Submitted to the Annals of Applied Probability

ROBUSTNESS OF LARGE-SCALE STOCHASTIC MATRICES TO LOCALIZED PERTURBATIONS

By Giacomo Como and Fabio Fagnani

Lund University and Politecnico di Torino

Upper bounds are derived on the total variation distance between the invariant distributions of two stochastic matrices differing on a subset $\mathcal{W}$ of rows. Such bounds depend on three parameters: the mixing time and the minimal expected hitting time on $\mathcal{W}$ for the Markov chain associated to one of the matrices; and the escape time from $\mathcal{W}$ 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 and consensus algorithms, and interacting particle systems.

1. 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 [6] 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 [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

AMS 2000 subject classifications: Primary 60J10; secondary 15B51, 93C73, 93A15, 68U35, 68U35

Keywords and phrases: Stochastic matrices, stationary probability distributions, robustness, large-scale networks

imsart-aap ver. 2010/08/03 file: "perturbation rapid mixing 5".tex date: August 30, 2013

for accidental variations of the WWW topology occurring, e.g., because of server failures or network congestion problems [20].

More generally, the problem is of central interest 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 analogous, 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-Frobenious 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 [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 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' [31, 32].

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 [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

$$||\tilde{\pi} - \pi|| \leq \theta(\tau\tilde{\chi}/\tau^*_{\mathcal{W}}) \,, \tag{2}$$

(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);

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

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 [1] 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 [3, 33, 24], 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 [29, 30], 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 [38, 39] estimating the condition number κ_P in terms of ergodicity coefficients. Also the estimates proposed in [2] for symmetric P, which 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 upper 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 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 2 introduces three motivating examples formalizing some of the applications mentioned at the beginning of this Introduction. In Section 3, we present our main result which is stated Theorem 3. Section 4 focuses on stochastic matrices whose support graph has high local connectivity and discusses lower bounds of the minimal hitting time $\tau^*_{\mathcal{W}}$. This allows for efficient application of Theorem 3 to grid-like graphs. Explicit examples on toroidal grid graphs are presented.

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 $\mathrm{supp}(A) := \{v :\ A_{v,\cdot} \neq 0\}$ 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., $\mu^{\mathcal{A}}_a = \mu_a/\mu(\mathcal{A})$ for $a \in \mathcal{A}$, and $\mu^{\mathcal{A}}_v = 0$ 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 $\tilde{V}(t)$, $t = 0, 1, \ldots$, 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 $\tau^v_{\mathcal{U}} := \mathbb{E}_v[T_{\mathcal{U}}]$ and $\tilde{\tau}^v_{\mathcal{U}} := \mathbb{E}_v[\tilde{T}_{\mathcal{U}}]$, respectively.

2. Three motivating examples.

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

2.1. *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 $\mathcal{E}_u := \{v : (u, v) \in \mathcal{E}\}$ are 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 [6] 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 [19], 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. Then, the unique invariant probability distribution $\tilde{\pi}$ of $\tilde{P} := (1-\beta)\tilde{Q} + \beta\mathbb{1}\mu'$ satisfies

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

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 2.1 of Section 3.

2.2. *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 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 [35] 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

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

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

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

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

so that

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

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

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

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

$$|\varepsilon_1| \le \frac{|\mathcal{F}|}{n}||y||_\infty\,, \quad |\varepsilon_2| \le \overline{d}||y||_\infty||\tilde{\pi} - \pi||\,, \quad |\varepsilon_3| \le \frac{|\mathcal{F}|}{n}\,, \quad |\varepsilon_4| \le \overline{d}||\tilde{\pi} - \pi||\,.$$

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 $||\tilde{\pi} - \pi|| = o(1)$.

