Distributed Algorithms for Consensus and Coordination in the Presence of Packet-Dropping Communication Links

Part I: Statistical Moments Analysis Approach

Alejandro D. Domínguez-García, *Member, IEEE*
Christoforos N. Hadjicostis, *Senior Member, IEEE*
Nitin H. Vaidya, *Fellow, IEEE*

September 28, 2011

Abstract

This two-part paper discusses robustification methodologies for linear-iterative distributed algorithms for consensus and coordination problems in multicomponent systems, in which unreliable communication links may drop packets. We consider a setup where communication links between components can be asymmetric (i.e., component j might be able to send information to component i, but not necessarily vice-versa), so that the information exchange between components in the system is in general described by a directed graph that is assumed to be strongly connected. In the absence of communication link failures, each component i maintains two auxiliary variables and updates each of their values to be a linear combination of their corresponding previous values and the corresponding previous values of neighboring components (i.e., components that send information to node i). By appropriately initializing these two (decoupled) iterations, the system components can asymptotically calculate variables of interest in a distributed fashion; in particular, the average of the initial conditions can be calculated as a function that involves the ratio of these two auxiliary variables. The focus of this paper to robustify this double-iteration algorithm against communication link failures. We achieve this by modifying the double-iteration algorithm (by introducing some additional auxiliary variables) and prove that the modified double-iteration converges almost surely to average consensus. In the first part of the paper, we study the first and second moments of the two iterations, and use them to establish convergence, and illustrate the performance of the algorithm with several numerical examples. In the second part, in order to establish the convergence of the algorithm, we use coefficients of ergodicity commonly used in analyzing inhomogeneous Markov chains.

University of Illinois at Urbana-Champaign. Coordinated Sciences Laboratory technical report UILU-ENG-11-2207 (CRHC-11-05)

A. D. Domínguez-García and N. H. Vaidya are with the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. E-mail: {aledan, nhv}@ILLINOIS.EDU.

C. N. Hadjicostis is with the Department of Electrical and Computer Engineering at the University of Cyprus, Nicosia, Cyprus, and also with the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. E-mail: chadjic@UCY.AC.CY.

The work of A. D. Domínguez-García was supported in part by NSF under Career Award ECCS-CAR-0954420. The work of C. N. Hadjicostis was supported in part by the European Commission (EC) Seventh Framework Programme (FP7/2007-2013) under grant agreements INFSO-ICT-223844 and PIRG02-GA-2007-224877. The work of N. H. Vaidya was supported in part by Army Research Office grant W-911-NF-0710287 and NSF award 1059540. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funding agencies, the U.S. government, or the European Commission.

I. INTRODUCTION

The design of protocols and algorithms for distributed computation and control/decision tasks has attracted significant attention by the computer science, communication, and control communities (e.g., [1], [2], [3], [4], [5], [6], [7] and references therein). For example, given i) a collection of robots moving in the plane, ii) a collection of sensors in a sensor network, or iii) a collection of distributed energy resources in an electrical grid, the components may be interested in, respectively, i) agreeing on a common direction to follow (this common direction could be provided by a leader robot), ii) measurement averaging (with each sensor providing a local measurement of a global quantity), or iii) collectively providing a predetermined total amount of active power subject to the constraints of each distributed resource. In the control literature, the first and second problems are respectively known as consensus and average consensus (see, e.g., [2]), whereas the third problem can be considered as a distributed resource coordination problem [8], [9].

In this two-part paper, we consider multicomponent systems in which each component can exchange information with other components in its neighborhood in order to compute, in a distributed fashion, some quantity of interest. In our setup, communication links between components (nodes) can be asymmetric (i.e., component j might be able to send information to component i, but not necessarily vice-versa), a situation that arises in a wireless setting if the transmission power available to different nodes are also different. In this setting, the information exchange between components in the system can be described by a directed graph which is assumed to be strongly connected. Through an iterative process, nodes in the network are required to compute (using only information made available by their neighbors) the quantity of interest. In particular, we study linear-iterative algorithms in which each node j maintains a value (or a set of values) that is updated to be a weighted linear combination of node j's own previous value and the previous values of its neighboring nodes (i.e., nodes that transmit information to node j). The main focus of the paper is to develop strategies to robustify the linear-iterative algorithms described above against communication links that may drop packets.

In the context of consensus and average-consensus problems, an extensive literature in the control community focuses on the linear-iterative algorithms described above (e.g., [7], [10], [2], [11], [12], [13], [14] and references therein). These works have revealed that if the network topology satisfies certain conditions, the weights for the linear iteration can be chosen so that all the nodes asymptotically converge to the same value (even if the network connections are time-varying). Additionally, if the interconnection topology is invariant and bidirectional (i.e., if node j can send information to node i, then node i can send information to node j), simple techniques can be used to choose the weights of the linear iteration so as to ensure that, after running the linear iteration, the nodes will asymptotically reach consensus to the average of their initial values [2], [11], [12]. Other works have looked at the consensus and average-consensus problems when the interconnection topology is described by a *directed* graph. In particular, the authors of [15] focus on continuous-time linear iterations and state necessary and sufficient conditions for a network of integrators to asymptotically reach agreement to a common value

(but not necessarily the average of their initial values). Similarly, the authors of [12] consider discrete-time iterations, and provide necessary and sufficient conditions on the weights that allow the nodes to asymptotically reach consensus to the average of their initial values. Additionally, the work in [16], [17] discusses how average-consensus can be reached asymptotically with linear-iterative algorithms in which the nodes use fixed weights in their linear updates and also develops linear-iterative algorithms where the nodes adapt their weights in a distributed fashion so that asymptotically average-consensus is reached. In the context of resource coordination, there is some recent work [18], [8], [9] that also focuses on linear-iterative algorithms, similar to those used to address consensus and average-consensus problems. Other recent work has addressed the related problem of achieving consensus and average-consensus in a multicomponent system where some nodes can exhibit malicious behavior [19], [20]. These works assume fault-free communication links, but are related to what we do in this paper in the sense that they can be used to handle unreliable nodes (as opposed to links).

In our development, we adopt a very general model for the communication modality between nodes, which allows asymmetric information structures, in the sense that if node i can transmit information to another node j, it is not necessarily true that node j can transmit information to node i. We only require that each node, apart from seeing incoming transmissions sent to it by neighboring nodes, knows the number of nodes that it can transmit information to, which in graph-theoretic terms is referred to as the out-degree of that node. In fact, in the proposed algorithm, each node will broadcast the same quantity to all receiving nodes, which simplifies the communication scheme between sending and receiving nodes (as it is not necessary for each sending node to separately communicate with each receiving node).

When the communication network is perfectly reliable (no packet drops), the collective dynamics of the linear iterations can be described by a discrete-time transition system with no inputs in which the transition matrix is column stochastic and primitive. Then, each node will run two identical copies of the linear iteration each of which, however is initialized differently depending on the problem to be solved.In this paper we mostly focus on the average consensus problem. Under proper initialization, it can be shown that each node will asymptotically calculate the desired value as a function of the outcomes of the two iterations. The details of these double-iteration approach are provided in [16], [17] for the average consensus case and in [8], [9] for the resource coordination problem. For the average-consensus problem, the double-iteration algorithm is a particular case of the algorithm in [21] (which is a generalization of the algorithm proposed in [22]), where the matrices describing each linear iteration are allowed to vary as time evolves, whereas in our setup (for the ideal case when there are no communication link failures) the transition matrix is fixed over time.

The focus of this paper is to robustify the double-iteration algorithm (informally described above and formally described in Section II) so that it can tolerate failures in communication links and converge to the average value. Our communication link reliability model assumes that at each time step, a communication link is unavailable with some probability. In other words, a packet containing information from node i to node j is dropped with some probability. Next we informally describe our robustification

approach. Consider two nodes i and j, and assume that j receives information from node i but not necessarily vice-versa. Let us refer to j as the receiving node (or receiver) and i as the sending node (or sender). An important requirement is for the graph describing the communication network to be strongly connected, which implies every node must be able to act both as a sender and as a receiver. Then, for each of the two iterations each node performs, node i (the sender) will keep track of the following quantities of interest: i) its own internal state (as captured by the state variables maintained in the double iteration scheme of [21], [18]; ii) the total mass broadcasted so far (to be described in detail soon); and iii) the total received mass from each node l that sends information to node i. Similarly, for both iterations, each node j (the receiver) updates the value of its internal state to be a linear combination of its own previous internal state value (weighted by the inverse of the number of nodes that have j as a neighbor) and the difference between the two most recently received mass values from each of its neighbors (also weighted by the inverse of the number of nodes that have j as a neighbor). At time instant k, the total broadcasted mass by node j is the sum up to (and including) time step k of the sequence of values of node j's internal value, weighted by the inverse of the number of nodes that receive values from node j). Additionally, node j updates the value of the received mass from each node l that sends information to node j as follows: the received mass from node l is the total broadcasted mass sent by node l up to time k if the communication link from node l to node j is available at time step k; otherwise, the received mass remains the same as the most recently received mass from node l. An implicit assumption here is that messages broadcasted by node l are tagged with the sender's identity so that the receiving node j can determine where different packages have originated from.

Recent work that has addressed the consensus and average-consensus problems in the presence of unreliable communication links [23], [24], [25] has employed a communication link availability model similar to ours. The work in [23] assumes that the graph describing the communication network is undirected and that when a communication link fails it affects communication in both directions. Additionally, nodes have some mechanism to detect link unavailability and compensate for it by rescaling their other weights (so that the resulting transition matrix remains column stochastic). Following this strategy, the authors show asymptotic convergence to the average of initial conditions and also calculate the rate at which the variance of the total deviation from the average converges to zero. The work in [24] does not require the graph describing the communication network to be undirected and proposes two compensation methods to account for communication link failures. In the first method, the so-called biased compensation method, the receiving node compensates for the unavailability of an incoming link by adding the weight associated to the unavailable link to its own weight (so that the resulting matrix remains row stochastic). In the second method, called the balanced compensation method, the receiving node compensates for link unavailability by rescaling all the incoming link weights so that the resulting matrix remains row stochastic. The key in both methods is the fact that at each time step, the resulting weight matrix is row stochastic; the authors show that the nodes converge almost surely to the same value, but this value is not necessarily the average of the initial conditions. The work in [25], which does

not require the communication graph to be undirected, proposes a correction strategy that corrects the errors in the quantity (state) iteratively calculated by each node, so that the nodes obtain the average of their initial values. This correction strategy is based on each node maintaining some auxiliary variable that accounts for the amount by which node i changes its state due to the updates from its neighbors, i.e., the nodes that can send information to node i. For their strategy to work and ensure that the nodes converge almost surely to the average consensus, the authors rely on the nodes sending acknowledgment messages and retransmitting information an appropriate number of times.

In [21], the authors proposed a gossip-based algorithm for average-consensus ver a directed graph where the transition matrices describing the nodes' collective dynamics change at every iteration step (depending on which node awakes). This scheme requires the node that is awake to perform an internal state update and send its internal state (weighted by the corresponding out-going link weight) to its neighbors. This approach results in generates a sequence of column stochastic matrices (not necessarily primitive) with the property that all the diagonal entries remain positive. The authors prove that by running two such iterations in parallel, one of them initialized with the values on which the average operation is to be performed and the other with the all-ones vector, each node will asymptotically achieve average consensus by taking the ratio of the two values in maintains. A key premise in their proof is that column stochasticity of the transition matrix is maintained over time, which requires sending nodes to know the number of nodes that are listening. This suggests that i) either the communication links are perfectly reliable, or ii) there is some acknowledgment and retransmission mechanism that ensures messages are delivered to the listening nodes at every round of information exchange. In this paper, we remove such assumptions and robustify the double-iteration algorithm against unreliable communication links using a pure broadcast-message model without any requirement for an acknowledgment/retransmission mechanism. Thus, despite the reliance of our algorithm on the ratio of two linear iterations, it is different both in the communication model we assume—a broadcast model in our case—and also in the nature of the protocol itself—our focus is on ensuring convergence in the presence of communication link failures.

An additional assumption made in [21] is that the diagonal entries of the transition matrix (at every step) remain positive. In our model, we originally consider that nodes do not drop self-packets. However, to ease the analysis, we remove this assumption and consider the case where self-packet drops are also allowed at every time step, which i) allows us to handle intermittent faults in the node processing device, and ii) removes the assumption that all diagonal entries must be positive at every step. Finally, the analysis machinery in [21] is quite different from the one used in this paper. We employ moment analysis of the two iterations to establish that they are linearly related as the number of steps goes to infinity, while [21] relies on establishing weak ergodicity of the product of the transition matrices as the number of steps goes to infinity. Finally, as it will be shown in the second part of this paper, our algorithm can be re-casted into a similar framework as the one in [21] by augmenting the dimension of the vector describing the collective dynamics to account for the packets that get dropped once there

is a communication failure. Note, However, that the resulting matrices will be column stochastic but will not necessarily have strictly positive entries on their diagonals. In the second part of the paper, we provide an analysis framework to establish the convergence of our algorithm and generalizes the ideas in [21] to the case when the matrices describing the system collective dynamics do not have strictly positive diagonals. In this regard, we will show that even in the case where self-packet drops are not allowed, the resulting transition matrices might still have zero diagonal entries.

The remainder of this paper is organized as follows. Section II provides background on graph theory, introduces the communication model, and briefly describes the non-robust version of the double-iteration algorithm we use in this work. Section III describes the proposed strategy to robustify the double-iteration algorithm against communication link failures and illustrates the use/performance of the algorithms via several examples. The convergence analysis of the robustified double-iteration algorithm is provided in Section V. Concluding remarks are presented in Section VI.

