



# Article Distributed Consensus Algorithms in Sensor Networks with Higher-Order Topology

Qianyi Chen<sup>1</sup>, Wenyuan Shi<sup>2</sup>, Dongyan Sui<sup>3,\*</sup> and Siyang Leng<sup>3,4</sup>

- <sup>1</sup> School of Information Science and Technology, Fudan University, Shanghai 200433, China
- <sup>2</sup> School of Mathematical Sciences, Fudan University, Shanghai 200433, China
- <sup>3</sup> Institute of AI and Robotics, Academy for Engineering and Technology, Fudan University, Shanghai 200433, China; syleng@fudan.edu.cn
- <sup>4</sup> Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China
- \* Correspondence: dysui22@m.fudan.edu.cn

Abstract: Information aggregation in distributed sensor networks has received significant attention from researchers in various disciplines. Distributed consensus algorithms are broadly developed to accelerate the convergence to consensus under different communication and/or energy limitations. Non-Bayesian social learning strategies are representative algorithms for distributed agents to learn progressively an underlying state of nature by information communications and evolutions. This work designs a new non-Bayesian social learning strategy named the hypergraph social learning by introducing the higher-order topology as the underlying communication network structure, with its convergence as well as the convergence rate theoretically analyzed. Extensive numerical examples are provided to demonstrate the effectiveness of the framework and reveal its superior performance when applying to sensor networks in tasks such as cooperative positioning. The designed framework can assist sensor network designers to develop more efficient communication topology, which can better resist environmental obstructions, and also has theoretical and applied values in broad areas such as distributed parameter estimation, dispersed information aggregation and social networks.

**Keywords:** distributed consensus algorithm; sensor network; higher-order topology; non-Bayesian social learning

## 1. Introduction

Sensor networks are widely deployed in environment for data gathering and monitoring purposes [1,2]. Originally motivated in military surveillance and then popularized to mobile and wireless communication, a large number of low-cost radars, endowed with communication capability, are distributedly located to fulfill tasks such as cooperative positioning and target recognition [3–7]. In typical applications, sensors collaborate to reach a consensus that accurately represents the correct classification of an event or the underlying true state of nature through a predefined communication network that is usually described by a connected graph and suffers from communication constraints [8]. Information processing in sensor networks has thus received significant attention from various disciplines such as computer science, signal processing and control theory [9,10].

The fusion-centric approaches assume each sensor has a communication link to a data fusion center, which is relatively energy consuming in large-scale networks [11]. Distributed consensus algorithms are then broadly developed to accelerate the convergence to consensus under different communication and/or energy limitations [12,13]. Social learning strategies, originating from social networks, are representative algorithms for distributed agents to learn progressively an underlying state of nature by information communications and evolutions [14]. Seminally, Banerjee [15] and Bikhchandani et al. [16] formulate the social learning paradigm in a fully Bayesian manner, which is further comprehensively analyzed and developed by Smith and Sørensen [17]. However, the Bayesian learning



Citation: Chen, Q.; Shi, W.; Sui, D.; Leng, S. Distributed Consensus Algorithms in Sensor Networks with Higher-Order Topology. *Entropy* **2023**, 25, 1200. https://doi.org/10.3390/ e25081200

Academic Editor: Shu-Chuan Chu

Received: 19 July 2023 Revised: 2 August 2023 Accepted: 8 August 2023 Published: 11 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). framework has a stringent requirement for a priori information and a high computational burden, preventing its usage even in simple networks [18]. Non-Bayesian algorithms are then introduced into social learning [19–22]; they have been flourishingly developing since the pioneering work of Jadbabaie et al. [14,23,24]. The non-Bayesian social learning framework consists of a *belief aggregation* step and a *Bayesian update* step, dealing with external communication and internal belief updating, respectively. In [25], the original model is extended using logarithmic aggregation, with its convergence and asymptotic normality being proved. Kar et al. [26] design a set of distributed parameter estimation algorithms by combining a consensus step and an innovation step in the update rule and apply them in sensor networks. The non-Bayesian social learning algorithms are subsequently generalized to various network structures. For example, Nedić et al. [27,28] analyze the learning rule for time-varying graphs, and the convergence result of a non-Bayesian learning algorithm for fixed graphs is provided in [29]. Authors in [30] consider the learning rules on weakly connected graphs, and social learning with time-varying weights is studied in [31]. Recently, a novel adaptive social learning strategy has been proposed by Bordignon et al. [32] to address the poor performance under nonstationary conditions. Sui et al. design a parametric social learning framework, introducing an agent stubbornness parameter to trade off the significance between its internal belief and external communication [33]. Notice that consensus networks, also known as opinion dynamics, have been extensively studied for many years [34,35]. Only paying attention to interpersonal communication, opinion dynamics focus on the formation, evolution and convergence of beliefs within a group [36,37]. Plentiful methods are proposed to judge/guarantee convergence in such networks. For example, a negative spectral gap is necessary to guarantee the convergence in the Influence Network (I-net) [38]. However, social learning frameworks proceed further than consensus networks by considering external communication and internal belief updating simultaneously, since they are more aligned with the reality of social networks. The difference also lies in the fact that social learning strategy converges to one of the preconceived hypotheses, while opinion dynamics converge to a certain consensus value.

Undoubtedly, the underlying network structure plays a key role in determining the convergence as well as its rate of the non-Bayesian social learning algorithms [39]. Practically, most sensors are physically restricted by their limited communication bandwidth, in which case they can only communicate with their local neighbors [8]. Moreover, reliability and connectivity are also constraints that the distributed consensus algorithms should address. In certain scenarios, a part of sensors is obstructed or misguided due to environmental factors. These drawbacks call for advanced design of the learning algorithms addressing more realistic circumstances. Real-world complex networks, including ecological networks [40], co-authorship networks [41] and communication networks [42], are proved to contain many-body interactions instead of merely pairwise interactions. In fact, a sum of two-body interactions is not able to capture the difference between the communication of three agents and three separate pairwise communications [43], where the former requires a smaller number of mutual interactions but with similar communication efficiency. For a real-world example in biochemistry, three proteins always interact with each other simultaneously and form a heterotrimer, which functions as a whole [44], and this can not be represented by traditional graphs. Hypergraph structure is then introduced into the non-Bayesian social learning strategies in this work, which extends the existing algorithms to more general situations and exhibits advantages not only in sensor networks, but also in other applications such as social networks.

In this work, we propose new non-Bayesian social learning algorithms based on higher-order topology called hypergraph social learning (HSL), whose convergence as well as the convergence rate are theoretically analyzed. Extensive numerical examples are provided to validate the effectiveness of the algorithms and unveil insights into the relationship between the consensus and hypergraph connectivity/structure. A practical technique for accelerating the convergence is designed and verified. We apply the proposed algorithms to a collective target location problem in sensor networks and analyze the difference with underlying graph and hypergraph structures. The remaining part of this paper is organized as follows: Section 2 provides a full description of the problem settings, reviewing necessary algorithms and proposing our HSL strategies. Section 3 presents sufficient assumptions/lemma and proves the convergence of the proposed algorithms. Section 4 provides extensive numerical examples illustrating the theoretical results and applies the algorithms in sensor networks. The findings are discussed and concluded in Section 5 with possible future work directions.

## 2. Preliminaries and Models

## 2.1. Problem Formulation

We consider a scenario where a group of *n* sensors or agents collaborate in order to arrive at a consensus that accurately represents the correct classification of an event or the underlying true state of nature, denoted as  $\theta^*$ , from a finite set of hypotheses  $\Theta = \{\theta_1, \theta_2, \dots, \theta_p\}$ . Each agent *i* receives an observation  $s_{i,t}$  of an environmental random process at each discrete time step  $t = 1, 2, \dots$ , where  $s_t = (s_{1,t}, s_{2,t}, \dots, s_{n,t})^\top$  is generated according to a given likelihood function  $\ell(\cdot|\theta^*)$ . The corresponding random variable of agent *i*'s observation at time *t* is denoted as  $S_{i,t}$  and  $S_t = (S_{1,t}, \dots, S_{n,t})^\top$ . Here, each  $S_{i,t}$ has its individual observation space  $S_i$  and is independently and identically distributed with respect to *t*.

The signal structure of agent *i* for state  $\theta$  is described by probability distribution  $\ell_i(\cdot|\theta)$ . In these settings,  $\ell_i(s_{i,t}|\theta)$  characterizes the probability that signal  $s_{i,t}$  can be observed by agent *i* at time *t* when it believes  $\theta$  is the true state. It is required that  $\ell_i(\cdot|\theta^*)$  coincides with the *i*th marginal distribution of  $\ell(\cdot|\theta^*)$ , which thus also describes the probability distribution of random variable  $S_{i,t}$ .

