

Article

# Analysis of Known Linear Distributed Average Consensus Algorithms on Cycles and Paths

Jesús Gutiérrez-Gutiérrez \*, Marta Zárraga-Rodríguez and Xabier Insausti

Department of Biomedical Engineering and Sciences, Tecnun, University of Navarra, Manuel Lardizábal 13, 20018 San Sebastián, Spain; mzarraga@tecnun.es (M.Z.-R.); xinsausti@tecnun.es (X.I.)

\* Correspondence: jgutierrez@tecnun.es; Tel.: +34-943-219-877 (ext. 2807)

Received: 22 February 2018; Accepted: 22 March 2018; Published: 24 March 2018



**Abstract:** In this paper, we compare six known linear distributed average consensus algorithms on a sensor network in terms of convergence time (and therefore, in terms of the number of transmissions required). The selected network topologies for the analysis (comparison) are the cycle and the path. Specifically, in the present paper, we compute closed-form expressions for the convergence time of four known deterministic algorithms and closed-form bounds for the convergence time of two known randomized algorithms on cycles and paths. Moreover, we also compute a closed-form expression for the convergence time of the fastest deterministic algorithm considered on grids.

**Keywords:** average consensus algorithms; distributed computation; sensor networks; convergence time; number of transmissions

## 1. Introduction

A distributed averaging (or average consensus) algorithm obtains in each sensor the average (arithmetic mean) of the values measured by all the sensors of a sensor network in a distributed way.

The most common distributed averaging algorithms are linear and iterative:

$$\mathbf{x}(t+1) = W(t) \mathbf{x}(t), \quad t \in \{0, 1, 2, \dots\}, \quad (1)$$

where:

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \quad (2)$$

is a real vector,  $n$  is the number of sensors of the network, which we label  $v_j$  with  $j \in \{1, \dots, n\}$ ,  $x_j(0)$  is the value measured by the sensor  $v_j$ ,  $x_j(t)$  is the value computed by the sensor  $v_j$  in time  $t \neq 0$  and the weighting matrix  $W(t)$  is an  $n \times n$  real sparse matrix satisfying that if two sensors  $v_j$  and  $v_k$  are not connected (i.e., if  $v_j$  and  $v_k$  cannot interchange information), then  $[W(t)]_{j,k} = 0$ . From the point of view of communication protocols, there exist efficient ways of implementing synchronous algorithms of the form of (1). (see, e.g., [1]). The linear distributed averaging algorithms can be classified as deterministic or randomized depending on the nature of the weighting matrices  $W(t)$ .

### 1.1. Deterministic Linear Distributed Averaging Algorithms

Several well-known deterministic linear distributed averaging algorithms can be found in [2] and [3]. Those algorithms are time-invariant and have symmetric weights, that is, the deterministic weighting matrix  $W(t)$  is symmetric and does not depend on  $t$  (and consequently,  $\mathbf{x}(t) = W^t \mathbf{x}(0)$ ).

In [2], the authors search among all the symmetric weighting matrices  $W$  the one that makes (1) the fastest possible and show that such a matrix can be obtained by numerically solving a convex

optimization problem. This algorithm is called the fastest linear time-invariant (LTI) distributed averaging algorithm for symmetric weights. It should be mentioned that in [4], the authors proposed an in-network algorithm for finding such an optimal weighting matrix.

In [2], the authors also give a slower algorithm: the fastest constant edge weights algorithm. In this other algorithm, they consider a particular structure of symmetric weighting matrices that depends on a single parameter and find the value of that parameter that makes (1) the fastest possible.

In [3], another two algorithms can be found: the maximum-degree weights algorithm and the Metropolis–Hastings algorithm.

For other deterministic linear distributed averaging algorithms, we refer the reader to [5] and the references therein.

### 1.2. Randomized Linear Distributed Averaging Algorithms

For the randomized case, a well-known linear distributed averaging algorithm was given in [6]. That algorithm is called the pairwise gossip algorithm because only two randomly-selected sensors interchange information at each time instant  $t$ .

Another well-known randomized algorithm can be found in [7]. That algorithm is called the broadcast gossip algorithm because a single sensor is randomly selected at each time instant  $t$  and broadcasts its value to all its neighboring sensors. The broadcast gossip algorithm is a linear distributed consensus algorithm rather than a linear distributed averaging algorithm. However, the broadcast gossip algorithm converges to a random consensus value, which is, in expectation, the average of the values measured by all the sensors of the network. If one uses the directed version of the broadcast gossip algorithm [8] in a symmetric graph, one would converge to the true average.

For other randomized linear distributed averaging algorithms, we refer the reader to [9] and the references therein. The linear distributed averaging algorithms reviewed in Sections 1.1 and 1.2 are the most cited algorithms in the literature on the topic.

### 1.3. Our Contribution

A key feature of a distributed averaging algorithm is its convergence time, because it allows one to establish the stopping criterion for the iterative algorithm. The convergence time is defined as the number of iterations  $t$  required in (1) until the effective value computed by the sensors,  $\mathbf{x}(t)$ , has approached the steady state sufficiently close (to a threshold  $\epsilon$ ). In the literature, we have not found closed-form expressions for the convergence time of the six linear distributed averaging algorithms mentioned in Sections 1.1 and 1.2. A mathematical expression is said to be a closed-form expression if it is written in terms of a finite number of elementary functions (i.e., in terms of a finite number of constants, arithmetic operations, roots, exponentials, natural logarithms and trigonometric functions). In the present paper, we compute closed-form expressions for the convergence time of the deterministic algorithms and closed-form upper bounds for the convergence time of the randomized algorithms on two common network topologies: the cycle and the path. Observe that these closed-form formulas give us upper bounds for the convergence time of the considered algorithms (stopping criteria) on any network that contains as a subgraph a cycle or a path with the same number of sensors. Specifically, in this paper, we compute:

- a closed-form expression for the convergence time of the fastest LTI distributed averaging algorithm for symmetric weights on the considered topologies (see Section 2.1); moreover, we also compute a closed-form expression for the convergence time of this algorithm on a grid;
- a closed-form expression for the convergence time of the fastest constant edge weights algorithm on the considered topologies (see Section 2.2);
- a closed-form expression for the convergence time of the maximum-degree weights algorithm on the considered topologies (see Section 2.3);

- a closed-form expression for the convergence time of the Metropolis–Hastings algorithm on the considered topologies (see Section 2.3);
- closed-form lower and upper bounds for the convergence time of the pairwise gossip algorithm on the considered topologies (see Section 3.1);
- closed-form lower and upper bounds for the convergence time of the broadcast gossip algorithm on the considered topologies (see Section 3.2).

From these closed-form formulas, we study the asymptotic behavior of the convergence time of the considered algorithms as the number of sensors of the network grows. The obtained asymptotic and non-asymptotic results allow us to compare the considered algorithms in terms of convergence time and, consequently, in terms of the number of transmissions required, as well (see Sections 4 and 5). The knowledge of the number of transmissions required lets us know the energy consumption of the distributed technique. The knowledge of the energy consumption is a key factor in the design of a new wireless sensor network (WSN), where one has to decide the number of nodes and the network topology. It should be mentioned that when designing new WSNs, cycles, paths and grids are topologies that are considered frequently.

## 2. Convergence Time of Deterministic Linear Distributed Averaging Algorithms

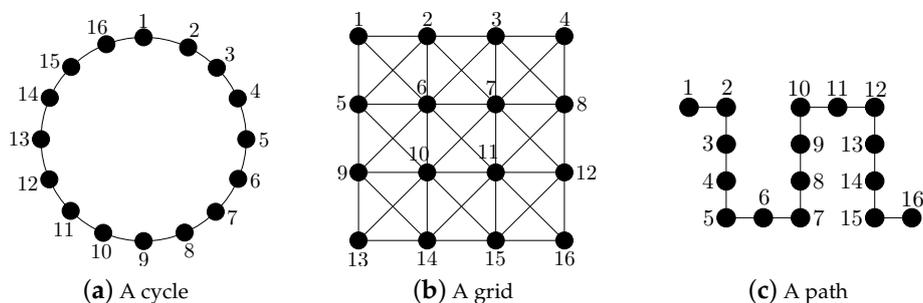
Different definitions of convergence time are used in the literature. We have found three different definitions for the convergence time of a deterministic linear distributed averaging algorithm (see [2,10,11]). In this paper, we consider the definition of  $\epsilon$ -convergence time given in [11]:

$$\tau(\epsilon, \{W(t)\}_{t \geq 0}) := \min \left\{ t_0 : \frac{\|\mathbf{x}(t) - P_n \mathbf{x}(0)\|_2}{\|\mathbf{x}(0) - P_n \mathbf{x}(0)\|_2} \leq \epsilon, \forall t \geq t_0, \forall \mathbf{x}(0) \neq P_n \mathbf{x}(0) \right\}, \quad (3)$$

where  $\epsilon \in (0, 1)$ ,  $\|\cdot\|_2$  is the spectral norm and  $P_n := \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ , with  $\mathbf{1}_n$  being the  $n \times 1$  matrix of ones and  $\top$  denoting the transpose. If we replace the spectral norm by the infinity norm in that definition, we obtain the definition of  $\epsilon$ -convergence time given in [10]. If the deterministic matrix  $W(t)$  in (1) does not depend on  $t$ , we denote the  $\epsilon$ -convergence time by  $\tau(\epsilon, W)$ .

### 2.1. Convergence Time of the Fastest LTI Distributed Averaging Algorithm for Symmetric Weights

In this section, we give a closed-form expression for the  $\epsilon$ -convergence time of the fastest LTI distributed averaging algorithm for symmetric weights, and we study its asymptotic behavior as the number of sensors of the network grows. We consider three common network topologies: the cycle, the grid and the path (see Figure 1).



**Figure 1.** Considered network topologies with 16 sensors.

## 2.1.1. The Cycle

Let:

$$\overset{\circ}{W}_n(\gamma) := \begin{pmatrix} 1-2\gamma & \gamma & 0 & \cdots & 0 & 0 & \gamma \\ \gamma & 1-2\gamma & \gamma & \cdots & 0 & 0 & 0 \\ 0 & \gamma & 1-2\gamma & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-2\gamma & \gamma & 0 \\ 0 & 0 & 0 & \cdots & \gamma & 1-2\gamma & \gamma \\ \gamma & 0 & 0 & \cdots & 0 & \gamma & 1-2\gamma \end{pmatrix}. \quad (4)$$

Using (4), Theorem 1 gives the expression of the weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights on a cycle with  $n$  sensors.

**Theorem 1.** Let  $n \in \mathbb{N}$ , with  $n > 3$ . Then,  $\overset{\circ}{W}_n(\gamma_0)$  is the weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights on a cycle with  $n$  sensors, where:

$$\gamma_0 = \frac{1}{2 - \cos \frac{2\pi}{n} - \cos \frac{2\pi(j_0-1)}{n}}, \quad (5)$$

with:

$$j_0 = \begin{cases} \frac{n}{2} + 1 & \text{if } n \text{ is even,} \\ \frac{n+1}{2} & \text{if } n \text{ is odd.} \end{cases} \quad (6)$$

**Proof.** See Appendix B.  $\square$

We now give a closed-form expression for the  $\epsilon$ -convergence time of the fastest LTI distributed averaging algorithm for symmetric weights on a cycle. We also study the asymptotic behavior of this convergence time as the number of sensors of the cycle grows.

We first introduce some notation: Two sequences of numbers  $\{a_n\}$  and  $\{b_n\}$  are said to be asymptotically equal, and write  $a_n \sim b_n$ , if and only if  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$  (see, e.g., [12] (p. 396)), and, consequently,

$$\tau(\epsilon, \overset{\circ}{W}_n(\gamma_0)) = \Theta(n^2 \log \epsilon^{-1}). \quad (7)$$

Let  $f, g : \mathbb{N} \rightarrow \mathbb{R}$  be two non-negative functions. We write  $f(n) = O(g(n))$  (respectively,  $f(n) = \Omega(g(n))$ ) if there exist  $K \in (0, \infty)$  and  $n_0 \in \mathbb{N}$  such that  $f(n) \leq Kg(n)$  (respectively,  $f(n) \geq Kg(n)$ ) for all  $n \geq n_0$ . If  $f(n) = O(g(n))$  and  $f(n) = \Omega(g(n))$ , then we write  $f(n) = \Theta(g(n))$ .

