Towards Differentially Private Inference on Exponential Random Graph Models

Alexander Goldberg

Abstract

This is my abstract.

Contents

Chapter 1: Introduction

This is my intro...

Chapter 2: Statistical Modeling of Network Data

An increasingly popular approach in quantitative analysis of networks is to fit statistical models to real world network data. Many of these models have generative interpretations, allowing researchers to understand the relative importance of multiple endogenous processes to the resulting structure of the network. The advantage of such an approach is best illustrated in contrast to computing statistics – like degree distributions, assortativity coefficients, and transitivity coefficients – to describe the network structure, without an explicit model of the network. While such metrics are incredibly useful in describing the structural properties of a given network, they cannot tease out the underlying processes that may give rise to such structures.

For example, one of the distinguishing characteristics of many real-world social networks is that they tend to have more triangles (sets of three connected nodes) than would be expected by drawing random edges of a graph [GKM09]. However, there are a number of different processes in the formation of a friend network that could give rise to this outcome. One potential explanation is the notion of "triangle closure," or the tendency for people to become friends with friends-of-friends, since they are easier to meet. Another, subtly different explanation, is that triangles arise out of "assortative matching," the propensity for people with the same attributes to become friends with one another, leading to more clustering in the network. Finally, a high number of triangles in a social network could arise for reasons of "sociality," the presence of only a few highly social individuals in the network, who are mutual friends to many people.

In order to consider what global or local processes best explain particular structures of a network, a statistical model of network data posits a probability distribution over the space of possible graphs (usually graphs with a fixed number of nodes.) The goal of inference over this distribution is to tune parameters of the distribution, such that the realized network is likely to be observed under the probability distribution.

A simple example of such a model is the Erdös-Rényi Random Graph Model, known as the $G(n, p)$ model, which proposes that edges are drawn independently with probability p between any two nodes of a network with n nodes. While this model has been studied in great depth by graph theorists, it does not capture many important features of real world networks, like the tendency for clustering or the power-law distribution of degrees of a graph, the pattern of many low-degree nodes and a few high-degree nodes.

In order to model more complicated structures in networks, a more general class of random graph models are Exponential Random Graph Models (sometimes known as p^* models), which we describe in Section 2.1. While these models arose out of the sociology literature, particularly in studying social networks, they have been applied to a broad range of problems, including analysis of interactions between proteins in the human body [RAS10], networks of neurons in the brain as people age [SDC+16], corporate management structures at Enron [UHH13], and the demographics of high school friendships [GKM09]. Thus, we study the specific ERGM model described in 2.2, both because it is one of the most widely used and generally applicable random graph models in practical network analysis and because, as we will see, it has robustness properties that motivate its amenability to analysis under differential privacy constraints.

2.1 Exponential Random Graph Models

Formally, a graph $G = (V, E)$ is defined by a set of vertices (or nodes) V, with $|V| = n$ and edges E, denoting the presence or absence of relationships between nodes. We will use the "adjacency matrix" representation of a graph, which we denote x, where $x_{ij} = 1$ if an edge exists between nodes i and j and $x_{ij} = 0$ otherwise. The models we consider are all defined over *undirected graphs*, so all the edges are bidirectional, and the adjacency matrix is therefore symmetric. Further, we consider graphs without self-loops, so that $x_{ii} = 0$ for all i.

We refer to the number of edges adjacent to node i as the *degree* of node i so $d_i = \sum_{j=1}^{n} x_{ij}$. Then, then the *degree distribution* of graph x is $D = (D_0, ..., D_{n-1})$ where $D_k = |\{i \in V : d_i = k\}|$.

Definition 2.1 (Exponential Random Graph [WP96])**.** A probability distribution over graphs of n vertices belongs to the family of *exponential random graph models* (henceforth referred to as ERGMs) if it takes the form:

$$\Pr(x|\theta) = \exp\left\{\theta^T u(x) - \psi(\theta)\right\}$$

where θ is a vector of parameters of the model, $u(x)$ is a vector of arbitrary sufficient statistics computed on graph x, and $\psi(\theta)$ is a normalization constant needed to ensure a valid probability distribution that integrate to one so that

$$\psi(\theta) = \log \sum_{x} \exp\left\{\theta^T u(x)\right\}$$

One advantage to this model is that it belongs to the *exponential family* of probability distributions, for which inference techniques are well studied in the statistics and machine learning literature. In general, the normalization constant $\psi(\theta)$ may be intractable to compute exactly since it requires summing over the space of all possible graphs on n vertices. Therefore, in practice, approximate inference methods, in particular sampling-based MCMC approaches, are used for parameter estimation of these models on realized data.

A further advantage of ERGMs, is that they describe a broad range of random graph models, with varying conditional dependence relationships between edges. For instance, the $G(n,p)$ graph we discussed can be viewed as an ERGM:

Example 2.1 ($G(n,p)$ graphs)**.** We can represent the Erdös-Rényi Random Graph ($G(n,p)$) model as an ERGM, by taking

$$u(x) = |E|, \quad \theta = \log \frac{p}{1-p}$$

$$\psi(\theta) = -\binom{n}{2} \log(1-p) = -\binom{n}{2} \log \frac{e^{-\theta}}{1+e^{-\theta}}$$

Then,

$$\begin{aligned} \Pr(x|\theta) &= \exp\left\{ |E| \log \frac{p}{1-p} + \binom{n}{2} \log(1-p) \right\} \\ &= p^{|E|}(1-p)^{\binom{n}{2}-|E|} \\ &= \prod_{i<j} p^{x_{ij}}(1-p)^{1-x_{ij}} \end{aligned}$$

so each possible edge is included independently with probability p as specified by the Erdös-Rényi Model.

In order to model the emergence of more complex structures in a network, sociologists have proposed various sufficient statistics of ERGMs that permit more general conditional independence assumptions than the Erdös-Rényi Model (which has the most restrictive independence assumption that any two edges are conditionally independent given the rest of the graph.) For instance, Frank and Strauss [FS86] consider "Markov" graphs, where two possible edges in a graph may be conditionally dependent given the rest of the graph if they share a common endpoint. The intuition behind this independence assumption is that the probabilities of any two different relationships formed in a network are only related through a shared individual who formed these relationships. By permitting such dependencies, it is possible to model node level effects on edge formation. Frank and Strauss showed that all Markov graphs can be described by ERGMs of the following form:

Example 2.2 (Markov graphs [FS86])**.** Any undirected *Markov graph* has probability distribution:

$$\Pr(x|\theta,\tau) = \exp\left\{ \sum_{k=1}^{n-1} \theta_k S_k(x) + \tau T(x) - \psi(\theta,\tau) \right\}$$

where the sufficient statistics are

$$\begin{aligned} &\text{number of edges:} && S_1(x) = \sum_{1\le i<j\le n} x_{ij} = |E| \\ &\text{number of } k\text{-stars } (k \ge 2): && S_k(x) = \sum_{i=1}^{n-1} \binom{i}{k} D_i(x) \\ &\text{number of triangles:} && T(x) = \sum_{1\le h<i<j<\le n} x_{hi}x_{ij}x_{hj} \end{aligned}$$

and the parameters are $\{\theta_k\}_{k=1}^n$ and τ.[1]

In practice, neither the $G(n,p)$ model nor the full Markov graph model are frequently used for inference over real-world data. As explained above, the $G(n,p)$ model fails to capture complex dependencies between edges in a network. More general Markov models suffer from poor statistical properties making them generally unsuitable for inference over real world networks.

