Robustness of large-scale stochastic matrices to localized perturbations

Giacomo Como and Fabio Fagnani

***Abstract*—Many notions of network centrality can be formulated in terms of invariant probability vectors of suitably defined stochastic matrices encoding the network structure. Analogously, invariant probability vectors of stochastic matrices allow one to characterize the asymptotic behavior of many linear network dynamics, e.g., arising in distributed averaging algorithms for estimation or control as well as opinion dynamics in social networks. Hence, a central problem in network science and engineering is that of assessing the robustness of such invariant probability vectors to perturbations possibly localized on some relatively small part of the network. In this work, upper bounds are derived on the total variation distance between the invariant probability vectors of two stochastic matrices differing on a subset $\mathcal{W}$ of rows. Such bounds depend on three parameters: the mixing time and the entrance time on the set $\mathcal{W}$ for the Markov chain associated to one of the matrices; and the escape probability from the set $\mathcal{W}$ for the Markov chain associated to the other matrix. These results, obtained through coupling techniques, prove particularly useful in scenarios where $\mathcal{W}$ is a small subset of the state space, even if the difference between the two matrices is not small in any norm. Several applications to large-scale network problems are discussed, including robustness of Google's PageRank algorithm, distributed averaging, consensus algorithms, and the voter model.**

***Index Terms*—Stochastic matrices, invariant probability vectors, robustness, resilience, large-scale networks, PageRank, centrality, distributed averaging, consensus, voter model.**

I. Introduction

How much can the invariant probability vector

$$\pi = \pi P$$

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 vectors of suitably defined stochastic matrices. In particular, Google's PageRank algorithm [6] assigns to webpages values corresponding to the entries of the invariant probability vector 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 [23], [10]. 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], [22], [13]. 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 [20].

The problem is of central interest also in the context of distributed averaging and consensus algorithms [35]. 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 [41], [42], [5], control [21], [34], [7], synchronization in sensor networks [36], or reputation management in ad-hoc networks [27], as well as behavioral models for flocking phenomena [43], [14], or opinion dynamics in social networks [15], [16], [18], [1]. 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 to a consensus vector with all entries equal to $\overline{x} = \pi x(0)$. Depending on the specific application, the natural question is to what extent the consensus value $\overline{x}$ 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 [2].

Other applications can be found in the context of interacting particle systems [25], [26]. In particular, in the voter model on a finite graph [11], [12], [3, Ch. 14], [17, Ch. 6.9], the probability vector of the final consensus value is determined by the invariant probability vector 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' [31], [32], namely agents with an asymmetric behavior in the way they influence and are influenced from their neighbor agents.

The above-described problems all boil down to estimating the distance between the invariant probability vector π of an irreducible stochastic matrix P and an invariant probability vector $\tilde{\pi} = \tilde{\pi}\tilde{P}$ of another stochastic matrix $\tilde{P}$, to be interpreted as a perturbed version of P. In some applications, P may be reversible, equivalently be obtained by normalizing the rows of a symmetric nonnegative matrix W, and π can be explicitly computed in terms of the row sums of W. 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 sensi-

G. Como is with the Department of Automatic Control, Lund University, SE-221 00 Lund, Sweden giacomo.como@control.lth.se. F. Fagnani is with the Department of Mathematical Sciences, Politecnico di Torino, Torino, 10129 fabio.fagnani@polito.it.

tivity analysis [37], [38], [39], [28], [8], [9], [29], [30], [2] do not provide a satisfactory answer to this problem. Indeed, they provide upper bounds of the form

$$||\tilde{\pi} - \pi||_p \leq \kappa_P ||\tilde{P} - P||_q \,, \tag{1}$$

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

$$||\tilde{\pi} - \pi|| \leq \theta \left(\frac{\tau_{\rm mix}}{\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}}} \right) , \tag{2}$$

(see Theorem 3) where:

- $\theta : [0, +\infty) \to [0, 1]$ is a continuous, nondecreasing function such that $\theta(0) = 0$ (see (22) for its definition and Fig. 1 for its graph);
- $\tau_{\rm mix}$ is the mixing time of the matrix P, defined as

$$\tau_{\rm mix} := \inf \left\{ t \geq 1 : \max_{u,v \in \mathcal{V}} ||P^t_{u,\cdot} - P^t_{v,\cdot}|| \leq \frac{1}{e} \right\} ; \tag{3}$$

- $\tau^*_{\mathcal{W}}$ is the entrance time on the set $\mathcal{W}$, defined as

$$\tau^*_{\mathcal{W}} := \min_{u \in \mathcal{V} \setminus \mathcal{W}} \tau^u_{\mathcal{W}} \tag{4}$$

  where $\tau^u_{\mathcal{W}}$, for $u \in \mathcal{V}$, are the solution of the linear system

$$\tau^u_{\mathcal{W}} = 0 \,,\ u \in \mathcal{W} \,, \qquad \tau^u_{\mathcal{W}} = 1 + \sum_{v \in \mathcal{V}} P_{uv} \tau^v_{\mathcal{W}} \,,\ u \in \mathcal{V} \setminus \mathcal{W} \tag{5}$$

  and thus coincide with the expected hitting times on the set $\mathcal{W}$ for a Markov chain with transition probability matrix P;

- $\tilde{\gamma}_{\mathcal{W}}$ stands for the escape probability from $\mathcal{W}$ defined as

$$\tilde{\gamma}_{\mathcal{W}} := \sup_{t \geq 0} \min_{\substack{w \in \mathcal{W}: \\ \tilde{\pi}_w > 0}} \frac{1}{t} \sum_{k=1}^{t} \sum_{\substack{\xi_0 = w, \xi_k \in \mathcal{V} \setminus \mathcal{W} \\ \xi_1, \ldots, \xi_{k-1} \in \mathcal{W}}} \prod_{1 \leq l \leq k} \tilde{P}_{\xi_{l-1} \xi_l} \,, \tag{6}$$

  where the second summation runs over all $(k+1)$-tuples ξ that start with $\xi_0 = w$, end with some $\xi_k \in \mathcal{V} \setminus \mathcal{W}$, and have all intermediate entries $\xi_l \in \mathcal{W}$, for $1 \leq l < k$. As shown in (24), the argument of the minimization in (6) coincides with the probability that a Markov chain with transition probability matrix $\tilde{P}$ started at w exits from $\mathcal{W}$ before time t, normalized by t.

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 [1] in the context of 'highly fluid' social networks. Because of the properties of $\theta(\,\cdot\,)$, the bound (2) implies that the total variation distance $||\tilde{\pi} - \pi||$ vanishes provided that $\tau_{\rm mix} / (\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}})$ 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 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 [3], [33], [24], and several results are available allowing one to estimate the mixing time $\tau_{\rm mix}$ 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 [29], [30], where (1) is proven for $p = 1$, $q = \infty$, and condition number κ_P proportional to $\tau_{\rm mix}$. Of a similar flavor are Seneta's results [38], [39] estimating the condition number κ_P in terms of ergodicity coefficients. Also the estimates proposed in [2] 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 $1/(\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}})$ 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 parameters appearing in the righthand side of (2), the escape probability $\tilde{\gamma}_{\mathcal{W}}$ 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 entrance time $\tau^*_{\mathcal{W}}$, which depends on P and $\mathcal{W}$ only, may result the hardest to get lower bounds on in typical applications where P is sparse and $\mathcal{W}$ remains small as the state space grows large. While Kac's formula ([24, Lemma 21.13])

$$\sum_{w \in \mathcal{W}} \sum_{v \in \mathcal{V}} \pi_w P_{wv} (\tau^v_{\mathcal{W}} + 1) = 1 \tag{7}$$

readily implies the upper bound $\tau^*_{\mathcal{W}} \leq 1/\pi(\mathcal{W})$, where $\pi(\mathcal{W}) := \sum_{w \in \mathcal{W}} \pi_w$, 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 $\mathcal{W}$ is localized and its size remains bounded (or grows very slowly) as the network size grows large. Results for more general graphs, in particular, 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. Section IV-A discusses in detail the application of our result to the PageRank manipulation problem. Section IV-B focuses on stochastic matrices whose support graph has high local connectivity and discusses lower bounds of the entrance time $\tau^*_{\mathcal{W}}$. This allows for efficient application of Theorem 3 to networks with a finite dimensional structure. Explicit examples on toroidal grid graphs are presented.

