ONE PLUS TWO MAY NOT EQUAL TWO PLUS ONE IN A SOCIAL SENSING NETWORK WITH UNKNOWN PARAMETERS

Stefano Marano, Vincenzo Matta

DIEM

University of Salerno via Giovanni Paolo II, I-84084, Fisciano (SA)

ABSTRACT

Parametric estimation for the generative social sensing model proposed in [19, 20] is addressed. First, we provide a detailed analysis of the estimation performance bounds, in terms of the Fisher information matrix, with emphasis on the fundamental scaling laws as the number of network agents and/or the number of monitored agents' activities is large. Then, we examine two viable estimation procedures that can be useful even in such large dataset applications: the Expectation-Maximization and the Fisher scoring algorithms, which both achieve the aforementioned performance bounds.

Index Terms— Social sensing, ML, Fisher information.

1. INTRODUCTION

Massive amounts of informations are being collected from across multiple data networks, for several motivations, including: commercial (e.g., advertising and marketing), security (e.g., detection of suspicious activities) or safety (e.g., occurrence of critical phenomena for hazard management) issues. The network agents' activity can be summarized in terms of different kinds of observations. There are "hard" data, such as binary votes; more structured forms of data, such as preferences or opinions; and "soft" data, such as environmental parameters of some phenomenon of interest. Making inference about the specific attributes characterizing the network agents is a critical task to perform agents' profiling in view of the aforementioned purposes. Two major needs emerge in this inferential setting: i) designing algorithms able to manage the vast amount of data arising from the sensing operations, and *ii*) finding the performance limits (estimation accuracy) and scaling laws (with the number of agents, with the number of activities, ...), to judge the goodness of a given algorithm.

Such fundamental questions have been addressed in the topical literature. A rough, though useful, categorization of the available results identifies two main research routes. The first one broadly refers to data dimensionality reduction techniques based on convex optimization — see, e.g., [12, 16]. Here, the bulk of data is typically represented by a "big" data matrix [3], Peter Willett*

ECE Department, U-2157 University of Connecticut Storrs, CT 06269 USA

which encode the agents' characteristics in some suitable structure. The inferential task consists of disentangling the components representative of different agents' behaviors, subject to a data-best-fit constraint.

In contrast, the second research route refers to statistical methods based on parametric/nonparametric modeling of users' data distributions. Here, the analyst first builds a statistical model for the generative mechanism of the agents' data, and consequently designs inference algorithms tailored to the chosen statistical model. Such methods include, to mention a few: graph-based models [17, 21], community detection [6], belief diffusion analysis in social networks [5], distributed strategies for multi-agent inference over social networks via diffusion adaptation [22], and social sensing problems where sensors are embedded in a social networks [1, 19, 20].

Especially relevant to our work is the generative social sensing model proposed in [19,20]: there are n agents, performing ttasks. During each task, each agent reports a claim among several possible claims. There is only one *true* claim, which is a latent, i.e., unobserved, variable. In dealing with *human agents*, *trustworthiness issues arise: conflicting observations about the same phenomenon might be reported, and it is critical to profile the agents and ascertain their credibility.* We shall start by exploiting and extending the aforementioned model in order to accommodate more general social data models, as detailed in the forthcoming section. This article is a focused version of [15].

Notation. Random objects are denoted by uppercase letters, their realizations by lowercase letters. Boldface fonts are used for vectors, while boldface *and* calligraphic fonts for matrices. The ℓ -th entry of a vector \boldsymbol{x} is denoted by x_{ℓ} , the (i, j)-th entry of a matrix $\boldsymbol{\mathcal{X}}$ is denoted by $[\boldsymbol{\mathcal{X}}]_{ij}$.

