Ergodic opinion dynamics over networks: learning influences from partial observations.

Chiara Ravazzi, *Member, IEEE*, Sarah Hojjatinia, *Member, IEEE*, Constantino M. Lagoa, *Member, IEEE*, Fabrizio Dabbene, *Senior Member, IEEE*

***Abstract*—In this paper we address the problem of inferring direct influences in social networks from partial samples of a class of opinion dynamics. The interest is motivated by the study of several complex systems arising in social sciences, where a population of agents interacts according to a communication graph. These dynamics over networks often exhibit an oscillatory behavior, given the stochastic effects or the random nature of the local interactions process.**

Inspired by recent results on estimation of vector autoregressive processes, we propose a method to estimate the social network topology and the strength of the interconnections starting from *partial observations* of the interactions, when the whole sample path cannot be observed due to limitations of the observation process. Besides the design of the method, our main contributions include a rigorous proof of the convergence of the proposed estimators and the evaluation of the performance in terms of complexity and number of sample. Extensive simulations on randomly generated networks show the effectiveness of the proposed technique.

I. Introduction

Recent years have witnessed the growth of a new research direction, at the boundary between social sciences and control theory, interested in studying dynamical social networks (DSN). As pointed out in the recent survey [1], "this trend was enabled by the introduction of new mathematical models describing dynamics of social groups, the advancement in complex networks theory and multi-agent systems, and the development of modern computational tools for big data analysis." Aim of this line of research is to study the mechanisms underlying *opinion formation*, that is to analyze how the individuals' opinions are modified and evolve as a consequence of the interactions among different agents connected together through a network of relationships.

To this end, several models have been proposed in the literature based on different communication mechanisms, see again [1] for a nice survey. These models were proven to be able to explain certain behaviors observed in the evolution of opinions, such as the emergence of consensus as in French-De-Groot models [2], [3], [4], or the persistent disagreement in social systems when stubborn agents are present [5], [6], [7]. Among these models, special attention has received the Friedkin and Johnsen (F&J) model [8], which has been experimentally validated for small and medium size groups [5], [9]. In the F&J model, the agents are influenced by the others' opinions, but are not completely open-minded, being persistently driven by their initial opinions. More precisely, at each round of communication the agents update their beliefs by taking a convex combination of the opinions coming from the neighbors, weighted with respect to an influence matrix, and their prejudices. It should be noted that this model extends the French-DeGroot model with stubborn agents, which is included as a particular case.

A common characteristic of most of the models presented in the literature is to assume synchronous interactions: in iterative rounds, all individuals interact with their neighbors (adjacent nodes in the relationship graph), and their opinions are updated taking into account others' opinion and their relative *influence* (strength of the interactions). Recently, the F&J model has been extended to situations in which the interactions occur following a *gossip* paradigm [10]. In particular, it was shown that the average opinion still exhibits the salient convergence properties of the original synchronous model.

The study of DSN has been very active in these years. On the one side, the theoretical properties of the proposed models have been thoroughly investigated [1]. On the other side, these models have been used to detect communities [11] and to define new centrality measures to identify social leaders [12], [13]. These latter have opened the way to the research on control of DSN, intended as the ability of placing influencers in an optimal way [14], [15].

However, most research in this field is based on the assumption that *the social network is given*, and in particular the influences among individuals are known. For instance, in the experiments performed by Friedkin and Johnsen in [16], the relative interpersonal influence was "measured" by introducing a mechanism in which actors were asked to distribute "chips" between the actors they interacted with. Clearly, these ad-hoc solutions are not viable in case of large networks. At the same time, however, there are now large amounts of data available, especially in the case of online social networks, and tools for measuring in real-time the individuals opinions are becoming available (as e.g. "likes" on Facebook or Twitter, instant polls, etc.).

These considerations have led to a new research direction, less explored so far, which aims at directly identifying the influence network based on collected observations. This problem has seen an increasing interest from various communities, such as computer science [17], [18], [19], [20], [21], signal processing [22], [23], and control community [24], [25], [26],

Chiara Ravazzi and Fabrizio Dabbene are with the Institute of Electronics, Computer and Telecommunication Engineering, National Research Council of Italy (CNR-IEIIT), c/o Politecnico di Torino, Corso Duca Degli Abruzzi, 10129, Italy. E-mail: chiara.ravazzi@ieiit.cnr.it, fabrizio.dabbene@ieiit.cnr.it.

Sarah Hojjatinia and Constantino M. Lagoa are with the Department of Electrical Engineering at the Pennsylvania State University, University Park, PA 16802, USA. E-mail: szh199@psu.edu, lagoa@psu.edu.

The authors' research has been partially supported by the PRIN 2017 Project, Prot. 2017S559BB, the National Institutes of Health (NIH) Grant R01 HL142732, and National Science Foundation (NSF) Grant 1808266.

to mention just a few. We refer the reader to [27] for an excellent survey of these techniques.

A. Relation to prior literature

One of the first works presenting the idea of using observed opinions to infer the topology of the underlying inference structure is given by [28], in which the authors consider French-De-Groot models with stubborn agents.

We observe that two different approaches, known as *finite-horizon* and *infinite horizon* identification procedures, can be adopted [29]. In the finite-horizon approach the opinions are observed for T subsequent rounds of conversation. Then, if enough observations are available, the influence matrix is estimated as the matrix best fitting the dynamics for $0 \leq k \leq T$, by employing classical identification techniques. This method however requires the knowledge of the discrete-time indices for the observations made, and the storage of the whole trajectory of the system. The loss of data from one of the agents in general requires to restart the experiment. More importantly, the updates usually occur at an unknown interaction rate, and the interaction times between agents are not observable in most scenarios, as observed in [30].

These considerations motivated the infinite-horizon identification procedure proposed in [28]. This approach performs the estimation based on the observations of the initial and final opinions' profile only, hence it is applicable only to asymptotically convergent dynamics. This technique was adopted in [31] to estimate the influence matrix of a F&J model, under the assumption that the matrix is sparse (i.e. agents are influenced by few friends only). In particular, by using tools from compressed sensing theory, under suitable assumptions, theoretical conditions to guarantee that the estimation problem is well posed and sufficient requirements on the number of observation ensuring perfect recovery are derived.

While the infinite horizon approach is surely innovative under various aspects, it suffers the clear drawback of being *static*. Indeed, the identification does not exploit the dynamical nature of the system, and it requires knowledge of initial and final opinion on several different topics to build up the necessary information to render the problem identifiable. Even if [31] shows that the number of topics that is necessary to correctly identify the network is strictly smaller than the size of the graph, and in many cases scales logarithmically with it, this information may be sometimes hard to collect.

The present work proposes a solution that goes in the direction of overcoming the main difficulties of both approaches: we propose a technique which exploits the dynamical evolution of the opinions, but at the same time does not require perfect knowledge of the interaction times, and it can be adapted to cases when some information is missing or partial. The main idea is to make recourse to tools recently developed in the context of identification of vector autoregressive (VAR) processes [32], [33]. The reader is referred to Section IV-C for a detailed discussion on the relationship among these results.

A preliminary version of this work appeared in [34]. This paper improves upon the results presented there both in terms of the proposed identification procedure and of theoretical analysis.

B. Paper organization

Section II introduces the main notation and some basic definitions that are used in the paper. In Section III, the adopted social interaction model is described, and a precise formulation of the estimation problem addressed in this paper is provided. The proposed approach to social network estimation is described in Section IV and the corresponding algorithms are presented in Section V. The theoretical results underpinning the proposed algorithms are provided in Section VI. Numerical results illustrating the performance of the proposed approach are presented in Section VII. Finally, concluding remarks are provided in Section VIII. The technical proofs are collected in the Appendix.

II. General notation

Throughout this paper, we use the following notation: The set of real numbers is denoted by $\mathbb{R}$ and the set of non-negative integers is denoted by $\mathbb{Z}_{\geq 0}$. The symbol $|\cdot|$ denotes either the cardinality of a set or the absolute value of a real number. We denote column vectors with lower case letters and matrices with upper case letters. The vector of all ones of appropriate dimension is represented by $\mathbb{1}$, and e_i is i-th vector of the standard basis; i.e., the vector that has 1 in its i-th entry and zeros for all other entries. Given a matrix A, $A^\dagger$ and $A^\top$ denote its pseudo-inverse and its transpose, respectively. Moreover, $\mathrm{sr}(A)$ is the spectral radius of the matrix A, and a square matrix A is said to be Schur stable if $\mathrm{sr}(A) < 1$. We denote the 2-norm, 1-norm, and 0-pseudo norm of a vector x with the symbols $\|x\|_2$, $\|x\|_1$, and $\|x\|_0$, respectively. The max norm of a matrix is defined as $\|A\|_{\max} \doteq \max_{ij} |A_{ij}|$, where A_{ij} is the entry of A in the i-th row and j-th column. For a matrix A, the symbol $\|A\|_2$ will stand for its induced 2-norm and $\|A\|_F$ denote the Frobenius norm. We define

$$\mathrm{supp}(A) = \{(i,j) : A_{i,j} \neq 0\}.$$

A matrix A with positive entries is said to be row stochastic if $A\mathbb{1} = \mathbb{1}$, and it is said to be row substochastic if $A\mathbb{1} \leq \mathbb{1}$, where the inequality is entry-wise. The symbol $\mathrm{vec}(A)$ denotes the vector composed of the columns of A stacked, and, for a given vector v, $\mathrm{diag}(v)$ is a matrix whose main diagonal is composed of the entries of the vector v and the rest of the matrix is zero. Given two matrices A and B, their Kronecker product is denoted by $A \otimes B$, and, for A and B of same dimension, their entry-wise (Hadamard) product is denoted by $A \circ B$. Given a matrix $A \in \mathbb{R}^{n \times n}$, we denote by

$$\mathcal{P}_+(A) = \underset{M \in \mathbb{R}^{n \times n} : M_{ij} \geq 0}{\operatorname{argmin}} \|A - M\|$$

the projection of the matrix onto the cone of matrices with non-negative entries, or, equivalently,

$$[\mathcal{P}_+(A)]_{ij} = \begin{cases} A_{ij} & \text{if } A_{ij} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

A directed graph is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. A path in a graph is a sequence of edges which joins a sequence of vertices. A directed graph $\mathcal{G}$ is called strongly connected if there is a path from each vertex in the graph to every other vertex. Given a matrix $W \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ with non-negative entries, the weighted graph associated to W is the graph $\mathcal{G}_W = (\mathcal{V}, \mathcal{E}_W, W)$ with node set $\mathcal{V}$, defined by drawing an edge $(i,j) \in \mathcal{E}_W$ if and only if $W_{i,j} > 0$ and putting weights $W_{i,j}$. Moreover, every node corresponding to a row which sums to less than one is said to be a deficiency node.

Given a probability space, we denote the expected value of random variable x by $\mathbb{E}[x]$.

III. Problem statement

In this section, we introduce the adopted opinion dynamics model, and we formulate the learning problem we are aiming to solve.

A. Randomized F&J opinion dynamics

We consider a finite population $\mathcal{V}$ of interacting individuals (agents) in a social network. To avoid trivialities, we assume that $|\mathcal{V}| > 2$. Mathematically, the social network is described by means of a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where $\mathcal{V}$ represents the agents, $\mathcal{E}$ are the potential interactions or communications, and the *influence matrix* $W \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ reflects the intensity of these interactions. The influence matrix is adapted to the graph. We denote the orientation of an agent $v \in \mathcal{V}$ towards some subject by a scalar value x_v. We will refer to x_v as the *opinion of agent* v.

