Locally Private Bayesian Inference for Count Models

Aaron Schein [1] **Zhiwei Steven Wu** [2] **Alexandra Schofield** [3] **Mingyuan Zhou** [4] **Hanna Wallach** [5]

Abstract

We present a general and modular method for privacy-preserving Bayesian inference for Poisson factorization, a broad class of models that includes some of the most widely used models in the social sciences. Our method satisfies limited-precision local privacy, a generalization of local differential privacy that we introduce to formulate appropriate privacy guarantees for sparse count data. We present an MCMC algorithm that approximates the posterior distribution over the latent variables conditioned on data that has been locally privatized by the geometric mechanism. Our method is based on two insights: 1) a novel reinterpretation of the geometric mechanism in terms of the Skellam distribution and 2) a general theorem that relates the Skellam and Bessel distributions. We demonstrate our method's utility using two case studies that involve real-world email data. We show that our method consistently outperforms the commonly used naïve approach, wherein inference proceeds as usual, treating the locally privatized data as if it were not privatized.

1. Introduction

Data from social processes often take the form of discrete observations (e.g., edges in a social network, word tokens in an email). These observations often contain sensitive information. As more aspects of social interaction are digitally recorded, the opportunities for social scientific insights grow; however, so too does the risk of unacceptable privacy violations. As a result, there is a growing need to develop privacy-preserving data analysis methods. In practice, social scientists will be more likely to adopt these methods if doing so entails minimal change to their current methodology. Toward that end, we present a method for privacy-preserving Bayesian inference for Poisson factorization (Titsias, 2008; Cemgil, 2009; Zhou & Carin, 2012; Gopalan & Blei, 2013; Paisley et al., 2014), a broad class of models for inferring latent structure from discrete data. This class contains some of the most widely used models in the social sciences, including topic models for text corpora (Blei et al., 2003; Buntine & Jakulin, 2004; Canny, 2004), population models for genetic data (Pritchard et al., 2000), stochastic block models for social networks (Ball et al., 2011; Gopalan & Blei, 2013; Zhou, 2015), and tensor factorization of dyadic data (Welling & Weber, 2001; Chi & Kolda, 2012; Schmidt & Morup, 2013; Schein et al., 2015; 2016b). It further includes deep hierarchical models (Ranganath et al., 2015; Zhou et al., 2015), dynamic models (Charlin et al., 2015; Acharya et al., 2015; Schein et al., 2016a), and many others.

Our method assumes that observations are privatized (or noised) via a randomized response method before they are aggregated into a data set for analysis. This ensures that no single location need ever store the non-privatized data set. We introduce limited-precision local privacy (LPLP)—a generalization of local differential privacy—in order to formulate appropriate privacy guarantees for sparse count data. We focus specifically on the geometric mechanism of Ghosh et al. (2012) and prove that it is a mechanism for LPLP.

Under local privacy, a data analysis algorithm sees only the privatized data set. Inferring latent structure (including model parameters) that accurately reflects the non-privatized data set is therefore a key statistical challenge. One option is a naïve approach, wherein inference proceeds as usual, treating the privatized data set as if it were not privatized. In the context of maximum likelihood estimation, the naïve approach has been shown to exhibit pathologies when observations are discrete or count-valued. Researchers have therefore advocated for treating the non-privatized observations as latent variables to be inferred (Yang et al., 2012; Karwa et al., 2014; Bernstein et al., 2017; Bernstein & Sheldon, 2018). We embrace this approach and extend it to Bayesian inference for Poisson factorization, where our goal is to approximate the locally private posterior distribution over the latent variables conditioned on the privatized data set and the randomized response method. Toward that goal, we present a Markov chain Monte Carlo (MCMC) algorithm that is asymptotically guaranteed to draw samples

[1] University of Massachusetts Amherst [2] University of Minnesota [3] Cornell University [4] University of Texas at Austin [5] Microsoft. Correspondence to: Aaron Schein <aschein@cs.umass.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

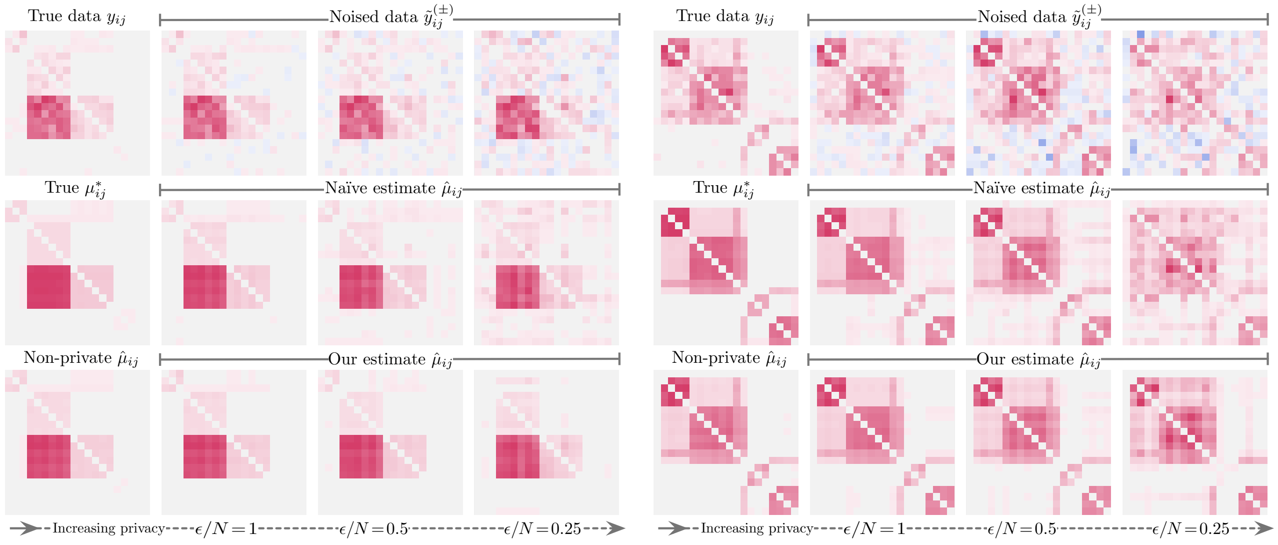


Figure 1. Block structure recovery: our method vs. the naïve approach. We generated the non-privatized data set synthetically. We then privatized the data set using three levels of noise. The top row depicts the data set, using red to denote positive counts and blue to denote negative counts. As the privacy level increases, the naïve approach overfits the noise and fails to recover the latent structure $\mu^\star_{ij}$, predicting high values even for sparse parts of the matrix. In contrast, our method recovers the latent structure, even for high levels of privacy.

from the locally private posterior. This algorithm is modular, allowing social scientists to extend (rather than replace) their implementations of non-private Poisson factorization.

Our main technical contribution is the derivation of a closed-form, computationally tractable way of "inverting" the randomized response method—i.e., sampling values of the non-privatized data set from its complete conditional. This derivation relies on two insights: 1) a novel reinterpretation of the geometric mechanism in terms of the Skellam distribution (Skellam, 1946) and 2) a general theorem that relates the Skellam and Bessel (Yuan & Kalbfleisch, 2000) distributions. These insights may be of independent interest.

We present two case studies applying our method to 1) topic modeling for text corpora and 2) overlapping community detection for social networks. Using real-world data from the Enron email corpus (Klimt & Yang, 2004), we show that our method consistently outperforms the naïve approach according to a variety of intrinsic and extrinsic evaluation metrics. We provide an illustrative example in figure 1. Finally, we note that our method sometimes outperforms Poisson factorization of the non-privatized data, suggesting that non-private Poisson factorization may be overfitting.

