

# Differentially Private Distributed Estimation and Learning

Marios Papachristou<sup>a</sup> and M. Amin Rahimian<sup>b</sup>

<sup>a</sup>Cornell University, papachristoumarios@cs.cornell.edu

<sup>b</sup>University of Pittsburgh, rahimian@pitt.edu

## Abstract

We study distributed estimation and learning problems in a networked environment where agents exchange information to estimate unknown statistical properties of random variables from their privately observed samples. The agents can collectively estimate the unknown quantities by exchanging information about their private observations, but they also face privacy risks. Our novel algorithms extend the existing distributed estimation literature and enable the participating agents to estimate a complete sufficient statistic from private signals acquired offline or online over time and to preserve the privacy of their signals and network neighborhoods. This is achieved through linear aggregation schemes with adjusted randomization schemes that add noise to the exchanged estimates subject to differential privacy (DP) constraints, both in an offline and online manner. We provide convergence rate analysis and tight finite-time convergence bounds. We show that the noise that minimizes the convergence time to the best estimates is the Laplace noise, with parameters corresponding to each agent’s sensitivity to their signal and network characteristics. Our algorithms are amenable to dynamic topologies and balancing privacy and accuracy trade-offs. Finally, to supplement and validate our theoretical results, we run experiments on real-world data from the US Power Grid Network and electric consumption data from German Households to estimate the average power consumption of power stations and households under all privacy regimes and show that our method outperforms existing first-order privacy-aware distributed optimization methods.

*Keywords:* Distributed Learning, Differential Privacy, Estimation

## 1 Introduction

Differential privacy (DP) is a gold standard in privacy-preserving algorithm design that limits what an adversary (or any observer) can learn about the inputs to an algorithm by observing its outputs (Dwork, 2011; Dwork and Roth, 2014), according to a privacy budget that is usually denoted by  $\epsilon$ . It requires that given the output, the probability that any pair of adjacent inputs generate the observed output should be virtually the same. Adding noise to the input data helps enforce this standard in different settings — e.g., for distributed

learning — but the added noise can also degrade our performance, e.g., lowering the quality of distributed estimation and collective learning for which agents exchange information.

This paper provides aggregation algorithms that facilitate distributed estimate and learning among networked agents while accommodating their privacy needs (e.g., protecting their private or private signals and network neighborhoods). Each algorithm implies a different tradeoff between its quality of collective learning and how much privacy protection it affords the participating agents (i.e., their privacy budgets). Our performance metrics reflect how distributional features of the private signals and the nature of privacy needs for individual agents determine the learning quality and requisite noise.

Decentralized decision-making and distributed learning problems arise naturally in a variety of applications ranging from sensor and robotic networks in precision agriculture, digital health, and military operations to the Internet of things and social networks (Bullo et al., 2009; Jackson, 2008); see Section 1.2 for a detailed literature review. We are particularly interested in distributed estimation problems that arise in smart grids with distributed generation and energy resources. Notably, a recent report from National Academies of Sciences, Engineering, and Medicine (2023a,b) suggests that net metering practices should be revised to reflect the value of distributed electricity generation, such as rooftop solar panels. Net metering compensates customers for the electricity they provide to the grid through distributed generation. The report notes that net metering has facilitated the embrace of distributed generation in states where it has been put into effect, resulting in levels surpassing 10% in a few states and projected to rise in both these and other states. Additionally, the report emphasizes the need to revisit and evolve net metering policies to support the deployment of distributed generation that adds value in reducing fossil fuel use, enhancing resilience, and improving equity. In this context, each customer faces an individual privacy risk in sharing their estimates since revealing exact measurements can pose security risks that can be leveraged by an adversary (e.g., understanding when someone is at their home, daily habits, family illness, etc.), and, therefore, developing privacy-preserving methods that support decentralized decision making in such setups is critical.

This paper introduces novel algorithms designed for the distributed estimation of the expected value of sufficient statistics in an exponential family distribution. The proposed methods leverage signals received by individual agents, who maintain and update estimates based on both these signals and information from their local neighborhood. Our contribu-

tions to the existing literature on distributed optimization include new privacy-aware distributed estimation algorithms that exhibit faster convergence rates compared to established first-order methods (cf. Rizk et al. (2023)). Notably, our algorithms safeguard the information in agents’ signals and local neighborhood estimates, ensuring optimal convergence times to true estimates. Furthermore, in contrast to existing approaches, our algorithms can support privacy-aware estimation within an online learning framework, accommodate dynamic topologies, and balance privacy and accuracy by distributing the privacy budget among agents. Finally, we verify our proposed algorithms on real-world datasets and show that they outperform existing first-order methods.

## 1.1 Main Results

**Summary of main notations.** We use barred bold letters to denote vectors (e.g.,  $\bar{\mathbf{x}}$ ) and bold letters to indicate vector components (e.g.,  $\mathbf{x}_i$  is a component of  $\bar{\mathbf{x}}$ ). We use capital letters to denote matrices (e.g.,  $A$ ) and small letters to represent their entries (e.g.,  $a_{ij}$ ). We use small letters to denote scalars. We denote the  $n \times 1$  column vector of all ones by  $\mathbf{1}$ .

**Summary of the Problem Setup (see Section 2.2).** We consider a network of  $n$  agents indexed by  $[n] = \{1, \dots, n\}$  whose interconnections are characterized by a symmetric, doubly-stochastic, adjacency matrix  $A$ . This adjacency structure, encoded by graph neighborhoods,  $\mathcal{N}_i = \{j : a_{ij} \neq 0\}, i \in [n]$ , may be a consequence of geospatial constraints such as sensing and communication range or geographic proximity; it can also be a reflection of how the network has evolved and other engineering considerations, e.g., which nodes belong to which countries or companies in a multi-party network. The adjacency weights may also result from geoeconomic constraints such as access to local trade and business intelligence (contracts, sales, and filled orders). In the case of social networks, they can also represent the presence of influence and mutual interactions among individuals.

Given the adjacency structure  $A$ , at every round  $t \in \{1, 2, 3, \dots\}$ , each agent  $i$  receives a private signal  $\mathbf{s}_{i,t}$  that is sampled from an exponential family distribution with the natural sufficient statistic  $\xi(\cdot) \in \mathbb{R}$  and a common, unknown parameter  $\theta$  belonging to a measurable set  $\Theta$ . The goal of the agents is to collectively estimate the common value of  $\mathbb{E}_\theta [\xi(\mathbf{s})]$  by combining their samples. This is achieved through a consensus algorithm by forming an estimate  $\boldsymbol{\nu}_{i,t}$  and exchanging it with their network neighbors in a distributed manner respecting the adjacency structure of  $A$ . The agents also want to control the information

that is leaked about their signals and the estimates of their neighbors  $\{\nu_{j,t-1}\}_{j \in \mathcal{N}_i}$ . They add noise  $\mathbf{d}_{i,t}$  to their updates. The noise level should be high enough not to violate an  $\varepsilon$  differential privacy budget. Briefly, we say that a mechanism  $\mathcal{M}$  is  $\varepsilon$ -DP when for any pair of “adjacent” inputs (i.e., private signals or private signals and neighboring estimates), the logarithm of the probability ratio of any output in the range space of  $\mathcal{M}$  being resulted from either of the adjacent inputs is bounded by  $\varepsilon$ :  $|\log(\mathbb{P}[\mathcal{M}(\mathbf{s}) \in R]/\mathbb{P}[\mathcal{M}(\mathbf{s}') \in R])| \leq \varepsilon$ , for all adjacent pairs  $(\mathbf{s}$  and  $\mathbf{s}')$  in the input domain and any subset  $R$  of the output range space. Our specific notion of adjacency between the input pairs will be determined by the nature of information leaks, i.e., private signals or private signals and network neighborhoods, against which the exchanged estimates  $(\nu_{i,t})$  are being protected (Figure 1).

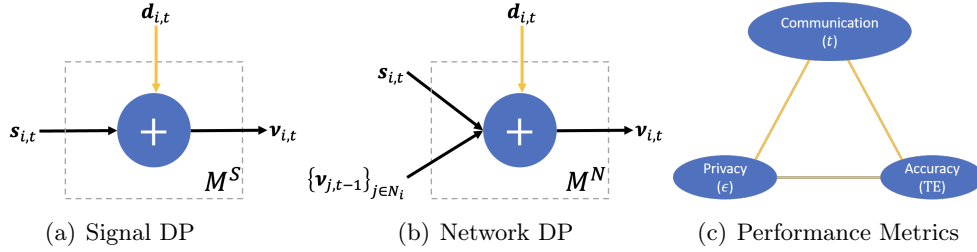


Figure 1: Two types of DP protections considered in this paper are signal DP,  $\mathcal{M}^S$ , and network DP,  $\mathcal{M}^N$ . The private signal of agent  $i$  at round  $t$  is denoted by  $\mathbf{s}_{i,t}$ ,  $\mathbf{d}_{i,t}$  corresponds to the noise added from agent  $i$  at round  $t$ , and  $\nu_{i,t}$  corresponds to the estimate of agent  $i$  at round  $t$ . Our theoretical guarantees delineate the relationship between communication resources ( $t$  rounds), privacy budget ( $\varepsilon$ -DP), and total error (TE). Signal and network DP imply different performance tradeoffs as detailed in Table 1.

**Theoretical Contributions (Sections 3.1 and 4).** In this paper, we provide bounds for the convergence of the DP estimates  $\bar{\nu}_t = (\nu_{i,t})_{i \in [n]}$  to the desired value  $\bar{\mathbf{m}}_\theta := \mathbf{1}\mathbf{m}_\theta$ , where  $\mathbf{m}_\theta = \mathbb{E}_\theta [\xi(\mathbf{s})]$ . We decompose the total error as follows:

$$\underbrace{\mathbb{E} [\|\bar{\nu}_t - \bar{\mathbf{m}}_\theta\|_2]}_{\text{Total Error (TE)}} \leq \underbrace{\mathbb{E} [\|\bar{\nu}_t - \bar{\boldsymbol{\mu}}_t\|_2]}_{\text{Cost of Privacy (CoP)}} + \underbrace{\mathbb{E} [\|\bar{\boldsymbol{\mu}}_t - \bar{\mathbf{m}}_\theta\|_2]}_{\text{Cost of Decentralization (CoD)}}$$

Here,  $\bar{\boldsymbol{\mu}}_t = (\boldsymbol{\mu}_{i,t})_{i \in [n]}$  corresponds to the vector of non-private estimates. The first term corresponds to the “Cost of Privacy” (CoP), the estimation cost incurred by the  $\varepsilon$ -DP noising. The second term corresponds to the expected error from running the non-private distributed learning algorithm and, therefore, measures the “Cost of Decentralization” (CoD). In Section 3, when we consider “offline” estimation of  $\mathbf{m}_\theta$  from a fixed collection of initial signals available at the beginning ( $t = 0$ ), we replace  $\mathbf{m}_\theta$  by the best possible estimate — the mini-

mum variance unbiased estimate (MVUE) — of an omniscient observer who has *centralized* access to all the private signals.

Our goal is to find differentially private aggregation mechanisms  $\mathcal{M}$  with fast convergence guarantees that minimize CoP parametrized by noise distributions  $\{\mathcal{D}_{i,t}\}_{i \in [n], t \geq 1}$  of the agents and subject to their  $\varepsilon$  differential privacy budget constraints, i.e., “s.t.  $\varepsilon$ -DP”:

$$\text{CoP}(\mathcal{M}) = \inf_{\{\mathcal{D}_{i,\tau}\}_{i \in [n], \tau \in [t]} \text{ s.t. } \varepsilon\text{-DP}} \mathbb{E}_{\{(\bar{\mathbf{v}}_\tau, \bar{\mathbf{s}}_\tau)\}_{\tau \in [t]}} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2]. \quad (1)$$

From now on, we will refer to this optimal value as the cost of privacy. In our analysis, the CoP and CoD are proportional to the signal and noise variance. The convergence rate also depends on the number of nodes  $n$  and the *spectral gap*  $\beta^*$  of the doubly-stochastic adjacency matrix  $A$ , which dictates the convergence rate of  $A^t$  to its limiting matrix  $(\mathbf{1}\mathbf{1}^T)/n$ , as  $t \rightarrow \infty$ .

Subsequently, the noise distribution can be optimized by minimizing the weighted variances of the noise terms in the upper bounds subject to DP constraints. Our main results are summarized in Table 1. These consist of minimum variance unbiased estimation (Section 3) and online learning (Section 4) of expected values under (i) protection of the private signals (Signal DP), and (ii) protection of the private signals and the local neighborhoods (Network DP). The following Informal Theorem summarizes our theoretical contributions along with Table 1:

**Informal Theorem.** If  $\Delta = \max_{\mathbf{s} \in \mathcal{S}} \left| \frac{d\xi(\mathbf{s}_j)}{ds_j} \right|$  is the global sensitivity of  $\xi$  at time  $t$ , and  $[a_{ij}]_{i,j=1}^n$  are the weights of the adjacency matrix, then the upper bound in CoP is minimized by the Laplace noise with parameters  $\frac{\Delta}{\varepsilon}$  for the case of Signal DP, and parameters  $\frac{\max\{\max_{j \in \mathcal{N}_i} a_{ij}, \Delta\}}{\varepsilon}$  for the case of network DP.

Moreover, whenever the global sensitivity  $\Delta$  is unbounded, something that happens to a variety of sufficient statistics  $\xi$ , we rely on the smooth sensitivity introduced by Nissim et al. (2007) and derive the same algorithms (but with parameters depending on the smooth sensitivity instead of  $\Delta$ ), which achieve  $(\varepsilon, \delta)$ -DP, namely using the smooth sensitivity introduces a compromise in the  $\varepsilon$ -DP guarantee by a small information leakage probability  $\delta$ , relaxing the DP constraint for a mechanism  $\mathcal{M}$  as follows:  $\mathbb{P}[\mathcal{M}(\mathbf{s}) \in R] \leq e^\varepsilon \mathbb{P}[\mathcal{M}(\mathbf{s}') \in R] + \delta$ , for all adjacent input pairs  $(\mathbf{s}$  and  $\mathbf{s}')$  and all subsets  $R$  of the output range space.

Table 1: Total Error Bounds.  $\Delta$  is the maximum signal sensitivity (absolute value of the derivative of  $\xi(\cdot)$ ),  $M_n = \max_{i \in [n]} |\xi(\mathbf{s}_i)|$ ,  $a = \max_{i \neq j} a_{ij}$  is the maximum non-diagonal entry of the adjacency matrix  $A$ , and  $\beta^* = \max\{\lambda_2(A), |\lambda_n(A)|\}$  is the spectral gap of  $A$ . The **blue** terms are due to privacy constraints (CoP), and the **red** terms are due to decentralization (CoD).

|            | Minimum Variance Unbiased Estimation                                                                                                               | Online Learning of Expected Values                                                                                                        |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Signal DP  | <p>Theorem 3.1</p> $O\left(n(\beta^*)^t \left(\frac{\Delta}{\varepsilon} + M_n\right) + \frac{\Delta}{\varepsilon}\right)$                         | <p>Theorem 4.1</p> $O\left(\frac{n}{\sqrt{t}} \left(\frac{\Delta}{\varepsilon} + \sqrt{\nabla[\xi(\mathbf{s})]}\right)\right)$            |
| Network DP | <p>Corollary 3.2</p> $O\left(n(\beta^*)^t \left(\frac{\max\{a, \Delta\}}{\varepsilon} + M_n\right) + \frac{\max\{a, \Delta\}}{\varepsilon}\right)$ | <p>Theorem 4.2</p> $O\left(\frac{n}{\sqrt{t}} \left(\frac{\max\{a, \Delta\}}{\varepsilon} + \sqrt{\nabla[\xi(\mathbf{s})]}\right)\right)$ |

**Experiments (Section 5).** We conduct experiments with two real-world datasets motivated by decentralized decision problems in power grids (see Section 1). The first dataset considers the daily consumption of several German Households over three years (Milojkovic, 2018), and the second one considers the US Power Grid network from Watts and Strogatz (1998). Our experiments show that we can achieve  $(\varepsilon, \delta)$ -DP while not significantly sacrificing convergence compared to the non-private baselines. The results also indicate the increased challenges in ensuring network DP compared to Signal DP and the importance of distributional features of the signals, in particular, having sufficient statistics with bounded derivatives.

**Code and Data.** For reproducibility, we supplement the manuscript with our code and data, which can be found at: <https://github.com/papachristoumarios/dp-distributed-estimation>.

## 1.2 Related Work

Our results relate to different bodies of literature across the engineering, statistics, and economics disciplines, and in what follows, we shall expand upon these relations.

**Decentralized Decision Making and Distributed Learning** have attracted a large body of literature over the years, with notable examples of Borkar and Varaiya (1982); Tsitsiklis and Athans (1984); Tsitsiklis (1993); Papachristou and Rahimian (2024). Recently, there has been a renewed interest in this topic due to its applications to sensor and robotic networks (Chamberland and Veeravalli, 2003; Olfati-Saber and Shamma, 2005; Kar et al., 2012; Atanasov et al., 2014b,a) and emergence of a new literature considering networks of sensor and computational units (Shahrampour et al., 2015; Nedić et al., 2015, 2017).

Other relevant results investigate the formation and evolution of estimates in social networks and subsequent shaping of the individual and mass behavior through social learning (Krishnamurthy and Poor, 2013; Wang and Djuric, 2015; Lalitha et al., 2014; Rahimian and Jadbabaie, 2016a,c). Obtaining a global consensus by combining noisy and unreliable locally sensed data is a crucial step in many wireless sensor network applications; subsequently, many sensor fusion schemes offer good recipes to address this requirement (Xiao et al., 2005, 2006). In many such applications, each sensor estimates the field using its local measurements, and then, the sensors initiate distributed optimization to fuse their local forecast. If all the data from every sensor in the network can be collected in a fusion center, then a jointly optimal decision is readily available by solving the global optimization problem given all the data (Alexandru et al., 2020). However, many practical considerations limit the applicability of such a centralized solution. This gives rise to the distributed sensing problems that include distributed network consensus or agreement (Aumann, 1976; Geanakoplos and Polemarchakis, 1982; Borkar and Varaiya, 1982), and distributed averaging (Dimakis et al., 2008); with close relations to the consensus and coordination problems that are studied in the distributed control theory (Jadbabaie et al., 2003; Mesbahi and Egerstedt, 2010; Bullo et al., 2009). Our work contributes to this body of work by providing a privacy-aware method for distributed estimation over a network.

