Estimating Unbiased Averages of Sensitive Attributes without Handshakes among Agents

Arijus Pleska and Jan Ramon

Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

arijus.pleska@inria.fr, jan.ramon@inria.fr

Abstract

We consider the problem of distributed averaging of sensitive attributes in a network of agents without central coordinators, where the graph of the network has an arbitrary degree sequence (degrees refer to numbers of neighbors of vertices). Usually, existing works solve this problem by assuming that either (i) the agents reveal their degrees to their neighbors or (ii) every two neighboring agents can perform handshakes (requests that rely on replies) in every exchange of information. However, the degrees suggest the profiles of the agents and the handshakes are impractical upon inactive agents. We propose a decentralized, handshake-free, asynchronous approach which is applicable upon privatized degrees. In particular, we use a simple gossip algorithm that computes averages that are biased when the graph is non-regular (upon unequal degrees) and then perform a procedure combining the biased averages for bias correction. We investigate a use case of the proposed approach for fitting a linear regression model upon privatizing the features and the degrees. We provide theoretical guarantees that the mean squared error (MSE) between the average of privatized attributes computed by our approach and the average of sensitive attributes is $\mathcal{O}(\frac{1}{n})$, where n is the number of agents. We show on a synthetic graph dataset that the theoretical error is sufficiently tight. Also, we show on the synthetic graph dataset and real graph datasets that, when fitting a regression model whose features are polynomials of degrees, our approach can outperform the centralized averaging of locally privatized attributes.

Introduction

Over the last decade there has been significant interest in self-organizing distributed systems, where the nodes (agents) in a communication network collaborate without central coordinators. One basic task corresponds to the problem where every agent has an individual value and every agent would like to know the average of those values. As we will argue, existing works usually assume a form of handshakes between every two neighboring agents in every exchange of information, i.e., when one agent uses information of a second agent, this second agent becomes aware of this and must actively help the process. However, there exist practical scenarios where such interaction is time consuming due to inactivity of some agents.

To give an illustrative example on communication without handshakes, let us consider a group of researchers who aim at solving a particular problem. The researchers can follow each other, and thus read each other's currently best strategy in each other's most recently published paper. The researchers work on their solution strategy individually and without necessarily directly contacting others, attempting to improve their current solution strategies based on their individual skills and the ideas read in the papers of the followed colleagues. At some point, some of them might find a fully satisfactory solution.

Motivation

Usually, there are two common types of information flow [1]: pulling, where an agent asks its neighboring agent for its value, and pushing, where an agent sends its value to a neighboring agent. In this work, we study a weak form of pulling where an agent obtains the current value of a neighbor without the neighbor being aware of this. In particular, every agent continuously publishes its current values so that its neighbors can obtain it, but there is no other communication (e.g., there is no communication process to build overlay networks [3] that can improve distributed computations). Since the agents do not exchange handshakes (like in the Transport Layer Security protocol), information dissemination is more robust against inactive agents. Further, we assume that the degree is a sensitive attribute because in some contexts it suggests the profile of an agent [2]. Finally, we mention that our approach applies for graphs with power-law degree sequences which, as suggested by Zipf's law, are common in real-world (e.g., computer, social, biological) networks.

Setting

Communication model. We model our network of agents by a graph $G = (V, E)$, where V is the set of vertices ($v \in V$ is a vertex) and E is the set of edges. We denote the order of the graph, i.e. $|V|$, by n. We denote the degree of a vertex v by d_v. We denote the degree sequence of G by $\mathbf{d} = (d_1, \ldots, d_n)$. For $k \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, we denote the kth raw moment $\frac{1}{n}\sum_{i=1}^{n} x_i^k$ by μ_{x^k}. We model G by a random graph drawn from the configuration model which is defined as follows:

Definition 1. *The configuration model is the probability distribution over graphs that is parametrized by* $\mathbf{d}$*, so that, for* $i \in [n-1]$ *and* $j \in \{i+1, \ldots, n\}$*, we have* $\Pr\left((v_i, v_j) \in E\right) = \frac{d_i d_j}{(\sum_{i'=1}^{n} d_{i'})-1}$.