2. MODEL

The activities of n agents are monitored by a network analyst. The agents focus on a random phenomenon H belonging to a discrete, finite alphabet \mathcal{H} , with probability mass function (pmf) π_h . Several instances (*tasks*) of the phenomenon of interest are observed. During the *m*-th task, $m = 1, \ldots, t$, the ℓ -th agent reacts to a particular state of the underlying phenomenon by producing an observation $x_\ell(m)$. The response of the ℓ -th agent is governed not only by the state h, but also

^{*}P. Willett was supported by the NPS via grant ONR N00244–14–1–0033.

by some personal agent attribute, represented by a real-valued parameter θ_i lying in an open set Θ . Given a state h, the *conditional* probability function of an agent output x, corresponding to an attribute θ , will be denoted by $p_h(x;\theta)$. Conditionally on the state of nature, the agents' responses are independent. For each task m, data are collected into the $n \times 1$ vectors $\boldsymbol{x}(m) = [x_1(m); \ldots; x_n(m)]$, which are assumed independent and identically distributed (i.i.d.). The goal of the network analyst is estimating the parameter vector $\boldsymbol{\theta} = [\theta_1; \ldots; \theta_n]$, based on the available data $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t)$. The probability function of the data vector \boldsymbol{x} collected in a single task is accordingly:

$$p(\boldsymbol{x};\boldsymbol{\theta}) \triangleq \sum_{h \in \mathcal{H}} \pi_h \prod_{\ell=1}^n p_h(x_\ell;\theta_\ell), \qquad (1)$$

and the joint probability function of the data is then given by: $p(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t); \boldsymbol{\theta}) \triangleq \prod_{m=1}^{t} p(\boldsymbol{x}(m); \boldsymbol{\theta})$. Special attention should be paid to the situation where the "dimensionality" of the problem is large, in terms of cardinality of data and/or parameter sets. In order to provide the guidelines for managing huge amounts of information, it is critical to understand how many data are needed to reach a target performance level, how the estimation error scales with the number of observed tasks t, and how beneficial is increasing the number of agents n. Our novel contribution is twofold: i) we provide rigorous results about the best asymptotic (Fisher information) performance and fundamental scaling laws with respect to the number of agents and tasks, and ii) we provide practical solutions to the estimation problem, which are affordable and deliver the promised best performance.

3. PERFORMANCE LIMITS

In order to capture the performance limits of the introduced social sensing model, here we provide a rigorous asymptotic characterization in terms of the Fisher Information Matrix (FIM), see, e.g., [13]. Let x denote the $n \times 1$ data vector collected during a single task. The score vector is:

$$\nabla \ln p(\boldsymbol{x}; \boldsymbol{\theta}) = \left[\frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \theta_1}; \dots; \frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \theta_n}\right].$$
 (2)

The corresponding *per-single-task* FIM will be denoted by $\mathcal{F}(\boldsymbol{\theta})$, the $n \times n$ matrix whose (i, j)-th entry is given by [13]:

$$[\mathbf{\mathcal{F}}(\boldsymbol{\theta})]_{ij} = \mathbb{E}\left[\frac{\partial \ln p(\boldsymbol{X};\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\boldsymbol{X};\boldsymbol{\theta})}{\partial \theta_j}\right].$$
 (3)

Now, since the observation vectors $\boldsymbol{x}(m)$ collected by the network analyst are i.i.d. across different tasks, the overall FIM is simply $t \mathcal{F}(\boldsymbol{\theta})$ [13]. In order to get useful insights, we now focus on three extreme cases, namely, *i*) the situation where only one¹ agent is monitored, *ii*) the performance of a *clairvoyant* system that knows the *true* state of nature *H*, and *iii*) the case where the number of agents *n* is large. For the case n = 1, we have $p(x; \theta) = \sum_{h \in \mathcal{H}} \pi_h p_h(x; \theta)$, so that the (scalar) Fisher information is:

$$\mathcal{F}(\theta) = \mathbb{E}\left[\left(\frac{\sum_{h \in \mathcal{H}} \pi_h \frac{\partial p_h(X;\theta)}{\partial \theta}}{\sum_{h \in \mathcal{H}} \pi_h p_h(X;\theta)}\right)^2\right]$$
(4)