II. PRELIMINARIES

This section provides background of graph-theoretic notions that are used to describe the communication network and the distributed consensus/coordination setup, introduces the basic communication link availability model, and reviews a previously proposed two-iteration algorithm that can be used to solve the class of problems addressed in this paper when the communication network is perfectly reliable.

A. Network Communication Model

Let discrete time instants be indexed $k = 0, 1, \ldots$; then, the information exchange between nodes (components) at each time instant k can be described by a directed graph $\mathcal{G}[k] = \{\mathcal{V}, \mathcal{E}[k]\}$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the vertex set (each vertex—or node—corresponds to a system component), and $\mathcal{E}[k] \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, where $(j, i) \in \mathcal{E}[k]$ if node j can receive information from node i at instant k. It is assumed that $\mathcal{E}[k] \subseteq \mathcal{E},\ \forall k \geq 0$, where $\mathcal{E}$ is the set of edges that describe all possibly available communication links between nodes; furthermore, the graph $(\mathcal{V}, \mathcal{E})$ is assumed to be strongly connected. All nodes that can possibly transmit information to node j are called its in-neighbors, and are represented by the set $\mathcal{N}_j^- = \{i \in \mathcal{V} : (j, i) \in \mathcal{E}\}$. Note that there are self-loops for all nodes in $\mathcal{G}$ (i.e., $(j, j) \in \mathcal{E}$ for all $j \in \mathcal{V}$). The number of neighbors of j (including itself) is called the in-degree of j and is denoted by $\mathcal{D}_j^- = |\mathcal{N}_j^-|$. The nodes that have j as neighbor (including itself) are called its out-neighbors and are denoted by $\mathcal{N}_j^+ = \{l \in \mathcal{V} : (l, j) \in \mathcal{E}\}$; the out-degree of node j is $\mathcal{D}_j^+ = |\mathcal{N}_j^+|$.

The existence of a communication link from node i to node j can be described in probabilistic terms as follows. At instant k, let $x_{ji}[k],\ \forall i, j \in \mathcal{V}$ be an indicator variable that takes value 1 with probability q and takes value zero with probability $1 - q$, i.e.,

$$\Pr\{x_{ji}[k] = m\} = \begin{cases} q, & \text{if } m = 1, \\ 1 - q, & \text{if } m = 0. \end{cases} \tag{1}$$

Then, for all $k \geq 0$, the existence of a communication link between nodes i and j can be described be another indicator variable $\ell_{ji}[k]$ defined as

$$\ell_{ji}[k] = \begin{cases} x_{ji}[k], & \text{if } (j,i) \in \mathcal{E}, \\ 0, & \text{if } (j,i) \notin \mathcal{E}. \end{cases} \tag{2}$$

It follows that $\mathcal{E}[k]$ contains the elements of $\mathcal{E}$ for which $\ell_{ji}[k] = x_{ji}[k] = 1$.

B. Double-Iteration Algorithm Formulation in Perfectly Reliable Communication Networks

When the communication network of a multi-component system is perfectly reliable, i.e., $\Pr\{\ell_{ji}[k] = 1\} = 1,\ \forall (j,i) \in \mathcal{E},\ \forall k \geq 0$, it was shown in [17], [9] that the components of the multi-component system can asymptotically solve average consensus and resource coordination problems in a distributed fashion by running two separate appropriately initialized linear iterations of the form

$$y_j[k+1] = \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} y_i[k], \tag{3}$$

$$z_j[k+1] = \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} z_i[k], \tag{4}$$

where $\mathcal{D}_j^+$ ($\mathcal{D}_i^+$) is the out-degree of node j (i). A requirement in all cases is that the underlying communication graph $(\mathcal{G}, \mathcal{E})$ is strongly connected.

1) Average Consensus Problem: In this problem, the nodes aim to obtain the average of the values $v_j,\ j = 1, \ldots, n$, they each posses. In [17], it was shown that if the initial conditions in (3) (referred to as iteration 1) are set to $y_j[0] = v_j$, and the initial conditions in (4) (referred to as iteration 2) are set to $z_j[0] = 1$, then the nodes can asymptotically calculate $\overline{v} := \sum_{j=1}^n v_j/n$ as

$$\overline{v} = \lim_{k \to \infty} \frac{y_j[k]}{z_j[k]}, \tag{5}$$

by running the two iterations in (3) and (4).

2) Resource Coordination Problem: In this problem, each node j can contribute a certain amount $\pi_j \geq 0$ of a given resource, which is upper and lower bounded by known capacity limits π_j^{max} and π_j^{min} respectively. The challenge is to coordinate the components so that they collectively provide a pre-determined total amount $\rho_d = \sum_{j=1}^n \pi_j$ of the resource[1] as specified by an external "leader." In [9], it was shown that i) if the initial conditions in (3) are set to $y_j[0] = \rho_d/m - \pi_j^{min}$ if j is an out-neighbor of the leader (where $m \geq 1$ is the number of nodes contacted initially by the external leader) and $y_j[0] = -\pi_j^{min}$ otherwise, and ii) if the initial conditions in (4) are set to $z_j[0] = \pi_j^{max} - \pi_j^{min}$, then the

[1]In the development in [18], [9], it is assumed that $\sum_{j=1}^n \pi_j^{min} \leq \rho_d \leq \sum_{j=1}^n \pi_j^{max}$; this is not a restrictive assumption because in the proposed algorithms, all nodes will be able to know if $\rho_d < \sum_{j=1}^n \pi_j^{min}$ or if $\rho_d > \sum_{j=1}^n \pi_j^{max}$ (which means that no feasible solution exists).

nodes can asymptotically calculate their own resource contribution π_j as

$$\pi_j := \lim_{k\to\infty} \left(\pi_j^{min} + \frac{y_j[k]}{z_j[k]}(\pi_j^{max} - \pi_j^{min})\right) = \pi_j^{min} + \frac{\rho_d - \sum_{l=1}^n \pi_l^{min}}{\sum_{l=1}^n \ell_l}(\pi_j^{max} - \pi_j^{min}), \tag{6}$$

which satisfies

$$\begin{aligned} &\pi_j^{min} \leq \pi_j \leq \pi_j^{max}, \ \forall j, \\ &\sum_{j=1}^n \pi_j = \rho_d. \end{aligned} \tag{7}$$

In this paper, we start with a double iteration of the form in (3)–(4) that is used for either average consensus or coordination, and develop systematic methodologies to handle packet drops in the communication links.

III. Robustification of Double-Iteration Algorithm

In this section, the algorithm described in Section II-B is modified so as to make it robust against communication link failures. As in (3)–(4), each node will run two iterations (which we refer to as iterations 1 and 2) to calculate quantities of interest and eventually solve the average consensus or coordination problems.

Consider the setup described in the previous section: we are given a (possible directed) strongly connected graph $(\mathcal{G}, \mathcal{E})$ representing a multicomponent system and its communication links between its components. For the sake of generality, let us refer to j as the receiving node (or receiver) and i as the sending node (or sender). For each of the two iterations, node i (the sender) will calculate several quantities of interest, which we refer to as: i) internal state; ii) total broadcasted mass; and iii) total received mass from each in-neighbor l of node i, i.e., for each node $l \in \mathcal{N}_i^-$. For both iterations 1 and 2, each node j updates the value of its internal state to be a linear combination of its own previous internal state value (weighted by the inverse of the number of nodes that have j as a neighbor, i.e., $1/\mathcal{D}_j^+$) and the sum (over all its in-neighbors) of the difference between the two most recently received mass values. At instant time k, the total broadcasted mass is the sum up to (and including) step k of the weighted value of node j's internal state (used to update the internal state of node j). Additionally, node j (the receiver) updates the value of the received mass from node l to be either the total broadcasted mass sent by node i if the communication link from i to j is available at instant k, or the most recently received mass value from node i, otherwise. An implicit assumption here is that messages broadcasted by node i are tagged with the sender's identity so that the receiving node j can determine where messages originated from.

For iteration 1, let $y_j[k]$ denote node j's internal state at time instant k, $\mu_{lj}[k]$ denote the mass broadcasted from node j to each of its out-neighbors l (this is a single value ad it is the quantity is the same for each out-neighbor l of node j, i.e., for each $l \in \mathcal{N}_j^-$), and $\nu_{ji}[k]$ denote the total mass received at node j from node $i \in \mathcal{N}_j^-$. Similarly, let $z_j[k]$ denote node j's internal state takes at time

instant k, $\sigma_{lj}[k]$ denote node j's broadcasted mass for each out-neighbor $l,\ l \in \mathcal{N}_j^+$, and $\tau_{ji}[k]$ denote the total mass received from node $i \in \mathcal{N}_j^-$. Then, the progress of iteration 1 is described by

$$y_j[k+1] = \frac{1}{\mathcal{D}_j^+} y_j[k] + \sum_{i \in \mathcal{N}_j^-} \big(\nu_{ji}[k] - \nu_{ji}[k-1]\big), \quad k \geq 0,$$

$$\mu_{lj}[k] = \mu_{lj}[k-1] + \frac{1}{\mathcal{D}_j^+} y_j[k] = \sum_{i=0}^{k} \frac{1}{\mathcal{D}_j^+} y_j[i], \quad k \geq 0, \tag{8}$$

where

$$\nu_{ji}[k] = \begin{cases} \mu_{ji}[k], & \text{if } (j,i) \in \mathcal{E}[k], \quad k \geq 0, \\ \nu_{ji}[k-1], & \text{if } (j,i) \notin \mathcal{E}[k], \quad k \geq 0. \end{cases}$$

Recall that $\mathcal{D}_j^+$ ($\mathcal{D}_i^+$) is the number of nodes that node j (i) can transmit information to. Similarly, the progress of iteration 2 is described by

$$z_j[k+1] = \frac{1}{\mathcal{D}_j^+} z_j[k] + \sum_{i \in \mathcal{N}_j^-} \big(\tau_{ji}[k] - \tau_{ji}[k-1]\big), \quad k \geq 0,$$

$$\sigma_{lj}[k] = \sigma_{lj}[k-1] + \frac{1}{\mathcal{D}_j^+} z_j[k] = \sum_{i=0}^{k} \frac{1}{\mathcal{D}_j^+} z_j[i], \quad k \geq 0, \tag{9}$$

where

$$\tau_{ji}[k] = \begin{cases} \sigma_{ji}[k], & \text{if } (j,i) \in \mathcal{E}[k], \quad k \geq 0, \\ \tau_{ji}[k-1], & \text{if } (j,i) \notin \mathcal{E}[k], \quad k \geq 0. \end{cases}$$

As mentioned earlier, for solving the average consensus problem, the initial conditions in (8) are set to $y_j[0] = v_j$, whereas the initial conditions in (9) are set to $z_j[0] = 1$. Similarly, for solving the resource coordination problem, the initial conditions in (8) are set to $y_j[0] = \rho_d/m - \pi_j^{min}$ if j is initially contacted by the leader and $y_j[0] = -\pi_j^{min}$ otherwise, whereas the initial conditions in (9) are set to $z_j[0] = \pi_j^{max} - \pi_j^{min} > 0$. In both the average consensus and coordination problems, $\mu_{ji}[-1] = 0$ and $\nu_{ji}[-1] = 0$ for all $(j,i) \in \mathcal{E}$, and $\sigma_{ji}[-1] = 0$ and $\tau_{ji}[-1] = 0$ for all $(j,i) \in \mathcal{E}$.

Main Result: We shall argue that with the proposed robustification strategy, despite the presence of unreliable communication links (at each time step, each link $(j,i) \in \mathcal{E}$, fails independently from other links and independently between time steps, with some probability $1 - q_{ji}$), nodes can asymptotically estimate the exact solution $\overline{v}$ to the average consensus by calculating, whenever $z_j[k] > 0$ the ratio $y_j[k]/z_j[k]$, i.e.,

$$\overline{v} = \lim_{k \to \infty} \frac{y_j[k]}{z_j[k]}, \tag{10}$$

whenever $z_j[k > 0]$. Similarly, exact solution to the resource coordination problem can be obtained as

$$\pi_j = \lim_{k\to\infty} \big(\pi_j^{min} + \frac{y_j[k]}{z_j[k]}(\pi_j^{max} - \pi_j^{min})\big), \tag{11}$$

whenever $z_j[k] > 0$. In both cases, we run the iterations in (8) and (9) and using the corresponding initial conditions as outlined above. In particular, we will show that, for every j, $z_j[k] - \frac{\sum_{j=1}^n z_0(j)}{\sum_{j=1}^n y_0(j)} y_j[k] \to 0$ as $k \to \infty$ almost surely. Additionally, we will show that $z_j[k] > 0$ occurs infinitely often.

A. Examples

We now illustrate how the proposed algorithm works for the case of average consensus in the presence of packet-dropping communication links. We start with the rather small network shown in Fig. 1 and assume that the packets on each link (including the self-links which are not drawn in the figure[2]) can be dropped with probability $1-q$, independently between different links and independently between different iterations. We also assume that the initial values of the five nodes are given by $v = [-4, 5, 6, -3, 1]^T$, with their average equal to 1. Thus, in the iterations (8) and (9)

$$y[0] = [-4, 5, 6, -3, 1]^T, \text{ and } z[0] = [1, 1, 1, 1, 1]^T,$$

with $\mu_{ji}[-1] = v_{ji}[-1] = \sigma_{ji}[-1] = \tau_{ji}[-1] = 0$ for all $(j, i) \in \mathcal{E}$.

