $\rho$  is simple and positive. Therefore the matrix  $(A - \rho I)$  has rank  $n - 1$ ,  
 206 where  $I$  denotes the  $n \times n$  identity matrix. Moreover, the corresponding left and right  
 207 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  
 208 Perron vectors are typically normalized using  $\mathbf{v}^{\top} \mathbf{1} = 1$  and  $\mathbf{1}^{\top} \mathbf{u} = 1$ , where  $\mathbf{1}$  is the  
 209  $n$ -vector of all ones. The non-negative matrix  $A$  is said to be stochastic if  $A \mathbf{1} = \mathbf{1}$ .  
 210 For such a matrix, the spectral radius  $\rho(A) = 1$ . We will denote the  $i$ -th canonical  
 211 basis vector of  $\mathbb{R}^n$  by  $e_i$ .

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

$$213 \quad (4.1) \quad \|y\|_* := \sup_{z \neq 0} \frac{|z^{\top} y|}{\|z\|} = \sup_{z \neq 0} \frac{|y^{\top} z|}{\|z\|}.$$

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

$$216 \quad (4.2) \quad \|A\| := \sup_{z \neq 0} \frac{\|Az\|}{\|z\|}.$$

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

$$221 \quad (4.3) \quad \|x\|_v := \sup_{1 \leq i \leq n} \frac{|x_i|}{v_i} = \|D_v^{-1}x\|_\infty \quad \text{where} \quad D_v = \text{diag}(v_1, \dots, v_n)$$

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

$$223 \quad v_i = \alpha^i, \quad 1 \leq i \leq n,$$

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

232 **5. General Rank-1 Perturbations.** In this section, we show that rank-1 per-  
233 turbations can be used to try to solve the TSDP. We will drop the constraint that  $\Delta$   
234 has to belong to  $\Delta^{\geq 0}$  and impose instead that  $\Delta$  is rank-1, that is, we consider

$$235 \quad (5.1) \quad \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta^{\text{rank-1}}} \|\Delta\|.$$

236 While the following theorem is fairly standard, we provide, for the sake of complete-  
237 ness, a proof in Appendix A.1.

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

$$240 \quad \Delta^{\hat{\mu}} \cap \Delta^{\text{rank-1}} = \left\{ \Delta : \Delta = \frac{x\hat{\mu}^\top}{\hat{\mu}^\top x}(I - G) \text{ for all } x \text{ with } \hat{\mu}^\top x \neq 0 \right\},$$

241 *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*  
242  *$\hat{\mu}$ . For any subordinate matrix norm, a minimum norm choice of  $\Delta \in \Delta^{\hat{\mu}} \cap \Delta^{\text{rank-1}}$*   
243 *is obtained by any  $x$  such that*

$$244 \quad (5.2) \quad \frac{|x^\top \hat{\mu}|}{\|x\|} = \|\hat{\mu}\|_*$$

245 *and the corresponding minimum norm  $\Delta$  has the norm*

$$246 \quad (5.3) \quad \|\Delta\| = \|(I - G)^\top \hat{\mu}\|_* / \|\hat{\mu}\|_*.$$

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

$$248 \quad \min_{\Delta \in \Delta^{\hat{\mu}}} \|\Delta\| = \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta^{\text{rank-1}}} \|\Delta\|.$$

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

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

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

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

255 **COROLLARY 5.3.** *The solutions to (5.1) for the 1-, 2-,  $v$ - and  $\infty$ -norms are given*  
 256 *by the vectors  $cx$  (with scale factor  $c \neq 0$ ), where  $x$  is defined as follows:*

- 257 • *for the 1-norm:  $x = e_i$  where  $i$  is any maximizing index of the vector  $\hat{\mu}$ , and*  
 258  $\|\Delta\|_1 = \|(I - G)^\top \hat{\mu}\|_\infty / \|\hat{\mu}\|_\infty$
- 259 • *for the 2-norm:  $x = \hat{\mu}$ , and  $\|\Delta\|_2 = \|(I - G)^\top \hat{\mu}\|_2 / \|\hat{\mu}\|_2$*
- 260 • *for the  $v$ -norm:  $x = D_v \mathbf{1}$ , and  $\|\Delta\|_v = \|D_v (I - G)^\top \hat{\mu}\|_1 / \|D_v \hat{\mu}\|_1$*
- 261 • *for the  $\infty$ -norm:  $x = \mathbf{1}$ , and  $\|\Delta\|_\infty = \|(I - G)^\top \hat{\mu}\|_1 / \|\hat{\mu}\|_1$ .*

262 We illustrate Corollary 5.3 with the following example.

263 **EXAMPLE 1.** *Let*

$$264 \quad G = \begin{bmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{bmatrix}$$

265 *with stationary distribution  $\mu^\top = (9/17, 8/17)$ . Following Corollary 5.3 for the*  
 266  *$\infty$ -norm, the smallest rank-1 perturbation to achieve a uniform distribution  $\hat{\mu}^\top =$*   
 267  *$(1/2, 1/2)$  is*

$$268 \quad \Delta = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2/3 & -2/3 \\ -3/4 & 3/4 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

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

$$270 \quad G + \Delta = \frac{1}{24} \left( \begin{bmatrix} 8 & 16 \\ 18 & 6 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \right) = \frac{1}{24} \begin{bmatrix} 7 & 17 \\ 17 & 7 \end{bmatrix}$$

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

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

277 **6. Rank-1 Perturbations Preserving Stochasticity.** In this section, we fo-  
 278 cus on solving

$$279 \quad (6.1) \quad \min_{\Delta \in \mathbf{\Delta}} \|\Delta\|$$

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

283 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  $\mu$ . 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  $\text{supp}(z_+) \cap \text{supp}(z_-) = \emptyset$ . Introduce vectors  $\ell := \ell(G, \hat{\mu})$  and  $u := u(G, \hat{\mu})$  defined through

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

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

$$(6.3) \quad \Delta(G, \hat{\mu}) = \left\{ \Delta = xz^\top : \forall x \in \mathbb{R}^n \text{ with } \hat{\mu}^\top x = 1 \text{ and } \ell \leq x \leq u \right\}.$$

It further holds:

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

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

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

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  $\alpha^*$  is effectively 0. In the following example, we consider the maximal feasible step-size  $\alpha^*$ .

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  $\alpha^*$ , the value of  $\|\Delta\|_\infty$  for  $\Delta$  from (6.4) with  $\alpha = 1$  in case  $\alpha^* \geq 1$ , and the minimum value of  $\|\Delta\|_\infty$  for  $\Delta \in \Delta^{\hat{\mu}}$  and  $\Delta \in \Delta(G, \hat{\mu})$ .

