Fast Distributed Estimation of Probability Mass Functions over Anonymous Networks

Håkan Terelius⋆, Damiano Varagnolo⋆, Carlos Baquero† and Karl Henrik Johansson⋆

***Abstract*— The aggregation and estimation of values over networks is fundamental for distributed applications, such as wireless sensor networks. Estimating the average, minimal and maximal values has already been extensively studied in the literature. In this paper, we focus on estimating the entire distribution of values in a network with anonymous agents. In particular, we compare two different estimation strategies in terms of their convergence speed, accuracy and communication cost. The first strategy is deterministic and based on the average consensus protocol, while the second strategy is probabilistic and based on the max consensus protocol. We characterize both strategies' statistical performance, and present trade-offs and guidelines for choosing the right estimation scheme.**

***Index Terms*— distributed computation, consensus, data aggregation, order statistics**

I. Introduction

The topology of networked systems, i.e., the structure of the local interactions among agents, has a crucial influence on the macroscopic properties of the whole system.

Here we follow this stream by proposing and characterizing a specific tool that estimates Probability Mass Functions (PMFs) over networks. We consider networks of collaborative anonymous agents, although anonymous, technically each agent is allowed to distinguish messages from its direct neighbors.

A. Literature review

The vast literature on the estimation of probability densities / mass functions over networks can be divided in the main classes of parametric and non-parametric approaches.

Parametric approaches generally assume the estimand to have an certain structure before obtaining observations, e.g., they assume the estimand to be a mixture of a certain number of Gaussians. Examples are the distributed implementations of the Expectation-Maximization (EM) algorithm [1], [2], [3], [4]. Nonparametric approaches instead do not fix the structure a priori, but rather select it from the observations. This class comprises the various distributed kernel density estimation [5], classification [6] and clustering approaches [7].

Additional to the parametric / non-parametric classification, characterizing the existing literature can also rely on how the information is propagated and aggregated over the network. We notice strategies based on more or less pre-estabilished hierarchical tree routing structures, where the various nodes compute summaries of the empirical distributions in their sub-trees and propagate them towards the root, eventually obtaining the approximated statistics of the whole network in a bottom-up fashion [8], [9], [10], [11]. Other strategies can instead be based on gossip communications, and exploit averaging techniques to explicitly compute Cumulative Distribution Functions (CDFs) [12], [13], [14]. Other techniques can eventually be based on applying estimation methods to directly estimate of many agents are in a certain specific state [15], [16].

B. Statement of contributions

We propose a novel algorithm fostering the properties: • be *symmetrically distributed*, i.e., without leaders / leader election steps, and with agents executing the same algorithm in parallel. • be *privacy preserving*, i.e., avoiding the possibility of tracing or characterizing a single agent. • exploit *aggregation techniques*, where the size of the exchanged information packets is constant in time. • be *fast*, i.e., s.t. the time for all the agents to share the same estimate is small.

More specifically, we propose and analyze a strategy that is based on max-consensus (see Sec. III). As this takes no more rounds to converge as the rounds needed to transmit information between arbitrary nodes in the network, this is the fastest aggregation mechanism possible over networks. This emphasis on *fast convergence techniques* is given by the consideration that time can be the crucial factor in many practical situations (e.g., in vehicular networks).

From an algorithmic point of view our strategy departs from [12], [13], [17] by substituting the average consensus schemes with max-consensus ones. This apparently minor modification actually makes the two estimators completely different, and opens a variety of novel problems. In fact, as will be clear later, while the average consensus scheme requires exchanging very few scalars per communication and where the agents computes the *exact* PMF only asymptotically in time, the max consensus scheme converges much faster than the average one, but not to the exact value. Indeed, the statistical performance depend on how many scalars are exchanged per communication. *Here we specifically characterize the temporal behavior of the performance of this max-consensus strategy, stating when it is preferable to the original one.*

⋆ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Emails: { hakante, damiano, kallej } @kth.se

†HASLab, INESC TEC & Universidade do Minho, Braga, Portugal. Email: cbm@di.uminho.pt

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, the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

C. Structure of the paper

The manuscript is organized as follows: we start with the formal problem definition in Sec. II, and then describe the two considered estimation strategies in Sec. III. Thus we provide the statistical characterization of the novel scheme in Sec. IV. We then compare the performances of the two schemes in Sec. V, and conclude with some remarks and future research directions in Sec. VI.

II. Statement of the estimation problem

Consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of $V = |\mathcal{V}|$ agents limited to the communication links $\mathcal{E}$. Let $\mathcal{V}_i$ denote the set of neighbors of agent i, and $\mathcal{V}_t^{(i)}$ the set of the t-steps neighbors of agent i. We recall that $\mathcal{V}_t^{(i)}$ can be defined for $t = 0$ as $\mathcal{V}_0^{(i)} = \{i\}$ and, for $t \geq 1$, through the recursion

$$\mathcal{V}_t^{(i)} \doteq \bigcup_{(j,i)\in\mathcal{E}} \mathcal{V}_{t-1}^{(j)} . \tag{1}$$

Let then every agent $i \in \mathcal{V}$ belong to a given discrete state $z^{(i)} \in \mathbb{N}_B \doteq \{0, \ldots, B-1\}$ ($\mathbb{N}_B$ being the set of plausible states). We are then interested in distributively estimating the relative frequencies of the local states $z^{(1)}, \ldots, z^{(V)}$, i.e., if $n_b \doteq \left|\{i \text{ s.t. } z^{(i)} = b\}\right|$ is the number of agents in state b, then we are estimating the PMF

$$p_b \doteq \frac{n_b}{V} \qquad b \in \mathbb{N}_B \tag{2}$$

given that the network size V is unknown while the state dimension B is known.