2.3. *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 $(v,u) \in \mathcal{E}$. This is an instance of the voter model []. 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 $X_v(t) \stackrel{t\to\infty}{\longrightarrow} Y$ 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 $\mathbb{E}[X(t)|X(0)] = P^t X(0) \stackrel{t\to\infty}{\longrightarrow} \mathbb{1}\pi' X(0)$, 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 $\mathcal{W} := \{u : (v,u) \in \mathcal{H} \text{ for some } v\}$ 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 continuous opinion dynamics. Observe that strong connectivity of $\tilde{\mathcal{G}}$ implies that, with probability one $\tilde{X}_v(t) \stackrel{t\to\infty}{\longrightarrow} \tilde{Y}$ for all v, where $Y \in \{0,1\}$ 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?

3. 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

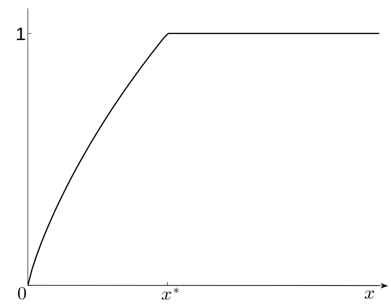

FIG 1. *Graph of the function $\theta(x)$ defined in (6).*

$$\theta : [0, +\infty) \to [0, 1] , \qquad \theta(x) := \begin{cases} x \ln\left(e^2/x\right) & x \le x^* \\ 1 & x \ge x^* , \end{cases} \tag{6}$$

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,*

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

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

PROOF. Let $V(t)$, $\tilde{V}(t)$ 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 since. Since P and $\tilde{P}$ coincide on $\mathcal{V} \setminus \mathcal{W}$, one has that $V(t)$ and $\tilde{V}(t)$ are Markov chains with transition probability matrix P and $\tilde{P}$, respectively. Then, for all $\mathcal{A} \subseteq \mathcal{V}$, and $t \ge 0$, one has that

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

where the first identity uses the invariance of $\tilde{\pi}$, and the last inequality follows from $||(\mu' P^t)' - \pi|| \le \exp(-\lfloor t/\tau \rfloor)$ (which is a standard consequence of the submultiplicativity of the maximal total variation distance, see, e.g., (4.31) in [24]) and a straightforward union bound $\mathbb{P}_{\tilde\pi}(\tilde T_{\mathcal W} < t) \le \sum_{i=0}^{t-1}\mathbb{P}_{\tilde\pi}(\tilde V(i)\in\mathcal W) = t\tilde\pi(\mathcal W)$. Therefore,

$$||\tilde\pi - \pi|| = \max_{\mathcal A\subseteq\mathcal V}\{\tilde\pi(\mathcal A)-\pi(\mathcal A)\} \le \exp(-\lfloor t/\tau\rfloor) + t\tilde\pi(\mathcal W)\,, \qquad \forall t\ge 0\,.$$

The claim now follows by choosing $t=\max\{\lfloor\tau\log\big((\tau\tilde\pi(\mathcal W))^{-1}e\big)\rfloor, 0\}$, such a choice being suggest by the minimization of $x\mapsto\exp(-x/\tau-1)+x\tilde\pi(\mathcal W)$ over continuous nonnegative values of x. ■

Lemma 1 shows that it is sufficient to have an upper bound on the product $\tau\tilde\pi(\mathcal W)$ 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

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

Observe that the minimal hitting time $\tau^*_{\mathcal W}$ 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 $\tilde P-P$.

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

$$\tilde\chi := \max_{\tilde\pi_w>0}\inf_{t\ge1}\frac{t}{\mathbb{P}_w(\tilde T_{\mathcal V\setminus\mathcal W}\le t)}\,. \tag{8}$$

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

$$\tilde\chi \le \max_{\tilde\pi_w>0}\frac{2\tilde\tau_{\mathcal V\setminus\mathcal W}}{1-\mathbb{P}_w(\tilde T_{\mathcal V\setminus\mathcal W}>2\tilde\tau^w_{\mathcal V\setminus\mathcal W})} \le 4\max_{\tilde\pi_w>0}\tilde\tau^w_{\mathcal V\setminus\mathcal W}\,,$$

which justifies the choice of the name escape time. The reason for introducing $\tilde\kappa$ instead of using $\max\{\tilde\tau^w_{\mathcal V\setminus\mathcal W}:\tilde\pi_w>0\}$ 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,*

$$\tilde{\pi}(\mathcal{W}) \leq \frac{\tilde{\chi}}{\tau^*_{\mathcal{W}}}\,, \tag{9}$$