**Theorem 2.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Let  $\overset{\circ}{W}_n(\gamma_0)$  be as in Theorem 1. Then,

$$\tau(\epsilon, \overset{\circ}{W}_n(\gamma_0)) = \begin{cases} \left\lceil \frac{\log \epsilon^{-1}}{-\log \frac{1 + \cos \frac{2\pi}{n}}{3 - \cos \frac{2\pi}{n}}} \right\rceil & \text{if } n \text{ is even,} \\ \left\lceil \frac{\log \epsilon^{-1}}{-\log \frac{\cos \frac{\pi}{n} + \cos \frac{2\pi}{n}}{2 + \cos \frac{\pi}{n} - \cos \frac{2\pi}{n}}} \right\rceil & \text{if } n \text{ is odd,} \end{cases} \quad (8)$$

where  $\log$  is the natural logarithm and  $\lceil x \rceil$  denotes the smallest integer not less than  $x$ . Moreover,

$$\tau(\epsilon, \overset{\circ}{W}_n(\gamma_0)) \sim \frac{n^2 \log \epsilon^{-1}}{2\pi^2}, \quad (9)$$

**Proof.** See Appendix C.  $\square$

Since the number of transmissions per iteration on a cycle with  $n$  sensors is  $n$  for the fastest LTI distributed averaging algorithm for symmetric weights, the total number of transmissions required for  $\tau(\epsilon, \overset{\circ}{W}_n(\gamma_0))$  iterations is  $T(\epsilon, \overset{\circ}{W}_n(\gamma_0)) := n\tau(\epsilon, \overset{\circ}{W}_n(\gamma_0))$ . From Theorem 2, we obtain:

$$T(\epsilon, \overset{\circ}{W}_n(\gamma_0)) \sim \frac{n^3 \log \epsilon^{-1}}{2\pi^2}, \tag{10}$$

and hence,  $T(\epsilon, \overset{\circ}{W}_n(\gamma_0)) = \Theta(n^3 \log \epsilon^{-1})$ .

2.1.2. The Grid

Let:

$$\tilde{W}_n(\alpha) := \begin{pmatrix} 1-\alpha & \alpha & & & & \\ \alpha & 1-2\alpha & \alpha & & & \\ & & \ddots & \ddots & \ddots & \\ & & & \alpha & 1-2\alpha & \alpha \\ & & & & \alpha & 1-\alpha \end{pmatrix} \tag{11}$$

be the  $n \times n$  matrix for  $n \geq 2$ , and  $\tilde{W}_1(\alpha) := 1$ . We define:

$$\boxtimes W_{r,c}(\alpha) := \tilde{W}_r(\alpha) \otimes \tilde{W}_c(\alpha), \tag{12}$$

where  $\otimes$  is the Kronecker product. Using (12), Theorem 3 gives the expression of the weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights on a grid of  $r$  rows and  $c$  columns.

**Theorem 3.** Let  $r, c \in \mathbb{N}$ , with  $rc > 2$ . Then, the  $rc \times rc$  matrix  $\boxtimes W_{r,c} \left( \frac{1}{2} \right)$  is the weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights on a grid of  $r$  rows and  $c$  columns.

**Proof.** See Appendix D.  $\square$

We now give a closed-form expression for the  $\epsilon$ -convergence time of the fastest LTI distributed averaging algorithm for symmetric weights on a grid of  $r$  rows and  $c$  columns. We also study the asymptotic behavior of this convergence time as the number of rows of the grid grows.

**Theorem 4.** Consider  $\epsilon \in (0, 1)$  and  $r, c \in \mathbb{N}$ , with  $rc > 2$ . Without loss of generality, we assume  $r \geq c$ . Then,

$$\tau \left( \epsilon, \boxtimes W_{r,c} \left( \frac{1}{2} \right) \right) = \left\lceil \frac{\log \epsilon^{-1}}{-\log \cos \frac{\pi}{r}} \right\rceil. \tag{13}$$

Moreover,

$$\tau \left( \epsilon, \boxtimes W_{r,c} \left( \frac{1}{2} \right) \right) \sim \frac{2r^2 \log \epsilon^{-1}}{\pi^2} \tag{14}$$

and consequently,

$$\tau \left( \epsilon, \boxtimes W_{r,c} \left( \frac{1}{2} \right) \right) = \Theta(r^2 \log \epsilon^{-1}). \tag{15}$$

**Proof.** From [2] (Theorem 1), Theorem A1 and (A64), we obtain (13). The rest of the proof runs as the proof of Theorem 2.  $\square$

Since the number of transmissions per iteration on a grid of  $r$  rows and  $c$  columns is  $rc$  for the fastest LTI distributed averaging algorithm for symmetric weights, the total number of transmissions required for  $\tau\left(\epsilon, \overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right)\right)$  iterations is:

$$T\left(\epsilon, \overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right)\right) := rc\tau\left(\epsilon, \overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right)\right). \quad (16)$$

If  $r = c = \sqrt{n}$ , from Theorem 4, we obtain:

$$T\left(\epsilon, \overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right)\right) \sim \frac{2n^2 \log \epsilon^{-1}}{\pi^2}, \quad (17)$$

and hence,  $T\left(\epsilon, \overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right)\right) = \Theta(n^2 \log \epsilon^{-1})$ . Observe that from (13), the optimal configuration for a grid with  $n$  sensors is obtained when  $r = c = \sqrt{n}$ .

### 2.1.3. The Path

Since the path with  $n$  sensors can be seen as a grid of  $n$  rows and one column, from Theorem 3, we conclude that  $\tilde{W}_n\left(\frac{1}{2}\right)$  is the weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights on a path of  $n$  sensors, and from Theorem 4, we conclude that:

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right) = \left\lceil \frac{\log \epsilon^{-1}}{-\log \cos \frac{\pi}{n}} \right\rceil. \quad (18)$$

Moreover,

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right) \sim \frac{2n^2 \log \epsilon^{-1}}{\pi^2} \quad (19)$$

and consequently,

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right) = \Theta(n^2 \log \epsilon^{-1}). \quad (20)$$

Finally, from (16), we obtain:

$$T\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right) \sim \frac{2n^3 \log \epsilon^{-1}}{\pi^2}, \quad (21)$$

and hence,  $T\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right) = \Theta(n^3 \log \epsilon^{-1})$ .

## 2.2. Convergence Time of the Fastest Constant Edge Weights Algorithm

In [2], the authors consider the real symmetric weighting matrices  $W_n(\rho)$  given by:

$$[W_n(\rho)]_{j,k} := \begin{cases} \rho & \text{if } j \neq k, \text{ and } v_j \text{ and } v_k \text{ are connected,} \\ 1 - d_j \rho & \text{if } j = k, \\ 0 & \text{otherwise,} \end{cases} \quad (22)$$

where  $d_j$  denotes the degree of the sensor  $v_j$  (i.e., the number of sensors different from  $v_j$  connected to  $v_j$ ).

Observe that the weighting matrices of the fastest LTI distributed averaging algorithms for symmetric weights given in Section 2.1 for a cycle and a path, namely  $\overset{\circ}{W}_n(\gamma_0)$  and  $\tilde{W}_n\left(\frac{1}{2}\right)$ , can be regarded as  $W_n(\rho)$  in (22) taking  $\rho = \gamma_0$  and  $\rho = \frac{1}{2}$ , respectively. Therefore, the closed-form expression for the  $\epsilon$ -convergence time of the fastest constant edge weights algorithm is given by Theorem 2 on

a cycle and by Theorem 4 on a path. That is, the  $\epsilon$ -convergence time of the fastest constant edge weights algorithm and the  $\epsilon$ -convergence time of the fastest LTI distributed averaging algorithm for symmetric weights is the same on a cycle and on a path.

### 2.3. Convergence Time of the Maximum-Degree Weights Algorithm and of the Metropolis–Hastings Algorithm

For the maximum-degree weights algorithm [3], the weighting matrix considered is the real symmetric matrix  $W_n(\rho)$  in (22) with:

$$\rho = \frac{1}{1 + \max_{j \in \{1, \dots, n\}} d_j}. \quad (23)$$

On the other hand, for the Metropolis–Hastings algorithm [3], the entries of the weighting matrix  $W_n$  are given by:

$$[W_n]_{j,k} = \begin{cases} \frac{[A]_{j,k}}{1 + \max\{d_j, d_k\}} & \text{if } j \neq k, \\ 1 - \sum_{h \in \{1, \dots, n\} \setminus \{j\}} [W_n]_{j,h} & \text{if } j = k, \end{cases} \quad (24)$$

where  $A$  is the adjacency matrix of the network, that is  $A$  is the  $n \times n$  real symmetric matrix given by:

$$[A]_{j,k} = \begin{cases} 1 & \text{if } j \neq k, \text{ and } v_j \text{ and } v_k \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

#### 2.3.1. The Cycle

Observe that the weighting matrices of the maximum-degree weights algorithm and the Metropolis–Hastings algorithm for a cycle with  $n$  sensors can be regarded as  $\overset{\circ}{W}_n(\gamma)$  in (4) taking  $\gamma = \frac{1}{3}$ .

We now give a closed-form expression for the  $\epsilon$ -convergence time of the maximum-degree weights algorithm and of the Metropolis–Hastings algorithm on a cycle. We also study the asymptotic behavior of this convergence time as the number of sensors of the cycle grows.

**Theorem 5.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Then:

$$\tau \left( \epsilon, \overset{\circ}{W}_n \left( \frac{1}{3} \right) \right) = \left\lceil \frac{\log \epsilon^{-1}}{-\log \frac{1 + 2 \cos \frac{2\pi}{n}}{3}} \right\rceil. \quad (26)$$

Moreover,

$$\tau \left( \epsilon, \overset{\circ}{W}_n \left( \frac{1}{3} \right) \right) \sim \frac{3n^2 \log \epsilon^{-1}}{4\pi^2}, \quad (27)$$

and therefore,

$$\tau \left( \epsilon, \overset{\circ}{W}_n \left( \frac{1}{3} \right) \right) = \Theta(n^2 \log \epsilon^{-1}). \quad (28)$$

**Proof.** Combining (A29) and (A30), we obtain:

$$\left\| \overset{\circ}{W}_n \left( \frac{1}{3} \right) - P_n \right\|_2 = \frac{1 + 2 \cos \frac{2\pi}{n}}{3}, \quad (29)$$

and applying [2] (Theorem 1) and Theorem A1, (26) holds. The rest of the proof runs as the proof of Theorem 2.  $\square$

Since the number of transmissions per iteration on a cycle with  $n$  sensors is  $n$  for both algorithms, the total number of transmissions required for  $\tau\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right)$  iterations is  $T\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right) := n\tau\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right)$ . From Theorem 5, we obtain:

$$T\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right) \sim \frac{3n^3 \log \epsilon^{-1}}{4\pi^2}, \quad (30)$$

and thus,  $T\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right) = \Theta(n^3 \log \epsilon^{-1})$ .

### 2.3.2. The Path

Observe that the weighting matrices of the maximum-degree weights algorithm and of the Metropolis–Hastings algorithm for a path with  $n$  sensors can be regarded as  $\tilde{W}_n(\alpha)$  in (11) taking  $\alpha = \frac{1}{3}$ .

We now give a closed-form expression for the  $\epsilon$ -convergence time of the maximum-degree weights algorithm and of the Metropolis–Hastings algorithm on a path. We also study the asymptotic behavior of this convergence time as the number of sensors of the path grows.

**Theorem 6.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Then:

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) = \left\lceil \frac{\log \epsilon^{-1}}{-\log \frac{1+2\cos\frac{\pi}{n}}{3}} \right\rceil. \quad (31)$$

Moreover,

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) \sim \frac{3n^2 \log \epsilon^{-1}}{\pi^2}, \quad (32)$$

and therefore,

$$\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) = \Theta(n^2 \log \epsilon^{-1}). \quad (33)$$

**Proof.** Combining (A63) and [4] (Lemma 1), we obtain:

$$\left\| \tilde{W}_n\left(\frac{1}{3}\right) - P_n \right\|_2 = \frac{1}{3} + \frac{2}{3} \cos \frac{\pi}{n}, \quad (34)$$

and applying [2] (Theorem 1) and Theorem A1, (31) holds. The rest of the proof runs as the proof of Theorem 2.  $\square$

Since the number of transmissions per iteration on a path with  $n$  sensors is  $n$  for both algorithms, the total number of transmissions required for  $\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right)$  iterations is  $T\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) := n\tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right)$ . From Theorem 6, we obtain:

$$T\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) \sim \frac{3n^3 \log \epsilon^{-1}}{\pi^2}, \quad (35)$$

and thus,  $T\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right) = \Theta(n^3 \log \epsilon^{-1})$ .

