

Asynchronous Distributed Hypothesis Testing in the Presence of Crash Failures ^{*}

Lili Su and Nitin H. Vaidya

Department of Electrical and Computer Engineering, and
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Email: {lilisu3, nhv} @illinois.edu

Technical Report

February 4, 2016

Abstract

This paper addresses the problem of distributed hypothesis testing in multi-agent networks, where agents repeatedly collect local observations about an *unknown* state of the world, and try to collaboratively detect the true state through information exchange. We focus on the impact of failures and asynchrony – two fundamental factors in distributed systems – on the performance of consensus-based non-Bayesian learning. In particular, we consider the scenario where the networked agents may suffer crash faults, and messages delay can be *arbitrarily* long but finite. We identify the minimal global detectability of the network for non-Bayesian rule to succeed. In addition, we obtain a generalization of a celebrated result by Wolfowitz and Hajnal to submatrices, which might be of independent interest.

1 Introduction

Decentralized hypothesis testing is an important component of many decision-making and learning algorithms for large-scale systems, and thus has received significant amount of attention [1,2,3,5,6,7,8]. The traditional decentralized detection framework consists of a collection of spatially distributed sensors/agents and a fusion center [5,6,7]. The sensors/agents independently collect *noisy* observations of the environment state, and send only *summary* of the private observations to the fusion center, where a final decision is made. In the case when the sensors/agents directly send all the private observations, the detection problem can be solved using a centralized scheme. However, the above framework does not scale well, since each sensor needs to be connected to the fusion center and full reliability of the fusion center is assumed, which may not be practical as the system scales. Distributed hypothesis testing in the absence of fusion center is introduced by Gale and Kariv [2] in the context of social learning, where fully Bayesian belief update rule is studied. Bayesian update rule is impractical in many applications due to memory and computation constraints of each agent, and the inter-agent coordination challenges.

Non-Bayesian learning rule is first proposed by Jadbabaie et al. [3] in the general setting where external signals are observed during each iteration of the algorithm execution. Specifically, the belief of each agent is repeatedly updated as the arithmetic mean of its local Bayesian update and

^{*} This research is supported in part by National Science Foundation award NSF 1329681. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funding agencies or the U.S. government.

the beliefs of its neighbors, combining iterative consensus algorithm with local Bayesian update. It is shown [3] that, under this learning rule, each agent learns the true state almost surely. Since the publication of [3], significant efforts have been devoted to designing and analyzing non-Bayesian learning rules with a particular focus on refining the fusion strategies and analyzing the (asymptotic and/or finite time) convergence rates of the refined algorithms [11,15,16,17,12,19,18,14].

Among the various proposed fusion rules, in this paper we are particularly interested in the log-linear form of the update rule, in which, essentially, each agent updates its belief as the geometric average of the local Bayesian update and its neighbors' beliefs [16]. The log-linear form (geometric averaging) update rule is proposed in [16], and is shown to converge exponentially fast [11,17]. Taking an axiomatic approach, the geometric averaging fusion is shown to be optimal [14]. An optimization-based interpretation of this rule is presented in [17], using dual averaging method with properly chosen proximal functions. Finite-time convergence rates are investigated independently in [15,12,18]. Both [15] and [19] consider time-varying networks, with slightly different network models. Specifically, [15] assumes that the union of every consecutive B networks is strongly connected, while [19] considers random networks. In this paper, for ease of exposition, we assume that the network topology is static. As can be seen later, our results can be easily generalized to time-varying networks.

All the above work implicitly assumes synchronous systems and reliable agents. However, in the context of decentralized hypothesis testing, asynchrony (message asynchrony and/or computation asynchrony) and failures (link failures and/or agent failures) – two fundamental factors in practical distributed systems – have not been addressed yet. In this paper, we consider the scenario where some agents may suffer crash faults (i.e., cease operating), and messages delay can be *arbitrarily* long but finite [13]. This model is commonly considered in distributed computing, and the main challenge is that when an agent i does not receive messages from the incoming neighbor agent j within some timeout interval, from agent i 's perspective, it is not possible to distinguish whether agent j has crashed, or the messages from agent j are delayed.

We identify the minimal global detectability of the network for non-Bayesian rule to succeed. In addition, we obtain a generalization of a celebrated result by Wolfowitz and Hajnal to submatrices, which might be of independent interest.

The rest of the paper is organized as follows. Section 2 presents the problem formulation and the learning rule. Section 3 introduces the notion of pseudo-belief vector, whose evolution captures the dynamics of the agents' beliefs, and admits a simple matrix representation. The limiting behavior of the product of the update matrices (the matrices in the matrix representation introduced in Section 3) is investigated in Section 4. Also in Section 4, we generalize a celebrated result by Wolfowitz and Hajnal. The necessary and sufficient network detectability is characterized in Section 5.

2 Problem Formulation

Network Model: We consider an asynchronous system, where the message delay can be *arbitrarily* long but finite. A collection of n agents are connected by a *directed* network $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, n\}$ and \mathcal{E} is the collection of *directed* edges. For each $i \in \mathcal{V}$, let \mathcal{I}_i denote the set of incoming neighbors of agent i . In any execution, up to f agents suffer crash faults (i.e., cease operating). Let $\mathcal{N} \subseteq \mathcal{V}$ be the set of agents that operate correctly during a given execution (i.e., they are non-faulty). Note that $|\mathcal{V} - \mathcal{N}| \leq f$, since at most f agents may suffer crash failure. As noted earlier, although we assume a static network topology, our results can be easily generalized to time-varying networks. Throughout this paper, we use the terms agent and node interchangeably.

Observation Model: Let $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ denote a finite set of m environmental states, which we call *hypotheses*. We consider asynchronous iterations. In the t -th iteration, each agent *independently* obtains private signal about the environmental state θ^* , which is initially unknown to every agent in the network. For ease of exposition, we assume that if multiple signals are observed, only one signal is used to update beliefs. Each agent i knows the structure of its private signal, which is represented by a collection of parameterized distributions $\mathcal{D}^i = \{\ell_i(w_i|\theta)|\theta \in \Theta, w_i \in \mathcal{S}_i\}$, where $\ell_i(\cdot|\theta)$ is a distribution with parameter $\theta \in \Theta$, and \mathcal{S}_i is the finite private signal space. For each $\theta \in \Theta$, and each $i \in \mathcal{V}$, the support of $\ell_i(\cdot|\theta)$ is the whole signal space, i.e., $\ell_i(w_i|\theta) > 0$, $\forall w_i \in \mathcal{S}_i$ and $\forall \theta \in \Theta$. Precisely, let s_t^i be the private signal observed by agent i in iteration t , and let $\mathbf{s}_t = \{s_t^1, s_t^2, \dots, s_t^n\}$ be the signal profile at time t (i.e., signals observed by the agents in iteration t). Given an environmental state θ , the signal profile \mathbf{s}_t is generated according to the joint distribution $\ell_1(\cdot|\theta) \times \ell_2(\cdot|\theta) \times \dots \times \ell_n(\cdot|\theta)$. The goal is to have all the non-faulty agents (agents in \mathcal{N}) collaboratively learn the environmental state θ^* .

Belief Update Rule: Each agent i keeps a belief μ^i , which is a distribution over the set Θ , with $\mu^i(\theta)$ being the probability with which the agent i *believes* that θ is the true environmental state. Since no signals are observed before the execution of an algorithm, the belief μ^i is often initially set to be uniform over the set Θ , i.e., $(\mu_0^i(\theta_1), \mu_0^i(\theta_2), \dots, \mu_0^i(\theta_m))^T = (\frac{1}{m}, \dots, \frac{1}{m})^T$.¹ In this work, we also adopt the above convention. (For our results to hold, it suffices to have $\mu_0^i(\theta) > 0$ for each $\theta \in \Theta$ and each $i \in \mathcal{V}$.)

In our algorithm, we will use a geometric averaging update rule that has been investigated in previous work [15,16,12,18]. Let $\mathcal{N}[t]$ be the set of agents that have not crashed *by the beginning* of iteration t , and let $\bar{\mathcal{N}}[t]$ be the set of agents that have not crashed *by the end* of iteration t . Note that $\mathcal{N}[t+1] \subseteq \mathcal{N}[t]$, $\mathcal{N}[t] - \bar{\mathcal{N}}[t]$ is the collection of agents that crash *during* iteration t , and that $\bar{\mathcal{N}}[t] = \mathcal{N}[t+1]$. In addition, $\lim_{t \rightarrow \infty} \mathcal{N}[t] = \mathcal{N} = \lim_{t \rightarrow \infty} \bar{\mathcal{N}}[t]$.

For $t \geq 1$, the steps to be performed by agent $i \in \mathcal{N}[t]$ in the t -th iteration are listed as follows, where messages are tagged with (asynchronous) iteration index.

1. *Transmit Step:* Transmit current belief vector μ_{t-1}^i on all outgoing edges.
2. *Receive Step:* Wait until a private signal s_t^i is observed and belief vectors are received from $|\mathcal{I}_i| - f$ incoming neighbors. Let $\mathcal{R}_i[t]$ be the set of incoming neighbors from whom agent i receives these belief vectors.
3. *Update Step:*² For each $\theta \in \Theta$, update $\mu^i(\theta)$ as

$$\mu_t^i(\theta) \triangleq \frac{\ell_i(s_t^i|\theta) \prod_{j \in \mathcal{R}_i[t] \cup \{i\}} \mu_{t-1}^j(\theta)^{\frac{1}{|\mathcal{I}_i| - f + 1}}}{\sum_{p=1}^m \ell_i(s_t^i|\theta_p) \prod_{j \in \mathcal{R}_i[t] \cup \{i\}} \mu_{t-1}^j(\theta_p)^{\frac{1}{|\mathcal{I}_i| - f + 1}}}. \quad (1)$$

Note that due to asynchrony and agent failures, $\mathcal{R}_i[t]$ may change over time and *is not* monotone. In contrast, in synchronous systems, $\mathcal{R}_i[t]$ is non-increasing, i.e., $\mathcal{R}_i[t+1] \subseteq \mathcal{R}_i[t]$ for any $t \geq 1$ and any $i \in \mathcal{V}$.

In iteration t , if an agent crashes after performing the update step in (1) for all $\theta \in \Theta$, without loss of generality, we say that this agent crashes in iteration $t+1$. Note that each agent in $\mathcal{N}[t] - \bar{\mathcal{N}}[t]$ may crash at any time during iteration t . In particular, it may crash before performing (1) or while performing (1). The above protocol is different from the original geometric averaging learning rule [15,16,12,18] in the receive step, where each agent waits to receive messages from $|\mathcal{I}_i| - f$ incoming neighbors instead of waiting to hear from all of its incoming neighbors. This modification

¹ In this paper, every vector considered is column vector.