Attack model. We assume that the agents are honest-but-curious, i.e., all agents follow the established protocols (they are honest), but they try to use the available information to infer sensitive information of other agents (they are curious).

We interpret a basic dataset as a table with instances over rows and (scalar) attributes over its columns. In local differential privacy, the common idea is to add noise to attributes, and such attributes are known as sensitive attributes. We define (ϵ, δ)-differential privacy for graphs:

Definition 2. *Let* $\epsilon, \delta \geq 0$*. A randomized algorithm* $\mathcal{A}$ *is* (ϵ, δ)*-differentially private if and only if, for all tuples* $(\mathcal{D}, \mathcal{D}')$ *in a collection where datasets* $\mathcal{D}$ *and* $\mathcal{D}'$ *differ only in the attribute of one instance and for all* $\mathcal{S} \subseteq \mathrm{image}(\mathcal{A})$*, we have* $\Pr(\mathcal{A}(\mathcal{D}) \in \mathcal{S}) \leq e^{\epsilon}\Pr(\mathcal{A}(\mathcal{D}') \in \mathcal{S}) + \delta$.

Approach

We intend to design a communication protocol for distributed averaging, where (i) individual values and degrees are kept differentially private, (ii) the computed averages are unbiased, (iii) the choice of the degree sequence is arbitrary, (iv) handshakes are absent, and (v) central coordinators are absent.

Our approach has two parts. In the first part, the agents commit to a simple gossip algorithm (`SiGo`) for distributed averaging. More specifically, every agent hides their sensitive attributes under differential privacy noise and makes them visible to the neighbors. Then, every agent repeatedly computes the averages the values revealed by their neighbors and updates the displayed value by the computed average. At some point this converges and the computed averages are biased when the graph is non-regular. In the second part, the agents perform a procedure combining the biased averages for bias correction.

Algorithm 1: `SimpleGossip(SiGo)`

Input : $\mathbf{A} \in [0,1]^{n \times n}$: adjacency matrix of the graph of agents
$\mathrm{it}_{\mathrm{go}} \in \mathbb{N}$: number of gossip iterations
$\mathbf{w} \in \mathbb{R}^n$
Output: $\mathbf{z} \in \mathbb{R}^n$
$\mathbf{d} \leftarrow \sum_{i=1}^{n} \mathbf{A}_{:,i}$
$\mathbf{T} \leftarrow \mathrm{diag}(\mathbf{d}^{\circ -1})\mathbf{A}$ ($\circ$ denotes the operator for the element-wise power)
$\mathbf{z} \leftarrow \mathbf{T}^{\mathrm{it}_{\mathrm{go}}}\mathbf{w}$

Theorem 1. *Let* $\mathbf{w} \in \mathbb{R}^n$*. Let* $\mathbf{A}$ *be the adjacency matrix of a simple, connected graph with at least one odd cycle. We have* $\lim_{\mathrm{it}_{\mathrm{go}} \to \infty} \texttt{SiGo}\left(\mathbf{A}, \mathrm{it}_{\mathrm{go}}, (w_i)_{i=1}^n\right) = \frac{1}{n}\frac{1}{\mu_d}\sum_{i=1}^{n} d_i w_i$*, where* $\texttt{SiGo}\left(\mathbf{A}, \mathrm{it}_{\mathrm{go}}, (w_i)_{i=1}^n\right)$ *denotes any element of the output of* `SiGo`*. In later references, we shorthand* $\lim_{\mathrm{it}_{\mathrm{go}} \to \infty} \texttt{SiGo}\left(\mathbf{A}, \mathrm{it}_{\mathrm{go}}, (w_i)_{i=1}^n\right)$ *by* $\texttt{SiGo}\left((w_i)_{i=1}^n\right)$.

