A Simple Counting Estimator of Network Agents' Behaviors: Asymptotics

Stefano Marano
DIEM, University of Salerno, Fisciano (SA), Italy
E-mail: marano@unisa.it

Peter Willett
ECE, University of Connecticut, Storrs, CT, USA
E-mail: willett@engr.uconn.edu

***Abstract*—Recent works address the problem of estimating agents' behaviors in complex networks, of which social networks are a prominent example. Many of the proposed techniques work but at the cost of a substantial computational complexity, which is not permitted when dealing with big data real-time analysis. This raises the question of whether a very simple nonparametric counting estimator works in practical problems. We propose such an estimator and investigate its asymptotic properties for large number of agents N and/or for large network observation time T. The asymptotic optimality of the estimator is proven and computer experiments are provided to assess its performance for finite values of N and T.**

I. Introduction

Many real-world inference problems over complex networks can be abstracted to a general model in which there are N agents each taking some course of action, independently of each other, but depending upon the network status. The network status changes over time, which reflects in different agents' behaviors. A network analyzer observes all the agents' courses of actions for a certain interval of time T. For $T, N \gg 1$ the problem is to profile all the N agents, in the sense of estimating what course of action they are likely to take, under each possible state of the network.

The problem formulation addressed in this paper is essentially borrowed from [1], [2], where a maximum likelihood approach is pursued by means of the EM (Expectation Maximization) algorithm. In contrast to the computationally expensive iterative method of [1], [2], our focus here is in showing that a simple estimator based on occurrence counting provides good performance for values of T and N of practical interest, and is asymptotically optimum. The proposed estimator is also naturally suited for a time-sequential implementation.

Our work complements that in [3], where a parametric approach is pursued: conditional on the state of network, each agent is characterized by some probability distribution, which is completely known but for an unknown scalar parameter. In contrast, we assume minimum a-priori knowledge and develop a *nonparametric* framework –in particular we assume no knowledge of the underlying probability distributions. As typical in nonparametric estimation, the lack of a-priori knowledge is paid in the coin of the number of observations needed: accordingly, we pay special attention to the asymptotic properties (consistency) of the developed estimator, in the regimes of large N and/or T.

Depending on the practical application, sometimes the goal is to estimate some coarse characteristic of the agents, such as the agent's most likely choice, or the expected value of its outputs, and so forth. We refer to the challenging problem of estimating the whole agents' distributions, from which estimators of coarser agents' characteristics can be straightforwardly derived, by a plug-in approach that employs statistical functionals of the empirical distribution. In this sense, we are approaching the more general nonparametric inference problem.

The research line pursued in this paper follows our recent investigation started with [4]. In turn, the general context in which this work naturally lies is that of inference in distributed systems, which has represented an important portion of the signal processing research in the last decades. Some very partial references to this vast literature, biased by our personal interests, are as follows: [5]–[7] (distributed inference), [8], [9] (learning), [10], [11] (data fusion), [12], [13] (energy efficiency), [14]–[16] (censoring), [17], [18] (inference vs. communication), [19], [20] (cooperation/adaptation), [21], [22] (consensus).

The remainder of the paper is organized as follows. Section II formalizes the problem. The main theoretical results are given in Sect. III, with a sketch of the proofs provided in Appendix A. Numerical investigations are reported in Sect. IV, and a summary is given in Sect. V.

II. Problem Setting

Consider N independent agents, each taking one course of action at discrete time $t \in \mathcal{T} := \{1, \dots, T\}$, a time at which the network *status* is H_t –a random variable taking values on a finite set $\mathcal{H}$, say $\mathcal{H} = \{1, \dots, |\mathcal{H}|\}$. The status H_t of the network changes with t, irrespective of time, according to a certain distribution $\pi = [\pi_1, \dots, \pi_{|\mathcal{H}|}]$, where $\pi_h := \mathbb{P}(H_t = h)$ is constant in t and strictly positive. Let $X_{i,t}$ be the random variable representing the course of action of agent $i \in \mathcal{N} := \{1, \dots, N\}$ at time t, taking values on the finite set $\mathcal{X}$, say $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$. For $i \in \mathcal{N}$, $j \in \mathcal{H}$, $x \in \mathcal{X}$, $t \in \mathcal{T}$, let us introduce the conditional probability mass function (PMF) $p_{i,j} = [p_{i,j}(1), p_{i,j}(2), \dots p_{i,j}(|\mathcal{X}|)]$, whose entries are