² In the notation $\mu_t^i(\theta)$, the superscript denotes the agents and subscript denotes iterations.

is necessary because of asynchrony and the possibility of up to f agents crashes. Requiring an agent to receive messages from more than $|\mathcal{I}_i| - f$ neighbors may lead to non-termination of the learning protocol (when f incoming neighbors have crashed).

Recall that θ^* is the true state. We say the networked agents collaboratively detect θ^* if for each non-faulty agent $i \in \mathcal{N}$,

$$\lim_{t \rightarrow \infty} \mu_t^i(\theta^*) = 1 \text{ a.s.} \quad (2)$$

3 Matrix Representation

In this section, we define a matrix representation of the agents' belief update. In synchronous and reliable networks, the update matrix $\mathbf{A}[t]$ is often chosen to be the weighted adjacency matrix of the network [15,12,18]. However, the above matrix representation is improper in the presence of message asynchrony and agent failure, observing that transmitted messages may not be used due to message delay, and agents *not* in $\bar{\mathcal{N}}[t]$ may not perform (1) for all $\theta \in \Theta$ (and may not even observe new private signals). To resolve the above complication, we introduce pseudo-belief vector $\tilde{\mu}^i$ for each $i \in \mathcal{V}$. The evolution of the pseudo-belief vectors admits a simple matrix representation, and captures the update of μ^i for each non-faulty agent $i \in \mathcal{N}$. For each $i \in \bar{\mathcal{N}}[t]$, define

$$\mathbf{A}_{ij}[t] \triangleq \begin{cases} \frac{1}{|\mathcal{I}_i| - f + 1}, & j \in \mathcal{R}_i[t] \cup \{i\} \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

and for each $i \notin \bar{\mathcal{N}}[t]$,

$$\mathbf{A}_{ij}[t] \triangleq \begin{cases} 1, & j = i \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Note that in iteration t , an agent can only receive messages tagged with (asynchronous) iteration index t from agents that have not crashed by the beginning of iteration t . Thus, $\mathbf{A}_{ij}[t] = 0$ for each $j \notin \bar{\mathcal{N}}[t]$, and

$$1 = \sum_{j=1}^n \mathbf{A}_{ij}[t] = \sum_{j \in \bar{\mathcal{N}}[t]} \mathbf{A}_{ij}[t] + \sum_{j \notin \bar{\mathcal{N}}[t]} \mathbf{A}_{ij}[t] = \sum_{j \in \bar{\mathcal{N}}[t]} \mathbf{A}_{ij}[t].$$

The pseudo-belief is defined as follows. For each $i \in \bar{\mathcal{N}}[t]$,

$$\tilde{\mu}_t^i(\theta) \triangleq \frac{\ell_i(s_t^i|\theta) \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]}}{\sum_{p=1}^m \ell_i(s_t^i|\theta_p) \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta_p)^{\mathbf{A}_{ij}[t]}}, \quad \forall \theta \in \Theta, \quad (5)$$

with $\tilde{\mu}_0^i(\theta) = \mu_0^i(\theta) = \frac{1}{m}$, and for each $i \notin \bar{\mathcal{N}}[t]$,

$$\tilde{\mu}_t^i(\theta) \triangleq \frac{\prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]}}{\sum_{p=1}^m \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta_p)^{\mathbf{A}_{ij}[t]}} = \tilde{\mu}_{t-1}^i(\theta), \quad \forall \theta \in \Theta, \quad (6)$$

with $\tilde{\mu}_0^i(\theta) = \mu_0^i(\theta) = \frac{1}{m}$. Note that, in contrast to (5), no private signal is involved in (6). In addition, (6) is equivalent to defining $\tilde{\mu}_t^i(\theta) = \tilde{\mu}_{t-1}^i(\theta)$. It is easy to see (by induction) that

$$\tilde{\mu}_t^i(\theta) = \mu_t^i(\theta), \quad \forall i \in \bar{\mathcal{N}}[t].$$

Note that $\tilde{\mu}^i(\theta)$ only captures the evolution of $\mu^i(\theta)$ for $i \in \bar{\mathcal{N}}[t]$, i.e., each agent that has not crashed *at the end* of iteration t . Since an agent $i \in \mathcal{N}[t] - \bar{\mathcal{N}}[t]$ may crash during the update step (1), $\tilde{\mu}^i(\theta)$ may not capture the real update performed by nodes in $\mathcal{N}[t] - \bar{\mathcal{N}}[t]$ for some subset of Θ . This inconsistency does not affect the accuracy of our analysis. Intuitively, since the nodes in $\mathcal{N}[t] - \bar{\mathcal{N}}[t]$ are crashing away, they will not affect further system evolution.

For any $\theta \in \Theta$, and any $i \in \mathcal{V}$, let

$$\psi_t^i(\theta) \triangleq \log \frac{\tilde{\mu}_t^i(\theta)}{\tilde{\mu}_t^i(\theta^*)}. \quad (7)$$

Note that for each $i \in \mathcal{N}$, $\mu_t^i(\theta^*) \xrightarrow{\text{a.s.}} 1$ if and only if $\psi_t^i(\theta) \xrightarrow{\text{a.s.}} -\infty$ for $\theta \neq \theta^*$. In addition, let $\mathcal{L}_t(\theta) \in \mathbb{R}^n$ such that

$$\mathcal{L}_t^i(\theta) \triangleq \begin{cases} \log \frac{\ell_i(s_t^i|\theta)}{\ell_i(s_t^i|\theta^*)}, & \forall i \in \bar{\mathcal{N}}[t], \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Then, for each $i \in \bar{\mathcal{N}}[t]$, we have

$$\begin{aligned} \psi_t^i(\theta) &= \log \frac{\tilde{\mu}_t^i(\theta)}{\tilde{\mu}_t^i(\theta^*)} \\ &= \log \left(\frac{\left(\prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]} \right) \ell_i(s_t^i|\theta)}{\sum_{p=1}^m \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]} \ell_i(s_t^i|\theta)} \times \frac{\sum_{p=1}^m \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]} \ell_i(s_t^i|\theta)}{\left(\prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta^*)^{\mathbf{A}_{ij}[t]} \right) \ell_i(s_t^i|\theta^*)} \right) \\ &= \log \left(\prod_{j=1}^n \left(\frac{\tilde{\mu}_{t-1}^j(\theta)}{\tilde{\mu}_{t-1}^j(\theta^*)} \right)^{\mathbf{A}_{ij}[t]} \times \frac{\ell_i(s_t^i|\theta)}{\ell_i(s_t^i|\theta^*)} \right) \\ &= \sum_{j=1}^n \mathbf{A}_{ij}[t] \log \frac{\tilde{\mu}_{t-1}^j(\theta)}{\tilde{\mu}_{t-1}^j(\theta^*)} + \log \frac{\ell_i(s_t^i|\theta)}{\ell_i(s_t^i|\theta^*)} \\ &= \sum_{j=1}^n \mathbf{A}_{ij}[t] \psi_{t-1}^j(\theta) + \mathcal{L}_t^i(\theta) \quad \text{by (8)} \end{aligned} \quad (9)$$

and for each $i \notin \bar{\mathcal{N}}[t]$, we have

$$\begin{aligned} \psi_t^i(\theta) &= \log \frac{\tilde{\mu}_t^i(\theta)}{\tilde{\mu}_t^i(\theta^*)} \\ &= \log \left(\frac{\left(\prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]} \right)}{\sum_{p=1}^m \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]}} \times \frac{\sum_{p=1}^m \prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta)^{\mathbf{A}_{ij}[t]}}{\left(\prod_{j=1}^n \tilde{\mu}_{t-1}^j(\theta^*)^{\mathbf{A}_{ij}[t]} \right)} \right) \quad \text{by (6)} \\ &= \log \left(\prod_{j=1}^n \left(\frac{\tilde{\mu}_{t-1}^j(\theta)}{\tilde{\mu}_{t-1}^j(\theta^*)} \right)^{\mathbf{A}_{ij}[t]} \right) \\ &= \sum_{j=1}^n \mathbf{A}_{ij}[t] \log \frac{\tilde{\mu}_{t-1}^j(\theta)}{\tilde{\mu}_{t-1}^j(\theta^*)} \\ &= \sum_{j=1}^n \mathbf{A}_{ij}[t] \psi_{t-1}^j(\theta) + \mathcal{L}_t^i(\theta) \quad \text{by (8)} \end{aligned} \quad (10)$$

Let $\psi_t(\theta) \in \mathbb{R}^n$ be the vector that stacks $\psi_t^i(\theta)$, with the i -th entry being $\psi_t^i(\theta)$ for all $i \in \mathcal{V}$. The evolution of $\psi(\theta)$ can be compactly written as

$$\psi_t(\theta) = \mathbf{A}[t]\psi_{t-1}(\theta) + \mathcal{L}_t(\theta). \quad (11)$$

Expanding (11), we get

$$\begin{aligned} \psi_t(\theta) &= \mathbf{A}[t]\psi_{t-1}(\theta) + \mathcal{L}_t(\theta) \\ &= \mathbf{A}[t](\mathbf{A}[t-1]\psi_{t-2}(\theta) + \mathcal{L}_{t-1}(\theta)) + \mathcal{L}_t(\theta) \\ &\dots \\ &= \mathbf{A}[t]\mathbf{A}[t-1]\cdots\mathbf{A}[1]\psi_0(\theta) + \sum_{r=1}^{t-1} \mathbf{A}[t]\mathbf{A}[t-1]\cdots\mathbf{A}[r+1]\mathcal{L}_r(\theta) + \mathcal{L}_t(\theta) \\ &= \Phi(t, 1)\psi_0(\theta) + \sum_{r=1}^t \Phi(t, r+1)\mathcal{L}_r(\theta), \end{aligned} \quad (12)$$

where $\Phi(t, r) = \mathbf{A}[t]\cdots\mathbf{A}[r]$ for $r \leq t+1$. By convention, $\Phi(t, t) = \mathbf{A}[t]$ and $\Phi(t, t+1) = \mathbf{I}$.