The clairvoyant FIM will be instead denoted by $\mathcal{F}^{\star}(\theta)$. Since the clairvoyant system observes H, we refer to the probability function of the *pair* (\mathbf{X}, H):

$$p(\boldsymbol{x},h;\boldsymbol{\theta}) \triangleq \pi_h \prod_{\ell=1}^n p_h(x_\ell;\theta_\ell),$$
(5)

which shows that the problem decouples across the agents, due to the conditional independence of the observations given H. Now, for $i \neq j$ we get $[\mathcal{F}^{\star}(\theta)]_{ij} = 0$, which follows from the conditional independence of X_i and X_j given H and from the fact that the score vector has zero mean [18]. As to the terms on the main diagonal, we have $[\mathcal{F}^{\star}(\theta)]_{ii} \triangleq \mathcal{F}^{\star}(\theta_i)$, where the scalar function $\mathcal{F}^{\star}(\theta)$ is defined by:

$$\left| \mathcal{F}^{\star}(\theta) \triangleq \sum_{h \in \mathcal{H}} \pi_h \mathbb{E} \left[\left(\frac{\partial \ln p_h(X;\theta)}{\partial \theta} \right)^2 \middle| H = h \right]$$
(6)

which is the average, over the pmf π_h , of the (scalar) Fisher informations corresponding to each state of nature.

Finally, we consider the case of a large number of agents. Let us preliminarily introduce the Kullback-Leibler (KL) divergences [7] between a pair of distinct states of nature h, k:

$$\mathcal{D}_{kh}(\theta) \triangleq \mathbb{E}\left[\ln \frac{p_k(X;\theta)}{p_h(X;\theta)} \middle| H = k\right].$$
 (7)

The following theorem characterizes the behavior of the FIM for a large number of agents n. The proof is omitted for space limitations, and is reported elsewhere [15].

THEOREM 1 (many-agents FIM). Let $\theta_1, \theta_2, ...$ be an infinite sequence of values in the set Θ , and consider the sequence (as n increases) of estimation problems for the parameter vectors $\boldsymbol{\theta} = [\theta_1; ...; \theta_n]$. Assume that, for all $h \in \mathcal{H}$, and $\theta \in \Theta$, one has $\left| \frac{\partial \ln p_h(x;\theta)}{\partial \theta} \right| \leq s(x; \theta)$, where $s(X; \theta)$ and $s^2(X; \theta)$ have finite expectation, and that:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^{n} \mathcal{D}_{kh}(\theta_{\ell}) = \bar{\mathcal{D}}_{kh} > 0, \quad \forall h, k \in \mathcal{H}, \quad h \neq k.$$
(8)

Then, for $i, j \in \mathbb{N}$ held fixed, the FIM performance approaches the clairvoyant system in the following sense:

$$\lim_{n \to \infty} [\mathcal{F}(\boldsymbol{\theta})]_{ii} = [\mathcal{F}^{\star}(\boldsymbol{\theta})]_{ii} = \mathcal{F}^{\star}(\theta_i)$$
(9)

$$\lim_{n \to \infty} [\mathcal{F}(\boldsymbol{\theta})]_{ij} = [\mathcal{F}^{\star}(\boldsymbol{\theta})]_{ij} = 0 \quad \text{for } i \neq j$$
(10)

¹Throughout the work, we shall assume that the estimation problem is not singular (i.e., is *identifiable*) at a single-agent level [18].

REMARK I. For the asymptotic regime considered in Theorem 1, the size *n* of the parameter vector $\boldsymbol{\theta}$ goes to infinity. In order to get meaningful asymptotic results about the estimation of $\boldsymbol{\theta}$, we have imposed to the patterns $\theta_1, \theta_2, \ldots$ a certain degree of asymptotic regularity through condition (8). Thus, it can be useful to describe some physical situations that correspond to such a condition. One is the extremely "regular" case where $\theta_i = \theta$ for all *i*. Another one is the extremely "irregular" case where the entries of θ_i look like realizations of i.i.d. random variables.