$$p_{i,j}(x) := \mathbb{P}\left(X_{i,t} = x \mid H_t = j\right), \tag{1}$$

which are assumed strictly positive, and the corresponding unconditional PMF: $p_i := \mathbb{P}(X_{i,t} = x) = \sum_{j \in \mathcal{H}} p_{i,j} \pi_j$.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

The observation matrix made available to the network analyzer is:

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,T} \\ X_{2,1} & X_{2,2} & \dots & X_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N,1} & X_{N,2} & \dots & X_{N,T} \end{pmatrix}. \quad (2)$$

In the above each row corresponds to a single agent and each column refers to a time instant. Therefore, the i-th row represents the T actions taken by agent i at successive time instants, and the t-th column represents the action taken by the N agents in the single time instant t. The N entries over any column are statistically dependent, but conditionally independent given the state of the network characterizing that column. Entries belonging to different columns are independent, and entries over the i-th row are independent and identically distributed with distribution p_i. By observing the NT values in (2), we want to estimate the $N|\mathcal{H}|\,|\mathcal{X}|$ probabilities in (1), where these PMFs are assumed completely unknown, as is the a-priori distribution π of the network status. The only condition that the PMFs are known to obey is the following: For all $i \in \mathcal{N}$ and $j \in \mathcal{H}$,

$$\mathbb{P}\left(X_{i,t} \in \mathcal{X}_j | H_t = j\right) > \mathbb{P}\left(X_{i,t} \in \mathcal{X}_k | H_t = j\right), \quad k \neq j. \quad (3)$$

In the above $\{\mathcal{X}_i\}_{i=1}^{|\mathcal{H}|}$ is a partition of the set $\mathcal{X}$, and obviously $\mathbb{P}\left(X_{i,t} \in \mathcal{X}_j\right)$ means $\sum_{x\in\mathcal{X}_j} \mathbb{P}(X_{i,t} = x)$. The assumption says that when the state of network is j, the probability of the associated event $\mathcal{X}_j$ is larger than that of all other events $\mathcal{X}_k$, $k \neq j$. Throughout the paper, condition (3) is always assumed.

III. Counting Estimator and its Asymptotic Performance

Let $\mathcal{I}(A)$ be the indicator function of A, and let $C_t(j;N) := \sum_{i\in\mathcal{N}} \mathcal{I}(X_{i,t} \in \mathcal{X}_j)$ be the number of agents whose output $\in \mathcal{X}_j$ (loosely speaking, "choose j") at time t. We want to study the asymptotic performance of the following simple estimator:

$$\text{(step 1)} \quad \hat{H}_t := \{j \,:\, C_t(j;N) \geq C_t(m;N),\ \forall j \neq m\}, \quad (4)$$

$$\text{(step 2)} \quad \hat{P}_{i,j}(x) := \frac{\sum_{t\in\mathcal{T}} \mathcal{I}(X_{i,t} = x \bigcap \hat{H}_t = j)}{\sum_{t\in\mathcal{T}} \mathcal{I}(\hat{H}_t = j)}, \quad (5)$$

where in the first step, ties are resolved by choosing j uniformly at random among the candidates, and in the second step $\hat{P}_{i,j}(x) := 1/|\mathcal{X}|$ whenever the denominator of the RHS of (5) is zero. The first step is a counting over the columns of (2) and provides an estimate of the network status at time t; the second step is a counting over the rows and provides the empirical estimate of (1).

