Exact Rate for Convergence in Probability of Averaging Processes via Generalized Min-Cut

Dragana Bajović[1,2], João Xavier[2], José M. F. Moura[1] and Bruno Sinopoli[1]

***Abstract*— We study the asymptotic exponential decay rate I for the convergence in probability of products $W_k W_{k-1}...W_1$ of random symmetric, stochastic matrices W_k. Albeit it is known that the probability P that the product $W_k W_{k-1}...W_1$ is ϵ away from its limit converges exponentially fast to zero, i.e., $P \sim e^{-kI}$, the asymptotic rate I has not been computed before. In this paper, assuming the positive entries of W_k are bounded away from zero, we explicitly characterize the rate I and show that it is a function of the underlying graphs that support the positive (non zero) entries of W_k. In particular, the rate I is given by a certain generalization of the min-cut problem. Although this min-cut problem is in general combinatorial, we show how to exactly compute I in polynomial time for the commonly used matrix models, gossip and link failure. Further, for a class of models for which I is difficult to compute, we give easily computable bounds: $\underline{I} \leq I \leq \overline{I}$, where $\underline{I}$ and $\overline{I}$ differ by a constant ratio. Finally, we show the relevance of I as a system design metric with the example of optimal power allocation in consensus+innovations distributed detection.**

I. Introduction

This paper calculates the exact exponential rate of convergence in probability for the products of random stochastic matrices. The convergence of products of stochastic matrices has recently received increased interest, e.g., [1], [2], [3], [4], in the study of consensus [5], [6], [4], and consensus+innovations algorithms [7], [8], [9].

Specifically, we consider the product $W_k W_{k-1}...W_1$ of independent identically distributed (i.i.d.) symmetric stochastic matrices. We assume that the positive entries of each realization W of W_k are bounded away from zero. A classical result is that, when the second largest in modulus eigenvalue of $\mathbb{E}[W_k]$ is less than one and the diagonal entries of W_k are almost surely positive, the product $W_k W_{k-1}...W_1$ converges in probability[1] to consensus, i.e., to the matrix $J := \frac{1}{N} 1 1^\top$, where 1 denotes the vector with unit entries. It is also well established that this convergence is exponentially fast [5], [10]. However, the *exact* decay rate has not yet been computed. In this paper, we calculate the exact decay rate I for the convergence in probability:

$$\lim_{k \to \infty} \frac{1}{k} \log \mathbb{P}\left(\|W_k...W_1 - J\| \geq \epsilon\right) = -I, \tag{1}$$

where $\|\cdot\|$ is the spectral norm and $\epsilon > 0$. Our results reveal that the rate I depends only on the graph realizations that support the non-zero entries of W_k, and that it does not depend on ϵ. The rate I in (1) is an important metric that plays a key role, e.g., with the consensus+innovations distributed detection [8], where the detection performance dramatically depends on the rate I; in particular, it exhibits a phase transition. When I is above a threshold, then distributed detection is asymptotically optimal and asymptotically equivalent to the centralized detection; below the threshold, distributed detection achieves only a fraction of the centralized detection performance.

Beyond the computation of I, we can draw the intuition that the rate I is determined by the most likely way in which the underlying graph stays disconnected for a long period of time. Although determining this most likely way and the rate I is, in general, a combinatorial problem, we show that, with several commonly used models of W_k, the rate I is easily computable. This is true, for example, for the link failure and gossip models, for which we explicitly calculate the rate I by solving a min-cut problem.

Calculation of the rate I is a useful tool in the design of distributed algorithms. We demonstrate this with the link failure model, where the link failure (fading) probability depends on the transmission power that nodes (sensors) use. We show that, under Rayleigh fading, the rate I is a convex function of the sensors' transmission power, which allows to optimally allocate the sensors' transmission power in consensus+innovations distributed detection.

The purpose of this paper is to provide new results on the rate I, and to extend the initial results established in [11]. These novel results are summarized as follows. First, we show that, for a generic model of W_k's, calculation of the rate I is equivalent to solving a generalized min-cut problem. Albeit solving the latter is computationally hard in general, we approximate the rate I efficiently for a class of gossip-like models that subsumes, e.g., standard pairwise gossip and symmetrized broadcast gossip. For this class, we provide easily computable tight approximations of I. While in [11] we considered spatially independent fading (failing) links, here we explicitly calculate the rate I for the

*The work of Dragana Bajović, João Xavier and Bruno Sinopoli is partially supported by grants CMU-PT/SIA/0026/2009 and SFRH/BD/33517/2008 (through the Carnegie Mellon/Portugal Program managed by ICTI) from Fundação para a Ciência e Tecnologia and also by ISR/IST plurianual funding (POSC program, FEDER). The work of José M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291. Dragana Bajović holds a fellowship from the Carnegie Mellon/Portugal Program.

[1]Dragana Bajović, José M. F. Moura and Bruno Sinopoli are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA dbajovic@andrew.cmu.edu, moura@ece.cmu.edu, brunos@ece.cmu.edu; ph: (412)268-6341; fax: (412)268-3890

[2] Dragana Bajović and João Xavier are with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal dragana@isr.ist.utl.pt, jxavier@isr.ist.utl.pt

[1]The convergence is also established in the almost sure sense, e.g., [4].

correlated fading. Namely, we show that, with this model, there is a single critical link that determines the rate; this link marks the transition between the connected and disconnected regime of network operation. Finally, we give a closed form solution for I for arbitrary type of averaging that runs on a tree.

Paper organization. Section II introduces the model for random matrices W_k and defines relevant quantities needed in the sequel. Section III states the main result on the existence of the limit (1) and computes the rate I for $\epsilon = 1$. Section IV formulates a generalized min-cut problem and shows that its solution gives the rate I. In Section V, we detail gossip and link failure averaging models, and in Section VI we address the optimal power allocation for distributed detection by maximizing the rate I. Section VII summarizes the paper.

II. Problem setup

In this section, we state the assumptions on the distributed averaging process and define relevant concepts that pertain to the graphs that underly the matrices W_k.

Let $\{W_t : t = 1, 2, ...\}$ be a discrete time (random) process, where W_t are independent and identically distributed (i.i.d.) stochastic $N \times N$ matrices. We assume that each realization W of W_t is symmetric, with positive diagonal entries and with positive entries bounded away from 0; that is, there exists a scalar δ, such that, for any realization W, if $W_{ij} > 0$, then $W_{ij} \geq \delta$. For a given averaging model, which satisfies the above assumption, we denote by $\mathcal{W}$ the set of all possible realizations of W_k. For example, with gossip on a graph G with N vertices, $\mathcal{W}^{\text{Gossip}}$ is the set of all $N \times N$ matrices with sparsity structure of the Laplacian matrix of a one link subgraph of G, with positive entries being arbitrary, but bounded away from zero.

