Gossip Algorithm for Multi-Agent Systems via Random Walk

Gabriele Oliva[a,c], Stefano Panzieri[b], Roberto Setola[a], Andrea Gasparri[b]

[a] *Complex Systems & Security Laboratory, University Campus Bio-Medico, via A. del Portillo 21, 00128, Rome, Italy.*
[b] *University Roma Tre, Via della Vasca Navale 79, 00146, Rome, Italy.*
[c] *Corresponding author. Email g.oliva@unicampus.it*

Abstract

This paper proposes an asynchronous gossip framework where agents move according to independent random walks over a location graph and interactions may occur only when two agents share the same location. Our goal is to investigate how average consensus may be achieved when agents' motion occurs over a set of discrete locations with topological constraints. This could be used to model the spreading of information across moving crowds or the coordination of agents monitoring a discrete set of points of interest.

Keywords:
Multi-Agent Systems, Gossip Algorithms, Random Walk, Distributed Averaging

1. Introduction

Over the last two decades, *distributed algorithms* have gained momentum in the scientific community as they represent an effective tool for coordinating multi-agent systems in different domains, ranging from for sensor networks [1, 2] to mobile robotic networks [3, 4].

An important classification of such algorithms arises from the nature of the coordination, i.e., whether the coordination is global, dictated for instance by availability of a common global clock or a common sense of time (i.e., *synchronous* algorithms [5]), or the coordination is completely local and pairwise, dictated for instance by the availability of a local (possibly diverse) clock for each agents (i.e., *asynchronous* algorithms [6]). Gossip-based protocols are a representative example of asynchronous algorithms [7]. The major advantage of gossip approaches comes from the fact that they do not require a global coordination, e.g., no common sense of time is required, and their implementation is, in general, significantly more accessible compared to synchronous protocols.

In this work, we consider a scenario where agents move according to independent random walks over a given location graph, where vertexes denote locations and edges describe the existence of a connection between them. Compared to classical gossip schemes, where the underlying interaction graph is fully connected and pair-wise interactions occur over time according to some probability distribution, in our setting the interaction graph is not determined a priori, but it is the result of proximity-based local interactions modeling the fact agents must necessarily be sharing the same location in order to exchange information (i.e., the topology changes over time, depending on the locations occupied by the agents).

In order to clarify the differences of the proposed approach with previous gossip literature, let us consider the diffusion of news in the immediate aftermath of a terroristic attack. Traditional gossip would be suitable to model a situation where individuals share information by means of cellphones, i.e., without proximity-constraints, while the proposed framework would be more adequate to model how information spreads by word of mouth, i.e., with proximity-constraints, as people escapes from the attack location and meets other individuals that are unaware of the ongoing situation.

1.1. Related Work

We point out that gossip schemes have been widely investigated in the literature [8, 9, 10, 7, 11, 2, 12]. Recent innovations in this field include, among others, fault detection based on gossip [13], finite-time convergent gossiping [14], gossip-based distributed Kalman filtering [15] and gossip-based distributed centroid and common reference frame estimation [3]. Interestingly, works can be found at the state of the art which consider explicitly the presence of topological constraints. Among the others, we mention [16] and [17], where agents' interaction is constrained according to the structure of a given graph; [18], where interactions occur over a graph via multi-hop communication; and [19] where finite-time consensus is achieved over ring networks.

1.2. Contribution

The main contributions of this work are as follows: (i) we characterize the time-varying pairwise communication probabilities among the agents, which are driven by independent random walks over the location graph; (ii) we prove the convergence of the proposed gossip dynamics in expectation (in terms of first moment and second moment) to the average of the initial conditions of the agents.

1.3. Paper Outline

The rest of the paper is organized as follows. In Section 2 preliminary definitions are collected and in Section 3 the problem statement is given and the agents' motion and communication behavior is characterized; in Section 4 the convergence to the average in first and second moments are given, while in Sections 5 and 6 simulation results are given and conclusions are drawn, respectively. Finally, we collect some technical results in the appendix.

2. Preliminaries

Notation: We denote vectors by boldface lowercase letters. The (i, j)-th entry of a matrix A is denoted by $A(i, j)$. We denote by I_n the $n \times n$ identity matrix and by $\mathbf{1}_n$ a vector with n components, all equal to one. We use $\boldsymbol{e}_i$ to denote i-th vector in the canonical basis in $\mathbb{R}^n$, i.e., a vector whose entries are all zeros, except the i-th entry which is equal to one. Moreover, we denote by $\mathrm{Ker}(A)$ the kernel of A and the eigenspace spanned by a vector $\boldsymbol{x} \in \mathbb{R}^n$ as $\mathrm{span}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^n \,|\boldsymbol{y} = \alpha\boldsymbol{x}\,,\, \alpha \in \mathbb{R}\}$. We denote by $\mathbb{N}_{\geq 0}$ the set of nonnegative integers. Finally, given a set X let 2^X be its *power set*, i.e., the set of all possible subsets of X.

Paracontracting Matrices: An $n \times n$ matrix W is said to be *non-expansive* [20] with respect to the Euclidean norm $||\cdot||$ if for all $\boldsymbol{x} \in \mathbb{R}^n$ $||W\boldsymbol{x}|| \leq ||\boldsymbol{x}||$, while it is said to be *paracontracting* [21] with respect to the Euclidean norm $||\cdot||$ if $W\boldsymbol{x} \neq \boldsymbol{x} \Leftrightarrow ||W\boldsymbol{x}|| < ||\boldsymbol{x}||$.

Remark 1. *As noted in [20], a symmetric matrix is paracontracting with respect to the Euclidean norm if and only if all its eigenvalues lie in the interval* $(-1, 1]$.

We now review a theorem [20] that characterizes the convergence of the infinite product of matrices.