We restrict our focus to distributed algorithms where each agent $i \in \mathcal{V}$ has a local variable $x^{(i)}(t)$ that can be modified at time $t+1$ by accessing the $x^{(j)}(t)$'s of the neighboring nodes and performing the aggregation operation

$$x^{(i)}(t+1) = f\left(x^{(i)}(t), x^{(j_1)}(t), x^{(j_2)}(t), \ldots\right), \qquad j_1, j_2, \ldots \in \mathcal{V}_i$$

that preserves the dimension of $x^{(i)}(t)$ (i.e., $x^{(i)}(t+1)$ and $x^{(i)}(t)$ have the same dimensionality). Furthermore, we assume that at every time t each agent can compute a local estimate of the PMF function only considering the local variable $x^{(i)}(t)$, i.e., we let

$$\widehat{p}_b^{(i)}(t) = g\left(x^{(i)}(t)\right)$$

for an opportune $g(\cdot)$.

The estimation strategy is thus defined by the initial local variables $x^{(i)}(0)$, the update function f and the estimation function g. To compare different estimation strategies we consider the Mean Squared Error (MSE) as a performance index J, i.e.,

$$J\big(\widehat{p}_1, \ldots, \widehat{p}_B\big) \doteq \mathbb{E}\left[\frac{1}{V \cdot B} \sum_{b\in\mathbb{N}_B, i\in\mathcal{V}} \left(p_b - \widehat{p}_b^{(i)}\right)^2\right] \tag{3}$$

where the expectation is taken over the initial conditions.

Remark 1 For notational simplicity we consider static networks. Nonetheless it is straightforward to handle time-varying topologies by substituting the edges $\mathcal{E}$ with a time-dependent set $\mathcal{E}(t)$, and the neighborhoods $\mathcal{V}^{(i)}$ with the time-dependent counterparts $\mathcal{V}^{(i)}(t)$.

The problem analyzed in this manuscript is thus to propose and compare different estimation schemes.

III. Estimators based on consensus protocols

We consider two particular estimators, one based on average consensus strategies (see also [12], [13], [17]), and a novel one based on max consensus strategies and structurally similar to the size estimation techniques in [18], [19].

In the following, we abstract away the message transmission and consider a distributed system where agents communicate by synchronous rounds that occur in locksteps. We also consider that at each round, and in each edge, only a constant size message is transmitted, and that no messages are lost. The need to distinguish neighbors can be attained with local IDs that do not depend on the total number of agents.

Remark 2 For notational simplicity we consider synchronous communications. Nonetheless this could be relaxed for both estimators, since they both can be adapted to operate with gossip transmissions.

A. Estimator based on Average consensus

In the average consensus protocol, the local variable is a B-dimensional real vector $x^{(i)}(t) \in \mathbb{R}^B$ containing the estimate of the PMF. At initialization, each node set its local variable based only on its own state,

$$x_b^{(i)}(0) = \begin{cases} 1, & \text{if } z^{(i)} = b \\ 0, & \text{otherwise.} \end{cases}$$

It is known that if at each time the local variables are updated with an average consensus update like

$$x_b^{(i)}(t) = \frac{\sum_{j\in\mathcal{V}_i} x_b^{(j)}(t-1)}{|\mathcal{V}_i|}, \quad b \in \mathbb{N}_B \tag{4}$$

then, assuming perfect computations[1], $x_b^{(i)}(t)$ converges to the average of the initial values [21]. Thus

$$x_b^{(i)}(t) \xrightarrow{t\to+\infty} \frac{1}{V} \sum_{j\in\mathcal{V}} x_b^{(i)}(0) = \frac{n_b}{V} = p_b.$$

The PMF estimate g can thus be set as

$$\widehat{p}_b^{(i)}(t) = x_b^{(i)}(t). \tag{5}$$

To describe the convergence properties of the algorithm, recall that the average consensus algorithm can be written on matrix form as

$$x_b(t) = W x_b(t-1) = W^t x_b(0)$$

[1] For simplicity we do not consider quantization effects. For the effects of quantization on the convergence properties of average consensus algorithms see, e.g., [20].

where W is the weight matrix (for example chosen as the so called Metropolis weights). The estimation error can then be bounded by

$$\begin{aligned}\left|\left|p_b - \widehat{p}_b(t)\right|\right|_2 = \left|\left|W p_b - W\widehat{p}_b(t-1)\right|\right|_2 = \\ \left|\left|W^t p_b - W^t\widehat{p}_b(0)\right|\right|_2 \leq ||W||_2^t \left|\left|p_b - \widehat{p}_b(0)\right|\right|_2,\end{aligned}$$

i.e., the error is bounded by an exponential function

$$\left|\left|p_b - \widehat{p}_b(t)\right|\right|_2 \leq c e^{-\alpha t}$$

where c and α depend on the initial condition, the network topology and the choice of the weights.

We notice that we do not consider more advanced protocols, such as accelerated average consensus, e.g., [22], or finite-time average consensus, e.g., [23]. The rationale for this choice is that we want to characterize the simplest averaging algorithm, with the smallest demands from both communication and computational points of view.

