

Consensus and Products of Random Stochastic Matrices: Exact Rate for Convergence in Probability

Dragana Bajović, *Student Member, IEEE*, João Xavier, *Member, IEEE*, José M. F. Moura, *Fellow, IEEE*, and Bruno Sinopoli, *Member, IEEE*

Abstract—We find the *exact rate* for convergence in probability of products of independent, identically distributed symmetric, stochastic matrices. It is well-known that if the matrices have positive diagonals almost surely and the support graph of the mean or expected value of the random matrices is connected, the products of the matrices converge almost surely to the average consensus matrix, and thus in probability. In this paper, we show that the convergence in probability is exponentially fast, and we explicitly characterize the exponential rate of this convergence. Our analysis reveals that the exponential rate of convergence in probability depends only on the statistics of the support graphs of the random matrices. Further, we show how to compute this rate for commonly used random models: gossip and link failure. With these models, the rate is found by solving a min-cut problem, and hence it is easily computable. Finally, as an illustration, we apply our results to solving power allocation among networked sensors in a consensus+innovations distributed detection problem.

Index Terms—Consensus, consensus+innovations, convergence in probability, exponential rate, performance analysis, random network.

I. INTRODUCTION

WE study the convergence in probability of products $W_k \cdots W_2 W_1$ of (doubly) stochastic symmetric $N \times N$ matrices W_k , $k \geq 1$. These products arise in many contexts; we consider a power allocation application in distributed detection in Section V. When 1) the matrices W_k are independent and identically distributed (i.i.d.), 2) the support graph of the expected matrix $\mathbb{E}[W_k]$ is connected, and 3) each W_k has positive

diagonals almost surely, it is well known that these products converge to the average consensus matrix $J = \frac{1}{N} \mathbf{1}\mathbf{1}^T$ almost surely [1], hence in probability. The goal of the current paper is to study the rate of this convergence in probability—namely, we establish that this convergence in probability is *exponentially fast*, and we determine the *exact exponential rate* of this convergence.

We explain our problem in intuitive terms. Consider the static (deterministic) case $W_k = A$, for all $k \geq 1$, where A is a (doubly) stochastic symmetric matrix with $|\lambda_2(A)| < 1$; let $\|\cdot\|$ denote the spectral norm. Then $\|W_k \cdots W_2 W_1 - J\| = |\lambda_2(A)|^k$, or, in words, the spectral norm of the error matrix $W_k \cdots W_2 W_1 - J = A^k - J$ decays exponentially fast with exponent $|\lambda_2(A)|$. When the W_k 's are random i.i.d. matrices, a similar behavior occurs, but now the role of $|\lambda_2(A)|$ is taken by the Lyapunov exponent $\gamma < 1$, i.e., the path of the norm $\|W_k \cdots W_2 W_1 - J\|$, $k \geq 1$, behaves as γ^k [2]–[4]¹. But, contrary to the deterministic case, because the W_k 's are random, there are paths of the norm $\|W_k \cdots W_2 W_1 - J\|$, $k \geq 1$, that decay slower than γ^k , although with vanishing probability as the size k of the product increases. To be specific, consider an arbitrary $\epsilon \in (0, 1]$ and, for large k , the *rare event* $\{\|W_k \cdots W_2 W_1 - J\| \geq \epsilon\}$. In this paper, we consider the probability of such rare events and the rate at which the sequence of these probabilities vanishes with k ; in particular, we show that the following *large deviation rate* I exists

$$I = \lim_{k \rightarrow \infty} -\frac{1}{k} \log \mathbb{P}(\|W_k \cdots W_1 - J\| \geq \epsilon) \quad (1)$$

and we show how it can be computed in terms of network parameters and the statistics of the W_k 's. Actually, we provide a stronger result on the rate I . We show that the same large deviation rate I holds for the following events. Let d_k , $k \geq 1$, be a sequence with a decay rate slower than exponential; e.g., $d_k = \frac{1}{k}$, for $k \geq 1$. Similarly to the case when $d_k \equiv \epsilon$, consider the rare event $\{\|W_k \cdots W_2 W_1 - J\| \geq d_k\}$. This is a rare event because $\|W_k \cdots W_2 W_1 - J\| \sim \gamma^k \ll d_k$. We show that the large deviation rate at which the probabilities of these rare events vanish with k is the same as the rate I in (1). More precisely, for any sequence d_k , $k \geq 1$, $d_k \in (0, 1]$, $\log d_k = o(k)$,

$$I = \lim_{k \rightarrow \infty} -\frac{1}{k} \log \mathbb{P}(\|W_k \cdots W_1 - J\| \geq d_k), \quad (2)$$

and the rate I is the same for any such sequence d_k .

¹More precisely, $\lim_{k \rightarrow \infty} (\|W_k \cdots W_2 W_1 - J\|)^{\frac{1}{k}} = \gamma$, almost surely. We also remark that γ is a constant that depends only on the statistics of the matrices W_k (and not on the particular choice of the sequence realization W_k, \dots, W_1), see also [2].

Manuscript received February 28, 2012; revised August 02, 2012, January 07, 2013, and January 14, 2013; accepted January 28, 2013. Date of publication February 20, 2013; date of current version April 24, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tongtong Li. The work of D. Bajović, J. Xavier, and B. Sinopoli was 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 J. M. F. Moura was partially supported by the NSF under Grants CCF-1011903 and CCF-1018509, and by an AFOSR grant FA95501010291. D. Bajović holds a fellowship from the Carnegie Mellon/Portugal Program.

D. Bajović is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal. She is also with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: dragana@isr.ist.utl.pt; dbajovic@andrew.cmu.edu).

J. Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal (e-mail: jxavier@isr.ist.utl.pt).

J. M. F. Moura and B. Sinopoli are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: brunos@ece.cmu.edu; moura@ece.cmu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2248003

Our results reveal that the large deviation rate I is solely a function of the graphs induced by the matrices W_k and the corresponding probabilities of occurrences of these graphs. In general, the computation of the rate I is a combinatorial problem. However, for special important cases, we can get particularly simple expressions. For example, when the matrices W_k are the weight matrices for gossip consensus on a tree, the rate I is equal to $|\log(1 - p_{ij})|$, where p_{ij} is the probability of the link $\{i, j\}$ that is least likely to occur in the gossip protocol. Another example is with gossip consensus over a regular graph when $p_{ij} \equiv p$ in which case we show that the rate I equals $|\log p_{\text{isol}}|$, where $p_{\text{isol}} = 1 - dp$ is the probability that a node is isolated from the rest of the network and d is the degree of a node. For gossip over more general graph structures, we show that the rate $I = |\log(1 - c)|$ where c is the min-cut value (or connectivity [5]) of a graph whose links are weighted by the gossip link probabilities; the higher the connectivity c is (the more costly or, equivalently, less likely it is to disconnect the graph) the larger the rate I and the faster the convergence are. Similarly, for consensus algorithms running on networks with link failures on general graphs, the rate is computed by solving a min-cut problem and is computable in polynomial time.

Review of the literature. There has been a large amount of work on linear systems driven by stochastic matrices. Early work includes [6], [7], and the topic received renewed interest in the past decade [8], [9]. Reference [9] analyzes convergence of the consensus algorithm under deterministic time-varying matrices W_k . Reference [10] provides a detailed study of the standard gossip model that has been further modified, e.g., in [11], [12]; for a recent survey, see [13]. Reference [1] analyzes convergence under random matrices W_k , not necessarily symmetric, and ergodic—hence not necessarily independent in time. Reference [14] studies effects of delays, while [15] studies the impact of quantization. Reference [16] considers random matrices W_k and addresses the issue of the communication complexity of consensus algorithms. The recent [17] surveys consensus and averaging algorithms and provides tight bounds on the worst case averaging times for deterministic time varying networks. In contrast with consensus (averaging) algorithms, *consensus+innovations* algorithms include both a local averaging term (consensus) and an innovation term (measurement) in the state update process. These algorithms find applications in distributed inference in sensor networks, see, e.g., [18]–[20] for distributed estimation, and, e.g., [21]–[23], for distributed detection. In this paper, we illustrate the usefulness of the large deviation rate I in the context of a *consensus+innovations* algorithms by allocating the transmission power of sensors for distributed detection.

Relation with the literature on products of stochastic matrices. Products of random matrices appear also in many other fields that use techniques drawn from Markov process theory. Examples include repeated interaction dynamics in quantum systems [24], inhomogeneous Markov chains with random transition matrices [25], [24], infinite horizon control strategies for Markov chains and non-autonomous linear differential equations [26], or discrete linear inclusions [27]. These papers are usually concerned with deriving convergence results on these products and determining the limiting matrix. Reference [24]

studies the product of matrices belonging to a class of complex contraction matrices and characterizes the limiting matrix by expressing the product as a sum of a decaying process, which exponentially converges to zero, and a fluctuating process. Reference [26] establishes conditions for strong and weak ergodicity for both forward and backward products of stochastic matrices, in terms of the limiting points of the matrix sequence. Using the concept of infinite flow graph, which the authors introduced in previous work, [25] characterizes the limiting matrix for the product of stochastic matrices in terms of the topology of the infinite flow graph. For more structured matrices, [28] studies products of nonnegative matrices. For nonnegative matrices, a comprehensive study of the asymptotic behavior of the products can be found in [29]. A different line of research, closer to our work, is concerned with the limiting distributions of the products (in the sense of the central limit theorem and large deviations). The classes of matrices studied are: invertible matrices [30], [31] and its subclass of matrices of determinant equal to 1, [32], and, also, positive matrices [33]. None of these apply to our case, as the matrices that we consider are not invertible ($W_k - J$ has a zero eigenvalue, for every realization of W_k) and, also, we allow the entries of W_k to be zero, and therefore the entries of $W_k - J$ might be negative with positive probability. Furthermore, as pointed out in [34], the results obtained in [30]–[32] do not provide ways to effectively compute the rates of convergence. Reference [34] improves on the existing literature in that sense by deriving more explicit *bounds* on the convergence rates, while showing that, under certain assumptions on the matrices, the convergence rates do not depend on the size of the matrices; the result is relevant from the perspective of large scale dynamical systems, as it shows that, in some sense, more complex systems are not slower than systems of smaller scale, but again it does not apply to our study.

When studying the products $W_k \cdots W_1$, most of the existing work, e.g., [10], [35], [36], [3], is concerned with (standard) consensus or gossip algorithms, and not with the *consensus+innovations* algorithms. References [10], [36] consider other metrics, e.g., the ϵ -averaging time, and $\lambda_2(\mathbb{E}[W_k^2])$. Further, [3] considers $\lim_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{E}[\|W_k \cdots W_1 - J\|]$. Note the difference to (2): [3] considers the rate for the average path of the product, while we consider the rate of the probability that the product has a worst-case path, with a sub-exponential dynamics. We note that, when studying *consensus+innovations* distributed detection, the large deviation rate I arises as a natural metric. (See Section V, and Fig. 2, left.) To our best knowledge, the *exact* large deviations rate I in (2) has not been computed for i.i.d. averaging matrices W_k , nor for the commonly used sub-classes of gossip and link failure models. From existing results, one can deduce upper bounds on I , but not the exact rate I . Consider, e.g., [10] and [36]; [10] is concerned with computing the ϵ -averaging times of gossip; in [36], as a means of showing the almost sure convergence of consensus, the authors prove that $\mathbb{P}(\|W_k \cdots W_1 - J\| \geq \epsilon)$ decays exponentially fast in k . With both [10], [36], the upper bounds on $\mathbb{P}(\|W_k \cdots W_1 - J\| \geq \epsilon)$ are derived by an application of Markov's (i.e., Chebyshev's) inequality (see, e.g., (32) in [10] or Theorem 6 in [36]); the latter yields a bound on I based on the second largest (in modulus) eigenvalue of

$\mathbb{E}[W_k]$ or $\mathbb{E}[W_k^2]$. However, from these bounds, we cannot see a very particular, distinguishing characteristic of the large deviation probabilities in (2), that is, the probabilities that the dynamics of consensus is sub-exponential, that we discover: in the asymptotic regime, these probabilities depend on the distribution of matrices W_k only through the distribution of their support graphs. For example, if each realization of W_k has a connected underlying support graph (the case studied in [9]), we calculate the rate I to be equal $+\infty$ (see Section III). That is, if each realization of W_k has a connected graph, then the convergence of the product $W_k \cdots W_1$ cannot be slower than exponential. On the other hand, the “rate” that would result from the bound based on $\lambda_2(\mathbb{E}[W_k^2])$ is finite unless $W_k \equiv J$.

One of our motivations for finding the limits in (2) and (1) comes from distributed inference in sensor networks, and, in particular, consensus+innovations distributed detection, e.g., [23], [22]. In [23], [22] we discovered that the rate I is the key network parameter that captures the effect of the network on distributed detection performance. Here we show that the rate I is a useful metric for the design of distributed algorithms by optimizing the allocation of the sensors’ transmission power in a sensor network with fading (failing) links in consensus+innovations detection [22], [23].

Paper organization. Section II introduces the model for random matrices W_k and defines relevant quantities needed in the sequel. Section III proves the result on the exact exponential rate I of consensus. Section IV shows how to compute the rate I for gossip and link failure models via a min-cut problem. Section V addresses the power allocation for distributed detection by maximizing the rate I . Finally, Section VI concludes the paper.

