Sample Complexity of Linear Regression Models for Opinion Formation in Networks

Haolin Liu[1*], Rajmohan Rajaraman [2], Ravi Sundaram[2],
Anil Vullikanti[1], Omer Wasim[2], Haifeng Xu[3]
[1]University of Virginia
[2]Northeastern University
[3] University of Chicago
srs8rh@virginia.edu, {r.rajaraman, r.sundaram}@northeastern.edu, asv9@virginia.edu, wasim.o@northeastern.edu, haifengxu@uchicago.edu

Abstract

Consider public health officials aiming to spread awareness about a new vaccine in a community interconnected by a social network. How can they distribute information with minimal resources, so as to avoid polarization and ensure community-wide convergence of opinion? To tackle such challenges, we initiate the study of sample complexity of opinion formation in networks. Our framework is built on the recognized *opinion formation game*, where we regard each agent's opinion as a data-derived model, unlike previous works that treat opinions as data-independent scalars. The opinion model for every agent is initially learned from its local samples and evolves game-theoretically as all agents communicate with neighbors and revise their models towards an equilibrium. Our focus is on the sample complexity needed to ensure that the opinions converge to an equilibrium such that every agent's final model has low generalization error.

Our paper has two main technical results. First, we present a novel polynomial time optimization framework to quantify the *total sample complexity* for arbitrary networks, when the underlying learning problem is (generalized) linear regression. Second, we leverage this optimization to study the *network gain* which measures the improvement of sample complexity when learning over a network compared to that in isolation. Towards this end, we derive network gain bounds for various network classes including cliques, star graphs, and random regular graphs. Additionally, our framework provides a method to study sample distribution within the network, suggesting that it is sufficient to allocate samples inversely to the degree. Empirical results on both synthetic and real-world networks strongly support our theoretical findings.

1 Introduction

In today's interconnected world, rapid dissemination of information plays a pivotal role in shaping public understanding and behavior, especially in areas of critical importance like public health. People learn and form opinions through personal investigations and interactions with neighbors. The heterogeneity of social networks often means that not every individual requires the same amount of information to form an informed opinion. Some may be heavily influenced by their peers, while others may need more direct information. This differential requirement presents both a challenge and an opportunity: How can one distribute information to ensure the entire community is well-informed, without unnecessary redundancies or gaps in knowledge dissemination?

To answer this question, the first step is to understand the dynamics of opinion formation within social networks. Starting from the seminal work of DeGroot (1974), one of the predominant models assumes that individuals shape their own beliefs by consistently observing and averaging the opinions of their network peers (DeGroot 1974; Friedkin and Johnsen 1990; DeMarzo, Vayanos, and Zwiebel 2003; Golub and Jackson 2008; Bindel, Kleinberg, and Oren 2015; Haddadan, Xin, and Gao 2024). We will adopt this seminal framework to model the opinion formation process among agents. However, these works always encapsulate opinions as single numeric values. While this provides a foundational understanding, it cannot capture the nuanced ways in which individuals process and integrate new information into their existing belief systems. To address this issue, our formulation generalizes the opinions to be a model parameter learned using a machine learning-based approach, as explored in (Haghtalab, Jackson, and Procaccia 2021) to study belief polarization. Such formulation aligns with the human decision-making procedure proposed by (Simon 2013) based on prior experiences and available data.

Specifically, we adopt the game theoretical formulation of this opinion dynamic process introduced by (Bindel, Kleinberg, and Oren 2015). Within this framework, an agent's equilibrium opinion emerges as a weighted average of everyone's initial opinions on the network (which are learned form their locally available data), where the weights are unique and determined by the network structure. It is not guaranteed, however, that a node's equilibrium opinion would have a small generalization error. Motivated by the work of (Blum et al. 2021), we study the conditions under this actually happens (i.e., each node has a model with a small generalization error): specifically, how should samples be distributed across the network such that at the equilibrium of the opinion formation game, everyone has a model that is close enough to the ground-truth, and the total number of samples is minimized; we are also interested in understanding the network gain, i.e., how much does the collaboration

[*]Alphabetical order

on a network help the sample complexity.

We note that when opinions are treated as data-driven model parameters, the equilibrium models share the same mathematical structure as the *fine-grained federated learning* model introduced in (Donahue and Kleinberg 2021). However, their study assumes fixed sample sizes without considering networks and focuses on collaboration incentives. Our work focuses on the optimal allocation of samples across networks to ensure that the generalization error of the equilibrium model remains low.

1.1 Problem Formulation

We consider a fixed set $V = \{1, 2, \cdots, n\}$ of agents (also referred to as nodes) connected by a network $G = (V, E)$. Let $N(i)$ denote the set of neighbors, and d_i the degree of agent i. We assume every agent i in a given network G has a dataset $S_i = \{(x_i^j, y_i^j)\}_{j\in[m_i]}$ allocated by a market-designer, where m_i is the number of samples at i, and $x_i^j \in \mathbb{R}^k$, $\forall j \in [m_i]$, and are independently drawn from a fixed distribution D with dimension k. Each agent i learns an initial model $\bar{\theta}_i$ (known as internal opinion); $\bar{\theta}_i$ need not be a fixed number (as in (Bindel, Kleinberg, and Oren 2015)), but can be learned through a machine learning approach, similar as (Haghtalab, Jackson, and Procaccia 2021). We assume that these datasets are allocated by a system designer and that $m_i \geq k$ (i.e., S_i has at least k samples), which ensures everyone has enough basic knowledge to form their own opinion before learning over the social network. Our full paper (Liu et al. 2023) includes a table of all notations.

Agent i's goal is to find a model θ_i that is close to its internal opinion, as well as to that of its neighbors, denoted by $N(i)$; i.e. compute θ_i that minimizes the loss function $\|\theta_i - \bar{\theta}_i\|^2 + \sum_{j\in N(i)} v_{ij}\|\theta_i - \theta_j\|^2$, where $v_{ij} \geq 0$ measures the influence of a neighboring agent j to agent i. We refer to $\boldsymbol{v}$, where $[\boldsymbol{v}]_{ij} = v_{ij}$, as the influence factor matrix. In Lemma 1, we show that the unique Nash equilibrium of this game is $(\theta_1^{eq}, \cdots, \theta_n^{eq})^T = W^{-1}(\bar{\theta}_1, \cdots, \bar{\theta}_n)^T$ where $W = \mathcal{L} + I$ and $\mathcal{L}$ is the weighted Laplacian matrix of G.