We run the iterations in (8) and (9) and plot the ratio $\frac{y_j[k]}{z_j[k]}$ as a function of the iteration step k for each node j ($j = 1, 2, 3, 4, 5$). Figure 2 shows the typical behavior that we observe for $q = 0.99$ (i.e., for a probability of a packet drop equal to 0.01). As can be seen in the figure, the ratio at each node quickly converges to the correct average, though the individual values for $y_j[k]$ and $z_j[k]$ do not converge.

In Figs. 3 and 4 we show typical behaviors of the same multicomponent system for $q = 0.5$ and $q = 0.1$. The behavior remains similar to the one observed before: even though $y_j[k]$ and $z_j[k]$ do not

[2]We make this assumption later in the paper for the purposes of simplifying notation.

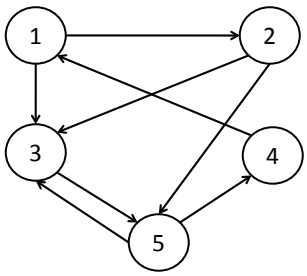
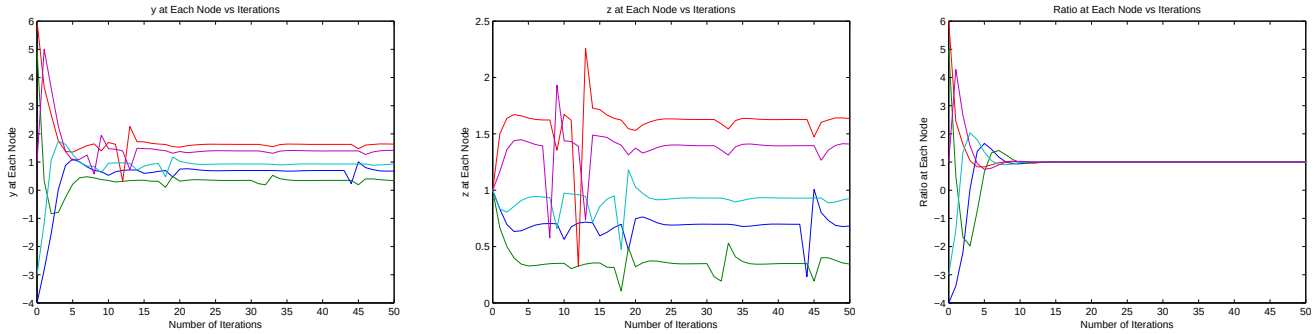
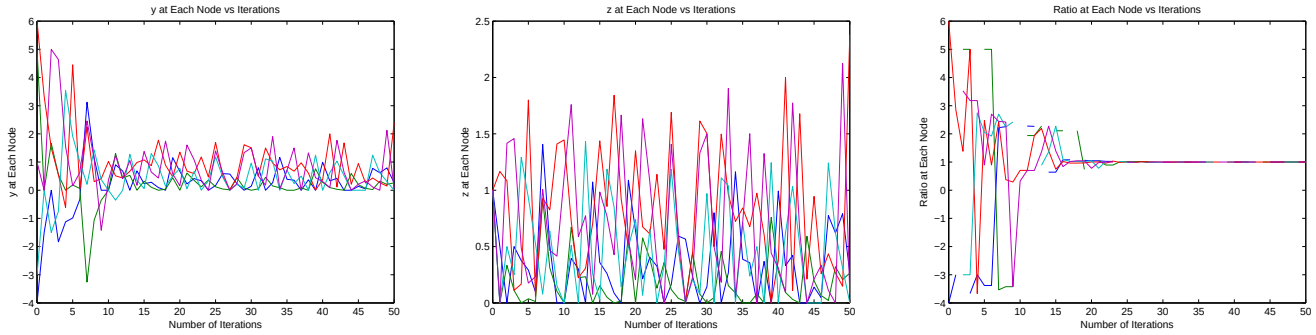


Fig. 1. Small directed graph used for illustration of the ratio algorithm for obtaining average consensus in the presence of packet-dropping communication links.



(a) (b) (c)

Fig. 2. Evolution of the values of $y_j[k]$ (left), $z_j[k]$ (middle) and $\frac{y_j[k]}{z_j[k]}$ (right) for $q = 0.99$.

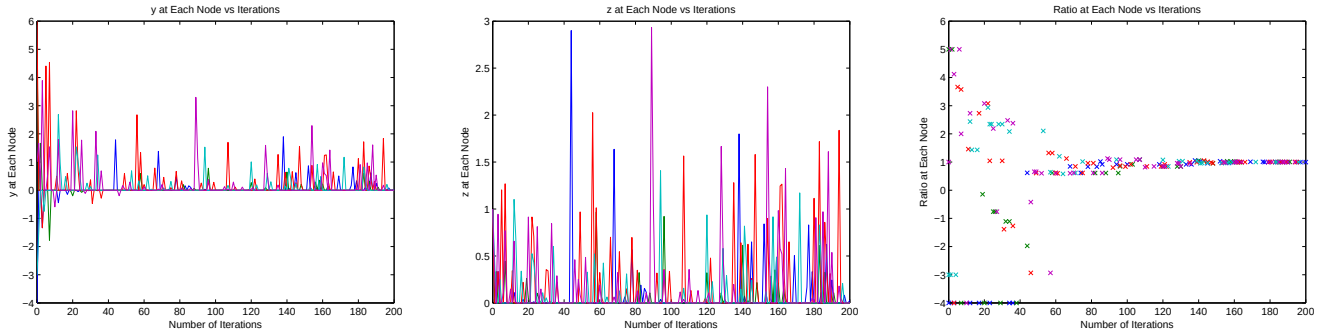


(a) (b) (c)

Fig. 3. Evolution of the values of $y_j[k]$ (left), $z_j[k]$ (middle), and $\frac{y_j[k]}{z_j[k]}$ (right) for $q = 0.5$, $j = 1, 2, 3, 4, 5$.

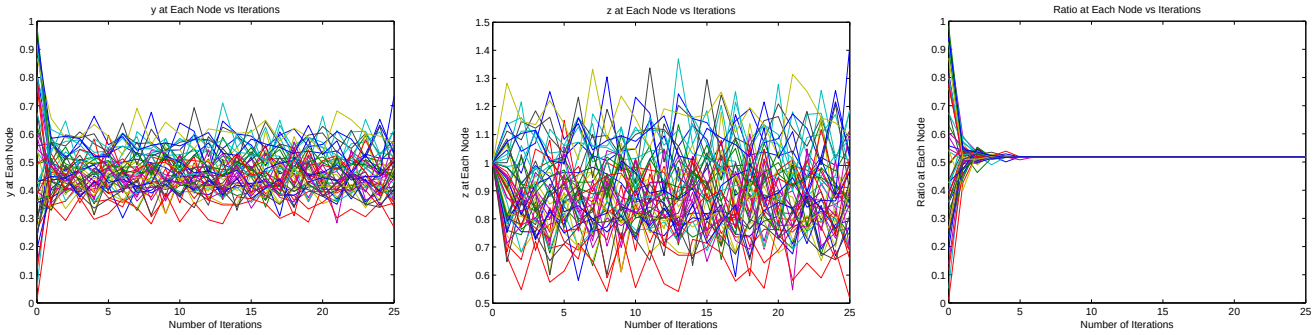
converge (in fact, they seem to behave more radically with decreasing q), the ratio $\frac{y_j[k]}{z_j[k]}$ does converge to the average of the initial values. Note that the plot of the ratio in Fig. 4 is quite different than the rest: in this case, q is small enough so that $z_j[k]$ (and simultaneously $y_j[k]$) can take the value zero (e.g., when all packets destined to node j are dropped at iteration k); thus, the ratio in such cases is not defined and is not plotted, resulting in a discontinuous set of points in the plot. Nevertheless, we can see that when packets are received (which happens frequently enough for each node), the ratio has the correct value. This is a point addressed later in the paper.

An example of what happens in larger graphs is shown in Fig. 5. Here we consider a graph with 50 nodes, randomly generated by choosing a directed edge from node i to node j, $1 \leq i, j \leq 50$, $i \neq j$, independently with probability $1/2$, and ensuring that the resulting graph is strongly connected. As can be seen the behavior remains similar to what we observed for the smaller graph: the ratio $\frac{y_j[k]}{z_j[k]}$ converges quickly to the average even though the individual $y_j[k]$ and $z_j[k]$ do not converge. For this particular plot, we used $q = 0.1$, which also justifies the fluctuation in the values of $y_j[k]$ and $z_j[k]$.



(a) (b) (c)

Fig. 4. Evolution of the values of $y_j[k]$ (left), $z_j[k]$ (middle), and $\frac{y_j[k]}{z_j[k]}$ (right) for $q = 0.1$, $j = 1, 2, 3, 4, 5$.



(a) (b) (c)

Fig. 5. Evolution of the values of $y_j[k]$ (left), $z_j[k]$ (middle), and $\frac{y_j[k]}{z_j[k]}$ (right) for $q = 0.1$ for a 50-component system.

IV. First and Second Moment Analysis

In this section, we obtain recurrence relations that describe the first and second moment of the iterations after (8) and (9); this analysis is used in Section V to establish the claims in (10) and (11). In order to ease the moment calculations, the expressions in (8)–(9) will be rewritten more compactly in vector form. Also, in order to facilitate notation, we will allow each node j to drop the packet carrying its own previous value when updating its value. This way, node j handles its own value in the same way as its neighbors' values and notation is simplified significantly.

A. Vectorized Description of Double-Iteration Algorithm

Using the definition for the indicator variable $x_{ji}[k]$ given in (1) and the resulting indicator variable $\ell_{ji}[k]$ given in (2), which describes the successful transmission of information from node i to node j

over an existing, unreliable communication link, iterations (8) and (9) can be rewritten as

$$\mu_{lj}[k] = \begin{cases} \mu_{lj}[k-1] + \frac{1}{\mathcal{D}_j^+} y_j[k], & \text{if } l \in \mathcal{N}_j^+, \quad k \geq 0, \\ 0, & \text{if } l \notin \mathcal{N}_j^+, \quad k \geq 0, \end{cases} \tag{12}$$

$$\nu_{ji}[k] = \begin{cases} \mu_{ji}[k] x_{ji}[k] + \nu_{ji}[k-1](1 - x_{ji}[k]), & \text{if } i \in \mathcal{N}_j^-, \quad k \geq 0, \\ 0, & \text{if } i \notin \mathcal{N}_j^-, \quad k \geq 0, \end{cases} \tag{13}$$

$$y_j[k+1] = \sum_{i=1}^{n} \big(\nu_{ji}[k] - \nu_{ji}[k-1]\big), \quad k \geq 0, \tag{14}$$

and

$$\sigma_{lj}[k] = \begin{cases} \sigma_{lj}[k-1] + \frac{1}{\mathcal{D}_j^+} z_j[k], & \text{if } l \in \mathcal{N}_j^+, \quad k \geq 0, \\ 0, & \text{if } l \notin \mathcal{N}_j^+, \quad k \geq 0, \end{cases} \tag{15}$$

$$\tau_{ji}[k] = \begin{cases} \sigma_{ji}[k] x_{ji}[k] + \tau_{ji}[k-1](1 - x_{ji}[k]), & \text{if } i \in \mathcal{N}_j^-, \quad k \geq 0, \\ 0, & \text{if } i \in \mathcal{N}_j^-, \quad k \geq 0, \end{cases} \tag{16}$$

$$z_j[k+1] = \sum_{i=1}^{n} \big(\tau_{ji}[k] - \tau_{ji}[k-1]\big), \quad k \geq 0, \tag{17}$$

where $\mu_{lj}[-1] = \nu_{ji}[-1] = \sigma_{lj}[-1] = \tau_{ji}[-1] = 0, \ \forall j, i$.

Let $A \circ B$ denote the Hadamard (entry-wise) product of a pair of matrices A and B of identical size. Then, for all $k \geq 0$, iteration (12)–(14) can be rewritten in matrix form as

$$M_k = M_{k-1} + P\text{diag}(y_k), \tag{18}$$

$$N_k = M_k \circ X_k + N_{k-1} \circ (U - X_k), \tag{19}$$

$$y_{k+1} = (N_k - N_{k-1})e = \big[(M_k - N_{k-1}) \circ X_k\big]e, \tag{20}$$

where $P = [p_{ji}] \in \mathbb{R}^{n \times n}$, with $p_{ji} = \frac{1}{\mathcal{D}_i^+}, \ \forall j \in \mathcal{N}_i^+$ and $p_{ji} = 0$ otherwise; $M_{-1} = N_{-1} = 0$; $y_k = y[k]$; $U \in \mathbb{R}^{n \times n}$, with $[U_{ji}] = 1, \ \forall i, j$; diag$(y_k)$ is the diagonal matrix that results by having the entries of y_k on the main diagonal; and $e = [1, 1, \ldots, 1]'$ (note that $U = ee^T$). Similarly, for $k \geq 0$, (15)–(17) can be rewritten in matrix form as

$$S_k = S_{k-1} + P\text{diag}(z_k), \tag{21}$$

$$T_k = S_k \circ X_k + T_{k-1} \circ (U - X_k), \tag{22}$$

$$z_{k+1} = (T_k - T_{k-1})e = \big[(S_k - T_{k-1}) \circ X_k\big]e, \tag{23}$$

where $S_{-1} = T_{-1} = 0$, $z_k = z[k]$, and diag(z_k) is the diagonal matrix that results by having the entries of z_k on the main diagonal.

