Robustness of large-scale stochastic matrices to localized perturbations

Giacomo Como and Fabio Fagnani

Abstract— **Many linear dynamics over networks can be related by duality to the evolution of a Markov chain with state space coinciding with the node set of the network. Examples include opinion dynamics over social networks as well as distributed averaging algorithms for estimation or control. When the transition probability matrix P associated to the Markov chain is irreducible, a key quantity is its invariant probability distribution $\pi = P'\pi$. In this work, we study how π is affected by, possibly non-reversible or non-irreducible, perturbations of P. In particular, we are interested in perturbations which are localized on a small fraction of nodes but are not necessarily small in any induced norm. While classical perturbation results based on matrix analysis can not be applied in this context, we present various bounds on the effect on π of changes of P obtained using coupling and other probabilistic techniques. Such results allow one to find sufficient conditions for the l_1-distance between π and its perturbed version to vanish in the large-scale limit, depending on the mixing time and one additional local property of the original chain P.**

Index Terms— **Robustness, resilience, large-scale networks, consensus, network centrality, stochastic matrices, stationary probability distributions.**

I. INTRODUCTION

How much can the invariant probability distribution π of an irreducible row-stochastic matrix P be affected by perturbations localized on a relatively small subset $\mathcal{W}$ of its state space $\mathcal{V}$? Such a question arises in an increasing number of applications, most notably in the emerging field of large-scale networks.

As an example, many notions of network centrality can be formulated in terms of invariant probability distributions of suitably defined stochastic matrices. In particular, Google's PageRank algorithm [1] assigns to webpages values corresponding to the entries of the invariant probability distribution π of the matrix P obtained as a convex combination of the normalized adjacency matrix of the directed graph describing the hyperlink structure of the World Wide Web (WWW), and of a matrix whose all entries equal the inverse of the total number of webpages [2], [3]. A well-known problem in this context is rank-manipulation, i.e., the intentional addition or removal of hyperlinks from some webpages (hence, the alteration of the corresponding rows of P) with the goal of modifying the PageRank vector [4], [5], [6]. A natural question is then, to what extent a small subset

$\mathcal{W}$ of webpages can alter the PageRank vector π. Similar robustness issues have been raised for accidental variations of the WWW topology occurring, e.g., because of server failures or network congestion problems [7].

More generally, the problem is of central interest in the context of distributed averaging and consensus algorithms [8]. There, linear systems of the form $x(t + 1) = Px(t)$, or their continuous-time analogues, are studied, e.g., as algorithms for distributed optimization [9], [10], control [11], [12], synchronization in sensor networks [13], or reputation management in ad-hoc networks [14], as well as behavioral models for flocking phenomena [15], [16], or opinion dynamics in social networks [17], [18], [19], [20]. Equilibria of such systems are consensus vectors, i.e., multiples of the all-one vector, and standard results following from Perron-Frobenius theory guarantee convergence (with the additional assumption of aperiodicity of P, in the discrete time case) to a consensus vector with all entries equal to $\pi'x(0)$. Depending on the specific applicative context, the natural question is to what extent the consensus value $\pi'x(0)$ is affected by perturbations of P corresponding, e.g., to malfunctioning of a small fraction of the sensors, or conservative/influential minorities in social networks [21].

Other applications can be found in the context of interacting particle systems [22], [23]. In particular, in the voter model on a finite graph [24], [25], [26, Ch. 14], [27, Ch. 6.9], the probability distribution of the final consensus value is determined by the invariant distribution of the stochastic matrix associated to the simple random walk on the graph. Perturbations in this case may model the presence of inhomogeneities or 'zealots' [28], [29].

The above-described problems all boil down to estimating the distance between the invariant probability distribution π of an irreducible stochastic matrix P and an invariant probability distribution $\tilde{\pi} = \tilde{P}'\tilde{\pi}$ of another stochastic matrix $\tilde{P}$, to be interpreted as a perturbed version of P. In some applications, P may be reversible, i.e., coincide with the normalization of a symmetric positive matrix, so that π can be easily computed in terms of the latter. However, even in these cases, the considered perturbations will typically be such that $\tilde{P}$ is not reversible and thus $\tilde{\pi}$ does not allow for a tractable explicit expression.

Remarkably, standard perturbation results based on sensitivity analysis [30]–[31], [21] do not provide a satisfactory answer to this problem. Indeed, they provide upper bounds of the form

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

for some $p, q \in [1, \infty]$, where κ_P is a condition number

depending on the original stochastic matrix P only. Such condition numbers are lower bounded by an absolute positive constant (e.g., $1/4$ for the smallest of those surveyed in [32]) and typically blow up as the state space $\mathcal{V}$ grows large. Therefore, such results do not allow one to prove that the distance $||\tilde{\pi} - \pi||_p$ vanishes in the limit of large network size, even if P and $\tilde{P}$ differ only in a single row, unless $||\tilde{P} - P||_q$ itself vanishes.