4 Convergence of $\Phi(t, r)$

In this section, we present the tight (necessary and sufficient) condition on $G(\mathcal{V}, \mathcal{E})$ [20] such that for $i, j \in \mathcal{N}$, the following holds.

$$\lim_{t \rightarrow \infty} |\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| = 0. \quad (13)$$

Note that (13) is weaker than requiring $\Phi(t, r)$ to be weakly ergodic, where (13) needs to hold for any $i, j \in \mathcal{V}$. If the infinite backward product $\lim_{t \rightarrow \infty} \Phi(t, r)$ satisfies (13), we say the product is weakly ergodic restricted to indices in \mathcal{N} .

Definition 1. A reduced graph \mathcal{H} of $G(\mathcal{V}, \mathcal{E})$ is obtained by (i) removing up to f incoming links for each $i \in \mathcal{V}$; (ii) in the obtained graph, removing up to f sinks, if any.³

Let \mathcal{C} be the collection of all reduced graph of $G(\mathcal{V}, \mathcal{E})$. Since $|\mathcal{V}| = n$ is finite, it holds that $\chi = |\mathcal{C}| < \infty$.

Definition 2. Given a graph $G(\mathcal{V}, \mathcal{E})$, a source component in $G(\mathcal{V}, \mathcal{E})$ is the strongly connected component that does not have an incoming link from outside the component.

Condition 1: Every reduced graph of $G(\mathcal{V}, \mathcal{E})$ contains a unique source component.

Let $\mathcal{H} \in \mathcal{C}$ be an arbitrary reduced of $G(\mathcal{V}, \mathcal{E})$ with source component $\mathcal{S}_{\mathcal{H}}$. Define

$$\gamma \triangleq \min_{\mathcal{H} \in \mathcal{C}} |\mathcal{S}_{\mathcal{H}}|, \quad (14)$$

i.e., γ is the minimum source component size in all reduced graphs. Note that $\gamma \geq 1$ if Condition 1 holds.

We first show that Condition 1 is necessary for (13) to hold. Note that if (13) holds, asynchronous crash consensus is achieved [13]. Thus the necessary condition for asynchronous consensus is also

³ For a given graph $H(\mathcal{V}, \mathcal{E})$, a node $s \in \mathcal{V}$ is called a sink if there is no outgoing links coming from node s .

necessary for (13). The following tight (necessary and sufficient) condition of asynchronous crash consensus is found in [20].

Condition 2: For any node partition L, R, C such that $L \neq \emptyset$ and $R \neq \emptyset$, at least one of the following holds:

- (1) there exists $i \in L$ such that $|\mathcal{I}_i \cap (R \cup C)| \geq f + 1$; or
- (2) there exists $j \in R$ such that $|\mathcal{I}_j \cap (L \cup C)| \geq f + 1$.

Theorem 1. *Condition 1 and Condition 2 are equivalent.*

Theorem 1 is proved in Appendix A.

Remark 1. Let \mathcal{H}' be the subgraph of $G(\mathcal{V}, \mathcal{E})$ obtained by removing up to f incoming links for each $i \in \mathcal{V}$. Indeed, we are able to show that Condition 2 holds if and only if \mathcal{H}' contains a single source component. That is the removal of sink nodes does not matter for the convergence of consensus because it does not affect the uniqueness of source component and the only state evolution of the non-faulty nodes matter in the consensus problem.

The sufficiency of Condition 1 is implied by several auxiliary propositions and lemmas. We first observe that the belief evolution is governed by reduced graphs.

Proposition 1. *Let $\xi = \frac{1}{1 + \max_{i \in \mathcal{V}} |\mathcal{I}_i|}$. For $t \geq 1$, there exists a reduced graph $\mathcal{H}[t]$ with adjacency matrix $\mathbf{H}[t]$ such that*

$$\mathbf{A}[t] \geq \xi \mathbf{H}[t].$$

The proof of Proposition 1 can be found in Appendix B.

Let \mathcal{H} be a reduced graph of $G(\mathcal{V}, \mathcal{E})$ with \mathbf{H} as the adjacency matrix. The convergence of the backward product $\Phi(t, r)$ can be analyzed using ergodic coefficients. We generalize some well-known results obtained by Wolfowitz [21] and Hajnal [10] to submatrices.

For $t \geq 1$, $t' \geq t$ and $r \geq 1$, define

$$\delta_r(\Phi(t', t)) \triangleq \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[r]} |\Phi_{ij}(t', t) - \Phi_{i'j}(t', t)|. \quad (15)$$

$$\eta_r(\Phi(t', t)) \triangleq \min_{i, i' \in \mathcal{N}[r]} \sum_{j \in \mathcal{V}} \min\{\Phi_{ij}(t', t), \Phi_{i'j}(t', t)\}. \quad (16)$$

Since $\mathcal{N}[t] \subseteq \mathcal{N}[r]$ for any $r \leq t$, by (15) and (16), we get it is easy to see that

$$\delta_t(\Phi(t', t)) \leq \delta_r(\Phi(t', t)), \text{ and } \eta_t(\Phi(t', t)) \geq \eta_r(\Phi(t', t)). \quad (17)$$

Note that the definition of $\delta_r(\Phi)(t', t)$ is symmetric in i and i' . Thus,

$$\begin{aligned} \delta_r(\Phi(t', t)) &= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[r]} |\Phi_{ij}(t', t) - \Phi_{i'j}(t', t)| \\ &= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[r]} (\Phi_{ij}(t', t) - \Phi_{i'j}(t', t)). \end{aligned} \quad (18)$$

Proposition 2. *Given $t \geq 1$, for $t' \geq t$, $i \in \mathcal{N}[t]$ and $j \notin \mathcal{N}[t]$, it holds that*

$$\Phi_{ij}(t', t) = 0.$$

Proof. We prove this proposition by inducting on t' .

(Base case) When $t' = t$, $\Phi(t', t) = \mathbf{A}[t]$. Since $i \in \mathcal{N}[t]$ and $j \notin \mathcal{N}[t]$, it holds that $i \neq j$, and $\mathbf{A}_{ij}[t] = 0$, proving Proposition 2 for $\Phi(t, t)$.

(Induction hypothesis) Suppose Proposition 2 holds for $t' \geq t$.

(Inductive step) Now we prove Proposition 2 for $\Phi(t' + 1, t)$. Recall that $\Phi(t' + 1, t) = \mathbf{A}[t' + 1]\Phi(t', t)$. For $i \in \mathcal{N}[t]$ and $j \notin \mathcal{N}[t]$, we get

$$\begin{aligned}
\Phi_{ij}(t' + 1, t) &= \sum_{k=1}^n \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) \\
&= \sum_{k \in \mathcal{N}[t]} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) + \sum_{k \notin \mathcal{N}[t]} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) \\
&= \sum_{k \in \mathcal{N}[t]} \mathbf{A}_{ik}[t' + 1]0 + \sum_{k \notin \mathcal{N}[t]} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) \quad \text{by induction hypothesis} \\
&\leq \sum_{k \notin \mathcal{N}[t'+1]} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) \quad \text{since } \mathcal{N}[t' + 1] \subseteq \mathcal{N}[t].
\end{aligned} \tag{19}$$

For $k \neq i$, by (4), we know $\mathbf{A}_{ik}[t' + 1] = 0$. For $k = i$, by induction hypothesis, we know that $\Phi_{ij}(t', t) = 0$. Thus, we have

$$\begin{aligned}
\sum_{k \notin \mathcal{N}[t'+1]} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) &= \mathbf{A}_{ii}[t' + 1]\Phi_{ij}(t', t) + \sum_{k \notin \mathcal{N}[t'+1], k \neq i} \mathbf{A}_{ik}[t' + 1]\Phi_{kj}(t', t) \\
&= 1 \times 0 + \sum_{k \notin \mathcal{N}[t'+1], k \neq i} 0 \times \Phi_{kj}(t', t) = 0,
\end{aligned}$$

i.e., the RHS of (19) equals 0. This completes the proof of the induction, and Proposition 2 is proved. \square

Recall that $\Phi(t', t)$ is row stochastic. As an immediate consequence of Proposition 2, for any t , $t' \geq t$, and $i \in \mathcal{N}[t]$, the following holds.

$$\sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t', t) = 1. \tag{20}$$

The following two lemmas generalize the results from Wolfowitz [21] and Hajnal [10].

Lemma 1. *For each $t \geq 0$ and $t' \geq t$, we have*

$$\delta_t(\Phi(t', t)) \leq 1 - \eta_t(\Phi(t', t)).$$

Proof.

$$\begin{aligned}
1 - \eta_t(\Phi(t', t)) &= 1 - \min_{i, i' \in \mathcal{N}[t]} \sum_{j \in \mathcal{V}} \min\{\Phi_{ij}(t', t), \Phi_{i'j}(t', t)\} \quad \text{by (16)} \\
&= \max_{i, i' \in \mathcal{N}[t]} \left(1 - \sum_{j \in \mathcal{V}} \min\{\Phi_{ij}(t', t), \Phi_{i'j}(t', t)\} \right) \\
&= \max_{i, i' \in \mathcal{N}[t]} \left(\sum_{j \in \mathcal{V}} \Phi_{ij}(t', t) - \sum_{j \in \mathcal{V}} \min\{\Phi_{ij}(t', t), \Phi_{i'j}(t', t)\} \right) \quad \text{since } \sum_{j \in \mathcal{V}} \Phi_{ij}(t', t) = 1 \\
&= \max_{i, i' \in \mathcal{N}[t]} \sum_{j: j \in \mathcal{V}, \Phi_{ij}(t', t) \geq \Phi_{i'j}(t', t)} (\Phi_{ij}(t', t) - \Phi_{i'j}(t', t)) \\
&\geq \max_{i, i' \in \mathcal{N}[t]} \max_{j \in \mathcal{V}} (\Phi_{ij}(t', t) - \Phi_{i'j}(t', t)) \\
&= \delta_t(\Phi(t', t)) \quad \text{by (18)}.
\end{aligned}$$

Lemma 2. For $t_2 > t_1 \geq t_0 \geq 1$, define $\mathbf{P} = \Phi(t_2, t_1 + 1)$, $\mathbf{G} = \Phi(t_1, t_0)$, and $\mathbf{F} = \Phi(t_2, t_0)$. Then it holds that

$$\delta_{t_1+1}(\mathbf{F}) \leq (1 - \eta_{t_1+1}(\mathbf{P})) \delta_{t_1+1}(\mathbf{G}).$$

Proof. Since $\mathbf{P} = \Phi(t_2, t_1 + 1)$, $\mathbf{G} = \Phi(t_1, t_0)$, and $\mathbf{F} = \Phi(t_2, t_0)$, it holds that

$$\mathbf{F} = \Phi(t_2, t_0) = \Phi(t_2, t_1 + 1)\Phi(t_1, t_0) = \mathbf{P}\mathbf{G}.$$