Usually, the agents interact with each other in a networked fashion, which is conventionally modelled by a directed graph describing pairwise interactions [45]. Due to a limited communication bandwidth, reliability and connectivity in a realistic scenario, most sensors are of the entailed constraint type to communicate with their local neighbors, and only a small number of delegates are endowed with mediate communication capacity. This brings in the need for higher-order topological structures that can capture the essence of local multi-agent interactions and centralized mediate communications [46,47]. In this work, hypergraph structure is introduced into the non-Bayesian social learning framework. The designed algorithms are demonstrated to significantly reduce the number of mutual interactions among sensors. A hypergraph is represented as  $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ . The set of vertices  $\mathcal{V} = \{1, 2, \dots, n\}$  denotes the *n* agents, and  $\mathcal{E} = \{e_i | i = 1, 2, \dots, m\}$  is the set of *m* hyperedges. Unlike a graph, a hyperedge may connect more than two nodes, leading to a different aggregation mechanism. We denote  $A = (a_{ii})_{m \times n}$  as the weight matrix with its row vector representing the weights of nodes in a single hyperedge. Moreover,  $B = (b_{ij})_{n \times m}$  denotes the weight matrix characterizing the weights of each node with respect to every hyperedge that it belongs to. Both A and B are assumed to be row stochastic, i.e.,  $\sum_{i=1}^{n} a_{ji} = 1, \forall j = 1, \cdots, m$ , and  $a_{ji} > 0$  if  $i \in e_j, \sum_{i=1}^{m} b_{ij} = 1, \forall i = 1, \cdots, n$ , and  $b_{ii} > 0$  if  $i \in e_i$ .

The *belief* of agent *i* at time *t* is denoted as  $\mu_{i,t}$ , which is a probability distribution over the set of possible states  $\Theta$ , i.e.,  $\sum_{k=1}^{p} \mu_{i,t}(\theta_k) = 1, \forall i = 1, \dots, n, \forall t = 0, 1, \dots$ . Here,  $\mu_{i,0}$  represents the *initial belief* of agent *i*. We further define the *belief of hyperedge*  $e_j$  at time *t* as  $\nu_{j,t}$  to record the belief aggregation from nodes to hyperedges, which is described detailedly in the following section.

#### 2.2. Hypergraph Social Learning

Traditional social learning generally consists of two key procedures at each time step for agents to update their beliefs, i.e., the Bayesian update step and the step of aggregating neighbor beliefs. Taking different orders of these two steps leads to the two basic social learning strategies called the LoAB (Logarithmic Aggregation and then Bayesian update) and the BLoA (Bayesian update and then Logarithmic Aggregation), respectively [48]. Notice that the hypergraph structure broadens the concept of neighbors, inducing the following new belief aggregation approaches.

At time step t + 1, the purpose of the Bayesian update step is to update each agent *i*'s prior belief  $\mu_{i,t}$  using environmental observation  $s_{i,t+1}$  and obtain its posterior belief  $\tilde{\mu}_{i,t+1}$ . This process can also be described as solving an optimization problem as follows:

$$\tilde{\mu}_{i,t+1} = \underset{f \in \mathbb{P}(\Theta)}{\arg\min} \{ D_{\mathrm{KL}}(f \parallel \mu_{i,t}) - \sum_{\theta \in \Theta} f(\theta) \log(\ell_i(s_{i,t+1}|\theta)) \}, \tag{1}$$

where  $D_{\text{KL}}(\cdot \| \cdot)$  is the Kullback–Leibler divergence (KL divergence) between two probability distributions. For instance, the first term on r.h.s. of (1) is  $D_{\text{KL}}(f \| \mu_{i,t}) = \sum_{\theta \in \Theta} f(\theta) \log \frac{f(\theta)}{\mu_{i,t}(\theta)}$ , describing the difference between the prior belief and the posterior belief. The second term on r.h.s. of (1) describes the maximum likelihood estimation given the latest observation  $s_{i,t+1}$ . By directly solving the optimization problem (1) with the Lagrange multiplier method, we obtain the update formula for the posterior belief as follows:  $\mu_{i,t}(\theta) \ell_i(s_{i,t+1}|\theta)$ 

$$\tilde{\mu}_{i,t+1}(\theta) = \frac{\mu_{i,t}(\theta)\ell_i(s_{i,t+1}|\theta)}{\sum\limits_{\theta'\in\Theta}\mu_{i,t}(\theta')\ell_i(s_{i,t+1}|\theta')}.$$
(2)

Inspired by the aggregation mechanism which is always used in Hypergraph Neural Networks (HGNN) [49,50], two steps are needed to achieve belief aggregation in hypergraph structures. The first step called *Node-to-Edge* is performed to aggregate the beliefs of all nodes in each hyperedge to form the defined belief of the hyperedge  $v_{j,t+1}$ . Then, the beliefs of hyperedges are aggregated to the nodes that are contained in their intersections, according to the weights in *B*, and this procedure is called *Edge-to-Node*. This two-step aggregation mechanism is consistent with the real-world prototype. For example, in a scenario of group discussions, participants exchange opinions to reach consensus within the group. If a person participates in multiple group meetings, they then update the personal belief by aggregating multiple consensus from different groups.

Practically, only performing a Bayesian update is not always feasible due to the identifiable problem that the agents encounter. This issue arises when the agents cannot differentiate between certain states based on their individual knowledge. Conversely, relying solely on the aggregation procedure poses a hindrance to agents in obtaining environmental information, consequently impeding their ability to reach consensus on the true state. By combining the Bayesian update step and the aggregation steps in hypergraphs in different orders, we derive two new hypergraph social learning algorithms, i.e.,

(a) HSL-NEB (Hypergraph Social Learning: Node-to-Edge, Edge-to-Node and then Bayesian update):

Step 1 (Node-to-Edge).

$$\nu_{j,t+1}(\theta) = \frac{\exp(\sum_{i=1}^{n} a_{ji} \log \mu_{i,t}(\theta))}{\sum_{\theta' \in \Theta} \exp(\sum_{i=1}^{n} a_{ji} \log \mu_{i,t}(\theta'))}, \forall \theta \in \Theta;$$
(3)

Step 2 (Edge-to-Node).

$$\tilde{\mu}_{i,t+1}(\theta) = \frac{\exp(\sum_{j=1}^{m} b_{ij} \log \nu_{j,t+1}(\theta))}{\sum_{\theta' \in \Theta} \exp(\sum_{j=1}^{m} b_{ij} \log \nu_{j,t+1}(\theta'))}, \forall \theta \in \Theta;$$
(4)

Step 3 (Bayesian update).

$$\mu_{i,t+1}(\theta) = \frac{\tilde{\mu}_{i,t+1}(\theta)\ell_i(s_{i,t+1}|\theta)}{\sum\limits_{\theta'\in\Theta}\tilde{\mu}_{i,t+1}(\theta')\ell_i(s_{i,t+1}|\theta')}, \forall \theta\in\Theta.$$
(5)

(b) HSL-BNE (Hypergraph Social Learning: Bayesian update, Node-to-Edge and then Edge-to-Node):

11

т

**Step 1** (Bayesian update).

$$\tilde{\mu}_{i,t+1}(\theta) = \frac{\mu_{i,t}(\theta)\ell_i(s_{i,t+1}|\theta)}{\sum\limits_{\theta' \in \Theta} \mu_{i,t}(\theta')\ell_i(s_{i,t+1}|\theta')}, \forall \theta \in \Theta;$$
(6)

Step 2 (Node-to-Edge).

$$\nu_{j,t+1}(\theta) = \frac{\exp(\sum_{i=1}^{n} a_{ji} \log \tilde{\mu}_{i,t+1}(\theta))}{\sum_{\theta' \in \Theta} \exp(\sum_{i=1}^{n} a_{ji} \log \tilde{\mu}_{i,t+1}(\theta'))}, \forall \theta \in \Theta;$$
(7)

Step 3 (Edge-to-Node).

$$\mu_{i,t+1}(\theta) = \frac{\exp(\sum_{j=1}^{m} b_{ij} \log \nu_{j,t+1}(\theta))}{\sum_{\theta' \in \Theta} \exp(\sum_{j=1}^{m} b_{ij} \log \nu_{j,t+1}(\theta'))}, \forall \theta \in \Theta.$$
(8)

## 3. Assumptions and Results

As widely discussed in previous works of social learning, we care about the convergence of the algorithms as well as the rate of convergence. The following assumptions are required to ensure the convergence of our HSL strategies:

**Assumption 1** (Communication network). Denote  $C = (c_{ij})_{n \times n} = BA$ , where  $A = (a_{ji})_{m \times n}$ and  $B = (b_{ij})_{n \times m}$ . The matrix C satisfies that it is the transition matrix of an irreducible, aperiodic Markov chain of finite states.