In this paper,[1] we obtain upper bounds on the total variation distance $||\tilde{\pi} - \pi|| := \frac{1}{2}||\tilde{\pi} - \pi||_1$ of the form

$$||\tilde{\pi} - \pi|| \leq \theta \left(\tau \frac{\tilde{\chi}}{\tau_{\mathcal{W}}^*} \right), \qquad (2)$$

(see Theorem 3) where: $\theta : [0, +\infty) \to [0, 1]$ is a continuous, nondecreasing function such that $\theta(0) = 0$ (see (6) for its definition);

$$\tau := \inf \left\{ t \geq 1 : ||P_{u,\cdot}^t - P_{v,\cdot}^t|| \leq 1/e \,, \forall u, v \in \mathcal{V} \right\} \quad (3)$$

is the mixing time of the original stochastic matrix P; $\tau_{\mathcal{W}}^*$ denotes the minimal expected hitting time on the set $\mathcal{W}$ for a Markov chain with transition probability matrix P (see (7)); and $\tilde{\chi}$ stands for the escape time from $\mathcal{W}$ for a Markov chain with transition probability matrix $\tilde{P}$ (see (8) for the exact definition). As opposed to the aforementioned sensitivity results, all derived from algebraic arguments, our proofs rely on coupling techniques, combined with an argument similar to the one developed in [20] for 'highly fluid' networks. Clearly, (2) implies that $||\tilde{\pi} - \pi||$ vanishes provided that $\tau \tilde{\chi} / \tau_{\mathcal{W}}^*$ does. As we will show, this finds immediate application in the PageRank manipulation problem. More in general, our results prove useful in many of those aforementioned large-scale network applications where classical sensitivity-based results fail to provide a satisfactory answer.

Mixing properties of stochastic matrices have been the object of extensive recent research [26], [34], [35], and several results are available allowing one to estimate the mixing time τ of a stochastic matrix P, e.g., in terms of the conductance or other geometrical properties of the graph associated to P. It is worth pointing out that a connection between mixing properties and robustness of stochastic matrices is already unveiled by the perturbation results of [36], [31], where (1) is proven for $p = 1$, $q = \infty$, and condition number κ_P proportional to τ. Of a similar flavor are Seneta's results [37], [38] estimating the condition number κ_P in terms of ergodicity coefficients. Also the estimates proposed in [21] for symmetric P can be rewritten as (1) with for $p = q = 2$ and κ_P equal to the inverse of the spectral gap of P. As compared to these references, the fundamental novelty of our bound (2) consists in measuring the size of the perturbation in terms of the ratio $\tilde{\chi} / \tau_{\mathcal{W}}^*$ instead of the distance $||\tilde{P} - P||_q$, thus enabling one to obtain significant results in scenarios where $\mathcal{W}$ is small but $\tilde{P} - P$ is not necessarily small in any norm.

In fact, of the last two parameters appearing in the righthand side of (2), the escape time $\tilde{\chi}$ is the only one truly depending on the perturbation $\tilde{P} - P$, and is indeed easily estimated in typical cases when $\mathcal{W}$ is a small subset of $\mathcal{V}$. On the other hand, the minimal hitting time $\tau_{\mathcal{W}}^*$, which depends on P and $\mathcal{W}$ only, turns out to be the hardest to get lower bounds on in typical applications where P is sparse and $\mathcal{W}$ remains small but not necessary localized as the state space grows large. While Kac's formula ([35, Lemma 21.13]) readily implies the upper bound $\tau_{\mathcal{W}}^* \leq 1/\pi(\mathcal{W})$, lower bounds on $\tau_{\mathcal{W}}^*$ typically involve finer details of P than just $\pi(\mathcal{W})$. In the last section of this paper, we will propose an analysis of $\tau_{\mathcal{W}}^*$ for networks with high local connectivity, which finds natural application when the graph associated to P is a d-dimensional grid, and the size of $\mathcal{W}$ remains bounded (or grows very slowly) as the network size grows large. Results for random, locally tree-like networks will be the object of a forthcoming work.

The rest of this paper is organized as follows. Section II introduces three motivating examples formalizing some of the applications mentioned at the beginning of this Introduction. In Section III, we present our main result which is stated as Theorem 3 and apply it in four examples. We refer the reader to [33] for applications to stochastic matrices whose support graph has high local connectivity such as grid-like graphs.

Before proceeding, let us collect here some notational conventions to be used throughout the paper. Vectors and matrices will be considered with entries from a set $\mathcal{V}$ of finite cardinality $n := |\mathcal{V}|$. The all-one column vector will be denoted by $\mathbb{1}$. For a matrix A, A' will stand for its transpose and $\text{supp}(A) := \{v : A_{v,\cdot} \neq 0\}$ for the set of its nonzero rows. Then, a probability distribution μ (i.e., a nonnegative-valued vector such that $\mu' \mathbb{1} = 1$) will be said invariant for a stochastic matrix A (i.e., a nonnegative-valued matrix such that $A\mathbb{1} = \mathbb{1}$) if $A'\mu = \mu$. The total variation distance between two probability distributions will be denoted by $||\mu - \pi|| := \frac{1}{2} \sum_v |\mu_v - \pi_v|$. For a probability distribution μ and a subset $\mathcal{A} \subseteq \mathcal{V}$ such that $\mu(\mathcal{A}) > 0$, $\mu^{\mathcal{A}}$ will stand for the conditional probability distribution on $\mathcal{A}$, i.e., $\mu_a^{\mathcal{A}} = \mu_a/\mu(\mathcal{A})$ for $a \in \mathcal{A}$, and $\mu_v^{\mathcal{A}} = 0$ for $v \in \mathcal{V} \setminus \mathcal{A}$. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ we shall use the convention that $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, so that $\mathcal{G}$ undirected means that if $(u, v) \in \mathcal{E}$ then $(v, u) \in \mathcal{E}$ as well. To every stochastic matrix P we shall associate the support graph $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}_P)$ where $(u, v) \in \mathcal{E}_P$ if and only if $P_{uv} > 0$. For stochastic matrices $P, \tilde{P}$, we will consider discrete-time Markov chains $V(t)$ and $\tilde{V}(t)$, $t = 0, 1, \dots$, with state space $\mathcal{V}$ and transition probability matrix P, and $\tilde{P}$, respectively. For $v \in \mathcal{V}$, $\mathbb{P}_v$ and $\mathbb{E}_v$ will stand for the probability and expectation conditioned on $V(0) = \tilde{V}(0) = v$, while for a probability distribution μ, $\mathbb{P}_\mu := \sum_v \mu_v \mathbb{P}_v$ and $\mathbb{E}_\mu := \sum_v \mu_v \mathbb{E}_v$. We will denote the corresponding hitting times on a subset $\mathcal{U} \subseteq \mathcal{V}$ by $T_{\mathcal{U}} := \inf\{t \geq 0 : V(t) \in \mathcal{U}\}$, and $\tilde{T}_{\mathcal{U}} := \inf\{t \geq 0 : \tilde{V}(t) \in \mathcal{U}\}$, and their expectations by $\tau_{\mathcal{U}}^v := \mathbb{E}_v[T_{\mathcal{U}}]$ and $\tilde{\tau}_{\mathcal{U}}^v := \mathbb{E}_v[\tilde{T}_{\mathcal{U}}]$, respectively.