Graph process. For a doubly stochastic symmetric matrix W, let $G(W)$ denote its induced undirected graph, i.e., $G(W) = (V, E(W))$, where $V = \{1, 2, \ldots, N\}$ is the set of all nodes and

$$E(W) = \left\{ \{i, j\} \in \binom{V}{2} : W_{ij} > 0 \right\}.$$

We define the random graph process $\{G_t : t = 1, 2, \ldots\}$ through the random matrix process $\{W_t : t = 1, 2, ...\}$ by: $G_t = G(W_t)$, for $t = 1, 2, \ldots$. As the matrix process is i.i.d., the graph process is i.i.d. as well. We collect the underlying graphs of all possible matrix realizations W in the set $\mathcal{G}$:

$$\mathcal{G} := \{G(W) : W \in \mathcal{W}\}. \tag{2}$$

Hence, each realization of G_t belongs to $\mathcal{G}$, and we, thus, refer to $\mathcal{G}$ as the set of realizable graphs (for the given averaging process). Also, for $H \in \mathcal{G}$, we denote by p_H its probability of occurrence, i.e., $p_H = \mathbb{P}(G_t = H)$. Note that $\sum_{H \in \mathcal{G}} p_H = 1$. Also, for a collection $\mathcal{H} \subseteq \mathcal{G}$ we use $p_{\mathcal{H}}$ to denote the probability that a graph realization G_t belongs to $\mathcal{H}$:

$$p_{\mathcal{H}} = \sum_{H \in \mathcal{H}} p_H. \tag{3}$$

Supergraph disconnected collections. For a collection of graphs $\mathcal{H} \subseteq \mathcal{G}$, let $\Gamma(\mathcal{H})$ denote the graph that contains all edges from all graphs in $\mathcal{H}$. That is, $\Gamma(\mathcal{H})$ is the graph with the minimal number of edges that is a supergraph of every $H \in \mathcal{H}$:

$$\Gamma(\mathcal{H}) := (V, \bigcup_{H \in \mathcal{H}} E(H)), \tag{4}$$

where $E(H)$ denotes the set of edges of graph H.

A special type of a collection supergraph, which will be important when determining the rate (see Lemma 2 in Section III), is the (random) supergraph of all graph realizations G_t until some fixed time k and we denote it by $\Gamma(k, 0)$:

$$\Gamma(k, 0) := \Gamma(\{G_1, G_2, \ldots, G_k\}).$$

We note that $\Gamma(k, 0)$ will contain all the links that occurred at least once from time $t = 1$ until time $t = k$.

Consider next the sets of realizable graphs $\mathcal{H} \subseteq \mathcal{G}$ whose supergraph $\Gamma(\mathcal{H})$ is disconnected and let $\Pi(\mathcal{G})$ collect all such sets:

$$\Pi(\mathcal{G}) = \{\mathcal{H} \subseteq \mathcal{G} : \Gamma(\mathcal{H}) \text{ is disconnected}\}.$$

We call en element $\mathcal{H}$ of $\Pi(\mathcal{G})$ a disconnected collection.

To relate $\Pi(\mathcal{G})$ with the graph process G_t, $t = 1, 2, \ldots$, we note that, as long as the graph process draws its realizations from some $\mathcal{H} \in \Pi(\mathcal{G})$, the supergraph $\Gamma(t, 0)$ stays disconnected, for $t = 1, 2, \ldots$. This means that there will exist two sets of vertices (nodes) $C_1, C_2 \subseteq V$, that remain isolated over time, i.e., the nodes from C_1 will not "communicate" with nodes from C_2, as long as $\Gamma(t, 0)$ stays disconnected.

III. Rate for convergence in probability of consensus

Denote $\Phi(k, 0) := W_k W_{k-1} \cdots W_1$, and $\widetilde{\Phi}(k, 0) := \Phi(k, 0) - J$, for $k \geq 1$. The norm of the matrix $\widetilde{\Phi}(k, 0)$, therefore, says how far the averaging process is from consensus. Theorem 1 gives the exponential rate of decay for the convergence in probability of $\widetilde{\Phi}(k, 0)$ to 0.

Theorem 1 Consider an i.i.d. random process $\{W_t : t = 1, 2, \ldots\}$ that takes realizations in the set of symmetric stochastic matrices. Let there exist $\delta > 0$, such that each realization W of W_t satisfies $W_{ii} > \delta$ and $W_{ij} > \delta$, if $W_{ij} > 0$. Then:

$$\lim_{k \to \infty} \frac{1}{k} \log \mathbb{P}\left(\left\| \widetilde{\Phi}(k, 0) \right\| \geq \epsilon \right) = -I, \ \forall \epsilon \in (0, 1] \tag{5}$$

where

$$I = \begin{cases} +\infty & \text{if } \Pi(\mathcal{G}) = \emptyset \\ |\log p_{\max}| & \text{otherwise} \end{cases}, \tag{6}$$

and $p_{\max} = \max_{\mathcal{H} \in \Pi(\mathcal{G})} p_{\mathcal{H}}$ is the probability of the most likely disconnected collection.

The proof of Theorem 1 can be found in [11]. However, to justify the claim of Theorem 1 on an intuitive level, we give here the key result behind Theorem 1, Lemma 2, which discovers the relation between the decay of the norm of the error matrix $\widetilde{\Phi}(k,0)$ and the Fiedler value of the associated supergraph $\Gamma(k,0)$. Also, based on Lemma 2, we compute here the rate I for $\epsilon = 1$. (As we show in [11], the rate is the same for all $\epsilon \in (0,1]$.) The proof of Lemma 2 can be found in [11].

Lemma 2 For any realization $W_1, \ldots, W_k$, $k \geq 1$:

$$\|\widetilde{\Phi}(k,0)\| \leq \left(1 - \delta^{2k}\lambda_{\mathrm{F}}\left(\Gamma(k,0)\right)\right)^{\frac{1}{2}},$$

where $\lambda_{\mathrm{F}}(G)$ denotes the Fiedler value of graph G, i.e., the second smallest eigenvalue of the Laplacian matrix of G.

