On Graphon Estimation

Nicolas Verzelen

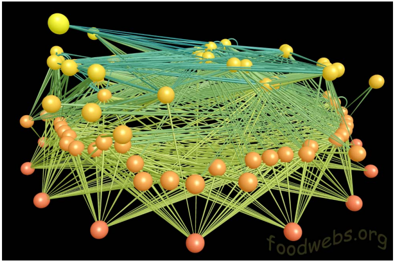
INRA, Montpellier

(Based on Joint Work with Olga Klopp and Alexander Tsybakov)

CIRM Luminy - July 10th

Network model

Network analysis is ubiquitous in social sciences, genomics, ecology,. . .



East-river trophic network [Yoon et al.(04)]

Objectives :

- Graph Visualization
- Backbone estimation
- Node clustering

Approach

- The modeling of real networks as random graphs.
- Model-based statistical analysis.

1 Graphon Model

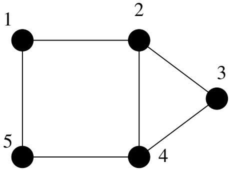
2 Towards Graphon Estimation

3 $\delta_\square$ and δ_2 Estimation of sparse Graphons

Graph Notation

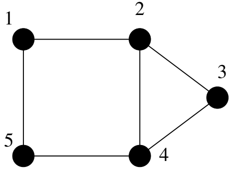
A (simple, undirected graph) $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ consists of

- a set of vertices $V = \{1, \ldots n\}$
- a set of edges $E \subset \{\{i, j\} : i, j \in V \text{ and } i \neq j\}$



A (simple, undirected graph) $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ consists of

- a set of vertices $V = \{1, \dots n\}$
- a set of edges $E \subset \{\{i,j\} : i,j \in V$ and $i \neq j\}$



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The corresponding adjacency matrix is denoted $\mathbf{A} = (\mathbf{A}_{i,j}) \in \{0,1\}^{n \times n}$, where $\mathbf{A}_{i,j} = 1 \Leftrightarrow (i,j) \in E$

Stochastic Block-Model (SBM) Holland et al. '80

A mixture model for random graphs : **K** classes.

SBM popular for clustering applications : generate graphs with a community structure

Latent labels : each node i belongs to class k with probability π_k :

$$\{\xi_i\}_i \text{ IID}, \ \xi_i \sim \mathcal{M}(1;\pi) \ ,$$

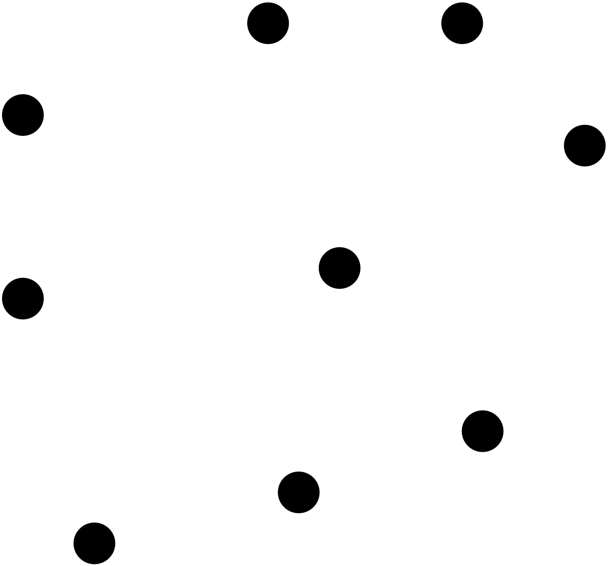
where $\pi = (\pi_1, \dots, \pi_K)$.

Observed edges : $(\mathbf{A}_{ij})$ are conditionally independent given the ξ_i's :

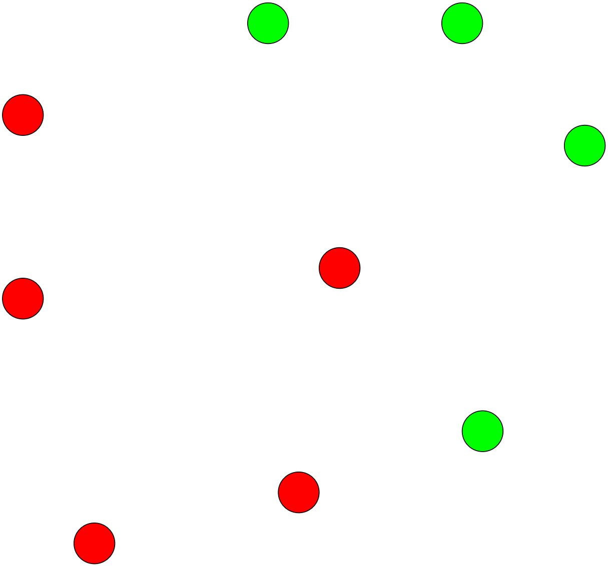
$$(\mathbf{A}_{ij} | \xi_i = k, \ \xi_j = l) \sim \mathcal{B}(\mathbf{Q}_{k,l})$$