In the F&J model [35], the opinions can attain a continuum of values and the times of interactions are discrete. The dynamics evolve as follows [36]: each agent $i \in \mathcal{V}$ starts from an initial belief $x_i(0) = u_i \in \mathbb{R}$, conceived a priori. Then, at each interaction time $k \in \mathbb{Z}_{\geq 0}$, a subset of nodes $\mathcal{V}_k$ of fixed cardinality is randomly selected from a uniform distribution over $\mathcal{V}$. If the node i is active at time k (i.e.; $i \in \mathcal{V}_k$), agent i interacts with a randomly chosen neighbor j and updates its belief according to a convex combination of its previous belief, the belief of j, and its initial belief. Namely,

$$\begin{aligned} x_i(k+1) &= \lambda_i\big((1-W_{ij})x_i(k) + W_{ij}x_j(k)\big) \\ &\quad +(1-\lambda_i)u_i \qquad \forall i \in \mathcal{V}_k \\ x_\ell(k+1) &= x_\ell(k) \qquad \forall \ell \in \mathcal{V}\setminus\mathcal{V}_k, \end{aligned} \tag{1}$$

where λ_i are parameters defining how sensitive each agent is to the opinions of the others based on interpersonal influences. We will refer to λ_i as the *susceptibility of agent* i. We assume that W is row-stochastic, i.e., $W\mathbb{1} = \mathbb{1}$, and we set $\Lambda \doteq \operatorname{diag}(\lambda)$. We assume that $\Lambda \neq I$ and $\lambda_i \in (0,1], \forall i \in \mathcal{V}$. We denote the set of neighbors of node $i \in \mathcal{V}$ by the notation $\mathcal{N}_i \doteq \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$, the degree $d_i \doteq |\mathcal{N}_i|$ and maximal degree $d_{\max} \doteq \max_{v\in\mathcal{V}} d_v$.

The dynamics (1) can be formally rewritten in the following form: given $\mathcal{V}_k$ and letting $\theta(k) \doteq \{\theta_i\}_{i\in\mathcal{V}_k}$, we have

$$x(k+1) = A(k)x(k) + b(k), \tag{2}$$

where

$$A(k) \doteq \left(I - \sum_{i\in\mathcal{V}_k} e_i e_i^\top (I-\Lambda)\right)\left(I + \sum_{i\in\mathcal{V}_k} W_{i\theta_i}(e_i e_{\theta_i}^\top - e_i e_i^\top)\right),$$

$\theta_i = j \in \mathcal{N}_i$ with probability $1/d_i$, $b(k) = B(k)u$, and

$$B(k) \doteq \sum_{i\in\mathcal{V}_k} e_i e_i^\top (I-\Lambda).$$

It should be noted that the opinions sequence $\{x(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ is a Markov process [37], i.e. the conditional distribution of $x(k+1)$ given the current state $x(k)$ does not depend on the past values. Due to the random nature of the dynamical system and to the pairwise interactions, the Markov process fails to converge in a deterministic sense, and shows persistent oscillations. However, under suitable conditions, we can guarantee the convergence of the expected dynamics and the ergodicity of the oscillations. More precisely, we make the following assumption involving the topology of the network.

Assumption 1. *From any $\ell \in \mathcal{V}$ there exists a path in $\mathcal{G}$ from node ℓ to a node i such that $\lambda_i < 1$.*

Then, the following two results are a direct consequence of the results in [10].

Proposition 1 (Convergence in expectation)**.** *Under Assumption 1, it holds*

$$\mathbb{E}[x(k+1)] = \overline{A}\mathbb{E}[x(k)] + \overline{b}$$

where

$$\begin{aligned} \overline{A} &\doteq \mathbb{E}[A(k)] = (1-\beta)\,I + \beta\Lambda(I - D^{-1}(I-W)), \\ \overline{b} &\doteq \beta(I-\Lambda)u, \end{aligned}$$

$\beta = |\mathcal{V}_k|/|\mathcal{V}|$, and D is the degree matrix of the network, a diagonal matrix whose diagonal entry is equal to the degree $d_i = |\mathcal{N}_i|$. Moreover, the sequence $\mathbb{E}[x(k)]$ converges to

$$\mathbb{E}[x(\infty)] = (I-\overline{A})^{-1}\overline{b}.$$

Theorem 1 (Ergodicity of F&J dynamics)**.** *Let Assumption 1 hold. Then*

1) *$x(k)$ converges in distribution to a random variable x_∞ and the distribution is the unique invariant distribution for* (1)*;*
2) *the process is ergodic, i.e. there exists a random variable x_∞ such that almost surely*

$$\lim_{k\to\infty} \frac{1}{k}\sum_{\ell=0}^{k-1} x(\ell) = \mathbb{E}[x_\infty]; \tag{3}$$

3) *the limit random variable satisfies $\mathbb{E}[x_\infty] = (I-\overline{A})^{-1}\overline{b}$.*

These properties play a crucial role for our developments and are illustrated through the following example.

Example 1. *We consider the Zachary's Karate Club dataset extracted from [38], whose graph is depicted in Figure 1. This network represents the friendships between 34 members of a karate club at a US university in the 1970s. The number of connections is equal to $|\mathcal{E}| = 78$, and the maximal degree of the network is $d_{\max} = 17$. The nonzero entries of influence matrix W are generated according to a uniform distribution in the range $[0,1]$, and then the rows are normalized to make*

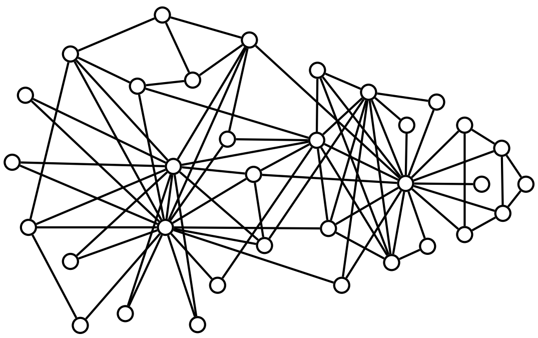
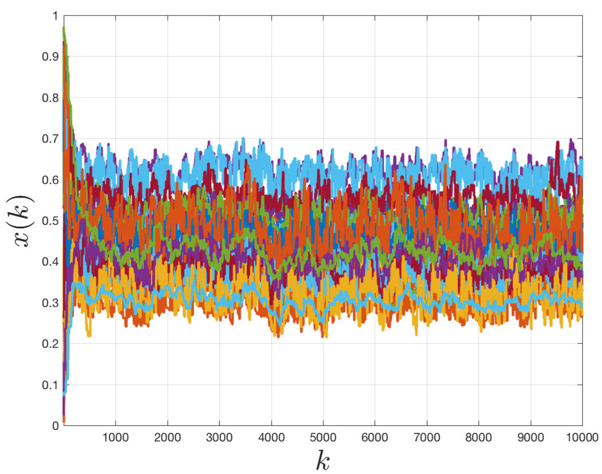
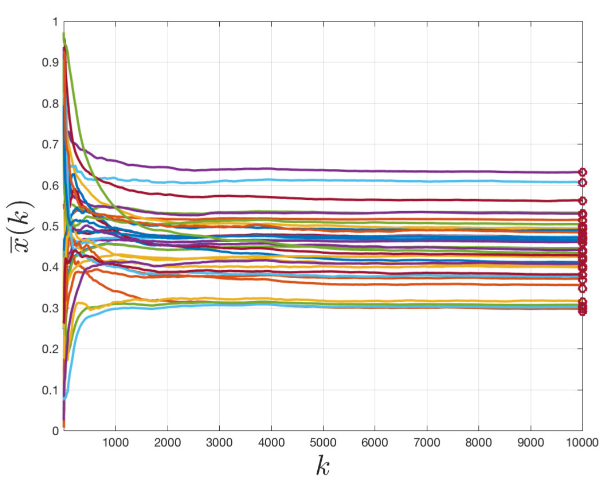


Fig. 1. Zachary karate club network.



(a) Evolution of the opinions.



(b) Evolution of the time averages.

Fig. 2. Friedkin and Johnsen dynamics with random interactions in Zachary karate club network

W row-stochastic. The sensitive parameters λ_i are extracted from a uniform distribution in the range of $[0.9, 1]$.

Figure 2(a) shows the evolution of the opinions. It should be noticed that the dynamics does not converge and the oscillations are persistent over the time. However, these fluctuations are ergodic and the time averages converge to a limit point corresponding to the limit point of the expected dynamics (see red circles in Figure 2(b)).

B. Sampling dynamics over networks

Let us consider the opinion dynamics in (1) where the influence matrix W is unknown. Our goal is to learn the influence matrix W, or, at least the topology of the influence network, that is $\mathrm{supp}(W)$, given access to partial observations from the F&J dynamics. More precisely, at each time we observe information of the form

$$z(k) = P(k)x(k), \tag{4}$$

where $P(k)$ is a random measurement matrix defined by

$$P(k) = \mathrm{diag}(p(k)),$$

and $p(k) \in \{0, 1\}^{\mathcal{V}}$ is a random selection vector independent of $x(k)$ with known distribution and $\mathbb{E}[p(k)] > 0$. This provides a general sampling model that can address many sampling schemes and theoretical results are proven in this general framework.

Example 2. *As an example, consider the so-called* intermittent observations scheme. *If we let*

$$p(k) = \begin{cases} \mathbb{1} & \textit{w.p. } \rho \\ 0 & \textit{otherwise} \end{cases}$$

then at $k \in \mathbb{Z}_{\geq 0}$ all observations are available with probability ρ, or no observations at all are performed. This model allows to capture the typical situation in which the actual rates at which the interactions occur are not perfectly known, and thus sampling time is different from interaction time.

Given the sequence of observation $\{z(k)\}_{k=1}^{t}$ we are interested in constructing an estimation of the matrix W, call it $\widehat{W}_t$, and in deriving theoretical conditions on the number of samples that are sufficient to have an error not larger than a fixed tolerance ϵ with high probability.

Remark 1 (On the observation scheme)**.** *Different observation schemes other than intermittent observations are possible. For instance, partial observation schemes as those discussed in [32], [34], in which at each step each opinion is observed independently with probability p, are captured by the proposed framework. However, it should be noted at this point that different schemes collect different "amounts of information." This can result in substantial changes in the performance of the proposed approach. Hence, although asymptotic results are proven for arbitrary sampling schemes, the number of measurements needed for meaningful results can differ substantially from scheme to scheme.*

IV. Learning algorithm

Before presenting our approach for the opinions model identification, we provide a second moments analysis of the evolution dynamics and its observations.

A. Second moment analysis

We start by introducing the opinions' cross-correlation matrix, which is defined as follows

$$\Sigma^{[\ell]}(k) \doteq \mathbb{E}\left[x(k)x(k+\ell)^\top\right].$$

The following theorem provides a description of the evolution of the covariance matrix $\Sigma^{[\ell]}(k)$, $\ell = 0, 1, 2, \ldots$ The proof is provided in Appendix A.

Theorem 2. *Assume that in the graph associated with W for any node $v \in \mathcal{V}$ there exists a path from v to a node i such that $\lambda_i < 1$. Then for all $k, \ell \in \mathbb{Z}_{\geq 0}$ we have*

$$\Sigma^{[\ell+1]}(k) = \Sigma^{[\ell]}(k)\overline{A}^\top + \mathbb{E}[x(k)]\overline{b}^\top. \tag{5}$$

Moreover $\Sigma^{[\ell]}(k)$ converges to $\Sigma^{[\ell]}(\infty)$ for all $\ell \in \mathbb{Z}_{\geq 0}$, satisfying

$$\Sigma^{[\ell+1]}(\infty) = \Sigma^{[\ell]}(\infty)\overline{A}^\top + \mathbb{E}[x(\infty)]\overline{b}^\top. \tag{6}$$

It should be noted that the relation in (6) is a sort of Yule-Walker equation, [32], used for estimation in autoregressive processes.

Using Proposition 1 and after some manipulations, equation (6) can be rewritten as an algebraic quadratic matrix equation.

Corollary 1. *Solving* (6) *is equivalent to finding the solutions of the following matrix equation*

$$\begin{aligned} &f^{[\ell]}(Q) \\ &:= Q\Sigma^{[\ell]}(\infty)Q^\top + Q\left(\Sigma^{[\ell+1]}(\infty) - \Sigma^{[\ell]}(\infty)\right) - \overline{b}\overline{b}^\top \\ &= 0, \end{aligned} \tag{7}$$

with $\overline{A} = I - Q$.