First, graphs instances of the general Markov graph model are susceptible to *model degeneracy*, where we refer to a probability distribution as degenerate if its mass is concentrated on a small subset of the space of possible graphs. For instance, consider the Markov graph with sufficient statistics S_1 and T (so $\theta_k = 0$ for $k \geq 2$). If we take $\theta_1 > 0$ and let τ be fairly large and positive, then this model puts almost all of its mass on the complete graph or nearly complete graphs, since the term in the exponent is extremely large for such graphs, as there are $\binom{n}{3}$ possible triangles. In fact, for $\tau > 0$, this model asymptotically (as $n \to \infty$) results in only three possible distributions: (1) all probability mass on the complete graph, (2) the $G(n,p)$ graph model or (3) a mixture distribution with some probability of the complete graph and some of $G(n,p)$ graphs [Jon99]. Thus, because we do not expect most interesting real world social or biological networks to be complete or $G(n,p)$ graphs, this model in its current form is not conducive to modeling real world networks.

Second, Markov graph models often suffer from *inferential degeneracy*, or the existence of many parameters that could maximize the likelihood $\Pr(x|\theta,\tau)$. As a simple example of such an issue, consider the case where the maximum degree of the graph is bounded by some $k < n-1$ so that there are no $(k+1)$-stars, but the model contains $S_{k+1}(x)$ as sufficient statistics and a corresponding parameter θ_{k+1}. Then the parameter θ_{k+1} could take on any value without changing the likelihood, so common techniques for maximum likelihood estimation may fail to converge. [Han03].

Lastly, the high sensitivity of the likelihood of general Markov graph models to addition or removal of edges makes common inference techniques challenging. As explained earlier, due to the intractability of computing the normalizing constant $\psi(\theta)$, sampling based inference methods are generally employed to perform inference over ERGMs. Roughly speaking, such methods proceed by sampling edges of a network in turn, holding the other edges constant, as repeating this procedure defines a Markov chain that converges asymptotically to the true distribution. At each sampling step, there is some probability of adding an edge to the graph which should be drawn according to the desired distribution. Now, because edges exhibit conditional dependencies, the addition of one edge may increase the probability of another edge being added to the graph. The difficulty with the specification of ERGMs given in *Example* 2.2 is that the likelihood can be highly sensitive to the addition or removal of an edge. For instance, including high-order k-stars all with positive θ, then for every additional edge added to a high degree node, the change to the likelihood grows exponentially since a d-degree node has $\binom{d}{k}$, k-stars. Thus, applying sampling based procedures to these Markov graph models tends to lead to "avalanche"

[1] Note that setting $\theta_2 = ... = \theta_k = \tau = 0$ in the Markov model, we recover the $G(n,p)$ model, which is an instance of a Markov graph $G(n,p)$ since any two edges are conditionally independent in the $G(n,p)$ model.

effects – as we add edges the conditional probability of other edges explosively increases, leading to convergence to the complete graph [SPRH06]. In fact, this is related to the issue of model degeneracy, since most of the probability mass of the distribution is on the complete graph. Thus, alternative models aim to use sufficient statistics that have smaller impact on the sufficient statistics and thus the likelihood of the model. This notion of robustness of the sufficient statistics to addition or removal of edges is closely related to differential privacy, suggesting that the subsequent model may be amenable to use under differential privacy constraints.

2.2 Alternating Sufficient Statistics for ERGMs

In response to the problems of degeneracy with Markov graphs, more robust sufficient "alternating" statistics are generally used in ERGMs. In this section, we present the alternating sufficient statistics commonly used for ERGMs. We will first provide definitions of the statistics and then explain the motivation and intuition behind them.

2.2.1 Definitions

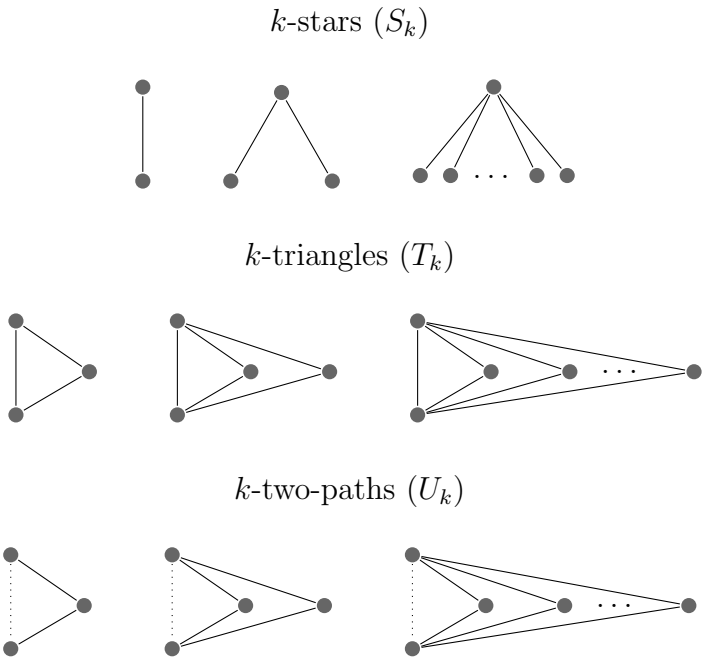

Figure 2.1: Subgraphs used in sufficient statistics of ERGMs.

Definition 2.2 (Alternating k-star statistic [SPRH06])**.** The *alternating k-star* statistic on graph x with weighting parameter $\lambda \geq 1$ is defined as

$$\begin{aligned} u_\lambda^{(s)}(x) &= S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \cdots + (-1)^{n-2}\frac{S_{n-1}}{\lambda^{n-3}} \\ &= \sum_{k=2}^{n-1} \frac{S_k}{\lambda^{k-2}} \end{aligned}$$

Now, we introduce the notion of "shared partners" of two nodes – the number of common neighbors that two nodes share. This will give us a clean way of representing k-triangles and k-two-paths:

Definition 2.3 (Shared partners)**.** We denote the *shared partner count* of nodes i and j by,

$$P_{ij}(x) = \sum_{\ell \in V} x_{i\ell} x_{j\ell} \tag{2.1}$$

Note, then that if two nodes i and j have a shared partner count of 1 and there is an edge between i and j, then the edge participates in one triangle.