It follows from Table 1 that the  $\hat{\mu}$ 's for  $\hat{\nu} = 0.2$  and  $\hat{\nu} = 2$  are too different from  $\mu$  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  $\mu$  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

$\hat{\nu}$	$\hat{\mu}$	$\alpha^*$	$\ \Delta\ _\infty$ of (6.4)	$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$
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  $\hat{\nu}$

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

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

$$326 \quad G_r(b) = \begin{bmatrix} 1-2b & b & 0 & 0 & \dots & b \\ b & 1-2b & b & 0 & \dots & 0 \\ 0 & b & 1-2b & b & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & b & 1-2b & b \\ b & 0 & \dots & 0 & b & 1-2b \end{bmatrix},$$

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

$$329 \quad G_s(\beta, \gamma) = \begin{bmatrix} 1-\beta & \frac{\beta}{n-1} & \frac{\beta}{n-1} & \frac{\beta}{n-1} & \dots & \frac{\beta}{n-1} \\ 1-\gamma & \gamma & 0 & 0 & \dots & 0 \\ 1-\gamma & 0 & \gamma & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1-\gamma & 0 & \dots & 0 & \gamma & 0 \\ 1-\gamma & 0 & \dots & 0 & 0 & \gamma \end{bmatrix},$$

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

$$331 \quad \mu_1 = \frac{1-\gamma}{1-\gamma+\beta} \quad \text{and} \quad \mu_i = \frac{\beta}{(n-1)(1-\gamma+\beta)}, \quad \text{for } i > 1.$$

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

$$336 \quad \Delta = \begin{bmatrix} -0.2182 & 0.1091 & 0 & 0.1091 \\ -0.1636 & 0.0818 & 0 & 0.0818 \\ 0 & 0 & 0 & 0 \\ -0.1636 & 0.0818 & 0 & 0.0818 \end{bmatrix},$$

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

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

$$342 \quad \Delta = \begin{bmatrix} 0.1571 & -0.0524 & -0.0524 & -0.0524 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

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

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

$$349 \quad (6.5) \quad \hat{\mu}^\top x = 1, \quad \ell \leq x \leq u,$$

350 to

$$351 \quad (6.6) \quad \hat{\mu}^\top x = 1, \quad 0 \leq x \leq u,$$

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

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

$$358 \quad (6.7) \quad \begin{aligned} & \min_x \|xz^\top\| \\ & \text{s.t.} \quad \hat{\mu}^\top (G + xz^\top) = \hat{\mu}^\top, \\ & \quad \hat{\mu}^\top x = 1, \\ & \quad G + xz^\top \geq 0, \\ & \quad 0 \leq x \leq u \end{aligned}$$

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

362 **THEOREM 6.2.** *Let  $G$  be an irreducible stochastic matrix with stationary distribu-*  
 363 *tion  $\mu$ , and let  $\hat{\mu}$  ( $\neq \mu$ ) be a target stationary distribution. Define the vectors  $z$ ,  $u$  and*  
 364  *$x$  as in Theorem 6.1, then Problem (6.7) has a solution if and only if  $\hat{\mu}^\top u \geq 1$ . This*  
 365 *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*  
 366 *a solution for every  $x$  solving the following convex optimization problems for the 1-,*  
 367 *2-,  $v$ - and  $\infty$ -norms, respectively:*

- 368 •  $\min \gamma$ , s.t.  $x^\top \mathbf{1} \leq \gamma$ ,  $0 \leq x \leq u$ ,  $\hat{\mu}^\top x = 1$  for the 1-norm
- 369 •  $\min \gamma$ , s.t.  $\begin{bmatrix} \gamma x^\top \\ x \gamma \mathbf{1} \end{bmatrix} \succeq 0$ ,  $0 \leq x \leq u$ ,  $\hat{\mu}^\top x = 1$  for the 2-norm
- 370 •  $\min \gamma$ , s.t.  $x \leq \gamma D_v \mathbf{1}$ ,  $0 \leq x \leq u$ ,  $\hat{\mu}^\top x = 1$  for the  $v$ -norm
- 371 •  $\min \gamma$ , s.t.  $x \leq \gamma \mathbf{1}$ ,  $0 \leq x \leq u$ ,  $\hat{\mu}^\top x = 1$  for the  $\infty$ -norm

372 In Appendix D, we provide explicit constructions for solving the convex optimiza-  
 373 tion problems given in Theorem 6.2.

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

$$380 \quad \Delta = e_i z^\top / \hat{\mu}_i,$$

where we need

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

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

388 **7.1. Increasing Only One Stationary Distribution Element.** Let us as-  
 389 sume that direction  $d$  is chosen such that the only positive value is  $d_i$  (corresponding to  
 390 the row we perturb in  $G$ ) and that the other values are negative or zero. A particular  
 391 choice that is useful is  $d = e_i - \mu$  which gives for the elements in  $\hat{\mu} := \hat{\mu}_\alpha = \mu + \alpha d$ :

$$392 \quad (7.1) \quad \hat{\mu}_i := (\hat{\mu}_\alpha)_i = \mu_i + \alpha(1 - \mu_i) \quad \text{and} \quad \hat{\mu}_j := (\hat{\mu}_\alpha)_j = (1 - \alpha)\mu_j, \quad \forall j \neq i,$$

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

$$394 \quad (7.2) \quad z^\top = d^\top (I - G) = \alpha e_i^\top (I - G), \quad G + \Delta = G + \frac{\alpha}{\mu_i + \alpha(1 - \mu_i)} e_i e_i^\top (I - G).$$

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

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

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

402 **EXAMPLE 4.** Consider a traffic network  $G$ , see Section 2.1, where  $i$  represents  
 403 a road segment that is of key importance for traffic congestion control. We tolerate  
 404 deviations from the traffic network as long as they do not increase the congestion at  
 405  $i$  above a pre-specified fraction  $\beta > 0$ , and we compute the minimal perturbation of  
 406 the given traffic network that reaches this tolerance bound. This minimal perturbation  
 407 gives robustness insights on, for example, the maximal measurement errors we can

408 accept to ensure that the current congestion at  $i$  does not exceed the pre-specified  
 409 fraction, or on which road segments are crucial to be accurately measured. We are  
 410 thus looking for a minimal perturbation  $\Delta$  of the  $i$ -th row such that  $\hat{\mu}_i \leq (1 + \beta)\mu_i$ .  
 411 For solving  $\Delta$ , we assume for ease of presentation that the mass that is shifted to  $i$  is  
 412 taken uniformly from the other nodes so that

$$413 \quad \hat{\mu}_i \leq (1 + \beta)\mu_i \Leftrightarrow \mu_i + \alpha(1 - \mu_i) \leq (1 + \beta)\mu_i \Leftrightarrow \alpha \leq \frac{\beta\mu_i}{1 - \mu_i}.$$