We define the ***total sample complexity***, $TSC(G, \boldsymbol{v}, k, \epsilon)$, as the minimum number of total samples, $\sum_i m_i$, subject to the constraint $m_i \geq k$, so that the expected square loss of θ_i^{eq} is at most ϵ for all i; here $m_i = |S_i|$. In the special case where the influence is uniform, i.e., $v_{ij} = \alpha, \forall i, j$, we use $TSC(G, \alpha, k, \epsilon)$ to denote the total sample complexity.

Let $\widetilde{M}(k, \epsilon)$ denote the minimum number of samples that any agent i would need to ensure that the best model $\bar{\theta}_i$ learned using only their samples ensures that the expected error is at most some target ϵ; if there is no interaction between the agents, a total of $n\widetilde{M}(k, \epsilon)$ samples would be needed to ensure that $\bar{\theta}_i$ has low error for each agent. Since every agent has at least k samples, we refer to

$$\mu(G, \boldsymbol{v}, k, \epsilon) = \frac{n(\widetilde{M}(k, \epsilon) - k)}{TSC(G, \boldsymbol{v}, k, \epsilon) - nk}, \tag{1}$$

the ratio of the additional samples needed to achieve error of ϵ for every agent under social learning to that needed under independent learning, as the ***network gain*** of G. We will show that for linear regression $\widetilde{M}(k, \epsilon) = \Theta(\frac{k}{\epsilon})$ (Theorem 2). We assume $\widetilde{M}(k, \epsilon) > k$ and the network gain can be infinite when $TSC(G, \boldsymbol{v}, k, \epsilon) = nk$.

1.2 Overview of Results

Our focus here is to estimate the $TSC(G, \boldsymbol{v}, k, \epsilon)$, and the network gain $\mu(G, \boldsymbol{v}, k, \epsilon)$, and characterize the distribution of the optimal m_is across the network. Most proofs are presented in our full paper (Liu et al. 2023).

Tight bounds on TSC. Using the structure of the Nash equilibrium of the opinion formation game (Lemma 1), and regression error bounds (Theorem 3), we derive tight bounds on $TSC(G, \boldsymbol{v}, k, \epsilon)$ for any graph G using a mathematical program (Theorem 4). We also show that the TSC can be estimated in polynomial time using second-order cone programming, allowing us to study it empirically.

Graph class	Network gain
Clique	$\Omega(n)$
Star	$O(1)$
Hypercube	$\Omega(d^2)$
Random d-regular	$\Omega(\min\{d^2, n\})$
d-Expander	$\Omega(\min\{d^2\tau(G)^4, n\})$

Table 1: Network gain $\mu(G, \alpha, k, \epsilon)$ for different graphs where $\tau(G)$ denotes the edge expansion of G (defined in Section 4).

Impact of graph structure on TSC, network gain, and sample distribution. We begin with the case of uniform influence factor α. We show that $\sum_{i=1}^n \frac{1}{(\alpha d_i+1)^2} \lesssim \frac{TSC(G,\alpha,k,\epsilon)}{\widetilde{M}(k,\epsilon)} \lesssim \sum_{i=1}^n \frac{\alpha+1}{\alpha d_i+1}$ (informal version of Theorem 5), where d_i denotes the degree of agent i. Assigning $\max\{k, \frac{\alpha+1}{\alpha d_i+1} \cdot \frac{k}{\epsilon}\}$ samples to agent i can ensure everyone learns a good model. Thus, ***It is sufficient to solve the TSC problem when the number of samples for an agent is inversely proportional to its degree***. In other words, low-degree nodes need more samples, in stark contrast to many network mining problems, e.g., influence maximization, where it suffices to choose high-degree nodes. Clearly, this result has policy implications. Building on Theorem 5, we derive tight bounds on the network gain for different classes of graphs, summarized in Table 1.

From Table 1, we can see a well-connected network offers a substantial reduction in TSC (high network gain), whereas a star graph provides almost no gain compared to individual learning. We also demonstrate that the lower bounds on network gain in Table 1 are tight. More detailed results for specific networks can be found in Table 2.

Finally, we consider general influence factors and derive upper and lower bounds on TSC for arbitrary graphs (Theorem 7). We find that these bounds are empirically tight.

Experimental evaluation. In our simulation experiments, we compute the total sample complexity and the near-optimal solutions for a large number of synthetic and real-world networks. We begin with the case of uniform influence factors and first validate the findings of Theorem 5 that sample size in the optimal allocation has negative correlation with degree. We then experimentally evaluate the d^2 network gains for random d-regular graphs. Overall, we find that experimentally computed solutions are consistent with our bounds in Table 1. Finally, we consider general influence factors and find that our theoretical bounds on TSC in Theorem 7 are quite tight empirically.

1.3 Related Work and Comparisons

Sample complexity of collaborative learning. Building on Blum et al. (2017), a series of papers (Chen, Zhang, and Zhou 2018; Nguyen and Zakynthinou 2018; Blum et al. 2021; Haghtalab, Jordan, and Zhao 2022) studied the minimum number of samples to ensure every agent has a low-error model in collaborative learning. In this setting, there is a center that can iteratively draw samples from different mixtures of agents' data distributions, and dispatch the final model to each agent. In contrast, we use the well-established decentralized opinion formation framework to describe the model exchange game; the final model of every agent is given by the equilibrium of this game.

Our formulation is similar to (Blum et al. 2021), which also considers the sample complexity of equilibrium, ensuring every agent has a sufficient utility guarantee. However, (Blum et al. 2021) study this problem in an incentive-aware setting without networks, which mainly focuses on the stability and fairness of the equilibrium. In contrast, our research is centered on the network effect of the equilibrium generated by the opinion formation model. Moreover, (Blum et al. 2021) assumes the agents' utility has certain structures that are not derived from error bounds while we directly minimize the generalization error of agents' final models.

Haddadan, Xin, and Gao (2024) is another related paper which also considers learning on networks under opinion dynamic process. However, they study selecting K agents to correct their prediction to maximize the overall accuracy in the entire network, rather than the sample complexity bound to ensure individual learner has a good model as our paper.