To satisfy the assumption, hypergraph  $\mathcal{H}$  should be a connected hypergraph without isolated vertices (a hypergraph is connected if for any pair of vertices, there is a path which connects these vertices; see Appendix A for detailed explanations). We recall the following lemma [51]:

**Lemma 1.** If a Markov chain of finite states is irreducible, then it has a unique stationary distribution  $\pi$ . Let *C* be the transition matrix of the Markov chain and further suppose it is aperiodic; then, we have  $\lim_{k\to\infty} [C^k]_{ij} = \pi_j$ , for  $1 \leq i, j \leq n$ .

The stationary distribution  $\pi$  can be interpreted as the normalized left eigenvector of *C* with respect to eigenvalue 1, which is known as the *eigenvector centrality* in related literatures. The Perron–Frobenius theorem ensures that all components of  $\pi$  are strictly positive.

**Assumption 2** (Belief and signal structure). For all agents  $i = 1, 2, \dots, n$ , (a) they have positive initial beliefs on all states, i.e.,  $\mu_{i,0}(\theta) > 0$  for all  $\theta \in \Theta$ ; (b) they have positive signal structures, i.e.,  $\ell_i(s_i|\theta) > 0$  for all  $s_i \in S_i$  and  $\theta \in \Theta$ .

Notice that if the initial belief of agent *i* on state  $\theta$  is zero, following our HSL algorithms, its belief remains at zero all the time. In this case,  $\theta$  is meaningless for agent *i*, and we thus eliminate the situation by imposing Assumption 2a. For the signal structures and Assumption 2b, the same explanation can be applied.

Two states,  $\theta_i$  and  $\theta_k$ , are called *observationally equivalent* for agent *i* if  $\ell_i(s_i|\theta_j) = \ell_i(s_i|\theta_k)$ ,  $\forall s_i \in S_i$ , in which case agent *i* is not able to distinguish these states only with its own information. Moreover, the true state is called *globally identifiable* if the set  $\Theta^* = \bigcap_{i=1}^n \Theta_i^*$  has only one element  $\theta^*$ , where  $\Theta_i^* = \{\theta \in \Theta | \ell_i(s_i|\theta) = \ell_i(s_i|\theta^*), \forall s_i \in S_i\}$ . This concept can be intuitively explained. If state  $\hat{\theta}$  is observationally equivalent to  $\theta^*$  for all agents, i.e.,  $\Theta^* = \{\theta^*, \hat{\theta}\}$ , then the two states are exactly the same from the view of all agents and they cannot identify the true state progressively and collectively, which, in addition, induces

**Assumption 3** (Globally identifiable). *The true state*  $\theta^*$  *is globally identifiable.* 

Under this assumption, for all  $\theta \neq \theta^*$ , there exists at least agent *i* satisfying the fact that  $D_{\text{KL}}(\ell_i(\cdot|\theta^*) \parallel \ell_i(\cdot|\theta))$  is strictly positive.

In the following, we denote  $K_i(\theta^*, \theta) = D_{\text{KL}}(\ell_i(\cdot|\theta^*) \parallel \ell_i(\cdot|\theta))$  and define a probability triple  $(\Omega, \mathcal{F}, \mathbb{P}^*)$ , where  $\Omega = \{\omega \mid \omega = (s_1, s_2, \cdots)\}$ ,  $\mathcal{F}$  is the  $\sigma$ -algebra generated by the observations, and  $\mathbb{P}^*$  is the probability measure induced by paths in  $\Omega$ , i.e.,  $\mathbb{P}^* = \prod_{t=1}^{\infty} \ell(\cdot|\theta^*)$ .

 $\mathbb{E}^*[\cdot]$  is used to denote the expectation operator associated with probability measure  $\mathbb{P}^*$ . Now, we can state the main results describing the convergence of the HSL strategies.

**Theorem 1.** Under Assumptions 1-3, the update rules (3)-(5) and (6)-(8) satisfy the following properties:

$$\lim_{t \to \infty} \frac{1}{t} \log \frac{\mu_{i,t}(\theta)}{\mu_{i,t}(\theta^*)} = -\sum_{j=1}^n \pi_j K_j(\theta^*, \theta), \quad \forall \theta \neq \theta^*$$
(9)

and

$$\lim_{t \to \infty} \mu_{i,t}(\theta^*) = 1, \quad \mathbb{P}^* - \text{a.s.}, \quad \forall i = 1, \cdots, n.$$
(10)

**Proof.** We consider the update rule (3)–(5) first. For each agent *i* and  $\theta \neq \theta^*$ , we have

$$\log \frac{\mu_{i,t+1}(\theta)}{\mu_{i,t+1}(\theta^*)} = \log \frac{\tilde{\mu}_{i,t+1}(\theta)}{\tilde{\mu}_{i,t+1}(\theta^*)} + \log \frac{\ell_i(s_{i,t+1}|\theta)}{\ell_i(s_{i,t+1}|\theta^*)} \\ = \sum_{j=1}^m b_{ij} \sum_{k=1}^n a_{jk} \log \frac{\mu_{k,t}(\theta)}{\mu_{k,t}(\theta^*)} + \log \frac{\ell_i(s_{i,t+1}|\theta)}{\ell_i(s_{i,t+1}|\theta^*)}.$$

By denoting  $\varphi_{i,t+1}(\theta) = \log \frac{\mu_{i,t+1}(\theta)}{\mu_{i,t+1}(\theta^*)}$  and  $L_{i,t+1}(\theta) = \log \frac{\ell_i(s_{i,t+1}|\theta)}{\ell_i(s_{i,t+1}|\theta^*)}$ , the above equation simplifies to

$$\varphi_{i,t+1}(\theta) = \sum_{j=1}^{m} b_{ij} \sum_{k=1}^{n} a_{jk} \varphi_{k,t}(\theta) + L_{i,t+1}(\theta).$$
(11)

We rewrite (11) in the matrix form:

$$\boldsymbol{\varphi}_{t+1}(\theta) = BA\boldsymbol{\varphi}_t(\theta) + \boldsymbol{L}_{t+1}(\theta).$$

Then, it follows that

$$\begin{aligned} \frac{1}{t}\boldsymbol{\varphi}_{t+1}(\theta) &= \frac{1}{t}BA\boldsymbol{\varphi}_t(\theta) + \frac{1}{t}\boldsymbol{L}_{t+1}(\theta) \\ &= \frac{1}{t}BA(BA\boldsymbol{\varphi}_{t-1}(\theta) + \boldsymbol{L}_t(\theta)) + \frac{1}{t}\boldsymbol{L}_{t+1}(\theta) = \cdots \\ &= \frac{1}{t}(BA)^{t+1}\boldsymbol{\varphi}_0(\theta) + \frac{1}{t}\sum_{k=1}^t (BA)^k \boldsymbol{L}_{t+1-k}(\theta) + \frac{1}{t}\boldsymbol{L}_{t+1}(\theta). \end{aligned}$$

Using the notation in Assumption 1, the above equation can be written as

$$\frac{1}{t}\boldsymbol{\varphi}_{t+1}(\theta) = \frac{1}{t}C^{t+1}\boldsymbol{\varphi}_0(\theta) + \frac{1}{t}\sum_{k=1}^t C^k \boldsymbol{L}_{t+1-k}(\theta) + \frac{1}{t}\boldsymbol{L}_{t+1}(\theta),$$
(12)

where *C* is an  $n \times n$  matrix. The assumptions admit that the first and the third terms on the right hand side of (12) converge to zero as  $t \to \infty$ . And the second term can be deformed as

$$\frac{1}{t} \sum_{k=1}^{t} C^{k} L_{t+1-k}(\theta) = \frac{1}{t} \sum_{k=1}^{t} (C^{k} - \mathbf{1}_{n} \pi) L_{t+1-k}(\theta) + \frac{1}{t} \sum_{k=1}^{t} \mathbf{1}_{n} \pi (L_{t+1-k}(\theta) + K(\theta^{*}, \theta)) - \frac{1}{t} \sum_{k=1}^{t} \mathbf{1}_{n} \pi K(\theta^{*}, \theta),$$
(13)

where  $\mathbf{1}_n$  is an *n*-dimensional column vector of ones. From Lemma 1, we know that  $\lim_{k\to\infty} C^k = \mathbf{1}_n \pi$ . Noticing that all elements of  $C^k(k = 1, 2, \dots)$  are bounded, the first term on the right hand side of (13) converges to zero as  $t \to \infty$ . Moreover,

$$\mathbb{E}^*[L_{i,t}(\theta)] = \mathbb{E}^*[\log \frac{\ell_i(s_{i,t}|\theta)}{\ell_i(s_{i,t}|\theta^*)}] = \int_{s \in S_i} \ell_i(s|\theta^*) \log \frac{\ell_i(s|\theta)}{\ell_i(s|\theta^*)} ds$$
$$= -D_{\mathrm{KL}}(\ell_i(\cdot|\theta^*) \parallel \ell_i(\cdot|\theta)) = -K_i(\theta^*, \theta).$$