### 3. Convergence Time of Randomized Linear Distributed Averaging Algorithms

#### 3.1. Lower and Upper Bounds for the Convergence Time of the Pairwise Gossip Algorithm

In the literature, we have found two different definitions for the convergence time of a randomized linear distributed averaging algorithm (see [6,7]). In this subsection, we consider the definition of  $\epsilon$ -convergence time for a randomized linear distributed averaging algorithm given in [6]:

$$\tau(\epsilon, \{W(t)\}_{t \geq 0}) := \sup_{\mathbf{x}(0) \neq 0_{n \times 1}} \inf \left\{ t : \Pr \left( \frac{\|\mathbf{x}(t) - P_n \mathbf{x}(0)\|_2}{\|\mathbf{x}(0)\|_2} \geq \epsilon \right) \leq \epsilon \right\}, \quad (36)$$

where  $\epsilon \in (0, 1)$  and  $\Pr$  denotes probability.

We prove in Theorem A1 (Appendix A) that the definitions of  $\epsilon$ -convergence time in (3) and (36) coincide when applied to deterministic LTI distributed averaging algorithms with symmetric weights (in particular, the four algorithms considered in Section 2). For those algorithms, we also obtain from Theorem A1 that:

$$\tau \left( \frac{1}{e}, W \right) = \lceil \tau'(W) \rceil, \quad (37)$$

where  $\tau'(W)$  denotes the definition of convergence time given in [2].

We recall here that in the pairwise gossip algorithm [6], only two sensors interchange information at each time instant  $t$ . These two sensors  $v_{j_t}$  and  $v_{k_t}$  are randomly selected at each time instant  $t$ , and the weighting matrix  $W(t)$ , which we denote by  $W_P(t)$ , is the symmetric matrix given by:

$$[W_P(t)]_{j,k} = \begin{cases} \frac{1}{2} & \text{if } j, k \in \{j_t, k_t\}, \\ 1 & \text{if } j = k \notin \{j_t, k_t\}, \\ 0 & \text{otherwise,} \end{cases} \quad (38)$$

for all  $j, k \in \{1, \dots, n\}$ .

In [6], a lower and an upper bound for the  $\epsilon$ -convergence time of the pairwise gossip algorithm were introduced. We now give a closed-form expression for those bounds on a cycle and on a path, and we study their asymptotic behavior as the number of sensors of the network grows.

##### 3.1.1. The Cycle

**Theorem 7.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Suppose that  $\overset{\circ}{W}_P(t)$  is the weighting matrix of the pairwise gossip algorithm given in (38) on a cycle with  $n$  sensors, where the edge  $\{v_{j_t}, v_{k_t}\}$  is randomly selected at each time instant  $t \in \mathbb{N} \cup \{0\}$  with probability  $\frac{1}{n}$ . Then:

$$\frac{1}{2} \overset{\circ}{l}_P(\epsilon) \leq \tau \left( \epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0} \right) \leq 3 \overset{\circ}{l}_P(\epsilon), \quad (39)$$

with:

$$\overset{\circ}{l}_P(\epsilon) = \frac{\log \epsilon^{-1}}{-\log \left( 1 + \frac{1}{n} (\cos \frac{2\pi}{n} - 1) \right)}. \quad (40)$$

Moreover,

$$\overset{\circ}{l}_P(\epsilon) \sim \frac{n^3 \log \epsilon^{-1}}{2\pi^2} \quad (41)$$

and:

$$\tau \left( \epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0} \right) = \Theta(n^3 \log \epsilon^{-1}) = \tau \left( \epsilon, \overset{\circ}{W}_n \left( \frac{1}{2n} \right) \right), \quad (42)$$

**Proof.** The entries of the expectation of  $\overset{\circ}{W}_P(0)$  are given by:

$$[E(\overset{\circ}{W}_P(0))]_{j,k} = \begin{cases} \frac{1}{2n} & \text{if } j-k \in \{-1, 1\}, \\ \frac{1}{2n} & \text{if } j-k \in \{1-n, n-1\}, \\ \frac{1}{n}(n-1) & \text{if } j=k, \\ 0 & \text{otherwise,} \end{cases} \quad (43)$$

for all  $j, k \in \{1, \dots, n\}$ . Thus,  $E(\overset{\circ}{W}_P(0)) = \overset{\circ}{W}_n(\frac{1}{2n})$ . Therefore, combining (A29) and [6] (Theorem 3), we obtain (39). The rest of the proof runs as the proof of Theorem 2.  $\square$

Since the number of transmissions per iteration on a cycle with  $n$  sensors is two for the pairwise gossip algorithm, the total number of transmissions required for  $\tau(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0})$  iterations is  $T(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0}) := 2\tau(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0})$ . From Theorem 7, we obtain  $\overset{\circ}{l}_P(\epsilon) \leq T(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0}) \leq 6\overset{\circ}{l}_P(\epsilon)$  and  $T(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0}) = \Theta(n^3 \log \epsilon^{-1})$ .

### 3.1.2. The Path

**Theorem 8.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Suppose that  $\tilde{W}_P(t)$  is the weighting matrix of the pairwise gossip algorithm given in (38) on a path with  $n$  sensors, where the edge  $\{v_{j_t}, v_{k_t}\}$  is randomly selected at each time instant  $t \in \mathbb{N} \cup \{0\}$  with probability  $\frac{1}{n-1}$ . Then:

$$\frac{1}{2}\tilde{l}_P(\epsilon) \leq \tau(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}) \leq 3\tilde{l}_P(\epsilon), \quad (44)$$

with:

$$\tilde{l}_P(\epsilon) = \frac{\log \epsilon^{-1}}{-\log\left(1 + \frac{1}{n-1}(\cos \frac{\pi}{n} - 1)\right)}. \quad (45)$$

Moreover,

$$\tilde{l}_P(\epsilon) \sim \frac{2n^3 \log \epsilon^{-1}}{\pi^2} \quad (46)$$

and:

$$\tau(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}) = \Theta(n^3 \log \epsilon^{-1}) = \tau\left(\epsilon, \tilde{W}_n\left(\frac{1}{2n-2}\right)\right). \quad (47)$$

**Proof.** The entries of the expectation of  $\tilde{W}_P(0)$  are given by:

$$[E(\tilde{W}_P(0))]_{j,k} = \begin{cases} \frac{1}{2n-2} & \text{if } j-k \in \{-1, 1\}, \\ 1 - \frac{1}{n-1} & \text{if } j=k, j \neq 1 \text{ and } j \neq n, \\ 1 - \frac{1}{2n-2} & \text{if } j=k, j \in \{1, n\}, \\ 0 & \text{otherwise,} \end{cases} \quad (48)$$

for all  $j, k \in \{1, \dots, n\}$ . Thus,  $E(\tilde{W}_P(0)) = \tilde{W}_n(\frac{1}{2n-2})$ . Therefore, combining (A63) and [6] (Theorem 3), we obtain (44). The rest of the proof runs as the proof of Theorem 2.  $\square$

Since the number of transmissions per iteration on a path with  $n$  sensors is two for the pairwise gossip algorithm, the total number of transmissions required for  $\tau(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0})$

iterations is  $T\left(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}\right) := 2\tau\left(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}\right)$ . From Theorem 8, we obtain  $\tilde{l}_P(\epsilon) \leq T\left(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}\right) \leq 6\tilde{l}_P(\epsilon)$  and  $T\left(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0}\right) = \Theta(n^3 \log \epsilon^{-1})$ .

### 3.2. Lower and Upper Bounds for the Convergence Time of the Broadcast Gossip Algorithm

We begin this subsection with the definition of  $\epsilon$ -convergence time for a randomized linear distributed averaging algorithm given in [7] (Equation (42)):

$$\tau(\epsilon, \{W(t)\}_{t \geq 0}) := \sup_{\mathbf{x}(0) \neq P_n \mathbf{x}(0)} \inf \left\{ t : \Pr \left( \frac{\|\mathbf{x}(t) - P_n \mathbf{x}(t)\|_2}{\|\mathbf{x}(0) - P_n \mathbf{x}(0)\|_2} \geq \epsilon \right) \leq \epsilon \right\}, \quad (49)$$

where  $\epsilon \in (0, 1)$ .

It can be proven that the definitions of  $\epsilon$ -convergence time in (36) and (49) coincide when applied to algorithms in which the matrix  $W(t)$  satisfies  $W(t)P_n = P_n W(t) = P_n$  for all  $t \in \mathbb{N} \cup \{0\}$  (in particular, the pairwise gossip algorithm and deterministic LTI distributed averaging algorithms with symmetric weights).

Observe that (49) is actually a definition for the convergence time of linear distributed consensus algorithms, not only of linear distributed averaging algorithms.

We recall here that in the broadcast gossip algorithm, a single sensor broadcasts at each time instant  $t$ . This sensor  $v_{j_t}$  is randomly selected at each time instant  $t$  with probability  $\frac{1}{n}$ , and the weighting matrix  $W(t)$  is given by:

$$[W(t)]_{j,k} = \begin{cases} 1 & \text{if } j = k \text{ and } [A]_{j,j_t} = 0, \\ \varphi & \text{if } j = k \text{ and } [A]_{j,j_t} = 1, \\ 1 - \varphi & \text{if } k = j_t \text{ and } [A]_{j,j_t} = 1, \\ 0 & \text{otherwise,} \end{cases} \quad (50)$$

for all  $j, k \in \{1, \dots, n\}$ , where  $\varphi \in (0, 1)$  and  $A$  is the adjacency matrix of the network. We denote by  $W_B(t)$  the weighting matrix in (50) when  $\varphi$  is the optimal parameter:  $\varphi_0$  (see [7] (Section V)).

In [7], a lower and an upper bound for the  $\epsilon$ -convergence time of the broadcast gossip algorithm were introduced. We now give a closed-form expression for  $\varphi_0$  and for those bounds on a cycle and on a path. We also study the asymptotic behavior of the bounds as the number of sensors of the network grows.

#### 3.2.1. The Cycle

**Theorem 9.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Suppose that  $\overset{\circ}{W}_B(t)$  is the weighting matrix in (50) when the network is a cycle with  $n$  sensors and  $\varphi$  is the optimal parameter:  $\overset{\circ}{\varphi}_0$ . Then:

$$\overset{\circ}{\varphi}_0 = 1 - \frac{n}{2(n + \cos \frac{2\pi}{n} - 1)} \quad (51)$$

and:

$$\overset{\circ}{l}_B(\epsilon) \leq \tau\left(\epsilon, \{\overset{\circ}{W}_B(t)\}_{t \geq 0}\right) \leq 6\overset{\circ}{l}_B(\epsilon), \quad (52)$$

with:

$$\overset{\circ}{l}_B(\epsilon) = \frac{\log \epsilon^{-1}}{-2 \log \left( \frac{n + 2 \cos \frac{2\pi}{n} - 2}{n + \cos \frac{2\pi}{n} - 1} \right)}. \quad (53)$$

Moreover,

$$\overset{\circ}{l}_B(\epsilon) \sim \frac{n^3 \log \epsilon^{-1}}{4\pi^2}, \quad (54)$$

and:

$$\tau\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right) = \Theta(n^3 \log \epsilon^{-1}) = \tau\left(\epsilon, \dot{W}_n\left(\frac{1 - \dot{\varphi}_0}{n}\right)\right). \quad (55)$$

**Proof.** See Appendix E.  $\square$

Since the number of transmissions per iteration on a cycle with  $n$  sensors is one for the broadcast gossip algorithm, the total number of transmissions required for  $\tau\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right)$  iterations is  $T\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right) := \tau\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right)$ . From Theorem 9, we obtain  $\dot{l}_B(\epsilon) \leq T\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right) \leq 6\dot{l}_B(\epsilon)$  and  $T\left(\epsilon, \{\dot{W}_B(t)\}_{t \geq 0}\right) = \Theta(n^3 \log \epsilon^{-1})$ .