REMARK II. Due to the diagonal structure exhibited in both the n = 1 and the $n \gg 1$ cases, the limiting Mean-Square-Error (MSE) performance can be captured in a simple way by examining the two *scalar* functions $\mathcal{F}(\theta)$ and $\mathcal{F}^*(\theta)$. Approximating the asymptotic (large number of tasks t) MSE of the single component θ_i as $(1/t)[\mathcal{F}(\theta)^{-1}]_{ii}$, we have [13]:

$$MSE \approx \frac{1}{t \mathcal{F}(\theta_1)}, \text{ [single agent, see (4)]}$$
(11)

$$MSE \approx \frac{1}{t \mathcal{F}^{\star}(\theta_i)}, \text{ [many agents, see (6)]}$$
(12)

4. ESTIMATOR DESIGN

Our goal is designing a practical estimator that keeps the promises of limiting-performance analysis. We start by introducing the definition of asymptotic efficiency: an estimator $\hat{\theta}(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t))$ is asymptotically efficient if [18]:

$$\sqrt{t} \left[\hat{\boldsymbol{\theta}}(\boldsymbol{X}(1), \dots, \boldsymbol{X}(t)) - \boldsymbol{\theta} \right] \stackrel{t \to \infty}{\rightsquigarrow} \mathcal{N} \left(0, \boldsymbol{\mathcal{F}}(\boldsymbol{\theta})^{-1} \right)$$
(13)

where \rightsquigarrow denotes convergence in distribution, and $\mathcal{N}(\mu, \mathfrak{C})$ denotes a Gaussian distribution with mean μ and covariance matrix C. Under classical regularity conditions, the ML estimator is asymptotically efficient [18]. In order to evaluate the ML, we need maximizing the likelihood over an n-dimensional parameter space. Since the dimension of the parameter space is dictated by the number of agents, in practical applications a brute-force grid-search over the parameter space is seldom advisable. As already noted in [19,20], one possibility to produce ML estimates is to implement the celebrated Expectation-Maximization (EM) algorithm [8], which naturally fits the latent-variable structure of the model, and basically amounts to an iterative optimization algorithm alternating between i) estimation of the underlying hypothesis given the current (estimated) vector parameter, and ii) estimation of the vector parameter given the current (estimated) underlying hypothesis.

In this work we exploit the peculiarities of the considered social sensing model to construct a different estimator based on the so-called Fisher scoring [18]. Such estimator will exhibit interesting features, particularly useful for large dataset applications. We start by observing that the statistics of the data the *i*-th agent depend only on the corresponding attribute θ_i : therefore, the network analyst can estimate the parameter θ_i from the data pertaining to agent *i* alone, for instance, by using the

single-agent ML estimator:

$$\tilde{\theta}(x_i(1), \dots, x_i(t)) \triangleq \arg\max_{\theta_i \in \Theta} p(x_i(1), \dots, x_i(t); \theta_i), \quad (14)$$

which will achieve the single-agent asymptotic efficiency [18]. Let us examine the good and the bad properties of the above estimator. The estimator improves its performance as t increases, i.e., as more data are collected by the network analyst, and its asymptotic MSE scales as $1/(t \mathcal{F}(\theta_i))$, see [18]. However, this is not necessarily the best that can be done starting from the whole dataset that aggregates information from all agents. Indeed, the single-agent estimator does not take advantage of the joint dependence across agents that is induced by the commonly monitored phenomenon. Nonetheless, we can leverage the good asymptotic properties of the single-agent estimate to obtain a (globally) asymptotically efficient estimator that admits a convenient analytical form. This result can be achieved by using an estimation procedure known as one-step ML, or Fisher scoring method [18]. We have the following result, whose proof is omitted for space limitations, and is reported elsewhere [15].

THEOREM 2 (proposed estimator). Let $\hat{\theta}$ be the $n \times 1$ vector collecting all the single-agent ML estimates, and introduce the Fisher scoring estimator:

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} + \frac{\boldsymbol{\mathcal{F}}(\tilde{\boldsymbol{\theta}})^{-1}}{t} \sum_{m=1}^{t} \nabla \ln p(\boldsymbol{x}(m); \tilde{\boldsymbol{\theta}})$$
(15)

