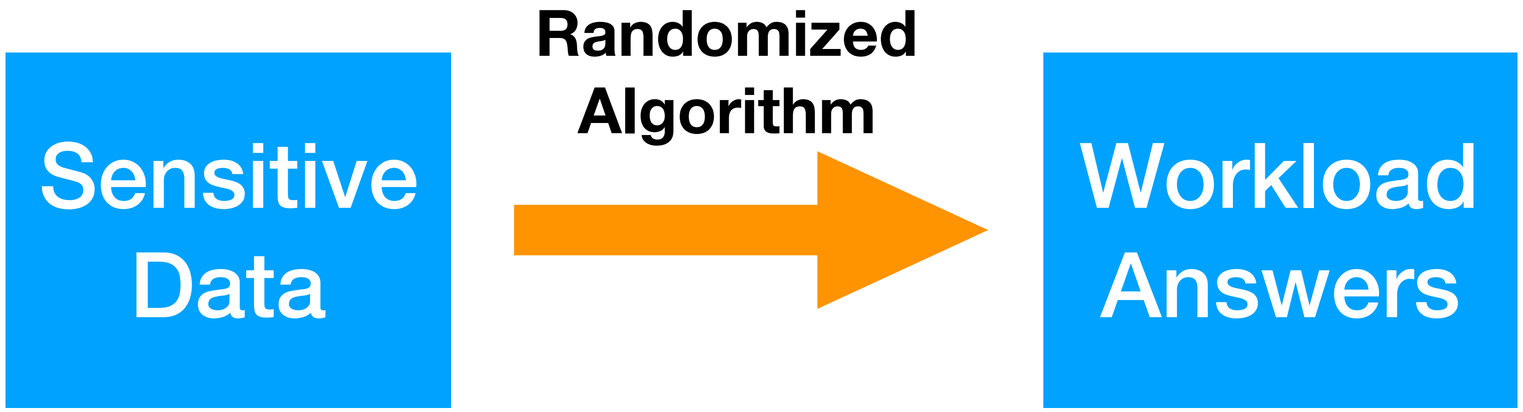
Graphical-model based estimation and inference for differential privacy

Ryan McKenna[1], Daniel Sheldon[1,2], Gerome Miklau[1]
[1]University of Massachusetts, Amherst
[2]Mount Holyoke College

Inference in Privacy Mechanisms

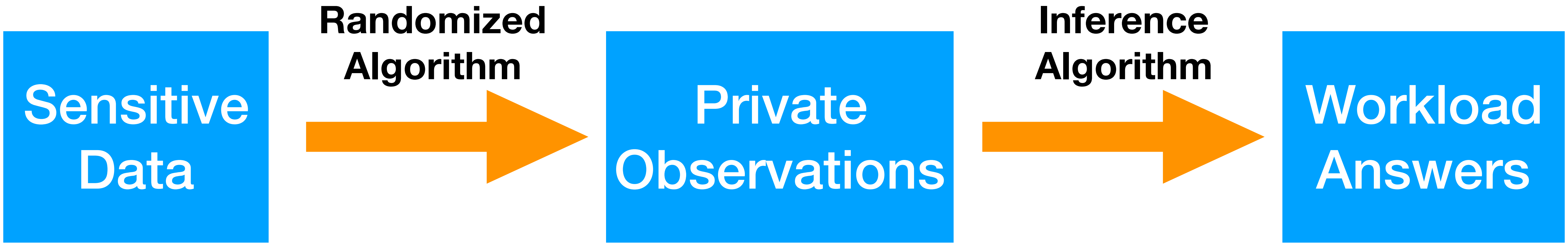


Sensitive Data

Randomized Algorithm

Workload Answers

Inference in Privacy Mechanisms



Sensitive Data

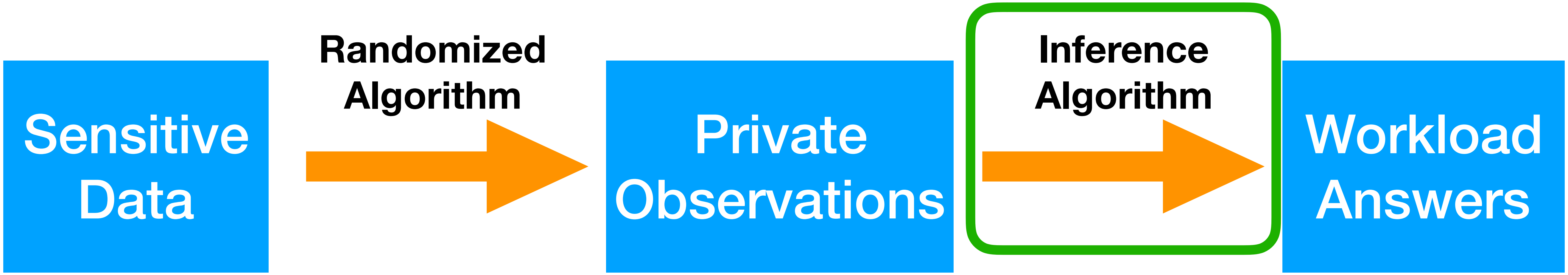
Randomized Algorithm

Private Observations

Inference Algorithm

Workload Answers

Inference in Privacy Mechanisms


Sensitive Data

Randomized Algorithm

Private Observations

Inference Algorithm

Workload Answers

- Existing techniques for inference either don’t scale or don’t extract the most utility from the private observations