Theorem 1 (Theorem 3.1, [20]). *Let us define $S = \{W_j \,|\, j \in \mathbb{J}\}$ as a set of possibly infinite $n \times n$ real matrices indexed by the index set $\mathbb{J} \subseteq \mathbb{N}_{\geq 0}$, and let $\{W_k\}_{k=0}^{\infty}$ be a sequence of matrices selected from S and consider the system $\boldsymbol{x}(k+1) = W_k\, \boldsymbol{x}(k)$, with $k \in \mathbb{N}_{\geq 0}$. Suppose that all $W_j \in S$ are non-expansive with respect to the same vector norm $||\cdot||$ and there exists a subsequence $\{W_{k_i}\}_{i=0}^{\infty}$ of the sequence $\{W_k\}_{k=0}^{\infty}$ such that $\lim_{i\to\infty} W_{k_i} = H$, where H satisfies the following properties: (i) H is paracontracting with respect to $\|\cdot\|$; (ii) $\mathrm{Ker}(I-H) \subseteq \cap_{j\in\mathbb{J}} \mathrm{Ker}(I-W_j)$. Then for any $\boldsymbol{x}(0) \in \mathbb{R}^n$ the sequence $\{\boldsymbol{x}(k)\}_{k=0}^{\infty}$ is convergent and it holds*

$$\lim_{i\to\infty} \boldsymbol{x}(i) \in \mathrm{Ker}(I-H) \subseteq \bigcap_{j\in\mathbb{J}} \mathrm{Ker}(I-W_j).$$

Graph Theory: Let $G = \{V, E\}$ denote a *graph* with a finite number m of nodes $v_i \in V$ with $i \in \{1, \ldots, m\}$ and edges $(v_i, v_j) \in E \subset V \times V$ from node v_i to node v_j. A graph is said to be *undirected* if $(v_i, v_j) \in E$ whenever $(v_j, v_i) \in E$, and it is said to be *directed* otherwise. In the following we will consider undirected graphs. Let the *neighborhood* $\mathcal{N}_i(G)$ of a node v_i over an undirected graph $G = \{V, E\}$ be the set $\mathcal{N}_i(G) = \{v_j \,|\, (v_i, v_j) \in E\}$. Let the *degree* d_i of a node v_i be the number of its incident edges, i.e., $d_i = |\mathcal{N}_i(G)|$. A *path* over a graph $G = \{V, E\}$, starting from a node $v_i \in V$ and ending in a node $v_j \in V$, is a subset of links in E that connect v_i and v_j; the *length* of the path is the cardinality of such set. A *minimum path* that connects v_i and v_j is the path from v_i to v_j of minimum length. The *diameter* δ of a graph G is the maximum length among the minimum paths that connect each possible pair of distinct nodes $v_i, v_j \in V$.
An undirected graph is *connected* if for each pair of nodes v_i, v_j there is a path over G that connects them. Let an $m \times m$ symmetric matrix W having the *same structure* as an undirected graph $G = \{V, E\}$ with m nodes, i.e., such that $(v_i, v_j) \in E$ implies $W_{ij} \neq 0$; it can be shown that W is irreducible if and only if the corresponding undirected graph G is connected. Let the *adjacency matrix* of a graph G be an $m \times m$ matrix A with the same structure as G and $A_{ij} \in \{0, 1\}$, i.e., $A_{ij} = 1$ if $(v_i, v_j) \in E$ and $A_{ij} = 0$, otherwise. Moreover, let the *degree matrix* be the $m \times m$ diagonal matrix D whose diagonal entries are $D_{ii} = d_i$, where d_i is the degree of node v_i over G.

Random Walks: A *random walk* over a graph $G = \{V, E\}$ is a path $\{v(0), v(1), \ldots, v(k), \ldots\}$ describing the uniformly random motion of an agent of the graph G with respect to time, where $v(k) = v_i$ denotes the node v_i occupied by the agent at time k. Briefly, in a random walk an agent moves at time k from a node $v(k)$ to a node $v(k+1) \in \mathcal{N}_{v(k)}(G)$ by selecting a neighboring location with uniform random probability, i.e., with a probability equal to $1/d_{v(k)}$. The aforementioned process can be conveniently described as a Markov chain [22, 23], by considering a matrix M of transition probabilities defined as $M(i,j) = 1/d_i$ if $(v_i, v_j) \in E$, and $M(i,j) = 0$, otherwise. It can be easily shown that $M = D^{-1}A$. The Markov chain representing the random walk can be expressed as $\boldsymbol{p}(k+1) = M^T \boldsymbol{p}(k)$ where $\boldsymbol{p}(k) \in \mathbb{R}^m$ is the probability distribution at time k, i.e., $p_i(k)$ is the probability that the node v_i is visited at time k. Note that $\boldsymbol{p}(k)$ has nonnegative entries and satisfies $\mathbf{1}_m^T \boldsymbol{p}(k) = 1$.

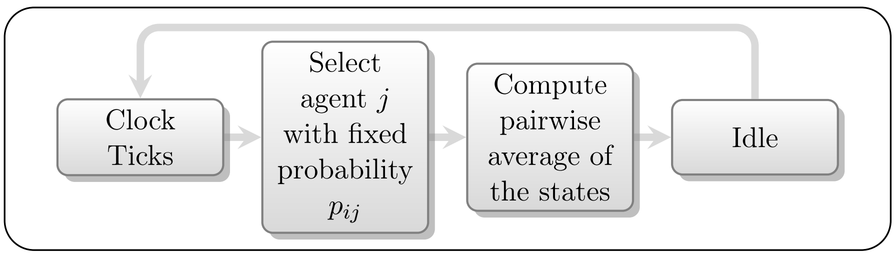

Figure 1: Flow chart representing one iteration within the gossip framework in [7], from the point of view of agent i.

Randomized Gossip: Let us briefly review the randomized gossip framework proposed in [7]; the logical steps that characterize each iteration of such an algorithm are reported in Figure 1. Specifically, let us consider a set of n agents $\{1, \ldots, n\}$, each holding a piece of information or value $x_i(0) \in \mathbb{R}$. Each agent has a clock which ticks at the times of a rate 1 Poisson process, independent across the agents and over time. In this way, although the agents' clock tick in an asynchronous way, the iterator k can be used to denote the k-th ticking. Notice that the probability that the clock of an agent i ticks at time k is equal to $1/n$. When the clock of an agent i ticks at time k, the agent selects another agent $j \in \{1, \ldots, n\} \setminus \{i\}$ with a probability P_{ij}. Then, both agents set their state equal to the average of their current states, according to $\boldsymbol{x}(k+1) = W(k)\,\boldsymbol{x}(k)$, where $\boldsymbol{x}(k) \in \mathbb{R}^n$ is the stacked vector of agents' states at iteration k and matrix $W(k)$ is a random matrix. In particular, let us define W_{ij} as the *pairwise interaction matrix* given by

$$W_{ij} = I_n - \frac{1}{2}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T, \tag{1}$$

with $\boldsymbol{e}_i$ being the i-th vector in the canonical basis in $\mathbb{R}^n$; the random matrix $W(k)$ is such that $W(k) = W_{ij}$ with probability P_{ij}/n. Let us denote with $x_{ave} = \frac{1}{n}\sum_{i=1}^{n} x_i(0)$ the average of the initial conditions and let us define the error $\boldsymbol{y}(k)$ between the state of the agents at iteration k and the average of the initial conditions as $\boldsymbol{y}(k) = \boldsymbol{x}(k) - x_{ave}\mathbf{1}_n$. Briefly, in [7], the authors demonstrate that the first order moment of the agents' state is $E[\boldsymbol{x}(k)] = \overline{W}^k \boldsymbol{x}(0)$, where $\overline{W} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}W_{ij}$ is a time invariant, doubly stochastic and irreducible dynamic matrix. Moreover, they prove that $E[\boldsymbol{y}(k)]$ converges to zero, thus $E[\boldsymbol{x}(k)]$ converges to the average of the initial conditions of the agents. Furthermore, the authors demonstrate that the second order moment $E[\boldsymbol{y}(k)^T\boldsymbol{y}(k)]$ of $\boldsymbol{y}(k)$ converges to zero, in that it holds $E[\boldsymbol{y}(k)^T\boldsymbol{y}(k)] \leq \lambda_2\left(E[W^T W]\right)\|\boldsymbol{y}(0)\|^2$, where, since all matrices $W(k)$ are identically distributed, W represents any of the matrices $W(k)$, while $\lambda_2(\cdot)$ is the eigenvalue with second largest magnitude.

