Weight Optimization for Consensus Algorithms With Correlated Switching Topology

Dušan Jakovetić, *Student Member, IEEE*, João Xavier, *Member, IEEE*, and José M. F. Moura, *Fellow, IEEE*

***Abstract*—We design the weights in consensus algorithms for spatially correlated random topologies. These arise with 1) networks with spatially correlated random link failures and 2) networks with randomized averaging protocols. We show that the weight optimization problem is convex for both symmetric and asymmetric random graphs. With symmetric random networks, we choose the consensus mean-square error (MSE) convergence rate as the optimization criterion and explicitly express this rate as a function of the link formation probabilities, the link formation spatial correlations, and the consensus weights. We prove that the MSE convergence rate is a convex, nonsmooth function of the weights, enabling global optimization of the weights for arbitrary link formation probabilities and link correlation structures. We extend our results to the case of asymmetric random links. We adopt as optimization criterion the mean-square deviation (MSdev) of the nodes' states from the current average state. We prove that MSdev is a convex function of the weights. Simulations show that significant performance gain is achieved with our weight design method when compared with other methods available in the literature.**

***Index Terms*—Broadcast gossip, consensus, correlated link failures, sensor networks, switching topology, unconstrained optimization, weight optimization.**

I. Introduction

THIS paper finds the optimal weights for the consensus algorithm in correlated random networks. Consensus is an iterative distributed algorithm that computes the global average of data distributed among a network of agents using only local communications. Consensus has renewed interest in distributed algorithms [1], [2], arising in many different areas from distributed data fusion [3]–[7] to coordination of mobile autonomous agents [8], [9]. A recent survey is [9].

This paper studies consensus algorithms in networks where the links (being online or off line) are random. We consider two scenarios: 1) the network is random, because links in the network may fail at random times and 2) the network protocol is randomized, i.e., the link states along time are controlled by a randomized protocol (e.g., standard gossip algorithm [10], broadcast gossip algorithm [11]). In both cases, we model the links as Bernoulli random variables. Each link (i,j) has some formation probability, i.e., probability of being active, equal to P_{ij}. Different links may be correlated at the same time, which can be expected in real applications. For example, in wireless sensor networks (WSNs), links can be spatially correlated due to interference among close links or electromagnetic shadows that may affect several nearby sensors.

References on consensus under time varying or random topology are [12]–[14] and [11], [15]–[19], among others, respectively. Most of the previous work is focussed on providing convergence conditions and/or characterizing the convergence rate under different assumptions on the network randomness [16]–[18]. For example, [16] and [20] study consensus algorithm with spatially and temporally independent link failures. They show that a necessary and sufficient condition for mean-square and almost sure convergence is for the communication graph to be connected on average.

We consider here the weight optimization problem: how to assign the weights W_{ij} with which the nodes mix their states across the network, so that the convergence towards consensus is the fastest possible. This problem has not been solved (with full generality) for consensus in random topologies. We study this problem for networks with symmetric and asymmetric random links separately, since the properties of the corresponding algorithm are different. For symmetric links (and connected network topology on average), the consensus algorithm converges to the average of the initial nodes' states almost surely. For asymmetric random links, all the nodes asymptotically reach agreement, but they only agree to a random variable in the neighborhood of the true initial average.

We refer to our weight solution as probability-based weights (PBWs). PBW are simple: we assume at each iteration that the weight of link (i,j) is W_{ij} (to be optimized) when the link is alive, or 0, otherwise. Self-weights are adapted such that the row-sums of the weight matrix at each iteration are one. This is suitable for distributed implementation. Each node updates its state after receiving messages from its current neighbors. No information about the number of nodes in the network or the neighbor's current degrees is needed. Hence, no additional online communication is required for computing

Manuscript received May 31, 2009; accepted March 01, 2010. Date of publication March 25, 2010; date of current version June 16, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anna Scaglione. This work was supported in part by the NSF under Grant CNS-0428404, by the Office of Naval Research under MURI N000140710747, and by the Carnegie Mellon/Portugal Program under a grant of the Fundação de Ciência e Tecnologia (FCT) from Portugal and in part supported by Grants SIPM PTDC/EEA-ACR/73749/2006 and SFRH/BD/33520/2008 (through the Carnegie Mellon/Portugal Program managed by ICTI) from FCT; and also by ISR/IST pluriannual funding (POSC program, FEDER). D. Jakovetić holds a fellowship from FCT.

D. Jakovetić is with the Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (IST), 1049-001 Lisboa, Portugal, and also with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890 (e-mail: djakovet@andrew.cmu.edu; djakovetic@isr.ist.utl.pt).

J. Xavier is with the SIPG-Signal and Image, Processing Group Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (IST), 1049-001 Lisboa, Portugal (e-mail: jxavier@isr.ist.utl.pt).

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890 USA (e-mail: 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.2010.2046635

weights, in contrast, for instance, to the case of the Metropolis weights (MWs) [14].

Our weight design method assumes that the link formation probabilities and their spatial correlations are known. With randomized protocols, the link formation probabilities and their correlations are induced by the protocol itself, and thus are known. For networks with random link failures, the link formation probabilities relate to the signal to noise ratio at the receiver and can be computed. In [21], the formation probabilities are designed in the presence of link communication costs and an overall network communication cost budget. When the WSN infrastructure is known, it is possible to estimate the link formation *probabilities* by measuring the reception rate of a link computed as the ratio between the number of received and the total number of sent packets. Another possibility is to estimate the link formation probabilities based on the received signal strength. Link formation *correlations* can also be estimated on actual WSNs [22]. If there is no training period to characterize quantitatively the links on an actual WSN, we can still model the probabilities and the correlations as a function of the transmitted power and the intersensor distances. Moreover, several empirical studies ([22] and [23] and references therein) on the quantitative properties of wireless communication in sensor networks have been done that provide models for packet delivery performance in WSNs.

Summary of the Paper: Section II lists our contributions, relates them with the existing literature, and introduces notation used in the paper. Section III describes our model of random networks and the consensus algorithm. Section IV and Section V study the weight optimization for symmetric random graphs and asymmetric random graphs, respectively. Section VI demonstrates the effectiveness of our approach by simulation. Finally, Section VII concludes the paper. We derive the proofs of some results in the Appendices A–C.

II. Contributions, Related Work, and Notation

1) Contribution: Our results extend previous studies of convergence conditions and rates for consensus algorithm (e.g.,[11], [15], and [21]). We address the problem of designing the optimal weights in consensus algorithms with *correlated* random topologies. Our method is applicable to 1) networks with correlated random link failures (see, e.g., [21]) and 2) networks with randomized algorithms (see, e.g., [10] and [11]). We first address the weight design problem for symmetric random links, and then we extend the results to asymmetric random links.

With symmetric random links, we use the mean-square consensus convergence rate $\phi(W)$ as the optimization criterion. We explicitly express the rate $\phi(W)$ as a function of the link formation probabilities, their correlations, and the weights. We prove that $\phi(W)$ is a convex, nonsmooth function of the weights. This enables global optimization of the weights for arbitrary link formation probabilities and arbitrary link correlation structures. We solve numerically the resulting optimization problem by the subgradient algorithm, showing also that the optimization computational cost grows tolerably with the network size. We provide insights into the weight design with a simple example of a complete random network. This example admits closed form solution for the optimal weights and for the convergence rate. This example shows how the optimal weights depend on the number of nodes, the link formation probabilities, and their correlations.

We extend our results to the case of asymmetric random links, adopting as optimization criterion the mean-square deviation (from the current average state) rate $\psi(W)$, and show that $\psi(W)$ is a convex function of the weights.

We provide comprehensive simulation experiments to demonstrate the effectiveness of our approach. We consider two different models of random networks with correlated link failures; in addition, we study the broadcast gossip algorithm [11], as an example of randomized protocol with asymmetric links. In all cases, simulations confirm that our method shows significant gain over other methods available in the literature. Also, we show that the gain increases with the network size.

2) Related Work: Weight optimization for consensus with switching topologies has not received much attention in the literature. Reference [21] studies the tradeoff between the convergence rate and the amount of communication that takes place in the network. This reference is mainly concerned with the design of the network topology, i.e., the design of the probabilities of reliable communication $\{P_{ij}\}$ and the weight α (assuming all nonzero weights are equal) assuming a communication cost C_{ij} per link, and an overall network communication budget. Reference [11] proposes the broadcast gossip algorithm, where at each time step, a single node, selected at random, broadcasts unidirectionally its state to all the neighbors within its wireless range. We detail the broadcast gossip in Section VI-B. This reference optimizes the weight for the broadcast gossip algorithm assuming equal weights for all links.

The problem of optimizing the weights for consensus under a random topology, when the weights for different links may be different, has not received much attention in the literature. Authors have proposed weight choices for random or time-varying networks [14], [24], but no claims to optimality are made. Reference [14] proposes the MWs, based on the Metropolis–Hastings algorithm for simulating a Markov chain with uniform equilibrium distribution [25]. The weights choice in [24] is based on the fastest mixing Markov chain problem studied in [26] and uses the information about the underlying supergraph. We refer to this weight choice as the supergraph based weights (SGBWs).

Notation: Vectors are denoted by a lower case letter (e.g., x), and it is understood from the context if x denotes a deterministic or random vector. Symbol $\mathbb{R}^N$ is the N-dimensional Euclidean space. Inequality $x \leq y$ is understood element wise, i.e., it is equivalent to $x_i \leq y_i$, for all i. Constant matrices are denoted by capital letters (e.g., X) and random matrices are denoted by calligraphic letters (e.g., $\mathcal{X}$.) A sequence of random matrices is denoted by $\{\mathcal{X}(k)\}_{k=0}^{\infty}$ and the random matrix indexed by k is denoted $\mathcal{X}(k)$. If the distribution of $\mathcal{X}(k)$ is the same for any k, we shorten the notation $\mathcal{X}(k)$ to $\mathcal{X}$ when the time instant k is not of interest. Symbol $\mathbb{R}^{N\times M}$ denotes the set of $N \times M$ real valued matrices and $\mathbb{S}^N$ denotes the set of symmetric real valued $N \times N$ matrices. The ith column of a matrix X is denoted by X_i. Matrix entries are denoted by X_{ij}. Quantities $X \otimes Y$, $X \odot Y$, and $X \oplus Y$ denote the Kronecker product, the Hadamard product, and the direct sum of the matrices X

and Y, respectively. Inequality $X \succeq Y (X \preceq Y)$ means that the matrix $X - Y$ is positive (negative) semidefinite. Inequality $X \geq Y (X \leq Y)$ is understood entry wise, i.e., it is equivalent to $X_{ij} \geq Y_{ij}$, for all i, j. Quantities $\|X\|$, $\lambda_{\max}(X)$, and $r(X)$ denote the matrix 2-norm, the maximal eigenvalue, and the spectral radius of X, respectively. The identity matrix is I. Given a matrix A, $\mathrm{Vec}(A)$ is the column vector that stacks the columns of A. For given scalars $x_1, \ldots, x_N$, $\mathrm{Diag}\,(x_1, \ldots, x_N)$ denotes the diagonal $N \times N$ matrix with the ith diagonal entry equal to x_i. Similarly, $\mathrm{Diag}(x)$ is the diagonal matrix whose diagonal entries are the elements of x. The matrix $\mathrm{Diag}(X)$ is a diagonal matrix with the diagonal equal to the diagonal of X. The N-dimensional column vector of ones is denoted with 1_N, and the subscript N is dropped when the dimension is understood from the context. We also let $J = (1/N)1_N 1_N^\top$. The ith canonical unit vector, i.e., the ith column of I, is denoted by e_i. Symbol $|S|$ denotes the cardinality of a set S.