B. Estimator based on Max consensus

In the max consensus protocol, the local variable is a $B \times M$-dimensional real matrix $x^{(i)}(t) \in \mathbb{R}^{B\times M}$ which scalar components are initially and locally set based only on the local state as

$$x_{b,m}^{(i)}(0) \sim \begin{cases} \mathcal{U}\left[0,1\right], & \text{if } z^{(i)} = b \\ 0, & \text{otherwise} \end{cases}$$

where $\mathcal{U}\left[0,1\right]$ is the uniform distribution between 0 and 1. Then at each time t, the local variables are updated with the max consensus update

$$x_{b,m}^{(i)}(t) = \max_{j\in\mathcal{V}_i} x_{b,m}^{(j)}(t-1), \quad b \in \mathbb{N}_B, m = 1, \ldots, M. \quad (6)$$

Notice that the definition of t-steps neighborhood $\mathcal{V}_t^{(i)}$ precisely captures which are the agents that contributed, from a statistical point of view, to the generation of $x_{b,m}^{(i)}(t)$, i.e.,

$$x_{b,m}^{(i)}(t) = \max_{j\in\mathcal{V}_t^{(i)}} \left\{ x_{b,m}^{(j)}(0) \right\}. \quad (7)$$

Let $V_t^{(i)} \doteq \left|\mathcal{V}_t^{(i)}\right|$,

$$p_b^{(i)}(t) \doteq \frac{\left|\left\{ i \in \mathcal{V}_t^{(i)} \text{ s.t. } z^{(i)} = b \right\}\right|}{V_t^{(i)}}, \quad (8)$$

and $n_b^{(i)}(t) \doteq p_b^{(i)}(t) V_t^{(i)}$. As shown in the following Sec. IV, the Maximum Likelihood (ML) estimator for $n_b^{(i)}(t)$ given the $x_{b,m}^{(i)}(t)$'s is

$$\widehat{n}_b^{(i)} = \left(\frac{1}{M} \sum_{m=1}^{M} -\ln\left(x_{b,m}^{(i)}\right) \right)^{-1}. \quad (9)$$

Since

$$p_b^{(i)}(t) = \frac{p_b^{(i)}(t)}{\sum_{\beta\in\mathbb{N}_B} p_\beta^{(i)}(t)} = \frac{n_b^{(i)}(t)}{\sum_{\beta\in\mathbb{N}_B} n_\beta^{(i)}(t)}$$

because of the functional invariance property of ML estimators [24, Thm. 7.2.10, p. 320], the ML estimate of $p_b^{(i)}(t)$ given the $x_{b,m}^{(i)}(t)$'s is

$$\widehat{p}_b^{(i)}(t) = \frac{\widehat{n}_b^{(i)}(t)}{\sum_{\beta\in\mathbb{N}_B} \widehat{n}_\beta^{(i)}(t)}. \quad (10)$$

Then, since for $t \geq d$ (d being the network diameter) $n_b^{(i)}(t) = n_b$, the PMF estimated $p_1^{(i)}(t), \ldots, p_B^{(i)}(t)$ converges to an estimate of the global PMF $p_1, \ldots, p_B$.

Remarkably, this estimator thus provide information not delivered by the average consensus scheme. In fact it provides meaningful estimates of the distributions of the states in every t-steps neighborhood. Considering a certain fixed agent i, the set of the $p_b^{(i)}(0), p_b^{(i)}(1), \ldots$ correspond to local views expanding up to the whole network that can be used by i to rapidly infer, e.g., if close neighbors tend to have the similar states.

We notice that estimator (10) has strong similarities with the size estimators proposed, e.g., in [25], [26], [27]. Nonetheless, as reported in the following section, its statistical properties are essentially different since each vector $\left[\widehat{p}_1^{(i)}(t), \ldots, \widehat{p}_B^{(i)}(t)\right]$ has correlated components.

We also notice that opportune termination rules can be based on estimates of the diameter d of the network, again obtainable exploiting max consensus approaches as in [19], [28].

C. Summary of the differences between the two estimators

The max consensus scheme 10 converges in d steps to an estimate of the true PMF. Given a fixed M, thus its MSE J (see (3)) will vary up to $t = d$ and then remain constant. Increasing M, the MSE curves are also expected to get closer and closer to zero, due to the consistency property of ML estimators. The average consensus scheme (5) is instead in general converging asymptotically for $t \to +\infty$. These comments are graphically represented in fig. 1.

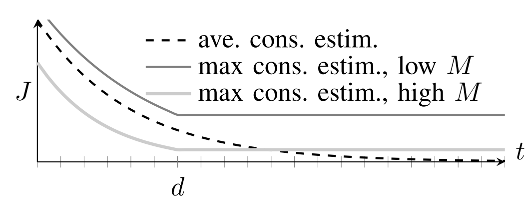


Fig. 1: Graphical representation of the properties from the estimators. By increasing M it is possible to let the max consensus estimator 10 perform better than the average consensus scheme (5) for $t \leq d$.

The aim is then to find conditions on M and on the network for which it is possible to state which algorithm is preferrable for $t \leq d$, i.e., when time is a concern. To solve this we first need to describe the statistical properties of the max consensus estimator.

IV. Statistical characterization of the max consensus PMF estimator

For notational simplicity we consider the a-consensus situation, where $x_{b,m}^{(i)}(t) = x_{b,m} \doteq \max_{i \in \mathcal{V}} \left\{ x_{b,m}^{(i)}(0) \right\}$.