Let us introduce some notation. We denote by $\boldsymbol{X} = [X_{t,1}, \dots, X_{t,N}]$ the t-th column of matrix (2) (subscript t omitted), and by $\boldsymbol{X}_i$ be the vector $\boldsymbol{X}$ with the i-th entry omitted. Lowercase letters $\boldsymbol{x}$ and $\boldsymbol{x}_i$ denote their realizations. We let:

$$\begin{aligned} q_h(\boldsymbol{x}) &= \mathbb{P}\left(\boldsymbol{X} = \boldsymbol{x} \mid H_t = h\right), \\ q_h(\boldsymbol{x}_i) &= \mathbb{P}\left(\boldsymbol{X}_i = \boldsymbol{x}_i \mid H_t = h\right), \\ n_j &= \{\text{No. of occurrences of } \mathcal{X}_j \text{ in } \boldsymbol{x}\}, \\ m_{j,x} &= \{\text{No. of occurrences of } \mathcal{X}_j \text{ in } \boldsymbol{x}_i\} + \mathcal{I}(x \in \mathcal{X}_j), \\ \tilde{n}_x &= \{\text{No. of occurrences of } x \text{ in } \boldsymbol{x}\}, \\ \tilde{m}_x &= \{\text{No. of occurrences of } x \text{ in } \boldsymbol{x}_i\}. \end{aligned}$$

Finally, let us introduce a short-hand for the multinomial distribution:

$$\mathcal{M}ul\left(k_1, \dots, k_{|\mathcal{X}|}; M, p\right) := \frac{M!}{\prod_{x\in\mathcal{X}} k_x} \prod_{x\in\mathcal{X}} [p(x)]^{k_x}$$

where $k_1, \dots, k_{|\mathcal{X}|}$ are integers whose sum is M, and p is a PMF with alphabet $\mathcal{X}$.

The following theorems represent our main result. They are valid under assumption (3), and the convergences are in the almost sure (a.s.) sense.

Theorem 1: Fix N. For $x \in \mathcal{X}$, $i \in \mathcal{N}$, $j \in \mathcal{H}$, we have

$$\lim_{T\to\infty} \hat{P}_{i,j}(x) = \frac{\displaystyle\sum_{h\in\mathcal{H}} p_{i,h}(x)\,\pi_h \sum_{\boldsymbol{x}_i:\left\{\substack{m_{j,x}\geq m_{l,x} \\ j\neq l}\right\}} \frac{q_h(\boldsymbol{x}_i)}{\sum_{l\in\mathcal{H}} \delta_{m_{j,x},m_{l,x}}}}{\displaystyle\sum_{h\in\mathcal{H}} \pi_h \sum_{\boldsymbol{x}:\left\{\substack{n_j\geq n_l \\ j\neq l}\right\}} \frac{q_h(\boldsymbol{x})}{\sum_{l\in\mathcal{H}} \delta_{n_j,n_l}}}. \quad (6)$$

Also, in the special case where $p_{i,j}$ is constant with i, $\forall j \in \mathcal{H}$,

$$\begin{aligned} q_h(\boldsymbol{x}) &= \mathcal{M}ul\left(\tilde{n}_1, \dots, \tilde{n}_{|\mathcal{X}|}; N, p_{i,h}\right), \\ q_h(\boldsymbol{x}_i) &= \mathcal{M}ul\left(\tilde{m}_1, \dots, \tilde{m}_{|\mathcal{X}|}; N-1, p_{i,h}\right). \end{aligned} \quad (7)$$

□

Theorem 2: Fix T. For $x \in \mathcal{X}$, $i \in \mathcal{N}$, $j \in \mathcal{H}$, we have

$$\lim_{N\to\infty} \hat{P}_{i,j}(x) = \frac{\sum_{t\in\mathcal{T}} \mathcal{I}(X_{i,t} = x \bigcap H_t = j)}{\sum_{t\in\mathcal{T}} \mathcal{I}(H_t = j)}, \quad (8)$$

for $\sum_{t\in\mathcal{T}} \mathcal{I}(H_t = j) > 0$, and the limit is $1/|\mathcal{X}|$ otherwise. Also, for any integer $r > 0$,