The work most importantly connected to our work is the work of Rizk et al. (2023), which introduces a first-order DP method for distributed optimization over a network of agents. While their work presents a general first-order method for DP distributed optimization, which is suitable for a larger family of optimization problems that include MVUE, adapting their method to our task comes with important trade-offs in the quality of the estimation, as running their method results in significantly higher error (up to  $1000\times$  more; see Figure 7) for the MVUE task due to the repeated inclusion of the signal in the belief updates. Moreover, contrary to ours, their method does not support the online learning regime. Finally, the concept of “graph-homomorphic noise” proposed in their paper is equivalent to the Network DP regime of our paper.

**Differential privacy** is a modern definition of data privacy encompassing many previous definitions, such as  $K$ -anonymity. A mechanism is differentially private if it maps similar records to the same value with equal probability. One consequence of the definition is that it guarantees that the outcome of a statistical analysis will be identical whether the individual

chooses to participate in the social learning process. Many previously proposed mechanisms can be shown to be differentially private, such as randomized response (Warner, 1965), the Laplace mechanism, and the Gaussian mechanism. The randomized response algorithm originally proposed by Warner (1965) consists of randomly perturbing binary responses, and it allows the population means to be recovered while giving individual respondents plausible deniability – an instance of adding noise to data.

Statistical disclosure control of donated data, e.g., submitting recommendations to a public policy agency or submitting online product reviews to an e-commerce platform, requires safeguarding donors’ privacy against adversarial attacks where standard anonymization techniques are shown to be vulnerable to various kinds of identification (Barbaro et al., 2006), linkage and cross-referencing (Sweeney, 2015; Narayanan and Shmatikov, 2008; Sweeney, 1997), and statistical difference and re-identification attacks (Kumar et al., 2007). Here, we propose to analyze the efficiency of distributed estimation where agents learn from each other’s actions protected by the gold standard of differential privacy (Dwork, 2011) to optimize statistical precision against the privacy risks to data donors who engage in social learning.

DP can be implemented using central or local schemes. In centralized DP implementations, individual data are collected, and privacy noise is added to data aggregates, used, e.g., in the U.S. Census Bureau’s Disclosure Avoidance system (US Census, 2020). In local implementations, DP noising is done as data is being collected. Local methods forego the need for any trusted parties and typically provide more fundamental protection that can withstand a broader range future infiltration, e.g., even a government subpoena for data cannot violate the privacy protection when collected data is itself subject to privacy noising — e.g., Google, LinkedIn, and Apple’s DP noising of their user data (Apple Differential Privacy Team, 2017; Erlingsson, 2014; Cardoso and Rogers, 2022; Rogers et al., 2020), cf. (Wilson et al., 2020; Guevara, 2019).

Regarding privacy and networks, Koufogiannis and Pappas (2017) present a privacy-preserving mechanism to enable private data to diffuse over social networks, where a user wants to access another user’s data and provide privacy guarantees on the privacy leak, which depends on the shortest path between two users in the network. Alexandru and Pappas (2021) study the problem of private weighted sum aggregation with secret weights, where a central authority wants to compute the weighted sum of the local data of some agents

under multiple privacy regimes. Rahimian et al. (2023a,b) study influence maximization using samples of network nodes that are collected in a DP manner.

Our work contributes to the above line of work by introducing a novel DP mechanism for distributed estimation and learning of exponential family distributions. Particularly, to the best of our knowledge, in the online learning regime, our algorithm introduces a novel weighting scheme that can protect both the individual signals and the neighboring beliefs, which can efficiently learn the expected value of the sufficient statistic. Moreover, we also provably derive the optimal distributions that minimize the convergence time of the algorithm and show that they are the Laplace distributions with appropriately chosen parameters.

**Cyber-Physical Systems** (e.g., energy, transportation systems, healthcare systems, etc.) correspond to the building blocks of modern information and communication technologies, whose privacy and security are crucial for the function of such technologies. There have been multiple methods, such as encryption and  $K$ -anonymity, to achieve privacy and security in cyber-physical systems (Hassan et al., 2019; Zhang et al., 2016; Kontar et al., 2021). By incorporating differential privacy techniques, such as noise injection or data aggregation, into the design and operation of cyber-physical systems, privacy risks can be mitigated while preserving the utility of the collected data. This ensures that individual privacy is protected, as the data released from these systems cannot be used to infer sensitive information about specific individuals. Moreover, the application of differential privacy to cyber-physical systems enables the collection and analysis of data at scale, allowing for improved system performance, anomaly detection, and predictive maintenance while maintaining the trust of individuals and protecting their privacy in an increasingly connected world (Li et al., 2010; Gowtham and Ahila, 2017; Xu et al., 2017). Our method’s efficiency, e.g., compared to Rizk et al. (2023); see Section 5 and Appendix B, makes it suitable for several large-scale data applications.

**Federated Learning (FL)** is a collaborative learning paradigm for decentralized optimization without the need to collect all data points in a central server for gradient calculations (McMahan and Thakurta, 2022; Kontar et al., 2021), with many applications in mind: distributed training of ML models (Bonawitz et al., 2021; Shi et al., 2023; Yue et al., 2022), healthcare (Kaissis et al., 2020), wireless communications (Niknam et al., 2020), etc. While more general than the setup we consider here, it suffers from issues in terms of com-

munication and privacy. Existing privacy-preserving FL methods (cf. Rizk et al. (2023)) usually adopt the instance-level differential privacy (DP), which provides a rigorous privacy guarantee but with several bottlenecks (Truong et al., 2021). Truex et al. (2019) proposed a privacy-aware FL system that combines DP with secure multiparty computation, which utilizes less noise without sacrificing privacy as the number of federating parties increases and produces models with high accuracy. Other FL methods, such as Zhang et al. (2022), accommodate differentially private updates via incorporating gradient clipping before adding privacy noise to achieve good performance subject to privacy constraints.

Contrary to most of these methods, which are first-order optimization methods that are suitable to a large variety of losses compared to our method, our zero-order belief updates for the MVUE are simple, more efficient, and have significantly lower error than first-order methods (see Figure 7 for comparison with Rizk et al. (2023)). Moreover, our method can learn from data that arrive in an online way, whereas methods such as Rizk et al. (2023); Zhang et al. (2022) are offline. Finally, most of these approaches rely on SGD. In contrast, our method focuses more on the decision-theoretic and statistical problem of estimating the expected value of the sufficient statistics of signals generated by an exponential family distribution.

## 2 Differential Privacy Protections in a Distributed Information Environment

### 2.1 The Distributed Information Aggregation Problem Setting

Let  $\Theta$  be any measurable set, and in particular, not necessarily finite. Consider a network of  $n$  agents and suppose that each agent  $i \in [n]$  observes an i.i.d. samples  $\mathbf{s}_i$  from a common distribution  $\ell(\cdot|\theta)$  over a measurable sample space  $\mathcal{S}$  (For simplicity in our proofs, we consider the simple case of 1D signals, i.e.,  $\mathcal{S} \subseteq \mathbb{R}$ . Extending to multi-dimensional signals (i.e.,  $\mathcal{S} \subseteq \mathbb{R}^s$ ) is straightforward and considers the  $\ell_\infty$  norm of the partial derivatives.). We assume that  $\ell(\cdot|\theta)$  belongs to a one-parameter exponential family so that it admits a probability density or mass function that can be expressed as

$$\ell(\mathbf{s}|\theta) = \tau(\mathbf{s})e^{\alpha(\theta)^T \xi(\mathbf{s}) - \varkappa(\alpha(\theta))}, \quad (2)$$

where  $\xi(\mathbf{s}) \in \mathbb{R}$  is a measurable function acting as a complete sufficient statistic for the i.i.d. random samples  $\mathbf{s}_i$ , and  $\alpha : \Theta \rightarrow \mathbb{R}$  is a mapping from the parameter space  $\Theta$  to the real line  $\mathbb{R}$ ,  $\tau(\mathbf{s}) > 0$  is a positive weighting function, and  $\varkappa(\alpha) := \ln \int_{\mathbf{s} \in \mathcal{S}} \tau(\mathbf{s}) e^{\alpha \xi(\mathbf{s})} d\mathbf{s}$  is a normalization factor known as the log-partition function. In (2),  $\xi(\cdot)$  is a complete sufficient statistic for  $\theta$ . It is further true that  $\sum_{i=1}^n \xi(\mathbf{s}_i)$  is a complete sufficient statistic given the  $n$  i.i.d. signals that the agents have received (Bickel and Doksum, 2015, Section 1.6.1). In particular, any inferences that involve the unknown parameter  $\theta$  based on the observed signals  $\bar{\mathbf{s}} = (\mathbf{s}_i)_{i \in [n]}$  can be equivalently performed given  $\sum_{i=1}^n \xi(\mathbf{s}_i)$ . The agents aim to estimate the expected value of  $\xi(\cdot)$ :  $\mathbf{m}_\theta = \mathbb{E}[\xi(\mathbf{s}_i)]$ , with as little variance as possible. The Lehmann-Scheffé theory — cf. (Casella and Berger, 2002, Theorem 7.5.1) — implies that any function of the complete sufficient statistic that is unbiased for  $\mathbf{m}_\theta$  is the almost surely unique minimum variance unbiased estimator of  $\mathbf{m}_\theta$ . In particular, the minimum variance unbiased estimator of  $\mathbf{m}_\theta$  given the initial data sets of all nodes in the network is given by:  $\widehat{\mathbf{m}}_\theta = (1/n) \sum_{i=1}^n \xi(\mathbf{s}_i)$ .

For concreteness, we can consider a group of  $n$  suppliers whose private signals consist of their contracts, sales orders, and fulfillment data. These suppliers would benefit from aggregating their private information to better estimate market conditions captured by the unknown parameter  $\theta$ , e.g., to predict future demand. However, sharing their private signals would violate the privacy of their customers and clients. In Section 2.2, we explain how the agents can compute the best (minimum variance) unbiased estimator of  $\mathbf{m}_\theta$  using average consensus algorithms Olshevsky (2014) that guarantee convergence to the average of the initial values without direct access to each other's private signals.

## 2.2 The Information Exchange Model

We consider an undirected network graph and let the undirected network  $\mathcal{G}(\mathcal{V} = [n], \mathcal{E})$  which corresponds to a Markov chain with a doubly-stochastic symmetric adjacency/transition matrix  $A = [a_{ij}]_{i,j=1}^n$  with the uniform stationary distribution. For instance, such an adjacency matrix  $A = [a_{ij}]_{i,j=1}^n$  can be defined according to the Metropolis-Hastings weights (Boyd

Algorithm 1: Non-Private Distributed Minimum Variance Unbiased Estimation

The agents initialize with:  $\boldsymbol{\mu}_{i,0} = \xi(\mathbf{s}_i)$ , and in any future time period the agents communicate their values and update them according to the following rule:

$$\boldsymbol{\mu}_{i,t} = a_{ii} \boldsymbol{\mu}_{i,t-1} + \sum_{j \in \mathcal{N}_i} a_{ij} \boldsymbol{\mu}_{j,t-1}. \quad (3)$$

Algorithm 2: Non-Private Online Learning of Expected Values

Initializing  $\boldsymbol{\mu}_{i,0}$  arbitrarily, in any future time period  $t \geq 1$  the agents observe a signal  $\mathbf{s}_{i,t}$ , communicate their current values  $\boldsymbol{\mu}_{i,t-1}$ , and update their beliefs to  $\boldsymbol{\mu}_{i,t}$ , according to the following rule:

$$\boldsymbol{\mu}_{i,t} = \frac{t-1}{t} \left( a_{ii} \boldsymbol{\mu}_{i,t-1} + \sum_{j \in \mathcal{N}_i} a_{ij} \boldsymbol{\mu}_{j,t-1} \right) + \frac{1}{t} \xi(\mathbf{s}_{i,t}). \quad (4)$$

et al., 2004):  $a_{ij} = 1/\max\{\deg(i), \deg(j)\}$  if  $(j, i) \in \mathcal{E}$ , and  $[A]_{ij} = 0$  otherwise for  $i \neq j$ ; furthermore,  $a_{ii} = 1 - \sum_{j \neq i} a_{ij}$ . This choice of weights leads to a Markov chain where the stationary distribution is the uniform distribution (Boyd et al., 2004), and the agents can set these weights locally based on their own and neighboring degrees without the global knowledge of the network structure. For choices of  $A$  that yield the fastest mixing Markov chain (but may not be locally adjustable), see Boyd et al. (2004).

The mechanisms for convergence, in this case, rely on the product of stochastic matrices, similar to the mixing of Markov chains (cf. Levin et al. (2009); Shahrampour et al. (2015)); hence, many available results on mixing rates of Markov chains can be employed to provide finite time guarantees after  $T$  iteration of the average consensus algorithm for fixed  $T$ . Such results often rely on the eigenstructure (eigenvalues/eigenvectors) of the communication matrix  $A$ , and the facts that it is a primitive matrix and its ordered eigenvalues satisfy  $-1 < \lambda_n(A) \leq \lambda_{n-1}(A) \leq \dots \leq \lambda_1(A) = 1$ , as a consequence of the Perron-Frobenius theory (Seneta, 2006, Theorems 1.5 and 1.7).

Moreover, another mechanism considers learning the expected values in an online way where agents receive signals at every round and then update their estimate by averaging the estimates of their neighbors, their own estimate, and the new signals (see Algorithm 2).

The  $1/t$  discounting provided in the above algorithm enables learning the expected values  $\mathbf{m}_\theta = \mathbb{E}_\theta[\xi(\mathbf{s})]$  asymptotically almost surely with a variance that scales as  $O(1/t)$ ; i.e.,

linearly in time. As shown in Rahimian and Jadbabaie (2016b), the variance upper bound comprises two terms. The former term considers the rate at which the Markov chain with transition matrix  $A$  is mixing and is governed by the spectral gap, i.e., the second largest magnitude of the eigenvalues of  $A$ . The latter term captures the diminishing variance of the estimates with the increasing number of samples gathered by all the agents in the network.

Now that we have the necessary background in distributed estimation, we present the two DP protection mechanisms that our paper considers: the Signal DP and the Network DP.

## 2.3 Differential Privacy Protections

### 2.3.1 Definitions and Mechanisms

In this paper, we consider two methods for differential privacy and refer to them as *Signal Differential Privacy (Signal DP)* and *Network Differential Privacy (Network DP)*. Both algorithms are local in principle; the agents simply add noise to their estimates to achieve a desired privacy guarantee. Roughly, Signal DP adds noise to protect the signal  $\mathbf{s}_{i,t}$  of each agent, and Network DP adds noise to protect the signal  $\mathbf{s}_{i,t}$  of each agent, as well as the estimates  $\{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}$  of her neighbors from round  $t - 1$ . We assume that the non-private network dynamics evolve as

$$\boldsymbol{\mu}_{i,t} = F_{i,t}(\boldsymbol{\mu}_{i,t-1}) + G_{i,t}(\{\boldsymbol{\mu}_{j,t-1}\}_{j \in \mathcal{N}_i}) + H_{i,t}(\mathbf{s}_{i,t}), \quad (5)$$

for each agent  $i \in [n]$ , and  $t \geq 1$ , where  $F_{i,t} : \mathbb{R} \rightarrow \mathbb{R}$ ,  $G_{i,t} : \mathbb{R}^{d_i} \rightarrow \mathbb{R}$ , and  $H_{i,t} : \mathcal{S} \rightarrow \mathbb{R}$  are functions determined by the learning algorithm, and correspond to the information from the agent's own estimate, the information from the neighboring estimates, and the information from the agent's private signal respectively. To achieve differential privacy, each agent adds some amount of noise  $\mathbf{d}_{i,t}$  drawn from a distribution  $\mathcal{D}_{i,t}$  to their estimate and reports the noisy estimate to their neighbors. The agent can either aim to protect only their private signal – which we call Signal DP and denote by  $\mathcal{M}^S$  –, or protect their network connections and their private signal – which we call Network DP and denote by  $\mathcal{M}^N$ . The noisy dynamics

are:

$$\boldsymbol{\nu}_{i,t} = F_{i,t}(\boldsymbol{\nu}_{i,t-1}) + \underbrace{G_{i,t}(\{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i})}_{\text{Network DP } \mathcal{M}_{i,t}^N} + \overbrace{H_{i,t}(\mathbf{s}_{i,t})}^{\text{Signal DP } \mathcal{M}_{i,t}^S} + \mathbf{d}_{i,t} \quad (6)$$

In Equation (6) and Figure 1, we have outlined the dynamics of the two types of privacy protections. Formally, the two types of mechanisms can also be written as

$$\begin{aligned} \psi_{\mathcal{M}_{i,t}^S}(\mathbf{s}_{i,t}) &= H_{i,t}(\mathbf{s}_{i,t}) + \mathbf{d}_{i,t}, & (\mathcal{M}_{i,t}^S) \\ \psi_{\mathcal{M}_{i,t}^N}(\mathbf{s}_{i,t}, \{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}) &= H_{i,t}(\mathbf{s}_{i,t}) + G_{i,t}(\{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}) + \mathbf{d}_{i,t}, & (\mathcal{M}_{i,t}^N) \end{aligned}$$

and the  $\varepsilon$ -DP requirement is denoted as

$$\begin{aligned} \left| \log \left( \frac{\mathbb{P}[\psi_{\mathcal{M}_{i,t}^S}(\mathbf{s}_{i,t}) = x]}{\mathbb{P}[\psi_{\mathcal{M}_{i,t}^S}(\mathbf{s}'_{i,t}) = x]} \right) \right| &\leq \varepsilon \text{ for all } \mathbf{s}_{i,t}, \mathbf{s}'_{i,t} \in \mathcal{S} \text{ s.t. } \|\mathbf{s}_{i,t} - \mathbf{s}'_{i,t}\|_1 \leq 1 \\ \left| \log \left( \frac{\mathbb{P}[\psi_{\mathcal{M}_{i,t}^N}(\mathbf{s}_{i,t}, \{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}) = x]}{\mathbb{P}[\psi_{\mathcal{M}_{i,t}^N}(\mathbf{s}'_{i,t}, \{\boldsymbol{\nu}'_{j,t-1}\}_{j \in \mathcal{N}_i}) = x]} \right) \right| &\leq \varepsilon \text{ for all } (\mathbf{s}_{i,t}, \{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}), (\mathbf{s}'_{i,t}, \{\boldsymbol{\nu}'_{j,t-1}\}_{j \in \mathcal{N}_i}) \in \mathcal{S} \times \mathbb{R}^{\deg(i)} \\ &\text{s.t. } \left\| (\mathbf{s}_{i,t}, \{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}) - (\mathbf{s}'_{i,t}, \{\boldsymbol{\nu}'_{j,t-1}\}_{j \in \mathcal{N}_i}) \right\|_1 \leq 1 \end{aligned}$$

for all  $x \in \mathbb{R}$ .