Then, $\hat{\theta}$ is asymptotically efficient. Moreover, in the largenetwork scenario, the estimator takes on the simple form:

$$\left| \hat{\theta}_{i} \stackrel{\text{large } n}{\approx} \tilde{\theta}_{i} + \frac{1}{t \, \mathcal{F}^{\star}(\tilde{\theta}_{i})} \sum_{m=1}^{t} \left. \frac{\partial \ln p(\boldsymbol{x}(m); \boldsymbol{\theta})}{\partial \theta_{i}} \right|_{\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}} \right|$$
(16)

We observe from (15) that, once the single-agent estimate has been obtained, the Fisher scoring estimator does not rely on any optimization routine. It admits a closed-form representation, additive w.r.t. the task index, and requires an $n \times n$ matrix computation and inversion, affordable up to network sizes $n \sim 10^3$. Remarkably, for the scenario of greatest interest where n is large, the Fisher scoring estimator is cast into the even more appealing form in (16): here also the complexity of an $n \times n$ matrix computation and inversion has totally disappeared. We stress that n in the order of 10 is typically enough for the clairvoyant approximation to be accurate, as we shall see in the forthcoming section. Such properties make the proposed estimator a good candidate even for distributed implementations, which are desirable in large dataset applications. We conclude that the additive correction term present in (15) is able to capture automatically the relevant dependence across agents in a compact and neat way, by simultaneously embodying the latent-variable detection and θ -estimation steps. In constrast, for small number of tasks, the Fisher scoring method is not expected to outperform the EM, provided that local maxima are not a concern for the EM, and that the EM converges toward the ML reasonably fast.



Fig. 1. Network of decision makers, example of Sect. 5.1.

5. EXAMPLES

Due to space constraints, we here skip all technical details (reported in [15]), and we prefer to offer a succinct report of some numerical experiments, aimed at elucidating how the carried analysis works in two relevant applicative scenarios.

5.1. Network of Decision Makers

Agents are tasked to perform a Gaussian shift-in-mean detection [14], with signal-to-noise ratio snr, about a binary state of nature H. The *i*-th agent compare its test statistic to a certain threshold θ_i . The analyst estimates the decision thresholds, in order to infer the agents' "way of thinking", e.g., to locate low-quality and/or anomalous agents. In Fig. 1, panel (a), we check Theorem 1: we display the asymptotic estimation error $[\mathcal{F}(\theta)^{-1}]_{11}$ corresponding to θ_1 (the other entries of θ are randomly generated), for estimation problems with increasing dimensions, and for several pairs $(\theta_1, \operatorname{snr})$. We see that the performance improves as n increases, and approaches the clairvoyant system. The error lies between the single-agent and the many-agents scalar performance figures $\mathcal{F}(\theta)$ and $\mathcal{F}^{\star}(\theta)$, which, therefore, offer a good summary of the overall inference system. In panel (b), we check Theorem 2: we display the empirical (Monte Carlo) as well as the theoretical (FIM trace) MSE performance, of EM, of the simplest Fisher scoring estimator in (16), and of the single-agent case. All the algorithms reach the pertinent asymptotic bounds: they are effective in delivering the best performance reasonably soon. Panel (c) illustrates how the algorithms reveal the hidden "nature" of the agents. Hot (red) circles denote a low-credibility, while cold (blue) circles denote a high-credibility. The red dashed circle surrounds low-quality agents that are in error more than 20% of times. We see that our algorithm is effective in unveiling the unreliable agents: the analyst viewpoint, arising from the estimated $\hat{\theta}$, matches well the ground truth.



Fig. 2. The agent-bias estimation problem, example of Sect. 5.2.