Consider then that the joint Probability Density Function (PDF) $p\left(\widehat{n}_b \; ; \; n_1, \ldots, n_B, M\right)$ characterizes also $p\left(\widehat{n}_b^{(i)}(t) \; ; \; n_1^{(i)}(t), \ldots, n_B^{(i)}(t), M\right)$ and the moments of $\widehat{n}_b$ and $\widehat{n}_b^{(i)}(t)$. To derive the former distribution we then consider that $b \neq \beta$ implies $x_{b,m}$ to be statistically independent on the the parameter n_β. Thus, from simple order-statistics arguments [29],

$$p\left(x_{b,m} \; ; \; n_1, \ldots, n_B\right) = p\left(x_{b,m} \; ; \; n_b\right) = n_b \left(x_{b,m}\right)^{n_b - 1}$$

for all m (we omit the dependency on the parameter M for notational brevity). Since the $x_{b,m}$'s are i.i.d.,

$$\begin{aligned} p\left(x_{b,1}, \ldots x_{b,M} \; ; \; n_b\right) &= \prod_{m=1}^{M} p\left(x_{b,m} \; ; \; n_b\right) \\ &= n_b^M \prod_{m=1}^{M} \left(x_{b,m}\right)^{n_b - 1} \end{aligned} \tag{11}$$

To derive $p\left(\widehat{n}_b \; ; \; n_b\right)$ we then consider that $z \doteq -\ln\left(\left(x_{b,m}\right)\right)$ is an exponential random variable with rate n_b, i.e.,

$$p\left(z \; ; \; n_b\right) = \begin{cases} n_b e^{-n_b z} & \text{if } z \geq 0 \\ 0 & \text{otherwise .} \end{cases} \tag{12}$$

Considering (9), $M\widehat{n}_b^{-1}$ is the sum of M i.i.d. exponential random variables with rate n_b, i.e., $M\widehat{n}_b^{-1}$ is a Γ variate with shape M and scale $\dfrac{1}{n_b}$. Thus $M^{-1}\widehat{n}_b \sim \text{I-}\Gamma\left(M, n_b\right)$, i.e.,

$$\begin{aligned} p\left(\widehat{n}_b \; ; \; n_b, M\right) &= \text{I-}\Gamma\left(M, M n_b\right) \\ &= \Gamma\left(M\right)^{-1} \frac{1}{\widehat{n}_b} \left(\frac{M n_b}{\widehat{n}_b}\right)^M \exp\left(-\frac{M n_b}{\widehat{n}_b}\right) \end{aligned}$$

where M is the shape and $M n_b$ the scale. Given (10), $\widehat{p}_b$ is thus the ratio of correlated sums of inverse-Gamma variates, each with its own scale.

Unfortunately to the best of our knowledge there exists no currently available literature describing the distribution of this kind of variates. The closest manuscripts in fact characterize ratios of the form $\frac{x}{x+y}$ where x and y are independent inverse Γ variates [30]. Moreover both the Gamma and inverse Gamma distributions are not closed, i.e., linear combinations of independent copies of these kind of variates have not the same original distribution up to location and scale parameters, see, e.g., [31]. This means that there is no possibility to reduce the fraction (10) to the case described in [30], and characterization of the statistical properties of $\widehat{p}_b$ must rely on Monte Carlo (MC) integration methods.

Case $\mathbb{N}_B = \{0, 1\}$

In this case $\widehat{p}_b^{(i)}(t)$ is a special ratio that is described in [30]. Considering the results therein and p_0 for simplicity,

$$p_{\widehat{p}_0}\left(x \; ; \; n_0, n_1, M\right) = \frac{\Big(x(1-x)\Big)^{M-1}}{\left(\dfrac{n_0}{n_1}\right)^M B\left(M, M\right)} \left(1 + \frac{n_1 - n_0}{n_0} x\right)^{-2M} \tag{13}$$

where $B\left(\cdot, \cdot\right)$ is the Beta function and $x \in [0, 1]$. Its cumulative distribution is given by (17) where

$${}_2F_1\left(a, b; c; x\right) \doteq \sum_{i=0}^{+\infty} \frac{\left(a\right)_i \left(b\right)_i}{\left(c\right)_i \cdot i!} x^i \tag{14}$$

is the Gauss hypergeometric function and

$$\left(x\right)_i \doteq x(x+1) \cdots (x+i-1) \tag{15}$$

is the so called Pochhammer symbol (with the convention that $\left(x\right)_0 = 1$). From this, it is possible to compute the moments of $\widehat{p}_0$ (and thus of $\widehat{p}_0 - \mathbb{E}\left[\widehat{p}_0\right]$) using the relation

$$\mathbb{E}\left[\left(\widehat{p}_0\right)^k\right] = \begin{cases} \dfrac{B\left(M+k, M\right)}{B\left(M, M\right)} \mathcal{F}(k, M, n_0, n_1) \\ \qquad\qquad\qquad \text{if } n_0 > n_1 \\ \left(\dfrac{n_0}{n_1}\right)^k \dfrac{B\left(M+k, M\right)}{B\left(M, M\right)} \mathcal{F}(k, M, n_1, n_0) \\ \qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{16}$$

where

$$\mathcal{F}(k, M, a, b) \doteq {}_2F_1\left(k, M; 2M + k; \frac{a - b}{a}\right)$$

(notice that n_0 and n_1 appear in inverted positions in the two cases in (16)).

It is possible to notice that when $n_0 = n_1$ then the estimators are unbiased for every M, otherwise –as expected– they are only asymptotically unbiased (for $M \to +\infty$).

Numerical evaluations of the dependency of the relative bias and MSE of the estimators on the design parameter M and on the distribution of the states are shown in figures 2 and 3. It can be noticed that the MSE performances follow the typical $O\left(\frac{1}{M}\right)$ proper of this kind of estimators.