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

Proof. For $k \geq 1$, let $\phi_w(k) := \mathbb{P}_w(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} = k)$. From Kac's formula, it follows that

$$\frac{1}{\tilde{\pi}(\mathcal{W})} - 1 = \frac{1}{\tilde{\pi}(\mathcal{W})} \sum_w \sum_v \tilde{\pi}_w \tilde{P}_{wv} \tau^v_{\mathcal{W}} \geq \frac{1}{\tilde{\pi}(\mathcal{W})} \tau^*_{\mathcal{W}} \sum_w \tilde{\pi}_w \phi_w(1)\,. \tag{10}$$

Now, observe that for all $w \in \mathcal{W}$, it holds $\sum_{w' \in \mathcal{W}} \tilde{\pi}_{w'} \tilde{P}_{w'w} \leq \tilde{\pi}_w$. 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 $t > 0$,

$$\sum_w \tilde{\pi}_w \phi_w(1) \geq \frac{1}{t} \sum_{1 \leq k \leq t} \sum_w \tilde{\pi}_w \phi_w(k) \geq \frac{1}{t} \sum_w \tilde{\pi}_w \mathbb{P}_w \left(\tilde{T}_{\mathcal{V}\setminus\mathcal{W}} \leq t \right)\,. \tag{11}$$

The claim now follows from (10), (11), and (8). ■

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,*

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

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 vanish as the network size grows large, it is sufficient that $\tau\tilde{\chi}/\tau^*_{\mathcal{W}}$ vanishes.

Example 1. *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., $\mathcal{V} = \mathbb{Z}^d_m$, $P_{uu} = 1/2$, $P_{uv} = 1/(4d)$ 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*

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

On the other hand, Kac's formula [24, Lemma 21.3] implies that

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

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

$$\tau \leq C_d n^{2/d}$$

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

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

The above guarantees that $||\pi - \tilde{\pi}||$ *vanishes as* n *grows large provided that* $d \geq 3$. *More general examples involving toroidal grids will be discussed in Section 4.*

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 2.1. Let* $\tilde{Q}$ *be a perturbation of* Q, *and* $\tilde{P} = (1-\beta)\tilde{Q} + \beta\mathbb{1}\mu'$. *Clearly* $\mathcal{W} := \mathrm{supp}(\tilde{Q} - Q) \supseteq \mathrm{supp}(\tilde{P} - P)$. *Moreover,*

$$\tilde{\chi} \leq \frac{1}{\max_w \mathbb{P}_w(\tilde{V}(1) \in \mathcal{V}\setminus\mathcal{W})} \leq \frac{1}{\beta(1-\mu(\mathcal{W}))} . \qquad (12)$$

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* $t \geq T_c := \inf\{t \geq 0 : U(t) = V(t)\}$, $U(t) = V(t)$ *move together with transition probability matrix* P. *Since*

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

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

$$\tau \leq \left\lceil \frac{-1}{\log(1-\beta)} \right\rceil \leq \frac{1}{\beta} + 1\,. \tag{13}$$

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

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

Using Kac's formula, the above implies that

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

It follows that

$$\tau_{\mathcal{W}}^{*} \geq \beta\tau_{\mathcal{W}}^{\mu} \geq \frac{\beta}{\pi(\mathcal{W})} - 1\,. \tag{14}$$

By combining (12), (13), and (14) 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).

