



Published in final edited form as:

J Am Stat Assoc. 2015 ; 110(509): 149–158. doi:10.1080/01621459.2014.893885.

Multi-Agent Inference in Social Networks: A Finite Population Learning Approach

Jianqing Fan,

Frederick L. Moore'18 Professor of Finance, Department of Operations Research and Finance Engineering, Princeton University, Princeton, NJ 08544 (jqfan@princeton.edu), and Adjunct Professor, School of International Economics and Management, Capital University of Economics and Business

Xin Tong [Assistant Professor], and

Marshall School of Business, University of Southern California, CA 90089 (xint@marshall.usc.edu)

Yao Zeng [Graduate Student]

Department of Economics, Harvard University, Cambridge, MA 02138 (yaozeng@fas.harvard.edu)

Abstract

When people in a society want to make inference about some parameter, each person may want to use data collected by other people. Information (data) exchange in social networks is usually costly, so to make reliable statistical decisions, people need to trade off the benefits and costs of information acquisition. Conflicts of interests and coordination problems will arise in the process. Classical statistics does not consider people's incentives and interactions in the data collection process. To address this imperfection, this work explores multi-agent Bayesian inference problems with a game theoretic social network model. Motivated by our interest in aggregate inference at the societal level, we propose a new concept, *finite population learning*, to address whether with high probability, a large fraction of people in a given finite population network can make “good” inference. Serving as a foundation, this concept enables us to study the long run trend of aggregate inference quality as population grows.

Keywords

Social networks; multi-agent inference; Bayesian learning; finite population learning; perfect learning; learning rates

1 Introduction

Statistical inference framework usually adopts a single-person perspective without taking costly information (data)¹ collection and interactions among people into account. Take the

Correspondence to: Jianqing Fan.

¹“Information”, “data”, and “signal” are used interchangeably in this paper.

Bayesian approach for example. Suppose an agent (person) is interested in a parameter θ . The Bayesian approach starts with a prior belief $\pi(\theta)$, which is a distribution on θ . Given the data $X_1, \dots, X_n \sim p(x|\theta)$, where $p(\cdot|\theta)$ is a conditional distribution given θ , the belief about θ is updated using the Bayes' Theorem,

$$\pi(\theta|X_1, \dots, X_n) = \frac{p(X_1, \dots, X_n|\theta)\pi(\theta)}{m(X_1, \dots, X_n)},$$

where $m(X_1, \dots, X_n) = \int p(X_1, \dots, X_n|\theta)\pi(\theta)d\theta$. Computation aside, such an update is clear once $\{X_1, \dots, X_n\}$ are given.

However, in a more realistic but complicated situation, the available data $\{X_1, \dots, X_n\}$ are scattered in the society, and agents' incentives for data collection are inter-dependent. Indeed, each agent has a choice between making inference based on her own data, and acquiring more data through costly exchange with others before making inference. A subtle point is that the amount of information an agent is willing to share depends on other agents' actions, and such interactions call for tools beyond mainstream statistics. Recall that in clinical trials, an agent weighs more accurate statistical inference against higher monetary cost of data acquisition. Similarly, in social networks, agents balance inference accuracy and time cost of data acquisition. Moreover, agents' inference in a network is inter-dependent in the sense that their decisions influence each other's data collection. Therefore, aggregate inference at the societal level is not a trivial problem that simply adds up each agent's inference. Our new multi-agent inference framework takes people's incentives and interactions in costly data acquisition into account, and rigorously characterizes the quality of inference at the societal level.

Multi-agent inference with endogenous data acquisition procedure makes two contributions to the Statistics field. First, it extends the "pipeline" of classical statistical inference, making agents' access to data an endogenous outcome of their incentives and interactions. From an individual agent's perspective, her inference results could be fundamentally different from those of classical statistical inference. Second, it provides a foundation to study aggregate inference at the societal level. More concretely, we address the question with a new *finite population learning* concept: can a large fraction of agents in a given finite population network make "good" inference about θ with high probability? Boosted by the internet and online social networks, this multi-agent perspective is particularly suited for understanding the informational efficiency of networks in the modern world.

Taking agents' incentives and interactions into account, this paper explores multi-agent Bayesian inference at the societal level using a game theoretic social network model. Our work differs from social network papers in existing statistical literature, in that the majority of those papers focus on graphical models (without considering agents' incentives), which are ideal to model network structures, but not individual agents' decisions. We refer interested readers to Kolaczyk (2009) and Newman (2010) for a general introduction to the existing literature. Our work supplements network structural modeling with an additional human incentive component using game theory. We demonstrate that communication mechanisms and agents' interactions, among other elements that lie outside of graphical

models are crucial to determine the aggregate inference quality at the societal level. We are inspired by the information exchange game in Acemoglu, Bimpikis and Ozdaglar (2012), which studies asymptotic learning and possible non-truthful communications when the network size (population in the society) goes to infinity. In contrast, our work focuses on networks of finite population, and predicts aggregate inference quality at the societal level. The finite population approach not only allows us to study the explicit interplay of parameters regarding aggregate inference results, but also provides a more solid benchmark to assess inference quality when network size grows.

The rest of the paper is organized as follows. In Section 2, we introduce the game theoretic social network model (Definition 1) and the associated equilibria (Definition 2), and illustrate the framework with examples. In Section 3, we construct the *finite population learning* criterion to capture the aggregate inference quality at the societal level (in an equilibrium), and derive necessary conditions and sufficient conditions for this criterion. These conditions involve only one equilibrium outcome—the number of signals each agent has acquired—in a clean and transparent manner. With the finite population learning criterion as a foundation, we study in Section 4 the long run trend of aggregate inference quality as population in a society grows. The trend will be captured by the *perfect learning* concept and the associated *learning rates*. Two conditions for perfect learning are proposed: one involves equilibrium, while the other (stronger condition) bypasses equilibrium, and only relies on information precision, model parameters, and network structures. Learning rates measure the quality of perfect learning, and are demonstrated in typical examples. Section 5 makes a few remarks and suggests future research lines. Proofs and discussions on related social learning literature are relegated to the Supplementary Materials.

2 The Model

We develop a game theoretic social network model, and introduce its components below. Examples and illustrations are also provided to facilitate understanding.

Agents and Network Structure

In a directed graph $G_n = (\mathcal{N}^n, \mathcal{E}^n)$, each node $i \in \mathcal{N}^n = \{1, 2, \dots, n\}$ represents an agent, and an ordered pair $(j, i) \in \mathcal{E}^n$ means agent j can send information to agent i directly (i.e., agent j is agent i 's in-degree neighbor, and agent i is agent j 's out-degree neighbor). When both $(j, i) \in \mathcal{E}^n$ and $(i, j) \in \mathcal{E}^n$, these two agents can communicate directly with each other.

Inference Problem and Information

Each agent would like to make inference about a parameter of common interest $\theta \in \mathbb{R}$ (the exact criterion will be introduced later in the section). Agents' common prior knowledge on θ is modeled by a normally distributed prior $\theta \sim \mathcal{N}(0, 1/\rho)$. At time $t = 0$, agent i is endowed with her own information (private signal) $s_i = \theta + z_i$, where $z_i \sim \mathcal{N}(0, 1/\rho)$ are independent with each other and also independent of θ .² Both ρ and $\bar{\rho}$ are assumed to be known parameters. The distributions of z_i 's are common knowledge to all agents and so is the

²Our results are not affected if θ and z_i have non-zero means.

network structure. Since the focus of our paper is not on technical issues related to Bayesian updates, we choose Gaussian distributions for simplicity.