Notation. We denote by A_{ij} or $[A]_{ij}$ the entry ij of a matrix A . For $N \in \mathbb{N}$, we denote by \mathbb{S}^N the set of stochastic symmetric N by N matrices; by \mathbb{G}^N the set of all undirected graphs on the set of vertices $V = \{1, \dots, N\}$; by I_N the identity matrix of size $N \times N$. For a graph $G \in \mathbb{G}^N$ we denote with $\lambda_F(G)$ the Fiedler value of G , i.e., the second smallest eigenvalue of the Laplacian matrix of G ; by $A(G)$ the adjacency matrix of G , defined by $[A(G)]_{ij} = 1$ if $\{i, j\}$ belongs to G , and $[A(G)]_{ij} = 0$ otherwise. $U(0, 1)$ denotes the uniform distribution on the interval $[0, 1]$; $\lceil x \rceil$ denotes the smallest integer not less than x . For a finite set S we denote by $\binom{V}{2}$ the set of all two-element subsets of V ; by $|S|$ the cardinality of S .

II. PROBLEM SETUP

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is the set of outcomes, \mathcal{F} is a sigma algebra on Ω , and \mathbb{P} is a probability measure on (Ω, \mathcal{F}) . Let $W_t : \Omega \mapsto \mathbb{S}^N$, $t \geq 1$, be a sequence of maps that are $(\mathcal{F}, \mathcal{B}(\mathbb{R}^{N \times N}) \cap \mathbb{S}^N)$ -measurable, that is, for any $B \in \mathcal{B}(\mathbb{R}^{N \times N}) \cap \mathbb{S}^N$, $\{W_t \in B\}$ belongs to \mathcal{F} , for all $t \geq 1$. In other words, $\{W_t\}_{t \geq 1}$ is a sequence of random matrices on $(\Omega, \mathcal{F}, \mathbb{P})$.

Assumption 1:

- 1) Random matrices W_t , $t \geq 1$, are independent and identically distributed (i.i.d.);
- 2) Diagonal entries of W_t are almost surely positive, i.e., for each t , almost surely $[W_t]_{ii} > 0$ for all $i = 1, \dots, N$.

Let $\Phi(s, t)$ denote the product of the matrices that occur from time $t+1$ until time s , $1 \leq t < s$, $\Phi(s, t) = W_s \cdots W_{t+1}$. Also, let $\tilde{\Phi}(s, t) = \Phi(s, t) - J$; we call $\tilde{\Phi}(s, t)$ the error matrix from time t until time s .

To analyze the products $\Phi(s, t)$, we introduce the induced graph operator $G : \mathbb{S}^N \mapsto \mathbb{G}^N$. For $W \in \mathbb{S}^N$, we define $G(W)$ by

$$G(W) = \left(V, \left\{ \{i, j\} \in \binom{V}{2} : W_{ij} > 0 \right\} \right). \quad (3)$$

Thus, the set of edges of the induced graph of a matrix encodes which entries of the matrix are positive.

Sequence of induced graphs. Using the induced graph operator, from the sequence $\{W_t\}_{t \geq 1}$, we derive the sequence of random graphs $\{G_t\}_{t \geq 1}$ by assigning $G_t = G(W_t)$, for $t = 1, 2, \dots$. More precisely, $G_t : \Omega \mapsto \mathbb{G}^N$, for $t \geq 1$, is defined by $G_t(\omega) = G(W_t(\omega))$, for any $\omega \in \Omega$. Note that, for any $t \geq 1$, G_t is $(\mathcal{F}, 2^{\mathbb{G}^N})$ -measurable, that is, for any $\mathcal{H} \subseteq \mathbb{G}^N$, the event $\{G_t \in \mathcal{H}\}$ belongs to \mathcal{F} .

As the random matrices W_t , $t \geq 1$ are independent, it follows by the disjoint block theorem [37] that the random graphs G_t , $t \geq 1$ are also independent. Furthermore, as W_t are identically distributed, it follows that, for any $H \in \mathbb{G}^N$, the probability $\mathbb{P}(G_t = H) = \mathbb{P}(G(W_t) = H)$ is the same at all times. Thus, the sequence $\{G_t\}_{t \geq 1}$ is i.i.d., and each G_t is distributed according to the same probability mass function p_H , $H \in \mathbb{G}^N$, where

$$p_H = \mathbb{P}(G(W_t) = H).$$

Further, for a collection $\mathcal{H} \subseteq \mathbb{G}^N$, let $p_{\mathcal{H}}$ denote the probability that the induced graph of W_t belongs to \mathcal{H} , that is, $p_{\mathcal{H}} = \mathbb{P}(G_t \in \mathcal{H})$. Then, $p_{\mathcal{H}} = \sum_{H \in \mathcal{H}} p_H$. Finally, we collect in the set \mathcal{G} all the graphs that occur with positive probability:

$$\mathcal{G} := \{H \in \mathbb{G}^N : p_H > 0\}, \quad (4)$$

and we call \mathcal{G} the set of realizable graphs. For example, if H contains a link $\{i, j\}$ such that $\mathbb{P}([W_t]_{ij} > 0) = 0$, then $H \notin \mathcal{G}$; similarly, if for some $\{i, j\}$ $\mathbb{P}([W_t]_{ij} > 0) = 1$, then all realizable graphs must contain this link. The complete graph $\mathcal{G} = (V, \binom{V}{2})$ is obtained whenever W_t has a joint probability density function that is strictly positive on \mathbb{S}^N . We next give examples of sequences of random matrices that satisfy Assumption 1 and, for each of the examples, we derive the set of realizable graphs and compute the distribution of the corresponding induced graphs.

Example 1 (Gossip With Uniformly Distributed Weights): Let $\hat{G} = (V, \hat{E})$ be an arbitrary connected graph on N vertices. At each time $t \geq 1$ a node in V is chosen independently from the previous choices and according to the probability mass function $r_u > 0$, $u \in V$, $\sum_{u \in V} r_u = 1$. The chosen node then randomly chooses a neighbor in \hat{G} according to the probability mass function $q_{uv} > 0$, $\{u, v\} \in \hat{E}$, $\sum_{\{u, v\} \in \hat{E}} q_{uv} = 1$, $u \in V$. Denote the node chosen at time t and its chosen neighbor by u_t and v_t , respectively. With gossip with uniformly distributed weights, averaging occurs only at the edge that is active at time

$t, \{u_t, v_t\}$, and with weight equal to the realization of a uniformly distributed parameter $\alpha_t \sim U[0, 1]$. Correspondingly, the weight matrix at time t is $W_t = I - \alpha_t(e_{u_t} - e_{v_t})(e_{u_t} - e_{v_t})^\top$. We assume that $\alpha_t, t \geq 1$, are independent random variables, and, also, that α_t is independent of u_s, v_s , for all s, t , implying that the sequence $W_t, t \geq 1$ is i.i.d. Also, since $\alpha_t = 1$ with probability zero, diagonal entries of W_t are almost surely positive, and we conclude that the sequence of random matrices $\{W_t\}_{t \geq 1}$ constructed in this way satisfies Assumption 1.

By construction, every realization of W_t is of the form $I - \alpha(e_u - e_v)(e_u - e_v)^\top$, for some $\alpha \in [0, 1]$ and $u, v \in V$ such that $\{u, v\} \in \hat{E}$. Thus, every realization of G_t is of the form: 1) (V, \emptyset) , when $\alpha_t = 0$; or 2) $(V, \{u, v\})$, for $\{u, v\} \in \hat{E}$. Since $\alpha_t = 0$ with probability 0, we have that $p_{(V, \emptyset)} = 0$, and, so, the potential candidates for realizable graphs are only graphs from the second category. Now, for $\{u, v\} \in \hat{E}$, $p_{(V, \{u, v\})} = \mathbb{P}(\alpha_t > 0, u_t = u \text{ and } v_t = v \text{ or } u_t = v \text{ and } v_t = u)$. Since α_t is independent of u_t and v_t , it follows that $p_{(V, \{u, v\})} = r_u q_{uv} + r_v q_{vu} > 0$, showing that $(V, \{u, v\})$ is a realizable graph. Summing up, the set of realizable graphs for gossip with uniformly distributed weights running on \hat{G} is the set of all one-link subgraphs of \hat{G}

$$\mathcal{G}^{\text{Gossip}}(\hat{G}) = \{(V, \{u, v\}) : \{u, v\} \in \hat{E}\}. \quad (5)$$

We remark that the same conclusions would be obtained if the uniform distribution, which generates α_t , was replaced by an arbitrary distribution $\mu : \mathcal{B}([0, 1]) \mapsto [0, 1]$ satisfying $\mu((0, 1)) = 1$.

Example 2 (Link Failure Model With Metropolis Weights): Consider a connected network defined by $\hat{G} = (V, \hat{E})$. We assume that, at any time $t \geq 1$, only edges in \hat{E} can occur, and, also, that occurrence of edge $e \in \hat{E}$ at time t is modeled as a Bernoulli random variable $Z_{e,t} \sim \text{Ber}(p_e)$, for $e \in \hat{E}$, where $p_e \in (0, 1)$. We assume that occurrences of edges are independent across space and in time. For $t \geq 1$ and $i = 1, \dots, N$, let $d_{i,t} = \sum_{j: \{i,j\} \in \hat{E}} Z_{\{i,j\},t}$, that is, $d_{i,t}$ is the degree of node i at time t . The weight matrix at time t is chosen by $[W_t]_{ij} = \frac{1}{1 + \max\{d_{i,t}, d_{j,t}\}}$, for all $\{i, j\} \in E_t$, $[W_t]_{ii} = 1 - \sum_{j=1}^N [W_t]_{ij}$, $i = 1, \dots, N$ and $[W_t]_{ij} = 0$, otherwise. It can be easily shown that, for every realization of $Z_{e,t}, e \in \hat{E}$, diagonal entries of W_t are positive. Further, since $\{Z_{e,t}\}_{e \in \hat{E}}$ are independent (in time), and for any $e \in E$, $Z_{e,t}$ for $t \geq 1$ are identically distributed, it follows that random matrices W_t are i.i.d. Thus, the sequence $\{W_t\}_{t \geq 1}$ satisfies Assumption 1.

For each time t , let E_t collect all the edges that are online at time t , $E_t = \{e : Z_{e,t} = 1\}$. Then, by construction of W_t , $G_t = (V, E_t)$, for all t . Using this fact, for any $H = (V, E) \in \mathcal{G}^N$ such that $E \subseteq \hat{E}$, we get $p_H = \mathbb{P}(Z_{e,t} = 1, e \in E \text{ and } Z_{e,t} = 0, e \notin E)$, which by the independence assumption yields $p_H = \prod_{e \in E} p_e \prod_{f \notin E} (1 - p_f) > 0$. We conclude that the set of realizable graphs for the link failure model on \hat{G} is the set of all subgraphs of \hat{G} :

$$\mathcal{G}^{\text{Link fail.}}(\hat{G}) = \{(V, E) : E \subseteq \hat{E}\}. \quad (6)$$

Accumulation graph and disconnected collections. For a collection of graphs $\mathcal{H} \subseteq \mathcal{G}^N$, we denote by $\Gamma(\mathcal{H})$ the graph that

contains all edges from all graphs in \mathcal{H} . That is, $\Gamma(\mathcal{H})$ is the minimal graph (i.e., the graph with the minimal number of edges) that is a supergraph of every graph in \mathcal{H} :

$$\Gamma(\mathcal{H}) := \left(V, \bigcup_{G \in \mathcal{H}} E(G) \right), \quad (7)$$

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

Specifically, for any $1 \leq t < s$, we denote by $\Gamma(s, t)^2$ the random graph that collects the edges from all the graphs G_r that appeared from time $r = t + 1$ to $r = s$, $s > t$, i.e.,

$$\Gamma(s, t) := \Gamma(\{G_s, G_{s-1}, \dots, G_{t+1}\}),$$

and we call $\Gamma(s, t)$ the accumulation graph from time t until time s .

We next define collections of realizable graphs of certain types that will be important in computing the rate in (2) and (1).

Definition 3: The collection $\mathcal{H} \subseteq \mathcal{G}$ is a disconnected collection on \mathcal{G} if its accumulation graph $\Gamma(\mathcal{H})$ is disconnected.

Thus, a disconnected collection is any collection of realizable graphs such that the union of all of its graphs yields a disconnected graph. We also define the set of all possible disconnected collections on \mathcal{G} :

$$\Pi(\mathcal{G}) = \{\mathcal{H} \subseteq \mathcal{G} : \mathcal{H} \text{ is a disconnected collection on } \mathcal{G}\}. \quad (8)$$

Example 4 (Gossip Model): Consider the gossip algorithm from Example 1 when \hat{G} is the complete graph on N vertices. In this case $\mathcal{G} = \{(V, \{i, j\}) : \{i, j\} \in \binom{V}{2}\}$, that is, \mathcal{G} is the set of all possible one-link graphs on N vertices. An example of a disconnected collection of \mathcal{G} is $\mathcal{G} \setminus \{(V, \{i, j\}) : j = 1, \dots, N\}$, where i is a fixed vertex, or, in words, the collection of all one-link graphs except of those whose link is adjacent to i . Another example is $\mathcal{G} \setminus (\{(V, \{i, k\}) : k = 1, \dots, N, k \neq j\} \cup \{(V, \{j, l\}) : l = 1, \dots, N, l \neq i\})$, where $\{i, j\}$ is a fixed link.