Using the Kolmogorov's strong law of large numbers, it follows that

$$\frac{1}{t}\sum_{k=1}^{t} \boldsymbol{L}_{t+1-k}(\boldsymbol{\theta}) - \frac{1}{t}\sum_{k=1}^{t} \mathbb{E}^*[\boldsymbol{L}_{t+1-k}(\boldsymbol{\theta})] \to 0, \quad \mathbb{P}^*-\text{a.s.},$$

as  $t \to \infty$ , which leads to

$$\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^t\mathbf{1}_n\pi(\boldsymbol{L}_{t+1-k}(\theta)+\boldsymbol{K}(\theta^*,\theta))=0,\quad\mathbb{P}^*-\text{a.s.}$$

Now, (13) provides

$$\lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} C^k L_{t+1-k}(\theta) = -\mathbf{1}_n \pi K(\theta^*, \theta), \quad \mathbb{P}^* - \text{a.s.}.$$
(14)

Therefore, property (9) holds, which can be directly induced from (12) and (14). Moreover, it follows that with probability one, for any  $\epsilon > 0$ , there exists an integer *T* such that  $\forall t > T$  and  $\forall \theta \neq \theta^*$ ,

$$\left|\frac{1}{t}\log\frac{\mu_{i,t}(\theta)}{\mu_{i,t}(\theta^*)}+\sum_{j=1}^n\pi_jK_j(\theta^*,\theta)\right|<\epsilon.$$

Noticing that  $\sum_{\theta \neq \theta^*} \mu_{i,t}(\theta) = 1 - \mu_{i,t}(\theta^*)$ , we have

$$\frac{1}{1 + \sum_{\theta \neq \theta^*} \exp\left(\left(\epsilon - \sum_{j=1}^n \pi_j K_j(\theta^*, \theta)\right) t\right)} < \mu_{i,t}(\theta^*) \leq 1.$$

Letting  $t \to \infty$ , property (10) is then proved because of the arbitrary selection of  $\epsilon$ .

For update rules (6)–(8), we can reach the same conclusions (9) and (10) via similar analysis.  $\Box$ 

Theorem 1 indicates that all agents interacting in the hypergraph structure eventually learn the underlying true state as long as the assumptions are satisfied. We provide extensive examples to validate the theoretical analyses and unveil more insights in the next section. Moreover, the following corollary describing the convergence rate can be obtained directly from Theorem 1:

**Corollary 1.** Under Assumptions 1–3, the update rules (3)–(5) and (6)–(8) satisfy that for all  $i = 1, 2, \dots, n$  and all  $\theta \neq \theta^*$ ,

$$\lim_{t\to\infty}\mu_{i,t}(\theta)\leqslant\exp(-\alpha_{\theta}t),\quad\mathbb{P}^*-\text{a.s.,}$$

where  $\alpha_{\theta} = \sum_{j=1}^{n} \pi_{j} K_{j}(\theta^{*}, \theta).$ 

In the following, the set of hyperedges where node *i* resides is denoted as  $d_i$ . We use  $|e_j|$  to denote the number of nodes being connected in the *j*th hyperedge and use  $|d_i|$  to denote the number of hyperedges where the *i*-th node resides. According to Assumption 1 and Lemma 1, C = BA has a unique stationary distribution  $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ , i.e.,  $\pi C = \pi$ . Under specific settings of the weight matrices *A* and *B*, we have further results summarized in the following remark:

**Remark 1.** If we consider a special case of A and B as follows,

$$a_{ji} = \begin{cases} \frac{1}{|e_j|}, & i \in e_j, \\ 0, & \text{otherwise,} \end{cases} \quad b_{ij} = \begin{cases} \frac{1}{|d_i|}, & i \in e_j, \\ 0, & \text{otherwise,} \end{cases}$$

we have  $\pi_i = rac{|d_i|}{\sum_{l=1}^n |d_l|}$ 

**Proof.** From the definitions, it is obvious that *A* and *B* are both row-stochastic matrices. It follows that C = BA is also row stochastic, i.e.,

$$\sum_{i=1}^n c_{ij} = 1, \quad \forall i = 1, \cdots, n.$$

The above equation can be rewritten as

$$\sum_{j=1}^{n} \sum_{e_k \in d_i \cap d_j} b_{ik} a_{kj} = \sum_{j=1}^{n} \sum_{e_k \in d_i \cap d_j} \frac{1}{|d_i|} \frac{1}{|e_k|} = 1, \quad \forall i = 1, \cdots, n.$$
(15)

Denoting  $\pi' = \pi C = \pi BA$ , it follows that

$$\pi'_{k} = \sum_{i=1}^{n} \pi_{i} c_{ik} = \sum_{i=1}^{n} \pi_{i} \sum_{j=1}^{m} b_{ij} a_{jk} = \sum_{i=1}^{n} \pi_{i} \sum_{e_{j} \in d_{i} \cap d_{k}} \frac{1}{|d_{i}|} \frac{1}{|e_{j}|}.$$
(16)

Substituting  $\pi_i = \frac{|d_i|}{\sum_{l=1}^n |d_l|}$  into the above equation provides

$$\pi'_{k} = \frac{1}{\sum\limits_{l=1}^{n} |d_{l}|} \sum_{i=1}^{n} |d_{i}| \sum_{e_{j} \in d_{i} \cap d_{k}} \frac{1}{|d_{i}|} \frac{1}{|e_{j}|} = \frac{1}{\sum_{l=1}^{n} |d_{l}|} \sum_{i=1}^{n} \sum_{e_{j} \in d_{i} \cap d_{k}} \frac{1}{|e_{j}|}.$$
(17)

From (15) and (17) and the uniqueness of the stationary distribution, we obtain

$$\pi'_{k} = \frac{|d_{k}|}{\sum_{l=1}^{n} |d_{l}|} = \pi_{k}.$$
(18)

Remark 1 indicates that increasing the degree  $|d_i|$  of the *i*th node also raises the value of  $\pi_i$ , demonstrating a greater importance of the role of node *i*. This mechanism provides us with viable solutions to accelerate the convergence rate by adjusting the hypergraph structure on nodes that have high eigenvector centrality, which is shown in the following examples.

#### 4. Numerical Examples and Applications

Here, we provide several numerical examples demonstrating the above theoretical analyses and apply the HSL algorithms in sensor networks.

### 4.1. Hypergraph Connectivity vs. Convergence

We first demonstrate the effectiveness of the HSL algorithms in reaching distributed consensus and the relationship with the underlying hypergraph connectivity. The structure of an unconnected hypergraph consisting of seven vertices and three hyperedges is depicted in Figure 1a. The non-zero values in the weight matrices *A* and *B* are randomly generated from (0, 1) and then normalized to satisfy the row-stochastic assumption. We assume there are three possible states  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , with  $\theta_1$  being the true state. The initial beliefs are also uniformly generated from interval (0, 1) and subject to

 $\sum_{k=1}^{5} \mu_{i,0}(\theta_k) = 1, \forall i = 1, \cdots, 7.$  Signals are assumed to be generated at each time from set

{H, T} and according to the probability distribution of  $\ell(H|\theta^*) = 0.2$  and  $\ell(T|\theta^*) = 0.8$ . Moreover, signal structures are set as  $\ell_i(H|\theta_1) = \ell_i(H|\theta_2) = 0.2$ ,  $\ell_i(T|\theta_1) = \ell_i(T|\theta_2) = 0.8$ ,  $\ell_i(H|\theta_3) = 0.6$ , and  $\ell_i(T|\theta_3) = 0.4$ , for i = 1, 2, 3. And for i = 4, 5, 6, 7, we assume that  $\ell_i(H|\theta_2) = 0.7$ ,  $\ell_i(T|\theta_2) = 0.3$ ,  $\ell_i(H|\theta_1) = \ell_i(H|\theta_3) = 0.2$ , and  $\ell_i(T|\theta_1) = \ell_i(T|\theta_3) = 0.8$ . To be clear, we denote the discrete probability distribution that agents 1–3 follow as  $\begin{bmatrix} 0 & 2 & 0 & 2 \\ 0 & 2 & 0 & 6 \end{bmatrix}$ 

 $L_{1-3} = \begin{bmatrix} 0.2 & 0.2 & 0.6 \\ 0.8 & 0.8 & 0.4 \end{bmatrix}$ , while agents 4–7 follow discrete probability distribution

 $L_{4-7} = \begin{bmatrix} 0.2 & 0.7 & 0.2 \\ 0.8 & 0.3 & 0.8 \end{bmatrix}$ . In this example, successful learning is considered to be reached if

 $\sum_{i=1}^{\prime} |\mu_{i,t}(\theta^*) - 1| \leq 10^{-3}$ . If the collective consensus is not reached after 100 iterations, the

learning ends on its own. Results in Figure 1b,c illustrate that an unconnected hypergraph structure may finally result in inconsistent learning results, which means the HSL strategies may not converge in such a case. This phenomenon accords with the intuition that completely separate groups may not reach a consensus.