III. Problem Model

This section introduces the random network model that we apply to networks with link failures and to networks with randomized algorithms. It also introduces the consensus algorithm and the corresponding weight rule assumed in this paper.

A. Random Network Model: Symmetric and Asymmetric Random Links

We consider random networks—networks with random links or with a random protocol. Random links arise because of packet loss or drop, or when a sensor is activated from sleep mode at a random time. Randomized protocols like standard pairwise gossip [10] or broadcast gossip [11] activate links randomly. This section describes the network model that applies to both problems. We assume that the links are up or down (link failures) or selected to use (randomized gossip) according to spatially correlated Bernoulli random variables.

To be specific, the network is modeled by a graph $G = (V, E)$, where the set of nodes V has cardinality $|V| = N$ and the set of directed edges E, with $|E| = 2M$, collects the ordered node pairs that are allowed to communicate, i.e., the realizable links. For example, with geometric graphs, realizable links connect nodes within their communication radius. The graph G is called the supergraph, e.g., [21]. The directed edge $(i, j) \in E$ if node j can transmit to node i.

The supergraph G is assumed to be symmetric, connected, and without self-loops. For the fully connected supergraph, the number of directed edges $2M$ is equal to $N(N-1)$. We are interested in supergraphs for which $M \ll 1/2N(N-1)$.

Associated with the graph G is its $N \times N$ adjacency matrix A:

$$A_{ij} = \begin{cases} 1, & \text{if } (i,j) \in E \\ 0, & \text{otherwise.} \end{cases}$$

The in-neighborhood set Ω_i (nodes that can transmit to node i) and the in-degree d_i of a node i are

$$\Omega_i = \{j : (i,j) \in E\}$$
$$d_i = |\Omega_i|.$$

We model the connectivity of a random WSN at time step k by a (possibly) directed random graph $\mathcal{G}(k) = (V, \mathcal{E}(k))$. The random edge set is

$$\mathcal{E}(k) = \{(i,j) \in E : (i,j) \text{ is online at time step } k\},$$

with $\mathcal{E}(k) \subseteq E$. The random adjacency matrix associated to $\mathcal{G}(k)$ is denoted by $\mathcal{A}(k)$ and the random in-neighborhood for sensor i by $\Omega_i(k)$.

We assume that link failures are *temporally independent* and *spatially correlated*. That is, we assume that the random matrices $\mathcal{A}(k)$, $k = 0, 1, 2, \ldots$ are independent identically distributed. The state of the link (i, j) at a time step k is a Bernoulli random variable, with mean P_{ij}, i.e., P_{ij} is the formation probability of link (i, j). At time step k, different edges (i, j) and (p, q) may be correlated, i.e., the entries $\mathcal{A}_{ij}(k)$ and $\mathcal{A}_{pq}(k)$ may be correlated. For the link r, by which node j transmits to node i, and for the link s, by which node q transmits to node p, the corresponding cross-variance is

$$[R_q]_{rs} = \mathrm{E}[\mathcal{A}_{ij}\mathcal{A}_{pq}] - P_{ij}P_{pq}.$$

Time correlation, as spatial correlation, arises naturally in many scenarios, such as when nodes awake according to a sleep schedule. However, time correlation between links requires an approach different than the one we pursue in this paper [20]. We plan to address the weight optimization with temporally correlated links in future work.

B. Consensus Algorithm

Let $x_i(0)$ represent some scalar measurement or initial data available at sensor i, $i = 1, \ldots, N$. Denote by x_{avg} the average

$$x_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} x_i(0).$$

The consensus algorithm computes x_{avg} iteratively at each sensor i by the distributed weighted average

$$x_i(k+1) = \mathcal{W}_{ii}(k)x_i(k) + \sum_{j \in \Omega_i(k)} \mathcal{W}_{ij}(k)x_j(k). \quad (1)$$

We assume that the random weights $\mathcal{W}_{ij}(k)$ at iteration k are given by

$$\mathcal{W}_{ij}(k) = \begin{cases} W_{ij}, & \text{if } j \in \Omega_i(k) \\ 1 - \sum_{m \in \Omega_i(k)} \mathcal{W}_{im}(k), & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

In (2), the quantities W_{ij} are nonrandom and are the variables to be optimized in our work. We also take $W_{ii} = 0$, for all i. By (2), when the link (i, j) is active, the weight is W_{ij}, and when not active it is zero. Note that W_{ij} are (possibly) nonzero only for edges (i, j) in the supergraph G. If an edge (i, j) is not in the supergraph the corresponding $W_{ij} = 0$ and $\mathcal{W}_{ij}(k) \equiv 0$.

We write the consensus algorithm in compact form. Let $x(k) = (x_1(k), x_2(k), \ldots, x_N(k))^\top$, $W = [W_{ij}]$, $\mathcal{W}(k) = [\mathcal{W}_{ij}(k)]$. The random weight matrix $\mathcal{W}(k)$ can be written in compact form as

$$\mathcal{W}(k) = W \odot \mathcal{A}(k) - \mathrm{Diag}\,(W\mathcal{A}(k)^\top) + I \quad (3)$$

and the consensus algorithm is simply stated with $x(k=0)=x(0)$ as

$$x(k+1)=\mathcal{W}(k)x(k),\ k\geq 0. \tag{4}$$

To implement the update rule, nodes need to know their random in-neighborhood $\Omega_i(k)$ at every iteration. In practice, nodes determine $\Omega_i(k)$ based on who they receive messages from at iteration k.

It is well known [11], [15] that, when the random matrix $\mathcal{W}(k)$ is symmetric, the consensus algorithm is average preserving, i.e., the sum of the states $x_i(k)$, and so the average state over time, does not change, even in the presence of random links. In that case the consensus algorithm converges almost surely to the true average x_{avg}. When the matrix $\mathcal{W}(k)$ is not symmetric, the average state is not preserved in time, and the state of each node converges to the same random variable with bounded mean-square error from x_{avg} [11]. For certain applications, where high precision on computing the average x_{avg} is required, average preserving, and thus a symmetric matrix $\mathcal{W}(k)$ is desirable. In practice, a symmetric matrix $\mathcal{W}(k)$ can be established by protocol design even if the underlying physical channels are asymmetric. This can be realized by ignoring unidirectional communication channels. This can be done, for instance, with a double acknowledgement protocol. In this scenario, effectively, the consensus algorithm sees the underlying random network as a symmetric network, and this scenario falls into the framework of our studies of symmetric links (Section IV).

When the physical communication channels are asymmetric, and the error on the asymptotic consensus limit c is tolerable, we can use consensus with an asymmetric weight matrix $\mathcal{W}(k)$. This type of algorithm is easier to implement, since there is no need for acknowledgement protocols. An example of such a protocol is the broadcast gossip algorithm proposed in [11]. Section V studies this type of algorithms.

Set of Possible Weight Choices: Symmetric Network: With symmetric random links ($\mathcal{A}(k)=\mathcal{A}(k)^\top$), we will always assume $W_{ij}=W_{ji}$. By doing this we easily achieve the desirable property that $\mathcal{W}(k)$ is symmetric. The set of all possible weight choices for symmetric random links S_W becomes

$$S_W=\Big\{W\in\mathbb{R}^{N\times N}: W_{ij}=W_{ji}, W_{ij}=0, \text{ if } (i,j)\notin E, W_{ii}=0,\ \forall\, i\Big\}. \tag{5}$$

Set of Possible Weight Choices: Asymmetric Network: With asymmetric random links, there is no good reason to require that $W_{ij}=W_{ji}$, and thus we drop the restriction $W_{ij}=W_{ji}$. The set of possible weight choices in this case becomes:

$$S_W^{\mathrm{asym}}=\Big\{W\in\mathbb{R}^{N\times N}: W_{ij}=0, \text{ if } (i,j)\notin E, W_{ii}=0,\ \forall\, i\Big\}. \tag{6}$$

Depending whether the random network is symmetric or asymmetric, there will be two error quantities that will play a role. These will be discussed in detail in Section IV and Section V, respectively. We introduce them here briefly, for reference.

Mean-Square Error: Symmetric Network: Define the consensus error vector $e(k)$ and the error covariance matrix $\Sigma(k)$:

$$e(k)=x(k)-x_{\mathrm{avg}}1 \tag{7}$$

$$\Sigma(k)=E[e(k)e(k)^\top]. \tag{8}$$

The mean-square consensus error mean-square error (MSE) is given by

$$\mathrm{MSE}(k)=\sum_{i=1}^{N}E\left[(x_i(k)-x_{\mathrm{avg}})^2\right] = E[e(k)^\top e(k)]=\mathrm{tr}\,\Sigma(k). \tag{9}$$

Mean-Square Deviation: Asymmetric Network: As explained, when the random links are asymmetric (i.e., when $\mathcal{W}(k)$ is not symmetric), and if the underlying supergraph is strongly connected, then the states of all nodes converge almost surely to a common value c that is in general a random variable that depends on the sequence of network realizations and on the initial state $x(0)$ (see [11] and [15]). In order to have $c=x_{\mathrm{avg}}$, almost surely, an additional condition must be satisfied:

$$1^\top\mathcal{W}(k)=1^\top,\ \text{a.s.} \tag{10}$$

See [11] and [15] for the details. We remark that (10) is a crucial assumption in the derivation of the MSE decay (25). Theoretically, equation (23) is still valid if the condition $\mathcal{W}(k)=\mathcal{W}(k)^\top$ is relaxed to $1^\top\mathcal{W}(k)=1^\top$. While this condition is trivially satisfied for symmetric links and symmetric weights $W_{ij}=W_{ji}$, it is very difficult to realize (10) in practice when the random links are asymmetric. So, in our work, we do not assume (10) with asymmetric links.