---

[1]The present is a short version of the manuscript [33] available on the arxiv and currently under review for journal publication. We will refer to [33] for the proofs of some of the statements.

II. THREE MOTIVATING EXAMPLES

In this section we present three motivating examples formalizing some of the application problems discussed in the Introduction.

A. PageRank manipulation

Let Q be a stochastic matrix, μ a probability distribution, and β a parameter in the interval $(0,1)$. Let $P := (1 - \beta)Q + \beta \mathbb{1}\mu'$, and observe that, irrespective of whether Q is reducible or not, the matrix $W := (I - (1 - \beta)Q')$ is strictly diagonally dominant, hence nonsingular, so that P has a unique invariant probability distribution $\pi = \beta W^{-1}\mu$.

Now, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the directed graph describing the WWW, whose nodes $v \in \mathcal{V}$ correspond to webpages and where there is a directed edge $(u, v) \in \mathcal{E}$ whenever page u has a hyperlink directed to page v. Let $d_u := |\mathcal{E}_u|$ and $\mathcal{E}_u := \{v : (u, v) \in \mathcal{E}\}$ be the number of hyperlinks and, respectively, the set of linked pages, from page u. Define the stochastic matrix Q by $Q_{uv} = 1/n$ for all v if $d_u = 0$, and, if $d_u \geq 1$, let $Q_{uv} = 0$ if $(u, v) \notin \mathcal{E}$ and $Q_{uv} = 1/d_u$ if $(u, v) \in \mathcal{E}$. Also, let μ be the uniform distribution over the set of webpages. Then, $\pi = (1 - \beta)Q'\pi + \beta\mu$ is the PageRank vector, first introduced by Brin and Page [1] to measure the relative importance of webpages. Typical values of β used in practice are about 0.15. For general probability distribution μ, the vector π is referred to as the personalized PageRank [39], and is used in context-sensitive searches.

Now, let $\mathcal{W} \subseteq \mathcal{V}$ be a (relatively small) set of webpages, and assume that the hyperlinks $\cup_{w \in \mathcal{W}}\mathcal{E}_w$ can be modified arbitrarily in order to change π. Let $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$ be the modified WWW graph, $\tilde{Q}$ the corresponding stochastic matrix. Let $\tilde{\pi}$ be the unique invariant probability distribution of $\tilde{P} := (1 - \beta)\tilde{Q} + \beta \mathbb{1}\mu'$. Then, one has that

$$||\tilde{\pi} - \pi|| = \max_{\mathcal{U} \subseteq \mathcal{V}} \{\tilde{\pi}(\mathcal{U}) - \pi(\mathcal{U})\} .$$

Hence, estimating the impact that the arbitrary change of the hyperlinks from a set of webpages $\mathcal{W}$ has on the aggregate PageRank of an arbitrary set of webpages $\mathcal{U}$ boils down to bounding the total variation distance between the invariant probability distributions π and $\tilde{\pi}$. Observe that the matrices Q and $\tilde{Q}$, and therefore P and $\tilde{P}$, differ only on the rows indexed by elements of $\mathcal{W}$. A solution to this problem will be discussed in Example II-A of Section III.

B. Faulty communication links in distributed averaging

Consider a sensor network described as a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, whose nodes and edges represent sensors and two-way communication links, respectively. Assume that each sensor v initially measures a scalar y_v and the goal is to design a distributed algorithm for the computation of the arithmetic average $\overline{y} := n^{-1}\sum_v y_v$.

A possible solution [8] is as follows. Let $d \in \mathbb{R}^{\mathcal{V}}$ be the degree vector in $\mathcal{G}$, and, for all $v \in \mathcal{V}$, put

$$x_v(0) = \frac{y_v}{d_v}, \qquad z_v(0) = \frac{1}{d_v}, \qquad (4)$$

$$
\begin{aligned}
[x_v(t+1), z_v(t+1)] &= \tfrac{1}{2}[x_v(t), z_v(t)] \\
&+ \tfrac{1}{2d_v}\sum_{u:(u,v)\in\mathcal{E}}[x_u(t), z_u(t)] .
\end{aligned}
\tag{5}
$$

What makes the above particularly appealing in large-scale network applications is the fact that it requires sensors to exchange information with their neighbors in $\mathcal{G}$ only, and that each sensor v needs to know its degree d_v only with no need for global knowledge about the network structure or size.