As a remark, the performances indicators summarized in figures 2 and 3 are valid for general $\widehat{p}_b^{(i)}(t)$'s when associated to the relative local $n_b^{(i)}(t)$'s. The derivations of this section thus characterize also the behavior of the estimators during the transient.

V. Comparisons

Here we compare the performance between the average consensus based estimator (5) and the max consensus based estimator (10) during also their transients. Our primary goal is to determine when to choose each algorithm, and how to tune the parameter M for the max consensus estimator.

We consider four different network topologies, i.e., the line topology (fig. 4a), the cyclic topology (fig. 4b), the cyclic

$$F_{\widehat{p}_0}\left(x\ ;\ n_0, n_1, M\right) = \frac{\left(1 + \dfrac{n_1}{n_0}\dfrac{1-x}{x}\right)^{-M}}{MB\left(M, M\right)} \cdot {}_2F_1\left(M, 1-M; M+1; \left(1 + \frac{n_1}{n_0}\frac{1-x}{x}\right)^{-1}\right) \tag{17}$$

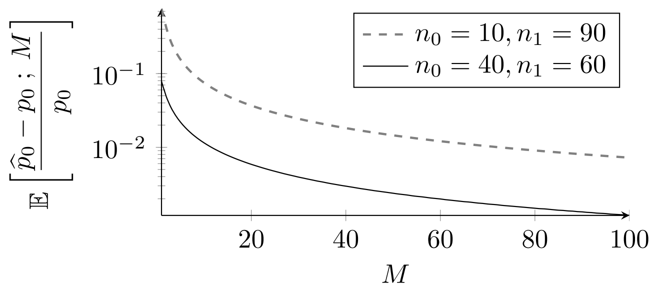


Fig. 2: Dependency of the relative bias $\mathbb{E}\left[\frac{\widehat{p}_0 - p_0}{p_0}\ ;\ M\right]$ on M for various values of n_0 and n_1. The estimators are unbiased for every M if $n_0 = n_1$.

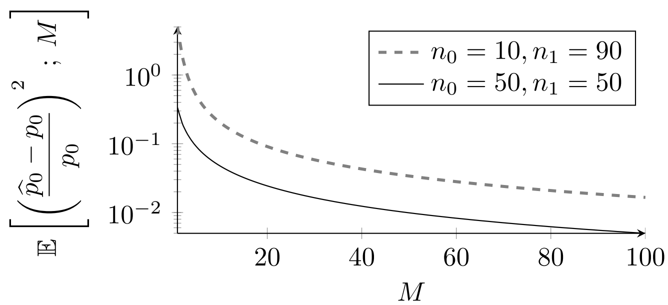
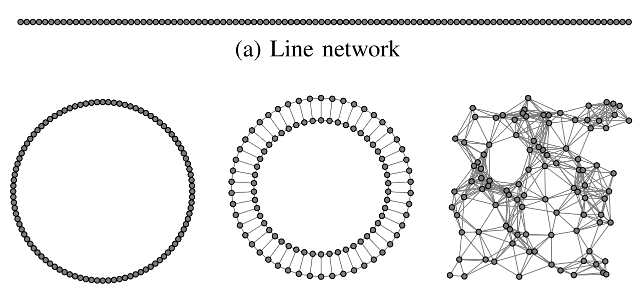


Fig. 3: Dependency of the relative MSE $\mathbb{E}\left[\left(\frac{\widehat{p}_0 - p_0}{p_0}\right)^2\ ;\ M\right]$ on M for various values of n_0 and n_1.



(a) Line network

(b) Cyclic network

(c) Cyclic grid network (2×50)

(d) Geometric random network

Fig. 4: Network topologies, with 100 nodes.

grid topology (fig. 4c), and the geometric random topology (fig. 4d), each network consisting of 100 agents.

We evaluate the algorithms with Monte Carlo (MC) simulations, using the MSE (3) as the estimation performance index, where the mean is taken over all agents and all MC runs. For each network the communication protocol proceeds in lock-step synchronous rounds, where nodes cyclically repeat the steps described in (4) and (6). We also assume the states $z^{(i)}$ to be fixed across the single MC execution.

• *First experiment - fig. 5:* here we select a random initial state for each MC run, where each agent is in state $z^{(i)} = 0$ or $z^{(i)} = 1$ with equal probability. The figure shows the 95% confidence intervals for both the average consensus based estimator as well as for the max consensus based estimator with $M = 10$, $M = 100$ and $M = 1000$.

As expected, the average consensus based estimator converges *asymptotically* to the true value. The max consensus based estimator converges instead *in a finite time* (after $d \doteq$ diameter of the graph steps) to an estimate which MSE decreases with M. In this scenario (notice that the a big portion of the area related to $M = 100$ is covered by the one related to $M = 1000$) the choices $M = 100$ and $M = 1000$ yield to similar and reasonable precisions that outperform the average consensus in most cases.

We also observe a noticeable phenomenon when M is too small, e.g., $M = 10$. In this case in fact it might happen that the MSE remarkably increases with the number of communications. This is induced by the following: the low M induces estimate with high statistical variance. Thus it is likely that some agents will have some of their $\widehat{p}_b^{(i)}(t)$'s noticeably overestimated. This then acts as a disturb that eventually arrives, thanks to the max consensus, to the other peers, that see their estimate changing considerably after some time. A factor contributing to this is also the uniform initial distribution of the states $z^{(i)}$'s, which makes any random subset of agents a good representation of the entire network's distribution, hence yielding good early estimates.