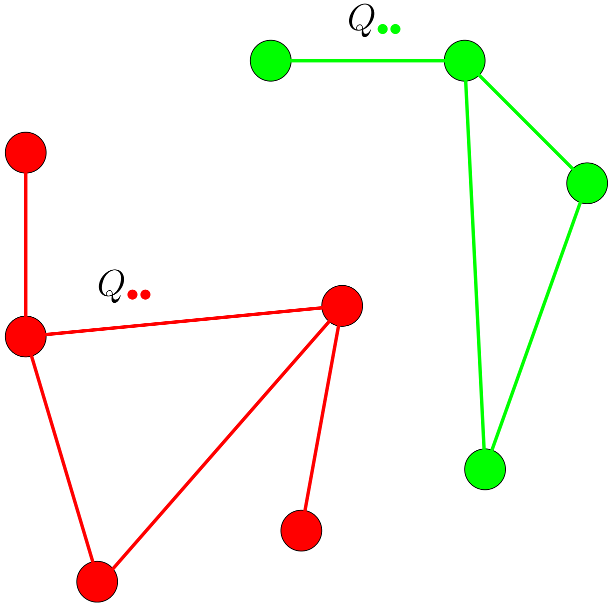
The symmetric $K \times K$ matrix $\mathbf{Q}$ is called the connectivity matrix.

(Basic approximation unit for more complex models)

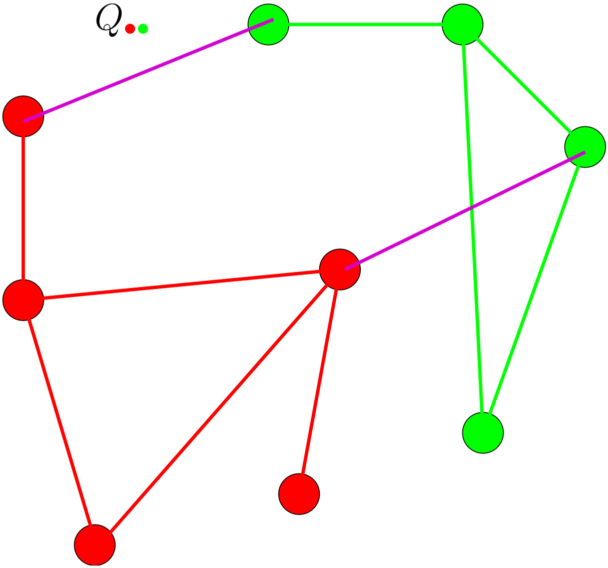


Example

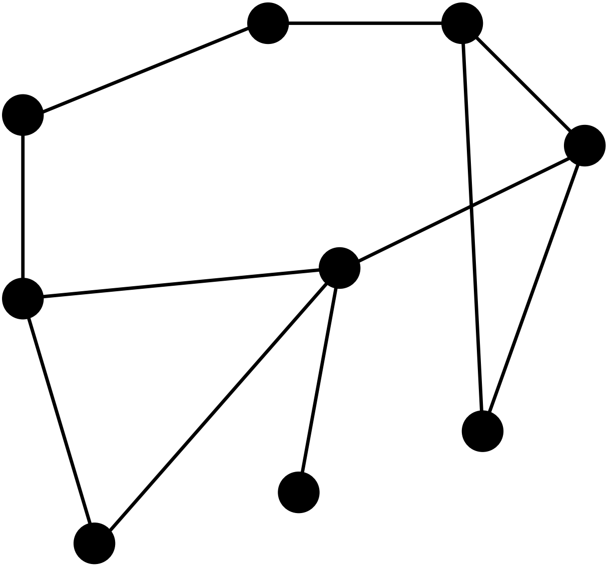



Q
Q

Example



Example

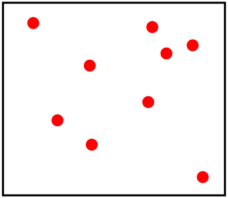


Definition

- ξ_i = unobserved position of node i in a latent space. e.g. $\xi_i \sim \mathcal{U}([0,1]^2)$
- Edges $\mathbf{A}_{ij}$ independent given ξ_i,

$$\mathbb{P}[\mathbf{A}_{ij} = 1] = \gamma(\|\xi_i - \xi_j\|_2$$

with $\gamma : \mathbb{R}^+ \to [0,1]$.



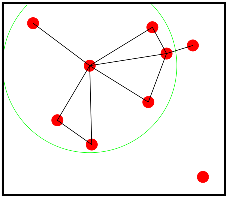
Latent space models (Hoff et al.,'02)

Definition

- ξ_i = unobserved position of node i in a latent space. e.g. $\xi_i \sim \mathcal{U}([0,1]^2)$
- Edges $\mathbf{A}_{ij}$ independent given ξ_i,

$$\mathbb{P}[\mathbf{A}_{ij} = 1] = \gamma(\|\xi_i - \xi_j\|_2$$

with $\gamma : \mathbb{R}^+ \to [0,1]$.