For asymmetric networks, we follow reference [11] and introduce the mean-square state deviation (MSdev) as a performance measure. Denote the current average of the node states by $x_{\mathrm{avg}}(k)=(1/N)1^\top x(k)$. Quantity MSdev describes how far apart different states $x_i(k)$ are; it is given by

$$\mathrm{MSdev}(k)=\sum_{i=1}^{N}E\left[(x_i(k)-x_{\mathrm{avg}}\,(k))^2\right]=E[\zeta(k)^\top\zeta(k)] \tag{11}$$

where

$$\zeta(k)=x(k)-x_{\mathrm{avg}}\,(k)1=(I-J)\,x(k).$$

C. Symmetric Links: Statistics of $\mathcal{W}(k)$

In this subsection, we derive closed form expressions for the first and the second order statistics of the random matrix $\mathcal{W}(k)$. Let $q(k)$ be the random vector that collects the nonredundant entries of $\mathcal{A}(k)$:

$$q_l(k)=\mathcal{A}_{ij}(k),\ i<j,\ (i,j)\in E \tag{12}$$

where the entries of $\mathcal{A}(k)$ are ordered in lexicographic order with respect to i and j, from left to right, top to bottom. For symmetric links, $\mathcal{A}_{ij}(k)=\mathcal{A}_{ji}(k)$, so the dimension of $q(k)$ is half of the number of directed links, i.e., M. We let the mean and the covariance of $q(k)$ and $\mathrm{Vec}\,(\mathcal{A}(k))$ be

$$\pi = E[q(k)] \tag{13}$$
$$\pi_l = E[q_l(k)] \tag{14}$$
$$R_q = \mathrm{Cov}\,(q(k)) = E[(q(k)-\pi)(q(k)-\pi)^\top] \tag{15}$$
$$R_A = \mathrm{Cov}\,(\mathrm{Vec}(\mathcal{A}(k))). \tag{16}$$

The relation between R_q and R_A can be written as

$$R_A = F R_q F^\top \tag{17}$$

where $F \in \mathbb{R}^{N^2 \times M}$ is the zero one selection matrix that linearly maps $q(k)$ to $\mathrm{Vec}\,(\mathcal{A}(k))$, i.e., $\mathrm{Vec}\,(\mathcal{A}(k)) = Fq(k)$. We introduce further notation. Let P be the $N \times N$ matrix of the link formation probabilities

$$P = [P_{ij}].$$

where we defined $P_{ij} = 0$, if $(i,j) \notin E$, and $P_{ii} = 0$, for all i. Define the matrix $B \in \mathbb{R}^{N^2 \times N^2}$ with $N \times N$ zero diagonal blocks and $N \times N$ off diagonal blocks B_{ij} equal to

$$B_{ij} = 1e_i^\top + e_j 1^\top$$

(e_i is the ith column of I_N) and write W in terms of its columns $W = [W_1 W_2 \ldots W_N]$. We let

$$W_C = W_1 \oplus W_2 \oplus \cdots \oplus W_N.$$

For symmetric random networks, the mean of the random weight matrix $\mathcal{W}(k)$ and of $\mathcal{W}^2(k)$ play an important role for the convergence rate of the consensus algorithm. Using the above notation, we can get compact representations for these quantities, as provided in Lemma 1 proved in Appendix A.

Lemma 1: Consider $\mathcal{W}(k)$ as defined in (3). Then, the mean and the second moment R_C of $\mathcal{W}$ defined below are

$$\overline{W} = E[\mathcal{W}] = W \odot P + I - \mathrm{Diag}\,(WP^\top) \tag{18}$$
$$R_C = E[\mathcal{W}^2] - \overline{W}^2 \tag{19}$$
$$= W_C^\top \{R_A \odot (I \otimes 11^\top + 11^\top \otimes I - B)\} W_C. \tag{20}$$

In the special case of spatially uncorrelated links, the second moment R_C of $\mathcal{W}$ are

$$\frac{1}{2}R_C = \mathrm{Diag}\{((11^\top - P)\odot P)(W \odot W)\} - (11^\top - P)\odot P \odot W \odot W. \tag{21}$$

For asymmetric random links, the expression for the mean of the random weight matrix $\mathcal{W}$ remains the same (as in Lemma 1.) With asymmetric random links, the quantity $E[\mathcal{W}^\top (I-J)\,\mathcal{W}]$ will play a role (as explained in Subsection V-A), instead of $E[\mathcal{W}^2]$ given in (19). Generally, $E[\mathcal{W}^\top(I-J)\mathcal{W}]$ does not admit the compact representation as given in (19), and we do not pursue here cumbersome entry wise representations. In the Appendix C, we do present the expressions for the matrix $E[\mathcal{W}^\top(I-J)\mathcal{W}]$ for the broadcast gossip algorithm [11] (that we study in Section VI-B).

IV. Weight Optimization: Symmetric Random Links

A. Optimization Criterion: Mean-Square Convergence Rate

We are interested in finding the rate at which $\mathrm{MSE}(k)$ (defined in (9)) decays to zero and to optimize this rate with respect to the weights W. First, we derive the recursion for the error $e(k)$. We have from (4)

$$1^\top x(k+1) = 1^\top \mathcal{W}(k)x(k) = 1^\top x(k) = 1^\top x(0) = N x_{\mathrm{avg}}$$
$$1^\top e(k) = 1^\top x(k) - 1^\top 1 x_{\mathrm{avg}} = N x_{\mathrm{avg}} - N x_{\mathrm{avg}} = 0.$$

We derive the error vector dynamics:

$$\begin{aligned} e(k+1) &= x(k+1) - x_{\mathrm{avg}}1 \\ &= \mathcal{W}(k)x(k) - \mathcal{W}(k)x_{\mathrm{avg}}1 \\ &= \mathcal{W}(k)e(k) = (\mathcal{W}(k) - J)e(k) \end{aligned} \tag{22}$$

where the last equality holds because $Je(k) = (1/N)11^\top e(k) = 0$.

Recall the definition of the mean-square consensus error $\mathrm{MSE}(k)$ in (9) and the error covariance matrix in (8). Introduce the quantity

$$\phi(W) = \lambda_{\max}(E[\mathcal{W}^2] - J). \tag{23}$$

The next lemma shows that the mean-square error decays at the rate $\phi(W)$.

Lemma 2 (m.s.s Convergence Rate): Consider the consensus algorithm given by (4). Then

$$\mathrm{MSE}(k+1) = \mathrm{tr}(\Sigma(k+1)) = \mathrm{tr}((E[\mathcal{W}^2] - J)\Sigma(k)) \tag{24}$$
$$\mathrm{tr}\,(\Sigma(k+1)) \leq \phi(W)\mathrm{tr}\,(\Sigma(k)),\ k \geq 0. \tag{25}$$

Proof: From the definition of the covariance $\Sigma(k+1)$, using the dynamics of the error $e(k+1)$, interchanging expectation with the tr operator, using properties of the trace, interchanging the expectation with the tr once again, using the independence of $e(k)$ and $\mathcal{W}(k)$, and, finally, noting that $\mathcal{W}(k)J = J$, we get (24). The independence between $e(k)$ and $\mathcal{W}(k)$ follows because $\mathcal{W}(k)$ is an i.i.d. sequence, and $e(k)$ depends on $\mathcal{W}(0), \ldots, W(k-1)$. Then $e(k)$ and $\mathcal{W}(k)$ are independent by the disjoint block theorem [27]. Having (24), (25), can be easily shown, for example, by [28, p. 423, exercise 18]. ■

We remark that, in the case of asymmetric random links, MSE does not asymptotically go to zero. For the case of asymmetric links, we use the MSdev(k) (given in (11)) as the performance metric. This will be detailed in Section V.

B. Symmetric Links: Weight Optimization Problem Formulation

We now resort to (25) which shows that $\phi(W)$ is an upper bound on the decay of $\mathrm{MSE}(k)$. We formulate the weight optimization problem as finding the weights W_{ij} that minimize $\phi(W)$ (optimize the mean-square rate of convergence):

$$\begin{aligned} \text{minimize} \quad & \phi(W) \\ \text{subject to} \quad & W \in S_W. \end{aligned} \tag{26}$$

The set S_W is defined in (5) and the rate $\phi(W)$ is given by (23). The optimization problem (26) is unconstrained, since effectively the optimization variables are $W_{ij} \in \mathbb{R}$, $(i,j) \in E$, other entries of W being zero.

A point $W^\bullet \in S_W$ such that $\phi(W^\bullet) < 1$ will always exist if the supergraph G is connected. Reference [29] studies the case when the random matrices $\mathcal{W}(k)$ are stochastic (row-sums are one.) This reference shows that $\phi(W^\bullet) < 1$ if the supergraph is connected, and all the realizations of the random matrix $\mathcal{W}(k)$ are stochastic symmetric matrices. Thus, to locate a point $W^\bullet \in S_W$ such that $\phi(W^\bullet) < 1$, we just take $W^\bullet$ in the set:

$$S_{\mathrm{stoch}} = \{W \in S_W : W_{ij} > 0, \text{ if } (i,j) \in E, W1 < 1\} \subseteq S_W. \tag{27}$$

It is trivial to show that, for any point $W^\bullet \in S_{\mathrm{stoch}}$, all the realizations of $\mathcal{W}(k)$ are stochastic and symmetric, and thus $\phi(W^\bullet) < 1$ if the supergraph is connected. We remark that the optimum $W^\star$ of (26) does not have to lie in the set S_{stoch}. In general, $W^\star$ lies in the set

$$S_{\mathrm{conv}} = \{W \in S_W : \phi(W) < 1\} \subseteq S_W. \tag{28}$$

The set S_{stoch} is a proper subset of S_{conv}. (If $W \in S_{\mathrm{stoch}}$ then $\phi(W) < 1$, but the converse statement is not true in general.) We also remark that the consensus algorithm (4) converges *almost surely* if $\phi(W) < 1$ (not only in mean-square sense). This can be shown, for instance, by the technique developed in [29].