**Central vs. Local Privacy.** In a local privacy scheme, the DP noise of the measurements can occur at the agent level by adding noise to the collected signals. Noise may be added to the signals after measurement to protect them against the revelations of belief exchange. The central scheme assumes a trusted environment where signal measurements can be collected without privacy concerns, but to the extent that protecting signals from the revelations of beliefs exchange is concerned, these methods would be equivalent.

### 3 Minimum Variance Unbiased Estimation

#### 3.1 Minimum Variance Unbiased Estimation with Signal DP

We present our first algorithm, which considers Minimum Variance Unbiased Estimation (MVUE). In this task, we aim to learn the MVUE of  $\mathbf{m}_\theta$ , i.e., to construct the estimate

$\widehat{\mathbf{m}}_\theta = (1/n) \sum_{i=1}^n \xi(\mathbf{s}_i)$  through local information exchange. The non-private version of this algorithm is presented in Algorithm 1 according to which, the agents start with some private signals  $\{\mathbf{s}_i\}_{i \in [n]}$ , calculate the sufficient statistics  $\{\xi(\mathbf{s}_i)\}_{i \in [n]}$  and then exchange these initial estimates with their local neighbors. Algorithm 1 converges to  $\widehat{\mathbf{m}}_\theta$  in  $t = O\left(\frac{\log(nM_n/\rho)}{\log(1/\beta^*)}\right)$  steps to  $\rho$ -accuracy, which depends on the number of nodes  $n$ , the maximum absolute value  $M_n$  of the sufficient statistics, and the spectral gap  $\beta^*$  of the transition matrix  $A$ .

In its DP version, the algorithm proceeds similarly to the non-DP case, except each agent  $i$  adds noise  $\mathbf{d}_i$  to their sufficient statistic  $\xi(\mathbf{s}_i)$ . As we show later, to respect  $\varepsilon$ -DP, the noise  $\mathbf{d}_i$  depends on the agent's realized signal  $\mathbf{s}_i$ , the sufficient statistics  $\xi(\cdot)$ , and the privacy budget  $\varepsilon$ . We provide the algorithm for the differentially private in Algorithm 3:

Algorithm 3: Minimum Variance Unbiased Estimation with Signal/Network DP

The agents initialize with  $\boldsymbol{\nu}_{i,0} = \xi(\mathbf{s}_i) + \mathbf{d}_i$  where  $\mathbf{d}_i \sim \mathcal{D}_i$  ( $\mathcal{D}_i$  is an appropriately chosen noise distribution), and in any future time period the agents communicate their values and update them according to the following rule:

$$\boldsymbol{\nu}_{i,t} = a_{ii} \boldsymbol{\nu}_{i,t-1} + \sum_{j \in \mathcal{N}_i} a_{ij} \boldsymbol{\nu}_{j,t-1}. \quad (7)$$

Regarding convergence, in Theorem 3.1, we prove that the convergence error is incurred due to two sources. The first source of error is the error due to the omniscient observer, which is the same as in the non-DP case, and the second source of error is incurred due to the privacy noise. Briefly, the latter term can be roughly decomposed to correspond to two terms: the former term is due to estimating the minimum variance unbiased estimator due to the noise, i.e.  $(1/n) \sum_{i=1}^n \mathbf{d}_i$  and is vanishing with a rate proportional to  $\log(\sum_{i=1}^n \mathbb{V}[\mathbf{d}_i]) / \log(1/\beta^*)$ , and an additional non-vanishing term which is due to the mean squared error of  $(1/n) \sum_{i=1}^n \mathbf{d}_i$  which corresponds to the sum of the variances of the signals.

To minimize the convergence error, it suffices to minimize each variance  $\mathbb{V}[\mathbf{d}_i]$  subject to  $\varepsilon$ -DP constraints. By following recent results on the DP literature (cf. Koufogiannis et al. (2015)) we deduce that the variance minimizing distribution under  $\varepsilon$ -DP constraints are the Laplace distributions with parameters  $\Delta/\varepsilon$ , where  $\Delta$  corresponds to the signal sensitivity (see below) and  $\varepsilon$  is the privacy budget. We present our Theorem (proved in Appendix A.1):

**Theorem 3.1** (Minimum Variance Unbiased Estimation with Signal DP). *The following hold for Algorithm 3:*

1. For all  $t$  and any zero-mean zero-mean distributions

$$\mathbb{E} [\|\bar{\mathbf{v}}_t - \mathbb{1}\widehat{\mathbf{m}}_\theta\|_2] \leq (1 + \sqrt{(n-1)}(\beta^*)^t) \sqrt{\sum_{j=1}^n \mathbb{V}[\mathbf{d}_j]} + \sqrt{n(n-1)}(\beta^*)^t M_n, \quad (8)$$

where  $M_n = \max_{i \in [n]} |\xi(\mathbf{s}_i)|$  and  $\beta^* = \max\{\lambda_2(A), |\lambda_n(A)|\}$ .

2. The optimal distributions  $\{\mathcal{D}_i^*\}_{i \in [n]}$  that minimize the MSE for each agent are the Laplace distribution with parameters  $\Delta/\varepsilon$  where  $\Delta$  is the global sensitivity of  $\xi$ . Subsequently,  $\text{TE}(\mathcal{M}^S) = O(n(\beta^*)^t (M_n + \frac{\Delta}{\varepsilon}) + \frac{\Delta}{\varepsilon})$ .

*Proof Sketch.* We do an eigendecomposition on  $A$  and prove that since the noise is independent among agents, we can show that  $\mathbb{E} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2] = \sum_{j=1}^n \mathbb{V}[\mathbf{d}_j] \sum_{i=1}^n \lambda_i^{2t}(A) \mathbf{q}_{ij}^2$ . By using Jensen's inequality, we can show that  $\mathbb{E} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2] \leq \sqrt{\sum_{j=1}^n \mathbb{V}[\mathbf{d}_j]} (1 + \sqrt{n-1}(\beta^*)^t)$ . To minimize the convergence time, it suffices to minimize the variance of  $\mathbf{d}_j \sim \mathcal{D}_j$  subject to privacy constraints for all  $j \in [n]$ , resulting in an optimization problem which we resolve by applying the main result of Koufogiannis et al. (2015).  $\square$

Finally, we note that similarly, in the multi-dimensional case, the sensitivity would be  $\Delta = \max_{\mathbf{s} \in \mathcal{S}} \|\nabla_{\mathbf{s}} \xi(\mathbf{s})\|_\infty$ .

**A Note about Global Sensitivity of  $\xi$ .** In the above Theorem, it is tempting to think that the local sensitivity of each agent, i.e.,  $\Delta \xi_i = |d\xi(\mathbf{s}_i)/d\mathbf{s}_i|$  can be used to calibrate the noise distribution that preserves differential privacy and has the minimum variance. However, as it has been shown in Nissim et al. (2007), releasing noise that depends on the local sensitivity can compromise the signal. However, there are cases where global sensitivity is unsuitable for the learning task. For instance, in many distributions, such as the log-normal distribution, the global sensitivity may be unbounded (e.g.,  $\xi(\mathbf{s}) = \log \mathbf{s}$  for the log-normal and the sensitivity is  $+\infty$  in this case). A possible solution for this situation and many exponential family distributions is to use another sensitivity instead of the global sensitivity. (Nissim et al., 2007, Definition 2.2) proposes the use of the  $\gamma$ -smooth sensitivity. The way of constructing the  $\gamma$ -smooth sensitivity is to calculate

$$S_{\xi, \gamma}^*(\mathbf{s}) = \max_{k > 0} \left\{ e^{-\gamma k} \max_{\mathbf{s}': \|\mathbf{s}' - \mathbf{s}\|_1 = k} |\xi(\mathbf{s}') - \xi(\mathbf{s})| \right\}.$$

Using the Laplace mechanism with parameters  $2S_{\xi, \gamma}^*(\mathbf{s})/\varepsilon$  for  $\gamma = \varepsilon/(2\log(2/\delta))$  would guarantee  $(\varepsilon, \delta)$ -DP; see (Nissim et al., 2007, Corollary 2.4).

For example, we will consider the case of mean estimation in a log-normal distribution with known variance, a common task in many sensor networks, as we argue in Section 5. The sufficient statistic in this case is  $\xi(\mathbf{s}) = \log \mathbf{s}$ , and the local sensitivity at distance  $k$  can be computed to be  $k/\mathbf{s}$ . Therefore, the smooth sensitivity is (note the global sensitivity in this case is unbounded):

$$S_{\xi, \gamma}^*(\mathbf{s}) = \max_{k>0} \left\{ e^{-\beta k} \frac{k}{\mathbf{s}} \right\} = \frac{1}{e\gamma\mathbf{s}} = \frac{2\log(2/\delta)}{e\epsilon\mathbf{s}}, \quad \text{for } \mathbf{s} > 0. \quad (9)$$

If agents have access to several signals, and the exact formula of  $\xi$  is not known, the sensitivity can still be approximated via samples as shown in Wood and Zhang (1996).

**Remark for Network DP.** Note that to achieve Network DP, one natural algorithm is to add noise  $\mathbf{d}_{i,t}$  at each round of Algorithm 3 at a high enough level to protect both the network neighborhoods and the private signals. The issue with this algorithm is that it is divergent, i.e.,  $\mathbb{E} \left[ \|\bar{\nu}_t\|_2^2 \right] \rightarrow \infty$ , because the estimate at time  $t$  is  $\bar{\nu}_t = A^t \bar{\xi} + \sum_{\tau=0}^{t-1} A^\tau \bar{\mathbf{d}}_{t-\tau}$ , and its mean squared error, i.e.,  $\mathbb{E} \left[ \left\| \sum_{\tau=0}^{t-1} A^\tau \bar{\mathbf{d}}_{t-\tau} \right\|_2^2 \right]$ , grows linearly with  $n$  and  $t$ . To avoid the accumulation of the DP noise, we should limit the DP noising to the initial step, which will achieve a bounded error because the mixing matrix,  $A$ , is doubly stochastic. To this end, we choose the noise level to satisfy  $\epsilon$ -DP with the aim of protecting both signals and network connections — and run Algorithm 3 at the new noise level. The error of this algorithm will be identical to the error bound of Theorem 3.1. However, the sensitivity of the noise should be set to accommodate both network and signal dependencies as follows:  $\Delta \nu_i^{\mathcal{M}^N} = \max \{ \Delta, \max_{j \in \mathcal{N}_i} a_{ij} \}$ , and the optimal distributions will be  $\mathcal{D}_i^* = \text{Lap}(\max \{ \Delta, \max_{j \in \mathcal{N}_i} a_{ij} \} / \epsilon)$ ; see Corollary 3.2. In the case that  $\Delta$  is unbounded, we can replace  $\Delta$  with the smooth sensitivity accounting for the network effects, i.e.,  $\max \left\{ \max_{j \neq i} a_{ij}, S_{\xi, \gamma}^*(\mathbf{s}_i) \right\}$  for each signal  $\mathbf{s}_i$  and get an  $(\epsilon, \delta)$ -DP algorithm. Note that private signals will remain DP-protected by the post-processing immunity of the Laplace mechanism. In Appendix A.2, we use induction to show that Network DP is preserved at the  $\epsilon$  level for all times  $t$  when the mixing matrix  $A$  is non-singular. The following corollary summarizes our results:

**Corollary 3.2** (Total Error of Minimum Variance Unbiased Estimation with Network DP).

*Assume the mixing matrix  $A$  is non-singular, let  $\Delta$  be the global sensitivity of  $\xi$ , and  $a =$*

$\max_{i \neq j} a_{ij}$ . The Total Error of Algorithm 3 with Network DP satisfies

$$\text{TE}(\mathcal{M}^N) = O\left(n(\beta^*)^t \left(M_n + \frac{\max\{a, \Delta\}}{\varepsilon}\right) + \frac{\max\{a, \Delta\}}{\varepsilon}\right).$$

**Tightness of Analysis.** We note that the above analysis is tight because there exists an instance for which the upper bound is precisely achieved (for any noise distribution). To show this, consider the complete graph  $K_n$ , which corresponds to weights  $a_{ij} = 1/n$  for all  $i, j \in [n]$  with a spectral gap of zero, and  $\mathbf{s}_i \sim \mathcal{N}(0, 1)$ . The network dynamics on the complete graph converge in one iteration, and it is straightforward to show that the total error (TE) converges to  $1/\varepsilon$  as  $n \rightarrow \infty$  for any noise distributions subject to  $\varepsilon$ -DP constraints (by the central limit theorem), i.e., Equation (A.1) in the Supplementary Materials holds with equality.

## 4 Online Learning of Expected Values

### 4.1 Privacy Frameworks

We consider the online learning framework where the agents aim to learn the common expected value  $\mathbf{m}_\theta = \mathbb{E}_\theta[\xi(\mathbf{s})]$  of the sufficient statistics of their signal distributions. In this regime, the agents observe signals  $\mathbf{s}_{i,t}$  at every time  $t \geq 1$  and update their estimates  $\boldsymbol{\nu}_{i,t}$  by weighing the information content of their most recent private signals,  $\xi(\mathbf{s}_{i,t})$ , their previous estimates,  $\boldsymbol{\nu}_{i,t-1}$ , and the estimates of their neighbors,  $\{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}$ . The mechanisms that we analyze in this section will accommodate two types of privacy needs with convergence and  $\varepsilon$ -DP guarantees for the agents:

$\mathcal{M}^S$ : *Signal DP* (Algorithm 4). Here the agent adds noise to privatize their belief  $\boldsymbol{\nu}_{i,t}$  with respect to their signal. To assert the consistency of the estimator, the agent averages her previous estimate and the estimates of her neighbors with weight  $(t-1)/t$  and her signal with weight  $1/t$ , similarly to Algorithm 2. Therefore, the local sensitivity of this mechanism is the sensitivity of the sufficient statistic weighted by  $1/t$  and equals  $\Delta \nu_{i,t}^{\mathcal{M}^S} = \frac{\Delta}{t}$ .

$\mathcal{M}^N$ : *Network DP* (Algorithm 5). Here we consider the protection of the agent's signals  $\mathbf{s}_{i,t}$  together with their local neighborhood, namely the neighboring beliefs  $\{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}$ . We note that when deciding the noise level at time  $t$ , we do not need to include the

agent's own belief from time  $t - 1$  in the DP protection;  $\nu_{i,t-1}$  including all the private signals up to and including at time  $t - 1$  remain protected by the property of the post-processing immunity. If we were to use *Algorithm 4* for Network DP the sensitivity of the mechanism would be  $\max \left\{ \frac{\Delta}{t}, \frac{t-1}{t} \max_{j \in \mathcal{N}_i} a_{ij} \right\}$  which approaches 1 as  $t \rightarrow \infty$  and violates the consistency of the estimator  $\nu_{i,t}$ . For this reason, we need to adapt the weighting scheme of Algorithm 4 to be consistent and respect Network DP. We give more details of the altered algorithm (Algorithm 5) in Section 4.3.

**Algorithm 4: Online Learning of Expected Values with Signal DP**

In any time period  $t \geq 1$  the agents observe a signal  $\mathbf{s}_{i,t}$ , and update their estimates according to the following rule:

$$\nu_{i,t} = \frac{t-1}{t} \left( a_{ii} \nu_{i,t-1} + \sum_{j \in \mathcal{N}_i} a_{ij} \nu_{j,t-1} \right) + \frac{1}{t} (\xi(\mathbf{s}_{i,t}) + \mathbf{d}_{i,t}), \quad (10)$$

where  $\mathbf{d}_{i,t} \sim \mathcal{D}_{i,t}$  is appropriately chosen noise.

## 4.2 Online Learning of Expected Values with Signal DP

Here we will analyze the performance of Algorithm 4 where agents only protect their signals and present the corresponding error analysis. The error of the estimates  $\nu_{i,t}$  compared to the expected value  $\mathbf{m}_\theta = \mathbb{E}_\theta [\xi(\mathbf{s})]$ , again consists of two terms; one due to decentralization and the statistics of the signals themselves (CoD), and another due to the variances of the added DP noise variables (CoP). Each of these two terms decays at a rate of  $1/\sqrt{t}$ . They, in turn, can be bounded by the sum of two terms: a constant that is due to the principal eigenvalue of  $A$  and represents the convergence of the sample average to  $\mathbf{m}_\theta$ , and  $n - 1$  terms due to  $|\lambda_i(A)|$  for  $2 \leq i \leq n$  which dictate the convergence of the estimates  $\nu_{i,t}$  to their sample average. The latter depends on the number of nodes  $n$  and the spectral gap  $\beta^*$  of matrix  $A$ . We formalize our results as follows (proved in Appendix A.3):

**Theorem 4.1** (Online Learning of Expected Values with Signal DP). *The following hold for Algorithm 4 and mechanism  $\mathcal{M}^S$ :*

1. For every time  $t$ , and all distributions  $\{\mathcal{D}_{i,t}\}_{i \in [n], t \geq 1}$  we have that

$$\mathbb{E} [\|\bar{\mathbf{v}}_t - \mathbf{1}\mathbf{m}_\theta\|_2] \leq \frac{1}{t} \left( \sqrt{nt\mathbb{V}[\xi(\mathbf{s})]} + \sqrt{\sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}]} \right) \left( 1 + \sqrt{\frac{n-1}{1-(\beta^*)^2}} \right).$$

2. The optimal distributions  $\{\mathcal{D}_{i,t}^*\}_{i \in [n], t \geq 1}$  are  $\mathcal{D}_{i,t}^* = \text{Lap}(\Delta/\varepsilon)$ , where  $\Delta$  is the global sensitivity. Moreover, we have:  $\text{TE}(\mathcal{M}^S) = O\left(\frac{n}{\sqrt{t}} \left( \sqrt{\mathbb{V}[\xi(\mathbf{s})]} + \frac{\Delta}{\varepsilon} \right)\right)$ .