Theorem 1 suggests that the resulting average is biased by μ_d and d_i when $d_i \neq \mu_d$. We define the bias-correcting gossip algorithm (`BCGo`) that combines two runs of `Sigo` for bias correction:

$$\frac{\texttt{SiGo}\left((w_i^k d_i^{-1})_{i=1}^n\right)}{\texttt{SiGo}\left((d_i^{-1})_{i=1}^n\right)} = \frac{\frac{1}{n}\frac{1}{\mu_d}\sum_{i=1}^{n} d_i (w_i^k d_i^{-1})}{\frac{1}{n}\frac{1}{\mu_d}\sum_{i=1}^{n} d_i (d_i^{-1})} = \frac{1}{n}\sum_{i=1}^{n} w_i^k = \mu_{w^k}. \tag{1}$$

Use Case on Linear Regression

Let $m \in \mathbb{N}$. We define a special case of multiple linear regression model with $m+1$ regression parameters and one-dimensional target value. For every $i \in [n]$,

$$y_i = \theta_0 + \theta_1 d_i^{k_1} + \ldots + \theta_m d_i^{k_m} + \xi_i^{\mathrm{reg}}, \tag{2}$$

where $\theta_0, \ldots, \theta_m \in \mathbb{R}$ are regression parameters, $d_i^{k_1}, \ldots, d_i^{k_m} \in \mathbb{R}$ are features, $k_1, \ldots, k_m \in \mathbb{R}$, y_i is a target value, $\Xi_i^{\mathrm{reg}} \sim \mathcal{N}(0, \sigma_{\mathrm{reg}}^2)$ is regression noise with variance σ_{reg}^2, and ξ_i^{reg} is an independent observation of Ξ_i^{reg}. Further, let $\mathbf{X} = \begin{bmatrix}\mathbf{1} & \mathbf{x}_1 & \ldots & \mathbf{x}_m\end{bmatrix} \in \mathbb{R}^{n \times (m+1)}$, where $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are feature vectors. Let $\mathbf{y}$ be the vector of target values. Let $\hat{\theta}$ be the vector of parameter estimates. By ordinary least squares, $\hat{\theta} = \left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{y}$, where $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ contains $\mu_{d^{k_1}}, \ldots, \mu_{d^{k_m}}$, $\mu_{d^{2k_1}}, \ldots, \mu_{d^{2k_m}}$, $\mu_{d^{k_1}.d^{k_m}}, \ldots, \mu_{d^{k_{m-1}}.d^{k_m}}$ and $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ contains $\mu_y, \mu_{y.d^{k_1}}, \ldots \mu_{y.d^{k_m}}$. The mentioned averages are U-statistics of degree 1 (definition omitted).

Privatization. Let $k \in \mathbb{R}$. Let $\Xi_{i,k}^{\mathrm{dp}} \sim \mathcal{N}\left(d_i^k, (\sigma_k^{\mathrm{dp}})^2\right)$, where $\sigma_k^{\mathrm{dp}} = \frac{\sqrt{2\log(1.25/\delta')}\Delta(d_i^k)}{\epsilon'}$, $\Delta(d^k) = \max_{d' \in [d_{\min}, d_{\max}-1]}\left|(d')^k - (d'+1)^k\right|$, and (ϵ', δ') is the privacy budget for one sensitive attribute. We denote a privatized attribute by $\nu_{i,k}$ (an independent observation of $\Xi_{i,k}^{\mathrm{dp}}$). We define the theoretical error between the average of sensitive attributes and the average of privatized attributes respectively computed by `BCGo` and the centralized averaging (`Cen`), i.e., the averaging operator $\frac{1}{n}\sum_{i=1}^{n}$:

$$e_k^{\texttt{BCGo}} = \mathbb{E}\left[\left(s_k - S_k^{\texttt{BCGo}}\right)^2\right], \quad e_k^{\texttt{Cen}} = \mathbb{E}\left[\left(s_k - S_k^{\texttt{Cen}}\right)^2\right],$$