We define *k-triangles* analogously to k-stars, so that a k-triangle consists of k triangles that all share an edge. We can count k-triangles using the number of common neighbors to any two nodes:

$$T_k(x) = \sum_{1 \le i < j \le n} x_{ij} \binom{P_{ij}}{k} \quad \text{for } (k \ge 2), \quad \text{and } T_1 = \frac{1}{3} \sum_{1 \le i < j \le n} x_{ij} P_{ij} \tag{2.2}$$

where T_1 has an extra factor of $\frac{1}{3}$ in front because of the symmetry of a 1-triangle for all three nodes included in the triangle.

Definition 2.4 (Alternating k-triangle statistic [SPRH06])**.** The *alternating k-triangle* statistic on graph x with weighting parameter $\gamma \ge 1$ is defined as

$$\begin{aligned} u_\gamma^{(t)}(x) &= 3T_1 - \frac{T_2}{\gamma} + \frac{T_3}{\gamma^2} - \cdots + (-1)^{n-3} \frac{T_{n-2}}{\gamma^{n-3}} \\ &= 3T_1 + \sum_{k=2}^{n-2} \left(\frac{-1}{\gamma} \right)^{k-1} T_k \end{aligned}$$

We define an *independent k-two-path* as a pair of nodes (possibly connected or unconnected) with k paths of length 2 connecting them. We can think of a k-two-path as a precondition for a k-triangle, since every k-triangle must contain an independent k-two-path. We can count the number independent k-two-paths in terms of shared partners as

$$U_k(x) = \sum_{1 \le i < j \le n} \binom{P_{ij}}{k} \text{ for } k \ne 2 \quad \text{and } U_2(x) = \frac{1}{2} \sum_{1 \le i < j \le n} \binom{P_{ij}}{2} \tag{2.3}$$

where U_2 is preceded by a factor of $\frac{1}{2}$, because a k-two-path with $k = 2$ is a 4-cycle and hence is symmetric with respect to the two pairs of non-adjacent nodes making up the cycle.

Definition 2.5 (Alternating k-two-path statistic [SPRH06])**.** The *alternating k-two-path* statistic on graph x with weighting parameter $\gamma \ge 1$ is defined as

$$\begin{aligned} u_\gamma^{(p)}(x) &= U_1 - \frac{2U_2}{\gamma} + \frac{U_3}{\gamma^2} - \cdots + (-1)^{n-3} \frac{U_{n-2}}{\gamma^{n-3}} \\ &= U_1 - \frac{2U_2}{\gamma} + \sum_{k=3}^{n-2} \left(\frac{-1}{\gamma} \right)^{k-1} U_k \end{aligned}$$

Now, having defined the "alternating" sufficient statistics, the proposed model has the form

$$\Pr(x|\theta) = \exp\left\{\theta_1 E(x) + \theta_2 u_\lambda^{(s)}(x) + \theta_3 u_\gamma^{(t)}(x) + \theta_4 u_\gamma^{(p)}(x) - \psi(\theta)\right\} \tag{2.4}$$

where $E(x)$ is the number of edges in graph x, the alternating k-two-path and k-triangle statistics generally use the same weighting parameter γ. In practice, a subset of the sufficient statistics can be used in the model, depending on what properties of a graph are pertinent to model for a given network.

2.2.2 Discussion

The overarching motivation behind introducing "alternating" sufficient statistics of the ERGMs, is that these statistics will be more robust to addition or removal of many edges adjacent to an individual node, alleviating issues of model and inferential degeneracy.

For instance, consider adding an edge to a high degree node with degree k. This addition leads to the addition of one $(k+1)$-star, $\binom{k}{k-1}$ k-stars, $\binom{k}{k-2}$ $(k-1)$-stars and so on, each using the new edge. Therefore, the total number of additional stars in the graph resulting from adding this edge is $\sum_{i=0}^{k}\binom{k}{i} = 2^k$. As discussed in Section 2.1, for Markov graphs including all stars as sufficient statistics, this could lead to a large change in the probability of the graph given the model (for arbitrary θ_k) causing a degenerate model that has almost all of its probability on near-complete graphs. However, by alternating the signs of k-stars, the additional $(k-1)$-stars and k-stars balance each-other out. In particular, in the case of k-stars, we can think of the alternating k-star statistic as imposing constraints on the θ_k in Example 2.2, namely that they must be alternating in sign and geometrically decreasing. In doing so, we enforce the property that adding edges to low degree nodes makes a significant difference in the likelihood of the graph, while adding edges to already high degree nodes makes less of a difference.

The same general reasoning applies to the use of alternating statistics for k-triangles and k-two-paths – alternation leads high degree nodes to become relatively less important in the likelihood of the graph, preventing the complete graph from having almost all of the probability mass of the distribution.

This interpretation of alternating statics as down-weighting high degree nodes can be understood by looking at an alternative representation of the statistics in terms of the degree distribution and the number of shared partners for nodes.

Alternating k-star

Note that using the relationship between k-stars and degrees given in Example 2.2 along with the binomial theorem we can rewrite the *alternating k-star* statistic as:

$$
\begin{aligned}
u_{\lambda}^{(s)}(x) &= \sum_{i=1}^{n-1} D_i(x) \sum_{k=2}^{n-1} \left(\frac{-1}{\lambda}\right)^{k-2} \binom{i}{k} \\
&= \lambda^2 \sum_{i=0}^{n-1} \left(\frac{\lambda-1}{\lambda}\right)^{i} D_i + 2\lambda|E| - n\lambda^2
\end{aligned} \tag{2.5}
$$

Thus, an ERGM using the alternating k-star statistic consists of a term representing the number of edges (akin to a $G(n,p)$ model) as well as a linear combination of the degrees where lower degree nodes are up-weighted exponentially compared to higher degree nodes, reflecting the tendency towards a power law degree distribution. Sociologically, the coefficient of the k-star statistic can be interpreted as the propensity for high degree nodes in the network. If the coefficient of the statistic is positive, then networks with a few high degree "hubs" are observed, while if it is negative, high degree nodes are discouraged and the network consists of low-degree nodes [SPRH06].