W-random Graph Model Diaconis & Janson('06)

SBM do not allow to analyze the fine structure of extremely large networks
$\rightsquigarrow$ **Non-parametric models**

Graphons

A graphon is a triplet (Ω, π, W) where :

- (Ω, π) is a Borel Probability space
- $W : \Omega \times \Omega \mapsto [0, 1]$ measurable

W-random Graph Model Diaconis & Janson('06)

SBM do not allow to analyze the fine structure of extremely large networks
$\rightsquigarrow$ **Non-parametric models**

Graphons

A graphon is a triplet (Ω, π, W) where :

- (Ω, π) is a Borel Probability space
- $W : \Omega \times \Omega \mapsto [0, 1]$ measurable

W-Random graph

W-random graph model of size n associated to (Ω, π, W) :

- $\xi = (\xi_1, \ldots, \xi_n)$ are sampled on Ω according to π.
- For each $i < j$, $\mathbf{A}_{ij} = 1$ with probability $W(\xi_i, \xi_j)$.

Notation : $\mathbb{P}^n_W$ xorresponding distribution ($\mathbb{P}_W := \mathbb{P}^\infty_W$)
$\boldsymbol{\Theta}_0$ defined by $\boldsymbol{\Theta}_{ij} = W(\xi_i, \xi_j)$ for $i \neq j$

Remarks

- $\mathbb{E}[\mathbf{A}|\xi] = \boldsymbol{\Theta}_0 \rightsquigarrow$ cond. to ξ, inhomogeneous random graph with Matrix $\boldsymbol{\Theta}_0$.
- If W is a k step-function, $\mathbf{A}$ SBM with k blocks

Universality of W-random graph model.

Let $\underline{\mathbf{A}} : \mathbb{N} \times \mathbb{N} \to \{0,1\}$ symmetric = the adjacency function of an infinite graph
For τ permutation, $\underline{\mathbf{A}}^\tau$ defined by $\underline{\mathbf{A}}^\tau[i,j] = \underline{\mathbf{A}}[\tau(i), \tau(j)]$.

Joint Exchangeability

The distribution of $\underline{\mathbf{A}}$ is jointly exchangeable if

$$\underline{\mathbf{A}} \sim \underline{\mathbf{A}}^\tau \ , \qquad \text{for any permutation } \tau.$$

Theorem (Aldous-Hoover Representation Theorem('79))

If the distribution of $\underline{\mathbf{A}}$ is jointly exchangeable, then there exists μ such that

$$(\Omega, \pi, W) \sim \mu$$

$$[\underline{\mathbf{A}} | (\Omega, \pi, W)] \sim \mathbb{P}_W$$

$\rightsquigarrow$ W-random graph distribution correspond to extremal points of the set of jointly exchangeable distributions.

Sparse Graphons

For W-random graph, $\mathbb{E}_W[\#E] \asymp n^2 \rightsquigarrow$ Realized graphs are dense

Sparse Graphon Models

(Ω, π, W, ρ_n) with $\rho_n \to_{n\to\infty} 0$.

1. Sample ξ according to π
2. For each $i < j$, draw an edge between i and j with probability $\rho_n W(\xi_i, \xi_j)$

Other methods :

- L_p graphon (e.g Borgs et al.('15))
- Graphex (e.g.Veitch & Roy('16))

Objective

Given an observation $\mathbf{A}$, goal = infering the graphon (Ω, π, W) in some sense...

Caveats :

- Identifiability
- Loss functions
- Approximation class

Objective

Given an observation $\mathbf{A}$, goal = infering the graphon (Ω, π, W) in some sense...

Caveats :

- Identifiability
- Loss functions
- Approximation class

Lemma

(Ω, π, W) *graphon*
$\tau : (\Omega', \pi') \mapsto (\Omega, \pi)$ *measure-preserving*
W^τ *be such that* $W^\tau(x, y) := W(\tau(x), \tau(y))$*. Then* $\mathbb{P}_W = \mathbb{P}_{W^\tau}$*.*

Two Consequences :

- Triplet (Ω, π, W) is not identifiable
- Sufficient to consider graphons on $([0, 1], \lambda)$ [but problematic]
  $\rightsquigarrow \mathcal{W}$: space of graphons on $([0, 1], \lambda)$.

Identifiability (fd)

Even Restricting to $([0,1],\lambda)$, the topology of a network **invariant** wrt node labeling change :

Weak isomorphism

Two graphons U and W are **weakly isomorphic** if there exist measure preserving maps ϕ, $\psi : [0,1] \to [0,1]$ such that $U^{\phi} = W^{\psi}$ almost everywhere.

Identifiability (fd)

Even Restricting to $([0,1],\lambda)$, the topology of a network **invariant** wrt node labeling change :

Weak isomorphism

Two graphons U and W are **weakly isomorphic** if there exist measure preserving maps ϕ, $\psi : [0,1] \to [0,1]$ such that $U^{\phi} = W^{\psi}$ almost everywhere.