Information Exchange

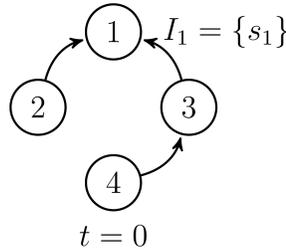
Agents can make inference based on their own signal, or they might also exchange signals with other agents before making inference. The information exchange process is described as follows. Suppose agents live in a world with continuous time $t \in [0, \infty)$. Waiting incurs a common exponential discount of the payoff with rate $r > 0$, i.e., $\exp(-rt)$. All agents communicate simultaneously at some time points following a homogeneous Poisson process with rate $\lambda > 0$, independent of θ and z_i . This Poisson process, common knowledge to all agents, defines some discrete communication rounds at which agents send off their private and acquired signals which are tagged with identities. After each communication round, agents update beliefs according to the Bayes' rule. For example, the posterior distribution of θ given k distinct signals is Gaussian with precision $\rho + k\rho$. So more acquired signals, i.e., a larger k , will increase the precision and lead to better inference results. Given the above, there is a natural trade-off between acting earlier to reduce the payoff discount, and waiting for more communication rounds to acquire more signals. This becomes an optimal stopping problem for each agent i : at any given time t , agent i either makes an estimate x_i of the parameter θ and “exits”, or “waits” for more signals.

By “exiting”, we mean that an agent no longer receives new signals, but continues to transmit signal(s) she has so far when new communication rounds take place. This assumption of “exiting” is important in capturing the agents' inter-dependent incentives in information acquisition. If an agent still acquires information after making inference, her neighbors do not have to take her decision into account. This assumption is also intuitive in reality: after an agent completes her decision, it makes no sense for her to further acquire costly information.³ In the following, we illustrate this information exchange scheme with an example.

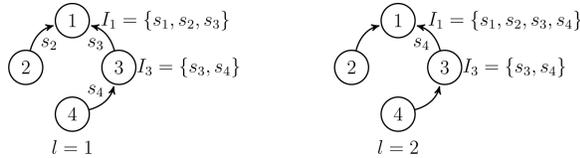
An Example of Information Exchange

Suppose there are four agents in a social network. At time $t = 0$, each agent i starts with her private signal s_i , and the total information endowment in the network is $\{s_1, s_2, s_3, s_4\}$. We only need to study communication rounds $l = 1, 2$, because the longest path in the network has length 2. We focus on the information set I_1 of agent 1.

³This assumption implies, however, even if an agent decides to “exit” at time $t = 0$, she is obliged to send her private signal to neighbors. In general, when she decides to “exit” at communication round j , she is still obliged to send in future rounds the signals she collected up to round j . This assumption encourages information exchange, and it would affect none of the results but Proposition 3.

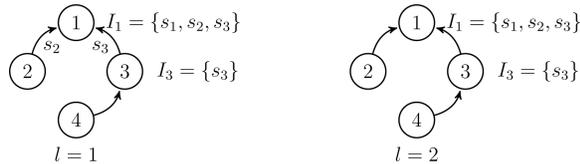


Two scenarios are studied. First, suppose no agent exits after time $t = 0$. The information flow is as follows:



After the first communication round, the information sets of agents 1 and 3 change to: $I_1 = \{s_1, s_2, s_3\}$, $I_3 = \{s_3, s_4\}$. After $l = 2$, I_1 is updated to $\{s_1, s_2, s_3, s_4\}$, while I_2 , I_3 and I_4 are unchanged.

In the second scenario, suppose agent 3 exits after time $t = 0$. Although she is still obliged to send all her signals (in this case, only her private signal s_3) to out-degree neighbors, she will not collect signals from in-degree neighbors. The information flow is as follows.



As agent 3 does not receive signals from agent 4 at $l = 1$, she does not have new signals to send to agent 1 at $l = 2$. Therefore, $I_1 = \{s_1, s_2, s_3\}$ at $l = 2$.

Comparing the two scenarios, individual agents' decisions can affect others' information acquisition, thus inference. This interdependence differentiates aggregate inference at the societal level from the trivial sum of individual inference, and motivates a game theoretical framework.

Payoff Function and Inference Decisions

We next specify the payoff function and agents' decision problems. Denote by $I_{i,t}^n$ the information set of agent i at time t . Suppose agent i estimates θ to be x_i at time t , her instantaneous payoff is defined as

$$u_i(x_i) = \psi - (x_i - \theta)^2,$$

where ψ is a real-valued constant. Note that the larger the ψ , the less sensitively the payoff depends on the squared error $(x_i - \theta)^2$, where x_i further depends on the agent's final

information set. Hence, in what follows we call ψ the information sensitiveness. While ψ plays no role in classical statistical inference, it is important in individual agent's decision making due to the exponential discount of the payoff in time. Agent i 's optimal expected instantaneous payoff with estimate x_i given information set $I_{i,t}^n$ (without considering discounting) is

$$U_{i,t}^n(I_{i,t}^n) = \max_{x_i} \mathbb{E}(u_i(x_i) | I_{i,t}^n).$$

It is easy to see that agent i 's optimal estimate is $x_{i,t}^{n,*} = \mathbb{E}[\theta | I_{i,t}^n]$ if she decides to exit at time t . Thanks to the normality assumption on θ and signals $\{s_i\}_{i=1}^n$, agent i 's optimal expected instantaneous payoff based on k signals can be calculated explicitly:

$$\mathbb{E}[\psi - (x_{i,t}^{n,*} - \theta)^2 | I_{i,t}^n] = \psi - \frac{1}{\rho + \bar{\rho}k}. \quad (2.1)$$

At any time t with information set $I_{i,t}^n$, agent i has to make a decision about whether to wait or to make an optimal estimate and exit. As a result, due to the common exponential discount in time, each agent should exit at some finite time point,⁴ and more precisely, right after a communication round.⁵ Therefore, to assess agents' inference, we only need to consider how many communication rounds they participated in before exiting. This simplification allows us to study the following game theoretic social network model, formally called a network game.

The Network Game

We specify the network game of information exchange, which captures agents' incentives and interactions in acquiring information. Denote l_i^n as agent i 's number of communication rounds before exiting, $i = 1, \dots, n$; $l^n = (l_1^n, \dots, l_n^n)$. Also denote by l_{-i}^n the vector l^n without the component l_i^n . Let τ_k be the time by which the k^{th} communication round occurs. Agent i 's payoff for choosing l_i^n , i.e., making an optimal estimate and exiting after l_i^n communication rounds, is

$$U_i^n(l_i^n, l_{-i}^n) = \mathbb{E} \left\{ e^{-r\tau_{l_i^n}} \max_{x_i} \mathbb{E}[\psi - (x_i - \theta)^2 | I_i^n(l^n)] \right\},$$

where $I_i^n(l^n)$ is agent i 's information set upon exiting, which not only depends on her own l_i^n , but also depends on other agents' l_{-i}^n , as we have seen in the previous four-agent example. By (2.1) and the exponential waiting time of the Poisson process, it is easy to show that

⁴This is because, waiting towards infinity would incur a zero payoff due to discounting.