$$\mathbb{E}\left[\left(\lim_{N\to\infty} \hat{P}_{i,j}(x)\right)^r\right] = \lim_{N\to\infty} \mathbb{E}\left[\left(\hat{P}_{i,j}(x)\right)^r\right], \quad (9)$$

and, in particular,

$$\begin{aligned} &\mathbb{E}\left[\lim_{N\to\infty} \hat{P}_{i,j}(x)\right] = p_{i,j}(x) + \left(\tfrac{1}{|\mathcal{X}|} - p_{i,j}(x)\right)(1-\pi_j)^T, \\ &\text{VAR}\left[\lim_{N\to\infty} \hat{P}_{i,j}(x)\right] = p_{i,j}(x)(1 - p_{i,j}(x)) \\ &\qquad\qquad \times \sum_{k=1}^{T} \tfrac{\left[(1-\pi_j)^{T-k} - (1-\pi_j)^T\right]}{k}. \end{aligned} \quad (10)$$

□

Theorem 3: The estimator $\hat{P}_{i,j}(x)$ in (4)-(5) is asymptotically consistent. Specifically, for $x \in \mathcal{X}$, $i \in \mathcal{N}$, $j \in \mathcal{H}$,

$$\lim_{N\to\infty}\lim_{T\to\infty} \hat{P}_{i,j}(x) = \lim_{T\to\infty}\lim_{N\to\infty} \hat{P}_{i,j}(x) = p_{i,j}(x).$$

□

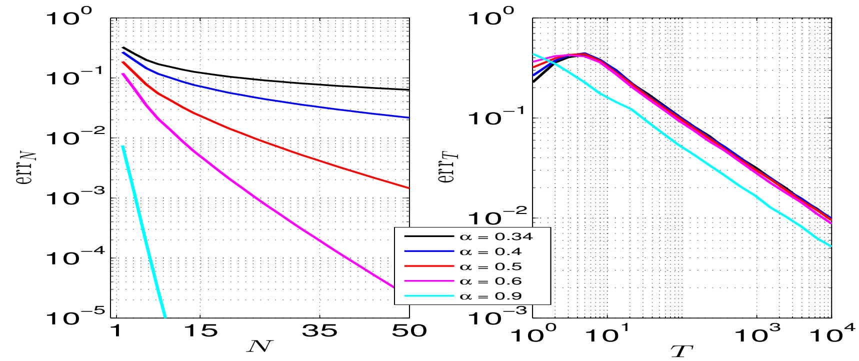

Fig. 1. *Left plot:* err_N versus the number of agents N. *Right plot:* err_T versus the time interval T. The curves are parametrized in α, see (11).

A. Discussion

Theorem 1 reveals that the limiting value of the estimator for $T \to \infty$ is deterministic: as T grows, the randomness is lost. The RHS of (6) depends on N and its computation becomes complicated for large N, because the expression involves counting the number of occurrences of certain events, over all the vectors of N (or $N-1$) entries. A simplification arises when all the agents share the same behavior, in the sense that, for all $j \in \mathcal{H}$, the PMFs $\{p_{i,j}\}$ are constant with the agent index $i \in \mathcal{N}$. In this case, the sums run not over all the vectors $\boldsymbol{x}$ and $\boldsymbol{x}_i$, but over vectors of indices $[\tilde{n}_1, \ldots, \tilde{n}_{|\mathcal{X}|}]$ and $[\tilde{m}_1, \ldots, \tilde{m}_{|\mathcal{X}|}]$. Clearly, the assumption of constant agents' behavior does not simplify the inference problem (the estimator is designed without using this information), but simplifies the presentation and interpretation of the results. This is why in the computer experiments of Sect. IV we make this assumption.