Lemma 2 asserts that, when the supergraph of the topologies underlying averaging matrices becomes connected, the spectral norm of the "error" matrix $\widetilde{\Phi}(k,0)$ drops below 1. This drop can be uniformly bounded (over all possible connected graphs $\Gamma(k,0)$ that are obtainable by the averaging model) by $\left(1 - c\delta^{2k}\right)^{\frac{1}{2}}$, where $c = 2(1-\cos\frac{\pi}{N})$ is the Fiedler value of the path on N vertices, i.e., the smallest Fiedler value of all connected graphs on N vertices [12]. Thus, if $\Gamma(k,0)$ is connected, then $\left\|\widetilde{\Phi}(k,0)\right\| \leq \left(1 - c\delta^{2k}\right)^{\frac{1}{2}}$, and hence is smaller than 1. Suppose, on the other hand, that $\Gamma(k,0)$ is disconnected. Then, $\Gamma(k,0)$ must have at least two components, say C_1 and C_2, $C_1, C_2 \subset V$, such that there are no edges in $\Gamma(k,0)$ connecting C_1 and C_2. Further, each graph G_t, that occurred from time $t=1$ through time $t=k$, will also lack edges between the sets of vertices C_1 and C_2. Then, each of the matrices W_t can be written in a block diagonal form (up to multiplication by a permutation matrix), with one block corresponding to each of the components C_1 and C_2. This implies that the product matrix $\Phi(k,0) = W_k \cdots W_1$ also has a block diagonal form, and, thus, has two eigenvalues equal to 1. Therefore, the spectral norm of $\widetilde{\Phi}(k,0) = \Phi(k,0) - J$ is equal to 1. Summarizing, the norm of $\widetilde{\Phi}(k,0)$ is less than 1 if and only if the graph $\Gamma(k,0)$ is connected, implying that

$$\left\|\widetilde{\Phi}(k,0)\right\| = 1 \;\Leftrightarrow\; G_t \in \mathcal{H},\, 1 \leq t \leq k \text{ for some } \mathcal{H} \in \Pi(\mathcal{G}).$$

In other words, as long as the graph process chooses its realizations from some disconnected collection, the norm of $\widetilde{\Phi}(k,0)$ will remain equal to 1. Using this observation, we can compute the probability in (5) for $\epsilon = 1$ by [2]

$$\mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right) = \mathbb{P}\left(\bigcup_{\mathcal{H}\in\Pi(\mathcal{G})}\left\{G_t \in \mathcal{H},\, 1 \leq t \leq k\right\}\right). \tag{7}$$

Bounding the probability of the union by the probability of a fixed event from the union, we get

$$\mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right) \geq \mathbb{P}\left(G_t \in \mathcal{H},\, t = 1,\ldots,k\right) = p_{\mathcal{H}}^k,$$

[2]Because the matrices W_t are stochastic with probability 1, the probability of the event $\left\{\left\|\widetilde{\Phi}(k,0)\right\| > 1\right\}$ is equal to zero, implying that $\mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| \geq 1\right) = \mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right)$.

where $\mathcal{H} \in \Pi(\mathcal{G})$ is some fixed disconnected collection and the last equality follows by the independence assumption on the graph realizations. To obtain the best bound, we choose the most likely disconnected collection $\mathcal{H}$ which gives

$$\mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right) \geq p_{\max}^k. \tag{8}$$

Next, applying the union bound in (7) yields

$$\begin{aligned}\mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right) &\leq \sum_{\mathcal{H}\in\Pi(\mathcal{G})} \mathbb{P}\left(G_t \in \mathcal{H},\, t = 1,\ldots,k\right)\\ &\leq |\Pi(\mathcal{G})| p_{\max}^k,\end{aligned} \tag{9}$$

where $|\Pi(\mathcal{G})|$ is the number of disconnected collections on $\mathcal{G}$. Combining (8) and (9), we get:

$$p_{\max}^k \leq \mathbb{P}\left(\left\|\widetilde{\Phi}(k,0)\right\| = 1\right) \leq |\Pi(\mathcal{G})| p_{\max}^k. \tag{10}$$

Taking the $\log$, dividing by k and taking the $\lim_{k\to\infty}$ on both sides of the previous inequality yields by the "sandwiching" argument that, for $\epsilon = 1$, the limit in (5) exists and equals $-|\log p_{\max}|$.

IV. Computation of $p_{\max}$ via generalized min-cut

This section introduces a generalization of the minimum cut (min-cut) problem and shows that computing $p_{\max}$ is equivalent to solving an instance of the generalized min-cut. For certain types of averaging, in which the number of graphs that "cover" an edge is relatively small, we show in Subsection IV-A that the generalized min-cut can be well approximated with the standard min-cut, and thus can be efficiently solved. We illustrate this with the broadcast gossip example in Subsection V-A, where we find a 2-approximation for $p_{\max}$ by solving two instances of the standard min-cut.

Generalization of the min-cut. Let $G = (V, E)$ be a given undirected graph, with the set of nodes V and the set of edges E. The generalization of the min-cut problem that is of interest to us assigns a cost to each set of edges $F \subseteq E$. This is different than the standard min-cut, as with the standard min-cut the costs are assigned to each edge of E and, thus, where the cost of F is simply the sum of the individual costs of edges in F. Similarly as with the standard min-cut, the goal is to find F that disconnects G with minimal cost. More formally, let the function $\mathcal{C}: 2^E \mapsto \mathbb{R}_+$ assign costs to subsets of E, i.e., the cost of F is $\mathcal{C}(F)$, for $F \subseteq E$. Then, the generalized min-cut problem is

$$\begin{array}{ll}\text{minimize} & \mathcal{C}(F)\\ \text{subject to} & F \subseteq E : (V, E \setminus F) \text{ is disconnected}\end{array}. \tag{11}$$

We denote by $\mathtt{gmc}(G, \mathcal{C})$ the optimal value of (11). We remark that, when the cost $\mathcal{C}(F)$ is decomposable over the edges of F, i.e., when for all $F \subseteq E$, $\mathcal{C}(F) = \sum_{e\in F} c(e)$, for some function $c: E \mapsto \mathbb{R}_+$, then the generalized min-cut simplifies to the standard min-cut. For this case, we denote the optimal value of (11) by $\mathtt{mc}(G, c)$.

Consider now a general averaging model on the set of nodes V and with the collection of realizable graphs $\mathcal{G}$. Let

$G = \Gamma(\mathcal{G})$, where $G = (V, E)$ and E collects all the edges that appear with positive probability. The following lemma shows that the rate I for the general averaging model can be computed by solving an instance of the generalized min-cut problem.