For any $i, j \in \mathcal{V}$, we have $\mathbf{F}_{ij} = \sum_{k=1}^n \mathbf{P}_{ik}\mathbf{G}_{kj}$. We get

$$\begin{aligned}
\delta_{t_1+1}(\mathbf{F}) &= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} (\mathbf{F}_{ij} - \mathbf{F}_{i'j}) \quad \text{by (18)} \\
&= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k=1}^n \mathbf{P}_{ik}\mathbf{G}_{kj} - \sum_{k=1}^n \mathbf{P}_{i'k}\mathbf{G}_{kj} \right) \\
&= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k=1}^n (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} \right) \\
&= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k \in \mathcal{N}[t_1+1]} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} + \sum_{k \notin \mathcal{N}[t_1+1]} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} \right). \quad (21)
\end{aligned}$$

Recall that $\mathcal{N}[t_1 + 1]$ is the collection of agents that have not crashed by the beginning of iteration $t_1 + 1$. If $k \notin \mathcal{N}[t_1 + 1]$, then k crashed in the first t_1 iterations. Intuitively speaking, for any $t > t_1$, since agent k has already crashed, it cannot influence any other agents – no other agents can receive any message from agent k after iteration t_1 . From Proposition 2, we know that $\mathbf{P}_{ik} = 0 = \mathbf{P}_{i'k}$ for all $k \notin \mathcal{N}[t_1 + 1]$, and $i, i' \in \mathcal{N}[t_1 + 1]$. Consequently, we have

$$\sum_{k \notin \mathcal{N}[t_1+1]} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} = \sum_{k \notin \mathcal{N}[t_1+1]} 0 \mathbf{G}_{kj} = 0, \quad \forall j \in \mathcal{V}. \quad (22)$$

Thus, equality (21) can be simplified as

$$\begin{aligned}
\delta_{t_1+1}(\mathbf{F}) &= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k \in \mathcal{N}[t_1+1]} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} \right) \text{ by (22)} \\
&= \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} \right. \\
&\quad \left. + \sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} < \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \mathbf{G}_{kj} \right) \\
&\leq \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \max_{k \in \mathcal{N}[t_1+1]} \mathbf{G}_{kj} \right. \\
&\quad \left. + \sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} < \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \min_{k \in \mathcal{N}[t_1+1]} \mathbf{G}_{kj} \right). \quad (23)
\end{aligned}$$

Recall that $\mathbf{P}_{ik} = 0 = \mathbf{P}_{i'k}$ for all $k \notin \mathcal{N}[t_1+1]$, and $i, i' \in \mathcal{N}[t_1+1]$. We have

$$\sum_{k \in \mathcal{N}[t_1+1]} \mathbf{P}_{ik} = 1 = \sum_{k \in \mathcal{N}[t_1+1]} \mathbf{P}_{i'k}.$$

Then,

$$\begin{aligned}
0 &= \sum_{k \in \mathcal{N}[t_1+1]} \mathbf{P}_{ik} - \sum_{k \in \mathcal{N}[t_1+1]} \mathbf{P}_{i'k} \\
&= \sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} < \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) + \sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}).
\end{aligned}$$

Thus, we have

$$\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} < \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) = - \sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}). \quad (24)$$

The RHS of (23) becomes

$$\begin{aligned}
\delta_{t_1+1}(\mathbf{F}) &\leq \max_{j \in \mathcal{V}} \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \right) \left(\max_{k \in \mathcal{N}[t_1+1]} \mathbf{G}_{kj} - \min_{k \in \mathcal{N}[t_1+1]} \mathbf{G}_{kj} \right) \\
&= \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \right) \max_{j \in \mathcal{V}} \max_{k, k' \in \mathcal{N}[t_1+1]} (\mathbf{G}_{kj} - \mathbf{G}_{k'j}) \\
&= \max_{i, i' \in \mathcal{N}[t_1+1]} \left(\sum_{k: k \in \mathcal{N}[t_1+1], \mathbf{P}_{ik} \geq \mathbf{P}_{i'k}} (\mathbf{P}_{ik} - \mathbf{P}_{i'k}) \right) \delta_{t_1+1}(\mathbf{G}) \text{ by (15)} \\
&= \max_{i, i' \in \mathcal{N}[t_1+1]} \left(1 - \sum_{k \in \mathcal{V}} \min\{\mathbf{P}_{ik}, \mathbf{P}_{i'k}\} \right) \delta_{t_1+1}(\mathbf{G}) \text{ since } \sum_{k \in \mathcal{N}[t_1+1]} \mathbf{P}_{ik} = 1 \\
&= (1 - \eta_{t_1+1}(\mathbf{P})) \delta_{t_1+1}(\mathbf{G}), \quad (25)
\end{aligned}$$

proving the lemma.

Recall that χ is the total number of reduced graphs of $G(\mathcal{V}, \mathcal{E})$. Using Proposition 1, Lemma 1 and Lemma 2, we can show that $|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)|$, where $i, j \in \mathcal{N}$, decays exponentially fast.

Theorem 2. *Suppose $r \leq t$. For any $i, j \in \bar{\mathcal{N}}[t]$ and for any $k \in \mathcal{V}$, it holds that*

$$|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| \leq \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\}.$$

Proof. For any $i, j \in \bar{\mathcal{N}}[t]$ and $k \in \mathcal{V}$, it trivially holds that

$$|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| \leq 1. \quad (26)$$

To prove this theorem, it is enough to show

$$|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| \leq (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}, \forall i, j \in \bar{\mathcal{N}}[t].$$

For $1 \leq k \leq \lfloor \frac{t-r+1}{n\chi} \rfloor$, let

$$\mathbf{Q}[k] = \Phi(r + n\chi k - 1, r + n\chi(k - 1)). \quad (27)$$

From Proposition 1, we have

$$\begin{aligned} \mathbf{Q}[k] &= \mathbf{A}[r + n\chi k - 1] \mathbf{A}[r + n\chi k - 2] \cdots \mathbf{A}[r + n\chi(k - 1)] \\ &\geq \xi^{n\chi} \prod_{\tau=r+n\chi(k-1)}^{r+n\chi k-1} \mathbf{H}[\tau]. \end{aligned} \quad (28)$$

The matrix corresponding to at least one reduced graph, say \mathcal{H}^* , appears at least n times in the product $\prod_{\tau=r+n\chi(k-1)}^{r+n\chi k-1} \mathbf{H}[\tau]$. Let $S \subseteq \mathcal{V}$ be the unique source component of \mathcal{H}^* . Due to the existence of self-loops, and the fact that only nodes *not* in $\bar{\mathcal{N}}[\tau]$ can be the removed sink nodes (as per Definition 1) in $\mathcal{H}[\tau]$, we know that each $j \in S$ reaches every node in $\bar{\mathcal{N}}[r + n\chi k - 1]$. Thus, for $j \in S$

$$\left[\prod_{\tau=r+n\chi(k-1)}^{r+n\chi k-1} \mathbf{H}[\tau] \right]_{ij} \geq 1, \forall i \in \bar{\mathcal{N}}[r + n\chi k - 1]. \quad (29)$$

By (28) and (29), for $i \in \bar{\mathcal{N}}[r + n\chi k - 1]$ and $j \in S$, we have

$$\mathbf{Q}_{ij}[k] \geq \xi^{n\chi}. \quad (30)$$

Now let us consider the case when $\bar{\mathcal{N}}[r + n\chi k - 1] = \mathcal{N}[r + n\chi(k - 1)]$, that is, no agents crash through iteration $r + n\chi(k - 1)$ to the end of iteration $r + n\chi k - 1$.

$$\begin{aligned} \eta_{r+n\chi(k-1)}(\mathbf{Q}[k]) &= \eta_{r+n\chi(k-1)}(\Phi(r + n\chi k - 1, r + n\chi(k - 1))) \\ &= \min_{i, i' \in \mathcal{N}[r+n\chi(k-1)]} \sum_{j \in \mathcal{V}} \min \{ \Phi_{ij}(r + n\chi k - 1, r + n\chi(k - 1)), \Phi_{i'j}(r + n\chi k - 1, r + n\chi(k - 1)) \} \\ &\geq \min_{i, i' \in \mathcal{N}[r+n\chi(k-1)]} \sum_{j \in S} \min \{ \Phi_{ij}(r + n\chi k - 1, r + n\chi(k - 1)), \Phi_{i'j}(r + n\chi k - 1, r + n\chi(k - 1)) \} \\ &= \min_{i, i' \in \bar{\mathcal{N}}[r+n\chi k-1]} \sum_{j \in S} \min \{ \Phi_{ij}(r + n\chi k - 1, r + n\chi(k - 1)), \Phi_{i'j}(r + n\chi k - 1, r + n\chi(k - 1)) \} \\ &\geq \min_{i, i' \in \bar{\mathcal{N}}[r+n\chi k-1]} \xi^{n\chi} = \xi^{n\chi}. \end{aligned} \quad (31)$$

For $i, i' \in \bar{\mathcal{N}}[t]$, it holds that

$$\begin{aligned}
& |\Phi_{ij}(t, r) - \Phi_{i'j}(t, r)| \leq \max_{j \in \mathcal{V}} \max_{k, k' \in \bar{\mathcal{N}}[t]} (\Phi_{kj}(t, r) - \Phi_{k'j}(t, r)) \\
& = \max_{j \in \mathcal{V}} \max_{k, k' \in \bar{\mathcal{N}}[t+1]} (\Phi_{kj}(t, r) - \Phi_{k'j}(t, r)) \quad \text{since } \bar{\mathcal{N}}[t] = \mathcal{N}[t+1] \\
& = \delta_{t+1}(\Phi(t, r)) \quad \text{by (15)} \\
& = \delta_{t+1} \left(\Phi \left(t, r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor \right) \Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \\
& \leq \delta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(t, r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor \right) \Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \quad \text{by (17)} \\
& \leq \left(1 - \eta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(t, r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor \right) \right) \right) \delta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \\
& \hspace{20em} \text{by Lemma 2} \\
& \leq \delta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right). \tag{32}
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
& \delta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \\
& \leq \delta_{r+n\chi(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1)} \left(\Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \quad \text{by (17)} \\
& \leq \delta_{r+n\chi(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1)} \left(\mathbf{Q}[\lfloor \frac{t-r+1}{n\chi} \rfloor] \Phi \left(r + n\chi \left(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1 \right) - 1, r \right) \right) \quad \text{by (27)} \\
& \leq \left(1 - \eta_{r+n\chi(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1)} \mathbf{Q}[\lfloor \frac{t-r+1}{n\chi} \rfloor] \right) \delta_{r+n\chi(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1)} \left(\Phi \left(r + n\chi \left(\lfloor \frac{t-r+1}{n\chi} \rfloor - 1 \right) - 1, r \right) \right) \\
& \hspace{20em} \text{by Lemma 2.} \tag{33}
\end{aligned}$$

Repeatedly apply (33) $\lfloor \frac{t-r+1}{n\chi} \rfloor$ times, we have

$$\delta_{r+n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor} \left(\Phi \left(r + n\chi \lfloor \frac{t-r+1}{n\chi} \rfloor - 1, r \right) \right) \leq \prod_{k=1}^{\lfloor \frac{t-r+1}{n\chi} \rfloor} (1 - \eta_{r+n\chi(k-1)}(\mathbf{Q}[k])), \tag{34}$$

where each $\mathbf{Q}[k]$, as defined in (27), is a product of $n\chi$ matrices.