⁵Suppose an agent exits at a time between two communication rounds. Because she does not get new signals between two adjacent communication rounds, it is always better to exit right after the earlier communication round, due to continuous time discounting.

$$U_i^n(l_i^n, l_{-i}^n) = \bar{r}^{l_i^n} \left(\psi - \frac{1}{\rho + \bar{\rho} k_i^{n, l_i^n}} \right),$$

where $\bar{r} = \lambda / (\lambda + r)$ and k_i^{n, l_i^n} is the number of signals agent i has upon exiting if agents in the network adopt \bar{r} . With this simplification, we consider the following game.

Definition 1—The network game $\Gamma(G_n)$ is a triplet $\{\mathcal{N}^n, \mathcal{L}^n, \mathcal{U}^n\}$, in which

- a. \mathcal{N}^n is the set of agents, i.e., $\mathcal{N}^n = \{1, 2, \dots, n\}$;
- b. $\mathcal{L}^n = (L_1^n, \dots, L_n^n)$ is the collection of agents' strategy spaces. For every agent $i \in \mathcal{N}^n$, her strategy space L_i^n is a finite set

$$L_i^n = \{0, 1, 2, \dots, (l_i^n)_{\max}\},$$

where $(l_i^n)_{\max} = \max_{j \in \mathcal{N}^n \setminus \{i\}} \{\text{length of shortest path from } j \text{ to } i\}$;

- c. $U_i^n \in \mathcal{U}^n$ is the payoff function for agent i :

$$U_i^n(l_i^n, l_{-i}^n) = \bar{r}^{l_i^n} \left(\psi - \frac{1}{\rho + \bar{\rho} k_i^{n, l_i^n}} \right) \quad (2.2)$$

⁶ For brevity, we use G_n to refer to both the network and the associated network game $\Gamma(G_n)$ when there is no confusion.

We restrict ourselves to pure-strategy (no randomization) Nash equilibria of this game, which is defined below for the readers' convenience.

Definition 2—In the network game $\Gamma(G_n) = \{\mathcal{N}^n, \mathcal{L}^n, \mathcal{U}^n\}$, a pure-strategy Nash equilibrium⁷ $\sigma^{n,*}$ is a vector $l^{n,*} = (l_1^{n,*}, \dots, l_n^{n,*}) \in \mathcal{L}^n$ such that for every $i \in \mathcal{N}^n$,

$$U_i^n(l_i^{n,*}, l_{-i}^{n,*}) \geq U_i^n(l_i^n, l_{-i}^{n,*}), \text{ for every } l_i^n \in L_i^n.$$

In other words, agents' strategies are a *pure-strategy Nash equilibrium* (or *equilibrium* for brevity) if for every agent, her strategy is optimal given other players' strategies. Our network game is closely related to that in Acemoglu, Bimpikis and Ozdaglar (2012), which used a more complicated model to accommodate possible non-truthful communications and the resulting asymptotic learning behavior. Their paper does not formally establish the existence or computability of an equilibrium. On the contrary, our goal is to study aggregate inference in a finite population network. We develop a simpler model to capture the

⁶If the signal precisions ρ and $\bar{\rho}$ are not known, we cannot simply replace the precisions by their estimates, as the estimates depend not only the number of signals, but also on the signals themselves. Generalization to the unknown precision cases would be interesting for further research.

⁷We refer interested readers to Fudenberg and Tirole (1991) that gives a textbook treatment of the equilibrium and the general proofs for its existence.

interdependence of agents' inference. Our model allows us to establish the existence of an equilibrium, which is also computationally feasible (Mckelvey and McLennan, 1996). The following lemma ensures the existence of equilibria in our network game.

Lemma 1—*The network game $\Gamma(G_n)$ has at least one pure-strategy Nash equilibrium.*

We offer some intuition for this lemma here (its formal proof is in the Supplementary Materials). When some agents stay longer, all other agents have (weakly) larger incentives to stay longer, because the amount of information that comes in the future (weakly) increases. This property is formally called “strategic complementarity” (Fudenberg and Tirole, 1991), and it essentially guarantees the existence of equilibria. This also plays an important role in making the equilibria easily computable (Mckelvey and McLennan, 1996).

An Example of the Network Game and Its Equilibrium

We offer an example to illustrate how agents' incentives and interactions affect their decisions of information acquisition and inference. In the four-agent network displayed previously, fix $r = 0.9$, $\psi = 1$ and $\rho = \bar{\rho} = 0.5$. Agents 2 and 4 should exit at $t = 0$, because they will not get any new signals due to the network structure, while incurring discounting penalty should they not act promptly. The payoff matrix for agent 1 (rows) and 3 (columns) is as follows, in which the first and the second value in each cell are respectively the payoffs of agent 1 and agent 3 [see (2.2)].

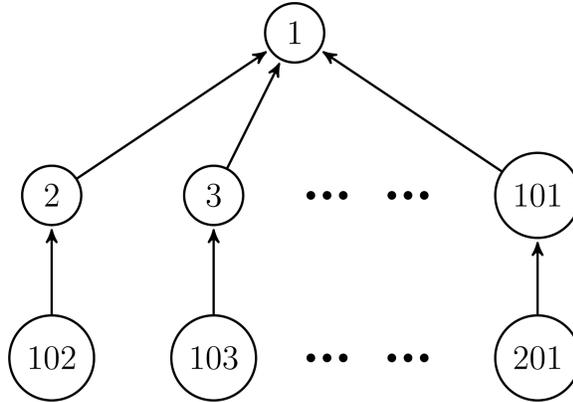
		Agent 3	
		Round 0	Round 1
Agent 1	Round 0	0, 0	0, 0.3
	Round 1	0.45, 0	0.45, 0.3
	Round 2	0.405, 0	0.486, 0.3

There is only one equilibrium of the game. In this equilibrium, agents 2 and 4 exit immediately after they receive their own private signals in Round 0, while agent 3 exits after Round 1, and agent 1 exits after Round 2. Such an outcome is hard to rationalize with any single-agent decision models. No matter what other agents choose, agent 3 always prefers Round 1 to Round 0 ($0.3 > 0$). But without taking into account agent 3's strategy, there is no clear best choice for agent 1: her best strategy is Round 1 should agent 3 choose Round 0, while her best strategy is Round 2 should agent 3 choose Round 1. Agent 1 is willing to wait longer in the equilibrium only because she believes that agent 3 would wait longer, and agent 3 will indeed do so.

Another Example of Interactions

In the equilibrium of the above example, all agents wait until their maximum rounds $(l_i^n)_{\max}$. There would not be much difference if we were to assume no exits (interactions) to start with. To highlight the effect of interactions, we provide another example. In the following network, 201 agents are organized in three layers. Agent 1 has 100 in-degrees, and agents 2

to 101 each has one in-degree. Fix $r = 0.95$, $\psi = 10$, $\rho = 0.2$ and $\bar{\rho} = 0.003$. From (2.2) one can check that there is only one equilibrium, in which agent 1 exits after Round 1, while all other agents exit immediately at Round 0. In this equilibrium, agent 1 has 101 signals upon existing. However, if agents 2 to 101 were forced not to exit, they would get the signals from their in-degree neighbors at Round 1, and this encourages agent 1 to wait until Round 2 and has 201 signals upon exiting. The latter hypothetical scenario would not happen in an equilibrium since agents 2 to 101 want to exit earlier. Note that in addition to the network structure, the game parameters also matter for the agents' inference, since they help determine agents' incentives and interactions in acquiring information.