Lemma 3 Let the cost function $\mathcal{C} : 2^E \mapsto \mathbb{R}_+$ be defined by $\mathcal{C}(F) = \mathbb{P}(\cup_{e \in F} \{e \in E(G_t)\})$, for $F \subset E$. Then,

$$I = -\log(1 - \mathrm{gmc}(G, \mathcal{C})) \tag{12}$$

Proof: For each $F \subseteq E$ such that $(V, E \setminus F)$ is disconnected, define S_F by: $S_F = \{\mathcal{H} \in \Pi(\mathcal{G}) : E(\Gamma(\mathcal{H})) \subseteq E \setminus F\}$. Note that $S_F \subseteq \Pi(\mathcal{G})$, for each F. We show that sets S_F cover $\Pi(\mathcal{G})$, i.e., that $\cup_{F \subseteq E : (V, E\setminus F) \text{ is disc.}} S_F = \Pi(\mathcal{G})$. To this end, pick an arbitrary $\mathcal{H} \in \Pi(\mathcal{G})$ and let $F^\star := E \setminus E(\Gamma(\mathcal{H}))$. Then, because supergraph $\Gamma(\mathcal{H})$ is disconnected, $F^\star$ must be a set of edges that disconnects G; if we now take the set $S_{F^\star}$ that is associated with $F^\star$, we have that $\mathcal{H}$ belongs to $S_{F^\star}$ proving the claim above. Since we established that $\cup_{F \subseteq E : (V, E\setminus F) \text{ is disc.}} S_F = \Pi(\mathcal{G})$, in order to find $p_{\max}$, we can branch the search over the S_F sets:

$$p_{\max} = \max_{\mathcal{H} \in \Pi(\mathcal{G})} p_{\mathcal{H}} = \max_{F \subseteq E : (V, E\setminus F) \text{ is disc.}} \max_{\mathcal{H} \in S_F} p_{\mathcal{H}} \tag{13}$$

(where, for every empty S_F, we define its corresponding value $\max_{\mathcal{H} \in S_F} p_{\mathcal{H}}$ to be 0). Next, pick a fixed set F for which S_F is nonempty and define $\mathcal{H}_F$ by:

$$\mathcal{H}_F = \{H \in \mathcal{G} : E(H) \subseteq E \setminus F\}; \tag{14}$$

that is, $\mathcal{H}_F$ collects all the realizable graphs whose edges do not intersect with F. Note that, by construction of $\mathcal{H}_F$, $E(\Gamma(\mathcal{H}_F)) \subseteq E \setminus F$, proving that $\mathcal{H}_F \in S_F$. Now, for an arbitrary fixed collection $\mathcal{H} \in S_F$, since any graph H that belongs to $\mathcal{H}$ must satisfy the property in (14), we have that $\mathcal{H} \subseteq \mathcal{H}_F$ and, consequently, $p_{\mathcal{H}} \leq p_{\mathcal{H}_F}$. This proves that, for every fixed non-empty S_F the maximum $\max_{\mathcal{H} \in S_F} p_{\mathcal{H}}$ is attained at $\mathcal{H}_F$ and equals $p_{\mathcal{H}_F} = \mathbb{P}(E(G_t) \subseteq E \setminus F)$. Combining the last remark with (13), yields:

$$p_{\max} = \max_{F \subseteq E : (V, E\setminus F) \text{ is disc.}} \mathbb{P}(E(G_t) \subseteq E \setminus F). \tag{15}$$

Finally, noting that $\mathbb{P}(E(G_t) \subseteq E \setminus F) = 1 - \mathbb{P}(\cup_{e \in F} \{e \in E(G_t)\})$ completes the proof of Lemma 3. ∎

Rate I for algorithms running on a tree. When the graph that collects all the links that appear with positive probability is a tree, we obtain a particularly simple solution for I using formula (12). To this end, let $T = (V, E)$ be the supergraph of all the realizable graphs and suppose that T is a tree. Then, removal of any edge from E disconnects T. This implies that, to find the rate, we can shrink the search space of the generalized min-cut problem in (12) (see also eq. (11)) to the set of edges of the tree:

$$\min_{F \subseteq E : (V, E\setminus F) \text{ is disc.}} \mathcal{C}(F) = \min_{e \in E} \mathcal{C}(e).$$

Now, $\mathcal{C}(e) = \mathbb{P}(e \in E(G_t))$ can be computed by summing up the probabilities of all graphs that cover e, i.e., $\mathcal{C}(e) = \sum_{H \in \mathcal{G} : e \in E(H)} p_H$. The minimum of $\mathcal{C}(e)$ is then achieved at the link that has the smallest probability of occurrence $p_{\mathrm{rare}} = \min_{e \in E} \sum_{H \in \mathcal{G} : e \in E(H)} p_H$. Thus, the rate I is determined by the probability of the "weakest" link in the tree, i.e., the link that is most rarely online and

$$I^{\mathrm{Tree}} = -\log(1 - p_{\mathrm{rare}}). \tag{16}$$

A. Approximation of I by min-cut based bounds

We now explain how we can compute the rate I by approximately solving the instance of the generalized min-cut in (12) via two instances of the standard min-cut. Our strategy to do this is to "sandwich" each cost $\mathcal{C}(F)$, $F \subseteq E$, by two functions, which are decomposable over the edges of F. To this end, fix F and observe that

$$\mathcal{C}(F) = \mathbb{P}\left(\bigcup_{e \in F} \{e \in E(G_t)\}\right) = \sum_{H \in \mathcal{G} : e \in E(H),\ \text{for } e \in F} p_H. \tag{17}$$

By the union bound, $\mathcal{C}(F)$ is upper bounded as follows:

$$\mathcal{C}(F) \leq \sum_{e \in F} \mathbb{P}(e \in E(G_t)) = \sum_{e \in F} \sum_{H \in \mathcal{G} : e \in E(H)} p_H. \tag{18}$$

We next assume that for every set of m edges we can find m distinct graphs, say $H_1, \ldots, H_m \in \mathcal{G}$, such that H_i covers e_i, $i = 1, \ldots, m$[3]. Then, for each $e \in F$, we can pick a different graph in the sum in (17), say H_e, such that $e \in E(H_e)$, until all the edges in F have its associated graph H_e. The sum of the probabilities of the chosen graphs $\sum_{e \in F} p_{H_e}$ is then smaller than $\mathcal{C}(F)$. Finally, we can bound each p_{H_e} by the probability of the least likely graph that covers e, thus yielding:

$$\sum_{e \in F} \min_{H \in \mathcal{G} : e \in E(H)} p_H \leq \mathcal{C}(F).$$

Motivated by the previous observations, we introduce $\overline{c}, \underline{c} : E \mapsto \mathbb{R}_+$ defined by

$$\overline{c}(e) = \sum_{H \in \mathcal{G} : e \in E(H)} p_H, \quad \underline{c}(e) = \min_{H \in \mathcal{G} : e \in E(H)} p_H. \tag{19}$$

Then, for each $F \subseteq E$, we have:

$$\sum_{e \in F} \underline{c}(e) \leq \mathcal{C}(F) \leq \sum_{e \in F} \overline{c}(e).$$

Because the inequality above holds for all $F \subseteq E$, we have that:

$$\mathrm{mc}(G, \underline{c}) \leq \mathrm{gmc}(G, \mathcal{C}) \leq \mathrm{mc}(G, \overline{c}). \tag{20}$$