Alternating k-triangle

Similarly, for the alternating k-triangle statistic, we can gain insight by rewriting in terms of the number of shared partners for pairs of nodes. By using this representation of k-triangles from Equation (2.2) along with the binomial theorem, we can rewrite the *alternating k-triangle* statistic as:

$$
\begin{aligned}
u_{\gamma}^{(t)}(x) &= \sum_{1\le i<j\le n} x_{ij} \sum_{k=1}^{n-2} \left(\frac{-1}{\gamma}\right)^{k-1} \binom{P_{ij}}{k} \\
&= \gamma \sum_{1\le i<j\le n} x_{ij} \left(1 - \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}\right) \\
&= \gamma|E| - \gamma \sum_{1\le i<j\le n} x_{ij} \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}
\end{aligned} \tag{2.6}
$$

Thus, the first term of the triangle statistic is just the number of edges. But note that in the second term if $P_{ij} = 0$ but $x_{ij} = 1$ so that we have an edge that does not participate in a triangle, then this term cancels with the added edge, while the second term geometrically decreases as we add additional shared partners for an edge, so the statistic does not change with the addition of an edge, but with the addition of triangles, although the change is smaller for higher-order k-triangles. Including this term has a sociological interpretation of taking into account the tendency for "triangle closure" as it increases with further closure of triangles in the graph. Thus, the corresponding coefficient for this term in the likelihood of the model can be interpreted as the importance of triad closure in the generation of the graph [GKM09]. In contrast to directly including the number of triangles in the graph, the alternating k-triangles statistic is much more stable, preventing the model degeneracies discussed earlier.

Alternating k-two-path

Using the representation of k-two-paths in terms of shared partners from Equation (2.3) and the binomial theorem, we can rewrite the *alternating k-two-path* statistic as:

$$\begin{aligned} u_\gamma^{(p)}(x) &= \sum_{1\le i<j\le n} \sum_{k=1}^{n-2} \left(\frac{-1}{\gamma}\right)^{k-1} \binom{P_{ij}}{k} \\ &= \gamma \sum_{1\le i<j\le n} \left(1 - \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}\right) \\ &= \gamma \binom{n}{2} - \gamma \sum_{1\le i<j\le n} \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}} \end{aligned} \tag{2.7}$$

Thus, the alternating k-two-path has an interpretation similar to that of the alternating k-triangle. As shared partners are added for any two nodes, the second term of the statistic increases, but the increase falls exponentially with additional partners. This term is generally only included in conjunction with the k-triangle statistic to try to separate out the effects of two-paths forming between unconnected nodes and mutual connections forming between already connected nodes.

In practice, the coefficients of these various sufficient statistics are usually inferred either over separate networks (for instance, networks of neurons at various life-stages) or over subpopulations of one network and by comparing the inferred coefficients, researchers can understand the relative importance of the different underlying processes discussed (such as triad closure and tendency for "hub" nodes) to the different networks.

2.3 Inference on ERGMs

While it can be of interest to consider the properties of ERGMs based on their parameters, generally ERGMs are of interest in modeling realized network datasets. In such an inferential approach, a data analyst wishes to finds parameters θ of a given class of ERGM that describe "well" the realized data and often to also decide what model best fits the data (which set of sufficient statistics best describe the data.)

2.3.1 Sampling Graphs

Write out Metropolis Hastings, (Gibbs?) samplers and motivate them and then describe and motivate Bayesian inference

2.3.2 Estimation of Parameters

Algorithm 1 Metropolis-Hastings Sampler for ERGMs

Input: parameter vector θ, initial graph $x^{(0)}$, number of iterations T
Output: sequence of graphs $x^{(1)}, ...x^{(T-1)}, x^{(T)}$ such that $x^{(T)} \sim p(X|\theta)$ as $T \to \infty$

For $t = 1, ..., T$:

1. Select nodes i and j at random
2. Propose graph x^* which is the same as $x^{(t-1)}$ except that we "toggle" the edge between i and j so $x^*_{ij} = 1 - x^{(t-1)}_{ij}$
3. Accept the proposed move with probability $\min\left\{1, \frac{p(x^*|\theta)}{p(x^{(t-1)}|\theta)}\right\}$. If the move is accepted set $x^{(t)} = x^*$. Otherwise, set $x^{(t)} = x^{(t-1)}$

Chapter 3: Differential Privacy over Graphs

3.1 Defining Differential Privacy over Graphs

In maintaining the privacy of individuals' data while analyzing a network, we employ the framework of differential privacy. We let $\mathcal{D}$ denote the space of all possible datasets. Then:

Definition 3.1. We say that two datasets $x, x' \in \mathcal{D}$ are *neighbors* if they differ in the record of one individual (we will discuss what it means for two graphs to be "neighbors" in Definitions 3.5 and 3.6). We denote that datasets x and x' are neighbors by $x \sim x'$.

Definition 3.2 (ϵ-differential privacy [DMNS06])**.** Let $\mathcal{A}$ be an algorithm over datasets in $\mathcal{D}$. Then $\mathcal{A}$ is *ϵ-differentially private* if for all $S \subseteq \text{Range}(\mathcal{A})$ and for every pair of neighboring datasets $x, x' \in \mathcal{D}$,

$$\Pr[\mathcal{A}(x) \in S] \leq e^{\epsilon} \Pr[\mathcal{A}(x') \in S]$$

Intuitively, differential privacy promises that the participation of any one individual in the dataset does not significantly change the outcome of a differentially private algorithm run on the dataset.

3.1.1 Properties

One of the desirable properties of differential privacy is its immunity to *post-processing* – an analyst cannot process the output of an ϵ-DP algorithm and make it less differentially private, without additional information about the private dataset. Immunity to post-processing will allow us to compute sufficient statistics of a statistical model in an ϵ-DP manner and then consider inference using these statistics to be a post-processing step that does not further degrade privacy.

Property 1 (Post-processing [DMNS06])**.** If $\mathcal{A}$ is an ϵ-differentially private algorithm, then for any algorithm f, the composition of the functions $f \circ \mathcal{A}$ is also ϵ-differentially private.

A second useful property of differential privacy is that multiple differentially private algorithms compose nicely. This allows us to use many DP algorithms as building blocks

to more complicated algorithms, and in particular to split a privacy budget across multiple private computations on the data, permitting us, for instance, to compute various sufficient statistics of a probabilistic model separately. Specifically, the privacy loss incurred by running multiple DP algorithms on a dataset grows linearly:

Property 2 (Basic composition [DMNS06])**.** Let $\mathcal{A}_i$ be an ϵ_i-differentially private algorithm for $i \in [k]$. Then, the algorithm releasing the result of running all k algorithms on the dataset $\mathcal{A}_{[k]}(x) = (\mathcal{A}_1(x), ..., A_k(x))$ is $(\sum_{i=1}^{k} \epsilon_i)$-DP.

3.1.2 Mechanisms

A query on a dataset is a function mapping the dataset to a real number, $f : \mathcal{D} \to \mathbb{R}$.