### 3.2.2. The Path

**Theorem 10.** Consider  $\epsilon \in (0, 1)$  and  $n \in \mathbb{N}$ , with  $n > 3$ . Suppose that  $\tilde{W}_B(t)$  is the weighting matrix in (50) when the network is a path with  $n$  sensors and  $\varphi$  is the optimal parameter:  $\tilde{\varphi}_0$ . Then:

$$\tilde{\varphi}_0 = 1 - \frac{n}{2\left(n + \cos \frac{\pi}{n} - 1\right)} \quad (56)$$

and:

$$\tilde{l}_B(\epsilon) \leq \tau\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right) \leq 6\tilde{l}_B(\epsilon), \quad (57)$$

with:

$$\tilde{l}_B(\epsilon) = \frac{\log \epsilon^{-1}}{-2 \log \left(\frac{n+2 \cos \frac{\pi}{n}-2}{n+\cos \frac{\pi}{n}-1}\right)}. \quad (58)$$

Moreover,

$$\tilde{l}_B(\epsilon) \sim \frac{n^3 \log \epsilon^{-1}}{\pi^2}, \quad (59)$$

and:

$$\tau\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right) = \Theta(n^3 \log \epsilon^{-1}) = \tau\left(\epsilon, \tilde{W}_n\left(\frac{1 - \tilde{\varphi}_0}{n}\right)\right). \quad (60)$$

**Proof.** See Appendix F.  $\square$

Since the number of transmissions per iteration on a path with  $n$  sensors is one for the broadcast gossip algorithm, the total number of transmissions required for  $\tau\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right)$  iterations is  $T\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right) := \tau\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right)$ . From Theorem 10, we obtain  $\tilde{l}_B(\epsilon) \leq T\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right) \leq 6\tilde{l}_B(\epsilon)$  and  $T\left(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0}\right) = \Theta(n^3 \log \epsilon^{-1})$ .

## 4. Discussion

As in this paper we have used the same definition of converge time for both deterministic and randomized linear distributed averaging algorithms (namely, the one in (49)), the results given in Sections 2 and 3 allow us to compare the considered algorithms on a cycle and on a path in terms of convergence time and, consequently, in terms of the number of transmissions required, as well. In particular, these results show the following:

- The behavior of the considered deterministic linear distributed averaging algorithms is as good as the behavior of the considered randomized ones in terms of the number of transmissions required on a cycle and on a path with  $n$  sensors:  $\Theta(n^3 \log \epsilon^{-1})$ .
- For a large enough number of sensors and regardless of the considered distributed averaging algorithm, the number of transmissions required on a path is four times larger than the number of transmissions required on a cycle.

Furthermore, regarding the cycle, from (10), (30), (41) and (54), we obtain the following enlightening asymptotic equalities:

$$\overset{\circ}{l}_B(\epsilon) \sim \frac{T(\epsilon, \overset{\circ}{W}_n(\gamma_0))}{2} \sim \frac{\overset{\circ}{l}_P(\epsilon)}{2} \sim \frac{T\left(\epsilon, \overset{\circ}{W}_n\left(\frac{1}{3}\right)\right)}{3}, \tag{61}$$

and regarding the path, from (21), (35), (46) and (59), we obtain:

$$\tilde{l}_B(\epsilon) \sim \frac{T\left(\epsilon, \tilde{W}_n\left(\frac{1}{2}\right)\right)}{2} \sim \frac{\tilde{l}_P(\epsilon)}{2} \sim \frac{T\left(\epsilon, \tilde{W}_n\left(\frac{1}{3}\right)\right)}{3}. \tag{62}$$

### 5. Numerical Examples

For the numerical examples, we first consider a cycle and a path with five and 10 sensors. For each network topology, we present a figure: Figure 2 for the cycle and Figure 3 for the path. Figure 2 (resp. Figure 3) shows the number of transmissions of the fastest LTI distributed averaging algorithm for symmetric weights  $T(\epsilon, \overset{\circ}{W}_n(\gamma_0))$  (resp.  $T(\epsilon, \tilde{W}_n(1/2))$ ) and of the Metropolis–Hastings algorithm  $T(\epsilon, \overset{\circ}{W}_n(1/3))$  (resp.  $T(\epsilon, \tilde{W}_n(1/3))$ ) with  $\epsilon \in (10^{-15}, 1)$ . The figure also shows the lower bound,  $\overset{\circ}{l}_P(\epsilon)$ , and upper bound,  $6\overset{\circ}{l}_P(\epsilon)$ , given for the number of transmissions of the pairwise gossip algorithm, and the lower bound,  $\overset{\circ}{l}_B(\epsilon)$ , and upper bound,  $6\overset{\circ}{l}_B(\epsilon)$ , given for the number of transmissions of the broadcast gossip algorithm (resp.  $\tilde{l}_P(\epsilon)$ ,  $6\tilde{l}_P(\epsilon)$ ,  $\tilde{l}_B(\epsilon)$  and  $6\tilde{l}_B(\epsilon)$ ). Furthermore, the figure shows the average number of transmissions of the pairwise gossip algorithm,  $\hat{T}(\epsilon, \{\overset{\circ}{W}_P(t)\}_{t \geq 0})$ , and of the broadcast gossip algorithm,  $\hat{T}(\epsilon, \{\overset{\circ}{W}_B(t)\}_{t \geq 0})$ , (resp.  $\hat{T}(\epsilon, \{\tilde{W}_P(t)\}_{t \geq 0})$  and  $\hat{T}(\epsilon, \{\tilde{W}_B(t)\}_{t \geq 0})$ ), that we have computed by using Monte Carlo simulations. In those simulations, we have performed 1000 repetitions of the corresponding algorithm for each  $\epsilon \in (10^{-15}, 1)$ , and we have considered that the values measured by the sensors,  $x_j(0)$  with  $j \in \{1, \dots, n\}$ , are independent identically distributed random variables with unit-variance, zero-mean and uniform distribution.

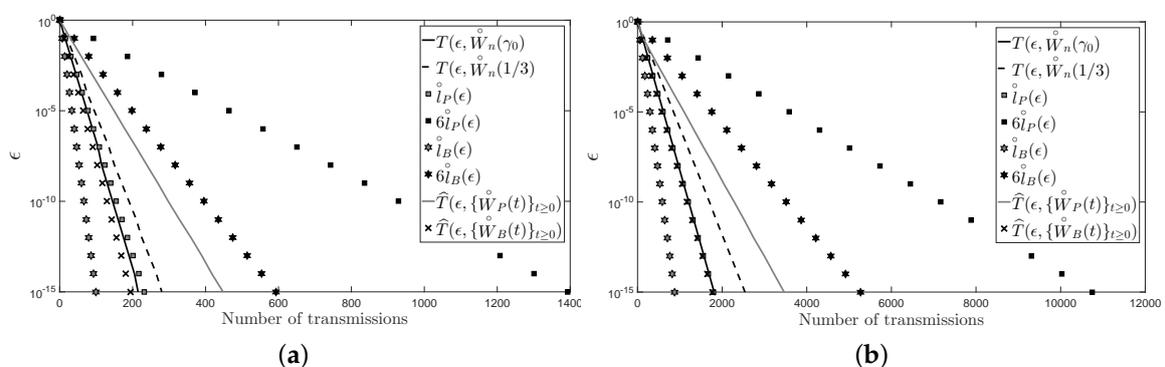


Figure 2. (a) A cycle with five sensors; (b) a cycle with 10 sensors.

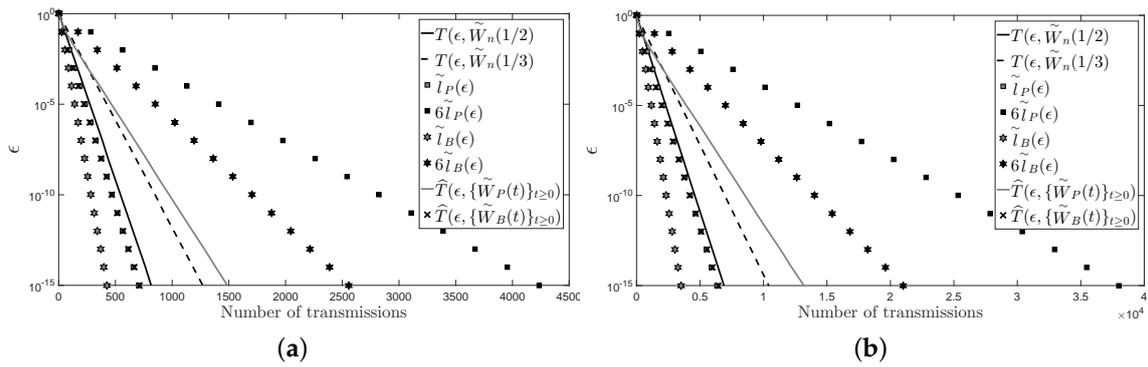


Figure 3. (a) A path with five sensors; (b) a path with 10 sensors.

In this section, we present another two figures: Figures 4 and 5. Unlike in Figures 2 and 3, in Figures 4 and 5, we have fixed  $\epsilon$  instead of the number of sensors  $n$  of the network. Specifically, we have chosen  $\epsilon = 10^{-3}$  and  $\epsilon = 10^{-6}$  with  $n \in \{5, \dots, 30\}$ .

In the figures, it can be observed that the Metropolis–Hastings algorithm behaves on average better than the pairwise gossip algorithm in terms of the number of transmissions required on the considered networks. It can also be observed that the broadcast gossip algorithm behaves on average approximately equal to the fastest LTI distributed averaging algorithm for symmetric weights in terms of the number of transmissions required on those networks. However, we recall here that the broadcast gossip algorithm converges to a random consensus value instead of to the average consensus value, and it should be executed several times in order to get that average value in every sensor.

The figures also bear evidence of the asymptotic equalities given in (61) and in (62).

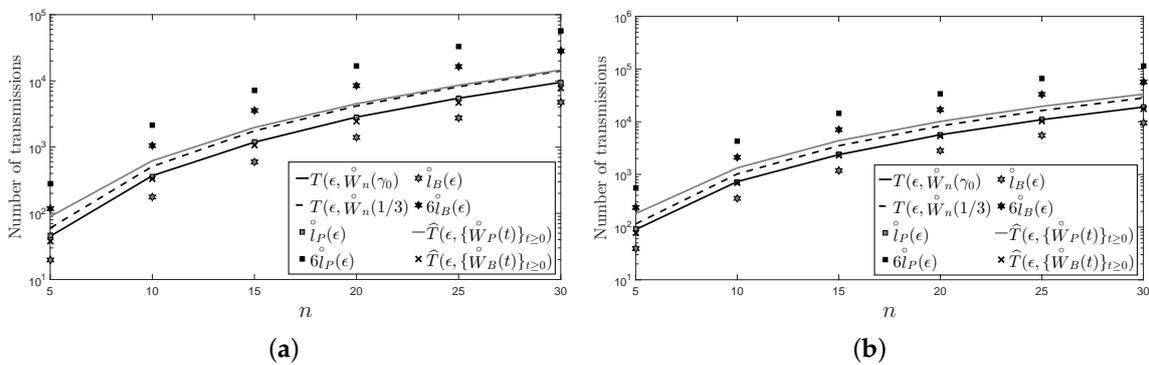


Figure 4. A cycle: (a)  $\epsilon = 10^{-3}$ ; (b)  $\epsilon = 10^{-6}$ .

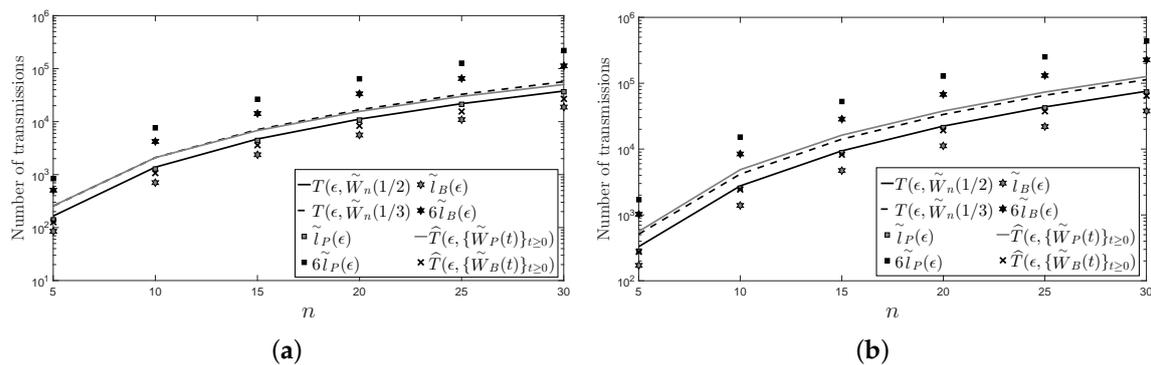


Figure 5. A path: (a)  $\epsilon = 10^{-3}$ ; (b)  $\epsilon = 10^{-6}$ .

