

# LEARNING GRAPH INFLUENCE FROM SOCIAL INTERACTIONS

Vincenzo Matta\* Virginia Bordignon† Augusto Santos Ali H. Sayed†

\*DIEM, University of Salerno. †EPFL, School of Engineering.

## ABSTRACT

In social learning, agents form their opinions or beliefs about certain hypotheses by exchanging local information. This work considers the recent paradigm of *weak graphs*, where the network is partitioned into sending and receiving components, with the former having the possibility of exerting a domineering effect on the latter. Such graph structures are prevalent over social platforms. We will *not* be focusing on the direct social learning problem (which examines *what* agents learn), but rather on the dual or reverse learning problem (which examines *how* agents learned). Specifically, from observations of the stream of beliefs at certain agents, we would like to examine whether it is possible to learn the strength of the connections (influences) from sending components in the network to these receiving agents.

**Index Terms**— Social learning, topology learning, weak graphs, Bayesian update, diffusion strategy.

## 1. INTRODUCTION AND RELATED WORK

Social learning is a collective process where agents construct their individual beliefs about certain hypotheses by integrating the beliefs of neighboring agents into their own through social interaction [1–4]. There exist several variants of social learning algorithms, which assume different protocols for the distributed propagation of information, as well as different ways of combining the neighbors’ beliefs. Most algorithms rely either on consensus [5] or diffusion strategies [6–10], with some works using linear combination of beliefs [5–8] and other works using logarithmic beliefs [9, 10]. However, with the exception [6–8], most prior works focus mainly on *strongly-connected* networks (i.e., graphs where there is a direct and reverse path between any two agents, in addition to some agents having a self-loop as a sign of confidence in their own information). Under this setting, the limiting (as time elapses) evolution of the individual agents’ belief has been shown to converge collectively to the same opinion, which can be the true underlying hypothesis [5, 6, 10], or a hypothesis minimizing a suitable objective function [9].

The relevant case of *weakly-connected* networks has received less attention in the literature, despite its relevance for information spread over social platforms. Over weak graphs, some sub-components of the graph send information in one direction towards receiving agents but do not necessarily pay attention to (or even receive) information back. For example, a celebrity on Twitter may have thousands or millions of followers, but may be following only a handful of these individuals. Another example is media networks broadcasting information to a large number of users and hardly receiving feedback from any of these users. More fundamentally, a weak graph is modeled as consisting of two components: *sending* sub-networks and *receiving* sub-networks [6–8]. This paradigm was considered in [7, 8] with reference to the linear-belief-combination

rule, and in [11] for the log-belief combination rule. These works showed that, over weak graphs, receiving agents can be strongly influenced by one or more sending sub-networks independent of their own local observations. In this way, receiving agents can be made to move towards wrong beliefs by domineering sending sub-networks.

The earlier works focused on the direct learning problem. They examined the following fundamental question. Given a weak graph, how does information diffuse through the network and what are the limiting beliefs that agents converge to? Will they all converge to the same opinion or to scattered opinions? This paper examines the *dual* or reverse learning problem. Assume we observe the evolution of the beliefs of certain receiving agents over time. Can we discover which sending sub-networks are most responsible for influencing the opinion formation of these agents? It is clear that this is a very relevant problem with many useful applications. It is also a challenging problem for reasons that will become clear as we progress with the presentation.

The inverse learning problem falls into the class of topology learning. However, in contrast with standard topology inference problems, there is an important element of novelty and distinction. In our formulation, we do *not* have access to the beliefs streaming from the sending agents. For this reason, we cannot exploit traditional methods where the estimation of connections between pairs of agents relies on comparison (e.g., correlation) between data streams coming from these *pairs* of agents [12–14]. We need to develop an alternative approach, which exploits to great effect our previous results on the direct learning problem. In particular, our analysis reveals a useful interplay between the two coexisting learning problems: *i*) the *direct* inferential problem of social learning; and *ii*) the *inverse* topology learning problem.