Consider now Theorem 2. We see that $\lim_{N\to\infty} \hat{P}_{i,j}(x)$ is a random variable and to gain some insight we refer to its expected value $\mathbb{E}[\lim_{N\to\infty} \hat{P}_{i,j}(x)]$. Suppose $|\mathcal{X}|\, p_{i,j}(x) \geq 1$. Inspection of the first equation in (10) reveals that such expected value is lower bounded by $1/|\mathcal{X}|$ and upper bounded by $p_{i,j}(x)$; thus, on the average, the limiting value underestimates the actual $p_{i,j}(x)$. The opposite is true when $|\mathcal{X}|\, p_{i,j}(x) \leq 1$, in which case we have $p_{i,j}(x) \leq \mathbb{E}[\lim_{N\to\infty} \hat{P}_{i,j}(x)] \leq 1/|\mathcal{X}|$. In both cases, as T grows the expected value $\mathbb{E}[\lim_{N\to\infty} \hat{P}_{i,j}(x)]$ converges monotonically to $p_{i,j}(x)$, with a rate of convergence that is ruled by $(1-\pi_j)^T$. The bias in the average value for small values of T is explained by the definition of the counting estimator that is equal to $1/|\mathcal{X}|$ when the denominator of (5) is zero, an event that happens with non negligible probability when T is small. It can be also shown that the variance of $\lim_{N\to\infty} \hat{P}_{i,j}(x)$ decreases with T, for large T.

Theorem 3 ensures that the deterministic number $\lim_{T\to\infty} \hat{P}_{i,j}(x)$ converges, as $N \to \infty$, to the correct value $p_{i,j}(x)$, and that the random variable $\lim_{N\to\infty} \hat{P}_{i,j}(x)$ also converges to $p_{i,j}(x)$ when $T \to \infty$. Thus, the simple counting estimator is strongly consistent: it converges almost surely to the true value $p_{i,j}(x)$ in the asymptotic regime where N and T go to infinity in an arbitrary order.

IV. Numerical Investigations

To simplify the analysis and the presentation of the results, in the computer experiments that follows we make the following assumptions.

- We set $\mathcal{X} = \mathcal{H}$ and $\mathcal{X}_i = \{i\}$, namely the agents' outputs are decisions about the network status. Note that agents are not biased: by assumption, $\mathbb{P}(X_{i,t} = j | H_t = j)$ is strictly larger than $\mathbb{P}(X_{i,t} = k | H_t = j)$, $k \neq j$.
- All agents have the same behavior: $\forall j \in \mathcal{H}$, we have $p_{i,j} = p_{k,j}$ for $i, k \in \mathcal{N}$. As already pointed out, this is not a simplification of the inference procedure, but only a way to simplify the presentation of the results.
- We assume that π is the uniform PMF: no network status is more likely than another.

Let us introduce two error terms:

$$\mathrm{err}_N = \sum_{j \in \mathcal{H}} \pi_j \max_{x \in \mathcal{X}} \left| \lim_{T\to\infty} \hat{P}_{i,j}(x) - p_{i,j}(x) \right|,$$

$$\mathrm{err}_T = \mathbb{E}\left[\sum_{j \in \mathcal{H}} \pi_j \max_{x \in \mathcal{X}} \left| \lim_{N\to\infty} \hat{P}_{i,j}(x) - p_{i,j}(x) \right| \right].$$

The former, a function of N, quantifies the difference between the true PMF $p_{i,j}$ and the limiting expression $\lim_{T\to\infty} \hat{P}_{i,j}(x)$ (see Theorem 1). The latter, a function of T, is a concise measure of the distance between the actual $p_{i,j}$ and $\lim_{N\to\infty} \hat{P}_{i,j}(x)$, computed in Theorem 2. The definition of err_T includes a statistical average, which is replaced by the arithmetic mean of the results of 10^3 Monte Carlo runs in the computer experiments.

Finally, we define

$$\mathrm{err}_{T,N} = \mathbb{E}\left[\sum_{j \in \mathcal{H}} \pi_j \max_{x \in \mathcal{X}} \left| \hat{P}_{i,j}(x) - p_{i,j}(x) \right| \right],$$

which provides a measure of how far is the estimated PMF $\hat{P}_{i,j}$ from the target $p_{i,j}$.