Therefore, we can efficiently approximate the rate I by solving two instances of the standard min-cut problem, with the respective costs $\underline{c}$ and $\overline{c}$. To further simplify the computation of I, we introduce D– the maximal number of graphs that "covers" an edge e, where the maximum is over all edges $e \in E$. We also introduce $\overline{p}$ and $\underline{p}$ to denote the probabilities of the most likely and least likely graph,

[3]The case when this is not true can be handled by splitting the probability p_H of a graph H into d equal parts, where d is the number of edges covered by H. The approximation bounds (that are derived further ahead) would then depend on d; we omit the details here due to lack of space

respectively, i.e., $\overline{p} = \max_{H \in \mathcal{G}} p_H$ and $\underline{p} = \min_{H \in \mathcal{G}} p_H$. Then, the function $\overline{c}$ can be uniformly bounded by $D\overline{p}$ and, similarly, function $\underline{c}$ can be uniformly bounded by $\underline{p}$, which combined with (20) yields[4]:

$$\underline{p}\, \mathrm{mc}(G,1) \leq \mathrm{gmc}(G,\mathcal{C}) \leq D\, \overline{p}\, \mathrm{mc}(G,1); \qquad (21)$$

The expression in (21) gives a $D\,\overline{p}/\underline{p}$-approximation for $\mathrm{gmc}(G,\mathcal{C})$, and it requires solving only one instance of the standard min-cut, with uniform (equal to 1) costs.

V. Examples: Randomized Gossip and Fading Model

This section computes the rate I for the commonly used averaging models: randomized gossip and link failure. Subsection V-A studies two types of the randomized gossip algorithm, namely pairwise gossip and symmetrized broadcast gossip and it shows that, for the pairwise gossip on a generic graph $G = (V, E)$, the corresponding rate can be computed by solving an instance of the standard min-cut; for broadcast gossip, we exploit the bounds derived in Subsection IV-A to arrive at a tight approximation for its corresponding rate. Subsection V-B studies the network with fading links for the cases when 1) all the links at a time experience the same fading (correlated fading), and 2) the fading is independent across different links (uncorrelated fading). Similarly as with the pairwise gossip, the rate for the uncorrelated fading can be computed by solving an instance of a min-cut problem. With the correlated fading, there exists a threshold on the fading coefficients, which induces two regimes of the network operation, such that if at a time t the fading coefficient is above the threshold, the network realization at time t is connected. We show that the rate is determined by the probability of the "critical" link that marks the transition between these two regimes.

A. Pairwise and broadcast gossip

Min-cut solution for pairwise gossip. Let $G = (V, E)$ be an arbitrary connected graph on N vertices. With pairwise gossip on graph G, at each averaging time, only one link from E can be active. Therefore, the set of realizable graphs $\mathcal{G}^{\mathrm{Gossip}}$ is the set of all one link graphs on G:

$$\mathcal{G}^{\mathrm{Gossip}} = \{(V,e) : e \in E\}\,.$$

Now, consider the probability $\mathbb{P}(\cup_{e\in F}\{e \in E(G_t)\})$, for a fixed subset of edges $F \subseteq E$. Because each realization of G_t can contain only one link, the events under the union are disjoint. Thus, the probability of the union equals the sum of the probabilities of individual events, yielding that the cost $\mathcal{C}(F)$ is decomposable for gossip, i.e.,

$$\mathcal{C}(F) = \sum_{e\in F} p_{(V,e)}$$

[4]We are using here the property of the min-cut with uniform positive costs by which $\mathrm{mc}(G, \alpha 1) = \alpha \mathrm{mc}(G, 1)$, for $\alpha \geq 0$ [13], where 1 denotes the cost function that has value 1 at each edge

Therefore, the rate for gossip is given by

$$I^{\mathrm{Gossip}} = -\log\left(1 - \mathrm{mc}(G, c^{\mathrm{Gossip}})\right), \qquad (22)$$

where $c^{\mathrm{Gossip}}(e) = p_{(V,e)}$. We remark here that, for pairwise gossip, functions $\underline{c}, \overline{c}$ in (19) are identical (each link e has exactly one graph (V, e) that covers it, hence $\underline{c}(e) = \overline{c}(e) = p_{(V,e)}$), which proves that bounds in (20) are touched for this problem instance. For the case when all links have the same activation probability equal to $1/|E|$, the edge costs $c^{\mathrm{Gossip}}(e)$ are uniform and equal to $1/|E|$, for all $e \in E$ and (22) yields the following simple formula for the rate for uniform gossip:

$$I^{\mathrm{Gossip}} = -\log\left(1 - 1/|E|\,\mathrm{mc}(G,1)\right). \qquad (23)$$

2-approximation for broadcast gossip. With bidirectional broadcast gossip on an arbitrary connected graph $G = (V, E)$, at each time a node $v \in V$ is chosen at random and the averaging is performed across the neighborhood of v. Thus, at each time t, the set of active edges is the set of all edges adjacent to the vertex that is chosen at time t; hence, the set of realizable graphs $\mathcal{G}^{\text{B-Gossip}}$ is

$$\mathcal{G}^{\text{B-Gossip}} = \{(V, \{\{u,v\} : \{u,v\} \in E\} : v \in V\}\,.$$

We can see that each edge $e = \{u, v\}$ can become active in two ways, when either node u or node v is active. In other words, each edge is covered by exactly two graphs. This gives $D = 2$ and using (20) we get the following approximation:

$$\underline{p}\,\mathrm{mc}(G,1) \leq \mathrm{gmc}(G,\mathcal{C}) \leq 2\overline{p}\,\mathrm{mc}(G,1),$$

where $\underline{p}$ and $\overline{p}$ are the probabilities of the least, resp., most, active node. For the case when all the nodes have the same activation probability equal to $1/N$, using (21) we get a 2-approximation:

$$\frac{1}{N}\mathrm{mc}(G,1) \leq \mathrm{mc}(G,\mathcal{C}) \leq \frac{2}{N}\mathrm{mc}(G,1).$$

Thus, the rate I for the broadcast gossip with uniform node activation probability satisfies:

$$I^{\text{B-Gossip}} \in [-\log(1-\frac{1}{N}\mathrm{mc}(G,1)), -\log(1-\frac{2}{N}\mathrm{mc}(G,1))]. \qquad (24)$$