Figure 1. Cont.



**Figure 1.** (**a**,**d**) The unconnected and connected hypergraph structures used in this example. (**b**,**c**) The evolution of beliefs on all possible states for agents interacting in structure of (**a**), with (**b**,**c**) denoting the results of HSL-NEB and HSL-BNE, respectively. (**e**,**f**) The evolution of beliefs on all possible states for agents interacting in structure of (**d**), with (**e**,**f**) denoting the results of HSL-NEB and HSL-BNE, respectively. (**e**,**f**) denoting the results of HSL-NEB and HSL-BNE, respectively.

However, by only adding a hyperedge linking Agent 1 and Agent 4—see Figure 1d the hypergraph becomes connected. HSL strategies soon converge after a few iterations see Figure 1e,f—validating the theoretical results in Theorem 1 and the effectiveness of the proposed algorithms.

# 4.2. Hypergraph Structure vs. Convergence Rate

As indicated in Theorem 1 and Corollary 1, the convergence rate is closely related to the eigenvector centrality of matrix *C* as well as the KL-divergence  $K(\theta^*, \theta)$ . This offers us the opportunity to increase the convergence rate via assigning a more important role, i.e., a larger eigenvector centrality, to the agent that is more informative, i.e., more helpful to distinguish between true and wrong states. We illustrate this idea by considering two similar hypergraphs shown in Figure 2a,b with four agents and three hyperedges. We define the weight matrices *A* and *B* in the same way as in Remark 1, i.e., for the hypergraph in Figure 2a,

$$A_{1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0\\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3}\\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad B_{1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

and for the hypergraph in Figure 2b,

$$A_{\rm r} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad B_{\rm r} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Clearly, Agents 1 and 4 have the largest eigenvector centrality, respectively, for hypergraphs in Figure 2a,b, demonstrating their important roles in the structures. We assume that there are two possible states  $\Theta = \{\theta_1, \theta_2\}$ , with  $\theta_1$  being the true state. The initial beliefs are uniformly generated from the interval (0, 1) and subject to  $\sum_{k=1}^{2} \mu_{i,0}(\theta_k) = 1, \forall i = 1, \dots, 4$ . At each time step *t*, signal *s*<sub>t</sub> is randomly generated following normal distribution  $\mathcal{N}(0, 1)$  and observed by all agents. As  $\theta_1$  is the underlying true state, from assumptions, we have  $\ell_i(\cdot|\theta_1) = \mathcal{N}(0, 1), \forall i = 1, \dots, 4$ . The likelihood functions of the other state are

have  $\ell_i(\cdot|\theta_1) = \mathcal{N}(0,1), \forall i = 1, \dots, 4$ . The likelihood functions of the other state are assigned as  $\ell_i(\cdot|\theta_2) = \mathcal{N}(\frac{i}{10},1), \forall i = 1, \dots, 4$ , while  $\theta^* = \theta_1$  is globally identifiable. These settings lead to  $K_4(\theta^*,\theta) > K_3(\theta^*,\theta) > K_2(\theta^*,\theta) > K_1(\theta^*,\theta) > 0$ , depicting that Agent 4 is the most informative agent and Agent 1 is the least informative one. We focus on the number of iterations for update rules HSL-NEB and HSL-BNE to collectively learn the

true state, where successful learning is considered to be reached if  $\sum_{i=1}^{4} |\mu_{i,t}(\theta^*) - 1| \le 10^{-3}$ .

Results in Figure 2c,d illustrate that consensus can be significantly faster reached with the structure in Figure 2b, where Agent 4 has both higher eigenvector centrality and is more informative, showing consistency with the theoretical analyses. Moreover, from Lemma 1, the convergence rate of the proposed algorithms and the convergence rate of transition matrix *C* are related to the second largest eigenvalue of *C*. That is to say, the convergence rate depends on the network sparsity (and the signal structures) instead of its size, which means no scaling problem exists for very large networks. However, practically, due to computational burden, we would like to accelerate the convergence rate to reach consensus.



**Figure 2.** (**a**,**b**) The two hypergraph structures used in this example. (**c**,**d**) The evolution of beliefs on all possible states, with (**c**,**d**) denoting the results of HSL-NEB and HSL-BNE, respectively.

## 4.3. Hypergraph Structure vs. Graph Structure

Here, we provide a preliminary survey on the relationship between traditional social learning on graphs and our HSL framework. As presented in Appendix A.2, every hypergraph  $\mathcal{H}$  can be mapped to a corresponding graph  $\mathcal{G}$  by connecting all pairs of nodes that belong to each hyperedge. This method is called *clique expansion* in the field of hypergraph representation learning [52]. We demonstrate the equivalence of a hypergraph and its clique expansion when performing HSL and traditional social learning, respectively, on them theoretically in Appendix A.2 and numerically as described further.

We consider a hypergraph with six vertices and three hyperedges; the structure is depicted in Figure 3a. We define the weight matrices A and B in the same way as in Remark 1, i.e.,

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2}\\ 1 & 0 & 0\\ 1 & 0 & 0\\ \frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & 1 \end{bmatrix}.$$



#### In this way, the matrix C = AB can be written as

**Figure 3.** (**a**,**b**) The hypergraph and graph structures used in this example. (**c**,**d**) The evolution of beliefs on all possible states for agents interacting in structures of (**a**,**b**), respectively, with the HSL-NEB algorithm.

The structure of the clique expansion is shown in Figure 3b, and we assume the weight matrix of this graph is identical to *C*. AWe asume there are four possible states  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$  with true state  $\theta^* = \theta_1$ . The initial beliefs are uniformly generated from the interval (0,1) and subject to  $\sum_{k=1}^{4} \mu_{i,0}(\theta_k) = 1, \forall i = 1, \cdots, 6$ . At each time step t, signal  $s_t$  is randomly generated, following normal distribution  $\mathcal{N}(0,1)$  and observed by all agents. As  $\theta_1$  is the underlying true state, from assumptions, we have  $\ell_i(\cdot|\theta_1) = \mathcal{N}(0,1), \forall i = 1, \cdots, 6$ . Moreover, all agents in the network are assumed to be equivalently informative to the true state, with  $\ell_i(\cdot|\theta_k) = \mathcal{N}(\frac{k-1}{5}, 1), \forall i = 1, \cdots, 6$  and  $\forall k = 2, 3, 4$ . For the hypergraph, we use the proposed update rules HSL-NEB and HSL-BNE, while the LoAB and BLoA algorithms are applied to the graph. In this example, we still regard  $\sum_{i=1}^{6} |\mu_{i,t}(\theta^*) - 1| \leq 10^{-3}$ , the same criterion as in the previous examples as the sign of achieving successful social learning.

Results in Figure 3c,d demonstrate that convergence to collective consensus are reached within almost the same number of iterations for the two cases, i.e., equivalent convergence rate. However, compared to the pairwise communications in graphs, which occur many times in one iteration, HSL significantly reduces the number of mutual interactions among agents and keeps the same convergence rate.