Consider the case $|\mathcal{H}| = |\mathcal{X}| = 3$, and the following three conditional PMFs characterizing the agents' behaviors:

$$\begin{aligned} p_{i,1} &= [\alpha,\ (1-\alpha)/2,\ (1-\alpha)/2], \\ p_{i,2} &= [(1-\alpha)/2,\ \alpha,\ (1-\alpha)/2], \\ p_{i,3} &= [(1-\alpha)/2,\ (1-\alpha)/2,\ \alpha]. \end{aligned} \qquad (11)$$

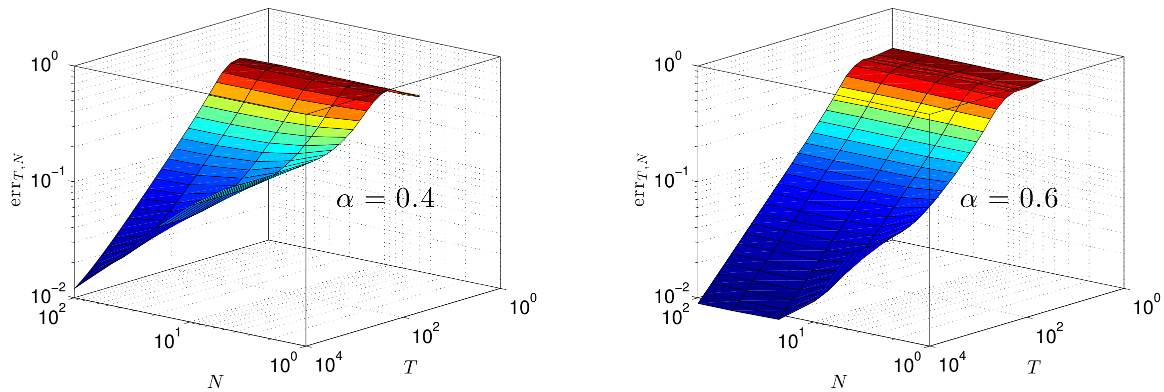

Fig. 2. The error $\mathrm{err}_{T,N}$ versus T and N for $\alpha = 0.4$ (left) and $\alpha = 0.6$ (right).

We assume $\alpha > 1/3$, in order to comply with the assumption of agents' unbiasedness.

The left plot in Fig. 1 shows the error err_N versus the number of agents N, for five different values of α. Similarly, the right plot in Fig. 1 shows the error err_T, versus T, for five different values of α.

Suppose T be sufficiently large and consider hence the left plot. For α close to $1/3$ there is a performance improvement by increasing N, but the error decreases very slowly: accurate agents' profiling is a challenging task. For larger values of α the agents' behavior under the different states of the network is markedly different, and the profiling task becomes simpler. Indeed, with very spiky PMFs, as those corresponding to $\alpha = 0.9$, the error decreases quite fast with increasing N.

Suppose now N be large so that the asymptotic regime of $N \to \infty$ comes into play. The right plot of Fig. 1 reveals that there is little difference in the error for different values of α, and only for $\alpha = 0.9$ the error is appreciably lower. The lesson learned by the combined analysis of the plots in Fig. 1 is that neither regime, $N \to \infty$ or $T \to \infty$, can be considered better performing than the other, and a case by case analysis is needed.

Let us now analyze the estimation performance when both N and T are finite. In Fig. 2, we use 10^3 Monte Carlo runs to compute different realizations of the estimator $\hat{P}_{i,j}(x)$, which are then averaged to compute $\mathrm{err}_{T,N}$. These values are shown as functions of both T and N, for two different values of α. As one expects, the estimation performance improves for larger values of α. This figure allows the network analyzer to set the proper values of T and N (if the number of agents can be controlled by the analyzer) to obtain a desired performance.

While it is obvious that the numbers shown in this section are valid only for the simple example (11), such example well highlights the general behavior of the estimator with respect to the main system parameters T, N, and to the smoothness or spiky character of the underlying PMFs.

Finally, we reiterate that the problem model addressed in this paper –originally inspired by [1], [2] and therein references– is similar to that of [3] with the important difference that the latter adopts a parametric model. For this reason a direct comparison with [3] in terms of estimation performance is inappropriate, but it can be certainly said that our counting approach is computationally simpler than the iterative procedures considered in [1]–[3].

