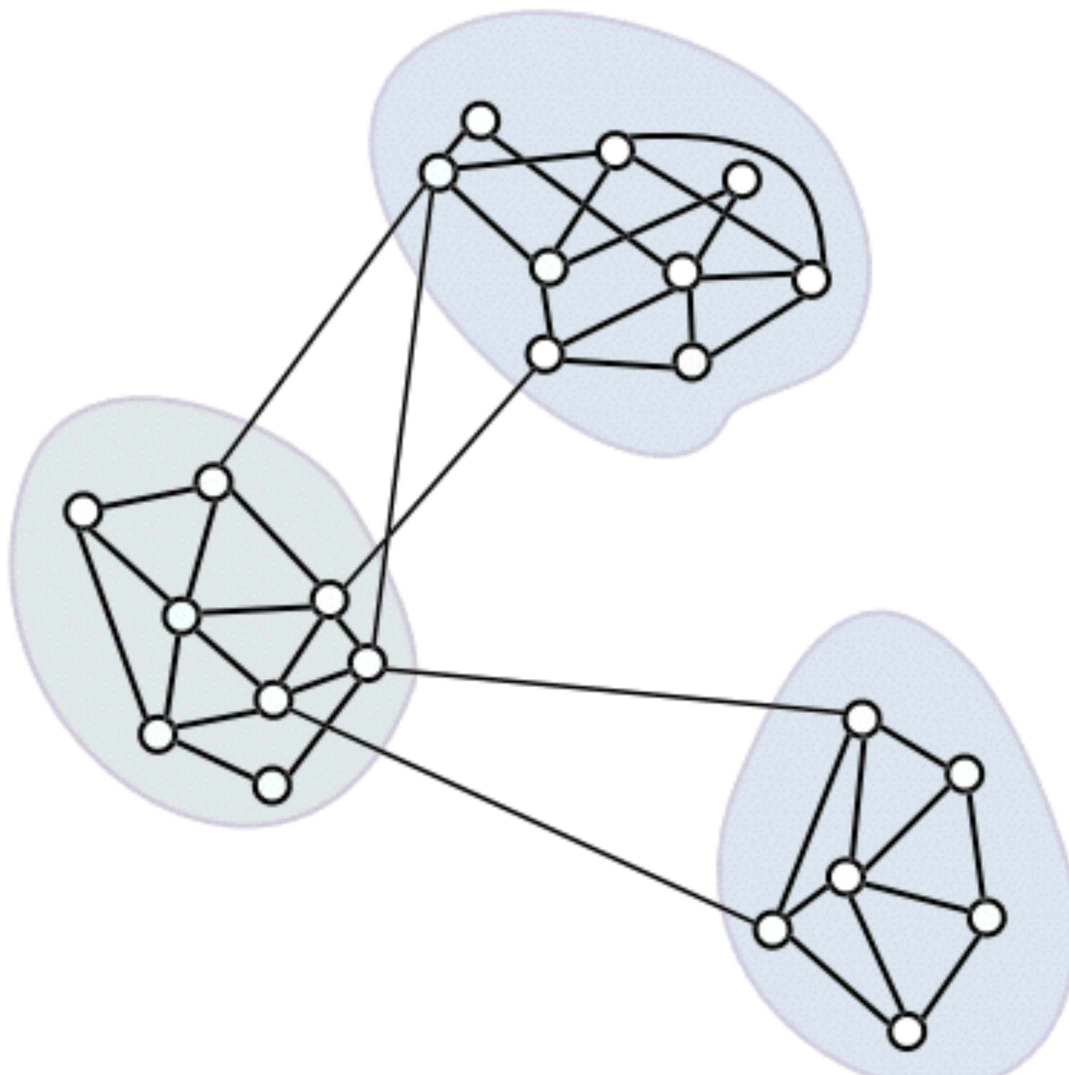
Graphon Estimation: Minimax Rates and Posterior Contraction

