Received 30 October 2019; revised 22 June 2020; accepted 22 June 2020. Date of publication 1 July 2020; date of current version 18 August 2020. The review of this article was arranged by Associate Editor Paolo Banelli.

Digital Object Identifier 10.1109/OJSP.2020.3006436

Interplay Between Topology and Social Learning Over Weak Graphs

VINCENZO MATTA [1], VIRGINIA BORDIGNON [2], AUGUSTO SANTOS [3], AND ALI H. SAYED [2]

[1]DIEM, University of Salerno, I-84084 Fisciano (SA), Italy
[2]School of Engineering, École Polytechnique Fédérale de Lausanne EPFL, CH-1015 Lausanne, Switzerland
[3]Adaptive System Laboratory, EPFL, CH-1015 Lausanne, Switzerland

CORRESPONDING AUTHOR: VINCENZO MATTA (e-mail: vmatta@unisa.it)

This work was supported in part by the Swiss National Science Foundation (SNSF) under Grant 205121-184999.
Partial results from this article were presented at ICASSP [1], [2].

ABSTRACT This work examines a distributed learning problem where the agents of a network form their beliefs about certain hypotheses of interest. Each agent collects streaming (private) data and updates continually its belief by means of a *diffusion* strategy, which blends the agent's data with the beliefs of its neighbors. We focus on *weakly-connected* graphs, where the network is partitioned into sending and receiving sub-networks, and we allow for heterogeneous models across the agents. First, we examine *what* agents learn (*social learning*) and provide an analytical characterization for the long-term beliefs at the agents. Among other effects, the analysis predicts when a leader-follower behavior is possible, where some sending agents control the beliefs of the receiving agents by forcing them to choose a particular and possibly fake hypothesis. Second, we consider the dual or reverse learning problem that reveals *how* agents learn: given the beliefs collected at a receiving agent, we would like to discover the influence that any sending sub-network might have exerted on this receiving agent (*topology learning*). An unexpected interplay between social and topology learning emerges: given H hypotheses and S sending sub-networks, topology learning can be feasible when $H \geq S$. The latter being only a necessary condition, we then examine the feasibility of topology learning for two useful classes of problems. The analysis reveals that a critical element to enable topology learning is a sufficient degree of *diversity* in the statistical models of the sending sub-networks.

INDEX TERMS Social learning, topology learning, weakly-connected networks, Bayesian update, diffusion strategy.

I. INTRODUCTION

In a social learning problem, several agents linked through a network topology form their individual opinions about a phenomenon of interest (learning process) by observing the beliefs of their neighboring agents (social interaction) [3]–[10]. One relevant paradigm for social learning is that of *weakly-connected* graphs, which are prevalent in social networks and have been studied in [11] and [12]. Under this model, there are two categories of sub-networks: the *sending* sub-networks, which feed information to the *receiving* sub-networks without getting back information from them. This scenario is common over social networks. For example, a celebrity may have a large number of followers, whose individual opinions are not necessarily followed by the celebrity. Another example is that of media networks, which promote the emergence of opinions by feeding data to users without paying attention to feedback from them.

This work addresses two fundamental challenges arising in the study of social learning problems. One challenge is to understand the fundamental mechanism and implications of specific social learning strategies on opinion formation. In particular, over weak graphs, it is critical to understand how the receiving agents are influenced by the sending sub-networks. It is not difficult to envisage that the network topology plays an important role in determining the opinion formation. This motivates the second problem, which can be regarded as a *dual* learning problem. Given observation of the receiving agents' behavior, we want to establish whether it is possible

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

to learn the strength of connections (weighted topology) from the sending components to the receiving agents. This second question is useful in identifying the main sources for opinion formation over a network.

A. RELATED WORK

There exist several useful paradigms and strategies for distributed decision-making over multi-agent networks — see [13]–[20] for a non-comprehensive list. One notable paradigm is that of social learning.

The goal of social learning is to let each individual agent create its own opinion (formally, choose one from among a finite set of hypotheses) through local consultation with its neighbors. One classical distinction for social learning models is Bayesian vs. non-Bayesian models. In the former category, the belief of each agent is obtained by computing posterior distributions through Bayes' rule [3]–[5], [21]–[24]. In order to accomplish this task, each agent must have some detailed knowledge of the other agents' likelihoods, and some prior knowledge about the phenomenon of interest. In the latter category, this level of knowledge is not assumed and the agents implement suitable distributed algorithms to interact with their neighbors and to aggregate their beliefs into their own [10]–[12], [25]–[31]. The present work considers non-Bayesian learning.

There are many useful implementations for non-Bayesian learning. The implementations differ in terms of the rule the agents adopt to update their beliefs. One first distinction concerns the type of distributed strategy. In particular, we can distinguish between consensus [28], [29] or diffusion [10]–[12], [30], [31] implementations. A second distinction regards the way the beliefs are combined. In [10], [12], [28], they are combined *linearly*, while a linear combination of the *logarithmic* beliefs is used in [30], [31]. This latter form is motivated by the fact that, in many detection problems, the best detection statistic is given by the linear combination of log-likelihoods and not of likelihoods. As a matter of fact, using the log-belief combination can help achieve an improved (i.e., faster) learning rate [31].

Once a particular learning rule has been chosen, the behavior of opinion formation will depend heavily on the type of network where the information propagates. In this respect, the majority of prior works in the literature focus on *strongly-connected* networks. In these networks, there always exist paths linking any two agents in both directions (which makes them connected) and, in addition, at least one agent has a self-loop and places some partial trusts in its own data (which makes them *strongly* connected). Under a *homogeneous* setting where the underlying true hypothesis is the same for the entire network, it has been shown that over strongly-connected networks all agents are able to discover and agree on the true hypothesis. This result is available in [10], [28] for diffusion and consensus rules with linear belief combination, and in [31] for the diffusion rule with linear log-belief combination. There are results available also for the *heterogeneous* setting, where different agents might have different data models, different likelihoods, and promote different opinions across the network. In particular, the diffusion rule with linear log-belief combination with a *doubly-stochastic* combination matrix is considered in [30], where it is shown that, over a strongly-connected network, all agents reach a common opinion, by minimizing cooperatively the sum of Kullback-Leibler (KL) divergences across the network.

In contrast, the important case of *weakly-connected* networks has received limited attention and was addressed more recently in [11], [12] by using the linear-belief-combination rule. Several interesting phenomena arise over weak graphs, which are absent from strongly-connected networks. The main difference in this work from [11], [12] is that we now consider the following general setting: *i*) *diffusion-type* algorithms with linear combination of *log-beliefs* (as opposed to the beliefs themselves), *ii*) *heterogeneous* data and inference model, and *iii*) *inverse* topology learning.

B. MAIN RESULTS

This work leads to two main contributions. First, we characterize (Theorem 1) the limiting (as learning time elapses) agents' belief through analytical formulas that depend in a transparent manner on inferential descriptors (Kullback-Leibler divergences) and network descriptors (limiting combination matrix power). Some revealing behaviors can be deduced from these formulas. For example, we will be able to characterize a *mind-control* effect in terms of the interplay between the *detection capacity* of each agent and the *centrality* of the different agents. We will see that useful analogies arise with what has been observed in [12]. For example, some of the effects shown in [12] will now be shown to hold under greater generality since our formulas can be obtained by relaxing some assumptions used in [12], in particular, the *all-truths-are-equal* assumption, and the *prevailing-signal* assumption. However, we will observe also some remarkable distinctions with respect to [12]. For instance, we will observe that the individual belief of each receiving agent will necessarily collapse to (or concentrate at) a *single hypothesis* (which might be different across the agents). This is in contrast with [12], where the beliefs of the receiving agents could end up assigning some probabilities to more than one hypothesis. The reason for this difference in behavior arises from the difference in the combination rule used in this work in comparison to [12].

The second contribution concerns the *topology learning* problem. We are interested to learn the topology linking the receiving agents to the sending components. This question is interesting because it would then allow us to identify the main sources of information in a network and how they influence opinion formation. Nevertheless, the inverse topology problem is challenging because we will assume that we can only observe the beliefs evolving at the receiving agents. In particular, we will be able to recover some macroscopic topology information, in terms of the limiting weights that each receiving agent sees from each sending component. We call this a macroscopic information since these weights incorporate: *i*)

the global effect coming from *all* agents belonging to a sending component, and *ii*) the effect of intermediate receiving agents linked to the receiving agent under consideration. The relevance in estimating these global weights relies on the fact that the limiting beliefs of the receiving agents depend solely on this aggregate information.

We will establish conditions under which topology inference becomes feasible. More specifically, given H hypotheses and S sending components, under the assumption of homogeneous statistical models within each sending component, we will ascertain that a necessary condition to achieve consistent topology learning is (Lemma 1):

$$\boxed{H \geq S} \tag{1}$$

Once established a necessary condition, we will examine some useful models to see whether topology learning *can be* in fact achieved. We consider first a *structured* Gaussian model where: *i*) the true underlying (Gaussian) distributions are distinct across the sending sub-networks; and *ii*) the (Gaussian) likelihoods are equal across the sending sub-networks, and contain the true distributions. For this setting, we will show in Theorem 2 that topology learning is feasible only when $S = 2$. We then recognize that one fundamental element for topology learning is the *diversity* between the sending sub-networks. Adding this further element, we will establish in Theorem 3 that the problem is feasible for *any* S provided that (1) holds, and even under more general (e.g., non-Gaussian) models.

In summary, we remark that there are two learning problems coexisting in our work: a social learning problem and a topology learning problem. The former is the primary, or direct inferential problem for which the agents are deployed. The latter is the *dual*, or *reverse* problem, which is in fact based on observation of the output (the beliefs) of the primary learning problem. One useful conclusion stemming from our analysis is to reveal some unexpected interplay between these two coexisting learning problems — see Section VIII further ahead.

Notation: We use bold font notation for random variables, normal fonts for their realizations. Capital letters are generally used for matrices, whereas small font letters for vectors or scalars. Given a matrix M, the symbol $M^{\dagger}$ denotes its Moore-Penrose pseudoinverse. The $L \times L$ identity matrix is denoted by I_L. Likewise, the $L \times 1$ vector of all ones is denoted by $\mathbb{1}_L$. The notation "a.s." signifies "almost-surely".

II. BACKGROUND

A. DATA MODEL AND INFERENCE

We consider a network of N agents that collect streaming data from the environment. Formally, the random variable $\boldsymbol{\xi}_{k,i} \in \mathcal{X}_k$ denotes the data at agent $k \in \{1, 2, \ldots, N\}$ collected at time $i \in \mathbb{N}$. The data are assumed to be independent over time, whereas they can be dependent[1] across the agents (i.e., over space).

We work under a *heterogeneous* setting. First, the space where the data are defined, $\mathcal{X}_k$, is allowed to vary across the agents. For example, the data at different agents can be random vectors of different sizes. In the social learning context, this scenario is not that uncommon since different types of attributes can be observed by different users across the network. Second, the data $\boldsymbol{\xi}_{k,i}$ are assumed to be generated according to certain distributions $f_k(\xi)$, which are allowed to vary across the agents as well, namely, for $k = 1, 2, \ldots, N$:

$$\boxed{\boldsymbol{\xi}_{k,i} \sim f_k \quad \text{[true distribution]}} \tag{2}$$

This type of heterogeneity can arise in practice for several reasons, for example, some agents may intentionally inject fake data to let the other agents have a bias towards fake hypotheses.

Based on the available data $\{\boldsymbol{\xi}_{k,i}\}$, the network agents aim to solve an inferential problem that consists in choosing one state of nature from among a finite collection $\Theta = \{1, 2, \ldots, H\}$, with H denoting the number of possible hypotheses. To solve such inferential problem, the agents rely on a family of *likelihood functions*. Specifically, for $k = 1, 2, \ldots, N$, the likelihood function of agent k is denoted by:

$$\boxed{L_k(\xi|\theta), \quad \xi \in \mathcal{X}_k, \ \theta \in \Theta \quad \text{[likelihoods]}} \tag{3}$$

We will often write $L_k(\theta)$ instead of $L_k(\xi|\theta)$ for simplicity. In our treatment, we assume that the data can be modeled either as continuous or discrete random variables, with the same (continuous or discrete) nature across all agents. Accordingly, both the true distributions and the likelihoods will be either probability density or probability mass functions, depending on the considered type of random variables.

We remark that the considered model includes the possibility that the true distribution $f_k(\xi)$ is equal to a certain likelihood $L_k(\xi|\theta_0)$ (as assumed, e.g., in [12], [31]) and in this case θ_0 can be considered as the true underlying hypothesis. More generally (e.g., in [30]) the true distribution $f_k(\xi)$ is not equal to any of the likelihoods $L_k(\xi|\theta)$, which might happen when the agents have an approximate knowledge of the statistical models, and even the likelihood more similar to the true distribution can be a mismatched version thereof. In order to quantify the dissimilarity between the true distribution and a certain likelihood, we will use the Kullback-Leibler (KL) divergence [32]:

$$D[f_k||L_k(\theta)] \triangleq \mathbb{E}_{f_k}\left[\log \frac{f_k(\boldsymbol{\xi}_k)}{L_k(\boldsymbol{\xi}_k|\theta)}\right], \quad k = 1, 2, \ldots, N. \tag{4}$$

We remark that we have written $\boldsymbol{\xi}_k$ in place of $\boldsymbol{\xi}_{k,i}$ to highlight identical distributions across time, and that the expectation is

[1] In social learning, it is usually assumed that no agent can know or model the N-dimensional (i.e., across the agents) joint distribution of the data. In contrast, the agents use only the *marginal* likelihood of $\boldsymbol{\xi}_{k,i}$, and only the marginal distribution of $\boldsymbol{\xi}_{k,i}$ will matter in the analysis [10], [12], [30], [31].

computed under the actual distribution of $\boldsymbol{\xi}_k$, i.e., under f_k. In the forthcoming treatment we use the following regularity condition [30], [31].