In order to analyze the algorithm let us rewrite (4) and (5) in matrix notation. Let P be the stochastic matrix associated to the lazy random walk on $\mathcal{G}$, i.e., $P = (I + Q)/2$, where I denotes the identity matrix and $Q_{uv} = 1/d_u$ if $(u, v) \in \mathcal{E}$. Let $x(0) = y/d$, $z(0) = \mathbb{1}/d$ (where division between two vectors is meant componentwise) and consider the iteration

$$x(t+1) = Px(t), \qquad z(t+1) = Pz(t) .$$

Observe that the unique invariant probability distribution $\pi = P'\pi$ is given by $\pi_u = d_u/(n\overline{d})$ where $\overline{d} := n^{-1}\sum_v d_v$ is the average degree. Moreover, irreducibility and acyclicity of P imply that

$$x(t) = P^t\frac{y}{d} \xrightarrow{t\to\infty} \mathbb{1}\pi'\frac{y}{d} = \mathbb{1}\frac{\overline{y}}{\overline{d}}, \quad z(t) = P^t\mathbb{1}/d \xrightarrow{t\to\infty} \mathbb{1}\frac{1}{\overline{d}},$$

so that

$$\frac{x_v(t)}{z_v(t)} \xrightarrow{t\to\infty} \overline{y}, \qquad \forall v \in \mathcal{V},$$

i.e., (4)-(5) effectively describe an iterative distributed algorithm for the computation of $\overline{y}$. The example can be easily generalized starting from an undirected weighted graph, thus preserving reversibility of P and an explicit form of the invariant distribution π.

Let $\mathcal{F} \subseteq \mathcal{E}$ be a subset of directed communication links which stop working. Let $\tilde{\mathcal{E}} := \mathcal{E} \setminus \mathcal{F}$, $\tilde{\mathcal{G}} := (\mathcal{V}, \tilde{\mathcal{E}})$, and $\tilde{d}$ be the vector of in-degrees in $\tilde{\mathcal{G}}$. Let $\tilde{P} = (I + \tilde{Q})/2$, where $\tilde{Q}$ is a stochastic matrix with $\tilde{Q}_{uv} = 1/\tilde{d}_u$ if $(v, u) \in \tilde{\mathcal{E}}$. Consider the analogous of (4) and (5) with d_v and $\mathcal{E}$ replaced by $\tilde{d}_v$ and $\tilde{\mathcal{E}}$, i.e., $\tilde{x}(0) = y/\tilde{d}$, $\tilde{z}(0) = \mathbb{1}/\tilde{d}$, $\tilde{x}(t+1) = \tilde{P}x(t)$, and $\tilde{z}(t+1) = \tilde{P}\tilde{z}(t)$. Then, provided that $\tilde{\mathcal{G}}$ remains strongly connected, arguing as before shows that

$$\frac{\tilde{x}_v(t)}{\tilde{z}_v(t)} \xrightarrow{t\to\infty} \frac{\tilde{\pi}'y/\tilde{d}}{\tilde{\pi}'\mathbb{1}/\tilde{d}} = \frac{\overline{y} + \varepsilon_1 + \varepsilon_2}{1 + \varepsilon_3 + \varepsilon_4}, \qquad \forall v \in \mathcal{V},$$

where $\tilde{\pi} = \tilde{P}'\tilde{\pi}$ is the unique invariant probability distribution of $\tilde{P}$ and

$$\varepsilon_1 := \frac{1}{n}\sum_v \left(\frac{d_v}{\tilde{d}_v} - 1\right)y_v, \qquad \varepsilon_2 := \overline{d}\sum_v (\tilde{\pi}_v - \pi_v)\frac{y_v}{\tilde{d}_v},$$

$$\varepsilon_3 := \frac{1}{n}\sum_v \left(\frac{d_v}{\tilde{d}_v} - 1\right), \qquad \varepsilon_4 := \overline{d}\sum_v (\tilde{\pi}_v - \pi_v)\frac{1}{\tilde{d}_v}.$$

Observe that $|\varepsilon_1| \leq \frac{|\mathcal{F}|}{n}||y||_\infty, |\varepsilon_2| \leq \overline{d}||y||_\infty||\tilde{\pi} - \pi||$, while $|\varepsilon_3| \leq \frac{|\mathcal{F}|}{n}$ and $|\varepsilon_4| \leq \overline{d}||\tilde{\pi} - \pi||$. Hence, provided that $|\mathcal{F}| = o(n)$, and that the average degree $\overline{d}$ and $||y||_\infty$ remain bounded as n grows large, a sufficient condition for $\tilde{y} = y + o(1)$ is that $||\tilde{\pi} - \pi|| = o(1)$.

C. Voter model with influential agents

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected undirected graph (with no self-loops). For $u \neq v \in \mathcal{V}$, let $E^{(u,v)} \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ have all entries equal to zero but for $E^{(u,v)}_{u,v} = -E^{(u,v)}_{u,u} = 1$. Consider the following Markov chain $X(t)$ over $\{0,1\}^{\mathcal{V}}$: given $X(t)$, $X(t+1) = (I + E^{(u,v)})X(t)$ with probability $1/|\mathcal{E}|$, for all $(u,v) \in \mathcal{E}$. This is an instance of the voter model [22], [23], [24], [25]. In a social network interpretation, this may be thought of modeling a society where every pair of individuals whose corresponding nodes are neighbors in $\mathcal{G}$ have the same chance to influence each other.

