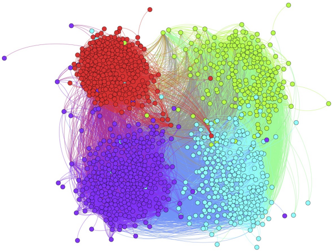
Rates of Convergence of Spectral Methods for Graphon Estimation

Jiaming Xu

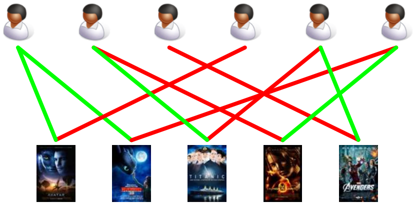
The Fuqua School of Business
Duke University

ICML, July 13, 2018

Statistical network analysis

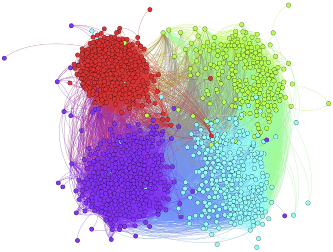


Simmons college friendship network

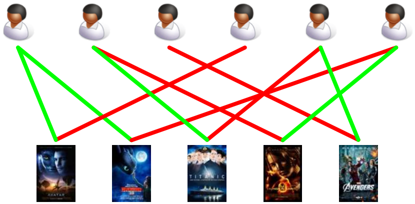


Moive recommender systems

Protein-protein interaction networks, DNA sequencing,...



Simmons college friendship network

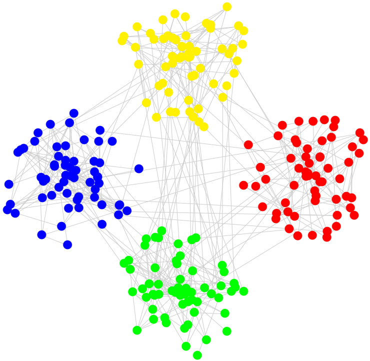


Moive recommender systems

Protein-protein interaction networks, DNA sequencing,...

A key problem

Learn the network generating mechanism – how edges are formed



Communities

$$x : \{1, 2, \dots, n\} \to \{1, 2, \dots k\}$$

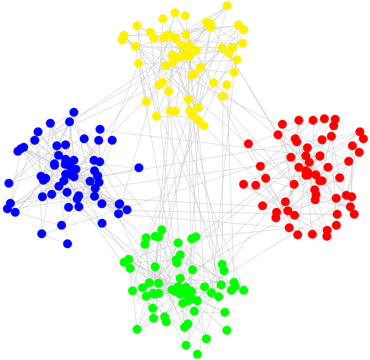
Connection probability

$$B = (b_{ab}) \in [0, 1]^{k \times k}$$
$$p_{ij} = b_{x(i)x(j)}$$

Observation of adjacency matrix

$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\,(p_{ij})$$

Stochastic block models [Holland-Laskey-Leinhardt '83]



Communities

$$x : \{1, 2, \dots, n\} \to \{1, 2, \dots k\}$$

Connection probability

$$B = (b_{ab}) \in [0,1]^{k \times k}$$
$$p_{ij} = b_{x(i)x(j)}$$

Observation of adjacency matrix

$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\,(p_{ij})$$

Limitation of SBM

Real networks may not display the clustered structure

Graphon models

Latent features

$$x_i \overset{\text{i.i.d.}}{\sim} \mu \quad \forall i \in [n]$$

Connection probability

$$p_{ij} = f\left(x_i, x_j\right) \in [0, 1]$$

Observation of adjacency matrix

$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\left(p_{ij}\right)$$

Graphon models

Latent features

$$x_i \overset{\text{i.i.d.}}{\sim} \mu \quad \forall i \in [n]$$

Connection probability

$$p_{ij} = f\left(x_i, x_j\right) \in [0, 1]$$

Observation of adjacency matrix

$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\left(p_{ij}\right)$$

- If $x_i \in \{1, \ldots, k\}$, reduces to SBM
- Originally proposed to study graph limits [Lovász '12]
- Popular framework for studying network generating mechanism estimation – estimate f or $P = \left(p_{ij}\right)$ from observation of A

Sparse graphon estimation

Partial observation of adjacency matrix

$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\left(\rho \times p_{ij}\right),$$