By defining $A_k := M_k - N_{k-1}$ and $B_k := S_k - T_{k-1}$, iteration (18)–(20) can be rewritten more

compactly as

$$A_k = A_{k-1} \circ (U - X_{k-1}) + P\text{diag}(y_k), \quad k \geq 1, \tag{24}$$

$$y_{k+1} = (A_k \circ X_k)e, \quad k \geq 0, \tag{25}$$

and iteration (21)–(23) as

$$B_k = B_{k-1} \circ (U - X_{k-1}) + P\text{diag}(z_k), \quad k \geq 1, \tag{26}$$

$$z_{k+1} = (B_k \circ X_k)e, \quad k \geq 0, \tag{27}$$

where $A_0 = M_0 - N_{-1} = P\text{diag}(y_0)$, and $B_0 = S_0 - T_{-1} = P\text{diag}(z_0)$.

For analysis purposes, each matrix in (24)–(25) and (26)–(27) will be rewritten in vector form by stacking up the corresponding columns.[3] Then, (24)–(25) and (26)–(27) can be rewritten in vector form as follows. Let $F = [I_n \ I_n \ \ldots \ I_n] \in \mathbb{R}^{n \times n^2}$, where I_n is the $n \times n$ identity matrix, and $\tilde{P} = [E_1 P^T \ E_2 P^T \ \ldots \ E_n P^T]^T \in \mathbb{R}^{n^2 \times n}$, where $E_i \in \mathbb{R}^{n \times n}$ has $E_i(i,i) = 1$ and all other entries equal zero. [The entries of $E_i P^T \in \mathbb{R}^{n \times n}$ ($PE_i^T = PE_i$) are all zero except for the i^{th} row (column) entries, which are those of the i^{th} row (column) of matrix P^T (P).] Then, (24)–(25) can be rewritten as

$$a_k = a_{k-1} \circ (u - x_{k-1}) + \tilde{P} y_k, \quad k \geq 1, \tag{28}$$

$$y_{k+1} = F(a_k \circ x_k), \quad k \geq 0, \tag{29}$$

where $a_k \in \mathbb{R}^{n^2}$, $x_k \in \mathbb{R}^{n^2}$, and $x_{k-1} \in \mathbb{R}^{n^2}$ result from stacking the columns of matrices A_k, X_k, and X_{k-1}, respectively. Similarly, (26)–(27) can be rewritten as

$$b_k = b_{k-1} \circ (u - x_{k-1}) + \tilde{P} z_k, \quad k \geq 1, \tag{30}$$

$$z_{k+1} = F(b_k \circ x_k), \quad k \geq 0, \tag{31}$$

where $b_k \in \mathbb{R}^{n^2}$ results from stacking the columns of matrix B_k.

Remark 1: It is important to note that matrices A_k and B_k, and their corresponding vectors a_k and b_k, have some entries that remain at zero for all $k \geq 0$. Specifically, the (j,i) entry of matrices A_k and B_k (and their corresponding entries in a_k and b_k) remain zero if there is no communication link from node i to node j, i.e., $(j,i) \notin \mathcal{E}$. The reason we keep these entries (despite the fact they are zero and do not play a role in the analysis) is because they facilitate matrix notation and calculations in subsequent developments. □

Since it will appear later at several points of the analysis, it is worth noting that when premultiplying $\tilde{P}$ by F, we recover the matrix P, i.e.,

$$P = F\tilde{P}. \tag{32}$$

[3]If we let $A = [A_{ij}] \in \mathbb{R}^{n \times n}$, then $a = [A_{11}, A_{21}, \ldots, A_{n1}, A_{12}, A_{22}, \ldots, A_{n2}, \ldots, A_{1n}, A_{2n}, \ldots, A_{nn}]^T$. Vectors defined by stacking the columns of a matrix will be denoted with the same small letter as the capital letter of the corresponding matrix.

B. First Moment Analysis

In this section, we describe the first moment dynamics of (28)–(31) via discrete-time transition systems with no inputs, where (as shown below) the corresponding transition matrices (which are obtained from P and q) are column stochastic and primitive. In both iterations, the sum of the entries of the first moment vectors for y_k and z_k is shown to remain constant over time and be respectively equal to the sum of $q\sum_i y_0(i)$ and $q\sum_i z_0(i)$. Furthermore, both first moments $\mathbf{E}[y_k]$ and $\mathbf{E}[z_k]$ are shown to reach a steady-state value as k goes to infinity. The above discussion is formalized in the following lemma.

Lemma 1: Let a_k, b_k, y_k, and z_k be described by the recurrence relations in (28)–(29), and (30)–(31) respectively. Let the first moments of a_k, y_k, b_k, and z_k (i.e., $\mathbf{E}[a_k]$, $\mathbf{E}[y_k]$, $\mathbf{E}[b_k]$, and $\mathbf{E}[z_k]$) be denoted by $\overline{a}_k$, $\overline{y}_k$, $\overline{b}_k$, and $\overline{z}_k$ respectively. Then the evolution of $\overline{a}_k$, $\overline{y}_k$, $\overline{b}_k$, and $\overline{z}_k$, $\forall k \geq 1$, is governed by

$$\overline{a}_k = \big[q\tilde{P}F + (1-q)I_{n^2}\big]\overline{a}_{k-1}, \tag{33}$$

$$\overline{y}_{k+1} = \big[qP + (1-q)I_n\big]\overline{y}_k, \tag{34}$$

$$\overline{b}_k = \big[q\tilde{P}F + (1-q)I_{n^2}\big]\overline{b}_{k-1}, \tag{35}$$

$$\overline{z}_{k+1} = \big[qP + (1-q)I_n\big]\overline{z}_k, \tag{36}$$

where I_m is the $m \times m$ identity matrix, with $\overline{a}_0 = \tilde{P}y_0$, $\overline{y}_1 = qPy_0$, $\overline{b}_0 = \tilde{P}z_0$, and $\overline{z}_1 = qPz_0$.

Proof: Since the development for obtaining $\overline{a}_k$ and $\overline{y}_k$ is parallel to that for obtaining $\overline{b}_k$ and $\overline{z}_k$, our analysis focuses on the first case. For $k = 0$ in (28)–(29), by taking expectations of both sides and noting that packet drops at time step $k = 0$ are independent of the initial values for a_0, it follows that

$$\overline{a}_0 = \tilde{P}y_0, \tag{37}$$

$$\overline{y}_1 = qF\overline{a}_0. \tag{38}$$

Substituting (37) into (38), we obtain $\overline{y}_1 = qF\tilde{P}\overline{y}_0 = qP\overline{y}_0$.

For $k \geq 1$ in (28)–(29), noting that packet drops at time step k are independent of previous packet drops and the initial values of a_0, it follows, by taking expectations on both sides, that

$$\overline{a}_k = \overline{a_{k-1} \circ (u - x_{k-1})} + \overline{\tilde{P}y_k} = \overline{a}_{k-1} \circ (u - \overline{x}_{k-1}) + \tilde{P}\overline{y}_k = (1-q)\overline{a}_{k-1} + \tilde{P}\overline{y}_k, \tag{39}$$

$$\overline{y}_{k+1} = \overline{F(a_k \circ x_k)} = F(\overline{a}_k \circ \overline{x}_k) = qF\overline{a}_k. \tag{40}$$

Substituting (40) into (39), we obtain

$$\overline{a}_k = (1-q)\overline{a}_{k-1} + q\tilde{P}F\overline{a}_{k-1} \tag{41}$$

$$= [q\tilde{P}F + (1-q)I_{n^2}]\overline{a}_{k-1}, \tag{42}$$

Similarly, substituting (39) into (40), we have

$$\overline{y}_{k+1} = (1-q)qF\overline{a}_{k-1} + qF\tilde{P}\overline{y}_k \tag{43}$$
$$= (1-q)\overline{y}_k + qF\tilde{P}\overline{y}_k \tag{44}$$
$$= [qP + (1-q)I_n]\overline{y}_k, \tag{45}$$

where I_n is the $n \times n$ identity matrix. ■

C. Second Moment Analysis

In the order to calculate the second moment dynamics for (28)–(31), we utilize in the following lemma.

Lemma 2: Let x, c and d be random vectors of dimension n. Furthermore, assume that the entries of x are Bernoulli *i.i.d.* random variables such that $\Pr\{x_i = 1\} = q$ and $\Pr\{x_i = 0\} = 1-q$, $\forall i = 1, 2, \ldots n$, and are independent from c and d. Then

$$S := \mathbf{E}\left[(c \circ x)(x \circ d)^T\right] = q^2\,\mathbf{E}[cd^T] + q(1-q)\,\mathbf{E}\left[\text{diag}(cd^T)\right], \tag{46}$$
$$T := \mathbf{E}\left[\big(c \circ x\big)\big((u-x) \circ d\big)^T\right] = q(1-q)\,\mathbf{E}[cd^T] - q(1-q)\,\mathbf{E}\left[\text{diag}(cd^T)\right], \tag{47}$$

where $\text{diag}(cd^T)$ is a diagonal matrix with the same diagonal as matrix cd^T.

Proof: The $(i,j), i \neq j$, entry of S can be obtained as follows:

$$S_{ij} = \mathbf{E}\left[c_i x_i d_j x_j\right]. \tag{48}$$

Since x_i and x_j are pairwise independent, and independent from c and d, it follows that

$$\mathbf{E}\left[c_i x_i d_j x_j\right] = q^2\,\mathbf{E}\left[c_i d_j\right]. \tag{49}$$

For $i = j$, observing that $\mathbf{E}[x_i x_i] = \mathbf{E}[x_i] = q,\ \forall i = 1, \ldots, n$, we obtain the corresponding entry of S as

$$S_{ii} = \mathbf{E}\left[c_i x_i d_i x_i\right] = \mathbf{E}\left[c_i d_i x_i\right] = q\,\mathbf{E}\left[c_i d_i\right]. \tag{50}$$

In (49), it is easy to see that $\mathbf{E}\left[c_i d_j\right]$ is the (i,j) entry of $\mathbf{E}[cd^T]$. Similarly, in (50), it is easy to see that $\mathbf{E}\left[c_i d_i\right]$ is the (i,i) entry of $\mathbf{E}[cd^T]$. From these observations, the result in (46) follows.

Similarly, the $(i,j), i \neq j$, entry of T can be obtained as follows:

$$T_{ij} = \mathbf{E}\left[c_i x_i d_j (1-x_j)\right]. \tag{51}$$

Since x_i and $(1-x_j)$ are independent, it follows that

$$\mathbf{E}\left[c_i x_i d_j (1-x_j)\right] = \mathbf{E}\left[c_i d_j\right]\mathbf{E}\left[x_i(1-x_j)\right] = q(1-q)\,\mathbf{E}\left[c_i d_j\right]. \tag{52}$$

For $i = j$, and observing that $\mathbf{E}[x_i(1-x_i)] = 0,\ \forall i = 1,\ldots,n$, the corresponding entry of T can obtained as follows;

$$T_{ii} = \mathbf{E}\left[c_i x_i d_i (1-x_i)\right] = \mathbf{E}\left[c_i d_i\right]\mathbf{E}\left[x_i(1-x_i)\right] = 0. \tag{53}$$

The result in (47) follows from (52) and (53). ■

The following lemma establishes that the evolution of $\mathbf{E}[a_k a_k^T]$, $\mathbf{E}[b_k b_k^T]$, and $\mathbf{E}[a_k b_k^T]$, and can be expressed as linear iterations with identical dynamics but different initial conditions. Similarly, the evolution of $\mathbf{E}[y_k y_k^T]$, $\mathbf{E}[z_k z_k^T]$, and $\mathbf{E}[y_k z_k^T]$ can also be expressed as linear iterations with identical dynamics but different initial conditions.