3 Finite Population Learning

In this section, we introduce *finite population learning* to assess aggregate inference quality at the societal level. This criterion answers whether with high probability, a large fraction of agents in the network can make “good” inference in equilibrium. We derive necessary conditions and sufficient conditions for this criterion, and discuss their implications. The finite population learning concept and its determination conditions echo the spirit of finite sample approach in the statistical learning literature, and they help reveal the explicit interplay among the population size, parameters of the network game and the learning tolerances. This methodological point distinguishes our paper from previous works in the social networks literature that focuses on the asymptotic effects of information aggregation (Acemoglu, Dahleh, Lobel and Ozdaglar, 2011; Acemoglu, Bimpikis and Ozdaglar, 2012).

Definition 3

Given a social network G_n with an equilibrium $\sigma^{n,*}$, we say G_n achieves $(\varepsilon, \bar{\varepsilon}, \delta)$ -learning under $\sigma^{n,*}$ if

$$\mathbb{P}_{\sigma^{n,*}} \left(\frac{1}{n} \sum_{i=1}^n (1 - M_i^{n,\varepsilon}) \geq \bar{\varepsilon} \right) \leq \delta,$$

in which $M_i^{n,\varepsilon} = \mathbf{1}(|x_i^* - \theta| \leq \varepsilon)$ where x_i^* is agent i 's optimal estimate upon exiting, and $\mathbb{P}_{\sigma^{n,*}}$ denotes the conditional probability given $\sigma^{n,*}$.

In this definition, the parameter ε defines a “good” estimate for individual agents, $1 - \varepsilon$ represents the fraction of agents who make such good estimates, and $1 - \delta$ represents the probability at which such a high fraction of agents make such good estimates. $(\varepsilon, \bar{\varepsilon}, \delta)$ -learning reflects a certain high quality of aggregate inference at the societal level, and these tolerance parameters can be tuned to different applications. Since our focus is on finite population networks, in verbal discussions and when there is no confusion, we call $(\varepsilon, \bar{\varepsilon}, \delta)$ -learning *finite population learning*.

A natural question to ask is whether such finite population learning occurs in a given network. If so, under what conditions? The following proposition provides a necessary

condition and a sufficient condition. Denote by $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ the error function of the standard normal distribution.

Proposition 1

For a given social network G_n under an equilibrium $\sigma^* (= \sigma^{n,*})$

a. $(\varepsilon, \bar{\varepsilon}, \delta)$ -learning does not occur if

$$\frac{1}{n} \sum_{i=1}^n \text{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho} k_i^{n, \sigma^*}}{2}} \right) < (1 - \bar{\varepsilon})(1 - \delta). \quad (3.1)$$

b. $(\varepsilon, \bar{\varepsilon}, \delta)$ -learning occurs if

$$\frac{1}{n} \sum_{i=1}^n \text{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho} k_i^{n, \sigma^*}}{2}} \right) \geq 1 - \bar{\varepsilon} \delta. \quad (3.2)$$

This proposition provides simple conditions for the occurrence of finite population learning: only one equilibrium outcome k_i^{n, σ^*} is involved. Hence, these conditions are more operative and transparent than their asymptotic counterparts in previous literature. The involvement of the equilibrium outcome k_i^{n, σ^*} also suggests that, our finite population learning criterion encodes not only graphical network structures but also agents’ incentives and interactions in information acquisition.

Conditions (3.1) and (3.2) also allow us to untangle the interplay among parameters. For example, we are able to answer the following question. Given the tolerances $\varepsilon, \bar{\varepsilon}, \delta$ and the signal precisions ρ and $\bar{\rho}$, how does the change in k_i^{n, σ^*} affect the occurrence of finite population learning in a given social network G_n ? When k_i^{n, σ^*} s are sufficiently small to validate condition (3.1), finite population learning does not occur. Similarly, when some of k_i^{n, σ^*} s are sufficiently large so that the condition (3.2) is satisfied, finite population learning occurs. Similar interpretations about unilateral changes also apply to parameters $\varepsilon, \bar{\varepsilon}, \delta, \rho$ and $\bar{\rho}$. As interplays among the parameters $\varepsilon, \bar{\varepsilon}, \delta, \rho, \bar{\rho}$ and k_i^{n, σ^*} are clear through (3.1) and (3.2), the two conditions help us better understand the quality of inference in different

circumstances. Also, a beauty of symmetry arises in our necessary conditions and sufficient conditions for finite population learning. The parameters $\bar{\varepsilon}$ and $\bar{\delta}$ are completely interchangeable in these conditions, which was not expected as they capture tolerances in different categories. On the other hand, in these two conditions, the parameter ε stands in a position unchangeable with $\bar{\varepsilon}$ or $\bar{\delta}$, which hints that ε plays a different role in determining the aggregate inference quality.

Conditions (3.1) and (3.2) have powerful implications. For example, the next corollary establishes a necessary condition and a sufficient condition without equilibrium outcomes.

Corollary 1

For a given social network G_n ,

a. $(\varepsilon, \bar{\varepsilon}, \bar{\delta})$ -learning does not occur if

$$\operatorname{erf}\left(\varepsilon\sqrt{\frac{\rho+\bar{\rho}n}{2}}\right) < (1-\bar{\varepsilon})(1-\bar{\delta}). \quad (3.3)$$

b. $(\varepsilon, \bar{\varepsilon}, \bar{\delta})$ -learning occurs if

$$\operatorname{erf}\left(\varepsilon\sqrt{\frac{\rho+\bar{\rho}}{2}}\right) \geq 1-\bar{\varepsilon}\bar{\delta}. \quad (3.4)$$

Corollary 1 follows from $1 \leq k_i^{n, \sigma^*} \leq n$. It is interesting because under some circumstances, we can determine an aggregate inference status without knowing either the structure of the social network or the equilibrium. Intuitively, if any tolerance parameter, information precision or the population size is too low, such that the condition (3.3) is satisfied, finite population learning does not occur no matter how effective the network is organized. Conversely, if any tolerance or information precision is sufficiently large such that condition (3.4) holds, finite population learning occurs even if all agents are isolated.

Finally, multiple equilibria emerge under some circumstances. In the following, we introduce a generalized (conservative) version of finite population learning to accommodate multiple equilibria without equilibrium selection.

Definition 4

Denote by $\Sigma^{n,*} = \{\sigma^*\}$ the set of equilibria of $\Gamma(G_n)$. $(\varepsilon, \bar{\varepsilon}, \bar{\delta})$ -learning occurs if

$$\sup_{\sigma^* \in \Sigma^{n,*}} \mathbb{P}_{\sigma^*} \left(\frac{1}{n} \sum_{i=1}^n (1 - M_i^{n,\varepsilon}) \geq \bar{\varepsilon} \right) \leq \bar{\delta}.$$

This definition offers a conservative standard in the sense that the least favorable equilibrium determines the aggregate inference status. When $\Sigma^{n,*}$ is a singleton, the above definition reduces to Definition 3. The proof of Proposition 1 can be recycled to deliver the next remark.