Example 5 (Toy Example): Suppose that, for some sequence of random matrices taking values in \mathbb{S}^5 , the set of realizable graphs is $\mathcal{G} = \{G_1, G_2, G_3\}$, where graphs $G_i, i = 1, 2, 3$ are given in Fig. 1. In this model each realizable graph is a two-link graph and the supergraph of all the realizable graphs $\Gamma(\{G_1, G_2, G_3\})$ is connected. If we scan over the supergraphs $\Gamma(\mathcal{H})$ of all subsets \mathcal{H} of \mathcal{G} , we see that $\Gamma(\{G_1, G_2\}), \Gamma(\{G_2, G_3\})$ and $\Gamma(\{G_1, G_2, G_3\})$ are connected, whereas $\Gamma(\{G_1, G_3\})$ and $\Gamma(G_i) = G_i, i = 1, 2, 3$, are disconnected. It follows that $\Pi(\mathcal{G}) = \{\{G_1\}, \{G_2\}, \{G_3\}, \{G_1, G_3\}\}$.

III. CONVERGENCE IN PROBABILITY—EXPONENTIAL RATE

This Section states and proves the main result of this paper, Theorem 6. We prove Theorem 6 by proving the corresponding large deviation upper and lower bound; the proof of the lower

²Graph $\Gamma(s, t)$ is associated with the matrix product $W_s \cdots W_{t+1}$ going from time $t + 1$ until time $s > t$. The notation $\Gamma(s, t)$ indicates that the product is backwards; see also the definition of the product matrix $\Phi(s, t)$ after Assumption 1 at the beginning of this section.

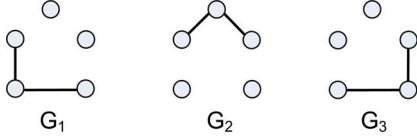


Fig. 1. Example of a five node network with three possible graph realizations, each being a two-link graph.

bound is given in Subsection III.A, whereas the proof of the upper bound is given in Subsection III.B.

Theorem 6: Let d_k be a sequence of real numbers such that $d_k \in (0, 1]$ and $\log d_k = o(k)$. Then:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) = -I,$$

where

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

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

We prove Theorem 6, by proving separately the lower bound (9) and the upper bound (10). We remark that we need to prove the lower bound (9) only for the case when $\Pi(\mathcal{G}) \neq \emptyset$, as the bound trivially holds when $I = +\infty$.

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) \geq -I \quad (9)$$

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) \leq -I. \quad (10)$$

Subsection III.A proves the lower bound (9), and Subsection III.B proves the upper bound (10).

A. Proof of the Lower Bound (9)

We first show that, for any $k \geq 1$, a sufficient condition for the norm $\|\tilde{\Phi}(k, 0)\|$ being above d_k is that the supergraph $\Gamma(k, 0)$ is disconnected. In fact, we prove the following stronger claim.

Lemma 7: For any fixed $\omega \in \Omega$ and any $k \geq 1$

$$\Gamma(k, 0) \text{ is disconnected} \Rightarrow \|\tilde{\Phi}(k, 0)\| = 1.$$

Proof: Fix $\omega \in \Omega$ and $k \geq 1$ and suppose that $\Gamma(k, 0)$ is not connected. Suppose further (without loss of generality) that $\Gamma(k, 0)$ has exactly two components and denote them by C_1 and C_2 . Then, for all i, j such that $i \in C_1$ and $j \in C_2$, we have $\{i, j\} \notin \Gamma(k, 0)$, and, consequently, $\{i, j\} \notin G_t$, for all $1 \leq t \leq k$. By definition of G_t , this implies that the corresponding entries in the matrices W_t , $1 \leq t \leq k$, are equal to zero, i.e.,

$$\forall t, 1 \leq t \leq k : [W_t]_{ij} = 0, \forall \{i, j\} \text{ s.t. } i \in C_1, j \in C_2.$$

Thus, every matrix realization W_t from time 1 to time k has a block diagonal form (up to a symmetric permutation of rows and columns, the same for all W_t)

$$W_t = \begin{bmatrix} [W_t]_{C_1} & 0 \\ 0 & [W_t]_{C_2} \end{bmatrix},$$

where $[W_t]_{C_1}$ is the block of W_t corresponding to the nodes in C_1 , and similarly for $[W_t]_{C_2}$. This implies that $\Phi(k, 0)$ has the same block diagonal form, which, in turn, proves that $\|\tilde{\Phi}(k, 0)\| = 1$. ■

Using the result of Lemma 7, we get:

$$\begin{aligned} \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) &\geq \mathbb{P}(\|\tilde{\Phi}(k, 0)\| = 1) \\ &\geq \mathbb{P}(\Gamma(k, 0) \text{ is disconnected}). \end{aligned} \quad (11)$$

We now focus on computing the probability of the event that $\Gamma(k, 0)$ is disconnected. For any fixed $k \geq 1$, a sufficient condition that guarantees that $\Gamma(k, 0)$ is disconnected is that every graph realization G_t from time 1 to time k is drawn from some disconnected collection $\mathcal{H} \in \Pi(\mathcal{G})$. More precisely, for every $\mathcal{H} \in \Pi(\mathcal{G})$ and every $k \geq 1$, it holds for all $\omega \in \Omega$:

$$G_t \in \mathcal{H}, \text{ for } 1 \leq t \leq k \Rightarrow \Gamma(k, 0) \text{ is disconnected}. \quad (12)$$

This can be easily shown by observing that if $\{G_1, \dots, G_k\} \subseteq \mathcal{H}$, then $\Gamma(k, 0) = \Gamma(\{G_1, \dots, G_k\})$ is a subgraph of $\Gamma(\mathcal{H})$, or, in other words, $\Gamma(k, 0)$ cannot contain any additional edge beyond the ones in $\Gamma(\mathcal{H})$. Now, since $\Gamma(\mathcal{H})$ is disconnected, it must be that $\Gamma(k, 0)$ is disconnected as well. Claim in (12) implies that for every $\mathcal{H} \in \Pi(\mathcal{G})$ and every $k \geq 1$

$$\begin{aligned} \mathbb{P}(\Gamma(k, 0) \text{ is disconnected}) &\geq \mathbb{P}(G_t \in \mathcal{H}, \\ &\text{for } 1 \leq t \leq k) = p_{\mathcal{H}}^k, \end{aligned} \quad (13)$$

where the last equality follows by the time independence assumption. Combining the previous bound with (11) and optimizing the bound over $\mathcal{H} \in \Pi(\mathcal{G})$ yields

$$\mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) \geq p_{\max}^k.$$

Finally, taking the log, dividing by k , and taking the \liminf over $k \rightarrow \infty$, the lower bound (9) follows.

B. Proof of the Upper Bound (10)

To prove the upper bound, we first extend the concept of the induced graph operator to the δ -induced graph operator which accounts only for those entries that are above some given $\delta > 0$, i.e., the entries that are sufficiently important. Using the definition of the δ -induced graph, we correspondingly extend the concepts of the accumulation graph, the set of realizable graphs and the most likely disconnected collection. We explain this next.

The family of δ -induced graph sequences.

Definition 8: For each $\delta > 0$ we define the δ -induced graph operator $G_{\delta} : \mathbb{S}^N \mapsto \mathbb{G}^N \cup \{E\}$ by (14), shown at the bottom of the page.

As we can see from the definition, if a matrix has all diagonal entries above δ , then its δ -induced graph contains all the edges whose corresponding entries of the matrix are above δ . On the other hand, any matrix that has a diagonal entry below

$$G_{\delta}(W) = \begin{cases} \left(V, \left\{ \{i, j\} \in \binom{V}{2} : W_{ij} \geq \delta \right\} \right), & \text{if } W_{ii} \geq \delta, \forall i \\ E, & \text{otherwise} \end{cases} \quad (14)$$

δ gets mapped by G_δ to the symbol E; note that, by doing this, we discard all the potential edges for such a matrix (no matter how large their corresponding entries are). Intuitively, δ -induced graph operator G_δ , compared to G , acts as an edge truncator that cuts off all the non-significant edges and, also, it discards all the matrices with low diagonals by mapping them to E. We will see later in the analysis that, whenever at some point in the sequence $W_t, t \geq 1$, a matrix with a small diagonal entry occurs, we cannot say much about the continuity of the “information flow” at that point. Thus, we introduce a special symbol, E, to indicate such matrices that cut (or “erase”) the information flow.

We now use operators $G_\delta, \delta > 0$, to construct from $\{W_t\}_{t \geq 1}$ new induced graph sequences. For every $\delta > 0, t \geq 1$, let $G_{t,\delta} : \Omega \mapsto \mathbb{G}^N \cup \{E\}$ be defined by $G_{t,\delta}(\omega) = G_\delta(W_t(\omega))$, for $\omega \in \Omega$. Thus, for every $\delta, G_{t,\delta}$ is the δ -induced graph of the matrix $W_t, t \geq 1$. Remark that, in contrast with the regular induced graph $G_t, G_{t,\delta}$ can take value E.

Each sequence $\{G_{t,\delta}\}_{t \geq 1}$ from this family indexed by δ is i.i.d., as the sequence $\{W_t\}_{t \geq 1}$ is i.i.d. For any $H \in \mathbb{G}^N$, denote by $p_{H,\delta}$ the probability that $G_{t,\delta}$ is equal to H , i.e., $p_{H,\delta} = \mathbb{P}(G_{t,\delta} = H)$. The probability that $G_{t,\delta}$ takes value E is denoted by $p_{E,\delta} = \mathbb{P}(G_{t,\delta} = E)$. We show in Lemma 9 that, for each $t, G_{t,\delta}$ converges almost surely to G_t as $\delta \rightarrow 0$, thus implying the corresponding convergence in distribution. For convenience, we state the result in terms of the adjacency matrices: for any $\omega \in \Omega, t \geq 1$ and $\delta > 0$, we define $A_t(\omega) = A(G_t(\omega)), A_{t,\delta}(\omega) = A(G_{t,\delta}(\omega))$, if $G_{t,\delta}(\omega) \neq E$, otherwise, we assign $A_{t,\delta}(\omega)$ to be the N by N matrix of all zeros.

Lemma 9: For any $t \geq 1$, almost surely $A_{t,\delta} \rightarrow A_t$, as $\delta \rightarrow 0$. Hence, for any $H \in \mathbb{G}^N$, $\lim_{\delta \rightarrow 0} p_{H,\delta} = p_H$ and also $\lim_{\delta \rightarrow 0} p_{E,\delta} = 0$.

Proof: For any $t \geq 1$, let $\Omega_t^* = \{[W_t]_{ii} > 0, \forall i\}$; note that, by Assumption 1, $\mathbb{P}(\Omega_t^*) = 1$. Now, fix t and $\omega \in \Omega_t^*$, and consider $W = W_t(\omega)$. Then, $W_{ii} > 0$ for all i and let $\delta_0 = \min_i W_{ii}$ (note that $\delta_0 > 0$ and also that it depends on ω). For all $\delta > \delta_0, G_\delta(W) = E$, whereas for all $\delta \in (0, \delta_0], G_\delta(W) \in \mathbb{G}^N$. Note that, to prove the claim, it is sufficient to consider only the case when $\delta \leq \delta_0$. First, for all $\{i, j\}$ such that $W_{ij} = 0$, we have $[A_t(\omega)]_{ij} = 0$ and also $[A_{t,\delta}(\omega)]_{ij} = 0$ for all $\delta \leq \delta_0$ (in fact, due to the definition of $A_{t,\delta}$, the latter holds for all δ), showing that $[A_{t,\delta}(\omega)]_{ij}$ converges to $[A_t(\omega)]_{ij}$. On the other hand, let α be the minimum over all positive entries of W , $\alpha = \min_{\{i,j\}: W_{ij} > 0} W_{ij}$ and note that $\alpha \leq \delta_0$ and $\alpha > 0$. Then, for all $\delta \leq \alpha, G_\delta(W)$ and $G(W)$ match, implying that $A_{t,\delta}(\omega) = A_t(\omega)$ for all such δ . As ω was an arbitrary point from Ω_t^* and since $\mathbb{P}(\Omega_t^*) = 1$, the almost sure convergence follows. The second part of the claim follows from the fact that almost sure convergence implies the convergence in distribution. ■

Similarly as with the set of realizable graphs, for each $\delta > 0$, we define the set of δ -realizable graphs

$$\mathcal{G}_\delta = \{H \in \mathbb{G}^N : p_{H,\delta} > 0\}. \quad (15)$$

For a collection of graphs $\mathcal{H} \subseteq \mathbb{G}^N$, we denote by $p_{\mathcal{H},\delta}$ the probability that $G_{t,\delta}$ belongs to \mathcal{H} , which is equal to $p_{\mathcal{H},\delta} = \sum_{H \in \mathcal{H}} p_{H,\delta}$. Similarly as before, $\Pi(\mathcal{G}_\delta)$ denotes the set of all possible disconnected collections on \mathcal{G}_δ .

For $\delta > 0$ such that $\Pi(\mathcal{G}_\delta) \neq \emptyset$, let $p_{\max,\delta} = \max_{\mathcal{H} \in \Pi(\mathcal{G}_\delta)} p_{\mathcal{H},\delta}$; that is, $p_{\max,\delta}$ is the probability that $G_{t,\delta}$ belongs to the most likely disconnected collection on \mathcal{G}_δ . The following corollary of Lemma 9 is one of the main steps in the proof of the upper bound (10). We omit the proof of Corollary 10 due to lack of space.