Proposition (Lovász('12))

$\mathbb{P}_U = \mathbb{P}_W$ *if and only if* U *and* W *are weakly isomorphic.*

$\rightsquigarrow$ one can only perform inference in $\widetilde{\mathcal{W}}$ (equivalence classes of $\mathcal{W}$ wrt weak isomorphism)

Distance betw. **graphs** $\rightsquigarrow$ Distance betw.**graphons** $\rightsquigarrow$ Distance bet. **equivalence classes**

Distance on Graphs :

$\|\mathbf{A}-\mathbf{B}\|_2 := \frac{1}{n}\sqrt{\sum_{ij}(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2}$ $\rightsquigarrow$ Frobenius distance

$\|\mathbf{A}-\mathbf{B}\|_\square := \frac{1}{n^2}\max_{S,T\subset[n]}\left|\sum_{i\in S,j\in T}\mathbf{A}_{ij}-\mathbf{B}_{ij}\right|$ $\rightsquigarrow$ Cut Distance

Metrics/Loss functions on $\widetilde{\mathcal{W}}$

Distance betw. **graphs** $\rightsquigarrow$ Distance betw.**graphons** $\rightsquigarrow$ Distance bet. **equivalence classes**

Distance on Graphs :

$\|\mathbf{A}-\mathbf{B}\|_2 := \frac{1}{n}\sqrt{\sum_{ij}(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2}$ $\rightsquigarrow$ Frobenius distance

$\|\mathbf{A}-\mathbf{B}\|_\square := \frac{1}{n^2}\max_{S,T\subset[n]}\left|\sum_{i\in S,j\in T}\mathbf{A}_{ij}-\mathbf{B}_{ij}\right|$ $\rightsquigarrow$ Cut Distance

Norms on Graphons :

$\|W\|_2 := [\int_{[0,1]^2} W^2(x,y)dxdy]^{1/2}$ $\rightsquigarrow$ Frobenius norm

$\|W\|_\square := \sup_{S,T\subset[0,1]}\left|\int_{S\times T} W(x,y)dxdy\right|$ $\rightsquigarrow$ Cut norm (cornerstone of graph limits)

Metrics/Loss functions on $\widetilde{\mathcal{W}}$

Distance betw. **graphs** $\rightsquigarrow$ Distance betw.**graphons** $\rightsquigarrow$ Distance bet. **equivalence classes**

Distance on Graphs :

$\|\mathbf{A}-\mathbf{B}\|_2 := \frac{1}{n}\sqrt{\sum_{ij}(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2} \rightsquigarrow$ Frobenius distance

$\|\mathbf{A}-\mathbf{B}\|_\square := \frac{1}{n^2}\max_{S,T\subset[n]}\left|\sum_{i\in S,j\in T}\mathbf{A}_{ij}-\mathbf{B}_{ij}\right| \rightsquigarrow$ Cut Distance

Norms on Graphons :

$\|W\|_2 := [\int_{[0,1]^2} W^2(x,y)dxdy]^{1/2} \rightsquigarrow$ Frobenius norm

$\|W\|_\square := \sup_{S,T\subset[0,1]}\left|\int_{S\times T} W(x,y)dxdy\right| \rightsquigarrow$ Cut norm (cornerstone of graph limits)

Distances on $\widetilde{\mathcal{W}}$.

$\mathcal{M}$: Measure-preserving bijections $\tau : [0,1]\to[0,1]$

l_2 distance $\delta_2(W,W_1) := \inf_{\tau\in\mathcal{M}}\|W-W_1^\tau\|_2$

Cut distance $\delta_\square(W,W_1) := \inf_{\tau\in\mathcal{M}}\|W-W_1^\tau\|_\square$

Metrics/Loss functions on $\widetilde{\mathcal{W}}$

Distance betw. **graphs** $\rightsquigarrow$ Distance betw.**graphons** $\rightsquigarrow$ Distance bet. **equivalence classes**

Distance on Graphs :

$\|\mathbf{A}-\mathbf{B}\|_2 := \frac{1}{n}\sqrt{\sum_{ij}(\mathbf{A}_{ij}-\mathbf{B}_{ij})^2}$ $\rightsquigarrow$ Frobenius distance

$\|\mathbf{A}-\mathbf{B}\|_\square := \frac{1}{n^2}\max_{S,T\subset[n]}\left|\sum_{i\in S, j\in T}\mathbf{A}_{ij}-\mathbf{B}_{ij}\right|$ $\rightsquigarrow$ Cut Distance

Norms on Graphons :

$\|W\|_2 := [\int_{[0,1]^2} W^2(x,y)dxdy]^{1/2}$ $\rightsquigarrow$ Frobenius norm

$\|W\|_\square := \sup_{S,T\subset[0,1]}\left|\int_{S\times T} W(x,y)dxdy\right|$ $\rightsquigarrow$ Cut norm (cornerstone of graph limits)

Distances on $\widetilde{\mathcal{W}}$.