Remark 1

For a given social network G_n ,

a. $(\varepsilon, \bar{\varepsilon}, \bar{\delta})$ -learning does not occur if

$$\min_{\sigma^* \in \Sigma^{n,*}} \frac{1}{n} \sum_{i=1}^n \text{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho} k_i^{n, \sigma^*}}{2}} \right) < (1 - \bar{\varepsilon})(1 - \delta).$$

b. $(\varepsilon, \bar{\varepsilon}, \bar{\delta})$ -learning occurs if

$$\min_{\sigma^* \in \Sigma^{n,*}} \frac{1}{n} \sum_{i=1}^n \text{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho} k_i^{n, \sigma^*}}{2}} \right) \geq 1 - \bar{\varepsilon} \delta.$$

4 Perfect Learning and the Rates

The finite population learning concept and Proposition 1 provide a solid foundation to investigate the aggregate inference quality as population in a network grows. This problem is important because the evolution of a society can impact the organization of information, the interdependence of individual inference, and the aggregate inference at the societal level. A new concept, *perfect learning*, determines whether in a sequence of growing networks $\{G_n\}_{n=1}^\infty$ (population goes to infinity), $(\varepsilon, \bar{\varepsilon}, \bar{\delta}_n)$ -learning can be achieved in G_n for all $n \in \mathbb{N}$ with $\bar{\delta}_n \rightarrow 0$. This means with increasing population, a large fraction of the people in a society can make “good” inference almost surely. The sequence $\{\bar{\delta}_n\}_{n=1}^\infty$ naturally induces a *learning rate*, which measures the speed towards perfect learning. This perfect learning concept and its associated learning rates also apply to the tolerance parameters ε and $\bar{\varepsilon}$. Our approach is different from previous social networks literature on learning that follows a direct asymptotic approach (Acemoglu, Dahleh, Lobel and Ozdaglar, 2011; Acemoglu, Bimpikis and Ozdaglar, 2012) or discusses other notions of learning rate under non-Bayesian inference context (Golub and Jackson, 2012a,b,c; Jadbabaie, Molavi and Tahbaz-Salehi, 2013).

4.1 Perfect Learning

Recall that the three tolerance parameters ε , $\bar{\varepsilon}$, and $\bar{\delta}$ tune the learning status of a society. We call by *an evolution path* a sequence of growing networks $\{G_n\}_{n=1}^\infty$ (population goes to infinity), where existing links are kept when networks grow. To investigate the limiting behavior under an evolution path, we can focus on one parameter at a time. The following definition introduces δ -perfect learning on a given evolution path $\{G_n\}_{n=1}^\infty$.

Definition 5— We say δ -perfect learning occurs on an evolution path $\{G_n\}_{n=1}^\infty$ under equilibria $\{\sigma^{n,*}\}_{n=1}^\infty$ if there exists a vanishing positive sequence $\{\bar{\delta}_n\}_{n=1}^\infty$ such that $(\varepsilon, \bar{\varepsilon}, \bar{\delta}_n)$ -learning occurs in G_n under its associated $\sigma^{n,*}$ for all n .

⁸Since we assume that the graph is evolved such that nodes/edges can only be added, for any agent i , the number of signals she gets increases in n . Concretely, if in a “smaller” graph i gets k_i signals in some equilibrium $\sigma^{n,*}$, then one can define a corresponding

This definition conveys the idea that as a society becomes larger, eventually we know for sure that a pre-specified large fraction of the people can make “good” inference. In verbal discussions, we call δ -perfect learning just *perfect learning*. Compared to its counterpart in previous works, our definition of perfect learning is both stronger and more general for the following reasons. First, we require networks on an evolution path to achieve a certain quality of inference not only in the limit but also all along the path towards the limit. Second, by focusing on different parameters ε , $\bar{\varepsilon}$ and δ , we could potentially address three different types of limiting behavior. As discussed in the previous section, these three parameters exhibit different impacts on finite population learning, so they play different roles in perfect learning as well. Third, this definition allows us to investigate learning rates (in the next subsection).

In the following, we derive two sufficient conditions for δ -perfect learning as Proposition 2 and Proposition 3, both taking Proposition 1 as a foundation. The first condition relies on the equilibrium outcome k_i^{n,σ^*} . The second condition relies only on the evolution path. To deliver the first sufficient condition, we define an *equilibrium informed agent*.

Definition 6 (Equilibrium Informed Agent)—*For agent i on a given evolution path $\{G_n\}_{n=1}^\infty$, she is equilibrium informed with respect to $\{G_n\}_{n=1}^\infty$ under equilibria $\{\sigma^{n,*}\}_{n=1}^\infty$ if*

$$\lim_{n \rightarrow \infty} k_i^{n,\sigma^*} = \infty.$$

Intuitively, an agent is equilibrium informed means that she enjoys increasing information advantage as population grows. The next proposition offers a sufficient condition for δ -perfect learning. In a similar spirit, a more general sufficient condition is derived as Lemma 3 in the Supplementary Materials. The proof of Proposition 2 is omitted as it is a corollary to Lemma 3.

Proposition 2— *δ -perfect learning occurs on an evolution path $\{G_n\}_{n=1}^\infty$ under equilibria $\{\sigma^{n,*}\}_{n=1}^\infty$ if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} |EI^{n,*}| = 1,$$

where $EI^{n,}$ is the set of equilibrium informed agents in the network G_n with respect to $\{G_n\}_{n=1}^\infty$ under $\{\sigma^{n,*}\}_{n=1}^\infty$.*

Proposition 2 states that perfect learning occurs when almost all agents are equilibrium informed. This reflects that most individuals should collect sufficient information for the society to achieve a high level of aggregate inference. We have such a transparent condition

equilibrium $\sigma^{t',*}$ ($t' > n$), in which no agents exit earlier than in $\sigma^{n,*}$ and hence i gets (weakly) more signals. As an alternative, we can restrict the perfect learning definition and associated learning rates on such equilibria sequences.

because our perfect learning concept is powered by finite population learning, a sufficient condition of which only involves one set of equilibrium variables: $\{k_i^{n,\sigma^*}\}$.

Next we consider the second sufficient condition that relies only on an evolution path. To streamline the presentation in the main texts, we assume that each agent enjoys a positive payoff even if she exits at $t=0$. From (2.1), this is equivalent to the following assumption, which will be held for the rest of this section. In the Supplementary Materials, we relax this assumption and show that none of the following results is affected.

Assumption 1— $(\rho + \bar{\rho})\bar{\psi} > 1$.

Before looking into the next sufficient condition for perfect learning, we point out an important observation. Although the number of signals an agent gets in equilibrium may diverge to infinity on an evolution path, the number of communication rounds she chooses will not increase unboundedly.

Lemma 2—*Under Assumption 1, given a network G_n under equilibrium σ^* , agent i 's optimal communication round before exiting is bounded from above by a constant independent of n . Mathematically,*

$$l_i^{n,\sigma^*} \leq l_i^n < \ln \left[1 - \frac{1}{(\rho + \bar{\rho})\bar{\psi}} \right] / \ln \bar{r}, \quad (4.1)$$

in which l_i^n stands for agent i 's optimal communication round given that other agents wait until their maximum rounds $(l_j^n)_{\max, j \neq i}$.

From condition (4.1), we see that the upper bound is exclusively determined by parameters of the network game. A more general version (Lemma 4) of Lemma 2 is in the Supplementary Materials. A key idea behind this lemma is that an agent's attainable payoff is bounded from above by $\bar{\psi}$, while waiting incurs a discounting of the payoff towards zero. Hence, when an agent gets sufficiently large number of signals within some finite communication rounds, even expecting infinite number of signals does not justify the discount of further waiting.

Lemma 2 plays an important role in shaping our next sufficient condition that directly links perfect learning status to formation of an evolution path. Recall that Proposition 2 states almost all agents' $k_i^{n,\sigma^*} \rightarrow \infty$ is sufficient for perfect learning. On the other hand, from Lemma 2 we know that no agent has an optimal unbounded communication round l_i^{n,σ^*} . Combining the two observations, the only possibility to validate the conditions in Proposition 2 is that almost all agents get unbounded number of signals within finite communication rounds. This consideration leads to the following definition of a *socially informed agent*.