## 2. BACKGROUND AND PROBLEM SETTING

A network of  $N$  agents collects streaming data from the environment. The random variable  $\boldsymbol{\xi}_{k,i} \in \mathcal{X}_k$  (we use bold fonts to emphasize randomness) describes the data at agent  $k \in \{1, 2, \dots, N\}$  at time  $i \in \mathbb{N}$ . Data are independent over time, but they can be dependent across agents. The space  $\mathcal{X}_k$  can vary across agents, and  $\boldsymbol{\xi}_{k,i}$  is generated according to  $f_k(\boldsymbol{\xi})$  (either a probability density function or a probability mass function), which is allowed to vary across agents as well. The goal of the agents is to choose one state of nature  $\theta$  belonging to a finite set  $\Theta = \{1, 2, \dots, H\}$ . To accomplish this task, the agents assume a family of *likelihood functions*  $L_k(\boldsymbol{\xi}|\theta)$  with  $\boldsymbol{\xi} \in \mathcal{X}_k$ . The dissimilarity between the true distribution  $f_k(\boldsymbol{\xi})$  and the likelihood  $L_k(\boldsymbol{\xi}|\theta)$  is quantified through the Kullback-Leibler (KL) divergence  $D[f_k||L_k(\theta)]$ , which will be assumed finite for all agents and hypotheses.

Let us now describe the social learning strategy. Since at time  $i = 0$ , the agents have no prior information to discard any hypothesis, we will assume that all agents assign nonzero probability mass to all hypotheses, namely,  $\boldsymbol{\mu}_{k,0}(\theta) > 0$  for all  $\theta \in \Theta$  [9, 10]. For any hypothesis  $\theta \in \Theta$ , agent  $k$  at time  $i$  employs its most recent *private* data,  $\boldsymbol{\xi}_{k,i}$ , to evaluate the likelihood  $L_k(\boldsymbol{\xi}_{k,i}|\theta)$ , which is in turn employed to update the local belief,  $\boldsymbol{\mu}_{k,i-1}(\theta)$ . This leads

This work was supported in part by the Swiss National Science Foundation under grant 205121-184999. E-mails: vmatta@unisa.it, virginia.bordignon@epfl.ch, augusto.pt@gmail.com, ali.sayed@epfl.ch

to an *intermediate* belief  $\psi_{k,i}(\theta)$  through the following Bayesian update:

$$\psi_{k,i}(\theta) = \frac{\boldsymbol{\mu}_{k,i-1}(\theta)L_k(\boldsymbol{\xi}_{k,i}|\theta)}{\sum_{\theta' \in \Theta} \boldsymbol{\mu}_{k,i-1}(\theta')L_k(\boldsymbol{\xi}_{k,i}|\theta')}. \quad (1)$$

Second, in a combination step, agent  $k$  aggregates the intermediate beliefs received from its neighbors by combining linearly the logarithm of these beliefs (exponentiation and normalization serve to give back an admissible belief):

$$\boldsymbol{\mu}_{k,i}(\theta) = \frac{\exp \left\{ \sum_{\ell=1}^N a_{\ell k} \log \psi_{\ell,i}(\theta) \right\}}{\sum_{\theta' \in \Theta} \exp \left\{ \sum_{\ell=1}^N a_{\ell k} \log \psi_{\ell,i}(\theta') \right\}}. \quad (2)$$

The matrix  $A = [a_{\ell k}]$  is left-stochastic since we assume that the weight  $a_{\ell k} \geq 0$  is necessarily equal to zero if  $k$  cannot receive data from  $\ell$ , and that the weights used by  $k$  to scale the received beliefs from its neighbors add up to one.

We focus on the case of a weak graph, which is defined as follows [7, 8]. The network  $\mathcal{N} = \{1, 2, \dots, N\}$  is divided into  $S$  sending sub-networks (denoted by  $\mathcal{N}_s$ , for  $s = 1, 2, \dots, S$ ) and  $R$  receiving sub-networks (denoted by  $\mathcal{N}_{S+r}$ , for  $r = 1, 2, \dots, R$ ):

$$\mathcal{N} = \mathcal{S} \cup \mathcal{R}, \quad \mathcal{S} \triangleq \bigcup_{s=1}^S \mathcal{N}_s, \quad \mathcal{R} \triangleq \bigcup_{r=1}^R \mathcal{N}_{S+r}. \quad (3)$$