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 $v \neq w$, 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 $v \neq w$. Hence, $||\tilde{\pi} - \pi|| = (1-\alpha-1/n)(n\alpha+1)$ which shows that $\alpha n \to \infty$ is indeed also a necessary condition for $||\tilde{\pi} - \pi|| \to 0$ as n grows large.*

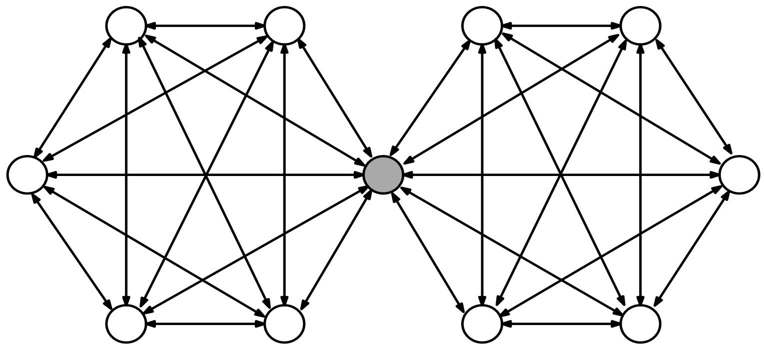

FIG 2. *The graph of Example 4, for $m = 7$. The perturbation set $\mathcal{W} = \{0\}$ is shaded in gray.*

EXAMPLE 4. *For $m > 1$, let $\mathcal{V} := \{-m, -m+1, \dots, m-1, m\}$ 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 [24, 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 $\tilde{\pi}_v - \pi_v = 2\alpha/(2m+1)sgn(v)$, which is arbitrarily close to 1 for large n and α close to $1/2$.*

4. Networks with high local connectivity. The minimal hitting time $\tau^*_{\mathcal{W}}$ can be, in general, the most difficult quantity to estimate in typical applications when P is sparse and $\mathcal{W}$ is a small subset of $\mathcal{V}$. In this section, we propose some results for graphs with high local connectivity where removing $\mathcal{W}$ does not drastically alter distances in the remaining part of the graph. We start with a result which turns out to be useful for localized perturbations, and we then propose a result for reversible stochastic matrices but with perturbations not necessarily localized. Such results find natural applications in structured graphs like d-dimensional toroidal grids (with $d \geq 3$) in contexts when $\mathcal{W}$ remains of bounded cardinality (or increases in a sub logarithmic way) with respect to the size of the graph approaching infinity.

4.1. *A simple bound for localized perturbations.* We start by considering a relatively simple case when $\mathcal{W}$ is localized and its boundary is sufficiently well connected in $\mathcal{V} \setminus \mathcal{W}$. Define the external boundaries of $\mathcal{W}$ as

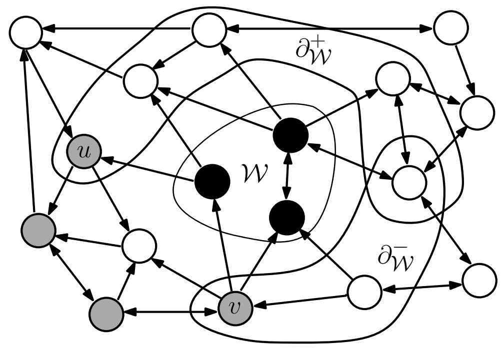

FIG 3. *The external boundaries $\partial^+_{\mathcal{W}}$ and $\partial^-_{\mathcal{W}}$ of a node set $\mathcal{W}$. A simple path in $\mathcal{V}\setminus\mathcal{W}$ from $u \in \partial^+_{\mathcal{W}}$ to $v \in \partial^-_{\mathcal{W}}$ is shaded in gray.*

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

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

(See Figure 3.) Clearly,

$$\tau^*_{\mathcal{W}} = \min\{\tau^v_{\mathcal{W}} : v \in \partial^-_{\mathcal{W}}\}\,. \tag{15}$$

On the the other hand, let

$$\tau^\circ_{\mathcal{W}} := \max\{\tau^v_{\mathcal{W}} \,:\, v \in \partial^+_{\mathcal{W}}\}\,, \tag{16}$$

and observe that, from Kac's formula,

$$\tau^\circ_{\mathcal{W}} \geq \sum_w \sum_v \frac{\pi_w}{\pi(\mathcal{W})} P_{wv} \tau^v_{\mathcal{W}} = \frac{1}{\pi(\mathcal{W})} - 1\,. \tag{17}$$

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 $\gamma = (u = v_0, v_1, \ldots, v_l = v) \in \Gamma_{u,v}$, let $P_\gamma := \prod_{1\leq i\leq l} P_{v_{i-1}v_i}$. Define

$$\lambda_{\mathcal{W}} := \min_{u,v} \max_{\gamma\in\Gamma_{u,v}} P_\gamma\,, \tag{18}$$

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. The following holds

LEMMA 4. *Let P be an irreducible stochastic matrix on a finite set $\mathcal{V}$, and π its invariant probability distribution. 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 (18).