Lemma 3: Consider the second moments of a_k, y_k, b_k, and z_k, and let $\mathbf{E}[a_k a_k^T]$, $\mathbf{E}[y_k y_k^T]$, $\mathbf{E}[b_k b_k^T]$, $\mathbf{E}[z_k z_k^T]$, $\mathbf{E}[a_k b_k^T]$, and $\mathbf{E}[y_k z_k^T]$) be denoted by Γ_k, Φ_k, Ψ_k, Λ_k, Ξ_k, and Υ_k respectively. Then, the evolutions of Γ_k, Φ_k, Ψ_k, Λ_k, Ξ_k, Υ_k, $\forall k \geq 1$, are described by the following iterations (where all I denote $n^2 \times n^2$ identity matrices):

$$\Gamma_k = \left[q\tilde{P}F + (1-q)I\right]\Gamma_{k-1}\left[q\tilde{P}F + (1-q)I\right]^T + q(1-q)[I - \tilde{P}F]\text{diag}(\Gamma_{k-1})[I-\tilde{P}F]^T, \tag{54}$$

$$\Phi_{k+1} = F\left[q^2\Gamma_k + q(1-q)\text{diag}(\Gamma_k)\right]F^T, \tag{55}$$

$$\Psi_k = \left[q\tilde{P}F + (1-q)I\right]\Psi_{k-1}\left[q\tilde{P}F + (1-q)I\right]^T + q(1-q)[I - \tilde{P}F]\text{diag}(\Psi_{k-1})[I-\tilde{P}F]^T, \tag{56}$$

$$\Lambda_{k+1} = F\left[q^2\Psi_k + q(1-q)\text{diag}(\Psi_k)\right]F^T, \tag{57}$$

$$\Xi_k = \left[q\tilde{P}F + (1-q)I\right]\Xi_{k-1}\left[q\tilde{P}F + (1-q)I\right]^T + q(1-q)[I - \tilde{P}F]\text{diag}(\Xi_{k-1})[I-\tilde{P}F]^T, \tag{58}$$

$$\Upsilon_{k+1} = F\left[q^2\Xi_k + q(1-q)\text{diag}(\Xi_k)\right]F^T, \tag{59}$$

with initial conditions

$$\Gamma_0 = \tilde{P}y_0 y_0^T\tilde{P}^T, \tag{60}$$

$$\Phi_1 = \overline{y}_1\overline{y}_1^T + q(1-q)F\text{diag}(\tilde{P}y_0 y_0^T\tilde{P}^T)F^T, \tag{61}$$

$$\Psi_0 = \tilde{P}z_0 z_0^T\tilde{P}^T, \tag{62}$$

$$\Lambda_1 = \overline{z}_1\overline{z}_1^T + q(1-q)F\text{diag}(\tilde{P}z_0 z_0^T\tilde{P}^T)F^T, \tag{63}$$

$$\Xi_0 = \tilde{P}y_0 z_0^T\tilde{P}^T, \tag{64}$$

$$\Upsilon_1 = \overline{y}_1\overline{z}_1^T + q(1-q)F\text{diag}(\tilde{P}y_0 z_0^T\tilde{P}^T)F^T. \tag{65}$$

Proof: The derivation of (54), (55), (60), and (61), is the same as the derivation of (56), (57), (62), and (63), thus the developments in the proof will only address the former. For $k = 0$, it follows from Lemma 1 and (28) that $a_0 = \tilde{P}y_0$. Then,

$$\Gamma_0 = \mathbf{E}[a_0 a_0^T] = \tilde{P}\,\mathbf{E}[y_0 y_0^T]\tilde{P}^T = \tilde{P}y_0 y_0^T\tilde{P}^T, \tag{66}$$

and

$$\Phi_1 = \mathbf{E}[y_1 y_1^T] = \mathbf{E}\left[F(a_0 \circ x_0)(x_0 \circ a_0)^T F^T\right] = F\,\mathbf{E}\left[(a_0 \circ x_0)(x_0 \circ a_0)^T\right] F^T. \tag{67}$$

Applying the results in Lemmas 1 and 2 to (67), it follows that

$$\begin{aligned}\Phi_1 &= q^2 F\,\mathbf{E}\left[a_0 a_0^T\right] F^T + q(1-q) F\,\mathbf{E}\left[\text{diag}(a_0 a_0^T)\right] F^T \\ &= (qF\tilde{P}y_0)(qF\tilde{P}y_0)^T + q(1-q)F\,\mathbf{E}[\text{diag}(\tilde{P}y_0 y_0^T \tilde{P}^T)]F^T \\ &= (qPy_0)(qPy_0)^T + q(1-q)F\text{diag}(\tilde{P}y_0 y_0^T \tilde{P}^T)F^T \\ &= \overline{y}_1 \overline{y}_1^T + q(1-q)F\text{diag}(\tilde{P}y_0 y_0^T \tilde{P}^T)F^T,\end{aligned} \tag{68}$$

where we used the fact that $F\tilde{P} = P$ $\big($refer to Eq. (32)$\big)$.

For $k \geq 1$, and taking into account that $y_k = F(a_{k-1} \circ x_{k-1})$, it follows that

$$\begin{aligned}\Gamma_k =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1}) + \tilde{P}y_k\big)\big(a_{k-1} \circ (u - x_{k-1}) + \tilde{P}y_k\big)^T\right] \\ =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(a_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(\tilde{P}y_k\big)^T\right] \\ &+ \mathbf{E}\left[\big(\tilde{P}y_k\big)\big(a_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(\tilde{P}y_k\big)\big(\tilde{P}y_k\big)^T\right] \\ =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(a_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(a_{k-1} \circ x_{k-1}\big)^T\right] F^T \tilde{P}^T \\ &+ \tilde{P}F\,\mathbf{E}\left[\big(a_{k-1} \circ x_{k-1}\big)\big(a_{k-1} \circ (u - x_{k-1})\big)^T\right] + \tilde{P}F\,\mathbf{E}\left[\big(a_{k-1} \circ x_{k-1}\big)\big(a_{k-1} \circ x_{k-1}\big)^T\right] F^T \tilde{P}^T.\end{aligned} \tag{69}$$

Then, from Lemma 2, (69) can be rewritten as

$$\begin{aligned}\Gamma_k =& (1-q)^2\,\mathbf{E}\left[a_{k-1}a_{k-1}^T\right] + q(1-q)\,\mathbf{E}\left[\text{diag}(a_{k-1}a_{k-1}^T)\right] \\ &+ q(1-q)\,\mathbf{E}\left[a_{k-1}a_{k-1}^T\right] F^T \tilde{P}^T - q(1-q)\,\mathbf{E}\left[\text{diag}(a_{k-1}a_{k-1}^T)\right] F^T \tilde{P}^T \\ &+ q(1-q)\tilde{P}F\,\mathbf{E}\left[a_{k-1}a_{k-1}^T\right] - q(1-q)\tilde{P}F\,\mathbf{E}\left[\text{diag}(a_{k-1}a_{k-1}^T)\right] \\ &+ q^2 \tilde{P}F\,\mathbf{E}\left[a_{k-1}a_{k-1}^T\right] F^T \tilde{P}^T + q(1-q)\tilde{P}F\,\mathbf{E}\left[\text{diag}(a_{k-1}a_{k-1}^T)\right] F^T \tilde{P}^T.\end{aligned} \tag{70}$$

By re-arranging terms in (70) and observing that $\Gamma_{k-1} = \mathbf{E}\left[a_{k-1}a_{k-1}^T\right]$ and $\text{diag}(\Gamma_{k-1}) = \mathbf{E}\left[\text{diag}(a_{k-1}a_{k-1}^T)\right]$, the result in (54) follows.

Additionally, from Lemma 2, it follows that

$$\begin{aligned}\Phi_{k+1} = \mathbf{E}\left[y_{k+1}y_{k+1}^T\right] &= F\,\mathbf{E}\left[(a_k \circ x_k)(x_k \circ a_k)^T\right] F^T \\ &= F\Big[q^2\,\mathbf{E}\left[a_k a_k^T\right] + q(1-q)\,\mathbf{E}\left[\text{diag}(a_k a_k^T)\right]\Big] F^T \\ &= F\big[q^2\Gamma_k + q(1-q)\text{diag}(\Gamma_k)\big] F^T.\end{aligned} \tag{71}$$

To obtain the iterations for Ξ_k and Υ_k, the developments are very similar to the ones above. For $k = 0$,

$$\Xi_0 = \mathbf{E}\left[a_0 b_0^T\right] = \tilde{P}\,\mathbf{E}[y_0 z_0^T]\tilde{P}^T = \tilde{P} y_0 z_0^T \tilde{P}, \tag{72}$$

and

$$\begin{aligned}\Upsilon_1 = \mathbf{E}[y_1 z_1^T] &= \mathbf{E}\left[F(a_0 \circ x_0)(x_0 \circ b_0)^T F^T\right] = F\,\mathbf{E}\left[(a_0 \circ x_0)(x_0 \circ b_0)^T\right] F^T \\ &= (qF\tilde{P}y_0)(qF\tilde{P}z_0)^T + q(1-q)F\,\mathbf{E}[\mathrm{diag}(\tilde{P}y_0 z_0^T \tilde{P}^T)]F^T \\ &= (qPy_0)(qPz_0)^T + q(1-q)F\mathrm{diag}(\tilde{P}y_0 z_0^T \tilde{P}^T)F^T \\ &= \overline{y}_1 \overline{z}_1^T + q(1-q)F\mathrm{diag}(\tilde{P}y_0 z_0^T \tilde{P}^T)F^T,\end{aligned} \tag{73}$$

where again we used the fact that $F\tilde{P} = P$ $\big($refer to Eq. (32)$\big)$.

For $k \geq 1$, from Lemma 2 and (28), and taking into account that $y_{k+1} = F(a_k \circ x_k)$ and $z_{k+1} = F(b_k \circ x_k)$, it follows that

$$\begin{aligned}\Xi_k =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1}) + \tilde{P}y_k\big)\big(b_{k-1} \circ (u - x_{k-1}) + \tilde{P}z_k\big)^T\right] \\ =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(b_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(\tilde{P}z_k\big)^T\right] \\ &+ \mathbf{E}\left[\big(\tilde{P}y_k\big)\big(b_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(\tilde{P}y_k\big)\big(\tilde{P}z_k\big)^T\right] \\ =& \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(b_{k-1} \circ (u - x_{k-1})\big)^T\right] + \mathbf{E}\left[\big(a_{k-1} \circ (u - x_{k-1})\big)\big(b_{k-1} \circ x_{k-1}\big)^T\right] F^T \tilde{P}^T \\ &+ \tilde{P}F\,\mathbf{E}\left[\big(a_{k-1} \circ x_{k-1}\big)\big(b_{k-1} \circ (u - x_{k-1})\big)^T\right] + \tilde{P}F\,\mathbf{E}\left[\big(a_{k-1} \circ x_{k-1}\big)\big(b_{k-1} \circ x_{k-1}\big)^T\right] F^T \tilde{P}^T \\ =& (1-q)^2\,\mathbf{E}\left[a_{k-1} b_{k-1}^T\right] + q(1-q)\,\mathbf{E}\left[\mathrm{diag}(a_{k-1} b_{k-1}^T)\right] \\ &+ q(1-q)\,\mathbf{E}\left[a_{k-1} b_{k-1}^T\right] F^T \tilde{P}^T - q(1-q)\,\mathbf{E}\left[\mathrm{diag}(a_{k-1} b_{k-1}^T)\right] F^T \tilde{P}^T \\ &+ q(1-q)\tilde{P}F\,\mathbf{E}\left[a_{k-1} b_{k-1}^T\right] - q(1-q)\tilde{P}F\,\mathbf{E}\left[\mathrm{diag}(a_{k-1} b_{k-1}^T)\right] \\ &+ q^2 \tilde{P}F\,\mathbf{E}\left[a_{k-1} b_{k-1}^T\right] F^T \tilde{P}^T + q(1-q)\tilde{P}F\,\mathbf{E}\left[\mathrm{diag}(a_{k-1} b_{k-1}^T)\right] F^T \tilde{P}^T.\end{aligned} \tag{74}$$

By re-arranging terms in (74) and observing that $\Xi_{k-1} = \mathbf{E}\left[a_{k-1} b_{k-1}^T\right]$ and $\mathrm{diag}(\Xi_{k-1}) = \mathbf{E}\left[\mathrm{diag}(a_{k-1} b_{k-1}^T)\right]$, the result in (58) follows. Finally,

$$\begin{aligned}\Upsilon_{k+1} = \mathbf{E}\left[y_{k+1} z_{k+1}^T\right] &= F\,\mathbf{E}\left[(a_k \circ x_k)(x_k \circ b_k)^T\right] F^T \\ &= F\Big[q^2\,\mathbf{E}\left[a_k b_k^T\right] + q(1-q)\,\mathbf{E}\left[\mathrm{diag}(a_k b_k^T)\right]\Big] F^T \\ &= F\left[q^2 \Xi_k + q(1-q)\mathrm{diag}(\Xi_k)\right] F^T,\end{aligned} \tag{75}$$

which completes the proof. ■

Although omitted in the statement of Lemma 3, it is easy to see that the dynamics of $\Delta_k = \mathbf{E}[b_k a_k^T]$ and $\Theta_k = \mathbf{E}[z_k y_k^T]$ can also be obtained by noting that $\Delta_k = \Psi_k^T$ and $\Theta_k = \Upsilon_k^T$.

V. Convergence Analysis of Robustified Double-Iteration Algorithm

The previous Section established that the iterations governing the evolution of Γ_k, Ψ_k and Ξ_k are identical except for the initial conditions. We will show next that the steady-state solutions of these iterations are also identical up to a multiplicative constant. To see this, we will rewrite (54), (56), and (58) in vector form using Kronecker products. For given matrices C, A, and B of appropriate dimensions, the matrix equation $C = AXB$ (where X is an unknown matrix) can be rewritten as a set of linear equations of the form $(B^T \otimes A)x = c$, where x and c are the vectors that result from stacking the columns of matrices X and C respectively, and $\otimes$ denotes the Kronecker product[4] of matrices [26]. Let γ_k be the vector that results from stacking the columns of Γ_k and $\tilde{\gamma}_k$ the vector that results from stacking the columns of $\text{diag}(\Gamma_k)$. Then, it can be easily seen that (54) can be rewritten as

$$\gamma_k = \big[[q\tilde{P}F + (1-q)I] \otimes [q\tilde{P}F + (1-q)I]\big]\gamma_{k-1} + \big[q(1-q)[I - \tilde{P}F] \otimes [I - \tilde{P}F]\big]\tilde{\gamma}_{k-1},\ k \geq 1. \tag{76}$$

Let G be a diagonal matrix with entries $G\big((l-1)n^2 + l, (l-1)n^2 + l\big) = 1,\ \forall l = 1, 2, \ldots, n^2$, and zero otherwise. Then, the second term on the right hand side of (76) can be written as

$$q(1-q)\big([I - \tilde{P}F] \otimes [I - \tilde{P}F]\big)\tilde{\gamma}_{k-1} = q(1-q)\big([I - \tilde{P}F] \otimes [I - \tilde{P}F]\big)G\gamma_{k-1},\ k \geq 1, \tag{77}$$

which leads us to

$$\gamma_k = \big[[q\tilde{P}F + (1-q)I] \otimes [q\tilde{P}F + (1-q)I] + q(1-q)\big([I - \tilde{P}F] \otimes [I - \tilde{P}F]\big)G\big]\gamma_{k-1},\ k \geq 1. \tag{78}$$

Let ψ_k and ξ_k and δ_k be the vectors that result from stacking the columns of Ψ_k, Ξ_k and Δ_k respectively. Then, it is easy to see that the same recurrence relation as in (78) governs the evolution of ψ_k and ξ_k.