## 6. Conclusions

In this paper, we have studied the convergence time of six known linear distributed averaging algorithms. We have considered both deterministic (the fastest LTI distributed averaging algorithm for symmetric weights, the fastest constant edge weights algorithm, the maximum-degree weights algorithm and the Metropolis–Hastings algorithm) and randomized (the pairwise gossip algorithm and the broadcast gossip algorithm) linear distributed averaging algorithms. In the literature, we have not found closed-form expressions for the convergence time of the considered algorithms. We have computed closed-form expressions for the convergence time of the deterministic algorithms and closed-form upper bounds for the convergence time of the randomized algorithms on two common network topologies: the cycle and the path. Moreover, we have also computed a closed-form expression for the convergence time of the fastest LTI algorithm on a grid. From the computed closed-form formulas, we have studied the asymptotic behavior of the convergence time of the considered algorithms as the number of sensors of the considered networks grows.

Although there exist different definitions of convergence time in the literature, in this paper, we have proven that one of them (namely, the one in (49)) encompasses all the others for the algorithms here considered. As we have used the definition of converge time in (49) for both deterministic and randomized linear distributed averaging algorithms, the obtained closed-form formulas and asymptotic results allow us to compare the considered algorithms on cycles and paths in terms of convergence time and, consequently, in terms of the number of transmissions required, as well.

We now summarize the most remarkable conclusions:

- The best algorithm among the considered deterministic distributed averaging algorithms is not worse than the best algorithm among the considered randomized distributed averaging algorithms for cycles and paths.
- The weighting matrix of the fastest LTI distributed averaging algorithm for symmetric weights and the weighting matrix of the fastest constant edge weights algorithm are the same on cycles and on paths.
- The number of transmissions required on a path with  $n$  sensors is asymptotically four-times larger than the number of transmissions required on a cycle with the same number of sensors.
- The number of transmissions required grows as  $n^3$  on cycles and on paths for the six algorithms considered.
- For the fastest LTI algorithm, the number of transmissions required grows as  $n^2$  on a square grid of  $n$  sensors (i.e.,  $r = c = \sqrt{n}$ ).

A future research direction of this work would be to generalize the analysis presented in the paper to other network topologies. In particular, networks that can be decomposed into cycles and paths could be studied.

**Acknowledgments:** This work was supported in part by the Spanish Ministry of Economy and Competitiveness through the RACHELproject (TEC2013-47141-C4-2-R), the CARMENproject (TEC2016-75067-C4-3-R) and the COMONSENSnetwork (TEC2015-69648-REDC).

**Author Contributions:** Jesús Gutiérrez-Gutiérrez conceived the research question. Jesús Gutiérrez-Gutiérrez, Marta Zárraga-Rodríguez and Xabier Insausti proved the main results. Xabier Insausti performed the simulations. Jesús Gutiérrez-Gutiérrez, Marta Zárraga-Rodríguez and Xabier Insausti wrote the paper. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Comparison of Several Definitions of Convergence Time

We begin by giving a property of the spectral norm. Its proof is implicit in the Appendix of [1].

**Lemma A1.** Let  $B$  and  $P$  be two  $n \times n$  real symmetric matrices with  $BP = P$  (or equivalently,  $PB = P$ ). Suppose that  $P$  is idempotent. Then:

1.  $B^t P = P$  for all  $t \in \mathbb{N}$ .
2.  $B^t - P = (B - P)^t$  for all  $t \in \mathbb{N}$ .
3.  $\|B^t - P\|_2 = \|B - P\|_2^t$  for all  $t \in \mathbb{N}$ .

We recall that an  $n \times n$  matrix  $A$  is idempotent if and only if  $A^2 = A$ . An example of idempotent matrix is  $P_n$  with  $n \in \mathbb{N}$ , since  $[P_n^2]_{j,k} = \sum_{h=1}^n [P_n]_{j,h} [P_n]_{h,k} = \sum_{h=1}^n \frac{1}{n} \frac{1}{n} = \frac{1}{n} = [P_n]_{j,k}$  for all  $j, k \in \{1, \dots, n\}$ .

The following result gives an eigenvalue decomposition for the matrix  $P_n$  for all  $n \in \mathbb{N}$ .

**Lemma A2.** If  $n \in \mathbb{N}$ , then  $P_n = V_n \text{diag}(1, 0, \dots, 0) V_n^*$ , where  $V_n$  is the  $n \times n$  Fourier unitary matrix.

**Proof.** From [13] (Lemma 2) or [14] (Lemma 3), we obtain that  $V_n \text{diag}(1, 0, \dots, 0) V_n^*$  is a circulant matrix with:

$$[V_n \text{diag}(1, 0, \dots, 0) V_n^*]_{j,1} = \left[ \frac{1}{\sqrt{n}} V_n \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right]_{j,1} = \frac{1}{\sqrt{n}} \sum_{k=1}^n [V_n]_{j,k} \left[ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right]_{k,1} = \frac{1}{\sqrt{n}} [V_n]_{j,1} = \frac{1}{n} \quad (\text{A1})$$

for all  $j \in \{1, \dots, n\}$ . Therefore,  $V_n \text{diag}(1, 0, \dots, 0) V_n^* = P_n$ .  $\square$

We finish this subsection with a result regarding the  $\epsilon$ -convergence time.

**Theorem A1.** Let  $B$  be an  $n \times n$  real symmetric matrix with  $B \neq P_n$  and  $B^t \rightarrow P_n$ . If  $\epsilon \in (0, 1)$ , then:

$$\min \left\{ t_0 \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \leq \epsilon, \forall t \geq t_0, \forall \mathbf{x} \neq P_n \mathbf{x} \right\} = \max_{\mathbf{x} \neq 0_{n \times 1}} \min \left\{ t \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \right\} \quad (\text{A2})$$

$$= \left\lceil \frac{\log \epsilon^{-1}}{-\log \|B - P_n\|_2} \right\rceil. \quad (\text{A3})$$

**Proof.** Let  $t \in \mathbb{N}$ . We first prove that the following statements are equivalent:

1.  $\frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \leq \epsilon$  for all  $\mathbf{x} \neq P_n \mathbf{x}$ .
2.  $\frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon$  for all  $\mathbf{x} \neq 0_{n \times 1}$ .

1⇒2 Fix  $\mathbf{x} \neq 0_{n \times 1}$ . If  $\mathbf{x} \neq P_n \mathbf{x}$ , applying Lemma A2 yields:

$$\frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\epsilon \|\mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\epsilon \|(I_n - P_n) \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \|I_n - P_n\|_2 = \epsilon \|V_n V_n^* - V_n \text{diag}(1, 0, \dots, 0) V_n^*\|_2 \tag{A4}$$

$$= \epsilon \|V_n I_n V_n^* - V_n \text{diag}(1, 0, \dots, 0) V_n^*\|_2 = \epsilon \|V_n \text{diag}(0, 1, \dots, 1) V_n^*\|_2 = \epsilon, \tag{A5}$$

where  $I_n$  is the  $n \times n$  identity matrix. If  $\mathbf{x} = P_n \mathbf{x}$ , from Lemma A1 and [2] (Theorem 1), we obtain:

$$\frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|B^t P_n \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|P_n \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = 0 < \epsilon. \tag{A6}$$

2⇒1 If  $\mathbf{x} \neq P_n \mathbf{x}$ , then:

$$\frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} = \frac{\|B^t \mathbf{x} - P_n \mathbf{x} - P_n \mathbf{x} + P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} = \frac{\|B^t \mathbf{x} - B^t P_n \mathbf{x} - P_n \mathbf{x} + P_n^2 \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \tag{A7}$$

$$= \frac{\|B^t (\mathbf{x} - P_n \mathbf{x}) - P_n (\mathbf{x} - P_n \mathbf{x})\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \leq \epsilon. \tag{A8}$$

Consequently,

$$\min \left\{ t_0 \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \leq \epsilon, \forall t \geq t_0, \forall \mathbf{x} \neq P_n \mathbf{x} \right\} \tag{A9}$$

$$= \min \left\{ t_0 \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon, \forall t \geq t_0, \forall \mathbf{x} \neq 0_{n \times 1} \right\} \tag{A10}$$

$$= \min \left\{ t_0 \in \mathbb{N} : \max_{\mathbf{x} \neq 0_{n \times 1}} \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon, \forall t \geq t_0 \right\} \tag{A11}$$

$$= \min \{ t_0 \in \mathbb{N} : \|B^t - P_n\|_2 \leq \epsilon, \forall t \geq t_0 \} \tag{A12}$$

$$= \min \{ t_0 \in \mathbb{N} : \|B - P_n\|_2^t \leq \epsilon, \forall t \geq t_0 \} \tag{A13}$$

$$= \min \left\{ t_0 \in \mathbb{N} : \|B - P_n\|_2^{t_0} \leq \epsilon \right\} \tag{A14}$$

$$= \min \left\{ t_0 \in \mathbb{N} : \log(\|B - P_n\|_2^{t_0}) \leq \log \epsilon \right\} \tag{A15}$$

$$= \min \{ t_0 \in \mathbb{N} : t_0 \log \|B - P_n\|_2 \leq \log \epsilon \} = \min \left\{ t_0 \in \mathbb{N} : t_0 \geq \frac{\log \epsilon}{\log \|B - P_n\|_2} \right\} \tag{A16}$$

$$= \min \left\{ t_0 \in \mathbb{N} : t_0 \geq \frac{\log \epsilon^{-1}}{-\log \|B - P_n\|_2} \right\} = \left\lceil \frac{\log \epsilon^{-1}}{-\log \|B - P_n\|_2} \right\rceil. \tag{A17}$$

To prove (A10), we have used the equivalence 1⇔2. To show (A12) and (A13), we have applied the definition of the spectral norm (see, e.g., [15] (pp. 603, 609)) and Assertion 3 of Lemma A1, respectively. To prove (A14), we have used [2] (Theorem 1) ( $\|B - P_n\|_2 < 1$ ).