Before proceeding, let us collect here some notational conventions to be used throughout the paper. When referring

to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we will always use the convention that $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, i.e., that its links are directed. Then, $\mathcal{G}$ undirected means that if $(u, v) \in \mathcal{E}$ then $(v, u) \in \mathcal{E}$ as well. Given $u \in \mathcal{V}$, put $\mathcal{E}_u := \{v : (u, v) \in \mathcal{E}\}$ and let $d_u := |\mathcal{E}_u|$ be the (out-) degree of node u. Vectors and matrices will be considered with entries from a set $\mathcal{V}$ of finite cardinality $n := |\mathcal{V}|$. A summation index v is always intended to run over the whole $\mathcal{V}$, while a summation index w is intended to run over a specified subset $\mathcal{W} \subseteq \mathcal{V}$. The all-one column vector will be denoted by $\mathbb{1}$. For a matrix A, A' will stand for its transpose and $\mathrm{supp}(A) := \{v : A_{v,\cdot} \neq 0\}$ for the set of its nonzero rows. We refer to a probability vector as a nonnegative row vector μ such that $\mu\mathbb{1} = 1$ and to a stochastic matrix P as a nonnegative square matrix P such that $P\mathbb{1} = \mathbb{1}$. A probability vector is said invariant for a stochastic matrix P if $\mu P = \mu$. A stochastic matrix P is said irreducible if the associated support graph $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}_P)$, where $(u, v) \in \mathcal{E}_P$ if and only if $P_{uv} > 0$, is connected. It is a standard result that every irreducible stochastic matrix P admits a unique probability vector $\pi = \pi P$. The total variation distance between two probability vectors μ and π is denoted by

$$||\mu - \pi|| := \frac{1}{2} \sum_v |\mu_v - \pi_v| .$$

Given a stochastic matrix P, it is natural to consider discrete-time Markov chains $V(t)$, $t = 0, 1, \ldots$, with state space $\mathcal{V}$ and transition probability matrix P. I.e., for all $u, v \in \mathcal{V}$ and $t \geq 0$, $\mathbb{P}(V(t+1) = v | V(t) = u) = P_{uv}$. For $u \in \mathcal{V}$, $\mathbb{P}_u$ and $\mathbb{E}_u$ will stand for the probability and expectation conditioned on $V(0) = u$. We will also use the notation $\mathbb{P}_\mu := \sum_v \mu_v \mathbb{P}_v$ for a probability vector μ. We will denote the hitting time on a subset $\mathcal{W} \subseteq \mathcal{V}$ by $T_\mathcal{W} := \inf\{t \geq 0 : V(t) \in \mathcal{W}\}$. It is a standard result that the expected hitting times $\mathbb{E}_u[T_\mathcal{W}]$ coincide with the solution $\tau^u_\mathcal{W}$ of equation (5).

II. Three motivating applications

In this section we present three motivating examples formalizing some of the application problems discussed in the Introduction. Througout, $n := |\mathcal{V}|$ will stand for the network size.

A. PageRank manipulation

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 link $(u, v) \in \mathcal{E}$ whenever page u has a hyperlink directed to page v. Define a stochastic matrix Q by putting $Q_{uv} = 1/n$ for all v if $d_u = 0$, and, if $d_u \geq 1$, letting $Q_{uv} = 0$ if $(u, v) \notin \mathcal{E}$ and $Q_{uv} = 1/d_u$ if $(u, v) \in \mathcal{E}$. Given an arbitrary probability vector μ and a parameter β in the interval $(0, 1)$, consider the equation

$$\pi = (1 - \beta)\pi Q + \beta\mu . \tag{8}$$

Since the matrix $W := (I - (1 - \beta)Q)$ is strictly diagonally dominant, hence nonsingular, equation (8) admits exactly one solution $\pi = \beta\mu W^{-1}$. As we will discuss in a moment, such π turns out to be a probability vector. It is known as the PageRank vector and was first introduced by Brin and Page [6] to measure the relative importance of webpages. In the original PageRank version, $\mu = n^{-1}\mathbb{1}$ is chosen as the uniform distribution over the set of webpages, while typical values of β used in practice are about 0.15. More general choices of the probability vector μ lead to the definition of the personalized PageRank[19], which is used in context-sensitive searches.

Consider now a (relatively small) set of webpages $\mathcal{W} \subseteq \mathcal{V}$, and assume that the set $\bigcup_{w \in \mathcal{W}} \mathcal{E}_w$ of hyperlinks originated from these webpages 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, and $\tilde{\pi}$ the corresponding modified PageRank vector solving the equation

$$\tilde{\pi} = (1 - \beta)\tilde{\pi}\tilde{Q} + \beta\mu . \tag{9}$$

A standard result [24, Proposition 4.2] allows one to write the total variation distance between π and $\tilde{\pi}$ as

$$||\tilde{\pi} - \pi|| = \max_{\mathcal{U} \subseteq \mathcal{V}} \{\tilde{\pi}(\mathcal{U}) - \pi(\mathcal{U})\} . \tag{10}$$

The identity above shows that the maximum, over all subsets of webpages $\mathcal{U}$, of the difference between the aggregate centralities that PageRank assigns to $\mathcal{U}$ in the WWW graph $\mathcal{G}$ and, respectively, in its modified version $\tilde{\mathcal{G}}$, coincides with the total variation distance between the PageRank vectors π and $\tilde{\pi}$.

We now give a different characterization of the PageRank vector and reformulate the perturbation problem. First, we introduce the stochastic matrix

$$P := (1 - \beta)Q + \beta\mathbb{1}\mu .$$

We claim that P has a unique invariant probability vector and that it coincides with the PageRank vector π. To see this equivalence, first notice that, if π solves (8), the fact that Q is a stochastic matrix and μ a probability vector, imply that

$$\pi\mathbb{1} = (1 - \beta)\pi Q\mathbb{1} + \beta\mu\mathbb{1} = (1 - \beta)\pi\mathbb{1} + \beta ,$$

so that $\pi\mathbb{1} = 1$. Now, if ν is any row vector such that $\nu\mathbb{1} = 1$, we have that

$$\nu P = (1 - \beta)\nu Q + \beta\nu\mathbb{1}\mu = (1 - \beta)\nu Q + \beta\mu ,$$

so that $\nu = \nu P$ if and only if ν coincides with the solution π of (8). An analogous argument shows that the modified PageRank vector $\tilde{\pi}$ coincides with the unique invariant probability vector of the stochastic matrix

$$\tilde{P} := (1 - \beta)\tilde{Q} + \beta\mathbb{1}\mu .$$

Hence, estimating the impact that an arbitrary change of the hyperlinks from a subset $\mathcal{W}$ of webpages has on the aggregate PageRank of an arbitrary subset $\mathcal{U}$ of webpages boils down to bounding the total variation distance between the invariant probability vectors π and $\tilde{\pi}$ of the stochastic matrices P and $\tilde{P}$, respectively. Observe that, since the matrices Q and $\tilde{Q}$ differ only on the rows indexed by elements of $\mathcal{W}$, so do P and $\tilde{P}$. In Example II-A of Section III we will prove an upper bound on $||\tilde{\pi} - \pi||$ depending only on the size of $\mathcal{W}$ (as measured by π and μ), and on the value of the parameter $\beta \in (0, 1)$.