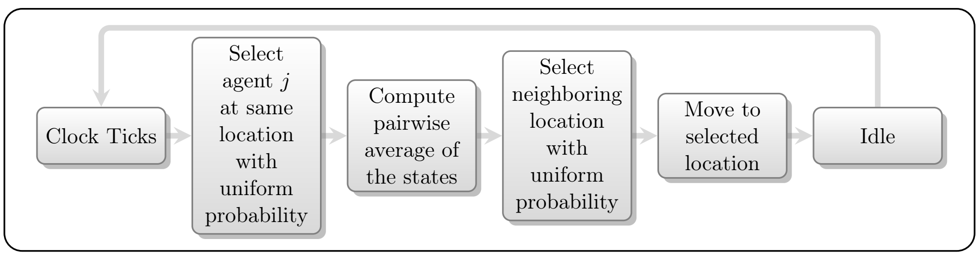

Figure 2: Flow chart representing one iteration within the proposed gossip framework, from the point of view of agent i.

3. Problem Statement and Agents' Behavior

Consider a set of m locations $\ell_1, \ldots, \ell_m$ and a connected and undirected *location graph* $G_\ell = \{V_\ell, E_\ell\}$, having the locations as nodes and featuring a link $(\ell_i, \ell_j) \in E_\ell$ if locations ℓ_i and ℓ_j are connected through a passage. We consider a set of n agents, each holding a value $x_i(0) \in \mathbb{R}$. We follow the theoretical clock ticking framework in [7]. However, differently from [7], the agents move over G_ℓ according to independent random walks; when the clock of an agent i ticks at the k-th iteration, the agent selects an agent j occupying its same location in a uniformly random way and agents i and j performs a standard gossip step. Then, agent i concludes its actions at iteration k by moving to a neighboring location, which is randomly selected with uniform probability.The different phases constituting an iteration of the proposed gossip process are reported in Figure 2. Our objective is to demonstrate the convergence of first and second moment of the proposed gossip scheme to the average of the initial conditions. To this end, we now characterize the agents' motion and communication within the proposed setting.

3.1. Agents' Motion

An agent i, after attempting to communicate, independently selects a neighboring location with uniform probability and moves to the new location, thus executing an iteration of a random walk over G_ℓ. Let $\boldsymbol{p}^{(i)}(k) \in \mathbb{R}^m$ be the probability distribution for the position of agent i over the m locations in V_ℓ at the k-th iteration. The motion of the i-th agent is represented by

$$\boldsymbol{p}^{(i)}(k+1) = H\boldsymbol{p}^{(i)}(k), \tag{2}$$

where

$$H = \frac{n-1}{n} I_m + \frac{1}{n} M^T \quad \text{and} \quad M = D^{-1}A.$$

Notice that the clock of agent i ticks with probability $1/n$ [7], and when it does, the agent moves according to a classical random walk; in particular, $H(i,i) = (n-1)/n$ represents the probability that the clock of the the i-th agent does not tick at iteration k. In other words, matrix H corresponds to the identity matrix (i.e., no motion) with probability $(n-1)/n$ and to M^T (i.e., random walk) with probability $1/n$. Let us define $\eta_{ij}(k)$ as the agents' *meeting probability*, i.e., the probability that agent i and agent j are in the same location at iteration k. We now characterize the structure of the meeting probability $\eta_{ij}(k)$ at each time k.

Proposition 1. *Consider agents i and j moving according to Eq.* (2)*. For all $k \geq 0$ it holds*

$$\eta_{ij}(k) = \begin{cases} \boldsymbol{p}^{(i)}(0)^T \,(H^k)^T\, H^k\, \boldsymbol{p}^{(j)}(0), & \text{if } i \neq j; \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

Proof Let $p_q^{(ij)}(k)$ be the probability that agents i and j with $i \neq j$ are in ℓ_q at iteration k. Since each random walk is independent, it holds $p_q^{(ij)}(k) = p_q^{(i)}(k)p_q^{(j)}(k)$,, while for $i = j$ it holds $p_q^{(ii)}(k) = p_q^{(i)}(k)$. The meeting probability $\eta_{ij}(k)$ is given by the sum of the $p_q^{(ij)}(k)$ terms over all locations ℓ_q, i.e.,

$$\eta_{ij}(k) = \begin{cases} \boldsymbol{p}^{(i)}(k)^T \boldsymbol{p}^{(j)}(k), & \text{if } i \neq j, \\ \mathbf{1}_m^T \boldsymbol{p}^{(j)}(k), & \text{otherwise.} \end{cases}$$

For $i \neq j$ we have that

$$\boldsymbol{p}^{(i)}(k)^T \boldsymbol{p}^{(j)}(k) = \boldsymbol{p}^{(i)}(0)^T \, (H^k)^T \, H^k \, \boldsymbol{p}^{(j)}(0);$$

for $i = j$, by definition, it holds $\mathbf{1}_m^T \boldsymbol{p}^{(j)}(k) = 1$. The proof is complete. ∎

We now show that the meeting probability $\eta_{ij}(k)$ for any couple of agents i and j becomes greater than zero after at most δ steps, where δ is the diameter of G_ℓ.