### 4.4. Application to Sensor Cooperative Positioning

In the following example, we demonstrate the applicability of our HSL algorithms in a sensor cooperative positioning problem [29]. We consider a radar network with six sensors located at  $(\pm 1, 0, 0)$ ,  $(0, \pm 1, 0)$  and  $(0, 0, \pm 1)$ . The sensors can communicate according to three hyperedge connections, as depicted in Figure 4a, where the *mediately communicating sensor* refers to the sensor belonging to multiple hyperedges, representing its role of communicating across hyperedges, and the *locally communicating sensor* denotes the sensor only communicating in one hyperedge. With this setting, the non-zero elements in the weight matrices *A* and *B* are randomly generated from the interval (0, 1) and then normalized to satisfy the row-stochastic assumption. We assume each sensor can sense the target's location along one dimension only, whereas the target location is a point in a three-dimensional space. Specifically, sensors located on the *x*-axis can sense whether the *x*-coordinate of the target lies in the (-1, 0) or [0, 1) or  $(-\infty, -1] \cup [1, \infty)$  interval. Similarly, sensors on the *y*-axis and the *z*-axis can each distinguish between three distinct non-intersecting intervals on the corresponding axis. Therefore, the total number of possible states is nine, i.e.,  $\Theta = \{\theta_1, \theta_2, \dots, \theta_9\}$ , where  $\theta_9$  denotes the outside infinite region.

The goal is to ascertain the area of the target aircraft by the HSL algorithms. We set  $\theta_1$  as the true position of the target. The initial beliefs are generated from the interval (0, 1) and subject to  $\sum_{k=1}^{9} \mu_{i,0}(\theta_k) = 1, \forall i = 1, \cdots, 6$ . Signals are assumed to be generated at each time from set {NEAR, MID, FAR}. Sensors are assumed to be more sensitive to near target, e.g.,

for sensor 5 on the positive *z*-axis. We set  $\begin{bmatrix} \ell_5(\text{NEAR}|\theta_k) \\ \ell_5(\text{MID}|\theta_k) \\ \ell_5(\text{FAR}|\theta_k) \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.05 \\ 0.05 \end{bmatrix}$  for k = 1, 2, 3, 4, and

 $\begin{bmatrix} \ell_{5}(\text{NEAR}|\theta_{k})\\ \ell_{5}(\text{MID}|\theta_{k})\\ \ell_{5}(\text{FAR}|\theta_{k}) \end{bmatrix} = \begin{bmatrix} 0.1\\ 0.85\\ 0.05 \end{bmatrix} \text{ for } k = 5, 6, 7, 8. \text{ Here, } z\text{-coordinates of } \theta_{1} \sim \theta_{4} \text{ lie in the interval}$ 

 $\begin{bmatrix} 0,1 \end{bmatrix}, \text{ and } (-1,0) \text{ for } \theta_5 \sim \theta_8. \text{ We additionally set } \begin{bmatrix} \ell_i(\text{NEAR}|\theta_9) \\ \ell_i(\text{MID}|\theta_9) \\ \ell_i(\text{FAR}|\theta_9) \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.8 \end{bmatrix}, \forall i = 1, 2, \cdots, 6.$ 

For other sensors, the signal structures are set up with the same logic. We further assume that the signals are generated following the same probability distribution as the signal

structure of  $\theta_1$ . Successful learning is considered to be reached if  $\sum_{i=1}^{6} |\mu_{i,t}(\theta^*) - 1| \le 10^{-3}$ .

Results in Figure 4b,c illustrate that the true collective consensus can be reached in less than 20 iterations, showing that the sensor network completes the positioning task efficiently and accurately by applying our HSL algorithms. In real scenarios, radar networks always suffer from resource waste due to redundant and repetitive communications especially when using traditional graph-based social learning algorithms. However, in our HSL framework, for example, in the sensor network considered here, the communication can be further reduced to one hyperedge covering sensors on x, y and z-axes, which is enough to guarantee successful positioning.



**Figure 4.** (a) The radar network with hypergraph topology considered in this example. (b,c) The evolution of beliefs on selected states, with (b,c) denoting the results of HSL-NEB and HSL-BNE, respectively.

### 4.5. Application to Consensus in Social Network

We validate the effectiveness of our HSL algorithms in a large-scale real-world higherorder social network, which is often used to model opinion formation [53]. Network topology is determined from public dataset Hypertext2009 provided by the SocioPatterns research collaboration (http://www.sociopatterns.org/). Three-body interactions are reconstructed by Wang et al. [54]. This dataset describes the scenario in which 85 participants exchange their views in a conference in order to form a consensus, e.g., to support or oppose a bill. We assume environmental information can be received by all participants and influences their beliefs. The network structure, consisting of 85 vertices and 225 hyperedges (192 edges + 33 triangles), are given in Figure 5a. The non-zero values in the weight matrices A and B are randomly generated from (0, 1) and then normalized to satisfy the rowstochastic assumption. We assume there are two possible states  $\Theta = \{\theta_1, \theta_2\}$ , with  $\theta_1$  being the true state. The initial beliefs are also uniformly generated from the interval (0, 1) and subject to  $\mu_{i,0}(\theta_1) + \mu_{i,0}(\theta_2) = 1, \forall i = 1, \dots, 85$ . Signals are assumed to be generated randomly, following the Gaussian distribution  $\mathcal{N}(0, 1)$  and observed by all agents. As  $\theta_1$  is the underlying true state, we have  $\ell_i(\cdot|\theta_1) = \mathcal{N}(0,1), \forall i = 1, \cdots, 85$ . Moreover, the likelihood functions of the other state are assigned as  $\ell_i(\cdot|\theta_2) = \mathcal{N}(\frac{i}{1000}, 1), \forall i = 1, \cdots, 85$ . We still use  $\sum_{i=1}^{\infty} |\mu_{i,t}(\theta^*) - 1| \leq 10^{-3}$  as the sign of successful social learning. Results in Figure 5b,c

illustrate that influenced by environmental signals and interpersonal communications, all participants finally reach a consensus and make the right decision, demonstrating the broad applicability of our HSL strategies.



**Figure 5.** (a) Visualization of the real-world hypergraph network Hypertext2009. Shaded triangles denote three-body interactions in the dataset. (b,c) The evolution of beliefs on all possible states, with (b,c) denoting the results of HSL-NEB and HSL-BNE, respectively.

#### 5. Discussion and Conclusions

Practically, in a sensor network, units may inevitably encounter sudden disconnection due to obstruction, prolonged operations, environmental damages, etc. [55,56]. Here, we discuss a scenario when a sensor loses its ability to sense environmental information. Following the design of the hypergraph considered in Section 4.3, including the network structure, weight matrices, possible states, signal structures, etc., we further assume one of the sensors is blocked from external information and examine the convergence rate, respectively. As shown in Figure 6, where node "0" represents the plain result with original settings and others represent the results with the corresponding node being blocked, the numbers of iterations to reaching consensus all increase when one sensor loses sensing ability. Interestingly, the sensors exhibit different roles in the network; the blocking of Sensors 1, 4 and 5 located at a mediate position across the hyperedges has significantly higher numbers of iterations than the blocking of Sensors 2, 3 and 6 lying only in one hyperedge. This phenomenon reveals that the mediately communicating sensors are more sensitive to environmental obstruction and should be providently cared. In future works, we will comprehensively discuss the relationship between network structure and convergence efficiency.



**Figure 6.** The histogram showing the number of iterations to reaching consensus when the corresponding node is blocked from sensing external information. Here, node "0" represents the plain result with original settings. Error bars denote the standard deviations over 1000 realizations. Notice that blocking the sensors located at mediate position across hyperedges has significantly higher numbers of iterations than blocking sensors lying only in one hyperedge.

In real-world scenarios, communication constraints seriously influence sensor networks. Based on the radar network considered in Section 4.4, we further introduce bandwidth limitation to the example. We assume that the unit of bandwidth is bits per hypothesis per unit time (we call it bits for short in the following), and the belief on each hypothesis  $\theta_k$  ( $k = 1, 2, \dots, 9$ ) is of size *I* bits. Therefore, by adjusting the value of *I*, we can investigate the effect of bandwidth limitation on HSL algorithms. Moreover, to simulate a highly disturbed environment, we choose the signal structures for Sensor 5 on the positive

z-axis to be 
$$\begin{bmatrix} \ell_5(\text{NEAR}|\theta_k)\\ \ell_5(\text{MID}|\theta_k)\\ \ell_5(\text{FAR}|\theta_k) \end{bmatrix} = \begin{bmatrix} 0.36\\ 0.32\\ 0.32 \end{bmatrix} \text{ for } k = 1, 2, 3, 4 \text{ and } \begin{bmatrix} \ell_5(\text{NEAR}|\theta_k)\\ \ell_5(\text{MID}|\theta_k)\\ \ell_5(\text{FAR}|\theta_k) \end{bmatrix} = \begin{bmatrix} 0.33\\ 0.35\\ 0.32 \end{bmatrix} \text{ for } k = 5, 6, 7, 8. \text{ We additionally set } \begin{bmatrix} \ell_i(\text{NEAR}|\theta_9)\\ \ell_i(\text{MID}|\theta_9) \end{bmatrix} = \begin{bmatrix} 0.33\\ 0.33\\ 0.33 \end{bmatrix}, \forall i = 1, 2, \cdots, 6. \text{ For other } k = 1, 2, 3, 4 \text{ and } k = 1, 3, 4 \text{ and } k = 1, 4, 5 \text$$

 $\lfloor \ell_i(FAR|\theta_9) \rfloor [0.34]$ sensors, the signal structures are recomposed with the same logic. Results in Figure 7a,b illustrate that wrong consensus ( $\theta_5$  in this example) is finally reached when limiting the bandwidth to 6 bits. When raising the bandwidth to 8 bits, as shown in Figure 7c,d, our algorithms can converge to the true collective consensus ( $\theta_1$  in this example). These results show that HSL algorithms have strong robustness; however, they inevitably make mistakes when the sensor networks suffer from severe interference and limited communication.