B. Faulty communication links in distributed averaging algorithms

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

$$\overline{y} := \frac{1}{n} \sum_v y_v \,.$$

A possible solution [35] is as follows. Let $d \in \mathbb{R}^{\mathcal{V}}$ be the degree vector in $\mathcal{G}$. Initialize the state of every sensor $v \in \mathcal{V}$ as

$$x_v(0) = \frac{y_v}{d_v} \,, \qquad z_v(0) = \frac{1}{d_v} \,. \tag{11}$$

Then, at every time instant $t = 0, 1, \ldots$, let every sensor $v \in \mathcal{V}$ update its state according to the recursion

$$[x_v(t+1), z_v(t+1)] = \frac{1}{2}[x_v(t), z_v(t)] + \frac{1}{2d_v} \sum_{(u,v)\in\mathcal{E}} [x_u(t), z_u(t)] \,. \tag{12}$$

What makes the above iteration 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 (11) and (12) 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 $(u, v) \in \mathcal{E}$ and $Q_{uv} = 0$ if $(u, v) \notin \mathcal{E}$. Let

$$x(0) = \frac{y}{d} \,, \qquad z(0) = \frac{\mathbb{1}}{d} \tag{13}$$

(where division between two vectors is meant componentwise) and consider the iteration

$$x(t+1) = Px(t) \,, \qquad z(t+1) = Pz(t) \,. \tag{14}$$

Observe that the unique invariant probability vector π of the matrix P is given by

$$\pi_v = \frac{d_v}{n\overline{d}} \,, \qquad v \in \mathcal{V} \,, \tag{15}$$

where

$$\overline{d} := \frac{1}{n} \sum_v d_v$$

is the average degree. Moreover, irreducibility and acyclicity of P (implied by $P_{uu} > 0$ for all u) imply that

$$x(t) = P^t \frac{y}{d} \stackrel{t\to\infty}{\longrightarrow} \mathbb{1}\pi\frac{y}{d} = \mathbb{1}\frac{\overline{y}}{\overline{d}} \,,$$

$$z(t) = P^t \frac{\mathbb{1}}{d} \stackrel{t\to\infty}{\longrightarrow} \mathbb{1}\pi\frac{\mathbb{1}}{d} = \mathbb{1}\frac{1}{\overline{d}} \,,$$

so that

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

Therefore, the iterative distributed algorithm defined by (13)-(14) effectively computes the average $\overline{y}$ of the vector y. The example can be generalized to those weighted graphs whose nodes all have in-degree equal to the out-degree (hence, in particular, undirected weighted graphs). Indeed, for these graphs, the invariant probability vector π of the associated stochastic matrix P admits the explicit form (15).

Now, let $\mathcal{F} \subseteq \mathcal{E}$ be a subset of directed communication links which stop working and $\tilde{\mathcal{G}} := (\mathcal{V}, \tilde{\mathcal{E}})$, where $\tilde{\mathcal{E}} := \mathcal{E} \setminus \mathcal{F}$, be the directed graph obtained from $\mathcal{G}$ by removing such links. Let $\tilde{d}$ be the vector of in-degrees in $\tilde{\mathcal{G}}$ and define $\tilde{P} = (I + \tilde{Q})/2$, where $\tilde{Q}$ is a stochastic matrix with $\tilde{Q}_{uv} = 1/\tilde{d}_u$ if $(v, u) \in \tilde{\mathcal{E}}$ and $\tilde{Q}_{uv} = 0$ otherwise. Consider the following recursion, analogous to (13) and (14), with d and $\mathcal{E}$ replaced by $\tilde{d}$ and $\tilde{\mathcal{E}}$, respectively:

$$\tilde{x}(0) = \frac{y}{\tilde{d}} \,, \qquad \tilde{z}(0) = \frac{\mathbb{1}}{\tilde{d}} \,, \tag{16}$$

$$\tilde{x}(t+1) = \tilde{P}x(t) \qquad \tilde{z}(t+1) = \tilde{P}\tilde{z}(t) \,. \tag{17}$$

Then, provided that $\tilde{\mathcal{G}}$ remains strongly connected, an argument as the one before shows that

$$\frac{\tilde{x}_v(t)}{\tilde{z}_v(t)} \stackrel{t\to\infty}{\longrightarrow} \tilde{y} \,, \qquad \forall v \in \mathcal{V} \,,$$

where

$$\tilde{y} = \frac{\tilde{\pi}(y/\tilde{d})}{\tilde{\pi}(\mathbb{1}/\tilde{d})}$$

and $\tilde{\pi}$ is the unique invariant probability vector of $\tilde{P}$. In other words, the perturbed dynamics (16)–(17) achieve consensus on a perturbed value $\tilde{y}$.

We are now going to show that the absolute error $|\tilde{y} - y|$ can be upper bounded in terms of the total variation $||\tilde{\pi} - \pi||$ and the fraction $|\mathcal{F}|/|\mathcal{E}|$ of failed communication links. To see this, first we express the perturbed consensus value as

$$\tilde{y} = \frac{\tilde{\pi}y/\tilde{d}}{\tilde{\pi}\mathbb{1}/\tilde{d}} = \frac{\overline{y} + \varepsilon_1 + \varepsilon_2}{1 + \varepsilon_3 + \varepsilon_4} \,,$$

where

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

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

Now, using the facts that $\tilde{d}_v \geq 1$ for all v (since $\tilde{\mathcal{G}}$ is connected) and that $|\mathcal{E}| = \sum_v d_v = n\overline{d}$, one gets that

$$|\varepsilon_1| \leq \overline{d} \frac{|\mathcal{F}|}{|\mathcal{E}|} ||y||_\infty \,, \quad |\varepsilon_2| \leq \overline{d} ||y||_\infty ||\tilde{\pi} - \pi|| \,,$$

$$|\varepsilon_3| \leq \overline{d} \frac{|\mathcal{F}|}{|\mathcal{E}|} \,, \quad |\varepsilon_4| \leq \overline{d} ||\tilde{\pi} - \pi|| \,, \quad |\tilde{y}| \leq ||y||_\infty \,.$$

It follows that

$$\begin{aligned} |\tilde{y} - \overline{y}| &= |\varepsilon_3\tilde{y} + \varepsilon_4\tilde{y} - \varepsilon_1 - \varepsilon_2| \\ &\leq |\tilde{y}|(|\varepsilon_3| + |\varepsilon_4|) + |\varepsilon_1| + |\varepsilon_2| \\ &\leq 2\overline{d} ||y||_\infty \left(|\mathcal{F}|/|\mathcal{E}| + ||\tilde{\pi} - \pi|| \right) \,, \end{aligned}$$

so that

$$\frac{|\tilde{y} - \overline{y}|}{||y||_\infty} \leq 2\overline{d} \left(\frac{|\mathcal{F}|}{|\mathcal{E}|} + ||\tilde{\pi} - \pi|| \right) \,. \tag{18}$$

Formula (18) shows that, provided an upper bound on the average degree $\overline{d}$, in order to guarantee that the value $\tilde{y}$ computed by the distributed averaging algorithm on the perturbed graph $\tilde{\mathcal{G}}$ is close to the average $\overline{y}$ of the sensors' measurements, it is sufficient that both the fraction $|\mathcal{F}|/|\mathcal{E}|$ of failed communication links and the total variation distance $||\tilde{\pi} - \pi||$ are small.