2. Background and problem formulation

Differential privacy. Differential privacy (Dwork et al., 2006) is a rigorous privacy criterion that guarantees that no single observation in a data set will have a significant influence on the information obtained by analyzing that data set.

Definition 1. *A randomized algorithm $\mathcal{A}(\cdot)$ satisfies ϵ-differential privacy if for all pairs of neighboring data sets Y and Y' that differ in only a single observation*

$$P\left(\mathcal{A}(Y) \in \mathcal{S}\right) \leq e^{\epsilon} P\left(\mathcal{A}(Y') \in \mathcal{S}\right), \tag{1}$$

for all subsets $\mathcal{S}$ in the range of $\mathcal{A}(\cdot)$.

Local differential privacy. We focus on local differential privacy, which we refer to as local privacy. Under this criterion, the observations remain private from even the data analysis algorithm. The algorithm sees only privatized versions of the observations, constructed by adding noise from specific distributions. The process of adding noise is known as randomized response—a reference to survey-sampling methods originally developed in the social sciences prior to the development of differential privacy (Warner, 1965). Satisfying this criterion means that no single location (e.g., a centralized server) need ever store the non-privatized data set.

Definition 2. *A randomized response method $\mathcal{R}(\cdot)$ is ϵ-private if for all pairs of observations $y, y' \in \mathcal{Y}$*

$$P\left(\mathcal{R}(y) \in \mathcal{S}\right) \leq e^{\epsilon} P\left(\mathcal{R}(y') \in \mathcal{S}\right), \tag{2}$$

for all subsets $\mathcal{S}$ in the range of $\mathcal{R}(\cdot)$. If a data analysis algorithm sees only the observations' ϵ-private responses, then the data analysis itself satisfies ϵ-local privacy.

The meaning of "observation" in definitions 1 and 2 varies depending on the context. For example, in the context of topic modeling, an observation is an entire document—i.e., a vector of counts representing the number of times each word type occurs in that document. To guarantee local privacy, a randomized response method must satisfy the condition in equation 2 for all pairs of observations. This typically involves adding noise that scales with the maximum difference between any pair of observations $N^{(\max)} = \max_{y,y'} \|y - y'\|_1$. When an observation is a document, $N^{(\max)}$ can be prohibitively large, meaning that the noise can overwhelm the signal in the data set. This challenge motivates the following alternative formulation of local privacy.

Limited-precision local privacy. Local privacy requires that a randomized response method render indistinguishable pairs of observations that are arbitrarily different. In some contexts, this requirement is unnecessarily strong. For example, the author of a document may only wish to conceal the occurrence of a handful of word types. To achieve this goal, a randomized response method need only render indistinguishable pairs of *similar* documents, such as a document in which those word types occur and an otherwise-identical document in which they do not. We operationalize this kind of privacy guarantee by generalizing definition 2 as follows.

Definition 3. *For any positive integer N, we say that a randomized response method $\mathcal{R}(\cdot)$ is (N, ϵ)-private if for all pairs of observations $y, y' \in \mathcal{Y}$ such that $\|y - y'\|_1 \leq N$*

$$P\left(\mathcal{R}(y) \in \mathcal{S}\right) \leq e^{\epsilon} P\left(\mathcal{R}(y') \in \mathcal{S}\right), \tag{3}$$

for all subsets $\mathcal{S}$ in the range of $\mathcal{R}(\cdot)$. If a data analysis algorithm sees only the observations' (N, ϵ)-private responses, then the data analysis itself satisfies (N, ϵ)-limited-precision local privacy. If $\|y\|_1 \leq N$ for all $y \in \mathcal{Y}$, then (N, ϵ)-limited-precision local privacy implies ϵ-local privacy.

Limited-precision local privacy (LPLP) is the local privacy analog of limited-precision differential privacy, originally proposed by Flood et al. (2013) and subsequently used to privatize analyses of geographic location data (Andrés et al., 2013) and financial network data (Papadimitriou et al., 2017). Like profile-based privacy (Geumlek & Chaudhuri, 2019), LPLP generalizes local privacy by flexibly specifying pairs of observations to be rendered indistinguishable. In section 3, we describe the geometric mechanism of Ghosh et al. (2012) and prove that it is a mechanism for LPLP.

Differentially private Bayesian inference. In Bayesian statistics, we begin with a probabilistic model $\mathcal{M}$ that relates observable variables Y to latent variables Z via a joint distribution $P_{\mathcal{M}}(Y, Z)$. The goal of inference is then to compute the posterior distribution $P_{\mathcal{M}}(Z \,|\, Y)$ over the latent variables conditioned on observed values of Y. The posterior is almost always analytically intractable and thus inference involves approximating it. The two most common methods of approximate Bayesian inference are variational inference, wherein we fit the parameters of an approximating distribution $Q(Z \,|\, Y)$, and Markov chain Monte Carlo (MCMC), wherein we approximate the posterior with a finite set of samples $\{Z^{(s)}\}_{s=1}^{S}$ generated via a Markov chain whose stationary distribution is the exact posterior.

We can conceptualize Bayesian inference as a randomized algorithm $\mathcal{A}(\cdot)$ that returns an approximation to the posterior distribution $P_{\mathcal{M}}(Z \,|\, Y)$. In general $\mathcal{A}(\cdot)$ does not satisfy ϵ-differential privacy. However, if $\mathcal{A}(\cdot)$ is an MCMC algorithm that returns a single sample from the posterior, it guarantees privacy (Dimitrakakis et al., 2014; Wang et al., 2015; Foulds et al., 2016; Dimitrakakis et al., 2017). Adding noise to posterior samples can also guarantee privacy (Zhang et al., 2016), though this set of noised samples $\{\tilde{Z}^{(s)}\}_{s=1}^{S}$ approximates some distribution $\tilde{P}_{\mathcal{M}}(Z \,|\, Y)$ that depends on ϵ and is different than the exact posterior (but close, in some sense, and equal when $\epsilon \to \infty$). For specific models, we can also noise the transition kernel of the MCMC algorithm to construct a Markov chain whose stationary distribution is again not the exact posterior, but something close that guarantees privacy (Foulds et al., 2016). We can also take an analogous approach to privatize variational inference, wherein we add noise to the sufficient statistics computed in each iteration (Park et al., 2016).