**Figure 7.** (**a**,**b**) The evolution of beliefs on selected states under bandwidth limit of 6 bits, with (**a**,**b**) denoting the results of HSL-NEB and HSL-BNE, respectively. (**c**,**d**) The evolution of beliefs on selected states under bandwidth limit of 8 bits, with (**c**,**d**) denoting the results of HSL-NEB and HSL-BNE, respectively.

There are practically numerous scenarios in which agents demonstrate intentional misbehavior or function in a faulty manner. Extensive studies have been conducted on the robustness of graph-based social learning against Byzantine attacks, wherein adversaries are capable of deviating the system from the prescribed protocol in an arbitrary fashion [57,58]. Nevertheless, the Byzantine attacks on higher-order-based social learning remain unexplored. Our subsequent study will encompass the examination of this particular case.

To conclude, we proposed a new non-Bayesian social learning strategy named the hypergraph social learning by introducing the higher-order topology as the underlying network structure, with its convergence as well as the convergence rate theoretically analyzed. Several numerical examples were provided to demonstrate the effectiveness of the HSL framework and reveal its superior performance when applying it to sensor networks in tasks such as cooperative positioning. Insights regarding the relationship between convergence rate and different positions in the network were given. This work can assist sensor network designers to develop more efficient communication topology, which can better resist environmental obstructions. The HSL framework also has theoretical and applied values in broad areas such as distributed parameter estimation, dispersed information aggregation and social networks. Future works include extending the framework to more complex topologies, such as multi-layer structures, and considering more realistic difficulties in sensor networks such as energy limitations, connectivity loss and inadequate coverage.

**Author Contributions:** Conceptualization, Q.C. and D.S.; methodology, Q.C.; validation, Q.C. and W.S.; formal analysis, Q.C.; investigation, Q.C.; writing—original draft preparation, Q.C.; writing—review and editing, Q.C., W.S. and S.L.; supervision, D.S.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (No. 12101133) and Shanghai Sailing Program (No. 21YF1402300). This research was also supported by Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX0103).

Institutional Review Board Statement: Not applicable.

**Data Availability Statement:** Data sharing is not applicable to this article as no new data were created or analyzed in this study.

**Acknowledgments:** We thank Chun Guan and Wei Lin from Fudan University for their helpful comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

## Appendix A. Properties of the Matrix C

We prove here that the matrix *C* is the transition matrix of an aperiodic and irreducible Markov chain of finite states under the assumption that hypergraph  $\mathcal{H}$  is connected without any isolated vertex.

## Appendix A.1. Aperiodicity

From the definition in Assumption 1,  $c_{ii} = \sum_{j=1}^{m} b_{ij}a_{ji}$ ,  $\forall i = 1, \dots, n$ . It is clear that if  $i \in e_j$  (or  $e_j \in d_i$ ), we have  $b_{ij} \neq 0$  and  $a_{ji} \neq 0$ , and vice versa. Therefore, unless  $\exists i$  such that  $d_i = \emptyset$ , we have  $\sum_{j=1}^{m} b_{ij}a_{ji} = \sum_{e_j \in d_i} b_{ij}a_{ji} > 0$ . The assumption that no isolated vertex exists in  $\mathcal{H}$  guarantees that  $d_i \neq \emptyset$ ,  $\forall i = 1, \dots, n$ , which leads to  $c_{ii} > 0$ ,  $\forall i = 1, \dots, n$  and the Markov chain is aperiodic.

## Appendix A.2. Irreducibility

For  $c_{ij} = \sum_{k=1}^{m} b_{ik}a_{kj}$ , where  $i \neq j$ , we have  $c_{ij} = \sum_{e_k \in d_i \cap d_j} b_{ik}a_{kj} > 0$  unless  $\exists i \neq j$  such that  $d_i \cap d_j = \emptyset$ . We define graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}')$ , where  $\mathcal{V} = \{1, 2, \dots, n\}$  and  $\mathcal{E}' = \{(i, j) | d_i \cap d_j \neq \emptyset\}$  and  $M = (m_{ij})_{n \times n}$  is the transition matrix corresponding to  $\mathcal{G}$ . Clearly,  $m_{ij} > 0$  is equivalent to  $d_i \cap d_j \neq \emptyset$ . Therefore, the connectivity of  $\mathcal{H}$  induces the connectivity of  $\mathcal{G}$ , which further ensures that C is irreducible.

#### References

- Olfati-Saber, R.; Franco, E.; Frazzoli, E.; Shamma, J.S. Belief consensus and distributed hypothesis testing in sensor networks. In Proceedings of the Networked Embedded Sensing and Control: Workshop NESC'05, University of Notre Dame, Notre Dame, IN, USA, 17–18 October 2005; Springer: Berlin/Heidelberg, Germany, 2006; pp. 169–182.
- Antonio, P.; Grimaccia, F.; Mussetta, M. Architecture and methods for innovative heterogeneous wireless sensor network applications. *Remote Sens.* 2012, 4, 1146–1161. [CrossRef]