C. Voter model with influential agents

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected undirected graph (with no self-loops). Nodes are to be interpreted as agents possessing a binary opinion. Opinions are changing with time as the consequence of interactions in the network. Precisely, for $u \in \mathcal{V}$ and $t = 0, 1, \ldots$, let $X_u(t) \in \{0, 1\}$ be the opinion of agent u at time t. Dynamics takes place as follows: at every time $t = 0, 1, \ldots$, a single directed link (u, v) is activated, chosen uniformly at random from $\mathcal{E}$, and its tail node u updates its state $X_u(t)$ by copying the head node v's current state $X_v(t)$. Assembling all opinions on one vector $X(t) \in \{0, 1\}^{\mathcal{V}}$ we obtain that $X(t)$ is a Markov chain whose transitions can be compactly described as follows. 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 $E^{(u,v)}_{u,v} = -E^{(u,v)}_{u,u} = 1$. Then, given $X(t)$, we have that

$$X(t+1) = (I + E^{(u,v)})X(t)$$

with probability $1/|\mathcal{E}|$, for all $(u, v) \in \mathcal{E}$. This is an instance of the voter model [25], [26], [11], [12]. 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 a standard result that, with probability one, this dynamics achieves consensus in some finite time. More precisely, there exists some random consensus time T, which is finite with probability one, and a random consensus value $Y \in \{0, 1\}$, such that

$$X_v(t) = Y\,, \qquad \forall v \in \mathcal{V}\,,\ \forall t \geq T\,. \tag{19}$$

The main asymptotic quantity of interest is the probability distribution of the consensus value Y conditioned to the initial condition $X(0)$. Specifically, we define

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

Now, let us consider the following variant to the model. Consider a direct subgraph $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$, where $\tilde{\mathcal{E}} = \mathcal{E} \setminus \mathcal{F}$ is obtained from $\mathcal{E}$ by removing a subset $\mathcal{F} \subseteq \mathcal{E}$ of directed links. We assume that $\tilde{\mathcal{G}}$ remains strongly connected. 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 [2] in the framework of opinion dynamics over continuous space. Observe that, analogously to the voter model on $\mathcal{G}$, strong connectivity of the graph $\tilde{\mathcal{G}}$ implies that, with probability one, the process $\tilde{X}(t)$ achieves a consensus in finite time on a binary random variable $\tilde{Y}$. We can similarly define the conditional probability

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

The absolute difference $|\tilde{y} - y|$ measures the effect of the influential individuals in the final consensus value. We now give a different characterization for y and $\tilde{y}$ in terms of invariant probability vectors of suitably defined stochastic matrices and propose a characterization of $|\tilde{y} - y|$ in terms of their total variation difference.

Let us define the stochastic matrix

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

Then, $\mathbb{E}[X(t+1)|X(t)] = PX(t)$ for all $t = 0, 1, \ldots$, so that an inductive argument proves that

$$\mathbb{E}[X(t)|X(0)] = P^t X(0) \qquad t \geq 0\,. \tag{20}$$

Since $\mathcal{G}$ is connected and undirected, P is irreducible and symmetric, so that its unique invariant probability vector is the uniform one

$$\pi = \frac{1}{n}\mathbb{1}'\,.$$

It then follows from (21) that, for all $t \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_v X_v(t)|X(0)\right] &= \pi\mathbb{E}[X(t)|X(0)] \\ &= \pi P^t X(0) \\ &= \pi X(0) \\ &= \frac{1}{n}\sum_v X_v(0)\,,\end{aligned} \tag{21}$$

a property that is sometimes referred to as as conservation of the average magnetization [40] in the statistical physics jargon. Finally, it follows from (19) and (21) that

$$y = \mathbb{E}[Y|X(0)] = \lim_{t\to+\infty}\frac{1}{n}\sum_v \mathbb{E}[X_v(t)|X(0)] = \frac{1}{n}\sum_v X_v(0)\,.$$

Similarly,

$$\tilde{y} = \tilde{\pi}\tilde{X}(0)\,,$$

where $\tilde{\pi} = \tilde{\pi}\tilde{P}$ is the unique invariant probability vector of the stochastic matrix

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

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

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

In fact, while the inequality above is valid for every initial state value $\tilde{X}(0) = X(0) \in \{0, 1\}^{\mathcal{V}}$, the identity (10) implies that such inequality is tight in the sense that there exists one value $x \in \{0, 1\}^{\mathcal{V}}$ (the one with $x_u = 1$ for $u \in \mathcal{U}$ and $x_v = 0$

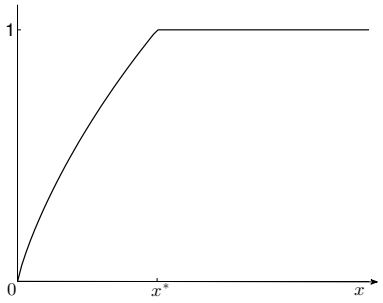

Fig. 1. Graph of the function $\theta(x)$ defined in (22).

for $v \in \mathcal{V} \setminus \mathcal{U}$, where $\mathcal{U}$ is such that $||\tilde{\pi} - \pi|| = \tilde{\pi}(\mathcal{U}) - \pi(\mathcal{U}))$ such that, if $\tilde{X}(0) = X(0) = x$, then $|\tilde{y} - y| = ||\tilde{\pi} - \pi||$.

It follows that the problem of estimating the difference between the probability vector of the eventual consensus value for the voter model on $\mathcal{G}$ and $\tilde{\mathcal{G}}$ is equivalent to the one of estimating the total variation distance between the invariant probability vectors of the stochastic matrices P and $\tilde{P}$, respectively.

III. Perturbation results

Let P be an irreducible stochastic matrix on the finite state space $\mathcal{V}$ and let $\pi = \pi P$ be its unique invariant probability vector. 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 $\tilde{\pi} = \tilde{\pi}\tilde{P}$ be an invariant probability vector 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 $\theta : [0, +\infty) \to [0, 1]$

$$\theta(x) := \begin{cases} x \ln\left(e^2/x\right) & x \le x^* \\ 1 & x \ge x^*, \end{cases} \tag{22}$$

where $x^* = 0.31784\ldots$ is the smallest positive solution of $e^2/x = \exp(1/x)$. (The graph of $\theta(\,\cdot\,)$ is plotted in Figure 1.)

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

$$||\tilde{\pi} - \pi|| \le \theta(\tau_{\rm mix} \cdot \tilde{\pi}(\mathcal{W}))\,,$$