It is standard result that with probability one $X_v(t) \overset{t \to \infty}{\Longrightarrow} Y$ for all v, where Y is a $\{0,1\}$-valued random variable. Moreover, it is not hard to see that

$$P := I + \frac{1}{|\mathcal{E}|} \sum_{(u,v) \in \mathcal{E}} E^{(u,v)}$$

is primitive and symmetric, so that $\mathbb{E}[X(t)|X(0)] = P^t X(0) \overset{t \to \infty}{\Longrightarrow} \mathbb{1}\pi' X(0)$, where $\pi = P'\pi$ is the uniform distribution over $\mathcal{V}$. In particular, this implies that

$$y := \mathbb{P}(Y = 1|X(0)) = \frac{1}{n} \sum_v X_v(0).$$

In the statistical physics jargon, the fact that the uniform distribution is invariant for P, so that $\sum_v \mathbb{E}[X_v(t)|X(0)]$ remains constant in t, is referred to as conservation of the average magnetization [40].

Now, let us consider the following variant to the model. Let $\mathcal{F} \subseteq \mathcal{E}$ be such that the directed graph $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$, where $\tilde{\mathcal{E}} := \mathcal{E} \setminus \mathcal{F}$ remains strongly connected, and consider the Markov chain $\tilde{X}(t)$ over $\{0,1\}^{\mathcal{V}}$ such that given $\tilde{X}(t)$, $\tilde{X}(t+1) = (I + E^{(u,v)})\tilde{X}(t)$ with probability $|\mathcal{E}|^{-1}$, for all $(u,v) \in \tilde{\mathcal{E}}$, and $\tilde{X}(t+1) = \tilde{X}(t)$ with probability $|\mathcal{F}|/|\mathcal{E}|$. The social network interpretation is that $\mathcal{W} := \{u : (u,v) \in \mathcal{F} \text{ for some } v\}$ is a set of influential individuals, whose interactions with some of their neighbors in $\mathcal{G}$ are asymmetric, as they influence such neighbors without being influenced in turn from them. A similar model is discussed in [21] in the framework of continuous opinion dynamics. Observe that strong connectivity of $\tilde{\mathcal{G}}$ implies that, with probability one $\tilde{X}_v(t) \overset{t \to \infty}{\Longrightarrow} \tilde{Y}$ for all v, where $\tilde{Y} \in \{0,1\}$ is a random variable such that

$$\tilde{y} := \mathbb{P}(\tilde{Y} = 1|\tilde{X}(0)) = \tilde{\pi}' \tilde{X}(0),$$

where $\tilde{\pi} = \tilde{P}'\tilde{\pi}$ is the unique invariant probability distribution of

$$\tilde{P} := I + \frac{1}{|\mathcal{E}|} \sum_{(u,v) \in \tilde{\mathcal{E}}} E^{(u,v)}.$$

Clearly, if the initial conditions of the two processes coincide, i.e., if $\tilde{X}(0) = X(0)$, then

$$|\tilde{y} - y| \leq ||\tilde{\pi} - \pi||,$$

with equality for at least one value of $\tilde{X}(0) = X(0) \in \{0,1\}^{\mathcal{V}}$. Will $|\tilde{y} - y|$ vanish as n grows large if the set of influential agents $\mathcal{W}$ (and hence $\mathcal{F}$) remains small?
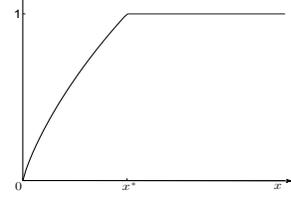

Fig. 1. Graph of the function $\theta(x)$ defined in (6).

III. PERTURBATION RESULTS

Let P be an irreducible stochastic matrix on the finite state space $\mathcal{V}$ and let $\pi = P'\pi$ be its unique invariant probability distribution. Let $\tilde{P}$ be another stochastic matrix (not necessarily irreducible) on the same state space $\mathcal{V}$, to be interpreted as a perturbation of P, and let $\tilde{\pi}$ be an invariant probability distribution of $\tilde{P}$ (not necessarily the unique one).

The following result provides an upper bound on the total variation distance between π and $\tilde{\pi}$. It is stated in terms of the function $\theta : [0, +\infty) \to [0, 1]$

$$\theta(x) := \begin{cases} x \ln\left(e^2/x\right) & x \leq x^* \\ 1 & x \geq x^*, \end{cases} \qquad (6)$$

where $x^* = 0.31784\ldots$ is the smallest positive solution of $e^2/x = \exp(1/x)$.

Lemma 1: Let P and $\tilde{P}$ be stochastic matrices on a finite set $\mathcal{V}$. Let P be irreducible with invariant probability measure π and mixing time τ (3), and $\tilde{\pi}$ be an invariant probability measure for $\tilde{P}$. Then,

$$||\tilde{\pi} - \pi|| \leq \theta(\tau \tilde{\pi}(\mathcal{W})),$$

for all $\mathcal{W} \subseteq \mathcal{V}$ such that $\mathcal{W} \supseteq \text{supp}(P - \tilde{P})$.

Proof: See [33]. ∎

Lemma 1 shows that it is sufficient to have an upper bound on the product $\tau \tilde{\pi}(\mathcal{W})$ in order to obtain an upper bound on $||\tilde{\pi} - \pi||$. In particular, assuming that an upper bound on the mixing time τ is available, e.g., from an estimate of the conductance of P, one is left with estimating $\tilde{\pi}(\mathcal{W})$. Observe that $\tilde{\pi}(\mathcal{W})$ is typically unknown in the applications. Below, we derive an upper bound on $\tilde{\pi}(\mathcal{W})$ in terms of two quantities.