V. Summary

Nowadays, there is a growing interest in classifying the users of social networks for commercial, security, or system optimization purposes. We assume that N agents of a network take some course of action (e.g., make choices) independently of each other, but depending on the status of the system at that time. By observing for a period T the N agents' outputs, a network analyzer is asked to profile all the agents, meaning that the probability of making any possible choice under any possible network status should be estimated. The problem is challenging when no a-priori knowledge of the status of the network nor of the agents' inclinations is available and, perhaps even more important, when the network analyzer operates under a real-time "big data" paradigm, with severe constraints on the computational complexity of the estimation procedure. A desirable feature, also, is the possibility of implementing the estimation algorithm in a sequential fashion.

In these situations, the design choice is biased towards very simple nonparametric estimators. Thus, our main motivation was to investigate the performance of a simple counting estimator, and the main results of our analysis are the proof of the asymptotic consistence of the estimator when both N and T diverges, and the derivation of analytical expressions for the two regimes of $N \to \infty$ and $T \to \infty$. Computer simulations are presented for performance assessments in the finite regime.

Appendix A
Sketch of the Proofs

The complete proofs are given in an extended journal version of this work [23]. Here we only provide the main arguments. The limits of random quantities are in the almost sure sense.

Proof of Theorem 1

For $T \to \infty$, it is not difficult to see that $\sum_{t \in \mathcal{T}} \mathcal{I}(\hat{H}_t = j) \to \infty, \forall j$, and the strong law of large numbers then implies

$$\hat{P}_{i,j}(x) \to \frac{\mathbb{P}(X_{i,t} = x \bigcap \hat{H}_t = j)}{\mathbb{P}(\hat{H}_t = j)}. \quad \text{(A.12)}$$

The denominator is obtained by writing $\mathbb{P}(\hat{H}_t = j) = \sum_h \pi_h \mathbb{P}(\hat{H}_t = j \mid H_t = h)$ and exploiting the definition of $\hat{H}_t$ in (4), which yields

$$\mathbb{P}(\hat{H}_t = j) = \sum_{h \in \mathcal{H}} \pi_h \sum_{\boldsymbol{x} : \left\{ \substack{n_j \geq n_l \\ j \neq l} \right\}} \frac{q_h(\boldsymbol{x})}{\sum_{l \in \mathcal{H}} \delta_{n_j, n_l}}. \quad \text{(A.13)}$$

Using the chain rule for probabilities the numerator of (A.12) can be rewritten as $\sum_{h \in \mathcal{H}} \mathbb{P}(\hat{H}_t = j | X_{i,t} = x, H_t = h) p_{i,h}(x) \pi_h$. Now, conditioning to $\{H_t = h\}$ makes the random variables $\{X_{k,t}\}_{k \in N}$ independent of each other, and conditioning to $\{X_{i,t} = x\}$ allows to count the occurrences of the different sets $\mathcal{X}_j$ only for the remaining $N-1$ variables $\{X_{k,t}\}_{k \in N \setminus i}$ in the vector $\boldsymbol{x}_i$. This yields $\mathbb{P}(\hat{H}_t = j | X_{i,t} = x, H_t = h)$ in the form

$$\sum_{\boldsymbol{x}_i : \left\{ \substack{m_{j,x} \geq m_{l,x} \\ j \neq l} \right\}} \frac{q_h(\boldsymbol{x}_i)}{\sum_{l \in \mathcal{H}} \delta_{m_{j,x}, m_{l,x}}},$$

and (6) follows. Expressions (7) are obvious.

Proof of Theorem 2

The proof of the first claim is straightforward. The RHS of (8) is nothing but the RHS of (5) with $\hat{H}_t$ replaced by H_t, and we only need to show that the former can be replaced by the latter, which intuitively follows by observing that $N \to \infty$ allows for exact estimation of the network status for any time instant t. Then, since $0 \leq \hat{P}_{i,j}(x) \leq 1$ a.s., these random variables are bounded, and a.s. convergence implies convergence of the moments [24], yielding (9). The computation of the average value and of the variance in (10) is omitted.