Definition 7 (Socially Informed Agent)—*For each agent i on a given evolution path $\{G_n\}_{n=1}^\infty$, let $L_i = \min\{l_0 \in \mathbb{N} : \lim_{n \rightarrow \infty} |B_{i,l_0}^n| < \infty\}$, where $B_{i,l}^n$ is the set of agents in G_n whose*

shortest path to i has length at most l . Agent i is socially informed with respect to $\{G_n\}_{n=1}^\infty$ if L_i is finite, and if there exists $N \in \mathbb{N}$ such that for $n \geq N$, we have

$$\psi - \frac{1}{\rho + \bar{\rho}|B_{i,L_i}^n|} > 0, \quad (4.2)$$

and

$$\bar{r}^{L_i} \left(\psi - \frac{1}{\rho + \bar{\rho}|B_{i,L_i}^n|} \right) > \bar{r}^l \left(\psi - \frac{1}{\rho + \bar{\rho}|B_{i,l}^n|} \right) \text{ for all } 0 \leq l < L_i. \quad (4.3)$$

Moreover, we denote by SI^n the set of socially informed agents in the network G_n .

The definition of a socially informed agent does not require knowledge of any specific equilibrium. It only depends on the topological structure of the graph and on the parameters of the network game. In Definition 7, condition (4.2) is automatically satisfied in view of Assumption 1. Intuitively, a socially informed agent can be reached by a large number of neighbors after some finite communication rounds L_i . Furthermore, condition (4.3) ensures that this agent strictly prefers to wait at least until she collects all signals from in-degree neighbors up to distance L_i , given other agents do not exit. With the help of socially informed agents, we bypass equilibrium and state the following sufficient condition for perfect learning.

Proposition 3— δ -perfect learning occurs on an evolution path $\{G_n\}_{n=1}^\infty$ (under any equilibria $\{\sigma^{n,*}\}_{n=1}^\infty$) if

$$\lim_{n \rightarrow \infty} \frac{1}{n} |SI^n| = 1.$$

Proposition 3 is interesting because an agent's socially informed status only depends on network structures and model parameters alone. Given the difficulty to get closed form solutions for equilibria in general scenarios, Proposition 3 is of more value.

One important intuition behind this result is that agent i 's socially informed status guarantees that from a certain point on an evolution path, there is a "hub" within her finite distance who can and will collect unbounded number of signals at the first communication round. Conditions (4.2) ensures that these hubs are willing to collect their immediate indegree neighbors' private signals regardless these neighbors' strategies, enabling us to bypass equilibrium. Moreover, conditions (4.2) and (4.3) also ensure a chain of agents willing to pass hubs' gathered signals to other agents to make them "socially informed". These conditions encode information on both network structures and agents' incentives and interactions. Even for an agent with infinite in-degrees, if she finds it optimal to exit before the first communication round with only her own signal, she cannot serve as a hub and thus does not help establish perfect learning. This suggests that the graphbased network models in classical statistics are not sufficient to fully understand the interdependence in multi-agent inference problems.

Hubs exist in important networks studied in literature. For example, a good representation of many real-world scenarios in politics and sociology is the island connection networks (Jackson, 2010; Easley and Kleinberg, 2010), which consist of nearly isolated subgraphs, but each subgraph is a nearly (two-way) complete graph.⁹ This may also represent more general social cliques or homophily as discussed in Golub and Jackson (2012a,b,c). Another typical class of networks with hubs is the (two-way) star networks, in which the star (hub) collects signals from all others and then send all signals together to every agent. In view of Proposition 3, a sequence of such networks with hubs may achieve perfect learning under any equilibria. On the negative side, the preferential attachment graphs, and Cayley trees where all nodes have k degrees, do not have such hubs.

4.2 Learning Rates

In this subsection, we define learning rates for δ -perfect learning, and derive them for some typical network classes. Again, finite population learning serves as a foundation.

Definition 8—*If δ -perfect learning occurs on an evolution path $\{G_n\}_{n=1}^\infty$ under equilibria $\{\sigma^{n,*}\}_{n=1}^\infty$, we call the corresponding sequence of tolerances $\{\delta_n\}_{n=1}^\infty$ a learning rate.*

Proposition 1 suggests a conservative approach to construct a learning rate. The sufficient condition part of this proposition prescribes that $(\varepsilon, \bar{\varepsilon}, \delta_n)$ -learning occurs if

$$\frac{1}{n} \sum_{i=1}^n \operatorname{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho} k_i^{n, \sigma^*}}{2}} \right) \geq 1 - \delta_n \bar{\varepsilon}.$$

Then we can solve the above inequalities with respect to δ_n , getting a sequence of lower bounds. This sequence, if converging to zero (δ -perfect learning achieved), can serve as a learning rate. However, without specific knowledge of network structures, it is hard in general to solve for k_i^{n, σ^*} in terms of n and other parameters. In the following, we will derive learning rates on some examples, and discuss general network classes when possible.

Example 1 (Isolated Agents)—*When all agents are isolated from each other in a network G_n , we have $k_i^{n, \sigma^*} = 1$ for every agent i .*

In Example 1, the necessary condition (3.1) is reduced to

$$\operatorname{erf} \left(\varepsilon \sqrt{\frac{\rho + \bar{\rho}}{2}} \right) < (1 - \bar{\varepsilon})(1 - \delta_n).$$

⁹On an evolution path, the populations of all but perhaps a finite number of subgraphs go to infinity. Real-world examples of island connection networks include the US Congress, in which the two major parties are well connected within themselves but there are only a few links between the two, and corporate email communication networks, in which internal email lists construct nearly complete graphs within individual companies, and a few connections are between companies. See Jackson (2010); Easley and Kleinberg (2010) for more examples.

If parameters are such that $\text{erf}\left(\varepsilon\sqrt{\frac{\rho+\bar{\rho}}{2}}\right) < (1-\bar{\varepsilon})$, the above inequality holds for large n for any vanishing sequence $\{\delta_n\}_{n=1}^\infty$. This tells us that in fairly general circumstances, an isolated evolution path cannot achieve δ -perfect learning.

Example 2 (Complete Graph)—When the network G_n is a (two-way) complete graph, and the benefit of getting $n-1$ new signals justifies the discount of one communication round, $k_i^{n,\sigma^*} = n$ for every agent i .

In Example 2,

$$\text{erf}\left(\varepsilon\sqrt{\frac{\rho+\bar{\rho}n}{2}}\right) \geq 1 - \delta_n\bar{\varepsilon}, \quad \forall n \in \mathbb{N},$$

is a sufficient condition for δ -perfect learning, which translates to

$$\delta_n \geq \frac{1}{\bar{\varepsilon}} \left(1 - \text{erf}\left(\varepsilon\sqrt{\frac{\rho+\bar{\rho}n}{2}}\right)\right). \quad (4.4)$$

The sequence of the right hand sides of inequality (4.4) can serve as a learning rate. We approximate the error function to get a conservative but more transparent rate. Note that the error function erf can be approximated by

$$1 - \text{erf}(x) < \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

Therefore a sufficient condition for δ -perfect learning is

$$\delta_n \geq \frac{1}{\sqrt{\pi}\bar{\varepsilon}} \frac{1}{\varepsilon\sqrt{\rho+\bar{\rho}n}} \exp\left(-\frac{\varepsilon^2(\rho+\bar{\rho}n)}{4}\right).$$

Keep other parameters fixed, and focus on the relation between population size n and δ_n . We see that δ_n could decrease in the order of $\exp(-\rho\varepsilon^2n/5)$. This implies that when population grows, the probability that at least ε fraction of people make a “bad” inference decreases very quickly to zero.