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

$$415 \quad \Delta = \frac{\beta}{1 + \beta} \frac{1}{1 - \mu_i} e_i e_i^\top (I - G).$$

**7.2. Maximal Weight Shift Between Two Elements of  $\mu$ .** 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  $\hat{\mu} := \hat{\mu}_\alpha = \mu + \alpha d$ :

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

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

$$417 \quad (7.3) \quad G + \Delta = G + \frac{\alpha}{\mu_i + \alpha} (e_i e_i^\top - e_i e_j^\top) (I - G).$$

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

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

and

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

418 The first and last of these inequalities hold for every  $\alpha \geq 0$ , but the second inequality  
 419 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  
 420 best choice for the index  $j$  is therefore to choose a maximal solution of

$$421 \quad (7.4) \quad \alpha^* = \max_{j \neq i} \mu_i G_{i,j} / (1 - G_{j,j}).$$

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

## 425 8. Heuristics for General-Rank Perturbations Preserving Stochasticity.

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

$$427 \quad \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|,$$

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

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

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

$$438 \quad \hat{\mu}_i = \begin{cases} 1/n + a & i = 1 \\ 1/n - a & i = 3 \\ 1/n & i \notin \{1, 3\} \end{cases}.$$

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

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

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

451 **8.1. Mathematical Programming Problems.** We reformulate our TSDP so  
452 that existing algorithms from literature can be used and which are implemented in  
453 (commercial) solvers. In the cases of the 1-norm, the  $v$ -norm, and the  $\infty$ -norm, we  
454 can cast the TSDP into a linear programming (LP) problem, and for the 2-norm, we  
455 can reformulate it as a linear matrix inequality (LMI) problem. Let  $Z := |\Delta|$ , then  
456 the TSDP can be written as the following LP problem in the variables  $\Delta$ ,  $Z$  and  $\gamma$ ,  
457 for the 1-,  $\infty$ - and  $v$ -norms, respectively:

$$458 \quad (8.1) \quad \begin{aligned} & \min_{\Delta, Z, \gamma} \quad \gamma \\ & \text{s.t.} \quad \hat{\mu}^\top (G + \Delta) = \hat{\mu}^\top \\ & \quad \quad G + \Delta \geq 0 \\ & \quad \quad Z \geq \Delta \\ & \quad \quad Z \geq -\Delta \\ & \quad \quad \begin{cases} \text{if } \gamma \geq \sum_i Z_{i,j}, \forall j & \text{for the 1-norm} \\ \text{if } \gamma \geq \sum_j Z_{i,j}, \forall i & \text{for the } \infty\text{-norm} \\ \text{if } v_i \gamma \geq \sum_j Z_{i,j} v_j, \forall i & \text{for the } v\text{-norm,} \end{cases} \end{aligned}$$

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

$$460 \quad (8.2) \quad \begin{aligned} & \min_{\Delta, \gamma} \quad \gamma \\ & \text{s.t.} \quad \hat{\mu}^\top (G + \Delta) = \hat{\mu}^\top \\ & \quad \quad G + \Delta \geq 0 \\ & \quad \quad \begin{bmatrix} \gamma I & \Delta^\top \\ \Delta & \gamma I \end{bmatrix} \succeq 0. \end{aligned}$$

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

464 Problems (8.1) and (8.2) always have a non-empty feasible set which can be  
 465 characterized using [14] as  $\Delta = \hat{G} - G$  for all  $\hat{G}$  in the convex polytope of stochastic  
 466 matrices with stationary distribution  $\hat{\mu}$ . Note that the feasible set includes  $\Delta =$   
 467  $\mathbf{1}\hat{\mu}^\top - G$ . Finding a general-rank solution is much more complex than finding a rank-  
 468 1 solution since the number of decision variables is quadratic in  $n$ , rather than linear.  
 469 Therefore, these programs have a worst-case time complexity of  $\mathcal{O}(n^6)$ . As a result,  
 470 (commercial) solvers are not recommended for large-scale problems.

471 **8.2. Rank-1 Steps Heuristics.** In this section, so-called rank-1 steps heuristics  
 472 are developed that compute approximate solutions of the TSDP. Starting from  $\mu$ ,  
 473 the idea is to iteratively reach intermediate stationary distributions that are getting  
 474 “closer and closer” to the target  $\hat{\mu}$  as illustrated in Figure 1. The  $i$ -th intermediate  
 475 stationary distribution after  $i$  perturbations/steps is denoted by  $\mu^{(i)}$ . The  $\mu^{(i)}$ ’s need  
 476 to be determined upfront or dynamically along the way. To make the heuristic com-  
 477 putationally efficient,  $\mu^{(i)}$  should be reachable from  $\mu^{(i-1)}$  via a rank-1 perturbation  
 478 that preserves stochasticity as analyzed in previous sections. The heuristic later on  
 479 prescribes how possible  $\mu^{(i)}$ ’s can be determined (for example, by fixing its elements  
 480 to that of  $\hat{\mu}$ , respectively, giving  $\mu^{(n-1)} = \hat{\mu}$ ). Although numerical experiments show  
 481 that it often works, there is no guarantee that the heuristic leads to a sequence of  
 482  $\mu^{(i)}$ ’s leading to  $\hat{\mu}$ . In case it cannot, one can fall back to the candidate  $\Delta = \mathbf{1}\hat{\mu}^\top - G$ .

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

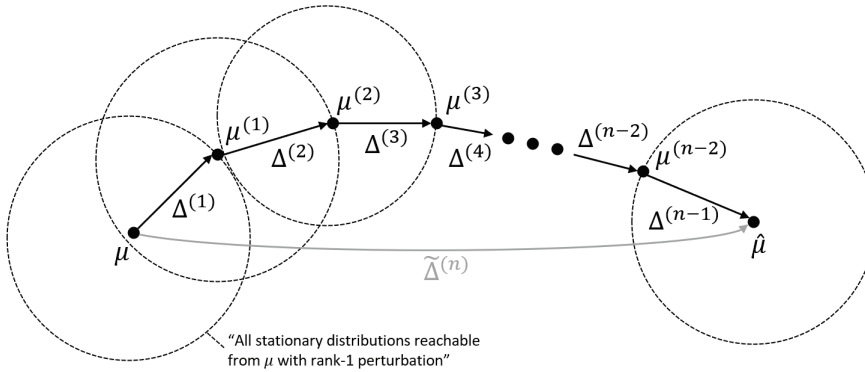


Fig. 1: Illustration of a rank-1 steps heuristic that takes  $n - 1$  rank-1 perturbations (or steps) towards  $\hat{\mu}$  to approximately solve  $\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|$ .