Definition 3.3 (Local sensitivity)**.** The *local sensitivity* of a query f on a dataset x is

$$LS_f(x) = \max_{x' \sim x} |f(D) - f(D')|$$

Definition 3.4 (Global sensitivity)**.** The *global sensitivity* of a query f is

$$GS_f = \max_{x \in \mathcal{D}} LS_f(x)$$

Theorem 3.1 (Laplace mechanism [DMNS06])**.** *Let f be a query on dataset x with global sensitivity GS_f and let* Lap *denote the zero-mean Laplace distribution*[1]*. Then, the Laplace mechanism $\mathcal{A}_L$ that outputs*

$$\mathcal{A}_L(x, f, \epsilon) = f(x) + Lap\left(\frac{GS_f}{\epsilon}\right)$$

is ϵ-differentially private.

Theorem 3.2 (Exponential mechanism [MT07])**.** *Let $u(x, r)$ be a utility function that maps database-outcome pair (x, r) to a real-valued score. Then, the Exponential mechanism $\mathcal{A}_E$ that produces a random outcome with probability:*

$$\Pr(\mathcal{A}_E(x, u, \epsilon) = r) \propto \exp\left(\frac{\epsilon u(x, r)}{2GS_u}\right)$$

is ϵ-differentially private. The global sensitivity of the utility function is the maximum change in the utility function on any fixed outcome over a pair of neighboring datasets:

$$GS_u = \max_{r, x' \sim x} |u(x, r) - u(x', r)|$$

Theorem 3.3 (Randomized response [War65],[KKS17])**.** *Let $\mathcal{D} = \{0, 1\}^n$ so $x \in \mathcal{D}$ consists of binary data. Then, randomized response flips each bit of x with probability $p \in (0, \frac{1}{2})$ and releases the resulting noisy bits.*[2] *This process provides ϵ-differential privacy with $\epsilon = -\log \frac{p}{1-p}$.*

[1]The Laplace distribution centered at 0 with scale parameter b has probability density function $p(x|b) = \frac{1}{2b} e^{-|x|/b}$ and the variance of the distribution is $\sigma^2 = 2b^2$.

[2]Note that a more general version of randomized response could in fact be used while collecting data (and indeed was initially proposed as such a survey method in the 60s) since a researcher could ask participants in a survey a question, and could flip a biased coin to decide whether to keep the true answer or not and if not, then flip a second coin to record.

3.1.3 Neighboring Graphs

We now turn to the question of how to define two graphs as "neighbors." There are two reasonable and widely used definitions, which provide privacy at very different granularities and thus may be appropriate in different circumstances, as we detail below:

Definition 3.5 (Edge-level adjacency)**.** We define two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ to be *edge-adjacent* if they have the same vertex set ($V_1 = V_2$) and they differ in only one edge ($|E_1 \triangle E_2| = 1$).

Edge-level differential privacy treats two graphs as adjacent if they differ in an individual edge, thereby protecting privacy of specific relationships between nodes. Thus, edge-level privacy could protect, for instance, an individual's Facebook friendship on Facebook with a controversial political leader. However, privacy at the edge-level could not guarantee protection from an adversary's discerning whether an individual has all Republican or Democratic friends on Facebook. Such concerns motivate a stronger definition of neighboring graphs:

Definition 3.6 (Node-level adjacency)**.** We define two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ to be *node-adjacent* if $G_1 - v_i = G_2 - v_i$ for some vertex v_i, where we use $G - v_i$ to mean that we remove all edges adjacent to node v_i. Thus, in the extreme case, one graphs may be obtained from the other by removing all edges incident to a single node.

Finally, we define the distance between two graphs as follows:

Definition 3.7 (Distance)**.** The *distance* between two graphs G_1 and G_2, denoted $d(G_1, G_2)$ is the minimal length of the sequence of graphs beginning with G_1 and ending with G_2 such that every two consecutive graphs on the path are neighbors (with respect to either node-level or edge-level adjacency). To go from G_1 to G_2 we will need to either add or remove every edge in the symmetric difference between the edge sets of G_1 and G_2 (all edges that are not the same in the two graphs.) Thus, under edge-level adjacency, we step from G_1 to G_2 by each edge, so $d(G_1, G_2) = |E_1 \triangle E_2|$. Under node-level adjacency, we can change all edges adjacent to a node at once, so in order to go from one graph to another, we can step through each node that touches edges that differ between the graphs. Therefore, the distance between two graphs is given by the size of the vertex cover of the symmetric difference graph (the graph including the edges in $E_1 \triangle E_2$): $d(G_1, G_2) = |VC(G_1 \triangle G_2)|$. Note that, finding the vertex cover of an arbitrary graph is NP hard, so computing the distance between two graphs under node-adjacency is an NP-hard problem[BBDS13].

3.2 Restricted Sensitivity

Now, it is clear that node-level privacy constitutes a much stronger guarantee than edge-level privacy. However, it also much more difficult to perform accurate analysis under node-level privacy. For instance, the global sensitivity of the degree distribution of a graph to the addition under the edge-level definition is only 2, because the degree of two

nodes will change by 1 due to the addition or removal of an edge, but under the node-level definition we could consider removing a node of degree $n-1$ which would affect n entries, so naive application of the Laplace mechanism would completely destroy the counts of the degree distribution.

This example also illustrates, that over graphs where we use node-level adjacency, the global sensitivity and even local sensitivity may be very high. However, if we hypothesized that our class of graphs has limited degree $k << n$, then the sensitivity might be much lower over these limited-degree graphs. This assumption would be reasonable for many real world social networks where there may be many individuals in the network, but people have a bounded number of friends. For instance, Facebook has billions of users on it, but an explicit restriction that users may not have more than 5,000 friends.

Motivated by the desire to perform accurate differential private analysis on potentially high-sensitivity graph statistics Blocki et al. [BBDS13] propose the notion of *restricted sensitivity*, where we use a hypothesis (such as the limited degree hypothesis above) to restrict the sensitivity of a query over the graph. The hypothesis $\mathcal{H}$ is a subset of all possible graphs and the hypothesis is true if the true graph is in this subset.

At a high level, if we project graphs into $\mathcal{H}$ and then calibrate added noise to the "restricted sensitivity" of graphs within this subset, then we can guarantee differential privacy. Further, if the hypothesis was true for our graph, then the projection did not alter the graph at all, so we preserve accuracy (up to the distortion of the noise-adding procedure). Note, however, that for graphs for which the hypothesis is false, we may not be able to make any accuracy guarantees, as this depends on how the projection into $\mathcal{H}$ alters the graph.