Fully decentralized federated learning. To reduce the communication cost in standard federated learning (McMahan et al. 2017), Lalitha et al. (2018, 2019) first studied fully decentralized federated learning on networks, where they use Bayesian approach to model agents' belief and give an algorithm that enables agents to learn a good model by belief exchange with neighbors. This setting can be regarded as a combination of Bayesian opinion formation models (Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992; Smith and Sørensen 2000; Acemoglu, Chernozhukov, and Yildiz 2006) and federated learning. In the literature regarding opinion formation on networks, besides those Bayesian models, non-Bayesian models are usually considered more flexible and practical (DeGroot 1974; Friedkin and Johnsen 1990; DeMarzo, Vayanos, and Zwiebel 2003; Golub and Jackson 2008; Bindel, Kleinberg, and Oren 2015; Haddadan, Xin, and Gao 2024). Comprehensive surveys of these two kinds of models can be found in Jackson (2010) and Acemoglu and Ozdaglar (2011). Our study makes connections between non-Bayesian opinion formation models and federated learning for the first time. Compared with Lalitha et al. (2018, 2019), we assume each agent can observe the model of neighbors, rather than a belief function. We do not restrict to specific algorithms but use game theoretical approaches to find the unique Nash equilibrium and analyze sample complexity at this equilibrium.

2 The Opinion Formation Game

As mentioned earlier, we utilize a variation on the seminal DeGroot model (DeGroot 1974) proposed by (Friedkin and Johnsen 1990), also studied by (Bindel, Kleinberg, and Oren 2015). Formally, agent i seeks θ_i which minimizes the loss

$$\|\theta_i - \bar{\theta}_i\|^2 + \sum_{j \in N(i)} v_{ij} \|\theta_i - \theta_j\|^2$$

where $v_{ij} \geq 0,\ \forall i, j \in [n], j \in N(i)$ measures the influence of agent j to agent i. In general, v_{ij}'s might not be known, and we will also study the simpler uniform case where $v_{ij} = \alpha \geq 0\ \forall i, j$. Lemma 1 gives the unique equilibrium of this game (also studied by Bindel, Kleinberg, and Oren (2015)); its proof is presented in Appendix B.2 of (Liu et al. 2023).

Lemma 1 (Nash equilibrium of opinion formation) *The unique Nash equilibrium* $\boldsymbol{\theta^{eq}} = (\theta_1^{eq}, \cdots, \theta_n^{eq})^T$ *of the above game is* $\boldsymbol{\theta^{eq}} = W^{-1}\bar{\boldsymbol{\theta}}$ *where* $W_{ij} = \begin{cases} \sum_{j \in N(i)} v_{ij} + 1 & j = i \\ -v_{ij} & j \in N(i) \\ 0 & j \notin N(i), j \neq i \end{cases}$ *and* $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \cdots, \bar{\theta}_n)$. *When all* $v_{ij} = \alpha \geq 0$, $W_{ij} = \begin{cases} \alpha d_i + 1 & j = i \\ -\alpha & j \in N(i) \\ 0 & j \notin N(i), j \neq i \end{cases}$. *Furthermore, we have* $\sum_{j=1}^{n} W_{ij}^{-1} = 1$ *and* $W_{ij}^{-1} \geq 0$ *for all* $i, j \in [n]$.

Thus, from Lemma 1, the equilibrium model of every agent is a convex combination of all the agents' internal opinion on the network. We show that $\boldsymbol{\theta^{eq}}$ can be computed in a federated manner using the algorithm of (Vanhaesebrouck, Bellet, and Tommasi 2017).

3 Total Sample Complexity of Opinion Formation

Recall that each agent i's dataset $S_i = \{(x_i^j, y_i^j)\}_{j \in [m_i]}$ has $m_i \geq k$ samples, where $x_i^j \in \mathbb{R}^k$. We assume there is a ground-truth model θ^* such that for any $x_i^j \sim D$, $y_i^j = (x_i^j)^\top \theta^* + \eta_i^j$ where $\eta_i^j \sim \eta_i(x_i^j)$ and $\eta_i : \mathbb{R}^k \to \Delta(\mathbb{R})$ is an agent-dependent noise function, mapping samples to a noise distribution. We consider unbiased noise with bounded variance, implying that for every agent i, noise function η_i is independently chosen from $\mathcal{N} = \{\eta : \mathbb{E}[\eta(x)] = 0, \mathrm{Var}[\eta(x)] \leq \sigma^2, \forall x \sim D\}$. Let $X_i = [x_i^1, \cdots, x_i^{m_i}]$ and $Y_i = [y_i^1, \cdots, y_i^{m_i}]^\top$. Every agent performs ordinary linear regression to output their initial opinion $\bar{\theta}_i = \mathrm{argmin}_\theta \sum_{j=1}^{m_i} \left((x_i^j)^\top \theta - y_i^j\right)^2$. For our study, we first make standard Assumption 1 on data distributions.

Assumption 1 (non-degeneracy) *For data distribution D over $\mathbb{R}^k$, if x is drawn from D, then for any linear hyperplane $H \subset \mathbb{R}^k$, we have $\mathbb{P}(x \in H) = 0$.*

Assumption 1 is standard to ensure the data distributions span over the whole $\mathbb{R}^k$ space. If it holds, from Fact 1 in (Mourtada 2022), for any X_i with $m_i \geq k$, $X_i^\top X_i$ is invertible almost surely. This implies the ordinary linear regression for every agent to form the initial opinion enjoys the closed-form solution $\bar{\theta}_i = \left(X_i^\top X_i\right)^{-1} X_i^\top Y_i$.

Loss measure. We use the expected square loss in the worst case of noise to measure the quality of agents' final opinions $(\theta_1^{eq}, \cdots, \theta_n^{eq})$ at the equilibrium. Namely, for agent i, we consider the loss

$$L(\theta_i^{eq}) = \sup_{\eta_{1:n} \in \mathcal{N}} \mathbb{E}_{x \sim D, \forall j, S_j \sim D^{m_j}(\eta_j)} \left[\left(x^\top \theta_i^{eq} - x^\top \theta^* \right)^2 \right] \quad (2)$$

where $D^{m_i}(\eta_i)$ denotes the joint distribution of $S_i = \{(x_i^j, y_i^j)\}_{j \in [m_i]}$ given noise function η_i. We take supremum over all possible noises and take expectation over all dataset because θ_i^{eq} is related to $\bar{\theta}_i$ for every agent i.