Locally private Bayesian inference. We first formalize the general objective of Bayesian inference under local privacy. Given a generative model $\mathcal{M}$ for non-privatized data set Y and latent variables Z with joint distribution $P_{\mathcal{M}}(Y, Z)$, we further assume a randomized response method $\mathcal{R}(\cdot)$ that generates privatized data sets: $\tilde{Y} \sim P_{\mathcal{R}}(\tilde{Y} \,|\, Y)$. The inference goal is then to approximate the locally private posterior

$$\begin{aligned} P_{\mathcal{M},\mathcal{R}}(Z \,|\, \tilde{Y}) &= \mathbb{E}_{P_{\mathcal{M},\mathcal{R}}(Y \,|\, \tilde{Y})}\left[P_{\mathcal{M}}(Z \,|\, Y)\right] \\ &= \int P_{\mathcal{M}}(Z \,|\, Y)\, P_{\mathcal{M},\mathcal{R}}(Y \,|\, \tilde{Y})\, dY. \end{aligned} \tag{4}$$

This distribution correctly characterizes our uncertainty about the latent variables Z, conditioned on all of our observations and assumptions—i.e., the privatized data set $\tilde{Y}$, the model $\mathcal{M}$, and the randomized response method $\mathcal{R}$. The expansion in equation 4 shows that this posterior implicitly treats the non-privatized data set Y as a latent variable and marginalizes over it using the mixing distribution $P_{\mathcal{M},\mathcal{R}}(Y \,|\, \tilde{Y})$ which is itself a posterior that characterizes our uncertainty about Y. The key observation here is that if we can generate samples from $P_{\mathcal{M},\mathcal{R}}(Y \,|\, \tilde{Y})$, we can use them to approximate the expectation in equation 4, assuming that we already have a method for approximating the non-private posterior $P_{\mathcal{M}}(Z \,|\, Y)$. In the context of MCMC, iteratively re-sampling values of the non-privatized data set from its complete conditional—i.e., $Y^{(s)} \sim P_{\mathcal{M},\mathcal{R}}(Y \,|\, Z^{(s-1)}, \tilde{Y})$—and then re-sampling values of the latent variables—i.e., $Z^{(s)} \sim P_{\mathcal{M}}(Z \,|\, Y^{(s)})$—constitutes a Markov chain whose stationary distribution is $P_{\mathcal{M},\mathcal{R}}(Z, Y \,|\, \tilde{Y})$. In scenarios where we already have derivations and implementations for sampling from $P_{\mathcal{M}}(Z \,|\, Y)$, we need only be able to sample efficiently from $P_{\mathcal{M},\mathcal{R}}(Y \mid Z, \tilde{Y})$ in order to obtain a locally private Bayesian inference algorithm; whether we can do this efficiently depends heavily on our choice of $\mathcal{M}$ and $\mathcal{R}$.

We note that the objective of Bayesian inference under local privacy, as defined in equation 4, is similar to that of Williams & McSherry (2010), who identify their key barrier to inference as being unable to analytically form the marginal likelihood that links the privatized data set to Z:

$$P_{\mathcal{M},\mathcal{R}}(\tilde{Y} \mid Z) = \int P_{\mathcal{R}}(\tilde{Y} \mid Y)\, P_{\mathcal{M}}(Y \mid Z)\, dY. \quad (5)$$

In the next sections, we show that if $\mathcal{M}$ is a Poisson factorization model and $\mathcal{R}$ is the geometric mechanism, then we can analytically form an augmented version of this marginal likelihood and derive an MCMC algorithm that samples efficiently from the locally private posterior in equation 4.

3. Locally private Poisson factorization

In this section, we describe a model $\mathcal{M}$—i.e., Poisson factorization—and a randomized response method $\mathcal{R}$—i.e., the geometric mechanism—each of which is a natural choice for count data. We prove two theorems about the geometric mechanism: 1) it is a mechanism for LPLP and 2) it can be re-interpreted in terms of the Skellam distribution (Skellam, 1946). We rely on the second theorem to show that Poisson factorization and the geometric mechanism combine to yield a novel generative process for privatized count data, which we then exploit in section 4 to derive our MCMC algorithm.

$\mathcal{M}$**: Poisson factorization.** We assume that Y is a data set that consists of counts, each of which $y_{\mathbf{i}} \in \mathbb{Z}_+$ is an independent Poisson random variable $y_{\mathbf{i}} \sim \text{Pois}(\mu_{\mathbf{i}})$ where the rate parameter $\mu_{\mathbf{i}}$ is defined to be a deterministic function of the latent variables Z. The subscript $\mathbf{i}$ is a multi-index. In Poisson matrix factorization, $\mathbf{i} \equiv (\mathbf{i}_1, \mathbf{i}_2)$; however, this notation also supports Poisson tensor factorization, where $\mathbf{i} \equiv (\mathbf{i}_1, \ldots, \mathbf{i}_M)$, and multiview models, where the multi-index may differ in the number of indices. This class of models includes some of the most widely used models in the social sciences, as described in section 1. In section 5, we present case studies involving two different models—specifically, the mixed-membership stochastic block model for social networks (Ball et al., 2011; Gopalan & Blei, 2013; Zhou, 2015) and latent Dirichlet allocation (Blei et al., 2003). Although both of these models are instances of Poisson matrix factorization, our method applies to any Poisson factorization model.

$\mathcal{R}$**: Geometric mechanism.** The two most commonly used randomized response methods in the differential privacy toolbox—the Gaussian and Laplace mechanisms—privatize observations by adding noise drawn from real-valued distributions. They are therefore unnatural choices for count data. Ghosh et al. (2012) introduced the geometric mechanism, which can be viewed as the discrete analog of the Laplace mechanism and involves adding integer-valued noise $\tau \in \mathbb{Z}$ drawn from the two-sided geometric distribution. The PMF for the two-sided geometric distribution is as follows:

$$2\text{Geo}(\tau; \alpha) = \frac{1-\alpha}{1+\alpha}\, \alpha^{|\tau|}. \quad (6)$$

Theorem 1. (Proof in appendix) *Let randomized response method $\mathcal{R}(\cdot)$ be the geometric mechanism with parameter α. Then for any positive integer N, and any pair of observations $y, y' \in \mathcal{Y}$ such that $\|y - y'\|_1 \leq N$, $\mathcal{R}(\cdot)$ satisfies*

$$P\left(\mathcal{R}(y) \in \mathcal{S}\right) \leq e^{\epsilon} P\left(\mathcal{R}(y') \in \mathcal{S}\right) \quad (7)$$

for all subsets $\mathcal{S}$ in the range of $\mathcal{R}(\cdot)$, where

$$\epsilon = N \ln\left(\frac{1}{\alpha}\right). \quad (8)$$

Therefore, for any positive integer N, the geometric mechanism with parameter α is an (N, ϵ)-private randomized response method with $\epsilon = N \ln\left(\frac{1}{\alpha}\right)$. If $\frac{\epsilon'}{N'} = \frac{\epsilon}{N}$, then the geometric mechanism with parameter α is also (N', ϵ')-private.

Theorem 2. (Proof in appendix) *A two-sided geometric random variable $\tau \sim 2Geo(\alpha)$ can be generated as follows:*

$$\tau \sim \textit{Skel}(\lambda^{(+)}, \lambda^{(-)}), \quad \lambda^{(*)} \sim \textit{Exp}(\tfrac{\alpha}{1-\alpha}), \quad (9)$$

where $Exp(\cdot)$ and $Skel(\cdot)$ are the exponential and Skellam distributions. The latter is the marginal distribution of the difference $\tau := g^{(+)} - g^{(-)}$ of two independent Poisson random variables $g^{(*)} \sim Pois(\lambda^{(*)})$, where $* \in \{+, -\}$.

Combining $\mathcal{M}$ and $\mathcal{R}$. We assume that each non-privatized count $y_{\mathbf{i}}$ is generated by $\mathcal{M}$ and then privatized by $\mathcal{R}$:

$$\tilde{y}_{\mathbf{i}}^{(\pm)} := y_{\mathbf{i}} + \tau_{\mathbf{i}}, \quad \tau_{\mathbf{i}} \sim 2\text{Geo}(\alpha), \quad (10)$$

where $\tilde{y}_{\mathbf{i}}^{(\pm)}$ is the privatized count, which we superscript with $(\pm)$ to denote that it may be non-negative or negative because the additive noise $\tau_{\mathbf{i}} \in \mathbb{Z}$ may itself be negative.