We now compare the rates for the uniform pairwise and uniform broadcast gossip when both algorithms are running on the same (connected) graph $G = (V, E)$. Consider first the case when G is a tree. Then, $E = N - 1$ and, since all the links have the same occurrence probability $1/(N - 1)$, the formula for gossip gives $I^{\mathrm{Gossip}} = -\log(1-1/(N-1))$. To obtain the exact rate for the broadcast gossip, we recall formula (16). As each link in the tree is covered by exactly two graphs, and the probability of a graph is $1/N$, we have that $p^{\mathrm{rare}} = 2/N$. Therefore, the rate for broadcast gossip on a tree is $I^{\mathrm{Gossip}} = -\log(1 - 2/N)$, which is higher than $I^{\mathrm{Gossip}} = -\log(1 - 1/(N - 1))$. Consider now the case when G is not a tree. Then, the number of edges $|E|$ in G is at least N and we have $I^{\mathrm{Gossip}} = -\log(1 - 1/|E|\mathrm{mc}(G, 1)) \leq -\log(1 - 1/N\mathrm{mc}(G, 1))$. On the other hand, by (24), $I^{\text{B-Gossip}} \geq -\log(1-1/N\mathrm{mc}(G,1))$. Combining the last two observations yields that the rate of broadcast gossip is always higher than the rate of pairwise gossip running on the same graph. This is in accordance with the intuition, as with broadcast gossip more links are active at a time, and, thus, we would expect that it performs the averaging faster.

B. Link failure: fading channels

Consider a network of N sensors described by graph $G = (V, E)$, where the set of edges E collects all the links $\{i, j\}$ that appear with positive probability, $i, j \in V$. To model the link failures, we adopt a symmetric fading channel model, a model similar to the one proposed in [14] (reference [14] assumes asymmetric channels). At time k, sensor j receives from sensor i $y_{ij,k} = g_{ij,k}\sqrt{S_{ij}/d_{ij}^{\alpha}}x_{i,k} + n_{ij,k}$, where S_{ij} is the transmission power that sensor i uses for transmission to sensor j, $g_{ij,k}$ is the channel fading coefficient, $n_{ij,k}$ is the zero mean additive Gaussian noise with variance σ_n^2, d_{ij} is the inter-sensor distance, and α is the path loss coefficient. We assume that $g_{ij,k}$, $k \geq 1$, are i.i.d. in time and that $g_{ij,t}$ and $g_{lm,s}$ are mutually independent for all $t \neq s$; also, the channels (i, j) and (j, i) at time k experience the same fade, i.e., $g_{ij,k} = g_{ji,k}$. We adopt the following link failure model. Sensor j successfully decodes the message from sensor i (link (i, j) is online) if the signal to noise ratio exceeds a threshold, i.e., if: $\mathrm{SNR} = \frac{S_{ij} g_{ij,k}^2}{\sigma_n^2 d_{ij}^{\alpha}} > \tau$, or, equivalently, if $g_{ij,k}^2 > \frac{\tau \sigma_n^2 d_{ij}^{\alpha}}{S_{ij}} := \gamma_{ij}$. Since link occurrences are "controlled" by the realizations of the fading coefficients, the set of realizable graphs in the link failure model depends on the joint distribution of $\{g_{ij,k}\}_{\{i,j\}\in E}$. In the sequel, we study the cases when the fading coefficients at some time k are either fully correlated or uncorrelated, and we compute the rate I for each of these cases.

Uncorrelated fading. With uncorrelated fading, $g_{ij,k}$ are independent across different links for all k. Therefore, in this model, the indicators of link occurrences are independent Bernoulli random variables, such that the indicator of link $\{i, j\}$ being online is 1 if the fading coefficient at link $\{i, j\}$ exceeds the communication threshold of $\{i, j\}$, i.e., if $g_{ij,k}^2 > \gamma_{ij}$, and is zero otherwise. Due to the independence, each subgraph $H = (V, E')$ of G, $E' \subseteq E$, is a realizable graph in this model, hence,

$$\mathcal{G}^{\text{Fail-uncorr}} = \{H = (V, E') : E' \subseteq E\}.$$

Denote with $P_{ij} = \mathbb{P}(g_{ij,k}^2 > \gamma_{ij})$ the probability that link $\{i, j\}$ is online. We compute the rate $I^{\text{Fail-uncorr}}$ for the uncorrelated link failure using the result of Lemma 3. To this end, let F be a fixed subset of E and consider the probability that defines $\mathcal{C}(F)$. Then,

$$\begin{aligned}\mathcal{C}(F) &= \mathbb{P}\left(\cup_{\{i,j\}\in F}\{\{i,j\} \in E(G_t)\}\right)\\ &= 1 - \mathbb{P}\left(\cap_{\{i,j\}\in F}\{\{i,j\} \notin E(G_t)\}\right)\\ &= 1 - \prod_{\{i,j\}\in F}(1 - P_{ij}),\end{aligned}$$

where the last equality follows by the independence of the link failures. To compute the rate, we follow formula (12):

$$\begin{aligned}1 - &\min_{F\subseteq E:(V,E\setminus F)\text{ is disc.}} \mathcal{C}(F) \qquad (25)\\ &= \max_{F\subseteq E:(V,E\setminus F)\text{ is disc.}} \prod_{\{i,j\}\in F}(1 - P_{ij})\\ &= \exp\Big(-\min_{F\subseteq E:(V,E\setminus F)\text{ is disc.}} \sum_{\{i,j\}\in F} -\log(1 - P_{ij})\Big). \qquad (26)\end{aligned}$$

Now, the optimization problem in the exponent is an instance of the standard min-cut problem with edge costs $c^{\text{Fail-uncorr}}(\{i, j\}) = -\log(1-P_{ij})$. By formula (12), the rate is obtained from the expression in line (25) by taking the $-\log$, which finally yields:

$$I^{\text{Fail-uncorr}} = \mathrm{mc}(G, c^{\text{Fail-uncorr}}). \qquad (27)$$

Correlated fading. With the correlated fading, at any time k each link experiences the same fading, i.e., $g_{ij,k} = g_k$ for all $\{i, j\} \in E$ and so the realization of the common fading g_k sets all the link occurrences at time k. For instance, if $g_k^2 = \bar{g}^2$, then all the links $\{i, j\}$ with $\gamma_{ij} < \bar{g}^2$ are online, and the rest of the links are offline. Therefore, the graph realization corresponding to the fading realization $\bar{g}^2$ is (V, E'), where $E' = \left\{\{i, j\} \in E : \gamma_{ij} < \bar{g}^2\right\}$. We can see that the higher the $\bar{g}^2$ is, the more links are online. Also, if we slide $\bar{g}^2$ from zero to $+\infty$, the corresponding graph realization gradually increases in size by adding one more link whenever $\bar{g}^2$ crosses some threshold γ_{ij} – starting from the empty graph ($\bar{g}^2 = 0$), until the full graph (V, E) is achieved, which occurs when $\bar{g}^2$ crosses the highest threshold. Therefore, if we order the links in the increasing order with respect to their thresholds γ_{ij}, such that $e_1 \in E$ has the lowest threshold, $\gamma_{e_1} = \min_{\{i,j\}\in E}\gamma_{ij}$, and $e_{|E|} \in E$ has the highest threshold, $\gamma_{e_{|E|}} = \max_{\{i,j\}\in E}\gamma_{ij}$, then the set of all realizable graphs with the correlated fading model is

