Tight bounds on sparse perturbations of Markov Chains*

Romain Hollanders[1,2], Giacomo Como[3], Jean-Charles Delvenne[2] and Raphaël M. Jungers[2]

***Abstract*— What influence can be exerted by one or a few nodes in the consensus or stationary distribution reached by a global network? We address this general question regarding the sensitivity of the invariant probability distribution of an irreducible row-stochastic matrix to the perturbation of a few of its rows, and provide tight bounds on the ∞-norm of the perturbation that can be computed in polynomial time.**

I. INTRODUCTION

How much can individual nodes of a network affect the properties of the whole network? In this work, we address this general question regarding the sensitivity of the invariant probability distribution π of an irreducible row-stochastic matrix P (i.e., its left dominant eigenvector) to the arbitrary modification of one or several rows.

Such invariant probability distributions arise in many fields of applications related to networks, of which a well known example is the concept of PageRank. First introduced by Brin and Page as a measure of centrality (i.e., importance) of the nodes in a graph, the notion of PageRank has been successfully applied to rank web pages in the well known search engine "Google" [BP98]. The idea is that the importance of a web page should be high if it is referred to by many important pages. Computing the PageRank of a network is rather cheap as it can be achieved by calculating the invariant probability distributions of a Markov chain corresponding to a uniform random walk on this network. In this context, a natural question is whether it is possible to optimize the PageRank of some nodes while only making local changes to the network [dKND08], [IT09], [CJB11]. To state the question otherwise, we ask how robust the PageRank of a node is to the manipulation of a few other nodes. These questions come down to studying the sensitivity of the PageRank vector to sparse (i.e., localized) perturbations of the Markov chain describing the uniform random walk.

A dual interpretation exists in terms of averaging and consensus algorithms [FD10], [FD13]. In the simplest setting, agents starting from an initial opinion replace it at each time step by a weighted average of their neighbors. The averaging weights of agent i are encoded in the i^{th} row of a row-stochastic matrix, and the repeated application of the process drives every agent to a common consensus value, provided that the matrix is irreducible and acyclic. The weight of every agent's initial opinion is given once again by the corresponding entry of the left eigenvector. In this interpretation, the question arises whether a few agents could change their averaging strategy so as to significantly modify the dominant eigenvector, e.g. in order to maximize their impact on the final consensus value or minimize the impact of some other nodes. Other applications of the same problem exist as well, e.g., in the context of interacting particle systems, as described in [CF13].

The general problem we are interested in can be formulated as follows. Let $P \in \mathbb{R}^{n\times n}$ be a row-stochastic matrix whose i^{th} row represents the transition probabilities of a state i of a Markov chain with state space $\mathcal{V}$ and assume that we are given control over a small subset $\mathcal{W} \subset \mathcal{V}$ of the states. This means that we are allowed to freely modify the rows of P corresponding to the states in $\mathcal{W}$. Let $\widetilde{P} := P + Z$ be the row-stochastic matrix obtained from P by applying a perturbation Z on the rows corresponding to $\mathcal{W}$. We are interested in bounding $\|\pi - \widetilde{\pi}\|$, where $\pi = P^T\pi$ and $\widetilde{\pi} = \widetilde{P}^T\widetilde{\pi}$ are the invariant probability distributions of P and $\widetilde{P}$ respectively, and where Z affects only the rows of P that correspond to the states in $\mathcal{W}$. In this general problem formulation, the choice of the norm to use has been deliberately left open for later.

As explained in [CF13], classical perturbation analysis fails to provide a satisfactory answer to the above question. Indeed, the bounds that can be obtained via these approaches are of the form:

$$\|\pi - \widetilde{\pi}\|_p \leq \kappa_P \|P - \widetilde{P}\|_q, \qquad (1)$$

for some $p, q \geq 1$, where κ_P is a condition number that only depends on the original matrix P. These bounds typically blow up when the size of the network in question increases. If these bounds were to be tight, it would mean that a single node of the network is eventually capable on its own to dramatically affect the whole network. This clearly indicates a need for better, tighter bounds. Moreover, it often makes sense in applications to consider not the magnitude of the perturbation, but rather its support (i.e., the set of perturbed entries). Typically, if the network represents agents, and a few of them decide to modify their connections, this might result in a change in the rows corresponding to these agents, not necessarily of small magnitude, but definitely restricted to a small set of entries.