- Proper inference has many benefits:
  - Resolves inconsistencies
  - Improves utility
  - Answers new queries
  - Supports synthetic data generation

UMASS AMHERST

Problem Statement

- Given:

  an unknown discrete data distribution $\mathbf{p} \in \mathbb{R}^{\mathrm{n}}$

  a query matrix $\mathbf{Q} \in \mathbb{R}^{\mathrm{m \times n}}$

- Our observation model is:

Random Laplace or Gaussian noise

$$\mathbf{y} = \mathbf{Q}\mathbf{p} + \varepsilon$$

- We want to recover an estimate of $\mathbf{p}$ from $\mathbf{y}$

$$\hat{\mathbf{p}} \in \arg\min_{\mathbf{p} \in S} ||\mathbf{Q}\mathbf{p} - \mathbf{y}||$$

Size of p is intractably large

UMASS AMHERST

Approach

- Reformulate problem to find a graphical model $\mathbf{p}_\theta$ instead

$$\hat{\theta} \in \arg\min_{\theta} ||\mathbf{Q}\mathbf{p}_{\theta} - \mathbf{y}||$$

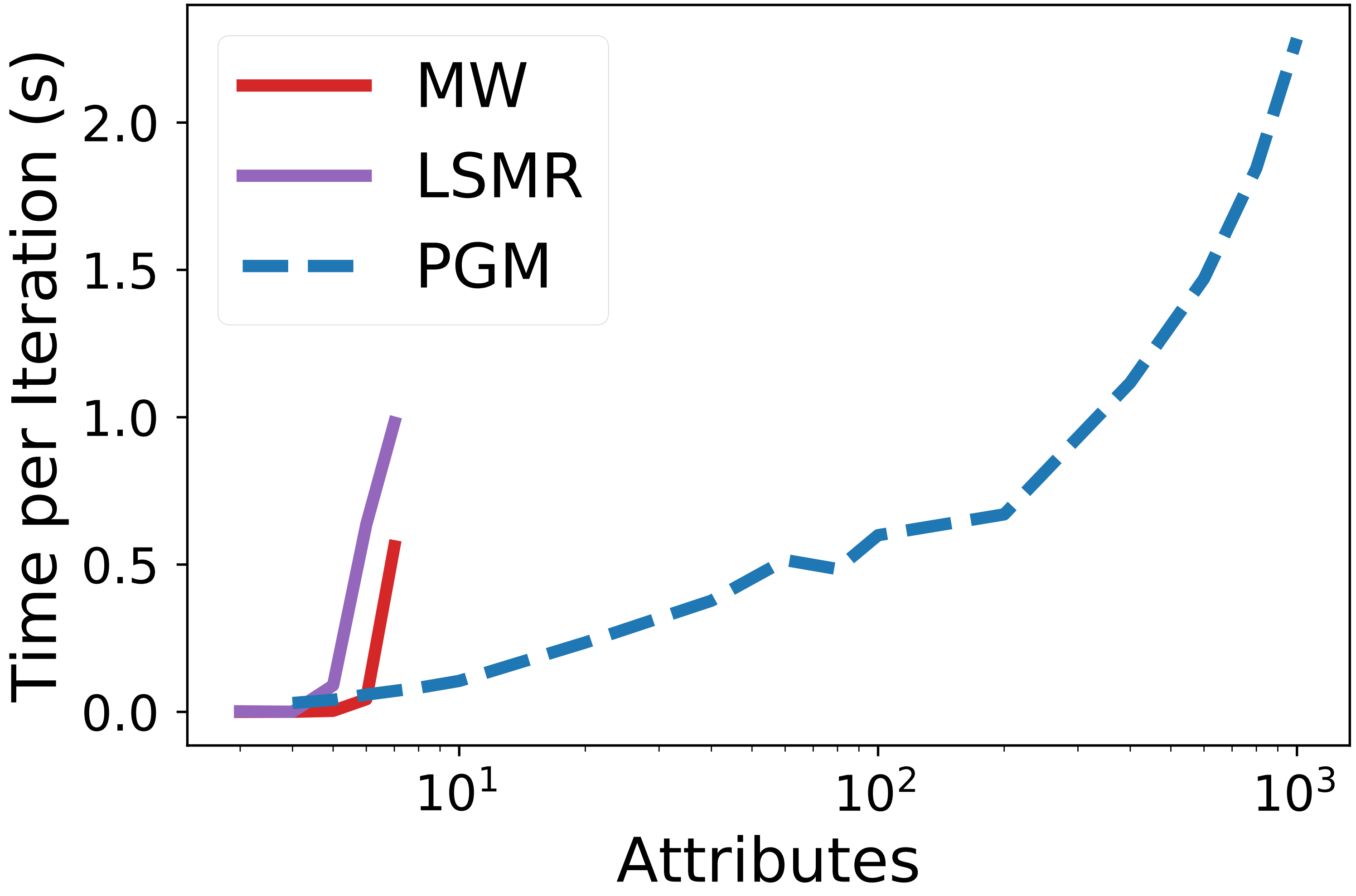
Much smaller than p

- If $\mathbf{Q}$ only depends on $\mathbf{p}$ though its marginals,
  - We can solve this problem efficiently
  - Solution to reformulated problem is the maximum entropy solution to the original problem

UMASS
AMHERST

Scalability Improvements of PGM

- Graphical-model inference scales much better than traditional approaches.