Theorem 2. *For all $k \geq \delta$ and $i \neq j$ it holds $\eta_{ij}(k) > 0$.*

Proof From Proposition 1, $\eta_{ij}(k)$ is the inner product between $\boldsymbol{p}^{(i)}(k)$ and $\boldsymbol{p}^{(j)}(k)$. Since all entries of $\boldsymbol{p}^{(i)}(k)$ and $\boldsymbol{p}^{(j)}(k)$ are nonnegative by definition, a sufficient condition for their inner product to be positive is that each entry of the two vectors is positive. From Eq. (2), it follows that if $p_q^{(i)}(k) > 0$, then $p_h^{(i)}(k+1) > 0$ for all $\ell_h \in \mathcal{N}_q(G_\ell)$. Note that $\frac{n-1}{n} I_m$ and $\frac{1}{n} M^T$ commute with respect to product, hence the *Binomial Theorem* applies and it holds

$$H^k = \sum_{h=0}^{m} \binom{k}{h} \left(\frac{n-1}{n}\right)^{k-n} \left(\frac{1}{n}\right)^h (M^T)^h.$$

Therefore, since M and $\boldsymbol{p}^{(i)}(0)$ have just nonnegative entries, it follows that $H^k \, \boldsymbol{p}^{(i)}(0)$ has all positive entries whenever $(M^T)^k \boldsymbol{p}^{(i)}(0)$ has all positive entries. The worst case scenario corresponds to an initial probability distribution for which only one location has probability one; this situation requires the longest propagation time to reach any other node. Since the diameter δ of G_ℓ indicates the length of such longest shortest path, then all entries of $\boldsymbol{p}^{(i)}(k)$ will be positive when $k \geq \delta$. ∎

3.2. Agents' Communication

Let $\boldsymbol{x}(k) \in \mathbb{R}^n$ be the stacked vector of agents' states at iteration k, and let us define $\omega_{ij}(k)$ as the *communication probability* for agents i and j at iteration k, i.e., the probability that agent i communicates[1] with agent j at iteration k. The agents update their states according to

$$\boldsymbol{x}(k+1) = W(k)\,\boldsymbol{x}(k) \tag{4}$$

where $W(k)$ is a random matrix. Specifically, with a time-varying probability $\omega_{ij}(k)$, which will be characterized in the following, $W(k) = W_{ij}$, where

$$W_{ij} = I_n - \frac{1}{2}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T$$

is the *pairwise interaction matrix* and $\boldsymbol{e}_i$ is the i-th vector in the canonical basis in $\mathbb{R}^n$.

We now characterize the communication probability $\omega_{ij}(k)$. To this end we first provide the following supporting lemma.

Lemma 1. *Let $\mathcal{I}_{ij}$ denote the set of nodes that are at the same location as the i-th and j-th agent at time k, and let us denote by $\mathcal{J}_{ij}$ a subset of $\mathcal{I}_{ij}$, i.e., $\mathcal{J}_{ij} \subseteq \mathcal{I}_{ij}$. The conditional probability $\theta_{ij}(k)$ that agent i selects[2] agent j at time k given the fact that: i) the clock of the i-th agent ticks at iteration k; and ii) both agents i and j are at the same location at time k is given by*

$$\theta_{ij}(k) = \sum_{\mathcal{J}_{ij} \subseteq \mathcal{I}_{ij}} \frac{\prod_{h \in \mathcal{J}_{ij}} \eta_{ih}(k) \prod_{h \in \overline{\mathcal{J}}_{ij}} (1 - \eta_{ih}(k))}{|\mathcal{J}_{ij}| + |\{i\} \cup \{j\}|} \tag{5}$$

Proof We define $\hat{\theta}(k, \mathcal{J}_{ij})$ as the the conditional probability that the agents indexed by $\mathcal{J}_{ij}$ are at the same location as agent i and agent j while each other agent is not, given the fact that agent i and agent j are at the same location. Such a probability corresponds to the product of the meeting probabilities for the agents indexed by $\mathcal{J}_{ij}$ and of the complement to one of the meeting probabilities for the agents indexed by $\overline{\mathcal{J}}_{ij}$, i.e.,

$$\hat{\theta}(k, \mathcal{J}_{ij}) = \prod_{h \in \mathcal{J}_{ij}} \eta_{ih}(k) \prod_{h \in \overline{\mathcal{J}}_{ij}} (1 - \eta_{ih}(k)).$$

[1]Notice that we explicitly model by $\omega_{ii}(k)$ the probability that the clock of agent i ticks, but no other agent is selected for communication.

[2]Notice that we allow i to select itself, and in this case no communication occurs.

Moreover, the conditional probability $\tilde{\theta}_{ij}(k, \mathcal{J}_{ij})$ that agent i selects agent j uniformly at random, given the occurrence of the above event is

$$\tilde{\theta}_{ij}(k, \mathcal{J}_{ij}) = \frac{1}{|\mathcal{J}_{ij}| + |\{i\} \cup \{j\}|}.$$

Therefore, the probability $\theta^*_{ij}(k, \mathcal{J}_{ij})$ that the agents indexed by $\mathcal{J}_{ij}$ are at the same location as agent i and agent j (while all agents indexed by $\overline{\mathcal{J}}_{ij}$ are not) and that agent j is selected at iteration k corresponds to

$$\theta^*_{ij}(k, \mathcal{J}_{ij}) = \hat{\theta}(k, J^*_{ij})\, \tilde{\theta}_{ij}(k, \mathcal{J}_{ij}).$$

The probability $\theta_{ij}(k)$ is the sum of the probabilities $\theta^*_{ij}(k, \mathcal{J}_{ij})$ for all possible $\mathcal{J}_{ij} \subseteq \mathcal{I}_{ij}$; in other words

$$\theta_{ij}(k) = \sum_{\mathcal{J}_{ij} \subseteq \mathcal{I}_{ij}} \theta^*_{i,j}(k, \mathcal{J}_{ij}),$$

which is the thesis. ■

Theorem 3 (Communication Probability). *For all $k \geq 0$ the communication probability $\omega_{ij}(k)$ is given by*

$$\omega_{ij}(k) = \frac{1}{\mathrm{n}} \eta_{ij}(k) \theta_{ij}(k) \tag{6}$$

Proof Let us point out that $\omega_{ij}(k)$ is the joint probability of two events occurring at iteration k: (i) the clock of the i-th agent ticks and (ii) agent i selects agent j for communication. Event (i) is independent and, as noted in [7], occurs with probability equal to $1/\,n$, independently on i and k. Event (ii) is dependent on event (i), hence we consider the conditional probability $\widehat{\omega}_{ij}(k)$ that agent i selects agent j at iteration k, given the occurrence of event (i); overall, it can be noted that, it holds $\omega_{ij}(k) = \widehat{\omega}_{ij}(k)/n$. Note that event (ii) can be further decomposed in two events occurring at iteration k: (ii.a) agent j is at the same location as agent i and (ii.b) agent j is chosen uniformly at random among the agents that are at the same location as agent i. Event (ii.a) is independent on (ii.b), and its probability corresponds to the meeting probability $\eta_{ij}(k)$. Conversely, event (ii.b) is conditional to the fact that agent i and agent j are at the same node in G_ℓ, and as shown in Lemma 1 corresponds to $\theta_{ij}(k)$. Hence, it holds $\widehat{\omega}_{ij}(k) = \eta_{ij}(k)\theta_{ij}(k)$. The proof follows. ■