$\mathcal{M}$: Measure-preserving bijections $\tau : [0,1]\to[0,1]$

l_2 distance $\delta_2(W,W_1) := \inf_{\tau\in\mathcal{M}}\|W-W_1^\tau\|_2$

Cut distance $\delta_\square(W,W_1) := \inf_{\tau\in\mathcal{M}}\|W-W_1^\tau\|_\square$

These metrics are not equivalent : $\delta_\square(W,\widehat{W}) \leq \delta_2(W,\widehat{W})$

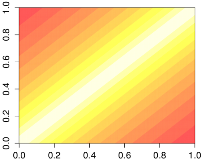
Regularity Lemma and approximation by SBMs

Proposition (Szemerédi ('75), Frieze and Kannan ('99))

For any $W \in \mathcal{W}$ and any k, there exists a k-step graphon W_k such that

$$\|W - W_k\|_{\square} \lesssim \frac{1}{\sqrt{\log(k)}}$$

This rate is universal!



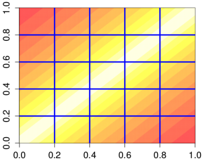
Proposition (Szemerédi ('75), Frieze and Kannan ('99))

For any $W \in \mathcal{W}$ and any k, there exists a k-step graphon W_k such that

$$\|W - W_k\|_{\square} \lesssim \frac{1}{\sqrt{\log(k)}}$$

This rate is universal !

SBM as basic stones for approximating graphons



Regularity Lemma and approximation by SBMs

Proposition (Szemerédi ('75), Frieze and Kannan ('99))

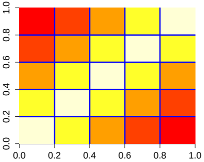
For any $W \in \mathcal{W}$ and any k, there exists a k-step graphon W_k such that

$$\|W - W_k\|_\square \lesssim \frac{1}{\sqrt{\log(k)}}$$

This rate is universal !

SBM as basic stones for approximating graphons

Obviously false for δ_2 : similar to histograms in classical Nonparametric Estimation



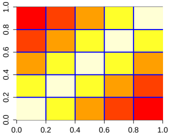
Estimating $f_0 = \rho_n W_0$

$$\rho_n W_0 \longrightarrow \mathbf{\Theta}_0 \longrightarrow \mathbf{A}$$

General Scheme :

1. Estimating the matrix $\mathbf{\Theta}_0 := \mathbb{E}[\mathbf{A}|\xi]$ by $\widehat{\mathbf{\Theta}}$.
2. From matrix to graphon. Given $\mathbf{\Theta}$, define the **empirical graphon** $\widetilde{f}_{\mathbf{\Theta}}$ as the n piecewise constant function :

$$\widetilde{f}_{\mathbf{\Theta}}(x,y) = \mathbf{\Theta}_{\lceil nx \rceil, \lceil ny \rceil}, \qquad x, y \in [0,1]$$



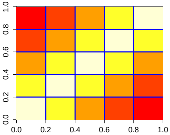
Estimating $f_0 = \rho_n W_0$

$$\rho_n W_0 \longrightarrow \mathbf{\Theta}_0 \longrightarrow \mathbf{A}$$

General Scheme :

1. Estimating the matrix $\mathbf{\Theta}_0 := \mathbb{E}[\mathbf{A}|\xi]$ by $\widehat{\mathbf{\Theta}}$.
2. From matrix to graphon. Given $\mathbf{\Theta}$, define the **empirical graphon** $\widetilde{f}_{\mathbf{\Theta}}$ as the n piecewise constant function :

$$\widetilde{f}_{\mathbf{\Theta}}(x, y) = \mathbf{\Theta}_{\lceil nx \rceil, \lceil ny \rceil}, \qquad x, y \in [0, 1]$$



Lemma

For any estimator $\widehat{\mathbf{\Theta}}$ and any norm N,

$$\mathbf{E}\left[\delta_N(\widetilde{f}_{\widehat{\mathbf{\Theta}}}, f_0)\right] \leq \mathbf{E}\left[\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0\|_N\right] + \underbrace{\mathbf{E}\left[\delta_N\left(\widetilde{f}_{\mathbf{\Theta}_0}, f_0\right)\right]}_{\text{agnostic error}}$$

Estimating Θ_0

- Raw data : $\mathbf{A}$
- Restricted Least Squares estimator : (RLS) Wolfe & Olhlede ('13), Borgs et al.('15), Klopp, Tsybakov, V.('17), Gao et al.('17)

$$\widetilde{\mathbf{\Theta}}_\lambda \in \arg \min_{\mathbf{\Theta} \in \mathrm{SBM}(k):\ \|\mathbf{\Theta}\|_\infty \leq \mathbf{r}} \|\mathbf{A} - \mathbf{\Theta}\|_2^2,$$

where $r \in (0,1)$ and $\mathrm{SBM}(k)$ space of k block-constant matrix.
($\widehat{\mathbf{\Theta}}_k^r$ is not polynomial-time computable)

