

Distributed estimation of diameter, radius and eccentricities in anonymous networks [★]

Federica Garin ^{*} Damiano Varagnolo ^{**} Karl H. Johansson ^{**}

^{*} *Inria Grenoble Rhône-Alpes, NECS team, Grenoble, France*

Email: federica.garin@inria.fr

^{**} *School of Electrical Engineering, Royal Institute of Technology,*

Osqudas väg 10, Stockholm, Sweden

Emails: { damiano | kallej }@kth.se

Abstract: We consider how a set of collaborating agents can distributedly infer some of the properties of the communication network that they form. We specifically focus on estimating quantities that can characterize the performance of other distributed algorithms, namely the eccentricities of the nodes, and the radius and diameter of the network. We propose a strategy that can be implemented in any network, even under anonymity constraints, and has the desirable properties of being fully distributed, parallel and scalable. We analytically characterize the statistics of the estimation error, and highlight how the performance of the algorithm depends on a parameter tuning the communication complexity.

Keywords: graph topology, network structure, graph structure, anonymous networks, consensus, distributed estimation, sensor networks

1. INTRODUCTION

The knowledge of the topology of a network plays a crucial role in achieving lasting and scalable Wireless Sensor Networks (WSNs) (Li and Yang, 2006): it is useful, e.g., to detect the presence of coverage or routing holes in WSNs (Ahmed et al., 2005), to improve the operativity of the network by maintaining a certain efficiency in communicating using less energy (Chen et al., 2002), and to implement better termination rules in distributed computations.

Distributed algorithms for topology reconstruction can either exploit the presence of IDs uniquely defining the various agents, e.g., through constructing and exchanging tables of IDs (Deb et al., 2004), or not, e.g., through suitable random-walk strategies (Hall, 2010).

Here we focus on the specific problem of estimating graphs' diameters, radii and eccentricities, paramount parameters that bound the speed of propagation of the information through the network.

The estimation of distances in a distributed system has a long tradition, and lists several different algorithms (see, e.g., the survey offered by Zwick (2001) and the references therein). Centralized approaches often use Breadth First Search (BFS) tree construction procedures that randomly choose a node, perform two BFSs and then return the maximal length among the shortest paths computed in this way (Lynch, 1996; Corneil et al., 2001; Crescenzi

et al., 2010). Due to the intrinsic computational complexity of the problem, generally the centralized approaches accept approximate solutions for faster computations (Elkin, 2001; Aingworth et al., 1996; Boitmanis et al., 2006).

Eccentricity and diameter estimation algorithms have been proposed also in distributed frameworks, see, e.g., Almeida et al. (2012). A natural approach is to let the agents propagate messages containing IDs and suitable hops counters: then from the values of these counters agents can infer the structure of the network.

Here we focus on the framework of distributed *anonymous* networks, where agents cannot or do not want to disclose their IDs, e.g., for privacy concerns. More specifically, we extend a well-known strategy for the estimation of the network size, based on mixing local random generation steps plus max-consensus procedures (Varagnolo et al., 2010; Cichon et al., 2011; Baquero et al., 2012).

From such works, our algorithm inherits several positive features: easy implementability, little communication / computation / memory resources requirements and little a-priori assumptions on the structure of the network. Moreover, the strategy is fully parallel, distributed, requiring no leader election steps and scalable, while mechanisms based on exchanging the IDs instead require an amount of memory and communication efforts that grow with network size. Another advantage w.r.t. the approaches proposed in the literature recalled above is that from our strategy it is possible to derive a modified algorithm capable to deal with dynamic frameworks, where agents may join / leave the system.

To the best of our knowledge, the only paper proposing a strategy similar to ours is the one by Cardoso et al. (2009). Despite sharing a similar algorithmic structure, our work is

[★] The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n. 257462 HYCON2 Network of excellence, from the Swedish Research Council and from the Knut and Alice Wallenberg Foundation.

essentially different: the former focuses on communication complexity from computer science perspectives, disregarding the statistical ones. This paper, instead, complements and completes the other one by providing both novel estimators and their full statistical descriptions. More importantly, we fully characterize the tradeoffs between the length of messages and the estimation performance.

This work is structured as follows: Sect. 2 collects the notation, while Sect. 3 gives an overview of the size estimation algorithm proposed in Varagnolo et al. (2010); Cichon et al. (2011). In Sect. 4 we introduce our estimators, while in Sect. 5 we analyze performance, both analytically and numerically. We end the work with Sect. 6 by collecting some conclusions and possible future works.