Denoting the space of all graphs $\mathcal{G}$, we define restricted sensitivity as follows:

Definition 3.8 (Restricted sensitivity [BBDS13])**.** For a given notion of adjacency (either edge or node), we define the *restricted sensitivity* of query f over hypothesis $\mathcal{H} \in \mathcal{G}$ as

$$RS_f(\mathcal{H}) = \max_{G_1, G_2 \in \mathcal{H}} \frac{|f(G_1) - f(G_2)|}{d(G_1, G_2)}$$

Note that the restricted sensitivity is defined over all graphs in $\mathcal{H}$, not just neighbors, since a projection of a graph and its neighbor to the hypothesis subset may not result in the two graphs being adjacent. However, if we use the hypothesis $\mathcal{H}_k$, that our class of graphs has a limited degree, formally $\mathcal{H}_k = \{G = (V, E) \in \mathcal{G} : deg(v) \leq k, \forall v \in V\}$, then, we can bound the restricted sensitivity by finding the global sensitivity restricted to adjacent graphs in $\mathcal{H}_k$:

Lemma 3.1 (Restricted sensitivity for $\mathcal{H}_k$)**.** *For the limited degree hypothesis we can bound restricted sensitivity by:*

$$RS_f(\mathcal{H}_k) \leq \max_{G_1, G_2 \in \mathcal{H}_k : G_1 \sim G_2} |f(G_1 - f(G_2)|$$

Proof. Note that if the distance between two k-degree graphs (in either the node or edge adjacency formulation) is m, then there must be a sequence of adjacent graphs all in $\mathcal{H}_k$

of length m starting with the first graph and ending with the second graph. This holds by always removing edges in the sequence (either on an individual level for edge-adjacency or at the node level for node-adjacency) before adding edges.[3]

Now, consider two graphs $G_0, G_m \in \mathcal{H}_k$ with $d(G_0, G_m) = m$ and let $G_i \in \mathcal{H}_k, i \in [m]$ be a sequence of adjacent graphs ($G_i \sim G_{i+1}$) beginning with G_0 and ending with G_m. Then, we have (applying the triangle inequality) that:

$$\begin{aligned} RS_f(\mathcal{H}) = \max_{G_0, G_m \in \mathcal{H}_k} \frac{|f(G_0) - f(G_m)|}{d(G_0, G_m)} &= \max_{G_0, G_m \in \mathcal{H}_k} \frac{|\sum_{i=0}^{m-1} f(G_i) - f(G_{i+1})|}{m} \\ &\leq \max_{G_0, G_m \in \mathcal{H}_k} \frac{\sum_{i=0}^{m-1} |f(G_i) - f(G_{i+1})|}{m} \\ &\leq \max_{G, G' \in \mathcal{H}_k : G \sim G'} |f(G) - f(G')| \end{aligned}$$

□

In many cases, the difference between neighboring graphs in $\mathcal{H}_k$ given in Lemma 3.1 will be much easier to bound than the general formulation of restricted sensitivity given in definition 3.8 over $\mathcal{H}_k$.

Now, to achieve differential privacy while calibrating noise to restricted sensitivity rather than global sensitivity, we will use a smooth projection $\mu : \mathcal{G} \to \mathcal{H}$ to project a graph G into $\mathcal{H}$ (so $\mu(G) = G$ if $G \in \mathcal{H}$) and then apply a differentially private algorithm adding noise proportional to the restricted sensitivity. We want this projection to be smooth so that two adjacent graphs are still close to each other in distance after being projected to $\mathcal{H}$. We define a smooth projection as:

Definition 3.9 (c-smooth projection [BBDS13])**.** A projection $\mu : \mathcal{G} \to \mathcal{H}$ is a c-smooth projection if for any pair of neighboring graphs $G \sim G'$, $d(\mu(G)), \mu(G') \leq c$.

Theorem 3.4. *Let $\mu : \mathcal{G} \to \mathcal{H}$ be a c-smooth projection. For a query f, define $f_{\mathcal{H}} = f \circ \mu$ to be the query applied to the projection. Then $GS_{f_{\mathcal{H}}} \leq c \cdot RS_f(\mathcal{H})$.*

In particular, this means that if we can find a c-smooth projection for a given hypothesis $\mathcal{H}$, then we can use ϵ-differentially private mechanisms that add noise proportional to $c \cdot RS_f(\mathcal{H})$.

For $\mathcal{H}_k$, Blocki et al. give an efficient 3-smooth projection to $\mathcal{H}_k$ in the edge-adjacency model and a projection for the node-adjacency model along with a 4-smooth estimator of the distance between graphs, which allows for (ϵ, δ)-privacy in the node-adjacency model. Thus, using $\mathcal{H}_k$ as our hypothesis, we will be able to add noise proportional to the restricted sensitivity over $\mathcal{H}_k$ of the statistics that we wish to compute. Therefore, our primary focus in proving privacy will be bounding the restricted sensitivity of the queries of interest over $\mathcal{H}_k$.

[3]Note that for an arbitrary hypothesis $\mathcal{H}$, however, the distance between two graphs in $\mathcal{H}$ may be realized only by a sequence that include graphs not in $\mathcal{H}$, which is why we give the general definition of restricted sensitivity as in Definition 3.8.

Chapter 4: Related Work on Differential Privacy and ERGMs

This is my intro... I will talk about "Birds of a Feather" [GKM09]

Chapter 5: Differentially Private Sufficient Statistics of ERGMs

In this section, we propose methods for releasing differentially private alternating sufficient statistics of the ERGM defined in Section 2.2. In particular, we take advantage of methods using restricted sensitivity (see Section 3.2) where we have hypothesis $\mathcal{H}_k$ that the network has degree limited to k. This seems like a reasonable assumption for two reasons. First, experimental results on ERGMs with the alternating statistics have demonstrated that for reasonable parameter values, the distribution tends to put low probability mass on very high-degree graphs [SPRH06]. Thus, given that we assume that an observed network is roughly drawn from the probability distribution specified by an ERGM, we believe with high probability that the graph has relatively low degree. Second, many real-world social networks that are relevant to study have bounded degree, although this may vary based on the specific network dataset.

Below, we characterize the restricted sensitivity of the alternating sufficient statistics of an ERGM under edge level privacy and node level privacy respectively. For reference, the "weighting parameters" of the alternating statistics γ and λ are generally set to be small constants between roughly 1 and 5 (most empirical work seems to find that values between 1 and 2 in fact suffice).

5.1 Edge Level Privacy

For the alternating k-star statistic under edge-level privacy, restricted sensitivity does not give any advantage over using global sensitivity, as the global sensitivity of this statistic is quite low:

Claim 5.1.1 (Global sensitivity of alternating k-star under edge-level privacy)**.** *The global sensitivity of the alternating k-star statistic is less than 2λ.*

Proof. We use the alternative formulation of the statistic given in Equation (2.5):

$$u_\lambda^{(s)}(x) = \lambda^2 \sum_{i=0}^{n-1} \left(\frac{\lambda-1}{\lambda}\right)^i D_i + 2\lambda|E| - n\lambda^2$$

Then, consider adjacent graphs x, x' differing in one edge where x has the additional edge. Then, the first term of the alternating k-statistic is larger for x' than for x and by at

most 2λ and at least 0, while the second term is larger for x than for x' by 2λ. Hence, the difference between the alternating k-star statistic computed on x and x' is at most $|2\lambda - 0| = 2\lambda$ and appealing to the bound on restricted sensitivity from Lemma 3.1, we have that $RS_{u_\gamma^{(t)}}(\mathcal{H}_k) \leq 2\lambda$. □

Note that we could also compute the k-star statistic by computing the degree distribution of the graph in a differentially private manner, which can be done with high accuracy using the Laplace mechanism and clever post-processing (see [HLMJ09]), and then using the degrees for the alternating-k-star statistic. However, adding noise proportional to global sensitivity of 2λ should give good accuracy, as the alternating k-star statistic is roughly on the order of $2\lambda|E|$.