Corollary 10: If $\Pi(\mathcal{G}) \neq \emptyset$, then there must exist $\bar{\delta} > 0$ such that, for every $0 < \delta \leq \bar{\delta}$, $\Pi(\mathcal{G}_\delta) \neq \emptyset$. Moreover, $\lim_{\delta \rightarrow 0} p_{\max,\delta} = p_{\max}$.

Similarly as with accumulation graph $\Gamma(s, t)$ that collects all the edges of the (regular) induced graphs G_{t+1}, \dots, G_s , for each $\delta > 0$, we define the δ -accumulation graph $\Gamma_\delta(s, t)$ to collect the edges of the δ -induced graphs $G_{t+1,\delta}, \dots, G_{s,\delta}$. In contrast with $\Gamma(s, t)$, here we have to take into account that, for some δ , realizations of the δ -induced graphs might be equal to E. To handle this, for each $\delta > 0$ and $t \geq 1$, we introduce $R_{t,\delta} : \Omega \mapsto \mathbb{N} \cup \{0\}$ which we define by (16), at the bottom of the page. Now, for any $1 \leq t < s$ and $\delta > 0$, we define $\Gamma_\delta(s, t)$ to be

$$\Gamma_\delta(s, t) = \begin{cases} \Gamma(\{G_{s,\delta}, \dots, G_{t+1,\delta}\}), & \text{if } R_{s,\delta} \leq t \\ \Gamma(\{G_{s,\delta}, \dots, G_{R_{s,\delta}+1,\delta}\}), & \text{if } t < R_{s,\delta} < s \\ E, & \text{if } R_{s,\delta} = s. \end{cases} \quad (17)$$

We now explain the intuition behind this construction of $\Gamma_\delta(s, t)$. If $R_{s,\delta} \leq t$, that is, if the interval from t until s is clear from the realization E, then we assign $\Gamma_\delta(s, t)$ to collect all the edges of all the δ -induced graph realizations that occurred from time $t + 1$ until time s . If, on the other hand, it happens that, starting from time t we encounter the realization E, i.e., if $G_{r,\delta} = E$ for some $r > t$, we consider this to be a bad event and we reset the number of collected edges so far to zero (formally, by assigning at time r $\Gamma_\delta(r, t) = E$). We repeat this until we hit time s . Since the last occurrence of the bad realization E was at time $R_{s,\delta}$, assuming that $R_{s,\delta} < s$, the δ -accumulation graph will contain all the edges of the δ -induced graph realizations that occurred from time $R_{s,\delta} + 1$ until time s .

We have seen in the proof of the lower bound in Lemma 7 that, if the accumulation graph $\Gamma(k, 0)$ is disconnected, then the norm of the error matrix is still equal to 1 at time k . Lemma 11

$$R_{t,\delta}(\omega) = \begin{cases} 0, & \text{if } G_{r,\delta}(\omega) \neq E, \text{ for all } 1 \leq r \leq t \\ \max\{1 \leq r \leq t : G_{r,\delta}(\omega) = E\}, & \text{otherwise} \end{cases}. \quad (16)$$

is, in a sense, a converse to this result, as it provides a sufficient condition in terms of $\Gamma_\delta(s, t)$ for the norm of the error matrix to drop on the time interval from t until s .

Lemma 11: For any fixed $\omega \in \Omega$, for all $1 \leq t < s$ and $\delta \in (0, 1)$ such that $\Gamma_\delta(s, t) \neq E$, it holds

$$\|\tilde{\Phi}(s, t)\|^2 \leq 1 - \lambda_F(\Gamma_\delta(s, t))\delta^{2(s-t)}. \quad (18)$$

Using the fact that the Fiedler value (algebraic connectivity) of a connected graph is positive [38], if $\Gamma_\delta(s, t)$ is connected (and $\Gamma_\delta(s, t) \neq E$), then the squared norm of the error matrix on this interval drops for at least $\lambda_F(\Gamma_\delta(s, t))\delta^{2(s-t)} > 0$. To get a uniform bound for this drop (that holds for all connected realizations of $\Gamma(s, t)$), we use the Fiedler value of the path graph on N vertices. This is stated next in Corollary 12.

Corollary 12: For any fixed $\omega \in \Omega$, for all $1 \leq t < s, \delta > 0$ such that $\Gamma_\delta(s, t) \neq E$ and $\Gamma_\delta(s, t)$ is connected

$$\|\tilde{\Phi}(s, t)\|^2 \leq 1 - c_N \delta^{2(s-t)}, \quad (19)$$

where $c_N = 2(1 - \cos \frac{\pi}{N})$ is the Fiedler value of the path graph on N vertices, i.e., the minimum of $\lambda_F(G) > 0$ over all connected graphs on N vertices [38].

We next prove Lemma 11.

Proof: We first prove Lemma 11 for all ω, δ, s, t such that $R_{s,\delta}(\omega) \leq t$. To this end, fix $\omega \in \Omega, \delta > 0$ and consider a fixed t, s $1 \leq t < s$, for which $R_{s,\delta}(\omega) \leq t$. Similarly to the proof of Lemma 1 a), b) in [39], it can be shown here that: 1) $[\Phi(s, t)(\omega)]_{ii} \geq \delta^{s-t}$, for all i ; and 2) $[\Phi(s, t)(\omega)]_{ij} \geq \delta^{s-t}$, for all $\{i, j\} \in E$, where we let E denote the set of edges of the graph $\Gamma_\delta(s, t)(\omega)$.

Notice that $\|\tilde{\Phi}(s, t)\|^2$ is the second largest eigenvalue of $\Phi(s, t)^\top \Phi(s, t)$, and, thus can be computed as:

$$\|\tilde{\Phi}(s, t)\|^2 = \max_{q^\top q = 1, q \perp 1} q^\top \Phi(s, t)^\top \Phi(s, t) q$$

Since $\Phi(s, t)^\top \Phi(s, t)$ is a symmetric stochastic matrix, it can be shown, e.g., [9], that its quadratic form, for a fixed vector $q \in \mathbb{R}^N$, can be written as:

$$q^\top \Phi(s, t)^\top \Phi(s, t) q = q^\top q - \sum_{\{i, j\}} [\Phi(s, t)^\top \Phi(s, t)]_{ij} (q_i - q_j)^2 \quad (20)$$

Now, combining the two auxiliary inequalities from the beginning of the proof, we get that, for all $\{i, j\} \in E$, $[\Phi(s, t)^\top \Phi(s, t)]_{ij} \geq \delta^{2(s-t)}$, where, we recall, E is the set of edges of $\Gamma_\delta(s, t)(\omega)$. Further, since all the entries of $\Phi(s, t)$ are non-negative (for every t , every realization of W_t is a stochastic matrix, and thus has non-negative entries), we can upper bound the sum in (20) over all $\{i, j\}$ by the sum over $\{i, j\} \in E$ only, yielding:

$$q^\top \Phi(s, t)^\top \Phi(s, t) q \leq q^\top q - \delta^{2(s-t)} \sum_{\{i, j\} \in E} (q_i - q_j)^2. \quad (21)$$

Finally, $\min_{q^\top q = 1, q \perp 1} \sum_{\{i, j\} \in E} (q_i - q_j)^2$ is equal to the Fiedler value (i.e., the second smallest eigenvalue of the Laplacian) of the graph $\Gamma_\delta(s, t)(\omega)$. This completes the proof of Lemma 11 for the case when ω, δ, s, t are such that $R_{s,\delta} \leq t$. The claim of Lemma 11 for the case when ω, δ, s, t are such that $t < R_{s,\delta}(\omega) < s$ essentially follows from the submultiplicativity of the spectral norm, the result of Lemma 11 for the case that we just proved (with $t' = R_{s,\delta}(\omega)$), and the fact that $\Gamma_\delta(s, R_{s,\delta}(\omega))(\omega) = \Gamma_\delta(s, t)(\omega)$. ■

Lemma 11 and Corollary 12 say that, for each fixed $\delta > 0$, whenever there is an interval in which the δ -accumulation graph is connected, then the norm of the error matrix on this interval improves by some finite amount (dependent on the interval size). We next introduce, for each $\delta > 0$, the sequence of δ -stopping times that registers these times at which we are certain that the error matrix makes an improvement.

Family of the sequences of δ -stopping times. For each $\delta > 0$, we define the sequence of δ -stopping times $T_{i,\delta} : \Omega \mapsto \mathbb{N} \cup \{+\infty\}$, $i = 1, 2, \dots$ by:

$$T_{i,\delta}(\omega) = \min\{t \geq T_{i-1,\delta}(\omega) + 1 : \Gamma_\delta(t, T_{i-1}(\omega)) \text{ is connected}\},$$

$$\text{for } i \geq 1, T_{0,\delta} \equiv 0.$$

By its construction, the sequence $\{T_{i,\delta}\}_{i \geq 1}$ defines the times that mark the right end point of “clear” intervals, without realization of δ -induced graphs equal to E , on which Γ_δ is connected. Using the result of Lemma 11, we have that at times $T_{i,\delta}$ the norm of $\tilde{\Phi}$ drops below 1 and the averaging process makes an improvement. Let further, for each $\delta > 0$ and $k \geq 1$, $M_{k,\delta} : \Omega \mapsto \mathbb{N} \cup \{0\}$ count the number of improvements with respect to the δ -stopping times until time k :

$$M_{k,\delta}(\omega) = \max\{i \geq 0 : T_{i,\delta}(\omega) \leq k\}. \quad (22)$$

We now explain how, at any given time k , we can use the knowledge of $M_{k,\delta}$ to bound the norm of the “error” matrix $\tilde{\Phi}(k, 0)$. Suppose that $M_{k,\delta} = m$. If we knew the locations of all the improvements until time k , $T_{i,\delta} = t_i, i = 1, \dots, m$ then, using Lemma 11, we could bound the norm of $\tilde{\Phi}(k, 0)$. Intuitively, since for fixed k and fixed m the number of allocations of $T_{i,\delta}$ ’s is finite, there will exist the one which yields the worst bound on $\|\tilde{\Phi}(k, 0)\|$. It turns out that the worst case allocation is the one with equidistant improvements, thus allowing for deriving a bound on $\|\tilde{\Phi}(k, 0)\|$ only in terms of $M_{k,\delta}$. This result is given in Lemma 13.

Lemma 13: For any fixed $\omega \in \Omega, \delta > 0$ and $k \geq 1$:

$$\|\tilde{\Phi}(k, 0)\| \leq \left(1 - c_N \delta^{\frac{2}{M_{k,\delta}}}\right)^{\frac{M_{k,\delta}}{2}}. \quad (23)$$

Proof: Fix $\omega \in \Omega, \delta > 0, k \geq 1$. If $M_{k,\delta}(\omega) = 0$, then the claim holds trivially. Thus, suppose $M_{k,\delta}(\omega) = m \geq 1$, and, suppose further $T_{1,\delta}(\omega) = t_1, T_{2,\delta}(\omega) = t_2, \dots, T_{m,\delta}(\omega) = t_m \leq k$ ($T_{i,\delta}(\omega) > k$, for $i > m$, because $M_{k,\delta}(\omega) = m$). By the construction of the δ -stopping times, we know that $\Gamma_\delta(t_i, t_{i-1})(\omega)$ is connected for all $i = 1, \dots, m$. Thus, we apply Lemma 11 on the intervals from t_{i-1} until t_i , for $i = 1, \dots, m$, to get $\|\tilde{\Phi}(t_i, t_{i-1})\| \leq (1 - c_N \delta^{2(t_i - t_{i-1})})^{\frac{1}{2}}$.

Combining this with submultiplicativity of the spectral norm, yields:

$$\begin{aligned} & \|\tilde{\Phi}(k, 0)(\omega)\| \\ &= \|\tilde{\Phi}(k, t_m)(\omega)\tilde{\Phi}(t_m, t_{m-1})(\omega) \cdots \tilde{\Phi}(t_1, 0)(\omega)\| \\ &\leq \|\tilde{\Phi}(k, t_m)(\omega)\| \|\tilde{\Phi}(t_m, t_{m-1})(\omega)\| \cdots \|\tilde{\Phi}(t_1, 0)(\omega)\| \\ &\leq \prod_{i=1}^m \left(1 - c_N \delta^{2(t_i - t_{i-1})}\right)^{\frac{1}{2}}. \end{aligned} \quad (24)$$

Denote $\Delta_i = t_i - t_{i-1}$ and note that $\sum_{i=1}^m \Delta_i \leq k$. Further, remark that $f(\Delta) = \log(1 - c_N \delta^{2\Delta})$ is a concave function. Taking the log in (24) and applying Jensen's inequality [40] for equal convex multipliers $\alpha_i = \frac{1}{m}$, $i = 1, \dots, m$, yields

$$\begin{aligned} \sum_{i=1}^m \alpha_i \log(1 - c_N \delta^{2\Delta_i}) &\leq \log\left(1 - c_N \delta^{2(\sum_{i=1}^m \alpha_i \Delta_i)}\right) \\ &= \log\left(1 - c_N \delta^{\frac{2}{m} \sum_{i=1}^m \Delta_i}\right). \end{aligned}$$

Finally, since f is increasing and $\sum_{i=1}^m \Delta_i \leq k$, $\sum_{i=1}^m \frac{1}{m} \log(1 - c_N \delta^{2\Delta_i}) \leq \log(1 - c_N \delta^{\frac{2k}{m}})$. Multiplying both sides of the last inequality with $\frac{m}{2}$, and computing the exponent yields (23). ■

Lemma 13 provides a bound on the norm of the “error” matrix $\tilde{\Phi}(k, 0)$ in terms of the number of improvements $M_{k,\delta}$ up to time k . Intuitively, if $M_{k,\delta}$ is high enough relative to k , then the norm of $\tilde{\Phi}(k, 0)$ decays exponentially fast (to see this, just take $M_{k,\delta} = k$ in (23)) and, thus, it cannot stay above d_k , which decays sub-exponentially as k increases. We show that this is indeed true for all $\omega \in \Omega$ for which $M_{k,\delta} = \alpha k$ or higher, for any choice of $\alpha \in (0, 1]$; this result is stated in Lemma 14, part 1. On the other hand, if the number of improvements is less than αk , then there must have been long intervals on which Γ_δ was disconnected. The probability that such an interval of length t occurs is essentially determined by the probability that the sequence of δ -induced graphs is “trapped” in some disconnected collection for time $t - 1$, and it equals $p_{\max,\delta}^{t-1}$. As the number of these intervals until time k is at most αk , this yields, in a crude approximation, the probability of $p_{\max,\delta}^{k-\alpha k}$ for the event $M_{k,\delta} \leq \alpha k$; this intuition is formalized in part 2 of Lemma 14.