493  
 494

There are two problems with this approach that need to be addressed. The first

495 problem is that it is unknown what a good feasible sequence of  $\mu^{(i)}$ 's are. The sec-  
 496 ond problem is that at every step the rank-1 perturbation does not take previous  
 497 perturbations into account which unnecessarily increases the norm of the accumu-  
 498 lated perturbation. Instead, ensuring that each perturbation “connects efficiently” to  
 499 the previous accumulated perturbation can significantly reduce the norm of the final  
 500 accumulated perturbation.

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

$$505 \quad (8.3) \quad \min_{\Delta \in \mathbf{\Delta}(G + \tilde{\Delta}^{(i)}, \mu^{(i)})} \|\tilde{\Delta}^{(i)} + \Delta\|.$$

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

$$510 \quad (8.4) \quad \min_x \|\tilde{\Delta}^{(i)} + xz^\top\| \quad \text{s.t.} \quad l \leq x \leq u, \quad \mu^{(i)\top} x = 1,$$

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

$$516 \quad \min_x \|\tilde{\Delta}^{(i)} + lz^\top + xz^\top\| \quad \text{s.t.} \quad 0 \leq x \leq u - l, \quad \mu^{(i)\top} x = 1 - \mu^{(i)\top} l.$$

517 In comparison to Theorem 6.1, there are three changes in this minimization problem:  
 518 (i) the upperbound of  $x$  has changed, (ii)  $\mu^{(i)\top} x$  should equal  $1 - \mu^{(i)\top} l$  instead of  
 519 1, and (iii) the objective now contains  $\tilde{\Delta}^{(i)} + lz^\top (\neq 0)$ . The first two differences are  
 520 not fundamental and the same algorithmic procedures from Section D apply. The  
 521 third difference demands a change in the algorithm: When we now start with  $x = 0$   
 522 and start increasing  $x$  to ensure  $\mu^{(i)\top} x = 1 - \mu^{(i)\top} l$ , we have to take into account  
 523 that different  $x_i$ -increases have different effects on the objective due to  $\tilde{\Delta}^{(i)} + lz^\top$ .  
 524 As a result, the algorithm should first focus on decreasing the objective as much as  
 525 possible, then increase  $x$  as much as possible without affecting the objective, and  
 526 then lastly, increase  $x$  proportionally to their effect on the objective until  $\mu^{(i)\top} x =$   
 527  $1 - \mu^{(i)\top} l$  is reached. While doing this, one has to take the upperbounds  $u - l$  of  
 528  $x$  into account and check throughout whether  $\mu^{(i)\top} x = 1 - \mu^{(i)\top} l$  is met. Once this  
 529 restriction is met, one can return the solution  $(l+x)z^\top$  (reversing the substitution) for  
 530  $\min_{\Delta \in \mathbf{\Delta}(G + \tilde{\Delta}^{(i)}, \mu^{(i)})} \|\tilde{\Delta}^{(i)} + \Delta\|$ . For notational convenience, we denote the solution  
 531 of this procedure by

$$532 \quad P(G, \mu^{(i)}, \tilde{\Delta}^{(i)}) := \arg \min_{\Delta \in \mathbf{\Delta}(G + \tilde{\Delta}^{(i)}, \mu^{(i)})} \|\tilde{\Delta}^{(i)} + \Delta\|.$$

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

535 The success of the rank-1 step heuristic depends on the chosen sequence of  $\mu^{(i)}$ 's.  
 536 A straightforward sequence, that will also be used below, is to iteratively set the

537 elements of  $\mu$  to the corresponding elements in  $\hat{\mu}$  (and keep those elements fixed in  
538 consecutive steps). Specifically,  $\mu^{(i)}$  has  $i$  elements fixed to those of  $\hat{\mu}$ , of which  $i - 1$   
539 elements were already fixed to  $\hat{\mu}$  in  $\mu^{(i-1)}$ , and the remaining elements of  $\mu^{(i)}$  divide  
540 the remaining mass proportionally to  $\mu$ . The elements can be set in a random order,  
541 but experiments show that it is better to consider the elements of  $|\hat{\mu} - \mu|$  in decreasing  
542 order (the ‘‘preparation’’ in the rank-1 steps heuristic below). The reasoning is that  
543 at the beginning you have the most flexibility to overcome the largest differences.  
544 After  $n - 1$  steps,  $\hat{\mu}$  will possibly be reached, but there is no guarantee that this  
545 sequence indeed reaches  $\hat{\mu}$ , i.e., it is not guaranteed that there is always a rank-1  
546 perturbation from  $\mu^{(i-1)}$  to  $\mu^{(i)}$ . However, numerical experiments showed that it  
547 often finds a ‘‘path’’ to  $\hat{\mu}$ , and if it does not, one can fall back to candidate  $\mathbf{1}\hat{\mu}^\top - G$ .  
548 In the following, we elaborated the rank-1 steps heuristic for this  $\mu^{(i)}$  sequence, named  
549 R1SH, that converges after  $n - 1$  steps if the sequence is feasible.

550 **Rank-1 steps heuristic (R1SH):** (approximately solving  $\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|$ )

551 Given  $G$ ,  $\mu$  and  $\hat{\mu}$  perform:

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

553 
$$|\hat{\mu}_1 - \mu_1| \geq |\hat{\mu}_2 - \mu_2| \geq \dots \geq |\hat{\mu}_n - \mu_n|,$$

554 i.e.,  $|\hat{\mu} - \mu|$  is sorted from large to small without loss of generality.

555 2. **Initialization:** Set  $\mu^{(0)} = \mu$ .

556 3. **For**  $i \in \{1, 2, \dots, n - 1\}$ , respectively, **do:**

557 (a) Determine  $\mu_j^{(i)}$  for  $j = 1, 2, \dots, n$  as follows:

558 
$$\mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } 1 \leq j \leq i \\ \frac{\mu_j}{\sum_{k=i+1}^n \mu_k} \left(1 - \sum_{k=1}^i \hat{\mu}_k\right), & \text{if } i + 1 \leq j \leq n \end{cases}$$

559 i.e., we fix  $\mu_j^{(i)}$  to  $\hat{\mu}_j$ , for  $j = 1, 2, \dots, i$ , and the remaining mass of  
560  $1 - \sum_{k=1}^i \hat{\mu}_k$  is distributed over  $\mu_j^{(i)}$ , for  $j = i + 1, \dots, n$ , in proportion  
561 to the corresponding values in  $\mu$ .