Claim 5.1.2 (Restricted sensitivity of alternating k-triangle under edge-level privacy)**.** *The restricted sensitivity of the alternating k-triangle statistic under $\mathcal{H}_k$ is less than $2(k-1)+\gamma$.*

Proof. Consider two adjacent graphs $x, x' \in \mathcal{H}_k$ differing in exactly one edge, so that $x_{ij} = 1$ and $x'_{ij} = 0$. Now, note that for nodes i and j, the number of shared partners is the same in x and x' since all edges are the same except for the edge between i and j. Then, let $P_{ij} = P'_{ij} = m \leq k-1$ by the limited degree hypothesis. Note that there are $2m$ edges for which $P'_e = P_e - 1$, since there are two other edges in each triangle. Then, recalling the definition of the alternating k-triangle statistic in terms of the shared partners of i and j given in Equation (2.6):

$$u_\gamma^{(t)}(x) = \gamma|E| - \gamma \sum_{1 \leq i < j \leq n} x_{ij} \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}$$

we have that

$$\begin{aligned}
|u_\gamma^{(t)}(x) - u_\gamma^{(t)}(x')| &= \left| \gamma - \gamma \left(\frac{\gamma-1}{\gamma}\right)^m + \gamma \sum_{e=1}^{2m} \left[\left(\frac{\gamma-1}{\gamma}\right)^{P_e - 1} - \left(\frac{\gamma-1}{\gamma}\right)^{P_e} \right] \right| \\
&= \left| \gamma - \gamma \left(\frac{\gamma-1}{\gamma}\right)^m + \sum_{e=1}^{2m} \left(\frac{\gamma-1}{\gamma}\right)^{P_e - 1} \right| \\
&\leq 2m + \gamma \\
&\leq 2(k-1) + \gamma
\end{aligned}$$

and again appealing to the bound on restricted sensitivity from Lemma 3.1, we have that $RS_{u_\gamma^{(t)}}(\mathcal{H}_k) \leq 2(k-1)+\gamma$. □

Note the usefulness of restricted sensitivity here, in contrast to global sensitivity. The global sensitivity of this statistic is $O(n)$, since there could conceivably be a graph with an $(n-1)$-triangle where removing the base of the triangle could lead to a very large change in the statistic, since it would lead to the removal of n triangles of which it was not the only base. However, if we restrict degrees, we can potentially add much less noise.

Claim 5.1.3 (Restricted sensitivity of alternating k-two-path under edge-level privacy)**.** *The restricted sensitivity of the alternating k-two-path statistic under $\mathcal{H}_k$ is less than $2(k-1)$.*

Proof. The proof will proceed in roughly the same way as for k-triangles. Define x and x' in the same way and recall the definition of the alternating k-two-path statistic in terms of shared partners as given in Equation (2.7):

$$u_\gamma^{(p)}(x) = \gamma\binom{n}{2} - \gamma \sum_{1\le i<j\le n} \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}$$

Then, the change between the statistic on x and x' is equal to

$$|u_\gamma^{(p)}(x) - u_\gamma^{(p)}(x')| = \sum_{e=1}^{2m} \left(\frac{\gamma-1}{\gamma}\right)^{P_e - 1} \le 2m \le 2(k-1)$$

□

5.2 Node Level Privacy

Claim 5.2.1 (Restricted sensitivity of alternating k-star under node-level privacy)**.** *The restricted sensitivity with hypothesis $\mathcal{H}_k$ of alternating k-star under node-level differential privacy is less than $3\lambda k$.*

Proof. We will again use the formulation of the alternating k-star statistic in terms of degree distribution from Equation (2.5). Now, consider two graphs $x, x' \in \mathcal{H}_k$ differing in one node i of degree $m \le k$, with all of its incident edges removed in x'. Then, the degree of node i is m in x and 0 in x', while the degrees of m other nodes are 1 lower in x' than in x, so:

$$\begin{aligned} |u_\lambda^{(s)}(x) - u_\lambda^{(s)}(x')| &= \left| 2\lambda m + \lambda^2 \left(\left(\frac{\lambda-1}{\lambda}\right)^m - 1 \right) + \sum_{j:x_{ij}=1} \lambda \left(\frac{\lambda-1}{\lambda}\right)^{d_j - 1} \right| \\ &\le \left| 3\lambda m + \lambda^2 \left(\left(\frac{\lambda-1}{\lambda}\right)^m - 1 \right) \right| \end{aligned}$$

and note that $0 \le \left(\frac{\lambda-1}{\lambda}\right)^m \le 1$ and that $|\lambda^2| \le 3\lambda m$ for reasonable choices of k and λ (since generally we choose $1 < \lambda < 5$, so in order to have the λ^2 term dominate the $3\lambda k$ term we would have to restrict k to 1, which would not be interesting or realistic, so the sensitivity is bounded by $3\lambda k$. □

Claim 5.2.2 (Restricted sensitivity of alternating k-triangle under node-level privacy)**.** *The restricted sensitivity with hypothesis $\mathcal{H}_k$ of the alternating k-triangle statistic under node-level differential privacy is less than $k^2 + (\gamma-1)k$.*

Proof. Consider two adjacent graphs $x, x' \in \mathcal{H}_k$ differing in one node i of degree m. Now, since each of the m edges incident to node i is removed this changes m edges $x_{ij} = 1$ to $x'_{ij} = 0$, so $E(x) - E(x') = m$ and for each of these m edges

$$x_{ij}\left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}} - x'_{ij}\left(\frac{\gamma-1}{\gamma}\right)^{P'_{ij}} = \left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}}$$

so the direct effect of removing the x_{ij} is that $u_\gamma^{(t)}(x') - u_\gamma^{(t)}(x') \leq m\gamma - 0$ (ignoring the effect on the shared partners of edges not adjacent to i.)