Assumption 1 (Finiteness of KL Divergences): All the KL divergences in (4) are well-posed, namely, for all $k = 1, 2, \ldots, N$, and all $\theta \in \Theta$ we have that:

$$\boxed{D[f_k || L_k(\theta)] < \infty} \tag{5}$$

□

Remark 1 (Likelihood Support): Assumption 1 implies that the likelihood $L_k(\boldsymbol{\xi}_{k,i}|\theta)$ can be equal to zero only for an ensemble of realizations $\boldsymbol{\xi}_{k,i}$ having zero probability under the distribution f_k. □

B. SOCIAL LEARNING ALGORITHM

Motivated by the diffusion strategy in [10], [12] for opinion formation over social networks, in this article we consider the useful variations studied in [30], [31]. The learning procedure employs a two-step algorithm that can be described as follows. For each admissible hypothesis $\theta \in \Theta$, at time i, each agent k uses its own fresh *private* observation, $\boldsymbol{\xi}_{k,i}$, to compute the local likelihood $L_k(\boldsymbol{\xi}_{k,i}|\theta)$. Using this likelihood, agent k updates its local belief, $\boldsymbol{\mu}_{k,i-1}(\theta)$, obtaining an intermediate belief $\boldsymbol{\psi}_{k,i}(\theta)$ through the following update rule:

$$\boxed{\boldsymbol{\psi}_{k,i}(\theta) = \frac{\boldsymbol{\mu}_{k,i-1}(\theta) L_k(\boldsymbol{\xi}_{k,i}|\theta)}{\sum_{\theta' \in \Theta} \boldsymbol{\mu}_{k,i-1}(\theta') L_k(\boldsymbol{\xi}_{k,i}|\theta')}} \tag{6}$$

Then, agent k aggregates the intermediate beliefs received from its neighbors through the following combination rule (the division by the denominator term in (7) is meant to ensure that $\boldsymbol{\mu}_{k,i}(\theta)$ is a probability measure with its entries adding up to one):

$$\boxed{\boldsymbol{\mu}_{k,i}(\theta) = \frac{\exp\left\{\sum_{\ell=1}^{N} a_{\ell k} \log \boldsymbol{\psi}_{\ell,i}(\theta)\right\}}{\sum_{\theta' \in \Theta} \exp\left\{\sum_{\ell=1}^{N} a_{\ell k} \log \boldsymbol{\psi}_{\ell,i}(\theta')\right\}}} \tag{7}$$

where $a_{\ell k}$ is the *nonnegative combination weight* that agent k uses to scale the intermediate log-belief received from agent ℓ. It is assumed that $a_{\ell k}$ is equal to zero if k does not receive information from ℓ, which means that agent k can combine only intermediate beliefs received from its neighbors. When collected into a *combination matrix* A (with (ℓ, k) entry equal to $a_{\ell k}$), these combination weights are assumed to obey the standard requirements that make A a left-stochastic matrix, namely, we have that:

$$A^{\top} \mathbb{1}_N = \mathbb{1}_N \Leftrightarrow \sum_{\ell=1}^{N} a_{\ell k} = 1, \quad \forall k = 1, 2, \ldots, N. \tag{8}$$

The value of $\boldsymbol{\mu}_{k,i}(\theta)$ provides an estimate for the likelihood by agent k at time i that the true hypothesis value is θ. We remark that, differently from [12], the second step in (7) combines linearly the *logarithm* of the intermediate beliefs, $\log \boldsymbol{\psi}_{\ell,i}(\theta)$, in the neighborhood of agent k. Exponentiation and normalization are used to construct the final belief.

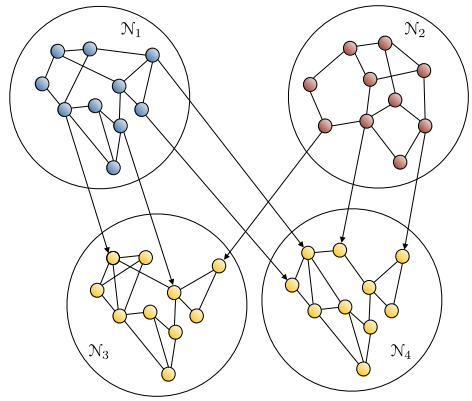

FIGURE 1. One example of weakly-connected network, with sending sub-networks $\mathcal{N}_1$ and $\mathcal{N}_2$, and receiving sub-networks $\mathcal{N}_3$ and $\mathcal{N}_4$.

We invoke the following standard initial condition, motivated by the fact that, at time $i = 0$, the agents have no elements to discard any hypothesis [30], [31].

Assumption 2 (Initial Beliefs): All agents start by assigning strictly positive probability mass to all hypotheses, namely, $\mu_{k,0}(\theta) > 0$ for all $\theta \in \Theta$ and $k = 1, 2, \ldots, N$. □

Remark 2 (Strict Positiveness of Beliefs): From (6) we see that, if $\boldsymbol{\mu}_{k,i-1}(\theta) > 0$, then $\boldsymbol{\psi}_{k,i}(\theta) > 0$ since $L_k(\boldsymbol{\xi}_{k,i}|\theta) > 0$ (but for zero-probability sets) — see Remark 1. Strict positiveness of $\boldsymbol{\mu}_{k,i}(\theta)$ now follows from (7) since the combination weights are nonnegative and convex. □

III. WEAK GRAPHS

In this section, we consider the case of a weak graph (or weakly-connected network), which is defined as follows [11], [12]. The overall network $\mathcal{N} = \{1, 2, \ldots N\}$ is divided into $S + R$ disjoint components — see Fig. 1. The first S sub-networks form the *sending* part $\mathcal{S}$, whereas the remaining R sub-networks form the *receiving* part $\mathcal{R}$:

$$\mathcal{N} = \mathcal{S} \cup \mathcal{R}, \qquad \mathcal{S} \triangleq \bigcup_{s=1}^{S} \mathcal{N}_s, \qquad \mathcal{R} \triangleq \bigcup_{r=1}^{R} \mathcal{N}_{S+r}. \tag{9}$$

We remark that S and R denote the number of sending and receiving *components*, respectively. They do *not* denote the cardinalities of $\mathcal{S}$ and $\mathcal{R}$, which are instead given by:

$$|\mathcal{S}| = \sum_{s=1}^{S} N_s, \qquad |\mathcal{R}| = \sum_{r=1}^{R} N_{S+r}, \tag{10}$$

where the notation N_j denotes the number of agents in sub-network $\mathcal{N}_j$. Communication from a sending component to a receiving component is permitted, whereas communication in the reverse direction is forbidden. Communication between sending agents is possible, but a sending sub-network is identified by the following two conditions: *i*) each sending sub-network is *strongly* connected [33]; and *ii*) agents belonging

to different sending sub-networks do not communicate with each other. If some sending sub-networks communicate with each other, then they can be blended into a larger sending sub-network. The R receiving sub-networks are all individually assumed to be *connected* (but not necessarily *strongly* connected), with communication among the R sub-networks being allowed. In particular, we assume that each receiving sub-network is connected to at least one agent in each sending sub-network.

Under the weak-graph abstraction, a sending sub-network evolves in isolation. A similar behavior is observed, for example, over social networks where an influential agent scarcely consults information from his/her followers; or in the presence of stubborn agents, which insist on their opinion regardless of the evidence provided by their own observations or by neighboring agents. In light of these features, weak-graphs provide a convenient mathematical model to approximate the behavior of some real networks. On the other hand, the model can be complemented with additional features. For example, one could assume some limited feedback from receiving to sending agents, or some one-way influence among sending sub-networks.

Without loss of generality, we assume that the network nodes across the $S+R$ components are listed in increasing order. According to the above description, the combination matrix corresponding to a weakly-connected network admits the following convenient block decomposition [11], [12]:

$$A = \left[\begin{array}{c|c} A_{\mathcal{S}} & A_{\mathcal{SR}} \\ \hline 0 & A_{\mathcal{R}} \end{array}\right], \tag{11}$$

where the matrix $A_{\mathcal{S}} = \mathsf{blockdiag}\{A_{\mathcal{N}_1}, A_{\mathcal{N}_2}, \ldots, A_{\mathcal{N}_S}\}$ contains the combination weights within the sending sub-networks, and has a block-diagonal form since communication between sending sub-networks is forbidden. The matrix block $A_{\mathcal{SR}}$ contains the combination weights for the communication that takes place *from sending* agents *to receiving* agents. The left-bottom matrix block is zero since there are no direct links from the receiving agents to the sending ones. Finally, the matrix block $A_{\mathcal{R}}$ contains the combination weights ruling the communication among the receiving agents. Figure 1 offers a graphical illustration of the weakly-connected paradigm.

It was shown in [11] that the limiting combination matrix power has the following structure:

$$A_\infty \triangleq \lim_{i\to\infty} A^i = \left[\begin{array}{c|c} E & EW \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} E & \Omega \\ \hline 0 & 0 \end{array}\right], \tag{12}$$

where

$$E = \mathsf{blockdiag}\left\{p^{(1)}\mathbb{1}_{N_1}^\top, p^{(2)}\mathbb{1}_{N_2}^\top, \ldots, p^{(S)}\mathbb{1}_{N_S}^\top\right\} \tag{13}$$

is a block diagonal matrix that stacks the $N_s \times 1$ Perron eigenvectors $p^{(s)}$ associated with the s-th sending sub-network,[2] and where

$$W = A_{\mathcal{SR}}\left(I_{|\mathcal{R}|} - A_{\mathcal{R}}\right)^{-1}, \quad \Omega = EW. \tag{15}$$

We denote the entries of Ω by $[\omega_{\ell k}]$ and we keep indexing the columns of the $|\mathcal{S}| \times |\mathcal{R}|$ matrix Ω with an index

$$k \in \{|\mathcal{S}|+1, |\mathcal{S}|+2, \ldots, |\mathcal{S}|+|\mathcal{R}|\}. \tag{16}$$

Since the limiting matrix power is left-stochastic and has a zero bottom block, the limiting weights $\omega_{\ell k}$ obey:

$$\sum_{\ell\in\mathcal{S}} \omega_{\ell k} = 1, \tag{17}$$

i.e., Ω is left-stochastic. From (15) we can also write:

$$\Omega = EA_{\mathcal{SR}}(I_{|\mathcal{R}|} + A_{\mathcal{R}} + A_{\mathcal{R}}^2 + \ldots), \tag{18}$$

whence we see that $\omega_{\ell k}$ embodies the sum of influences over all paths from sending agent ℓ to receiving agent k.

IV. LIMITING BELIEFS OF RECEIVING AGENTS

Let us momentarily consider a single-agent scenario where agent ℓ operates alone. A natural way for agent ℓ to choose a hypothesis would be to choose the θ that gives the best match between a model $L_\ell(\theta)$ and the distribution of the observed data, f_ℓ. One measure of the match between f_ℓ and $L_\ell(\theta)$ is the KL divergence $D[f_\ell||L_\ell(\theta)]$. The smaller the value of this divergence is, the higher the match between the data and the model. For this reason, a strategy could be that of choosing the θ that minimizes the divergence $D[f_\ell||L_\ell(\theta)]$.

In the social learning context, this optimization problem turns into a *distributed* optimization problem. In particular, under our social learning setting over weak graphs, we will show soon (Theorem 1) that the log-belief diffusion strategy in (6)–(7) will end up minimizing (without knowing the true distributions) the following *average* divergence at *receiving* agent $k \in \mathcal{R}$:

$$\boxed{\mathscr{D}_k(\theta) \triangleq \sum_{\ell\in\mathcal{S}} \omega_{\ell k} D[f_\ell||L_\ell(\theta)]} \tag{19}$$

which is a weighted combination, through the limiting combination weights $\{\omega_{\ell k}\}$, of the KL divergences of the *sending* agents reaching k. The role of average divergence measures like the one in (19) already arose in the case of strongly-connected networks. For example, it was shown in [30], [31] that with the log-belief diffusion strategy in (6)–(7), each agent ends up minimizing the *same* weighted combination of divergences. Under classical identifiability conditions, such minimization leads each individual agent to discover the true underlying hypothesis [31] or the best available approximation thereof [30]. In our weak-graph setting, however, the effect of minimizing $\mathscr{D}_k(\theta)$ (which depends on the particular receiving agent k) will be less obvious. We already see from (19) that the average divergence combines *topological* attributes, encoded in the limiting combination weights, with *inferential attributes*, encoded in the local KL divergences. The interplay arising between the network topology and social learning will be critical in determining the choices of the receiving agents.

[2]For $s = 1, 2, \ldots, S$, the Perron eigenvector of the sub-matrix $A_{\mathcal{N}_s}$ corresponding to the s-th sending sub-network is given by:

$$A_{\mathcal{N}_s} p^{(s)} = p^{(s)}, \;\; \mathbb{1}_{N_s}^\top p^{(s)} = 1, \;\; p_\ell^{(s)} > 0, \;\; \ell = 1, 2, \ldots, N_s. \tag{14}$$

Throughout the work, we will invoke the following classical identifiability assumption that, as we will see in our examples, arises naturally in several models of interest.

Assumption 3 (Unique Minimizer): For each $k = 1, 2, \ldots, N$, the function $\mathscr{D}_k(\theta)$ has a unique minimizer:

$$\boxed{\theta_k^\star \triangleq \operatorname*{argmin}_{\theta\in\Theta} \mathscr{D}_k(\theta)} \tag{20}$$

□

We are now ready to characterize the limiting belief of the receiving agents. The following theorem is an extension to the case of weakly-connected graphs of similar theorems proved in [30], [31] for the case of strongly-connected graphs.

Theorem 1 (Belief Collapse at Receiving Agents): Let $k \in \mathcal{R}$. Under Assumptions 1–3 we have that:

$$\boxed{\lim_{i\to\infty} \boldsymbol{\mu}_{k,i}(\theta_k^\star) \overset{\text{a.s.}}{=} 1} \tag{21}$$

where the symbol $\overset{\text{a.s.}}{=}$ denotes that the pertinent limit exists almost-surely. Moreover, for all $\theta \neq \theta_k^\star$, the convergence of the belief to zero takes place at an exponential rate as:

$$\boxed{\lim_{i\to\infty} \frac{\log \boldsymbol{\mu}_{k,i}(\theta)}{i} \overset{\text{a.s.}}{=} \mathscr{D}_k(\theta_k^\star) - \mathscr{D}_k(\theta)} \tag{22}$$

Proof: The proof combines the techniques to establish the convergence of the social learning algorithm used, e.g., in [30], [31] for strongly-connected graphs, with the convergence results of the combination matrix over weak graphs used in [11], [12]. The detailed steps are reported in Appendix A. ■

Several insightful conclusions arise from Theorem 1.

Remark 3 (Collapse): The limiting belief *of each receiving agent* is always degenerate, meaning that it collapses to a single hypothesis, when sufficient time for learning is allowed. □

Remark 4 (Discord): Different agents can in principle be in *discord*, since they can converge to different hypotheses. The particular behavior (who chooses what) will depend on a weighted combination of KL divergences. □

Remark 5 (Mind Control): We see from (19) and (20) that only the local divergences corresponding to the sending agents, $\ell \in \mathcal{S}$, determine the value of $\mathscr{D}_k(\theta)$ and, hence, of $\theta_k^\star$. Therefore, the limiting hypothesis $\theta_k^\star$ at agent k is determined by the KL divergences pertaining only to the statistical models *within the sending sub-networks*, and, hence, *irrespective of the data sensed at agent k within its receiving sub-network.* In a nutshell, we see the emergence of a mind-control effect: *i*) the final states of the receiving agents are dependent only upon the properties of the detection problems at the *sending* agents; and *ii*) different network topologies allow the sending agents to drive the receiving agents to potentially different decisions. The emergence of a mind-control effect over weakly-connected networks was already discovered in [11], [12]. Here, we establish a similar effect albeit one where the receiving agents attain degenerate beliefs. In comparison, in [11], [12], receiving agents end up assigning nonzero probabilities to more than one belief. □

Remark 6 (Distinctions relative to [12]): Besides these commonalities, there are nevertheless important distinctions between the behavior observed in our setting and what was observed in [12]. First, for the linear-belief-combination algorithm used in [12], the limiting belief of a receiving agent was shown to be a convex combination of the limiting beliefs of the sending agents, with the convex weights coming from the matrix W in (15). This means that if two sending sub-networks have, e.g., limiting beliefs that collapse to different hypotheses, then the limiting belief of a receiving agent can have nonzero values at these two different locations. In comparison, in the log-belief-combination algorithm considered here the limiting beliefs are always concentrated at a *single* hypothesis.