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 every $\gamma = (u = v_0, v_1, \ldots, v_{l-1}, v_l = v) \in \Gamma_{u,v}$, let $\mathbb{1}_\gamma$ be the indicator function of the event $\cap_{t=0}^{l} \{V(t) = v_t\}$. Then,

$$\tau^*_{\mathcal{W}} = \tau^u_{\mathcal{W}} \geq \mathbb{E}_u[T_{\mathcal{W}} \mathbb{1}_\gamma] = P_\gamma(\tau^v_{\mathcal{W}} + l) \geq P_\gamma \tau^\circ_{\mathcal{W}} \,. \tag{19}$$

The claim now follows from (17), (19), and the arbitrariness of γ. ∎

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 5. *Let P be the lazy simple random walk on a d-toroidal grid as in Example 1 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 Lemma 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, by [24, Theorem 5.5], $\tau \leq C_d n^{2/d}$ for some positive constant C_d independent from n, we have that

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

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

We now consider the escape time 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

$$\tilde{\chi} \leq s^d \delta^{-s^d} .$$

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.

4.2. *Estimations through the effective resistance for reversible matrices.* We now extend Lemma 4 to sets $\mathcal{W}$ which are not necessary localized. Throughout this subsection, we shall restrict to the case when the stochastic matrix P is reversible, i.e., when $\pi_u P_{uv} = \pi_v P_{vu}$ for all $u, v \in \mathcal{V}$. Observe that reversibility implies that $\partial^+_{\mathcal{W}} = \partial^-_{\mathcal{W}} =: \partial_{\mathcal{W}}$. Consider the following modification of (18):

$$\rho_{\mathcal{W}} = \min_{w \in \mathcal{W}} \min_{\substack{u \neq v \in \partial_{\mathcal{W}}: \\ P_{wu} P_{wv} > 0}} \max_{\gamma \in \Gamma_{u,v}} \mathbb{P}_\gamma \,, \tag{20}$$

and observe that $\rho_{\mathcal{W}} \geq \lambda_{\mathcal{W}}$.

For $u \in \mathcal{V}$, let $T_u^+ := \inf\{t \geq 1 : V(t) = u\}$ be the return time, and let

$$R_{\text{eff}}(u \leftrightarrow v) := \frac{1}{n \pi_u \mathbb{P}_u(T_v < T_u^+)} \,,$$

be the effective resistance between u and v. Let

$$R_{\text{eff}}^{\max}(P) := \max\{R_{\text{eff}}(u \leftrightarrow v) : u \neq v \in \mathcal{V}\}$$

denote the maximal effective resistance. Then, the following result holds.

Lemma 5. *Let P be an irreducible reversible stochastic matrix over a finite set $\mathcal{V}$, and let π be its invariant probability distribution. Then,*

$$\tau^*_{\mathcal{W}} \geq \min\left\{\rho_{\mathcal{W}}, \frac{1}{n\pi(\mathcal{W}) R_{\text{eff}}^{\max}(P)}\right\}^{|\partial_{\mathcal{W}}|+1} \left(\frac{1}{\pi(\mathcal{W})} - 1\right) \,,$$

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

Proof. Let $\tau^*_{\mathcal{W}}$ and $\tau^\circ_{\mathcal{W}}$ be as defined by (7) and (16), respectively, and

$$\theta := (\tau^\circ_{\mathcal{W}} / \tau^*_{\mathcal{W}})^{1/(|\partial_{\mathcal{W}}|+1)} \,.$$