In [CF13], the authors proposed another bound of the form:

$$\|\pi - \widetilde{\pi}\|_1 \leq \theta\left(\tau \cdot \frac{\tau_{\mathcal{W}\to\mathcal{V}}}{\tau_{\mathcal{V}\to\mathcal{W}}}\right), \qquad (2)$$

*This work was supported by an ARC grant from the French Community of Belgium and by the IAP network "Dysco" funded by the office of the Prime Minister of Belgium. The scientific responsibility rests with the authors.

[1]Contact author: romain.hollanders@uclouvain.be
[2]catholic University of Louvain, Belgium
[3]Lund University, Sweden

where θ is a continuous, non-decreasing function, τ is the mixing time of the original Markov chain defined by P, $\tau_{\mathcal{W}\to\mathcal{V}}$ is the expected escape time of $\mathcal{W}$ and $\tau_{\mathcal{V}\to\mathcal{W}}$ is the expected hitting time of the states in $\mathcal{W}$. This bound behaves much better than the ones of the form (1) in terms of scalability. Additionally, it provides interesting insight about the dynamics generated by the perturbation and relates it to physical concepts. For instance, using (2), it is easy to quantify the impact of fixing the escape time of $\mathcal{W}$, which could be interpreted in the context of PageRank as a way to prevent being detected as a link farm by Google. On the other hand, it is not supposed to be tight and it only works to bound the 1-norm of the perturbation. It is indeed not straightforward to adapt the bound to other important norms, like the ∞-norm, using the same kind of techniques.

In this work we consider another approach that aims to bound the ∞-norm of $\pi - \widetilde{\pi}$. In contrast to the bounds above, our bound is algorithmic and has no proper explicit form, hence it does not have the same explicative power as, e.g., (2). On the other hand, it is tight and fairly cheap to compute. Additionally, our algorithm provides the optimal perturbation that meets the bound. We also explore how fixing the escape time of $\mathcal{W}$ can be included into the picture.

II. A POLYNOMIAL TIME ALGORITHM FOR COMPUTING THE MAXIMUM PERTURBATION OF THE INFINITE NORM

To bound the ∞-norm of $\pi-\widetilde{\pi}$, our approach is to compute the minimum and the maximum value of $\widetilde{\pi}$ for every state v and then look at the maximum difference with π which is achieved for some v.

We first reformulate the problem in terms of Markov chains. The invariant probability distributions $\pi = P^T\pi$ quantifies the stationary frequency of visit of every state of a Markov chain when following a random walk with the transition probabilities in P. For some state v, π_v can also be seen as the inverse of the *first return time* λ_v of v, i.e., the expected number of steps between two visits of v. Another interesting measure is the *first hitting time* to v from the other states of the Markov chain, i.e., the expected number of steps that the other states need to reach v for the first time. We will denote the vector of first hitting times to v by φ^v. The first hitting time to v of state u is the weighted sum of the first hitting times of its neighbors plus 1 (for performing a step) and hence, φ^v can be computed by solving the following linear system:

$$\varphi^v = P\varphi^v + 1, \quad \text{with} \quad \varphi^v_v = 0.$$

Then the first return time of v is obtained from:

$$\lambda_v = P_{v,:}\,\varphi^v + 1,$$

where we used standard Matlab notations, and we have $\pi_v = 1/\lambda_v$. Hence, to maximize (resp. minimize) π_v, one should minimize (resp. maximize) λ_v.

Coming back to our original problem, we should choose a probability distribution on the outgoing edges of the states in $\mathcal{W}$ in order to maximize or minimize λ_v for all v. This is a similar approach to the one described in [CJB11] to optimize the PageRank of a node in a network. Furthermore, since the decision states are grouped into a set $\mathcal{W}$, it is straightforward to see that optimizing the first return time of v comes back to optimizing the first hitting time to v of the states in $\mathcal{W}$.

Minimizing the first hitting time to v of any control state, and hence maximizing its invariant probability, is easy: the control state should simply jump towards v with probability 1. As for minimizing $\widetilde{\pi}_v$, it is easy to see that all the states of $\mathcal{W}$ should stay in $\mathcal{W}$ with probability 1 in order to maximize their distance to v. However, if no regulation is applied to $\widetilde{P}$, this irremediably creates a reducible matrix. To prevent this, a natural thing to do is to impose a fixed escape time T for the states of $\mathcal{W}$. It turns out that a solution of the resulting problem can still be found in polynomial time. Indeed in that case, it can be shown using a convexity argument that all the control states of $\mathcal{W}$ should remain in $\mathcal{W}$ at the next step with some probability $p(T) < 1$ and go "all in" to some unique other state $u \notin \mathcal{W}$ with probability $1 - p(T)$.