• *Second experiment - fig. 6:* we now consider a single worst-case initial distribution of the states $z^{(i)}$, where the leftmost half of the agents in fig. 4 are in state 0 and the rightmost half are in state 1. Notice that this is actually not an unreasonable distribution, since for estimation applications in wireless sensor networks the communication topology and the measured physical quantities might be spatially correlated.

Since there is only one fixed initial state, the average consensus based estimator is deterministic and unique. The figure thus compares the confidence intervals of the max consensus estimators (depending upon the realization of the $x_{b,m}^{(i)}$'s) against the performance of the deterministic average consensus estimate.

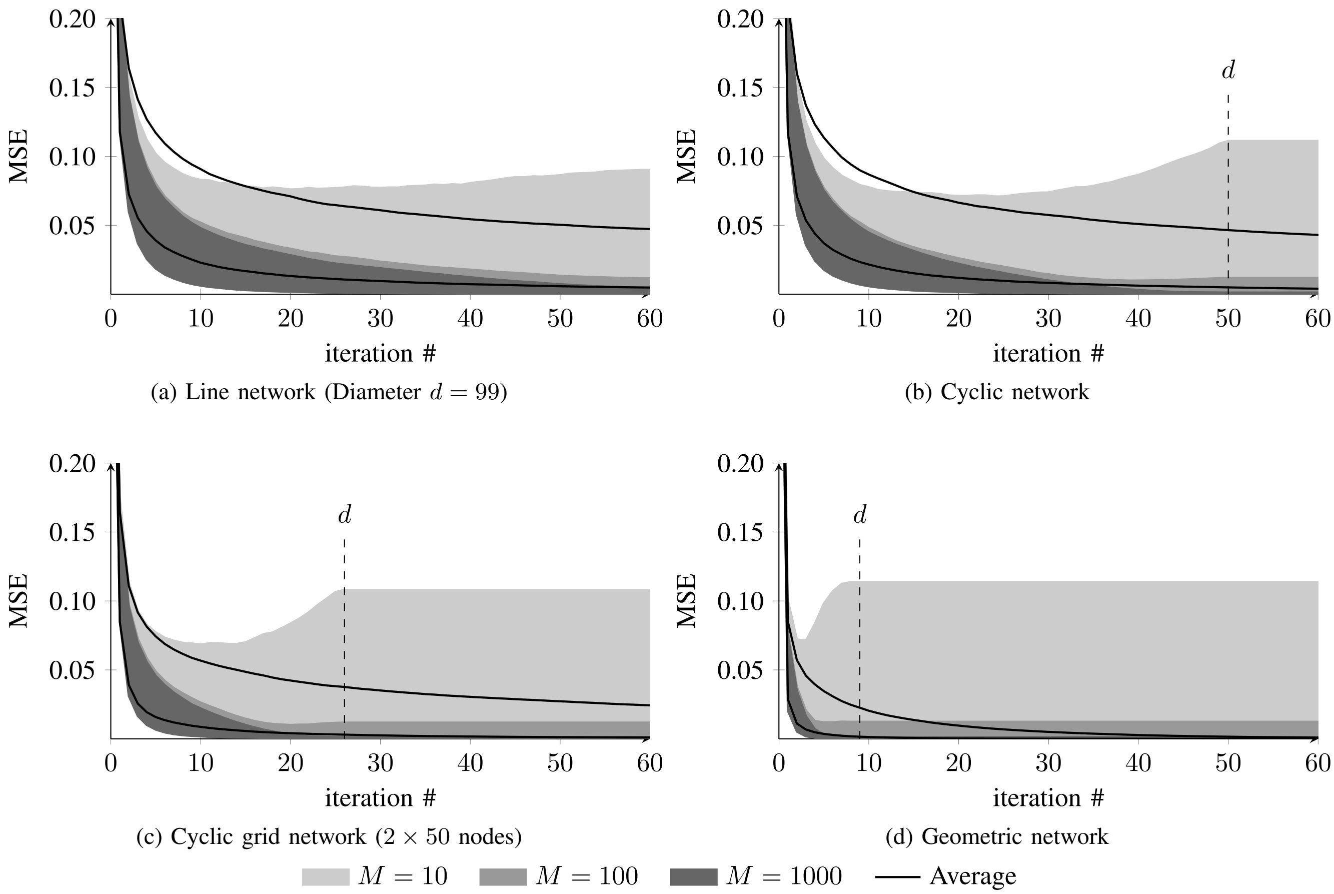


Fig. 5: Comparison of max-consensus based estimator against the average consensus based estimator. Each network consists of 100 nodes, and the network diameter d is marked in the figures. The shaded regions mark the 95% confidence interval for the max-consensus estimator, while the two solid lines mark the upper and lower end of the 95% confidence interval for the average-consensus estimator.