where $s_k = \frac{1}{n}\sum_{i=1}^{n} d_i^k$, $S_k^{\texttt{Cen}} = \frac{1}{n}\sum_{i=1}^{n}\Xi_{i,k}^{\mathrm{dp}}$, $S_k^{\texttt{BCGo}} = \frac{\texttt{SiGo}\left((\Xi_{i,k-1}^{\mathrm{dp}})_{i=1}^n\right)}{\texttt{SiGo}\left((\Xi_{i,-1}^{\mathrm{dp}})_{i=1}^n\right)}$. Both errors are $\mathcal{O}\left(\frac{1}{n}\right)$ (proof omitted). We define the corresponding empirical errors as follows:

$$\hat{e}_k^{\texttt{BCGo}} = \left(\frac{1}{n}\sum_{i=1}^{n} d_i^k - \frac{\texttt{SiGo}\left((\nu_{i,k-1})_{i=1}^n\right)}{\texttt{SiGo}\left((\nu_{i,-1})_{i=1}^n\right)}\right)^2, \quad \hat{e}_k^{\texttt{Cen}} = \left(\frac{1}{n}\sum_{i=1}^{n} d_i^k - \frac{1}{n}\sum_{i=1}^{n}\nu_{i,k}\right)^2.$$

Experiments

Experiment 1. *We compare the precision of the averages of privatized attributes respectively computed by* `BCGo` *and* `Cen`*. Secondly, we evaluate if the theoretical errors are tight (within factor of* 10*) with respect to their empirical counterparts. Thirdly, we investigate if the theoretical errors indeed display the tendency to decrease linearly in* n*.*

Experiment 2. *We compare the utility of the regression model respectively fitted using the averages computed by* `BCGo`*,* `SiGo`*, and* `Cen`*. In particular, we evaluate the MSE between true and predicted target values (of a test set of size* 10^3*).*

Our synthetic graph dataset is created by generating a power-law degree sequence and generating a graph that follows the configuration model. For real graph datasets, we use the graphs of the email network dataset (1005 vertices and 25571 edges) and the autonomous systems dataset (6474 vertices and 13895 edges), both of which are part of SNAP [4].

For every $i \in [n]$, let $y_i = \theta_0 + \theta_1 d_i^{-1} + \theta_2 d_i^{1/2} + \theta_3 d_i^2 + \xi_i^{\mathrm{reg}}$, where $\theta_0 = 0$, $\theta_1 = \theta_2 = \theta_3 = 1$, and $\sigma_{\mathrm{reg}} = 1$. Let $\delta = 10^{-5}$, $\mathrm{it}_{\mathrm{go}} = 2^{10}$, $d_{\min} = 3$, and $d_{\max} = 10^2$. We fix the number of experiment repetitions to $\mathrm{it}_{\mathrm{exp}} = 2^{10}$ (details of the randomization of experiments omitted).

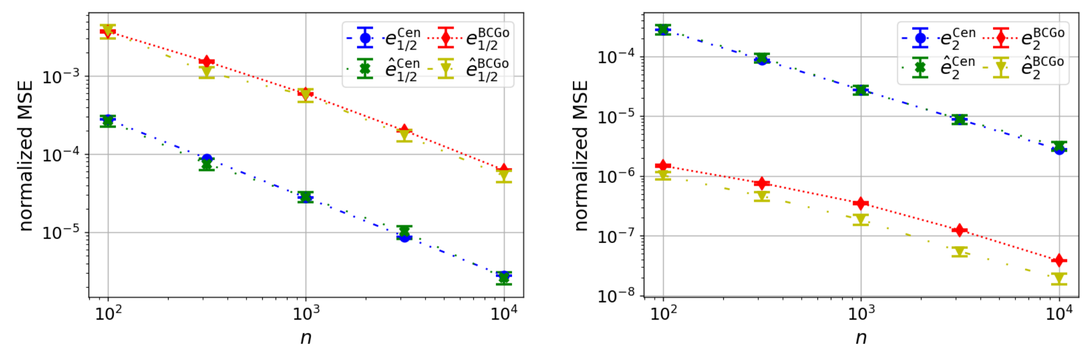

Figure 1: Results of Experiment 1 (synthetic dataset), when $\epsilon = 2^2$. $k = 1/2$ (left) and $k = 2$ (right) respectively corresponds to estimation of $\mu_{d^{1/2}}$ and μ_{d^2}. The MSE is normalized dividing by $(6\sigma_k^{\mathrm{dp}})^2$