The combination matrix over weak graphs has the following block form (with increasing node ordering across the  $S + R$  components):

$$A = \begin{bmatrix} A_{\mathcal{S}} & A_{\mathcal{S}\mathcal{R}} \\ 0 & A_{\mathcal{R}} \end{bmatrix} \quad (4)$$

where the matrix  $A_{\mathcal{S}} = \text{blockdiag} \{A_{\mathcal{N}_1}, A_{\mathcal{N}_2}, \dots, A_{\mathcal{N}_S}\}$  contains the weights within the sending sub-networks, and has a block-diagonal form since communication between distinct sending sub-networks is not necessary (otherwise, sending sub-networks can be grouped into a larger sending sub-network). Likewise, the left-bottom block of  $A$  is zero since communication from receiving to sending sub-networks is forbidden. The  $S$  sending sub-networks (resp., the  $R$  receiving sub-networks) are all individually assumed *strongly connected* (resp., connected; meaning that self-loops are not necessary). Communication among the  $R$  sub-networks is allowed. Finally, we assume that each receiving sub-network is connected to at least one sending agent.

It was shown in [7] that the limiting combination matrix power has the following structure:

$$A_{\infty} \triangleq \lim_{i \rightarrow \infty} A^i = \begin{bmatrix} E & EW \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} E & \Omega \\ 0 & 0 \end{bmatrix}, \quad (5)$$

where  $E = \text{blockdiag} \{p^{(1)} \mathbb{1}_{N_1}^{\top}, p^{(2)} \mathbb{1}_{N_2}^{\top}, \dots, p^{(S)} \mathbb{1}_{N_S}^{\top}\}$  is a block diagonal matrix that stacks the  $N_s \times 1$  Perron eigenvectors  $p^{(s)}$  associated with the  $s$ -th sending sub-network,  $\mathbb{1}_L$  is an all-ones vector of size  $L \times 1$ , and where:

$$W = A_{\mathcal{S}\mathcal{R}} (I - A_{\mathcal{R}})^{-1}, \quad \Omega = EW. \quad (6)$$

The entries of  $\Omega$  are denoted by  $[\omega_{\ell k}]$  and we keep indexing the columns of  $\Omega$  with an index  $k = |\mathcal{S}| + 1, \dots, |\mathcal{S}| + |\mathcal{R}|$ . Since the limiting matrix power is left-stochastic and has a zero right-bottom block,  $\Omega$  is left-stochastic as well. From (6) we can also write  $\Omega = EA_{\mathcal{S}\mathcal{R}}(I + A_{\mathcal{R}} + A_{\mathcal{R}}^2 + \dots)$ , and we see that  $\omega_{\ell k}$

embodies the sum of influences over all paths from sending agent  $\ell$  to receiving agent  $k$ .

Let us now introduce the following *average divergence* at receiving agent  $k \in \mathcal{R}$ :

$$\mathcal{D}_k(\theta) \triangleq \sum_{\ell \in \mathcal{S}} \omega_{\ell k} D[f_{\ell} || L_{\ell}(\theta)], \quad (7)$$

which is a weighted combination of the KL divergences pertaining *only* to the *sending* agents. Throughout the work, we will invoke the following standard identifiability assumption.

**Assumption 1 (Unique Minimizer).** For  $k = 1, 2, \dots, N$ , the function  $\mathcal{D}_k(\theta)$  has the unique minimizer:

$$\theta_k^* \triangleq \underset{\theta \in \Theta}{\text{argmin}} \mathcal{D}_k(\theta). \quad (8)$$

□

It was shown in [11] that, under Assumption 1, the diffusion strategy in (1)–(2) minimizes the divergence in (7), namely, that:

$$\lim_{i \rightarrow \infty} \boldsymbol{\mu}_{k,i}(\theta_k^*) \stackrel{\text{a.s.}}{=} 1, \quad (9)$$

where  $\stackrel{\text{a.s.}}{=}$  denotes almost-sure convergence. Moreover, for all  $\theta \neq \theta_k^*$ , the belief goes to zero exponentially as:

$$\lim_{i \rightarrow \infty} \frac{\log \boldsymbol{\mu}_{k,i}(\theta)}{i} \stackrel{\text{a.s.}}{=} \mathcal{D}_k(\theta_k^*) - \mathcal{D}_k(\theta). \quad (10)$$

### 3. TOPOLOGY LEARNING

In light of (8), the particular opinion  $\theta_k^*$  that will be chosen by the  $k$ -th receiving agent is ultimately determined by the average divergence in (7). This dependence creates a strong tie between the network topology (through the limiting combination weights  $\omega_{\ell k}$ ), and the shape of the beliefs. We now examine the reverse problem. Assume the belief evolution of a receiving agent is monitored. This is a reasonable assumption since the information shared by the agents in the social learning strategy is actually constituted by the beliefs. We want to use this information to infer the underlying links between the receiving agent and the sending sub-networks. This problem will be addressed under the following homogeneity assumption.