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

Proof Let $V(t)$ and $\tilde{V}(t)$ be two Markov chains on $\mathcal{V}$ which start and move together with transition probabilities P_{uv} until the first time $T_{\mathcal{W}} = \tilde{T}_{\mathcal{W}}$ they hit $\mathcal{W}$, and move independently with transition probabilities P_{uv} and $\tilde{P}_{uv}$, respectively, ever after. Since P and $\tilde{P}$ coincide on $\mathcal{V} \setminus \mathcal{W}$, one has that the marginal transition probability matrices of $V(t)$ and $\tilde{V}(t)$ coincide with P and $\tilde{P}$, respectively. Then, for all $\mathcal{A} \subseteq \mathcal{V}$, and $t \ge 0$, one has that

$$\begin{aligned} \tilde{\pi}(\mathcal{A}) &= \mathbb{P}_{\tilde{\pi}}(\tilde{V}(t) \in \mathcal{A}) \\ &= \mathbb{P}_{\tilde{\pi}}(V(t) \in \mathcal{A}, \tilde{T}_{\mathcal{W}} \ge t) + \mathbb{P}_{\tilde{\pi}}(\tilde{V}(t) \in \mathcal{A}, \tilde{T}_{\mathcal{W}} < t) \\ &\le \mathbb{P}_{\tilde{\pi}}(V(t) \in \mathcal{A}) + \mathbb{P}_{\tilde{\pi}}(\tilde{T}_{\mathcal{W}} < t) \\ &\le \pi(\mathcal{A}) + \exp(-\lfloor t/\,\tau_{\rm mix}\rfloor) + t\tilde{\pi}(\mathcal{W})\,, \end{aligned}$$

where the first identity uses the invariance of $\tilde{\pi}$, and the last inequality follows from $||\mu P^t - \pi|| \le \exp(-\lfloor t/\,\tau_{\rm mix}\rfloor)$ (which is a standard consequence of the submultiplicativity property of the maximal total variation distance, see, e.g., (4.31) in [24]) and the bound

$$\mathbb{P}_{\tilde{\pi}}(\tilde{T}_{\mathcal{W}} < t) \le \sum_{i=0}^{t-1} \mathbb{P}_{\tilde{\pi}}(\tilde{V}(i) \in \mathcal{W}) = t\tilde{\pi}(\mathcal{W})\,,$$

which is implied by the union bound and, again, invariance of $\tilde{\pi}$ for $\tilde{P}$. Therefore, using the characterization (10) of the total variation distance, one gets that

$$||\tilde{\pi} - \pi|| = \max_{\mathcal{A} \subseteq \mathcal{V}} \{\tilde{\pi}(\mathcal{A}) - \pi(\mathcal{A})\} \le \exp(-\lfloor t/\,\tau_{\rm mix}\rfloor) + t\tilde{\pi}(\mathcal{W})\,,$$

for all $t \ge 0$. The claim now follows by choosing

$$t = \max\left\{\left\lfloor \tau_{\rm mix} \log \frac{e}{\tau_{\rm mix} \cdot \tilde{\pi}(\mathcal{W})} \right\rfloor, 0\right\}\,,$$

such a choice being suggested by the minimization of the function

$$x \mapsto \exp(-x/\,\tau_{\rm mix} - 1) + x\tilde{\pi}(\mathcal{W})$$

over continuous nonnegative values of x. ■

Lemma 1 shows that it is sufficient to have an upper bound on the product $\tau_{\rm mix} \cdot \tilde{\pi}(\mathcal{W})$ in order to obtain an upper bound on the total variation distance $||\tilde{\pi} - \pi||$. In particular, assuming that an upper bound on the mixing time $\tau_{\rm mix}$ 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 the entrance time $\tau^*_{\mathcal{W}}$ and of the escape probability $\tilde{\gamma}_{\mathcal{W}}$, defined in (4) and (6), respectively. These two quantities can be given the following probabilistic interpretation. Consider a Markov chain $\tilde{V}(t)$ on $\mathcal{V}$ with transition probability matrix $\tilde{P}$, and let

$$\tilde{T}_{\mathcal{W}} := \inf\{t \ge 0 :\ \tilde{V}(t) \in \mathcal{W}\}$$

and

$$\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} := \inf\{t \ge 0 :\ \tilde{V}(t) \in \mathcal{V} \setminus \mathcal{W}\}$$

be, respectively, the hitting time on, and the exit time from, the set $\mathcal{W}$. Then, since P and $\tilde{P}$ coincide outside $\mathcal{W}$, one has that the expected hitting times satisfy

$$\mathbb{E}_v[\tilde{T}_{\mathcal{W}}] = \tau^v_{\mathcal{W}} = \mathbb{E}_v[T_{\mathcal{W}}]\,, \qquad v \in \mathcal{V}\,. \tag{23}$$

In fact, the entrance time $\tau^*_{\mathcal{W}} = \min\{\tau^u_{\mathcal{W}} :\ v \in \mathcal{V} \setminus \mathcal{W}\}$ only depends on the choice of the subset $\mathcal{W} \supseteq \mathrm{supp}(\tilde{P} - P)$ 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 $\tilde{P} - P$. On the other hand, for every $w \in \mathcal{W}$ and $k \geq 1$ one has that

$$\phi_w(k) := \mathbb{P}_w(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} = k) = \sum_{\substack{\xi_0 = w, \xi_k \in \mathcal{V}\setminus\mathcal{W} \\ \xi_1, \ldots, \xi_{k-1} \in \mathcal{W}}} \prod_{1 \leq l \leq k} \tilde{P}_{\xi_{l-1}\xi_l} \,,$$

so that the escape probability defined in (6) satisfies

$$\tilde{\gamma}_{\mathcal{W}} := \sup_{t \geq 1} \min_{\substack{w \in \mathcal{W}: \\ \tilde{\pi}_w > 0}} \frac{1}{t} \sum_{1 \leq k \leq t} \phi_w(k) = \sup_{t \geq 1} \min_{\substack{w \in \mathcal{W}: \\ \tilde{\pi}_w > 0}} \frac{\mathbb{P}_w(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} \leq t)}{t} . \tag{24}$$

Notice that the escape probability $\tilde{\gamma}_{\mathcal{W}}$ depends only on those rows of the perturbed matrix $\tilde{P}$ whose indices lie in $\mathcal{W}$ (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, one has that $\tilde{\gamma}_{\mathcal{W}} = 0$ if and only if $\mathcal{V} \setminus \mathcal{W}$ is not accessible under $\tilde{P}$ from some state $w \in \mathcal{W}$ such that $\tilde{\pi}_w > 0$.

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 $\tilde{\pi} = \tilde{\pi}\tilde{P}$ an invariant probability measure. Then,*

$$\tilde{\pi}(\mathcal{W}) \leq \frac{1}{\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}}} \,, \tag{25}$$

for all $\mathcal{W} \subseteq \mathcal{V}$.

Proof Observe that, for $k \geq 1$ and $w \in \mathcal{W}$,

$$\sum_v \tilde{P}_{wv} \tau^v_{\mathcal{W}} = \sum_{u \in \mathcal{V}\setminus\mathcal{W}} \tilde{P}_{wu} \tau^u_{\mathcal{W}} \geq \tau^*_{\mathcal{W}} \phi_w(1) \,. \tag{26}$$

Then, it follows from Kac's formula (7) applied to $\tilde{P}$ and $\tilde{\pi}$, the identity (23), and the inequality (26), that

$$\frac{1}{\tilde{\pi}(\mathcal{W})} - 1 = \frac{1}{\tilde{\pi}(\mathcal{W})} \sum_w \sum_v \tilde{\pi}_w \tilde{P}_{wv} \tau^v_{\mathcal{W}} \geq \frac{\tau^*_{\mathcal{W}}}{\tilde{\pi}(\mathcal{W})} \sum_w \tilde{\pi}_w \phi_w(1) \,. \tag{27}$$

Now, observe that

$$\sum_{w' \in \mathcal{W}} \tilde{\pi}_{w'} \tilde{P}_{w'w} \leq \sum_{v \in \mathcal{V}} \tilde{\pi}_v \tilde{P}_{vw} = \tilde{\pi}_w \,, \qquad w \in \mathcal{W} \,.$$

Then, for all $k \geq 1$, one gets that

$$\sum_{w'} \tilde{\pi}_{w'} \phi_{w'}(k+1) = \sum_{w'} \sum_w \tilde{\pi}_{w'} \tilde{P}_{w'w} \phi_w(k) \leq \sum_w \tilde{\pi}_w \phi_w(k) \,.$$