- Singular Value Thresholding : $\widetilde{\mathbf{\Theta}}_\lambda$ e.g. Chatterjee('12), Klopp & V.('17)

$$\widetilde{\mathbf{\Theta}}_\lambda := \sum_{j:\sigma_j(\mathbf{A}) \geq \lambda} \sigma_j(\mathbf{A}) u_j(\mathbf{A}) v_j(\mathbf{A})^T \ ,$$

Probability Matrix Estimation in $\|.\|_2$

Theorem (Oracle inequality Klopp, Tsybakov, V.('17))

For any $\mathbf{\Theta}_0$ *such that* $\|\mathbf{\Theta}_0\|_\infty \leq r$*,*

$$\mathbb{E}\left[\|\widehat{\mathbf{\Theta}}^r_k - \mathbf{\Theta}_0\|_2^2\right] \lesssim \min_{\mathbf{\Theta}\in\mathrm{SBM}(k)} \|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_2^2 + r\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right)$$

(Minimax optimal over $\mathrm{SBM}(k) \cap \mathcal{B}_\infty(r)$) (Gao et al. ('15))

Two terms :

- $\frac{k^2}{n^2} \rightsquigarrow$ parametric rate $(k(k+1)/2$ parameter to estimate)
- $\frac{n\log(k)}{n^2} \rightsquigarrow$ clustering rate (of order k^n possible partitions)

Probability Matrix Estimation in $\|.\|_2$

Theorem (Oracle inequality Klopp, Tsybakov, V.('17))

For any $\mathbf{\Theta}_0$ *such that* $\|\mathbf{\Theta}_0\|_\infty \leq r$*,*

$$\mathbb{E}\left[\|\widehat{\mathbf{\Theta}}_k^r - \mathbf{\Theta}_0\|_2^2\right] \lesssim \min_{\mathbf{\Theta} \in \mathrm{SBM}(k)} \|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_2^2 + r\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right)$$

(Minimax optimal over $\mathrm{SBM}(k) \cap \mathcal{B}_\infty(r)$) (Gao et al. ('15))

Two terms :

- $\frac{k^2}{n^2} \rightsquigarrow$ parametric rate ($k(k+1)/2$ parameter to estimate)
- $\frac{n\log(k)}{n^2} \rightsquigarrow$ clustering rate (of order k^n possible partitions)

SVT estimator

Fix $\lambda = c\sqrt{n\|\mathbf{\Theta}_0\|_\infty}$. For all k,

$$\mathbb{E}[\|\widetilde{\mathbf{\Theta}}_\lambda - \mathbf{\Theta}_0\|_2^2] \lesssim \min_{\mathbf{\Theta} \in \mathrm{SBM}(k)} \|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_2^2 + \frac{\|\mathbf{\Theta}_0\|_\infty k}{n},$$

Loss of order $\frac{k}{\log(k)} \wedge \frac{n}{k}$ wrt RLS estimators

Best known polynomial time bound

δ_2 Graphon Estimation for k-step functions

$\mathcal{W}[k]$: Collection of k-**Step function graphons**
Here $f_0 = \rho_n W_0$ with $W \in \mathcal{W}[k]$

Proposition

$$\mathbb{E}\left[\delta^2\left(\widetilde{f}_{\boldsymbol{\Theta}_0}, f_0\right)\right] \lesssim \rho_n^2 \sqrt{\frac{k}{n}}$$

If $\rho_n \leq r$ then for RLS

$$\mathbb{E}\left[\delta^2\left(f_{\widehat{\boldsymbol{\Theta}}_k^r}, f_0\right)\right] \lesssim \rho_n \left(\frac{k^2}{n^2} + \frac{\log(k)}{n}\right) + \rho_n^2 \sqrt{\frac{k}{n}}$$

δ_2 Graphon Estimation for k-step functions

$\mathcal{W}[k]$: Collection of k-**Step function graphons**
Here $f_0 = \rho_n W_0$ with $W \in \mathcal{W}[k]$

Proposition

$$\mathbb{E}\left[\delta^2\left(\widetilde{f}_{\mathbf{\Theta}_0}, f_0\right)\right] \lesssim \rho_n^2 \sqrt{\frac{k}{n}}$$

If $\rho_n \leq r$ then for RLS

$$\mathbb{E}\left[\delta^2\left(f_{\widehat{\mathbf{\Theta}}_k^r}, f_0\right)\right] \lesssim \rho_n \left(\frac{k^2}{n^2} + \frac{\log(k)}{n}\right) + \rho_n^2 \sqrt{\frac{k}{n}}$$

For SVT

$$\mathbb{E}\left[\delta^2\left(f_{\widetilde{\mathbf{\Theta}}_\lambda}, f_0\right)\right] \lesssim \rho_n \frac{k}{n} + \rho_n^2 \sqrt{\frac{k}{n}}$$