We now relate (26) to [30]. This reference studies the weight optimization for the case of a static topology. In this case the topology is deterministic, described by the supergraph G. The link formation probability matrix P reduces to the supergraph adjacency (zero-one) matrix A, since the links occur always if they are realizable. Also, the link covariance matrix R_q becomes zero. The weight matrix $\mathcal{W}$ is deterministic and equal to

$$\mathcal{W} = \overline{W} = W \cdot A - \mathrm{Diag}\,(WA) + I.$$

Recall that $r(X)$ denotes the spectral radius of X. Then, the quantities $(r(\mathcal{W} - J))^2$ and $\phi(W)$ coincide. Thus, for the case of static topology, the optimization problem (26) that we address reduces to the optimization problem proposed in [30].

C. Convexity of the Weight Optimization Problem

We show that $\phi : S_W \to \mathbb{R}_+$ is convex, where S_W is defined in (6) and $\phi(W)$ by (23). Recall the closed form expression of $E[\mathcal{W}^2]$ in Lemma 1. We remark that convexity of $\phi(W)$ is not obvious and requires proof. The function $\phi(W)$ is a concatenation of a matrix quadratic function and $\lambda_{\max}(\cdot)$. Although the function $\lambda_{\max}(\cdot)$ is a convex function of *its argument*, one still has to show that the following concatenation is convex: $W \mapsto E[\mathcal{W}^2] - J \mapsto \phi(W) = \lambda_{\max} E[\mathcal{W}^2] - J)$.

Lemma 3 (Convexity of $\phi(W)$): The function $\phi : S_W \to \mathbb{R}_+$ is convex.

Proof: Choose arbitrary $X, Y \in S_W$. We restrict our attention to matrices W of the form

$$W = X + tY, \quad t \in \mathbb{R}. \tag{29}$$

Recall the expression for $\mathcal{W}$ given by (2) and (4). For the matrix W given by (29), we have for $\mathcal{W} = \mathcal{W}(t)$

$$\begin{aligned} \mathcal{W}(t) &= I - \mathrm{Diag}[(X+tY)\mathcal{A}] + (X+tY) \odot \mathcal{A} \\ &= \mathcal{X} + t\mathcal{Y}, \mathcal{X} = X \odot \mathcal{A} + I - \mathrm{Diag}\,(X\mathcal{A}), \\ \mathcal{Y} &= Y \odot \mathcal{A} - \mathrm{Diag}\,(Y\mathcal{A}). \end{aligned} \tag{30}$$

Introduce the auxiliary function $\eta : \mathbb{R} \to \mathbb{R}_+$,

$$\eta(t) = \lambda_{\max}(E[\mathcal{W}(t)^2] - J). \tag{31}$$

To prove that $\phi(W)$ is convex, it suffices to prove that the function $\eta(t)$ is convex. Introduce $\mathcal{Z}(t)$ and compute successively

$$\mathcal{Z}(t) = \mathcal{W}(t)^2 - J \tag{32}$$
$$= (\mathcal{X} + t\mathcal{Y})^2 - J \tag{33}$$
$$= t^2\mathcal{Y}^2 + t(\mathcal{X}\mathcal{Y} + \mathcal{Y}\mathcal{X}) + \mathcal{X}^2 - J \tag{34}$$
$$= t^2\mathcal{Z}_2 + t\mathcal{Z}_1 + \mathcal{Z}_0. \tag{35}$$

The random matrices $\mathcal{Z}_2$, $\mathcal{Z}_1$, and $\mathcal{Z}_0$ do not depend on t. Also, $\mathcal{Z}_2$ is semidefinite positive. The function $\eta(t)$ can be expressed as

$$\eta(t) = \lambda_{\max}(E[\mathcal{Z}(t)]).$$

We will now derive that

$$\begin{aligned} \mathcal{Z}((1-\alpha)t + \alpha u) &\preceq (1-\alpha)\mathcal{Z}(t) + \alpha\mathcal{Z}(u), \\ &\forall\, \alpha \in [0,1], \quad \forall\, t, u \in \mathbb{R}. \end{aligned} \tag{36}$$

Since $\xi(t) = t^2$ is convex, the following inequality holds:

$$[(1-\alpha)t + \alpha u]^2 \le (1-\alpha)t^2 + \alpha u^2, \quad \alpha \in [0,1]. \tag{37}$$

Since the matrix $\mathcal{Z}_2$ is positive semidefinite, (37) implies that

$$(((1-\alpha)t + \alpha u)^2)\mathcal{Z}_2 \preceq (1-\alpha)t^2\mathcal{Z}_2 + \alpha u^2\mathcal{Z}_2, \quad \alpha \in [0,1].$$

After adding to both sides $((1-\alpha)t + \alpha u)\mathcal{Z}_1 + \mathcal{Z}_0$, we get (36). Taking the expectation to both sides of (36) get

$$\begin{aligned} E[\mathcal{Z}((1-\alpha)t + \alpha u)] &\preceq E[(1-\alpha)\mathcal{Z}(t) + \alpha\mathcal{Z}(u)] \\ &= (1-\alpha)E[\mathcal{Z}(t)] + \alpha E[\mathcal{Z}(u)], \\ &\alpha \in [0,1]. \end{aligned}$$

Now, we have that

$$\begin{aligned} \eta((1-\alpha)t + \alpha u) &= \lambda_{\max}(E[\mathcal{Z}((1-\alpha)t + \alpha u)]) \\ &\le \lambda_{\max}((1-\alpha)E[\mathcal{Z}(t)] + \alpha E[\mathcal{Z}(u)]) \\ &\le (1-\alpha)\lambda_{\max}(E[\mathcal{Z}(t)]) \\ &\quad + \alpha\lambda_{\max}(E[\mathcal{Z}(u)]) \\ &= (1-\alpha)\eta(t) + \alpha\eta(u), \alpha \in [0,1]. \end{aligned}$$

The last inequality holds since $\lambda_{\max}(\cdot)$ is convex. This implies $\eta(t)$ is convex, and hence $\phi(W)$ is convex ■

D. Fully Connected Random Network: Closed Form Solution

To get some insight on how the optimal weights depend on the network parameters, we consider the impractical, but simple geometry of a random complete symmetric graph. For this example, the optimization problem (26) admits a closed form solution, while, in general, numerical optimization is needed to solve (26). Although not practical, this example provides insight on how the optimal weights depend on the network size N, the link formation probabilities, and the link formation spatial correlations. The supergraph is symmetric, fully connected, with N nodes and $M = N(N-1)/2$ undirected links. We assume that $N > 2$ and all the links have the same formation probability, i.e., that $\mathrm{Prob}(q_l = 1) = \pi_l = p, p \in (0,1], l = 1, \ldots, M$. We assume that the cross-variance between any pair of links i and j equals $[R_q]_{ij} = \beta p(1-p)$, where β is the correlation coefficient. The matrix R_q is given by

$$R_q = p(1-p)[(1-\beta)I + \beta 1 1^\top].$$

The eigenvalues of R_q are $\lambda_1(R_q) = p(1-p)(1+(M-1)\beta)$, and $\lambda_i(R_q) = p(1-p)(1-\beta) \geq 0$, $i = 2, \ldots, M$. The condition that $R_q \succeq 0$ implies that $\beta \geq -1/(M-1)$. Also, we have that

$$\begin{aligned}\beta :=& \frac{E[q_i q_j] - E[q_i]E[q_j]}{\sqrt{\mathrm{Var}(q_i)}\sqrt{\mathrm{Var}(q_j)}} \qquad (38)\\ =& \frac{\mathrm{Prob}(q_i = 1, q_j = 1) - p^2}{p(1-p)}\\ \geq& -\frac{p}{1-p}. \qquad (39)\end{aligned}$$

Thus, the range of β is restricted to

$$\max\left(-\frac{1}{M-1}, -\frac{p}{1-p}\right) \leq \beta \leq 1. \qquad (40)$$

Due to the problem symmetry, the optimal weights for all links are the same, say $w^\star$. The expressions for the optimal weight $w^\star$ and for the optimal convergence rate $\phi^\star$ can be obtained after careful manipulations and expressing the matrix $E[\mathcal{W}^2] - J$ explicitly in terms of p and β; then, it is easy to show that

$$w^\star = \frac{1}{Np + (1-p)(2 + \beta(N-2))} \qquad (41)$$

$$\phi^\star = 1 - \frac{1}{1 + \frac{1-p}{p}\left(\frac{2}{N}(1-\beta) + \beta\right)}. \qquad (42)$$

The optimal weight $w^\star$ decreases as β increases. This is also intuitive, since positive correlations imply that the links emanating from the same node tend to occur simultaneously, and thus the weight should be smaller. Similarly, negative correlations imply that the links emanating from the same node tend to occur exclusively, which results in larger weights. Finally, we observe that, in the uncorrelated case ($\beta = 0$), as N becomes very large, the optimal weight behaves as $1/(Np)$. Thus, for the uncorrelated links and large network, the optimal strategy (at least for this example) is to rescale the supergraph-optimal weight $1/N$ by its formation probability p. Finally, for fixed p and N, the fastest rate is achieved when β is as negative as possible.

E. Numerical Optimization: Subgradient Algorithm

We solve the optimization problem in (26) for generic networks by the subgradient algorithm, [31]. In this subsection, we consider spatially uncorrelated links, and we comment on extensions for spatially correlated links. Expressions for spatially correlated links are provided in Appendix B.