It follows that, for all $t > 0$,

$$\sum_w \tilde{\pi}_w \phi_w(1) \geq \sum_w \tilde{\pi}_w \cdot \frac{1}{t} \sum_{1 \leq k \leq t} \phi_w(k) \,. \tag{28}$$

Then, (27) and (28) imply that

$$\begin{aligned} \frac{1}{\tilde{\pi}(\mathcal{W})} &\geq \tau^*_{\mathcal{W}} \sum_w \frac{\tilde{\pi}_w}{\tilde{\pi}(\mathcal{W})} \cdot \frac{1}{t} \sum_{k=1}^t \phi_w(k) \\ &\geq \tau^*_{\mathcal{W}} \min_{\substack{w \in \mathcal{W}: \\ \tilde{\pi}_w > 0}} \frac{1}{t} \sum_{k=1}^t \phi_w(k) \,. \end{aligned}$$

Since $t \geq 1$ is arbitrary, the inequality above implies that

$$\frac{1}{\tilde{\pi}(\mathcal{W})} \geq \tau^*_{\mathcal{W}} \cdot \tilde{\gamma}_{\mathcal{W}} \,,$$

thus proving the claim. ■

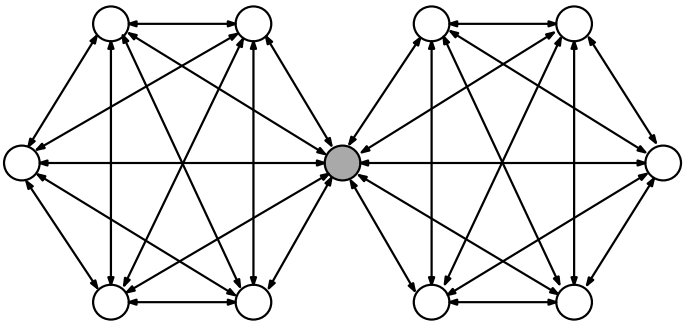

Fig. 2. The graph of Example 2, for $m = 7$. The perturbation set $\mathcal{W} = \{0\}$ is shaded in gray.

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 $\tau_{\rm mix}$, and $\tilde{\pi}$ be an invariant probability measure for $\tilde{P}$. Then,*

$$||\tilde{\pi} - \pi|| \leq \theta \left(\frac{\tau_{\rm mix}}{\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}}} \right) \,,$$

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

Theorem 3 implies that, in order for the total variation distance $||\tilde{\pi} - \pi||$ to be small, it is sufficient that, for some set $\mathcal{W} \supseteq \mathrm{supp}(\tilde{P} - P)$, the ratio

$$\frac{\tau_{\rm mix}}{\tilde{\gamma}_{\mathcal{W}} \cdot \tau^*_{\mathcal{W}}}$$

is small. While the term $1/\tau^*_{\mathcal{W}}$ is a measure of 'how large' the set $\mathcal{W}$ is, hence it is expected to play a central role in estimating $||\tilde{\pi} - \pi||$, one might wonder what the roles of the mixing time $\tau_{\rm mix}$ and of the escape probability $\tilde{\gamma}_{\mathcal{W}}$ are. The following two simple examples show that having control of each of the terms $\tilde{\gamma}_{\mathcal{W}}$ and $\tau_{\rm mix}$ is indeed necessary in order to bound the total variation distance $||\tilde{\pi} - \pi||$.

Example 1. *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 $v \neq w$, for some $\alpha \in (0, 1 - 1/n)$. Then, $\tau_{\rm mix} = 1$, $\tau^*_{\mathcal{W}} = n$, and $\tilde{\gamma}_{\mathcal{W}} = \alpha$, 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 $v \neq w$. Hence,*

$$||\tilde{\pi} - \pi|| = \frac{1 - \alpha - 1/n}{n\alpha + 1}$$

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

Example 2. *For a positive integer m, define the stochastic matrix P on the set $\mathcal{V} := \{-m, -m+1, \ldots, m-1, m\}$ by putting $P_{uv} = 1/m$ if $u \neq v$ and $u \cdot v \geq 0$, $P_{uv} = 0$ if $u \cdot v < 0$ or $u = v$, and $P_{0v} = 1/(2m)$ for all $v \neq 0$. Then, one has that*

$$\pi_0 = \frac{1}{m+1} \,, \qquad \pi_v = \frac{1}{2m+2} \,, \quad v \neq 0 \,.$$

Perturb P on $\mathcal{W} = \{0\}$ by putting, for some $0 < \alpha < 1/2$, $\tilde{P}_{0v} = (1/2 + \alpha \mathrm{sgn}(v))/m$ for $v \neq 0$ and $\tilde{P}_{00} = 0$. Straightforward computations show that $\tau^*_{\mathcal{W}} = m$, while $\tilde{\gamma}_{\mathcal{W}} = 1$. On the other hand, the bottleneck bound [24, Theorem 7.3] implies that $\tau_{\mathrm{mix}} \geq 1/(4\pi_0) \geq m/2$, so that Theorem 3 is useless as it only provides the trivial conclusion that $||\tilde{\pi} - \pi|| \leq 1$. In fact, observe that

$$\tilde{\pi}_v - \pi_v = \frac{\alpha}{m+1} \mathrm{sgn}(v)\,, \qquad v \in \mathcal{V}\,,$$

so that

$$||\tilde{\pi} - \pi|| = m \cdot \frac{\alpha}{m+1}$$

is arbitrarily close to α for large m.

IV. Back to the applications

In this section, we discuss applications of Theorem 3 first to the PageRank manipulation problem, and then to stochastic matrices associated to networks with a finite-dimensional grid structure.

A. PageRank manipulation (continued)

For a stochastic matrix Q, a probability vector 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, one has that $\mathcal{W} := \mathrm{supp}(\tilde{Q} - Q) \supseteq \mathrm{supp}(\tilde{P} - P)$. Moreover, one easily gets the following estimate of the escape probability

$$\tilde{\gamma}_{\mathcal{W}} \geq \min_w \sum_{v \in \mathcal{V} \setminus \mathcal{W}} P_{wv} \geq \beta(1 - \mu(\mathcal{W}))\,. \tag{29}$$

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. Then, starting from the first time they meet, i.e., for

$$t \geq T_c := \inf\{t \geq 0 : U(t) = V(t)\}\,,$$

$U(t) = V(t)$ move together with transition probability matrix P. For every $t \geq 0$ and $u, v \in \mathcal{V}$, [24, Theorem 5.2] implies that

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

so that

$$\tau_{\mathrm{mix}} \leq \left\lceil \frac{-1}{\log(1-\beta)} \right\rceil \leq \frac{1}{\beta} + 1\,. \tag{30}$$

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

$$\tau^v_{\mathcal{W}} \leq \sum_{t \geq 0} (1-\beta)^t \beta (t + \tau^\mu_{\mathcal{W}}) = \frac{1-\beta}{\beta} + \tau^\mu_{\mathcal{W}}\,.$$

Using Kac's formula (7), the above implies that

$$\frac{1}{\pi(\mathcal{W})} = 1 + \sum_w \sum_v \frac{\pi_w}{\pi(\mathcal{W})} P_{wv} \tau^v_{\mathcal{W}} \leq \frac{1}{\beta} + \tau^\mu_{\mathcal{W}}\,.$$

It follows that

$$\tau^*_{\mathcal{W}} \geq \beta \tau^\mu_{\mathcal{W}} \geq \frac{\beta}{\pi(\mathcal{W})} - 1\,. \tag{31}$$

By combining (29), (30), and (31) with Theorem 3, one gets that

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

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).