(RLS is Minimax optimal (up to possible $\log(k)$ term)) Klopp et al. ('17)

δ_2 Graphon Estimation for k-step functions

$\mathcal{W}[k]$: Collection of **k-Step function graphons**
Here $f_0 = \rho_n W_0$ with $W \in \mathcal{W}[k]$

Proposition

$$\mathbb{E}\left[\delta^2\left(\widetilde{f}_{\mathbf{\Theta}_0}, f_0\right)\right] \lesssim \rho_n^2\sqrt{\frac{k}{n}}$$

If $\rho_n \leq r$ then for RLS

$$\mathbb{E}\left[\delta^2\left(f_{\widehat{\mathbf{\Theta}}_k^r}, f_0\right)\right] \lesssim \rho_n\left(\frac{k^2}{n^2} + \frac{\log(k)}{n}\right) + \rho_n^2\sqrt{\frac{k}{n}}$$

For SVT

$$\mathbb{E}\left[\delta^2\left(f_{\widetilde{\mathbf{\Theta}}_\lambda}, f_0\right)\right] \lesssim \rho_n\frac{k}{n} + \rho_n^2\sqrt{\frac{k}{n}}$$

(RLS is Minimax optimal (up to possible $\log(k)$ term)) Klopp et al. ('17)

(i) **Weakly sparse graphs :** $\Longrightarrow$ the agnostic error dominates.

(ii) **Moderately sparse graphs :** $\Longrightarrow$ the Probability matrix estimation error dominates

(iii) **Highly sparse graphs :** The null estimator $\tilde{f} \equiv 0$ is of smaller order

Probability Matrix Estimation in cut norm

Proposition

For any probability matrix $\mathbf{\Theta}_0$ such that $\|\mathbf{\Theta}_0\|_\infty \geq 1/n$,

$$\mathbf{E}\left[\|\mathbf{A} - \mathbf{\Theta}_0\|_\square\right] \leq 12\sqrt{\frac{\|\mathbf{\Theta}_0\|_\infty}{n}}$$

- Valid for all matrices $\mathbf{\Theta}_0$. Optimal convergence rate (even for simple classes such as two-block matrices)
- More refined estimators (SVT) do not decrease the performances but RLS may be biased.

Graphon Estimation in Cut distance

Theorem ((Consequence of Szemeredi's Lemma) Lovász, '12)

For all W_0 with $\rho_n = 1$, one has whp

$$\delta_{\square}\left(\tilde{f}_{\mathbf{A}}, W_0\right) \lesssim \frac{1}{\sqrt{\log(n)}} .$$

Valid for all graphons !

Theorem (Klopp and V., '17)

For all $W_0 \in \mathcal{W}[k]$ and $\rho_n > 0$, we have

$$\mathbf{E}_{W_0}\left[\delta_{\square}\left(\widetilde{f}_{\mathbf{A}}, f_0\right)\right] \lesssim \rho_n \min\left(\sqrt{\frac{k}{n\log(k)}}, \frac{1}{\sqrt{\log(n)}}\right) + \sqrt{\frac{\rho_n}{n}}$$

- Similar bound for the SVT estimator $\widetilde{f}_{\widetilde{\mathbf{\Theta}}_\lambda}$
- **This convergence rate is optimal**

Graphon Estimation in Cut distance

Theorem ((Consequence of Szemeredi's Lemma) Lovász, '12)

For all W_0 with $\rho_n = 1$, one has whp

$$\delta_{\square}\left(\tilde{f}_{\mathbf{A}}, W_0\right) \lesssim \frac{1}{\sqrt{\log(n)}}\ .$$

Valid for all graphons !

Theorem (Klopp and V., '17)

For all $W_0 \in \mathcal{W}[k]$ and $\rho_n > 0$, we have

$$\mathbf{E}_{W_0}\left[\delta_{\square}\left(\widetilde{f}_{\mathbf{A}}, f_0\right)\right] \lesssim \rho_n \min\left(\sqrt{\frac{k}{n\log(k)}}, \frac{1}{\sqrt{\log(n)}}\right) + \sqrt{\frac{\rho_n}{n}}$$

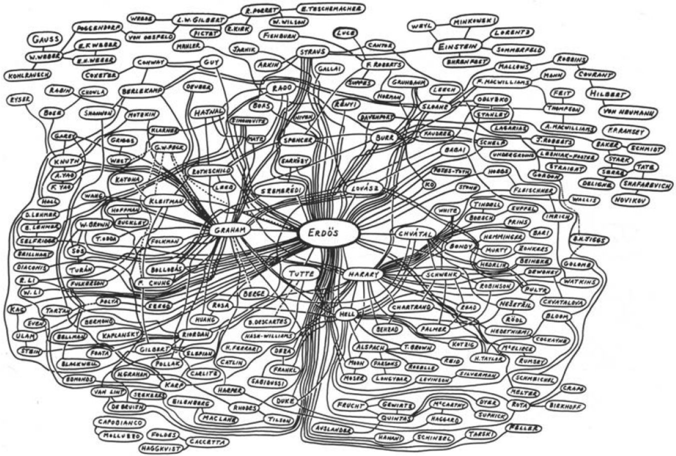
- Similar bound for the SVT estimator $\widetilde{f}_{\widetilde{\mathbf{\Theta}}_\lambda}$
- **This convergence rate is optimal**

(i) **Weakly sparse graphs** Agnostic error dominates.