562 (b) Calculate  $u$  (see Theorem 6.1) for stochastic matrix  $G + \tilde{\Delta}^{(i)}$  and new  
563 stationary vector  $\mu^{(i)}$ .

564 (c) **If**  $u^\top \mu^{(i)} \geq 1$ :

565 Calculate  $\Delta^{(i)} = P(G, \mu^{(i)}, \tilde{\Delta}^{(i)})$ .

566 **Else:**

567 **Return**  $\Delta = \mathbf{1}\hat{\mu}^\top - G$  (intended sequence is infeasible).

568 4. **Return**  $\Delta = \tilde{\Delta}^{(n)}$  as approximate solution to  $\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|$ .

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

573

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

579 becomes

$$580 \quad \mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } j \in \cup_{k=1}^i P_k \\ \frac{\mu_j}{\sum_{k=i+1}^n \mu_k} \left(1 - \sum_{k=1}^i \hat{\mu}_k\right), & \text{if } j \notin \cup_{k=1}^i P_k \end{cases}.$$

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

583

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

588

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

$$594 \quad \mu_j^{(i)} = \begin{cases} \hat{\mu}_j, & \text{if } j = i \bmod n \\ \frac{\mu_j^{(i-1)}}{\sum_{k \neq i} \mu_k} (1 - \hat{\mu}_j), & \text{if } j \neq i \bmod n \end{cases},$$

595 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  
596  $1 - \hat{\mu}_{i \bmod n}$  proportionally over the other elements of  $\mu^{(i)}$ . One can repeat the for-  
597 loops until  $\|\mu^{(i)} - \mu^{(i-1)}\| < \phi$ , where  $\phi > 0$  is a given precision. Then, we hope to  
598 reach  $\hat{\mu}$  from  $\mu^{(i)}$  with a single rank-1 perturbation. Note that in contrast to R1SH  
599 only  $\mu_{i \bmod n}^{(i)}$  is fixed to  $\hat{\mu}_{i \bmod n}$  in the  $i$ -th for-loop of FR1SH( $\phi$ ). The time com-  
600 plexity of FR1SH( $\phi$ ) is  $\mathcal{O}(\log_2(\mathcal{C}/\xi)n^3/\phi)$ .

601

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

612 **8.3. Numerical Experiments.** We present in this section experiments for lar-  
613 ger numerical instances. In particular, the tests in Section 8.3.1 make use of randomly  
614 generated dense matrices, whereas Section 8.3.2 performs tests on real-life sparse  
615 matrices. Throughout this section, the  $\infty$ -norm is considered.

616 **8.3.1. Dense Random Matrices.** To test R1SH and its variants for larger  
617 examples, we generated random problem instances of  $\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta_{\geq 0}} \|\Delta\|$ . A random  
618  $G$  is generated by drawing a  $n \times n$  matrix with random values in  $(0, 1)$  and scaling  
619 the rows such that row sums are all one. Similarly, to generate  $\hat{\mu}$ , a random  $n \times 1$   
620 vector  $v$  of random values in  $(0, 1)$  is first generated and scaled so that it sums up  
621 to one, and then a 0.1-fraction of this random vector  $v$  is then mixed with  $\mu$  (of  $G$ )

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

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

630 The following can be observed from the numerical experiments:

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

644 **8.3.2. Sparse Matrices.** The applicability of our heuristics for sparse  $G$  is  
 645 hindered because  $\Delta(G, \hat{\mu})$  is empty for many  $\hat{\mu}$ . As a result, there is limited flexibility  
 646 in jumping to intermediate stationary distributions. More specifically, in a rank-1  
 647 perturbation, the same vector (such as  $z^{\top} = \hat{\mu}(I - G)$ ) is used to modify every row.  
 648 Since the perturbation “transfers” mass within a row, some elements will be positive  
 649 and some negative. In a sparse setting, there are many zeros, meaning that a single  
 650 vector can often only be used for perturbing a single row (or a few rows at most).  
 651 Being able to perturb only one row in a stochastic matrix, it is not hard to imagine  
 652 that the number of reachable stochastic matrices is limited. In other words, finding a  
 653 rank-1 perturbation towards a *specific* stationary distribution (the main focus of this  
 654 paper) is often infeasible. Example 5 demonstrates this for the (sparse) ring network.

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

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

671 and allows the comparison of the two approaches.

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

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

683 To further explore the applicability of rank-1 steps heuristics in sparse networks,  
 684 we will consider the following three real-life networks from three different domains  
 685 with different objectives regarding  $\hat{\mu}$  (see also the applications overview in Section 2):  
 686 **Social network:** A high-school network of student relationships where we aim to  
 687 increase the popularity of a clique of students by 10%. This could potentially enhance  
 688 the group’s cohesion.

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

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

695 The weighted adjacency matrices of all networks are normalized so that they  
 696 are stochastic and we only considered the largest strongly connected component (so  
 697 that the stationary distribution exists). More details about the considered datasets  
 698 can be found in Table 2. For each real-life network, we search for the 25 largest  
 699 cliques (its computation time turns out to be negligible in our examples, probably  
 700 due to sparsity), and for each clique we apply the rank-1 steps heuristics. We did not  
 701 solve the LP with Gurobi for these networks because of the scalability issues of that  
 702 approach (after computing for a relatively long time, it still did not find a candidate  
 703 solution). Table 6 and Table 7 (in the appendix) present the average results over the  
 704 25 cliques for the different real-life networks with a time limit of 60 and 600 seconds,  
 705 respectively. Note that the time limit does not necessarily have to be reached because  
 706 the rank-1 steps heuristics are terminated once a full loop over the nodes did not lead  
 707 to an improvement.

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

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

Name	Description	Goal $\hat{\mu}$	Ref.
<b>Social network</b>	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]
<b>Road network</b>	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]
<b>Org. network</b>	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.

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

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

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

742 **Acknowledgement.** The authors want to express their gratitude to the anony-  
743 mous reviewers for their valuable and constructive suggestions.