3.1 Derivation of Error Bounds

We first define the error for initial opinion as

$$L(\bar{\theta}_i) = \sup_{\eta_i \in \mathcal{N}} \mathbb{E}_{x \sim D, S_i \sim D^{m_i}(\eta_i)} \left[\left(x^\top \bar{\theta}_i - x^\top \theta^* \right)^2 \right].$$

To quantify the upper bound of $L(\bar{\theta}_i)$, we additionally consider the following assumption on data distribution.

Assumption 2 (small-ball condition) *For data distribution D, there exists constant $C_i \geq 1$ and $\alpha_i \in (0, 1]$ such that for every hyperplane $H \subset \mathbb{R}^k$ and $t > 0$, if x is drawn from D, we have $\mathbb{P}\left(\text{dist}(\Sigma_i^{-\frac{1}{2}} x, H) \leq t\right) \leq (C_i t)^{\alpha_i}$ where $\Sigma_i = \mathbb{E}_{x \sim D}\left[xx^\top\right]$.*

Given Assumption 1, Σ_i is guaranteed to be invertible. Assumption 2 ensures $\Sigma_i^{-\frac{1}{2}} x$ is not too close to any fixed hyperplane, which is introduced in (Mourtada 2022) and is a variant of the small-ball condition in (Koltchinskii and Mendelson 2015; Mendelson 2015; Lecué and Mendelson 2016). From Proposition 5 in (Mourtada 2022) (see also Theorem 1.2 in (Rudelson and Vershynin 2015)), if every coordinate of D are independent and have bounded density ratio, then Assumption 2 holds. More discussions on this assumption could be found in Section 3.3 in (Mourtada 2022).

Theorem 2 (Thm 1 and Prop 2 in (Mourtada 2022)) *For every agent i with m_i samples, ordinary linear regression attains loss $L(\bar{\theta}_i) = \Theta\left(\frac{k}{m_i}\right)$.*

For certain data distributions, tighter closed-form bounds have been derived. For instance, if D is a k-dimensional multivariate Gaussian distribution with zero mean, then $L(\bar{\theta}_i) = \frac{\sigma^2 k}{m_i - k - 1}$ (Anderson et al. 1958; Breiman and Freedman 1983; Donahue and Kleinberg 2021). For mean estimation, $L(\bar{\theta}_i) = \frac{\sigma^2}{m_i}$ (Donahue and Kleinberg 2021). Theorem 2 shows that $L(\bar{\theta}_i)$ scales with $\frac{k}{m_i}$, assuming the data distributions satisfy non-degeneracy and the small-ball condition. Thus, $\widetilde{M}(k, \epsilon) = \Theta(\frac{k}{\epsilon})$ samples suffice for a single agent learning a model with error ϵ.

We next present our main technique for bounding the generalization error of an equilibrium model, which uses Theorem 2 to express the error as a function of the matrix W and the number of samples at each agent.

Theorem 3 (Bound on generalization error) *For every agent i, we have $L(\theta_i^{eq}) = \Theta\left(k \sum_{j=1}^n \frac{(W_{ij}^{-1})^2}{m_j}\right)$ where θ_i^{eq} and matrix W is defined in Lemma 1.*

Remark. In Appendix B.2 of (Liu et al. 2023), we show that Theorem 3 also holds for generalized linear regression with adapted assumptions, where a mapping function ϕ exists such that $\mathbb{E}[y] = \phi(x)^\top \theta^*$ for any possible data (x, y).

3.2 Total Sample Complexity

Armed with Theorem 3, we initiate the study of total sample complexity of opinion convergence. Recall that we want to ensure that at the equilibrium, every node on the network has a model with a small generalization error. Specifically, we want to ensure $L(\theta_i^{eq}) \leq \epsilon$ for any $i \in [n]$ and a given $\epsilon > 0$. Theorem 3 precisely gives us the mechanism to achieve a desired error bound.

The Total Sample Complexity (TSC) problem. Recall that for a given graph G, influence factor $\boldsymbol{v}$, dimension k and error ϵ, the total sample complexity $TSC(G, \boldsymbol{v}, k, \epsilon)$ is the minimum value of $\sum_i m_i$, under the constraint $m_i \geq k$, $m_i \in \mathbb{Z}^+$ and $L(\theta_i^{eq}) \leq \epsilon$ for every $i \in [n]$. Our central result, Theorem 4, derives near-tight bounds on the total sample complexity.

Theorem 4 (Bounds on TSC) *For any $\epsilon > 0$, let $(m_i^*, i = 1, \ldots, n)$ denote an optimal solution of the following optimization problem as a measure of the minimum samples for opinion formation on graph G with influence factor $\boldsymbol{v}$.*

$$\begin{aligned} \min_{m_1, \cdots, m_n} \quad & \sum_{i=1}^n m_i \\ \text{s.t.} \quad & \sum_{j=1}^n \frac{(W_{ij}^{-1})^2}{m_j} \leq \frac{\epsilon}{k}, \ \forall i \\ & m_i > 0, \ \forall i \end{aligned} \quad (3)$$

where W is defined in Lemma 1. Then, $TSC(G, \boldsymbol{v}, k, \epsilon) = \Theta(\sum_{i=1}^n m_i^* + nk)$. Assigning $\max\{O(\lceil m_i^* \rceil), k\}$ samples to agent i suffices for $L(\theta_i^{eq}) \leq \epsilon$ for every $i \in [n]$. Moreover, $\frac{\epsilon m_i^*}{k}$ is a fixed value for any k and ϵ.

From Theorem 4, $TSC(G, \boldsymbol{v}, k, \epsilon)$ has order $\sum_{i=1}^n m_i^* + nk$ and assigning $\max\{\lceil m_i^* \rceil, k\}$ samples to agent i is sufficient to solve the TSC problem up to some constant. Thus, we only need to focus on the solution of Equation 3, $m_i^*, \forall i \in [n]$. In the following sections, we will characterize properties of $m_i^*, \forall i \in [n]$ and use it to prove network gain for different graphs.

4 Network Effects on Total Sample Complexity

In this section, we analyze the network effects on the total sample complexity (TSC) of opinion convergence. We derive bounds on network gain, demonstrating how network learning reduces sample requirements. For uniform influence weights, we establish tight asymptotic bounds for key network classes and determine optimal sample distribution by node degree. We then generalize to arbitrary weights, providing empirically tight TSC bounds.