(ii) **Moderately sparse graphs** Probability Matrix Estimation error dominates.

Discussion

- Non-parametric viewpoint on network analysis
- Identifiability Caveats
- Importance of the metric choice
- Good behavior of universal Singular Value Thresholding estimator.
- Computational barriers for estimation in δ_2 ?
- Less results for L_p graphons (Borgs et al.'16) and graphex
- Incorporating some geometry into estimation
  $\rightsquigarrow$ Functional Estimation (e.g. Issartel'17+)



ERDÖS
GRAHAM
CHVÁTAL
HARARY
TUTTE
HELL
LOVÁSZ
SZEMERÉDI
KLEITMAN
GAUSS
W. WEBER
E.F. WEBER
E.H. WEBER
KOHLRAUSCH
POGGENDORF
VON OESFELD
WEBER
L.W. GILBERT
R. PORRET
E. TESCHEMACHER
W. WILSON
E. KIRK
PICTET
FISHBURN
LUCE
CANTOR
WEYL
MINKOWSKI
LORENTZ
EINSTEIN
SOMMERFELD
EHRENFEST
MALLOWS
ROBBINS
COURANT
HILBERT
VON NEUMANN
F.P. RAMSEY
BAKER
SCHMIDT
STARR
TATE
SERRE
DELIGNE
SHAFAREVICH
NOVIKOV
WALLIS
MAHLER
JARNIK
STRAUS
ARKIN
GALLAI
F. ROBERTS
SUPPES
GRUNBAUM
NORMAN
LEECH
F. MACWILLIAMS
MANN
FEIT
THOMPSON
A. MACWILLIAMS
J. ROBERTS
STRAIGHT
GORDON
CONWAY
GUY
COXETER
RADO
DEUBER
BERLEKAMP
RABIN
CHOWLA
RYSER
BOSE
SHANNON
MOTZKIN
HAJNAL
BOAS
RÉNYI
DAVENPORT
SLOANE
ODLYZKO
STANLEY
LAGARIAS
FAUDREE
BURR
SPENCER
SIMONOVITS
NIVEN
KLARNER
GAREY
GRIGGS
G.W. PECK
KNUTH
WEST
A. YAO
F. YAO
KATONA
ROTHSCHILD
ERDÖS
LEEB
SCHELP
UNDERGROUND
BABAI
HOBBS
FLEISCHNER
STONE
KO
HALL
D. LEHMER
E. LEHMER
SELFRIDGE
BRILLHART
DIACONIS
R. LI
W. LI
KAC
EVEN
ULAM
STEIN
WANG
HOFFMAN
W. BROWN
BUCKLEY
T. ODDA
FOLKMAN
SÓS
TURÁN
BOLLOBÁS
F. CHUNG
FULKERSON
POLYÁ
TARTAN
SZEGÖ
ROSA
BERGE
HUANG
B. DESCARTES
NASH-WILLIAMS
BERMOND
KAPLANSKY
BELLMAN
RIORDAN
GILBERT
FOATA
SLEPIAN
H. FERRARI
BLACKWELL
POLLAK
CATLIN
DEZA
FRANKL
EDMONDS
N. GRAHAM
CARLITZ
KARP
SABIDUSSI
VAN LINT
SEEKAMP
HARPER
DE BRUIJN
EILENBERG
RHODES
DUKE
MAC LANE
TILSON
CAPOBIANCO
MOLLUZZO
FOLDES
CACCETTA
HAGGKVIST
WHITE
TINDELL
EGGLETON
IMRICH
BOESCH
PRINS
BARI
HEMMINGER
B.N. JIGGS
BONDY
MURTY
ZONKERS
HEDRLIN
SEINSCHE
DEWDNEY
GOLOMB
SCHWENK
WATKINS
ROBINSON
PALTZ
CHVATALOVÁ
CHARTRAND
READ
NEŠETŘIL
BLOOM
RÖDL
BEHZAD
PALMER
HEDETNIEMI
COCKAYNE
ALSPACH
T. BROWN
KOTZIG
McELIECE
H. TAYLOR
REID
MOON
PARSONS
ROSELLE
RUMSEY
MOSER
LONGYEAR
LEVINSON
SILVERMAN
SCHMEICHEL
MELTER
CRAPO
FRUCHT
GEWIRTZ
McCARTHY
DYER
ROTA
BIRKHOFF
QUINTAS
HAGGARD
SUPNICK
AUSLANDER
HANANI
SCHINZEL
TARSKI
FELLER

Thank You !