Now, we consider edges e such that the endpoints of e have i as a shared partner. Note that there are $\binom{m}{2} = m^2 - m$ such edges, because we can choose any 2 edges of i and the endpoints of these edges have i as a shared partner. Now, each of these edges still exists in x'_{ij} but has its number of shared partners decrease by 1. Then, we have

$$\begin{aligned}
|u_\gamma^{(t)}(x) - u_\gamma^{(t)}(x')| &= \left|\gamma m - \gamma \sum_{j:x_{ij}=1}\left(\frac{\gamma-1}{\gamma}\right)^{P_{ij}} + \sum_{e=1}^{m^2-m}\left(\frac{\gamma-1}{\gamma}\right)^{P_e-1}\right| \\
&\leq |\gamma m + (m^2 - m)| \\
&\leq k^2 + (\gamma - 1)k
\end{aligned}$$

□

Claim 5.2.3 (Restricted sensitivity of alternating k-two-path under node-level privacy)**.** *The restricted sensitivity with hypothesis $\mathcal{H}_k$ of the alternating k-two-path statistic under node-level differential privacy is less than k^2.*

Proof. As for k-triangles, consider two adjacent graphs $x, x' \in \mathcal{H}_k$ differing in node i of degree m. Then, the removal of these m edges impacts the shared partners of m^2 edges, the m incident to i and the $\binom{m}{2} = m^2 - m$ that have i as a shared partner and the decrease in shared partners for each of these edges can change the statistic by at most 1 so the overall change is at most $m^2 \leq k^2$. □

5.3 Using the Noisy Statistics

Now, by projecting our graph into $\mathcal{H}_k$ using the projections of Blocki et al. [BBDS13] and then applying the Laplace mechanism (3.1), we can release the sufficient statistics of the ERGM in a differentially private manner by calibrating the noise of the Laplace mechanism to the restricted sensitivity by 3.4. In theory, we could now release these sufficient statistics to analysts who wish to study the network, since the likelihood of the ERGM depends on the data only through the sufficient statistics. However, using these noisy statistics directly for standard inference techniques may lead to poor accuracy. It may be an interesting research question in itself how adding noise to the sufficient statistics of the model degrades inference, although the hope is to perform inference in such a way as to take into account the noise-adding procedure, which has worked well for MCMC

techniques akin to those used for inference on ERGMs (for instance, in [LM14], [FGWC16]). Because the convergence properties and accuracy of non-private inference methods for inference over ERGMs are primarily understood from an experimental standpoint, we propose experimental evaluation of methods for inference over private sufficient statistics.

Chapter 6: Inference

This is my intro... I will talk about "Birds of a Feather" [GKM09]

Chapter 7: Conclusion

Bibliography

[AGZF09] E. M. Airoldi, A. Goldenberg, A. Zheng, and S. Fienberg, "A survey of statistical network models", eng, *Machine Learning -Boston-*, vol. 2, no. 2, 2009, ISSN: 0885-6125.

[BBDS13] J. Blocki, A. Blum, A. Datta, and O. Sheffet, "Differentially private data analysis of social networks via restricted sensitivity", eng, in *Proceedings of the 4th conference on innovations in theoretical computer science*, ser. ITCS '13, ACM, Jan. 2013, pp. 87–96, ISBN: 9781450318594.

[DMNS06] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis", in *Proceedings of the 3rd Theroy of Cryptography Conference*, Springer, 2006, pp. 265–284.

[FGWC16] J. Foulds, J. Geumlek, M. Welling, and K. Chaudhuri, "On the theory and practice of privacy-preserving bayesian data analysis", Mar. 2016.

[FS86] O. Frank and D. Strauss, "Markov graphs", *Journal of the American Statistical Association*, vol. 81, Sep. 1986, ISSN: 0162-1459.

[GKM09] S. M. Goodreau, J. A. Kitts, and M. Morris, "Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks", *Demography*, vol. 46, no. 1, pp. 103–125, 2009.

[Han03] M. S. Handcock, "Assessing degeneracy in statistical models of social networks", Tech. Rep., Dec. 2003, Working Paper no. 39, Center for Statistics and the Social Sciences, Univeristy of Washington, Seattle.

[HLMJ09] M. Hay, C. Li, G. Miklau, and D. Jensen, "Accurate estimation of the degree distribution of private networks", eng, IEEE Publishing, Dec. 2009, pp. 169–178, ISBN: 978-1-4244-5242-2.

[HH06] D. R. Hunter and M. S. Handcock, "Inference in curved exponential family models for networks", eng, *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 565–583, Sep. 2006.

[Jon99] J. Jonasson, "The random triangle model", eng, *Journal of Applied Probability*, vol. 36, no. 3, pp. 852–867, Sep. 1999.

[KKS17] V. Karwa, P. N. Krivitsky, and A. B. Slavkovic, "Sharing social network data: Differentially private estimation of exponential family random graph models", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 3, pp. 481–500, Apr. 2017.

[KRSY14] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure", eng, *ACM Transactions on Database Systems (TODS)*, vol. 39, no. 3, pp. 1–33, Oct. 2014.

[KS16] V. Karwa and A. Slavkovic, "Inference using noisy degrees: Differentially private β-model and synthetic graphs", eng, *Annals of Statistics*, vol. 44, no. 1, Feb. 2016.

[LM14] W. Lu and G. Miklau, "Exponential random graph estimation under differential privacy", eng, in *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, Aug. 2014, pp. 921–930.

[MT07] F. Mcsherry and K. Talwar, "Mechanism design via differential privacy", eng, IEEE, Oct. 2007, pp. 94–103.

[MW12] D. Mir and R. Wright, "A differentially private estimator for the stochastic kronecker graph model", eng, in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, ser. EDBT-ICDT '12, ACM, Mar. 2012, pp. 167–176.

[RAS10] E. Roland, B. Alla, and B. Svetlana, "Bayesian statistical modelling of human protein interaction network incorporating protein disorder information", eng, *BMC Bioinformatics*, vol. 11, no. 1, Jan. 2010.

[SDC+16] M. R. Sinke, R. M. Dijkhuizen, A. Caimo, C. J. Stam, and W. M. Otte, "Bayesian exponential random graph modeling of whole-brain structural networks across lifespan", eng, *NeuroImage*, vol. 135, pp. 79–91, Jul. 2016.

[SPRH06] T. Snijders, P. Pattison, G. Robins, and M. Handcock, "New specifications for exponential random graph models", eng, *Sociological Methodology*, vol. 36, pp. 99–153, Jan. 2006.

[UHH13] S. Uddin, J. Hamra, and L. Hossain, "Exploring communication networks to understand organizational crisis using exponential random graph models", eng, *Computational and Mathematical Organization Theory*, vol. 19, no. 1, pp. 25–41, Mar. 2013.

[War65] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias", *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, Mar. 1965.

[WP96] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks", eng, *Psychometrika*, vol. 61, no. 3, pp. 401–425, Sep. 1996.

Appendix A: Smooth Projections to $\mathcal{H}_k$

The contents...