Corollary 2. *Solving* (6) *is equivalent to finding the solutions of the following matrix equation*

$$\Sigma^{[\ell+h+1]}(\infty) - \Sigma^{[\ell+1]}(\infty) = (\Sigma^{[\ell+h]}(\infty) - \Sigma^{[\ell]}(\infty))\overline{A}^\top \tag{8}$$

for any given $h \in \mathbb{Z}_{\geq 0}$

Theorem 2, Corollary 1 and Corollary 2 provide some hints on how to identify the influence matrix W. The main stream of the methodology is summarized in Figure 3.

B. Proposed methodology

Our approach consists in the following steps (see also Figure 3):

(a) Sampling the opinion dynamics according to (4).
(b) Estimation of the cross-correlation matrices $\Sigma^{[\ell]}(\infty)$, $\ell = 0, 1, \ldots, N_\Sigma$ from partial observations $\{z(k)\}_{k=1}^t$ exploiting ergodicity of the dynamics;
(c) Use of the cross-correlation matrices estimations to approximate $\overline{A}$ exploiting the relation (6), (7), or (8) and project onto the cone of matrices with non-negative entries.
(d) Use of the transition matrix estimation to recover the influence network W.

Before presenting the main algorithm some considerations are in order. Assume that the cross correlation matrices are known exactly. Then, we need to solve equation (7) in order to get an estimation of average transition matrix A (see point (c)). This approach requires to solve a quadratic matrix equation which is similar to an Algebraic Riccati Equation [39] but with the main difference that the unknown matrix Q is not symmetric. However, we can attempt to find a solution to (7) by solving the following optimization problem

$$\min_{Q \in \mathbb{R}^{n\times n}} \frac{1}{2}\|f^{[\ell]}(Q)\|_F^2. \tag{9}$$

If $f^{[\ell]}(Q)$ were convex, the minimization problem would admit a unique solution [40] and several iterative algorithms could be used for the minimization, as conjugate gradient methods (see Polak & Ribiére version and Fletcher & Reeves version in [41]). The main feature of these iterative methods is that each step requires the evaluation of the gradient and the exact line searches and the complexity is of order $O(|\mathcal{V}|^6)$ for each iteration. An alternative is to use a Matlab function, `nleqn`, using the Gauss-Newton and the Levenberg-Marquardt methods. However, these would need to evaluate the Jacobian matrix that can be very demanding for large networks.

Remark 2. *It is worth remarking that relations in* (6) *and* (7) *require the knowledge of parameters u, β and Λ in order to compute term $\overline{b}$. In case of missing knowledge of these parameters one can exploit the relation in* (8).

In next section we propose two algorithms for social influence learning that will be compared in terms of accuracy and prior information on parameters of the dynamics required. The main difference consists in how to perform point (c), i.e. the average transition matrix estimation.

- Social Influence lEarNiNg Algorithm I (SIENNA I). The estimation of average transition matrix $\overline{A}$ is based on the relation in (6) to avoid nonlinear optimization. This will require an estimator of the expected opinions $\mathbb{E}[x(\infty)]$ and allow us to perform estimation by solving a linear matrix equation at the cost of a certain approximation error that will be evaluated in Theorem 3.
- Social Influence lEarNiNg Algorithm II (SIENNA II). The estimation of average transition matrix $\overline{A}$ will leverage on the relation (8) in Corollary 2 for a fixed $h \in \mathbb{Z}_{\geq 0}$.

C. Relationship with VAR processes identification

The approach presented in this paper takes inspiration upon recent works on the identification of vector autoregressive (VAR) processes from partial measurements.

In particular, the works [32], [33] consider VAR processes of the form

$$x(k+1) = Ax(k) + w(k), \text{ with } w(k) \sim \mathcal{N}(0, Q_w) \tag{10}$$

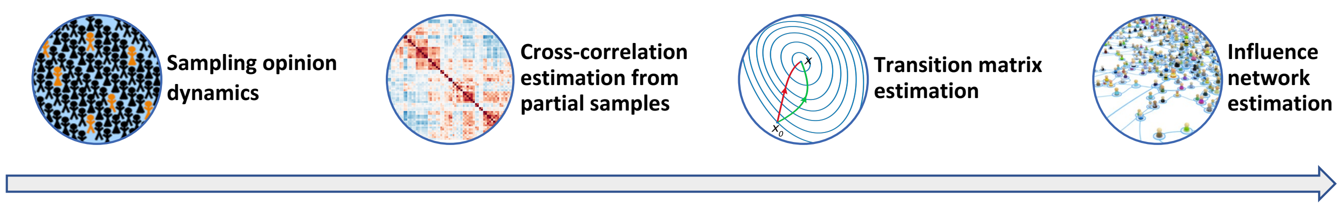


Fig. 3. SIENNA: Main stream of the methodology.

with $x(k) \in \mathbb{R}^n$. The goal there is to recover the state transition matrix A from partial observations of the state x. We observe that the observation framework considered in [32], [33] bears some similarities with the one discussed in Section III-B. Moreover, the flow of the estimation and the methods proposed in [32], [33] share some common features (covariance estimation and solution of a sort of Yule-Walker equations). However, the system (2) considered in this study is not captured in the VAR setup. Indeed, while (10) considers a *fixed* transition matrix A, and the only source of randomness is given by the additive Gaussian noise, in the setup of this paper the whole matrix $A = A(k)$ is random. Moreover, as it can be noticed in the Appendix the tools used for the proofs are very different.

V. SIENNA: Social Influence lEarNiNg Algorithm

The estimators presented in this work (cross-correlation matrices, transition matrix estimator, etc.) will be computed on-line, in the sense that, each time a new sample set is aquired, the estimators will be updated. In the rest of the paper, the index k will be used for the time representing the evolution of opinions and the index t is related to the evolution of estimators.

A. Estimating the expected opinion profile

In order to estimate the expected opinion profile $\mathbb{E}[x(\infty)]$, we consider time-averaged observations of $z(k)$. The proof of the next proposition is a direct consequence of the independence assumption and is omitted for brevity.

Proposition 2. *The following relation holds*

$$\mathbb{E}[z(k)] = \pi \circ \mathbb{E}[x(k)]$$

where $\pi = \mathbb{E}[p(k)] > 0$.

We estimate $\mathbb{E}[z(\infty)]$ using time averages $\overline{z}(t) \doteq \frac{1}{t}\sum_{k=1}^{t} z(k)$, and $\mathbb{E}[x(\infty)]$ by leveraging on Proposition 2

$$\widehat{\overline{x}}_i(t) \doteq \overline{z}_i(t)/\pi_i. \tag{11}$$

B. Estimating the cross-correlation matrices

In order to estimate the cross correlation matrices $\Sigma^{[\ell]}(\infty)$, we consider the empirical covariance matrix of the observations $z(k)$. Let us denote

$$S^{[\ell]}(k) \doteq \mathbb{E}[z(k)z(k+\ell)^\top].$$

We have the following proposition, whose proof follows from basic arguments, and is given in Appendix B for completeness.

Proposition 3. *The following relation holds*

$$S^{[\ell]}(k) = \Pi^{[\ell]}(k) \circ \Sigma^{[\ell]}(k)$$

where $\Pi^{[\ell]} \doteq \mathbb{E}[p(k)p(k+\ell)^\top]$.

Since $S^{[\ell]}(k)$ is unknown, we construct an estimate of it using time averages

$$\widehat{S}^{[\ell]}(t) \doteq \frac{1}{t-\ell}\sum_{k=1}^{t-\ell} z(k)z(k+\ell)^\top$$

from which, using the relation in Proposition 3, we compute

$$\widehat{\Sigma}^{[\ell]}_{ij}(t) \doteq \widehat{S}^{[\ell]}_{ij}(t)/\Pi^{[\ell]}_{ij}. \tag{12}$$

Example 3. *It should be noticed that in the special case of intermittent observations we have* $\pi = \rho\mathbb{1}$

$$\Pi^{[0]} = \rho\mathbb{1}\mathbb{1}^\top \quad \textit{and} \quad \Pi^{[\ell]} = \rho^2\mathbb{1}\mathbb{1}^\top \quad \textit{if } \ell \neq 0$$

from which $\widehat{\overline{x}}(t) = \overline{z}(t)/\rho$ *and* $\widehat{\Sigma}^{[\ell]}(t) = \widehat{S}^{[\ell]}(t)/\rho^2$.

C. Estimating the average transition matrix

Given estimations $\widehat{\Sigma}^{[\ell]}(t)$ of $\Sigma^{[\ell]}(t)$, we use them to approximate $\overline{A}$ by exploiting relation in (6). More precisely, we start by choosing the number N_Σ of covariance matrices that are going to be used in the estimation of dynamics. In particular, given estimates $\widehat{\Sigma}^{[\ell]}(t)$, define

$$\widehat{\Sigma}_-(t) \doteq \frac{1}{N_\Sigma}\sum_{\ell=0}^{N_\Sigma-1}\widehat{\Sigma}^{[\ell]}(t); \qquad \widehat{\Sigma}_+(t) \doteq \frac{1}{N_\Sigma}\sum_{\ell=1}^{N_\Sigma}\widehat{\Sigma}^{[\ell]}(t).$$

Using relation (6) and projecting onto the cone of matrices with non-negative entries. one obtains

$$\widehat{\overline{A}}(t)^\top = \mathcal{P}_+\left[\widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right]$$

or, equivalently,

$$\widehat{\overline{A}}(t)^\top = \mathcal{P}_+\left[I + \widehat{\Sigma}_-(t)^\dagger\left(\frac{1}{N_\Sigma}\left(\widehat{\Sigma}^{[N_\Sigma]} - \widehat{\Sigma}^{[0]}(t)\right) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right] \tag{13}$$

Remark 3 (On N_Σ)**.** *It should be remarked that, from a theoretical point of view, one could simply use* $N_\Sigma = 1$ *to perform the identification. This is exactly the approach we originally introduced in [34]. However, we observe that larger values of* N_Σ *can significantly improve performance, especially when the probability of observation decreases. Indeed, as the probability of observation decreases, the probability of the entries of* $z(k)z(k+1)^T$ *being zero increases, leading to* *poorer estimates of covariance matrices when using $N_\Sigma = 1$. By using not only $z(k)z(k+1)^T$ for the identification of the influence matrix, but also $z(k)z(k+2)^T$, ..., $z(k)z(k+N_\Sigma)^T$, we have more information and are able to better estimate the dynamics of the social network.*

Remark 4 (On prior information)**.** *The procedure presented in this paper can be easily modified to incorporate prior information available on the network to be identified. More precisely, instead of* (13)*, the following estimator can be used*

$$\min_{A} \quad \left\{ \widehat{\Sigma}_-(t)A - \left[\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right]\right\} + \gamma f(A)$$
$$\text{s.t. } A \in \Omega$$

where f is used to "encourage" some property of the network, γ provides a trade-off between accuracy and the desired property and Ω enforces a priori *known structure. For example, if one wants to determine a sparse network that is compatible with the data collected, then a good choice for $f(A)$ is the ℓ_1 norm of the off-diagonal elements of A [42]. Moreover, if it is known that there is a set of connections that do not exist, then Ω is the set of matrices whose entry (i,j) is zero for the pairs (i,j) that are not connected.*

D. Estimating the network topology and the influence matrix

Once an estimate of the average transition matrix $\overline{A}$ has been obtained, we can retrieve the topology of the influence network in a straightforward manner, by noticing that $\text{supp}(\overline{A}) = \text{supp}(W)$. Hence, we can reconstruct the support of W by taking the elements of the estimated matrix $\widehat{\overline{A}}$ that are significantly larger than zero.