**Assumption 2 (Homogeneity in sending sub-networks).** For  $s = 1, 2, \dots, S$ , 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  $\ell \in \mathcal{N}_s$ :

$$f_{\ell} = f^{(s)}, \quad L_{\ell}(\theta) = L^{(s)}(\theta). \quad (11)$$

□

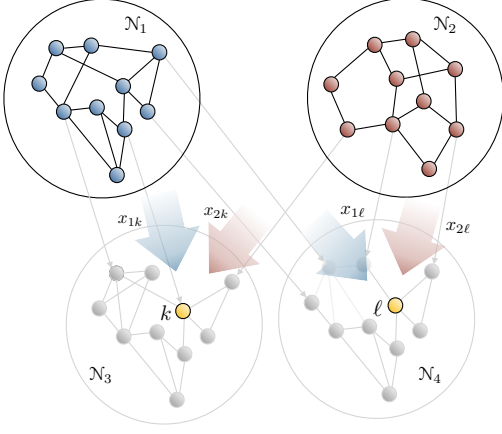
Assumption 2 implies that (7) becomes:

$$\mathcal{D}_k(\theta) = \sum_{s=1}^S \left( D[f^{(s)} || L^{(s)}(\theta)] \sum_{\ell \in \mathcal{N}_s} \omega_{\ell k} \right), \quad (12)$$

which means that the topology influences  $\mathcal{D}_k(\theta)$  only through an *aggregate* weight:

$$x_{sk} \triangleq \sum_{\ell \in \mathcal{N}_s} \omega_{\ell k} = \sum_{\ell \in \mathcal{N}_s} w_{\ell k}, \quad (13)$$

where the latter equality comes from (6) and the definition of  $E$ . Now, while a weight  $a_{\ell k}$  accounts for a *local* pairwise or *microscopic* interaction between  $\ell$  and  $k$ , the aggregate weight  $x_{sk}$  accounts for *macroscopic* topology effects, since: *i*)  $x_{sk}$  is determined



**Fig. 1.** The topology inference goal is to estimate the *global weights*  $x_{sk}$  linking sending sub-network  $s$  to receiving agent  $k$ . The weight  $x_{1k}$  in the figure embodies the influence of *all* sending agents in  $\mathcal{N}_1$ , from *all* paths (including intermediate receiving agents) leading to the receiving agent  $k \in \mathcal{N}_3$ .

by the *limiting weights*  $\omega_{\ell k}$ , which embody also effects *mediated* by multi-hop paths connecting  $\ell$  and  $k$ ; and *ii*)  $x_{sk}$  embodies the *global effect* coming from all agents belonging to the  $s$ -th sending component. Since we know that  $\mathcal{D}_k(\theta)$  determines the behavior of the limiting belief, Eq. (12) reveals that the topology ultimately determines the opinion chosen by a receiving agent only through the *global weights*  $\{x_{sk}\}$ .

Regarding the data used for topology inference, we assume the shared intermediate beliefs,  $\psi_{k,i}(\theta)$ , are available. We will say that *consistent* topology learning is achievable if the  $\{x_{sk}\}$  can be correctly estimated when sufficient time is given for learning. We focus accordingly on the *limiting data*:<sup>1</sup>

$$y_k(\theta) \triangleq \lim_{i \rightarrow \infty} \frac{\log \psi_{k,i}(\theta)}{i} \stackrel{\text{a.s.}}{=} \mathcal{D}_k(\theta_k^*) - \mathcal{D}_k(\theta), \quad (14)$$

and formulate the following topology inference problem, which is illustrated in Fig. 1. Introduce the global-weight vector  $x_k \triangleq [x_{1k}, x_{2k}, \dots, x_{Sk}]^\top$  and stack the  $H$  limiting beliefs  $y_k(\theta)$  as  $y_k \triangleq [y_k(1), y_k(2), \dots, y_k(H)]^\top$ . We would like to know whether we can estimate  $x_k$  consistently from observation of  $y_k$ .