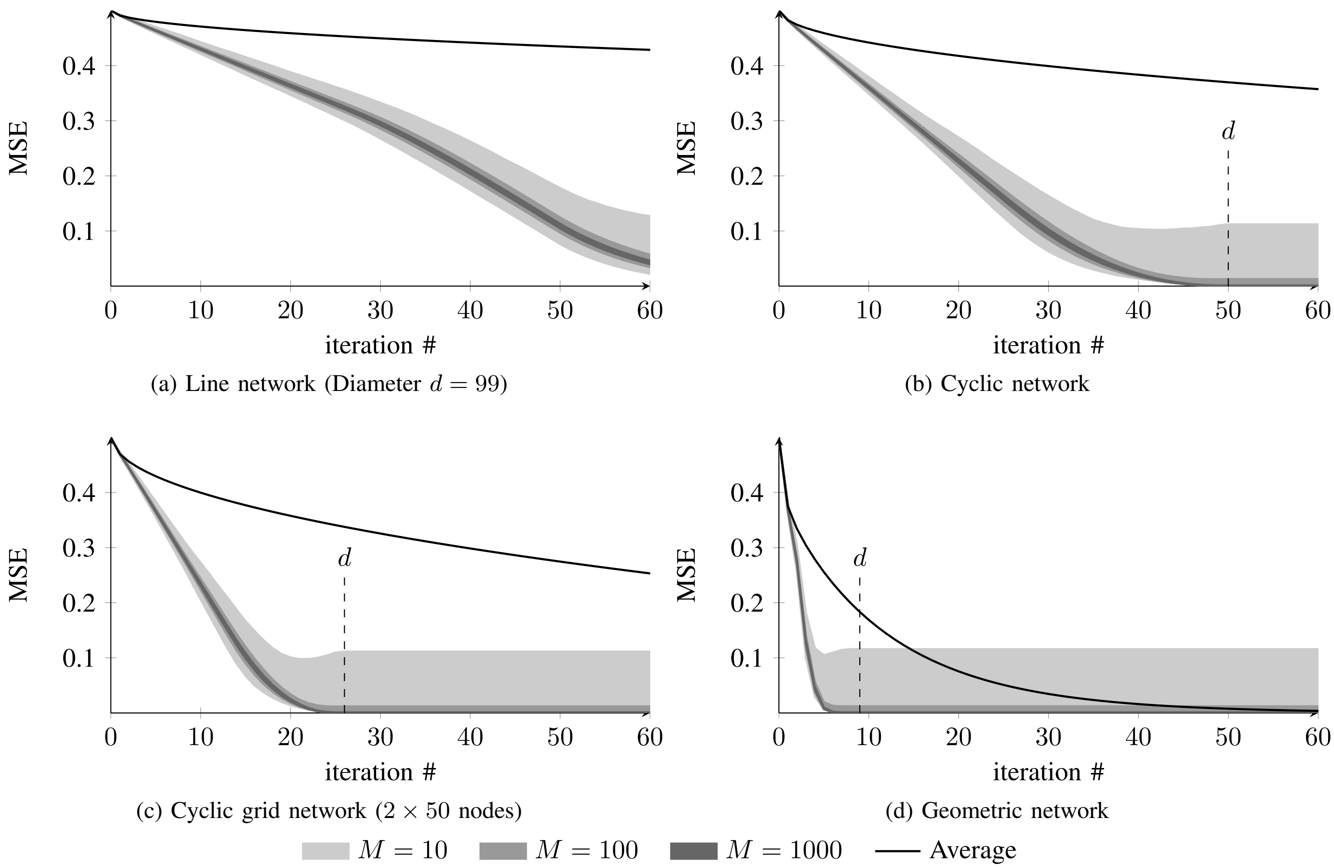
The outcome is then that for various networks the max consensus based estimator (10) can be much faster and more accurate than the average consensus counterpart (5), even for very small M's (even though a larger M improves the accuracy). The motivation is then that if the distribution of the states is not geographically homogeneous (assuming for simplicity communications links that follow Euclidean rules) then the max consensus is much more efficient to bring information of the existence of certain states to the other part of the network than the average.

VI. Concluding remarks and future directions

The two distributed estimators of PMFs over networks considered here, the one based on max consensus protocols and the one based on average consensus schemes, have several intrinsical differences. With the average, agents exchange messages containing only 1 scalar. With the max, instead, they exchange messages containing M scalars (with M a design parameter to be chosen by the user). With the average, convergences is (discarding quantization issues) asymptotical in time. With the max, instead, convergence is in finite time. With the average, the final estimate is (discarding communications failures issues) equal to the true value. With the max, the final estimate has a statistical precision that is direcly related to M.

The results shown here clearly indicate that there exists no uniformly better algorithm and no uniform rule to choose M: while in certain situations the average consensus strategy is the most reasonable strategy, in some others it is outperformed by the max consensus one. The rationale is induced by how the states of the peers are distributed across the network. If these are geographically clustered, then the max consensus scheme is preferrable because of its faster mixing properties, i.e., the capability of bringing information about the existence of other states across the network in a much faster way than the average one.

This work opens thus a variety of research directions. The first one is a more precise characterization of when the strategy performs better than the average-consensus one and of how to tune M, for certain fixed categories of graphs and communication protocols. An other one is on how to exploit the strategy to perform fast detection of changes in the aggregated network state. An other important ones is to associate the state with local topological properties, e.g., by setting it to be equal to the number of neighbors, and build



(a) Line network (Diameter $d = 99$)
(b) Cyclic network
(c) Cyclic grid network (2×50 nodes)
(d) Geometric network

Fig. 6: Comparison of max-consensus based estimator against the average consensus based estimator for a single worst-case initial condition. Each network consists of 100 nodes, and the initial state is determined by the agents spatial configuration. The shaded regions mark the 95% confidence interval for the max-consensus estimator, while the solid line mark the deterministic estimation for the average-consensus estimator.

on top of the proposed PMF estimators schemes that detect the most likely shape of the network.

References

[1] H. Jiang and S. Jin, "Scalable and Robust Aggregation Techniques for Extracting Statistical Information in Sensor Networks," in *IEEE International Conference on Distributed Computing Systems*, 2006.