When parameters β and Λ are also known and $\lambda_i \in (0,1]$, we can estimate the intensity of the influence by exploiting Proposition 1, writing

$$\widehat{W} = \widehat{D}\Lambda^{-1}(\overline{A} - (1-\beta)I - \beta\Lambda(I - \widehat{D}^{-1})),$$

where $\widehat{D}$ represents an estimate of the degree matrix D obtained from the reconstructed support. That is, $\widehat{D}$ is the diagonal matrix with elements

$$\widehat{D}_{ii} = \|a_i\|_0,$$

with $a_i^\top$ being the i-th row of matrix $\widehat{\overline{A}}$.

VI. Performance analysis

In this section, we provide a theoretical analysis on the performance of the proposed estimators.

Theorem 3 (Error in expected opinion profile)**.** *Let $\Delta x \doteq \widehat{\overline{x}}(t) - \mathbb{E}[x(\infty)]$ and $\Delta\Sigma^{[\ell]}(t) \doteq \widehat{\Sigma}^{[\ell]}(t) - \Sigma^{[\ell]}(\infty)$. We have the following bounds*

$$\mathbb{P}(\|\Delta x(t)\|_2 \geq \epsilon_1) \leq \frac{C_1 n}{\epsilon_1^2(t+1)(1-\text{sr}(\overline{A}))(\pi^\star)^2}$$
$$\mathbb{P}(\|\Delta\Sigma^{[0]}(t)\|_F \geq \epsilon_2) \leq \frac{C_2 n^2}{\epsilon_2^2(t+1)(1-\text{sr}(\overline{Q}))(\Pi^\star)^2}$$

where C_1 and C_2 are some positive constants (independent of t and n), $\Pi^\star = \min_{ij}\Pi_{ij}^{[\ell]}$, $\pi^\star = \min_{i\in\mathcal{V}}\pi_i$, and

$$\overline{Q} \doteq \begin{bmatrix} \mathbb{E}[A(k)\otimes A(k)] & \mathbb{E}[A(k)\otimes B(k)] & \mathbb{E}[B(k)\otimes A(k)] \\ 0 & \overline{A}\otimes I & 0 \\ 0 & 0 & I\otimes\overline{A} \end{bmatrix}.$$

Moreover, if $\ell \neq 0$,

$$\mathbb{P}(\|\Delta\Sigma^{[\ell]}(t)\|_F \geq \epsilon_3) \leq \frac{C_\ell n^2}{\epsilon_3^2(t+1)(1-\text{sr}(\overline{\Gamma}))(\Pi^\star)^2}$$

with

$$\overline{\Gamma} \doteq \begin{bmatrix} \overline{A}\otimes\overline{A} & \overline{A}\otimes\overline{B} & \overline{B}\otimes\overline{B} \\ 0 & \overline{B}\otimes I & 0 \\ 0 & 0 & I\otimes\overline{A} \end{bmatrix}$$

and $\overline{B} = \mathbb{E}[B(k)]$.

The proof of Theorem 3 is given in Appendix C. We can rewrite the results in the following equivalent form that expresses the speed of convergence of the proposed estimators as a function of the size of the network and number of samples.

Corollary 3. *With probability at least $1-\delta$ we have*

$$\|\Delta x(t)\|_2 \leq \frac{C_1\sqrt{n}}{\pi^\star\sqrt{\delta(t+1)(1-\text{sr}(\overline{A}))}},$$

$$\|\Delta\Sigma^{[0]}(t)\|_F \leq \frac{C_2 n}{\Pi^\star\sqrt{\delta(t+1)(1-\text{sr}(\overline{Q}))}},$$

and, for any $\ell \neq 0$,

$$\|\Delta\Sigma^{[\ell]}(t)\|_F \leq \frac{C_3 n}{\Pi^\star\sqrt{\delta(t+1)(1-\text{sr}(\overline{\Gamma}))}}.$$

where C_1, C_2, and C_3 are some positive constants (independent of t and n). Finally, there are positive constants C_+, C_- (independent of t and n) such that

$$\|\Delta\Sigma_+(t)\|_F \leq \frac{C_+ n}{\Pi^\star\sqrt{\delta(t+1)(1-\text{sr}(\overline{\Gamma}))}}$$

and

$$\|\Delta\Sigma_-(t)\|_F \leq \frac{C_- n}{\Pi^\star\sqrt{\delta(t+1)(1-\max(\text{sr}(\overline{\Gamma}),\text{sr}(\overline{Q}))}}.$$

Note that we can roughly estimate $\text{sr}(\overline{A}) \leq 1-\beta+\beta\lambda_{\max}$ where $\lambda_{\max} = \max_j \lambda_j$ (see proof of Lemma 2 in Appendix A) and, from Kronecker properties and Schur stability of $\overline{A}$, $\text{sr}(\overline{\Gamma}) \leq (1-\beta+\beta\lambda_{\max})$. In particular, for the intermittent sampling framework, we have $\pi^\star = \rho$ and $\Pi^\star = \rho^2$. Hence, we obtain that in this case the errors in the estimation of $\mathbb{E}[x(\infty)]$ and $\Sigma^{[\ell]}(\infty)$ are inversely proportional to the sampling probability and to the square of the sampling probability, respectively.

Leveraging on this bounds we can roughly estimate the error in the average transition matrix as follows. The proof is given in Appendix D

Theorem 4. *Let $\Delta\overline{A}(t) \doteq \overline{A}(t) - \widehat{\overline{A}}(t)$. With probability at least $1-\delta$ we have*

$$\|\Delta\overline{A}(t)\|_F = O\left(\frac{n^{3/2}(\sigma^+_{\max}+n)}{(\underline{\sigma}^-_{\min})^2\Pi^\star\sqrt{\delta(t+1)(1-\max(\mathrm{sr}(\overline{\Gamma}),\mathrm{sr}(\overline{Q}),\mathrm{sr}(\overline{A}))}}\right). \tag{14}$$

where $\sigma^+_{\max} = \|\Sigma_+\|_2$ and $\underline{\sigma}^-_{\min} \doteq \min(\sigma^-_{\min}, \widehat{\sigma}^-_{\min})$, being $\sigma^-_{\min}$, $\widehat{\sigma}^-_{\min}$ the minimum singular value of Σ_- and $\widehat{\Sigma}_-$, respectively

It should be noted that the estimation error on the transition matrix is based on the previous estimation of the cross-correlation matrices. In particular, in order to compute (11) we have to invert $(\hat{\Sigma}_+)$ and the estimation error depends on the minimum singular value (see also the Proof in Appendix D).

VII. Numerical Experiments

In this section, we provide numerical results that illustrate the performance of the proposed approach. As mentioned before, in this paper we mainly focus on a specific observation scheme; i.e., intermittent observations. In all the examples we used $\beta = 1$ in order to observe a social system where all agents have updated their opinion. This choice does not affect the analysis since the rate of convergence will be affected only by a constant. Diagonal values in matrix Λ were randomly chosen uniformly between 0.9 and 1. Initial conditions were also generated randomly with a value between 0 and 1. Moreover, when not mentioned otherwise, we have chosen $N_\Sigma = 5$. The quantity used to measure the performance of covariance estimation is the residual

$$R_{\widehat{\Sigma}_+}(k) \doteq \frac{\|\widehat{\Sigma}_+(k) - \widehat{\Sigma}_-(k)\,\overline{A}^\top + \widehat{x}(k)\,\overline{b}^\top\|_F}{\|\widehat{\Sigma}_+(k)\|_F}.$$

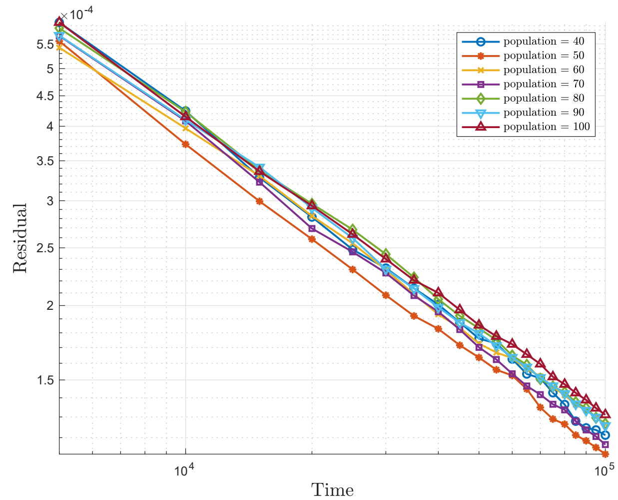
Finally, distance/similarity between two finite sets is measured using the so-called Jaccard index. The Jaccard index J of two finite sets A and B is defined as

$$J(A,B) \doteq |A\cap B|/|A\cup B|,$$

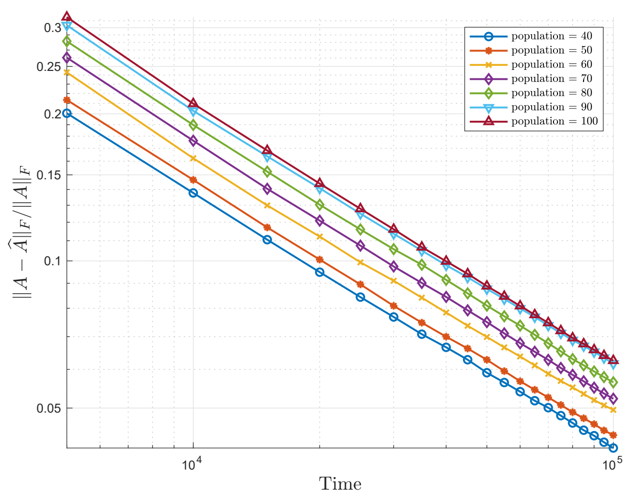
and in our case was used to measure the different between the real and the identified set of connections.

We start this section by analyzing the performance of the algorithm SIENNA I. First, to test the scalability of the proposed method, we performed simulations for several network sizes. More precisely we considered random networks with node degree 3 with a number of nodes between 40 and 100. Again, we assumed that the opinions were measured at every time instant. The averaged results for 20 simulations are depicted in Figures 4. As expected, the rate of convergence is similar for all network sizes with "graceful degradation" of performance with increasing size. This shows that the proposed approach scales well with the dimension of the problem. Moreover, the relative error in the estimate of the influence matrix decreases approximately at a rate of $t^{-1/2}$, as predicted by the results in Theorem 4. Again, performance of both methods is similar.

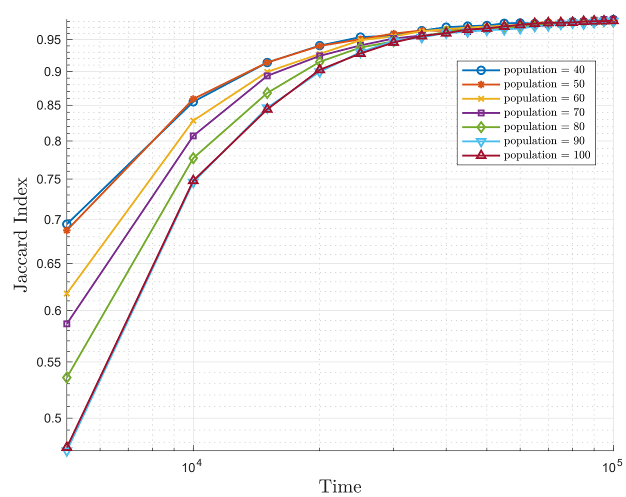
In the results described above it is assumed that the state was measured at all time instants. To test how the algorithm



(a) Residual in estimating the covariance.



(b) Error in estimating the influence matrix.



(c) Jaccard index.

Fig. 4. Convergence of the proposed scheme in a network with $\rho = 1$, degree 3 using intermittent measurements for different population sizes. Average of 20 simulations.

performs under intermittent observations, we performed simulations for different values of the probability of observation ρ. The results of 20 simulations were averaged and these are depicted in Figure 5. As expected, as the probability of observation decreases the accuracy of the estimate also decreases. However, there is a "graceful" and predictable degradation of performance with decreasing ρ, where the decrease of information collected can be compensated by increasing the number of measurements.