As:

$$\min \left\{ t_0 \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x} - P_n \mathbf{x}\|_2} \leq \epsilon, \forall t \geq t_0, \forall \mathbf{x} \neq P_n \mathbf{x} \right\} = \min \{ t \in \mathbb{N} : \|B - P_n\|_2^t \leq \epsilon \} \tag{A18}$$

$$= \min \{ t \in \mathbb{N} : \|B^t - P_n\|_2 \leq \epsilon \}, \tag{A19}$$

we only need to show that  $T_1 = T_2$  to finish the proof, where:

$$T_1 = \min \left\{ t \in \mathbb{N} : \max_{\mathbf{x} \neq 0_{n \times 1}} \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \right\} \tag{A20}$$

and:

$$T_2 = \max_{\mathbf{x} \neq 0_{n \times 1}} t_{\mathbf{x}}, \tag{A21}$$

with:

$$t_x = \min \left\{ t \in \mathbb{N} : \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \right\}. \quad (\text{A22})$$

Since:

$$\frac{\|B^{T_1} \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \quad \forall \mathbf{x} \neq \mathbf{0}_{n \times 1} \quad (\text{A23})$$

we have  $t_x \leq T_1$  for all  $\mathbf{x} \neq \mathbf{0}_{n \times 1}$  and, consequently,  $T_2 \leq T_1$ . If  $\left\{ \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \right\}$  is a decreasing sequence for all  $\mathbf{x} \neq \mathbf{0}_{n \times 1}$ , then:

$$\frac{\|B^{T_2} \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\|B^{t_x} \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \quad \forall \mathbf{x} \neq \mathbf{0}_{n \times 1} \quad (\text{A24})$$

and therefore,

$$\max_{\mathbf{x} \neq \mathbf{0}_{n \times 1}} \frac{\|B^{T_2} \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \epsilon \quad (\text{A25})$$

and  $T_1 \leq T_2$ . Thus, if we prove that these sequences are decreasing, the proof is complete. Given  $\mathbf{x} \neq \mathbf{0}_{n \times 1}$ , from Lemma A1 and [2] (Theorem 1), we conclude that:

$$\frac{\|B^{t+1} \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|(B - P_n)^{t+1} \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\|(B - P_n)\|_2 \|(B - P_n)^t \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\|(B - P_n)^t \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|B^t \mathbf{x} - P_n \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \quad (\text{A26})$$

for all  $t \in \mathbb{N}$ . To prove the two equalities in (A26), we have used Assertion 2 of Lemma A1. To show the first inequality in (A26), we have applied a well-known inequality on the spectral norm (see, e.g., [15] (p. 611)), and to prove the second inequality in (A26), we have used [2] (Theorem 1) ( $\|B - P_n\|_2 < 1$ ).  $\square$

## Appendix B. Proof of Theorem 1

Let  $B(\gamma_1, \dots, \gamma_n)$  be the  $n \times n$  real symmetric matrix given by:

$$\begin{pmatrix} 1-\gamma_1-\gamma_n & \gamma_1 & 0 & \dots & 0 & 0 & \gamma_n \\ \gamma_1 & 1-\gamma_1-\gamma_2 & \gamma_2 & \dots & 0 & 0 & 0 \\ 0 & \gamma_2 & 1-\gamma_2-\gamma_3 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-\gamma_{n-3}-\gamma_{n-2} & \gamma_{n-2} & 0 \\ 0 & 0 & 0 & \dots & \gamma_{n-2} & 1-\gamma_{n-2}-\gamma_{n-1} & \gamma_{n-1} \\ \gamma_n & 0 & 0 & \dots & 0 & \gamma_{n-1} & 1-\gamma_{n-1}-\gamma_n \end{pmatrix}. \quad (\text{A27})$$

Observe that the matrix in (A27) satisfies  $B(\gamma_1, \dots, \gamma_n)P_n = P_n$ .

We define the function  $f: \mathbb{R}^n \mapsto [0, \infty)$  as  $f(\gamma_1, \dots, \gamma_n) := \|B(\gamma_1, \dots, \gamma_n) - P_n\|_2$ . We next prove that:

$$\left\| \overset{\circ}{W}_n(\gamma_0) - P_n \right\|_2 \leq \|B(\gamma_1, \dots, \gamma_n) - P_n\|_2 \quad \forall \gamma_1, \dots, \gamma_n \in \mathbb{R} \quad (\text{A28})$$

Observe that  $\overset{\circ}{W}_n(\gamma_0) = B(\gamma_0, \dots, \gamma_0)$ . As  $\overset{\circ}{W}_n(\gamma)$  is circulant, its eigenvalues are (see, e.g., [16] (Equation (3.7)) or [17] (Equation (5.2))):

$$a_j := 1 + 2\gamma \left( \cos \frac{2\pi(j-1)}{n} - 1 \right), \quad j \in \{1, \dots, n\} \quad (\text{A29})$$

Let  $V_n = (\mathbf{v}_1 | \cdots | \mathbf{v}_n)$  be the  $n \times n$  Fourier unitary matrix. It is well known (see, e.g., [16] (Equation (3.11)) or [17] (Lemma 5.1)) that  $\mathbf{v}_j$  is an unit eigenvector of  $\overset{\circ}{W}(\gamma)$  associated with the eigenvalue  $a_j$  for all  $j \in \{1, \dots, n\}$ . From Lemma A2:

$$\left\| \overset{\circ}{W}_n(\gamma) - P_n \right\|_2 = \max\{|a_j| : j \in \{2, \dots, n\}\}. \quad (\text{A30})$$

Case 1: Assume that  $n$  is even. Then,

$$\left\| \overset{\circ}{W}(\gamma_0) - P_n \right\|_2 = a_2 = -a_{\frac{n}{2}+1} = a_n = \frac{1 + \cos \frac{2\pi}{n}}{3 - \cos \frac{2\pi}{n}} \in (0, 1). \quad (\text{A31})$$

Therefore,  $\mathbf{y}_2 = \frac{\sqrt{2}}{2}(\mathbf{v}_n + \mathbf{v}_2)$  and  $\mathbf{y}_n = \frac{\sqrt{2}}{2\sqrt{-1}}(\mathbf{v}_n - \mathbf{v}_2)$  are unit eigenvectors of  $\overset{\circ}{W}(\gamma_0)$  associated to  $a_2 = a_n$ . As:

$$[\mathbf{y}_2]_{j,1} = \sqrt{\frac{2}{n}} \cos \frac{2\pi(j-1)}{n}, \quad (\text{A32})$$

$$[\mathbf{y}_n]_{j,1} = \sqrt{\frac{2}{n}} \sin \frac{2\pi(j-1)}{n}, \quad (\text{A33})$$

$$[\mathbf{v}_{\frac{n}{2}+1}]_{j,1} = \sqrt{\frac{1}{n}} (-1)^{j-1} \quad (\text{A34})$$

for all  $j \in \{1, \dots, n\}$ , from [4] (Theorem 1), we obtain three subgradients of  $f$  at  $(\gamma_0, \dots, \gamma_0) \in \mathbb{R}^n$ , namely  $\mathbf{g}_1$ ,  $\mathbf{g}_2$  and  $\mathbf{g}_3$ , given by:

$$[\mathbf{g}_1]_{j,1} = -\frac{2}{n} \left( \cos \frac{2\pi(j-1)}{n} - \cos \frac{2\pi j}{n} \right)^2, \quad (\text{A35})$$

$$[\mathbf{g}_3]_{j,1} = -\frac{2}{n} \left( \sin \frac{2\pi(j-1)}{n} - \sin \frac{2\pi j}{n} \right)^2, \quad (\text{A36})$$

$$[\mathbf{g}_2]_{j,1} = \frac{4}{n} \quad (\text{A37})$$

for all  $j \in \{1, \dots, n\}$ . If  $\mu = \frac{1}{3 - \cos \frac{2\pi}{n}}$ , we have that  $\mu \mathbf{g}_1 + \mu \mathbf{g}_2 + (1 - 2\mu) \mathbf{g}_3 = \mathbf{0}_{n \times 1}$ , where  $\mathbf{0}_{n \times 1}$  is the  $n \times 1$  zero matrix. The result now follows from [18] (p. 12) and the fact that a convex combination of subgradients of  $f$  at  $(\gamma_0, \dots, \gamma_0)$  is also a subgradient of  $f$  at  $(\gamma_0, \dots, \gamma_0)$ .

Case 2: Assume that  $n$  is odd. Then,

$$\left\| \overset{\circ}{W}(\gamma_0) - P_n \right\|_2 = a_2 = -a_{\frac{n+1}{2}} = -a_{\frac{n+3}{2}} = a_n = \frac{\cos \frac{2\pi}{n} + \cos \frac{\pi}{n}}{2 - \cos \frac{2\pi}{n} + \cos \frac{\pi}{n}} \in (0, 1). \quad (\text{A38})$$

Therefore,  $\mathbf{y}_2 = \frac{\sqrt{2}}{2}(\mathbf{v}_n + \mathbf{v}_2)$  and  $\mathbf{y}_n = \frac{\sqrt{2}}{2\sqrt{-1}}(\mathbf{v}_n - \mathbf{v}_2)$  are unit eigenvectors of  $\overset{\circ}{W}(\gamma_0)$  associated with  $a_2 = a_n$ , and  $\mathbf{y}_{\frac{n+1}{2}} = \frac{\sqrt{2}}{2}(\mathbf{v}_{\frac{n+1}{2}} + \mathbf{v}_{\frac{n+3}{2}})$  and  $\mathbf{y}_{\frac{n+3}{2}} = \frac{\sqrt{2}}{2\sqrt{-1}}(\mathbf{v}_{\frac{n+1}{2}} - \mathbf{v}_{\frac{n+3}{2}})$  are unit eigenvectors of  $\overset{\circ}{W}(\gamma_0)$  associated with  $a_{\frac{n+1}{2}} = a_{\frac{n+3}{2}}$ . As:

$$[\mathbf{y}_2]_{j,1} = \sqrt{\frac{2}{n}} \cos \frac{2\pi(j-1)}{n}, \tag{A39}$$

$$[\mathbf{y}_n]_{j,1} = \sqrt{\frac{2}{n}} \sin \frac{2\pi(j-1)}{n}, \tag{A40}$$

$$[\mathbf{y}_{\frac{n+1}{2}}]_{j,1} = \sqrt{\frac{2}{n}} (-1)^{j-1} \cos \frac{\pi(j-1)}{n}, \tag{A41}$$

$$[\mathbf{y}_{\frac{n+3}{2}}]_{j,1} = \sqrt{\frac{2}{n}} (-1)^{j-1} \sin \frac{\pi(j-1)}{n} \tag{A42}$$

for all  $j \in \{1, \dots, n\}$ , from [4] (Theorem 1), we obtain four subgradients of  $f$  at  $(\gamma_0, \dots, \gamma_0) \in \mathbb{R}^n$ , namely  $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$  and  $\mathbf{g}_4$  given by:

$$[\mathbf{g}_1]_{j,1} = -\frac{2}{n} \left( \cos \frac{2\pi(j-1)}{n} - \cos \frac{2\pi j}{n} \right)^2, \tag{A43}$$

$$[\mathbf{g}_2]_{j,1} = -\frac{2}{n} \left( \sin \frac{2\pi(j-1)}{n} - \sin \frac{2\pi j}{n} \right)^2, \tag{A44}$$

$$[\mathbf{g}_3]_{j,1} = \frac{2}{n} \left( \cos \frac{\pi(j-1)}{n} + \cos \frac{\pi j}{n} \right)^2, \tag{A45}$$

$$[\mathbf{g}_4]_{j,1} = \frac{2}{n} \left( \sin \frac{\pi(j-1)}{n} + \sin \frac{\pi j}{n} \right)^2 \tag{A46}$$

for all  $j \in \{1, \dots, n\}$ . If  $\mu = \frac{1 + \cos \frac{\pi}{n}}{2 - \cos \frac{2\pi}{n} + \cos \frac{\pi}{n}}$ , we have that  $\mu \mathbf{g}_1 + \mu \mathbf{g}_2 + (\frac{1}{2} - \mu) \mathbf{g}_3 + (\frac{1}{2} - \mu) \mathbf{g}_4 = \mathbf{0}_{n \times 1}$ . The result now follows from [18] (p. 12) and the fact that a convex combination of subgradients of  $f$  at  $(\gamma_0, \dots, \gamma_0)$  is also a subgradient of  $f$  at  $(\gamma_0, \dots, \gamma_0)$ .

Since  $\|\overset{\circ}{W}(\gamma_0) - P_n\|_2 < 1$ , applying [2] (Theorem 1) and Theorem A1, Theorem 1 holds.

### Appendix C. Proof of Theorem 2

From [2] (Theorem 1), Theorem A1, (A31) and (A38), we obtain (8).

To finish the proof, we only need to show (9) and (7).