Remark 2. *According to the above Theorem, $\omega_{ij}(k)$ is the probability that agent i communicates with agent j at iteration k. The terms $\omega_{ij}(k)$ capture all possible cases for the communication of two agents, including the case $i = j$ which models the fact that the clock of agent i ticks, but no other agent is selected for communication. Hence, by construction, it holds*

$$\sum_{i=1}^{\mathrm{n}} \sum_{j=1}^{\mathrm{n}} \omega_{ij}(k) = 1, \quad \forall k \geq 0.$$

Corollary 1. *For all iterations $k \geq \delta$, where δ is the diameter of G_ℓ it holds $\omega_{ij}(k) > 0$, with $i, j \in \{1, \dots, \mathrm{n}\}$.*

4. Convergence Analysis

In this section, we prove the convergence to zero of the error $\boldsymbol{y}(k)$ defined as

$$\boldsymbol{y}(k) = \boldsymbol{x}(k) - x_{ave}\mathbf{1}_n, \tag{7}$$

where

$$x_{ave} = \frac{1}{n}\sum_{i=1}^{n} x_i(0).$$

Specifically, we prove convergence with respect to the terms of the first and second moments; to this end, we rely on the technical results provided in the appendix, as well as on the following support lemma.

Lemma 2. *The error dynamics evolves according to*

$$\boldsymbol{y}(k+1) = W(k)\,\boldsymbol{y}(k). \tag{8}$$

Proof The result follows from [7]. Specifically, since by construction $W(k)$ is doubly stochastic, it holds $W(k)\mathbf{1}_n = \mathbf{1}_n$; hence, we have that

$$\boldsymbol{y}(k+1) = x(k+1) - x_{ave}\mathbf{1}_n = W(k)x(k) - W(k)x_{ave}\mathbf{1}_n = W(k)\boldsymbol{y}(k).$$

The thesis follows. ∎

4.1. Convergence of First Moment

We now prove convergence of $\boldsymbol{y}(k)$ to zero in expectation.

Theorem 4. *Consider $\boldsymbol{y}(k)$ defined as in (7), evolving according to (8). Then, for all $k \geq 0$, it holds*

$$\lim_{k\to\infty} \{\boldsymbol{E}\left[\boldsymbol{y}(k)\right]\} = \mathbf{0}_{\mathrm{n}}. \tag{9}$$

Proof We observe that, by Lemma 5, $\boldsymbol{E}[\boldsymbol{y}(k)]$ is in the form of Eq. (12) and, by Lemma 3, it is paracontracting; hence, point (i) in Theorem 1 is satisfied. Moreover, by Corollary 1, all $\omega_{ij}(k) > 0$ for $k \geq \delta$. Therefore, for $k \geq \delta$ each $\boldsymbol{E}[W(k)]$ has only positive entries and thus it is irreducible. Since $\boldsymbol{E}[W(k)]$ is doubly stochastic by construction, by the Perron-Frobenius Theorem, it follows that $\lambda_1(\boldsymbol{E}[W(k)]) = 1$, while $|\lambda_i(\boldsymbol{E}[W(k)])| < 1$ for $i = 2, \ldots, n$, where we denote the i-th largest eigenvalue of $\boldsymbol{E}[W(k)]$ by $\lambda_i(\boldsymbol{E}[W(k)])$. As a consequence, it holds

$$\mathrm{Ker}(I - \boldsymbol{E}[W(k)]) = \mathrm{span}(\mathbf{1}_n), \quad \forall k \geq \delta.$$

As for the case $k \leq \delta$, let us recall that according to Lemma 3, $\boldsymbol{E}[W(k)]$ is a doubly stochastic but not necessarily irreducible matrix, for which it holds

$$(I - \boldsymbol{E}[W(k)])\, \alpha \mathbf{1}_n = \alpha\ (\mathbf{1}_n - \boldsymbol{E}[W(k)]\mathbf{1}_n) = \mathbf{0}_n.$$

Since this holds true for any $\alpha \in \mathbb{R}$ it follows that, by construction,

$$\mathrm{span}(\mathbf{1}_n) \subseteq \mathrm{Ker}(I - \boldsymbol{E}[W(k)]).$$

Hence, also point (ii) of Theorem 1 is satisfied. Therefore, setting

$$H = \lim_{k\to\infty} E[W(k)],$$

we have that it holds

$$\lim_{k\to\infty} \boldsymbol{E}[\boldsymbol{y}(k)] \in \bigcap_{j\in J} \mathrm{Ker}(I - \boldsymbol{E}[W(k)]) = \mathrm{span}(\mathbf{1}_n).$$

We now show that it must hold $\alpha = 0$. By Theorem 1, we have that

$$\lim_{k\to\infty} \boldsymbol{E}[\boldsymbol{y}(k)] = \alpha \mathbf{1}_n,$$

with $\alpha \in \mathbb{R}$. Since, by Lemma 4, it holds

$$\mathbf{1}_n^T \boldsymbol{y}(k) = 0,$$

it follows that

$$\boldsymbol{E}[\mathbf{1}_n^T \boldsymbol{y}(k)] = 0.$$

Finally, by definition, it holds

$$\boldsymbol{E}[\mathbf{1}_n^T \boldsymbol{y}(k)] = \mathbf{1}_n^T \boldsymbol{E}[\boldsymbol{y}(k)];$$

hence, we have that

$$\lim_{k\to\infty} \mathbf{1}_n^T \boldsymbol{E}[\boldsymbol{y}(k)] = \alpha\, \mathbf{1}_n^T \mathbf{1}_n = 0.$$

Thus, it must be $\alpha = 0$. ■

4.2. Convergence of Second Moment

We now prove the convergence of $\boldsymbol{E}\left[\boldsymbol{y}(k)^T \boldsymbol{y}(k)\right]$ to zero as k approaches infinity.