Lemma 14: For any fixed $\delta \in (0, 1)$, $\alpha \in (0, 1]$:

1) there exists sufficiently large $k_0 = k_0(\delta, \alpha, \{d_k\})$ such that

$$\mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k, M_{k,\delta} \geq \alpha k) = 0, \quad \forall k \geq k_0(\delta, \alpha, \{d_k\}); \quad (25)$$

2) for every $\lambda \in (0, \bar{I}_\delta)$

$$\mathbb{P}(M_{k,\delta} < \alpha k) \leq \exp(-\lambda(k - \lceil \alpha k \rceil))(1 - a_\delta(\lambda))^{-\lceil \alpha k \rceil}, \quad (26)$$

where $a_\delta(\lambda) = \exp(\lambda - \bar{I}_\delta) < 1$ and \bar{I}_δ is defined as $\bar{I}_\delta = |\log(p_{\max,\delta} + |\Pi(\mathcal{G}_\delta)|p_{E,\delta})|$, for δ such that $\Pi(\mathcal{G}_\delta) \neq \emptyset$, and $\bar{I}_\delta = |\log p_{E,\delta}|$, otherwise.

Proof: Fix $\delta \in (0, 1)$, $\alpha \in (0, 1]$. To prove 1, we first note that by Lemma 13 we have:

$$\{\|\tilde{\Phi}(k, 0)\| \geq d_k\} \subseteq \left\{\left(1 - c_N \delta^{2\frac{k}{M_{k,\delta}}}\right)^{\frac{M_{k,\delta}}{2}} \geq d_k\right\}. \quad (27)$$

This gives

$$\begin{aligned} & \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k, M_{k,\delta} \geq \alpha k) \\ &\leq \mathbb{P}\left(\left(1 - c_N \delta^{2\frac{k}{M_{k,\delta}}}\right)^{\frac{M_{k,\delta}}{2}} \geq d_k, M_{k,\delta} \geq \alpha k\right) \\ &= \sum_{m=\lceil \alpha k \rceil}^k \mathbb{P}\left(\left(1 - c_N \delta^{2\frac{k}{m}}\right)^{\frac{m}{2}} \geq d_k, M_{k,\delta} = m\right) \\ &= \sum_{m=\lceil \alpha k \rceil}^k \mathbb{P}\left(g(k, M_{k,\delta}) \geq \frac{\log d_k}{k}, M_{k,\delta} = m\right), \end{aligned} \quad (28)$$

where $g(k, m) := \frac{m}{2k} \log(1 - c_N \delta^{2\frac{k}{m}})$, for $m > 0$. For fixed k , each of the probabilities in the sum above is equal to 0 for those m such that $g(k, m) < \frac{\log d_k}{k}$. This yields:

$$\begin{aligned} & \sum_{m=\lceil \alpha k \rceil}^k \mathbb{P}\left(g(k, M_{k,\delta}) \geq \frac{\log d_k}{k}, M_{k,\delta} = m\right) \\ &\leq \sum_{m=\lceil \alpha k \rceil}^k s(k, m), \end{aligned} \quad (29)$$

where $s(k, m)$ is the switch function defined by:

$$s(k, m) := \begin{cases} 0, & \text{if } g(k, m) < \frac{\log d_k}{k} \\ 1, & \text{otherwise} \end{cases}$$

Also, as $g(k, \cdot)$ is, for fixed k , decreasing in m , it follows that $s(k, m) \leq s(k, \alpha k)$ for $m \geq \alpha k$. Combining this with (28) and (29), we get:

$$\mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k, M_{k,\delta} \geq \alpha k) \leq (k - \lceil \alpha k \rceil + 1)s(k, \alpha k).$$

We now show that $s(k, \alpha k)$ will eventually become 0, as k increases, which would yield part 1 of Lemma 14. To show this, we observe that g has a constant negative value at $(k, \alpha k)$:

$$g(k, \alpha k) = \frac{\alpha}{2} \log\left(1 - c_N \delta^{\frac{2}{\alpha}}\right).$$

Since $\frac{\log d_k}{k} \rightarrow 0$, as $k \rightarrow \infty$, there exists $k_0 = k_0(\delta, \alpha, \{d_k\})$ such that $g(k, \alpha k) < \frac{\log d_k}{k}$, for every $k \geq k_0$. Thus, $s(k, \alpha k) = 0$ for every $k \geq k_0$. This completes the proof of part 1.

To prove part 2, we first prove the following result which is the main argument in the proof of part 2.

Lemma 15: For any $\delta > 0$, $t \geq 1$

$$\mathbb{P}(T_{1,\delta} > t) \leq e^{-\bar{I}_\delta t}. \quad (30)$$

Proof: Fix $\delta > 0$, $t \geq 1$. For the case when $\Pi(\mathcal{G}_\delta) = \emptyset$, the claim easily follows by noting that $\mathbb{P}(T_{1,\delta} > t) = \mathbb{P}(G_{r,\delta} = E, 1 \leq r \leq t)$. (The latter is true because each realization of $G_{r,\delta}$, which has a positive probability of occurrence is either a connected graph or equal

to E .) Suppose now that δ is such that $\Pi(\mathcal{G}_\delta) \neq \emptyset$. Define $S_l, l \geq 1$ to be the (random) locations of the realization E in the sequence of δ -induced graphs and let also Q_t be the number of such realizations until time t ; for convenience, let also $S_0 \equiv 0$. By definition of Γ_δ , the event $\{T_{1,\delta} > t\}$ is equivalent to the event that Γ_δ is disconnected on each block in the sequence of $G_{r,\delta}, 1 \leq r \leq t$ that is clear from realizations of E . Partitioning this event over all possible number of realizations of E on the interval from time 1 until time t , Q_t , and, also, over all possible locations of E , S_l , we get

$$\begin{aligned} \mathbb{P}(T_{1,\delta} > t) &= \sum_{L=1}^t \sum_{1 \leq s_1 < \dots < s_L \leq t} \mathbb{P}(Q_t = L, S_l = s_l, \\ &\quad \Gamma_\delta(s_l - 1, s_{l-1}) \text{ is disc.}, l = 1, \dots, L, \Gamma_\delta(t, s_L) \text{ is disc.}) \\ &= \sum_{L=0}^t p_{E,\delta}^L \sum_{1 \leq s_1 < \dots < s_L \leq t} \mathbb{P}(\Gamma_\delta(t, s_L) \text{ is disc.}) \\ &\quad \times \prod_{l=1}^L \mathbb{P}(\Gamma_\delta(s_l - 1, s_{l-1}) \text{ is disc.}), \end{aligned} \quad (31)$$

where the last equality follows from the fact that realizations of $G_{r,\delta}$ belonging to disjoint blocks are independent, and, also, we implicitly assume that the statement $\Gamma_\delta(s_l - 1, s_{l-1})$ is disc. implies that $G_{r,\delta} \neq E$, for $s_{l-1} < r < s_l$. We now fix l, s_l, s_{l-1} and focus on computing $\mathbb{P}(\Gamma_\delta(s_l - 1, s_{l-1}) \text{ is disc.})$. To this end, let $\Omega_\delta^* = \cap_{t \geq 1} \{G_{t,\delta} \in \{\mathcal{G}_\delta \cup \{E\}\}\}$ and note that, since each of the events in the intersection has probability 1, the event Ω_δ^* also has probability 1. We show that $\{\Gamma_\delta(s_l - 1, s_{l-1}) \text{ is disc.}\} \cap \Omega_\delta^* \subseteq \cup_{\mathcal{H} \in \Pi(\mathcal{G}_\delta)} \{G_{r,\delta} \in \mathcal{H}, s_{l-1} < r < s_l\}$, or, in words, if Γ_δ is disconnected on some interval and all the graph realizations that occurred during this interval belong to \mathcal{G}_δ , then there must exist a disconnected collection on \mathcal{G}_δ to which all the graph realizations belong to; the last claim, since $\mathbb{P}(\Omega_\delta^*) = 1$, would yield

$$\begin{aligned} \mathbb{P}(\Gamma_\delta(s_l - 1, s_{l-1}) \text{ is disc.}) &\leq \sum_{\mathcal{H} \in \Pi(\mathcal{G}_\delta)} p_{\mathcal{H},\delta}^{s_l - s_{l-1} - 1} \\ &\leq |\Pi(\mathcal{G}_\delta)| p_{\max,\delta}^{s_l - s_{l-1} - 1}. \end{aligned} \quad (32)$$

To prove the claim above, consider fixed $\omega \in \Omega_\delta^*$ such that $\Gamma_\delta(s_l - 1, s_{l-1})$ is disconnected, and let $\mathcal{H}_l = \{G_{s_{l-1}+1,\delta}(\omega), \dots, G_{s_l-1,\delta}(\omega)\}$. Since $\omega \in \Omega_\delta^*$, and we assume that $G_{r,\delta}(\omega) \neq E$, then it must be that $G_{r,\delta}(\omega) \in \mathcal{G}_\delta$, for all $s_{l-1} < r < s_l$. On the other hand, since $\Gamma_\delta(s_l - 1, s_{l-1})(\omega) = \Gamma(\mathcal{H}_l)$ is disconnected, it follows that \mathcal{H}_l is a disconnected collection on \mathcal{G}_δ , thus proving the claim. Combining now (31) and (32) yields (30)

$$\begin{aligned} \mathbb{P}(T_{1,\delta} > t) &\leq \sum_{L=0}^t p_{E,\delta}^L \sum_{1 \leq s_1 < \dots < s_L \leq t} |\Pi(\mathcal{G}_\delta)| p_{\max,\delta}^{t-s_L} \prod_{l=1}^L |\Pi(\mathcal{G}_\delta)| p_{\max,\delta}^{s_l - s_{l-1} - 1} \\ &= \sum_{L=0}^t \binom{t}{L} p_{E,\delta}^L |\Pi(\mathcal{G}_\delta)|^{L+1} p_{\max,\delta}^{t-L} \\ &= (p_{\max,\delta} + |\Pi(\mathcal{G}_\delta)| p_{E,\delta})^t. \end{aligned}$$

Now, notice that we can express the event that $M_{k,\delta} < \alpha k$ through increments of δ -stopping times: $\{M_{k,\delta} < \alpha k\} = \{T_{\lceil \alpha k \rceil, \delta} > k\} = \{\sum_{i=1}^{\lceil \alpha k \rceil} T_{i,\delta} - T_{i-1,\delta} > k\}$. Applying the exponential Markov inequality [41] with parameter $\lambda > 0$

$$\begin{aligned} \mathbb{P}(M_{k,\delta} < \alpha k) &\leq \exp(-\lambda k) \mathbb{E} \left[\exp \left(\sum_{i=1}^{\lceil \alpha k \rceil} \lambda (T_{i,\delta} - T_{i-1,\delta}) \right) \right] \\ &= \exp(-\lambda k) (\mathbb{E}[\exp(\lambda T_{1,\delta})])^{\lceil \alpha k \rceil}, \end{aligned} \quad (33)$$

where the equality follows from the fact that the increments of δ -stopping times are i.i.d. We now focus on computing the expectation in the equation above. Using the result of Lemma 15

$$\begin{aligned} \mathbb{E}[\exp(\lambda T_{1,\delta})] &= \sum_{t=1}^{\infty} \exp(\lambda t) \mathbb{P}(T_{1,\delta} = t) \\ &\leq \sum_{t=1}^{\infty} \exp(\lambda t) \mathbb{P}(T_{1,\delta} > t-1) \\ &\leq \exp(\lambda) \sum_{t=1}^{\infty} \exp(\lambda(t-1)) \\ &\quad \times (p_{\max,\delta} + |\Pi(\mathcal{G}_\delta)| p_{E,\delta})^{t-1}. \end{aligned} \quad (34)$$

The sum in the previous equation converges for all $\lambda < \bar{I}_\delta$ to $1/(1 - a_\delta(\lambda))$. Combining this with (33) completes the proof of part 2. ■

From parts 1 and 2 of Lemma 14 it follows that for any fixed $\alpha \in (0, 1]$, $\delta \in (0, 1)$ and $\lambda \in (0, \bar{I}_\delta)$:

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) &\leq -\lambda(1 - \alpha) - \alpha \log(1 - a_\delta(\lambda)). \end{aligned} \quad (35)$$

Now, taking first the infimum over α and then the infimum over λ yields:

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{P}(\|\tilde{\Phi}(k, 0)\| \geq d_k) &\leq \inf_{\lambda \in (0, \bar{I}_\delta)} \inf_{\alpha \in (0, 1]} -\lambda(1 - \alpha) - \alpha \log(1 - a_\delta(\lambda)) \\ &= \inf_{\lambda \in (0, \bar{I}_\delta)} -\lambda = -\bar{I}_\delta. \end{aligned} \quad (36)$$

Finally, if $\Pi(\mathcal{G}) \neq \emptyset$, then, by Lemma 9 and Corollary 10, \bar{I}_δ converges to $|\log p_{\max}|$, as $\delta \rightarrow 0$. On the other hand, if $\Pi(\mathcal{G}) = \emptyset$, it can be easily shown that \bar{I}_δ goes to $+\infty$, as $\delta \rightarrow 0$. Taking the limit $\delta \rightarrow 0$ in eq. (36) yields the upper bound (10).