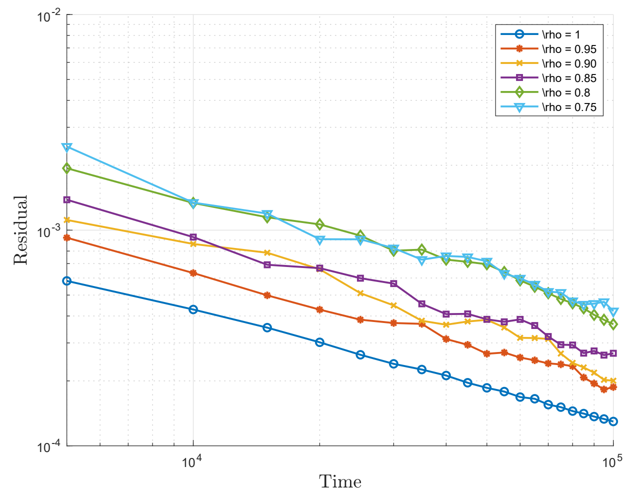
Again, distance between identified and true network structure was measured using Jaccard index. We see that the proposed method recovers the true network structure for high enough number of measurements.

To further test the performance of the proposed approach, the algorithm was applied to Erdős Rényi networks of size $N = 50$ and probability of connection $p = 0.08$. This value of p ensures that the networks are connected (with high probability). The results obtained are depicted in Figure 6. As it can be seen, the performance is similar to the networks with homogeneous degree used in the previous examples.

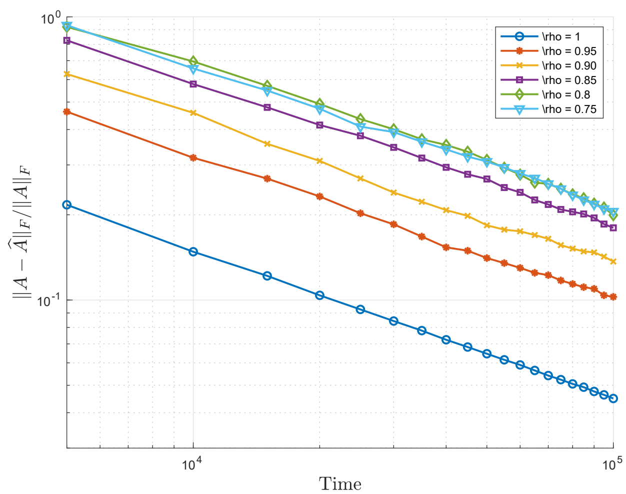
We now provide results on the influence of the number of covariance matrices N_Σ used on the performance of the proposed approach. To this end, we considered random networks of size 50 and a probability of observation of $\rho = 0.9$ and varied N_Σ between 1 and 10. The results of 20 averaged simulations are depicted in Figure 7. The plots clearly show the advantage of using N_Σ greater than one. There is a very large jump in performance from $N_\Sigma = 1$ to $N_\Sigma = 2$. Moreover, performance increases until $N_\Sigma = 5$ and then more or less stabilizes. This clearly illustrates that additional covariance information can compensate for a decrease in collected data.

We now study how the stubbornness of agents affects the performance of the algorithm. Recall that λ_i defines the sensitivity of agent i to the opinions of others and to test its influence we assumed that all agents had the same value; i.e., we used $\Lambda = \lambda I$. The results are depicted in Figure 8. Although the theoretical results provided in this paper show that the smaller the value of λ the faster the convergence of the mean and covariance estimates, the numerical simulations show better convergence of the algorithm for larger values of λ. Our conjecture is that, since equation (5) is satisfied for all values of k, the algorithm can also leverage the transient values of the estimates to estimate the network structure. In the case of large (close to 1) values of λ there is a rich transient behavior that seems to contain a lot of information on the network and our approach appears to be able to use it. Effort is now being put in carefully studying this phenomenon.

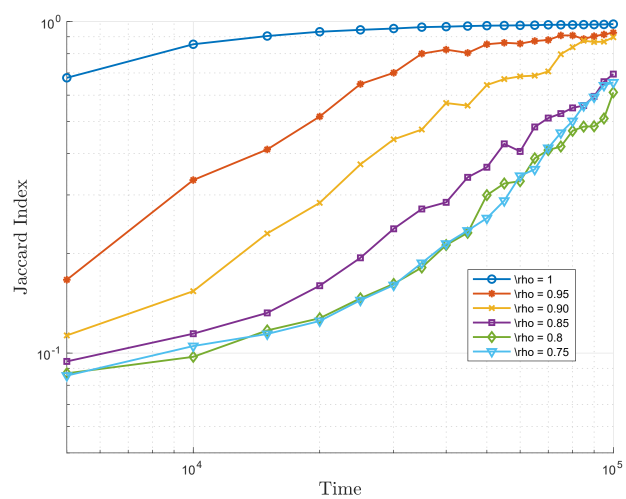
Finally, we provide simulation results comparing the performance of SIENNA I and SIENNA II. We start be recalling that SIENNA II uses less *a priori* information to compute estimates of the influence matrix. More precisely, as compared to SIENNA I, SIENNA II does not require knowledge of the initial condition of the agents nor of their stubbornness. This is compensated by estimating one more term associated with the covariance of the agents' opinions. In Figure 9 we compare the performance of these two algorithms by averaging the results of 20 simulations. As it can be seen, although both algorithms converge as the number of measurements increases,



(a) Residual in estimating the covariance.

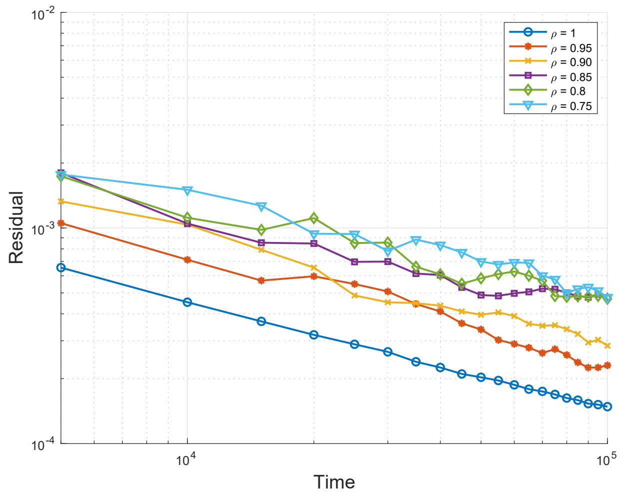


(b) Error in estimating the influence matrix.

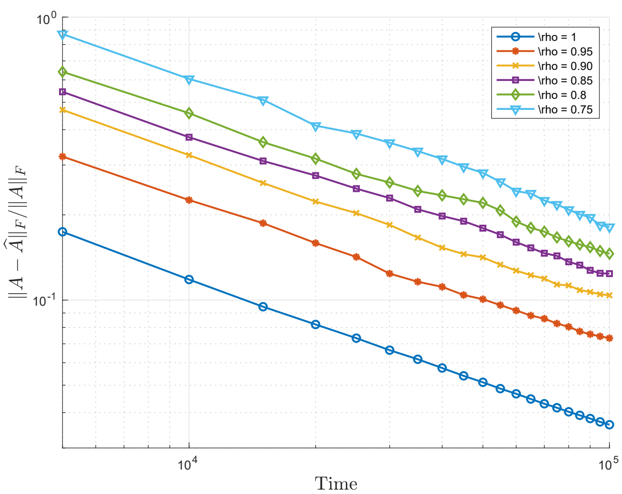


(c) Jaccard index.

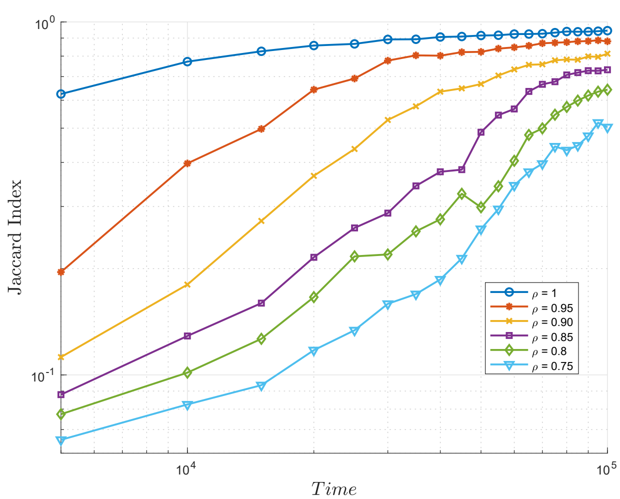
Fig. 5. Convergence of the proposed scheme in a network with $N = 50$, degree 3 using intermittent measurements for different values of ρ. Average of 20 simulations.



(a) Residual in estimating the covariance.



(b) Error in estimating the influence matrix.



(c) Jaccard index.

Fig. 6. Convergence of the proposed scheme for random Erdős Rényi graph with probability 0.08 in a network with $N = 50$, using intermittent measurements for different values of ρ. Average of 20 simulations.

SIENNA II has a larger initial error and a slower speed of convergence. Hence, we believe that SIENNA II should only be used in cases where the information on initial conditions and/or stubbornness is not available.

VIII. Concluding Remarks

In this paper we have addressed the problem of estimating the influence matrix of randomized opinion dynamics over networks from intermittent observations. The method well adapts to the realistic situation in which it is not always possible to know the instants of local interactions between agents. More precisely, we have considered cases in which the system is sampled at times which may differ from the actual system update times. Convergence of the proposed methods is proven and their performance is illustrated using randomly generated gossip dynamics over networks.

It should be noted that the methodology proposed in this work immediately extends to the general problem of identifying ergodic systems starting from partial observation, that is systems of the form

$$x(k+1) = A(k)x(k) + b(k), \tag{15}$$

with $(A(k), b(k))$ independent and identically distributed random variables; $A(k)$ is a $n \times n$ matrix and $b(k)$ is a $n \times 1$ vector [43].

IX. Acknowledgment

The authors would like to thank F. Bullo, G. Como, F. Fagnani, P. Frasca, N. E. Friedkin, and A. Proskurnikov for interesting conversations on topics related to this paper. The authors are also grateful to one anonymous reviewer for pointing out equation (8), which allowed to derive the SIENNA II version of the algorithm.

Appendix

A. *Proof of Theorem 2*

We first provide the following two lemmas, whose introduction is instrumental for proving Theorem 2.

Lemma 1. *Consider a substochastic matrix $M \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$. If in the graph associated to M there is a path from every node to a deficiency node, i.e. a node corresponding to a row which sums to less than one, then M is Schur stable.*

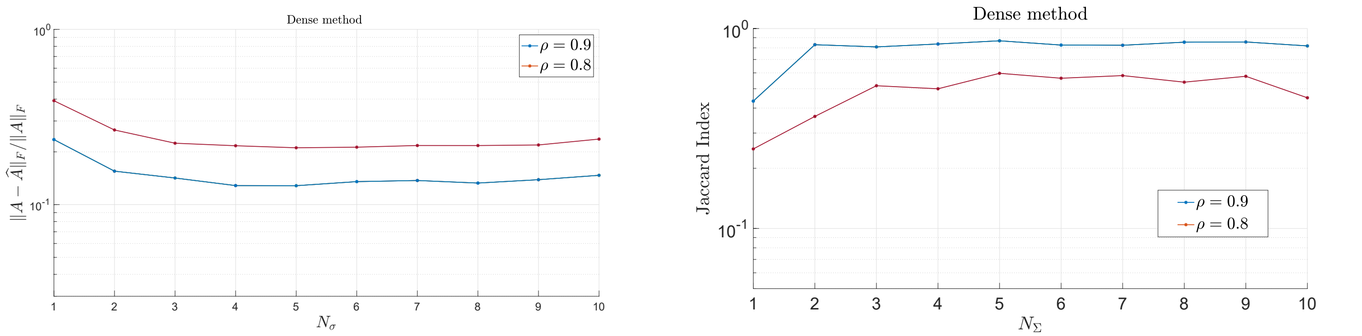
The interested reader can find the proof in [36].