Theorem 5. *Consider* $\boldsymbol{y}(k)$ *defined as in* (7)*, evolving according to* (8)*. Then, it holds*

$$\lim_{k\to\infty} \{\boldsymbol{E}\left[\boldsymbol{y}(k)^T \boldsymbol{y}(k)\right]\} = 0. \tag{10}$$

Proof Let us characterize $\boldsymbol{E}[\boldsymbol{y}(k)^T \boldsymbol{y}(k) | \boldsymbol{y}(k-1)]$. It holds

$$\begin{aligned}\boldsymbol{E}[\boldsymbol{y}(k)^T \boldsymbol{y}(k) | \boldsymbol{y}(k-1)] &= \boldsymbol{y}(k-1)^T \boldsymbol{E}[W(k-1)^T W(k-1)] \boldsymbol{y}(k-1) \\ &= \boldsymbol{y}(k-1)^T \boldsymbol{E}[W(k-1)] \boldsymbol{y}(k-1),\end{aligned}$$

where the last equality follows from Proposition 2. At this point, since $\boldsymbol{E}[W(k-1)]$ is symmetric by construction, we express

$$\boldsymbol{E}[W(k-1)] = \sum_{i=1}^{n} \lambda_i(k-1) \boldsymbol{q}_i(k-1) \boldsymbol{q}_i^T(k-1),$$

where $\boldsymbol{q}_i(k)$ is the eigenvector associated to the i-th largest eigenvalue $\lambda_i(k)$ of $\boldsymbol{E}[W(k)]$. As shown in Theorem 4, it holds $\lambda_1(k-1) = 1$ with $\boldsymbol{q}_1(k-1) = \mathbf{1}_n$ and $|\lambda_i(k-1)| < 1$ for all $i \geq 2$. By Lemma 4, for all k, it holds $\mathbf{1}_n^T \boldsymbol{y}(k) = 0$; therefore, we have that

$$\boldsymbol{E}[\boldsymbol{y}(k)^T \boldsymbol{y}(k) | \boldsymbol{y}(k-1)] = \boldsymbol{y}(k-1)^T Q(k-1) \boldsymbol{y}(k-1),$$

with

$$Q(k-1) = \sum_{i=2}^{n} \lambda_i(k-1)\boldsymbol{q}_i(k-1)\boldsymbol{q}_i^T(k-1).$$

Notice that, by construction, $Q(k-1)$ is symmetric for all k; moreover, by construction, it holds $\boldsymbol{E}[\boldsymbol{z}^T\boldsymbol{z}|\boldsymbol{g}] = |\boldsymbol{E}[\boldsymbol{z}^T\boldsymbol{z}|\boldsymbol{g}]|$ for all $\boldsymbol{z}, \boldsymbol{g} \in \mathbb{R}^n$. Therefore, we have that

$$\begin{aligned}
\boldsymbol{E}[\boldsymbol{y}(k)^T\boldsymbol{y}(k)|\boldsymbol{y}(k-1)] &= \left|\boldsymbol{E}[\boldsymbol{y}(k)^T\boldsymbol{y}(k)|\boldsymbol{y}(k-1)]\right| \\
&= \left|\boldsymbol{y}(k-1)^T Q(k-1)\boldsymbol{y}(k-1)\right| \\
&= |< \boldsymbol{y}(k-1), Q(k-1)\boldsymbol{y}(k-1) >| \\
&\leq \|\boldsymbol{y}(k-1)\|\,\|Q(k-1)\boldsymbol{y}(k-1)\| \\
&\leq \|Q(k-1)\|\,\|\boldsymbol{y}(k-1)\|^2 \\
&\leq |\lambda_2(k-1)|\|\boldsymbol{y}(k-1)\|^2 \\
&= |\lambda_2(k-1)|\,\boldsymbol{y}(k-1)^T\boldsymbol{y}(k-1),
\end{aligned}$$

where we used the Cauchy-Schwarz inequality, i.e., $| < \boldsymbol{x}, \boldsymbol{y} > | \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|$. By resorting to the Law of Total Expectation, we obtain

$$\begin{aligned}
\boldsymbol{E}\left[\boldsymbol{y}(k)^T\boldsymbol{y}(k)\right] &= \boldsymbol{E}\left[\boldsymbol{E}\left[\boldsymbol{y}(k)^T\boldsymbol{y}(k)|\boldsymbol{y}(k-1)\right]\right] \\
&\leq \boldsymbol{E}\left[|\lambda_2(k-1)|\,\boldsymbol{y}(k-1)^T\boldsymbol{y}(k-1)\right] \\
&= |\lambda_2(k-1)|\,\boldsymbol{E}[\boldsymbol{y}(k-1)^T\boldsymbol{y}(k-1)].
\end{aligned}$$

Thus, iterating for all k, we have that

$$\boldsymbol{E}[\boldsymbol{y}(k)^T\boldsymbol{y}(k)] \leq \prod_{h=0}^{k-1}|\lambda_2(h)|\,\boldsymbol{E}[\boldsymbol{y}(0)^T\boldsymbol{y}(0)] = \prod_{h=0}^{k-1}|\lambda_2(h)|\|\boldsymbol{y}(0)\|^2 \tag{11}$$

where the last equality holds true since $\boldsymbol{y}(0)$ is deterministic by definition. By construction, it holds $|\lambda_2(k)| \leq 1$ for all $k \geq 0$, while, by Corollary 1, $|\lambda_2(k)| < 1$, for $k \geq \delta$. The proof follows. ■

5. Simulations

In this section we provide an example to numerically demonstrate the convergence of the proposed gossip scheme. Figure (3) describes an instance of the proposed framework where a location graph G_ℓ with $m = 40$ locations, 161 links and $n = 10$ agents over it is considered, as shown in Figure (3a),