2. NOTATION AND DEFINITIONS

In the following, bold italic fonts indicate vectors and plain italic fonts indicate scalars. Moreover $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ indicates a graph, intended as a set \mathcal{V} of agents and a set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ of links among agents, i.e., $(u, v) \in \mathcal{E}$ if u can successfully communicate with v . In this paper we will consider only undirected graphs, although some results can be extended to directed ones. Moreover, we will always assume that the graph \mathcal{G} is connected. We will denote by $\text{dist}(u, v)$ the distance from u to v , defined as the length (number of edges) of the shortest path from u to v . The eccentricity of an agent u is indicated with $e(u)$ and is defined as the longest shortest path starting from u , i.e., $e(u) := \max_{v \in \mathcal{V}} \text{dist}(u, v)$. The diameter of \mathcal{G} is indicated with d and is defined as the maximal eccentricity, i.e., $d := \max_{u \in \mathcal{V}} e(u) = \max_{u, v \in \mathcal{V}} \text{dist}(u, v)$, while the radius of \mathcal{G} is the minimum eccentricity, $r := \min_{u \in \mathcal{V}} e(u)$. The k -hops neighbors of u are the agents v such that $\text{dist}(u, v) = k$. The set of k -hops neighbors is called the k -hops neighborhood of u and is indicated with $\mathcal{D}_k^{(u)}$, while its cardinality is indicated with $D_k^{(u)}$. The total number of nodes in the network is indicated with V . We will also use the notation \mathcal{D}_k for the set of all nodes u such that $e(u) = k$, and D_k for its cardinality. Finally, we use widehats to indicate estimates, for instance, \widehat{V} indicates an estimate of V .

3. ESTIMATION UNDER SYNCHRONOUS COMMUNICATIONS ASSUMPTIONS

We will endow a classical algorithm for the estimation of the size of a network with the capability of estimating the local eccentricity, the diameter and the radius. We assume the following synchronous communications protocol: the time is divided in epochs, indexed by $t = 0, 1, 2, \dots$. Every agent broadcasts its information exactly once per epoch. The order of the broadcasting operations is irrelevant, and can change in time. When an agent broadcasts its information, it broadcasts the information that it had at the beginning of the epoch. Thus, the time index t does *not* denote a physical quantity (e.g., seconds), but rather the index of the various epochs.

Choice of the convergence time: the assumed communications protocol ensures that any max-consensus algorithm

will converge exactly after d communication steps. The problem is thus how to instruct max-consensus algorithms with a proper termination rule, d being both the unknown quantity to be estimated *and* the optimal termination rule.

A classical approach is to choose a terminating time t^* sufficiently large to ensure $t^* \geq d$ for all the cases under consideration. This choice is obviously critical and represents a major issue. Practical selection rules are:

- if V_{\max} is a (known) upper bound on the number of agents in the network (i.e., $V \leq V_{\max}$), then set $t^* = V_{\max}$, which ensures $t^* \geq d$. A similar strategy could be implemented knowing an upper bound d_{\max} on the diameter d ;
- if \widehat{V} is an estimate of the number of agents in the network V , then set $t^* = \alpha \widehat{V}$ where $\alpha \geq 1$ accounts for the uncertainties on \widehat{V} . A similar strategy can be implemented knowing an estimate \widehat{d} of the diameter d .

In the following, we will assume that t^* has already been chosen and that the chances for inexact convergence are sufficiently small to be negligible.

The original size estimation algorithm: size estimation can be performed by the means of Alg. 1. In it, each agent u is endowed with a vector $\mathbf{x}^{(u)}(t) \in \mathbb{R}^M$, $\mathbf{x}^{(u)}(t) = [x_1^{(u)}(t), \dots, x_M^{(u)}(t)]$, where M is a fixed positive integer and t is the time index.

Algorithm 1 Size estimation in anonymous networks

- 1: (storage allocation) each agent u stores a vector $\mathbf{x}^{(u)} \in \mathbb{R}^M$, $\mathbf{x}^{(u)} = [x_1^{(u)}, \dots, x_M^{(u)}]$;
- 2: (initialization) $x_m^{(u)}(0) \sim \mathcal{U}[0, 1]$, $m = 1, \dots, M$, i.i.d., for all $u \in \mathcal{V}$;
- 3: **for** $t = 1, \dots, t^*$ **do**
- 4: (communication) every agent broadcasts $\mathbf{x}^{(u)}(t-1)$ to its neighbors;
- 5: (information mixing) every agent computes $\mathbf{x}^{(u)}(t)$ by means of $\mathbf{x}^{(u)}(t-1)$ and the $\mathbf{x}^{(v)}(t-1)$'s received from its 1-hop neighbors through

$$x_m^{(u)}(t) = \max \left(x_m^{(u)}(t-1), \max_{v \in \mathcal{D}_1^{(u)}} x_m^{(v)}(t-1) \right) \quad (1)$$

for $m = 1, \dots, M$.