Lemma 2. *Assume that in the graph associated to W for any node $v \in \mathcal{V}$ there exists a path from v to a node i such that $\lambda_i < 1$. Then*

$$\overline{A} \doteq \mathbb{E}[A(k)] = (1-\beta)I + \beta\Lambda(I - D^{-1}(I - W)) \tag{16}$$

and $\overline{A^{\otimes 2}} \doteq \mathbb{E}[A(k) \otimes A(k)]$ are Schur stable. Moreover,

$$\mathrm{sr}(\overline{A}) \leq 1 - \beta + \beta\lambda_{\max}.$$



(a) Relative error in the estimation of the average transition matrix.

(b) Jaccard index between the original influence matrix and reconstructed influence matrix.

Fig. 7. Comparison of performance for different values of N_Σ in a network with degree 3, population 50 and for $\rho = 0.8$ and $\rho = 0.9$. The relative error in the estimation of the average transition matrix and the Jaccard index are averaged over 20 simulations.

Proof. Equation (16) follows from the definition of expectation. It should be noticed that $\text{supp}(\overline{A}) = \text{supp}(W)$ and

$$\overline{A}_{ii} = (1-\beta) + \beta\lambda_i(1-(1-W_{ii})/d_i)$$
$$\overline{A}_{ij} = \beta\lambda_i W_{ij}/d_i \quad \text{if } i \neq j.$$

We get $\sum_j \overline{A}_{ij} = (1-\beta) + \beta\lambda_i$ which is strictly less than 1 if $\lambda_i < 1$. Hence, under the hypothesis, we have that $\overline{A}$ is a substochastic matrix corresponding to a graph with a path from any node v to a node m whose row sums up to less than one. By Lemma 1 $\overline{A}$ is Schur stable and from (16) we get that $\text{sr}(\overline{A}) \leq 1 - \beta + \beta\lambda_{\max}$.

We now prove that $\overline{A^{\otimes 2}}$ is Schur stable. We observe that $\overline{A^{\otimes 2}}$ is a substochastic matrix as $\overline{A^{\otimes 2}}\mathbb{1}_{|\mathcal{V}|^2} = \text{vec}(\mathbb{E}[A(k)\mathbb{1}_{|\mathcal{V}|}\mathbb{1}_{|\mathcal{V}|}^\top A(k)^\top])$. Notice that

$$[\overline{A^{\otimes 2}}]_{(i-1)|\mathcal{V}|+j,(h-1)|\mathcal{V}|+\ell} = \mathbb{E}[A_{ih}(k)A_{j\ell}(k)].$$

Since $\lambda_i > 0$, then $A(k)_{ii} > 0$ with probability one. Therefore we have $\mathbb{E}[A_{ii}(k)A_{j\ell}(k)] > 0$ iff $\mathbb{E}[A_{j\ell}] > 0$ and, with the same argument, we get $\mathbb{E}[A_{ih}(k)A_{jj}(k)] > 0$ iff $\mathbb{E}[A_{ih}] > 0$. Given $(i_0, j_0) \in \mathcal{V} \times \mathcal{V}$, under the hypothesis, we have that there exists a sequence $(i_0, i_1, i_2, \ldots i_n, h)$ from i_0 to a node h such that $\lambda_h < 1$ and a sequence $(j_0, j_1, j_2, \ldots j_n, \ell)$ from j_0 to ℓ with $\lambda_\ell < 1$. It should be noticed that (h, ℓ) is a deficiency node for the product graph associated to $\overline{A^{\otimes 2}}$. This implies that there exists an admissible path $(i_0, j_0), (i_0, j_1), \ldots, (i_0, \ell), (i_1, \ell), \ldots, (h, \ell)$ in the product graph associated to $\overline{A^{\otimes 2}}$. Using Lemma 1 we conclude that $\mathbb{E}[A(k) \times A(k)]$ is Schur stable. Moreover, we have

$$\sum_{h,\ell} \mathbb{E}[A_{ih}A_{j\ell}] = \begin{cases} 1-\beta+\beta\lambda_i^2 & \text{if } i = j \\ (1-\beta(1-\lambda_i))(1-\beta(1-\lambda_j)) & \text{if } i \neq j. \end{cases}$$

□

Proof of Theorem 2: It should be noticed that the opinions $x(k)$ are bounded, as they satisfy

$$\min_{v\in\mathcal{V}} u_v \leq x_i(k) \leq \max_{v\in\mathcal{V}} u_v =: u_{\max} \tag{17}$$

for all $i \in \mathcal{V}$ and $k \geq 0$. Then $\mathbb{E}[x(k)]$ and $\mathbb{E}[x(k)x(k)^\top]$ exist and are bounded.

By defining

$$y(k) \doteq [y_1(k)^\top, y_2(k)^\top, y_3(k)^\top]^\top,$$

with $y_1(k) = \text{vec}(x(k)x(k)^\top)$, $y_2(k) = \text{vec}(ux(k)^\top)$, $y_3(k) = \text{vec}(x(k)u^\top)$,

$$Q(k) \doteq \begin{bmatrix} A(k)\otimes A(k) & A(k)\otimes B(k) & B(k)\otimes A(k) \\ 0 & A(k)\otimes I & 0 \\ 0 & 0 & I\otimes A(k) \end{bmatrix},$$

and $M(k) \doteq [(B(k)\otimes B(k))^\top, (B(k)\otimes I)^\top, (I\otimes B(k))^\top]^\top$, we can easily check that

$$y(k+1) = Q(k)y(k) + M(k)\text{vec}(uu^\top) \tag{18}$$

from which

$$\begin{aligned}\mathbb{E}[y(k+1)] &= \mathbb{E}[\mathbb{E}[y(k+1)|y(k)]] \\ &= \mathbb{E}[Q(k)]\mathbb{E}[y(k)] + \mathbb{E}[M(k)]\text{vec}(uu^\top)\end{aligned}$$

with

$$\begin{aligned}\overline{Q} &\doteq \mathbb{E}[Q(k)] \\ &= \begin{bmatrix} \mathbb{E}[A(k)\otimes A(k)] & \mathbb{E}[A(k)\otimes B(k)] & \mathbb{E}[B(k)\otimes A(k)] \\ 0 & \mathbb{E}[A(k)]\otimes I & 0 \\ 0 & 0 & I\otimes \mathbb{E}[A(k)] \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{E}[A(k)\otimes A(k)] & \mathbb{E}[A(k)\otimes B(k)] & \mathbb{E}[B(k)\otimes A(k)] \\ 0 & \overline{A}\otimes I & 0 \\ 0 & 0 & I\otimes \overline{A} \end{bmatrix}\end{aligned}$$

and $\overline{M} \doteq \mathbb{E}[M(k)] = (\mathbb{E}[B(k)\otimes B(k)]^\top, (\overline{B}\otimes I)^\top, (I\otimes \overline{B})^\top]^\top$. By Lemma 2 the matrix $\mathbb{E}[Q(k)]$ is Schur stable and, consequently,

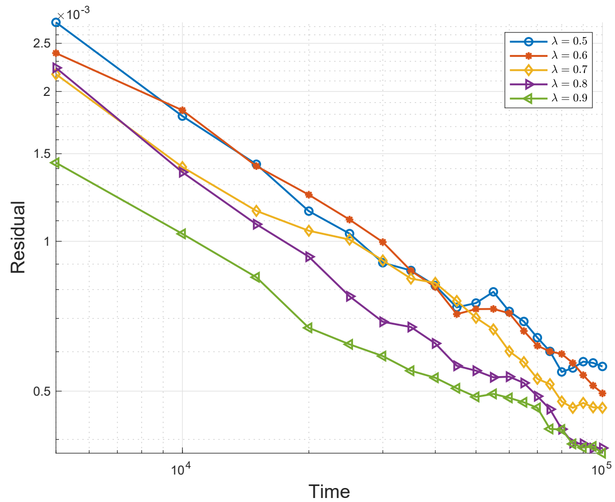
$$\lim_{k\to\infty} \mathbb{E}[y(k)] = (I-\overline{Q})^{-1}\overline{M}\text{vec}(uu^\top),$$

from which we conclude that the sequence $\mathbb{E}[y_1(k)]$ and $\Sigma^{[0]}(k) = \mathbb{E}[x(k)x(k)^\top]$ are convergent as $k \to \infty$. From definition it is easy to verify that

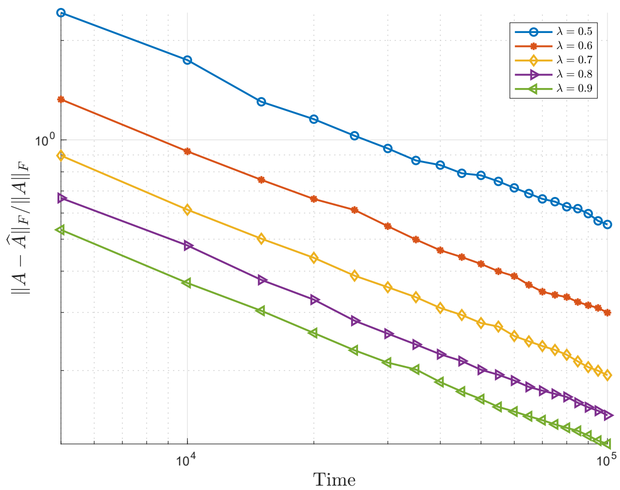
$$\Sigma^{[1]}(k) = \Sigma^{[0]}(k)\overline{A}^\top + \mathbb{E}[x(k)]\overline{b}^\top$$

from which, letting $k \to \infty$ and using Proposition 1, we get that also $\Sigma^{[1]}(k)$ converges to a limit point satisfying

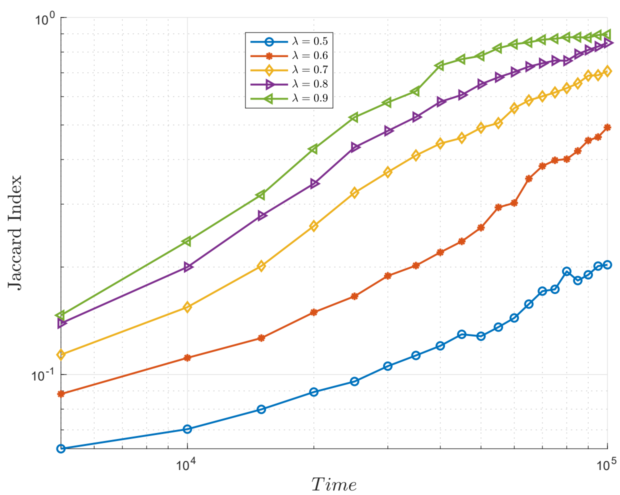
$$\Sigma^{[1]}(\infty) = \Sigma^{[0]}(\infty)\overline{A}^\top + \mathbb{E}[x(\infty)]\overline{b}^\top.$$



(a) Residual in estimating the covariance.

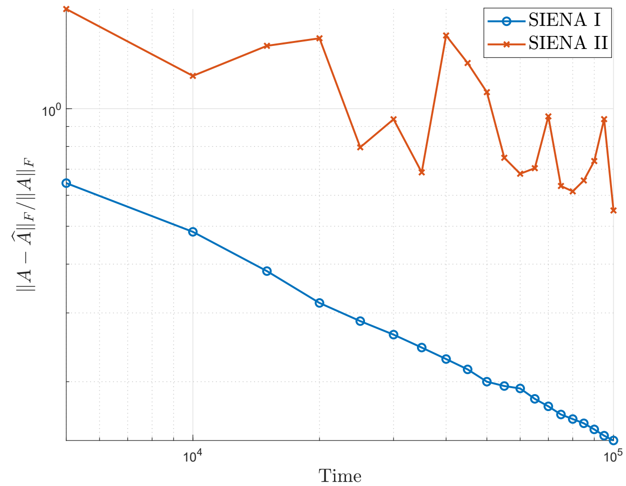


(b) Error in estimating the influence matrix.