*Proof Sketch.* We note that  $\bar{\mathbf{v}}_t = \frac{1}{t} \sum_{\tau=0}^{t-1} A^\tau (\bar{\boldsymbol{\xi}}_{t-\tau} + \bar{\mathbf{d}}_{t-\tau})$ . We decompose  $A$  as  $A = Q\Lambda Q^T$  where  $\Lambda$  is the eigenvalue matrix and  $Q$  is the orthonormal eigenvector matrix, and upper bound  $\mathbb{E} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2]$  as a weighted sum of powers of the eigenvalues of  $A$  and the sum of the variances of  $\bar{\mathbf{d}}_{t-\tau}$ , i.e.,  $\mathbb{E} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2] \leq \frac{1}{t^2} \sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i^{2\tau}(A) \mathbb{E} [\|\bar{\mathbf{d}}_{t-\tau}\|_2^2]$ . We apply the Cauchy-Schwarz inequality again and bound this term with the sum of all variances across all agents  $j \in [n]$  and all the rounds  $0 \leq \tau \leq t-1$ , plus the sum of the powers of the eigenvalues of  $A$ , i.e.,  $\sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i^{2\tau}(A)$ . The latter term is decomposed into a term that depends on the principal eigenvalue  $\lambda_1(A) = 1$  and  $n-1$  terms that are dominated by the spectral gap  $\beta^*$ . We do the same analysis for  $\bar{\boldsymbol{\mu}}_t - \mathbf{1}\mathbf{m}_\theta$ , and we finally apply Jensen's inequality and the triangle inequality to get the final bound. To optimize the bound, it suffices to minimize the variances  $\mathbb{V}[\mathbf{d}_{i,t-\tau}]$  for all  $0 \leq \tau \leq t-1$  and  $i \in [n]$  subject to  $\varepsilon$ -DP constraints. The optimal noise distributions are derived by solving the same optimization problems as in Theorem 3.1 for all  $i \in [n]$  and all  $t \geq 1$ . Calculating the bound for Laplace noise with parameters  $\Delta/\varepsilon$  gives the upper bound on  $\text{TE}(\mathcal{M}^S)$ .  $\square$

### 4.3 Online Learning of Expected Values with Network DP

Above, we briefly discussed why a learning algorithm that puts weights  $\frac{t-1}{t}$  on the network predictions and  $\frac{1}{t}$  on the private signal would not work (and in fact, the dynamics become divergent in that case). In Algorithm 5, we present a different learning scheme that uses weights  $\frac{1}{t}$  for the private signals and neighboring observations and  $1 - \frac{1}{t}(2 - a_{ii})$  for the previous beliefs of the agent. The motivation behind this learning scheme is that the sensitivity is now going to be  $\Delta \nu_{i,t}^{\mathcal{M}^N} = \frac{1}{t} \max \{\max_{j \in \mathcal{N}_i} a_{ij}, \Delta\}$  which goes to zero, instead of  $\max \{\frac{t-1}{t} \max_{j \in \mathcal{N}_i} a_{ij}, \frac{1}{t} \Delta\}$  which approaches 1. The drawback is that the added self-weight on one's own beliefs at every time step will slow down the mixing time and convergence. In the sequel, we present Algorithm 5 and analyze its performance.

Algorithm 5: Online Learning of Expected Values with Network DP

In any time period  $t \geq 1$  the agents observe a signal  $\mathbf{s}_{i,t}$  and update their estimates according to the following rule:

$$\boldsymbol{\nu}_{i,t} = \left(1 - \frac{1}{t}(2 - a_{ii})\right) \boldsymbol{\nu}_{i,t-1} + \frac{1}{t} \left( \sum_{j \in \mathcal{N}_i} a_{ij} \boldsymbol{\nu}_{j,t-1} + \xi(\mathbf{s}_{i,t}) \right) + \frac{1}{t} \mathbf{d}_{i,t}, \quad (11)$$

where  $\mathbf{d}_{i,t} \sim \mathcal{D}_{i,t}$  is appropriately chosen noise.

We can write the above system in matrix notation as  $\bar{\boldsymbol{\nu}}_t = (B(t) - \frac{1}{t}I) \bar{\boldsymbol{\nu}}_{t-1} + \frac{1}{t}\bar{\boldsymbol{\xi}}_t + \frac{1}{t}\bar{\mathbf{d}}_t$  where  $b_{ii}(t) = 1 - \frac{1}{t}(1 - a_{ii})$  and  $b_{ij}(t) = \frac{1}{t}a_{ij}$  for all  $j \neq i$ . First, we study the convergence of Algorithm 5 when no noise is added, i.e.,  $\bar{\mathbf{d}}_t = 0$ . Note that similarly to Algorithm 4, the error term is comprised of two terms, one owing to the principal eigenvalues of  $C(t) = B(t) - \frac{1}{t}I$ , i.e.,  $\lambda_1(C(t)) = 1 - 1/t$  which controls the convergence of the sample average of the estimates to  $\mathbf{m}_\theta$ , and  $n-1$  terms due to  $|\lambda_i(C(t))|$  which control the convergence of the estimates  $\boldsymbol{\nu}_{i,t}$  to their sample average.

**Theorem 4.2** (Online Learning of Expected Values with Network DP). *For Algorithm 5, the following hold:*

1. For all  $t \geq 1$  and all distributions  $\{\mathcal{D}_{i,t}\}_{i \in [n], t \geq 1}$ , we have that

$$\mathbb{E} [\|\bar{\boldsymbol{\nu}}_t - \mathbf{1}\mathbf{m}_\theta\|_2] \leq \frac{1}{t} \left( \sqrt{nt\mathbb{V}[\xi(\mathbf{s})]} + \sqrt{\sum_{i=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{i,t-\tau}]} \right) \left( 1 + \sqrt{\frac{n-1}{3-2\beta^*}} \right).$$

2. The optimal distributions  $\{\mathcal{D}_{i,t}^*\}_{i \in [n], t \geq 1}$  that minimize the MSE bound subject to  $\varepsilon$ -DP are the Laplace Distributions with parameters  $\max\{\max_{j \in \mathcal{N}_i} a_{ij}, \Delta\}/\varepsilon$ . Moreover, if  $\Delta$  is the global sensitivity and  $a = \max_{i \neq j} a_{ij}$ , then

$$\text{TE}(\mathcal{M}^N) = O \left( \frac{n}{\sqrt{t}} \left( \frac{\max\{a, \Delta\}}{\varepsilon} + \sqrt{\mathbb{V}[\xi(\mathbf{s})]} \right) \right).$$

*Proof Sketch.* Our proof (Appendix A.4) follows a similar analysis to Theorem 4.1. We first note that  $C(t)$  can be written as  $C(t) = \frac{t-2}{t}I + \frac{1}{t}A$ , and therefore shares the same eigenvectors with  $A$  and has eigenvalues  $\lambda_i(C(t)) = \frac{t-2}{t} + \frac{1}{t}\lambda_i(A)$ . We define  $\Phi(t) = \prod_{\tau=0}^{t-1} C(\tau)$ , and show that  $\lambda_i(\Phi(t)) \leq t^{\lambda_i(A)-2}$ . Moreover,  $\bar{\boldsymbol{\nu}}_t = \frac{1}{t} \sum_{\tau=0}^{t-1} \Phi(\tau)(\bar{\boldsymbol{\xi}}_{t-\tau} + \bar{\mathbf{d}}_{t-\tau})$ .

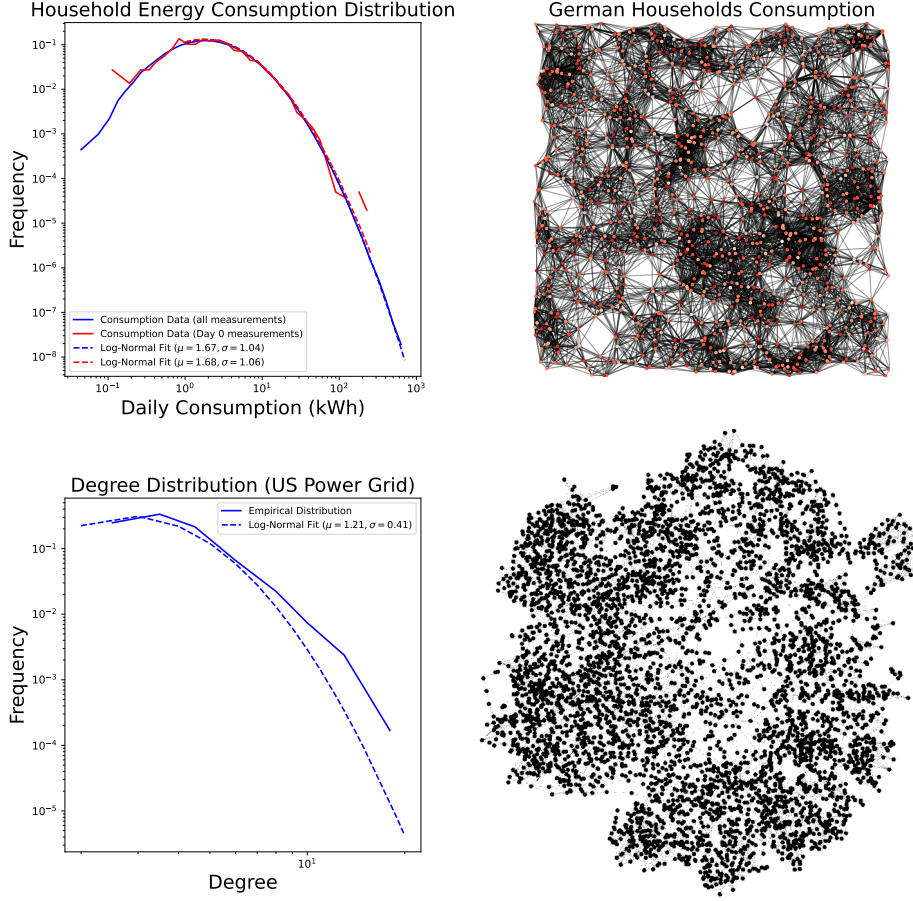


Figure 2: Distribution of Daily Consumption (in kWh) for the GEM House openData with log-normal fits is shown on the left (for all measurements and Day 0 measurements), followed by a visualization of the generated random geometric network with  $\rho = 0.1$ . The next two figures show the US Power Network degree distribution with log-normal fit, followed by its visualization.

By using the bound on the eigenvalues of  $\Phi(\tau)$  for  $0 \leq \tau \leq t-1$  and by applying the same analysis as in Theorem 4.1, we decompose  $\mathbb{E} \left[ \|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2 \right]$  and get that the sum of the powers of the eigenvalues consists of a term due to the principal eigenvalues, which decays with the rate  $1/t^2$  and  $n-1$  terms that decay as  $\frac{n-1}{t^2(3-2\beta^*)}$ . We finally deduce that  $\mathbb{E} [\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2] \leq \frac{1}{t} \sqrt{\sum_{i=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{i,t-\tau}]} \left( 1 + \sqrt{\frac{n-1}{3-2\beta^*}} \right)$ . The same bound holds for  $\bar{\mathbf{q}}_t = \bar{\boldsymbol{\mu}}_t - \mathbf{1}\mathbf{m}_\theta$ , and by applying Jensen's inequality and the triangle inequality, we get the error bound. The optimization of the noise variables follows similar logic to Theorem 4.1, with a different sensitivity.  $\square$

**Tightness of analysis.** We note that the analysis is tight, and the tight example is precisely the same as the example we provided for MVUE.

## 5 Real-World Experiments

**Datasets and Real-World Scenarios.** To showcase the effectiveness of our algorithms, i.e., convergence to the actual estimates subject to  $(\varepsilon, \delta)$ -DP, we conduct two experiments that correspond to the estimation of power consumption in the electric grid.

The first case considers estimating power consumption via electricity measurements of individual households. Consumption behavior is considered highly sensitive. It can reveal compromising information about daily habits and family illnesses or pose a security threat if exploited by an adversary, e.g., to coordinate attack time. Here, we assume that each household faces a privacy risk in sharing their measurements, and they may decide to mitigate the privacy risks by adding noise to their estimates. The ability to estimate average consumption in a distributed manner is useful for distributed load balancing and deciding generation plans.

For this scenario, we consider the GEM House openData dataset Milojkovic (2018), which contains power consumption measurements of  $n = 969$  individual households over  $T = 1096$  days (i.e., three years). The dataset contains cumulative power consumption measurements  $c_{i,t}$  for each particular day. To extract the actual measurements (in kWh)  $s_{i,t}$  we take the differences between consecutive days and divide by  $10^{10}$  (see Section 3.2.1 of Milojkovic (2018)), i.e.,  $s_{i,t} = \frac{c_{i,t} - c_{i,t-1}}{10^{10}}$ . We observe that  $s_{i,t}$  follows a log-normal distribution with mean  $\mu = 1.67$  and standard deviation  $\sigma = 1.04$ , as shown in Figure 2. Moreover, in this dataset, the network structure is absent. For this reason, we generate a random geometric graph, i.e., we generate  $n = 969$  nodes randomly distributed in  $[0, 1]^2$  and connect nodes with a distance at most  $\rho = 0.1$ . Random geometric graphs have been used to model sensor networks Kenniche and Ravelomananana (2010) and correspond to a straightforward criterion for determining links since they connect nearby households. For a fixed random seed (seed = 0), the network contains  $m = 13,236$  edges and is visualized in Figure 2.

The second dataset examines estimating power consumption in the US Power Grid Network from Watts and Strogatz (1998). In this case, we hypothesize that each power station faces a privacy risk – for example, vulnerability to a cyber attack – in sharing their measurements and decides to reduce its privacy risk by adding noise. The network contains  $n = 4,941$  nodes and  $m = 6,594$  edges. Figure 2 shows the power network and its degree distribution. Here we artificially generate i.i.d. signals for  $T = 100$  as

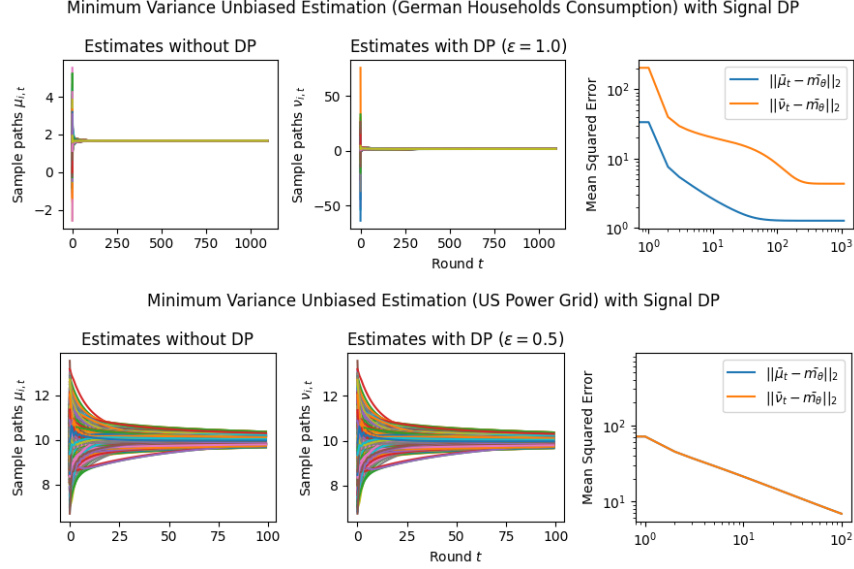


Figure 3: Sample Paths for MVUE with Signal DP. Note the large error in the case of the German household dataset is because protecting households with low (near zero) consumption rates even at a relatively high privacy budget ( $\epsilon = 10$ ) comes at a huge cost to accuracy.

$\mathbf{s}_{i,t} \sim \text{LogNormal}(\mu = 10, \sigma = 1)$ .

**Estimation Tasks.** In both cases, the task is to estimate each log-normal distribution’s mean  $\mu$  in two scenarios. In the first scenario, we estimate the mean only from the initial measurements, i.e., estimate  $\hat{\mu}_{\text{MVUE}} = \frac{\sum_{i=1}^n \log \mathbf{s}_{i,1}}{n}$ . Figure 3 presents some sample paths for this task as the horizon  $t$  varies, and Figure 7 presents the final MSE after  $T$  for various values of the privacy budget  $\epsilon$ . In the second scenario, we estimate the mean with online learning (OL) to estimate  $\hat{\mu}_{\text{OL}} = \frac{\sum_{i=1}^n \sum_{t=1}^T \log \mathbf{s}_{i,t}}{nT} \rightarrow \mathbb{E}[\log \mathbf{s}] = \mu_{\text{OL}}$ . We run simulations in both regimes where we want to protect the signal and the network connections. Because in this case the global sensitivity is unbounded, we use the smooth signal sensitivities  $S_{\xi,\gamma}^*(\mathbf{s}_{i,t}) = \frac{2 \log(2/\delta)}{\epsilon \epsilon \mathbf{s}_{i,t}}$  for each signal  $\mathbf{s}_{i,t}$  with  $\delta = 0.01$ . The resulting algorithms are  $(\epsilon, \delta)$ -DP (see Section 3.1).

**Comparison with Rizk et al. (2023).** Finally, in Appendix B we explain the adaptation of Rizk et al. (2023) first-order DP consensus algorithm to both MVUE and OL tasks. Figure 7 shows the comparison of MSE performances between our algorithms and Rizk et al. (2023) given the same privacy budget per agent. Compared to Rizk et al. (2023), our method is able to achieve significantly smaller MSE under the same total privacy budget;  $\approx 1000\times$  smaller for both datasets. While Rizk et al. (2023)’s algorithm are applicable to a broader set of tasks than the MVUE and OL estimation setups presented here, their

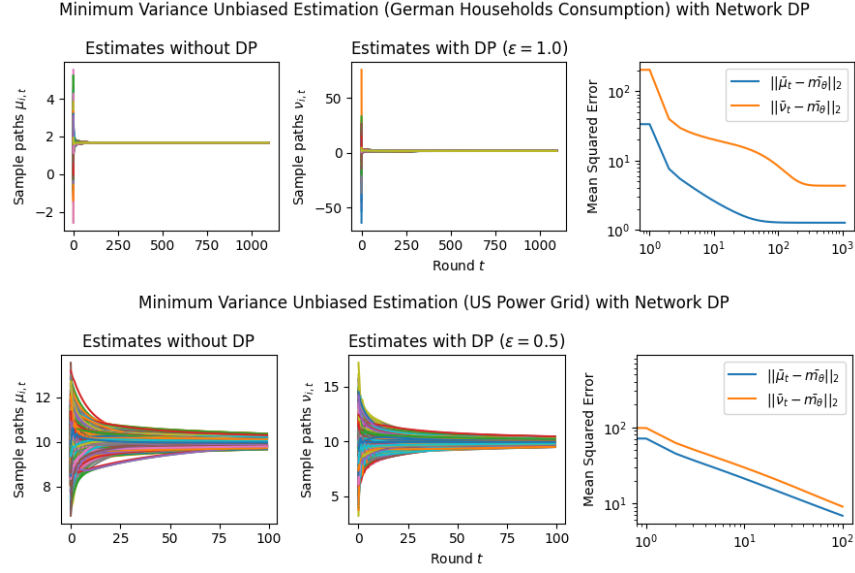


Figure 4: Sample Paths for MVUE with Network DP. Note that even requiring a moderate accuracy in the case of the German household dataset comes at a high cost to privacy ( $\epsilon = 1$ ), pointing to the challenges of maintaining privacy when sensitivities cannot be locally bounded (some household consumption values are close to zero).