We recall that the function $\phi(W)$ is convex (proved in Section IV-C). It is nonsmooth because $\lambda_{\max}(\cdot)$ is nonsmooth. Let $H \in \mathbb{S}^N$ be the subgradient of the function $\phi(W)$. To derive the expression for the subgradient of $\phi(W)$, we use the variational interpretation of $\phi(W)$:

$$\phi(W) = \max_{v^\top v = 1} v^\top (E[\mathcal{W}^2] - J)v = \max_{v^\top v=1} f_v(W). \qquad (43)$$

By the subgradient calculus, a subgradient H_u of the function $f_u(W)$, for which the maximum of the optimization problem (43) is attained, is also a subgradient of $\phi(W)$ at point W (see, e.g., [31]). The maximum of $f_v(W)$ (with respect to v) is attained at $v = u$, where u is the eigenvector of the matrix $E[\mathcal{W}^2] - J$ that corresponds to its maximal eigenvalue, i.e., the maximal eigenvector. In our case, the function $f_u(W)$ is differentiable (quadratic function), and hence the subgradient of $f_u(W)$ (and also the subgradient of $\phi(W)$) is equal to the gradient of $f_u(W)$ [31]:

$$H_{ij} = \begin{cases} u^\top \frac{\partial(E[\mathcal{W}^2] - J)}{\partial W_{ij}} u, & \text{if } (i,j) \in E \\ 0, & \text{otherwise.} \end{cases} \qquad (44)$$

We compute for $(i,j) \in E$

$$\begin{aligned}H_{ij} =& u^\top \frac{\partial(\overline{W}^2 - J + R_C)}{\partial W_{ij}} u \qquad (45)\\ =& u^\top \Big(-2\overline{W} P_{ij}(e_i - e_j)(e_i - e_j)^\top\\ &+ 4W_{ij}P_{ij}(1-P_{ij})(e_i - e_j)(e_i - e_j)^\top\Big) u\\ =& 2P_{ij}(u_i - u_j)u^\top(\overline{W}_j - \overline{W}_i)\\ &+ 4P_{ij}(1-P_{ij})W_{ij}(u_i - u_j)^2. \qquad (46)\end{aligned}$$

The subgradient algorithm is given by algorithm 1. The stepsize α_k is nonnegative, diminishing, and nonsummable: $\lim_{k\to\infty} \alpha_k = 0$, $\sum_{k=1}^{\infty} \alpha_k = \infty$. We choose $\alpha_k = 1/\sqrt{k}$, $k = 1, 2, \ldots$, similarly as in [30].

Algorithm 1: Subgradient Algorithm

Set initial $W^{(1)} \in S_W$

Set $k = 1$

Repeat

  Compute a subgradient $H^{(k)}$ of ϕ at $W^{(k)}$, and set $W^{(k+1)} = W^{(k)} - \alpha_k H^{(k)}$

  $k := k + 1$

V. Weight Optimization: Asymmetric Random Links

We now address the weight optimization for asymmetric random networks. Section V-A and Section V-B introduce the optimization criterion and the corresponding weight optimization problem, respectively. Subsection V-C shows that this optimization problem is convex.

A. Optimization Criterion: Mean-Square Deviation Convergence Rate

Introduce now

$$\psi(W) := \lambda_{\max}(E[\mathcal{W}^\top(I-J)\mathcal{W}]). \tag{47}$$

Reference [11] shows that the MSdev satisfies the following equation:

$$\text{MSdev}(k+1) \leq \psi(W)\ \text{MSdev}(k). \tag{48}$$

Thus, if the quantity $\psi(W)$ is strictly less than one, then MSdev converges to zero asymptotically, with the worst case rate equal to $\psi(W)$. We remark that the condition (10) is not needed for (48) to hold, i.e., MSdev converges to zero even if condition (10) is not satisfied; this condition is needed only for (25) to hold, i.e., only to have the MSE to converge to zero.

B. Asymmetric Network: Weight Optimization Problem Formulation

In the case of asymmetric links, we propose to optimize the mean-square deviation convergence rate, i.e., to solve the following optimization problem:

$$\begin{aligned} &\text{minimize} && \psi(W) \\ &\text{subject to} && W \in S_W^{\text{asym}} \\ & && \sum_{i=1}^N P_{ij}W_{ij} = \sum_{i=1}^N P_{ji}W_{ji}, \ \ j=1,\dots,N. \end{aligned} \tag{49}$$

The constraints in the optimization problem (49) assure that, in expectation, condition (10) is satisfied, i.e., that

$$1^\top E[\mathcal{W}] = 1^\top. \tag{50}$$

If (50) is satisfied, then the consensus algorithm converges to the true average x_{avg} in expectation [11].

Equation (50) is a linear constraint with respect to the weights W_{ij}, and thus does not violate the convexity of the optimization problem (49). We emphasize that, in the case of asymmetric links, we do not assume the weights W_{ij} and W_{ji} to be equal. In Section VI-B, we show that allowing W_{ij} and W_{ji} to be different leads to better solutions (lower optimal value of (49)) in the case of asymmetric networks.

C. Convexity of the Weight Optimization Problem

We show that the function $\psi(W)$ is convex. We remark that [11] shows that the function is convex, when all the weights W_{ij} are equal to g. We show here that this function is convex even when the weights are different.

Lemma 4 (Convexity of $\psi(W)$): The function $\psi: S_W^{\text{asym}} \to \mathbb{R}_+$ is convex.

Proof: The proof is very similar to the proof of Lemma 3. The proof starts with introducing W as in (29) and with introducing $\mathcal{W}(t)$ as in (30). The difference is that, instead of considering the matrix $\mathcal{W}^2 - J$, we consider now the matrix $\mathcal{W}^\top(I-J)\mathcal{W}$. In the proof of Lemma 3, we introduced the auxiliary function $\eta(t)$ given by (31); here, we introduce the auxiliary function $\kappa(t)$, given by

$$\kappa(t) = \lambda_{\max}(\mathcal{W}(t)^\top(I-J)\mathcal{W}(t)) \tag{51}$$

and show that $\psi(W)$ is convex by proving that $\kappa(t)$ is convex. Then, we proceed as in the proof of Lemma 3. In (35) the matrix $\mathcal{Z}_2$ becomes $\mathcal{Z}_2 := \mathcal{Y}^\top(I-J)\mathcal{Y}$. The random matrix $\mathcal{Z}_2$ is obviously positive semidefinite. The proof then proceeds as in Lemma 3. ∎

VI. Simulations

We demonstrate the effectiveness of our approach with a comprehensive set of simulations. These simulations cover both examples of asymmetric and symmetric networks and both networks with random link failures and with randomized protocols. In particular, we consider the following two standard sets of experiments with random networks: 1) spatially correlated link failures and symmetric links and 2) randomized protocols, in particular, the broadcast gossip algorithm [11]. With respect to the first set, we consider correlated link failures with two types of correlation structure. We are particularly interested in studying the dependence of the performance and of the gains on the size of the network N and on the link correlation structure.

In all these experiments, we consider geometric random graphs. Nodes communicate among themselves if the nodes are within their radius of communication, r. The nodes are uniformly distributed on a unit square. The number of nodes is $N=100$, and the average relative degree $(=2M/N(N-1))$ is 15% In Section VI-A, the random instantiations of the networks are undirected; in Section VI-B, the random instantiations of the networks are directed.

In the first set of experiments with correlated link failures, the link formation probabilities P_{ij} are chosen such that they decay quadratically with the distance δ_{ij} between the nodes i and j:

$$P_{ij} = 1 - k\left(\frac{\delta_{ij}}{r}\right)^2 \tag{52}$$

where we choose $k=0.7$. We see that, with (52), a link will be active with high probability if the nodes are close $(\delta_{ij} \simeq 0)$, while the link will be down with probability at most 0.7, if the nodes are apart by r.

We recall that we refer to our weight design, i.e., to the solutions of the weight optimization problems (26), (49), as PBWs. We study the performance of PBW, comparing it with the standard weight choices available in the literature: Section VI-A, we compare it with the MWs, discussed in [30], and the SGBWs. The SGBW are the optimal (nonnegative) weights designed for a static (nonrandom) graph G, which are then applied to a random network when the underlying supergraph is G. This is the strategy used in [24]. For asymmetric links (and for asymmetric weights $W_{ij} \neq W_{ji}$), in Section VI-B, we compare PBW with the optimal weight choice in [11] for broadcast gossip that considers all the weights to be equal.

In the first set of experiments in Section VI-A, we quantify the performance gain of PBW over SGBW and MW by the gains

$$\Gamma_s^\tau = \frac{\tau_{\rm SGBW}}{\tau_{\rm PBW}} \tag{53}$$

where τ is a time constant defined as

$$\tau = \frac{1}{0.5 \ln \phi(W)}. \tag{54}$$

We also compare PBW with SGBW and MW with the following measure:

$$\Gamma_s^\eta = \frac{\eta_{\rm SGBW}}{\eta_{\rm PBW}} \tag{55}$$

$$\Gamma_m^\eta = \frac{\eta_{\rm MW}}{\eta_{\rm PBW}} \tag{56}$$

where η is the asymptotic time constant defined by

$$\eta = \frac{1}{|\gamma|} \tag{57}$$

$$\gamma = \lim_{k\to\infty} \frac{1}{k} \ln\left(\frac{\|e(k)\|}{\|e(0)\|}\right). \tag{58}$$

Reference [24] shows that for random networks η is an almost sure constant and τ is an upper bound on η.

Section VI-A and Subsection VI-B will provide further details on the experiments.

A. Symmetric Links: Random Networks With Correlated Link Failures

To completely define the probability distribution of the random link vector $q \in \mathbb{R}^M$, we must assign probability to each of the 2^M possible realizations of q, $q = (\alpha_1, \ldots, \alpha_M)^\top$, $\alpha_i \in \{0,1\}$. Since in networks of practical interest M may be very large, of order 1000 or larger, specifying the complete distribution of the vector q is most likely infeasible. Hence, we work with the second moment description and specify only the first two moments of its distribution, the mean and the covariance, π and R_q. Without loss of generality, order the links so that $\pi_1 \le \pi_2 \le \cdots \le \pi_M$.

Lemma 5: The mean and the variance (π, R_q) of a Bernoulli random vector satisfy equations (59)–(61) (see the bottom of the page).

Proof: Equations (59) and (60) must hold because π_l's are probabilities and R_q is a covariance matrix. Recall that

$$\begin{aligned}[R_q]_{ij} &= E[q_i q_j] - E[q_i]E[q_j] \\ &= \mathrm{Prob}(q_i = 1, q_j = 1) - \pi_i \pi_j.\end{aligned} \tag{62}$$

To prove the lower bound in (61), observe that

$$\begin{aligned}\mathrm{Prob}(q_i=1, q_j=1) &= \mathrm{Prob}(q_i = 1) + \mathrm{Prob}(q_j = 1) \\ &\quad - \mathrm{Prob}(\{q_i = 1\} \text{ or } \{q_j = 1\}) \\ &= \pi_i + \pi_j - \mathrm{Prob}(\{q_i = 1\} \text{ or } \{q_j = 1\}) \\ &\ge \pi_i + \pi_j - 1.\end{aligned} \tag{63}$$

In view of the fact that $\mathrm{Prob}(q_i = 1, q_j = 1) \ge 0$, (63), and (62), the proof for the lower bound in (61) follows. The upper bound in (61) holds because $\mathrm{Prob}(q_i = 1, q_j = 1) \le \pi_i$, $i < j$ and (62). ■

If we choose a pair (π, R_q) that satisfies (59), (60), and (61), one cannot guarantee that (π, R_q) is a valid pair, in the sense that there exists a probability distribution on q with its first and second moments being equal to (π, R_q) [32]. Furthermore, if (π, R_q) is given, to simulate binary random variables with the marginal probabilities and correlations equal to (π, R_q) is challenging. These questions have been studied; see [32] and [33]. We use the results in [32] and [33] to generate our correlation models. In particular, we use the result that $\overline{R} = [\overline{R}_{ij}]$ (see (61)) is a valid correlation structure for any π, [33]. We simulate the correlated links by the method proposed in [32]; this method handles a wide range of different correlation structures and has a small computational cost.