Moreover, in [12], the analysis required some regularity assumptions called *all-truths-are-equal* and *prevailing-signal* assumptions. These assumptions are not required in Theorem 1. In a sense, the lack of these assumptions ascertains that some relevant effects, such as mind control, hold under greater generality and more relaxed settings.

Finally, it is useful to observe that one fundamental role in our setting is played by the weighting matrix $\Omega = EW$, while in [12] the main role was played by W alone. □

A. CANONICAL EXAMPLES

In order to examine in more detail the implications of Theorem 1, we consider a simple yet insightful example. The sending and receiving components are:

$$\mathcal{S} = \mathcal{N}_1 \cup \mathcal{N}_2, \qquad \mathcal{R} = \mathcal{N}_3, \tag{23}$$

namely, we have two sending sub-networks, $\mathcal{N}_1$ and $\mathcal{N}_2$, and one receiving sub-network $\mathcal{N}_3$.

For what concerns the inferential model, we assume there are three possible hypotheses, $\theta \in \{1, 2, 3\}$. The likelihood functions are the same across *all* agents. In particular, we assume that, for all $\xi \in \mathbb{R}$, and for $\theta \in \{1, 2, 3\}$:

$$L(\xi|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\xi - \mathsf{m}_\theta)^2}{2}\right\}, \tag{24}$$

where the means corresponding to the different hypotheses are chosen as, for some $\Delta > 0$:

$$\mathsf{m}_1 = -\Delta,\ \mathsf{m}_2 = 0,\ \mathsf{m}_3 = +\Delta. \tag{25}$$

We further assume that the *true* distributions of the sending sub-networks (recall that only the sending sub-networks determine the limiting beliefs of the receiving sub-network) are Gaussian distributions, with expectations chosen among the expectations in (25). In particular, we assume that agents belonging to sub-network $\mathcal{N}_1$ generate data according to model $\theta = 1$, i.e., with expectation equal to $-\Delta$, whereas agents belonging to sub-network $\mathcal{N}_2$ generate data according to model

$\theta = 3$, i.e., with expectation equal to $+\Delta$. Formally we write:

$$f_\ell(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\xi+\Delta)^2}{2}\right\}, \quad \forall \ell \in \mathcal{N}_1, \tag{26}$$

$$f_\ell(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\xi-\Delta)^2}{2}\right\}, \quad \forall \ell \in \mathcal{N}_2. \tag{27}$$

Recalling that the KL divergence between two unit-variance Gaussian distributions of expectations a and b is given by $0.5(a-b)^2$, under the setting described above we can write, for all $k \in \mathcal{R}$:

$$\begin{aligned}\mathscr{D}_k(\theta) &= \sum_{\ell\in\mathcal{S}} \omega_{\ell k} D[f_\ell || L_\ell(\theta)] \\ &= \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} D[f_\ell || L(\theta)] + \sum_{\ell\in\mathcal{N}_2} \omega_{\ell k} D[f_\ell || L(\theta)] \\ &= \frac{(-\Delta-\mathsf{m}_\theta)^2}{2} \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} + \frac{(\Delta-\mathsf{m}_\theta)^2}{2} \sum_{\ell\in\mathcal{N}_2} \omega_{\ell k},\end{aligned} \tag{28}$$

which further implies:

$$\begin{aligned}\mathscr{D}_k(1) &= 2\Delta^2 \sum_{\ell\in\mathcal{N}_2} \omega_{\ell k}, \\ \mathscr{D}_k(2) &= \frac{\Delta^2}{2}, \\ \mathscr{D}_k(3) &= 2\Delta^2 \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k},\end{aligned} \tag{29}$$

where, in the intermediate equality, we used (17). As a result, we can compute the limiting hypothesis, for each $k \in \mathcal{R}$, as:

$$\theta_k^\star = \operatorname{argmin}\left\{4\sum_{\ell\in\mathcal{N}_2} \omega_{\ell k},\ 1,\ 4\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k}\right\}. \tag{30}$$

From (18), one can argue that $\sum_{\ell\in\mathcal{N}_s} \omega_{\ell k}$ reflects the sum of influences over *all* paths connecting *all* sending agents in sub-network s to receiving agent k.

In order to find the minimizer in (30), we start by using (17) in (30), which yields:

$$\theta_k^\star = \operatorname{argmin}\left\{1-\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k},\ 0.25,\ \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k}\right\}. \tag{31}$$

In view of Theorem 1, the belief of the k-th receiving agent will converge to $\theta_k^\star = 1$ if the following two conditions are simultaneously verified:

$$\begin{aligned}1-\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} < 0.25 \quad &\Leftrightarrow \quad \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} > 0.75, \\ 1-\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} < \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} \quad &\Leftrightarrow \quad \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} > 0.5.\end{aligned} \tag{32}$$

Taking the most stringent condition in (32) reveals that:

$$\theta_k^\star = 1 \Leftrightarrow \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} > 0.75. \tag{33}$$

In summary, we conclude that agent k follows the opinion promoted by sending sub-network $\mathcal{N}_1$ if the influence of sub-network $\mathcal{N}_1$ on agent k is "sufficiently large".

The situation is reversed if the influence of sub-network $\mathcal{N}_2$ is sufficiently large, namely,

$$\theta_k^\star = 3 \Leftrightarrow \sum_{\ell\in\mathcal{N}_2} \omega_{\ell k} > 0.75, \tag{34}$$

where we recall that hypothesis $\theta = 3$ is promoted by sub-network $\mathcal{N}_2$. However, there is another possibility. It occurs when:

$$\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k} < 0.75 \ \text{ and } \ \sum_{\ell\in\mathcal{N}_2} \omega_{\ell k} < 0.75. \tag{35}$$

In this case, no clear dominance from one sub-network can be ascertained, and each receiving agent will choose $\theta_k^\star = 2$, i.e., an *opinion that does not coincide with any of the opinions promoted by the sending sub-networks*.

From (33) and (34), we see that the dominance of one of the sending sub-networks is determined by the aggregate influence $\sum_{\ell\in\mathcal{N}_1} \omega_{\ell k}$, with the complementary aggregate influence being $\sum_{\ell\in\mathcal{N}_2} \omega_{\ell k} = 1 - \sum_{\ell\in\mathcal{N}_1} \omega_{\ell k}$. The main way to manipulate these factors consists in varying the *sizes* of the sending sub-networks or their connections with the receiving agents.

In order to illustrate more carefully the possible scenarios, we consider the following simulation framework:

- The strongly-connected sending components $\mathcal{N}_1$ and $\mathcal{N}_2$ are generated as Erdős-Rényi random graphs with connection probability q, and the entries of the corresponding combination matrix are determined by the averaging rule, namely,[3]

$$a_{\ell k} = \begin{cases} 1/n_k, & \text{if } k \neq \ell \text{ are neighbors or } k = \ell \\ 0, & \text{otherwise} \end{cases} \tag{36}$$

  where n_k is the number of neighbors of node k (including node k itself). In our experiments we set $q = 0.7$.
- An agent k is connected to a sending agent through a Bernoulli distribution with parameter π_s, which depends on the sending sub-network s. Given the total number d_k, of directed edges from sending agents to agent k, we initially set $a_{\ell k} = 1/d_k$. The combination matrix A of the overall network $\mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3$ is normalized so that it is left-stochastic.

It is now possible to examine different scenarios by manipulating the size of the sending sub-networks as well as the send-receive connection probabilities π_s.

— Setup 1 or "*How majorities build a majority*". In Fig. 2, we set $\pi_1 = \pi_2 = 0.5$, i.e. it is equally probable that a receiving agent connects to any sending agent, irrespective of the sending sub-network. In view of this uniformity, we can expect that the limiting weights $\omega_{\ell k}$ are sufficiently uniform across the two sending sub-networks and, hence, that the value

[3]When drawing the random graph, we have verified that there exists at least one self-loop.

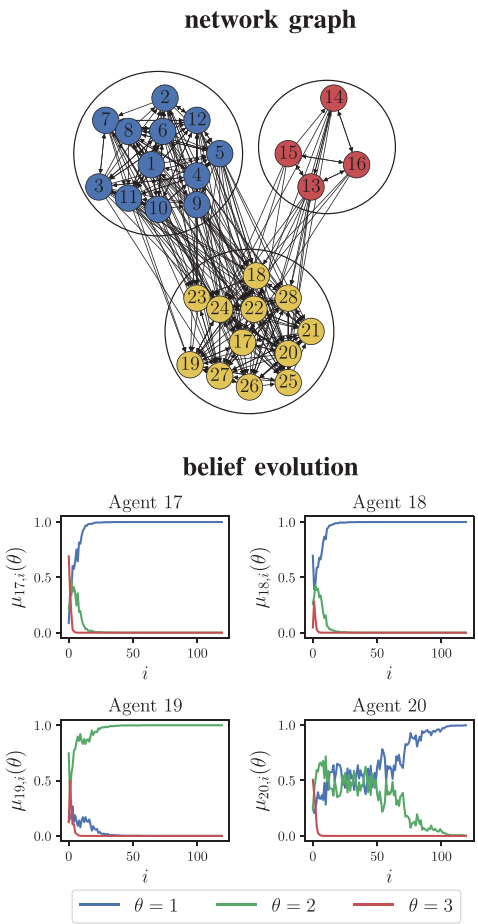

FIGURE 2. ***How majorities build a majority*: Convergence of beliefs when the size of sending sub-network $\mathcal{N}_1$ is dominant.**

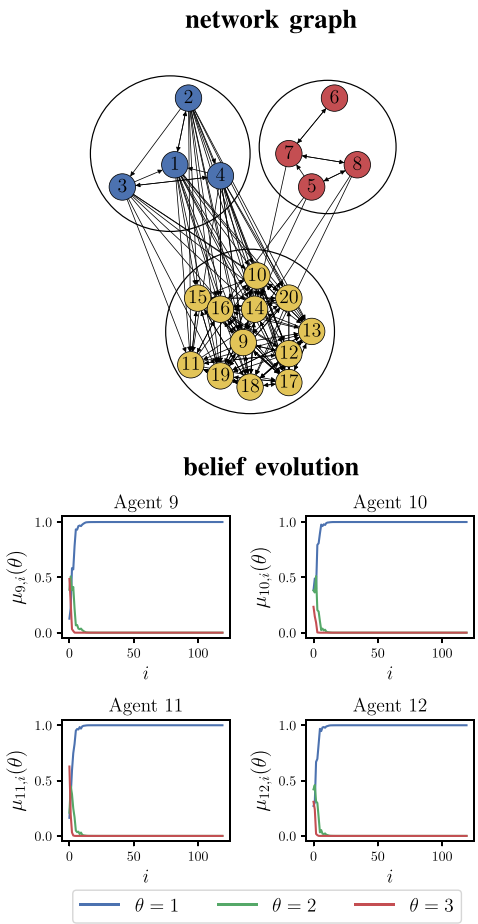

FIGURE 3. ***How filter bubbles build a majority*: Convergence of beliefs when the connectivity from sending sub-network $\mathcal{N}_1$ is dominant.**

of $\sum_{\ell \in \mathcal{N}_1} \omega_{\ell k}$ is primarily determined by the sub-network size N_1. In the example we are going to illustrate, we assume that the number of agents in sub-network $\mathcal{N}_1$ is three times larger than the size of sub-network $\mathcal{N}_2$. For clarity of visualization, we display only the belief of four receiving agents. From the lowermost panel in Fig. 2, we observe that receiving agents (17, 18, 20) converge to $\theta = 1$, i.e., to the opinion promoted by $\mathcal{N}_1$. We see also that agent 19 takes a minority position and opts for $\theta = 2$, i.e., it does follow neither the opinion promoted by $\mathcal{N}_1$ nor by $\mathcal{N}_2$. This shows the following interesting effect. Even if sub-network $\mathcal{N}_1$ is bigger, for the specific topology shown in the example (see the uppermost panel of Fig. 2), the aggregate weight of agent 19 is $\sum_{\ell \in \mathcal{N}_1} \omega_{\ell\, 19} = 0.6885$. This means that condition (35) is actually verified, which explains why agent 19 opts for $\theta = 2$. Including in the analysis also the agents that are not displayed, in this example we have that 2/3 of the receiving agents in the network opt for $\theta = 1$. In summary, we observed that building a majority of agents in $\mathcal{N}_1$ relative to $\mathcal{N}_2$ yields a majority of receiving agents opting for the hypothesis promoted by $\mathcal{N}_1$.

— Setup 2 or "*How filter bubbles build a majority*". Under this setup, we assume that both sending components have the same size, however π_s is different for each of the two components. We set $\pi_1 = 0.9$ and $\pi_2 = 0.1$ in order to motivate agent k to have more connections with sub-network $\mathcal{N}_1$ than with $\mathcal{N}_2$. This scenario is considered in Fig. 3, where we see that the displayed receiving agents end up agreeing with opinion $\theta = 1$, i.e., with the opinion promoted by the sending component $\mathcal{N}_1$. Including in the analysis also the receiving agents that are not displayed, in this example we have that *all* the receiving agents in the network opt for $\theta = 1$. Therefore, closing a receiving agent into the "*filter bubble*" determined by the overwhelming flow of data coming from $\mathcal{N}_1$ essentially makes these agents blind to the solicitations coming from $\mathcal{N}_2$.

— Setup 3 or "*Truth is somewhere in between*". We now address the balanced case where the sending sub-networks have the same size and similar number of connections to the receiving sub-network ($\pi_1 = \pi_2 = 0.5$). Under this setting, it is expected that no dominant behavior emerges, and (35) holds. We see in Fig. 4 that the opinions of receiving agents (9, 10, 11, 12) tend to converge with full confidence to hypothesis $\theta = 2$ ($\mathsf{m}_\theta = 0$), which is an opinion pushed by none of the sending agents. How can we explain this effect? One interpretation is that, in the presence of *conflicting suggestions* coming from the two sub-networks, the receiving agent opts

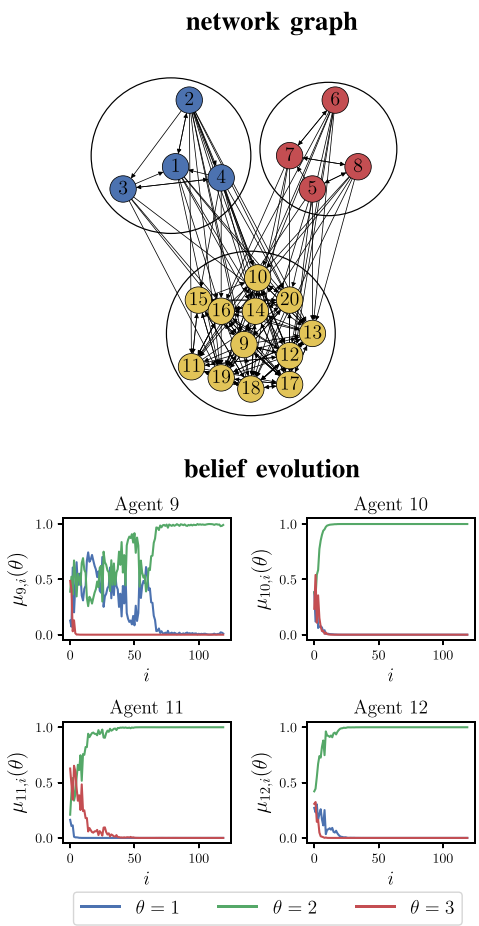

FIGURE 4. ***Truth is somewhere in between*: Convergence of beliefs under balanced influences.**

for a conservative choice. If sending sub-network $\mathcal{N}_1$ says "choose $-\Delta$," while sending sub-network $\mathcal{N}_2$ says "choose $+\Delta$," then the receiving agent prefers to be agnostic and stays in the middle, i.e., it chooses 0. Referring to real-life situations, we can think of one person betting on a soccer match between teams A and B. Assuming that discordant solicitations come from the environment, i.e., the person receives data suggesting to bet on the victory of team A, as well as data suggesting to bet on the victory of team B. If there is no sufficient evidence to let one suggestion prevail, then the most probable choice would be betting on a draw! This "*truth-is-somewhere-in-between*" effect is a remarkable effect that is peculiar to the weakly-connected setting, and that has been not observed before, e.g., it was not present in [12].