It is useful to introduce the  $H \times S$  matrix  $[D]_{\theta_s} = d_{\theta_s} = D[f^{(s)} | L^{(s)}(\theta)]$ , which allows rewriting the limiting data as:

$$y_k(\theta) = \mathcal{D}(\theta_k^*) - \mathcal{D}(\theta) = \sum_{s=1}^S (d_{\theta_k^* s} - d_{\theta s}) x_{sk}. \quad (15)$$

Accordingly, we see that the topology inference problem can be recast in terms of the following constrained linear system:

$$\text{Find } \tilde{x}_k \in \mathbb{R}^S : \tilde{y}_k = C_k \tilde{x}_k, \quad \tilde{x}_k > 0, \quad (16)$$

where we defined:

$$B_k \triangleq \left( \mathbb{1}_H e_{\theta_k^*}^\top - I_H \right) D, \quad C_k \triangleq \begin{bmatrix} B_k \\ \mathbb{1}_S^\top \end{bmatrix}, \quad \tilde{y}_k \triangleq \begin{bmatrix} y_k \\ 1 \end{bmatrix}, \quad (17)$$

with  $e_m$  being an  $H \times 1$  vector with all zeros and a one in the  $m$ -th position, and where the last row in  $C_k$  and the last entry in  $\tilde{y}_k$  serve to embody the convexity constraint  $\sum_{s=1}^S \tilde{x}_{sk} = 1$ . We want

<sup>1</sup>We remark that  $\psi_{k,i}(\theta)$  and  $\mu_{k,i}(\theta)$  have the same limiting properties.

to examine the achievability of consistent topology learning. We study this problem under the assumption that the matrices  $D$  and  $B_k$  are known.<sup>2</sup> Now, achievability of consistent topology learning translates into the condition that the linear system in (16) admits a unique solution. We observe that the augmented matrix  $C_k$  is an  $(H+1) \times S$  matrix with an all-zeros row. Under the assumption that the global weight vector does not contain zeros, the following lemma can be proved (proof omitted due to space limitations).

**Lemma 1 (Necessary Condition for Topology Learning).** *The system in (16) admits a unique solution if, and only if,  $\text{rank}(C_k) = S$ . Thus, a necessary condition for topology learning is:*

$$H \geq S, \quad (18)$$

*i.e., that the number of hypotheses is at least equal to the number of sending sub-networks.* ■

Lemma 1 reveals a remarkable interplay between Social Learning (SL) and Topology Learning (TL). One interpretation of the condition  $H \geq S$  is that the TL problem becomes feasible when its complexity (number of sending components) is not greater than the SL complexity (number of hypotheses). Lemma 1 reveals also that TL consistency is not easily granted. For example, if the agents want to solve a binary detection problem, the maximum number of sending sub-networks that could allow consistent TL is  $S = 2$ .

### 3.1. Structured Gaussian Models

We now introduce a useful Gaussian model that can arise in many applications. We assume that all agents use the same family of likelihoods  $\{L(\theta)\}$ , for  $\theta = 1, 2, \dots, H$ , which are unit-variance Gaussian likelihoods with different means  $\{m_\theta\}$ . Each true distribution coincides with one of the likelihoods, which means that  $f^{(s)}$  is a unit-variance Gaussian distribution with mean  $\nu_s$  that is chosen among the means  $\{m_\theta\}$ , namely,  $\nu_s \in \{m_1, m_2, \dots, m_H\}$ . The sending sub-networks have different means. Without loss of generality, we assume that the sending sub-networks are numbered so that the means of the true distributions are  $\nu_1 = m_1, \dots, \nu_S = m_S$ , which implies that the divergence matrix  $D$  is equal to:

$$\frac{1}{2} \begin{bmatrix} 0 & (m_1 - m_2)^2 & \dots & (m_1 - m_S)^2 \\ (m_2 - m_1)^2 & 0 & \dots & (m_2 - m_S)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (m_H - m_1)^2 & (m_H - m_2)^2 & \dots & (m_H - m_S)^2 \end{bmatrix}. \quad (19)$$