Proof of Theorem 3

It is not difficult to show that $\hat{P}_{i,j}(x)$ converges to $\mathbb{P}(X_{i,t} = x \mid \hat{H}_t = j)$ when $T \to \infty$, yielding,

$$\begin{aligned}
&\lim_{N \to \infty} \lim_{T \to \infty} \hat{P}_{i,j}(x) \\
&= \lim_{N \to \infty} \mathbb{P}\left(X_{i,t} = x \mid \hat{H}_t = j \right) \\
&= \lim_{N \to \infty} \frac{\sum_{h \in \mathcal{H}} \mathbb{P}\left(\hat{H}_t = j \mid X_{i,t} = x, H_t = h \right) p_{i,h}(x)\, \pi_h}{\sum_{h \in \mathcal{H}} \mathbb{P}\left(\hat{H}_t = j \mid H_t = h \right) \pi_h}
\end{aligned}$$

where the last equality follows by an application of the chain rule for probabilities. Now conditioning on both the events $\{H_t = h\}$ and $\{X_{i,t} = x\}$, one can easily prove that $\hat{H}_t \to h$ when $N \to \infty$, which implies $\lim_{N \to \infty} \mathbb{P}(\hat{H}_t = j \mid X_{i,t} = x, H_t = h) = \delta_{j,h}$. The same result holds for the denominator of the above limit. Plugging these results into the above expression gives $\lim_{N \to \infty} \lim_{T \to \infty} \hat{P}_{i,j}(x) = p_{i,j}(x)$. The limit the other way around follows by letting $T \to \infty$ in (8) and observing that $\sum_{t \in \mathcal{T}} \mathcal{I}(H_t = j) \to \infty$.

REFERENCES

[1] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1026–1037, Jun. 2013.

[2] D. Wang, L. Kaplan, and T. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Trans. on Sensor Networks*, vol. 10, no. 2, Jan. 2014.

[3] S. Marano, V. Matta, and P. Willett, "The importance of being earnest: Social sensing with unknown agent quality," *IEEE Trans. Signal and Info. Process. over Networks*, in print, DOI: 10.1109/TSIPN.2016.2555799.

[4] ——, "Sensor network tomography: The revenge of the detected," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4329–4338, Aug. 2015.

[5] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I – fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[6] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.

[7] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Signal Processing*, H. V. Poor and J. B. Thomas, Eds. JAI Press, 1993, pp. 297–344.

[8] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 56–69, Jul. 2006.

[9] S. Marano, V. Matta, and P. Willett, "Nearest-neighbor distributed learning by ordered transmissions," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5217–5230, Nov. 2013.

[10] P. K. Varshney, *Distributed Detection and Data Fusion.* New York, NY: Springer, 1997.

[11] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized MMSE estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5485–5496, Nov. 2007.

[12] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3229–3235, Jul. 2008.

[13] P. Braca, S. Marano, and V. Matta, "Single-transmission distributed detection via order statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2042–2048, Apr. 2012.

[14] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[15] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5497–5505, Nov. 2007.

[16] S. Marano, V. Matta, and P. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1976–1986, 2009.

[17] B. Chen, L. Tong, and P. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.

[18] S. Marano, V. Matta, P. Willett, and L. Tong, "Cross-layer design of sequential detectors in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4105–4117, Nov. 2006.

[19] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[20] P. Braca, S. Marano, V. Matta, and A. H. Sayed, "Asymptotic performance of adaptive distributed detection over networks," *IEEE Trans. Inf. Theory*, submitted, available at arXiv:1401.5742v2 [cs.IT].

[21] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[22] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.

[23] S. Marano and P. Willett, "Profiling Agents in Networks by Simply Knowing How to Count," *submitted*, 2016.

[24] H. Shao, *Mathematical Statistics*, 2nd ed. Springer, 2003.