In summary, it is the cumulative influence of a sending group over a receiving agent that determines whether it will follow the group's opinion or not. This situation emulates the social phenomenon of herd behavior: agents choose to ignore their private signal in order to follow the most influencing group of agents. When none of the above dominance situations occurs, the receiving agent can opt for an opinion that is not promoted by any of the sending agents.

V. TOPOLOGY LEARNING

In the previous section we examined the effect of the network topology on the social learning of the agents. In particular, we discovered how the topology and the states of the sending agents determine the opinion formation by the receiving agents. The way the information is delivered across the network ultimately determines the minimizers in (20), i.e., the value that each receiving agent's belief will converge to. We now examine the reverse problem. Assume we observe the belief evolution of part of the network. We would like to use this information to infer the underlying influences and topology. This is a useful question to consider because understanding the topology can help us understand why a particular agent adopts a certain opinion. The main question we consider now is this: given some measurements collected at the receiving agents, can we estimate their connections to the sending sub-networks?

We shall answer this question under the following assumption of homogeneity of likelihoods and true distributions inside the individual sending sub-networks.

Assumption 4 (Homogeneity within sending sub-networks): For $s = 1, 2, \ldots, S$, we assume that the distribution and the likelihood functions within the s-th sending sub-network are equal across all agents in that sub-network, namely, for all $\ell \in \mathcal{N}_s$:

$$\boxed{f_\ell = f^{(s)}, \qquad L_\ell(\theta) = L^{(s)}(\theta)} \tag{37}$$

□

One main consequence of Assumption 4 is that (19) becomes:

$$\begin{aligned}\mathscr{D}_k(\theta) &= \sum_{\ell \in \mathcal{S}} \omega_{\ell k} D[f_\ell || L_\ell(\theta)] \\ &= \sum_{s=1}^{S} \left(D[f^{(s)} || L^{(s)}(\theta)] \sum_{\ell \in \mathcal{N}_s} \omega_{\ell k} \right),\end{aligned} \tag{38}$$

where $\mathcal{N}_s$ denotes the collection of agents in the s-th sending sub-network. Equation (38) has the following relevant implication. Under Assumption 4, the network topology influences the average divergence $\mathscr{D}_k(\theta)$ only through an *aggregate* weight:

$$\boxed{x_{sk} \triangleq \sum_{\ell \in \mathcal{N}_s} \omega_{\ell k} = \sum_{\ell \in \mathcal{N}_s} w_{\ell k}} \tag{39}$$

The latter equality, using $w_{\ell k}$ instead of $\omega_{\ell k}$, comes straightforwardly from (13) and (15). This equality reveals that the aggregate weights depend solely on the matrix W, and not on the matrix E of Perron eigenvectors. In other words, the inner structure of the pertinent sending sub-network s does not influence the aggregate weight x_{sk}. We notice that, while a combination weight $a_{\ell k}$ accounts for a *local, small-scale* pairwise interaction between agent ℓ and agent k, the aggregate weight x_{sk} accounts for *macroscopic* topology effects, for two reasons. First of all, x_{sk} is determined by the *limiting*

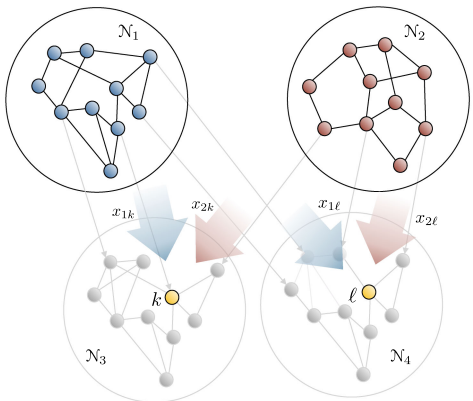

FIGURE 5. ***Macroscopic*** **topology inference problem. The object of topology inference is constituted by the *global* weights x_{sk} from sending sub-network s to receiving agent k. For example, the weight x_{1k} in the figure embodies the influence of *all* sending agents in $\mathcal{N}_1$, from *all* paths (possibly including intermediate receiving agents) leading to receiving agent $k \in \mathcal{N}_3$.**

weights $\omega_{\ell k}$, which embody not only direct connection effects between ℓ and k, but also effects *mediated* by multi-hop paths connecting ℓ and k. Second, from (39) we see that x_{sk} embodies the *global* effect coming from all agents belonging to the s-th sending component. In other words, x_{sk} is a measure of the effect from all agents in sending sub-network s on agent k. Since, in view of Theorem 1, the average divergence determines the behavior of the limiting belief, we conclude from (38) that the network topology ultimately influences the particular hypothesis chosen by a receiving agent only through these *global* weights $\{x_{sk}\}$.

We assume that the data available for estimating x_{sk} are the shared (intermediate) beliefs, $\boldsymbol{\psi}_{k,i}(\theta)$. We will say that *consistent* topology learning is achievable if the x_{sk} can be correctly guessed when sufficient time is given for learning, i.e., we will focus on the *limiting* data, for all $\theta \neq \theta_k^\star$:[4]

$$y_k(\theta) \triangleq \lim_{i\to\infty} \frac{\log \boldsymbol{\psi}_{k,i}(\theta)}{i} \overset{\text{a.s.}}{=} \mathscr{D}_k(\theta_k^\star) - \mathscr{D}_k(\theta). \tag{40}$$

Accordingly, the topology inference problem we are interested in can be formally stated as follows. For any receiving agent k, introduce its global-weight vector:

$$x_k \triangleq [x_{1k}, x_{2k}, \ldots, x_{Sk}]^\top, \tag{41}$$

and consider the vector stacking the H limiting beliefs $y_k(\theta)$ (i.e., the data):

$$y_k \triangleq [y_k(1), y_k(2), \ldots, y_k(H)]^\top. \tag{42}$$

The main question is whether we can estimate x_k consistently from observation of y_k. In the sequel we will sometimes refer to this problem as a *macroscopic* topology inference problem — see Fig. 5 for an illustration. In order to avoid confusion, we remark that the method proposed in this work does not allow retrieving the topology of the network (for that purpose, we refer the reader instead to [34], [35]), but the influence (quantified by the aggregate weights x_{sk}) that each sending sub-network exerts on each receiving agent. While this information has the real topology of the network embedded in it, some other information is missing. For instance, topology inside the sending sub-networks and inside the receiving sub-networks is not considered.

As compared to other topology inference problems, we are faced here with one critical element of novelty. We have no data coming from the sending agents. This means that correlation between sending and receiving agent pairs cannot be performed. This is in sharp contrast with traditional topology inference problems, where the estimation of connections between pairs of agents is heavily based on comparison (e.g., correlation) between data streams coming from these *pairs* of agents [34]–[36]. In contrast, we focus here on the asymmetrical case where, when estimating the weights x_{sk} from sending to receiving agents, no data are available from the sending agents. For this reason, the topology learning problem addressed in this work is significantly different from other traditional topology problems studied in the literature.

VI. IS MACROSCOPIC TOPOLOGY LEARNING FEASIBLE?

We now examine the feasibility of the topology learning problem illustrated in the previous section.

Let us preliminarily introduce a matrix $D = [d_{\theta s}]$, which collects the $H \times S$ divergences between any true distribution in the sending sub-networks and any likelihood, and whose (θ, s) entry is:

$$[D]_{\theta s} = d_{\theta s} = D[f^{(s)} || L^{(s)}(\theta)]. \tag{43}$$

Using (41) and (43) in (38), the network divergence of receiving agent k, evaluated at θ, can be written as:

$$\mathscr{D}_k(\theta) = \sum_{s=1}^{S} d_{\theta s} x_{sk}. \tag{44}$$

Through (42) we can rewrite the limiting data in (40) as:

$$y_k(\theta) = \mathscr{D}(\theta_k^\star) - \mathscr{D}(\theta) = \sum_{s=1}^{S} \left(d_{\theta_k^\star s} - d_{\theta s}\right) x_{sk}. \tag{45}$$

It is useful to introduce the matrix:

$$B_k \triangleq \left(\mathbb{1}_H e_{\theta_k^\star}^\top - I_H\right) D, \tag{46}$$

where e_m is an $H \times 1$ vector with all zeros and a one in the m-th position. It is important to note that B_k has its $\theta_k^\star$-th row equal to zero. We can now formulate the topology problem in terms of the following constrained system:

$$\text{Find } \widetilde{x}_k \in \mathbb{R}^S \text{ such that } \begin{cases} y_k = B_k \widetilde{x}_k, \\ \sum_{s=1}^{S} \widetilde{x}_{sk} = 1, \\ \widetilde{x}_k > 0, \end{cases} \tag{47}$$

[4]We remark that, in view of (71) in Appendix A, the asymptotic properties of $\boldsymbol{\psi}_{k,i}(\cdot)$ are the same as $\boldsymbol{\mu}_{k,i}(\cdot)$.

where we remark that the notation $\widetilde{x}_k > 0$ signifies that all entries in the solution vector $\widetilde{x}_k$ must be strictly positive. This positivity constraint is enforced because by assumption, each receiving sub-network is connected to at least one agent from each sending sub-network, which implies that the true vector we are looking for, x_k, has all positive entries. The equality constraint in (47) can be readily included in matrix form by introducing the augmented matrix and vector:

$$C_k \triangleq \begin{bmatrix} B_k \\ \mathbb{1}_S^\top \end{bmatrix}, \qquad \widetilde{y}_k \triangleq \begin{bmatrix} y_k \\ 1 \end{bmatrix}, \tag{48}$$

which allow rewriting (47) as:

$$\boxed{\text{Find } \widetilde{x}_k \in \mathbb{R}^S : \ \widetilde{y}_k = C_k \widetilde{x}_k, \ \ \widetilde{x}_k > 0} \tag{49}$$

We are now ready to state formally the concept of feasibility for the topology learning problem. First, we want to solve the problem under the assumption that the matrix of divergences, D, is known, i.e., that sufficient knowledge is available about the underlying statistical models (likelihoods and true distributions). In this respect, we remark that the matrix B_k in (46) depends on $\theta_k^\star$, which in turn depends on the unknowns x_{sk} as well through (20). However, from Theorem 1 we know that the beliefs (and also the intermediate beliefs) converge to 1 at $\theta_k^\star$. Therefore, we can safely estimate $\theta_k^\star$ from the limiting data $y_k(\theta)$, which is tantamount to assuming that the matrix B_k is known.

Therefore, achievability of a consistent solution for the topology learning problem translates into the condition that the linear system in (49) should admit a unique solution. We will now prove the following result.

Lemma 1 (Necessary Condition for Macroscopic Topology Learning): The topology learning problem described by the system in (49) admits a unique solution if, and only if:

$$\boxed{\text{rank}(C_k) = S} \tag{50}$$

Thus, a necessary condition for topology learning is that the number of hypotheses is at least equal to the number of sending sub-networks, namely, that:

$$\boxed{H \geq S} \tag{51}$$

■

Proof: We remark that we are not concerned with the existence of a solution for the constrained linear system (49). In fact, this system admits at least a solution, namely, the true weight vector, $x_k \in \mathbb{R}_+^S$, which by assumption fulfills the equation $\widetilde{y}_k = C_k x_k$.

Let us now focus on the unconstrained system (i.e., the system in (49) *without the inequality constraints*), whose set of solutions is given by [38]:

$$\widetilde{x}_k = C_k^\dagger \widetilde{y}_k + (I_S - C_k^\dagger C_k) z, \tag{52}$$

where $z \in \mathbb{R}^S$ is an arbitrary vector, and $C_k^\dagger$ is the Moore-Penrose pseudoinverse of C_k. If $\text{rank}(C_k) = S$, it is well known [38] that $C_k^\dagger = (C_k^\top C_k)^{-1} C_k^\top$, which implies that the second term on the RHS in (52) is zero, which in turn implies that the unconstrained system has the unique solution:

$$\boxed{\widetilde{x}_k = C_k^\dagger \widetilde{y}_k = (C_k^\top C_k)^{-1} C_k^\top \widetilde{y}_k = x_k} \tag{53}$$

The latter equality holds because, if the unconstrained system has a *unique* solution, this is also the unique solution for the constrained system, i.e., it coincides with x_k and satisfies the positivity constraints. Accordingly, we have proved that whenever $\text{rank}(C_k) = S$, the *constrained* system has the unique solution corresponding to the true vector x_k.

We now show that when $\text{rank}(C_k) < S$ the constrained system has infinite solutions. Since any solution of the unconstrained system takes on the form (52), and since x_k *is* a particular solution, there will exist a certain vector z_0 such that x_k can be written as:

$$x_k = C_k^\dagger \widetilde{y}_k + (I_S - C_k^\dagger C_k) z_0. \tag{54}$$

Consider a solution $\widetilde{x}_k$ in (52) that corresponds to another vector, $z = z_0 + \epsilon$, where ϵ is a perturbation vector:

$$\widetilde{x}_k = C_k^\dagger \widetilde{y}_k + (I_S - C_k^\dagger C_k)(z_0 + \epsilon) = x_k + (I_S - C_k^\dagger C_k)\epsilon. \tag{55}$$

Since by assumption $x_k > 0$, we conclude from (55) that for sufficiently small perturbations it is always possible to obtain a distinct $\widetilde{x}_k > 0$, which implies that the *constrained* system in (49) has infinite solutions.

In summary, we conclude that the topology learning problem is feasible if, and only if, $\text{rank}(C_k) = S$. Finally, by observing that the augmented matrix C_k is an $(H+1) \times S$ matrix with an all-zeros row, we have in fact proved the claim of the lemma. ■

Lemma 1 has at least three useful implications. First, it reveals a fundamental interplay between social learning and topology learning: the possibility of estimating x_k depends on the comparison between two seemingly unrelated quantities, the number of hypotheses H (an attribute of the social inferential problem) and the number of sending sub-networks S (an attribute of the network topology).

Second, the necessary condition in (51) highlights that topology learning over social networks is challenging. For example, if the agents of the social network want to solve a binary detection problem ($H = 2$), then the maximum number of sending sub-networks that could allow faithful topology estimation is $S = 2$. Increasing the complexity of the social learning problem (i.e., increasing H) is beneficial to topology estimation, since it allows to increase also S.