For  $H = S$ , the matrix  $D$  is a Euclidean distance matrix (but for the constant  $1/2$ ) [15]. These matrices are constructed as follows. Given points  $r_1, r_2, \dots, r_L$ , belonging to  $\mathbb{R}^{\text{dim}}$ , the  $(i, j)$ -th entry of the matrix  $\text{EDM}(r_1, r_2, \dots, r_L)$  is given by the squared Euclidean distance between points  $r_i$  and  $r_j$ . We see then from (19) that, for  $H = S$ :

$$D = \frac{1}{2} \text{EDM}(m_1, m_2, \dots, m_H). \quad (20)$$

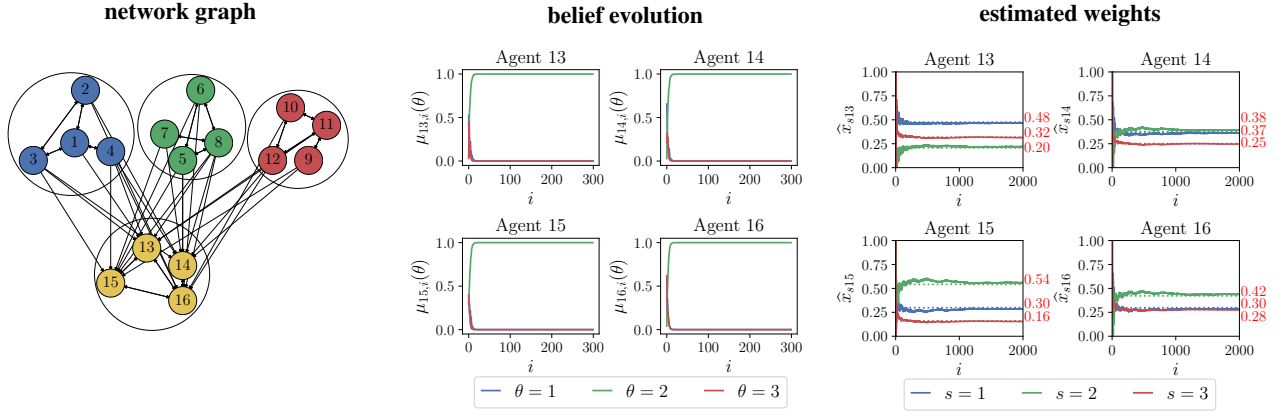
In the case  $H > S$ , the matrix  $D$  can be described as an *extended* Euclidean distance matrix:

$$D = \begin{bmatrix} E_S \\ F \end{bmatrix}, \quad E_H = \begin{bmatrix} E_S & F^\top \\ F & E_{H-S} \end{bmatrix}, \quad (21)$$

where:

$$\begin{aligned} E_S &\triangleq \frac{1}{2} \text{EDM}(m_1, m_2, \dots, m_S), \\ E_H &\triangleq \frac{1}{2} \text{EDM}(m_1, m_2, \dots, m_H), \\ E_{H-S} &\triangleq \frac{1}{2} \text{EDM}(m_{S+1}, m_{S+2}, \dots, m_H), \end{aligned} \quad (22)$$

<sup>2</sup> $B_k$  depends on  $\theta_k^*$ , which can be estimated consistently from  $y_k(\theta)$ .



**Fig. 2.** Randomly perturbed Gaussian model. *Leftmost panel.* Network topology. *Middle panel.* Belief convergence of receiving agents. *Rightmost panel.* Estimated limiting topology, with the red numbers denoting the true values  $\{x_{sk}\}$ .

and where  $F$  is the  $(H - S) \times S$  matrix with entries, for  $\theta = S + 1, S + 2, \dots, H$  and  $s = 1, 2, \dots, S$ :

$$[F]_{\theta s} = \frac{1}{2}(\mathfrak{m}_\theta - \mathfrak{m}_s)^2. \quad (23)$$

The following theorem ascertains the feasibility of the TL problem for the structured Gaussian model. The proof relies heavily on some fundamental properties of Euclidean distance matrices, and is omitted for space limitations.

**Theorem 1 (Topology Learning under Structured Gaussian Models).** *Let  $S \geq 2$  and  $H \geq S$ . Under the structured Gaussian model and Assumption 1, for all  $k \in \mathcal{R}$  we have that  $\text{rank}(C_k) = 2$*  ■

In view of Lemma 1, Theorem 1 implies that under the structured Gaussian model topology learning is very challenging, as it is feasible only when  $S = 2$ .