Chao Gao
Yale University

@Leiden, March 2015

Stochastic Block Model



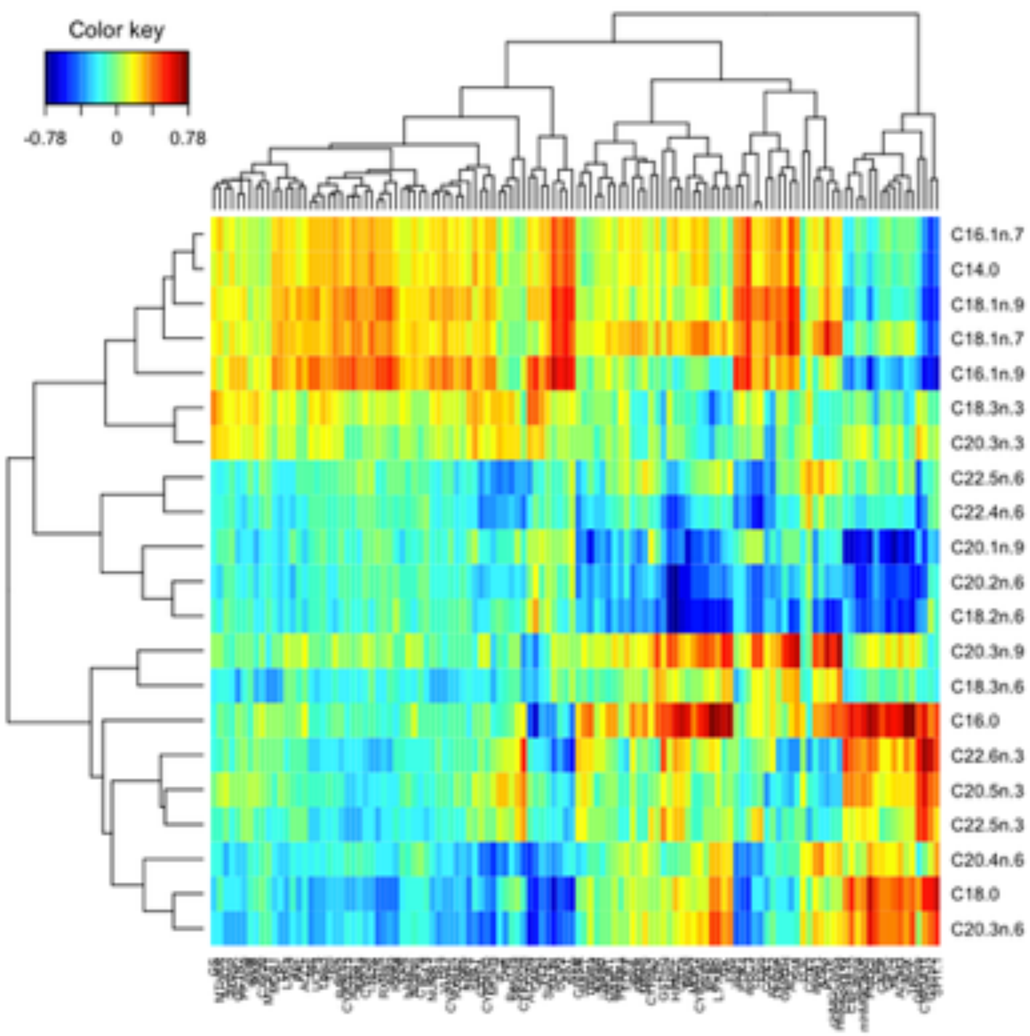
$$z : \{1, 2, ..., n\} \to \{1, 2, ..., k\}$$

$$A_{ij} \sim \text{Bernoulli}(\theta_{ij})$$

$$\theta_{ij} = Q_{z(i)z(j)}$$

Goal: recover θ_{ij}

Biclustering (Hartigan, 1972)



$$z_1 : \{1, 2, ..., n\} \to \{1, 2, ..., k\}$$

$$z_2 : \{1, 2, ..., m\} \to \{1, 2, ..., l\}$$

$$\mathbb{E}(A_{ij}) = \theta_{ij} = Q_{z_1(i) z_2(j)}$$

Goal: recover θ_{ij}

Nonparametric Regression

$$y_i = f(x_i) + \epsilon_i$$

$$x_i \in \mathcal{D}, \quad \epsilon_i \sim N(0,1)$$

Common assumption: f is smooth on $\mathcal{D}$.

Goal: recover f from both x and y

A More Challenging Problem

$$y_i = f(x_i) + \epsilon_i$$

$$x_i \in \mathcal{D}, \quad \epsilon_i \sim N(0,1)$$

Common assumption: f is smooth on $\mathcal{D}$.