Following the idea of error function approximations, we go beyond Example 2 to consider a more general case in which $k_i^{n,\sigma^*} \geq f(n)$ for every agent i where $f(n)$ is a deterministic sequence. A sufficient condition for δ -perfect learning is then

$$\delta_n \geq \frac{1}{\sqrt{\pi}\bar{\varepsilon}} \frac{1}{\varepsilon\sqrt{\rho+\bar{\rho}f(n)}} \exp\left(-\frac{\varepsilon^2(\rho+\bar{\rho}f(n))}{4}\right). \quad (4.5)$$

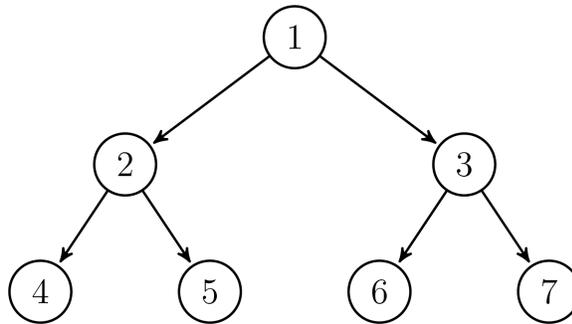
If $f(n)$ diverges to infinity as n goes to infinity, the right hand side of inequality (4.5) converges to 0. Keeping other parameters fixed, this implies δ_n could decrease in the order of $\exp(-\rho \varepsilon^2 f(n)/5)$. Formally, we have the next proposition.

Proposition 4—*Suppose there exists a diverging sequence $f(n)$ such that $k_i^{n,\sigma^*} \geq f(n)$ for every agent i in network G_n with associated equilibrium $\sigma^{n,*}$, then δ -perfect learning could occur with learning rate $\{\delta_n\}_{n=1}^\infty$, where each δ_n is in the order of $\exp(-\rho \varepsilon^2 f(n)/5)$.*

An interpretation of this proposition is that, even if each agent can only get a small proportion of information scattered in the network, perfect learning can still be reached at a fast rate. This speaks to the commonly observed class of island connection networks discussed above, and is also related to interesting results pertaining to social cliques or homophily as discussed in Golub and Jackson (2012a,b,c). Lemma 3 in the Supplementary Materials further renders Proposition 4 as a special case and provides more general implications on learning rates.

Next, we consider the binomial trees, which are an axiomatic representation of various hierarchical social structures (Jackson, 2010). In particular, as the information flow within a binomial tree can be either from the root to the leafs or from the leafs to the root, binomial trees can accommodate both the top-down and the bottom-up cases of information exchange in various real-world scenarios. Hence, it is instructive to analyze the binomial tree with a few different settings, where we generalize our network game by allowing the information sensitiveness $\psi = \psi_n$ to vary along an evolution path $\{G_n\}_{n=1}^\infty$.

Example 3 (Binomial Tree: Information Flow from Root to Leafs)—*The agents in the communication network G_n form a binomial tree, where information can only flow from root to leafs. For simplicity, consider only the number of agents n such that $n = 1 + 2 + 4 + \dots + 2^{(m-1)}$, where m_n is the number of layers in the binomial tree. The following graph illustrates such a binomial tree with three layers.*



We will study two scenarios of this binomial tree, in both of which $\lambda = r$ so that $\bar{r} = 1/2$.

- i) $\psi_n = \rho = \bar{\rho} = 1$. For agent 1 on the top layer, he should exit right after round 0 because he does not have any chance to receive others' private information. For agent 2 and 3, who are on the second top layer, they decide between round 0 and 1. A simple calculation on their pay off functions reveals that they should exit

after round 0. Agents 4, 5, 6, 7 who are on the third layer potentially should decide between 0,1 and 2 rounds. But since agents 2 and 3 cannot not pass through agent 1's info, round 2 is eliminated before any calculation. So agents on the third layer actually faces same decision problems as agents on the second layer. Continue with the same argument till the m_n 'th layer, we learn that everyone in the network exits right after she gets her private signal. Therefore, this scenario is the same as *isolated agents* in terms of information exchange.

In general, as depicted in this subcase i), when the game is less information sensitive (i.e., ψ higher), the precision of the prior ρ is higher, or $\bar{\rho}$ is lower, it is less likely to achieve δ -perfect learning, even if the agents are well connected.

ii) $\psi_n < \frac{2}{\rho+(m_n-1)\bar{\rho}} - \frac{1}{\rho+m_n\bar{\rho}}$ and $\varepsilon^2 < -\frac{4}{\bar{\rho}} \log\left(\frac{1}{2} \sqrt{\frac{\rho+2\bar{\rho}}{\rho+\bar{\rho}}}\right)$. Same as subcase i), agent 1 does not have a choice. For agents on the second layer to choose exit

at round 1, we need $\psi_n < \frac{2}{\rho+\bar{\rho}} - \frac{1}{\rho+2\bar{\rho}}$. For agents on the third layer to exit at round 2, we need

$$\psi_n < \min \left\{ \frac{2}{\rho+\bar{\rho}} - \frac{1}{\rho+2\bar{\rho}}, \frac{2}{\rho+2\bar{\rho}} - \frac{1}{\rho+3\bar{\rho}} \right\} = \frac{2}{\rho+2\bar{\rho}} - \frac{1}{\rho+3\bar{\rho}}.$$

In general, an agent on layer j prefers to wait till the $j-1$ round if

$$\psi_n < \min \{g(1), \dots, g(j-1)\} = g(j-1).$$

where $g(x) = \frac{2}{\rho+x\bar{\rho}} - \frac{1}{\rho+(x+1)\bar{\rho}}$. The last equality holds because $g(x)$ is a decreasing function, thanks to $g'(x) < 0$. Hence under equilibrium, agents on layer j have j signals. In

particular, agents in the last layer each has $m_n = \log_2(n+1)$ signals. Note that there are $\frac{n+1}{2}$ agents in this layer. Using (3.2), a similar derivation to that in Example 2 leads to that a learning rate $\{\delta_n\}$ can be

$$\delta_n \geq \frac{1}{n\varepsilon\bar{\varepsilon}\sqrt{\pi}} \sum_{i=1}^{\log_2(n+1)} 2^{i-1} \frac{1}{\sqrt{\rho+\bar{\rho}i}} \exp\left(-\frac{\varepsilon^2(\rho+\bar{\rho}i)}{4}\right).$$

To unravel the right hand side of the above inequality, we let

$$h(x) = 2^{x-1} \frac{1}{\sqrt{\rho+\bar{\rho}x}} \exp\left(-\frac{\varepsilon^2(\rho+\bar{\rho}x)}{4}\right),$$

which is monotone increasing, because $h(x+1)/h(x) > 1$ under our condition. Therefore, it is sufficient to have

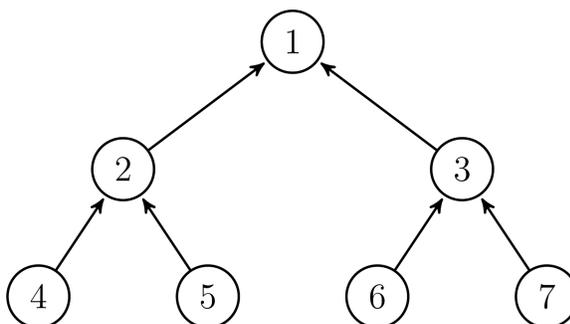
$$\delta_n \geq \frac{1}{n\varepsilon\bar{\varepsilon}\sqrt{\pi}} \log_2(n+1)h(\log_2(n+1)).$$

Therefore δ_n could decay in the order of $\sqrt{\log_2(n+1)} \cdot (n+1)^{-\varepsilon^2\bar{\rho}/4}$, which is a polynomial rate.