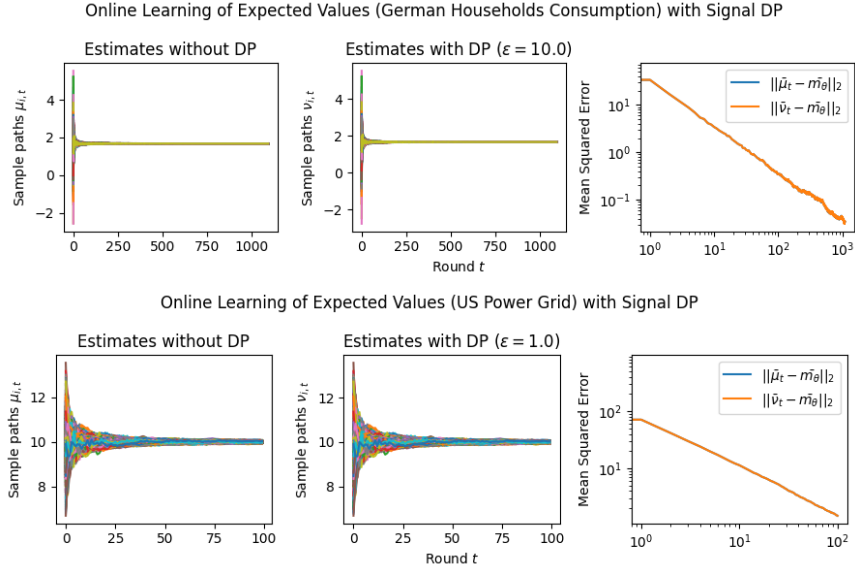


Figure 5: Sample Paths for Online Learning of Expected Values with Signal DP. Choosing  $\epsilon$  large enough leads to a convergent behavior for the German household dataset, but no meaningful privacy protection can be afforded in that case ( $\epsilon = 10$ ).

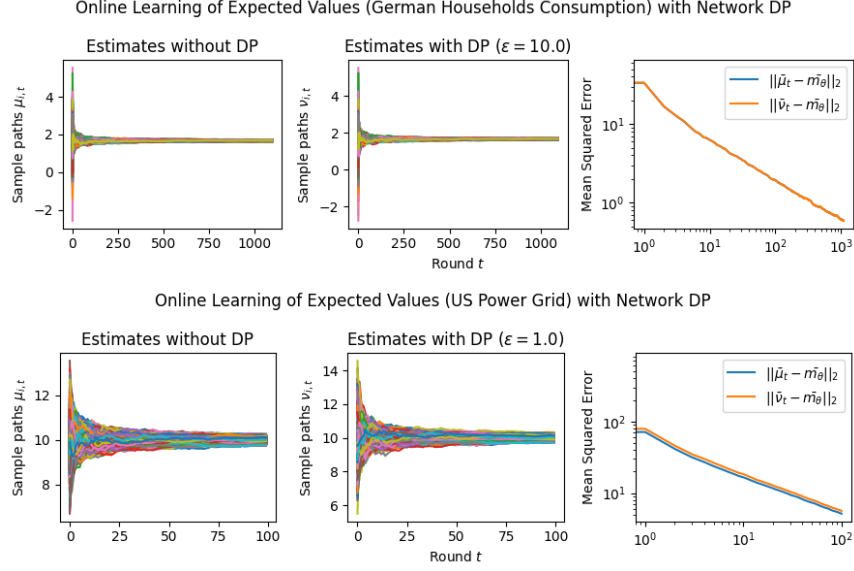


Figure 6: Sample Paths for the Online Learning of Expected Values with Network DP. Protecting network neighborhoods is a harder task than protecting private signals. While almost perfect signal DP can be achieved with reasonable accuracy for the US power grid network ( $\epsilon = 1$  in Figure 5), even moderate protection of network neighborhoods ( $\epsilon = 1$ ) come at a noticeable cost to accuracy. Privacy protection for network neighborhoods in the case of German households is further complicated by the existence of almost zero signals with locally unbounded sensitivity and no meaningful protections is accomplished ( $\epsilon = 10$ ). Privacy and accuracy, in this case, become conflicting criteria that cannot be reconciled.

inclusion of private signals at every iteration entails DP noising at every step of the iteration and comes at a higher cost to accuracy.

## 6 Discussion and Conclusion

**Results and Insights.** In all cases of signal DP with the US power grid network, the DP noise did not affect the convergence rate in practice for this choice of signals, privacy budget  $\epsilon$ , and information leakage probability  $\delta$ . Also, we observe that Algorithm 4 converges faster than Algorithm 5 (even in the absence of DP noise) because of the underlying mixing matrices, which are  $\frac{t-1}{t}A$ , and  $C(t) = \frac{t-2}{t}I + \frac{1}{t}A$  respectively. Moreover, both of these algorithms converge faster (with and without DP noise) than Algorithm 3. This is expected since Algorithm 3 has access to  $n$  samples in total, while Algorithms 4 and 5 have access to  $nt$  signals and can bring the estimation error down by a  $1/t$  factor. Comparison of Signal DP (Figures 3 and 5) with Network DP (Figures 4 and 6) for MVUE and online learning tasks points to the increased difficulty of ensuring network DP: network privacy

protections are harder to achieve and they imply signal protections automatically. On the other hand, when the local sensitivities can grow large — as with the German household dataset — maintaining privacy for households with low consumption comes at a huge cost to accuracy (see, e.g., Figure 5). This is because for the log-normal distribution,  $d\xi(\mathbf{s})/d\mathbf{s}$  grows unbounded as  $\mathbf{s} \rightarrow 0$ .

**Extensions.** We extend our algorithms to address additional forms of heterogeneity. Specifically, in Appendix C, we show how our algorithms can provably converge under minimal assumptions when the network topology is changing dynamically (Appendix C.1) and when the corresponding topology is directed (Appendix C.2). These scenarios are pertinent to real-world sensor networks since sensors and communications can fail, corresponding to dynamically changing networks with asymmetries (cf. Touri and Nédic (2009)). Moreover, to balance the trade-offs between accuracy and privacy, the agents can resort to heterogeneous privacy budgets  $\{\varepsilon_i\}_{i \in [n]}$  and improve their collective estimation performance while maintaining a minimum privacy protection (capping the individual privacy budgets at  $\varepsilon_{i,\max}$ ). The possibility to accommodate heterogeneous budgets in the local DP setting leads to interesting design choices for improving the collective learning performance, e.g., using personalized DP methods Acharya et al. (2024); Jorgensen et al. (2015). In Appendix C.3, we provide both centralized and decentralized schemes to allocate privacy budgets to optimize their collective accuracy subject to individual privacy budget caps and test their performance on the German Households dataset (cf. Supplementary Figure 3).

**Conclusion.** Our paper focuses on distributed estimation and learning in a networked environment subject to privacy constraints. The aim is to estimate the statistical properties of unknown random variables based on observed data. Our aggregation methods aim to combine the observed data efficiently without requiring explicit coordination beyond the local neighborhood of each agent. This allows for estimating a complete sufficient statistic using either offline or online signals provided to the agents. To preserve privacy, agents add noise to their estimates, adhering to a differential privacy budget ( $\varepsilon$ -DP) to safeguard the privacy of either their signals (signal DP) or their signals and network neighborhoods (Network DP). Our algorithms employ linear aggregation schemes that combine the observations of all agents while incorporating the added noise, either online or offline. We prove that the estimation error bounds depend on two terms: the first term corresponds to the error incurred due to the aggregation scheme (which we call Cost of Decentralization) and can be

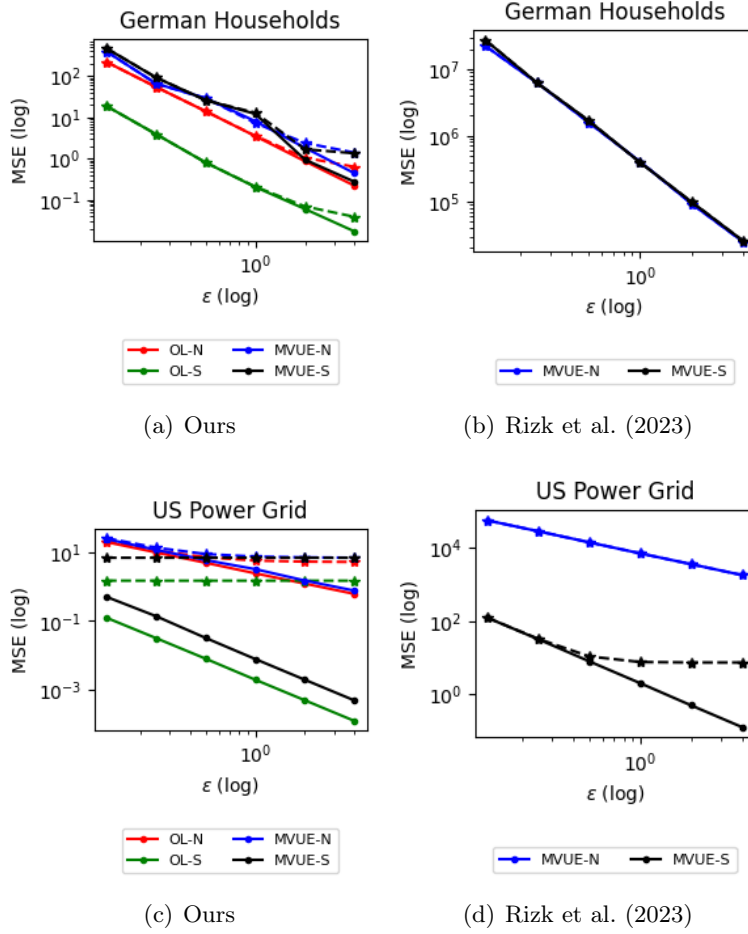


Figure 7: MSE plots vs. varying privacy budget  $\epsilon$  for the German Households dataset and the US Power Grid Dataset. We compare with the first-order method of Rizk et al. (2023) with a learning rate of  $\eta = 0.001$  (see Appendix B). The solid lines represent the CoP, and the dashed lines represent the Total Error.

controlled by the mixing rate of the doubly-stochastic adjacency weights, and the second term corresponds to the error due to the DP noising (which we call Cost of Privacy). We prove that under all cases (see also Section 1.1), the noise distributions that minimize the convergence rate correspond to the Laplace distributions with parameters that depend on the (local or global) signal sensitivities, the network structure, and the differential privacy budget  $\epsilon$ . Finally, we test our algorithms and validate our theory in numerical experiments.

When sensitivities are locally bounded, signal DP can be achieved efficiently with a graceful accuracy loss over a decreasing privacy budget. This is facilitated by the post-processing immunity of DP (Dwork and Roth, 2014, Proposition 2.1) that no future leaks are possible after adequate noising of the private signals and indicates the resilience of linear

aggregation schemes to DP noising. However, achieving network DP with noising of estimates is significantly more challenging, and while individual noisy estimates are protected against a one-time attack, network information can still leak over time across multiple estimates. The composition property (Dwork and Roth, 2014, Chapter 3) implies that we can protect the network neighborhoods at  $\varepsilon$ -DP level against an adversary who eavesdrops  $k$  times by protecting individual estimates at  $\varepsilon/k$ -DP level. Such protection can be challenging if an adversary can eavesdrop on the estimates for a long time or has simultaneous access to estimates of multiple agents. In such cases, a fast convergence rate (using the fastest mixing weights) can limit communications and help agents maintain privacy without completely sacrificing the accuracy of their estimates.

## References

- Acharya, K., Boenisch, F., Naidu, R., and Ziani, J. (2024). Personalized differential privacy for ridge regression. *arXiv preprint arXiv:2401.17127*.
- Alexandru, A. B. and Pappas, G. J. (2021). Private Weighted Sum Aggregation. *IEEE Transactions on Control of Network Systems*, 9(1):219–230.
- Alexandru, A. B., Tsiamis, A., and Pappas, G. J. (2020). Towards Private Data-driven Control. In *IEEE Conference on Decision and Control (CDC 2020)*, pages 5449–5456. IEEE.
- Apple Differential Privacy Team (2017). Learning with privacy at scale. <https://machinelearning.apple.com/research/learning-with-privacy-at-scale>. Accessed: 2023-05-18.
- Atanasov, N., Tron, R., Preciado, V. M., and Pappas, G. J. (2014a). Joint Estimation and Localization in Sensor Networks. *IEEE Conference on Decision and Control (CDC 2014)*, pages 6875–6882.
- Atanasov, N. A., Le Ny, J., and Pappas, G. J. (2014b). Distributed Algorithms for Stochastic Source Seeking with Mobile Robot Networks. *Journal of Dynamic Systems, Measurement, and Control*.
- Aumann, R. J. (1976). Agreeing to Disagree. *The Annals of Statistics*, pages 1236–1239.
- Barbaro, M., Zeller, T., and Hansell, S. (2006). A face is exposed for AOL searcher no. 4417749. *New York Times*, 9(2008):8.
- Bickel, P. J. and Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics*, volume I. CRC Press.
- Bonawitz, K., Kairouz, P., McMahan, B., and Ramage, D. (2021). Federated Learning and Privacy: Building Privacy-preserving Systems for Machine Learning and Data Science on Decentralized Data. *Queue*, 19(5):87–114.
- Borkar, V. and Varaiya, P. (1982). Asymptotic Agreement in Distributed Estimation. *IEEE Transactions on Automatic Control*, 27(3):650–655.
- Boyd, S., Diaconis, P., and Xiao, L. (2004). Fastest Mixing Markov Chain on a Graph. *SIAM Review*, 46(4):667–689.
- Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D. (2005). Gossip Algorithms: Design, Analysis and Applications. In *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005)*, volume 3, pages 1653–1664. IEEE.

- Bullo, F., Cortés, J., and Martinez, S. (2009). *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*, volume 27. Princeton University Press.
- Cardoso, A. R. and Rogers, R. (2022). Differentially Private Histograms under Continual Observation: Streaming Selection into the Unknown. In *International Conference on Artificial Intelligence and Statistics (AISTATS 2022)*, pages 2397–2419. PMLR.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*, volume 2. Duxbury Pacific Grove, CA.
- Chamberland, J.-F. and Veeravalli, V. V. (2003). Decentralized Detection in Sensor Networks. *IEEE Transactions on Signal Processing*, 51(2):407–416.
- Chatterjee, S. (2023). Spectral Gap of Nonreversible Markov Chains. *arXiv preprint arXiv:2310.10876*.
- Chazelle, B. (2011). The Total  $s$ -energy of a Multiagent System. *SIAM Journal on Control and Optimization*, 49(4):1680–1706.
- Dimakis, A. D., Sarwate, A. D., and Wainwright, M. J. (2008). Geographic Gossip: Efficient Averaging for Sensor Networks. *IEEE Transactions on Signal Processing*, 56(3):1205–1216.
- Dwork, C. (2011). A Firm Foundation for Private Data Analysis. *Communications of the ACM*, 54(1):86–95.
- Dwork, C. and Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.
- Erlingsson, U. (2014). Learning statistics with privacy, aided by the flip of a coin. <https://ai.googleblog.com/2014/10/learning-statistics-with-privacy-aided.html>. Accessed: 2023-05-18.
- Geanakoplos, J. D. and Polemarchakis, H. M. (1982). We can’t Disagree Forever. *Journal of Economic Theory*, 28(1):192–200.
- Gowtham, M. and Ahila, S. S. (2017). Privacy Enhanced Data Communication Protocol for Wireless Body Area Network. In *International Conference on Advanced Computing and Communication Systems (ICACCS 2017)*, pages 1–5. IEEE.
- Guevara, M. (2019). Enabling developers and organizations to use differential privacy. <https://developers.googleblog.com/2019/09/enabling-developers-and-organizations.html>. Accessed: 2023-05-18.
- Hassan, M. U., Rehmani, M. H., and Chen, J. (2019). Differential Privacy Techniques for Cyber Physical Systems: a Survey. *IEEE Communications Surveys & Tutorials*, 22(1):746–789.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton, NJ.
- Jadbabaie, A., Lin, J., and Morse, A. S. (2003). Coordination of Groups of Mobile Autonomous Agents using Nearest Neighbor Rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001.
- Jorgensen, Z., Yu, T., and Cormode, G. (2015). Conservative or Liberal? personalized Differential Privacy. In *IEEE International Conference on Data Engineering (ICDE 2015)*, pages 1023–1034. IEEE.
- Kaissis, G. A., Makowski, M. R., Rückert, D., and Braren, R. F. (2020). Secure, Privacy-preserving and Federated Machine Learning in Medical Imaging. *Nature Machine Intelligence*, 2(6):305–311.
- Kar, S., Moura, J., and Ramanan, K. (2012). Distributed parameter estimation in sensor networks: Non-linear observation models and imperfect communication. *IEEE Transactions on Information Theory*, 58, no. 6, pp. 3575–3605.
- Kenniche, H. and Ravelomananana, V. (2010). Random Geometric Graphs as Model of Wireless Sensor Networks. In *International Conference on Computer and Automation Engineering (ICCAE 2010)*, volume 4, pages 103–107. IEEE.
- Kontar, R., Shi, N., Yue, X., Chung, S., Byon, E., Chowdhury, M., Jin, J., Kontar, W., Masoud, N., Nouiehed, M., et al. (2021). The internet of federated things (IoFT). *IEEE Access*, 9:156071–156113.

