Differential Privacy and Statistical Inference

Vishesh Karwa
Affiliation: Harvard University and Carnegie Mellon University

Sharing Social Network Data
Goal: Enable sharing of social network data under rigorous privacy guarantees and maintain data utility for statistical inference.
Motivating Example - Epidemiological studies on sexual networks → Survey number of partners: Degree Sequence
→ Estimate parameters, reconstruct typical networks, test hypothesis
Privacy Release “noisy” data
Utility Inference using “noisy” data

Exponential Random Graph Models (ERGM)
Any model class for a network $X$ can be parametrized in the form:

$$p_\theta(X = x) = \frac{\exp(\theta g(x))}{c(\theta, X)}, \quad x \in \mathcal{X}$$

→ $\theta \in \mathbb{R}^q$ a q-vector of parameters
→ $g(x)$ a q-vector of sufficient statistics
→ $c(\theta, X)$ distribution normalizing constant

Beyond Degree Sequences
What about other sufficient statistics?
What if we don’t know what set of sufficient statistics are needed?

Randomized Response
Old Wine in new Bottle

Population → Sample $X$ → Private Sample $Z$

Differentially Private Estimator:

$$\hat{\theta}_{DPS} = \arg\max_{\theta} \sum_{x \in \mathcal{X}} P(Z = z | X = x, \gamma) P(X; \theta)$$

The Degree Sequence ERGM

Theorem (Informal). If $g(X)$ is the degree sequence, there exists an efficient differentially private estimator of $\theta$ that is also asymptotically optimal, in particular, consistent and asymptotically normal.

The Degree Sequence ERGM

Theorem (Informal). The above differentially private estimator for any Exponential Random Graph model runs at only twice the computational cost of the non-private estimator.

Conclusions
→ Sharing relationship data for reproducibility and new scientific discoveries.
→ Design differentially private algorithms for statistical inference.
→ Privacy preserving inference (in some cases optimal) for random graph models.
→ Extensive experiments show that approach can work in practise.

References
→ Karwa, Krivitsky and Slavkovic, Sharing Social Networks using Edge Differential Privacy. Forthcoming.