Link correlation structures: We consider two different correlation structures for any pair of links i and j in the supergraph

$$[R_q]_{ij} = c_1 \overline{R}_{ij} \tag{64}$$

$$[R_q]_{ij} = c_2 \theta^{\kappa_{ij}} \overline{R}_{ij} \tag{65}$$

where $c_1 \in (0,1]$, $\theta \in (0,1)$ and $c_2 \in (0,1]$ are parameters, and κ_{ij} is the distance between links i and j defined as the length of the shortest path that connects them in the supergraph.

The correlation structure (64) assumes that the correlation between any pair of links is a fraction of the maximal possible correlation, for the given π (see (61) to recall $\overline{R}_{ij}$). Reference [33] constructs a method for generating the correlation structure (64).

The correlation structure (65) assumes that the correlation between the links decays geometrically with this distance. In our simulations, we set $\theta = 0.95$, and find the maximal c_2, such that the resulting correlation structure can be simulated by the method in [32]. For all the networks that we simulated in the paper, c_2 is between 0.09 and 0.11.

Results: We want to address the following two questions: 1) What is the performance gain (Γ_s^η, Γ_m^η in (55) and (56)) of PBW over SGBW and MW; and 2) How does this gain scale with the network size, i.e., the number of nodes N?

Performance gain of PBW over SGBW and MW: We consider question 1) for both correlation structures (64), (65). We generate 20 instantiations of our standard supergraphs (with 100 nodes each and approximately the same average relative degree, equal to 15%). Then, for each supergraph, we generate formation probabilities according to rule (52). For each supergraph with the given formation probabilities, we generate two link correlation structures, (64) and (65). We evaluate the convergence rate ϕ_j given by (25), time constants η_j given by (57), and τ_j, given by (54), and the performance gains $[\Gamma_s^\eta]_j$, $[\Gamma_m^\eta]_j$ for each supergraph ($j = 1, \ldots, 20$). We compute the mean

$$0 \le \pi_i \le 1, \quad i = 1, \ldots, N \tag{59}$$

$$R_q \succeq 0 \tag{60}$$

$$\max(-\pi_i\pi_j, \pi_i + \pi_j - 1 - \pi_i\pi_j) \le [R_q]_{ij} \le \pi_i(1-\pi_j) = \overline{R}_{ij}, \quad i < j. \tag{61}$$

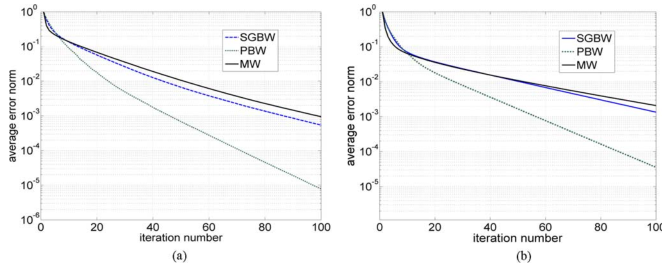

Fig. 1. Average error norm versus iteration number. Left: Correlation structure (64); right: correlation structure (65).

TABLE I
CORRELATION STRUCTURE (64): AVERAGE $\overline{(\cdot)}$, MAXIMAL $(\cdot)^+$, AND MINIMAL $(\cdot)^-$ VALUES OF THE MSE CONVERGENCE RATE ϕ (23), AND CORRESPONDING TIME CONSTANTS τ (54) AND η (57), FOR 20 GENERATED SUPERGRAPHS

	SGBW	PBW	MW
$\overline{\phi}$	0.91	0.87	
ϕ^+	0.95	0.92	
ϕ^-	0.89	0.83	
$\overline{\tau}$	22.7	15.4	
τ^+	28	19	
τ^-	20	14	
$\overline{\eta}$	20	13	29
η^+	25	16	38
η^-	19	12	27

$\overline{\phi}$, the maximum ϕ^+ and the minimum ϕ^- from the list $\{\phi_j\}$, $j = 1, \ldots, 20$ (and similarly for $\{\eta_j\}$ and $\{\tau_j\}$, $j = 1, \ldots, 20$). Results for the correlation structure (64) are given in Table I and for the correlation structure (65), in Table II. The performance gains Γ_s, Γ_m, for both correlation structures are in Table III. In addition, Fig. 1 depicts the averaged error norm over 100 sample paths. We can see that the PBW outperform the SGBW and the MW for both correlation structures (64) and (65). For example, for the correlation (64), the PBW take less than 40 iterations to achieve 0.2% precision, while the SGBW take more than 70, and the MW take more than 80 iterations. For correlation (65), to achieve 0.2% precision, the PBW take about 47 iterations, while the SGBW and the MW take more than 90 and 100 iterations, respectively. The average performance gain of PBW over MW is larger than the performance gain over SGBW, for both (64) and (65). The gain over SGBW, Γ_s, is significant, being 1.54 for (64) and 1.73 for (65). The gain with the correlation structure (65) is larger than the gain with (64), suggesting that larger gain over SGBW is achieved with smaller correlations. This is intuitive, since large positive correlations imply that the random links tend to occur simultaneously, i.e., in a certain sense random network realizations are more similar to the underlying supergraph.

Notice that the networks with R_q as in (65) achieve faster rate than for (64) (having at the same time similar supergraphs and formation probabilities). This is in accordance with the analytical studies in Section IV-D that suggest that faster rates can be achieved for smaller (or negative correlations) if G and π are fixed.

Performance Gain of PBW Over SGBW as a Function of the Network Size: To answer question 2), we generate the supergraphs with N ranging from 30 up to 140 (one supergraph for each N), keeping the average relative degree of the supergraph almost the same (18%); it varies for different N's less

TABLE II
CORRELATION STRUCTURE (65): AVERAGE $\overline{(\cdot)}$, MAXIMAL $(\cdot)^+$, AND MINIMAL $(\cdot)^-$ VALUES OF THE MSE CONVERGENCE RATE ϕ (23), AND CORRESPONDING TIME CONSTANTS τ (54) AND η (57), FOR 20 GENERATED SUPERGRAPHS

	SGBW	PBW	MW
$\overline{\phi}$	0.92	0.86	
ϕ^+	0.94	0.90	
ϕ^-	0.91	0.84	
$\overline{\tau}$	25.5	14.3	
τ^+	34	19	
τ^-	21	12	
$\overline{\eta}$	20	11.5	24.4
η^+	23	14	29
η^-	16	9	19

TABLE III
AVERAGE $\overline{(\cdot)}$, MAXIMAL $(\cdot)^+$, AND MINIMAL $(\cdot)^-$ PERFORMANCE GAINS Γ_s^η AND Γ_m^η (55) FOR THE TWO CORRELATION STRUCTURES (64) AND (65) FOR 20 GENERATED SUPERGRAPHS

	Correlation (64)	Correlation (65)
$\overline{(\Gamma_s^\eta)}$	1.54	1.73
$(\Gamma_s^\eta)^+$	1.66	1.91
$(\Gamma_s^\eta)^-$	1.46	1.58
$\overline{(\Gamma_m^\eta)}$	2.22	2.11
$(\Gamma_m^\eta)^+$	2.42	2.45
$(\Gamma_m^\eta)^-$	2.07	1.92

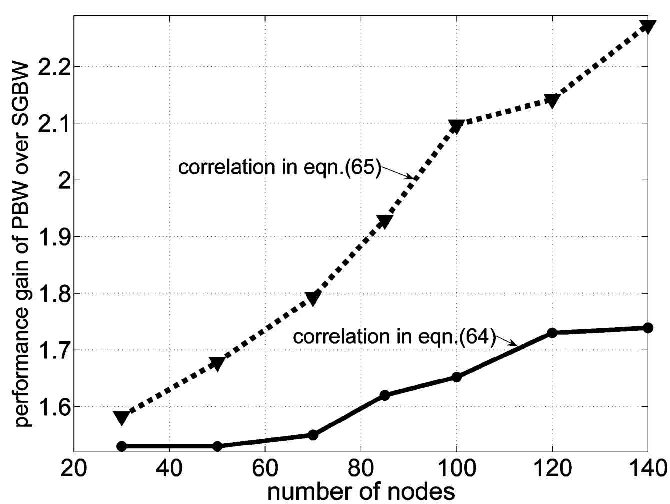

Fig. 2. Performance gain of PBW over SGBW (Γ_s^η, (55)) as a function of the number of nodes in the network.

than 2 percent of its nominal value. Quantities Γ_s are estimated based on $K = 20$ simulation runs; we previously verified that $K = 20$ gives a very good estimate for Γ_s by comparing the estimates for $K = 20$ with the estimates for larger K (namely, $K = 100$). Again, PBW performs better than MW, so we focus on the dependence of Γ_s on N, since it is more critical.

Fig. 2 plots Γ_s versus N, for the two correlation structures. The gain Γ_s increases with N for both (65) and (64). We notice that the gains for $N = 100$ are larger in Fig. 2 than in Table III due to larger average relative degree in Fig. 2.

B. Broadcast Gossip Algorithm [11]: Asymmetric Random Links

In Section III-B, we demonstrated the effectiveness of our approach in networks with random symmetric link failures. This

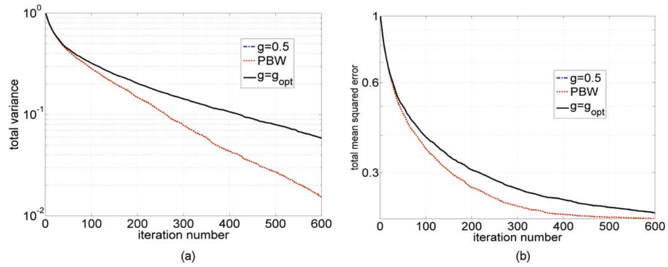

Fig. 3. Broadcast gossip algorithm with different weight choices. Left: Total variance; right: total mean-square error.

section demonstrates the validity of our approach in randomized protocols with asymmetric links. We study the broadcast gossip algorithm [11]. Although the optimization problem (49) is convex for generic spatially correlated directed random links, we pursue here numerical optimization of the broadcast gossip algorithm proposed in [11], where, at each time step, node i is selected at random, with probability $1/N$. Node i then broadcasts its state to all its neighbors within its wireless range. The neighbors then update their state by performing the weighted average of the received state with their own state. The nodes outside the set Ω_i and the node i itself keep their previous state unchanged. The broadcast gossip algorithm is well suited for WSN applications, since it exploits the broadcast nature of wireless media and avoids bidirectional communication [11].