Goal: recover f from only y

A nonparametric view of network models and Newman–Girvan and other modularities

Peter J. Bickel[a,1] and Aiyou Chen[b]

[a]University of California, Berkeley, CA 94720; and [b]Alcatel-Lucent Bell Labs, Murray Hill, NJ 07974

Edited by Stephen E. Fienberg, Carnegie Mellon University, Pittsburgh, PA, and approved October 13, 2009 (received for review July 2, 2009)

Prompted by the increasing interest in networks in many fields, we present an attempt at unifying points of view and analyses of these objects coming from the social sciences, statistics, probability and physics communities. We apply our approach to the Newman–Girvan modularity, widely used for "community" detection, among others. Our analysis is asymptotic but we show by simulation and application to real examples that the theory is a reasonable guide to practice.

modularity | profile likelihood | ergodic model | spectral clustering

The social sciences have investigated the structure of small networks since the 1970s, and have come up with elaborate modeling strategies, both deterministic, see Doreian et al. (1) for a view, and stochastic, see Airoldi et al. (2) for a view and recent work. During the same period, starting with the work of Erdös and Rényi (3), a rich literature has developed on the probabilistic properties of stochastic models for graphs. A major contribution to this work is Bollobás et al. (4). On the whole, the goals of the analyses of ref. 4, such as emergence of the giant component, are not aimed at the statistical goals of the social science literature we have cited.

principle, "fail-safe" for rich enough models. Moreover, our point of view has the virtue of enabling us to think in terms of "strength of relations" between individuals not necessarily clustering them into communities beforehand.

We begin, using results of Aldous and Hoover (9), by introducing what we view as the analogues of arbitrary infinite population models on infinite unlabeled graphs which are "ergodic" and from which a subgraph with n vertices can be viewed as a piece. This development of Aldous and Hoover can be viewed as a generalization of deFinetti's famous characterization of exchangeable sequences as mixtures of i.i.d. ones. Thus, our approach can also be viewed as a first step in the generalization of the classical construction of complex statistical models out of i.i.d. ones using covariates, information about labels and relationships.

It turns out that natural classes of parametric models which approximate the nonparametric models we introduce are the "blockmodels" introduced by Holland, Laskey and Leinhardt ref. 10; see also refs. 2 and 11, which are generalizations of the Erdös–Rényi model. These can be described as follows.

In a possibly (at least conceptually) infinite population (of vertices) there are K unknown subcommunities. Unlabeled individuals (vertices) relate to each other through edges which for this

- 1D Problem
- 2D Problem
- Minimax Rate for Stochastic Block Model
- Minimax Rate for Graphon Estimation
- Adaptive Bayes Estimation

1D Problem

$$y_i = f(x_i) + \epsilon_i, \quad x_i = \frac{i}{n}, \quad i = 1, 2, .., n$$

$$\mathcal{F} = \Big\{ f : f(x) = q_1 \text{ for } x \in (0, 1/2], f(x) = q_2 \text{ for } x \in (1/2, 1] \Big\}$$

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^{n} (\hat{f}(x_i) - f(x_i))^2 \right) \asymp \frac{1}{n}.$$

1D Problem

$$y_i = f(x_i) + \epsilon_i, \quad x_i = \frac{i}{n}, \quad i = 1, 2, .., n$$

Without observing x, the problem is equivalent to

$$y_i = \theta_i + \epsilon_i. \qquad \Theta = \{\theta : \text{half } \theta_i \text{ is } q_1, \text{ half } \theta_i \text{ is } q_2\}$$

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2\right) \asymp 1.$$

2D Problem

$$y_{ij} = f(\xi_i, \xi_j) + \epsilon_{ij}, \quad \xi_i = \frac{i}{n}, \quad i, j = 1, 2, .., n$$

$\mathcal{F}$ collects f such that

$$f(x,y) = \begin{cases} q_1 & (x,y) \in [0,1/2) \times [0,1/2) \\ q_2 & (x,y) \in [0,1/2) \times [1/2,1] \\ q_3 & (x,y) \in [1/2,1] \times [0,1/2) \\ q_4 & (x,y) \in [1/2,1] \times [1/2,1] \end{cases}$$

2D Problem

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left(\frac{1}{n^2} \sum_{1 \leq i,j \leq n} (\hat{f}(\xi_i, \xi_j) - f(\xi_i, \xi_j))^2 \right) \asymp \frac{1}{n^2}.$$

How about without knowing the design?

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left(\frac{1}{n^2} \sum_{1 \leq i,j \leq n} (\hat{f}(\xi_i, \xi_j) - f(\xi_i, \xi_j))^2 \right) \asymp \frac{1}{n}.$$

2D Problem

Let $\theta_{ij} = f(\xi_i, \xi_j)$. Does θ_{ij} have any structure?