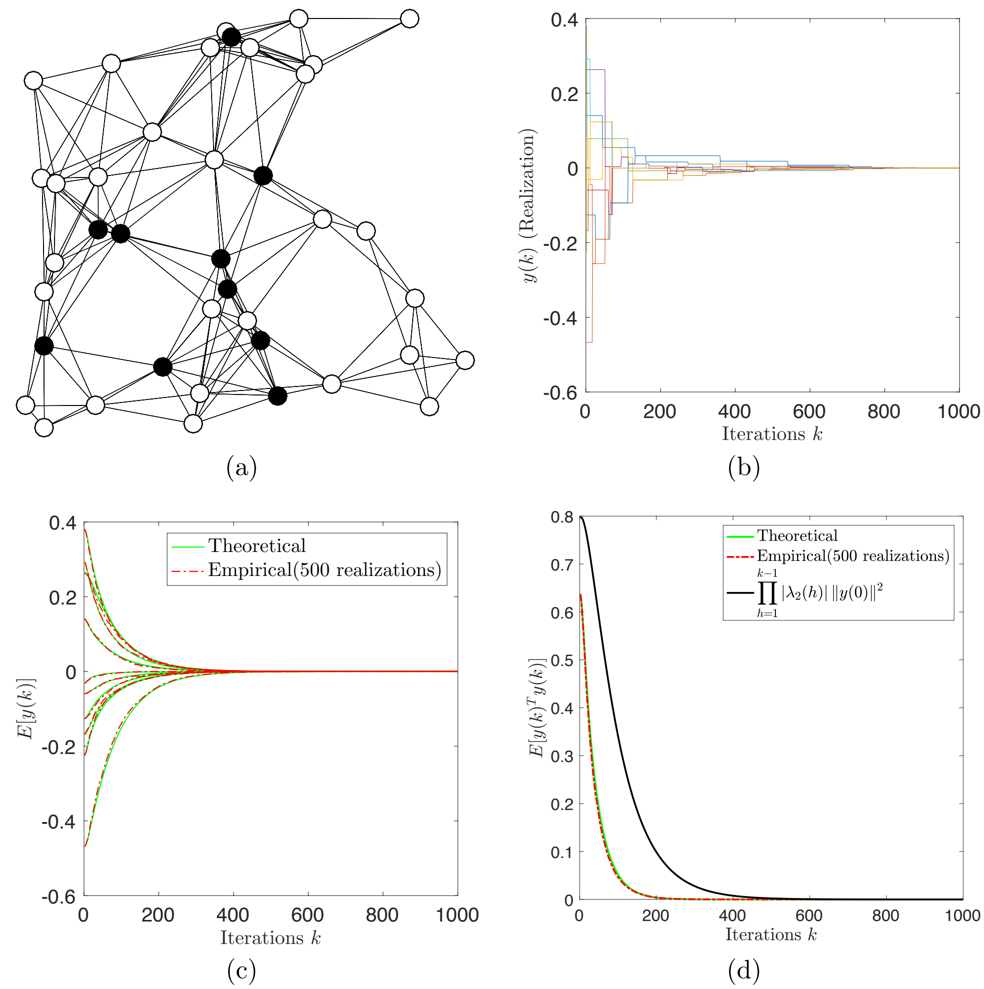

Figure 3: Figure (3a) shows the location graph where nodes filled in black denotes the agents' initial locations. Figure (3b) depicts the error $\boldsymbol{y}(k)$ evolution for a particular realization. Figures (3c) and (3d) illustrate a comparison of the theoretical and empirical first moment $E[\boldsymbol{y}(k)]$ and second moment $E[\boldsymbol{y}(k)^T\boldsymbol{y}(k)]$, respectively, where the empirical moments are obtained by computing an approximation of the expected dynamical matrix $\boldsymbol{E}[W(k)]$ over 500 instances

where black nodes denote locations occupied by the agents at time $k = 0$. Figure (3b) depicts the evolution of the error dynamics $\boldsymbol{y}(k)$ for a particular instance, where it can be noticed that $\boldsymbol{y}(k)$ converges to zero according to Theorems 4 and 5. In this regard, Figures (3c) and (3d) describe the empiric and theoretical convergence of the first and second moments, respectively;

where the empiric moments are obtained by computing an approximation of the expected dynamical matrix $\boldsymbol{E}[W(k)]$ over 500 instances. Notably, Figure (3d) also numerically demonstrates that the expression in Eq. (11), given in Theorem 5, represents an upper-bound on the second moment.

6. Conclusions

In this work we proposed an asynchronous framework for distributed averaging where each agent moves according to independent random walks over a location graph and pairwise interactions may occur only when two agents share a common location. A theoretical analysis to demonstrate the convergence properties of the proposed gossip scheme has been provided, along with numerical simulations to corroborate the theoretical findings. Future work will be mainly focused on two directions: i) characterizing the relation between the topology of the location graph, the initial distribution of the agents' location and the convergence rate of the gossip process; ii) considering a scenario where the walk of the agents is influenced by the gossip process, e.g., in order to implement dynamic patrolling schemes based on the information perceived by the agents during their visit of the different locations.

Appendix

Lemma 3. *Let $W(k)$ a random matrix such that $W(k) = W_{ij}$ with probability $\omega_{ij}(k)$, where W_{ij} is the pairwise interaction matrix and $\omega_{ij}(k)$ is defined in Eq (6). Then, the expected matrix $\boldsymbol{E}[W(k)]$ is paracontracting for all iterations k.*

Proof To prove the lemma, we notice that by definition the expected value of $W(k)$ at iteration k can be obtained according to the communication probability $\omega_{ij}(k)$ as

$$\boldsymbol{E}[W(k)] = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij}(k) W_{ij}.$$

In particular, it holds

$$\boldsymbol{E}[W(k)](i,j) = \begin{cases} \dfrac{\omega_{ij}(k) + \omega_{ji}(k)}{2}, & \text{if } i \neq j, \\ 1 - \sum_{h=1,\, h\neq i}^{n} \dfrac{\omega_{ih}(k) + \omega_{hi}(k)}{2}, & \text{otherwise.} \end{cases}$$

Since $\omega_{ij}(k) \geq 0$ and, by construction, it holds $\sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij}(k) = 1$, we conclude that $\boldsymbol{E}[W(k)](i,j) \geq 0$, for all i, j and the sum over each row/column is equal to one. Moreover, $\boldsymbol{E}[W(k)]$ is symmetric and therefore it has real eigenvalues $\lambda_i(\boldsymbol{E}[W(k)])$. Finally, from the Gershgorin Circle Theorem, if follows that all $\lambda_i(\boldsymbol{E}[W(k)]) \in [0,1]$. Therefore, as discussed in Section 2, matrix $\boldsymbol{E}[W(k)]$ is paracontracting. ∎

Lemma 4. *Consider* $\boldsymbol{y}(k)$ *defined as in* (7)*, evolving according to* (8)*. Then, for all* $k \geq 0$*, it holds* $\mathbf{1}_{\mathrm{n}}^T \boldsymbol{y}(k) = 0$.

Proof We notice that $W(k)$ is doubly stochastic at each iteration k, as pointed out in Lemma 2. Therefore it holds $\mathbf{1}_n^T W(k) = \mathbf{1}_n^T$, for all $k = 0, 1, \ldots$ and

$$\mathbf{1}_n^T \boldsymbol{y}(k) = \mathbf{1}_n^T \prod_{h=0}^{k-1} W(h) \boldsymbol{y}(0) = \mathbf{1}_n^T W(k-1) \prod_{h=0}^{k-2} W(h) \boldsymbol{y}(0);$$

by iterating for all $h = 1, \ldots, k$, we get $\mathbf{1}_n^T \boldsymbol{y}(k) = \mathbf{1}_n^T W(0) \boldsymbol{y}(0) = \mathbf{1}_n^T \boldsymbol{y}(0)$. At this point, since we have that $\boldsymbol{y}(0) = \boldsymbol{x}(0) - x_{ave} \mathbf{1}_n$, we conclude that it holds $\mathbf{1}_n^T \boldsymbol{y}(k) = \mathbf{1}_n^T \boldsymbol{x}(0) - x_{ave} \mathbf{1}_n^T \mathbf{1}_n = 0$. The proof is complete. ∎

Lemma 5. *Consider* $\boldsymbol{y}(k)$ *defined as in* (7)*, evolving according to* (8)*. Then, for all* $k \geq 0$*, it holds*

$$\boldsymbol{E}[\boldsymbol{y}(k)] = \prod_{h=0}^{k-1} \boldsymbol{E}[W(h)] \boldsymbol{y}(0). \tag{12}$$