- 6: **end for**
-

- 7: (estimation) set $x_m = x_m^{(u)}(t^*)$. Then

$$\widehat{V} = \left(-\frac{1}{M} \sum_{m=1}^M \log(x_m) \right)^{-1}. \quad (2)$$

The estimator \widehat{V} in (2) is the Maximum Likelihood (ML) estimator for the size of the network V given $\mathbf{x}^{(u)}(t^*)$ (see, e.g., Varagnolo et al. (2010) for its statistical description).

Remark 1. Alg. 1 gives an estimate of the size V of the set \mathcal{V} of all nodes, based on the maxima (over all nodes $u \in \mathcal{V}$) of the M entries of the initial vectors $\mathbf{x}^{(u)}(0)$. Beside providing $x_m^{(u)}(t^*) = \max_{v \in \mathcal{V}} \{x_m^{(v)}(0)\}$ for all u ,

the max-consensus algorithm gives some useful information also along the iterations $t < t^*$. Indeed, $x_m^{(u)}(t) = \max \left\{ x^{(v)}(0) : v \in \bigcup_{k=0}^t \mathcal{D}_k^{(u)} \right\}$. Hence, at iteration t , node u can compute the ML estimate \widehat{V}_t of $V_t^{(u)} := \sum_{k=0}^t D_k^{(u)}$ (the number of nodes within distance t from itself) as

$$\widehat{V}_t^{(u)} := \left(-\frac{1}{M} \sum_{m=1}^M \log \left(x_m^{(u)}(t) \right) \right)^{-1}, \quad (3)$$

equivalent to (2) where \widehat{V}, V are substituted by $\widehat{V}_t^{(u)}$ and $V_t^{(u)}$, respectively.

4. ESTIMATION OF ECCENTRICITIES, DIAMETER AND RADIUS

Remark 2. The following algorithms are ensured to converge exactly only if $t^* \geq 2d$. Thus, from now on we assume that the termination rule has been chosen sufficiently large to ensure $t^* \geq 2d$, and not just $t^* \geq d$.

Under our synchronicity assumptions, every max-consensus algorithm will converge after at most d communication steps. Moreover, the value at any given node u will converge after at most $e(u)$ communication steps. This can be used to let node u estimate its own eccentricity and also the diameter and the radius of the network¹.

More precisely, we define the following counters, for all $m = 1, \dots, M$ and $u \in \mathcal{V}$:

$$T_m^{(u)}(t) := \max \left\{ k \leq t \text{ s.t. } x_m^{(u)}(k) > x_m^{(u)}(k-1) \right\}. \quad (4)$$

Namely, $T_m^{(u)}(t)$ is the last time (before the current time t) where node u changed the value of the m -th scalar of its vector $\mathbf{x}^{(u)}$. Clearly, $T_m^{(u)}(t)$ can be obtained recursively by setting $T_m^{(u)}(0) = 0$ and then updating $T_m^{(u)}(t) = t - 1$ if $x_m^{(u)}(t) > x_m^{(u)}(t-1)$ and $T_m^{(u)}(t) = T_m^{(u)}(t-1)$ otherwise.

Then, we define

$$\widehat{e}^{(u)}(t) := \max_m T_m^{(u)}(t), \quad (5)$$

and we notice that, for any t , $\widehat{e}^{(u)}(t)$ is a lower bound for the eccentricity $e(u)$. Moreover, the estimate improves over time, being non-decreasing for all $t \leq t^*$, and clearly constant later. This remark means that each node u can compute an estimate of its own eccentricity, which is always a lower bound, using a slight modification of Alg. 1, without any additional communication complexity and with very little additional computation complexity.

Moreover, such local estimates can be combined in order to obtain an estimate of the diameter, with a little additional communication effort. Indeed,

$$d = \max_{u \in \mathcal{V}} e(u) \geq \max_{u \in \mathcal{V}} \widehat{e}^{(u)}(t^*) =: \widehat{d}. \quad (6)$$

Because $\widehat{e}^{(u)}(t)$ is non-decreasing with t , there is no need to wait for the final time t^* before running a max-consensus-like iteration capable of giving \widehat{d} .

¹ In this framework the necessary conditions for computability expressed in Hendrickx et al. (2011) are not valid, since the outcomes of the algorithms depend both on the initial data and on the structure of the graph.