Reference [11] shows that, in broadcast gossiping, all the nodes converge a.s. to a common random value c with mean x_{avg} and bounded mean-square error. Reference [11] studies the case when the weights $W_{ij} = g, \forall\,(i,j) \in E$, and finds the optimal $g = g^\star$ that optimizes the mean-square deviation MSdev (see (49)). We optimize the same objective function (see (49)) as in [11], but allowing different weights for different directed links. We detail the numerical optimization for the broadcast gossip in Appendix C. We consider again the supergraph G from our standard experiment with $N = 100$ and average relative degree 15%. For the broadcast gossip, we compare the performance of PBW with 1) the optimal equal weights in [11] with $W_{ij} = g^\star, (i,j) \in E$; and 2) broadcast gossip with $W_{ij} = 0.5$, $(i,j) \in E$ (as in [10]).

Fig. 3 (left) plots the consensus MSdev for the three different weight choices. The decay of MSdev is much faster for the PBW than for $W_{ij} = 0.5, \forall\,(i,j)$ and $W_{ij} = g^\star, \forall\,(i,j)$. For example, the MSdev falls below 10% after 260 iterations for PBW (i.e., 260 broadcast transmissions); broadcast gossip with $W_{ij} = g^\star$ and $W_{ij} = 0.5$ takes 420 transmissions to achieve the same precision. This is to be expected, since PBW has many more degrees of freedom to optimize than the broadcast gossip in [11] with all equal weights $W_{ij} = g^\star$. Fig. 3 (right) plots the MSE, i.e., the deviation of the true average x_{avg}, for the three weight choices. PBW shows faster MSE decay than the broadcast gossip with $W_{ij} = g^\star$ and $W_{ij} = 0.5$. The weights provided by PBW are different among themselves, varying from 0.3 to 0.95. The weights W_{ij} and W_{ji} are also different, where the maximal difference between W_{ij} and W_{ji}, $(i,j) \in E$, is 0.6. Thus, for the case of directed random networks, faster convergence is obtained with an asymmetric matrix W.

VII. Conclusion

In this paper, we studied the optimization of the weights for the consensus algorithm under random topology spatially correlated links. We considered both networks with random link failures and randomized algorithms; from the weights optimization point of view, both fit into the same framework. We showed that, for symmetric random links, optimizing the MSE convergence rate is a convex optimization problem, and, for asymmetric links, optimizing the mean-square deviation from the current average state is also a convex optimization problem. We illustrated with simulations that the PBWs presented in this paper outperform previously proposed weight design strategies that do not use the statistics of the network randomness. The simulations also show that using the link quality estimates and the link correlations for designing the weights significantly improves the convergence speed, typically reducing the time to consensus by one third to one half, compared to choices previously proposed in the literature.

Appendix A

Proof of Lemma 1 (a Sketch): Equation (18) follows from the expectation of (3). To prove the remaining of the Lemma, we find $\mathcal{W}^2$, $\overline{W}^2$, and the expectation of $\mathcal{W}^2$. We obtain successively

$$\begin{aligned}
\mathcal{W}^2 &= (W \odot \mathcal{A} + I - \text{Diag}\,(W\mathcal{A}))^2 \\
&= (W \odot \mathcal{A})^2 + \text{Diag}^2(W\mathcal{A}) + I + 2W \odot \mathcal{A} \\
&\quad - 2\,\text{Diag}\,(W\mathcal{A}) - (W \odot \mathcal{A})\,\text{Diag}\,(W\mathcal{A}) \\
&\quad - \text{Diag}\,(W\mathcal{A})(W \odot \mathcal{A}) \\
\overline{W}^2 &= (W \odot P)^2 + \text{Diag}^2(WP) \\
&\quad + I + 2W \odot P - 2\,\text{Diag}\,(WP) \\
&\quad - [(W \odot P)\,\text{Diag}\,(WP) + \text{Diag}\,(WP)(W \odot P)] \\
E[\mathcal{W}^2] &= E[(W \odot \mathcal{A})^2] + E[\text{Diag}^2(W\mathcal{A})] + I + 2W \odot P \\
&\quad - 2\,\text{Diag}\,(WP) - E[(W \odot \mathcal{A})\,\text{Diag}\,(W\mathcal{A}) \\
&\qquad + \text{Diag}\,(W\mathcal{A})(W \odot \mathcal{A})].
\end{aligned}$$

We will next show the following three equalities:

$$E[(W \odot \mathcal{A})^2] = (W \odot P)^2 + {W_C}^\top \{R_A \odot (11^\top \otimes I)\} W_C \tag{66}$$

$$E[\text{Diag}^2(W\mathcal{A})] = \text{Diag}^2(W\,P) + {W_C}^\top \{R_A \odot (I \otimes 11^\top)\} W_C \tag{67}$$

$$\begin{aligned}
E[(W \odot \mathcal{A})\,\text{Diag}\,(W\mathcal{A}) &+ \text{Diag}\,(W\mathcal{A})(W \odot \mathcal{A})] \\
&= (W \odot P)\,\text{Diag}\,(WP) + \text{Diag}\,(WP)(W \odot P) \\
&\quad - {W_C}^\top \{R_A \odot B\} W_C.
\end{aligned} \tag{68}$$

First, consider (66) and find $E\left[(W \odot \mathcal{A})^2\right]$. Algebraic manipulations allow to write $(W \odot \mathcal{A})^2$ as follows:

$$\begin{aligned}
(W \odot \mathcal{A})^2 &= {W_C}^\top \{\mathcal{A}_2 \odot (11^\top \otimes I)\} W_C, \\
\mathcal{A}_2 &= \text{Vec}(\mathcal{A})\text{Vec}^\top(\mathcal{A}).
\end{aligned} \tag{69}$$

To compute the expectation of (69), we need $E[\mathcal{A}_2]$ that can be written as

$$E[\mathcal{A}_2] = P_2 + R_A, \text{with } P_2 = \text{Vec}(P)\text{Vec}^\top(P).$$

Equation (66) follows, realizing that

$$W_C{}^\top\{P_2 \odot (11^\top \otimes I)\}W_C = (W \odot P)^2.$$

Now consider (67) and (68). After algebraic manipulations, it can be shown that

$$\begin{aligned}&\mathrm{Diag}^2(W\mathcal{A})\\&\quad = W_C{}^\top\{\mathcal{A}_2 \odot (I \otimes 11^\top)\}W_C\\&(W \odot \mathcal{A})\mathrm{Diag}\,(W\mathcal{A}) + \mathrm{Diag}\,(W\mathcal{A})(W \odot \mathcal{A})\\&\quad = W_C{}^\top\{\mathcal{A}_2 \odot B\}W_C.\end{aligned}$$

Computing the expectations in the last two equations leads to (67) and (68).

Using equalities (66), (67), and (68) and comparing the expressions for $\overline{W}^2$ and $E[\mathcal{W}^2]$ leads to

$$\begin{aligned}R_C &= E[\mathcal{W}^2] - \overline{W}^2\\&= W_C{}^\top\{R_A \odot (I \otimes 11^\top + 11^\top \otimes I - B)\}W_C. \quad (70)\end{aligned}$$

This completes the proof of Lemma 1.

Appendix B
Subgradient Step Calculation for the Case of Spatially Correlated Links

To compute the subgradient H, from (44) and (45) we consider the computation of $E[\mathcal{W}^2 - J] = \overline{W}^2 - J + R_C$. Matrix $\overline{W}^2 - J$ is computed in the same way as for the uncorrelated case. To compute R_C, from (70), partition the matrix R_A into $N \times N$ blocks:

$$R_A = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1N}\\ R_{21} & R_{22} & \dots & R_{2N}\\ \vdots & \dots & \dots & \vdots\\ R_{N1} & R_{N2} & \dots & R_{NN}\end{pmatrix}.$$

Denote by d_{ij}, by c_{ij}^l, and by r_{ij}^l the diagonal, the lth column, and the lth row of the block R_{ij}. It can be shown that the matrix R_C can be computed as follows:

$$\begin{aligned}[R_C]_{ij} &= W_i^\top(d_{ij} \odot W_j) - W_{ij}(W_i^\top c_{ij}^i + W_j^\top r_{ij}^j), \quad i \neq j\\ [R_C]_{ii} &= W_i^\top(d_{ii} \odot W_i) + W_i^\top R_{ii} W_i.\end{aligned}$$

Denote by $R_A(:,k)$ the kth column of the matrix R_A and by

$$\begin{aligned}k_1 &= \left(e_j^\top \otimes I_N\right) R_A(:,(i-1)N+j)\\ k_2 &= \left(e_i^\top \otimes I_N\right) R_A(:,(j-1)N+i)\\ k_3 &= \left(e_i^\top \otimes I_N\right) R_A(:,(i-1)N+j)\\ k_4 &= \left(e_j^\top \otimes I_N\right) R_A(:,(j-1)N+i).\end{aligned}$$

Quantities k_1, k_2, k_3 and k_4 depend on (i,j) but for the sake of the notation simplicity indexes are omitted. It can be shown that the computation of H_{ij}, $(i,j) \in E$ boils down to

$$\begin{aligned}H_{ij} &= 2u_i^2 W_i^T c_{ii}^j + 2u_j^2 W_j^\top c_{jj}^i\\ &\quad + 2u_i W_j^\top(u \odot k_1) + 2u_j W_i^\top(u \odot k_2)\\ &\quad - 2u_i u_j W_j^\top c_{ji}^j - 2u_i u_j W_i^\top c_{ij}^i - 2u_i W_i^T(u \odot k_3)\\ &\quad - 2u_j W_j^\top(u \odot k_4) + 2P_{ij}(u_i - u_j)u^\top(\overline{W}_j - \overline{W}_i).\end{aligned}$$

Appendix C
Numerical Optimization for the Broadcast Gossip Algorithm

With broadcast gossip, the matrix $\mathcal{W}(k)$ can take N different realizations, corresponding to the broadcast cycles of each of the N sensors. We denote these realizations by $\mathcal{W}^{(i)}$, where i indexes the broadcasting node. We can write the random realization of the broadcast gossip matrix $\mathcal{W}^{(i)}$, $i = 1, \dots, N$, as follows:

$$\mathcal{W}^{(i)}(k) = W \odot \mathcal{A}^{(i)}(k) + I - \mathrm{Diag}(W\mathcal{A}^{(i)}(k)^\top) \quad (71)$$

where $\mathcal{A}_{li}^{(i)}(k) = 1$, if $l \in \Omega_i$. Other entries of $\mathcal{A}^{(i)}(k)$ are zero.