Third, we see that having more sending sub-networks makes topology learning more complicated. This is because increasing the number of sending sub-networks increases the number of unknowns (i.e., the dimension of x_k), while not adding information since in our setting we are not allowed to probe the sending nodes. Remarkably, when examining jointly the social learning and the topology learning problems, *the role of the data and of the unknowns is exchanged.* In the social learning problem, more hypotheses means more unknowns and more sending sub-networks means more data;

in the topology learning problem, the situation is exactly reversed.

A. STRUCTURED GAUSSIAN MODELS

In this section we consider the practical case of a Gaussian model, defined as follows.

- All agents use the same family of likelihood functions $\{L(\theta)\}$, for $\theta = 1, 2, \ldots, H$.
- These likelihoods are unit-variance Gaussian likelihoods with different means $\{\mathsf{m}_\theta\}$.
- Each true distribution coincides with one of the likelihoods. This implies that the distribution of the s-th sending sub-network, $f^{(s)}$, is a unit-variance Gaussian distribution with mean ν_s that is chosen among the means $\{\mathsf{m}_\theta\}$, namely, for $s = 1, 2, \ldots, S$:

$$\nu_s \in \{\mathsf{m}_1, \mathsf{m}_2, \ldots, \mathsf{m}_H\}. \tag{56}$$

- The sending sub-networks have different means.

Using (43) and the definition of KL divergence between Gaussian distributions, the matrix D is given by:

$$D = \frac{1}{2}\begin{bmatrix} (\mathsf{m}_1 - \nu_1)^2 & (\mathsf{m}_1 - \nu_2)^2 & \ldots & (\mathsf{m}_1 - \nu_S)^2 \\ (\mathsf{m}_2 - \nu_1)^2 & (\mathsf{m}_2 - \nu_2)^2 & \ldots & (\mathsf{m}_2 - \nu_S)^2 \\ \vdots & & & \vdots \\ (\mathsf{m}_H - \nu_1)^2 & (\mathsf{m}_H - \nu_2)^2 & \ldots & (\mathsf{m}_H - \nu_S)^2 \end{bmatrix}. \tag{57}$$

From (57) it is readily seen that, if the sending sub-networks share the same true distribution (i.e., if $\nu_1 = \nu_2 = \cdots = \nu_S$), then the matrix D has rank 1, and, hence, the topology learning problem is obviously not feasible. As said, we will instead focus on the opposite case where the true expectations are all distinct.

For ease of presentation, and without loss of generality we can assume that the sending sub-networks are numbered so that the expectations of the true distributions are:

$$\nu_1 = \mathsf{m}_1,\ \nu_2 = \mathsf{m}_2, \ldots, \nu_S = \mathsf{m}_S, \tag{58}$$

which implies that (57) takes on the form:

$$D = \frac{1}{2}\begin{bmatrix} 0 & (\mathsf{m}_1 - \mathsf{m}_2)^2 & \ldots & (\mathsf{m}_1 - \mathsf{m}_S)^2 \\ (\mathsf{m}_2 - \mathsf{m}_1)^2 & 0 & \ldots & (\mathsf{m}_2 - \mathsf{m}_S)^2 \\ \vdots & & & \vdots \\ (\mathsf{m}_H - \mathsf{m}_1)^2 & (\mathsf{m}_H - \mathsf{m}_2)^2 & \ldots & (\mathsf{m}_H - \mathsf{m}_S)^2 \end{bmatrix}. \tag{59}$$

The structure in (59) implies that, for $H = S$, the matrix D is a Euclidean distance matrix (but for the constant 1/2) [37]. These matrices are constructed as follows. Given points $r_1, r_2, \ldots, r_L$, belonging to $\mathbb{R}^{\mathsf{dim}}$, the (i, j) entry of the matrix $\mathsf{EDM}(r_1, r_2, \ldots, r_L)$ is given by the squared Euclidean distance between points r_i and r_j. Accordingly, we see from (59) that, for $H = S$:

$$D = \frac{1}{2}\,\mathsf{EDM}(\mathsf{m}_1, \mathsf{m}_2, \ldots, \mathsf{m}_H). \tag{60}$$

For $H > S$, the matrix D can be described as an *extended* Euclidean distance matrix, constructed as follows. Let:

$$E_S \triangleq \frac{1}{2}\,\mathsf{EDM}(\mathsf{m}_1, \mathsf{m}_2, \ldots, \mathsf{m}_S),$$
$$E_H \triangleq \frac{1}{2}\,\mathsf{EDM}(\mathsf{m}_1, \mathsf{m}_2, \ldots, \mathsf{m}_H),$$
$$E_{H-S} \triangleq \frac{1}{2}\,\mathsf{EDM}(\mathsf{m}_{S+1}, \mathsf{m}_{S+2}, \ldots, \mathsf{m}_H), \tag{61}$$

and let F be the $(H - S) \times S$ matrix with entries, for $\theta = S + 1, S + 2, \ldots, H$ and $s = 1, 2, \ldots, S$:

$$[F]_{\theta s} = \frac{1}{2}(\mathsf{m}_\theta - \mathsf{m}_s)^2. \tag{62}$$

Then, we have the following representation:

$$D = \begin{bmatrix} E_S \\ F \end{bmatrix}, \qquad E_H = \begin{bmatrix} E_S & F^\top \\ F & E_{H-S} \end{bmatrix}. \tag{63}$$

The following theorem, which establishes the feasibility of the topology learning problem for the considered Gaussian model, relies heavily on some fundamental properties of Euclidean distance matrices.

Theorem 2 (Macroscopic Topology Learning under Structured Gaussian Models): Let $S \geq 2$ and $H \geq S$. Assume that all sending sub-networks have the same family of unit-variance Gaussian likelihood functions $L(\theta)$ with distinct means $\{\mathsf{m}_\theta\}$, for $\theta = 1, 2, \ldots, H$. Assume that the true distributions $f^{(s)}$, within the sending sub-networks $s = 1, 2, \ldots, S$, are unit-variance Gaussian with distinct means ν_s, chosen from the collection $\{\mathsf{m}_\theta\}$. Then, under Assumption 3 (so that the matrix B_k in (46) is well defined), for all receiving agents $k \in \mathcal{R}$ we have that:

$$\boxed{\mathrm{rank}(C_k) = 2} \tag{64}$$

Proof: The proof is reported in Appendix B. ■

Remark 7 (Topology Learning under Structured Gaussian Models is Challenging): In view of Lemma 1, Eq. (64) has the following implication. Under the considered Gaussian model, topology learning is feasible only when $S = 2$. We remark also that, when $S = 2$, condition (51) plays no role, since any meaningful classification problem has at least $H = 2$. In summary, Theorem 2 reveals that the structure of the Gaussian model makes topology learning very challenging, as this problem is not solvable for networks with more than 2 sending sub-networks. Thus, the theorem reveals that $H \geq S$ is *not* a sufficient condition for consistent topology learning. □

B. DIVERSITY MODELS

We can now examine the effect that diversity in the models of the sending sub-networks can have on topology learning. Since the limiting beliefs are essentially determined by the divergence matrix D, it is meaningful to impose a form of diversity in terms of the divergences between distributions and likelihoods. In other words, differently from the Gaussian case illustrated in the previous section, we now require that the entries of D are not tightly related to each other, namely,

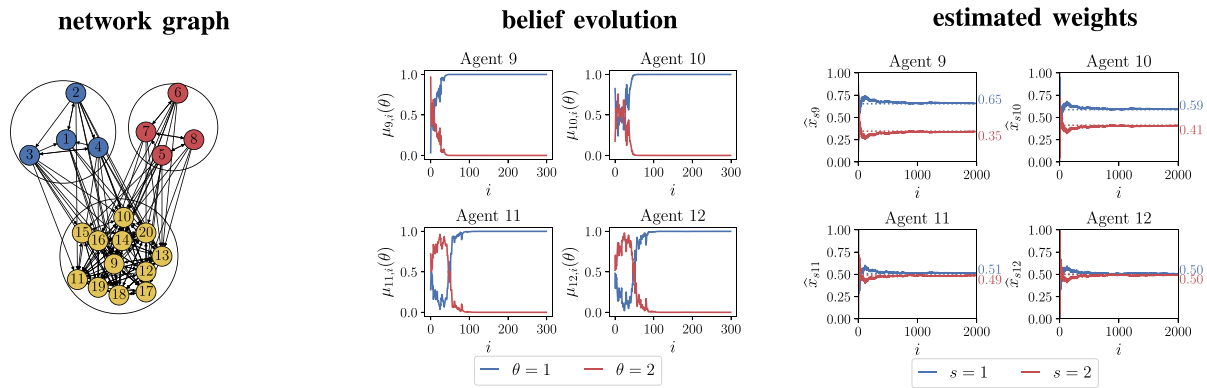

FIGURE 6. **Unperturbed Gaussian model. *Left*. Network topology. *Middle*. Belief convergence at the receiving agents. *Right*. Estimated macroscopic topology. For each of the four panels, the numbers on the right denote the true values $\{x_{sk}\}$, with different colors denoting different s, according to the legend.**

we allow them to assume values in $\mathbb{R}_+^{H\times S}$ (where we denote by $\mathbb{R}_+$ the nonnegative reals) with no strong structure linking them.

One typical model for this type of diversity is that the divergences perceived by the different agents (i.e., across index s), and corresponding to different hypotheses (i.e., across index h), are modeled as absolutely continuous random variables. This randomness is a formal way to embody some degree of variability in how the agents "see" the world. For example, this is a useful model to consider when the agents, due to imperfect knowledge, have likelihoods that are slightly *perturbed* versions of some nominal model. Examples of this type are illustrated in the next section.

In order to avoid confusion, it is important to remark one fundamental property. Under the diversity setting, the matrix $\boldsymbol{D}$ is random[5] with entries modeled as absolutely continuous random variables. The full-rank property for this type of matrices is a classical result. However, we observe from (46) that the matrix $\boldsymbol{B}_k$ is obtained from $\boldsymbol{D}$ by multiplying a matrix that depends on a random variable $\boldsymbol{\theta}_k^\star$, which in turn depends statistically upon the entries of $\boldsymbol{D}$. Finally, we know from (48) that $\boldsymbol{C}_k$ is obtained from $\boldsymbol{B}_k$ by adding an all-ones row. Accordingly, to determine the rank of $\boldsymbol{C}_k$ we need to address carefully these intricate dependencies. This is accomplished in the proof of the forthcoming Theorem 3.

Theorem 3 (Macroscopic Topology Learning under General Models with Diversity): Let $H \geq S$, and assume that the array $\{\boldsymbol{d}_{\theta s}\}$, with $\theta = 1, 2, \ldots, H$ and $s = 1, 2, \ldots, S$, is made of random variables that are jointly absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+^{H\times S}$. Then, for all receiving agents $k \in \mathcal{R}$ we have that, with probability 1, Assumption 3 is verified and the matrix $\boldsymbol{C}_k$ is full column rank, namely,

$$\boxed{\mathbb{P}\left[\boldsymbol{\theta}_k^\star \text{ is unique and } \operatorname{rank}(\boldsymbol{C}_k) = S\right] = 1} \tag{65}$$

Proof: The proof is reported in Appendix C. ■

The meaning of Theorem 3 is that *configurations of KL divergence that lead to a rank-deficient matrix C_k are rare*. In other words, if some diversity exists in the statistical models of the sending components, then the topology inference problem is feasible for almost all configurations.

VII. SIMULATION RESULTS

We now present some illustrative examples. The first example refers to the Gaussian model presented in Section VI-A. The other two examples refer to the setting with diversity presented in Section VI-B.

a) Gaussian with $H = S = 2$. We consider the topology shown in the leftmost panel of Fig. 6. The likelihoods and true distributions for the sending sub-networks are unit-variance Gaussian with means $\nu_1 = \mathsf{m}_1 = 1$, $\nu_2 = \mathsf{m}_2 = 2$. The receiving agents[6] employ the same likelihoods of the sending agents, and their true distributions are unit-variance Gaussian with mean equal to 1. In Fig. 6 (middle) we show the belief convergence for four receiving agents.

Next, we address the topology learning problem. First, for an observation time i, we construct the empirical data $\widehat{y}_k(\theta) = (1/i)\log \boldsymbol{\psi}_{k,i}(\theta)$, and construct an estimate $\widehat{\theta}_k^\star$ as the value of θ that maximizes $\widehat{y}_k(\theta)$ (i.e., the hypothesis where $\widehat{y}_k(\theta)$ will collapse to 1). We can then construct an estimate for B_k as:

$$\widehat{B}_k = \left(\mathbb{1}_H e_{\widehat{\theta}_k^\star}^\top - I_H\right) D, \tag{66}$$

from which we obtain $\widehat{C}_k$ by adding an all-ones row, according to (48). At this point, we have verified on the simulated data that, for any receiving agent $k \in \{9, 10, 11, 12\}$, the matrices $\widehat{C}_k$ are full column rank. Then, we used (53) with empirical

[5] Accordingly, we will now use the bold notation for the matrix entries, $\boldsymbol{d}_{\theta s}$, as well as for other related quantities.

[6] We recall that the models of the receiving agents will be ultimately immaterial as regards their limiting beliefs.

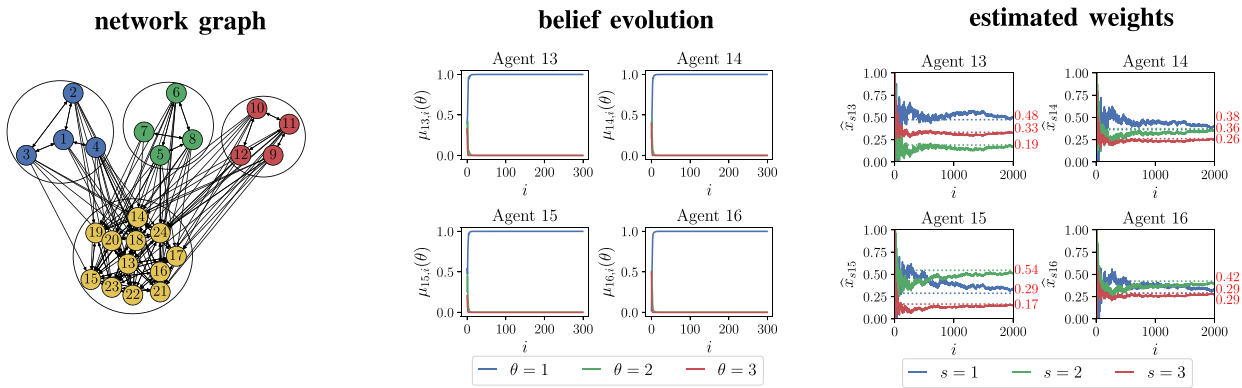

FIGURE 7. Perturbed Gaussian model. *Left*. Network topology. *Middle*. Belief convergence at the receiving agents. *Right*. Estimated macroscopic topology. For each of the four panels, the numbers on the right denote the true values $\{x_{sk}\}$, with different colors denoting different s, according to the legend.

matrices replacing the exact ones to estimate the connection-weight vector x_k as:[7]

$$\widehat{x}_k = \widehat{C}_k^{\dagger} \begin{bmatrix} \widehat{y}_k \\ 1 \end{bmatrix} = (\widehat{C}_k^{\top} \widehat{C}_k)^{-1} \widehat{C}_k^{\top} \begin{bmatrix} \widehat{y}_k \\ 1 \end{bmatrix}. \tag{67}$$

We see from Fig. 6 (right) that this procedure allows us to retrieve the topology coefficients $\{x_{sk}\}$, provided that the system evolves for a sufficiently long time.

b) Randomly perturbed Gaussian with $H = S = 3$. The network topology has three sending sub-networks and one receiving sub-network as shown in the leftmost panel of Fig. 7. When $S > 2$, we know from Theorem 2 that for the *structured* Gaussian model, diversity in the sending components is not enough to ensure the full column rank of the matrix C_k. In order to increase diversity, we consider a *randomly perturbed* model for the likelihood functions, where the likelihood of the s-th sending sub-network, evaluated at hypothesis θ, is unit-variance Gaussian with mean $\theta + \boldsymbol{\epsilon}_{\theta s}$. The random variables $\{\boldsymbol{\epsilon}_{\theta s}\}$ are equally correlated zero-mean Gaussian with variance equal to 0.02 and Pearson correlation coefficient equal to 0.5. For the receiving sub-network we use the same type of random perturbation of the likelihoods. The true distributions for *all* sending and receiving agents are unit-variance Gaussian with mean equal to 1. The belief convergence for four receiving agents can be seen in the middle group of panels of Fig. 7. In the rightmost group of panels, we see how the estimates $\{\widehat{x}_{sk}\}$ of the topology weights converge to the true values $\{x_{sk}\}$. In contrast with the *structured* Gaussian case, topology learning is now feasible for $S > 2$ and even if the true distributions are equal across all sending components. This change in behavior is due to the *diversity* in the models of the sending sub-networks, represented by the different means of the likelihoods. Moreover, we see from the parameters of the random variables $\{\boldsymbol{\epsilon}_{\theta s}\}$ that a relatively small perturbation is already sufficient to enable consistent topology learning.