IV. COMPUTATION OF I VIA MIN-CUT: GOSSIP AND LINK FAILURE MODELS

Motivated by the applications of averaging in sensor networks and distributed dynamical systems, we consider two frequently used types of random averaging models: gossip and link failure models. For a generic graph $G = (V, E)$, we show that p_{\max} for both models can be found by solving an instance of a min-cut problem over the same graph G . The corresponding link costs are simple functions of the link occurrence probabilities. In this section, we detail the relation between the minimum

cut (min-cut) problem and computation of p_{\max} . We first explain the min-cut problem.

Minimum cut (min-cut) problem. Let $G = (V, E, C)$ be given undirected weighted graph where V is the set of N nodes, E is the set of edges, and $C = [c_{ij}]$ is the $N \times N$ matrix of the edge nonnegative costs; by convention, we set $c_{ii} = 0$, for all i , and $c_{ij} = 0$, for $\{i, j\} \notin E$. The min-cut problem is to find the subset of edges F such that $(V, E \setminus F)$ is disconnected and the sum $\sum_{\{i,j\} \in F} c_{ij}$ is minimal; we denote this minimal value, also referred to as the connectivity, by $\text{mincut}(V, E, C)$. The min-cut problem is easy to solve, and there exist efficient algorithms to solve it, e.g., [42], [5].

A. Gossip Model

Consider the network of N nodes, collected in the set V and with the set $E \subseteq \binom{V}{2}$ defining communication links between the nodes, such that if $\{i, j\} \in E$ then nodes $i, j \in V$ can communicate. In the gossip algorithm, only one link $\{i, j\} \in E$ is active at a time. Let p_{ij} be the probability of occurrence of link $\{i, j\} \in E$: $p_{ij} = \mathbb{P}(G_t = (V, \{i, j\}))$. We note that $\sum_{\{i,j\} \in E} p_{ij} = 1$.

Lemma 16: Consider a gossip model on a graph $G = (V, E)$ with link probabilities $p_{ij}, \{i, j\} \in E$. Construct a mincut problem instance with the graph G and the cost assigned to link $\{i, j\}$ equal p_{ij} . Then:

$$p_{\max}^{\text{Gossip}}(V, E, P) = 1 - \text{mincut}(V, E, P) \quad (37)$$

$$I^{\text{Gossip}}(V, E, P) = -\log(1 - \text{mincut}(V, E, P)), \quad (38)$$

where P is the symmetric matrix that collects link occurrence probabilities, $P_{ij} = p_{ij}, \{i, j\} \in E, P_{ii} = 0$, for $i = 1, \dots, N$ and $P_{ij} = 0, \{i, j\} \notin E$.

Proof: For the gossip model, the set of all possible graph realizations $\mathcal{G}^{\text{Gossip}}$ is the set of all one-link subgraphs of (V, E) : $\mathcal{G}^{\text{Gossip}} = \{(V, \{i, j\}) : \{i, j\} \in E\}$. Also, there is a one to one correspondence between the set of collections of realizable graphs and the set of subgraphs of G : a collection $\mathcal{H} \subseteq \mathcal{G}$ corresponds to the subgraph H of G if and only if $H = \Gamma(\mathcal{H})$. Thus, if we assign to each link in G a cost equal to p_{ij} , then searching over the set $\Pi(\mathcal{G})$ of all disconnected collections to find the most likely one is equivalent to searching over all disconnected subgraphs of G with the maximal total cost:

$$p_{\max}^{\text{Gossip}} = \max_{\mathcal{H} \in \Pi(\mathcal{G})} p_{\mathcal{H}} = \max_{E' \subseteq E, (V, E') \text{ is disc.}} \sum_{\{i,j\} \in E'} p_{ij}. \quad (39)$$

Using the fact that $\sum_{\{i,j\} \in E} p_{ij} = 1$, (39) can be written as:

$$\begin{aligned} & \max_{E' \subseteq E, (V, E') \text{ is disc.}} \sum_{\{i,j\} \in E'} p_{ij} \\ &= \max_{F \subseteq E, (V, E \setminus F) \text{ is disc.}} 1 - \sum_{\{i,j\} \in F} p_{ij} \\ &= 1 - \min_{F \subseteq E, (V, E \setminus F) \text{ is disc.}} \sum_{\{i,j\} \in F} p_{ij}. \end{aligned}$$

The minimization problem in the last equation is the min-cut problem $\text{mincut}(V, E, P)$. ■

Gossip on a regular network. We now consider a special case of the uniform gossip model on a connected regular graph with degree $d, d = 2, \dots, N - 1$, and the uniform link occurrence probability $p := p_{ij} = \frac{2}{Nd}$. It can be easily seen that the value of the min-cut is p times the minimal number of edges that disconnects the graph, which equals $pd = 2/N$; this corresponds to cutting all the edges of a fixed node, i.e., isolating a fixed node. Hence,

$$\begin{aligned} p_{\max} &= \mathbb{P}(\text{node } i \text{ is isolated}) = 1 - 2/N, \\ I &= -\log(1 - 2/N). \end{aligned}$$

Note that the asymptotic rate I is determined by the probability that a fixed node is isolated; and the rate I does not depend on the degree d .

B. Link Failure Model

Similarly as with the gossip model, we introduce a graph $G = (V, E)$ to model the communication links between the nodes. In contrast with the gossip model, the link failure model assumes that each feasible link $\{i, j\} \in E$ occurs independently from all the others links in the network, as explained in detail in Example 2. Given the realization of the network topology, weight matrices W_k can, for example, be chosen according to the Metropolis rule, see Example 2. Let again p_{ij} denote the probability of occurrence of link $\{i, j\} \in E$. (Remark that, due to the independence assumption, we now do not have any condition on the link occurrence probabilities p_{ij} .)

Lemma 17: Consider a link failure model on a graph $G = (V, E)$ with link probabilities $p_{ij}, \{i, j\} \in E$. Construct a mincut problem instance with the graph G and the cost of link $\{i, j\}$ equal to $-\log(1 - p_{ij})$. Then:

$$p_{\max}^{\text{Link fail.}}(V, E, P) = e^{-\text{mincut}(V, E, -\log(1-P))} \quad (40)$$

$$I^{\text{Link fail.}}(V, E, P) = \text{mincut}(V, E, -\log(1 - P)), \quad (41)$$

where P is the symmetric matrix that collects the link occurrence probabilities, $P_{ij} = p_{ij}, \{i, j\} \in E, P_{ii} = 0$, for $i = 1, \dots, N$ and $P_{ij} = 0, \{i, j\} \notin E$ and $\log X$ denotes the entry wise logarithm of a matrix X .

Proof: Since the links occur independently, any subgraph $H = (V, E')$ of G can occur at a given time, therefore yielding that the collection of realizable graphs $\mathcal{G}^{\text{Link fail.}}$ is the collection of all subgraphs of G :

$$\mathcal{G}^{\text{Link fail.}} = \{(V, E') : E' \in 2^E\}; \quad (42)$$

here 2^E denotes the power set of E , i.e., the collection of all possible subsets of the set of feasible links E .

This implies that for any fixed set $F \subseteq E$ of edges that disconnect $G = (V, E)$ we can find a disconnected collection $\mathcal{H} \subseteq \mathcal{G}$ such that $\Gamma(\mathcal{H}) = (V, E \setminus F)$ (recall that $\Gamma(\mathcal{H})$ is the minimal supergraph of all the graphs contained in \mathcal{H}). On the other hand, any disconnected collection will map by Γ to one disconnected subgraph of G . Therefore, in order to find

$p_{\max}^{\text{Link fail.}}$ we can split the search over disconnected collections \mathcal{H} as follows:

$$\begin{aligned} p_{\max}^{\text{Link fail.}} &= \max_{\mathcal{H} \subseteq \mathcal{G} \text{ is disc.}} p_{\mathcal{H}} \\ &= \max_{F \subseteq E, F \text{ disconnects } (V, E)} \max_{\mathcal{H} \subseteq \mathcal{G} \text{ } \Gamma(\mathcal{H}) = (V, E \setminus F)} p_{\mathcal{H}}. \end{aligned} \quad (43)$$

Next, we fix a disconnecting set of edges $F \subseteq E$ and consider all $\mathcal{H} \subseteq \mathcal{G}$ such that $\Gamma(\mathcal{H}) = (V, E \setminus F)$. We claim that, among all such collections, the one with maximal probability is $\mathcal{H}_F := \{(V, E') : E' \subseteq E \setminus F\}$. To show this, we observe that if $H = (V, E') \in \mathcal{H}$, then $E' \cap F = \emptyset$, thus implying:

$$\begin{aligned} p_{\mathcal{H}} &= \sum_{H \in \mathcal{H}} P(G_t = H) \\ &\leq \sum_{H=(V, E') : E' \subseteq E \setminus F} P(G_t = H) = p_{\mathcal{H}_F}. \end{aligned}$$

Therefore, the expression (43) simplifies to:

$$\max_{F \subseteq E, F \text{ disconnects } G=(V, E)} p_{\mathcal{H}_F}.$$

We next compute $p_{\mathcal{H}_F}$ for given $F \subseteq E$:

$$\begin{aligned} p_{\mathcal{H}_F} &= \mathbb{P}(E(G_t) \cap F = \emptyset) \\ &= \mathbb{P}(\{i, j\} \notin E(G_t), \text{ for all } \{i, j\} \in F) \\ &= \prod_{\{i, j\} \in F} (1 - p_{ij}), \end{aligned}$$

where the last equality follows by the independence assumption on the link occurrence probabilities. This implies that $p_{\max}^{\text{Link fail.}}$ can be computed by

$$\begin{aligned} p_{\max}^{\text{Link fail.}} &= \max_{F \subseteq E, F \text{ disconnects } G=(V, E)} \prod_{\{i, j\} \in F} (1 - p_{ij}) \\ &= e^{-\min_{F \subseteq E, F \text{ disconnects } (V, E)} \text{mincut}(V, E, -\log(1-P))}. \end{aligned}$$

Regular graph and uniform link failures. We now consider the special case when the underlying graph is a connected regular graph with degree $d, d = 2, \dots, N-1$, and the uniform link occurrence probabilities $p_{ij} = p$. It is easy to see that p_{\max} and I simplify to:

$$\begin{aligned} p_{\max} &= \mathbb{P}(\text{node } i \text{ is isolated}) = (1 - p)^d, \\ I &= -d \log(1 - p). \end{aligned}$$

V. APPLICATION: POWER ALLOCATION FOR DISTRIBUTED DETECTION

We now demonstrate the usefulness of our Theorem 6 by applying it to consensus+innovations distributed detection, [23], [22], over networks with symmetric fading links. We first show that the asymptotic performance (error exponent of the detection error probability) of distributed detection explicitly depends on the large deviation rate $|\log p_{\max}|$. Further, we note

that $|\log p_{\max}|$ is a function of the link fading (failure) probabilities, and, consequently, of the sensors' transmission power. We exploit this fact to formulate the optimization problem of minimizing the transmission power subject to a lower bound on the guaranteed detection performance; the latter translates into the requirement that $|\log p_{\max}|$ exceeds a threshold. We show that the corresponding optimization problem is convex. Finally, we illustrate by simulation the gains achieved by optimizing the transmission power.

A. Consensus+Innovations Distributed Detection

Detection problem. We now briefly explain the distributed detection problem that we consider. We consider a network of N sensors that cooperate to detect an event of interest, i.e., face a binary hypothesis test H_1 versus H_0 . Each sensor i , at each time step $t, t = 1, 2, \dots$, performs a measurement $Y_{i,t}$. We assume that the measurements are i.i.d., both in time and across sensors, where under hypothesis $H_l, Y_{i,t}$ has the density function $f_{l,l} = 0, 1$, for $i = 1, \dots, N$ and $t = 1, 2, \dots$.