Thus, at any iteration t and node u it is possible to compute the following estimates

$$\widehat{V}^{(u)}(t) = \left(-\frac{1}{M} \sum_{m=1}^M \log \left(x_m^{(u)}(t) \right) \right)^{-1} \quad (7)$$

$$\widehat{e}^{(u)}(t) = \begin{cases} \widehat{e}^{(u)}(t-1) & \text{if } \mathbf{x}^{(u)}(t) = \mathbf{x}^{(u)}(t-1), \\ t & \text{otherwise.} \end{cases} \quad (8)$$

$$\widehat{d}^{(u)}(t) = \max \left(\widehat{e}^{(u)}(t), \max_{v \in \mathcal{D}_1^{(u)}} \widehat{d}^{(v)}(t-1) \right). \quad (9)$$

$\widehat{V}^{(u)}(t)$ is an estimate of the number of nodes whose distance from u does not exceed t (see Remark 1), while $\widehat{e}^{(u)}(t)$ and $\widehat{d}^{(u)}(t)$ are lower bounds for the eccentricity of u and for the diameter, respectively. Both lower bounds are non-decreasing with t .

Notice that $T_m^{(u)}(t^*) = \text{dist}(u, u_m^*)$, where u_m^* is the node where the initial $x_m^{(v)}(0)$ is maximal, if there is a unique such node, and in case of multiple nodes having the same maximal initial condition it is the nearest one to u . Hence, the bound on the eccentricity $e(u)$ converges to $\widehat{e}^{(u)}(t^*) = \max_m \text{dist}(u, u_m^*)$, as it is illustrated in Fig. 1.

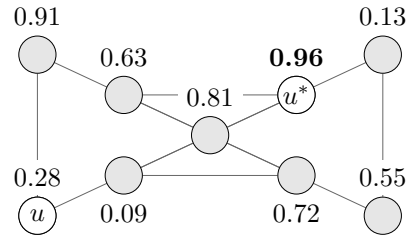


Fig. 1. In this example $M = 1$. The label near a node v represents $x_1^{(v)}(0)$, and u^* is the node with maximal initial condition. Considering agent u , its estimated eccentricity $\widehat{e}^{(u)}(t)$ will converge to 3, i.e., its distance from u^* .

Remark 3. The continuity of the distribution of the initial conditions $x_m^{(v)}(0)$ (here chosen to be uniform on $[0, 1]$) ensures that, with probability one, for each $m = 1, \dots, M$, there is a unique node $u_m^* = \arg \max_v x_m^{(v)}(0)$.

Hence, with probability one,

$$T_m^{(u)}(t^*) = \text{dist}(u, u_m^*). \quad (10)$$

Let \mathcal{U}^* to denote the (random) set $\mathcal{U}^* = \{u_1^*, \dots, u_M^*\}$. We stress that the random set \mathcal{U}^* has a paramount role and dictates the final estimation outcomes.

Eq. (10) implies that the estimators for the eccentricity and the diameter given before will converge to

$$\widehat{e}^{(u)} = \widehat{e}^{(u)}(t^*) = \max_{m=1, \dots, M} \text{dist}(u, u_m^*), \quad (11)$$

$$\widehat{d} = \widehat{d}^{(u)}(t^*) = \max_{v \in \mathcal{V}} \max_{m=1, \dots, M} \text{dist}(v, u_m^*), \quad (12)$$

where we notice that, under our assumptions, $\widehat{d} = \widehat{d}^{(u)}(t^*)$ for all $u \in \mathcal{V}$.

We also notice that (10) suggests a way to compute an upper bound for the radius. Consider in fact that, from the definition of radius,

$$r = \min_{u \in \mathcal{V}} e(u) \leq \min_{u_m^* \in \mathcal{V}} e(u_m^*) = \min_{m=1, \dots, M} e(u_m^*). \quad (13)$$

Now, for a fixed m , by definition it holds that $e(u_m^*) = \max_{u \in \mathcal{V}} \text{dist}(u, u_m^*)$. We notice that there exists a node \bar{u} s.t. $T_m^{\bar{u}}(t^*) = \text{dist}(\bar{u}, u_m^*) = \max_{u \in \mathcal{V}} \text{dist}(u, u_m^*) = e(u_m^*)$. I.e., the protocol ensures that every $e(u_m^*)$, $m = 1, \dots, M$ has been computed by some node. It follows that, having the knowledge of all the various $T_m^u(t^*)$'s, one could estimate r through

$$r \leq \min_{m=1, \dots, M} \max_{u \in \mathcal{V}} T_m^u(t^*) =: \hat{r}. \quad (14)$$

Unfortunately, this simultaneous maximization / minimization has two disadvantages with respect to the double maximization performed in (12) to obtain \hat{d} : first, it requires us to use M different counters and M separate maximizations, instead of the scalar value used before to find \hat{d} ; second, it provides an estimate which is not monotone along iterations, and which is not an upper bound at any given t , but only at the final time t^* . The details are given in Alg. 2.