- 3. Tenney, R.R.; Sandell, N.R. Detection with distributed sensors. *IEEE Trans. Aerosp. Electron. Syst.* **1981**, *AES-17*, 501–510. [CrossRef]
- Chen, B.; Jiang, R.; Kasetkasem, T.; Varshney, P.K. Channel aware decision fusion in wireless sensor networks. *IEEE Trans. Signal Process.* 2004, 52, 3454–3458. [CrossRef]
- Alanyali, M.; Venkatesh, S.; Savas, O.; Aeron, S. Distributed Bayesian hypothesis testing in sensor networks. In Proceedings of the 2004 American Control Conference, Boston, MA, USA, 30 June–2 July 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 6, pp. 5369–5374.
- 6. Chakrabarty, K.; Iyengar, S.S.; Qi, H.; Cho, E. Grid coverage for surveillance and target location in distributed sensor networks. *IEEE Trans. Comput.* **2002**, *51*, 1448–1453. [CrossRef]
- 7. Shi, C.; Teng, W.; Zhang, Y.; Yu, Y.; Chen, L.; Chen, R.; Li, Q. Autonomous Multi-Floor Localization Based on Smartphone-Integrated Sensors and Pedestrian Indoor Network. *Remote Sens.* **2023**, *15*, 2933. [CrossRef]
- 8. Kar, S.; Moura, J.M. Distributed consensus algorithms in sensor networks: Quantized data and random link failures. *IEEE Trans. Signal Process.* **2009**, *58*, 1383–1400. [CrossRef]
- 9. Chong, C.Y.; Kumar, S.P. Sensor networks: Evolution, opportunities, and challenges. Proc. IEEE 2003, 91, 1247–1256. [CrossRef]
- 10. Leng, S.; Aihara, K. Common stochastic inputs induce neuronal transient synchronization with partial reset. *Neural Netw.* **2020**, *128*, 13–21. [CrossRef] [PubMed]
- Chamberland, J.F.; Veeravalli, V.V. Decentralized detection in sensor networks. *IEEE Trans. Signal Process.* 2003, 51, 407–416. [CrossRef]
- 12. Borkar, V.; Varaiya, P. Asymptotic agreement in distributed estimation. IEEE Trans. Autom. Control 1982, 27, 650–655. [CrossRef]
- 13. Appadwedula, S.; Veeravalli, V.V.; Jones, D.L. Energy-efficient detection in sensor networks. *IEEE J. Sel. Areas Commun.* 2005, 23, 693–702. [CrossRef]
- 14. Jadbabaie, A.; Molavi, P.; Sandroni, A.; Tahbaz-Salehi, A. Non-Bayesian social learning. *Games Econ. Behav.* 2012, *76*, 210–225. [CrossRef]
- 15. Banerjee, A.V. A simple model of herd behavior. Q. J. Econ. 1992, 107, 797–817. [CrossRef]
- 16. Bikhchandani, S.; Hirshleifer, D.; Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **1992**, *100*, 992–1026. [CrossRef]
- 17. Smith, L.; Sørensen, P. Pathological outcomes of observational learning. Econometrica 2000, 68, 371–398. [CrossRef]
- 18. Gale, D.; Kariv, S. Bayesian learning in social networks. *Games Econ. Behav.* 2003, 45, 329–346. [CrossRef]
- 19. Bordignon, V.; Vlaski, S.; Matta, V.; Sayed, A.H. Learning from heterogeneous data based on social interactions over graphs. *IEEE Trans. Inf. Theory* **2022**, *69*, 3347–3371. [CrossRef]
- Bordignon, V.; Matta, V.; Sayed, A.H. Partial information sharing over social learning networks. *IEEE Trans. Inf. Theory* 2022, 69, 2033–2058. [CrossRef]
- Hare, J.Z.; Uribe, C.A.; Kaplan, L.; Jadbabaie, A. Non-Bayesian social learning with uncertain models. *IEEE Trans. Signal Process.* 2020, 68, 4178–4193. [CrossRef]
- 22. Ntemos, K.; Bordignon, V.; Vlaski, S.; Sayed, A.H. Social Learning with Disparate Hypotheses. In Proceedings of the 2022 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 29 August–2 September 2022; pp. 2171–2175.
- 23. Molavi, P.; Tahbaz-Salehi, A.; Jadbabaie, A. *Foundations of Non-Bayesian Social Learning*; Columbia Business School Research Paper; Columbia Business School: New York, NY, USA, 2017.
- Molavi, P.; Rad, K.R.; Tahbaz-Salehi, A.; Jadbabaie, A. On consensus and exponentially fast social learning. In Proceedings of the 2012 American Control Conference (ACC), Montreal, QC, Canada, 27–29 June 2012; pp. 2165–2170.
- Rad, K.R.; Tahbaz-Salehi, A. Distributed parameter estimation in networks. In Proceedings of the 49th IEEE Conference on Decision and Control (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 5050–5055.
- 26. Kar, S.; Moura, J.M.; Ramanan, K. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *IEEE Trans. Inf. Theory* **2012**, *58*, 3575–3605. [CrossRef]
- 27. Nedić, A.; Olshevsky, A.; Uribe, C.A. Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs. In Proceedings of the 2015 American Control Conference (ACC), Chicago, IL, USA, 1–3 July 2015; pp. 5884–5889.
- Nedić, A.; Olshevsky, A.; Uribe, C.A. Fast convergence rates for distributed non-Bayesian learning. *IEEE Trans. Autom. Control* 2017, 62, 5538–5553. [CrossRef]
- 29. Lalitha, A.; Sarwate, A.; Javidi, T. Social learning and distributed hypothesis testing. *IEEE Int. Symp. Inf. Theory* 2014, 64, 6161–6179. [CrossRef]
- Salami, H.; Ying, B.; Sayed, A.H. Social learning over weakly connected graphs. *IEEE Trans. Signal Inf. Process. Netw.* 2017, 3, 222–238. [CrossRef]
- 31. Liu, Q.; Fang, A.; Wang, L.; Wang, X. Social learning with time-varying weights. J. Syst. Sci. Complex. 2014, 27, 581–593. [CrossRef]
- 32. Bordignon, V.; Matta, V.; Sayed, A.H. Adaptive social learning. IEEE Trans. Inf. Theory 2021, 67, 6053–6081. [CrossRef]
- 33. Sui, D.; Guan, C.; Gan, Z.; Lin, W.; Leng, S. Tuning Convergence Rate via Non-Bayesian Social Learning: A Trade-Off between Internal Belief and External Information. In Proceedings of the 2023 62nd IEEE Conference on Decision and Control (CDC), Marina Bay Sands, Singapore, 13–15 December 2023, accepted.
- 34. Olfati-Saber, R.; Murray, R.M. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Autom. Control* **2004**, *49*, 1520–1533. [CrossRef]

- 35. Kar, S.; Moura, J.M. Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise. *IEEE Trans. Signal Process.* **2008**, *57*, 355–369. [CrossRef]
- 36. Anderson, B.D.; Ye, M. Recent advances in the modelling and analysis of opinion dynamics on influence networks. *Int. J. Autom. Comput.* **2019**, *16*, 129–149. [CrossRef]
- Proskurnikov, A.V.; Tempo, R. A tutorial on modeling and analysis of dynamic social networks. Part I. Annu. Rev. Control 2017, 43, 65–79. [CrossRef]
- 38. Lewis, T.G. Network Science: Theory and Applications; John Wiley & Sons: Hoboken, NJ, USA, 2011.
- Ha, Y.; Guo, Y.; Lin, W. Non-Bayesian social learning model with periodically switching structures. *Chaos Interdiscip. J. Nonlinear Sci.* 2021, *31*, 043137. [CrossRef]
- 40. Bairey, E.; Kelsic, E.D.; Kishony, R. High-order species interactions shape ecosystem diversity. *Nat. Commun.* 2016, 7, 12285. [CrossRef]
- 41. Patania, A.; Petri, G.; Vaccarino, F. The shape of collaborations. EPJ Data Sci. 2017, 6, 18. [CrossRef]
- 42. Lambiotte, R.; Rosvall, M.; Scholtes, I. From networks to optimal higher-order models of complex systems. *Nat. Phys.* 2019, 15, 313–320. [CrossRef]
- Qiu, X.; Yang, L.; Guan, C.; Leng, S. Closed-loop control of higher-order complex networks: Finite-time and pinning strategies. Chaos Solitons Fractals 2023, 173, 113677. [CrossRef]
- 44. Estrada, E.; Ross, G.J. Centralities in simplicial complexes. Applications to protein interaction networks. *J. Theor. Biol.* **2018**, 438, 46–60. [CrossRef]
- 45. Yang, L.; Lin, W.; Leng, S. Conditional cross-map-based technique: From pairwise dynamical causality to causal network reconstruction. *Chaos Interdiscip. J. Nonlinear Sci.* **2023**, *33*, 063101. [CrossRef]
- 46. Zhang, H.; Song, L.; Han, Z.; Zhang, Y. Hypergraph Theory in Wireless Communication Networks; Springer: Cham, Switzerland, 2018.
- Yan, M.; Lam, K.Y.; Han, S.; Chan, E.; Chen, Q.; Fan, P.; Chen, D.; Nixon, M. Hypergraph-based data link layer scheduling for reliable packet delivery in wireless sensing and control networks with end-to-end delay constraints. *Inf. Sci.* 2014, 278, 34–55. [CrossRef]
- Liu, Q.; Zhao, J.; Wang, X. Distributed detection via Bayesian updates and consensus. In Proceedings of the 2015 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015; pp. 6992–6997.
- Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. Proc. AAAI Conf. Artif. Intell. 2019, 33, 3558–3565. [CrossRef]
- 50. Zou, M.; Gan, Z.; Cao, R.; Guan, C.; Leng, S. Similarity-navigated graph neural networks for node classification. *Inf. Sci.* 2023, 633, 41–69. [CrossRef]
- 51. Hoel, P.G.; Port, S.C.; Stone, C.J. Introduction to Stochastic Processes; Waveland Press: Long Grove, IL, USA, 1986.
- 52. Surana, A.; Chen, C.; Rajapakse, I. Hypergraph similarity measures. IEEE Trans. Netw. Sci. Eng. 2022, 10, 658–674. [CrossRef]
- 53. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* 2009, *81*, 591. [CrossRef]
- Wang, H.; Ma, C.; Chen, H.S.; Lai, Y.C.; Zhang, H.F. Full reconstruction of simplicial complexes from binary contagion and Ising data. *Nat. Commun.* 2022, 13, 3043. [CrossRef]
- 55. Cao, R.; Guan, C.; Gan, Z.; Leng, S. Reviving the Dynamics of Attacked Reservoir Computers. Entropy 2023, 25, 515. [CrossRef]
- Liang, X.; Zhong, Y.; Tang, J.; Liu, Z.; Yao, P.; Sun, K.; Zhang, Q.; Gao, B.; Heidari, H.; Qian, H.; et al. Rotating neurons for all-analog implementation of cyclic reservoir computing. *Nat. Commun.* 2022, 13, 1549. [CrossRef]
- Su, L.; Vaidya, N.H. Defending non-Bayesian learning against adversarial attacks. *Distrib. Comput.* 2019, *32*, 277–289. [CrossRef]
   Hare, J.Z.; Uribe, C.A.; Kaplan, L.M.; Jadbabaie, A. On malicious agents in non-Bayesian social learning with uncertain models. In Proceedings of the 2019 22nd International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019; pp. 1–8.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.