[2] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2245–2253, Aug. 2003.

[3] P. A. . Forero, A. Cano, and G. B. . Giannakis, "Consensus-based distributed expectation-maximization algorithm for density estimation and classification using wireless sensor networks," in *Acoustics, Speech and Signal Processing*, Las Vegas, Nevada, 2008.

[4] N. Vlassis, Y. Sfakianakis, and W. Kowalczyk, "Gossip-based greedy Gaussian mixture learning," *Advances in Informatics*, vol. 3746, pp. 349–359, 2005.

[5] Y. Hu, J.-g. Lou, H. Chen, and J. Li, "Distributed density estimation using non-parametric statistics," in *Distributed Computing Systems*, no. 49, Toronto, Canada, 2007.

[6] M. Klusch, S. Lodi, and G. Moro, "Distributed clustering based on sampling local density estimates," in *International joint conference on Artificial intelligence*, 2003.

[7] X. Nguyen, M. Wainwright, and M. Jordan, "Nonparametric decentralized detection using kernel methods," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4053–4066, Nov. 2005.

[8] M. B. Greenwald and S. Khanna, "Power-conserving computation of order-statistics over sensor networks," in *ACM Symposium on Principles of Database Systems*, no. Pods. New York, New York, USA: ACM Press, 2004, pp. 275–285.

[9] N. Shrivastava, C. Buragohain, D. Agrawal, and S. Suri, "Medians and beyond: new aggregation techniques for sensor networks," in *International Conference on Embedded Networked Sensor Systems*, 2004, pp. 239–249.

[10] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a Tiny AGgregation service for ad-hoc sensor networks," *ACM SIGOPS Operating Systems Review*, vol. 36, no. SI, p. 131, Dec. 2002.

[11] S. Motegi, K. Yoshihara, and H. Horiuchi, "DAG based in-network aggregation for sensor network monitoring," in *International Symposium on Applications and the Internet*, 2006.

[12] M. Borges, P. Jesus, C. Baquero, and P. S. Almeida, "Spectra: Robust Estimation of Distribution Functions in Networks," *Distributed Applications and Interoperable Systems*, vol. 7272, pp. 96 – 103, 2012.

[13] J. Sacha, J. Napper, C. Stratan, and G. Pierre, "Adam2: Reliable distribution estimation in decentralised environments," in *Distributed Computing Systems*, Genova, Italy, 2010.

[14] M. Haridasan and R. V. Renesse, "Gossip-based distribution estimation in peer-to-peer networks," in *International Conference on Peer-to-Peer Systems*, 2008.

[15] S. Cheng, J. Li, Q. Ren, and L. Yu, "Bernoulli Sampling Based (e, d)-Approximate Aggregation in Large-Scale Sensor Networks," in *IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[16] L. Massouliè, E. L. Merrer, A.-M. Kermarrec, and A. Ganesh, "Peer counting and sampling in overlay networks: random walk methods," in *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, 2006, pp. 123–132.

[17] P. C. d. O. Jesus, "Robust Distributed Data Aggregation," Ph.D. Thesis, Universidade do Minho, 2011.

[18] D. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed statistical estimation of the number of nodes in Sensor Networks," in *IEEE*

Conference on Decision and Control, Atlanta, USA, Dec. 2010, pp. 1498–1503.

[19] J. C. S. Cardoso, C. Baquero, and P. S. Almeida, "Probabilistic Estimation of Network Size and Diameter," in *Fourth Latin-American Symposium on Dependable Computing*. João Pessoa, Brasil: IEEE, Sept. 2009, pp. 33–40.

[20] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. September, pp. 70–80, 2010.

[21] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, May 2008.

[22] T. Aysal, B. Oreshkin, and M. Coates, "Accelerated Distributed Average Consensus via Localized Node State Prediction," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1563–1576, Apr. 2009.

[23] Y. Yuan, G.-B. Stan, M. Barahona, L. Shi, and J. Goncalves, "Decentralised minimal-time consensus," in *IEEE Conference on Decision and Control and European Control Conference*, Dec. 2011, pp. 4282–4289.

[24] G. Casella and R. L. Berger, *Statistical inference*, 2nd ed. Duxbury Thomson Learning, 2001.

[25] C. Baquero, P. S. S. Almeida, R. Menezes, and P. Jesus, "Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 668 – 675, Apr. 2012.

[26] H. Terelius, D. Varagnolo, and K. H. Johansson, "Distributed size estimation of dynamic anonymous networks." in *IEEE Conference on Decision and Control*, 2012.

[27] J. Cichon, J. Lemiesz, W. Szpankowski, and M. Zawada, "Two-Phase Cardinality Estimation Protocols for Sensor Networks with Provable Precision," in *IEEE Wireless Communications and Networking Conference*, Paris, France, Apr. 2012.

[28] F. Garin, D. Varagnolo, and K. H. Johansson, "Distributed estimation of diameter, radius and eccentricities in anonymous networks," in *3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems*, 2012.

[29] H. A. A. David and H. N. N. Nagaraja, *Order Statistics*. Wiley series in Probability and Statistics, 2003.

[30] M. Ali, M. Pal, and J. Woo, "On the Ratio of Inverted Gamma Variates," *Austrian Journal of Statistics*, vol. 36, no. 2, pp. 153–159, 2007.

[31] V. Witkovsky, "Computing the distribution of a linear combination of inverted gamma variables," *Kybernetika*, vol. 37, no. 1, pp. 79–90, 2001.