Consensus+innovations distributed detector. To resolve between the two hypothesis, each sensor i maintains over time k its local decision variable $x_{i,k}$ and compares it with a threshold; if $x_{i,k} > 0$, sensor i accepts H_1 ; otherwise, it accepts H_0 . Sensor i updates its decision variable $x_{i,k}$ by exchanging the decision variable locally with its neighbors, by computing the weighted average of its own and the neighbors' variables, and by incorporating its new measurement through a log-likelihood ratio $L_{i,k} = \log \frac{f_1(Y_{i,k})}{f_0(Y_{i,k})}$:

$$\begin{aligned} x_{i,k} &= \sum_{j \in O_{i,k}} W_{ij,k} \left(\frac{k-1}{k} x_{j,k-1} + \frac{1}{k} L_{j,k} \right), \\ k &= 1, 2, \dots, x_{i,0} = 0. \end{aligned} \quad (44)$$

Here $O_{i,k}$ is the (random) neighborhood of sensor i at time k (including i), and $W_{ij,k}$ is the (random) averaging weight that sensor i assigns to sensor j at time k .

Let $x_k = (x_{1,k}, x_{2,k}, \dots, x_{N,k})^\top$ and $L_k = (L_{1,k}, \dots, L_{N,k})^\top$. Also, collect the averaging weights $W_{ij,k}$ in the $N \times N$ matrix W_k , where, clearly, $W_{ij,k} = 0$ if the sensors i and j do not communicate at time step k . For the weight matrices W_k we assume the Metropolis weight model explained in Example 2. Then, using the definition of $\Phi(k, t)$, writing (44) in matrix form, and unwinding the recursion, we get:

$$x_k = \frac{1}{k} \sum_{t=1}^k \Phi(k, t-1) L_t, \quad k = 1, 2, \dots \quad (45)$$

Equation (45) shows the significance of the matrices $\Phi(k, t)$ to the distributed detection performance, and, in particular, on the significance of how much the matrices $\Phi(k, t)$ are close to J . Indeed, when $\Phi(k, t) = J$, the contribution of L_t to $x_{i,k}$ is $[\Phi(k, t) L_t]_i = \frac{1}{N} \sum_{i=1}^N L_{i,t}$, and hence sensor i effectively uses the local likelihood ratios of all the sensors. In the other extreme, when $\Phi(k, t) = I_N$, $[\Phi(k, t) L_t]_i = L_{i,t}$, and hence sensor i effectively uses only its own likelihood ratio. In fact, it

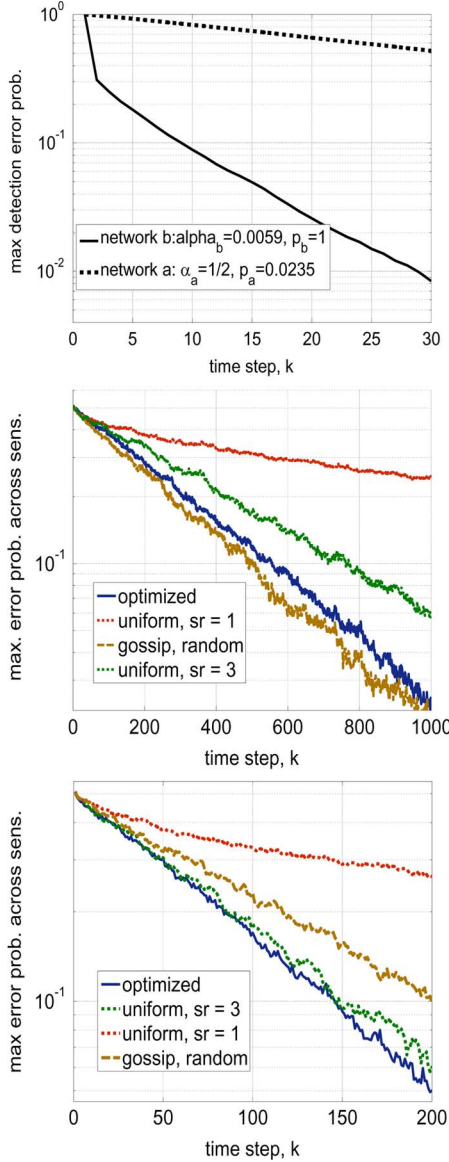


Fig. 2. **Top:** Detection error probability for the worst sensor versus time step k for the 2 node network. **Center and Bottom:** Detection error probability of the worst sensor versus time step k for the optimized power allocation, the uniform power allocation with $sr = 1, 3$, ($sr = \frac{\text{total power per k for uniform allocation}}{\text{total power per k for optimal allocation}}$) and for the random, gossip allocation; **Center:** lower SNR; **Bottom:** higher SNR.

can be shown that, when the rate I exceeds a certain threshold, then the error exponent of the detection error probability at each sensor i is optimal, i.e., equal to the error exponent of the best centralized detector. Specifically, the optimality threshold on I depends on the sensor observations distributions f_1 and f_0 and is given by³:

$$I \geq I^*(f_1, f_0, N). \quad (46)$$

Furthermore, when $I < I^*(f_1, f_0, N)$, the detection error exponent is a highly nonlinear function of I . (See Fig. 2, right in [23].)

Remark: We now provide a simple example to demonstrate that the large deviations rate I is a more relevant metric than

the existing aggregate metrics, like $\lambda_2(\mathbb{E}[W_k^2])$ and $\|\mathbb{E}[W_k - J]\|$, when studying the *consensus+innovations* detector's performance. Consider a $N = 2$ -node network with the underlying 2×2 i.i.d. symmetric, stochastic matrices W_k , with $[W_k]_{12} = \alpha$, with probability p , and $[W_k]_{12} = 0$, with probability $1 - p$, and $\alpha \in (0, 1)$ is constant. Sensor $i, i = 1, 2$, has Gaussian measurements under hypothesis l : $f_l \sim \mathcal{N}(m_{i,l}, \sigma^2)$, with $\sigma^2 = 1$, $m_{i,1} = 1$, and $m_{i,0} = 0$, else. To show that $\lambda_2(\mathbb{E}[W_k^2])$ does not capture sufficient information to describe detection performance, we make two different models of W_k with equal $\mathbb{E}[W_k^2]$, and hence, with equal $\lambda_2(\mathbb{E}[W_k^2])$. With the first model a , we set $p_a = 0.2(1 - e^{-(1/8)m_{11}^2}) = 0.0235$, a small value so that distributed detector does not have an optimal error exponent, and $\alpha_a = 1/2$. With the second model b , we set $p_b = 1$, and $\alpha_b = 0.5(1 - (1 - p_a)^{1/2})$, so that $\mathbb{E}[W_k^2]$ is equal for both models a and b . Now, using $\lambda_2(\mathbb{E}[W_k^2])$, we would infer that, with both networks a and b , distributed detector achieves the same performance. On the other hand, calculating the large deviations rate $I_a = -\log(1 - p) = 0.02378$ and $I_b = +\infty$, we expect a very different detection performance. Fig. 2, top plots the maximal probability of error (across nodes 1 and 2) versus time step k , with models a and b , estimated via 20,000 Monte Carlo runs. We can see, as predicted by the large deviations rate I , that the performance of models a and b is very different. Moreover, $\lambda_2(\mathbb{E}[W_k^2])$ fails to shed light on detection performance. We note that we can show similarly that $\|\mathbb{E}[W_k] - J\|$ also fails to measure the consensus+innovations distributed detector performance.

B. Power Allocation

Equation (46) says that there is a sufficient large deviation rate I^* such that the distributed detector is asymptotically optimal; a further increase of I above I^* does not improve the exponential decay rate of the error probability. Also, as we have shown in Subsection IV.B, the large deviation rate I is a function of the link occurrence probabilities, which are further dependent on the sensors' transmission power. In summary, (46) suggests that there is a sufficient (minimal required) transmission power that achieves detection with the optimal exponential decay rate. This discussion motivates us to formulate the optimal power allocation problem of minimizing the total transmission power per time k subject to the optimality condition $I \geq I^*$. Before presenting the optimization problem, we detail the inter-sensor communication model.

Inter-sensor communication model. We adopt a symmetric Rayleigh fading channel model, a model similar to the one proposed in [43] ([43] assumes asymmetric channels). At time k , sensor j receives from sensor i : $y_{ij,k} = g_{ij,k} \sqrt{\frac{S_{ij}}{d_{ij}^\alpha}} x_{i,k} + n_{ij,k}$, 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 the channels (i, j) and (j, i) at time k experience the same fade, i.e., $g_{ij,k} = g_{ji,k}$; $g_{ij,k}$ is i.i.d. in time; and $g_{ij,t}$ and $g_{lm,s}$ are mutually independent for all t, s . We adopt the following link failure model. Sensor j successfully decodes the message from sensor i (i.e., the link (i, j) is online) if the signal

³See [23] for the precise expression of the threshold.

to noise ratio exceeds a threshold, i.e., if: $\text{SNR} = \frac{S_{ij} g_{ij,k}^2}{\sigma_n^2 d_{ij}^\alpha} > \tau$, or, equivalently, if $g_{ij,k}^2 > \frac{\tau \sigma_n^2 d_{ij}^\alpha}{S_{ij}} := \frac{K_{ij}}{S_{ij}}$. The quantity $g_{ij,k}^2$ is, for the Rayleigh fading channel, exponentially distributed with parameter 1. Hence, we arrive at the expression for the probability of the link (i, j) being online:

$$P_{ij} = \mathbb{P} \left(g_{ij,k}^2 > \frac{K_{ij}}{S_{ij}} \right) = e^{-\frac{K_{ij}}{S_{ij}}}. \quad (47)$$

We constrain the choice of the transmission powers by $S_{ij} = S_{ji}$ ⁴, so that the link (i, j) is online if and only if the link (j, i) is online, i.e., the graph realizations are undirected graphs. Hence, the underlying communication model is the link failure model, with the link occurrence probabilities P_{ij} in (47) that are dependent on the transmission powers S_{ij} .

With this model, the large deviation rate I is given by (40), where the weight c_{ij} associated with link (i, j) is: $c_{ij}(S_{ij}) = -\log(1 - e^{-K_{ij}/S_{ij}})$. We denote by $\{S_{ij}\}$ the set of all powers S_{ij} , $\{i, j\} \in E$.

Lemma 18: The function $I(\{S_{ij}\}) = \text{mincut}(V, E, C)$, with $c_{ij} = -\log(1 - e^{-K_{ij}/S_{ij}})$, for $\{i, j\} \in E$, and $c_{ij} = 0$ else, is concave.

Proof: Note that the function $I(\{S_{ij}\}) = \text{mincut}(V, E, C)$ can be expressed as

$$\min_{E' \subset E: G' = (V, E') \text{ is disconnected}} \sum_{\{i, j\} \in E \setminus E'} c_{ij}(S_{ij}).$$

On the other hand, $c_{ij}(S_{ij})$ is concave in S_{ij} for $S_{ij} \geq 0$, which can be shown by computing the second derivative and noting that it is non-positive. Hence, $I(\{S_{ij}\})$ is a pointwise minimum of concave functions, and thus it is concave. ■

Power allocation problem formulation. We now formulate the power allocation problem as the problem of minimizing the total transmission power used at time k , $2 \sum_{\{i, j\} \in E} S_{ij}$, so that the distributed detector achieves asymptotic optimality. This translates into the following optimization problem:

$$\begin{aligned} & \text{minimize} && \sum_{\{i, j\} \in E} S_{ij} \\ & \text{subject to} && I(\{S_{ij}\}) \geq I^*. \end{aligned} \quad (48)$$

The cost function in (48) is linear, and hence convex. Also, the constraint set $\{\{S_{ij}\} : I(\{S_{ij}\}) \geq I^*\} = \{\{S_{ij}\} : -I(\{S_{ij}\}) \leq -I^*\}$ is convex, as a sub level set of the convex function $-I(\{S_{ij}\})$. (See Lemma 18.) Hence, we have just proved the following Lemma.

Lemma 19: The optimization problem (48) is convex. Convexity of (48) allows us to find a globally optimal solution.

C. Simulation Example

We first describe the simulation setup. We consider a geometric network with $N = 14$ sensors. We place the sensors uniformly over a unit square, and connect those sensors whose distance d_{ij} is less than a radius. The total number of (undirected)

⁴We assumed equal noise variances $\sigma_n^2 = \text{Var}(n_{ij,k}) = \text{Var}(n_{ji,k})$ so that $K_{ij} = K_{ji}$, which implies the constraint $S_{ij} = S_{ji}$. Our analysis easily extends to unequal noise variances, in which case we would require $\frac{K_{ij}}{S_{ij}} = \frac{K_{ji}}{S_{ji}}$, this is not considered here.

links is 38. (These 38 links are the failing links, for which we want to allocate the transmission powers S_{ij} .) We set the coefficients $K_{ij} = 6.25d_{ij}^\alpha$, with $\alpha = 2$. For the averaging weights, we use Metropolis weights, i.e., if link $\{i, j\}$ is online, we assign $W_{ij,k} = 1/(1 + \max\{d_{i,k}, d_{j,k}\})$, where $d_{i,k}$ is the degree of node i at time k and $W_{ij,k} = 0$ otherwise; also, $W_{ii,k} = 1 - \sum_{j \in O_{i,k}} W_{ij,k}$. For the sensors' measurements, we use the Gaussian distribution $f_1 \sim \mathcal{N}(m, \sigma^2)$, $f_0 \sim \mathcal{N}(0, \sigma^2)$, with $\sigma^2 = 1$. For a lower signal-to-noise ratio (SNR) case, we set $m = 0.0447$, and for a higher SNR case, we set $m = 2 \cdot 0.0447$. The corresponding values are $I^* = (N - 1)N \frac{m^2}{8\sigma^2} = 0.0455$, for a lower SNR, and $I^* = 0.182$, for a higher SNR; see [23].