5.2. An Agent-Bias Estimation Problem

In many multi-agents situations a common, time-varying, phenomenon is observed by the suite. Each agent comes loaded with a systemic error, which represents the *biased* perspective of each individual agent. Here, the time-varying phenomenon is that each agent's observation is Gaussian coming from one of a set of variances; and the "bias" is that each agent i is confused by its individual mean θ_i (the bias), which is the object of the estimation task. In Fig. 2, panel (a), we examine the FIM behavior. For this particular example, it can be shown (details omitted for space limitations) that the FIM is diagonal, with all equal entries on the main diagonal, which are further independent of θ . Accordingly, it suffices to focus on the inverse FIM entry $[\mathcal{F}(\boldsymbol{\theta})^{-1}]_{11}$, for a certain parameter vector $\boldsymbol{\theta}$. We consider a uniform distribution among $|\mathcal{H}|$ states, for several values of $|\mathcal{H}|$. We see that the performance improves as n increases approaching the clairvoyant system. The performance worsens as the cardinality of the underlying state of nature increases, accounting for the increased uncertainty about the collected data increases. In panel (b), we see our estimation algorithms in operation. The general trends are similar to those made in the comments to Fig. 1, while here the EM algorithm provides some advantages for smaller numbers of tasks, a behavior that should come with no surprise, in the light of the discussion made at the end of Sect. 4. Finally, in panel (c), we show how the proposed algorithms are useful in performing identification of the biased agents and subsequent de-biasing. In the leftmost plot of the panel, we display the output observations x, with a different color for each agent. The curves show clearly how the different biases impact on the observations produced by the agents. In the rightmost plot, we display the observations of agent 1, after a de-biasing operation based on the estimated value θ_1 . We can appreciate how our algorithm allows an accurate de-biasing operation.

6. REFERENCES

- C. Aggarwal and T. Abdelzaher, "Integrating Sensors and Social Networks," *Social Network Data Analytics*, Springer, 2011.
- [2] B. Baingana, G. Mateos, and G. B. Giannakis, "Proximal-Gradient Algorithms for Tracking Cascades over Social Networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 563–575, Aug. 2014.
- [3] "Big Data: Theoretical and Algorithmic Foundations," IEEE Signal Process. Mag., vol. 31, no. 5, Sep. 2014.
- [4] E. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [5] C. Chamley, A. Scaglione, and Lin Li, "Models for the Diffusion of Beliefs in Social Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 16–29, May 2013.
- [6] P. Y. Chen and A. O. Hero, "Universal Phase Transition in Community Detectability Under a Stochastic Block Model," *Physical Review E* 91, 032804, 2015.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., John Wiley & Sons, 2006.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *J. Royal Statist. Soc.*, vol. B39, no. 1, pp. 1–38, 1977.
- [9] D. L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [10] Y. C. Eldar and M. Mishali, "Robust Recovery of Signals From a Structured Union of Subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [11] W. Feller, An Introduction to Probability and Its Applications, vol. 2, Wiley, NY, 1971.

- [12] A. O. Hero and B. Rajaratnam, "Foundational Principles for Large Scale Inference: Illustrations Through Correlation Mining," *Proc. of the IEEE*, in press, 2015. Also available as arxiv 1505.02475.
- [13] S. M. Kay, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory. Prentice Hall, 1993.
- [14] S. M. Kay, Fundamentals of Statistical Signal Processing, Volume II: Detection Theory. Prentice Hall, 1998.
- [15] S. Marano, V. Matta, and P. Willett, "The Importance of Being Earnest: Social Sensing With Unknown Agent Quality," *IEEE Trans. Signal and Inf. Process. Over Netw.*, submitted, Aug. 2015.
- [16] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace Learning and Imputation for Streaming Big Data Matrices and Tensors," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2663–2677, May 2015.
- [17] B. Oselio, J. Kulezsa, and A. O. Hero, "Multi-Layer Graph Analysis for Dynamic Social Networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 514–523, Aug. 2014.
- [18] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.
- [19] D. Wang, L. Kaplan, and T. Abdelzaher, "Maximum Likelihood Analysis of Conflicting Observations in Social Sensing," ACM Trans. Sensor Networks, vol. 10, no. 2, article 30, pp. 1–27, Jan. 2014.
- [20] 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.
- [21] K. Xu and A. O. Hero, "Dynamic Stochastic Blockmodels for Time-Evolving Social Networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 552–562, Aug. 2014.
- [22] X. Zhao and A. H. Sayed, "Learning over Social Networks via Diffusion Adaptation," *Proc. Asilomar Conference on Signals, Systems, and Computers*, pp. 709–713, Pacific Grove, CA, Nov. 2012.