Via theorem 2, we can express the generative process for $\tilde{y}_{\mathbf{i}}^{(\pm)}$ in four equivalent ways, shown in figure 2, each of which provides a unique and necessary insight. Process 1 is a graphical representation of the generative process as described above. Process 2 represents the two-sided geometric noise in terms of a pair of Poisson random variables with exponentially distributed rates; in so doing, it reveals the auxiliary variables that facilitate inference. Process 3 represents the sum of the non-privatized count and the positive component of the noise as a single Poisson random variable $\tilde{y}_{\mathbf{i}}^{(+)} = y_{\mathbf{i}} + g_{\mathbf{i}}^{(+)}$. Process 4 marginalizes out the remaining Poisson random variables to yield a representation of $\tilde{y}_{\mathbf{i}}^{(\pm)}$ as an exponentially randomized Skellam random variable:

$$\tilde{y}_{\mathbf{i}}^{(\pm)} \sim \text{Skel}\left(\lambda_{\mathbf{i}}^{(+)} + \mu_{\mathbf{i}}, \lambda_{\mathbf{i}}^{(-)}\right), \quad \lambda_{\mathbf{i}}^{(*)} \sim \text{Exp}(\tfrac{\alpha}{1-\alpha}). \quad (11)$$

The derivations of generative processes 2, 3, and 4 rely on properties of the two-sided geometric, Skellam, and Poisson distributions. We provide these properties in the appendix.

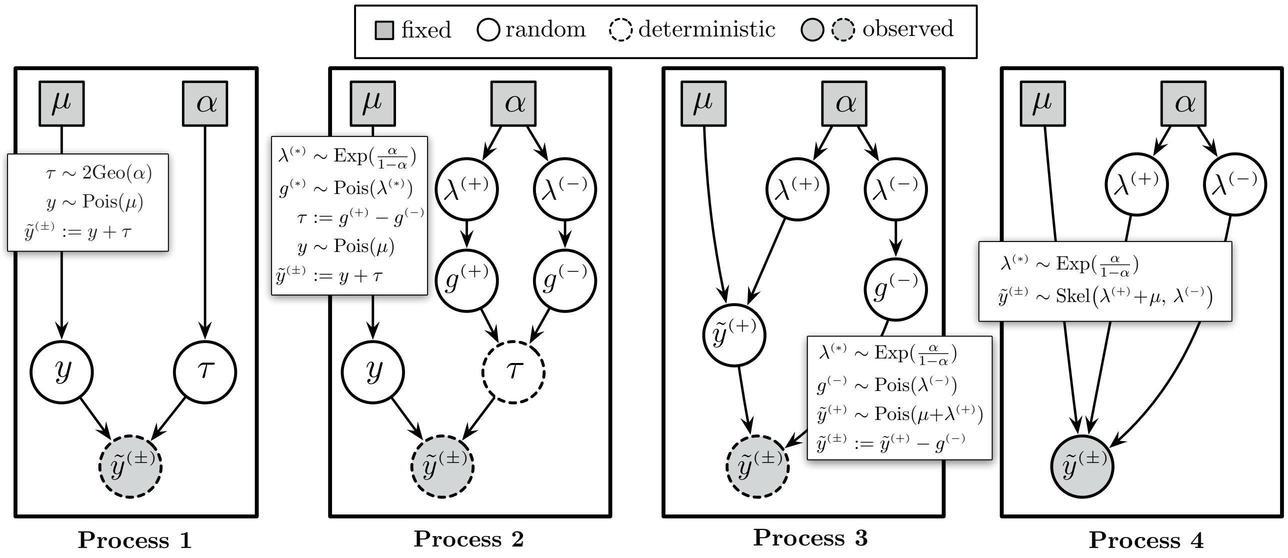

■ fixed
○ random
◌ deterministic
● ◌ observed
μ
α
$\tau \sim 2\mathrm{Geo}(\alpha)$
$y \sim \mathrm{Pois}(\mu)$
$\tilde{y}^{(\pm)} := y + \tau$
y
τ
$\tilde{y}^{(\pm)}$
Process 1
$\lambda^{(*)} \sim \mathrm{Exp}(\frac{\alpha}{1-\alpha})$
$g^{(*)} \sim \mathrm{Pois}(\lambda^{(*)})$
$\tau := g^{(+)} - g^{(-)}$
$y \sim \mathrm{Pois}(\mu)$
$\tilde{y}^{(\pm)} := y + \tau$
$\lambda^{(+)}$
$\lambda^{(-)}$
$g^{(+)}$
$g^{(-)}$
Process 2
$\lambda^{(*)} \sim \mathrm{Exp}(\frac{\alpha}{1-\alpha})$
$g^{(-)} \sim \mathrm{Pois}(\lambda^{(-)})$
$\tilde{y}^{(+)} \sim \mathrm{Pois}(\mu+\lambda^{(+)})$
$\tilde{y}^{(\pm)} := \tilde{y}^{(+)} - g^{(-)}$
$\tilde{y}^{(+)}$
Process 3
$\lambda^{(*)} \sim \mathrm{Exp}(\frac{\alpha}{1-\alpha})$
$\tilde{y}^{(\pm)} \sim \mathrm{Skel}(\lambda^{(+)}+\mu, \lambda^{(-)})$
Process 4

Figure 2. Four generative processes that yield the same marginal distribution $P(\tilde{y}^{(\pm)} \mid \mu, \alpha)$.

4. MCMC algorithm

Given a privatized data set $\tilde{Y}^{(\pm)}$, our inference goal is to approximate the locally private posterior using MCMC. To do this, we need to be able to sample values of the non-privatized data set Y from its complete conditional, as explained in section 2. By assuming that each privatized count $\tilde{y}_{\mathbf{i}}^{(\pm)}$ is a Skellam random variable, as in equation 11, we can exploit the following general theorem that relates the Skellam and Bessel (Yuan & Kalbfleisch, 2000) distributions.

Theorem 3. (Proof in appendix) *Consider two Poisson random variables $y_1 \sim Pois(\lambda^{(+)})$ and $y_2 \sim Pois(\lambda^{(-)})$. Their minimum $m := min\{y_1, y_2\}$ and their difference $\tau := y_1 - y_2$ are deterministic functions of y_1 and y_2. However, if not conditioned on y_1 and y_2, the random variables m and τ can be marginally generated as follows:*

$$\tau \sim \textit{Skel}(\lambda^{(+)}, \lambda^{(-)}),\; m \sim \textit{Bes}\left(|\tau|, 2\sqrt{\lambda^{(+)}\lambda^{(-)}}\right). \quad (12)$$

Theorem 3 means that we can generate two independent Poisson random variables by first generating their difference τ and then their minimum m. Because $\tau = y_1 - y_2$, if τ is positive, then y_2 must be the minimum and thus $y_1 = \tau + m$. In practice, this means that if we only get to observe the difference of two Poisson-distributed counts, we can still "recover" the counts by sampling a Bessel auxiliary variable.