$\rho = \rho_n$ may converge to 0 as $n \to \infty$

Sparse graphon estimation

Partial observation of adjacency matrix

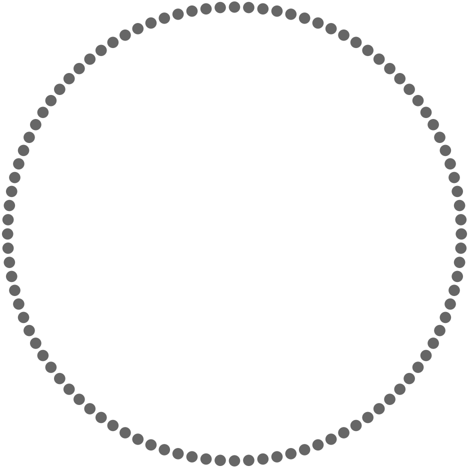
$$A_{ij} \mid x_i, x_j \sim \mathrm{Bern}\left(\rho \times p_{ij}\right),$$

$\rho = \rho_n$ may converge to 0 as $n \to \infty$

Focus of this talk: estimating $P = (p_{ij})$

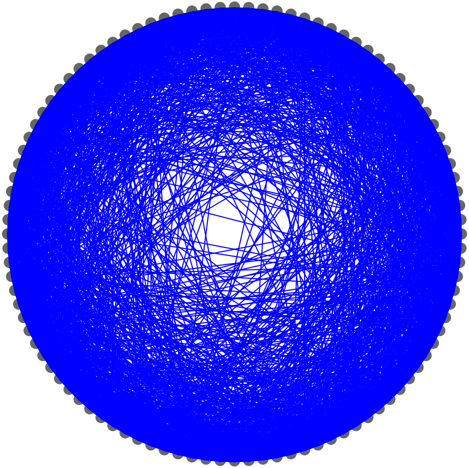
$$\mathsf{MSE}(\widehat{P}) = \frac{1}{n^2}\mathbb{E}\left[\|P - \widehat{P}\|_{\mathrm{F}}^2\right]$$

An Illustrating example



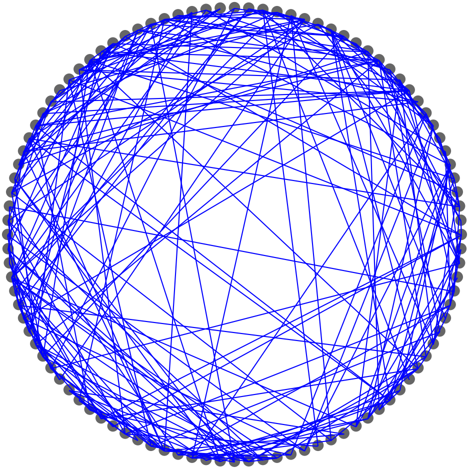
$n = 100$ and $f(x, y) = 1 - |x - y|$ and $\rho = 0.1$

An Illustrating example



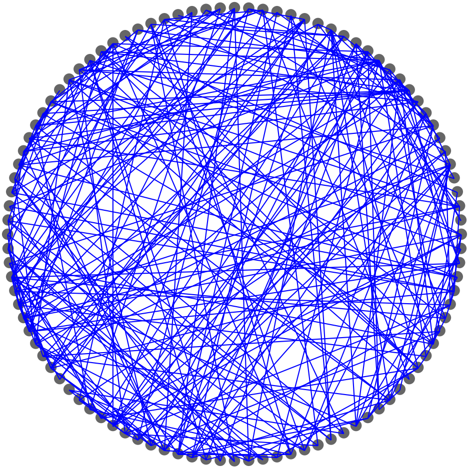
$n = 100$ and $f(x, y) = 1 - |x - y|$ and $\rho = 0.1$

An Illustrating example



$n = 100$ and $f(x, y) = 1 - |x - y|$ and $\rho = 0.1$

An Illustrating example



$n = 100$ and $f(x, y) = 1 - |x - y|$ and $\rho = 0.1$

Universal singular value thresholding (USVT) [Chatterjee '15]

1: Take singular value decomposition of A:

$$A = \sum_{i=1}^{n} s_i u_i v_i^\top$$

2: Threshold singular values:

$$\widehat{A} = \sum_{i=1}^{n} \mathbf{1}_{\{s_i \geq \tau\}} s_i u_i v_i^\top$$

3: Output

$$\widehat{P} = \frac{\widehat{A}}{\rho}$$

Universal singular value thresholding (USVT) [Chatterjee '15]

1: Take singular value decomposition of A:

$$A = \sum_{i=1}^{n} s_i u_i v_i^\top$$

2: Threshold singular values:

$$\widehat{A} = \sum_{i=1}^{n} \mathbf{1}_{\{s_i \geq \tau\}} s_i u_i v_i^\top$$

3: Output

$$\widehat{P} = \frac{\widehat{A}}{\rho}$$

Remarks: Threshold at $\tau = (1+\delta)\|A - \mathbb{E}[A]\|_2$ to suppress noise

Rates of Convergence of USVT for α-smooth graphons

Theorem (X. 17')