At first glance, it could be tempting to choose u as the state with the highest first hitting time to v in the Markov chain defined by P, i.e., the state which is originally the furthest away from v. This is unfortunately wrong in general as the example of Figure 1 illustrates. Therefore, finding u a priori requires to scan every candidate state and pick the one that minimizes $\widetilde{\pi}_v$.

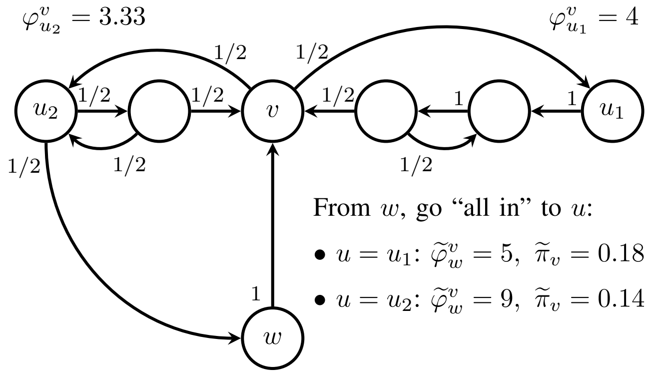

Fig. 1. An example showing that pointing to the node maximizing the first hitting time to v may be suboptimal to minimize $\widetilde{\pi}_v$. Here $\varphi^v_{u_2} = 3.33 < 4 = \varphi^v_{u_1}$. However, in order to minimize $\widetilde{\pi}_v$, the unique outgoing edge of w should point to u_2 instead of u_1. (the pictured Markov chain is the unperturbed P).

Algorithm 1 MAXIMIZATION OF $\|\pi - \widetilde{\pi}\|_\infty$

Ensure: The maximum value of $\|\pi - \widetilde{\pi}\|_\infty$.

1: Compute $\pi = P^T\pi,\ \mathbb{1}^T\pi = 1$.
2: For all states v, compute $\widetilde{\pi}^{\max}_v$ (known perturbation).
3: For all states v, compute $\widetilde{\pi}^{\min}_v$ (test $\sim|\mathcal{V}|$ perturbations).
4: **return** $\|\pi - \widetilde{\pi}\|_\infty = \max_v \max\{\widetilde{\pi}^{\max}_v - \pi_v, \pi_v - \widetilde{\pi}^{\min}_v\}$.

Based on the above observation, we can now propose an algorithm to maximize $\|\pi - \widetilde{\pi}\|_\infty$. The main steps are described in Algorithm 1. Note that the minimization of $\widetilde{\pi}$ requires looping through every state $u \in \mathcal{V}\backslash\mathcal{W}$, each time computing $\widetilde{\pi}$ for the Markov chain in which the control states point towards themselves with probability $p(T)$ and towards u with probability $1 - p(T)$. The number of times that we need to compute an invariant probability distribution is polynomial, hence the result stated in Theorem 1.

Theorem 1: It is possible to compute an exact upper bound on $\|\pi - \widetilde{\pi}\|_\infty$ in polynomial time.

III. CONCLUSION AND PERSPECTIVES

In this work, we have chosen an algorithmic approach to compute upper bounds on the maximal perturbation of the invariant probability distribution of a Markov chain for the ∞-norm in polynomial time. In future works, we would like to investigate whether the same approach could be applied to the 1-norm. In this case, we would like to quantify empirically the improvement of our tight bound against the one given by (2).

REFERENCES

[BP98] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

[CF13] G. Como and F. Fagnani. Robustness of large-scale stochastic matrices to localized perturbations. *arXiv preprint arXiv:1309.0158*, 2013.

[CJB11] B. C. Csáji, R. M. Jungers, and V. D. Blondel. PageRank Optimization by Edge Selection. *Discrete Applied Mathematics*, 2011.

[dKND08] C. de Kerchove, L. Ninove, and P. Van Dooren. Maximizing PageRank via outlinks. *Linear Algebra and its Applications*, 429:1254-1276, 2008.

[FD10] F. Fagnani and J.-C. Delvenne. Democracy in markov chains and its preservation under local perturbations. In *proceedings of the 49th IEEE Conference on Decision and Control (CDC)*, pages 6620–6625, 2010.

[FD13] F. Fagnani and J.-C. Delvenne. The robustness of democratic consensus. *submitted*, 2013.

[IT09] H. Ishii and R. Tempo. Computing the PageRank Variation for Fragile Web Data. *SICE Journal of Control, Measurement, and System Integration*, 2(1):1-9, 2009.