- Koufogiannis, F., Han, S., and Pappas, G. J. (2015). Optimality of the laplace mechanism in differential privacy. *arXiv preprint arXiv:1504.00065*.
- Koufogiannis, F. and Pappas, G. J. (2017). Diffusing Private Data over Networks. *IEEE Transactions on Control of Network Systems*, 5(3):1027–1037.
- Krishnamurthy, V. and Poor, H. V. (2013). Social learning and bayesian games in multiagent signal processing: How do local and global decision makers interact? *IEEE Signal Processing Magazine*, 30(3):43–57.
- Kumar, R., Novak, J., Pang, B., and Tomkins, A. (2007). On Anonymizing Query Logs via Token-based Hashing. In *International conference on World Wide Web (WWW 2017)*, pages 629–638.
- Lalitha, A., Sarwate, A., and Javidi, T. (2014). Social Learning and Distributed Hypothesis Testing. *IEEE International Symposium on Information Theory (ISIT 2014)*, pages 551–555.
- Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society.
- Li, M., Lou, W., and Ren, K. (2010). Data Security and Privacy in Wireless Body Area Networks. *IEEE Wireless communications*, 17(1):51–58.
- McMahan, B. and Thakurta, A. (2022). Federated learning with formal differential privacy guarantees. <https://blog.research.google/2022/02/federated-learning-with-formal.html>. Accessed: 2024-03-26.
- Mesbahi, M. and Egerstedt, M. (2010). *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press.
- Milojkovic, F. (2018). GEM House Opendata: German Electricity Consumption in Many Households over Three Years 2018–2020 (Fresh Energy).
- Narayanan, A. and Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In *IEEE Symposium on Security and Privacy (SP 2008)*, pages 111–125.
- National Academies of Sciences, Engineering, and Medicine (2023a). Net metering practices should be revised to better reflect the value of integrating distributed electricity generation into the nation’s power grid. <https://www.nationalacademies.org/news/2023/05/net-metering-practices-should-be-revised-to-better-reflect-the-value-of-integrating-distributed-electricity-generation-into-the-nations-power-grid>. Accessed: 2023-05-20.
- National Academies of Sciences, Engineering, and Medicine (2023b). The Role of Net Metering in the Evolving Electricity System. <https://www.nationalacademies.org/our-work/the-role-of-net-metering-in-the-evolving-electricity-system>. Accessed: 2023-05-20.
- Nedić, A., Olshevsky, A., and Uribe, C. A. (2015). Nonasymptotic Convergence Rates for Cooperative Learning over Time-varying Directed Graphs. In *American Control Conference (ACC 2015)*, pages 5884–5889. IEEE.
- Nedić, A., Olshevsky, A., and Uribe, C. A. (2017). Fast Convergence Rates for Distributed non-Bayesian Learning. *IEEE Transactions on Automatic Control*, 62(11):5538–5553.
- Niknam, S., Dhillon, H. S., and Reed, J. H. (2020). Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges. *IEEE Communications Magazine*, 58(6):46–51.
- Nissim, K., Raskhodnikova, S., and Smith, A. (2007). Smooth Sensitivity and Sampling in Private Data Analysis. In *ACM Symposium on Theory of Computing (STOC 2007)*, pages 75–84.
- Olfati-Saber, R. and Shamma, J. (2005). Consensus Filters for Sensor Networks and Distributed Sensor Fusion. *IEEE Conference on Decision and Control (CDC 2005)*, pages 6698 – 6703.
- Olshevsky, A. (2014). Linear Time Average Consensus on Fixed Graphs and Implications for Decentralized Optimization and Multi-agent Control. *arXiv preprint arXiv:1411.4186*.

- Papachristou, M. and Rahimian, M. A. (2024). Group Decision-Making among Privacy-Aware Agents. *AAAI Workshop on Privacy-preserving Artificial Intelligence (PPAI)*.
- Rahimian, M. A. and Jadbabaie, A. (2016a). Bayesian learning without recall. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):592–606.
- Rahimian, M. A. and Jadbabaie, A. (2016b). Distributed Estimation and Learning over Heterogeneous Networks. In *Communication, Control, and Computing (Allerton 2016)*, pages 1314–1321. IEEE.
- Rahimian, M. A. and Jadbabaie, A. (2016c). Group decision making and social learning. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 6783–6794. IEEE.
- Rahimian, M. A., Yu, F.-Y., and Hurtado, C. (2023a). Differentially private network data collection for influence maximization. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 2795–2797.
- Rahimian, M. A., Yu, F.-Y., and Hurtado, C. (2023b). Seeding with differentially private network information. *arXiv preprint arXiv:2305.16590*.
- Rizk, E., Vlaskiy, S., and Sayed, A. H. (2023). Enforcing Privacy in Distributed Learning with Performance Guarantees. *IEEE Transactions on Signal Processing*.
- Rogers, R., Subramaniam, S., Peng, S., Durfee, D., Lee, S., Kancha, S. K., Sahay, S., and Ahammad, P. (2020). LinkedIn’s Audience Engagements API: A Privacy Preserving Data Analytics System at Scale. *arXiv preprint arXiv:2002.05839*.
- Sayed, A. H. et al. (2014). Adaptation, Learning, and Optimization over Networks. *Foundations and Trends® in Machine Learning*, 7(4-5):311–801.
- Seneta, E. (2006). *Non-negative Matrices and Markov Chains*. Springer.
- Shahrampour, S., Rakhlin, A., and Jadbabaie, A. (2015). Distributed detection: Finite-time analysis and impact of network topology. *IEEE Transactions on Automatic Control*, 61(11):3256–3268.
- Shi, N., Lai, F., Al Kontar, R., and Chowdhury, M. (2023). Ensemble Models in Federated Learning for Improved Generalization and Uncertainty Quantification. *IEEE Transactions on Automation Science and Engineering*.
- Sweeney, L. (1997). Weaving Technology and Policy Together to Maintain Confidentiality. *The Journal of Law, Medicine & Ethics*, 25(2-3):98–110.
- Sweeney, L. (2015). Only you, your Doctor, and many others may know. *Technology Science*, 2015092903(9):29.
- Touri, B. and Nedic, A. (2009). Distributed consensus over Network with Noisy Links. In *International Conference on Information Fusion (FUSION 2009)*, pages 146–154. IEEE.
- Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., and Zhou, Y. (2019). A Hybrid Approach to Privacy-preserving Federated Learning. In *ACM Workshop on Artificial Intelligence and Security*, pages 1–11.
- Truong, N., Sun, K., Wang, S., Guitton, F., and Guo, Y. (2021). Privacy preservation in Federated Learning: An insightful Survey from the GDPR Perspective. *Computers & Security*, 110:102402.
- Tsitsiklis, J. N. (1993). Decentralized Detection. *Advances in Statistical Signal Processing*, 2(2):297–344.
- Tsitsiklis, J. N. and Athans, M. (1984). Convergence and asymptotic agreement in distributed decision problems. *Automatic Control, IEEE Transactions on*, 29(1):42–50.
- US Census (2020). 2020 decennial census: Processing the count: Disclosure avoidance modernization. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html>. Accessed: 2023-05-18.

- Wang, Y. and Djuric, P. M. (2015). Social Learning with Bayesian Agents and Random Decision Making. *IEEE Transactions on Signal Processing*, 63(12):3241–3250.
- Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309):63–69.
- Watts, D. J. and Strogatz, S. H. (1998). Collective Dynamics of “Small-world” Networks. *Nature*, 393(6684):440–442.
- Wilson, R. J., Zhang, C. Y., Lam, W., Desfontaines, D., Simmons-Marengo, D., and Gipson, B. (2020). Differentially Private SQL with Bounded User Contribution. *Proceedings on privacy enhancing technologies*, 2020(2):230–250.
- Wood, G. and Zhang, B. (1996). Estimation of the lipschitz constant of a function. *Journal of Global Optimization*, 8:91–103.
- Xiao, L., Boyd, S., and Lall, S. (2005). A Scheme for Robust Distributed Sensor Fusion based on Average Consensus. In *International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pages 63–70.
- Xiao, L., Boyd, S., and Lall, S. (2006). A Space-time Diffusion Scheme for Peer-to-peer Least-squares Estimation. In *International Conference on Information Processing in Sensor Networks (IPSN 2006)*, pages 168–176.
- Xu, Q., Ren, P., Song, H., and Du, Q. (2017). Security-aware Waveforms for Enhancing Wireless Communications Privacy in Cyber-physical systems via Multipath Receptions. *IEEE Internet of Things Journal*, 4(6):1924–1933.
- Yue, X., Kontar, R. A., and Gómez, A. M. E. (2022). Federated Data Analytics: A Study on Linear Models. *IJSE Transactions*, pages 1–25. in-press.
- Zhang, H., Shu, Y., Cheng, P., and Chen, J. (2016). Privacy and Performance Trade-off in Cyber-physical Systems. *IEEE Network*, 30(2):62–66.
- Zhang, X., Chen, X., Hong, M., Wu, Z. S., and Yi, J. (2022). Understanding Clipping for Federated Learning: Convergence and Client-level Differential Privacy. In *International Conference on Machine Learning (ICML 2022)*.

## Acknowledgements

M.P. was partially supported by a LinkedIn Ph.D. Fellowship, an Onassis Fellowship (ID: F ZT 056-1/2023-2024), and grants from the A.G. Leventis Foundation and the Gerondelis Foundation. M.A.R. was partially supported by NSF SaTC-2318844. The authors would like to thank the seminar participants at Rutgers Business School, Jalaj Upadhyay, Saeed Sharifi-Malvajerdi, Jon Kleinberg, Kate Donahue, and Vasilis Charisopoulos for their valuable discussions and feedback.

## Data Availability Statement

The data that support the findings of this study are openly available in the following repositories:

1. **GEM House openData.** URL: <https://dx.doi.org/10.21227/4821-vf03> (Accessed at 1-6-2023), Reference: Milojkovic (2018).
2. **US Power Grid Network.** URL: <https://toreopsahl.com/datasets/#uspowergrid> (Accessed at 1-6-2023), Reference: Watts and Strogatz (1998).

Our code and data can be found at: <https://github.com/papachristoumarios/dp-distributed-estimation>.

## Biographical Sketches

**Marios Papachristou.** Marios Papachristou is a 4th year PhD student at the Department of Computer Science at Cornell University and is advised by Jon Kleinberg. His research interests span the theoretical and applied aspects of social and information networks, exploring their roles within large-scale social and information systems and understanding their wider societal implications. His research is supported by the Onassis Scholarship and has been supported in the past by a LinkedIn Ph.D. Fellowship, a grant from the A.G. Leventis Foundation, a grant from the Gerondelis Foundation, and a Cornell University Fellowship.

**M. Amin Rahimian.** Amin Rahimian has been an assistant professor of industrial engineering at the University of Pittsburgh since 2020, where he leads the sociotechnical systems research lab. Prior to that, he was a postdoc with joint appointments at MIT Institute for Data, Systems, and Society (IDSS) and MIT Sloan School of Management. He received his PhD in Electrical and Systems Engineering from the University of Pennsylvania, and Master's in Statistics from Wharton School. Broadly speaking his works are at the intersection of networks, data, and decision sciences, and have been published in the Proceedings of the National Academy of Sciences, Nature Human Behaviour, Nature Communications, and the Operations Research journal among others. His research interests are in applied probability, applied statistics, algorithms, decision and game theory, with applications ranging from online social networks, public health, and e-commerce to modern civilian cyberinfrastructure and future warfare. His research is currently supported by NSF, CDC, and the Department of the Army.

# Supplementary Material

## A Proofs

### A.1 Proof of Theorem 3.1

*Proof.* Since  $A$  is real and symmetric, we can do an eigen decomposition of  $A$  as  $A = Q\Lambda Q^T$  where  $Q$  is an orthonormal eigenvector matrix, and  $\Lambda$  is the diagonal eigenvalue matrix. We have that  $\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2 = \|Q\Lambda^t Q^T \bar{\mathbf{d}}\|_2^2 = \sum_{i=1}^n \lambda_i^{2t}(A) (\bar{\mathbf{q}}_i^T \bar{\mathbf{d}})^2$ . Note that  $\mathbb{E}[(\bar{\mathbf{q}}_i^T \bar{\mathbf{d}})^2] = \mathbb{E}\left[\sum_{j=1}^n \mathbf{q}_{ij}^2 \mathbf{d}_j^2 + \sum_{1 \leq j < k \leq n} \mathbf{q}_{ij} \mathbf{q}_{ik} \mathbf{d}_j \mathbf{d}_k\right] = \mathbb{E}\left[\sum_{j=1}^n \mathbf{q}_{ij}^2 \mathbf{d}_j^2\right] = \sum_{i=1}^n \mathbf{q}_{ij}^2 \mathbb{V}[\mathbf{d}_j]$ . Therefore  $\mathbb{E}[\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2] = \sum_{j=1}^n \mathbb{V}[\mathbf{d}_j] \sum_{i=1}^n \lambda_i^{2t}(A) \mathbf{q}_{ij}^2$ . Since for all  $2 \leq i \leq n$  we have that  $|\lambda_i(A)| \leq \beta^*$ , and by using Jensen's inequality we get that

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{v}}_t - \bar{\boldsymbol{\mu}}_t\|_2] &\leq \sqrt{\sum_{i,j=1}^n \mathbb{V}[\mathbf{d}_j] \lambda_i^{2t}(A) \mathbf{q}_{ij}^2} \leq \sum_{i,j=1}^n \sqrt{\mathbb{V}[\mathbf{d}_j]} |\lambda_i^t(A)| |\mathbf{q}_{ij}| \\ &\leq \sum_{j=1}^n \sqrt{\mathbb{V}[\mathbf{d}_j]} |\mathbf{q}_{1j}| + (\beta^*)^t \sum_{i,j=1}^n \sqrt{\mathbb{V}[\mathbf{d}_j]} |\mathbf{q}_{ij}| \\ &\leq \sqrt{\sum_{j=1}^n \mathbb{V}[\mathbf{d}_j]} (1 + \sqrt{n-1}(\beta^*)^t). \end{aligned}$$

To minimize the upper bound on the MSE for each agent  $j$ , it suffices to minimize the variance of  $\mathbf{d}_j \sim \mathcal{D}_j$ , subject to differential privacy constraints. We assume that the PDF of  $\mathcal{D}_j$  – denoted by  $p_{\mathbf{d}_j}(\cdot) \in \Delta(\mathbb{R})$  – is differentiable everywhere in  $\mathbb{R}$ . The differential privacy constraint is equivalent to

$$\begin{aligned} \left| \frac{d}{d\mathbf{s}_j} \log \mathbb{P}[\psi_{\mathcal{M}_j^S}(\mathbf{s}_j) = t] \right| &\leq \varepsilon \iff \\ \left| \frac{d}{d\mathbf{s}_j} \log p_{\mathbf{d}_j}(t - \xi(\mathbf{s}_j)) \right| &\leq \varepsilon \iff \\ \left| \frac{d}{d\mathbf{s}_j} p_{\mathbf{d}_j}(t - \xi(\mathbf{s})) \right| &\leq \varepsilon p_{\mathbf{d}_j}(t - \xi(\mathbf{s}_j)), \end{aligned}$$

for all  $t \in \mathbb{R}$  and  $\mathbf{s}_j \in \mathcal{S}$ . Letting  $u = t - \xi(\mathbf{s}_j)$  we get that, in order to satisfy  $\varepsilon$ -DP,

$$\left| \frac{dp_{\mathbf{d}_j}(u)}{du} \right| \leq \frac{\varepsilon}{\Delta} p_{\mathbf{d}_j}(u),$$

where  $\Delta = \max_{\mathbf{s} \in \mathcal{S}} \left| \frac{d\xi(\mathbf{s}_j)}{d\mathbf{s}_j} \right|$  is the global sensitivity of  $\xi$ . We have that

$$\begin{aligned} \min_{p_{\mathbf{d}_j}(\cdot) \in \Delta(\mathbb{R})} \quad & \mathbb{E}_{\mathbf{d}_j \sim \mathcal{D}_j} [\mathbf{d}_j^2] = \int_{\mathbb{R}} t^2 p_{\mathbf{d}_j}(t) dt \\ \text{s.t.} \quad & \int_{\mathbb{R}} p_{\mathbf{d}_j}(t) dt = 1, \\ & |p'_{\mathbf{d}_j}(t)| \leq \frac{\varepsilon}{\Delta} p_{\mathbf{d}_j}(t), \quad \forall t \in \mathbb{R}. \end{aligned}$$

From Theorem 6 of Koufogiannis et al. (2015) we get that the optimal solution to the above problem is the Laplace distribution with scale  $\lambda_j = \Delta/\varepsilon$ ,

$$p_{\mathbf{d}_j}(t) = \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon}{\Delta}|t|\right), \quad \forall t \in \mathbb{R}.$$

To derive the upper bound on the error note that by Theorem 3 of Rahimian and Jadbabaie (2016b) we have that  $\mathbb{E} [\|\bar{\boldsymbol{\mu}}_t - \mathbf{1}\widehat{\mathbf{m}}_\theta\|_2] \leq \sqrt{n(n-1)}(\beta^*)^t M_n$ , and also  $\mathbb{E} [\|\bar{\boldsymbol{\mu}}_t - \bar{\boldsymbol{\nu}}_t\|_2] \leq \sqrt{\sum_{j=1}^n \mathbb{V}[\mathbf{d}_j]}(1 + \sqrt{n-1}(\beta^*)^t)$ . Applying the triangle inequality yields the final result, i.e.,

$$\mathbb{E} [\|\bar{\boldsymbol{\nu}}_t - \mathbf{1}\widehat{\mathbf{m}}_\theta\|_2] \leq \sqrt{n(n-1)}(\beta^*)^t M_n + \sqrt{\sum_{j=1}^n \mathbb{V}[\mathbf{d}_j]}(1 + \sqrt{n-1}(\beta^*)^t). \quad (\text{A.1})$$

Using the optimal distributions  $\mathcal{D}_i^* = \text{Lap}(\Delta/\varepsilon)$  in (8) gives the claimed upper bound on  $\text{TE}(\mathcal{M}^S)$ .  $\square$

## A.2 Proof of Corollary 3.2 (DP preservation across time)

To derive the DP guarantee for the MVUE for round  $t$ , we will do an induction. Specifically, we want to prove that for all  $\bar{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^n$  we have for all  $i \in [n]$ ,

$$\left| \log \left( \frac{\mathbb{P}[\boldsymbol{\nu}_{i,t} = \mathbf{x}_i]}{\mathbb{P}[\boldsymbol{\nu}'_{i,t} = \mathbf{x}_i]} \right) \right| \leq \varepsilon,$$

for all adjacent pairs of signals and beliefs, i.e.,  $\left\| (\mathbf{s}_i, \{\boldsymbol{\nu}_{j,t-1}\}_{j \in \mathcal{N}_i}) - (\mathbf{s}'_i, \{\boldsymbol{\nu}'_{j,t-1}\}_{j \in \mathcal{N}_i}) \right\|_1 \leq 1$ .