The first quantity we need to introduce is the minimal hitting time

$$\tau^*_{\mathcal{W}} := \min\{\tau^v_{\mathcal{W}} : v \in \mathcal{V} \setminus \mathcal{W}\}. \qquad (7)$$

Observe that the minimal hitting time $\tau^*_{\mathcal{W}}$ only depends on the choice of the subset $\mathcal{W} \supseteq \text{supp}(\tilde{P} - P)$ and on the original matrix P (in particular, on the rows of P indexed by $v \notin \mathcal{W}$), but not on finer details of the perturbation $\tilde{P} - P$.

The second quantity we shall need is the escape time from $\mathcal{W}$ with respect to $\tilde{P}$ and $\tilde{\pi}$, defined as

$$\tilde{\chi} := \max_{w : \tilde{\pi}_w > 0} \inf_{t \geq 1} \frac{t}{\mathbb{P}_w(\tilde{T}_{\mathcal{V} \setminus \mathcal{W}} \leq t)}. \qquad (8)$$

Notice that the escape time $\tilde{\chi}$ depends only on the rows of the perturbed matrix $\tilde{P}$ whose index lies in the set $\mathcal{W}$

(because so does the distribution of $\tilde{T}_{\mathcal{V}\setminus\mathcal{W}}$) and, when $\tilde{P}$ is not irreducible, on the choice of the invariant measure $\tilde{\pi}$. In particular, $\tilde{\chi} = +\infty$ if and only if the set $\mathcal{V}\setminus\mathcal{W}$ is not accessible under $\tilde{P}$ from some state $w \in \mathcal{W}$ in the support of $\tilde{\pi}$. Observe that Markov's inequality implies that

$$\tilde{\chi} \le \max_{w:\,\tilde{\pi}_w > 0} \frac{2\tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}}}{1 - \mathbb{P}_w(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} > 2\tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}}^w)} \le 4\max_{\tilde{\pi}_w > 0} \tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}}^w,$$

which justifies the choice of the name escape time. The reason for introducing $\tilde{\chi}$ instead of using $\max\{\tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}}^w : \tilde{\pi}_w > 0\}$ directly is that in some cases the former is more easily estimated than the latter.

We are now in a position to prove the following result.

Lemma 2: Let $\tilde{P}$ be a stochastic matrix on a finite set $\mathcal{V}$, and $\tilde{\pi}$ an invariant probability measure. Then,

$$\tilde{\pi}(\mathcal{W}) \le \frac{\tilde{\chi}}{\tau_{\mathcal{W}}^*}, \qquad \forall \mathcal{W} \subseteq \mathcal{V}. \qquad (9)$$

Proof: See [33]. $\blacksquare$

Lemmas 1 and 2 immediately imply the following result:

Theorem 3: Let P and $\tilde{P}$ be stochastic matrices on a finite set $\mathcal{V}$. Let P be irreducible with invariant probability measure π and mixing time τ, and $\tilde{\pi}$ be an invariant probability measure for $\tilde{P}$. Then,

$$||\tilde{\pi} - \pi|| \le \theta\left(\tau \frac{\tilde{\chi}}{\tau_{\mathcal{W}}^*}\right),$$

for all $\mathcal{W} \subseteq \mathcal{V}$ such that $\text{supp}(\tilde{P} - P) \subseteq \mathcal{W}$.

Theorem 3 implies that, in order for the total variation distance $||\tilde{\pi} - \pi||$ to vanish as the network size grows large, it is sufficient that $\tau\tilde{\chi}/\tau_{\mathcal{W}}^*$ vanishes.

Example 1: For integers $m \ge 2$ and $d \ge 1$, let P be the transition probability matrix of the lazy random walk on a d-dimensional toroidal grid of size $n = m^d$, i.e., $\mathcal{V} = \mathbb{Z}_m^d$, $P_{uu} = 1/2$, $P_{uv} = 1/(4d)$ if $\sum_{1 \le i \le d} |u_i - v_i| = 1$, and $P_{uv} = 0$ if $\sum_{1 \le i \le d} |u_i - v_i| \ge 2$. For some $w \in \mathcal{V}$ and $\alpha \in (0, 1)$, consider a perturbed stochastic matrix $\tilde{P}$ coinciding with P outside w, and such that $\tilde{P}_{ww} < 1$. Put $\mathcal{W} = \{w\}$. It is immediate to verify that

$$\tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}} = (1 - \tilde{P}_{ww})^{-1}.$$

On the other hand, Kac's formula [35, Lemma 21.3] yields

$$n = \frac{1}{\pi_w} = 1 + \frac{1}{4d}\sum_{v:|v-w|=1}\tau_w^v = 1 + \frac{1}{2}\tau_{\mathcal{W}}^*,$$

where last equality follows from a basic symmetry argument. Moreover, standard results [35, Theorem 5.5] imply that $\tau \le C_d n^{2/d}$ for some constant C_d depending on d but not on n. Then, Theorem 3 implies that

$$||\pi - \tilde{\pi}|| \le \theta\left(\frac{2C_d}{1 - \tilde{P}_{ww}}\frac{n^{2/d}}{n-1}\right).$$

The above guarantees that $||\pi - \tilde{\pi}||$ vanishes as n grows large provided that $d \ge 3$. More general examples involving toroidal grids will are discussed in [33].