4.1 Uniform Influence Factors

We model uniform influence factors by setting $v_{ij} = \alpha$ for every agent i and neighbor j of i, for a given real $\alpha \geq 0$. Let $TSC(G, \alpha, k, \epsilon)$ be the solution of the optimization in Theorem 4 for this case. Theorem 5 provides interpretable bounds for $TSC(G, \alpha, k, \epsilon)$ related to degree, and serves as the first step to derive tight bounds for specific graph classes. The proof of Theorem 5 leverages the dual form of Equation 3, together with a careful analysis of matrix W defined in Lemma 1. Detailed proof can be found in (Liu et al. 2023).

Theorem 5 (Sample allocation and degree distribution) *The optimal solution $\{m_i^*\}$ to Equation 3 satisfies*

$$\max\left\{\sum_{i=1}^{n}\frac{1}{(\alpha d_i+1)^2}, 1\right\} \leq \sum_{i\in[n]} m_i^* \cdot \frac{\epsilon}{k} \leq \sum_{i=1}^{n}\frac{\alpha+1}{\alpha d_i+1}.$$

We have $TSC(G, \alpha, k, \epsilon) = \Theta(\sum_{i\in[n]} m_i^* + nk)$.

Assigning $\max\{O\left(\lceil\frac{\alpha+1}{\alpha d_i+1}\cdot\frac{k}{\epsilon}\rceil\right), k\}$ samples to every agent i suffices for $L(\theta_i^{eq}) \leq \epsilon,\ \forall i \in [n]$.

Theorem 5 suggests that it is sufficient to allocate samples *inversely proportional* to the node degrees. This has interesting policy implications; in contrast to other social network models, e.g., the classic work on influence maximization (Kempe, Kleinberg, and Tardos 2003), our result advocates that allocating more resources to low-connectivity agents benefit the network at large. We provide empirical validation of this phenomenon in Section 5.

With the help of Theorem 5, we could derive the bounds of network gain for any graph in Corollary 6.

Corollary 6 *For any graph G, if $\epsilon \leq O(\frac{1}{\alpha \max_i d_i+1})$, then $\mu(G, \alpha, k, \epsilon) \geq \Omega\left(\frac{n}{\sum_{i=1}^{n}\frac{\alpha+1}{\alpha d_i+1}}\right)$. If $\epsilon \leq O(\frac{1}{\max_i(\alpha d_i+1)^2})$, then $\mu(G, \alpha, k, \epsilon) \leq O\left(\min\left\{\frac{n}{\sum_{i\in[n]}\frac{1}{(\alpha d_i+1)^2}}, n\right\}\right)$.*

For small ϵ, Corollary 6 demonstrates that the network gain is at least $\Omega\left(\frac{n}{\sum_{i\in[n]}\frac{1}{d_i}}\right)$, implying that networks with more high-degree nodes result in higher network gain, but the gain is limited to $O\left(\min\left\{\frac{n}{\sum_{i\in[n]}\frac{1}{d_i^2}}, n\right\}\right)$.

Now we are ready to analyze the tight network gain for special graphs classes. Our main results are shown in Table 2, with detailed lemmas and proofs deferred to (Liu et al. 2023). The third column of Table 2 gives $\sum_{i\in[n]} m_i^*$, which suffices to characterize total sample complexity (TSC) because $TSC(G, \alpha, k, \epsilon) = \Theta\left(\sum_{i\in[n]} m_i^* + nk\right)$ from Theorem 4. The last column indicates the number of samples required for each agent to ensure that all constraints of the TSC problem are satisfied. For each kind of graph, if the sample size for every agent is the maximum of k and the term in the last column, it is sufficient to guarantee $L(\theta_i^{eq}) \leq \epsilon,\ \forall i$. To prove the results in Table 2, we leverage the spectral properties of graph Laplacian $\mathcal{L}$ for different networks, together with the lower bound in Theorem 5. Note that the lower bound in Table 2 does not require any constraint on ϵ. This contrasts with the bounds in Corrollary 6 and requires a more refined analysis. On the other hand, there is no general upper bound for the network gain because it can be infinite.

We now give a more detailed discussion of the results in Table 2 as follows. **(1)** For a clique, the network gain is $\Omega(n)$ and $TSC(G, \alpha, k, \epsilon) = \Theta\left(\frac{k}{\epsilon} + nk\right)$. From Corollary 6, this is the best lower bound of network gain for any graph. **(2)** For a star, when ϵ is less than some constant, the network gain is $O(1)$ and $TSC(G, \alpha, k, \epsilon) = \Theta\left(\frac{nk}{\epsilon}\right)$. This implies a star graph almost cannot get any gain compared with learning individually. **(3)** For a hypercube, when $\alpha \geq \frac{3}{8}$, the network gain is $\Omega(d^2)$ and $TSC(G, \alpha, k, \epsilon) = \Theta\left(\frac{nk}{d^2\epsilon} + nk\right)$. **(4)** For a random d-regular graph, with high probability, the network gain is $\Omega(d^2)$ when $d \leq \sqrt{n}$ while it is $\Omega(n)$ when $d > \sqrt{n}$. The total sample complexity is $TSC(G, \alpha, k, \epsilon) = \Theta\left(\frac{nk}{\min\{d^2,n\}\epsilon} + nk\right)$. **(5)** For a d-regular graph G, the edge expansion $\tau(G)$ equals $\min_{|S|:|S|\leq n/2}\frac{|\partial S|}{d|S|}$, where ∂S is the set $\{(u, v) \in E : u \in S, v \in V \setminus S\}$. Edge expansion measures graph connectivity and can be viewed as a lower bound on the probability that a randomly chosen edge from any subset of vertices S has one endpoint outside S. A d-expander denotes a d-regular graph with edge expansion $\tau(G)$. For this kind of graph, the network gain is $\Omega(\min\{d^2\tau(G)^4, n\})$ and $TSC(G, \alpha, k, \epsilon) = \Theta(\frac{nk}{\min\{d^2\tau(G)^4,n\}\epsilon} + nk)$. Note that the $\Omega\left(\min\{d^2, n\}\right)$ lower bound is also optimal for these $d-$regular graphs from the upper bound in Corollary 6.