Observe that (15) implies that there exists at least one $k \in \{0, \ldots, |\partial_{\mathcal{W}}|\}$ such that $\tau^v_{\mathcal{W}} \notin (\tau^*_{\mathcal{W}} \theta^k, \tau^*_{\mathcal{W}} \theta^{k+1})$ for all $v \in \partial_{\mathcal{W}}$. Fix any such k, and define

$$\partial_1 := \left\{v \in \partial_{\mathcal{W}} : \tau^v_{\mathcal{W}} \geq \tau^*_{\mathcal{W}} \theta^{k+1}\right\} \,, \qquad \partial_2 := \left\{v \in \partial_{\mathcal{W}} : \tau^v_{\mathcal{W}} \leq \tau^*_{\mathcal{W}} \theta^k\right\} \,,$$

$$\mathcal{W}_2 := \{w \in \mathcal{W} \mid P_{wv} > 0 \text{ for some } v \in \partial_2\}\,.$$

If there exists $w \in \mathcal{W}_2$ such that $P_{wv_1} > 0$ for some $v_1 \in \partial_1$, then arguing as in the proof of Lemma 4 one gets that $1/\theta \geq \rho$, which in turn yields

$$\tau^*_{\mathcal{W}} \geq \tau^\circ_{\mathcal{W}} \rho_{\mathcal{W}}^{|\partial_{\mathcal{W}}|+1}\,. \tag{21}$$

If, instead, $P_{wv} = 0$ for every $w \in \mathcal{W}_2$ and $v \in \partial_1$, then, for every $v_2 \in \partial_2$, one has that

$$\begin{aligned} \tau^{v_2}_{\mathcal{W}} &\geq \sum_{v_1 \in \partial_1} \mathbb{E}_{v_2}[T_{\mathcal{W}} | T_{\partial_1} = T_{v_1} < T_{\mathcal{W}_2}] \mathbb{P}_{v_2}(T_{\partial_1} = T_{v_1} < T_{\mathcal{W}_2}) \\ &\geq \sum_{v_1 \in \partial_1} \tau^{v_1}_{\mathcal{W}} \mathbb{P}_{v_2}(T_{\partial_1} = T_{v_1} < T_{\mathcal{W}_2}) \\ &\geq \tau^*_{\mathcal{W}} \theta^{k+1} \mathbb{P}_{v_2}(T_{\partial_1} < T_{\mathcal{W}_2})\,. \end{aligned} \tag{22}$$

Let now $\pi^{\mathcal{W}_2}$ be π conditioned on $\mathcal{W}_2$Since

$$\mathbb{P}_{\pi^{\mathcal{W}_2}}(T_{\partial_1} < T^+_{\mathcal{W}_2}) = \sum_{v_2 \in \partial_2} \sum_{w \in \mathcal{W}_2} \frac{\pi_w}{\pi(\mathcal{W}_2)} P_{wv_2} \mathbb{P}_{v_2}(T_{\partial_1} < T^+_{\mathcal{W}_2})$$

there exists some $v_2 \in \partial_2$ such that $\mathbb{P}_{v_2}(T_{\partial_1} < T^+_{\mathcal{W}_2}) \geq \mathbb{P}_{\pi^{\mathcal{W}_2}}(T_{\partial_1} < T^+_{\mathcal{W}_2})$. Using (22), one gets $\tau^*_{\mathcal{W}}\theta^k \geq \tau^*_{\mathcal{W}}\theta^{k+1}\mathbb{P}_{\pi_{W_2}}(\tau_{\partial W_1} < \tau^+_{W_2})$, so that

$$\tau^*_{\mathcal{W}} \geq \tau^\circ_{\mathcal{W}} \mathbb{P}_{\pi_{\mathcal{W}_2}}(T_{\partial_1} < T^+_{\mathcal{W}_2})^{|\partial_{\mathcal{W}}|+1}\,. \tag{23}$$

Using techniques of electrical networks (essentially Thompson's principle [24, Theorem 9.10]), one can prove that

$$\mathbb{P}_{\pi_{\mathcal{W}_2}}(T_{\partial_1} < T^+_{\mathcal{W}_2}) \geq \frac{1}{n\pi(\mathcal{W}_2) R^{\max}_{\text{eff}}(P)}\,. \tag{24}$$

(For the sake of completeness, a proof is sketched in Sect. **??**.) The claim now follows by applying the above to the righthand side of (23), and combining that with (21). ∎

It worth pointing out that reversibility of P was used only for proving (24). A critical look at the proof of Lemma 5 also reveals that the choose $\theta = (\tau^\circ/\tau^*)^{1/|\partial_{\mathcal{W}}|}$, sufficient to guarantee there exists in interval $(A, A\theta)$ to which no $\tau^v_{\mathcal{W}}$ belongs, could result quite conservative, and be improved upon if more information is available.