$$\mathcal{G}^{\text{Fail-corr}} = \{H_l = (V, \{e_1, e_2, \ldots, e_l\}) : 0 \leq l \leq |E|\},$$

where the graph realization corresponding to $l = 0$ is the empty graph $(V, \emptyset)$. For fixed l, $0 \leq l \leq |E|$, let $p_l = \mathbb{P}(g_k^2 > \gamma_l)$ denote the probability that link e_l is online. Then, the probability p_{H_l} of graph realization H_l is $p_{H_l} = p_l - p_{l+1} = \mathbb{P}(\gamma_l < g_k^2 \leq \gamma_{l+1})$. Let l^{c} be the index corresponding to the link that marks the connectedness transition of graphs H_l, such that H_l is disconnected for all $l < l^{\mathrm{c}}$, and H_l is disconnected, for all $l \geq l^{\mathrm{c}}$. Then, any disconnected collection on $\mathcal{G}^{\text{Fail-corr}}$ is of the form $\{H_1, H_2, \ldots, H_l\}$, where $l < l^{\mathrm{c}}$. The most likely one is $\{H_1, H_2, \ldots, H_{l^{\mathrm{c}}-1}\}$, and its probability is $p_{H_1} + p_{H_2} + \ldots + p_{H_{l^{\mathrm{c}}-1}} = 1 - p^{\text{crit}}$, where we use p^{crit} to denote the probability of the "critical" link $e_{l^{\mathrm{c}}}$ (i.e., $p^{\text{crit}} = p_{l^{\mathrm{c}}}$). Therefore, $p_{\max} = 1 - p^{\text{crit}}$ and the rate for the correlated fading model is:

$$I^{\text{Fail-corr}} = -\log\left(1 - p^{\text{crit}}\right).$$

VI. Application: Optimal power allocation for distributed detection

We demonstrate the usefulness of the rate I by optimally allocating the sensors' transmission power for consensus+innovations distributed detection.

We briefly explain the distributed detection problem. The goal is that each of N sensors make the decision on the valid hypothesis, H_1 versus H_0, by acquiring measurements $Y_{i,k}$ over time k, and by cooperating with its neighbors. The measurement $Y_{i,k}$ has the density $f_1(\cdot)$ under H_1, and $f_0(\cdot)$ under H_0.

To resolve the detection problem above, each sensor i performs the local decision at time k by comparing its local decision variable $x_i(k)$ against the zero threshold. To update $x_i(k)$, sensor i exchanges its current decision variable with its neighbors, and assimilates the new measurement through the log-likelihood ratio $L_{i,k} = \log \frac{f_1(Y_{i,k})}{f_0(Y_{i,k})}$:

$$x_{i,k} = \sum_{j \in \mathcal{N}_{i,k}} W_{ij,k} \left(\frac{k-1}{k} x_{j,k-1} + \frac{1}{k} L_{j,k} \right), \tag{28}$$

where $x_{i,0} = 0$, $\mathcal{N}_{i,k}$ is the (random) neighborhood of sensor i at time k (including i), and $W_{ij,k}$ is the (random) averaging weight, for $k = 1, 2, ...$

To re-write (28) in matrix form, we let $x_k = (x_{1,k}, x_{2,k}, ..., x_{N,k})^\top$ and $L_k = (L_{1,k}, ..., L_{N,k})^\top$, and collect $W_{ij,k}$'s in the $N \times N$ matrix W_k, such that $W_{ij,k} = 0$ if i and j do not communicate at time k. By unwinding the recursion (28) and using the vector quantities, we obtain:

$$x_k = \frac{1}{k} \sum_{t=1}^{k} \Phi(k, t-1) L_t, \; k = 1, 2, ... \tag{29}$$

Clearly, the dynamics of x_k, and hence the distributed detection performance, depend heavily on the convergence speed of products $\Phi(k, t-1)$, and thus on the rate of consensus I. In fact, reference [9] shows that, when I is above a threshold $I^\star = I^\star(f_1, f_0, N)$, then distributed detector at any sensor i is asymptotically optimal, i.e., it achieves the highest possible exponential rate of decay of the error probability. Below the threshold, distributed detector at any sensor i achieves only a fraction of this best possible rate.

The threshold level $I = I^\star$ has a significance in a practical wireless sensor network scenario, where the inter-sensor communication channels experience fading. The probability of successful communication, and thus the rate I, increases with the increased transmission power. However, at the point $I = I^\star$, distributed detector achieves the optimal detection rate; further increase of the transmission power (and the increase of rate I) does not pay off in terms of the detection performance. Thus, the transmission power that sets I to $I^\star$ leads to the optimal power allocation.

We address the optimal power allocation problem for the symmetric Rayleigh fading channels. Denote by S_{ij} the transmission power that sensor i uses to transmit $x_i(k)$ to sensor j. Then, the probability of successful decoding is modeled as (see [14] for details):

$$P_{ij} = e^{-\frac{K_{ij}}{S_{ij}}}, \tag{30}$$

where the parameter $K_{ij} > 0$ depends on the distance between i and j ($K_{ij} = \tau \sigma_n^2 d_{ij}^\alpha$, see Subsection V-B for details).

Mathematically, the symmetric Rayleigh fading is the link failure model where links fail independently, and the link occurrence probability is given by (30). Hence, the rate I is calculated as in (27), where the weight c_{ij} associated with link (i,j) is:

$$c_{ij}(S_{ij}) = -\log \left(1 - e^{-K_{ij}/S_{ij}} \right).$$

Denote by $\{S_{ij}\}$ the set of all S_{ij}'s, $\{i,j\} \in E$; then, the rate $I(\{S_{ij}\}) = \mathrm{mc}(G, c)$, with $c_{ij} = -\log(1 - e^{-K_{ij}/S_{ij}})$, for $\{i,j\} \in E$, and $c_{ij} = 0$ else. We can now formulate the optimal power allocation problem as minimizing the total power consumption per k, $2 \sum_{\{i,j\} \in E} S_{ij}$, such that the rate I exceeds the optimality threshold $I^\star$.

$$\begin{array}{ll} \text{minimize} & \sum_{\{i,j\} \in E} S_{ij} \\ \text{subject to} & I(\{S_{ij}\}) \geq I^\star \end{array}. \tag{31}$$