$\{\theta_{i1}, \theta_{i2}, ..., \theta_{in}\}$ are from the same row for each i.

$\{\theta_{1j}, \theta_{2j}, ..., \theta_{nj}\}$ are from the same column for each j.

2D Problem

$$y_{ij} = f(\xi_{ij}) + \epsilon_{ij}, \quad \xi_{ij} \in [0,1]^2, \quad i,j = 1,2,..,n$$

Without knowing the design?

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left(\frac{1}{n^2} \sum_{1 \le i,j \le n} (\hat{f}(\xi_{ij}) - f(\xi_{ij}))^2 \right) \asymp 1.$$

Stochastic Block Model

$$A_{ij} \sim \mathrm{Bernoulli}(\theta_{ij})$$

$$\Theta_2 = \left\{ \theta : \theta_{ij} = Q_{z(i)z(j)},\ \text{with } z : [n] \to [2] \right\}$$

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_2} \mathbb{E} \left(\frac{1}{n^2} \sum_{1 \leq i,j \leq n} (\hat{\theta}_{ij} - \theta_{ij})^2 \right) \asymp \frac{1}{n}.$$

Stochastic Block Model

$$A_{ij} \sim \text{Bernoulli}(\theta_{ij})$$

$$\Theta_k = \left\{ \theta : \theta_{ij} = Q_{z(i)z(j)}, \ \text{with } z : [n] \to [k] \right\}$$

Theorem 1.1. *Under the stochastic block model, we have*

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_k} \mathbb{E} \left\{ \frac{1}{n^2} \sum_{i,j \in [n]} (\hat{\theta}_{ij} - \theta_{ij})^2 \right\} \asymp \frac{k^2}{n^2} + \frac{\log k}{n},$$

for any $1 \leq k \leq n$.

Stochastic Block Model

Let $k \asymp n^\delta$, for $\delta \in [0, 1]$.

$$\frac{k^2}{n^2} + \frac{\log k}{n} \asymp \begin{cases} n^{-2} & \delta = 0, k = 1, \\ n^{-1} & \delta = 0, k > 1, \\ n^{-1} \log n & \delta \in (0, 1/2], \\ n^{-2(1-\delta)} & \delta \in (1/2, 1]. \end{cases}$$

Graphon Estimation

Theorem (Aldous-Hoover). *A random array* $\{A_{ij}\}$ *is jointly exchangeable in the sense that*

$$\{A_{ij}\}\overset{d}{=}\{A_{\sigma(i)\sigma(j)}\} \text{ for all permutation } \sigma,$$

if and only if it can be represented as follows: there is a random function $F:[0,1]^3 \to \mathbb{R}$ *such that*

$$A_{ij}\overset{d}{=}F(\xi_i,\xi_j,\xi_{ij}),$$

where $\{\xi_i\}$ *and* $\{\xi_{ij}\}$ *are i.i.d. Unif*$[0,1]$.

Graphon Estimation

When the graph is undirected and has no self-loop,

$$A_{ij}|\xi_i, \xi_j \sim \text{Bernoulli}(\theta_{ij}), \quad \theta_{ij} = f(\xi_i, \xi_j).$$

$$\xi_i \sim \text{Unif}(0, 1) \text{ i.i.d.}$$

$$\boxed{\text{Goal: recover } f.}$$

Graphon Estimation

$$A_{ij}|\xi_i,\xi_j \sim \text{Bernoulli}(\theta_{ij}), \quad \theta_{ij} = f(\xi_i,\xi_j).$$

$$(\xi_1,...,\xi_n) \sim \mathbb{P}_\xi$$

$$\text{Assumption: } f \in \mathcal{F}_\alpha(M).$$

Theorem 1.2. *Consider the Hölder class $\mathcal{F}_\alpha(M)$,*

$$\inf_{\hat{\theta}} \sup_{f\in\mathcal{F}_\alpha(M)} \sup_{\xi\sim\mathbb{P}_\xi} \mathbb{E}\left\{\frac{1}{n^2}\sum_{i,j\in[n]}(\hat{\theta}_{ij}-\theta_{ij})^2\right\} \asymp \begin{cases} n^{-\frac{2\alpha}{\alpha+1}}, & 0<\alpha<1, \\ \frac{\log n}{n}, & \alpha\geq 1. \end{cases}$$

The expectation is jointly over $\{A_{ij}\}$ and $\{\xi_i\}$.