- 745 [1] K. Abbas, J. Berkhout and B. Heidergott (2016) A critical account of perturbation analysis of  
746 Markov chains. *Markov Processes and Related Fields* **22** 227–266.
- 747 [2] P. Alquier, N. Friel, R. Everitt and A. Boland (2016) Noisy Monte Carlo: convergence of Markov  
748 chains with approximate transition kernels. *Statistics and Computing* **26** 29–47.
- 749 [3] C. Andrieu and G. Roberts (2009) The pseudo-marginal approach for efficient Monte Carlo  
750 computations. *Annals of Statistics* **37** 697–725.
- 751 [4] V. Anisimov (1988) Estimates for the deviations of the transition characteristics of nonhomogeneous  
752 Markov processes. *Ukrainian Math. J.* **40** 588–592.
- 753 [5] L. Bortolussi and R. Hayden (2013) Bounds on the deviation of discrete-time Markov chains from  
754 their mean-field model. *Performance Evaluation* **70** 736–749.
- 755 [6] J. Breen and S. Kirkland (2015) Stationary vectors of stochastic matrices subject to combinatorial  
756 constraints. *Electronic Journal of Linear Algebra* **28**: 25–53.
- 757 [7] P. Carrington, J. Scott and S. Wasserman (2005) *Models and methods in social network analysis*.  
758 Cambridge University Press.
- 759 [8] H. Caswell (2013) Sensitivity analysis of discrete Markov chains via matrix calculus. *Linear*  
760 *Algebra and its Applications* **438** 1727–1745.
- 761 [9] M. Chekroun, J. Neelin, D. Kondrashov, McWilliams and M. Ghil (2014) Rough parameter  
762 dependence in climate models and the role of Ruelle-Pollicott resonances. *Proceedings of the*  
763 *National Academy of Sciences of the United States of America* **111** 1684–1690.
- 764 [10] G. Cho and C. Meyer (2001) Comparison of perturbation bounds for the stationary distribution  
765 of a Markov chain. *Linear Algebra and its Applications* **335** 137–150.
- 766 [11] E. Crisostomi, S. Kirkland and R. Stephen (2011) A Google-like model of road network dynamics  
767 and its application to regulation and control. *International Journal of Control* **84** 633–651.
- 768 [12] B. Golub and M. Jackson (2010) Naive learning in social networks and the wisdom of crowds.  
769 *American Economic Journal: Microeconomics* **2** 112–49.
- 770 [13] A. Greenbaum, R.C. Li and M. Overton (2020) First-order perturbation theory for eigenvalues  
771 and eigenvectors. *SIAM Review* **62(2)** 463–482.
- 772 [14] D. Gregory, S. Kirkland and N. Pullman (1992) Row-stochastic matrices with a common left  
773 fixed vector. *Linear Algebra and its Applications* **169** 131–149.
- 774 [15] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas (2003) Self-similar community  
775 structure in a network of human interactions. *Physical Review E: Statistical, Nonlinear and*  
776 *Soft Matter Physics* **68(6)** 065103.
- 777 [16] B. Heidergott, A. Hordijk and M. Van Uitert (2007) Series expansions for finite-state Markov  
778 chains. *Probability in the Engineering and Informational Sciences* **21(3)** 381–400.
- 779 [17] B. Heidergott and A. Hordijk (2003) Taylor series expansions for stationary Markov chains.  
780 *Advances in Applied Probability* **35(4)** 1046–1070.
- 781 [18] I. Ipsen and T. Selee (2011) Ergodicity coefficients defined by vector norms. *SIAM Journal on*  
782 *Matrix Analysis and Applications* **32** 153–200.
- 783 [19] M. Jackson (2010) *Social and economic networks*. Princeton University Press.
- 784 [20] C. de Kerchove, L. Ninove and P. Van Dooren (2008) Maximizing PageRank via outlinks. *Linear*  
785 *Algebra and its Applications* **429** 1254–1276.
- 786 [21] S. Kirkland (2002) On a question concerning condition numbers for Markov chain. *SIAM Journal*  
787 *on Matrix Analysis and Applications* **23** 1109–1119.
- 788 [22] KONECT (2016) Adolescent health network dataset – KONECT. [http://konect.uni-koblenz.  
789 de/networks/moreno\\_health](http://konect.uni-koblenz.de/networks/moreno_health).
- 790 [23] KONECT (2017) Euroroads network dataset – KONECT. [http://konect.cc/networks/subelj\\_  
791 euroroad](http://konect.cc/networks/subelj_euroroad).
- 792 [24] KONECT (2017) U. Rovira i Virgili network dataset – KONECT. [http://konect.uni-koblenz.  
793 de/networks/moreno\\_health](http://konect.uni-koblenz.de/networks/moreno_health).
- 794 [25] J. Kunegis (2013) KONECT – The Koblenz Network Collection. *WWW '13 Companion: Pro-*  
795 *ceedings of the 22nd International Conference on World Wide Web*, 1343–1350.
- 796 [26] W. Li, L. Jiang, W.-K. Ching and L.-B. Cui (2013) On perturbation bounds for the joint  
797 stationary distribution of multivariate Markov chain models. *East Asian Journal on Applied*  
798 *Mathematics* **3** 1–17.
- 799 [27] C.D. Meyer (1980) The condition of a finite Markov chain and perturbation bounds for the  
800 limiting probabilities. *SIAM Journal of Algebraic Discrete Methods* **1** 273–283.
- 801 [28] S. Meyn and R. Tweedie (2012) *Markov chains and stochastic stability*. Springer Science &  
802 Business Media.
- 803 [29] A. Mitrophanov, A. Lomsadze and M. Borodovsky (2005) Sensitivity of hidden Markov models.  
804 *Journal of Applied Probability* **42** 632–642.
- 805 [30] A. Mitrophanov (2005) Sensitivity and convergence of uniformly ergodic Markov chains. *Journal*  
806 *of Applied Probability* **42** 1003–1014.