Theorem 1: Let $P \in \mathbb{R}^{n\times n}$ be a column stochastic and primitive weight matrix associated with a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with $\mathcal{V} = \{1, 2, \ldots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Let $F = [I_n\ I_n\ \ldots\ I_n] \in \mathbb{R}^{n\times n^2}$, where I_n is the $n \times n$ identity matrix, and $\tilde{P} = [E_1P^T\ E_2P^T\ \ldots\ E_nP^T]^T \in \mathbb{R}^{n^2\times n}$, where each $E_i \in \mathbb{R}^{n\times n},\ i \in \{1, 2, \ldots, n\}$, satisfies $E_i(i, i) = 1$ and has all other entries equal to zero. Then, for any q, $0 < q \leq 1$, the matrix Π defined as

$$\Pi \equiv [q\tilde{P}F + (1-q)I] \otimes [q\tilde{P}F + (1-q)I] + q(1-q)\big([I - \tilde{P}F] \otimes [I - \tilde{P}F]\big)G \tag{79}$$

is column stochastic, and it has a single eigenvalue of maximum magnitude at value one.

Proof: We show first column stochasticity of matrix Π. Let $C = q\tilde{P}F + (1-q)I$ and $D = I - \tilde{P}F$, so that $\Pi = C \otimes C + q(1-q)(D \otimes D)G$. We will establish that $C \otimes C$ is column stochastic and also show

[4]The Kronecker product of matrices $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{R}^{p\times q}$ is defined (see, e.g., [26]) as the block matrix

$$A \otimes B := \begin{bmatrix} a_{11}B & \ldots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \ldots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp\times nq}.$$

that the column sums of $D \otimes D$ are all zero. By construction, the entries of the i^{th} column of $\tilde{P} \in \mathbb{R}^{n^2 \times n}$ are all zero, with the possible exception of the ones indexed by $\big((i-1)n + j, i\big),\ i, j = 1, 2, \ldots, n$, each of which corresponds to the (j, i) entry of matrix P. Then, it follows that $\sum_{l=1}^{n^2} \tilde{P}_{li} = \sum_{j=1}^{n} P_{ji} = 1, \forall i = 1, 2, \ldots, n^2$. The matrix $\tilde{P}F \in \mathbb{R}^{n^2 \times n^2}$ is also column stochastic by construction, as it results from horizontally concatenating n times the matrix $\tilde{P}$, i.e., $\tilde{P}F = [\tilde{P}\ \tilde{P}\ \ldots\ \tilde{P}]$; therefore, the matrix C is also column stochastic. The kronecker product of C with itself, results in an $n^4 \times n^4$ block matrix of the form $C \otimes C = [C_1\ C_2 \ldots\ C_{n^2}]$, where $C_j = [c_{1j}C^T\ c_{2j}C^T\ \ldots\ c_{n^2 j}C^T]^T$. Then, it follows that the sum of the entries of the l^{th} column of C_j is $\sum_{m=1}^{n^4} C_j(m, l) = (\sum_{i=1}^{n^2} c_{ij})(\sum_{r=1}^{n^2} c_{rl})$. Since $\sum_{i=1}^{n^2} c_{ij}$ and $\sum_{r=1}^{n^2} c_{rl}$ are the sum of the entries of the j^{th} and l^{th} columns of $C = q\tilde{P}F + (1-q)I_{n^4}$ (which is column stochastic), it follows that $\sum_{m=1}^{n^4} C_j(m, l) = 1$; therefore, $C \otimes C$ is also column stochastic.

Since $\tilde{P}F$ is column stochastic, the column-sums of $D = I - \tilde{P}F$ are zero. The kronecker product of D with itself is of the form $D \otimes D = [D_1\ D_2 \ldots\ D_{n^2}]$, where $D_j = [d_{1j}D^T\ d_{2j}D^T\ \ldots\ d_{n^2 j}D^T]^T$. Using similar arguments as above, it follows that $\sum_{m=1}^{n^4} D_j(m, l) = (\sum_{i=1}^{n^2} d_{ij})(\sum_{r=1}^{n^2} d_{rl}) = 0$, which implies that the column-sums of $D \otimes D$ are zero. The only thing left to establish that Π is column stochastic is to show that all entries of Π are nonnegative (from where it immediately follows that $\Pi = C \otimes C + q(1-q)(D \otimes D)G$ is column stochastic). We argue nonnegativity of Π as follows: due to the sparsity structure of G in (79), the only nonzero entries of $(D \otimes D)G$ will be in columns $(k-1)n^2 + k,\ k = 1, 2, \ldots, n^2$; thus except for entries in these columns, the entries of Π will be identical to the corresponding entries in $C \otimes C$. From the structure of $\tilde{P}F$, entries of $C \otimes C$ and $q(1-q)(D \otimes D)$ can, respectively, take one of the following three forms:

$$\big(qp_{ij} + (1-q)\big)\big(qp_{lm} + (1-q)\big), \tag{80}$$

$$q(1-q)(1-p_{ij})(1-p_{lm}), \tag{81}$$

or

$$qp_{ij}\big(qp_{lm} + (1-q)\big), \tag{82}$$

$$-q(1-q)p_{ij}(1-p_{lm}), \tag{83}$$

or

$$q^2 p_{ij} p_{lm}, \tag{84}$$

$$q(1-q)p_{ij}p_{lm}, \tag{85}$$

where $p_{ij} \geq 0$ and $p_{lm} \geq 0$ are the (i, j) and (l, m) entries of matrix P. For (80) and (81), the corresponding entry of Π is of the form

$$\big(qp_{ij} + (1-q)\big)\big(qp_{lm} + (1-q)\big) + q(1-q)(1-p_{ij})(1-p_{lm}) = qp_{ij}p_{lm} + (1-q), \tag{86}$$

and satisfies $0 \leq qp_{ij}p_{lm} + (1-q) \leq 1$ For (82) and (83), the corresponding entry of Π is of the form

$$qp_{ij}\big(qp_{lm} + (1-q)\big) - q(1-q)p_{ij}(1-p_{lm}) = qp_{ij}p_{lm}, \tag{87}$$

and satisfies $0 \leq qp_{ij}p_{lm} \leq 1$. For (84) and (85), the corresponding entry of Π is of the form

$$q^2 p_{ij}p_{lm} + q(1-q)p_{ij}p_{lm} = qp_{ij}p_{lm}, \tag{88}$$

and satisfies $0 \leq qp_{ij}p_{lm} \leq 1$.

To prove the second assertion, we will show first that matrix $\tilde{P}F$ can be written via a permutation of its indices in the form

$$\begin{bmatrix} U & V \\ 0 & W \end{bmatrix}, \tag{89}$$

where U is an irreducible column stochastic matrix and $\lim_{k\to\infty} W^k = 0$. Since $\tilde{P}F$ is column stochastic, we can assume that it corresponds to the weight matrix of some graph $\tilde{\mathcal{G}} = \{\tilde{\mathcal{V}}, \tilde{\mathcal{E}}\}$. We will show that this graph has a single recurrent class plus a few transient states, from which the decomposition of $\tilde{P}F$ in (89) follows. Let

$$\tilde{\mathcal{V}} = \{(1,1),(2,1),\ldots,(n,1),(1,2),(2,2),\ldots,(n,2),\ldots,(n,n-1),(1,n),(2,n),\ldots,(n,n)\}. \tag{90}$$

From the structure of $\tilde{P}F$, it follows that for any node $(i,j) \in \tilde{\mathcal{V}}$, one-step transitions out of (i,j) are to nodes of the form (m,i), with $i \in \mathcal{N}_m^-$, where $\mathcal{N}_m^-$ is the set in-neighbors of node m in the graph $\mathcal{G}$ (with weight matrix P). From the structure of $\tilde{P}F$, it also follows that there are possibly several rows of $\tilde{P}F$ with all entries equal to zero, which means that a node (i,j) that is associated with such row cannot be reached from any other node; however, as already argued, from nodes of the form (i,j), it is possible to reach nodes of the form (m,i), where $i \in \mathcal{N}_m^-$. Clearly, the nodes corresponding to rows with all entries being zero are transient. Note that the possibility of individual nodes that cannot be reached from any other node being disconnected is ruled out as it is easy to see the only nonzero diagonal entries of $\tilde{P}F$ correspond to diagonal entries of P, which are strictly smaller than one.

Next we will show that from a node (i,j) whose corresponding row in $\tilde{P}F$ has some nonzero entries one can reach any other node (m,l) whose corresponding row in $\tilde{P}F$ has some nonzero entries. This means that all non-transient nodes form a single recurrent class (as already argued all nonzero diagonal entries are strictly smaller than one which means there cannot be absorbing nodes). This follows from the fact that the graph $\mathcal{G}$ is strongly connected, which means that for any $j,l \in \mathcal{V}$, there exists a path between j and l. Let $i_1, i_2, \ldots, i_t$ denote the nodes traversed along the path between j and l. We will show next that for any two non-transient nodes $(i,j),(r,l) \in \tilde{\mathcal{V}}$ there exists a path. As already argued, from (i,j) one can reach in a single hop any node of the form (m,i), where m is a neighbor of node i in the graph $\mathcal{G}$. Since i_1 is the first node traversed in the path between j and l, it follows that $(i_1,i) \in \tilde{\mathcal{V}}$ can be reached in one step from (i,j). By repeatedly using this argument, it follows that the sequence

of nodes $(i_1, i), (i_2, i_1), \ldots, (i_t, i_{t-1}), (r, i_t)$ forms a path between (i, j) and (r, l), which means that any non-transient node can be reached by any other non-transient node; thus, the set of non-transient nodes forms a single recurrent class. Clearly, the vertex set $\tilde{\mathcal{V}}$ can be decomposed into a single recurrent class and possibly several transient nodes. By re-ordering the nodes, it follows that $\tilde{P}F$ can be rewritten as in (89) (see, e.g., [27, p. 126]). Furthermore, since Q in (89) is irreducible, it follows that $qQ+(1-q)I$ (where I is the identity matrix) is primitive. It follows that $C = q\tilde{P}F + (1-q)I$ has a unique largest eigenvalue of value one, i.e., $\lambda_1 = 1$, and $1 > |\lambda_2| \geq \cdots \geq |\lambda_{n^2}|$. Let $\sigma(C) = \{\lambda_1, \lambda_2, \ldots, \lambda_{n^2}\}$. Then, $\sigma(C \otimes C) = \{\lambda_i \lambda_j,\ i = 1, \ldots, n,\ j = 1, \ldots, n\}$, including algebraic multiplicities in both cases [26, p. 245]. Since $\lambda_1 = 1$ is unique (multiplicity one) and $|\lambda_i| < 1,\ i = 2, \ldots, n^2$, it follows that the eigenvalue of $C \otimes C = [q\tilde{P}F + (1-q)I] \otimes [q\tilde{P}F + (1-q)I]$ of largest magnitude also takes value 1 and is unique. Since $C \otimes C$ is column stochastic and $\lambda_1 = 1$ is unique, we know that either $C \otimes C$ is also primitive or it can be decomposed following a permutation of indices to the form [27, p. 126]:

$$\begin{bmatrix} L & M \\ 0 & N \end{bmatrix}, \tag{91}$$

where L is a primitive matrix and $\lim_{k\to\infty} N^k = 0$.

We will show next that $\Pi = C \otimes C + q(1-q)(D \otimes D)G$ has exactly the same nonzero entries as $C \otimes C$ and therefore can be decomposed following the same permutation of indices to the form in (91). As argued before, due to the sparsity structure of G in (79), the only nonzero entries of $(D \otimes D)G$ will be in columns $(k-1)n^2 + k,\ k = 1, 2, \ldots, n^2$, thus except for entries in the aforementioned columns, the nonzero entries of Π will be the same as those in $C \otimes C$. For all other columns in Π (that include nonzero entries in $(D \otimes D)G$), it was shown in (86)–(88) that the nonzero entries of Π are strictly positive, from where it follows that Π has the same sparsity structure as $C \otimes C$, which means that Π can also be decomposed in the form of (91) (for some matrices L', M', N'), and the resulting upper-right block is also a primitive matrix. Therefore, Π has a unique largest eigenvalue at one. ■

The following two lemmas establish that the first and second moments of a_k and b_k, and y_k and z_k converge to the same solution up to a scalar multiplication. These two lemmas will be used to show that as $k \to \infty$, the random vector $v_k = z_k - \alpha y_k$, for $\alpha = \dfrac{\sum_{j=1}^n z_0(j)}{\sum_{j=1}^n y_0(j)}$, will converge almost surely to $v = 0$. This suggests that, as $k \to \infty$, and whenever z_k is nonzero, each node i can obtain an estimate of $\alpha = \frac{\sum_{j=1}^n z_0(j)}{\sum_{j=1}^n y_0(j)}$ by calculating the ratio $y_k(i)/z_k(i)$. We will also show that, in fact, z_k will be larger than some threshold infinitely often.