Algorithm 2 Simultaneous size / eccentricities / diameter / radius estimation in anonymous networks

1: (storage allocation) each agent u stores the vectors $\mathbf{x}^{(u)} = [x_1^{(u)}, \dots, x_M^{(u)}]$, $\mathbf{C}^{(u)} = [C_1^{(u)}, \dots, C_M^{(u)}]$ and the scalars $\hat{V}^{(u)}$, $\hat{e}^{(u)}$, $\hat{d}^{(u)}$ and $\hat{r}^{(u)}$;

2: (initialization), $x_m^{(u)}(0) \sim \mathcal{U}[0, 1]$, i.i.d., $m = 1, \dots, M$, $\hat{V}^{(u)}(0) = 1$, $\hat{e}^{(u)} = 0$ and $\hat{d}^{(u)} = 0$ for all $u \in \mathcal{V}$;

3: **for** $t = 1, \dots, t^*$ **do**

4: (communication) every agent u broadcasts

$$\mathbf{x}^{(u)}(t-1) \quad \mathbf{C}^{(u)}(t-1)$$

to its neighbors;

5: (information mixing – max-consensus) every agent u computes $\mathbf{x}^{(u)}(t)$ from the vectors $\mathbf{x}^{(v)}(t-1)$ received from its neighbors: for $m = 1, \dots, M$,

$$x_m^{(u)}(t) = \max \left(x_m^{(u)}(t-1), \max_{v \in \mathcal{D}_1^{(u)}} x_m^{(v)}(t-1) \right) \quad (15)$$

6: (size estimation)

$$\hat{V}^{(u)}(t) = \left(-\frac{1}{M} \sum_{m=1}^M \log \left(x_m^{(u)}(t) \right) \right)^{-1}$$

7: (eccentricity estimation)

$$\hat{e}^{(u)}(t) = \begin{cases} \hat{e}^{(u)}(t-1) & \text{if } \mathbf{x}^{(u)}(t) = \mathbf{x}^{(u)}(t-1), \\ t & \text{otherwise.} \end{cases}$$

8: (diameter and radius estimation)

$$C_m^{(u)}(t) = \begin{cases} \max \left(C_m^{(u)}(t-1), \max_{v \in \mathcal{D}_1^{(u)}} C_m^{(v)}(t-1) \right) \\ \quad \text{if } x_m^{(u)}(t) = x_m^{(u)}(t-1), \\ t & \text{otherwise;} \end{cases}$$

$$\hat{d}^{(u)}(t) = \max_{m=1, \dots, M} C_m^{(u)}(t), \quad \hat{r}^{(u)}(t) = \min_{m=1, \dots, M} C_m^{(u)}(t).$$

9: **end for**

Finally notice that every upper bound for the radius translates immediately into an upper bound for the diameter as well, since $d \leq 2r$ (indeed, if a node w has eccentricity

$e(w) = r$, then $\text{dist}(u, v) \leq \text{dist}(u, w) + \text{dist}(w, v) \leq 2r$ for any pair of nodes u, v).

5. PERFORMANCE ANALYSIS

Statistical properties of the estimators: the statistical properties of the size estimation have been analyzed in Varagnolo et al. (2010) and are recalled in Sect. 3.

Here, we discuss properties of the estimators of the eccentricities, diameter and radius defined in Sect. 4. *In particular we focus on the final estimates* $\hat{e}^{(u)} = \hat{e}^{(u)}(t^*)$, and $\hat{d} = \hat{e}^{(u)}(t^*)$, $\hat{r} = \hat{r}^{(u)}(t^*)$ (the final diameter and radius estimates being the same for all u 's).

Recall that $\hat{e}^{(u)}(t) \leq e(u)$ and $\hat{d}^{(u)}(t) \leq d$ for all $u \in \mathcal{V}$ and for all the possible realizations of the random initial conditions. Moreover, with probability one, $\hat{r} \geq r$, and $\hat{d} \leq d \leq 2\hat{r}$. Clearly, the good property of having lower (resp. upper) bounds comes at the price of having estimators which are usually biased. The statistical description of such estimators depends on two factors:

- (1) the parameter M , which gives a trade-off between communication, computation and memory complexity and accuracy;
- (2) the topology of the network.

Throughout this section we make use of Remark 3: we compute probabilities and expectations conditioned on the event that, for each $m = 1, \dots, M$, there is a unique $u_m^* = \text{argmax}_v x_m^{(v)}(0)$. Then, such conditional probabilities and expectations are the same as the unconditioned ones, because such event has probability one.