Graph	Network gain	$\sum_{i\in[n]} m_i^*$	Samples / node
Clique (Lemma 13)	$\Omega(n)$	$\Theta(\frac{k}{\epsilon})$	$O(\frac{k}{\epsilon n})$
Star (Lemma 14)	$O(1)$	$\Theta(\frac{nk}{\epsilon})$	$O(\frac{k}{\epsilon})$ (leaf) $O(\frac{k}{n\epsilon})$ (center)
Hypercube (Lemma 16)	$\Omega(d^2)$	$\Theta(\frac{nk}{d^2\epsilon})$	$O(\frac{k}{d^2\epsilon})$
Random d-regular (Lemma 17)	$\Omega(\min\{n, d^2\})$	$\Theta(\frac{nk}{\min\{n,d^2\}\epsilon})$	$O(\frac{k}{\min\{n,d^2\}\epsilon})$
d-Expander (Lemma 18)	$\Omega(\min\{n, d^2\tau^4\})$	$\Theta(\frac{nk}{\min\{n,d^2\tau^4\}\epsilon})$	$O(\frac{k}{\min\{n,d^2\tau^4\}\epsilon})$

Table 2: Network gain $\mu(G, \alpha, \epsilon)$, $\sum_{i\in[n]} m_i^*$ and distribution of samples in different network classes and constant α. The number of samples at a node is the maximum of k and the term in the third column. In the last row, $\tau = \tau(G)$ is the edge expansion.

It follows from Table 2 and Corollary 6 that both random and expander d-regular graphs (with constant $\tau(G)$) have the optimal lower bound for network gain. Interestingly, this property also holds for the hypercube, which has regular degree $d = \log n$, even though its expansion is $O(\frac{1}{\log n}) = o(1)$. A natural open question is to characterize other degree-bounded network families which also achieve near-optimal network gains.

4.2 General Influence Factors

Going beyond uniform influence factors, we investigate the total sample complexity of opinion convergence with general influence factors. The non-uniformity of the influence factors implies that the sample complexity is not just dependent on the network structure, but also on how each individual agent weighs the influence of its neighbors. A trivial bound is $\Omega\left(\frac{k}{\epsilon}\right) \leq TSC(G, \boldsymbol{v}, k, \epsilon) \leq O\left(\frac{nk}{\epsilon}\right)$ which means learning through the game is always more beneficial than learning individually but needs at least the samples for one agent to learn a good model.

To derive more sophisticated and tighter bounds, we need to analyze the matrix W with general weights, which captures both the network topology and influence factors. Since W^{-1} is positive-definite, from the Schur product theorem, $(W^{-1}) \circ (W^{-1})$ is also positive-definite where $\circ$ is the Hadamard product (element-wise product). Define $B = ((W^{-1})\circ(W^{-1}))^{-1}$ and $[B]_{ij} = b_{ij}$. Theorem 7 establishes a bound for graphs with general influence factors, which is empirically tight from our experiments in Section 5. The proof for the upper bound utilizes a thorough analysis of the property of b_{ij}s, and the lower bound is derived based on the dual form of Equation 3 with general weights. We refer the reader to (Liu et al. 2023) for the whole proof.

Theorem 7 (TSC under general influence factors) *Let* $\gamma_i = \max\{0, \sum_{j=1}^n \frac{b_{ij}}{(\sum_{k=1}^n b_{jk})^2}\}$ *for every* $i \in [n]$*, then the optimal solution* $\{m_i^*\}$ *to Equation 3 satisfies*

$$\sum_{i=1}^n \left(2\sqrt{\sum_{j=1}^n \gamma_j (W_{ij}^{-1})^2} - \gamma_i\right) \leq \sum_i m_i^* \cdot \frac{\epsilon}{k} \leq \sum_{i=1}^n \frac{1}{\sum_{j=1}^n b_{ij}}$$

and $TSC(G, \boldsymbol{v}, k, \epsilon)$ *is* $\Theta(\sum_i m_i^* + nk)$. *Assigning* $\max\left\{k, O\left(\lceil \frac{k}{\epsilon \sum_{j=1}^n b_{ij}} \rceil\right)\right\}$ *samples to agent* i *suffices for* $L(\theta_i^{eq}) \leq \epsilon$ *for every* $i \in [n]$.

Although the above two bounds give good empirical estimations to the TSC, deriving closed-form solution for Equation 3 is still interesting, and is investigated in Lemma 8 under a certain condition.

Lemma 8 *If* $\sum_{j=1}^n \frac{b_{ij}}{(\sum_{k=1}^n b_{jk})^2} \geq 0$ *for all* $i \in [n]$, $\sum_{i\in[n]} m_i^* = \sum_{i=1}^n \frac{1}{\sum_{j=1}^n b_{ij}} \cdot \frac{k}{\epsilon}$ *where the optimal* $m_i^* = \frac{1}{\sum_{j=1}^n b_{ij}} \cdot \frac{k}{\epsilon}$.

We only know Lemma 8 holds for cliques now. Characterizing the graphs that meet this condition remains an interesting open problem.

5 Experiments

We use experiments on both synthetic and real-world networks to further understand the distribution of samples (Theorem 5), the quality of our bounds for general influence weights (Theorem 7), and the network gains for d-regular graph (Table 2 and Lemma 17). We observe good agreement with our theoretical bounds. We only list part of results here and more experiments and be found in (Liu et al. 2023).

From Theorem 4, given a network and influence factors, $\frac{\epsilon m_i^*}{k}$ is fixed for any k and ϵ. This value can be solved directly by reformulating Equation 3. Since $\widetilde{M}(k, \epsilon) = \Theta(\frac{k}{\epsilon})$, $\frac{\epsilon m_i^*}{k}$ indicates the percentage of samples one agent needs at the equilibrium compared to the number of samples required to learn independently, which is the main factor we are interested in. Thus, we only consider $\frac{\epsilon m_i^*}{k}$ for experiments of Theorem 5 and Theorem 7 without specific ϵ and k.

To solve $m_i^\star$ or $\frac{\epsilon m_i^*}{k}$ for every $i \in [n]$ exactly, we reformulate Equation 3 to a second-order cone programming (SOCP) and then use the solver CVXOPT ((Andersen et al. 2013)) to solve it. More details are in (Liu et al. 2023). We now describe the networks used in our experiments.

Synthetic networks. We use three types of synthetic networks: scale-free (**SF**) networks (Barabási and Albert 1999), random d-regular graphs (**RR**), and Erdös-Renyi random graphs (**ER**). The exact parameters for these graphs used in different experiments are given in (Liu et al. 2023).