Graphon Estimation

Proof:

$$\min_k \left\{ \frac{1}{k^{2\alpha}} + \frac{k^2}{n^2} + \frac{\log k}{n} \right\}$$

$\square$

Lower Bound Proof

When $1 < k \leq O(1)$, the minimax rate is $\dfrac{1}{n}$.

Sufficient to prove for $k = 2$.

Lower Bound Proof

Proposition (Fano)**.** *Let* (Θ, ρ) *be a metric space and* $\{P_\theta : \theta \in \Theta\}$ *a collection of probability measures. For any* $T \subset \Theta$*, denote by* $\mathcal{M}(\epsilon, T, \rho)$ *the* ϵ*-packing number of* T *w.r.t.* ρ*. Define the KL diameter of* T *by*

$$d_{\mathrm{KL}}(T) = \sup_{\theta, \theta' \in T} D(P_\theta || P_{\theta'}).$$

Then

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \rho^2 \left(\hat{\theta}(X), \theta\right) \geq \sup_{\epsilon > 0} \frac{\epsilon^2}{4} \left(1 - \frac{d_{\mathrm{KL}}(T) + \log 2}{\log \mathcal{M}(\epsilon, T, \rho)}\right)$$

Lower Bound Proof

- Construct a subset
- Upper bound the KL-diameter
- Lower bound the packing number

Lower Bound Proof

$$T = \Bigg\{ \{\theta_{ij}\} \in [0,1]^{n\times n} : \theta_{ij} = \frac{1}{2} \text{ for } (i,j) \in (S\times S)\cup(S^c\times S^c),$$

$$\theta_{ij} = \frac{1}{2} + \frac{c}{\sqrt{n}} \text{ for } (i,j) \in (S\times S^c)\cup(S^c\times S), \text{ with some } S \in \mathcal{S} \Bigg\}.$$

$$\begin{array}{c|c|c|}
\cline{2-3}
S & \frac{1}{2} & \frac{1}{2} + \frac{c}{\sqrt{n}} \\
\cline{2-3}
S^c & \frac{1}{2} + \frac{c}{\sqrt{n}} & \frac{1}{2} \\
\cline{2-3}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{S} & \multicolumn{1}{c}{S^c}
\end{array}$$

Lower Bound Proof

$$\rho^2(\theta, \theta') = \frac{1}{n^2} \sum_{1 \le i,j \le n} (\theta_{ij} - \theta'_{ij})^2 = \frac{2c^2}{n} \frac{|\mathbb{I}_S - \mathbb{I}_{S'}|}{n} \frac{(n - |\mathbb{I}_S - \mathbb{I}_{S'}|)}{n}.$$

$$\begin{array}{c|c|c|}
\cline{2-3}
S & \frac{1}{2} & \frac{1}{2} + \frac{c}{\sqrt{n}} \\
\cline{2-3}
S^c & \frac{1}{2} + \frac{c}{\sqrt{n}} & \frac{1}{2} \\
\cline{2-3}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{S} & \multicolumn{1}{c}{S^c}
\end{array}$$

Lower Bound Proof

Construct a subset:

$$T \subset \Theta_k$$

Upper bound the KL diameter

$$\sup_{\theta,\theta' \in T} D(\mathbb{P}_\theta || \mathbb{P}_{\theta'}) \leq \sup_{\theta,\theta' \in T} 8||\theta - \theta'||^2 \leq 8c^2 n.$$

Lower bound the packing number

Lower Bound Proof

Lower bound the packing number

$$\text{pick } \; S_1, \ldots, S_N \; \subset \; \mathcal{S} \quad \text{s.t.} \quad \frac{1}{4}n \leq |\mathbb{I}_S - \mathbb{I}_{S'}| \leq \frac{3}{4}n,$$

$$\rho^2(\theta, \theta') = \frac{2c^2}{n} \frac{|\mathbb{I}_S - \mathbb{I}_{S'}|}{n} \frac{(n - |\mathbb{I}_S - \mathbb{I}_{S'}|)}{n} \geq \frac{c^2}{8n} =: \epsilon^2.$$

$$\mathcal{M}(\epsilon, T, \rho) \geq N \geq \exp(c_1 n)$$

Lower Bound Proof

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_2} \mathbb{E} \left(\frac{1}{n^2} \sum_{1 \leq i,j \leq n} (\hat{\theta}_{ij} - \theta_{ij})^2 \right) \geq \frac{c^2}{32n} \left(1 - \frac{8c^2 n + \log 2}{c_1 n} \right) \gtrsim \frac{1}{n}.$$