Define $k \in \{1, \dots, \lfloor \frac{t-r+1}{n\chi} \rfloor\} \triangleq \mathcal{K}$. Let $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that for each $k \in \bar{\mathcal{K}}$, it holds that $\bar{\mathcal{N}}[r + n\chi k - 1] = \mathcal{N}[r + n\chi(k-1)]$. Since at most f agents crash, $|\bar{\mathcal{K}}| \geq \lfloor \frac{t-r+1}{n\chi} \rfloor - f$. Thus, for $i, i' \in \bar{\mathcal{N}}[t]$ and for $j \in \mathcal{V}$, we have

$$\begin{aligned}
& |\Phi_{ij}(t, r) - \Phi_{i'j}(t, r)| \leq \prod_{k=1}^{\lfloor \frac{t-r+1}{n\chi} \rfloor} (1 - \eta_{r+n\chi(k-1)}(\mathbf{Q}[k])) \quad \text{by (34)} \\
& \leq \prod_{k \in \bar{\mathcal{K}}} (1 - \eta_{r+n\chi(k-1)}(\mathbf{Q}[k])) \\
& \leq \prod_{k \in \bar{\mathcal{K}}} (1 - \xi^{n\chi}) \quad \text{by (31)} \\
& \leq (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f} \quad \text{since } |\bar{\mathcal{K}}| \geq \lfloor \frac{t-r+1}{n\chi} \rfloor - f \tag{35}
\end{aligned}$$

Combining (26) and (35), for $i, i' \in \bar{\mathcal{N}}[t]$ and for $j \in \mathcal{V}$, we conclude that

$$|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| \leq \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\},$$

proving the theorem. □

An immediate consequence of Theorem 2 is that (13) holds. Thus Condition 2 is also sufficient for (13) to hold.

Our next lemma states that rows of $\Phi(t, r)$ with indices in \mathcal{N} have a common limit.

Lemma 3. *For all $i \in \mathcal{N}$, the i -th row of $\Phi(t, r)$, denoted by $\Phi_i(t, r)$ converges to a stochastic vector $\pi^T(r)$, i.e.,*

$$\Phi_i(t, r) \rightarrow \pi^T(r), \forall r. \quad (36)$$

Indeed, for each $i \in \mathcal{N}$, the row $\Phi_i(t, r)$ converges to $\pi(r)$ exponentially fast.

Proposition 3. *Suppose $r \leq t$. For any $i \in \mathcal{N}$ and for any $k \in \mathcal{V}$, it holds that*

$$|\Phi_{ik}(t, r) - \pi_k(r)| \leq \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\}.$$

In addition, the limit vector $\pi(r)$ has the following property.

Lemma 4. *For any r , there exists a reduced graph $\tilde{\mathcal{H}}[r]$ with source component S_r such that $\pi_j(r) \geq \xi^{n\chi}$, for each $j \in S_r$. In addition, $|S_r| \geq \gamma$.*

The proofs of Lemma 3, Proposition 3 and Lemma 4 can be found in Appendix C.

5 Convergence of Non-Bayesian Learning

In the absence of agent failures and messages asynchrony [3], for the networked agents to detect the true hypothesis θ^* , it is enough to assume that $G(\mathcal{V}, \mathcal{E})$ is strongly connected, and that θ^* is globally identifiable, i.e., for any $\theta, \theta^* \in \Theta$ and $\theta \neq \theta^*$, there exists a node $j \in \mathcal{V}$ such that

$$D(\ell_j(\cdot|\theta^*)||\ell_j(\cdot|\theta)) \triangleq \sum_{w_j \in \mathcal{S}_j} \ell_j(w_j|\theta^*) \log \frac{\ell_j(w_j|\theta^*)}{\ell_j(w_j|\theta)} \neq 0,$$

or equivalently,

$$\sum_{i \in \mathcal{V}} D(\ell_i(\cdot|\theta^*)||\ell_i(\cdot|\theta)) \neq 0. \quad (37)$$

Since θ^* may change from execution to execution, (37) is required to hold for any choice of θ and θ^* such that $\theta \neq \theta^*$. Intuitively speaking, if any pair of states θ_1 and θ_2 can be distinguished by at least one agent in the network, then sufficient exchange of local beliefs over strongly connected network will enable every agent to distinguish θ_1 from θ_2 . However, in the presence of failures and asynchrony, the effective influence network (reduced graphs) may not be strongly connected. Thus, stronger global identifiability of the network is required.

Assumption 1 For any $\theta, \theta^* \in \Theta$, $\theta \neq \theta^*$, and for any reduced graph \mathcal{H} of $G(\mathcal{V}, \mathcal{E})$ with $\mathcal{S}_{\mathcal{H}}$ denoting the unique source component, the following holds.

$$\sum_{i \in \mathcal{S}_{\mathcal{H}}} D(\ell_i(\cdot|\theta^*)||\ell_i(\cdot|\theta)) \neq 0. \quad (38)$$

In contrast to (37), where the summation is taken over all the agents in the network, in (38), the summation is taken over the source component of an arbitrary reduced graph. Intuitively, the condition imposed by Assumption 1 is that all the agents in the source component can detect the true state θ^* collaboratively. Then if iterative consensus, specified in (13), is achieved, the accurate belief can be propagated from the source component to every non-faulty agent in the network. Assumption 1 is necessary for (2) to be achievable. This is because if Assumption 1 is violated, no information outside the source component is available to the agents in the component to distinguish two hypotheses, and θ^* cannot be learned. In addition, Assumption 1 is also sufficient for θ^* to be learned. Our convergence analysis of the proposed protocol has a similar structure to that in [15].

Recall that \mathcal{C} is the collection of all reduced graphs of $G(\mathcal{V}, \mathcal{E})$. Let $\mathcal{H} \in \mathcal{C}$ with source component $\mathcal{S}_{\mathcal{H}}$. Define C_0 and C_1 as

$$-C_0 \triangleq \min_{i \in \mathcal{V}} \min_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \min_{w_i \in \mathcal{S}_i} \left(\log \frac{\ell_i(w_i|\theta_1)}{\ell_i(w_i|\theta_2)} \right), \quad (39)$$

and

$$C_1 \triangleq \min_{\mathcal{H} \in \mathcal{C}} \min_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \sum_{i \in \mathcal{S}_{\mathcal{H}}} D(\ell_i(\cdot|\theta_1)||\ell_i(\cdot|\theta_2)). \quad (40)$$

Since $|\Theta| = m < \infty$ and the finiteness of $|\mathcal{S}_i|$ for each $i \in \mathcal{V}$, we know that $-C_0 > -\infty$. In addition, it is easy to see that $-C_0 \leq 0$ (thus, $C_0 \geq 0$). From Assumption 1 and the fact that $|\mathcal{C}| = \chi < \infty$, we get $C_1 > 0$.

Theorem 3. When $G(\mathcal{V}, \mathcal{E})$ satisfies Condition 1, and Assumption 1 holds, each non-faulty agent $i \in \mathcal{N}$ learns the true hypothesis θ^* almost surely, i.e.,

$$\mu_i^i(\theta^*) \xrightarrow{\text{a.s.}} 1, \quad \forall i \in \mathcal{N}.$$

Proof. Consider any $\theta \in \Theta$ such that $\theta \neq \theta^*$, where θ^* is the underlying true state. Recall from (12) that

$$\psi_t(\theta) = \Phi(t, 1)\psi_0(\theta) + \sum_{r=1}^t \Phi(t, r+1)\mathcal{L}_r(\theta)$$

Define $H(\theta) \in \mathbb{R}^n$ such that

$$H_i(\theta) \triangleq \sum_{w_i \in \mathcal{S}_i} \ell_i(w_i|\theta^*) \log \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} = -D(\ell_i(\cdot|\theta^*)||\ell_i(\cdot|\theta)). \quad (41)$$

Since $D(\ell_i(\cdot|\theta^*)||\ell_i(\cdot|\theta)) \geq 0$, it follows that $H_i(\theta) \leq 0$.⁴ Note that, for each $t \geq 1$ and for each $i \in \mathcal{N}[t]$, it holds that

$$H_i(\theta) = \mathbb{E}^*[\mathcal{L}_t^i(\theta)], \quad (42)$$

⁴ Alternatively, since $\log(\cdot)$ is concave, by Jensen's inequality, we get

$$H_i(\theta) = \sum_{w_i \in \mathcal{S}_i} \ell_i(w_i|\theta^*) \log \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} \leq \log \left(\sum_{w_i \in \mathcal{S}_i} \ell_i(w_i|\theta^*) \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} \right) = \log 1 = 0.$$

where the expectation $\mathbb{E}^*[\cdot]$ is taken over $\ell_i(\cdot|\theta^*)$. Since the support of $\ell_i(\cdot|\theta)$ is the whole signal space \mathcal{S}_i for each agent $i \in \mathcal{V}$, it holds that $\left| \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} \right| < \infty$ for each $w_i \in \mathcal{S}_i$, and

$$\begin{aligned} H_i(\theta) &= \sum_{w_i \in \mathcal{S}_i} \ell_i(w_i|\theta^*) \log \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} \geq \min_{w_i \in \mathcal{S}_i} \left(\log \frac{\ell_i(w_i|\theta)}{\ell_i(w_i|\theta^*)} \right) \quad \text{since } \sum_{w_i \in \mathcal{S}_i} \ell_i(w_i|\theta^*) = 1 \\ &\geq \min_{i \in \mathcal{V}} \min_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \min_{w_i \in \mathcal{S}_i} \left(\log \frac{\ell_i(w_i|\theta_1)}{\ell_i(w_i|\theta_2)} \right) = -C_0 > -\infty. \end{aligned} \quad (43)$$