Probability distribution of $\hat{e}^{(u)}$: consider a given agent u . By (11), $\hat{e}^{(u)} = \max_{u_m^* \in \mathcal{U}^*} \text{dist}(u, u_m^*)$. Hence, $\hat{e}^{(u)}$ is exactly the eccentricity $e(u)$ if and only if there is a $u_m^* \in \mathcal{U}^*$ having $\text{dist}(u, u_m^*) = e(u)$. Moreover $\hat{e}^{(u)}$ is at least k if and only if there is a $u_m^* \in \mathcal{U}^*$ s.t. $\text{dist}(u, u_m^*) \geq k$. Thus,

$$\begin{aligned} \mathbb{P}[\hat{e}^{(u)} \geq k] &= \mathbb{P} \left[\mathcal{U}^* \cap \left(\bigcup_{h=k}^d \mathcal{D}_h^{(u)} \right) \neq \emptyset \right] \\ &= 1 - \left(1 - \frac{\sum_{h \geq k} D_h^{(u)}}{V} \right)^M, \end{aligned}$$

$$\begin{aligned} \mathbb{P}[\hat{e}^{(u)} = k] &= \mathbb{P}[\hat{e}^{(u)} \geq k] - \mathbb{P}[\hat{e}^{(u)} \geq k+1] \\ &= \left(1 - \frac{\sum_{h \geq k+1} D_h^{(u)}}{V} \right)^M - \left(1 - \frac{\sum_{h \geq k} D_h^{(u)}}{V} \right)^M. \end{aligned}$$

Probability distribution of \hat{d} : by (12),

$$\hat{d} = \max_{u \in \mathcal{V}} \max_{u_m^* \in \mathcal{U}^*} \text{dist}(u, u_m^*) = \max_{u_m^* \in \mathcal{U}^*} e(u_m^*).$$

This implies that $\hat{d} = d$ if and only if there is a $u_m^* \in \mathcal{U}^*$ having $e(u_m^*) = d$. More generally, $\hat{d} \geq k$ if and only if there is a $u_m^* \in \mathcal{U}^*$ in the union of the sets \mathcal{D}_h with $h \geq k$. Thus, as before,

$$\mathbb{P}[d \geq k] = 1 - \left(1 - \frac{\sum_{h \geq k} D_h}{V} \right)^M,$$

$$\mathbb{P}[d = k] = \left(1 - \frac{\sum_{h \geq k+1} D_h}{V}\right)^M - \left(1 - \frac{\sum_{h \geq k} D_h}{V}\right)^M.$$

Probability distribution of \hat{r} : Eq. (14) implies that $\hat{r} = r$ if and only if there is a u_m^* s.t. $e(u_m^*) = r$, i.e., s.t. $u_m^* \in \mathcal{D}_r$. More generally, $\hat{r} \leq k$ if and only if there is a $u_m^* \in \mathcal{U}^*$ in the union of the sets \mathcal{D}_h with $h \leq k$. Thus, as before,

$$\mathbb{P}[\hat{r} \leq k] = 1 - \left(1 - \frac{\sum_{h \leq k} D_h}{V}\right)^M,$$

$$\mathbb{P}[\hat{r} = k] = \left(1 - \frac{\sum_{h \leq k-1} D_h}{V}\right)^M - \left(1 - \frac{\sum_{h \leq k} D_h}{V}\right)^M.$$

Expected errors: from the previous expressions we can compute the bias of the estimators with the use of the following very simple remarks: $D_h^{(u)} = 0$ for all $h > e(u)$, $D_h = 0$ for all $h < r$ and $h > d$, and $\sum_{0 \leq h \leq e(u)} D_h^{(u)} = \sum_{r \leq h \leq d} D_h = V$. We obtain

$$\mathbb{E}[e(u) - \hat{e}^{(u)}] = \sum_{k=1}^{e(u)} \left(1 - \frac{\sum_{h \geq k} D_h^{(u)}}{V}\right)^M, \quad (16)$$

$$\mathbb{E}[d - \hat{d}] = \sum_{k=1}^d \left(1 - \frac{\sum_{h \geq k} D_h}{V}\right)^M, \quad (17)$$

$$\mathbb{E}[\hat{r} - r] = \sum_{k=r}^{d-1} \left(1 - \frac{\sum_{h \leq k} D_h}{V}\right)^M. \quad (18)$$

Bounds on the probabilities of correct estimates and on the expected errors: all the above expressions show that the quality of the estimates $\hat{e}^{(u)}$, \hat{d} and \hat{r} heavily depends on the graph topology. In the following, we give bounds that are valid for all graphs, as they describe the worse and best performance for a given size V . Due to space limitation, we omit the proofs, which can be found in Garin et al. (2012).