We begin by proving (9). Applying Taylor’s theorem (see, e.g., [19] (p. 113)), there exist two bounded functions  $f, g : (0, \frac{\pi}{2}] \rightarrow \mathbb{R}$  such that:

$$-\log \frac{1 + \cos x}{3 - \cos x} = \frac{x^2}{2} + f(x)x^3 \tag{A47}$$

and:

$$-\log \frac{\cos \frac{x}{2} + \cos x}{2 + \cos \frac{x}{2} - \cos x} = \frac{x^2}{2} + g(x)x^3 \tag{A48}$$

for all  $x \in (0, \frac{\pi}{2}]$ . Therefore, from (A31) and (A38), we have:

$$\frac{\log \epsilon^{-1}}{-\log \|\overset{\circ}{W}(\gamma_0) - P_n\|_2} = \frac{\log \epsilon^{-1}}{\frac{(\frac{2\pi}{n})^2}{2} + y_n (\frac{2\pi}{n})^3}, \tag{A49}$$

where  $\{y_n\}_{n \geq 4}$  is the bounded sequence of real numbers given by:

$$\{y_n\}_{n \geq 4} = \begin{cases} f\left(\frac{2\pi}{n}\right) & \text{if } n \text{ is even,} \\ g\left(\frac{2\pi}{n}\right) & \text{if } n \text{ is odd.} \end{cases} \tag{A50}$$

Thus,

$$\lim_{n \rightarrow \infty} \frac{\log \epsilon^{-1}}{-n^2 \log \|\overset{\circ}{W}(\gamma_0) - P_n\|_2} = \lim_{n \rightarrow \infty} \frac{\log \epsilon^{-1}}{2\pi^2 + y_n \frac{8\pi^3}{n}} = \frac{\log \epsilon^{-1}}{2\pi^2}. \tag{A51}$$

Hence, as  $a \leq [a] < a + 1$ , with  $a \in \mathbb{R}$ , applying Theorem A1, we obtain:

$$\frac{\log \epsilon^{-1}}{2\pi^2} = \lim_{n \rightarrow \infty} \frac{\log \epsilon^{-1}}{-n^2 \log \|\overset{\circ}{W}(\gamma_0) - P_n\|_2} \leq \lim_{n \rightarrow \infty} \frac{\tau(\epsilon, \overset{\circ}{W}(\gamma_0))}{n^2} \tag{A52}$$

$$\leq \lim_{n \rightarrow \infty} \frac{\log \epsilon^{-1}}{-n^2 \log \|\overset{\circ}{W}(\gamma_0) - P_n\|_2} + \lim_{n \rightarrow \infty} \frac{1}{n^2} = \frac{\log \epsilon^{-1}}{2\pi^2}, \tag{A53}$$

and consequently, (9) holds.

Finally, we prove (7). If  $\delta \in \left(0, \frac{\log \epsilon^{-1}}{2\pi^2}\right)$ , then there exists  $n_0 \in \mathbb{N}$ , with  $n_0 > 3$ , such that:

$$\left| \frac{\tau(\epsilon, \overset{\circ}{W}(\gamma_0))}{n^2} - \frac{\log \epsilon^{-1}}{2\pi^2} \right| < \delta \quad \forall n \geq n_0. \tag{A54}$$

Thus, if  $n \geq n_0$ , then:

$$-\delta < \frac{\tau(\epsilon, \overset{\circ}{W}(\gamma_0))}{n^2} - \frac{\log \epsilon^{-1}}{2\pi^2} < \delta, \tag{A55}$$

or equivalently,

$$\left(\frac{1}{2\pi^2} - \frac{\delta}{\log \epsilon^{-1}}\right) n^2 \log \epsilon^{-1} < \tau(\epsilon, \overset{\circ}{W}(\gamma_0)) < \left(\frac{1}{2\pi^2} + \frac{\delta}{\log \epsilon^{-1}}\right) n^2 \log \epsilon^{-1}. \tag{A56}$$

### Appendix D. Proof of Theorem 3

We denote with  $\overset{\boxtimes}{\mathcal{W}}$  the set of all the  $rc \times rc$  real symmetric matrices such that:

$$\overset{\boxtimes}{\mathcal{W}}_{r,c} = \left\{ B \in \mathbb{R}^{rc \times rc}, \quad B = B^\top, \quad BP_n = P_n, \quad [B]_{j,k} = 0 \text{ if } j \neq k \text{ and } [A]_{j,k} = 0 \right\}, \tag{A57}$$

where  $\overset{\boxtimes}{A}$  is the adjacency matrix of a grid of  $r$  rows and  $c$  columns. Consider the bijection  $B : \mathbb{R}^q \mapsto \overset{\boxtimes}{\mathcal{W}}_{r,c}$  defined in [4] (Equation (8)), where  $q = 4rc - 3c - 3r + 2$  (i.e.,  $q$  is the number of edges when the network is viewed as an undirected graph).

We define the function  $f : \mathbb{R}^q \mapsto [0, \infty)$  as  $f(w_1, \dots, w_q) := \|B(w_1, \dots, w_q) - P_n\|_2$ . We next prove that:

$$\left\| \overset{\boxtimes}{\mathcal{W}}_{r,c} \left(\frac{1}{2}\right) - P_n \right\|_2 \leq \|B(w_1, \dots, w_q) - P_n\|_2 \quad \forall w_1, \dots, w_q \in \mathbb{R}. \tag{A58}$$

Without loss of generality, we can assume that  $r \geq c$ . We first show that  $\overset{\boxtimes}{W}_{r,c} \left(\frac{1}{2}\right) \in \overset{\boxtimes}{\mathcal{W}}_{r,c}$ :

$$\left(\overset{\sim}{W}_r \left(\frac{1}{2}\right) \otimes \overset{\sim}{W}_c \left(\frac{1}{2}\right)\right)^\top = \left(\overset{\sim}{W}_r \left(\frac{1}{2}\right)\right)^\top \otimes \left(\overset{\sim}{W}_c \left(\frac{1}{2}\right)\right)^\top = \overset{\sim}{W}_r \left(\frac{1}{2}\right) \otimes \overset{\sim}{W}_c \left(\frac{1}{2}\right), \tag{A59}$$

and:

$$\left(\tilde{W}_r \left(\frac{1}{2}\right) \otimes \tilde{W}_c \left(\frac{1}{2}\right)\right) P_n = \left(\tilde{W}_r \left(\frac{1}{2}\right) \otimes \tilde{W}_c \left(\frac{1}{2}\right)\right) (\sqrt{n}P_r \otimes \sqrt{n}P_c) \tag{A60}$$

$$= n \left( \left(\tilde{W}_r \left(\frac{1}{2}\right) P_r\right) \otimes \left(\tilde{W}_c \left(\frac{1}{2}\right) P_c\right) \right) \tag{A61}$$

$$= n (P_r \otimes P_c) = \sqrt{n}P_r \otimes \sqrt{n}P_c = P_n. \tag{A62}$$

The eigenvalues of  $\tilde{W}_n(\alpha)$  are (see, e.g., [20]):

$$a_j := 1 - 2\alpha + 2\alpha \left( \cos \frac{\pi(j-1)}{n} \right), \quad j \in \{1, \dots, n\} \tag{A63}$$

and therefore, the eigenvalues of  $\tilde{W}_n \left(\frac{1}{2}\right)$  are given by  $a_j(n) := \cos \frac{(j-1)\pi}{n}$  with  $j \in \{1, \dots, n\}$ .

Their associated orthonormal eigenvectors are given by  $[\mathbf{v}_1(n)]_{k,1} = \frac{1}{\sqrt{n}}$  and  $[\mathbf{v}_j(n)]_{k,1} = \sqrt{\frac{2}{n}} \cos \frac{(2k-1)(j-1)\pi}{2n}$

with  $j \in \{2, \dots, n\}, k \in \{1, \dots, n\}$  (see, e.g., [20]). Consequently, the eigenvalues of  $\tilde{W}_r \left(\frac{1}{2}\right) \otimes \tilde{W}_c \left(\frac{1}{2}\right)$  are  $a_j(r)a_k(c)$  and associated orthonormal eigenvectors are  $\mathbf{v}_j(r) \otimes \mathbf{v}_k(c)$  with  $j \in \{1, \dots, r\}$  and  $k \in \{1, \dots, c\}$ .

From [4] (Lemma 1),

$$\left\| \overset{\boxtimes}{W}_{r,c} \left(\frac{1}{2}\right) - P_n \right\|_2 = a_2(r)a_1(c) = -a_r(r)a_1(c) = \cos \frac{\pi}{r} \in (0, 1). \tag{A64}$$

Then,  $\mathbf{y}_1 = \mathbf{v}_2(r) \otimes \mathbf{v}_1(c)$  and  $\mathbf{y}_2 = \mathbf{v}_r(r) \otimes \mathbf{v}_1(c)$  are unit eigenvectors of  $\overset{\boxtimes}{W}_{r,c} \left(\frac{1}{2}\right)$  associated with  $a_2(r)a_1(c)$  and  $a_r(r)a_1(c)$ , respectively, and their entries are given by:

$$[\mathbf{y}_1]_{c(j-1)+k,1} = \sqrt{\frac{2}{rc}} \cos \frac{(2j-1)\pi}{2r}, \tag{A65}$$

$$[\mathbf{y}_2]_{c(j-1)+k,1} = (-1)^{j-1} \sqrt{\frac{2}{rc}} \sin \frac{(2j-1)\pi}{2r}, \tag{A66}$$

for all  $j \in \{1, \dots, r\}$  and  $k \in \{1, \dots, c\}$ .

Let  $\varepsilon$  be the set of edges of the grid. An edge  $e = \{j, k\}$  connects the sensors  $v_j, v_k$ , and we enumerate the edges such that  $e_l = \{j_l, k_l\}$  for all  $l \in \{1, \dots, q\}$ . We consider that the edges of the grid are sorted as follows:  $\varepsilon_H, \varepsilon_V, \varepsilon_{NW}$  and  $\varepsilon_{NE}$  are the set of horizontal, vertical, northwest-southeast diagonal and northeast-southwest diagonal edges, respectively. Moreover, if  $e_{l_1} = \{j_{l_1}, k_{l_1}\}, e_{l_2} = \{j_{l_2}, k_{l_2}\} \in \varepsilon_h$  with  $h \in \{H, V, NW, NE\}$  and  $\min\{j_{l_1}, k_{l_1}\} < \min\{j_{l_2}, k_{l_2}\}$ , then the edge  $e_{l_1}$  precedes the edge  $e_{l_2}$  in  $\varepsilon_h$ .

From [4] (Theorem 1), we obtain two subgradients of  $f$ ,  $\mathbf{g}_1$  and  $\mathbf{g}_2$  given by:

$$\mathbf{g}_1 = \left( \mathbf{g}_1^{(H)} | \mathbf{g}_1^{(V)} | \mathbf{g}_1^{(NW)} | \mathbf{g}_1^{(NE)} \right)^\top, \tag{A67}$$

$$\mathbf{g}_2 = \left( \mathbf{g}_2^{(H)} | \mathbf{g}_2^{(V)} | \mathbf{g}_2^{(NW)} | \mathbf{g}_2^{(NE)} \right)^\top, \tag{A68}$$

where  $\mathbf{g}_1^{(H)} = \mathbf{g}_2^{(H)} = \mathbf{0}_{1 \times r(c-1)}$ ,

$$\left[ \mathbf{g}_1^{(V)} \right]_{1,c(j-1)+k} = -\frac{8}{rc} \sin^2 \left( \frac{\pi}{2r} \right) \sin^2 \left( \frac{j\pi}{r} \right), \tag{A69}$$

$$\left[ \mathbf{g}_2^{(V)} \right]_{1,c(j-1)+k} = \frac{8}{rc} \cos^2 \left( \frac{\pi}{2r} \right) \sin^2 \left( \frac{j\pi}{r} \right), \tag{A70}$$

for all  $j \in \{1, \dots, r-1\}, k \in \{1, \dots, c\}$ ,

$$\left[ \mathbf{g}_1^{(NW)} \right]_{1, (c-1)(j-1)+k} = -\frac{8}{rc} \sin^2\left(\frac{\pi}{2r}\right) \sin^2\left(\frac{j\pi}{r}\right), \tag{A71}$$

$$\left[ \mathbf{g}_2^{(NW)} \right]_{1, (c-1)(j-1)+k} = \frac{8}{rc} \cos^2\left(\frac{\pi}{2r}\right) \sin^2\left(\frac{j\pi}{r}\right), \tag{A72}$$

for all  $j \in \{1, \dots, r-1\}, k \in \{1, \dots, c-1\}$ , and:

$$\left[ \mathbf{g}_1^{(NE)} \right]_{1, (c-1)(j-1)+k-1} = -\frac{8}{rc} \sin^2\left(\frac{\pi}{2r}\right) \sin^2\left(\frac{j\pi}{r}\right), \tag{A73}$$

$$\left[ \mathbf{g}_2^{(NE)} \right]_{1, (c-1)(j-1)+k-1} = \frac{8}{rc} \cos^2\left(\frac{\pi}{2r}\right) \sin^2\left(\frac{j\pi}{r}\right), \tag{A74}$$

for all  $j \in \{1, \dots, r-1\}, k \in \{2, \dots, c\}$ . If  $\mu = \cos^2\left(\frac{\pi}{2r}\right)$ , we have that  $\mu \mathbf{g}_1 + (1-\mu) \mathbf{g}_2 = \mathbf{0}_{(4rc-3c-3r+2) \times 1}$ . The result now follows from [18] (p. 12) and the fact that a convex combination of subgradients of  $f$  at a certain point is also a subgradient of  $f$  at that point.