B. Networks with high local connectivity

Applications of our results to examples like the distributed averaging algorithm with faulty links or to the voter model with influential agents amount to working with perturbations of lazy random walks on graphs, i.e., of stochastic matrices of the form $P = (I+Q)/2$, where I is the identity matrix and Q is the stochastic matrix defined by $Q_{uv} = 1/d_u$ if $(u,v) \in \mathcal{E}$ and $Q_{uv} = 0$ otherwise. The minimal hitting time $\tau^*_{\mathcal{W}}$ can be, in general, difficult to be estimated in typical applications when P is sparse and $\mathcal{W}$ is a small subset of $\mathcal{V}$. In this section, we propose some initial results under two assumptions: one is that the set $\mathcal{W}$ is not only small but localized in the graph. The second one is that the graph has high local connectivity so that removing $\mathcal{W}$ does not drastically alter distances in the remaining part of the graph. The typical graphs for which this holds are the d-dimensional grids (with $d \geq 3$). We believe that both assumptions can be considerably weakened at the price of a deeper analysis. This is the subject of undergoing research which we aim at presenting in another paper.

We start with a simple example to be generalized later on.

Example 3. *For integers $m \geq 2$ and $d \geq 1$, 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., the node set $\mathcal{V} = \mathbb{Z}^d_m$ coincides with the direct product of d copies of the group of integers modulo m, and, for all $u, v \in \mathcal{V}$, $P_{uu} = 1/2$, $P_{uv} = 1/(4d)$ if $\sum_{1 \leq i \leq d} |u_i - v_i| = 1$, and $P_{uv} = 0$ if $\sum_{1 \leq i \leq d} |u_i - v_i| \geq 2$. 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*

$$\tilde{\gamma}_{\mathcal{W}} = 1 - \tilde{P}_{ww}\,.$$

On the other hand, Kac's formula (7) implies that

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

where last equality follows from a basic symmetry argument. Moreover, standard results [24, Theorem 5.5] imply that

$$\tau_{\mathrm{mix}} \leq C_d n^{2/d}$$

for some constant C_d depending on d but not on n. Then, Theorem 3 implies that

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

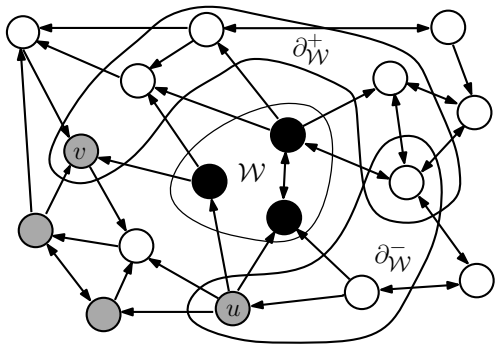

Fig. 3. The external boundaries $\partial^+_{\mathcal{W}}$ and $\partial^-_{\mathcal{W}}$ of a node set $\mathcal{W}$. A simple path in $\mathcal{V} \setminus \mathcal{W}$ from $u \in \partial^-_{\mathcal{W}}$ to $v \in \partial^+_{\mathcal{W}}$ is shaded in gray.

The above guarantees that the total variation distance $||\pi - \tilde{\pi}||$ vanishes as n grows large provided that $d \geq 3$.

In the previous example, $\tau^*_{\mathcal{W}}$ was exactly computed in terms of $\pi(\mathcal{W})$ using Kac's formula (7) and the spatial symmetry in the neighborhood of the perturbed set $\mathcal{W} = \{w\}$. For general $\mathcal{W}$, such symmetry argument breaks down. Below we propose a way to overcome this difficulty in a general situation where $\mathcal{W}$ is localized and its boundary is sufficiently well connected in $\mathcal{V} \setminus \mathcal{W}$. Define the external boundaries of $\mathcal{W}$ as

$$\partial^+_{\mathcal{W}} := \{v \in \mathcal{V} \setminus \mathcal{W} \,:\, P_{wv} > 0 \text{ for some } w \in \mathcal{W}\}\,,$$

$$\partial^-_{\mathcal{W}} := \{v \in \mathcal{V} \setminus \mathcal{W} \,:\, P_{vw} > 0 \text{ for some } w \in \mathcal{W}\}\,.$$

(See Figure 3.) Clearly,

$$\tau^*_{\mathcal{W}} = \min\{\tau^v_{\mathcal{W}} \,:\, v \in \partial^-_{\mathcal{W}}\}\,. \tag{32}$$

On the the other hand, let

$$\tau^\circ_{\mathcal{W}} := \max\{\tau^v_{\mathcal{W}} \,:\, v \in \partial^+_{\mathcal{W}}\}\,, \tag{33}$$

and observe that, from Kac's formula (7),

$$\tau^\circ_{\mathcal{W}} \geq \sum_w \sum_v \frac{\pi_w}{\pi(\mathcal{W})} P_{wv} \tau^v_{\mathcal{W}} = \frac{1}{\pi(\mathcal{W})} - 1\,. \tag{34}$$

Now, for all $u \in \partial^-_{\mathcal{W}}$ and $v \in \partial^+_{\mathcal{W}}$, let $\Gamma_{u,v}$ be the (possibly empty) set of simple paths in $\mathcal{V} \setminus \mathcal{W}$ starting in u and ending in v. For all paths $\xi = (u = \xi_0, \xi_1, \ldots, \xi_l = v) \in \Gamma_{u,v}$, let $P_\xi := \prod_{1 \leq i \leq l} P_{\xi_{i-1}\xi_i}$. Define

$$\lambda_{\mathcal{W}} := \min_{u,v} \max_{\xi \in \Gamma_{u,v}} P_\xi\,, \tag{35}$$

where the minimization is intended to run over all $u \in \partial^+_{\mathcal{W}}$ and $v \in \partial^-_{\mathcal{W}}$ such that $u \neq v$, and we use the convention that the minimum over an empty set equals 1, and the maximum over an empty set equals 0. Then, the following result holds true.

Proposition 4. *Let P be an irreducible stochastic matrix on a finite set $\mathcal{V}$, and π its invariant probability vector. Then, for all $\mathcal{W} \subseteq \mathcal{V}$,*

$$\tau^*_{\mathcal{W}} \geq \lambda_{\mathcal{W}} \left(\frac{1}{\pi(\mathcal{W})} - 1\right)\,,$$

where $\lambda_{\mathcal{W}}$ is defined as in (35).

Proof Let $u \in \partial^-_{\mathcal{W}}$ and $v \in \partial^+_{\mathcal{W}}$ be such that $\tau^u_{\mathcal{W}} = \tau^*_{\mathcal{W}}$ and $\tau^v_{\mathcal{W}} = \tau^\circ_{\mathcal{W}}$. For $\xi = (u = \xi_0, \xi_1, \ldots, \xi_{l-1}, \xi_l = v) \in \Gamma_{u,v}$, let $\mathbb{1}_\xi$ be the indicator function of the event $\cap_{t=0}^l \{V(t) = \xi_t\}$. I.e., $\mathbb{1}_\xi$ is the indicator function of the event that the first l steps of the Markov chain $V(t)$ started at u are along the path ξ. Then,

$$\tau^*_{\mathcal{W}} = \tau^u_{\mathcal{W}} \geq \mathbb{E}_u[T_{\mathcal{W}} \mathbb{1}_\xi] = P_\xi(\tau^v_{\mathcal{W}} + l) \geq P_\xi \tau^\circ_{\mathcal{W}}\,. \tag{36}$$

The claim now follows from (34), (36), and the arbitrariness of $\xi \in \Gamma_{u,v}$. ■

The above result turns out to be useful in those contexts where the set $\mathcal{W}$ is sufficiently localized so that its boundary is tightly connected outside of $\mathcal{W}$ and $\lambda_{\mathcal{W}}$ remains bounded away from 0.