[7]The symbol $\widehat{\ }$ is used for quantities *estimated* from the data, to be not confused with the symbol $\widetilde{\ }$ used for the *exact* quantities appearing in (49).

c) Beta with $H = S = 3$. Finally, we consider a non-Gaussian example. Moreover, since in the previous examples (motivated by what is typically observed in many networks) we have considered a number of receiving agents fairly larger than the size of the sending sub-networks, we now explore a case where the size of the receiving sub-network is equal to the size of the sending sub-networks.

The non-Gaussian setting used in Fig. 8 considers likelihood functions following a Beta distribution with scale parameter equal to 2 and with shape parameters given by $\theta + 1 + \boldsymbol{u}_{\theta s}$, where $\{\boldsymbol{u}_{\theta s}\}$, for $\theta \in \{1, 2, 3\}$ and $s \in \{1, 2, 3\}$, are independent random variables sampled from a uniform distribution with support $[-0.1, 0.1]$. The true distributions coincide with the unperturbed likelihoods, i.e., the true distribution of the s-th sending sub-network is a Beta distribution with scale parameter equal to 2 and shape parameter equal to $s + 1$. For the receiving sub-network we apply the same type of random perturbation of the likelihoods, whereas the true distributions are Beta with scale and shape parameters equal to 2. The belief convergence for the receiving agents can be seen in the middle group of panels of Fig. 8. In the rightmost group of panels, we see the convergence of the topology estimates.

A. AN EXAMPLE OF NOISY TOPOLOGY RECOVERY

Let us consider a topology learning problem that is feasible according to our previous models and results. In practice, different sources of error can alter these models (and possibly the results). In this section we focus on a relevant source of error and show that the proposed strategy is stable with respect to it.

In the previous treatment, the divergence matrix D was assumed known. However, in some applications this knowledge can be approximate, and D can be known up to a certain error $\delta D \in \mathbb{R}^{H \times S}$. Under this assumption, the solution in (53) is replaced by the following noisy version (agent index k suppressed for ease of notation):

$$x + \delta x = (C + \delta C)^{\dagger} \widetilde{y}, \tag{68}$$

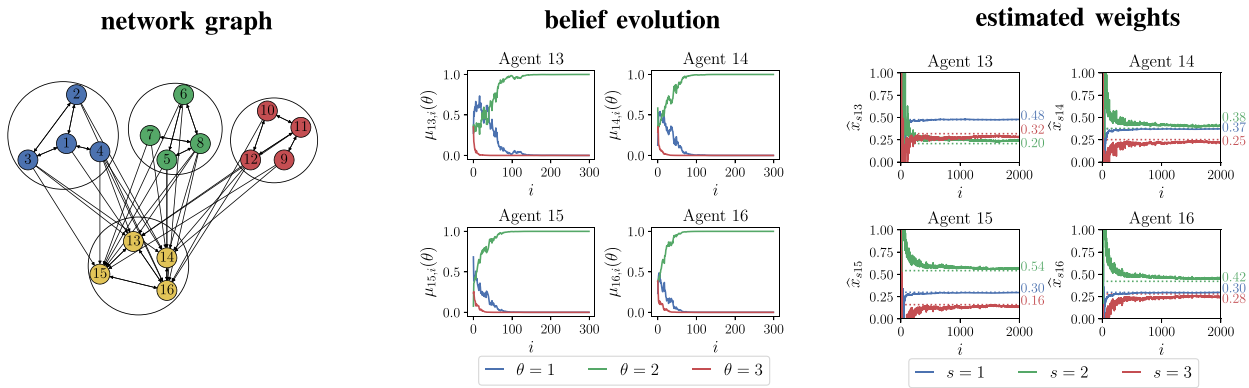

FIGURE 8. Perturbed Beta model. *Left*. Network topology. *Middle*. Belief convergence at the receiving agents. *Right*. Estimated macroscopic topology. For each of the four panels, the numbers on the right denote the true values $\{x_{sk}\}$, with different colors denoting different s, according to the legend.

where $\delta C \in \mathbb{R}^{(H+1)\times S}$ is the error induced by δD on C — see (46) and (48) — and $\delta x \in \mathbb{R}^S$ is the error induced by δC on the true solution x. We now quantify the error δx.

Since we are considering a feasible topology learning problem, we have $H \geq S$ and $\text{rank}(C) = S$. We also know that a noisy matrix δD would typically preserve the rank[8] of C, and, hence, we assume that $\text{rank}(C + \delta C) = S$. Finally, we introduce the condition number $\kappa \triangleq \|C\|_2 \|C^\dagger\|_2$ (where $\|\cdot\|_2$ is the spectral norm), and assume that the matrix C is well-conditioned and the noise is small such that $\|\delta C\|_2 \|C^\dagger\|_2 < 1$. Under these assumptions, Theorem 5.1 in [39] provides the following bound on the relative error:

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \frac{\kappa}{1 - \kappa \frac{\|\delta C\|_2}{\|C\|_2}} \frac{\|\delta C\|_2}{\|C\|_2}, \tag{69}$$

which reveals that, for sufficiently small deviations δC, the relative error $\|\delta x\|_2 / \|x\|_2$ is on the same order as the relative error $\|\delta C\|_2 / \|C\|_2$ [38].

Let us now provide a numerical example to illustrate how (69) works in practice. We consider the same setting of Fig. 6, focusing on receiving agent 10, for which the exact weight vector is given by $x = [0.59,\ 0.41]^\top$. Now, in the considered example we have:

$$D = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}. \tag{70}$$

We generate the matrix δD as follows. The off-diagonal entries of δD are independent random variables following a zero-mean Gaussian distribution with standard deviation σ (when the resulting off-diagonal entries of $D + \delta D$ are negative we resample until nonnegative entries are obtained). The main-diagonal entries of δD are independent random variables distributed as the absolute value of zero-mean Gaussian random variables with standard deviation σ. Then we apply the topology estimation procedure described in Section VII. In Table 1 we report, for several values of σ, the root-mean-square error, $\|\delta x\|_2^{\text{rms}}$, computed over 10^3 Monte Carlo iterations for each value of σ. Examining Table 1, we see that the topology learning strategy is in fact stable w.r.t. to the noise introduced on the divergence matrix.[9]

TABLE I. Root-Mean-Square Error for Different Values of σ. The True Solution is $x = [0.59,\ 0.41]^\top$.

σ	0.001	0.026	0.051	0.075	0.100
$\|\delta x\|_2^{\text{rms}}$	0.001	0.033	0.069	0.105	0.156

VIII. SOCIAL LEARNING VS. TOPOLOGY LEARNING

In this work we have considered two learning problems. The first problem is the Social Learning (SL) problem, which is the goal of the agents in the network. These agents aim at forming their opinions after consulting the beliefs of their neighbors through an iterative update-and-combine SL algorithm. The second problem is the Topology Learning (TL) problem, where a receiving agent (or some entity monitoring its behavior) attempts to get knowledge about the connections between that receiving agent and the sending sub-networks. We can refer to the SL problem as the *direct* learning problem, in the sense that it is the original inferential problem the network is deployed for. Likewise, we can refer to the TL problem as the *dual* learning problem, since it is an inferential procedure that takes as input data *the output of the direct TL problem*.

The analysis conducted in this work has revealed some interesting interplay between SL and TL problems. Let us make a summary of the main results. We recall that S denotes the number of sending sub-networks, and H the number of

[8] For example, if the entries of δD are modeled as jointly absolutely continuous random variables, reasoning as in Theorem 3 we have $\text{rank}(C + \delta C) = S$ with probability 1.

[9] In the considered example, it is straightforward to relate the error $\|\delta x\|_2^{\text{rms}}$ to the errors relative to the individual entries of the true solution, $x = [0.59, 0.41]^\top$. Since both the perturbed and true solution have sum equal to 1, the entries of δx have sum equal to 0, which implies that their (common) root-mean-square value is $\|\delta x\|_2^{\text{rms}} / \sqrt{2}$.

hypotheses. First, we established in Lemma 1 that $H \geq S$ is a necessary condition to achieve consistent TL. This condition has a remarkable interpretation. In a sense, the number of hypotheses is an index (even if not the only one) of complexity associated to the SL problem since, other conditions being equal, more hypotheses make the SL problem more complicated. Likewise, the number of sending components represents an index of complexity of the TL problem, since, other conditions being equal, estimating more links is more complicated. According to these remarks, the condition $H \geq S$ *implies that the TL problem can be feasible when its complexity is not greater than the complexity of the SL problem*. Such an interplay appears to be not obvious at all. As a matter of fact, in the traditional topology inference problems, the connections between agents are inferred from some kind of pairwise measure of their dependency. In our setting, since we cannot measure the output of the sending sub-network, we cannot get direct data quantifying dependency between a receiving and a sending agent. Our TL inference is based instead on the belief functions. The belief function contains some richness of information in that it is evaluated for the H different values of θ. This richness (i.e., H) is critical to enable feasibility of the TL problem. In particular, $H \geq S$ means that the richness of information in the belief function should be greater than or equal to the number of unknown topology weights to be estimated, S.

Having established a necessary condition for consistent TL, we moved on to examine some useful models to see whether and when consistent TL is in fact achievable. First, we have considered a structured Gaussian model where *all* sending sub-networks use the same family of Gaussian likelihoods, and the sending sub-networks have distinct true distributions, each one coinciding with one of the likelihoods. We have shown in Theorem 2 that the TL problem is feasible only if $S = 2$, for any $H \geq 2$. The limited possibility of achieving consistent TL can be ascribed to the limited diversity existing between the different sub-networks (which all use the same family of likelihoods). This observation motivated the analysis of more general models with a certain degree of *diversity*, a condition formalized by saying that the KL divergences between true distributions and likelihoods are not structured, i.e., they are nonnegative real numbers with no particular relationship among them. Under this setting we have ascertained that, if $H \geq S$, the TL problem becomes feasible for almost all configurations, in a precise mathematical sense as stated in Theorem 3. In summary, two critical features that enable consistent TL are: *at least as many hypotheses as sending components* and *a sufficient degree of diversity*.

APPENDIX A
PROOF OF THEOREM 1

Exploiting (6) we can write, for $\theta, \theta' \in \Theta$:

$$\log \frac{\boldsymbol{\psi}_{\ell,i}(\theta)}{\boldsymbol{\psi}_{\ell,i}(\theta')} = \log \frac{\boldsymbol{\mu}_{\ell,i-1}(\theta)}{\boldsymbol{\mu}_{\ell,i-1}(\theta')} + \log \frac{L_\ell(\boldsymbol{\xi}_{\ell,i}|\theta)}{L_\ell(\boldsymbol{\xi}_{\ell,i}|\theta')}. \tag{71}$$

Using (7) we have:

$$\log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} = \sum_{\ell=1}^{N} a_{\ell k} \left(\log \frac{\boldsymbol{\mu}_{\ell,i-1}(\theta)}{\boldsymbol{\mu}_{\ell,i-1}(\theta')} + \log \frac{L_\ell(\boldsymbol{\xi}_{\ell,i}|\theta)}{L_\ell(\boldsymbol{\xi}_{\ell,i}|\theta')} \right). \tag{72}$$

By iterating over i, we can write:

$$\frac{1}{i} \log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} = \frac{1}{i} \sum_{\ell=1}^{N} \sum_{t=1}^{i} [A^{i-t+1}]_{\ell k} \log \frac{L_\ell(\boldsymbol{\xi}_{\ell,t}|\theta)}{L_\ell(\boldsymbol{\xi}_{\ell,t}|\theta')} + \frac{1}{i} \sum_{\ell=1}^{N} [A^i]_{\ell k} \log \frac{\boldsymbol{\mu}_{\ell,0}(\theta)}{\boldsymbol{\mu}_{\ell,0}(\theta')}. \tag{73}$$

Under Assumptions 1–2, thanks to the integrability of the log-ratios between the true distributions and the likelihoods implied by (5), through standard limiting arguments (see, e.g., [30], [31]) it is possible to determine the asymptotic behavior of (73) by: *i*) replacing the powers of matrix A with their limit A_∞ in (12); and *ii*) applying the strong law of large numbers to conclude that:

$$\begin{aligned} \lim_{i\to\infty} \frac{1}{i} \log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} &\overset{\text{a.s.}}{=} \sum_{\ell=1}^{N} [A_\infty]_{\ell k} \mathbb{E}\left[\log \frac{f_\ell(\boldsymbol{\xi}_{\ell,t})}{L_\ell(\boldsymbol{\xi}_{\ell,t}|\theta')} - \log \frac{f_\ell(\boldsymbol{\xi}_{\ell,t})}{L_\ell(\boldsymbol{\xi}_{\ell,t}|\theta)} \right] \\ &= \sum_{\ell\in\mathcal{S}} \omega_{\ell k} D[f_\ell || L_\ell(\theta')] - \sum_{\ell\in\mathcal{S}} \omega_{\ell k} D[f_\ell || L_\ell(\theta)] \\ &= \mathscr{D}_k(\theta') - \mathscr{D}_k(\theta), \end{aligned} \tag{74}$$

where in the second-last equality we used (4), and we performed the replacement $[A_\infty]_{\ell k} = \omega_{\ell k}$, which holds in view of the block representation in (12) since k is a receiving agent and since we adopt the indexing in (16). Now, in light of Assumption 3, we conclude that:

$$\lim_{i\to\infty} \frac{1}{i} \log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta_k^\star)} \overset{\text{a.s.}}{=} \mathscr{D}_k(\theta_k^\star) - \mathscr{D}_k(\theta) < 0 \tag{75}$$

for all $\theta \neq \theta_k^\star$. Since the denominator of $\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta_k^\star)}$ is bounded by 1, Eq. (75) implies that the numerator $\boldsymbol{\mu}_{k,i}(\theta)$ is converging to zero. Since the belief function must sum to 1, the result in (21) holds.