Since  $\|\overset{\boxtimes}{W}_{r,c}\left(\frac{1}{2}\right) - P_n\|_2 < 1$ , applying [2] (Theorem 1) and Theorem A1, Theorem 3 holds.

### Appendix E. Proof of Theorem 9

We begin by proving (51). The Laplacian matrix of a cycle with  $n$  sensors is:

$$\overset{\circ}{L} = \text{diag}(\mathbf{1}_n^\top \text{circ}_n(0, 1, 0, \dots, 0, 1)) - \text{circ}_n(0, 1, 0, \dots, 0, 1) \tag{A75}$$

$$= \text{diag}(2, 2, \dots, 2) - \text{circ}_n(0, 1, 0, \dots, 0, 1) = \text{circ}_n(2, -1, 0, \dots, 0, -1). \tag{A76}$$

From [21] (Equation (3.4a)), the eigenvalues of  $\overset{\circ}{L}$  are given by  $\left\{ 2\left(1 - \cos\frac{2\pi(j-1)}{n}\right) : j \in \{1, \dots, n\} \right\}$  and, consequently,  $\lambda_{n-1}(\overset{\circ}{L}) = 2\left(1 - \cos\frac{2\pi}{n}\right)$ . From [7] (Corollary 1), we have:

$$\overset{\circ}{\varphi}_0 = \frac{n - \lambda_{n-1}(\overset{\circ}{L})}{2n - \lambda_{n-1}(\overset{\circ}{L})} = 1 - \frac{n}{2\left(n + \cos\frac{2\pi}{n} - 1\right)}, \tag{A77}$$

and therefore, (51) holds. The entries of the expectation of  $\overset{\circ}{W}_B(0)$  are given by:

$$[E(\overset{\circ}{W}_B(0))]_{j,k} = \begin{cases} \frac{1}{n}(1 - \overset{\circ}{\varphi}_0) & \text{if } j - k \in \{-1, 1\}, \\ \frac{1}{n}(1 - \overset{\circ}{\varphi}_0) & \text{if } j - k \in \{1 - n, n - 1\}, \\ \frac{1}{n}(2\overset{\circ}{\varphi}_0 + n - 2) & \text{if } j = k, \\ 0 & \text{otherwise,} \end{cases} \tag{A78}$$

for all  $j, k \in \{1, \dots, n\}$ . Thus,  $E(\overset{\circ}{W}_B(0)) = \overset{\circ}{W}_n\left(\frac{1 - \overset{\circ}{\varphi}_0}{n}\right)$ . Therefore, combining (A29) and (A30) yields:

$$\left\| \overset{\circ}{W}_n\left(\frac{1 - \overset{\circ}{\varphi}_0}{n}\right) - P_n \right\|_2 = \frac{n + 2\cos\frac{2\pi}{n} - 2}{n + \cos\frac{2\pi}{n} - 1}. \tag{A79}$$

As:

$$\overset{\circ}{\varphi}_0 = \frac{n - \lambda_{n-1}(\overset{\circ}{L})}{2n - \lambda_{n-1}(\overset{\circ}{L})} = 1 - \frac{n}{2n - \lambda_{n-1}(\overset{\circ}{L})} \tag{A80}$$

we get:

$$\lambda_{n-1}(\overset{\circ}{L}) = n \left( 2 - \frac{1}{1 - \overset{\circ}{\varphi}_0} \right) = n \frac{1 - 2\overset{\circ}{\varphi}_0}{1 - \overset{\circ}{\varphi}_0} \quad (\text{A81})$$

and consequently,

$$1 - \frac{2\overset{\circ}{\varphi}_0(1 - \overset{\circ}{\varphi}_0)}{n} \lambda_{n-1}(\overset{\circ}{L}) - \frac{(1 - \overset{\circ}{\varphi}_0)^2}{n^2} \left( \lambda_{n-1}(\overset{\circ}{L}) \right)^2 = 1 - 2\overset{\circ}{\varphi}_0(1 - 2\overset{\circ}{\varphi}_0) - (1 - 2\overset{\circ}{\varphi}_0)^2 \quad (\text{A82})$$

$$= 1 - 2\overset{\circ}{\varphi}_0 + 4\overset{\circ}{\varphi}_0^2 - 1 + 4\overset{\circ}{\varphi}_0 - 4\overset{\circ}{\varphi}_0^2 = 2\overset{\circ}{\varphi}_0 = 2 - \frac{n}{n + \cos \frac{2\pi}{n} - 1} = \frac{n + 2 \cos \frac{2\pi}{n} - 2}{n + \cos \frac{2\pi}{n} - 1}. \quad (\text{A83})$$

Now, applying (A79), (A82) and [7] (Equations (28) and (46)), we obtain (52). The rest of the proof runs as the proof of Theorem 2.

## Appendix F. Proof of Theorem 10

We begin by proving (56). The Laplacian matrix of a path with  $n$  sensors is:

$$[\tilde{L}]_{j,k} = \begin{cases} -1 & \text{if } j - k \in \{-1, 1\}, \\ 2 & \text{if } j = k, j \neq 1 \text{ and } j \neq n, \\ 1 & \text{if } j = k, j \in \{1, n\}, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A84})$$

From [20], the eigenvalues of  $\tilde{L}$  are given by  $\left\{ 2 \left( 1 - \cos \frac{\pi(j-1)}{n} \right) : j \in \{1, \dots, n\} \right\}$  and, consequently,  $\lambda_{n-1}(\tilde{L}) = 2 \left( 1 - \cos \frac{\pi}{n} \right)$ . From [7] (Corollary 1), we have:

$$\tilde{\varphi}_0 = \frac{n - \lambda_{n-1}(\tilde{L})}{2n - \lambda_{n-1}(\tilde{L})} = 1 - \frac{n}{2(n + \cos \frac{\pi}{n} - 1)}, \quad (\text{A85})$$

and therefore, (56) holds. The entries of the expectation of  $\tilde{W}_B(0)$  are given by:

$$[E(\tilde{W}_B(0))]_{j,k} = \begin{cases} \frac{1}{n}(1 - \tilde{\varphi}_0) & \text{if } j - k \in \{-1, 1\}, \\ \frac{1}{n}(2\tilde{\varphi}_0 + n - 2) & \text{if } j = k, j \neq 1 \text{ and } j \neq n, \\ \frac{1}{n}(\tilde{\varphi}_0 + n - 1) & \text{if } j = k, j \in \{1, n\}, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A86})$$

for all  $j, k \in \{1, \dots, n\}$ . Thus,  $E(\tilde{W}_B(0)) = \tilde{W}_n \left( \frac{1 - \tilde{\varphi}_0}{n} \right)$ . Therefore, combining (A63) and [4] (Lemma 1) yields:

$$\left\| \tilde{W}_n \left( \frac{1 - \tilde{\varphi}_0}{n} \right) - P_n \right\|_2 = \frac{n + 2 \cos \frac{\pi}{n} - 2}{n + \cos \frac{\pi}{n} - 1}. \quad (\text{A87})$$

As:

$$\tilde{\varphi}_0 = \frac{n - \lambda_{n-1}(\tilde{L})}{2n - \lambda_{n-1}(\tilde{L})} = 1 - \frac{n}{2n - \lambda_{n-1}(\tilde{L})} \quad (\text{A88})$$

we get:

$$\lambda_{n-1}(\tilde{L}) = n \left( 2 - \frac{1}{1 - \tilde{\varphi}_0} \right) = n \frac{1 - 2\tilde{\varphi}_0}{1 - \tilde{\varphi}_0} \quad (\text{A89})$$

and consequently,

$$1 - \frac{2\tilde{\varphi}_0(1-\tilde{\varphi}_0)}{n} \lambda_{n-1}(\tilde{L}) - \frac{(1-\tilde{\varphi}_0)^2}{n^2} \left( \lambda_{n-1}(\tilde{L}) \right)^2 = 1 - 2\tilde{\varphi}_0(1-2\tilde{\varphi}_0) - (1-2\tilde{\varphi}_0)^2 \quad (\text{A90})$$

$$= 1 - 2\tilde{\varphi}_0 + 4\tilde{\varphi}_0^2 - 1 + 4\tilde{\varphi}_0 - 4\tilde{\varphi}_0^2 = 2\tilde{\varphi}_0 = 2 - \frac{n}{n + \cos \frac{2\pi}{n} - 1} = \frac{n + 2 \cos \frac{\pi}{n} - 2}{n + \cos \frac{\pi}{n} - 1}. \quad (\text{A91})$$

Now, applying (A87), (A90) and [7] (Equations (28) and (46)), we obtain (57). The rest of the proof runs as the proof of Theorem 2.

## References

- Insausti, X.; Camaró, F.; Crespo, P.M.; Beferull-Lozano, B.; Gutiérrez-Gutiérrez, J. Distributed pseudo-gossip algorithm and finite-length computational codes for efficient in-network subspace projection. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 163–174.
- Xiao, L.; Boyd, S. Fast linear iterations for distributed averaging. *Syst. Control Lett.* **2004**, *53*, 65–78.
- Xiao, L.; Boyd, S.; Kimb, S.J. Distributed average consensus with least-mean-square deviation. *J. Parallel Distrib. Comput.* **2007**, *67*, 33–46.
- Insausti, X.; Gutiérrez-Gutiérrez, J.; Zárraga-Rodríguez, M.; Crespo, P.M. In-network computation of the optimal weighting matrix for distributed consensus on wireless sensor networks. *Sensors* **2017**, *17*, 1702.
- Olshevsky, A.; Tsitsiklis, J. Convergence speed in distributed consensus and averaging. *SIAM Rev.* **2011**, *53*, 747–772.
- Boyd, S.; Ghosh, A.; Prabhakar, B.; Shah, D. Randomized gossip algorithms. *IEEE Trans. Inf. Theory* **2006**, *52*, 2508–2530.
- Aysal, T.C.; Yildiz, M.E.; Sarwate, A.D.; Scaglione, A. Broadcast gossip algorithms for consensus. *IEEE Trans. Signal Process.* **2009**, *57*, 2748–2761.
- Wu, S.; Rabbat, M.G. Broadcast Gossip Algorithms for Consensus on Strongly Connected Digraphs. *IEEE Trans. Signal Process.* **2013**, *61*, 3959–3971.
- Dimakis, A.D.G.; Kar, S.; Moura, J.M.F.; Rabbat, M.G.; Scaglione, A. Gossip algorithms for distributed signal processing. *Proceed. IEEE* **2010**, *98*, 1847–1864.
- Olshevsky, A.; Tsitsiklis, J. Convergence speed in distributed consensus and averaging. *SIAM J. Control Optim.* **2009**, *48*, 33–55.
- Olshevsky, A.; Tsitsiklis, J. A lower bound for distributed averaging algorithms on the line graph. *IEEE Trans. Autom. Control* **2011**, *56*, 2694–2698.
- Apostol, T.M. *Calculus*; John Wiley & Sons: Hoboken, NJ, USA, 1967; Volume 1.
- Gutiérrez-Gutiérrez, J.; Crespo, P.M. Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems. *IEEE Trans. Inf. Theory* **2008**, *54*, 5671–5680.
- Gutiérrez-Gutiérrez, J.; Crespo, P.M. Asymptotically equivalent sequences of matrices and multivariate ARMA processes. *IEEE Trans. Inf. Theory* **2011**, *57*, 5444–5454.
- Bernstein, D.S. *Matrix Mathematics*; Princeton University Press: Princeton, NJ, USA, 2009.
- Gray, R.M. Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory* **2006**, *2*, 155–239.
- Gutiérrez-Gutiérrez, J.; Crespo, P.M. Block Toeplitz matrices: Asymptotic results and applications. *Found. Trends Commun. Inf. Theory* **2011**, *8*, 179–257.
- Shor, N.Z. *Minimization Methods for Non-Differentiable Functions*; Springer: Berlin, Germany, 1985.
- Apostol, T.M. *Mathematical Analysis*; Addison-Wesley: Boston, MA, USA, 1974.
- Yueh, W.C.; Cheng, S.S. Explicit Eigenvalues and inverses of tridiagonal Toeplitz matrices with four perturbed corners. *ANZIAM J.* **2008**, *49*, 361–387.
- Gray, R.M. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Trans. Inf. Theory* **1972**, *18*, 725–730.