We proceed with the induction as follows:

- For  $t = 1$ , the result is held by the construction of the noise and the definition of DP.
- For time  $t \in \mathbb{N}$ , we assume that  $\left| \log \left( \frac{\mathbb{P}[\boldsymbol{\nu}_{i,t} = \mathbf{x}_i]}{\mathbb{P}[\boldsymbol{\nu}'_{i,t} = \mathbf{x}_i]} \right) \right| \leq \varepsilon$  for all  $\bar{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^n$ .
- For time  $t + 1$ , we have that for all  $i \in [n]$ ,

$$\left| \log \left( \frac{\mathbb{P}[\boldsymbol{\nu}_{i,t+1} = \mathbf{x}_i]}{\mathbb{P}[\boldsymbol{\nu}'_{i,t+1} = \mathbf{x}_i]} \right) \right| = \left| \log \left( \frac{\mathbb{P}[\boldsymbol{\nu}_{i,t} = (A^{-1}\bar{\mathbf{x}})_i]}{\mathbb{P}[\boldsymbol{\nu}'_{i,t} = (A^{-1}\bar{\mathbf{x}})_i]} \right) \right| \leq \varepsilon.$$

which holds by applying the definition of the MVUE update, the fact that  $A$  is non-singular, and the inductive hypothesis for  $t$ .

### A.3 Proof of Theorem 4.1

*Proof.* Similarly to Theorem 3.1, we decompose  $A$  as  $A = Q\Lambda Q^T$  and get that

$$\begin{aligned}\|\bar{\nu}_t - \bar{\mu}_t\|_2^2 &= \left\| \frac{1}{t} Q \sum_{\tau=0}^{t-1} \Lambda^\tau \bar{\mathbf{d}}_{t-\tau} Q^T \right\|_2^2 \\ &= \frac{1}{t^2} \sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i^{2\tau}(A) (\bar{\mathbf{q}}_i^T \bar{\mathbf{d}}_{t-\tau})^2.\end{aligned}$$

We take expectations and apply Cauchy-Schwarz to get

$$\begin{aligned}\mathbb{E} [\|\bar{\nu}_t - \bar{\mu}_t\|_2^2] &\leq \frac{1}{t^2} \sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i^{2\tau}(A) \mathbb{E} [\|\bar{\mathbf{d}}_{t-\tau}\|_2^2] \\ &\leq \frac{1}{t^2} \left( 1 + \sum_{i=2}^n \sum_{\tau=0}^{t-1} \lambda_i^{2\tau}(A) \right) \left( \sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}] \right) \\ &\leq \frac{1}{t^2} \left( 1 + (n-1) \sum_{\tau=0}^{t-1} (\beta^*)^{2\tau} \right) \left( \sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}] \right) \\ &\leq \frac{1}{t^2} \left( 1 + \frac{n-1}{1 - (\beta^*)^2} \right) \left( \sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}] \right).\end{aligned}\tag{A.2}$$

By Jensen's inequality, we get that

$$\mathbb{E} [\|\bar{\nu}_t - \bar{\mu}_t\|_2] \leq \frac{1}{t} \left( 1 + \sqrt{\frac{n-1}{1 - (\beta^*)^2}} \right) \sqrt{\sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}]}.$$

Also note that the dynamics of  $\bar{\mathbf{q}}_t = \bar{\mu}_t - \mathbf{1}\mathbf{m}_\theta$  obey

$$\bar{\mathbf{q}}_t = \frac{t-1}{t} A \bar{\mathbf{q}}_{t-1} + \frac{1}{t} (\bar{\boldsymbol{\xi}}_t - \mathbf{1}\mathbf{m}_\theta).$$

By following the same analysis as Equation (A.2) we get that

$$\mathbb{E} [\|\bar{\nu}_t - \bar{\mu}_t\|_2] \leq \sqrt{\frac{n}{t}} \left( 1 + \sqrt{\frac{n-1}{1 - (\beta^*)^2}} \right) \sqrt{\mathbb{V}[\boldsymbol{\xi}(\mathbf{s})]},$$

and then, the triangle inequality yields

$$\mathbb{E} [\|\bar{\mathbf{v}}_t - \mathbf{1} \mathbf{m}_\theta\|_2] \leq \frac{1}{t} \left( 1 + \sqrt{\frac{n-1}{1-(\beta^*)^2}} \right) \left( \sqrt{nt \mathbb{V}[\xi(\mathbf{s})]} + \sqrt{\sum_{j=1}^n \sum_{\tau=0}^{t-1} \mathbb{V}[\mathbf{d}_{j,t-\tau}]} \right).$$

To optimize the upper bound of Equation (A.2), for every index  $0 \leq \tau \leq t-1$  and agent  $j \in [n]$  we need to find the zero mean distribution  $\mathcal{D}_{j,\tau+1}$  with minimum variance subject to differential privacy constraints. We follow the same methodology as Theorem 3.1 and arrive at the optimization problem

$$\begin{aligned} \min_{p_{\mathbf{d}_{j,\tau+1}}(\cdot) \in \Delta(\mathbb{R})} \quad & \mathbb{E}_{\mathbf{d}_{j,\tau+1} \sim \mathcal{D}_j} [\mathbf{d}_{j,\tau+1}^2] = \int_{\mathbb{R}} u^2 p_{\mathbf{d}_{j,\tau+1}}(u) du \\ \text{s.t.} \quad & \int_{\mathbb{R}} p_{\mathbf{d}_{j,\tau+1}}(u) du = 1, \\ & |p'_{\mathbf{d}_{j,\tau+1}}(u)| \leq \frac{\varepsilon}{\Delta} p_{\mathbf{d}_{j,\tau+1}}(u), \quad \forall u \in \mathbb{R}. \end{aligned}$$

The optimal distribution is derived identically to Theorem 3.1, and equals  $\mathcal{D}_{j,\tau+1}^* = \text{Lap}\left(\frac{\Delta}{\varepsilon}\right)$  for all  $j \in [n]$  and  $0 \leq \tau \leq t-1$ . □

## A.4 Proof of Theorem 4.2

*Proof.* Let  $C(t) = B(t) - \frac{1}{t}I$ , and let  $\Phi(t) = \prod_{\tau=0}^{t-1} C(\tau)$ . Note that  $C(t)$  can be written as  $tC(t) = (t-2)I + A$ , and we can infer that the eigenvalues of  $C(t)$  satisfy  $\lambda_i(C(t)) = 1 + \frac{\lambda_i(A)-2}{t}$ , and that  $\{C(\tau)\}_{\tau \in [t]}$  and  $A$  have the same eigenvectors. Therefore,

$$\lambda_i(\Phi(t)) = \prod_{\tau=0}^{t-1} \lambda_i(C(\tau)) \leq \exp \left( \sum_{\tau=1}^t \frac{\lambda_i(A)-2}{\tau} \right) \leq t^{\lambda_i(A)-2}.$$

Similarly to Theorem 4.1 we have that

$$\begin{aligned}
\mathbb{E} \left[ \|\bar{\boldsymbol{\nu}}_t - \bar{\boldsymbol{\mu}}_t\|_2^2 \right] &= \frac{1}{t^2} \sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i(\Phi(\tau))^2 \mathbb{E} \left[ (\bar{\mathbf{q}}_i^T \bar{\mathbf{d}}_{t-\tau})^2 \right] \\
&\leq \frac{1}{t^2} \left( \sum_{i=1}^n \sum_{\tau=0}^{t-1} \lambda_i(\Phi(\tau))^2 \right) \left( \sum_{\tau=0}^{t-1} \mathbb{E} \left[ \|\bar{\mathbf{d}}_{t-\tau}\|_2^2 \right] \right) \\
&\leq \frac{1}{t^2} \left( \int_1^t \tau^{2\lambda_i(A)-4} d\tau \right) \sum_{\tau=0}^{t-1} \mathbb{E} \left[ \|\bar{\mathbf{d}}_{t-\tau}\|_2^2 \right] \\
&\leq \frac{1}{t^2} \left( \sum_{i=1}^n \int_1^t \tau^{2\lambda_i(A)-4} d\tau \right) \left( \sum_{i=1}^n \sum_{\tau=0}^{t-1} \mathbb{E} \left[ \|\bar{\mathbf{d}}_{t-\tau}\|_2^2 \right] \right) \\
&\leq \frac{1}{t^2} \left( \sum_{i=1}^n \frac{1}{3-2\lambda_i(A)} \right) \sum_{\tau=0}^{t-1} \mathbb{E} \left[ \|\bar{\mathbf{d}}_{t-\tau}\|_2^2 \right] \\
&\leq \frac{1}{t^2} \left( 1 + \frac{n-1}{3-2\beta^\star} \right) \sum_{\tau=0}^{t-1} \mathbb{E} \left[ \|\bar{\mathbf{d}}_{t-\tau}\|_2^2 \right] \\
&\leq \frac{1}{t^2} \left( 1 + \frac{n-1}{3-2\beta^\star} \right) \left( \sum_{i=1}^n \sum_{\tau=0}^{t-1} \mathbb{V} [\mathbf{d}_{i,t-\tau}] \right).
\end{aligned}$$

Applying Jensen's inequality, we get that

$$\mathbb{E} [\|\bar{\boldsymbol{\nu}}_t - \bar{\boldsymbol{\mu}}_t\|_2] \leq \frac{1}{t} \sqrt{\sum_{i=1}^n \sum_{\tau=0}^{t-1} \mathbb{V} [\mathbf{d}_{i,t-\tau}]} \left( 1 + \sqrt{\frac{n-1}{3-2\beta^\star}} \right).$$

Similarly, by considering the dynamics of  $\bar{\mathbf{q}}_t = \bar{\boldsymbol{\mu}}_t - \mathbf{1}\mathbf{m}_\theta$  we get that

$$\mathbb{E} [\|\bar{\boldsymbol{\mu}}_t - \mathbf{1}\mathbf{m}_\theta\|_2] \leq \sqrt{\frac{n}{t}} \sqrt{\mathbb{V} [\xi(\mathbf{s})]} \left( 1 + \sqrt{\frac{n-1}{3-2\beta^\star}} \right).$$

The triangle inequality yields the final error bound.

To derive the optimal distributions, note that at each round  $t$ , the optimal action for agent  $i$  is to minimize  $\mathbb{V} [\mathbf{d}_{i,t}]$  subject to DP constraints. By following the analysis similar to Theorem 3.1, we deduce that the optimal noise to add is Laplace with parameter  $\max\{\max_{j \in \mathcal{N}_i} a_{ij}, \Delta\}/\varepsilon$ .

□

## B Algorithm of Rizk et al. (2023)

We adapt the framework of Rizk et al. (2023) to our problem, for which the identification of the MVUE  $\widehat{\mathbf{m}}_\theta$  can be formulated as

$$\widehat{\mathbf{m}}_\theta = \operatorname{argmin}_{\mathbf{m} \in \mathbb{R}} \frac{1}{2n} \sum_{i=1}^n \underbrace{(\mathbf{m} - \xi(\mathbf{s}_i))^2}_{J_i(\mathbf{m})}.$$

The private dynamics for updating the beliefs  $\bar{\nu}_t$  can be found by simplifying the consensus algorithm given in Equations (24)-(26) of Rizk et al. (2023):

$$\nu_{i,t} = a_{ii}(\nu_{i,t-1} + g_{ii,t}) + \sum_{j \in \mathcal{N}_i} a_{ij}(\nu_{j,t-1} + g_{ij,t}) + d_{i,t} - \eta(\nu_{i,t-1} - \xi(s_i)).$$

Here,  $\eta$  is the learning rate,  $d_{i,t}$  is noise used to protect the private signal, and  $g_{ij,t}$ ,  $\{g_{ij,t}\}_{j \in \mathcal{N}_i}$  are noise terms used to protect own and the neighboring beliefs. As a first difference, we observe that Rizk et al. (2023) uses  $(n + m)T$  noise variables, whereas our method uses just  $n$  noise variables, which makes our method easier to implement for the MVUE task. It is easy to observe that these dynamics converge slower than our dynamics for two reasons: (i) the privacy protections are added separately for the signal and each neighboring belief, and (ii) the beliefs are always using information from the signals since the method is first-order, thus requiring noise to be added at each iteration.

For this reason, the authors consider graph-homomorphic noise, i.e., noise of the form  $g_{ij,t} = q_{i,t}$  for all  $j \neq i$  and  $g_{ii,t} = -\frac{1-a_{ii}}{a_{ii}}q_{i,t}$  where  $q_{i,t}$  are noise variables. Rewriting the dynamics in this form, we get the following update:

$$\nu_{i,t} = (a_{ii} - \eta)\nu_{i,t-1} + \sum_{j \in \mathcal{N}_i} a_{ij}\nu_{j,t-1} + \eta\xi(s_i) + d_{i,t}.$$

Given a privacy budget  $\varepsilon$ , in order to make a fair comparison with our algorithm, the noise variable should be chosen as  $d_{i,t} \sim \text{Lap}\left(\frac{\eta T S_{\xi,\gamma}^*(s_i)}{\varepsilon}\right)$  for Signal DP, and  $d_{i,t} \sim \text{Lap}\left(\frac{T \max\{\max_{j \neq i} a_{ij}, \eta S_{\xi,\gamma}^*(s_i)\}}{\varepsilon}\right)$  for Network DP. Here, since the per-agent privacy budget is  $\varepsilon$ , and the noise is added  $T$  times at each iteration, the initial budget needs to be divided by  $T$ .

We run the same experiments as in Section 5 with the method of Rizk et al. (2023) and compare with our method using the same values of the privacy budget  $\varepsilon$  and a learning rate  $\eta = 0.001$  to get the results in Figure 7.

## C Extensions to Dynamic and Directed Networks and Heterogeneous Privacy Budgets

### C.1 Dynamic Networks

In this problem, the agents observe a sequence of dynamic networks  $\{G(t)\}_{t \in \mathbb{N}}$ , for example, due to corrupted links, noisy communications, other agents choosing not to share their measurements, power failures etc. These dynamic networks correspond to a sequence of doubly stochastic matrices  $\{A(t)\}_{t \in \mathbb{N}}$ . A choice for the weights are the modified Metropolis-Hastings (MH) weights:

$$a_{ij}(t) = \begin{cases} \frac{1}{2 \max\{\deg_t(i), \deg_t(j)\}}, & j \neq i \\ 1 - \sum_{j \in \mathcal{N}_i} a_{ij}(t), & j = i \end{cases}. \quad (\text{C.3})$$

This is a natural choice of weights since they can be easily and efficiently computed by the agents in a distributed manner (e.g., in a sensor network) from the knowledge of own and neighboring degrees, thus requiring minimal memory to be stored for each agent. Below we will show that the MVUE and the OL algorithms converge under minimal assumptions for the time-varying MH weights.

**MVUE.** We can prove that if  $G_t$  contain no isolated nodes, the beliefs would converge to  $\widehat{\mathbf{m}}_\theta$  as a direct consequence of (Chazelle, 2011, Theorem 1.4), i.e.

**Proposition C.1.** *If  $G_t$  contains no isolated nodes for all  $t \in \mathbb{N}$ , for accuracy  $0 < \rho < 1/2$ , the dynamics with the modified MH weights of Equation (C.3) will converge to the MVUE,  $\widehat{\mathbf{m}}_\theta$ , in*

$$t = \min \left\{ \frac{2^{O(n)}(M_n + \Delta/\varepsilon)}{\rho}, \left( \log \frac{M_n + \Delta/\varepsilon}{\rho} \right)^{n-1} 2^{n^2 - O(n)} \right\}$$

steps.

*Proof.* The proof follows from applying Theorem 1.4 of Chazelle (2011), since  $\max_{j \neq i} a_{ij} \leq 1/2$ , the graph is undirected, and the diameter of the points is at most  $2(M_n + \Delta/\varepsilon)$ .  $\square$

Note that the above convergence rate is exponentially worse than the

$$t = O \left( \frac{\log(n(M_n + \Delta/\varepsilon)/\rho)}{\log(1/\beta^*)} \right)$$

convergence time that we obtain for static networks.

**Online Learning.** For the online learning regime, we can follow the approach presented in Touri and Nédic (2009). For this, we consider the following update rule:

$$\bar{\mathbf{v}}_t = \frac{t-1}{t} \bar{\mathbf{v}}_{t-1} + \frac{1}{t} A(t) \bar{\mathbf{v}}_{t-1} + \frac{1}{t} (\bar{\boldsymbol{\xi}}_t + \bar{\mathbf{d}}_t). \quad (\text{C.4})$$

We show that under reasonable assumptions on the graph sequence, the above algorithm converges to  $\mathbf{1}\mathbf{m}_\theta$  almost surely.

**Proposition C.2.** *If  $G_t$  contains no isolated nodes for all  $t \in \mathbb{N}$ , and there exists an integer  $B$  such that  $\bigcup_{\tau=tB}^{t(B+1)-1} G_\tau$  is strongly connected for every  $t \in \mathbb{N}$ , then the dynamics of Equation (C.4) with the modified MH weights of Equation (C.3) converge to  $\mathbf{1}\mathbf{m}_\theta$  almost surely.*

*Proof.* The dynamics have the form of the dynamics of Touri and Nédic (2009). We will show the result by showing that the assumptions of Touri and Nédic (2009) hold. Specifically, all non-zero elements of  $A(t)$  have value at least  $1/(2n)$ ,  $a_{ii}(t) \geq 1/2$  for all  $i \in [n], t \in \mathbb{N}$  by our hypothesis, there exists an integer  $B$  such that  $\bigcup_{\tau=tB}^{t(B+1)-1} G_\tau$  is strongly connected for every  $t \in \mathbb{N}$  by our hypothesis,  $\sum_{t=1}^{\infty} 1/t = \infty$ , and  $\sum_{t=1}^{\infty} 1/t^2 < \infty$ . Thus, by Proposition 2 of Touri and Nédic (2009), the beliefs converge almost surely to  $\mathbf{1}\mathbf{m}_\theta$  as  $t \rightarrow \infty$ . Furthermore, we can show that the dependence on  $t$  becomes slower, i.e., from  $t^{-1/2}$  when the network is static to  $t^{-1/n^3}$  when the network is dynamic.  $\square$

## C.2 Directed Networks

We can also consider extensions of our results to directed graphs, i.e., when  $a_{ij} \neq a_{ji}$  in general. Such asymmetries can arise due to systemic heterogeneities, e.g., varying accuracy and reliability of sensors in a sensor network or imbalances caused by influence dynamics in a social network. The adjacency weights in such cases are no longer doubly stochastic, making the MVUE dynamics in Equations (3) and (7) to converge to  $\widehat{\mathbf{m}}_\theta = \bar{\mathbf{q}}_1^T \bar{\boldsymbol{\xi}}$  where  $\bar{\mathbf{q}}_1$  is the stationary distribution of the Markov chain with transition matrix  $A$ , satisfying  $A\bar{\mathbf{q}}_1 = \bar{\mathbf{q}}_1$  and  $\bar{\mathbf{q}}_1^T \mathbf{1} = 1$ . Unlike the doubly stochastic case, the stationary distribution is not necessarily uniform, in which case the aggregate converges to a weighted average of  $\xi(\mathbf{s}_i)$  that is no longer minimum variance. To analyze the convergence rate of the algorithms with asymmetric adjacency weights, we cannot use tools from the theory of doubly-stochastic matrices anymore, and we need to rely on different tools from the analysis of convergence of non-reversible Markov chains. The results of Chatterjee (2023) extend the notion of spectral gap and can make this analysis possible. However, these results come with significant technicalities, using substantially different analytical steps that are beyond the scope of our present work. More limited conclusions about the asymptotic rates and finite-time convergence can be asserted in special cases by applying existing results such as Touri and Nédic (2009).