APPENDIX B
PROOF OF THEOREM 2

Preliminarily, it is useful to introduce some auxiliary matrices. We let, for all $\theta = 1, 2, \ldots, H$:

$$\mathcal{I}(\theta) \triangleq \mathbb{1}_H e_\theta^\top - I_H, \tag{76}$$

and

$$B(\theta) \triangleq \mathcal{I}(\theta) D, \quad C(\theta) = \begin{bmatrix} B(\theta) \\ \mathbb{1}_S^\top \end{bmatrix}. \tag{77}$$

In view of Eqs. (46) and (48), the definitions in (76) and (77) imply:

$$B_k = B(\theta_k^\star), \quad C_k = C(\theta_k^\star). \tag{78}$$

We continue by showing some useful properties of the matrix D under the considered Gaussian model. Let us focus on the representation in (63). It is a known result that the rank of a Euclidean distance matrix with n points in $\mathbb{R}^{\mathsf{dim}}$ is at most $\mathsf{dim}+2$ [37]. Since in our case $\mathsf{dim}=1$, we, can write:

$$\text{rank}(E_S) \leq 3. \tag{79}$$

Moreover, for the cases $S=2$ and $S=3$ we have that:

$$E_2 = \frac{1}{2}\begin{bmatrix} 0 & (\mathsf{m}_1-\mathsf{m}_2)^2 \\ (\mathsf{m}_2-\mathsf{m}_1)^2 & 0 \end{bmatrix}, \tag{80}$$

$$E_3 = \frac{1}{2}\begin{bmatrix} 0 & (\mathsf{m}_1-\mathsf{m}_2)^2 & (\mathsf{m}_1-\mathsf{m}_3)^2 \\ (\mathsf{m}_2-\mathsf{m}_1)^2 & 0 & (\mathsf{m}_2-\mathsf{m}_3)^2 \\ (\mathsf{m}_3-\mathsf{m}_1)^2 & (\mathsf{m}_3-\mathsf{m}_2)^2 & 0 \end{bmatrix}, \tag{81}$$

and, hence:

$$\det(\mathrm{E}_2) = -\frac{1}{4}(\mathsf{m}_1-\mathsf{m}_2)^2,$$
$$\det(\mathrm{E}_3) = \frac{1}{4}(\mathsf{m}_1-\mathsf{m}_2)^2(\mathsf{m}_1-\mathsf{m}_3)^2(\mathsf{m}_2-\mathsf{m}_3)^2. \tag{82}$$

Therefore, when the points that determine the Euclidean distance matrix are all distinct, both the above matrices are full rank. Thus, when $S=2$, we have that $\text{rank}(E_S)=2$. When $S>2$, since E_3 is full rank, and in view of (79), we have instead $\text{rank}(E_S)=3$. From the representation of D in (63), we then conclude that:

$$\text{rank}(D) = \begin{cases} 2, & \text{if } S=2, \\ 3, & \text{if } S>2. \end{cases} \tag{83}$$

Next we state and proof a useful lemma.

Lemma 2: Let $\mathcal{I}(\theta)$ be defined as in (76). Then, for all $\theta = 1,2,\ldots,H$ we have that:

$$\boxed{I_H - \mathcal{I}^\dagger(\theta)\mathcal{I}(\theta) = \frac{1}{H}\mathbb{1}\mathbb{1}^\top} \tag{84}$$

Proof of Lemma 2: For ease of notation, in the following proof the explicit dependence on θ is suppressed, and we write $\mathcal{I}$ in place of $\mathcal{I}(\theta)$. By definition of the Moore-Penrose inverse, matrix $\mathcal{I}^\dagger$ satisfies:

$$\mathcal{I}\mathcal{I}^\dagger\mathcal{I} = \mathcal{I}, \qquad (\mathcal{I}^\dagger\mathcal{I})^\top = \mathcal{I}^\dagger\mathcal{I}. \tag{85}$$

Then we note that:

$$\mathcal{I}(I_H - \mathcal{I}^\dagger\mathcal{I}) = \mathcal{I} - \mathcal{I}\mathcal{I}^\dagger\mathcal{I} = \mathcal{I} - \mathcal{I} = 0, \tag{86}$$

where in the second equality we used the first identity in (85). Equation (86) implies that the columns of $(I_H - \mathcal{I}^\dagger\mathcal{I})$ belong to the null space of $\mathcal{I}$, denoted by $\mathcal{N}(\mathcal{I}) = \{v : \mathcal{I}v = 0\}$. On the other hand, in view of (76) we can write:

$$\mathcal{I}v = \mathbb{1}_H e_\theta^\top v - v = \mathbb{1}_H v_\theta - v = 0, \tag{87}$$

with v_θ the θ-th element of v. As a result, Eq. (87) will be satisfied only if $v_h = v_\theta$ for all $h = 1,\ldots,H$. Therefore, we obtain:

$$\mathcal{N}(\mathcal{I}) = \{\alpha\mathbb{1}_H : \alpha \in \mathbb{R}\}, \tag{88}$$

further implying, in light of (86), that, for each $h = 1,2,\ldots,H$, the h-th column of $I_H - \mathcal{I}^\dagger\mathcal{I}$ is of form $\alpha_h\mathbb{1}_H$ for some $\{\alpha_h\}$. On the other hand, since $I_H - \mathcal{I}^\dagger\mathcal{I}$ is symmetric in view of the second identity in (85), we conclude that $\alpha_h = \overline{\alpha}$ for all h, namely,

$$I_H - \mathcal{I}^\dagger\mathcal{I} = \overline{\alpha}\mathbb{1}_H\mathbb{1}_H^\top \tag{89}$$

for some $\overline{\alpha} \in \mathbb{R}$. Finally, since in particular $\mathbb{1}_H \in \mathcal{N}(\mathcal{I})$, we can write:

$$(I_H - \mathcal{I}^\dagger\mathcal{I})\mathbb{1}_H = \mathbb{1}_H - \mathcal{I}^\dagger\mathcal{I}\mathbb{1}_H = \mathbb{1}_H, \tag{90}$$

which, in view of (89), yields:

$$\overline{\alpha}\,\mathbb{1}_H\mathbb{1}_H^\top\mathbb{1}_H = \overline{\alpha}H\mathbb{1}_H = \mathbb{1}_H \Rightarrow \overline{\alpha} = \frac{1}{H}, \tag{91}$$

and we have in fact proved (84). ■

We are now ready to prove Theorem 2.

Proof of Theorem 2: We will now show that

$$\text{rank}(C(\theta)) = 2 \;\text{ for all}\, \theta = 1,2,\ldots,H, \tag{92}$$

which clearly implies the claim of the theorem in view of the second equation in (78).

For the case $H=S=2$, it is immediately seen that the matrix $C(\theta)$ (assuming, e.g., $\theta=1$) takes on the form:

$$C(1) = \begin{bmatrix} 0 & 0 \\ -\dfrac{(\mathsf{m}_1-\mathsf{m}_2)^2}{2} & \dfrac{(\mathsf{m}_1-\mathsf{m}_2)^2}{2} \\ 1 & 1 \end{bmatrix}, \tag{93}$$

which reveals that $\text{rank}(C(\theta)) = 2$.

Let us move on to examine the other cases where $H \geq S$ (excluding $H=S=2$). We will examine first the properties of the matrix $B(\theta)$ in (77). As done before, the dependence on θ is suppressed for ease of notation, and, in particular, we write B, C, and $\mathcal{I}$ in place of $B(\theta)$, $C(\theta)$, and $\mathcal{I}(\theta)$, respectively. Applying Sylvester's inequality to the first equation in (77) we can write [38]:

$$\text{rank}(B) \geq \text{rank}(D) + \text{rank}(\mathcal{I}) - H = \text{rank}(D) - 1, \tag{94}$$

where in the latter equality we used the fact that $\text{rank}(\mathcal{I}) = H-1$. Therefore, from (83) and (94) we conclude that:

$$\text{rank}(B) \geq 1, \quad \text{if } S=2, \tag{95}$$

$$\text{rank}(B) \geq 2, \quad \text{if } S>2. \tag{96}$$

Now we would like to see if equality is satisfied for the cases $S=2$ (with $H>2$) and $S>2$ (with $H \geq S$).

To this end, we start by noticing that equality in Sylvester's inequality holds if, and only if, there exist matrices X and Y that solve [38]:

$$DX + Y\mathcal{I} = I_H, \tag{97}$$

which in turn admits a solution if, and only if, [40]:

$$(I_H - DD^\dagger)(I_H - \mathcal{I}^\dagger\mathcal{I}) = 0. \tag{98}$$

Applying Lemma 2, from (98) we get:

$$(I_H - DD^\dagger)\frac{1}{H}\mathbb{1}_H\mathbb{1}_H^\top = 0, \tag{99}$$

which means that the equality sign in (95) or (96) holds if, and only if:

$$\boxed{DD^\dagger\mathbb{1}_H = \mathbb{1}_H} \tag{100}$$

In particular, we will now show that (100) does not hold for $S = 2$, while it holds for $S > 2$.

Let us start with the case $S = 2$ (and $H > 2$). We will appeal to the representation of D in (63), which for the case $S = 2$ can be written as:

$$D = \frac{1}{2}\begin{bmatrix} 0 & (\mathsf{m}_1 - \mathsf{m}_2)^2 \\ (\mathsf{m}_2 - \mathsf{m}_1)^2 & 0 \\ (\mathsf{m}_3 - \mathsf{m}_1)^2 & (\mathsf{m}_3 - \mathsf{m}_2)^2 \\ \vdots & \vdots \\ (\mathsf{m}_H - \mathsf{m}_1)^2 & (\mathsf{m}_H - \mathsf{m}_2)^2 \end{bmatrix}. \tag{101}$$

Let us now consider the linear system $Dv = \mathbb{1}_H$. From the first two rows of D, we get the unique solution: $v = 2(\mathsf{m}_1 - \mathsf{m}_2)^{-2}\mathbb{1}_2$. Considering now the third row, we get the identity $(\mathsf{m}_3 - \mathsf{m}_1)^2 + (\mathsf{m}_3 - \mathsf{m}_2)^2 = (\mathsf{m}_1 - \mathsf{m}_2)^2$, which is true only if the third point, m_3, is equal to one of the previous points. We conclude that there exist no v such that $Dv = \mathbb{1}_H$, which further implies that $DD^\dagger\mathbb{1}_H \neq \mathbb{1}_H$ Therefore, for $S = 2$ Eq. (95) gives $\text{rank}(B) > 1$, which since B is of dimension $H \times 2$, with $H > 2$, implies that $\text{rank}(B) = 2$.

Let us move on to examine the case $S > 2$ and $H \geq 2$. It is known that, for an $L \times L$ Euclidean distance matrix M, one has $MM^\dagger\mathbb{1}_L = \mathbb{1}_L$, implying that $\mathbb{1}_L$ belongs to the range space of M [41]. We can apply this result to the matrices E_S and E_H in (63), since they are proportional to Euclidean distance matrices. In particular, we can say that there exist vectors u_S and u_H such that $E_S u_S = \mathbb{1}_S,\ E_H u_H = \mathbb{1}_H$. In particular, one of the (infinite) solutions is given by

$$u_H^\star = \begin{bmatrix} u_S \\ 0 \end{bmatrix}. \tag{102}$$

Applying now (102) into (63), we can write:

$$\mathbb{1}_H = E_H u_H^\star = \begin{bmatrix} E_S & F^\top \\ F & E_{H-S} \end{bmatrix}\begin{bmatrix} u_S \\ 0 \end{bmatrix} = \begin{bmatrix} E_S \\ F \end{bmatrix} u_S = Du_S. \tag{103}$$

Equation (100) now follows by observing that:

$$DD^\dagger \underbrace{\mathbb{1}_H}_{Du_S} = \underbrace{DD^\dagger D}_{D} u_S = Du_S = \mathbb{1}_H. \tag{104}$$

We have in fact shown that (100) holds true for $S > 2$, which implies that (96) becomes an equality for $S > 2$.

In summary, we have shown so far that $\text{rank}(B) = 2$ for all $H \geq S$ (but for the case $H = S = 2$, which has been examined separately). We will now use this result to prove the claim of the theorem, namely, that $\text{rank}(C) = 2$. Since C is obtained from B by adding an all-ones row, determining the rank of C from that of B amounts to check whether the row vector $\mathbb{1}_S^\top$ lies in the row space of B, which is tantamount to ascertaining whether there exists z such that:

$$z^\top\mathcal{I}D = \mathbb{1}_S^\top. \tag{105}$$

Since we exclude the case $H = S = 2$, we have always $H \geq 3$. Now, let us consider an EDM E_3 defined on 3 distinct points p_1, p_2, p_3. Since in this case E_3 is full rank, the system $v_3^\top E_3 = \mathbb{1}_3^\top$ has the following (unique) solution:

$$v_3^\top = \begin{bmatrix} \frac{e_{13}+e_{12}-e_{23}}{e_{13}e_{12}} & \frac{e_{12}+e_{23}-e_{13}}{e_{12}e_{23}} & \frac{e_{13}+e_{23}-e_{12}}{e_{13}e_{23}} \end{bmatrix}, \tag{106}$$

where we denoted by $e_{ij} = 1/2(p_i - p_j)^2$ the (i, j) entry of E_3. Let us now introduce the vector:

$$v_H^\top = \begin{bmatrix} v_3^\top & 0_{H-3}^\top \end{bmatrix}. \tag{107}$$

Since, for $H \geq 3$, we know that $\text{rank}(E_H) = 3$, we conclude that:

$$v_3^\top E_3 = \mathbb{1}_3^\top \Rightarrow v_H^\top E_H = \mathbb{1}_H^\top, \tag{108}$$

which, using the block representation of D in (63), yields:

$$v_H^\top D = \mathbb{1}_S^\top. \tag{109}$$

In view of (109), one solution z to (105) exists if $z^\top\mathcal{I} = v_H^\top$, that is, if $v_H^\top$ lies in the row space of $\mathcal{I}$.