Real-world networks. The networks we use (labeled as **RN**) are shown in Table 3 with their features (err_u/err_l will be defined later). We use the 130bit network, the econ-mahindas network ((Rossi and Ahmed 2015)), the ego-Facebook network ((Leskovec and Mcauley 2012)), and the email-Eu-core temporal network ((Paranjape, Benson, and Leskovec 2017)). The last two networks are also in Leskovec and Krevl (2014).

Network	**Nodes**	**Edges**	err_u	err_l
ego-Facebook	4039	88234	27%	24%
Econ	1258	7620	6.6%	1.3%
Email-Eu	986	16064	30%	21%
130bit	584	6067	41%	36%

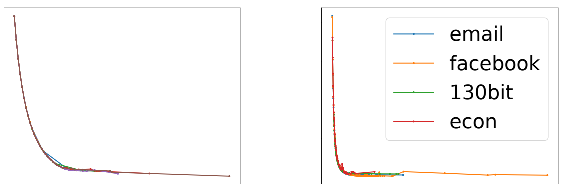
Table 3: Real-world networks and bound performance

Our results are summarized below.

Distribution of samples (Theorem 5). We set the uniform influence factor $\alpha = 0.1$, showing how the sample assignment is related to degrees at the solution of Equation 3 on synthesized scale-free networks (Figures 1(a)) and real-world networks (Figure 1(b)). Figures 1(a) and 1(b) show that the average number of samples (divided by $\frac{k}{\epsilon}$) for each degree, where the x-axis is degree and y-axis is the average sample for each degree (i.e average of $N^d = \{\frac{\epsilon m_i^*}{k} : d_i = d, i \in [n]\}$ where $m_i^*, \forall i \in [n]$ is the solution of Equation 3). We also record the variance of samples for each degree (variance of N^d). For synthesized scale-free networks, the variance is bounded by $1e-5$ while for real-world networks, the variance is bounded by $1e-3$. More network instances and detailed variance reprot can be found in (Liu et al. 2023).

Our main observations are: (1) The samples assigned to nodes with the same degree tend to be almost the same (i.e., the variance is small), (2) Fewer samples tend to be assigned to high-degree agents. These observations are consistent with Theorem 5, which states that sample assignment has negative correlation with degree.

Tightness of bounds (Theorem 7). We empirically show the performance of the bounds in Theorem 7. Here, all influence factors v_{ij}s are generated randomly from $[0, 1]$. For each kind of synthetic network, we generate 150 different instances with different number of nodes n and v_{ij}s. We meausure the performance of bounds in Theorem 7 by relative error (i.e $\frac{|U \text{ or } L - \sum_{i\in[n]} m_i^*|}{\sum_{i\in[n]} m_i^*}$ where U is the upper bound in Theorem 7 and L is the max of 1 (trivial lower bound) and the lower bound in Theorem 7). We visualize the relative error through frequency distributions of generated networks. Our observation is that the bounds are very tight for scale-free graphs (Figure 2). In Liu et al. (2023), we also show they are tight for other graph instances.



(a) SF Average (b) RN Average

Figure 1: Relationship between sample distribution and degree for different networks when $\alpha = 1$. The x-axis is node degree and the y-axis is the average samples for each degree. Lines with different colors represent different graphs.

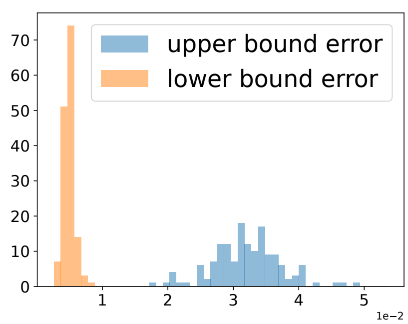


Figure 2: Bound tightness for synthetic networks for random v_{ij}s. The x-axis is the relative errors of upper and lower bounds and the y-axis is the frequency.

For real-world networks, we generate 50 different $v_{ij} \in [0, 1]$ and denote the maximum relative error of upper/lower bounds as err_u/err_l. From Table 3, the relative errors are still relatively small for real-world networks, showing the tightness of our bounds.

Network Gain (Table 2 and Lemma 17). We empirically demonstrate the network gain results for random d-regular graphs (Lemma 17). To calculate the network gain, we replace the $TSC(G, \boldsymbol{v}, k, \epsilon)$ in network gain (Equation 1) by its upper bound $\sum_{i=1}^{n} m_i^* + nk + n$. This allows us to calculate a lower bound of the network gain, which aligns with the trend of lower bound in Lemma 17. We set $k = 5$ and $\epsilon = 0.01$, making $\frac{k}{\epsilon} = 500$. Additionally, we set $\alpha = 1$ and consider degree ranging from 2 to 10. We choose different number of nodes n such that degree $d \leq \sqrt{n}$, which implies a d^2 gain in theory (4^{th} row of Table 2 and Lemma 17). The empirical result, shown in Figure 3, exactly confirms the gain is $\Theta(d^2)$ in our setup. Moreover, we also found that when degree $d = 11$, the gain becomes infinity. This is because $\frac{k}{d^2\epsilon} = \frac{500}{121} < k = 5$. According to the 4^{th} row of Table 2 (Lemma 17), in this scenario, it is sufficient for each node to have k samples, resulting in an infinite network gain.

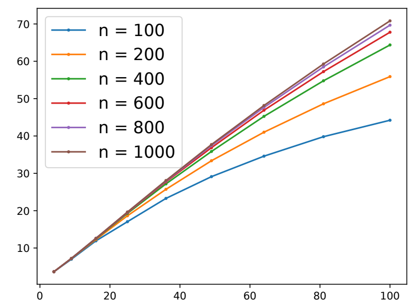


Figure 3: Network Gain of RR. The x-axis is the square of degree and the y-axis is a lower bound of network gain.

Impact of influence factors. When α is small, sample assignment is mostly independent of n; however, for large α, increasing n leads to fewer samples for nodes of the same degree. Additionally, the bounds in Theorem 7 get worse as influence factors increase. See (Liu et al. 2023) for details.

6 Discussion and Future Works

We initiate a study on the total sample complexity (TSC) required for opinion formation over networks. By extending the standard opinion formation game into a machine-learning framework, we characterize the minimal samples needed to ensure low generalization error for equilibrium models. We provide tight bounds on TSC and analyze sample distributions, showing that network structure significantly reduces TSC, with an inverse relationship to node degree being an interesting result.