By assuming that $\tilde{y}_{\mathbf{i}}^{(\pm)} \sim \mathrm{Skel}(\lambda_{\mathbf{i}}^{(+)} + \mu_{\mathbf{i}}, \lambda_{\mathbf{i}}^{(-)})$ via theorem 2, we can sample an auxiliary Bessel random variable $m_{\mathbf{i}}$:

$$(m_{\mathbf{i}} \mid -) \sim \mathrm{Bes}\left(|\tilde{y}_{\mathbf{i}}^{(\pm)}|, 2\sqrt{(\lambda_{\mathbf{i}}^{(+)} + \mu_{\mathbf{i}})\lambda_{\mathbf{i}}^{(-)}}\right). \quad (13)$$

Yuan & Kalbfleisch (2000) give details of the Bessel distribution, which can be sampled efficiently (Devroye, 2002).

Via theorem 3, $m_{\mathbf{i}}$ represents the minimum of two latent Poisson random variables whose difference equals $\tilde{y}_{\mathbf{i}}^{(\pm)}$; these latent variables are depicted in process 3 of figure 2—i.e., $\tilde{y}_{\mathbf{i}}^{(\pm)} := \tilde{y}_{\mathbf{i}}^{(+)} - g_{\mathbf{i}}^{(-)}$ and $m_{\mathbf{i}} = \min\{\tilde{y}_{\mathbf{i}}^{(+)}, g_{\mathbf{i}}^{(-)}\}$. Given $\tilde{y}_{\mathbf{i}}^{(\pm)}$ and a sampled value of $m_{\mathbf{i}}$, we can compute $\tilde{y}_{\mathbf{i}}^{(+)}$, $g_{\mathbf{i}}^{(-)}$:

$$\begin{aligned} &\tilde{y}_{\mathbf{i}}^{(+)} := m_{\mathbf{i}},\;\; g_{\mathbf{i}}^{(-)} := \tilde{y}_{\mathbf{i}}^{(+)} - \tilde{y}_{\mathbf{i}}^{(\pm)} \;\text{ if } \tilde{y}_{\mathbf{i}}^{(\pm)} \leq 0 \\ &g_{\mathbf{i}}^{(-)} := m_{\mathbf{i}},\;\; \tilde{y}_{\mathbf{i}}^{(+)} := g_{\mathbf{i}}^{(-)} + \tilde{y}_{\mathbf{i}}^{(\pm)} \;\text{ otherwise.} \end{aligned} \quad (14)$$

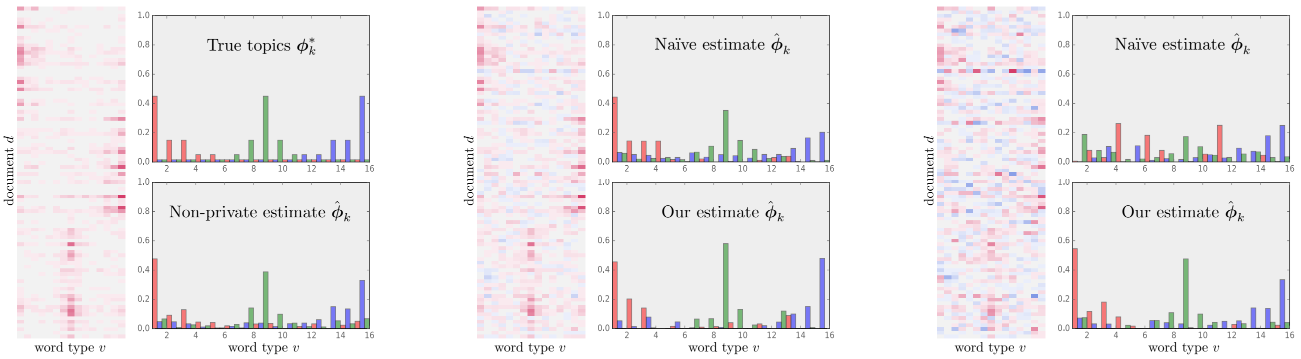
Because $\tilde{y}_{\mathbf{i}}^{(+)} = y_{\mathbf{i}} + g_{\mathbf{i}}^{(+)}$ is itself the sum of two independent Poisson random variables, we can then sample $y_{\mathbf{i}}$ from its conditional posterior, which is a binomial distribution:

$$(y_{\mathbf{i}} \mid -) \sim \mathrm{Binom}\left(\tilde{y}_{\mathbf{i}}^{(+)}, \frac{\mu_{\mathbf{i}}}{\mu_{\mathbf{i}} + \lambda_{\mathbf{i}}^{(+)}}\right). \quad (15)$$

Equations 13 through 15 constitute a way to draw samples from $P_{\mathcal{M},\mathcal{R}}(y_{\mathbf{i}} \mid \mu_{\mathbf{i}}, \tilde{y}_{\mathbf{i}}^{(\pm)}, \boldsymbol{\lambda}_{\mathbf{i}})$. We can also sample the auxiliary variables $\lambda_{\mathbf{i}}^{(+)}, \lambda_{\mathbf{i}}^{(-)}$ from their complete conditional:

$$(\lambda_{\mathbf{i}}^{(*)} \mid -) \sim \Gamma\left(1 + g_{\mathbf{i}}^{(*)}, \frac{\alpha}{1-\alpha} + 1\right). \quad (16)$$

Iteratively re-sampling $y_{\mathbf{i}}$ and $\boldsymbol{\lambda}_{\mathbf{i}}$ constitutes a chain whose stationary distribution over $y_{\mathbf{i}}$ is $P_{\mathcal{M},\mathcal{R}}(y_{\mathbf{i}} \mid \mu_{\mathbf{i}}, \tilde{y}_{\mathbf{i}}^{(\pm)})$, as desired. Given a sampled non-privatized data set Y, we can then sample values of the latent variables Z from their complete conditionals, which are the same as in non-private Poisson factorization. Equations 13–16, along with the complete conditionals for Z, define an efficient MCMC algorithm that is asymptotically guaranteed to generate samples from the locally private posterior $P_{\mathcal{M},\mathcal{R}}(Z \mid \tilde{Y}^{(\pm)})$.



(a) No noise.

(b) Low noise.

(c) High noise.

Figure 3. Topic recovery: our method vs. the naïve approach. (a) We generated the non-privatized data set synthetically so that the true topics were known. We then privatized the data set using (b) a low noise level and (c) a high noise level. The heatmap in each subfigure visualizes the data set, using red to denote positive counts and blue to denote negative counts. With a high level of noise, the naïve approach overfits the noise and therefore fails to recover the true topics. In contrast, our method is still able to recover the true topics.

5. Case studies

We present two case studies applying our method to 1) topic modeling for text corpora and 2) overlapping community detection for social networks. In each one, we formulate privacy guarantees, ground them in examples, and demonstrate our method's utility using synthetic and real-world data.

Real-world data. For our experiments using real-world data, we derived count matrices from the Enron email corpus (Klimt & Yang, 2004). For the topic modeling case study, we randomly selected $D = 10,000$ emails with at least 50 word tokens. We limited the vocabulary to $V = 10,000$ word types by selecting the most frequent word types with document frequencies less than 0.3. For the community detection case study, we obtained a $V \times V$ adjacency matrix Y where y_{ij} is the number of emails sent from actor i to actor j. We included an actor if they sent at least one email and sent or received at least one hundred emails, yielding $V = 161$ actors. When an email included multiple recipients, we incremented the corresponding counts by one.

Reference methods. We compare the performance of our method to two references methods: 1) non-private Poisson factorization of the non-privatized data and 2) the naïve approach, wherein inference treats the privatized data as if it were not privatized. The naïve approach must first truncate any negative counts $\tilde{y}_{\mathbf{i}}^{(\pm)} < 0$ to zero and thus implicitly uses the *truncated* geometric mechanism (Ghosh et al., 2012).