Example 2: For a stochastic matrix Q, a probability distribution μ, and some $\beta \in (0, 1)$, let P and π be as in Section II-A. Let $\tilde{Q}$ be a perturbation of Q, and $\tilde{P} = (1-\beta)\tilde{Q} + \beta\mathbb{1}\mu'$. Clearly $\mathcal{W} := \text{supp}(\tilde{Q} - Q) \supseteq \text{supp}(\tilde{P} - P)$. Moreover,

$$\tilde{\chi} \le \frac{1}{\max_w \mathbb{P}_w(\tilde{V}(1) \in \mathcal{V}\setminus\mathcal{W})} \le \frac{1}{\beta(1 - \mu(\mathcal{W}))}. \quad (10)$$

On the other hand, the mixing time can be easily bounded by considering a coupling of two Markov chains, $U(t)$ and $V(t)$ defined as follows. Before meeting, $U(t)$ and $V(t)$ move independently according to the transition probability matrix Q with probability $(1-\beta)$ and jump to a common new state chosen according to μ with probability β. From the first time they meet, i.e., for $t \ge T_c := \inf\{t \ge 0 : U(t) = V(t)\}$, $U(t) = V(t)$ move together with transition probability matrix P. Since

$$||P_{u,\cdot}^t - P_{v,\cdot}^t|| \le \mathbb{P}(T_c > t | U(0) = u, V(0) = v) \le (1 - \beta)^t$$

for every $t \ge 0$ and $u, v \in \mathcal{V}$, one gets that

$$\tau \le \left\lceil\frac{-1}{\log(1-\beta)}\right\rceil \le \frac{1}{\beta} + 1. \quad (11)$$

Finally, let $\tau_{\mathcal{W}}^\mu := \sum_v \mu_v \tau_{\mathcal{W}}^v$ be the expected hitting time of the Markov chain with initial distribution μ and transition probability matrix P. For all v, one has that

$$\tau_{\mathcal{W}}^v \le \sum_{t \ge 0}(1-\beta)^t\beta(t + \tau_{\mathcal{W}}^\mu) = \frac{1-\beta}{\beta} + \tau_{\mathcal{W}}^\mu.$$

Using Kac's formula, the above implies that

$$\frac{1}{\pi(\mathcal{W})} = 1 + \sum_w \sum_v \frac{\pi_w}{\pi(\mathcal{W})}P_{wv}\tau_{\mathcal{W}}^v \le \frac{1}{\beta} + \tau_{\mathcal{W}}^\mu.$$

It follows that

$$\tau_{\mathcal{W}}^* \ge \beta\tau_{\mathcal{W}}^\mu \ge \frac{\beta}{\pi(\mathcal{W})} - 1. \quad (12)$$

By combining (10), (11), and (12) with Theorem 3, one gets that

$$||\tilde{\pi} - \pi|| \le \theta\left(\frac{(1+\beta)\pi(\mathcal{W})}{\beta^2(1 - \mu(\mathcal{W}))}\right).$$

In particular, the above implies that the alteration of a set of rows $\mathcal{W}$ of vanishing aggregate PageRank $\pi(\mathcal{W})$, and $\mu(\mathcal{W})$ bounded away from 1, has a negligible effect on the whole PageRank vector π (in total variation distance).

We conclude this section with the following two simple examples, showing that having control of each of the terms $\tilde{\chi}$ and τ is necessary in order to estimate $||\tilde{\pi} - \pi||$.

Example 3: Consider the stochastic matrix P with all entries equal to $1/n$, and perturb it in a single node w by putting $\tilde{P}_{ww} = 1 - \alpha$, and $\tilde{P}_{wv} = \alpha/(n-1)$ for all $v \ne w$, for some $\alpha \in (0, 1 - 1/n)$. Then, $\tau = 1$, $\tau_{\mathcal{W}}^* = n$, and $\tilde{\tau}_{\mathcal{V}\setminus\mathcal{W}} = 1/\alpha$, so that Theorem 3 guarantees that $\alpha n \to \infty$ is a sufficient condition for $||\tilde{\pi} - \pi|| \to 0$ as n grows large. On the other hand, it is easily verified that $\pi_v = 1/n$ for all v, while $\tilde{\pi}_w = 1/(n\alpha+1)$, and $\tilde{\pi}_v = n\alpha/((n-1)(n\alpha+1))$, for all $v \ne w$. Hence, $||\tilde{\pi}-\pi|| = (1-\alpha-1/n)(n\alpha+1)$ which
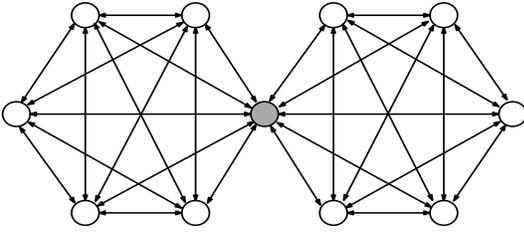
Fig. 2. The graph of Example 4, for $m = 7$. The perturbation set $\mathcal{W} = \{0\}$ is shaded in gray.

shows that $\alpha n \to \infty$ is indeed also a necessary condition for $\|\tilde{\pi} - \pi\| \to 0$ as n grows large.

Example 4: For $m > 1$, let $\mathcal{V} := \{-m, -m+1, \ldots, m-1, m\}$ and $P_{uv} = 1/m$ if $u \neq v$ and $uv \geq 0$, $P_{uv} = 0$ if $uv < 0$ or $u = v$, and $P_{0v} = 1/(2m)$ for all $v \neq 0$. Then, one has $\pi_0 = 2/(2m+1)$ while $\pi_v = 1/(2m+1)$ for all $v \neq 0$. Now perturb P on $\mathcal{W} = \{0\}$ only, by putting $P_{0v} = (1/2 - \alpha)/m$ if $v < 0$ and $P_{0v} = (1/2 + \alpha)/m$ if $v > 0$, for some $\alpha \in (0, 1/2)$. Observe that $\tau_{\mathcal{W}}^* = m$, while $\tilde{\tau}_{\mathcal{V} \setminus \mathcal{W}} = 1$. On the other hand, the bottleneck bound [35, Theorem 7.3] implies that $\tau \geq 1/(4\pi_0) \geq m/2$, so that Theorem 3 is useless as it only provides the trivial conclusion that $\|\tilde{\pi} - \pi\| \leq 1$. However, observe that $\tilde{\pi}_v - \pi_v = 2\alpha/(2m+1)\mathrm{sgn}(v)$, which is arbitrarily close to 1 for large n and α close to $1/2$.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the 7th International World Wide Web Conference*, 1998, pp. 107–117.