On the other hand, from the definition in (76), we see that the matrix $\mathcal{I}$ can be represented as:

$$\mathcal{I} = \begin{bmatrix} -1 & 0 & \dots & 0 & \mathbf{1} & 0\dots0 \\ 0 & -1 & \dots & 0 & \mathbf{1} & 0\dots0 \\ \vdots & \vdots & & \vdots & & \\ 0 & 0 & \dots & -1 & \mathbf{1} & 0\dots0 \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0}\dots\mathbf{0} \\ 0 & 0 & \dots & 0 & \mathbf{1} & -1\dots0 \\ \vdots & \vdots & & \vdots & & \\ 0 & 0 & \dots & 0 & \mathbf{1} & 0\dots-1 \end{bmatrix}, \tag{110}$$

where the bold notation highlights the θ-th row and column. According to (110), the row space of $\mathcal{I}$ is:

$$\text{Row}(\mathcal{I}) = \left\{ [\alpha_1\,\alpha_2\dots\alpha_{\text{H}}] : \alpha_\theta = -\sum_{\text{h}\neq\theta}\alpha_{\text{h}} \right\}, \tag{111}$$

which is equivalent to:

$$\mathrm{Row}(\mathcal{J}) = \left\{[\alpha_1\, \alpha_2 \ldots \alpha_{\mathrm{H}}] : \alpha^\top \mathbb{1}_{\mathrm{H}} = 0\right\}. \tag{112}$$

Examining (106), from straightforward algebra it can be shown that $v_3^\top \mathbb{1}_3 = 0$, which, in light of (107), implies that $v_H^\top \mathbb{1}_H = 0$. Using (112), we conclude that $v_H^\top$ lies in fact in the row space of $\mathcal{J}$, which finally implies, for $H \geq S$ (excluding the case $H = S = 2$) that $\mathrm{rank}(C) = 2$. ■

APPENDIX C
PROOF OF THEOREM 3

We remark that in our setting the divergences are modeled as random variables, which implies that the value of $\boldsymbol{\theta}_k^\star$ is random as well. We should take this into account when proving the claim of the theorem. First, we observe that:

$$\begin{aligned}
&\mathbb{P}[\boldsymbol{\theta}_k^\star \text{ is unique and } \mathrm{rank}(\boldsymbol{C}_k) = S] \\
&= \mathbb{P}[\boldsymbol{\theta}_k^\star \text{ is unique and } \mathrm{rank}(\boldsymbol{C}(\boldsymbol{\theta}_k^\star)) = S] \\
&= \sum_{\theta=1}^{H} \mathbb{P}[\boldsymbol{\theta}_k^\star = \theta, \mathrm{rank}(\boldsymbol{C}(\theta)) = S].
\end{aligned} \tag{113}$$

We now show that, for all $\theta = 1, 2, \ldots, H$:

$$\boxed{\mathbb{P}[\mathrm{rank}(\boldsymbol{C}(\theta)) = S] = 1} \tag{114}$$

It is useful to visualize the matrix $\boldsymbol{C}(\theta)$ as follows:

$$\begin{bmatrix}
\boldsymbol{d}_{\theta 1} - \boldsymbol{d}_{11} & \boldsymbol{d}_{\theta 2} - \boldsymbol{d}_{12} & \ldots & \boldsymbol{d}_{\theta S} - \boldsymbol{d}_{1S} \\
\boldsymbol{d}_{\theta 1} - \boldsymbol{d}_{21} & \boldsymbol{d}_{\theta 2} - \boldsymbol{d}_{22} & \ldots & \boldsymbol{d}_{\theta S} - \boldsymbol{d}_{2S} \\
\vdots & \vdots & & \vdots \\
\boldsymbol{d}_{\theta 1} - \boldsymbol{d}_{(\theta-1)1} & \boldsymbol{d}_{\theta 2} - \boldsymbol{d}_{(\theta-1)2} & \ldots & \boldsymbol{d}_{\theta S} - \boldsymbol{d}_{(\theta-1)S} \\
0 & 0 & \ldots & 0 \\
\boldsymbol{d}_{\theta 1} - \boldsymbol{d}_{(\theta+1)1} & \boldsymbol{d}_{\theta 2} - \boldsymbol{d}_{(\theta+1)2} & \ldots & \boldsymbol{d}_{\theta S} - \boldsymbol{d}_{(\theta+1)S} \\
\vdots & \vdots & & \vdots \\
\boldsymbol{d}_{\theta 1} - \boldsymbol{d}_{H1} & \boldsymbol{d}_{\theta 2} - \boldsymbol{d}_{H2} & \ldots & \boldsymbol{d}_{\theta S} - \boldsymbol{d}_{HS} \\
1 & 1 & \ldots & 1
\end{bmatrix}. \tag{115}$$

The matrix $\boldsymbol{C}(\theta)$ has $H - 1$ *random* rows (i.e., excluding the all-zeros and all-ones rows). Thus, when $H > S$ there are at least S rows with random entries. These random entries are jointly absolutely continuous since *i*) so are the entries of $\boldsymbol{D}$; and *ii*) the mapping from $\boldsymbol{D}$ to (the random entries of) $\boldsymbol{C}(\theta)$ is non-singular.[10] This implies that, for $H > S$:

$$\mathbb{P}[\mathrm{rank}(\boldsymbol{C}(\theta)) = S] = 1, \tag{116}$$

which proves (114) for the case $H > S$.

We switch to the case $H = S$. Let us denote by $\boldsymbol{B}_{S-1}(\theta)$ the sub-matrix of $\boldsymbol{B}(\theta)$ obtained by deleting its last column, and with $\boldsymbol{b}_S(\theta)$ the last column of $\boldsymbol{B}(\theta)$. We can write:

$$\boldsymbol{C}(\theta) = \begin{bmatrix} \boldsymbol{B}_{S-1}(\theta) & \boldsymbol{b}_S(\theta) \\ \mathbb{1}_{S-1}^\top & 1 \end{bmatrix}. \tag{117}$$

We notice that $\boldsymbol{B}_{S-1}(\theta)$ depends only on the sub-matrix $\boldsymbol{D}_{S-1}$ that is obtained by deleting from $\boldsymbol{D}$ the last column. It is thus meaningful to introduce the set of matrices:

$$\mathcal{E} \triangleq \{D_{S-1} : \mathrm{rank}(B_{S-1}(\theta)) = S - 1\}. \tag{118}$$

Recalling that $B_{S-1}(\theta)$ contains an all-zeros row, we see that, given a matrix $D_{S-1} \in \mathcal{E}$, there exists a unique sequence of weights:

$$w_1, w_2, \ldots, w_{\theta-1}, w_{\theta+1}, \ldots, w_S, \tag{119}$$

to obtain the row vector $\mathbb{1}_{S-1}^\top$ as a weighted linear combination of the rows of $B_{S-1}(\theta)$. Accordingly, given a matrix $D_{S-1} \in \mathcal{E}$, the rank of $\boldsymbol{C}(\theta)$ will be equal to S if the last row in $\boldsymbol{C}(\theta)$ cannot be obtained as a linear combination of the rows of $\boldsymbol{B}(\theta)$. In view of (117), this corresponds to check whether the linear combination of the elements in $\boldsymbol{b}_S$ with the same weights is equal to 1, namely, if:

$$\sum_{h \neq \theta} w_h (\boldsymbol{d}_{\theta S} - \boldsymbol{d}_{hS}) = 1. \tag{120}$$

Consider now a matrix $D_{S-1} \in \mathcal{E}$ We have that:

$$\mathbb{P}\left[\sum_{h \neq \theta} w_h (\boldsymbol{d}_{\theta S} - \boldsymbol{d}_{hS}) = 1 \,\middle|\, D_{S-1}\right] = 0, \tag{121}$$

since (also conditioned on $\boldsymbol{D}_{S-1}$) the random variables $\{\boldsymbol{d}_{hS}\}$, with $h = 1, 2, \ldots, H$, are jointly absolutely continuous. We then conclude that:

$$\mathbb{P}[\mathrm{rank}(\boldsymbol{C}(\theta)) = S | D_{S-1}] = 1, \tag{122}$$

which implies (114) since, in view of the joint absolute continuity of the entries in $\boldsymbol{D}$, we have that:

$$\mathbb{P}[\mathrm{rank}(\boldsymbol{B}_{S-1}(\theta)) = S - 1] = 1 \Rightarrow \mathbb{P}[\boldsymbol{D}_{S-1} \in \mathcal{E}] = 1. \tag{123}$$

If we now apply (114) in (113), we conclude that:

$$\begin{aligned}
&\mathbb{P}[\boldsymbol{\theta}_k^\star \text{ is unique and } \mathrm{rank}(\boldsymbol{C}_k) = S] \\
&= \sum_{\theta=1}^{H} \mathbb{P}[\boldsymbol{\theta}_k^\star = \theta] = \mathbb{P}[\boldsymbol{\theta}_k^\star \text{ is unique }].
\end{aligned} \tag{124}$$

The proof of the theorem will be now complete if we show that the probability of having a unique $\boldsymbol{\theta}_k^\star$ is equal to 1. To this aim, by using (20) and (44), we see that:

$$\boldsymbol{\theta}_k^\star = \underset{\theta \in \Theta}{\mathrm{argmin}} \sum_{s=1}^{S} x_{sk} \boldsymbol{d}_{\theta s}. \tag{125}$$

Let us consider the summations in (125) corresponding to different values of θ. Since the random variables $\{\boldsymbol{d}_{\theta s}\}$ are jointly absolutely continuous (and since x_k is not an all-zeros vector), the probability that two or more summations are equal is zero, which finally implies that $\boldsymbol{\theta}_k^\star$ is unique. ■

[10] For example, property *ii*) can be grasped by noting that, conditioned on $d_{\theta 1} \ldots, d_{\theta S}$, the random entries in (115) are jointly absolutely continuous.

REFERENCES

[1] V. Matta, A. Santos, and A. H. Sayed, "Exponential collapse of social beliefs over weakly-connected heterogeneous networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 5267–5271.

[2] V. Matta, V. Bordignon, A. Santos, and A. H. Sayed, "Learning graph influence from social interactions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2020, pp. 5535–5539.

[3] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, 2011.

[5] V. Krishnamurthy and H. V. Poor, "Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact?" *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 43–57, May 2013.

[6] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Bus. School Res. Paper*, no. 13–28, pp. 1–38, May 2013.

[7] C. Chamley, A. Scaglione, and L. 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.

[8] 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.

[9] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[10] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 709–713.

[11] B. Ying and A. H. Sayed, "Information exchange and learning dynamics over weakly connected adaptive networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1396–1414, Mar. 2016.

[12] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 222–238, Jun. 2017.

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

[14] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II–Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

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

[16] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4118–4132, Nov. 2006.

[17] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.

[18] G. Scutari, S. Barbarossa, and L. Pescosolido, "Distributed decision through self-synchronizing sensor networks in the presence of propagation delays and asymmetric channels," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1667–1684, Apr. 2008.

[19] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[20] V. Matta and A. H. Sayed, "Estimation and detection over adaptive networks," in *Cooperative and Graph Signal Processing*, P. Djuric and C. Richard, Eds. New York, NY, USA: Elsevier, 2018, pp. 69–106.

[21] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, 2000.

[22] D. Acemoglu, M. Dahleh, A. Ozdaglar, and A. Tahbaz-Salehi, "Observational learning in an uncertain world," in *Proc. IEEE Conf. Decis. Control*, Dec. 2010, pp. 6645–6650.

[23] V. Krishnamurthy, O. N. Gharehshiran, and M. Hamdi, "Interactive sensing and decision making in social networks," *Found. Trends Signal Process.*, vol. 7, no. 1–2, pp. 1–196, Apr. 2014.

[24] D. Acemoglu, M. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Rev. Econ. Stud.*, vol. 78, no. 4, pp. 1201–1236, Oct. 2011.

[25] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.

[26] L. G. Epstein, J. Noor, and A. Sandroni, "Non-Bayesian learning," *BE J. Theor. Econ.*, vol. 10, no. 1, pp. 1–20, 2010.

[27] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis)information in social networks," *Games Econ. Behav.*, vol. 70, no. 2, pp. 194–227, Nov. 2010.

[28] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games Econ. Behav.*, vol. 76, no. 1, pp. 210–225, Sep. 2012.

[29] P. Molavi, A. Jadbabaie, K. R. Rad, and A. Tahbaz-Salehi, "Reaching consensus with increasing information," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 358–369, Apr. 2013.

[30] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed non-Bayesian learning," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5538–5553, Nov. 2017.

[31] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, pp. 6161–6179, Sep. 2018.

[32] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[33] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, no. 4/5, pp. 311–801, 2014.

[34] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 622–646, Jan. 2019.

[35] A. Santos, V. Matta, and A. H. Sayed, "Local tomography of large networks under the low-observability regime," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 587–613, Jan. 2020.

[36] G. Mateos, S. Segarra, A. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.

[37] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 12–30, Nov. 2015.

[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[39] P. Wedin, "Perturbation theory for pseudo-inverses," *BIT Numer. Math.*, vol. 13, no. 2, pp. 217–232, Jun. 1973.

[40] J. Baksalary and R. Kala, "The matrix equation $AX - YB = C$," *Linear Algebra Appl.*, vol. 25, pp. 41–43, Jun. 1979.

[41] J. C. Gower, "Properties of Euclidean and non-Euclidean distance matrices," *Linear Algebra Appl.*, vol. 67, pp. 81–97, Jun. 1985.

VINCENZO MATTA received the Laurea degree (*cum laude*) in electronic engineering and the Ph.D. degree in information engineering from the University of Salerno, Fisciano, Italy, in 2001 and 2005, respectively. He is currently an Associate Professor with the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno. His research interests cover the wide area of statistical signal processing and information theory, with current emphasis on: Adaptation and learning over networks; the interplay between inference, communications and security in distributed systems; multiobject/multisensor tracking and data fusion; detection of gravitational waves. He has published more than 120 articles on international journals and proceedings of international conferences. He serves as an Associate Editor for the IEEE OPEN JOURNAL OF SIGNAL PROCESSING, as a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS, and formerly served as an Associate Editor for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS.

VIRGINIA BORDIGNON received engineering degrees from Ecole Centrale de Lyon, France, and the Federal University of Rio Grande do Sul (UFRGS), Brazil, and the M.S. degree in electrical engineering from UFRGS. She is currently working toward the Ph.D. degree at the Adaptive Systems Laboratory in Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Her research interests include social learning, distributed learning and statistical signal processing.

AUGUSTO SANTOS received the B.Sc. and M.Sc. degrees from Instituto Superior Técnico (IST), Lisbon, Portugal, in 2007 and 2008, respectively and the Ph.D. degree from IST and Carnegie Mellon University (CMU), Pittsburgh, USA, in 2014, all in electrical and computer engineering. He held a postdoctoral scholar position at CMU during 2015–2017 and another at the Adaptive Systems Laboratory at École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, during 2017–2019.

ALI H. SAYED is Dean of Engineering at EPFL, Switzerland, and has served as Distinguished Professor and Chairman of Electrical Engineering at the University of California, Los Angeles. He is a member of the U.S. National Academy of Engineering and recognized as a Highly Cited Researcher. He served as President of the IEEE Signal Processing Society in 2018 and 2019. An author of over 550 scholarly publications and six books, his research involves several areas including adaptation and learning theories, data and network sciences, and statistical inference. Dr. Sayed has received several awards including the 2015 Education Award from the IEEE Signal Processing Society, 2014 Athanasios Papoulis Award from the European Association for Signal Processing, the 2013 Meritorious Service Award, and the 2012 Technical Achievement Award from the IEEE Signal Processing Society. He also received the 2005 Terman Award from the American Society for Engineering Education, the 2003 Kuwait Prize, and the 1996 IEEE Donald G. Fink Prize. He served as Distinguished Lecturer for the IEEE Signal Processing Society in 2005 and as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2003–2005). His articles received several Best Paper Awards from the IEEE (2002, 2005, 2012, 2014). He is a Fellow of IEEE, EURASIP, and the American Association for the Advancement of Science (AAAS).