To obtain the optimized power allocation, we solve the optimization problem (48) by applying the subgradient algorithm with constant stepsize $\beta = 0.0001$ on the unconstrained exact penalty reformulation of (48), see, e.g., [40], which is to minimize $\sum_{\{i, j\} \in E} S_{ij} + \mu \max\{0, -\text{mincut}(V, E, C) + I^*\}$, where $C = [c_{ij}]$, $c_{ij} = -\log(1 - e^{-K_{ij}/S_{ij}})$, for $\{i, j\} \in E$, and zero else; and μ is the penalty parameter that we set to $\mu = 500$. We used the MATLAB implementation [44] of the min-cut algorithm from [42]. Note that the resulting power allocation is optimal over the class of *deterministic* power allocations, i.e., the power allocations that: 1) use the same total power across all links per each time step; and 2) use *deterministic* power assignment policy at each time step.

Results. Fig. 2 (center) plots the detection error probability for a lower SNR case, of the worst sensor $\max_{i=1, \dots, N} P_i^e(k)$ versus time k . We compare: 1) the optimized power allocation $\{S_{ij}^*\}$ (solid blue line); 2) the uniform power allocation $S_{ij} = S$ across all links, such that the total power per k over all links $2 \sum_{\{i, j\} \in E} S_{ij} = 2 \sum_{\{i, j\} \in E} S_{ij}^* =: S$; and 3) a *random*, gossip like, power allocation, where, at a time step k , only one out of all links is activated (uniformly across all links) such that the power S is invested in it (half of S in each direction of the communication.) Note that this allocation is *random*, hence outside of the class that we optimize over. The optimized power allocation significantly outperforms the uniform power allocation. For example, to achieve the error probability 0.1, the optimized power allocation scheme requires about 550 time steps, hence the total consumed power is $550S$; in contrast, the uniform power allocation needs more than $2000S$ for the same target error 0.1. In addition, Fig. 2 plots the detection performance for the uniform power allocation with the total power per k equal to $\text{sr} \times 3S$. This scheme takes more than 700 time steps to achieve an error of 0.1, hence requiring the total power of $700 \times 3 \times S = 2100S$ to achieve an error of 0.1. Further, we can see that, for a lower SNR case, the random, gossip policy achieves—exactly as the optimized policy—the best detection error exponent \mathcal{D} . (Note that the two corresponding lines are parallel.) This is not a contradiction as the random policy is outside of the class of deterministic allocations that we optimize over. Furthermore, the randomized gossip policy is slightly better than the optimized policy (It has a better constant C in the detection error $P_e(k) \approx C e^{-k\mathcal{D}}$). However, for a larger SNR (Fig. 2, bottom), the gossip policy no longer achieves the optimal slope \mathcal{D} , and the optimized policy becomes better. In particular, for the 0.1 detection error, the optimized policy

saves about 50 time steps (from 200 to 150), with respect to gossip, hence saving 25% of total required power.

VI. CONCLUSION

In this paper, we studied the products $W_k \cdots W_1$ of i.i.d. symmetric, stochastic matrices W_k . Under the assumption of almost surely positive diagonals of the W_k 's and $\|E[W_k] - J\| < 1$, we found the large deviation rate I for $\mathbb{P}(\|W_k \cdots W_1 - J\| \geq d_k)$, $\log d_k = o(k)$, i.e., the large deviation rate for the probability that the convergence of $W_k \cdots W_1$ to J is sub-exponential. We showed that the rate I depends solely on the probabilities of the graphs that underly the matrices W_k . In general, calculating the rate I is a combinatorial problem. However, we show that, for the two commonly used averaging models, gossip and link failure, the rate I is obtained by solving an instance of the min-cut problem, and is hence easily computable. Further, for certain simple structures, we compute the rate I in closed form: for gossip over a tree, $I = |\log(1 - p_{ij})|$, where p_{ij} is the occurrence probability of the "weakest" link, i.e., the smallest-probability link; for both gossip and link failure models over a regular network, the rate $I = |\log p_{\text{isol}}|$, where p_{isol} is the probability that a node is isolated from the rest of the network at a time. Intuitively, our results show that the rate I is determined by the most likely way in which the network stays disconnected over a long period of time. Finally, we illustrated the usefulness of rate I by allocating the sensors' transmission power for consensus+innovations distributed detection.

REFERENCES

- [1] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, Jan. 2010.
- [2] H. Furstenberg and H. Kesten, "Products of random matrices," *Ann. Math. Statist.*, vol. 31, no. 2, pp. 457–469, Jun. 1960.
- [3] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 634–649, May 2008.
- [4] P. Denantes, F. Benezit, P. Thiran, and M. Vetterli, "Which distributed averaging algorithm should I choose for my sensor network?," in *Proc. INFOCOM 2008, 27th IEEE Conf. Comput. Commun.*, Phoenix, AZ, USA, Mar. 2008, pp. 986–994.
- [5] R. D. Carr, G. Konjevod, G. Little, V. Natarajan, and O. Parekh, "Compacting cuts: A new linear formulation for minimum cut," *ACM Trans. Algorithms*, vol. 5, no. 3, Jul. 2009, DOI: 10.1145/1541885.1541888.
- [6] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, MIT, Cambridge, MA, USA, 1984.
- [7] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, pp. 118–121, 1974.
- [8] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. AC-48, no. 6, pp. 988–1001, Jun. 2003.
- [9] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
- [10] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
- [11] A. Dimakis, A. Sarwate, and M. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1205–1216, 2008.
- [12] D. Üstebay, B. Oreshkin, M. Coates, and M. Rabbat, "Greedy gossip with eavesdropping," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3765–3776, 2010.
- [13] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010, DOI: 10.1109/JPROC.2010.2052531.
- [14] A. Nedić and A. Ozdaglar, "Convergence rate for consensus with delays," *J. Global Optimiz.*, vol. 47, no. 3, pp. 437–456, 2008.
- [15] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2506–2517, 2009.
- [16] Y. Mo and B. Sinopoli, "Communication complexity and energy efficient consensus algorithm," in *Proc. 2nd IFAC Workshop on Distrib. Estim. Control in Netw. Syst.*, France, Sep. 2010, DOI: 10.3182/20100913-2-FR-4014.00057.
- [17] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM Rev.*, vol. 53, pp. 747–772, Nov. 2011.
- [18] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Aug. 2008.
- [19] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.
- [20] S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimator," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 410–415, Feb. 2009.
- [21] P. Braca, S. Marano, V. Matta, and P. Willet, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.
- [22] D. Bajović, D. Jakovetić, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.
- [23] 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 [Online]. Available: <http://arxiv.org/abs/1111.4555>
- [24] L. Bruneau, A. Joye, and M. Merkli, "Infinite products of random matrices and repeated interaction dynamics," *Ann. l'Inst. Henri Poincaré, Probab. Statist.*, vol. 46, no. 2, pp. 442–464, 2010.
- [25] B. Touri and A. Nedić, "Product of random stochastic matrices," *Ann. Probabil.* 2011 [Online]. Available: <http://arxiv.org/pdf/1110.1751v1.pdf>, submitted for publication
- [26] A. Leizarowitz, "On infinite products of stochastic matrices," *Linear Algebra Appl.*, vol. 168, pp. 189–219, Apr. 1992.
- [27] B. Touri and A. Nedić, "On backward product of stochastic matrices," 2011 [Online]. Available: <http://arxiv.org/abs/1102.0244>
- [28] P. Diaconis and P. M. Wood, "Random doubly stochastic tridiagonal Matrices," 2011 [Online]. Available: <http://stat.stanford.edu/cgates/PERSI/papers/TriDiag1.pdf>
- [29] E. Seneta, *Nonnegative Matrices and Markov Chains*. New York, NY, USA: Springer, 1981.
- [30] V. N. Tutubalin, "On limit theorems for products of random matrices," *Theory Probabil. Appl.*, vol. 10, pp. 15–27, 1965.
- [31] Y. Guivarc'h and A. Raugi, "Products of random matrices: Convergence theorems," *Contemp. Math.*, vol. 50, pp. 31–54, 1986.
- [32] É. Le Page, "Théorèmes limites pour les produits de matrices aléatoires," *Probabil. Meas. Groups (Oberwolfach, 1981), Lecture Notes in Math.*, vol. 928, pp. 258–303, 1982.
- [33] H. Hennion, "Limit theorems for products of positive random matrices," *Ann. Probabil.*, vol. 25, no. 4, pp. 1545–1587, 1997.
- [34] V. Kargin, "Products of random matrices: Dimension and growth in norm," *Ann. Probabil.*, vol. 20, no. 3, pp. 890–906, 2010.
- [35] D. Jakovetić, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3788–3801, Jul. 2010.
- [36] A. Tahbaz-Salehi and A. Jadbabaie, "On consensus over random networks," in *Proc. 44th Ann. Allerton Conf. Commun., Contr., Comput.*, IL, USA, Sep. 2006, pp. 1315–1321.
- [37] A. F. Karr, *Probability Theory*. New York, NY, USA: Springer-Verlag, 1993.
- [38] M. Fiedler, "Algebraic connectivity of graphs," *Czech. Math. J.*, vol. 23, no. 98, pp. 298–305, 1973.
- [39] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [40] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*, ser. Grundlehren der Math. Wissenschaften. Berlin, Germany: Springer-Verlag, 1993, vol. 305 and 306.
- [41] A. F. Karr, *Probability Theory*. New York, NY, USA: Springer-Verlag, 1993, Springer Texts in Statist..

- [42] M. Stoer and F. Wagner, "A simple min-cut algorithm," *J. ACM*, vol. 44, no. 4, pp. 585–591, Jul. 1997.
- [43] K. Chan, A. Swami, Q. Zhao, and A. Scaglione, "Consensus algorithms over fading channels," in *Proc. Military Commun. Conf. (MILCOM)*, San Jose, CA, USA, Oct. 2010, pp. 549–554.
- [44] Y. Devir, in *Matlab m-File for the Min-Cut Algorithm*, 2006 [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/13892-a-simple-min-cut-algorithm>



Dragana Bajović (S'11) received the Engineer diploma from the School of Electrical Engineering, Department of Automatic Control, University of Belgrade, in 2007.

She is currently working toward the Ph.D. degree in the Dual Ph.D. program between Carnegie-Mellon University, Pittsburgh, PA, and Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal. Her research interests are in the areas of optimization, dimensionality reduction, and information processing in sensor networks.



João Xavier (S'97–M'03) received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2002.

Currently, he is an Assistant Professor with the Department of Electrical and Computer Engineering, IST. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon. His current research interests are in the area of optimization and statistical inference for distributed systems.



José M. F. Moura (S'71–M'75–SM'90–F'94) received the engenheiro electrotécnico degree from the Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He is University Professor at Carnegie Mellon University (CMU), Pittsburgh, PA, with the Departments of Electrical and Computer Engineering and, by courtesy, BioMedical Engineering. He was on the faculty at IST, has held visiting faculty appointments at MIT, and was a Visiting Research Scholar with the University of Southern California, Los Angeles.

He is a founding codirector of the Center for Sensed Critical Infrastructures Research (CenSCIR) and manages a large education and research program between CMU and Portugal, www.cmuportugal.org. His research interests include statistical and algebraic signal processing, image, bioimaging, and video processing, and digital communications. He has published over 450 technical Journal and Conference papers, filed 11 patents with eight issued by the U.S. Patent Office, is the co-editor of two books, and has given numerous invited seminars at international conferences, U.S. and European Universities, and industrial and government Laboratories.

Dr. Moura is the Director of IEEE Division IX (2012–2013) and a member of the IEEE Board of Directors. He served on the IEEE Publications Services and Products Board. He was the President (2008–2009) of the IEEE Signal Processing Society (SPS), Editor-in-Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING, interim Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, and was on the Editorial Board of several Journals, including the IEEE PROCEEDINGS, the IEEE SIGNAL PROCESSING MAGAZINE, and the ACM *Transactions on Sensor Networks*. He has been on the steering and technical committees of several Conferences. He is a member of the U.S. National Academy of Engineering, a Fellow of the American Association for the Advancement of Science (AAAS), and a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society Meritorious Service Award and the IEEE Millennium Medal in 2000. In 2007, he received the CMU's College of Engineering Outstanding Research Award and in 2009, the 2008 Philip L. Dowd Fellowship Award for Contributions to Engineering Education. In 2010, he was elected University Professor at CMU and received the IEEE Signal Processing Technical Achievement Award. In 2012, he received the IEEE Signal Processing Society Award.



Bruno Sinopoli (M'03) received the Dr.Eng. degree from the University of Padova in 1998 and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, in 2003 and 2005 respectively.

After a postdoctoral position with Stanford University, Stanford, CA, he joined the faculty at Carnegie Mellon University, Pittsburgh, PA, where he is an Assistant Professor with the Department of Electrical and Computer Engineering with courtesy appointments in Mechanical Engineering and in the Robotics Institute. His research interests include networked embedded control systems, distributed estimation and control over wireless sensor-actuator networks, and cyberphysical systems security.

Dr. Sinopoli was awarded the 2006 Eli Jury Award for Outstanding Research Achievement in the areas of systems, communications, control, and signal processing at UC Berkeley and the NSF Career Award in 2010.