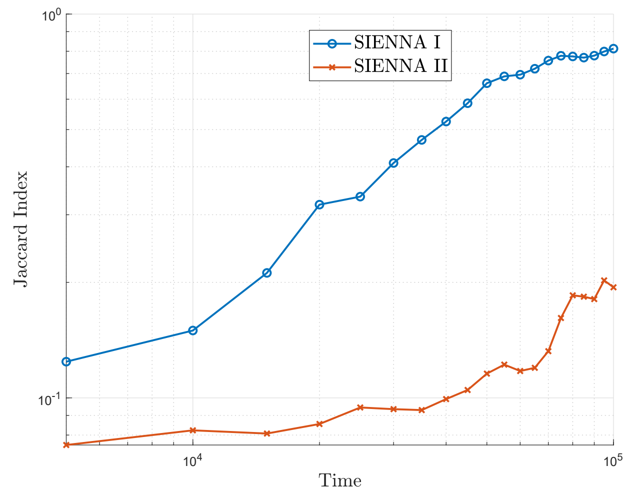


(c) Jaccard index.

Fig. 8. Convergence of the proposed scheme in a network with $N = 50$, degree 3, $\rho = 0.9$ using intermittent measurements for different values of λ_i. Average of 40 simulations.



(a) Error in estimating the influence matrix.



(b) Jaccard index.

Fig. 9. Comparison between SIENNA I and SIENNA II in networks of size 20 with $\rho = 0.9$ and degree 3 using intermittent measurements. Average of 20 simulations.

Moreover, since for all $\ell \in \mathbb{Z}_{\geq 0}$ we have

$$\Sigma^{[\ell+1]}(k) = \Sigma^{[\ell]}(k)\overline{A}^\top + \mathbb{E}[x(k)]\overline{b}^\top.$$

Then, by induction, $\Sigma^{[\ell]}(k)$ converges to some $\Sigma^{[\ell]}(\infty)$ as $k \to \infty$ and the limit satisfies

$$\Sigma^{[\ell+1]}(\infty) = \Sigma^{[\ell]}(\infty)\overline{A}^\top + \mathbb{E}[x(\infty)]\overline{b}^\top.$$

□

Remark 5. *In alternative way, the result in Theorem 2 for $\ell \neq 0$ can be shown by defining*

$$y^{[\ell]}(k) \doteq [y_1^{[\ell]}(k)^\top, y_2^{[\ell]}(k)^\top, y_3^{[\ell]}(k)^\top]$$

with $y_1^{[\ell]}(k) = \mathrm{vec}(x(k)x(k+\ell)^\top)$, $y_2^{[\ell]}(k) = \mathrm{vec}(ux(k+\ell)^\top)$, *and* $y_3^{[\ell]}(k) = \mathrm{vec}(x(k)u^\top)$ *and noticing that there exists matrices* $\Gamma^{[\ell]}(k)$ *and* $\Upsilon^{[\ell]}(k)$ *such that*

$$y^{[\ell]}(k+1) = \Gamma^{[\ell]}(k)y^{[\ell]}(k) + \Upsilon^{[\ell]}(k)\mathrm{vec}(uu^\top)$$

with

$$\overline{\Gamma} \doteq \mathbb{E}[\Gamma^{[\ell]}(k)] \begin{bmatrix} \overline{A}\otimes\overline{A} & \overline{A}\otimes\overline{B} & \overline{B}\otimes\overline{A} \\ 0 & \overline{A}\otimes I & 0 \\ 0 & 0 & I\otimes\overline{A} \end{bmatrix},$$

that is Schur stable, and

$$\overline{\Upsilon} \doteq \mathbb{E}[\Upsilon^{[\ell]}(k)] = \begin{bmatrix} \overline{B}\otimes\overline{B} \\ \overline{B}\otimes I \\ I\otimes\overline{B} \end{bmatrix}.$$

B. Proof of Proposition 3

By definition we have

$$\begin{aligned} S^{[\ell]} &= \mathbb{E}[z(k)z(k+\ell)^\top] \\ &= \mathbb{E}\left[P(k)x(k)x(k+\ell)^\top P(k+\ell)^\top\right]. \end{aligned}$$

By stacking the columns of $S^{[\ell]}$ into a single column vector, we obtain

$$\begin{aligned} &\text{vec}\left(S^{[\ell]}\right) \\ &\quad= \mathbb{E}\left[(P(k)\otimes P(k+\ell))\text{vec}\left(x(k)x(k+\ell)^\top\right)\right] \\ &\quad= \mathbb{E}\left[\mathbb{E}\left[P(k)\otimes P(k+\ell)\text{vec}\left(x(k)x(k+\ell)^\top\right)\middle| x(k)\right]\right] \\ &\quad= \mathbb{E}\left[P(k)\otimes P(k+\ell)\right]\mathbb{E}\left[\text{vec}\left(x(k)x(k+\ell)^\top\right)\right] \\ &\quad= \mathbb{E}\left[P(k)\otimes P(k+\ell)\right]\text{vec}\left(\mathbb{E}\left[x(k)x(k+\ell)^\top\right]\right) \\ &\quad= \mathbb{E}\left[P(k)\otimes P(k+\ell)\right]\text{vec}\left(\Sigma^{[\ell]}(k)\right) \\ &\quad= \mathbb{E}\left[\text{diag}(p(k))\otimes\text{diag}(p(k+\ell))\right]\text{vec}\left(\Sigma^{[\ell]}(k)\right) \end{aligned}$$

from which we conclude the statement. □

C. Proof of Theorem 3

We now estimate $\mathbb{E}[\|\Delta x(k)\|_2^2]$. It should be noticed that by the definition of (1), the opinions $x(k)$ are bounded, as they satisfy (17). As a consequence, partial observations $z(k)$ are bounded and all moments of $x(k)$ and $z(k)$ are uniformly bounded. Let us denote $x^\star := \mathbb{E}[x(\infty)]$ and $e(\ell) := (z(\ell) - \pi\circ x^\star)$ and observe that $\pi\circ(\widehat{x}(k) - x^\star) = \overline{z} - \pi\circ x^\star = \frac{1}{k+1}\sum_{\ell=0}^{k} e(\ell)$. We thus have

$$\mathbb{E}\|\pi\circ(\widehat{x}(k)-x^\star)\|^2 = \mathbb{E}\left\|\frac{1}{(k+1)}\sum_{\ell=0}^{k} e(\ell)\right\|^2$$

$$= \frac{1}{(k+1)^2}\sum_{\ell=0}^{k}\mathbb{E}\left[e(\ell)^\top e(\ell)\right] + 2\sum_{\ell=0}^{k}\sum_{\ell=r}^{k-\ell}\mathbb{E}\left[e(\ell)^\top e(\ell+r)\right].$$

From (17) we can ensure that there exists a constant $\eta\in\mathbb{R}$ such that $\frac{1}{(k+1)}\sum_{\ell=0}^{k}\mathbb{E}\left[\|e(\ell)\|^2\right] \le \eta \doteq (u_{\max}-u_{\min})^2 n,\ \forall k$, where $u_{\min} = \min_v u_v$ and $u_{\max} = \max_v u_v$. Now, it should be observed that

$$\begin{aligned} &\mathbb{E}\left[e(\ell)^\top e(\ell+r)\right] \\ &\quad= \mathbb{E}\left[\mathbb{E}\left[e(\ell)^\top e(\ell+r)|P(\ell),x(\ell)\right]\right] \\ &\quad= \mathbb{E}\left[e(\ell)^\top\mathbb{E}\left[e(\ell+r)|P(\ell),x(\ell)\right]\right] \\ &\quad= \mathbb{E}\left[e(\ell)^\top\left(\mathbb{E}\left[z(\ell+r)|P(\ell),x(\ell)\right]-\pi\circ x^\star\right)\right] \\ &\quad= \mathbb{E}\left[e(\ell)^\top\pi\circ\left(\mathbb{E}\left[x(\ell+r)|x(\ell)\right]-x^\star\right)\right]. \end{aligned} \tag{19}$$

By repeated conditioning on $x(\ell), x(\ell+1),\ldots,x(\ell+r-1)$, we obtain

$$\mathbb{E}\big[x(\ell{+}r)|x(\ell)\big] = \mathbb{E}\left[A(k)\right]^r x(\ell) + \sum_{s=0}^{r-1}\mathbb{E}[A(k)]^s\mathbb{E}[B]u,$$

and by recalling that $x^\star$ is a fixed point for the expected dynamics we get

$$x^\star = \mathbb{E}\left[A(k)\right]^r x^\star + \sum_{s=0}^{r-1}\mathbb{E}[A(k)]^s\mathbb{E}[B(k)]u. \tag{20}$$

From equations (19) and (20) we obtain

$$\begin{aligned} &\mathbb{E}\left[e(\ell)^\top e(\ell+r)\right] \\ &\quad= \mathbb{E}\left[e(\ell)^\top\pi\circ\left(\mathbb{E}\left[A(k)\right]^r(x(\ell)-x^\star)\right)\right] \\ &\quad= \mathbb{E}\left[(P(\ell)x(\ell)-\pi\circ x^\star)^\top\pi\circ\left(\mathbb{E}\left[A(k)\right]^r(x(\ell)-x^\star)\right)\right] \\ &\quad= \mathbb{E}\left[\pi\circ(x(\ell)-x^\star)^\top\pi\circ(\overline{A}^r(x(\ell)-x^\star))\right] \\ &\quad\le \eta\nu^r, \end{aligned}$$

where, by Lemma 2, $\nu = \text{sr}(\overline{A}) < 1$. Finally, we have

$$\begin{aligned} \mathbb{E}\left[\|\pi\circ(\widehat{x}(k)-x^\star)\|^2\right] &\le \frac{\eta}{(k+1)^2}\left(k+1+2\sum_{\ell=0}^{k-1}\sum_{r=0}^{k-\ell}\nu^r\right) \\ &\le \frac{\eta}{(k+1)}\left(1+\frac{2}{1-\nu}\right). \end{aligned}$$

Since $\nu < 1$ we have there exists $C_1 > 0$ such that

$$\mathbb{E}\left[\|\pi\circ(\widehat{x}(k)-x^\star)\|^2\right] \le \frac{1}{(k+1)}\frac{Cn}{1-\nu}$$

and, consequently, we get

$$\mathbb{E}\left[\|\widehat{x}(k)-x^\star\|^2\right] \le \frac{Cn}{(k+1)(1-\nu)(\pi^\star)^2},$$

where $\pi^\star = \min_{v\in\mathcal{V}}\pi_v$.

We thus have for any $\epsilon_1 > 0$

$$\begin{aligned} \mathbb{P}(\|\Delta x(t)\|_2 \ge \epsilon_1) &= \mathbb{P}(\|\Delta x(t)\|_2^2 \ge \epsilon_1^2) \\ &\le \frac{\mathbb{E}[\|\Delta x(t)\|_2^2]}{\epsilon_1^2} \\ &\le \frac{C_1 n}{\epsilon_1^2(k+1)(1-\nu)(\pi^\star)^2} \end{aligned}$$

where the first inequality follows from Markov inequality [44].