Example 4. *Let P be the lazy simple random walk on a d-toroidal grid as in Example 3 and let $\mathcal{W} = \prod_{i=1}^d [\alpha_i, \alpha_i + s - 1]$ be a hypercube. It is immediate to check that any pair of nodes in $\partial^+_{\mathcal{W}} = \partial^-_{\mathcal{W}}$ can be connected by a path of length $s + d$ outside $\mathcal{W}$, so that $\lambda_{\mathcal{W}} \geq (4d)^{-(s+d)}$. On the other hand, $n\pi(\mathcal{W}) = |\mathcal{W}| = s^d$, so that Proposition 4 implies that*

$$\tau^*_{\mathcal{W}} \geq \frac{1}{(4d)^{(s+d)}} \left(\frac{1}{\pi(\mathcal{W})} - 1\right) = \frac{1}{(4d)^{(s+d)}} \left(\frac{n}{s^d} - 1\right)\,.$$

Since the mixing time satisfies $\tau_{\mathrm{mix}} \leq C_d n^{2/d}$ for some positive constant C_d independent from n [24, Theorem 5.5], we have that

$$\frac{\tau_{\mathrm{mix}}}{\tau^*_{\mathcal{W}}} \leq C'_d \frac{(4d)^s s^d}{1 - s^d/n} n^{2/d - 1}\,,$$

with $C'_d := C_d (4d)^d$.

It remains to be estimated the escape probability from $\mathcal{W}$ which is the (only) term depending on the perturbation. Assume that $\tilde{P}$ is irreducible, and put

$$\delta = \min\left\{\tilde{P}_{wv} \,:\, w \in \mathcal{W}\,,\ \tilde{P}_{wv} > 0\right\}\,.$$

Since from every $w \in \mathcal{W}$ there is a path leading to $\partial\mathcal{W}$ of length at most $|\mathcal{W}| = s^d$, one gets that

$$\tilde{\gamma}_{\mathcal{W}} \geq \min_{w:\, \tilde{\pi}_w > 0} \frac{1}{s^d} \mathbb{P}_w(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} \leq s^d) \geq \frac{\delta^{s^d}}{s^d}\,.$$

Multiplying the two estimations and noting that the dominating term in the size of the perturbation is given by δ^{-s^d}, we immediately obtain from Theorem 3 that, if

$$\limsup_n \frac{|\mathcal{W}|}{\log n} < \frac{d-2}{d \log \delta^{-1}}$$

then $||\tilde{\pi} - \pi|| \to 0$ as n grows large.

V. Conclusion

Invariant probability vectors of stochastic matrices play a central role in a large number of multi-agent network models including distributed averaging algorithms, opinion dynamics, and centrality measures such as PageRank. This paper investigates the fundamental question of how such invariant probability vectors are resilient to perturbations of the network. The main result provides an estimate of the total

variation distance between the invariant probability vectors of two stochastic matrices in terms of the mixing time of one of the matrices and of the size of the perturbation set $\mathcal{W}$ measured as the product of two quantities: the entrance time on $\mathcal{W}$ and the escape probability from $\mathcal{W}$. Explicit applications to network models have also been discussed in detail. Among the relevant issues which have not been addressed by this paper and deserve to be considered for future research are:

- The estimation of the entrance time of the perturbation set remains the most challenging problem in applying our result. In particular, we would like to extend our estimation to small but scattered perturbation sets as well to other general classes of networks such as locally tree-like graphs.
- In many applications of network centrality, the total variation distance between two probability vectors may not be the most relevant measure of the effect of a perturbation. E.g., the maximal ratio of the centralities assigned to the same node in the unperturbed and in the perturbed network would be of great potential interest in such cases.
- When a network is perturbed locally, we expect the effect of the perturbation to decay as a function of the distance from the perturbation set. This is not captured by the total variation analysis and may require an essentially different approach.

ACKNOWLEDGEMENTS

The first author is a member of the LCCC Linnaeus Center and the ELLIIT Excellence Center at Lund University. His research has been partially supported by the Swedish Research Council (VR) through the the junior research grant Information Dynamics in Large-Scale Networks.

REFERENCES

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

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

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

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

[5] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont, MA, 1997.

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

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

[8] G.E. Cho and C.D. Meyer, *Markov chain sensitivity measured by mean first passage times*, Linear Algebra and its Applications **316** (2000), 21–28.

[9] ———, *Comparison of perturbation bounds for the stationary distribution of a Markov chain*, Linear Algebra and its Applications **335** (2001), 137–150.

[10] F. Chung, *The heat kernel as the pagerank of a graph*, Proceedings of the National Academy of Science **104** (2007), no. 50, 19735–19740.

[11] P. Connely and D. Welsh, *Finite particle systems and infection models*, Mathematical Proceedings of the Cambridge Philosophical Society **94** (1982), 167–182.

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

[13] B.C. Csáji, R.M. Jungers, and V.D. Blondel, *PageRank optimization by edge selection*, Applied Discrete Mathematics (2013).

[14] F. Cucker and S. Smale, *The mathematics of emergence*, Japanese Journal of Mathematics **2** (2007), no. 1, 197–227.

[15] M.H. DeGroot, *Reaching a consensus*, Journal of the American Statistical Association **69** (1974), no. 345, 118–121.

[16] P.M. DeMarzo, D. Vayanos, and J. Zwiebel, *Persuasion bias, social influence, and unidimensional opinions*, The Quarterly Journal of Economics **118** (2003), no. 3, 909–968.

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

[18] B. Golub and M.O. Jackson, *Naive learning in social networks and the wisdom of crowds*, American Economic Journal: Microeconomics **2** (2010), no. 1, 112–149.

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

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

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

[22] C. De Kerchove, L. Ninove, and P. Van Dooren, *Maximizing PageRank via outlinks*, Linear Algebra and its Applications **429** (2008), 1254–1276.

[23] A.N. Langville and C.D. Meyer, *Google's pagerank and beyond: The science of search engine rankings*, Princeton University Press, 2006.

[24] D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2010.

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

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

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

[28] C.D. Meyer, *Sensitivity of the stationary distribution of a Markov chain*, SIAM Journal on Matrix Analysis and Applications **15** (1994), 715–728.

[29] A. Mitrophanov, *Stability and exponential convergence of continuous-time Markov chains*, Journal of Applied Probability **40** (2003), 970–979.

[30] ———, *Sensitivity and convergence of uniformly ergodic Markov chains*, Journal of Applied Probability **42** (2005), 1003–1014.

[31] M. Mobilia, *Does a single zealot affect an infinite group of voters?*, Physical Review Letters **91** (2003), no. 2, 028701.

[32] M. Mobilia and I.T. Georgiev, *Voting and catalytic processes with inhomogeneities*, Physical Review E **71** (2005), 046102.

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

[34] R. Olfati-Saber and R.M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control **49** (2004), no. 9, 1520–1533.

[35] A. Olshevsky and J.N. Tsitsiklis, *Convergence speed in distributed consensus and averaging*, SIAM Review **53** (2011), no. 4, 747–772.

[36] L. Qun and D. Rus, *Global clock synchronization in sensor networks*, IEEE Journal on Computers **55** (2006), no. 2, 214–226.

[37] P. J. Schweitzer, *Perturbation theory and finite Markov chains*, Journal of Applied Probability **5** (1968), 401–413.

[38] E. Seneta, *Perturbation of the stationary distribution measured by ergodicity constraints*, Advances in Applied Probability **20** (1988), 228–230.

[39] ———, *Sensitivity of finite Markov chains under perturbation*, Statistics & Probability Letters **17** (1993), 163–168.

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

[41] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

[42] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.

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