Recall that $\boldsymbol{\pi}(r+1) \in \mathbb{R}^n$ is the limit of $[\boldsymbol{\Phi}(t, r+1)]_i$ for each $i \in \mathcal{N}$, and is stochastic. By (43), we know that

$$-\infty < -C_0 \leq \boldsymbol{\pi}^T(r+1)H(\theta) \leq 0.$$

Due to the finiteness of $\boldsymbol{\pi}^T(r+1)H(\theta)$, we are able to add $\sum_{r=1}^t \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta)$ and subtract $\sum_{r=1}^t \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta)$ from (12), where $\mathbf{1}_n \in \mathbb{R}^n$ is an all one vector of dimension n . So we get

$$\begin{aligned} \boldsymbol{\psi}_t(\theta) &= \boldsymbol{\Phi}(t, 1)\boldsymbol{\psi}_0(\theta) + \sum_{r=1}^t \boldsymbol{\Phi}(t, r+1)\mathcal{L}_r(\theta) \\ &= \boldsymbol{\Phi}(t, 1)\boldsymbol{\psi}_0(\theta) + \sum_{r=1}^t \boldsymbol{\Phi}(t, r+1)\mathcal{L}_r(\theta) - \sum_{r=1}^t \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta) + \sum_{r=1}^t \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta) \\ &= \boldsymbol{\Phi}(t, 1)\boldsymbol{\psi}_0(\theta) + \sum_{r=1}^t (\boldsymbol{\Phi}(t, r+1)\mathcal{L}_r(\theta) - \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta)) + \sum_{r=1}^t \mathbf{1}_n \boldsymbol{\pi}^T(r+1)H(\theta). \end{aligned} \quad (44)$$

For each $i \in \mathcal{N}$, we have

$$\begin{aligned} \boldsymbol{\psi}_t^i(\theta) &= \sum_{k=1}^n \boldsymbol{\Phi}_{ik}(t, 1)\boldsymbol{\psi}_0^i(\theta) + \sum_{r=1}^t \left(\sum_{k=1}^n \boldsymbol{\Phi}_{ik}(t, r+1)\mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1)H_k(\theta) \right) \\ &\quad + \sum_{r=1}^t \sum_{k=1}^n \pi_k(r+1)H_k(\theta). \end{aligned} \quad (45)$$

To show $\lim_{t \rightarrow \infty} \mu_t^i(\theta^*) \xrightarrow{\text{a.s.}} 1$, it is enough to show $\boldsymbol{\psi}_t^i(\theta) \xrightarrow{\text{a.s.}} -\infty$ for $\theta \neq \theta^*$. The remaining proof of convergence has similar structure as the analysis in [15]. We will bound the three terms in the right hand side of (45) separately. Since μ_0^i is uniform, $\boldsymbol{\psi}_0^i(\theta) = 0$. We show that the second term of the RHS of (45) decreases linearly in t . For each $r \leq t$, let S_r be the set of agents that has the property stated in Lemma 4. Then, we have

$$\begin{aligned} \sum_{k=1}^n \pi_k(r+1)H_k(\theta) &\leq \sum_{k \in S_{r+1}} \pi_k(r+1)H_k(\theta) \quad \text{since } H_k(\theta) \leq 0 \\ &= - \sum_{k \in S_{r+1}} \pi_k(r+1)D(\ell_k(\cdot|\theta^*)||\ell_k(\cdot|\theta)) \quad \text{by (41)} \\ &\leq -\xi^{n\chi} \sum_{k \in S_{r+1}} D(\ell_k(\cdot|\theta^*)||\ell_k(\cdot|\theta)) \quad \text{by Lemma 4} \\ &\leq -\xi^{n\chi} C_1 \quad \text{by (40)} \end{aligned}$$

Thus, we get

$$\sum_{r=1}^t \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \leq - \sum_{r=1}^t \sum_{j \in S_r} \xi^{n\chi} |H_j(\theta)| \leq -C_1 \xi^{n\chi} t. \quad (46)$$

Using Kolmogorov's strong law of large number, as shown below, we can prove that

$$\frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \xrightarrow{\text{a.s.}} 0.$$

Specifically,

$$\begin{aligned} & \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \\ &= \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) + \sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \\ &= \frac{1}{t} \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) + \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right). \end{aligned} \quad (47)$$

For each $k \notin \bar{\mathcal{N}}[t]$, by (8), we know that

$$\left| \mathcal{L}_r^k(\theta) \right| = 0. \quad (48)$$

For each $k \in \bar{\mathcal{N}}[t]$, by (8), we have

$$\left| \mathcal{L}_r^k(\theta) \right| = \left| \log \frac{\ell_i(s_t^i | \theta)}{\ell_i(s_t^i | \theta^*)} \right| \leq \max_{i \in \mathcal{V}} \max_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \max_{w_i \in \mathcal{S}_i} \left| \log \frac{\ell_i(w_i | \theta_1)}{\ell_i(w_i | \theta_2)} \right|.$$

Note that $\max_{i \in \mathcal{V}} \max_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \max_{w_i \in \mathcal{S}_i} \left| \log \frac{\ell_i(w_i | \theta_1)}{\ell_i(w_i | \theta_2)} \right|$ is symmetric in θ_1 and θ_2 . Thus,

$$\begin{aligned} \left| \mathcal{L}_r^k(\theta) \right| &\leq \max_{i \in \mathcal{V}} \max_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \max_{w_i \in \mathcal{S}_i} \left| \log \frac{\ell_i(w_i | \theta_1)}{\ell_i(w_i | \theta_2)} \right| = \max_{i \in \mathcal{V}} \max_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \max_{w_i \in \mathcal{S}_i} \log \frac{\ell_i(w_i | \theta_1)}{\ell_i(w_i | \theta_2)} \\ &= \max_{i \in \mathcal{V}} \max_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \max_{w_i \in \mathcal{S}_i} - \log \frac{\ell_i(w_i | \theta_2)}{\ell_i(w_i | \theta_1)} \\ &= - \min_{i \in \mathcal{V}} \min_{\theta_1, \theta_2 \in \Theta; \theta_1 \neq \theta_2} \min_{w_i \in \mathcal{S}_i} \log \frac{\ell_i(w_i | \theta_2)}{\ell_i(w_i | \theta_1)} = -(-C_0) = C_0. \end{aligned} \quad (49)$$

By (48) and (49), we get

$$\left| \mathcal{L}_r^k(\theta) \right| \leq C_0, \quad \forall k \in \mathcal{N}. \quad (50)$$

Note that $\sum_{r=1}^t \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r}{n\chi} \rfloor - f}\}$ is a geometric series for sufficient large t , and is convergent. In particular,

$$\sum_{r=1}^t \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r}{n\chi} \rfloor - f}\} \leq \sum_{r=1}^{\infty} \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r}{n\chi} \rfloor - f}\} \triangleq C < \infty. \quad (51)$$

For each $i \in \mathcal{N}$, the absolute value of the first term in the right hand side of (47) can be bounded as follows.

$$\begin{aligned}
& \frac{1}{t} \left| \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) \right| \\
& \leq \frac{1}{t} \sum_{r=1}^t \sum_{k=1}^n |\Phi_{ik}(t, r+1) - \pi_k(r+1)| \left| \mathcal{L}_r^k(\theta) \right| \\
& \leq \frac{1}{t} \sum_{r=1}^t \sum_{k=1}^n |\Phi_{ik}(t, r+1) - \pi_k(r+1)| C_0 \text{ by (50)} \\
& = \frac{1}{t} \left(\sum_{r=1}^{t-1} \sum_{k=1}^n |\Phi_{ik}(t, r+1) - \pi_k(r+1)| C_0 + \sum_{k=1}^n |\Phi_{ik}(t, t+1) - \pi_k(t+1)| C_0 \right) \\
& \leq \frac{1}{t} \left(\sum_{r=1}^{t-1} \sum_{k=1}^n |\Phi_{ik}(t, r+1) - \pi_k(r+1)| C_0 + nC_0 \right) \text{ since } |\Phi_{ik}(t, r+1) - \pi_k(r+1)| \leq 1 \\
& \leq \frac{1}{t} \left(\sum_{r=1}^{t-1} n \cdot \min\{1, (1 - \xi^{n\lambda})^{\lfloor \frac{t-r}{n\lambda} \rfloor - f}\} C_0 + nC_0 \right) \text{ by Proposition 3} \\
& \leq \frac{1}{t} \sum_{r=1}^t n \cdot \min\{1, (1 - \xi^{n\lambda})^{\lfloor \frac{t-r}{n\lambda} \rfloor - f}\} C_0 \text{ since } \min\{1, (1 - \xi^{n\lambda})^{\lfloor \frac{t-t}{n\lambda} \rfloor - f}\} = 1 \\
& \leq \frac{1}{t} nCC_0 \text{ by (51)} \tag{52}
\end{aligned}$$

Taking limsup on both sides of (52), we get

$$\begin{aligned}
\limsup_{t \rightarrow \infty} \frac{1}{t} \left| \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) \right| & \leq \limsup_{t \rightarrow \infty} \frac{1}{t} nCC_0 \\
& = 0 \\
& \leq \liminf_{t \rightarrow \infty} \frac{1}{t} \left| \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) \right|.
\end{aligned}$$

Thus, the limit of $\frac{1}{t} \left| \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) \right|$ exists, and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \left| \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) \right| = 0. \tag{53}$$

Therefore,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{r=1}^t \sum_{k=1}^n (\Phi_{ik}(t, r+1) - \pi_k(r+1)) \mathcal{L}_r^k(\theta) = 0, \tag{54}$$

i.e., the first term in the right hand side of (47) goes to 0.