The proof of the second part of theorem follows the same arguments using the recursion in (18) and Remark 6 and we omit for brevity. □

D. Proof of Theorem 4

From definition we have

$$\|\Delta\overline{A}(t)\|_F \le \left\|\overline{A}(t) - \mathcal{P}_+\left[\widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right]\right\|_F.$$

Using triangular inequality and the fact that $\mathcal{P}_+$ is the operator that project a matrix onto the cone of positive matrices, we get

$$\begin{aligned}
\|\Delta\overline{A}(t)\|_F &\leq \left\|\overline{A}(t) - \widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right\|_F + \\
&+ \Big\|\widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right) \\
&\quad -\mathcal{P}_+\left[\widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right]\Big\|_F \\
&\leq 2\left\|\overline{A}(t) - \widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right\|_F \\
&\leq 2\sqrt{n}\left\|\overline{A}(t) - \widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right\|_2.
\end{aligned}$$

Let us compute

$$\begin{aligned}
&\left\|\overline{A}(t) - \widehat{\Sigma}_-(t)^\dagger\left(\widehat{\Sigma}_+(t) - \widehat{\overline{x}}(t)\overline{b}^\top\right)\right\|_2 \\
&\quad\leq \|\Sigma_-^\dagger(\Sigma_+ - \overline{x}(t)\overline{b}^\top) - \widehat{\Sigma}_-^\dagger(\widehat{\Sigma}_+ - \widehat{\overline{x}}(t)\overline{b}^\top)\|_2 \\
&\quad\leq \underbrace{\|\Delta\Sigma_-^\dagger\|_2\left(\|\Sigma_+\|_2 + \overline{b}^\top\overline{x}(t)\right)}_{T_1} \\
&\quad+ \underbrace{\|\widehat{\Sigma}_-^\dagger\|_2\left(\|\Delta\Sigma_+\|_2 + \|\Delta\overline{x}(t)\|_2\|\overline{b}\|_2\right)}_{T_2}.
\end{aligned}$$

Using the inequality derived in [45], and observing that $\|\Delta\Sigma_-\|_2 \leq \|\Delta\Sigma_-\|_F$ one obtains

$$\begin{aligned}
\|\Delta\Sigma_-^\dagger\|_2 &\leq \|\Sigma_-^\dagger\|_2^2\|\Delta\Sigma_-\|_2 \leq \|\Sigma_-^\dagger\|_2^2\|\Delta\Sigma_-\|_F \\
&= \frac{C_- n}{\sigma_{\min}^2\Pi^\star\sqrt{\delta(t+1)(1-\max(\mathrm{sr}(\overline{\Gamma}),\mathrm{sr}(\overline{Q}))}}.
\end{aligned}$$

Notice that $\overline{b}^\top\overline{x}(t) \in [0, \beta u_{\max}^2 n]$ from which

$$T_1 = O\left(\frac{n(\sigma_{\max}^+ + n)}{\Pi^\star\sigma_{\min}^2\sqrt{\delta(t+1)(1-\max(\mathrm{sr}(\overline{\Gamma}),\mathrm{sr}(\overline{Q}))}}\right).$$

For the estimation of term T_2 we observe

$$\|\Delta\overline{x}(t)\|_2\|\overline{b}\|_2 \leq \frac{C_1\beta u_{\max} n}{\pi^\star\sqrt{\delta(t+1)(1-\mathrm{sr}(\overline{A}))}}$$

and

$$\|\Delta\Sigma_+\|_2 \leq \|\Delta\Sigma_+\|_F \leq \frac{C_+ n}{\Pi^\star\sqrt{\delta(t+1)(1-\mathrm{sr}(\overline{\Gamma}))}}$$

from which

$$T_2 \leq \frac{Cn(\sigma_{\max}^+ + n)}{(\widehat{\sigma}_{\min}^-)^2\Pi^\star\sqrt{\delta(t+1)(1-\mathrm{sr}(\overline{\Gamma}))}}$$

where C is a positive constant. By summation of T_1 and T_2 we get

$$\begin{aligned}
&\|\Delta\overline{A}(t)\|_F \\
&= O\left(\frac{n^{3/2}(\sigma_{\max}^+ + n)}{\min(\sigma_{\min}^-, \widehat{\sigma}_{\min}^-)^2\Pi^\star\sqrt{\delta(t+1)(1-\max(\mathrm{sr}(\overline{\Gamma}),\mathrm{sr}(\overline{Q},\mathrm{sr}(\overline{A}))}}\right).
\end{aligned}$$

□

REFERENCES

[1] A. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

[2] J. French, "A formal theory of social power," *Psychological Review*, vol. 63, pp. 181–194, 1956.

[3] M. H. DeGroot, "Reaching a Consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[4] S. Chatterjee and E. Seneta, "Towards Consensus: Some Convergence Theorems on Repeated Averaging," *J. Appl. Prob.*, vol. 14, pp. 159–164, 1977.

[5] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinion change," *Advances in Group Processes*, vol. 16, pp. 1–29, 1999.

[6] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence, models, analysis and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, p. 2, 2002.

[7] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Math. Oper. Res.*, vol. 38, no. 1, pp. 1–27, Feb. 2013.

[8] N. Friedkin and E. Johnsen, "Social-influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, p. 193–205, 1990.

[9] N. Friedkin, "A formal theory of reflected appraisals in the evolution of power," *Administrative Science Quarterly*, vol. 56, no. 4, pp. 501– 529, Dec 2011.

[10] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *IEEE Trans. Control of Network Syst.*, vol. 2, no. 1, pp. 78–87, 2015.

[11] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1 – 44, 2016, community detection in networks: A user guide.

[12] N. E. Friedkin, "Theoretical Foundations for Centrality Measures," *The American Journal of Sociology*, vol. 96, no. 6, pp. 1478–1504, 1991.

[13] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, p. 056131, Nov. 2004.

[14] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, "Binary opinion dynamics with stubborn agents," *ACM Trans. Econ. Comput.*, vol. 1, no. 4, pp. 19:1–19:30, Dec. 2013.

[15] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabasi, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, May 2011.

[16] N. Friedkin and E. Johnsen, "Social influence networks and opinion change," in *Advances in Group Processes*, S. Thye, E. Lawler, M. Macy, and H. Walker, Eds., 1999, vol. 16, pp. 1–29.

[17] S. Daitch, J. Kelner, and D. Spielman, "Fitting a graph to vector data," 01 2009, p. 26.

[18] O. Banerjee and L. Ghaoui, "Model selection through sparse max likelihood estimation model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, 08 2007.

[19] G. Bresler, D. Gamarnik, and D. Shah, "Hardness of parameter estimation in graphical models," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 1062–1070.

[20] A. Montanari and J. A. Pereira, "Which graphical models are difficult to learn?" in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 1303–1311.

[21] G. Bresler, D. Gamarnik, and D. Shah, "Learning graphical models from the glauber dynamics," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4072–4080, 2018.

[22] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[23] S. P. Chepuri, S. Liu, G. Leus, and A. Hero, "Learning sparse graphs under smoothness prior," 03 2017, pp. 6508–6512.

[24] D. Materassi, G. Innocenti, L. Giarr, and M. Salapaka, "Model identification of a network as compressing sensing," *Systems & Control Letters*, vol. 62, no. 8, pp. 664 – 672, 2013.

[25] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[26] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[27] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.
[28] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks." *IEEE Trans. Signal and Information Processing over Networks*, vol. 2, no. 3, pp. 406–419, 2016.
[29] S. Parsegov, A. Proskurnikov, R. Tempo, and N. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2270–2285, 2017.
[30] X. Wang, X. Chen, and L. Wang, "Evolutionary dynamics of fairness on graphs with migration," *Journal of Theoretical Biology*, vol. 380, pp. 103 – 114, 2015.
[31] C. Ravazzi, R. Tempo, and F. Dabbene, "Learning influence structure in sparse social networks," *IEEE Trans. Control of Network Syst.*, vol. 5, no. 4, pp. 1976-1986, Dec. 2018.
[32] M. Rao, T. Javidi, Y. C. Eldar, and A. Goldsmith, "Estimation in autoregressive processes with partial observations," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4212–4216.
[33] M. Rao, A. Kipnis, T. Javidi, Y. C. Eldar, and A. Goldsmith, "System identification from partial samples: Non-asymptotic analysis," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 2938–2944.
[34] C. Ravazzi, S. Hojjatinia, C. Lagoa, and F. Dabbene, "Randomized opinion dynamics over networks: influence estimation from partial observations," in *Proc. of IEEE Conference on Decision and Control (CDC)*, 2018, pp. 2452–2457.
[35] N. Friedkin and E. Johnsen, "Social influence and opinions," *J. Math. Sociol.*, vol. 15, no. 3-4, pp. 193–205, 1990.
[36] P. Frasca, C. Ravazzi, R. Tempo, and H. Ishii, "Gossips and prejudices: Ergodic randomized dynamics in social networks," in *Proc. of IFAC NecSys 2013 Workshop*, Koblenz, Germany, 2013, pp. 212–219.
[37] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, ser. Probability and Random Processes. OUP Oxford, 2001.
[38] W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, pp. 452–473, 1977.
[39] P. Lancaster, P. Rodman, D. Lancaster, L. Rodman, and D. Rodman, *Algebraic Riccati Equations*, ser. Oxford Mathematical Monographs. Clarendon Press, 1995.
[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.
[41] H. Kim, N. Higham, and H.-M. Kim, "Numerical analysis of a quadratic matrix equation," *IMA Journal of Numerical Analysis*, vol. 20, 02 2001.
[42] M. Rao, T. Javidi, Y. C. Eldar, and A. Goldsmith, "Fundamental estimation limits in autoregressive processes with compressive measurements," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2895–2899.
[43] P. Diaconis and D. Freedman, "Iterated random functions," *SIAM Rev.*, vol. 41, no. 1, pp. 45–76, Mar. 1999.
[44] D. W. Stroock, *Probability Theory: An Analytic View*, 2nd ed. New York, NY, USA: Cambridge University Press, 2010.
[45] S. M. Rump, "Bounds for the componentwise distance to the nearest singular matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, pp. 83–103, 1997.

Chiara Ravazzi (M'13) received the Ph.D. degree in mathematics for engineering sciences from Politecnico di Torino, in 2011. From 2011 to 2016, she held Postdoctoral positions with the Departments of Mathematics and Electronics and Telecommunications, Politecnico di Torino. Since 2017, she has been a Tenured Researcher with the CNR-IEIIT, working with the Systems Modeling and Control Group. She has been serving as Member of the Conference Editorial Board of the IEEE CSS and as Associate Editor of the IEEE Transactions on Signal Processing since 2019. Her research interests include signal processing, optimization, and control of network systems.

Sarah Hojjatinia is currently a Ph.D. candidate in electrical engineering at the Pennsylvania State University. She received the B.S. and M.S. degrees in electrical engineering from K. N. Toosi University of Technology, Tehran in 2009 and 2013 respectively, She started her Ph.D. program in electrical engineering with dual M.S. degree in mechanical engineering at the Pennsylvania State University, in 2015, where she received the M.S. degree in mechanical engineering in 2018. Her research interests include system identification, machine learning, control and optimization with applications in behavioral sciences and data analysis.

Constantino M. Lagoa received the Ph.D. degree from the University of Wisconsin at Madison in 1998. He joined the Electrical Engineering Department of Pennsylvania State University, in August 1998, where he currently holds the position of Professor. He has a wide range of research interests including robust optimization and control, chance constrained optimization, controller design under risk specifications, system identification and control of computer networks. Dr. Lagoa has served as Associate Editor of IEEE Transactions on Automatic Control (2012-2017) and IEEE Transactions on Control systems Technology (2009-2013) and he is currently Associate Editor of Automatica.

Fabrizio Dabbene (SM'09) is a Director of Research at the institute IEIIT of the National Research Council of Italy (CNR), where he is the coordinator of the Information & Systems Engineering Group. He has held visiting and research positions at The University of Iowa, at Penn State University and at the Russian Academy of Sciences, Institute of Control Science, Moscow. He published more than 100 research papers and two books. He served as Associate Editor for Automatica (2008-2014) and for the IEEE Transactions on Automatic Control (2008- 2012), and he is currently Senior Editor of the IEEE Control Systems Society Letters. He was elected member of the Board of Governors (2014-2016) and he served as Vice-President for Publications (2015-2016). He is currently chairing the IEEE-CSS Italy Chapter.