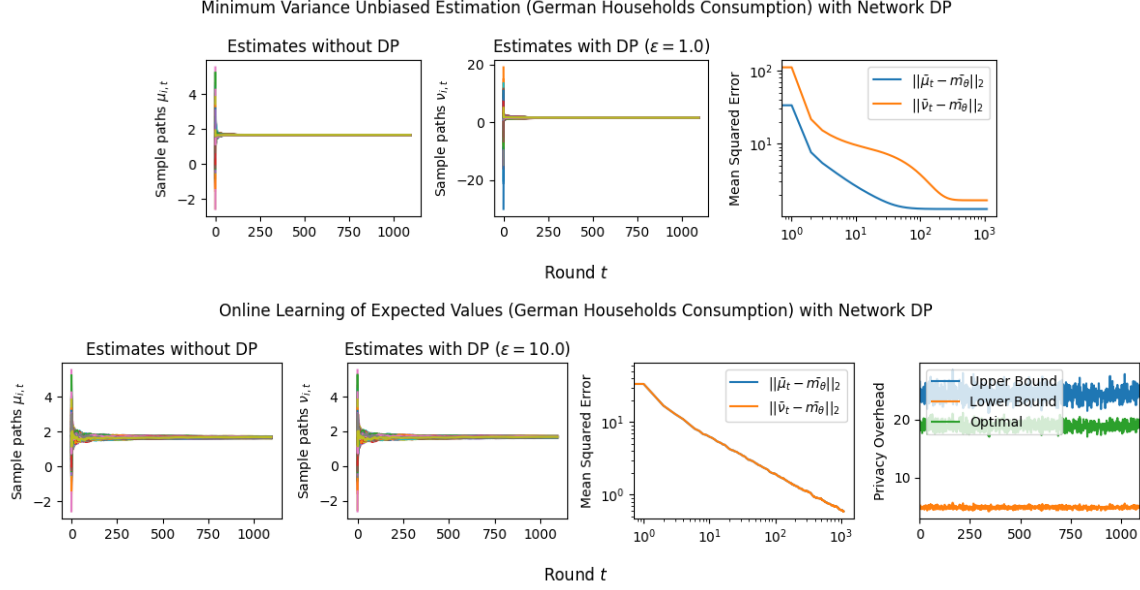
## C.3 Heterogeneous Privacy Budgets

Our algorithms extend easily to the case where each agent has their own privacy budget  $\varepsilon_i$ , e.g., due to their differing energy consumption levels. In this case, it is easy to show that all of the results can be updated to accommodate heterogeneous  $\varepsilon_i$  and the smooth sensitivities  $\Delta_i$ , as shown in Table 2 below.

Table 2: Total Error Bounds for heterogeneous privacy budgets  $\{\varepsilon_i\}_{i \in [n]}$ . The blue terms are due to privacy constraints (CoP), and the red terms are due to decentralization (CoD). Here  $M_n$  and  $\beta^*$  are the same as in Table 1, and  $\{\Delta_i\}_{i \in [n]}$  are the smooth sensitivities for each agent which can be set as  $\Delta_i = S_{\xi, \gamma}^*(\mathbf{s}_i)$  for the case of Signal DP and  $\Delta_i = \max \left\{ \max_{j \neq i} a_{ij}, S_{\xi, \gamma}^*(\mathbf{s}_i) \right\}$  for the case of Network DP.

| Minimum Variance Unbiased Estimation                                                               | Online Learning of Expected Values                                                                                                        |
|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| $O \left( (\beta^*)^t M_n + (1 + (\beta^*)^t) \sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i} \right)$ | $O \left( \frac{n}{\sqrt{t}} \sqrt{\mathbb{V}[\xi(\mathbf{s})]} + \frac{1}{\sqrt{t}} \sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i} \right)$ |

The possibility of heterogeneous privacy budgets opens new avenues to explore the allocation of an overall privacy budget ( $n\varepsilon$ ) while respecting individual budgets  $\{\varepsilon_{i, \max}\}_{i \in [n]}$ . Such a situation may arise, e.g., to limit overall information leakage against an adversary that can eavesdrop on some or all of the communications. Suppose that each agent wants to maintain  $\varepsilon_i$ -DP and suppose that there is an adversary that eavesdrops on all of the beliefs (this also covers the cases where the adversary has access to a subset  $W$  of the  $[n]$ , such as in the case of a targeted attack to a large part of the network). Note that all of the results presented so far protect against information leakage about a single agent's signals (or their network neighborhoods) when their reports  $\nu_{i,t}$  are compromised. However, if the goal is to protect the private signals against an adversary that can eavesdrop simultaneously on all or a subset of agents, then each individual agent's report can be regarded as a data release



Supplementary Figure 1: Sample Paths for MVUE and OL for the German Households Dataset with heterogeneous budgets (centralized solution). For the OL case, we plot the optimal privacy overhead  $\sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i^*}$  which we compare with the lower bound  $\sum_{i=1}^n \frac{\Delta_i}{\varepsilon}$ , and the upper bound  $\sum_{i=1}^n \frac{\Delta_i}{\varepsilon_{i,\max}}$ .

about the vector of all signals that needs to be protected at the  $n\varepsilon$ -DP level, in addition to the individual-level  $\varepsilon_{i,\max}$ -DP protections required by each agent. If the adversary can eavesdrop on all of the signals, then the resulting mechanism that protects the joint distribution of the signals can be thought of as a mechanism  $\Psi^{\mathcal{M}^S}$  which adds  $n$ -dimensional noise  $\bar{\mathbf{d}}$  to the sufficient statistics and each dimension  $i \in [n]$  of the noise corresponds to  $\mathbf{d}_{i,t}$ , i.e.,

$$\Psi^{\mathcal{M}^S}(\mathbf{s}_{1,t}, \dots, \mathbf{s}_{n,t}) = \bar{\boldsymbol{\xi}}_t + \bar{\mathbf{d}}_t$$

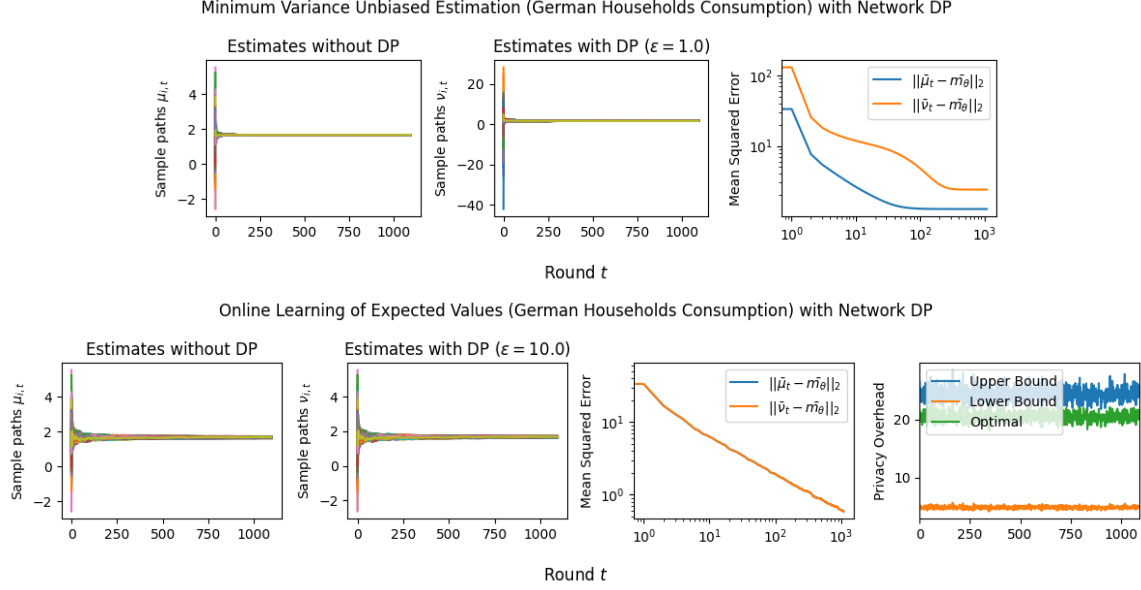
Then, if  $\Delta_i$  is the sensitivity for each agent  $i$ , the noise has PDF:

$$p_{\bar{\mathbf{d}}_t}(\mathbf{u}_1, \dots, \mathbf{u}_n) = \prod_{i=1}^n \frac{\varepsilon_i}{2\Delta_i} \exp\left(-\frac{\varepsilon_i}{\Delta_i} |\mathbf{u}_i|\right) \propto \exp\left(-\sum_{i=1}^n \frac{\varepsilon_i}{\Delta_i} |\mathbf{u}_i|\right)$$

Now, consider a pair  $(\mathbf{s}_{1,t}, \dots, \mathbf{s}_{n,t}), (\mathbf{s}'_{1,t}, \dots, \mathbf{s}'_{n,t})$  of sets of signals such that

$$\|(\mathbf{s}_{1,t}, \dots, \mathbf{s}_{n,t}) - (\mathbf{s}'_{1,t}, \dots, \mathbf{s}'_{n,t})\|_1 \leq 1.$$

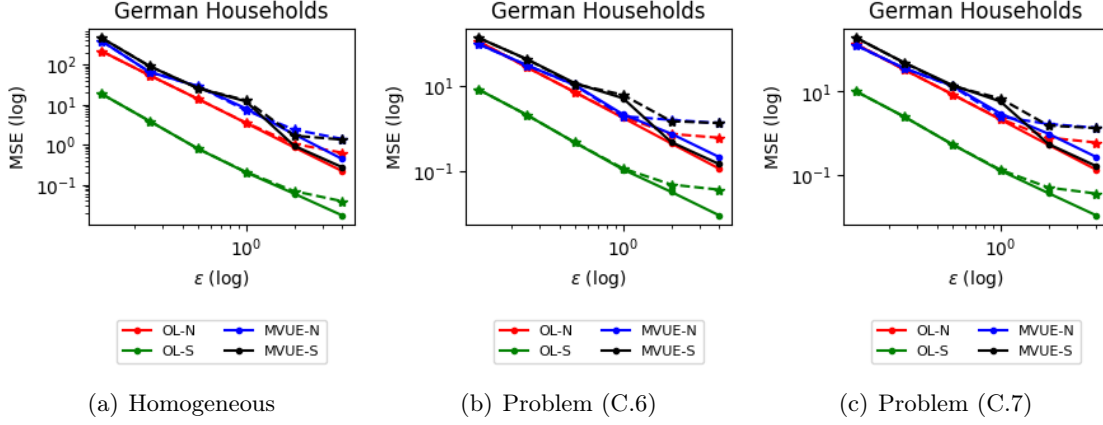
Then, we have that for all  $\bar{\mathbf{x}} \in \mathbb{R}^n$ :



Supplementary Figure 2: Sample Paths for MVUE and OL for the German Households Dataset with heterogeneous budgets (decentralized solution). For the OL case, we plot the optimal privacy overhead  $\sum_{i=1}^n \frac{\Delta_i}{\epsilon_i^*}$  which we compare with the lower bound  $\sum_{i=1}^n \frac{\Delta_i}{\epsilon}$ , and the upper bound  $\sum_{i=1}^n \frac{\Delta_i}{\epsilon_{i,\max}}$ .

$$\begin{aligned}
\left| \log \left( \frac{\mathbb{P}[\Psi^{\mathcal{M}_{i,t}^S}(\mathbf{s}_{1,t}, \dots, \mathbf{s}_{n,t}) = \bar{\mathbf{x}}]}{\mathbb{P}[\Psi^{\mathcal{M}_{i,t}^S}(\mathbf{s}'_{1,t}, \dots, \mathbf{s}'_{n,t}) = \bar{\mathbf{x}}]} \right) \right| &= \left| \log \left( \frac{p_{\bar{\mathbf{d}}_t}(\bar{\boldsymbol{\xi}}_t - \bar{\mathbf{x}})}{p_{\bar{\mathbf{d}}_t}(\bar{\boldsymbol{\xi}}'_t - \bar{\mathbf{x}})} \right) \right| \\
&= \left| \sum_{i=1}^n \frac{\epsilon_i}{\Delta_i} (|\xi(\mathbf{s}'_{i,t}) - \bar{\mathbf{x}}| - |\xi(\mathbf{s}_{i,t}) - \bar{\mathbf{x}}|) \right| \\
&\leq \sum_{i=1}^n \frac{\epsilon_i}{\Delta_i} |\xi(\mathbf{s}_{i,t}) - \xi(\mathbf{s}'_{i,t})| \\
&\leq \sum_{i=1}^n \epsilon_i \|\mathbf{s}_{i,t} - \mathbf{s}'_{i,t}\|_1 \\
&\leq \left( \sum_{i=1}^n \epsilon_i \right) \max_{i \in [n]} \|\mathbf{s}_{i,t} - \mathbf{s}'_{i,t}\|_1 \\
&\leq \sum_{i=1}^n \epsilon_i. \tag{C.5}
\end{aligned}$$

Now suppose we want to protect the vector of all beliefs against the eavesdropper at the  $n\epsilon$ -DP level (assuming an average privacy budget of  $\epsilon$  per agent for the overall protection of the vector of all private signals). The noise that is added to individual estimates also works to protect the entire vector of all estimates against the eavesdropper and to achieve the latter at the  $n\epsilon$  level, Equation (C.5) indicates that it is sufficient to ensure that  $\sum \epsilon_i \leq n\epsilon$ . On the other hand, given  $\epsilon_i$  privacy level at every agent  $i$ , we also want to minimize the



Supplementary Figure 3: MSE Plots for the German Households Dataset with heterogeneous privacy budgets. We note that compared to the homogeneous case, using heterogeneous budgets reduces the MSE.

accuracy loss by reducing  $\sum_{i=1}^n \Delta_i / \varepsilon_i$  while ensuring that individual privacy budgets do not exceed a preset maximum:  $\text{vec} \varepsilon_i \leq \varepsilon_{i,\max}$  for all  $i$ . The subsequent optimization problem to allocate individual privacy budgets can be formulated as follows:

$$\begin{aligned}
 \min_{\varepsilon_1, \dots, \varepsilon_n > 0} \quad & \sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i} \\
 \text{s.t.} \quad & \sum_{i=1}^n \varepsilon_i \leq n\varepsilon. \\
 & \varepsilon_i \leq \varepsilon_{i,\max} \quad \forall i \in [n]
 \end{aligned} \tag{C.6}$$

Following KKT conditions Boyd et al. (2005), Equation (C.6) admits a closed-form solution. The solution indicates that by allowing heterogeneous privacy budgets and taking into account individual smooth sensitivities in the optimal allocation, we can improve the total error. These observations are summarized below:

**Proposition C.3.** *The following hold:*

1. If  $\sum_{i=1}^n \varepsilon_{i,\max} \geq n\varepsilon$ , then the optimal solution to Equation (C.6) is

$$\varepsilon_i^* = \min \left\{ \varepsilon_{i,\max}, \frac{n\varepsilon \sqrt{\Delta_i}}{\sum_{j \in [n]} \sqrt{\Delta_j}} \right\}$$

for all  $i \in [n]$ . Moreover, the improvement over  $\sum_{i=1}^n \frac{\Delta_i}{\varepsilon}$  satisfies  $\frac{\min_{i \in [n]} \Delta_i}{\max_{i \in [n]} \Delta_i} \leq$

$$\frac{\sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i^*}}{\sum_{i=1}^n \frac{\Delta_i}{\varepsilon}} \leq 1.$$

2. If  $\sum_{i=1}^n \varepsilon_{i,\max} < n\varepsilon$  then the optimal solution is  $\varepsilon_i^* = \varepsilon_{i,\max}$  for all  $i \in [n]$ .

In a large network making individual nodes aware of their allocated budgets based on their smooth sensitivities is difficult to achieve in a central manner. The following formulation of the allocation problem arrives at a sub-optimal solution that satisfies the  $n\varepsilon$  global privacy budget constraint by imposing  $n$  additional constraints in the local neighborhoods:  $a_{ii}\varepsilon_i + \sum_{j \in \mathcal{N}_i} a_{ij}\varepsilon_j \leq \varepsilon$ ,  $\forall i \in [n]$ . The advantage of these constraints is that they can be verified locally, and satisfying them implies the  $n\varepsilon$  global budget constraint in Equation (C.6). The subsequent optimization problem is given in Equation (C.7). It allows individuals to learn their allocated budgets in a distributed manner by running distributed gradient descent, which is guaranteed to converge since the problem is convex (Sayed et al., 2014, Chapter 7).

$$\begin{aligned}
& \min_{\varepsilon_1, \dots, \varepsilon_n > 0} \quad \sum_{i=1}^n \frac{\Delta_i}{\varepsilon_i} \\
& \text{s.t.} \quad a_{ii}\varepsilon_i + \sum_{j \in \mathcal{N}_i} a_{ij}\varepsilon_j \leq \varepsilon, \quad \forall i \in [n] \\
& \quad \varepsilon_i \leq \varepsilon_{i,\max}, \quad \forall i \in [n].
\end{aligned} \tag{C.7}$$

**Numerical Experiment.** We test our method with the German Households dataset. Specifically, we set a per-agent average budget of  $\varepsilon = 1$  for MVUE and  $\varepsilon = 10$  for OL. We put a maximum individual budget cap of  $\varepsilon_{i,\max} = 10\varepsilon$  in both cases. We report the sample paths in Supplementary Figures 1 and 2, and observe that the dynamics converge faster compared to the homogeneous case (cf. Figure 4). In Supplementary Figure 3, we present an MSE plot where the MSE is plotted as a function of  $\varepsilon$ , and observe that the algorithm with the heterogeneous thresholds has smaller MSE compared to the homogeneous thresholds. These results confirm our theoretical observations in Proposition C.3.