Now we bound the second term in the right hand side of (47), using Kolmogorov's strong law of large number. Kolmogorov's strong law of large number states that if $\{X_r\}_{r=1}^\infty$ is a sequence of

independent random variables such that $\sum_{r=1}^{\infty} \frac{\text{Var}(X_r)}{r^2} < \infty$, then

$$\frac{1}{t} \sum_{r=1}^t X_r - \frac{1}{t} \sum_{r=1}^t \mathbb{E}[X_r] \xrightarrow{\text{a.s.}} 0.$$

For each r , let

$$X_r \triangleq \sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta). \quad (55)$$

Since the private signals s_r^i 's are i.i.d. across iterations, X_r 's are independent. In addition, since

$$\mathbb{E}[X_r^2] = \mathbb{E} \left[\left(\sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) \right)^2 \right] \leq \mathbb{E} \left[\max_{k \in \mathcal{V}} |\mathcal{L}_r^k(\theta)|^2 \right] \leq C_0^2,$$

where the last inequality follows from (50), we get

$$\sum_{r=1}^{\infty} \frac{\text{Var}(X_r)}{r^2} \leq \sum_{r=1}^{\infty} \frac{\mathbb{E}[X_r^2]}{r^2} \leq \sum_{r=1}^{\infty} \frac{1}{r^2} C_0^2 < \infty.$$

For each $i \in \mathcal{N}$ and each $k \in \mathcal{V}$, by Lemma 3, it holds that $\pi_k(r+1) = \lim_{t \rightarrow \infty} \Phi_{ik}(t, r+1)$. In addition, by (2), for each $k \notin \mathcal{N}[t+1]$, we have

$$\pi_k(r+1) = \lim_{t \rightarrow \infty} \Phi_{ik}(t, r+1) = 0. \quad (56)$$

Thus, we get

$$X_r = \sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) = \sum_{k \in \mathcal{N}[t+1]} \pi_k(r+1) \mathcal{L}_r^k(\theta). \quad (57)$$

Thus, we have, for each $i \in \mathcal{N}$

$$\begin{aligned} \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \pi_k(r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) &= \frac{1}{t} \sum_{r=1}^t \left(X_r - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \quad \text{by (55)} \\ &= \frac{1}{t} \sum_{r=1}^t \left(X_r - \sum_{k \in \mathcal{N}[r+1]} \pi_k(r+1) H_k(\theta) \right) \quad \text{by (56)} \\ &= \frac{1}{t} \sum_{r=1}^t \left(X_r - \sum_{k \in \mathcal{N}[r+1]} \pi_k(r+1) \mathbb{E}^*[\mathcal{L}_r^k(\theta)] \right) \quad \text{by (42)} \\ &= \frac{1}{t} \sum_{r=1}^t (X_r - \mathbb{E}^*[X_r]) \quad \text{by (57)} \\ &\xrightarrow{\text{a.s.}} 0. \quad (58) \end{aligned}$$

By (54), (58) and (47), we obtain that

$$\frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \xrightarrow{\text{a.s.}} 0. \quad (59)$$

From (45), for each $i \in \mathcal{N}$ and $\theta \neq \theta^*$, we get

$$\begin{aligned}
\limsup_{t \rightarrow \infty} \frac{1}{t} \psi_t^i(\theta) &= \limsup_{t \rightarrow \infty} \Phi_{ik}(t, 1) \psi_0^i(\theta) + \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \\
&\quad + \frac{1}{t} \limsup_{t \rightarrow \infty} \sum_{r=1}^t \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \\
&= 0 + \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r+1) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) \quad \text{since } \psi_0^i(\theta) = 0 \\
&\quad + \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{r=1}^t \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \\
&\leq \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{r=1}^t \left(\sum_{k=1}^n \Phi_{ik}(t, r) \mathcal{L}_r^k(\theta) - \sum_{k=1}^n \pi_k(r+1) H_k(\theta) \right) - \liminf_{t \rightarrow \infty} \left(\frac{1}{t} t \xi^{n\chi} C_1 \right) \quad \text{by (46)} \\
&= -\xi^{n\chi} C_1 \quad \text{by (59) a.s.}
\end{aligned}$$

Consequently, for each $i \in \mathcal{N}$ and each $\theta \neq \theta^*$, we have

$$\psi_t^i(\theta) \xrightarrow{\text{a.s.}} -\infty. \quad \text{and} \quad \lim_{t \rightarrow \infty} \mu_t^i(\theta) \xrightarrow{\text{a.s.}} 0.$$

Due to (7) and the fact that $\tilde{\mu}_t^i(\theta) = \mu_t^i(\theta), \forall i \in \mathcal{N}$, we get

$$\lim_{t \rightarrow \infty} \mu_t^i(\theta) \xrightarrow{\text{a.s.}} 0.$$

Thus, for each $i \in \mathcal{N}$,

$$\lim_{t \rightarrow \infty} \mu_t^i(\theta^*) \xrightarrow{\text{a.s.}} 1.$$

Therefore, we have shown that all non-faulty agents learn the true parameter θ^* almost surely. \square

Remark 2. As it can be seen from the above analysis, our results can be trivially adapted to time-varying graphs. In addition, most of the existing results on (asymptotic and finite-time) convergence rates can be generalized to our setting.

Appendices

A Equivalence of Condition 1 and Condition 2

The proof of Theorem 1 relies on the following proposition.

Proposition 4. *Every subgraph \mathcal{H}' of $G(\mathcal{V}, \mathcal{E})$ obtained by removing up to f incoming links for each $i \in \mathcal{V}$ contains a unique source component.*

Proof. Suppose \mathcal{H}' contains at least two source components. Let L and R be two source components in \mathcal{H}' , and let $C = \mathcal{V} - L - R$. It is easy to see that $L \neq \emptyset$ and $R \neq \emptyset$. Since L is a source component, no nodes in L have incoming links from $R \cup C$. In addition, by the construction of \mathcal{H}' , we know that in $G(\mathcal{V}, \mathcal{E})$, every node in L has at most f incoming neighbors in $R \cup C$, i.e., $|N_i^- \cap (R \cup C)| \leq f$ for each $i \in L$. Similarly, we have $|N_j^- \cap (L \cup C)| \leq f$ for each $j \in R$. This contradicts the assumption that $G(\mathcal{V}, \mathcal{E})$ satisfies Condition 1.

Therefore, \mathcal{H}' contains only one source component. □

Proof (Proof of Theorem 1). We first show that Condition 1 implies Condition 2, i.e., if the graph $G(\mathcal{V}, \mathcal{E})$ satisfies Condition 1, then by Proposition 4, every reduced graph \mathcal{H} of $G(\mathcal{V}, \mathcal{E})$, defined as per Definition 1, has only one source component.

Let \mathcal{H}' be an arbitrary subgraph of $G(\mathcal{V}, \mathcal{E})$ obtained by removing up to f incoming links for each node. From Definition 1, it can be seen that to show Condition 1 implies Condition 2, it is enough to show that every reduced graph obtained from \mathcal{H}' contains only one source component.

From Definition 1, we know that when \mathcal{H}' contains no sink nodes, \mathcal{H}' is the only reduced graph associated with \mathcal{H}' . By Proposition 4, \mathcal{H}' contains only one source component. Similarly, from Definition 1, when \mathcal{H}' contains at least one sink node – a node that does not have any outgoing links, multiple reduced graphs can be obtained from \mathcal{H}' . Let \mathcal{H} be an arbitrary reduced graph associated with \mathcal{H}' such that $\mathcal{H} \neq \mathcal{H}'$, with $\{i_1, \dots, i_p\}$ being the p (where $1 \leq p \leq f$) sink nodes removed from \mathcal{H}' . The uniqueness of source component in \mathcal{H}' implies that there exists a node that can reach any other node in \mathcal{H}' . That is, there exists a node j such that for any $i \in \mathcal{V}$, $i \neq j$, there is a j, i -path in \mathcal{H}' . Since a sink node does not have any outgoing link, any intermediate node of a j, i -path in \mathcal{H}' cannot be a sink node. Then, if $i \notin \{i_1, \dots, i_p\}$, any j, i -path in \mathcal{H}' is remained in \mathcal{H} . Thus, node j can reach any node $i \notin \{i_1, \dots, i_p\}$ in \mathcal{H} , proving \mathcal{H} contains only one source component. By the arbitrariness of \mathcal{H} , we conclude that every reduced graph obtained from \mathcal{H}' contains only one source component.

Since \mathcal{H}' is also chosen arbitrarily, therefore, we conclude that Condition 1 implies Condition 2.

Next we show that Condition 2 implies Condition 1, i.e., if every reduced graph of $G(\mathcal{V}, \mathcal{E})$ contains only one source component, then $G(\mathcal{V}, \mathcal{E})$ satisfies Condition 1. We prove this by contradiction. Suppose $G(\mathcal{V}, \mathcal{E})$ satisfies Condition 2, but does not satisfy Condition 1, i.e., there exists a node partition L, R, C such that $L \neq \emptyset$, $R \neq \emptyset$, and the following holds: for $i \in L$, $|N_i^- \cap (R \cup C)| \leq f$, and for $j \in R$, $|N_j^- \cap (L \cup C)| \leq f$. Now, consider the reduced graph constructed as follows: for each $i \in L$, remove all incoming links from $R \cup C$, and for each $j \in R$, remove all incoming links

form $L \cup C$. Denote the obtained subgraph as \mathcal{H} . From Definition 1, we know that \mathcal{H} is a reduced graph. Since nodes in L do not have any incoming link from $R \cup C$, and nodes in R do not have any incoming link from $L \cup C$, thus both L and R contain source components. Consequently, there are at least two source components in \mathcal{H} , contradicting the fact that \mathcal{H} contains only one source component. Thus, we conclude that Condition 2 implies Condition 1.

Therefore, Condition 1 and Condition 2 are equivalent. □

B Proof of Proposition 1

Proof. We consider two cases separately: (1) every agent participates in iteration t , i.e., $\mathcal{N}[t] = \mathcal{V}$; and (2) there exists an agent that does not participate in iteration t , i.e., $\mathcal{V} - \mathcal{N}[t] \neq \emptyset$.