Performance measures. Each method uses MCMC to approximate the posterior with a set of S samples of the latent variables. We can therefore use these samples to approximate each method's posterior expectation of $\mu_{\mathbf{i}}$ as follows:

$$\hat{\mu}_{\mathbf{i}} = \frac{1}{S} \sum_{s=1}^{S} \mu_{\mathbf{i}}^{(s)} \approx \mathbb{E}_{P_{\mathcal{M},\mathcal{R}}(Z \mid \tilde{Y}^{(\pm)})} [\mu_{\mathbf{i}}] \,. \tag{17}$$

We can then calculate the mean absolute error (MAE) of $\hat{\mu}_{\mathbf{i}}$ with respect to $y_{\mathbf{i}}$—i.e., the reconstruction error. In the topic modeling case study, we also compare the quality of the inferred topics using two standard metrics: 1) normalized pointwise mutual information (NPMI; Lau et al., 2014) and 2) coherence (Mimno et al., 2011). We use the non-privatized data set as the reference corpus for both. For each metric, we use only the top ten word types for each topic, and we average the topics' scores to yield a single value.

5.1. Case study 1: Topic modeling

Topic models (Blei et al., 2003) are widely used in the social sciences (Ramage et al., 2009; Grimmer & Stewart, 2013; Mohr & Bogdanov, 2013; Roberts et al., 2013) to characterize the high-level thematic structure of text corpora via latent "topics"—i.e., probability distributions over the word types in some vocabulary. However, in many scenarios, documents can contain sensitive information, so people may be unwilling to share them without privacy guarantees.

Limited-precision local privacy. In the context of topic modeling, a data set Y is a $D \times V$ count matrix, where each element $y_{dv} \in \mathbb{Z}_+$ represents the number of times that word type $v \in [V]$ occurs in document $d \in [D]$. It is natural to consider each document $\boldsymbol{y}_d \equiv (y_{d1}, \ldots, y_{dV})$ to be a single observation. Under LPLP, N determines the neighborhood of documents within which ϵ-local privacy holds. For example, if $N = 4$, then a document in which a word type occurs four times and an otherwise-identical document in which it does not occur at all would be rendered indistinguishable after privatization, assuming ϵ is small.

Poisson factorization. Latent Dirichlet allocation (Blei et al., 2003), the most commonly used topic model, can be thought of as a special case of Poisson factorization where Y is a $D \times V$ count matrix and $\mu_{dv} = \sum_{k=1}^{K} \theta_{dk}\, \phi_{kv}$. The factor θ_{dk} represents how much topic k is used in document

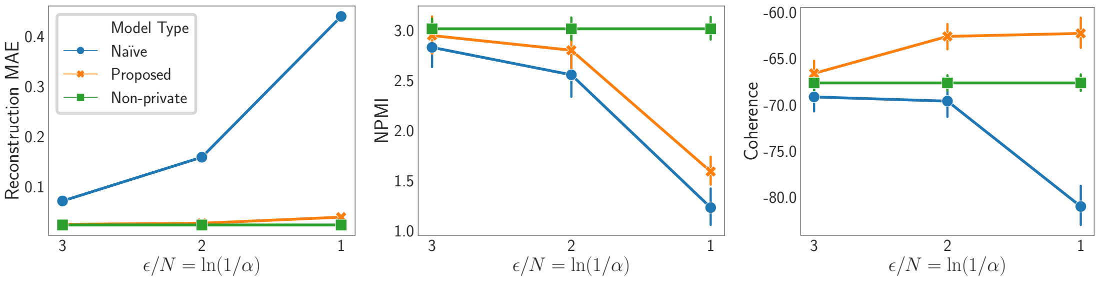


Figure 4. Each subplot compares our method to the two reference methods for increasing levels of privacy. The left subplot depicts reconstruction error (lower values are better), the center subplot depicts NPMI (higher values are better), and the right subplot depicts coherence (higher values are better). Our method has lower reconstruction error and higher quality topics than the naïve approach.

d, while the factor ϕ_{kv} represents how much word type v is used in topic k. The set of latent variables is thus $Z = \{\Theta, \Phi\}$, where Θ and Φ are $D \times K$ and $K \times V$ non-negative, real-valued matrices, respectively. It is standard to assume independent gamma priors over the factors—i.e., $\theta_{dk}, \phi_{kv} \sim \Gamma(a_0, b_0)$, where a_0 and b_0 are shape and rate hyperparameters respectively. We set $a_0 = 0.1$ and $b_0 = 1$.

Experiments using synthetic data. We generated a synthetic data set of $D = 90$ documents, with $K = 3$ topics and $V = 15$ word types. We set Φ^* so that the topics were well separated, with each putting the majority of its mass on five different word types. We also ensured that the documents were well separated into three equal groups of thirty, with each document putting the majority of its mass on a different topic. We then sampled a data set $y^*_{dv} \sim \text{Pois}(\mu^*_{dv})$ where $\mu^*_{dv} = \sum_{k=1}^{K} \theta^*_{dk} \phi^*_{kv}$. We then generated a heterogeneously-noised data set by sampling the d^{th} document's noise level $\alpha_d \sim \text{Beta}\big(c\,\alpha_0,\, c\,(1-\alpha_0)\big)$ from a Beta distribution with mean α_0 and concentration parameter $c = 10$ and then sampling $\tau_{dv} \sim 2\text{Geo}(\alpha_d)$ for each word type v. We repeated this for a small and large value of α_0. For each method and data set, we ran 6,000 MCMC iterations, saving every 25th sample after the first 1,000. We selected $\hat{\Phi}$ to be the sample from the posterior with the highest joint probability. (Due to label switching, we could not average samples of Φ.) Following Newman et al. (2009), we then aligned the topic indices of $\hat{\Phi}$ to Φ^* using the Hungarian bipartite matching algorithm. We visualize the results in figure 3. The naïve approach performs poorly at recovering the topics under high levels of privacy.

Experiments using real-world data. We used three privacy levels $\epsilon/N \in \{3, 2, 1\}$. We generated five privatized data sets for each privacy level by adding noise drawn from a two-sided geometric distribution with $\alpha = -\exp(\epsilon/N)$ independently to each element of the non-privatized data set. We applied our method and naïve approach to each of the fifteen privatized data sets and applied non-privatized Poisson factorization to the non-privatized data set. For each method and data set, we used $K = 50$ topics and ran 7,500 iterations of MCMC, saving every 100th sample of the latent variables after the first 2,500. We used the fifty saved samples to compute $\hat{\mu}_{dv} = \frac{1}{S} \sum_{s=1}^{S} \sum_{k=1}^{K} \theta^{(s)}_{dk} \phi^{(s)}_{kv}$. We also computed NPMI and coherence using each saved sample, and averaged the resulting values over the samples.

Results. We find that our method has both lower reconstruction error and higher quality topics than the naïve approach. The reconstruction error for each method and data set is shown in left subplot of figure 4. Our method has almost the same reconstruction error as non-private Poisson factorization of the non-privatized data. In contrast, the naïve approach has a higher reconstruction error that increases dramatically as the privacy level increases. The center and right subplots of figure 4 depict NPMI and coherence, respectively. According to both metrics, our method yields higher quality topics than the naïve approach. Surprisingly, the topics inferred by our method have better coherence than the topics inferred by Poisson factorization of the non-privatized data. This result suggests that non-private Poisson factorization may be overfitting; in contrast, our method may avoid overfitting by attributing small counts to the added noise and ignoring them. Because NPMI places more emphasis on rarer word types, excluding less reliable rare word types from topics does not benefit the metric, as is reflected in the center subplot of figure 4.