Our formulation opens several research directions. We currently focus on linear regression, but it's worth investigating other problems like kernel methods or soft SVMs. Moreover, it turns out that our problem formulation is similar to the best arm identification problem with fixed confidence (Karnin, Koren, and Somekh 2013). It is interesting to extend our results analog to the fixed budget setting. Namely, assuming the number of samples is fixed, how to assign them to minimize the total error? Another future direction is to consider individual-level incentives for sample collection, building on the results of (Blum et al. 2021).

Acknowledgments

RR and OW were partially supported by NSF grant CCF-2335187. AV was partially supported by NSF grants CCF-1918656, IIS-1955797, CNS 2317193 and IIS 2331315. RS was partially supported by NSF grant D-ISN 2146502.

References

Acemoglu, D.; Chernozhukov, V.; and Yildiz, M. 2006. Learning and disagreement in an uncertain world.

Acemoglu, D.; and Ozdaglar, A. 2011. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1): 3–49.

Andersen, M. S.; Dahl, J.; Vandenberghe, L.; et al. 2013. CVXOPT: A Python package for convex optimization. *abel. ee. ucla. edu/cvxopt*, 88.

Anderson, T. W.; Anderson, T. W.; Anderson, T. W.; and Anderson, T. W. 1958. *An introduction to multivariate statistical analysis*, volume 2. Wiley New York.

Banerjee, A. V. 1992. A simple model of herd behavior. *The quarterly journal of economics*, 107(3): 797–817.

Barabási, A.-L.; and Albert, R. 1999. Emergence of scaling in random networks. *science*, 286(5439): 509–512.

Bikhchandani, S.; Hirshleifer, D.; and Welch, I. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5): 992–1026.

Bindel, D.; Kleinberg, J.; and Oren, S. 2015. How bad is forming your own opinion? *Games and Economic Behavior*, 92: 248–265.

Blum, A.; Haghtalab, N.; Phillips, R. L.; and Shao, H. 2021. One for one, or all for all: Equilibria and optimality of collaboration in federated learning. In *International Conference on Machine Learning*, 1005–1014. PMLR.

Blum, A.; Haghtalab, N.; Procaccia, A. D.; and Qiao, M. 2017. Collaborative PAC Learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 2389–2398. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Breiman, L.; and Freedman, D. 1983. How many variables should be entered in a regression equation? *Journal of the American Statistical Association*, 78(381): 131–136.

Chen, J.; Zhang, Q.; and Zhou, Y. 2018. Tight bounds for collaborative pac learning via multiplicative weights. *Advances in neural information processing systems*, 31.

DeGroot, M. H. 1974. Reaching a consensus. *Journal of the American Statistical association*, 69(345): 118–121.

DeMarzo, P. M.; Vayanos, D.; and Zwiebel, J. 2003. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly journal of economics*, 118(3): 909–968.

Donahue, K.; and Kleinberg, J. 2021. Model-sharing games: Analyzing federated learning under voluntary participation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 5303–5311.

Friedkin, N. E.; and Johnsen, E. C. 1990. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4): 193–206.

Golub, B.; and Jackson, M. O. 2008. How homophily affects diffusion and learning in networks. *arXiv preprint arXiv:0811.4013*.

Haddadan, S.; Xin, C.; and Gao, J. 2024. Optimally Improving Cooperative Learning in a Social Setting. In *International Conference on Machine Learning*. PMLR.

Haghtalab, N.; Jackson, M. O.; and Procaccia, A. D. 2021. Belief polarization in a complex world: A learning theory perspective. *Proceedings of the National Academy of Sciences*, 118(19): e2010144118.

Haghtalab, N.; Jordan, M.; and Zhao, E. 2022. On-demand sampling: Learning optimally from multiple distributions. *Advances in Neural Information Processing Systems*, 35: 406–419.

Jackson, M. O. 2010. *Social and economic networks*. Princeton university press.

Karnin, Z.; Koren, T.; and Somekh, O. 2013. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, 1238–1246. PMLR.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the Spread of Influence through a Social Network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, 137–146. New York, NY, USA: Association for Computing Machinery. ISBN 1581137370.

Koltchinskii, V.; and Mendelson, S. 2015. Bounding the smallest singular value of a random matrix without concentration. *International Mathematics Research Notices*, 2015(23): 12991–13008.

Lalitha, A.; Kilinc, O. C.; Javidi, T.; and Koushanfar, F. 2019. Peer-to-peer federated learning on graphs. *arXiv preprint arXiv:1901.11173*.

Lalitha, A.; Shekhar, S.; Javidi, T.; and Koushanfar, F. 2018. Fully decentralized federated learning. In *Third workshop on Bayesian Deep Learning (NeurIPS)*.

Lecué, G.; and Mendelson, S. 2016. Performance of empirical risk minimization in linear aggregation.

Leskovec, J.; and Krevl, A. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

Leskovec, J.; and Mcauley, J. 2012. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25.

Liu, H.; Rajaraman, R.; Sundaram, R.; Vullikanti, A.; Wasim, O.; and Xu, H. 2023. Sample Complexity of Opinion Formation on Networks. *arXiv preprint arXiv:2311.02349*.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Mendelson, S. 2015. Learning without concentration. *Journal of the ACM (JACM)*, 62(3): 1–25.

Mourtada, J. 2022. Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *The Annals of Statistics*, 50(4): 2157–2178.

Nguyen, H.; and Zakynthinou, L. 2018. Improved algorithms for collaborative PAC learning. *Advances in Neural Information Processing Systems*, 31.

Paranjape, A.; Benson, A. R.; and Leskovec, J. 2017. Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining*, 601–610.

Rossi, R. A.; and Ahmed, N. K. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.

Rudelson, M.; and Vershynin, R. 2015. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19): 9594–9617.

Simon, H. A. 2013. *Administrative behavior*. Simon and Schuster.

Smith, L.; and Sørensen, P. 2000. Pathological outcomes of observational learning. *Econometrica*, 68(2): 371–398.

Vanhaesebrouck, P.; Bellet, A.; and Tommasi, M. 2017. Decentralized Collaborative Learning of Personalized Models over Networks. In Singh, A.; and Zhu, X. J., eds., *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, volume 54 of *Proceedings of Machine Learning Research*, 509–517. PMLR.