[2] A. Langville and C. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.

[3] F. Chung, "The heat kernel as the pagerank of a graph," *Proceedings of the National Academy of Science*, vol. 104, no. 50, pp. 19 735–19 740, 2007.

[4] K. Avrachenkov and N. Litvak, "The effect of new links on Google PageRank," *Stochastic Models*, vol. 22, pp. 319–331, 2006.

[5] C. D. Kerchove, L. Ninove, and P. V. Dooren, "Maximizing PageRank via outlinks," *Linear Algebra and its Applications*, vol. 429, pp. 1254–1276, 2008.

[6] B. Csáji, R. Jungers, and V. Blondel, "PageRank optimization by edge selection," *Applied Discrete Mathematics*, 2013.

[7] H. Ishii and R. Tempo, "Computing the pagerank variation for fragile web data," *SICE Journal of Control, Measurement, and System Integration*, vol. 2, no. 1, pp. 1–9, 2006.

[8] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM Review*, vol. 53, no. 4, pp. 747–772, 2011.

[9] J. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

[10] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[11] A. Jadbabaie, J. Lin, and S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[12] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, "Communication constraints in coordinated consensus problems," in *Proceedings of IEEE ACC*, 2006.

[13] L. Qun and D. Rus, "Global clock synchronization in sensor networks," *IEEE Journal on Computers*, vol. 55, no. 2, pp. 214–226, 2006.

[14] Y. Liu and Y. Yang, "Reputation propagation and agreement in mobile ad hoc networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2003, pp. 1510–1515.

[15] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and S. O., "Novel type of phase transitions in a system of self-driven particles," *Physical Review Letters*, vol. 75, no. 6, pp. 1226–1229, 1995.

[16] F. Cucker and S. Smale, "The mathematics of emergence," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 197–227, 2007.

[17] M. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[18] P. DeMarzo, D. Vayanos, and J. Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly Journal of Economics*, vol. 118, no. 3, pp. 909–968, 2003.

[19] B. Golub and M. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, vol. 2, no. 1, pp. 112–149, 2010.

[20] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operation Research*, vol. 38, no. 1, pp. 1–27, 2013.

[21] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis)information in social networks," *Games and Economic Behavior*, vol. 70, no. 2, pp. 194–227, 2010.

[22] T. Liggett, *Interacting particle systems*. Springer-Verlag, 1985.

[23] ——, *Stochastic interacting systems: Contact, voter, and exclusion processes*. Springer Berlin, 1999.

[24] P. Connely and D. Welsh, "Finite particle systems and infection models," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 94, pp. 167–182, 1982.

[25] J. Cox, "Coalescing random walks and voter model consensus times on the torus in $\mathbb{Z}^d$," *The Annals of Probability*, vol. 17, no. 4, pp. 1333–1366, 1989.

[26] D. Aldous and J. Fill, "Reversible Markov chains and random walks on graphs," 2002, monograph in preparation, http://www.stat.berkeley.edu/ aldous/RWG/book.html.

[27] R. Durrett, *Random graph dynamics*. Cambridge University Press, 2006.

[28] M. Mobilia, "Does a single zealot affect an infinite group of voters?" *Physical Review Letters*, vol. 91, no. 2, p. 028701, 2003.

[29] M. Mobilia and I. Georgiev, "Voting and catalytic processes with inhomogeneities," *Physical Review E*, vol. 71, p. 046102, 2005.

[30] P. J. Schweitzer, "Perturbation theory and finite Markov chains," *Journal of Applied Probability*, vol. 5, pp. 401–413, 1968.

[31] A. Mitrophanov, "Sensitivity and convergence of uniformly ergodic Markov chains," *Journal of Applied Probability*, vol. 42, pp. 1003–1014, 2005.

[32] G. Cho and C. Meyer, "Comparison of perturbation bounds for the stationary distribution of a Markov chain," *Linear Algebra and its Applications*, vol. 335, pp. 137–150, 2001.

[33] G. Como and F. Fagnani, "Robustness of large-scale stochastic matrices to localized perturbations," *Submitted*, 2014.

[34] R. Montenegro and P. Tetali, "Mathematical aspects of mixing times in Markov chains," *Foundations and Trends in Theoretical Computer Science*, vol. 1, no. 3, pp. 237–354, 2006.

[35] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, 2010.

[36] A. Mitrophanov, "Stability and exponential convergence of continuous-time Markov chains," *Journal of Applied Probability*, vol. 40, pp. 970–979, 2003.

[37] E. Seneta, "Perturbation of the stationary distribution measured by ergodicity constraints," *Advances in Applied Probability*, vol. 20, pp. 228–230, 1988.

[38] ——, "Sensitivity of finite Markov chains under perturbation," *Statistics & Probability Letters*, vol. 17, pp. 163–168, 1993.

[39] H. Haveliwala, "Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search," *IEEE Transactions on Knowledge Data Engineering*, vol. 15, no. 4, pp. 784–796, 2003.

[40] K. Suchecki, V. Eguíluz, and M. S. Miguel, "Conservation laws for the voter model in complex networks," *Europhysics Letters*, vol. 69, no. 2, p. 228, 2005.