Similarly as in Appendix A, we can arrive at the expressions for $E[\mathcal{W}^\top\mathcal{W}] := E[\mathcal{W}^\top(k)\mathcal{W}(k)]$ and for $E\left[\mathcal{W}^\top J\mathcal{W}\right] := E\left[\mathcal{W}^\top(k)J\mathcal{W}(k)\right]$, for all k. We remark that the matrix W needs not to be symmetric for the broadcast gossip and that $W_{ij} = 0$, if $(i,j) \notin E$.

$$\begin{aligned}E\left[(\mathcal{W}^\top\mathcal{W})_{ii}\right] &= \frac{1}{N}\sum_{l=1,l\neq i}^{N} W_{li}^2 + \frac{1}{N}\sum_{l=1,l\neq i}^{N}(1-W_{il})^2\\ E\left[(\mathcal{W}^\top\mathcal{W})_{ij}\right] &= \frac{1}{N}W_{ij}(1-W_{ij}) + \frac{1}{N}W_{ji}(1-W_{ji}),\ i \neq j\\ E\left[(\mathcal{W}^\top J\mathcal{W})_{ii}\right] &= \frac{1}{N^2}\left(1+\sum_{l\neq i} W_{li}\right)^2 + \frac{1}{N^2}\sum_{l=1,l\neq i}^{N}(1-W_{il})^2\\ \left[E[\mathcal{W}^\top J\mathcal{W})_{ij}\right] &= \frac{1}{N^2}(1-W_{ji})\left(1+\sum_{l=1,l\neq i}^{N} W_{li}\right)\\ &\quad + \frac{1}{N^2}(1-W_{ij})\left(1+\sum_{l=1,l\neq j}^{N} W_{lj}\right)\\ &\quad + \frac{1}{N^2}\sum_{l=1,l\neq i,l\neq j}^{N}(1-W_{il})(1-W_{jl}),\ \ i \neq j.\end{aligned}$$

Denote by $W^{\mathrm{BG}} := E[\mathcal{W}^\top\mathcal{W}] - E[\mathcal{W}^\top J\mathcal{W}]$ and recall the definition of the MSdev rate $\psi(W)$ given by (47). We have that $\psi(W) = \lambda_{\max}(W^{\mathrm{BG}})$. We proceed with the calculation of the subgradient of $\psi(W)$ similarly as in Section IV-E. The subgradient entry H_{ij}^{BG}, $(i,j) \in E$, of the function $\psi(W)$ is given by ($H_{ij} = 0$, if $(i,j) \notin E$):

$$H_{ij}^{\mathbf{BG}} = q^\top\left(\frac{\partial}{\partial W_{i,j}}W^{\mathrm{BG}}\right)q$$

where q is the eigenvector associated with the maximal eigenvalue of the matrix W^{BG}. Finally, the $N \times N$ partial derivative of the matrix W^{BG} with respect to the weight W_{ij}, $(i,j) \in E$, has the entries given by the following set of equations:

$$\begin{aligned}\frac{\partial}{\partial W_{i,j}}W_{i,i}^{\mathrm{BG}} &= -2\frac{N-1}{N}(1-W_{i,j})\\ \frac{\partial}{\partial W_{i,j}}W_{j,j}^{\mathrm{BG}} &= \frac{2}{N}W_{i,j} - \frac{2}{N^2}\left(1+\sum_{l=1,l\neq j}^{N} W_{l,j}\right)\\ \frac{\partial}{\partial W_{i,j}}W_{i,j}^{\mathrm{BG}} &= \frac{1}{N}(1-2W_{i,j})\end{aligned}$$

$$-\frac{1}{N^2}\left(-2-\sum_{l=1,l\neq j}^{N} W_{l,j}+W_{i,j}\right)$$

$$\frac{\partial}{\partial W_{i,j}} W_{i,l}^{\mathrm{BG}} = \frac{1}{N^2}(1-W_{l,j}),\ l\neq i,\ l\neq j,$$

$$\frac{\partial}{\partial W_{i,j}} W_{l,j}^{\mathrm{BG}} = \frac{\partial}{\partial W_{i,j}} W_{j,l}^{\mathrm{BG}} = -\frac{1}{N^2}(1-W_{l,j}),\ l\neq i,\ l\neq j$$

$$\frac{\partial}{\partial W_{i,j}} W_{l,m}^{\mathrm{BG}} = 0,\ \text{otherwise}.$$

References

[1] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[2] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. disseration, Massachusetts Inst. Technol., Cambridge, MA, 1984.

[3] S. Kar and J. M. F. Moura, "Ramanujan topologies for decision making in sensor networks (Invited Paper)," in *44th Annu. Allerton Conf. Commun., Control, Comput.*, Allerton, IL, Monticello, Sep. 2006, pp. 145–150.

[4] S. Kar, S. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Trans. Signal Process.*, vol. 56, pp. 2609–2613, Jun. 2008.

[5] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Conf. Inf. Processing Sensor Networks (IPSN)*, Los Angeles, CA, Apr. 2005, pp. 63–70.

[6] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[7] S. Kar and J. M. F. Moura, *Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication*, pp. 51–51, Aug. 2008, submitted for publication.

[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, pp. 988–1001, Jun. 2003.

[9] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, pp. 215–233, Jan. 2007.

[10] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508–2530, Jun. 2006.

[11] T. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.

[12] V. Blondel, J. Hendrickx, A. Olshevsky, and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proc. 44th IEEE Conf. Decision Control (CDC)*, Seville, Spain, Dec. 2005, pp. 2996–3000.

[13] 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, pp. 1520–1533, Sep. 2004.

[14] L. Xiao, S. Boyd, and S. Lall, "Distributed average consensus with time-varying metropolis weights," 2006 [Online]. Available: http://www.stanford.edu/boyd/papers/avg_metropolis

[15] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, pp. 225–230, Jan. 2010.

[16] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with random link failures," in *Proc.. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Pacific Grove, CA, Apr. 2007, vol. 2, pp. II–1013.

[17] Y. Hatano and M. Mesbahi, "Agreement over random networks," in *'Proc. 43rd IEEE Conf. Decision Control (CDC)*, Paradise Island, Bahamas, Dec. 2004, vol. 2, pp. 2010–2015.

[18] M. Porfiri and D. Stilwell, "Stochastic consensus over weighted directed networks," in *Proc. Amer. Control Conf.*, New York, Jul. 2007, pp. 1425–1430.

[19] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pt. 1, pp. 1383–1400, Mar. 2010.

[20] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[21] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3315–3326, Jul. 2008.

[22] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *Proc. 1st Int. ACM Conf. Embedded Networked Sensor Systems*, Los Angeles, CA, Nov. 2003, pp. 1–13.

[23] A. Cerpa, J. Wong, L. Kuang, M. Potkonjak, and D. Estrin, "Statistical model of lossy links in wireless sensor networks," in *Proc. 4th Int. Symp. Information Processing in Sensor Networks (IPSN)*, Los Angeles, CA, Apr. 2005, pp. 81–88.

[24] P. Denantes, F. Benezit, P. Thiran, and M. Vetterli, "Which distributed averaging algorithm should I choose for my sensor network," in *Proc. 7th IEEE Conf. Computer Communications (INFOCOM)*, Phoenix, AZ, Mar. 2008, pp. 986–994.

[25] W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[26] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Rev.*, vol. 46, no. 4, pp. 667–689, Dec. 2004.

[27] A. F. Karr, *Probability Theory*, ser. Springer Texts in Statistics. New York: Springer-Verlag, 1993.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[29] A. Tahbaz-Salehi and A. Jadbabaie, "On consensus in random networks," in *Proc. 44th Annu. Allerton Conf. Communication, Control, omputing*, Allerton House, IL, Sep. 2006, pp. 1315–1321.

[30] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[31] H. U. H. Urruty and C. Lemarechal, *Convex Analysis And Minimization Algorithms: Part 1: Fundamentals*. Berlin, Germany: Springer-Verlag, 1993.

[32] B. Quaqish, "A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations," *Biometrika*, vol. 90, no. 2, pp. 455–463, Jun. 2003.

[33] S. D. Oman and D. M. Zucker, "Modelling and generating correlated binary variables," *Biometrika*, vol. 88, no. 1, pp. 287–290, Mar. 2001.

Dušan Jakovetić (S'10) received the Dipl.Ing. degree from the School of Electrical Engineering, University of Belgrade, in 2007. Currently, he is working towards the Ph.D. degree in electrical and computer engineering within the joint program of Carnegie Mellon University, Pittsburgh, PA, and Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal.

His research interests include consensus algorithms in sensor networks and distributed optimization in the context of sensor networks.

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

Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering, IST. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon, Portugal. His current research interests are in the area of optimization, sensor networks and signal processing on manifolds.

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 of Electrical and Computer Engineering and, by courtesy, of BioMedical Engineering, at Carnegie Mellon University (CMU). He was on the faculty at IST, has held visiting faculty appointments at MIT, and was a visiting research scholar at the University of Southern California. He is a founding co-director of the Center for Sensed Critical Infrastructures Research (CenSCIR) and manages a large education and research program between CMU and Portugal, www.icti.cmu.edu. His research interests include statistical and algebraic signal processing, image, bioimaging, and video processing, and digital communications. He has published over 400 technical journal and conference papers, is the co-editor of two books, holds six patents, and has given numerous invited seminars at international conferences, U.S. and European universities, and industrial and government Laboratories.

Dr. Moura was the President (2008–2009) of the IEEE Signal Processing Society (SPS). He was 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 was on the steering and technical committees of several conferences. He is 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 in 2000 the IEEE Millennium Medal. In 2007, he received the CMU's College of Engineering Outstanding Research Award, in 2008 the Philip L. Dowd Fellowship Award for Contributions to Engineering Education, and in 2010 he was elected University Professor. He is affiliated with several IEEE societies, Sigma Xi, AMS, IMS, and SIAM.