The probabilities that the estimates are exact can be bounded as follows:

$$1 - \left(1 - \frac{1}{V}\right)^M \leq \mathbb{P}[\hat{e}^{(u)} = e(u)] \leq 1 - \frac{1}{V^M}, \quad (19)$$

$$1 - \left(1 - \frac{2}{V}\right)^M \leq \mathbb{P}[\hat{d} = d] \leq 1, \quad (20)$$

$$1 - \left(1 - \frac{1}{V}\right)^M \leq \mathbb{P}[\hat{r} = r] \leq 1, \quad (21)$$

while the expected errors can be bounded as follows:

$$\frac{1}{V^M} \leq \mathbb{E}[e(u) - \hat{e}^{(u)}] \leq \frac{1}{V^M} \sum_{k=1}^{V-1} k^M, \quad (22)$$

$$0 \leq \mathbb{E}[d - \hat{d}] \leq \sum_{k=1}^{\lfloor \frac{V-1}{2} \rfloor} \left(1 - \frac{2k}{V}\right)^M, \quad (23)$$

$$0 \leq \mathbb{E}[\hat{r} - r] \leq \sum_{k=1}^{\lfloor \frac{V-1}{2} \rfloor} \left(1 - \frac{2k-1}{V}\right)^M. \quad (24)$$

We notice that *all the previous bounds are tight*, in the sense that for each bound and for any V there exists at least one graph achieving that bound (at least at one leaf, for the eccentricity), as clearly shown by the following examples:

- a circle graph achieves the upper bounds on $\mathbb{P}[\hat{d} = d]$ and $\mathbb{P}[\hat{r} = r]$, and the lower bounds on $\mathbb{E}[d - \hat{d}]$ and $\mathbb{E}[\hat{r} - r]$;
- a line graph achieves the lower bound for $\mathbb{P}[\hat{d} = d]$ and the upper bound for $\mathbb{E}[d - \hat{d}]$. The two leafs of a line graph also achieve the lower bound for $\mathbb{P}[\hat{e}^{(u)} = e(u)]$ and the upper bound for $\mathbb{E}[e(u) - \hat{e}^{(u)}]$. Moreover, the lower bound on $\mathbb{P}[\hat{r} = r]$ and the upper bound on $\mathbb{E}[\hat{r} - r]$ are achieved by a line graph if V is odd, and by a graph such as in Fig. 2 if V is even;
- a complete graph achieves all upper bounds for probability of exact estimation and all lower bounds for expected errors.

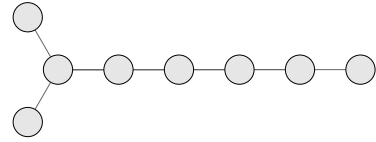


Fig. 2. Example of a graph with an even number of nodes that achieves the lower bound on $\mathbb{P}[\hat{r} = r]$ and the upper bound on $\mathbb{E}[\hat{r} - r]$.

Remark 4. Eq. (22) implies that, even with the best graph topology, the estimator for the eccentricity can make wrong estimates. These errors correspond to the case where an agent u generates all the maximal initial values, an event having non-zero probability.

Simulation results: here we consider 10^4 connected realizations of a random geometric graph, with 30 nodes, randomly and independently deployed in $[0, 1] \times [0, 1]$ and with communication radius 0.3 (see, e.g., Fig. 3). For each graph we run Alg. 2 for various values of M and plot the dependence of the expected errors $\mathbb{E}[d - \hat{d}]$ and $\mathbb{E}[\hat{r} - r]$ on M in Fig. 4. It is immediate to recognize that the expected errors decay with M , although not exponentially fast.

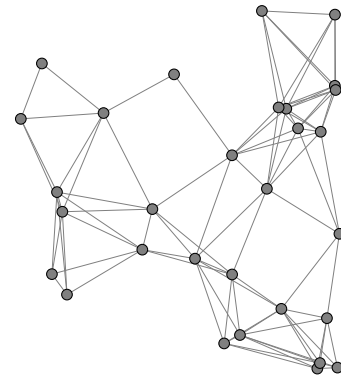


Fig. 3. A typical network considered in our analysis.

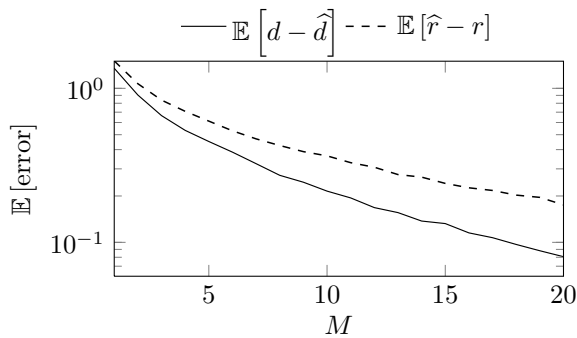


Fig. 4. Dependence of $\mathbb{E} [d - \hat{d}]$ and $\mathbb{E} [\hat{r} - r]$ on M .

6. CONCLUSIONS

The availability of information on the topology of a communication network is vital for diagnosis and self-configuration purposes. In this paper we proposed and statistically analyzed a distributed and privacy-preserving algorithm that allows the agents of a distributed system to estimate some fundamental parameters, such as the diameter and radius of the network and their own eccentricity.