### 3.2. Diversity Models

Once ascertained that the topology over a structured Gaussian model is difficult to learn, we now examine the effect that diversity in the models of the sending sub-networks can have on TL. Differently from the previous section, we require that the entries of  $D$  are not tightly related, and we allow them to assume values in  $\mathbb{R}_+^{H \times S}$  ( $\mathbb{R}_+$  collects the nonnegative reals) with no structure linking them. As a formal way to embody this degree of variability in how the agents “see” the world, we model the divergences as jointly absolutely continuous random variables (bold notation  $\mathbf{d}_{\theta s}$ ). Under this framework, it is possible to establish the following result, whose proof is omitted for space constraints.

**Theorem 2 (Topology Learning under General Models with Diversity).** *Assume that the array  $\{\mathbf{d}_{\theta s}\}$  is made of random variables that are jointly absolutely continuous w.r.t. the Lebesgue measure on  $\mathbb{R}_+^{H \times S}$ . If  $H \geq S$ , Assumption 1 is verified and the matrix  $C_k$  is full column rank with probability 1, for all  $k \in \mathcal{R}$ .* ■

Theorem 2 reveals that *divergence configurations leading to a rank-deficient matrix  $C_k$  are rare* if sufficient diversity exists in the models of the sending components, i.e., the TL problem is feasible for most configurations.

## 4. ILLUSTRATIVE EXAMPLE

We show an example pertaining to Theorem 2, for a case with  $H = S = 3$ . The network topology is shown in the leftmost panel of Fig. 2. The true distribution of sub-network  $s = 1, 2, 3$  is a unit-variance Gaussian with mean  $s$ . The likelihood of the  $s$ -th sending sub-network, evaluated at hypothesis  $\theta$ , is unit-variance Gaussian with expectation  $\theta + \mathbf{u}_{\theta s}$ , with  $\mathbf{u}_{\theta s}$  being independent random variables uniformly distributed in  $[-0.1, 0.1]$ . The middle panel of Fig. 2 pertains to the SL problem, as it displays the convergence of the receiving agents’ beliefs. In the considered example, sub-network  $s = 2$  (green agents) exerts a domineering role, since the beliefs of the receiving agents converge to opinion  $\theta = 2$ .

We move on to the TL problem. First, for an observation time  $i$ , we construct the empirical data  $\hat{\gamma}_k(\theta) = (1/i) \log \psi_{k,i}(\theta)$ , and estimate  $\hat{\theta}_k^*$  as the value that maximizes  $\hat{\gamma}_k(\theta)$ . Then, we solve (16) with empirical matrices replacing the exact ones to estimate the connection-weight vector  $x_k$ . Provided that the system evolves for a sufficiently long time, this procedure allows to retrieve the true  $x_k$ , as shown in the rightmost panel of Fig. 2.

## 5. CONCLUSION

This work considered the following *dual* problem of social learning over weakly-connected networks. Given observation of *what* the agents are learning (Social Learning, SL), we want to discover *how* they are being influenced from the sending agents (Topology Learning, TL). We established that a necessary condition for consistent TL is that the number of hypotheses  $H$  is at least equal to the number of sending components  $S$ . In other words, the complexity of the TL problem (number of sub-networks) must be not greater than the complexity of the SL problem (number of hypotheses). We examined two models. A structured Gaussian model where *all* sending sub-networks use the same family of Gaussian likelihoods, and the true distributions are chosen within this family and are distinct across the sending sub-networks. We showed that for this model TL is feasible only when  $S = 2$ , due to the limited diversity across the sending sub-networks. Accordingly, we examined another model, where the likelihoods and the true distributions exhibit a certain *diversity*. For this case, we showed that the TL problem is feasible with probability one provided that  $H \geq S$ . In summary, the two critical features to enable consistent TL are: *more hypotheses than sending components and a sufficient degree of diversity*.

## 6. REFERENCES

- [1] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [2] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Business School Research Paper*, pp. 13–28, May 2013.
- [3] 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.
- [4] 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.
- [5] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, Sep. 2012.
- [6] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2012, pp. 709–713.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] V. Matta, A. Santos, and A. H. Sayed, "Exponential collapse of social beliefs over weakly-connected heterogeneous networks," in *Proc. IEEE ICASSP*, Brighton, UK, May 2019, pp. 5267–5271.
- [12] 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.
- [13] A. Santos, V. Matta, and A. H. Sayed, "Local tomography of large networks under the low-observability regime," *IEEE Trans. Inf. Theory*, available in early access, Oct. 2019, doi: 10.1109/TIT.2019.2945033.
- [14] 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.
- [15] 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.