Lemma 4: The first moments of a_k and b_k (also y_k and z_k asymptotically converge to the same

solution up to scalar multiplication:

$$\lim_{k\to\infty} \overline{z}_k = \alpha \lim_{k\to\infty} \overline{y}_k, \tag{92}$$

$$\lim_{k\to\infty} \overline{b}_k = \alpha \lim_{k\to\infty} \overline{a}_k, \tag{93}$$

where $\alpha = \dfrac{\sum_{j=1}^{n} z_0(j)}{\sum_{j=1}^{n} y_0(j)}$.

Proof: In Lemma 1, it was shown that $\overline{y}_{k+1} = \big[qP+(1-q)I\big]\overline{y}_k$ and $\overline{z}_{k+1} = \big[qP+(1-q)I\big]\overline{z}_k$ with $\overline{y}_1 = qy_0$, and $\overline{z}_1 = qz_0$. Since P is column stochastic and primitive, it follows that $[qP+(1-q)I\big]$ is also column stochastic and primitive. Thus, $\lim_{k\to\infty} \overline{z}_k = \alpha \lim_{k\to\infty} \overline{y}_k$, where from the column stochasticity property it follows that $\sum_{j=1}^{n} \overline{z}_k(j) = q(\sum_{j=1}^{n} z_0(j))$ and $\sum_{j=1}^{n} \overline{y}_k(j) = q\big(\sum_{j=1}^{n} y_0(j)\big),\ \forall k \geq 1$; this implies that $\alpha = \frac{\sum_{j=1}^{n} z_0(j)}{\sum_{j=1}^{n} y_0(j)}$, which establishes (92).

By noting that $\overline{a}_0 = \tilde{P} y_0$, and $\overline{b}_0 = \tilde{P} y_0$, and using the fact that $\tilde{P}$ is column stochastic, it follows that $\sum_{j=1}^{n} \overline{a}_k(j) = \sum_{j=1}^{n} \overline{a}_0(j) = \sum_{j=1}^{n} y_0(j)$ and $\sum_{j=1}^{n} \overline{b}_k(j) = \sum_{j=1}^{n} \overline{b}_0(j) = \sum_{j=1}^{n} z_0(j)$. Since $q\tilde{P}F + (1-q)I$ (i.e., the matrix that governs the dynamics of $\overline{a}_k$ and $\overline{b}_k$) is column stochastic and, as shown in the proof of Theorem 1, has a single largest eigenvalue at value 1, a similar development to the one above can be used to show (93). ■

Lemma 5: Define $w_k = b_k - \alpha a_k$ and denote by χ_k the vector that results from stacking the columns of $X_k := \mathbf{E}[w_k w_k^T]$. Then, it follows that

$$\chi_k = \Pi \chi_{k-1}, \tag{94}$$

with $\chi_0 = \psi_0 + \alpha^2 \gamma_0 - \alpha(\xi_0 + \delta_0)$ and $\sum_{l=1}^{n^4} \chi_0(l) = 0$.

Proof: Since $X_k := \mathbf{E}[w_k w_k^T] = \mathbf{E}[b_k b_k^T] + \alpha^2\, \mathbf{E}[a_k a_k^T] - \alpha(\mathbf{E}[a_k b_k^T] + \mathbf{E}[b_k a_k^T]) = \Psi_k + \alpha^2 \Gamma_k - \alpha(\Xi_k + \Delta_k)$, it follows that $\chi_k = \psi_k + \alpha^2 \gamma_k - \alpha(\xi_k + \delta_k)$. From (78) and subsequent discussion, it follows that $\gamma_k = \Pi \gamma_{k-1}$, $\psi_k = \Pi \psi_{k-1}$, $\xi_k = \Pi \xi_{k-1}$, and $\delta_k = \Pi \delta_{k-1}$, thus $\chi_k = \Pi \psi_{k-1} + \alpha^2 \Pi \gamma_{k-1} - \alpha(\Pi \xi_{k-1} + \Pi \delta_{k-1}) = \Pi(\psi_{k-1} + \alpha^2 \gamma_{k-1} - \alpha(\xi_{k-1} + \delta_{k-1})) = \Pi \chi_{k-1}$.

In Lemma 3, it was shown that $\Gamma_0 = \tilde{P} y_0 y_0^T \tilde{P}^T$, $\Psi_0 = \tilde{P} z_0 z_0^T \tilde{P}^T$, and $\Xi_0 = \tilde{P} y_0 z_0^T \tilde{P}^T = \Delta_0^T$. Since γ_0, ψ_0, ξ_0, and δ_0 result from stacking the columns of Γ_0, Ψ_0, Ξ_0, and Δ_0, it follows that

$$\sum_{l=1}^{n^4} \gamma_0(l) = \sum_{i=1}^{n^2}\sum_{j=1}^{n^2} \Gamma_0(i,j) = \left(\sum_{i=1}^{n} y_0(i)\right)^2, \tag{95}$$

$$\sum_{l=1}^{n^4} \psi_0(l) = \sum_{i=1}^{n^2}\sum_{j=1}^{n^2} \Psi_0(i,j) = \left(\sum_{i=1}^{n} z_0(i)\right)^2, \tag{96}$$

$$\sum_{l=1}^{n^4} \xi_0(l) = \sum_{i=1}^{n^2}\sum_{j=1}^{n^2} \Xi_0(i,j) = \left(\sum_{i=1}^{n} y_0(i)\right)\left(\sum_{i=1}^{n} z_0(i)\right),$$

$$\sum_{l=1}^{n^4} \delta_0(l) = \sum_{i=1}^{n^2}\sum_{j=1}^{n^2} \Delta_0(i,j) = \left(\sum_{i=1}^{n} z_0(i)\right)\left(\sum_{i=1}^{n} y_0(i)\right), \tag{97}$$

where the last equality is obtained by taking into account that i) matrix $\tilde{P}$ is column stochastic by construction, and ii) for any $a, b \in \mathbb{R}^n$, we have that $\sum_{i=1}^n \sum_{j=1}^n ab^T(i,j) = (\sum_{l=1}^n a_l)(\sum_{l=1}^n b_l)$. Since $\alpha = \frac{\sum_{j=1}^n z_0(j)}{\sum_{j=1}^n y_0(j)}$, it follows that $\sum_{l=1}^{n^4} \chi_0(l) = \sum_{l=1}^{n^4} (\psi_0(l) + \alpha^2 \gamma_0(l) - \alpha(\xi_0(l) + \delta_0(l))) = 0$. ■

Theorem 2: Let y_k and z_k be the random vectors that result from iterations (28)–(29) and (30)–(31). Define $v_k = z_k - \alpha y_k$, where $\alpha = \frac{\sum_{j=1}^n z_0(j)}{\sum_{j=1}^n y_0(j)}$. Then, $\|v_k\|_\infty \to 0$ almost surely. Furthermore, for every j, $v_k(j) \to 0$ as $k \to \infty$ almost surely (i.e., for every j, $\lim_{k\to\infty} v_k(j) = 0$ with probability one).

Proof: The result follows from the first Borel-Cantelli lemma [28, Theorem 7.3.10]. For all $k \geq 0$ and all $\epsilon > 0$, define the event $E_k(\epsilon) = \{\|v_k\|_\infty > \epsilon\}$. We will first establish an upper bound on $\sum_{k=0}^\infty \Pr\{E_k(\epsilon)\}$ by noting that $\Pr\{E_k(\epsilon)\} = \Pr\{\|v_k\|_\infty > \epsilon\} \leq \frac{\mathbf{E}[\|v_k\|_\infty]}{\epsilon}$, thus $\sum_{k=0}^\infty \Pr\{E_k(\epsilon)\} \leq \frac{1}{\epsilon}\sum_{k=0}^\infty \mathbf{E}\left[\|v_k\|_\infty\right] \leq \frac{1}{\epsilon}\sum_{k=0}^\infty \mathbf{E}\left[\|v_k\|_2\right]$. Note that $\mathbf{E}\left[\|v_k\|_2\right] = (\mathbf{E}[v_k^T v_k]))^{1/2} = (\text{trace}(\mathbf{E}[v_k v_k^T]))^{1/2} = (\text{trace}(\mathbf{E}[z_k z_k^T]) + \alpha^2 \text{trace}(\mathbf{E}[y_k y_k^T]) - 2\alpha \text{trace}(\mathbf{E}[y_k z_k^T])^{1/2}$. We will next show that $\mathbf{E}\left[\|v_k\|_2\right] \to 0$ as $k \to \infty$ geometrically fast. Using Lemma 3, it can be established that $\mathbf{E}[v_k v_k^T] = \mathbf{E}[z_k z_k^T] + \alpha^2 \mathbf{E}[y_k y_k^T] - \alpha(\mathbf{E}[y_k z_k^T] + \mathbf{E}[z_k y_k^T])) = F\left[q^2 X_{k-1} + q(1-q)\text{diag}(X_{k-1})\right]F^T$ where $X_{k-1} = \mathbf{E}[w_k w_k^T]$ as defined in Lemma 5, thus the evolution of $\mathbf{E}[v_k v_k^T]$ is governed by the evolution of X_{k-1} or by χ_{k-1} (the vector that results from stacking the columns of X_{k-1}). In Theorem 1, we showed that Π has a unique eigenvector (with all entries strictly positive) associated to the largest eigenvalue $\lambda_1 = 1$. Then, the solution of (94) is unique and equal to this eigenvector (up to scalar multiplication). Since Π is a column stochastic matrix, and Lemma 5 established that $\sum_{l=1}^{n^4} \chi_0(l) = 0$, it follows that $\sum_{l=1}^{n^4} \chi_k(l) = 0, k \geq 0$, and therefore $\lim_{k\to\infty} \chi_k(l) = 0, \ \forall l$. Additionally, it is well-known that the convergence of (94) is geometric with a rate of convergence given by the eigenvalue λ_2 of Π with the second largest modulus, which satisfies $|\lambda_2| < \lambda_1 = 1$ (see, e.g., [27]). Thus, we have established that $\chi_k(l) \to 0, \ \forall l$, geometrically fast, from where it follows that all the entries of $\mathbf{E}[v_k v_k^T]$ go to zero also geometrically fast. Therefore, the trace$(\mathbf{E}[v_k v_k^T])$ also goes to zero geometrically fast, so that $\mathbf{E}\left[\|v_k\|_2\right]$ also goes to 0 geometrically fast. It immediately follows that $\sum_{k=0}^\infty \mathbf{E}\left[\|v_k\|_2\right] < \infty$ and therefore $\sum_{k=0}^\infty \Pr\{\|v_k\|_\infty \geq \epsilon\} < \infty$. Then, from the first Borel-Cantelli lemma $\Pr\{\|v_k\|_\infty \geq \epsilon$ infinitely often$\} = 0$ (or $\Pr\{\|v_k\|_\infty \geq \epsilon$ i.o.$\} = 0$).

Finally, since, for every j, $\|v_k\|_\infty \geq |v_k(j)|$, then, for every j, $\Pr\{\|v_k\|_\infty \geq \epsilon\} \geq \Pr\{|v_k(j)| \geq \epsilon\}$, and thus, for every j, $\sum_{k=0}^{\infty} \Pr\{|v_k(j)| \geq \epsilon\} < \infty$. Then, by Theorem 7.2.4.c of [28], for every j, $v_k(j) \to 0$ almost surely. ■

Theorem 2 has established that, in the limit as the number of iterations k becomes large, the values of vectors y_k and z_k will be perfectly aligned so that $z_k - \alpha y_k = 0$ with probability one. Thus, in this limiting case, each node j can calculate the value of $\frac{1}{\alpha}$ by taking the ratio $\frac{y_k(j)}{z_k(j)}$, *as long as* $z_k(j) \neq 0$. Note that, as also evidenced by the simulations provided for the small network of Fig. 1 (e.g., the plots on the left and in the middle for Figure 3), the vectors y_k and z_k do not converge in any way;[5] however, the values y_k and z_k become perfectly aligned (with probability one), allowing each node j to calculate $\frac{1}{\alpha} = \frac{y_k(j)}{z_k(j)}$. The only problem here arises when $y_k(j)$ and $z_k(j)$ have both value zero, which does not constitute a violation of $z_k - \alpha y_k = 0$, but clearly does not allow node j to calculate the desired value $\frac{1}{\alpha}$. This is evidenced also in the simulations provided for the small network of Fig. 1: for example, in the plots in Fig. 4, the values of $y_k(j)$ and $z_k(j)$ often go to zero (simultaneously) leaving their ratio undefined.[6] The next two theorems essentially establish that $z_k(j),\ j = 1, 2, \ldots, n$, will be greater than zero (in fact, greater than a constant C that will be specified) infinitely often. Note that, in subsequent developments, $z_k(j)$ is denoted with $z_j[k]$ in order to remain close to the notation in (15)–(17).