Upper Bound

Oracle solution

When the clustering z is known, an obvious estimator

$$\hat{\theta}_{ij} = \frac{1}{|z^{-1}(a)||z^{-1}(b)|} \sum_{(i,j)\in z^{-1}(a)\times z^{-1}(b)} A_{ij}, \quad \text{for } (i,j) \in z^{-1}(a) \times z^{-1}(b)$$

achieves the rate

$$\|\hat{\theta} - \theta\|_{\mathrm{F}}^2 \leq O_P\left(k^2\right).$$

Upper Bound

An equivalent form (least squares)

Fixing the known z, then solve

$$\begin{aligned} \min_{\theta} \quad & \|A-\theta\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \theta_{ij} = Q_{z(i)z(j)} \text{ for some } Q = Q^T \in [0,1]^{k\times k} \end{aligned}$$

A natural estimator

Solve

$$\begin{aligned} \min_{\theta} \quad & \|A-\theta\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \theta_{ij} = Q_{z(i)z(j)} \text{ for some } Q = Q^T \in [0,1]^{k\times k} \\ & \text{and some } z : \{1,2,...,n\} \to \{1,2,...,k\}. \end{aligned}$$

$$\|\hat{\theta}-\theta\|_{\mathrm{F}}^2 \le O_P\left(k^2 + n\log k\right)$$

Bayes Estimation

1. Sample $k \sim \pi$. $\pi(k) \propto \exp\left(-D(k^2 + n \log k)\right)$

2. Sample $z \in \{z : [n] \to [k]\}$. uniform

3. Sample $Q \sim f$. ?

4. Let $\theta_{ij} = Q_{z(i)z(j)}$.

Bayes Estimation

1. Sample $k \sim \pi$. $\pi(k) \propto \exp\left(-D(k^2 + n\log k)\right)$

2. Sample $z \in \{z : [n] \to [k]\}$. uniform

3. Sample $Q \sim f$. $f(Q) = \frac{1}{2}\left(\frac{\lambda_k}{\sqrt{\pi}}\right)^{k^2} \boxed{\frac{\Gamma(k^2/2)}{\Gamma(k^2)}} e^{-\lambda_k ||Q||}$

4. Let $\theta_{ij} = Q_{z(i)z(j)}$.

Bayes Estimation

1. Sample $k \sim \pi$. $\pi(k) \propto \frac{\Gamma(k^2)}{\Gamma(k^2/2)} \exp\left(-D(k^2 + n\log k)\right)$

2. Sample $z \in \{z : [n] \to [k]\}$. uniform

3. Sample $Q \sim f$. $f(Q) = \frac{1}{2}\left(\frac{\lambda_k}{\sqrt{\pi}}\right)^{k^2} \boxed{\frac{\Gamma(k^2/2)}{\Gamma(k^2)}} e^{-\lambda_k ||Q||}$

4. Let $\theta_{ij} = Q_{z(i)z(j)}$.

Bayes Estimation

1. Sample $k \sim \pi$. $\pi(k) \propto \exp\left(-D(k^2 + n \log k)\right)$

2. Sample $z \in \{z : [n] \to [k]\}$. uniform

3. Sample $Q \sim f$. $f(Q) = \frac{1}{2}\left(\frac{\lambda_k}{\sqrt{\pi}}\right)^{k^2} e^{-\lambda_k ||Q||}$

4. Let $\theta_{ij} = Q_{z(i)z(j)}$.

Bayes Estimation

Theorem 1.3. *Consider* $\lambda_k = \beta\frac{n}{k}$ *for some constant* $\beta > 0$. *Then*

$$\mathbb{E}_{\theta^*}\Pi\left(\frac{1}{n^2}\sum_{i,j}(\theta_{ij}-\theta^*_{ij})^2 > M\left(\frac{k^2}{n^2}+\frac{\log k}{n}\right)\Big|A\right) \leq \exp\left(-C'\left(k^2+n\log k\right)\right),$$

for some constants $M, C' > 0$.

Reference

Gao, Chao, Yu Lu, and Harrison H. Zhou. "Rate-optimal Graphon Estimation." arXiv preprint arXiv:1410.5837 (2014).

Thank you