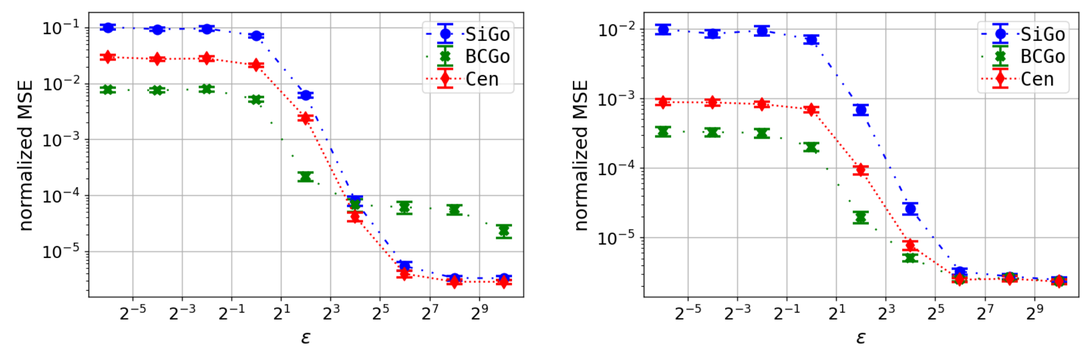

Figure 2: Results of Experiment 2 (the email network dataset on the left and the autonomous systems dataset on the right) The MSE is normalized dividing by the variance of true target values

Conclusion and Future Work

- `BCGo` can outperform `Cen` since $\frac{e_k^{\texttt{BCGo}}}{e_k^{\texttt{Cen}}} \approx \left(\frac{m+2}{m}\right)^2 \frac{\log(\frac{5(m+2)}{4\delta})}{\log(\frac{5m}{4\delta})}\left(\mu_{d^2}\left(\frac{\Delta(d_i^{k-1})}{\Delta(d_i^k)}\right)^2 + \frac{\mu_{d^k}}{\mu_d^2}\left(\frac{\Delta(d_i^{-1})}{\Delta(d_i^k)}\right)^2\right)$, and, for $k \geq 2$, we have $\frac{\Delta(d_i^{k-1})}{\Delta(d_i^k)} = \frac{d_{\max}^{k-1} - (d_{\max}-1)^{k-1}}{d_{\max}^k - (d_{\max}-1)^k}$
- `BCGo` can estimate the unbiased sample variance which is a U-statistic of degree 2 as, for $\mathbf{z} \in \mathbb{R}^n$, the unbiased sample variance is $\frac{1}{n(n-1)}\sum_{j>i}(z_i - z_j)^2 = \frac{1}{n-1}\sum_i^n (z_i - \mu_z)^2$, and we can compute μ_z and $\frac{1}{n}\sum_i^n (z_i - \mu_z)^2$. This way, it remains to identify a strategy to estimate $n-1$

References

[1] George Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *28th International Symposium on Theoretical Aspects of Computer Science, STACS 2011, March 10-12, 2011, Dortmund, Germany*, pages 57–68, 2011.

[2] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the accuracy of differentially private histograms through consistency. *Proc. VLDB Endow.*, 3(1):1021–1032, 2010.

[3] Márk Jelasity, Alberto Montresor, and Özalp Babaoglu. T-man: Gossip-based fast overlay topology construction. *Computer Networks*, 53(13):2321–2339, 2009.

[4] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection, June 2014.