Suppose $n\rho = \Omega(\log n)$. *If* $f : [0,1]^d \times [0,1]^d \to [0,1]$ *is* α*-Hölder or Sobolev smooth, then*

$$\mathsf{MSE}(\widehat{P}) \lesssim (n\rho)^{-\frac{2\alpha}{2\alpha+d}}$$

Rates of Convergence of USVT for α-smooth graphons

Theorem (X. 17')

Suppose $n\rho = \Omega(\log n)$. If $f : [0,1]^d \times [0,1]^d \to [0,1]$ is α-Hölder or Sobolev smooth, then

$$\mathsf{MSE}(\widehat{P}) \lesssim (n\rho)^{-\frac{2\alpha}{2\alpha+d}}$$

- Need $n\rho = \Omega(\log n)$ to ensure $\| A - \mathbb{E}[A] \|_2 \lesssim \sqrt{n\rho}$
- α-smooth $\approx (\alpha - 1)$-derivatives of f are well-behaved
- Suboptimal comparing to minimax rate $(d = 1)$ [Gao-Lu-Zhou '15, Klopp-Tsybakov-Verzelen '15, Gao-Lu-Ma-Zhou '16]

$$\inf_{\widehat{P}} \sup_{P \in \mathcal{P}} \mathsf{MSE}(\widehat{P}) \asymp \begin{cases} \frac{\log(n\rho)}{n\rho} & n\rho \le n^{\alpha} \log^{\alpha+1}(n) \\ (n^2\rho)^{-\frac{\alpha}{\alpha+1}} & n\rho \ge n^{\alpha} \log^{\alpha+1}(n) \end{cases}$$

- Gap to minimax rate shrinks as α increases

Rates of Convergence of USVT for analytic graphons

Theorem (X. 17')

Suppose $n\rho = \Omega(\log n)$. *If* $f : [0,1]^d \times [0,1]^d \to [0,1]$ *is analytic, then*

$$\mathsf{MSE}(\widehat{P}) \lesssim \frac{\log^d(n\rho)}{n\rho}$$

Rates of Convergence of USVT for analytic graphons

Theorem (X. 17')

Suppose $n\rho = \Omega(\log n)$. *If* $f : [0,1]^d \times [0,1]^d \to [0,1]$ *is analytic, then*

$$\mathsf{MSE}(\widehat{P}) \lesssim \frac{\log^d(n\rho)}{n\rho}$$

- Analytic ≈ infinitely differentiable with locally converging Taylor series
- When $d = 1$, matches the minimax rate for α-smooth graphons with $\alpha \geq 1$

Proof ideas

1. To show [Shah-Balakrishnan-Guntuboyina-Wainwright '16, Koltchinskii-Lounici-Tsybakov '11, Klopp'11]

$$\mathsf{MSE}(\widehat{P}) \lesssim \min_{0 \le r \le n} \left(\frac{r}{n\rho} + \frac{1}{n^2} \sum_{i \ge r+1} \mathbb{E}\left[\lambda_i^2(P)\right] \right)$$

$\frac{r}{n\rho}$: estimation error of rank-r matrix

$\sum_{i \ge r+1} \lambda_i^2(P)$: approximation error of rank-r matrix

2. To show

$$\frac{1}{n^2} \sum_{i \ge r+1} \mathbb{E}\left[\lambda_i^2(P)\right] \lesssim \begin{cases} r^{-2\alpha/d} & \text{if } f \text{ is } \alpha\text{-smooth} \\ e^{-r^{1/d}} & \text{if } f \text{ is analytic} \end{cases}$$

using piecewise polynomial approximation of f [Birman-Solomyak '67 '77]

Summary and concluding remarks

When $n\rho = \Omega(\log n)$: the rates of convergence of USVT

$$\mathsf{MSE}(\widehat{P}) \lesssim \begin{cases} (n\rho)^{-\frac{2\alpha}{2\alpha+d}} & \text{if } f \text{ is } \alpha\text{-smooth} \\ \frac{\log^d(n\rho)}{n\rho} & \text{if } f \text{ is analytic} \end{cases}$$

When $n\rho = \Omega(1)$: Trim high-degree vertices or use convex programs

Summary and concluding remarks

When $n\rho = \Omega(\log n)$: the rates of convergence of USVT

$$\mathrm{MSE}(\widehat{P}) \lesssim \begin{cases} (n\rho)^{-\frac{2\alpha}{2\alpha+d}} & \text{if } f \text{ is } \alpha\text{-smooth} \\ \frac{\log^d(n\rho)}{n\rho} & \text{if } f \text{ is analytic} \end{cases}$$

When $n\rho = \Omega(1)$: Trim high-degree vertices or use convex programs

Future work

- Minimax optimal rate for high-dimensional feature space $d > 1$
- Applications of graphon estimation in real data