- 807 [31] J. Moody (2001) Peer influence groups: Identifying dense clusters in large networks. *Social*  
808 *Networks*, **23(4)** 261–283.
- 809 [32] Z. Mouhoubi and D. Aissani (2010) New perturbation bounds for denumerable Markov chains.  
810 *Linear Algebra and its Applications* **432** 1627–1649.
- 811 [33] M. Neumann and J. Xu (2003) Improved bounds for a condition number of Markov chains.  
812 *Linear Algebra and its Applications* **386** 225–241.
- 813 [34] R. Pal, A. Datta and E.R. Dougherty (2008) Robust intervention in probabilistic Boolean net-  
814 works. *IEEE Transactions on Signal Processing* **56** 1280–1294.
- 815 [35] P. Schweitzer (1968) Perturbation theory and finite Markov chains. *Journal of Applied Proba-*  
816 *bility* **5** 410–413.
- 817 [36] J. Scott(1988) Social network analysis. *Sociology* **22** 109–127.
- 818 [37] E. Seneta (2001) Sensitivity analysis, ergodicity coefficients, and rank-one updates of finite  
819 Markov chains. In: Stewart, W. J. (ed.) *Numerical Solutions of Markov Chains*, Marcel  
820 Dekker, New York.
- 821 [38] G. Stewart (1973) Error and perturbation bounds for subspaces associated with certain eigen-  
822 value problems. *SIAM Review* **15(4)** 727–764.
- 823 [39] L. Šubelj and M. Bajec (2011) Robust network community detection using balanced propagation.  
824 *The European Physical Journal B*, **81(3)** 353–362.
- 825 [40] O. Szehr (2015) Eigenvalue estimates for the resolvent of a non-normal matrix. *Journal of*  
826 *Spectral Theory* **4** 783–813.
- 827 [41] O. Szehr and M. Wolf (2013) Perturbation bounds for quantum Markov processes and their  
828 fixed points. *Journal of Mathematical Physics* **54** 032203.
- 829 [42] S. Tuteja (2013) Enhancement in weighted PageRank algorithm using VOL. *Journal of Com-*  
830 *puter Engineering*, **14** 135–141.
- 831 [43] N. Tyagi and S. Sharma (2012) Weighted Page Rank algorithm based on number of visits of  
832 links of web page. *International Journal of Soft Computing and Engineering (IJSCE)* **2(3)**  
833 441–446.
- 834 [44] S. Wasserman and K. Faust (1994) *Social network analysis: Methods and applications*. Cam-  
835 bridge University Press.

836

**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

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

This is also the minimum norm solution for arbitrary matrices  $\Delta$  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

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

837 which completes the proof.

838

**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  $\Delta$  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  $\Delta$  simplifies to

841

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

842

with  $z := (\mu - \hat{\mu})^\top(I - G)$ . 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  $z = z_+ + z_-$ ,  $\text{supp}(z_+) \cap \text{supp}(z_-) = \emptyset$  both have non-empty support. The above equations then yield the intervals

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

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

$$846 \quad (\text{A.1}) \quad \hat{\mu}^\top u \geq 1,$$

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

$$849 \quad (\text{A.2}) \quad y = z, \quad x = \frac{u}{\hat{\mu}^\top u} \implies \Delta = \frac{u}{\hat{\mu}^\top u} z^\top.$$

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

853 **A.3. Proof of Theorem 6.2.** It follows from Theorem 6.1 that if  $\hat{\mu}^\top u < 1$ ,  
854 the feasible set is empty, and that if  $\hat{\mu}^\top u = 1$ , the feasible set is a single point  
855  $x = u$ . If  $\hat{\mu}^\top u > 1$ , then the optimization problems merely express that one should  
856 minimize the norm  $\gamma$  of the vector  $x$  over the set of constraints. This is formulated  
857 as a convex optimization problem that is feasible, as was pointed out in Theorem  
858 6.1. The problems listed above can be solved using a descent method, and details are  
859 provided in Section D.

## 860 Appendix B. Queuing Networks.

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

$$867 \quad \Delta = \begin{bmatrix} -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \\ -0.0435 & -0.0978 & 0 & 0.1413 \end{bmatrix},$$

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

$$869 \quad G + \Delta = \begin{bmatrix} 0.7391 & 0.1196 & 0 & 0.1413 \\ 0.3478 & 0.2935 & 0.2174 & 0.1413 \\ -0.0435 & 0.6848 & 0 & 0.3587 \\ -0.0435 & -0.0978 & 0.7826 & 0.3587 \end{bmatrix}.$$

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

872 **B.2. Rank-1 Heuristics.** Consider again the experiment from Section B.1 Ta-  
 873 ble 3 shows the minimum  $\infty$ -norms for different optimization problems and the results  
 874 for the R1SH, R1SH(2), FR1SH( $10^{-3}$ ) and the Riesz projector candidate  $\mathbf{1}\hat{\mu}^\top - G$   
 875 (in short “Riesz”). For comparison, we also show the minimum norms of the opti-  
 876 mization problem with different feasible sets:  $P_1 := \min_{\Delta \in \Delta^{\hat{\mu}}} \|\Delta\|_\infty$  is found with  
 877 Theorem 5.1,  $P_2 := \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|_\infty$  is found by solving the LP from Section 8.1  
 878 using Gurobi 9.1.2 and  $P_3 := \min_{\Delta \in \Delta(G, \hat{\mu})} \|\Delta\|_\infty$  is found by applying the algorithm  
 from Subsection D.3. It follows from the results that R1SH is able to solve all in-

$\hat{\nu}$	$P_1$	$P_2$	$P_3$	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  $\hat{\nu}$

879

880 stances, also those for which  $\Delta(G, \hat{\mu}) = \emptyset$ . This means that it successfully finds a  
 881 sequence of rank-1 perturbations leading to  $\hat{\mu}$ . In contrast, R1SH(2) did not find a  
 882 candidate for  $\hat{\nu} = 0.2$ . Though better than the Riesz projector, R1SH is not that  
 883 successful in finding a candidate near  $P_2 = \min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \|\Delta\|_\infty$  for  $\hat{\nu} \in \{0.2, 2\}$ . For  
 884  $\hat{\nu} \in \{0.2, 2\}$ , FR1SH( $10^{-3}$ ) finds candidates that are significantly better than the can-  
 885 didates found by R1SH. Also R1SH(2) finds a better candidate for  $\hat{\nu} = 2$  compared to  
 886 R1SH, which shows that more subsets (R1SH = R1SH(3)) does not necessarily lead  
 887 to better candidates.

888

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

893

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

895

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

896

898

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

899

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

901

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

902

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

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

$$906 \quad \mathbf{\Delta}^{\text{supp}}(G, \hat{\mu}) := \left\{ \Delta = xz^\top \in \mathbf{\Delta}(G, \hat{\mu}) : x_i = 0 \text{ for } i \in S_0^{\text{supp}} \text{ and} \right. \\
 907 \quad \left. l_i < x_i < u_i \text{ for } i \in S_{\neq 0}^{\text{supp}} \right\}.$$

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

$$911 \quad \bar{x}_i = \begin{cases} 0, & \text{if } i \in S_0^{\text{supp}} \\ u_i, & \text{if } i \notin S_0^{\text{supp}} \end{cases}.$$

912 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 com-  
 913 ponents, and the implied candidate  $G + \Delta$  will have the same support as  $G$  and will  
 914 therefore be irreducible if  $G$  was irreducible.