Example 6. *For two integers $d \geq 3$ and $m \geq 2$, let P be the lazy simple random walk on a d-toroidal grid of size $n = m^d$ as in Examples 1 and 5* *and let $\mathcal{W} \subseteq \mathbb{Z}_m^d$ be any subset. We say that $R = \prod_{i=1}^d [\alpha_i, \beta_i] \subseteq \mathbb{Z}_m^d$ is a separating rectangle for $\mathcal{W}$ if the augmented rectangle $R' := \prod_{i=1}^d [\alpha_i - 1, \beta_i + 1] \subseteq \mathbb{Z}_m^d$ is such that $(R' \setminus R) \cap \mathcal{W} = \emptyset$. A separating rectangle R is said minimal if no strictly included rectangle is separating. It is immediate to verify that there exists a family $\{R^k = \prod_{i=1}^d [\alpha_i^k, \beta_i^k]\}_{1 \le k \le r}$ of minimal separating rectangles such that $\mathcal{W} \subseteq \mathcal{W}' := \cup_k R_k$. As a consequence of minimality one gets that*

$$\max_{i,k} \left\{ |\beta_i^k - \alpha_i^k| \right\} + 1 \le \max_k |\mathcal{W} \cap R_k| \le |\mathcal{W}|$$

It then follows that

$$\rho_{\mathcal{W}'} \ge \frac{1}{(4d)^{d|\mathcal{W}|}}$$

Similarly, one gets that $n\pi(\mathcal{W}') = |\mathcal{W}'| \le |\mathcal{W}|^d$ and $|\partial_{\mathcal{W}'}| \le (|\mathcal{W}| + 1)2d$. On the other hand, it is known (see, e.g., [24, Proposition 10.13]) that, in dimension $d \ge 3$,

$$R_{\text{eff}}^{\max}(P) \le k_d \,,$$

for some constant k_d independent of n. Then, Lemma 5 implies that, for large enough n,

$$\tau^*_{\mathcal{W}'} \ge \frac{1}{(4d)^{2d^2|\mathcal{W}|(|\mathcal{W}|+1)}} \left(\frac{n}{|\mathcal{W}|^d} - 1 \right) .$$

Then, arguing as in Example 5, one gets that if

$$\limsup_n \frac{|\mathcal{W}|}{(\log n)^{1/2}} < \left(\frac{d-2}{4d^3 \log(4d)} \right)^{1/2}$$

then $||\tilde{\pi} - \pi|| \to 0$ as n grows large. As compared to Example 5, where $\mathcal{W}$ was assumed to be a rectangle, the admissible growth rate of $\mathcal{W}$ has of factor $\sqrt{\log n}$.

4.3. *Proof of* (24).

LEMMA 6. *Let P be an irreducible reversible stochastic matrix over a finite set $\mathcal{V}$, and let π be its invariant probability distribution. Then, for every nonempty $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$,*

$$\mathbb{P}_{\pi^{\mathcal{B}}} \left(T_{\mathcal{A}} < T_{\mathcal{B}}^+ \right) \ge \frac{1}{n\pi(\mathcal{B}) R_{\text{eff}}^{max}(P)} \,, \tag{25}$$

where $T_{\mathcal{B}}^+ := \inf\{t \ge 1 : V(t) \in \mathcal{B}\}$ be the return time.