Case (1): Suppose every agent participates in iteration t , i.e., $\mathcal{N}[t] = \mathcal{V}$. By the construction of matrix $\mathbf{A}[t]$ in (3) and (4), for each $i \in \mathcal{V}$, $\mathbf{A}_{ij}[t] \geq \xi$ for $|N_i^-| - f + 1$ incoming neighbors of node i . This corresponds to the operation of removing f incoming links of node i . Thus, there exists a reduced graph $\mathcal{H}[t]$ with adjacency matrix $\mathbf{H}[t]$ such that

$$\mathbf{A}[t] \geq \xi \mathbf{H}[t].$$

Case (2): Suppose there exists agent that does not participate iteration t , i.e., $\mathcal{V} - \mathcal{N}[t] \neq \emptyset$. In this case, since each agent in $\mathcal{V} - \mathcal{N}[t]$ has already crashed during the first $t - 1$ iterations, it will no longer send out messages as well as receiving messages. That is, nodes in $\mathcal{V} - \mathcal{N}[t]$ are isolated nodes in iteration t . By construction of $\mathbf{A}[t]$ in (3) and (4), we know that $\mathbf{A}_{ij}[t] = 0$ for each $j \in \mathcal{V} - \mathcal{N}[t]$, where $i \neq j$. Intuitively speaking, this corresponds to the fact that a crashed agent cannot influence any other agent, and thus is a sink node. In addition, the fact that $\mathbf{A}_{ji}[t] = 0$ for any $j \neq i$ and $j \in \mathcal{V} - \mathcal{N}[t]$ corresponds to the operation that the sink nodes in $\mathcal{V} - \mathcal{N}[t]$ are removed. Therefore, there exists a reduced graph $\mathcal{H}[t]$ with adjacency matrix $\mathbf{H}[t]$ such that

$$\mathbf{A}[t] \geq \xi \mathbf{H}[t].$$

Therefore, the proof of Proposition 1 is complete. □

C Proofs of Lemma 3, Proposition 3 and Lemma 4

Proof (Proof of Lemma 3). Since \mathbb{R} is complete, it is enough to show that $\{\Phi_{ij}(t, r)\}_{t=r}^{\infty}$ is a Cauchy sequence. That is, we need to show that for any $\epsilon > 0$, there exists $t^* \geq r$ such that for any $t \geq t^*$ and any $p \in \mathbb{N}$

$$|\Phi_{ij}(t + p, r) - \Phi_{ij}(t, r)| < \epsilon, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{V}.$$

By Theorem 2, we know for any $\epsilon > 0$, there exists $t^* \geq r$ such that for all $t \geq t^*$

$$|\Phi_{ij}(t, r) - \Phi_{kj}(t, r)| < \epsilon, \quad \forall i, k \in \bar{\mathcal{N}}[t], \forall j \in \mathcal{V}. \quad (60)$$

For any $p \in \mathbb{N}$, we have

$$\begin{aligned}
|\Phi_{ij}(t+p, r) - \Phi_{ij}(t, r)| &= \left| \sum_{k=1}^n \Phi_{ik}(t+p, t+1) \Phi_{kj}(t, r) - \Phi_{ij}(t, r) \right| \\
&= \left| \sum_{k=1}^n \Phi_{ik}(t+p, t+1) (\Phi_{kj}(t, r) - \Phi_{ij}(t, r)) \right| \quad \text{since } \sum_{k=1}^n \Phi_{ik}(t+p, t+1) = 1 \\
&= \left| \sum_{k \in \mathcal{N}[t+1]} \Phi_{ik}(t+p, t+1) (\Phi_{kj}(t, r) - \Phi_{ij}(t, r)) \right| \quad \text{by Proposition 2} \\
&\leq \sum_{k \in \mathcal{N}[t+1]} \Phi_{ik}(t+p, t+1) |\Phi_{kj}(t, r) - \Phi_{ij}(t, r)| \\
&= \sum_{k \in \mathcal{N}[t]} \Phi_{ik}(t+p, t+1) |\Phi_{kj}(t, r) - \Phi_{ij}(t, r)| \\
&\leq \left(\sum_{k \in \mathcal{N}[t]} \Phi_{ik}(t+p, t+1) \right) \cdot \epsilon \quad \text{by (60)} \\
&= \epsilon.
\end{aligned}$$

Thus, there exists $\pi(r)$ such that (36) holds. Since $\Phi(t, r)$ is row-stochastic, the limiting vector $\pi(r)$ is also stochastic, proving the lemma. \square

Proof (Proof of Proposition 3). We now show that $\Phi_{ik}(t, r) - \pi_k(r) \leq \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\}$, for $i \in \mathcal{N}$ and $k \in \mathcal{V}$. For any $p \in \mathbb{N}$,

$$\begin{aligned}
\Phi_{ik}(t, r) - \pi_k(r) &= \Phi_{ik}(t, r) - \lim_{p \rightarrow \infty} \Phi_{ik}(t+p, r) \quad \text{by Lemma 3} \\
&= \lim_{p \rightarrow \infty} (\Phi_{ik}(t, r) - \Phi_{ik}(t+p, r)) \\
&= \lim_{p \rightarrow \infty} \left(\Phi_{ik}(t, r) - \sum_{j=1}^n \Phi_{ij}(t+p, t+1) \Phi_{jk}(t, r) \right) \\
&= \lim_{p \rightarrow \infty} \sum_{j=1}^n \Phi_{ij}(t+p, t+1) (\Phi_{ik}(t, r) - \Phi_{jk}(t, r)) \quad \text{since } \sum_{j=1}^n \Phi_{ij}(t+p, t+1) = 1 \\
&= \lim_{p \rightarrow \infty} \sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t+p, t+1) (\Phi_{ik}(t, r) - \Phi_{jk}(t, r)) \\
&\quad + \lim_{p \rightarrow \infty} \sum_{j \notin \mathcal{N}[t]} \Phi_{ij}(t+p, t+1) (\Phi_{ik}(t, r) - \Phi_{jk}(t, r)) \\
&= \lim_{p \rightarrow \infty} \sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t+p, t+1) (\Phi_{ik}(t, r) - \Phi_{jk}(t, r)) \quad \text{by Proposition 2} \quad (61) \\
&\leq \sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t+p, t+1) \cdot \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\} \quad \text{by Theorem 2 and } \mathcal{N}[t] = \mathcal{N}[t+1] \\
&= \min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\} \quad \text{by (20)}
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
\Phi_{ik}(t, r) - \pi_k(r) &= \lim_{p \rightarrow \infty} \sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t + p, t + 1) (\Phi_{ik}(t, r) - \Phi_{jk}(t, r)) \quad \text{as per (61)} \\
&\geq \sum_{j \in \mathcal{N}[t]} \Phi_{ij}(t + p, t + 1) \cdot \left(-\min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\} \right) \quad \text{by Theorem 2} \\
&= -\min\{1, (1 - \xi^{n\chi})^{\lfloor \frac{t-r+1}{n\chi} \rfloor - f}\}.
\end{aligned}$$

The proof of Proposition 3 is complete. \square

In addition, the limit vector $\pi(r)$ has the following property.

Proof (Proof of Lemma 4). For each $j \in \mathcal{V}$, we have

$$\begin{aligned}
\pi_j(r) &= \lim_{t \rightarrow \infty} \Phi_{ij}(t, r) \\
&= \lim_{t \rightarrow \infty} \sum_{k=1}^n \Phi_{ik}(t, r + n\chi) \Phi_{kj}(r + n\chi - 1, r) \\
&= \lim_{t \rightarrow \infty} \left(\sum_{k \in \mathcal{N}[r+n\chi]} \Phi_{ik}(t, r + n\chi) \Phi_{kj}(r + n\chi - 1, r) + \sum_{k \notin \mathcal{N}[r+n\chi]} \Phi_{ik}(t, r + n\chi) \Phi_{kj}(r + n\chi - 1, r) \right) \\
&= \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}[r+n\chi]} \Phi_{ik}(t, r + n\chi) \Phi_{kj}(r + n\chi - 1, r).
\end{aligned}$$

From the proof of Theorem 2, we know that there exists a reduced graph $\tilde{\mathcal{H}}[r]$ with source component S_r such that for each $j \in S_r$,

$$\Phi_{kj}(r + n\chi - 1, r) \geq \xi^{n\chi}, \quad \forall k \in \tilde{\mathcal{N}}[r + n\chi - 1].$$

Thus, for each $j \in S_r$ we have

$$\begin{aligned}
\pi_j(r) &= \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}[r+n\chi]} \Phi_{ik}(t, r + n\chi) \Phi_{kj}(r + n\chi - 1, r) \\
&\geq \lim_{t \rightarrow \infty} \left(\sum_{k \in \mathcal{N}[r+n\chi]} \Phi_{ik}(t, r + n\chi) \right) \xi^{n\chi} \\
&= \xi^{n\chi} \quad \text{by Proposition 2.}
\end{aligned}$$

Recall that γ is the minimum source component size in all reduced graph. So $|S_r| \geq \gamma$. The proof is complete. \square

References

1. J.-F. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Transactions on Signal Processing*, 2003.

2. D. Gale and S. Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 2003.
3. A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. *Games and Economic Behavior*, 2012.
4. B. C. Levy. *Principles of Signal Detection and Parameter Estimation*. Springer Science & Business Media, 2008.
5. J. Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals and Systems*, 1988.
6. J. N. Tsitsiklis. Decentralized detection. *Advances in Statistical Signal Processing*, 1993.
7. P. K. Varshney. *Distributed Detection and Data Fusion*. Springer Science & Business Media, 2012.
8. E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems*. Springer Science & Business Media, 2012.
9. H. Attiya and J. Welch. *Distributed computing: fundamentals, simulations, and advanced topics*, volume 19. John Wiley & Sons, 2004.
10. J. Hajnal and M. Bartlett. Weak ergodicity in non-homogeneous markov chains. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 233–246. Cambridge Univ Press, 1958.
11. A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi. Information heterogeneity and the speed of learning in social networks. *Columbia Business School Research Paper*, (13-28), 2013.
12. A. Lalitha, A. Sarwate, and T. Javidi. Social learning and distributed hypothesis testing. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 551–555, June 2014.
13. N. A. Lynch. *Distributed Algorithms*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996.
14. P. Molavi and A. Jadbabaie. Foundations of non-bayesian social learning. *Columbia Business School Research Paper*, 2015.
15. A. Nedić, A. Olshevsky, and C. A. Uribe. Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs. *arXiv preprint arXiv:1410.1977*, 2014.
16. K. R. Rad and A. Tahbaz-Salehi. Distributed parameter estimation in networks. In *IEEE Conference on Decision and Control (CDC)*, pages 5050–5055. IEEE, 2010.
17. S. Shahrampour and A. Jadbabaie. Exponentially fast parameter estimation in networks using distributed dual averaging. In *IEEE Conference on Decision and Control (CDC)*, pages 6196–6201. IEEE, 2013.
18. S. Shahrampour, A. Rakhlin, and A. Jadbabaie. Distributed detection: Finite-time analysis and impact of network topology. *arXiv preprint arXiv:1409.8606*, 2014.
19. S. Shahrampour, A. Rakhlin, and A. Jadbabaie. Finite-time analysis of the distributed detection problem. *arXiv preprint arXiv:1512.09311*, 2015.
20. L. Tseng. Fault-tolerant Consensus and Shared Memory Consistency Model. Ph.D dissertation *University of Illinois at Urbana-Champaign*, 2015.
21. J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. In *Proceedings of the American Mathematical Society*, pages 733–737. JSTOR, 1963.