Proof We point out that, by construction, it holds

$$\boldsymbol{E}[\boldsymbol{y}(k) | \boldsymbol{y}(k-1)] = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij}(k-1) W_{ij} \boldsymbol{y}(k-1) = \boldsymbol{E}\left[W(k-1)\right] \boldsymbol{y}(k-1).$$

By the *Law of Total Expectation* [24], we conclude that it holds

$$\begin{aligned} \boldsymbol{E}[\boldsymbol{y}(k)] &= \boldsymbol{E}[\boldsymbol{E}[\boldsymbol{y}(k) | \boldsymbol{y}(k-1)]] = \boldsymbol{E}[\boldsymbol{E}[W(k-1)] \boldsymbol{y}(k-1)] \\ &= \boldsymbol{E}[W(k-1)] \boldsymbol{E}[\boldsymbol{y}(k-1)]. \end{aligned}$$

At this point, we observe that it holds

$$\boldsymbol{E}[\boldsymbol{y}(k)] = \boldsymbol{E}[W(k-1)] \boldsymbol{E}[\boldsymbol{y}(k-1)] = \boldsymbol{E}[W(k-1)] \boldsymbol{E}[W(k-2)] \boldsymbol{E}[\boldsymbol{y}(k-2)];$$

by iterating this reasoning for all k we conclude that it holds

$$\boldsymbol{E}[\boldsymbol{y}(k)] = \prod_{h=0}^{k-1} \boldsymbol{E}[W(h)]\boldsymbol{E}[\boldsymbol{y}(0)].$$

Being $\boldsymbol{y}(0)$ deterministic, it holds $\boldsymbol{E}[\boldsymbol{y}(0)] = \boldsymbol{y}(0)$. The proof is complete. ■

Proposition 2. *Consider $W(k)$ defined as in Lemma 3, then, for all $k \geq 0$ it holds*

$$\boldsymbol{E}[W(k)^T W(k)] = \boldsymbol{E}[W(k)] \tag{13}$$

Proof By construction, it holds

$$\boldsymbol{E}[W(k)^T W(k)] = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij}(k) W_{ij}^T W_{ij}. \tag{14}$$

Moreover, from the definition of W_{ij}, it follows that

$$W_{ij}^T W_{ij} = W_{ij}^2 = W_{ij}. \tag{15}$$

By plugging Eq. (15) into Eq. (14), we get

$$\boldsymbol{E}[W(k)^T W(k)] = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij}(k) W_{ij} = \boldsymbol{E}[W(k)].$$

The proof is complete. ■

References

[1] P. Ogren, E. Fiorelli, N. E. Leonard, Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment, IEEE Transactions on Automatic control 49 (8) (2004) 1292–1302.

[2] E. Garone, A. Gasparri, F. Lamonaca, Clock synchronization protocol for wireless sensor networks with bounded communication delays, Automatica 59 (C) (2015) 60–72.

[3] M. Franceschelli, A. Gasparri, Gossip-based centroid and common reference frame estimation in multiagent systems, IEEE Transactions on Robotics 30 (2) (2014) 524–531.

[4] A. R. Mosteo, E. Montijano, D. Tardioli, Optimal role and position assignment in multi-robot freely reachable formations, Automatica 81 (2017) 305–313. doi:https://doi.org/10.1016/j.automatica.2017.03.040.

[5] R. Olfati-Saber, J. A. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems, Proceedings of the IEEE 95 (1) (2007) 215–233.

[6] F. Xiao, L. Wang, Asynchronous consensus in continuous-time multi-agent systems with switching topology and time-varying delays, IEEE Transactions on Automatic Control 53 (8) (2008) 1804–1816.

[7] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Randomized gossip algorithms, IEEE/ACM Transactions on Networking (TON) 14 (SI) (2006) 2508–2530.

[8] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Analysis and optimization of randomized gossip algorithms, in: Decision and Control, 2004. CDC. 43rd IEEE Conference on, Vol. 5, IEEE, 2004, pp. 5310–5315.

[9] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Gossip algorithms: Design, analysis and applications, in: 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, Vol. 3, IEEE, 2005, pp. 1653–1664.

[10] L. Xiao, S. Boyd, S. Lall, A scheme for robust distributed sensor fusion based on average consensus, in: Proceedings of the 4th international symposium on Information processing in sensor networks, IEEE Press, 2005, p. 9.

[11] F. Fagnani, S. Zampieri, Asymmetric randomized gossip algorithms for consensus, IFAC Proceedings Volumes 41 (2) (2008) 9051–9056.

[12] M. Blot, D. Picard, M. Cord, N. Thome, Gossip training for deep learning, arXiv preprint arXiv:1611.09726.

[13] D. Silvestre, P. Rosa, J. P. Hespanha, C. Silvestre, Stochastic and deterministic fault detection for randomized gossip algorithms, Automatica 78 (2017) 46–60.

[14] G. Shi, B. Li, M. Johansson, K. H. Johansson, Finite-time convergent gossiping, IEEE/ACM Transactions on Networking 24 (5) (2016) 2782–2794.

[15] K. Ma, S. Wu, Y. Wei, et al., Gossip based distributed tracking in networks of heterogeneous agents, IEEE Communications Letters.

[16] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Gossip and mixing times of random walks on random graphs, Preprint available at http://www. stanford. edu/boyd/gossip gnr. html.

[17] M. Mehyar, D. Spanos, J. Pongsajapan, S. H. Low, R. M. Murray, Distributed averaging on asynchronous communication networks, in: Decision and Control, and European Control Conference. CDC-ECC'05. 44th IEEE Conference on, IEEE, 2005.

[18] A. G. Dimakis, A. D. Sarwate, M. J. Wainwright, Geographic gossip: Efficient aggregation for sensor networks, in: Proceedings of the 5th international conference on Information processing in sensor networks, ACM, 2006, pp. 69–76.

[19] A. Falsone, K. Margellos, S. Garatti, M. Prandini, Finite time distributed averaging over gossip-constrained ring networks, IEEE Transactions on Control of Network Systems.

[20] L. Elsner, R. Bru, M. M. Neumann, Convergence of infinite products of matrices and inner-outer iteration schemes, Electronic Transactions on Numerical Analysis 2 (1994) 183–193.

[21] S. Nelson, M. Neumann, Generalizations of the projection method with applications to sor theory for hermitian positive semidefinite linear systems, Numerische Mathematik 51 (2) (1987) 123–141.

[22] P. G. Snell, P. Doyle, Random walks and electric networks, Free Software Foundation.

[23] P. Diaconis, Group representations in probability and statistics, Lecture Notes-Monograph Series 11 (1988) i–192.

[24] N. A. Weiss, A course in probability, Addison-Wesley, 2006.