5.2. Case study 2: Overlapping community detection

Organizations often want to know whether their employees are interacting as productively as possible. For example, are there missing links between their employees that, if present, would reduce duplication of effort? Do the “communities” that emerge naturally from employee interactions match the formal organizational structure? Although social scientists may be able gain answer such questions using employee interaction data, sharing such data increases the risk of privacy violations. Moreover, standard anonymization procedures can be reverse-engineered adversarially and thus do not provide privacy guarantees (Narayanan & Shmatikov, 2009).

Limited-precision local privacy. In this context, a data set Y is a $V \times V$ count matrix, where each element $y_{ij} \in \mathbb{Z}_+$

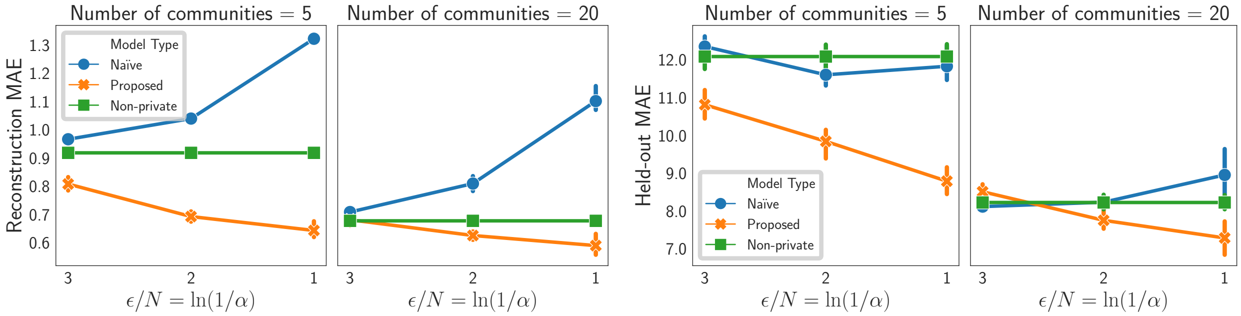


Figure 5. Each subplot compares our method to the two reference methods for increasing levels of privacy. The left subplot and the center left subplot depict reconstruction error (lower values are better) for $C = 5$ and $C = 20$; the center right subplot and the right subplot depict held-out reconstruction error (lower values are better) for $C = 5$ and $C = 20$. We provide results for $C = 10$ in the appendix.

represents the number of interactions from actor $i \in [V]$ to actor $j \in [V]$. It is natural to consider each count y_{ij} to be a single observation. Under LPLP, if $y_{ij} \leq N$, then its privatized version will be indistinguishable from the privatized version of $y_{ij} = 0$. In other words, an adversary will be unable to tell from the privatized count whether i had interacted with j at all, assuming ϵ is small. If $y_{ij} \gg N$, then only the exact number of interactions will be concealed.

Poisson factorization model. The mixed-membership stochastic block model (Ball et al., 2011; Gopalan & Blei, 2013; Zhou, 2015) is a special case of Poisson factorization where Y is a $V \times V$ count matrix and $\mu_{ij} = \sum_{c=1}^{C} \sum_{d=1}^{C} \theta_{ic}\, \theta_{jd}\, \pi_{cd}$. The factors θ_{ic} and θ_{jd} represent how much actors i and j participate in communities c and d, respectively, while the factor π_{cd} represents how much actors in community c interact with actors in community d. The set of latent variables is thus $Z = \{\Theta, \Pi\}$ where Θ and Π are $V \times C$ and $C \times C$ non-negative, real-valued matrices, respectively. As with latent Dirichlet allocation, it is standard to assume independent gamma priors over the factors—i.e., $\theta_{ic}, \pi_{cd} \sim \Gamma(a_0, b_0)$, where a_0 and b_0 are shape and rate hyperparameters, respectively. We set $a_0 = 0.1$ and $b_0 = 1$.

Experiments using synthetic data. We generated social networks of $V = 20$ actors with $C = 5$ communities. We randomly generated the true parameters $\theta^*_{ic}, \pi^*_{cd} \sim \Gamma(a_0, b_0)$ with $a_0 = 0.01$ and $b_0 = 0.5$ to encourage sparsity; doing so exaggerates the block structure in the network. We then sampled a data set $y_{ij} \sim \text{Pois}(\mu^*_{ij})$ and added noise $\tau_{ij} \sim \text{2Geo}(\alpha)$ for three increasing values of α. For each data set, we set $N = \hat{\mathbb{E}}[y_{ij}]$ and then set $\alpha = \exp(-\epsilon/N)$ for $\epsilon \in \{2.5, 1, 0.75\}$. For each method and data set, we ran 8,500 MCMC iterations, saving every 25th sample after the first 1,000 and using these samples to compute $\hat{\mu_{ij}}$. In figure 1, we visually compare our method's estimates of $\hat{\mu}_{ij}$ to those of the reference methods. The naïve approach overfits the noise, predicting high values even for sparse parts of the matrix. In contrast, our method approach reproduces the sparse block structure even for high levels of privacy.

Experiments using real-world data. We used the same experimental design that we used in the topic modeling case study. For each method and data set, we used $C \in \{5, 10, 20\}$ and we used the saved samples to compute $\hat{\mu}_{ij} = \frac{1}{S} \sum_{s=1}^{S} \sum_{c=1}^{C} \sum_{d=1}^{C} \theta^{(s)}_{ic} \theta^{(s)}_{jd} \pi^{(s)}_{cd}$. However, instead of computing NPMI and coherence, we ran link-prediction experiments. Specifically, prior to inference, we held out all elements of the count matrix that involved the top fifty senders or recipients. After inference, we used the saved samples to compute $\hat{\mu}_{ij}$, but only for the held-out elements, and then calculated the corresponding reconstruction error.

Results. We find that our method generally obtains lower reconstruction and lower held-out error than either the naïve approach or non-private Poisson factorization of the non-privatized data. The results for $C \in \{5, 20\}$ are shown in figure 5. We provide results for $C = 10$ in the appendix.

6. Conclusion and future work

We presented a general and modular method for privacy-preserving Bayesian inference for Poisson factorization. Our method satisfies limited-precision local privacy, a generalization of local differential privacy that we introduced to formulate appropriate privacy guarantees for sparse count data. In two case studies, we demonstrated that our method generally outperforms the commonly used naïve approach, wherein inference treats the privatized data as if it were not privatized. Surprisingly, our method also outperformed non-private Poisson factorization. In the context of topic modeling, it inferred more coherent topics; in the context of overlapping community detection, it obtained lower held-out error in link-prediction experiments. These findings are consistent with known connections between privacy-preserving mechanisms and regularization (Chaudhuri & Monteleoni, 2009). Because our method can attribute small counts to the added noise, it can therefore ignore them. In turn, it may be less susceptible to inferring spurious structure. Initial results suggest that $y_{\mathbf{i}} \sim \text{Skel}(\lambda^{(+)}_{\mathbf{i}} + \mu_{\mathbf{i}},\, \lambda^{(-)}_{\mathbf{i}})$ may be a robust alternative to non-private Poisson factorization. We therefore highlight this direction for future work.