Theorem 3: Consider a (possibly directed) strongly connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the iteration in (15)–(17), where $x_{ji}[k]$, $(j,i) \in \mathcal{E}$, $k = 0, 1, 2, ...$, are independent identically distributed (i.i.d.) indicator R.V.'s as defined in (1), i.e., $x_{ji}[k] = 1$ with probability q and $x_{ji}[k] = 0$ with probability $1 - q$, independently between $(j,i) \in \mathcal{E}$ and independently for different k. For every $j = 1, 2, \ldots, n$, define the event $E_k^j = \{z_j[kn] \geq C\},\ k \geq 1$, where $C = \frac{n}{(n+m)(\mathcal{D}_{\max}^+)^{n-1}}$, $\mathcal{D}_{\max}^+ = \max_{j\in\mathcal{V}}\{\mathcal{D}_j^+\}$, $n = |\mathcal{V}|$, and $m = |\mathcal{E}|$. Let ζ_k^j denote the indicator of the event $E_k^j,\ k \geq 1$, i.e., $\zeta_k^j = 1$ whenever $E_k^j,\ k \geq 1$ occurs, and $\zeta_k^j = 0$ otherwise. Then, whatever $\zeta_1, \zeta_2, \ldots, \zeta_{k-1}$, we have that

$$\Pr\{z_j[(k+1)n] \geq C \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} \geq q^n,\ \forall j. \tag{98}$$

Proof: Note that the iteration in (15) to (17) involves nonnegative quantities: since for every j, $z_j[0] > 0,\ \forall j$, it follows from (30)–(31) that, for every j, $z_j[k] \geq 0,\ k \geq 0$. Then, it is not hard to establish that the total mass[7] $\mathcal{M}_{k+1}$ in the system, defined as

$$\mathcal{M}_{k+1} := \sum_{j=1}^{n} z_j[k+1] + \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k] - \tau_{ji}[k-1])(1 - x_{ji}[k])\ , \tag{99}$$

[5]Earlier, we established that, for large k, the quantities $E[y_k]$, $E[z_k]$, $E[y_k y_k^T]$, $E[z_k z_k^T]$ and $E[y_k z_k^T]$ converge, but this does not imply any convergence for the values of y_k or z_k.

[6]Since in the simulations for the plots in Fig. 4, each packet (including self-packets) can be dropped with probability $1 - q$ at iteration k, there is a nonzero probability that all packets destined for node j will be dropped, causing both of its values at the next iteration ($y_{k+1}(j)$ and $z_{k+1}(j)$) to be zero. For instance, in the simulation of Fig. 4, $z_k(1)$ will be zero with probability at least $(1-q)^2 = 0.81$ because node 1 will have value zero if both packets destined for it (including the self-packet) are dropped.

[7]This notion is discussed in great detail in Part II of this paper.

satisfies

$$\mathcal{M}_{k+1} = n \ , \ \text{for all } k = 0, 1, 2, \ldots \ .$$

[This follows from the fact that $M_0 = \sum_{j=1}^{n} z_j[0] = n$ and the observation that

$$\begin{aligned}
\mathcal{M}_{k+1} &:= \sum_{j=1}^{n} z_j[k+1] + \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k] - \tau_{ji}[k-1])(1 - x_{ji}[k]) \\
&= \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k] - \tau_{ji}[k-1]) x_{ji}[k] + \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k] - \tau_{ji}[k-1])(1 - x_{ji}[k]) \\
&= \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k] - \tau_{ji}[k-1]) \\
&= \sum_{(j,i)\in\mathcal{E}} \left(\sigma_{ji}[k-1] + \frac{1}{\mathcal{D}_j^+} z_j[k] - \sigma_{ji}[k-1] x_{ji}[k-1] - \tau_{ji}[k-2](1 - x_{ji}[k-1]) \right) \\
&= \sum_{j=1}^{n} z_j[k] + \sum_{(j,i)\in\mathcal{E}} (\sigma_{ji}[k-1] - \tau_{ji}[k-2])(1 - x_{ji}[k-1]) \ ,
\end{aligned}$$

which is equal to $\mathcal{M}_k$.]

The definition of $\mathcal{M}_{k+1}$ in (99) involves the summation of $n + m$ *nonnegative* quantities, namely, $z_j[k+1]$ for $j = 1, 2, ..., n$ and $m_{ji}[k+1] := (\sigma_{ji}[k] - \tau_{ji}[k-1])(1 - x_{ji}[k])$ for $(j, i) \in \mathcal{E}$. We can think of these quantities as follows: $z_j[k+1]$ is the mass at node j, whereas $m_{ji}[k+1]$ is the mass waiting to get transferred to node j from node i. Since all of these quantities are nonnegative, at least one of them is larger or equal to $\frac{n}{n+m}$. Regardless of whether this quantity is associated with a node (say node j^*) or a link (say link (j^*, i^*)), this mass has at least one way of reaching any node i of interest in graph $\mathcal{G}$ via a path of length at most $n - 1$ (because the graph $\mathcal{G}$ is strongly connected): in particular, there is at least one path of length at most $n - 1$ from node j^* to node i and all the links in this path have weight at least $\frac{1}{\mathcal{D}^+_{\max}}$. If all these links are activated, which occurs with probability q^{n-1} (q^n in the case of link (j^*, i^*) because the mass needs to first transfer to j^*), then a fraction $(\frac{1}{\mathcal{D}^+_{\max}})^{n-1}$ of the mass will transfer to node i in at most n steps. Then, since for every j, $z_j[k] \geq 0, \ k \geq 0$, independently of the values of $z_j[ln], \ l = 1, 2, \ldots, k$, $\Pr\{z_j[(k+1)n] \geq C \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} \geq q^n$ obtains, whatever $\zeta_1, \zeta_2, \ldots, \zeta_{k-1}$. Finally, for every j , $\Pr\{z_j[(k+1)n] \geq C \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} = 1 - \Pr\{z_j[(k+1)n] < C \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} \leq 1 - Pr\{z_j[(k+1)n] = 0 \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} \leq 1 - q^{\mathcal{D}_j^-}$, where $\mathcal{D}_j^-$ is the in-degree of node j. ■

Given a sequence of events $E_1, \ E_2, \ldots, E_n, \ldots$ defined on some probability space, the next theorem (which we do not prove) states the 1912 Borel criterion for establishing whether the event that infinitely many of the E_k occur, denoted by $\{E_k \text{ i.o}\}$, will occur with probability one or zero (see, e.g., [29], [30]). This result, together with the result in Theorem 3 will be used to establish that, for every j, the event $E_k^j = \{z_j[kn] \geq C\}, \ k \geq 1$ occurs infinitely often.

Theorem 4: Let $\{E_k\}, \ k = 1, 2, \ldots$, be a sequence of events defined on some probability space. Let

ζ_k be the indicator function of the event E_k. Let $\Pr\{E_{k+1} \mid \zeta_k, \zeta_{k-1}, \ldots, \zeta_1\}$ denote the conditional probability of the event E_{k+1} given the outcome of previous trials. If $0 < p'_k \leq \Pr\{E_{k+1} \mid \zeta_k, \zeta_{k-1}, \ldots, \zeta_1\} \leq p''_k$ for every k, whatever $\zeta_1, \zeta_2, \ldots, \zeta_k$, then i) $\Pr\{E_k \text{ i.o.}\} = 0$ if $\sum_{k=1}^{\infty} p''_k \leq \infty$, and ii) $\Pr\{E_k \text{ i.o.}\} = 1$ if $\sum_{k=1}^{\infty} p'_k = \infty$.

Theorem 5: Consider a (possibly directed) strongly connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the iteration in (15)–(17). For every $j = 1, 2, \ldots, n$, define the event $E_k^j = \{z_j[kn] \geq C\},\ k \geq 1$, where $C = \frac{n}{(n+m)(\mathcal{D}^+_{\max})^{n-1}}$, $\mathcal{D}^+_{\max} = \max_{j\in\mathcal{V}}\{\mathcal{D}^+_j\}$, $n = |\mathcal{V}|$, and $m = |\mathcal{E}|$. Then, $\Pr\{E_k \text{ i.o.}\} = 1$.

Proof: Theorem 3 established that, for every j, $\Pr\{z_j[(k+1)n] \geq C \mid \zeta_k^j, \zeta_{k-1}^j, \ldots, \zeta_1^j\} \geq q^n$. Define $p'_k = q^n$, then it follows that $\sum_{k=1}^{\infty} p'_k = \infty$, and by the second assertion in Theorem 4, we conclude that, for every j, $\Pr\{E_k^j \text{ i.o.}\} = 1$. ■

The final piece is to establish that whenever $z_j[k] \geq C$, which occurs infinitely often, each node will be able to calculate an estimate of $\overline{v}$ by calculating the ratio $y_j[k]/z_j[k]$ and this estimate will converge to $1/\alpha$ as k goes to infinity.

Theorem 6: For each j, let $k = t_1, t_2, \ldots$ be an increase sequence of time steps for which $z_j[k] > C$. Then, almost surely

$$\lim_{n\to\infty} \left| \frac{y_j[t_n]}{z_j[t_n]} - \frac{1}{\alpha} \right| = 0. \tag{100}$$

Proof: Since $z_j[k] \geq C$ for $k = t_1, t_2, \ldots$, it follows that $\frac{y_j[t_n]}{z_j[t_n]} - \frac{1}{\alpha} \leq \frac{\alpha y_j[t_n] - z_j[t_n]}{\alpha C}$. Also, in the proof of Theorem 3, we established that $\mathcal{M}_k = n, k \geq 0$, from where it follows that $z_j[t_n] \leq n$, therefore $\frac{y_j[t_n]}{z_j[t_n]} - \frac{1}{\alpha} \geq \frac{\alpha y_j[t_n] - z_j[t_n]}{\alpha n}$. In Theorem 2, we established that $|\alpha y_j[k] - z_j[k]| \to 0$ almost surely, which implies that the subsequence $|\alpha y_j[t_n] - z_j[t_n]| \to 0$ almost surely, then since $C < n$, we have that

$$\lim_{n\to\infty} \left| \frac{y_j[t_n]}{z_j[t_n]} - \frac{1}{\alpha} \right| \leq \lim_{n\to\infty} \left| \frac{\alpha y_j[t_n] - z_j[t_n]}{\alpha C} \right| = 0 \tag{101}$$

almost surely. ■

VI. Concluding Remarks

In this paper, we proposed a method to ensure robustness of a class of linear-iterative distributed algorithms against unreliable communication links that may drop packets. We used statistical-moment analysis and the Borel-Cantelli lemmas to establish the correctness of the proposed robustified algorithm. In Part II of this paper, we establish similar convergence properties by recasting the problem as a finite inhomogeneous Markov and using coefficients of ergodicity commonly to used in analyzing this type of Markov chains.

References

[1] N. Lynch, *Distributed Algorithms*. San Mateo, CA: Morgan Kaufmann Publishers, 1996.

[2] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[4] M. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN)*, 2004, pp. 20–27.

[5] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[6] J. Hromkovic, R. Klasing, A. Pelc, P. Ruzicka, and W. Unger, *Dissemination of Information in Communication Networks*. Springer-Verlag, 2005.

[7] J. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[8] A. D. Domínguez-García and C. N. Hadjicostis, "Distributed algorithms for control of demand response and distributed energy resources," in *Proc. IEEE Conference on Decision and Control*, 2011 (to appear).

[9] ——, "Distributed algorithms for resource coordination in networked systems described by directed graph," under review.

[10] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

[11] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation and consensus using linear iterative strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 650–660, May 2008.

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

[13] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, Feb. 2005.

[14] A. Olshevsky and J. Tsitsklis, "Convergence speed in distributed consensus and averaging," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 33–55, 2009.

[15] R. Olfati-Saber and R. M. Murray, "Agreement problems in networks with directed graphs and switching topology," in *Proceedings of American Control Conference*, vol. 4, 2003, pp. 4123–4132.

[16] A. D. Domínguez-García and C. N. Hadjicostis, "Distributed strategies for average consensus in directed graphs," in *Proc. IEEE Conference on Decision and Control*, 2011 (to appear).

[17] ——, "Average consensus on directed graphs," under review.

[18] A. D. Domínguez-García and C. N. Hadjicostis, "Coordination and control of distributed energy resources for provision of ancillary services," in *Proc. IEEE SmartGridComm*, 2010, pp. 537 – 542.

[19] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495 –1508, July 2011.

[20] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, 2011, accepted. [Online]. Available: http://motion.me.ucsb.edu/pdf/2009b-pbb.pdf

[21] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proc. of IEEE International Symposium on Information Theory*, June 2010, pp. 1753 –1757.

[22] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. IEEE Symposium on Foundations of Computer Science*, Oct. 2003, pp. 482 – 491.

[23] S. Patterson, B. Bamieh, and A. El Abbadi, "Distributed average consensus with stochastic communication failures," in *Proc. of IEEE Conference on Decision and Control*, Dec. 2007, pp. 4215–4220.

[24] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 102–133, 2009.

[25] Y. Chen, R. Tron, A. Terzis, and R. Vidal, "Corrective consensus: Converging to the exact average," in *Proc. IEEE Conference on Decision and Control*, Dec. 2010, pp. 1221–1228.

[26] R. Horn and C. Johnson, *Topics in Matrix Analysis*. New York, NY: Cambridge University Press, 1991.

[27] E. Seneta, *Non-negative Matrices and Markov Chains*, revised printing ed. New York, NY: Springer, 2006.

[28] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press, 1992.

[29] E. Borel, "Sur un probleme de probabilites relatif aux fractions continues," *Mathematische Annalen*, vol. 72, pp. pp. 578–587, 1912.

[30] S. W. Nash, "An extension of the borel-cantelli lemma," *The Annals of Mathematical Statistics*, vol. 25, no. 1, pp. 165–167.