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

## 919 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:

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

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  $0 \leq x$ ,  $\delta$  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,

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

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

$$922 \quad \hat{u}_1 = u_1 - u_k \geq \hat{u}_2 = u_2 - u_k \geq \dots \geq \hat{u}_k = u_k - u_k = \hat{u}_{k+1} = \dots = \hat{u}_n = 0,$$

923 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  
 924 procedure with a shorter vector of nonzero upper bounds. It is clear that we achieve  
 925 a maximum decrease of  $\gamma$  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

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

926 where we assume  $\hat{\mu}^\top u > 1$ . Even though this is a convex problem that can be solved  
 927 via LMI techniques, the quadratic inequality makes it harder to characterize the  
 928 solution in analytic form. But the solution is unique since the level sets of the 2-norm  
 929 form a strictly convex set. Also, if  $x = \hat{\mu}/(\hat{\mu}^\top \hat{\mu})$  satisfies the constraints  $0 \leq x \leq u$ ,  
 930 then it is the minimum norm solution of our problem since it is already the minimum  
 931 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  $0 \leq x \leq u$ . Moreover, the convex combinations

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

all satisfy  $\hat{\mu}^\top x_c = 1$ . Therefore the largest value of  $c$  for which  $0 \leq x_c \leq u$ , 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  $\infty$ -Norm Rank-1 Perturbation.** The minimum  $\infty$ -norm problem is given by

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

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:

$$u_1 = \dots = u_\ell > u_{\ell+1} \geq \dots \geq u_n,$$

where there are  $\ell$  elements of maximal size. In order to decrease the  $\infty$ -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  $1 \leq i \leq \ell$ , and bound  $\delta$  by  $u_1 - u_{\ell+1}$  so that  $x_i$  for  $1 \leq i \leq \ell$  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

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

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  $\hat{u}_i = u_{\ell+1}$ ,  $1 \leq i \leq \ell$ , keep the other ones unchanged, i.e.,  $\hat{u}_i = u_i$  for  $\ell + 1 \leq i \leq n$ , and change the quantity  $\hat{\mu}^\top u$  to  $\hat{\mu}^\top \hat{u}$ . This yields

$$\hat{u}_1 = \dots = \hat{u}_\ell = \hat{u}_{\ell+1} \geq \dots \geq \hat{u}_n.$$

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

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

$$939 \quad \Delta = \begin{bmatrix} -0.087 & -0.1957 & 0 & 0.2826 \\ -0.087 & -0.1957 & 0 & 0.2826 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

940 *which gives*

$$941 \quad G + \Delta = \begin{bmatrix} 0.6957 & 0.0217 & 0 & 0.2826 \\ 0.3043 & 0.1957 & 0.2174 & 0.2826 \\ 0 & 0.7826 & 0 & 0.2174 \\ 0 & 0 & 0.7826 & 0.2174 \end{bmatrix}$$

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

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

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

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

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

$$956 \quad \alpha^* = \frac{1}{n} \frac{b}{1 - (1 - 2b)} = \frac{1}{2n},$$

957 *where we can choose either node 2 or node  $n$  to shift the mass from. Suppose we*  
 958 *shift mass from the link of node 1 to node 2. This then gives a new stationary weight*  
 959  *$3/(2n)$  for node 1, a weight of  $1/(2n)$  for node 2, and the weight of the rest of the*  
 960 *nodes remains  $1/n$ .*

961 From (7.3), the corresponding perturbation matrix  $\Delta$  can be found as follows:

$$962 \quad \frac{\alpha}{\mu_i + \alpha} = \frac{\alpha^*}{\mu_1 + \alpha^*} = \frac{1}{3},$$

963

$$964 \quad \Delta = \frac{\alpha^*}{\mu_1 + \alpha^*} (e_1 e_1^\top - e_1 e_2^\top) (I - G) = \frac{1}{3} \begin{bmatrix} 3b & -3b & b & \cdots & -b \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

965 which gives

$$966 \quad G + \Delta = \begin{bmatrix} 1-b & 0 & \frac{1}{3}b & 0 & \cdots & \frac{2}{3}b \\ b & 1-2b & b & 0 & \cdots & 0 \\ 0 & b & 1-2b & b & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & b & 1-2b & b \\ b & 0 & \cdots & 0 & b & 1-2b \end{bmatrix}.$$

967 Note that  $\Delta$  is structural.

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

$$970 \quad \alpha^* = \frac{\beta}{(n-1)(1-\gamma+\beta)}.$$

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

$$973 \quad \beta < (n-1)(1-\gamma) < 2\beta.$$

974 This can be seen as follows

$$975 \quad \frac{1-\gamma}{1-\gamma+\beta} > \frac{\beta}{(n-1)(1-\gamma+\beta)}$$

976 being equivalent to

$$977 \quad (n-1)(1-\gamma) > \beta.$$

978 The highest increase node  $i$  can realize is  $\alpha^*$ , which gives the new stationary value

$$979 \quad \hat{\mu}_i = \mu_i + \alpha^* = \frac{2\beta}{(n-1)(1-\gamma+\beta)}.$$

980 This value exceeds  $\mu_1$  if

$$981 \quad \frac{2\beta}{(n-1)(1-\gamma+\beta)} > \frac{1-\gamma}{1-\gamma+\beta},$$

982 which proves the claim.

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

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

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

Method	mean( $\ \Delta\ _\infty$ )	ff	mean run time	mean rank
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0506	0	0.0002	1
$\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \ \Delta\ _\infty$	0.0506	1	6.632	98.4
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.049871	0	0.006601	1
$\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \ \Delta\ _\infty$	0.04974	0.56	611.633	497.929
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.005014	0	0.049828	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0587	0	0.0001	1
$\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \ \Delta\ _\infty$	0.0683	1	3.7114	78.68
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	nan	0	0.0001	nan
R1SH( $K$ ) for $K = 2, 4, 8, 16$	nan	0	$\approx 0.0001$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0287	0	0.0002	1
$\min_{\Delta \in \Delta^{\hat{\mu}} \cap \Delta \geq 0} \ \Delta\ _\infty$	0.0347	1	4.5041	85
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	nan	0	0.0018	nan
R1SH( $K$ ) for $K = 2, 4, 8, 16$	nan	0	$\approx 0.0004$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0008	0	0.3725	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0005	0	0.0661	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0026	0	0.0857	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0008	0	0.3194	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0005	0	0.071	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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
$\min_{\Delta \in \Delta^{\hat{\mu}}} \ \Delta\ _\infty$	0.0026	0	0.0856	1
$\min_{\Delta \in \Delta(G, \hat{\mu})} \ \Delta\ _\infty$	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.