PROOF. Fix $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Put $\bar{\mathcal{V}} := (\mathcal{V} \setminus \mathcal{B}) \cup \{b\}$ and consider the map $\varphi : \mathcal{V} \to \bar{\mathcal{V}}$ which is the identity on $\mathcal{V} \setminus \mathcal{B}$ and maps $\mathcal{B}$ in b. For a matrix $M \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$, let $\varphi_{\#}M \in \mathbb{R}^{\overline{\mathcal{V}}\times\overline{\mathcal{V}}}$ be the matrix with entries defined by

$$(\varphi_{\#}M)_{\bar{v}\bar{v}'} = \sum_{\varphi(v)=\bar{v},\varphi(v')=\bar{v}'} M_{vv'} \,.$$

Let $C \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ be the conductance matrix with entries $C_{uv} = \pi_u P_{uv} n$, and $\bar{C} := \varphi_{\#}C$. Let $\bar{P}$ be the stochastic matrix on $\bar{\mathcal{V}}$ defined by $\overline{P}_{uv} = \overline{C}_{uv}/C_u$, where $\overline{C}_u := \sum_{\overline{v}\in\overline{\mathcal{V}}} \overline{C}_{u\overline{v}}$. Use $\bar{\mathbb{P}}$ to denote probabilities with respect to the Markov chain with transition probability matrix $\bar{P}$. A straightforward computation shows that

$$\bar{\mathbb{P}}_b(T_a < T_b^+) = \mathbb{P}_{\pi_{\mathcal{B}}}(T_a < T_{\mathcal{B}}^+) \leq \mathbb{P}_{\pi^{\mathcal{B}}}\left(T_{\mathcal{A}} < T_{\mathcal{B}}^+\right) \,. \tag{26}$$

Using the electrical interpretation we can write

$$\bar{\mathbb{P}}_b(T_a < T_b^+) = \frac{1}{\bar{C}_b \bar{R}_{\text{eff}}(b \leftrightarrow a)} \,, \tag{27}$$

where $\bar{R}_{\text{eff}}(b \leftrightarrow a)$ is the effective resistance between b and a on the electrical network induced by $\bar{C}$. Let $\lambda \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ be a unit flow for P from b to a (as defined in [24, Section 9.3]), and $\overline{\lambda} = \varphi_{\#}\lambda \in \mathbb{R}^{\overline{\mathcal{V}}\times\overline{\mathcal{V}}}$. It is immediate to check that $\overline{\lambda}$ is a unitary flow for $\overline{P}$ from b to a. Moreover,

$$\begin{aligned}
|\overline{\lambda}| &= \frac{1}{2}\sum_{\overline{v},\overline{v}'} \frac{1}{\overline{C}_{\overline{v},\overline{v}'}} \overline{\lambda}^2_{\overline{v},\overline{v}'} \\
&= \frac{1}{2}\sum_{\overline{v},\overline{v}'} \frac{1}{\overline{C}_{\overline{v},\overline{v}'}} \left(\sum_{v,v':\varphi(v)=\overline{v},\varphi(v')=\overline{v}'} C_{vv'}^{1/2} \frac{\lambda_{vv'}}{C_{vv'}^{1/2}} \right)^2 \\
&\leq \frac{1}{2}\sum_{\overline{v},\overline{v}'} \frac{1}{\overline{C}_{\overline{v},\overline{v}'}} \sum_{v,v':\varphi(v)=\overline{v},\varphi(v')=\overline{v}'} C_{vv'} \sum_{v,v':\varphi(v)=\overline{v},\varphi(v')=\overline{v}'} \frac{\lambda_{vv'}^2}{C_{vv'}} = |\lambda|
\end{aligned}$$

Using Thompson's principle [24, Theorem 9.10], this yields $\bar{R}_{\text{eff}}(b \leftrightarrow a) \leq R_{\text{eff}}(b \leftrightarrow a)$, thus completing the proof. ■

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.

Department of Automatic Control,
Lund University,
BOX 118, 22100, Lund, Sweden
E-mail: giacomo.como@control.lth.se

Dipartimento di Matematica,
Politecnico di Torino,
Corso Stati Uniti 24, 10129, Torino, Italy
E-mail: fabio.fagnani@polito.it