This study of optimal power allocation across sensors such that the rate I is at the optimal detection operating point $I^\star$ is also considered in [11], where we show that (31) is a convex problem. Here we address the case of a larger network and where the distributed detection algorithm is defined by constant averaging weights. We solve an instance of (31) on a random geometric graph $G = (V, E)$ with $N = 20$ sensors and $|E| = 86$ edges (we obtain G by placing 20 sensors uniformly over a square and connecting those whose distance is less than a radius). We remark that each link $\{i,j\} \in E$ is a failing links; thus, the number of variables S_{ij} in the optimization problem that we are solving is 86. The coefficients K_{ij} are dependent on the inter-sensor distances, and we choose $K_{ij} = 6.25 d_{ij}^2$, where d_{ij} is the distance between nodes i and j in the obtained geometric graph. We solve the corresponding instance of (31) by applying the subgradient algorithm on the unconstrained exact penalty reformulation of (31) (see, e.g., [15]). Denote with $\{S_{ij}^\star\}$ the obtained optimal solution.

Simulations. This section compares by simulations performance of two detection algorithms – one running on a network with uniform transmission powers and the other running on a network with optimally allocated sensors' transmission powers. The common ground for the comparison is the total power "invested" at each iteration of the algorithm which is equal to $S^{\mathrm{TOT}} := 2 \sum_{\{i,j\} \in E} S_{ij}^\star$. (Note that factor 2 accounts for transmissions of both nodes i and j.) Then, with uniform powers, the transmission power at each sensor is S_0, where $S^{\mathrm{TOT}} = 2 \sum_{\{i,j\} \in E} S_0$. The averaging process is for both algorithms defined by the constant weights $\delta = 1/(d_{\max} + 1)$, where $d_{\max}$ is the maximum degree of G; that is, if at a time k link $\{i,j\} \in E$ is online, then $W_{ij,k} = \delta$, and $W_{ij,k} = 0$ otherwise; also, $W_{ii,k} =$

$1-\sum_{j\in O_{i,k}} W_{ij,k}$. We draw sensors' measurements from the Gaussian distribution with parameters $m = 0.15$, and $\sigma^2 = 1$; the value of $I^\star$ is, for the Gaussian model equal to $(N-1)N\frac{m^2}{8\sigma^2}$ (see [8]), and equals 0.0027 for this simulation setup. As the performance metric of the detection algorithms, we adopt detection error probability of the worst sensor at a time k, $P^{\max}(k) := \max_{i=1,\ldots,N} P_i^e(k)$; $P^{\max}(k)$ is numerically estimated by performing 5000 Monte Carlo runs. Figure 1 plots $P^{\max}(k)$ versus time k for the optimal power allocation (solid red line), and the uniform power allocation (dotted blue line). The figure shows clearly that the optimal power allocation scheme outperforms the uniform power allocation. For example, to achieve the target error probability 0.01, the uniform power allocation needs 400 iterations; this is 2.5 times more than what the optimal power allocation needs to achieve the target accuracy (150). Therefore, the total power that is needed to invest with the uniform allocation to achieve the desired accuracy of 0.01 is $400S^{\mathrm{TOT}}$, implying that the optimal allocation saves about 60% of this power.

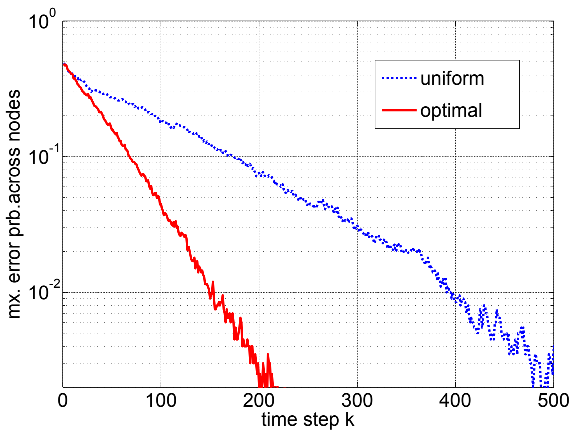

Fig. 1. Detection error probability of the worst sensor versus time k for the optimal and uniform power allocations

VII. Conclusion

We studied convergence in probability of products $W_k \cdots W_1$ of i.i.d. random stochastic, symmetric matrices. Existing literature already established that the convergence in probability of such products is exponentially fast, but did not provide the exact asymptotic convergence rate I. In this paper, we explicitly characterize the rate I for generic distributions of W_k, as long as the positive entries of W_k are bounded away from zero. We demonstrated the connection between the rate I and a generalized min-cut problem. The relation with the min-cut enabled us to find tight approximations or even the exact value I, for: 1) gossip-like models; and 2) models with (possibly spatially correlated) link failures. Finally, we demonstrated the usefulness of rate I by optimally allocating sensors' transmission power in consensus+innovations distributed detection.

References

[1] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM Rev.*, vol. 53, pp. 747–772, November 2011.

[2] B. Touri and A. Nedić, "Product of random stochastic matrices," *Submitted to the Annals of Probability*, 2011, available at: http://arxiv.org/pdf/1110.1751v1.pdf.

[3] ——, "On backward product of stochastic matrices," 2011, available at: http://arxiv.org/abs/1102.0244.

[4] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 225–230, January 2010.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[6] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, November 2010, digital Object Identifier: 10.1109/JPROC.2010.2052531.

[7] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *Accepted for publication in IEEE Transactions on Information Theory, 51 pages*, August 2008. [Online]. Available: arXiv:0809.0009v1 [cs.MA]

[8] D. Bajović, D. Jakovetić, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[9] D. Bajović, D. Jakovetić, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," 2011, available at: http://arxiv.org/abs/1111.4555.

[10] A. Tahbaz-Salehi and A. Jadbabaie, "On consensus over random networks," in *44th Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, Illinois, USA, September 2006, pp. 1315–1321.

[11] D. Bajović, J. Xavier, J. M. F. Moura, and B. Sinopoli, "Consensus and products of random stochastic matrices: Exact rate for convergence in probability," *submitted to IEEE Transactions on Signal Processing*, Feb. 2012.

[12] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 98, pp. 298–305, 1973.

[13] M. Stoer and F. Wagner, "A simple min-cut algorithm," *Journal of the ACM*, vol. 44, no. 4, pp. 585–591, July 1997.

[14] K. Chan, A. Swami, Q. Zhao, and A. Scaglione, "Consensus algorithms over fading channels," in *Proc. MILCOM 2010, Military communications conference*, San Jose, CA, October 2010, pp. 549–554.

[15] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*, ser. Grundlehren der Mathematischen Wissenschaften, Vols. 305 and 306. Berlin, Germany: Springer-Verlag, 1993.