Acknowledgments

We acknowledge the support of NSF awards #1812699 and #1652536. We also thank Dan Sheldon, David Mimno, and Antti Honkela. This research was partially conducted while Alexandra Schofield was a summer intern at Microsoft.

References

Acharya, A., Ghosh, J., and Zhou, M. Nonparametric Bayesian factor analysis for dynamic count matrices. arXiv:1512.08996, 2015.

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., and Palamidessi, C. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer &#38; Communications Security*, CCS '13, pp. 901–914, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2477-9.

Ball, B., Karrer, B., and Newman, M. E. J. Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103, 2011.

Bernstein, G. and Sheldon, D. R. Differentially private bayesian inference for exponential families. In *Advances in Neural Information Processing Systems*, pp. 2919–2929, 2018.

Bernstein, G., McKenna, R., Sun, T., Sheldon, D., Hay, M., and Miklau, G. Differentially private learning of undirected graphical models using collective graphical models. *arXiv preprint arXiv:1706.04646*, 2017.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3: 993–1022, 2003.

Buntine, W. and Jakulin, A. Applying discrete PCA in data analysis. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pp. 59–66, 2004.

Canny, J. GaP: a factor model for discrete data. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 122–129, 2004.

Cemgil, A. T. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009.

Charlin, L., Ranganath, R., McInerney, J., and Blei, D. M. Dynamic Poisson factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 155–162, 2015.

Chaudhuri, K. and Monteleoni, C. Privacy-preserving logistic regression. In *Advances in Neural Information Processing Systems*, pp. 289–296, 2009.

Chi, E. C. and Kolda, T. G. On tensors, sparsity, and nonnegative factorizations. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1272–1299, 2012.

Devroye, L. Simulating Bessel random variables. *Statistics & Probability Letters*, 57(3):249–257, 2002.

Dimitrakakis, C., Nelson, B., Mitrokotsa, A., and Rubinstein, B. I. P. Robust and private Bayesian inference. In *International Conference on Algorithmic Learning Theory*, pp. 291–305, 2014.

Dimitrakakis, C., Nelson, B., Zhang, Z., Mitrokotsa, A., and Rubinstein, B. Differential privacy for bayesian inference through posterior sampling. *Journal of Machine Learning Research*, 18(11):1–39, 2017.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, volume 3876, pp. 265–284, 2006.

Flood, M., Katz, J., Ong, S., and Smith, A. Cryptography and the economics of supervisory information: Balancing transparency and confidentiality. 2013.

Foulds, J., Geumlek, J., Welling, M., and Chaudhuri, K. On the theory and practice of privacy-preserving Bayesian data analysis. 2016.

Geumlek, J. and Chaudhuri, K. Profile-based privacy for locally private computations. *CoRR*, abs/1903.09084, 2019. URL http://arxiv.org/abs/1903.09084.

Ghosh, A., Roughgarden, T., and Sundararajan, M. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

Gopalan, P. K. and Blei, D. M. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 110(36):14534–14539, 2013.

Grimmer, J. and Stewart, B. M. Text as data: The promise and pitfalls fo automatic content analysis methods for political texts. *Political Analysis*, pp. 1–31, 2013.

Karwa, V., Slavković, A. B., and Krivitsky, P. Differentially private exponential random graphs. In *International Conference on Privacy in Statistical Databases*, pp. 143–155. Springer, 2014.

Klimt, B. and Yang, Y. The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*, pp. 217–226. Springer, 2004.

Lau, J. H., Newman, D., and Baldwin, T. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 530–539, 2014.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., and McCallum, A. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 262–272. Association for Computational Linguistics, 2011.

Mohr, J. and Bogdanov, P. (eds.). *Poetics: Topic Models and the Cultural Sciences*, volume 41. 2013.

Narayanan, A. and Shmatikov, V. De-anonymizing social networks. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pp. 173–187, 2009.

Newman, D., Asuncion, A., Smyth, P., and Welling, M. Distributed algorithms for topic models. *Journal of Machine Learning Research*, 10(Aug):1801–1828, 2009.

Paisley, J., Blei, D. M., and Jordan, M. I. Bayesian nonnegative matrix factorization with stochastic variational inference. In Airoldi, E. M., Blei, D. M., Erosheva, E. A., and Fienberg, S. E. (eds.), *Handbook of Mixed Membership Models and Their Applications*, pp. 203–222. 2014.

Papadimitriou, A., Narayan, A., and Haeberlen, A. Dstress: Efficient differentially private computations on distributed data. In *Proceedings of the Twelfth European Conference on Computer Systems*, EuroSys '17, pp. 560–574, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4938-3.

Park, M., Foulds, J., Chaudhuri, K., and Welling, M. Private topic modeling. *arXiv:1609.04120*, 2016.

Pritchard, J. K., Stephens, M., and Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

Ramage, D., Rosen, E., Chuang, J., Manning, C. D., and McFarland, D. A. Topic modeling for the social sciences. In *Neural Information Processing Systems Workshop on Applications for Topic Models*, 2009.

Ranganath, R., Tang, L., Charlin, L., and Blei, D. Deep exponential families. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pp. 762–771, 2015.

Roberts, M. E., Stewart, B. M., Tingley, D., and Airoldi, E. M. The structural topic model and applied social science. In *Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*, 2013.

Schein, A., Paisley, J., Blei, D. M., and Wallach, H. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1045–1054, 2015.

Schein, A., Wallach, H., and Zhou, M. Poisson–gamma dynamical systems. In *Advances in Neural Information Processing Systems 29*, pp. 5005–5013, 2016a.

Schein, A., Zhou, M., Blei, D. M., and Wallach, H. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *Proceedings of the 33rd International Conference on Machine Learning*, 2016b.

Schmidt, M. N. and Morup, M. Nonparametric Bayesian modeling of complex networks: an introduction. *IEEE Signal Processing Magazine*, 30(3):110–128, 2013.

Skellam, J. G. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A (General)*, 109:296, 1946.

Titsias, M. K. The infinite gamma–Poisson feature model. In *Advances in Neural Information Processing Systems 21*, pp. 1513–1520, 2008.

Wang, Y.-X., Fienberg, S., and Smola, A. Privacy for free: posterior sampling and stochastic gradient Monte Carlo. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 2493–2502, 2015.

Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Welling, M. and Weber, M. Positive tensor factorization. *Pattern Recognition Letters*, 22(12):1255–1261, 2001.

Williams, O. and McSherry, F. Probabilistic inference and differential privacy. In *Advances in Neural Information Processing Systems 23*, pp. 2451–2459, 2010.

Yang, X., Fienberg, S. E., and Rinaldo, A. Differential privacy for protecting multi-dimensional contingency table data: Extensions and applications. *Journal of Privacy and Confidentiality*, 4(1):5, 2012.

Yuan, L. and Kalbfleisch, J. D. On the Bessel distribution and related problems. *Annals of the Institute of Statistical Mathematics*, 52(3):438–447, 2000.

Zhang, Z., Rubinstein, B. I. P., and Dimitrakakis, C. On the differential privacy of Bayesian inference. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pp. 2365–2371, 2016.

Zhou, M. Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the 18th Conference on Artificial Intelligence and Statistics*, pp. 1135–1143, 2015.

Zhou, M. and Carin, L. Augment-and-conquer negative binomial processes. In *Advances in Neural Information Processing Systems 25*, pp. 2546–2554, 2012.

Zhou, M., Cong, Y., and Chen, B. The Poisson gamma belief network. In *Advances in Neural Information Processing Systems 28*, pp. 3043–3051, 2015.