The considered strategy has several advantages. Namely, it is fully distributed and easy to be implemented, it has small computational and memory complexities, it can be extended to time-varying situations and it relies on max-consensus protocols, which are intrinsically fast information exchange mechanisms.

This work proposed a full statistical description of the performance of these estimators. We provided the analytical characterization of the probabilities of errors and the expected errors, and offered several bounds for these quantities. These results are essential to understand the effects of the topology on the outcomes, and to evaluate the intrinsic trade-off between the performance of the estimators and the corresponding communication complexity.

Despite providing full statistical descriptions, the offered results refer to a specific framework and are thus partial. Namely, the algorithm assumes synchronized and perfect communications (i.e., considers neither quantization issues nor packets losses), and an ideal random numbers generation mechanism (i.e., assumes sampling to be from an absolutely continuous distribution). Future extensions should thus address these limitations and extend the previous results, in sight of an holistic analysis of the fundamental limits of topology identification in anonymous networks.

REFERENCES

Ahmed, N., Kanhere, S.S., and Jha, S. (2005). The holes problem in wireless sensor networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(2), 1 – 14.

Aingworth, D., Chekuri, C., Motwani, R., and Indyk, P. (1996). Fast estimation of diameter and shortest paths (without matrix multiplication). In *Proceedings of the seventh annual ACM-SIAM symposium on Discrete algorithms*, 547–553.

Almeida, P.S.S., Baquero, C., and Cunha, A. (2012). Fast Distributed Computation of Distances in Networks. In

IEEE Conference on Decision and Control (submitted), 1–13 (arXiv:111.6087 [csDC]).

Baquero, C., Almeida, P.S.S., Menezes, R., and Jesus, P. (2012). Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes. *IEEE Transactions on Parallel and Distributed Systems*, 23(4), 668 – 675.

Boittmanis, K., Freivalds, K., Ledinsk, P., and Opmanis, R. (2006). Fast and Simple Approximation of the Diameter and Radius of a Graph. *Experimental Algorithms*, 4007, 98 – 108.

Cardoso, J.C.S., Baquero, C., and Almeida, P.S. (2009). Probabilistic Estimation of Network Size and Diameter. In *Fourth Latin-American Symposium on Dependable Computing*, 33–40. IEEE, João Pessoa, Brasil.

Chen, B., Jamieson, K., Balakrishnan, H., and Morris, R. (2002). Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks. *Wireless Networks*, 8(5), 481–494.

Cichon, J., Lemiesz, J., and Zawada, M. (2011). On Cardinality Estimation Protocols for Wireless Sensor Networks. *Ad-hoc, mobile, and wireless networks*, 6811, 322–331.

Corneil, D.G., Dragan, F.F., Habib, M., and Paul, C. (2001). Diameter determination on restricted graph families. *Discrete Applied Mathematics*, 113(2-3), 143–166.

Crescenzi, P., Grossi, R., Imbrenda, C., Lanzi, L., and Marino, A. (2010). Finding the Diameter in Real-World Graphs - Experimentally Turning a Lower Bound into an Upper Bound. In *European Conference on Algorithms*.

Deb, B., Bhatnagar, S., and Nath, B. (2004). STREAM: Sensor Topology Retrieval at Multiple Resolutions. *Telecommunication Systems*, 26(2-4), 285–320.

Elkin, M. (2001). Computing almost shortest paths. In *Proceedings of the twentieth annual ACM symposium on Principles of distributed computing - PODC '01*, 53–62. ACM Press, New York, New York, USA.

Garin, F., Varagnolo, D., and Johansson, K.H. (2012). Distributed estimation of diameter, radius and eccentricities in anonymous networks. Technical report, HAL Inria: <http://hal.inria.fr/hal-00717580/en>.

Hall, C.P. (2010). *Peer-to-Peer Algorithms for Sampling Generic Topologies*. Ph.D. thesis, Università della Svizzera Italiana.

Hendrickx, J.M., Olshevsky, A., and Tsitsiklis, J.N. (2011). Distributed anonymous discrete function computation. *IEEE Transactions on Automatic Control*, 56(10), 2276–2289.

Li, M. and Yang, B. (2006). A Survey on Topology issues in Wireless Sensor Network. In *Proceedings of the International Conference on Wireless Networks*.

Lynch, N.A. (1996). *Distributed Algorithms*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Varagnolo, D., Pilonetto, G., and Schenato, L. (2010). Distributed statistical estimation of the number of nodes in Sensor Networks. In *IEEE Conference on Decision and Control*, 1498–1503. Atlanta, USA.

Zwick, U. (2001). Exact and Approximate Distances in Graphs – A Survey. In F.M. Heide (ed.), *Algorithms - ESA 2001*, volume 2161 of *Lecture Notes in Computer Science*, 33–48. Springer Berlin Heidelberg, Berlin, Heidelberg.