Compared to the complete graph, the binomial tree achieves perfect learning more slowly. The difference in learning rates arises not only from the physical network structures, but also from different interactions among agents in the two environments.

Next, we consider two cases in which information flows in the opposite direction.

Example 4 (Binomial Tree: Information Flow from Leafs to Root)—Now let information flow from leafs to root, i.e., reverse all the directed edges in Example 3. The following graph illustrates such a binomial tree with three layers.



We give the following results. The detailed analysis is similar to Example 3.

- i. $\psi_n = \rho = \bar{\rho} = 1$. All agents exit after time 0.
- ii. $\psi_n < \frac{2}{\rho+2^{(m_n-1)}\bar{\rho}} - \frac{1}{\rho+2^{m_n}\bar{\rho}}$. All agents get the maximum number of signals that they could possibly get, then δ_n can be such that

$$\delta_n \geq \frac{1}{n\varepsilon\bar{\varepsilon}\sqrt{\pi}} \sum_{j=1}^{\log_2(n+1)} 2^{j-1} \frac{1}{\sqrt{\rho+2^{(m_n-j+1)}\bar{\rho}}} \exp\left(-\frac{\varepsilon^2(\rho+2^{(m_n-j+1)}\bar{\rho})}{4}\right).$$

A conservative estimate on the summation on the right hand side would give $\delta_n \sim n^{-3/4}$, a much faster rate than that in Example 3 when $\varepsilon^2\bar{\rho} \ll 3$ (a typical case as we have in mind very small ε).

Note that information flow directions matter for learning rates. When parameters are in a comparable range, the bottom-up case exhibits a higher learning rate than the top-down case. In other words, the bottom-up organization of information flow within a binomial tree is more efficient. This result is consistent with early economics and sociology literature; Hayek

(1945) for example, highlight the importance of effectively organizing dispersed information sources.

5 Remarks and Further Research

We have explored multi-agent Bayesian inference at the societal level with a game theoretic social network model, highlighting agents' incentives and interactions in information acquisition. The inference quality is captured by a new concept, finite population learning, which answers whether with high probability, a large fraction of people in a finite population network can make "good" inference. Echoing the spirit of the finite sample method in statistics, our new concept helps reveal explicit interplays among people's preferences, network characteristics, and tolerances on inference quality. With finite population learning as a foundation, we also provide conditions to determine the long term trend of aggregate inference quality as population in a society grows. Our work offers the statistics community a tractable framework to assess the inference quality at the societal level, taking the interdependence of agents' inference and information acquisition into account.

A further question to ask is: what specific topologies of social networks would improve inference quality at the societal level? To answer this question might open a new direction for statistical research. Our paper makes an initial attempt to investigate the interplay between network topology and aggregate inference, but a generic solution is difficult to reach. The difficulty lies in the effect of the game parameters on the inference results; more concretely, slight perturbation of parameter realizations might lead to drastically different inference results in the same network. The fundamental reason is that our approach calls for complete knowledge on the network structure to determine the quality of aggregate inference. It would be interesting to develop new criteria for aggregate inference on network classes with only some summary statistics regarding topologies. To achieve this, we need to look for novel statistical properties of networks. Golub and Jackson (2012a,b,c) are promising attempts towards this direction, but their results are currently limited to non-Bayesian decision problems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Jianqing Fan research is supported by NSF grant DMS-1206464 and the National Institute of General Medical Sciences of NIH through Grant Number R01-GM072611. The authors thank the Editor, the Associate Editor, anonymous referees, Daron Acemoglu, Sébastien Bubeck, John Campbell, Darrell Duffie, Emmanuel Farhi, Drew Fudenberg, Benjamin Golub, Ning Hao, Matthew Jackson, Gareth James, Josh Lerner, Jingyi Jessica Li, Philip Reny, Philippe Rigollet, Andrei Shleifer, Alp Simsek, and Yiqing Xing, and numerous conference and seminar participants for valuable comments and helpful discussions.

References

- Acemoglu D, Bimpikis K, Ozdaglar A. Dynamics of information exchange in endogenous social networks. *Theoretical Economics*. 2012 forthcoming.
- Acemoglu D, Dahleh M, Lobel I, Ozdaglar A. Bayesian learning in social networks. *Review of Economic Studies*. 2011; 78:1201–1236.

- Banerjee A. A simple model of herd behavior. *Quarterly Journal of Economics*. 1992; 107:797–817.
- Banerjee A, Fudenberg D. Word-of-mouth learning. *Games and Economic Behavior*. 2004; 46:1–22.
- Bikhchandani S, Hirshleifer D, Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*. 1992; 100:992–1026.
- Crawford V, Sobel J. Strategic information transmission. *Econometrica*. 1982; 50:1431–1451.
- DeGroot M. Reaching a consensus. *Journal of the American Statistical Association*. 1974; 69:118–121.
- DeGroot M. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*. 2003; 118:909–968.
- Duffie D, Malanud S, Manso G. Information percolation with equilibrium search dynamics. *Econometrica*. 2009; 77:1513–1574.
- Easley, D.; Kleinberg, J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press; 2010.
- Fudenberg, D.; Tirole, J. *Game Theory*. MIT Press; 1991.
- Galeotti A, Ghiglino C, Squintani F. Strategic information transmission in networks. *Journal of Economic Theory*. 2013; 148:1751–1769.
- Golub B, Jackson M. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*. 2010; 2:112–149.
- Golub B, Jackson M. Does homophily predict consensus times? testing a model of network structure via a dynamic process. *Review of Network Economics*. 2012a; 11:1–31.
- Golub B, Jackson M. How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*. 2012b; 127:1287–1338.
- Golub B, Jackson M. Network structure and the speed of learning: Measuring homophily based on its consequences. *Annals of Economics and Statistics*. 2012c; 107/108:33–50.
- Hagenbach J, Koessler F. Strategic communication networks. *Review of Economic Studies*. 2010; 77:1072–1099.
- Hayek F. The use of knowledge in society. *The American Economic Review*. 1945; 35:519–530.
- Jackson, M. *An Overview of Social Networks and Economic Applications*. North Holland: 2010.
- Jadbabaie A, Molavi P, Tahbaz-Salehi A. Information heterogeneity and the speed of learning in social networks. working paper. 2013
- Kolaczyk, E. *Statistical Analysis of Network Data: Methods and Models*. Springer; 2009.
- Mckelvey, R.; McLennan, A. *Computation of Equilibria in Finite Games*. Elsevier; 1996.
- Mobius M, Phan T, Szeidl A. Treasure hunt. working paper. 2010
- Muller-Frank M. A general framework for rational learning in social networks. *Theoretical Economics*. 2012; 8:1–40.
- Newman, M. *Networks: an Introduction*. Oxford University Press; 2010.
- Smith L, Sorensen P. Pathological outcomes of observational learning. *Econometrica*. 2000; 68:371–398.
- Topkis D. Equilibrium points in non-zero sum n-person submodular games. *SIAM J. of Control and Optimization*. 1979; 17:773–787.