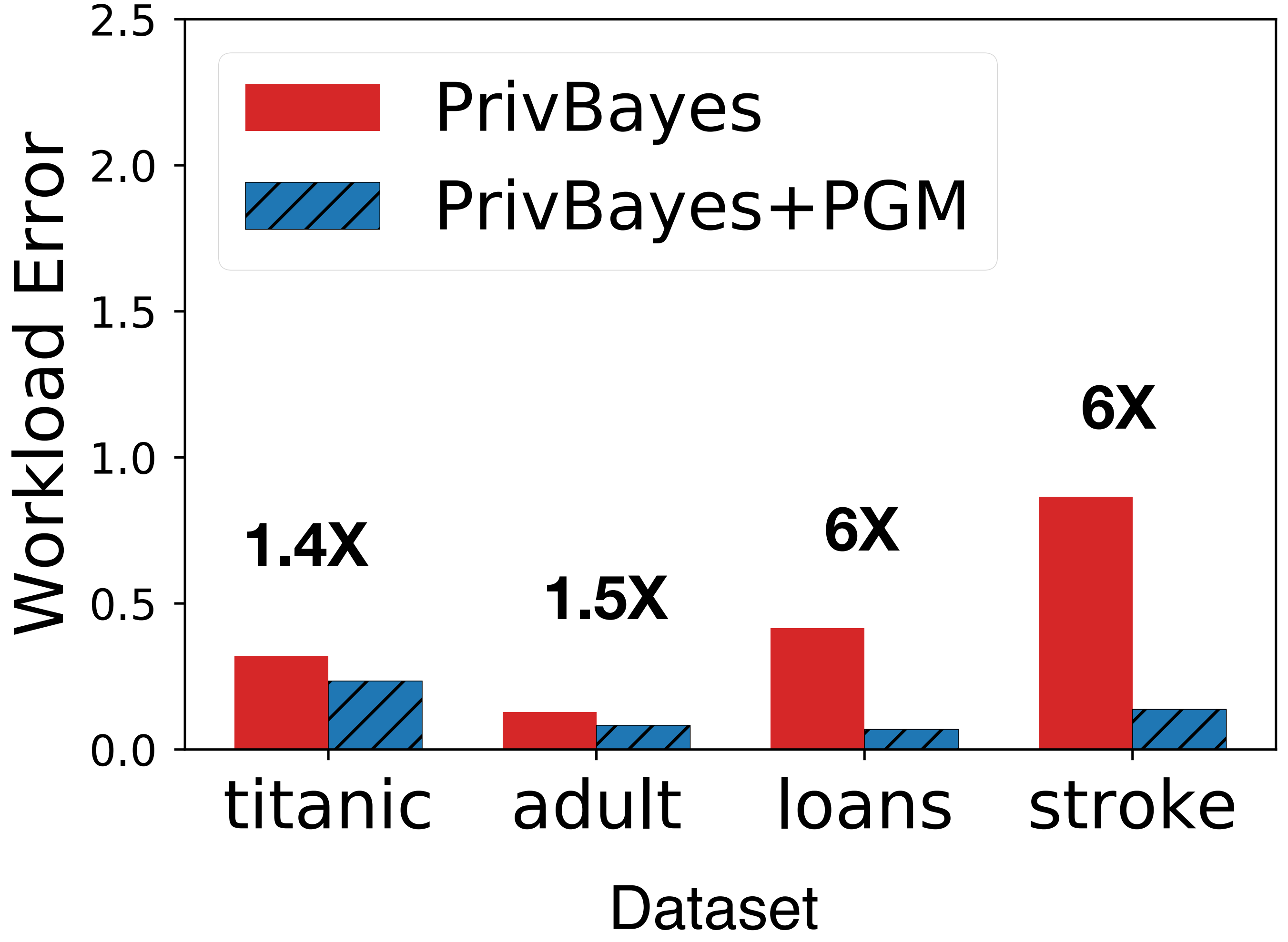


Traditional approaches fail at 10 dimensions

PGM scales to 1000 dimensions

Utility Improvements of PGM

- Graphical-model inference improves the utility of several state-of-the-art privacy mechanisms.



Error reduction up to 6X

We offer similar improvements for DualQuery, HDMM, and MWEM as well (see poster)

UMASS AMHERST

Graphical-model based estimation and inference for differential privacy

Poster #171

Code available on GitHub:

https://github.com/ryan112358/private-pgm

UMASS
AMHERST