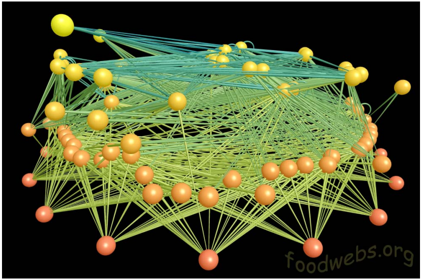
Oracle inequalties for network models and sparse graphon estimation

Alexandre Tsybakov
ENSAE

joint work with **Olga Klopp**, **Nicolas Verzelen** and with **Pierre Bellec**

Luminy, February 3, 2016

Networks arise in many different fields: social sciences, computer science, statistical physics, biology,...



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

Approach

- Modeling of real networks as random graphs.
- Statistical analysis of random graphs.

Blog Network

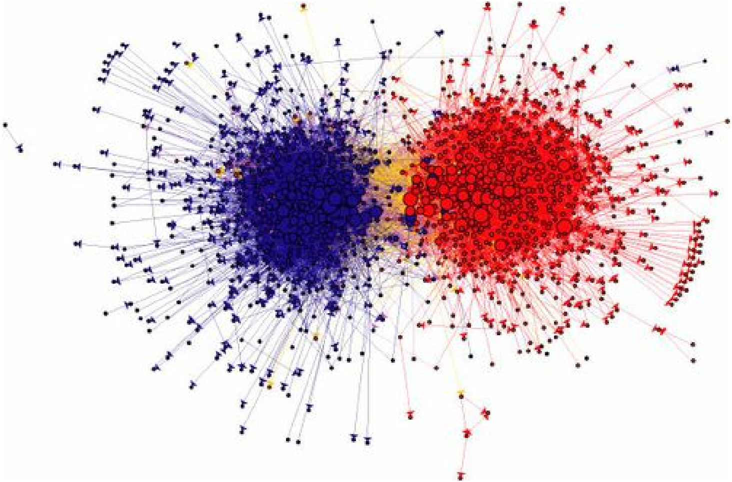
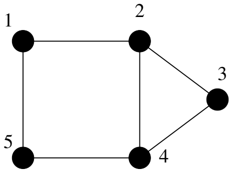


Figure : Mutual citations in blogs of politicians. Red: democrats, Blue: republicans.

Graph Notation

A simple, undirected graph consists of

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



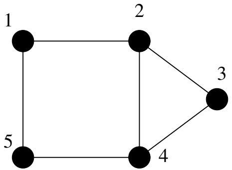
$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Adjacency matrix of a graph is defined as $\boldsymbol{A} = (\boldsymbol{A}_{ij}) \in \{0,1\}^{n \times n}$, where $\boldsymbol{A}_{ij} = 1 \Leftrightarrow (i,j) \in E$. Symmetric matrix with zeros on the diagonal.

Graph Notation

A simple, undirected graph consists of

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



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

Adjacency matrix of a graph is defined as $\boldsymbol{A} = (\boldsymbol{A}_{ij}) \in \{0,1\}^{n \times n}$, where $\boldsymbol{A}_{ij} = 1 \Leftrightarrow (i,j) \in E$. Symmetric matrix with zeros on the diagonal.

Network Sequence Model

- We observe the entries $\boldsymbol{A}_{ij} \in \{0,1\}$, $1 \le j < i \le n$, of the **adjacency matrix** $\boldsymbol{A}$.
- $\boldsymbol{A}_{ij} = 1$ means that the nodes i and j are connected and $\boldsymbol{A}_{ij} = 0$ otherwise. Set $\boldsymbol{A}_{ii} = 0$ for all diagonal entries.
- $\boldsymbol{A}_{ij}$ are independent Bernoulli r. v. with **connection probabilities**

$$\boldsymbol{\Theta}_{ij} = \boldsymbol{P}(\boldsymbol{A}_{ij} = 1), \quad 1 \le j < i \le n.$$

- $\boldsymbol{\Theta}_0$ is a $n \times n$ symmetric matrix with entries $\boldsymbol{\Theta}_{ij}$ for $1 \le j < i \le n$ and zero diagonal entries.
- The model with such observations $\boldsymbol{A}_{ij}$, $1 \le j < i \le n$ is called the **network sequence model**.

Problem 1:

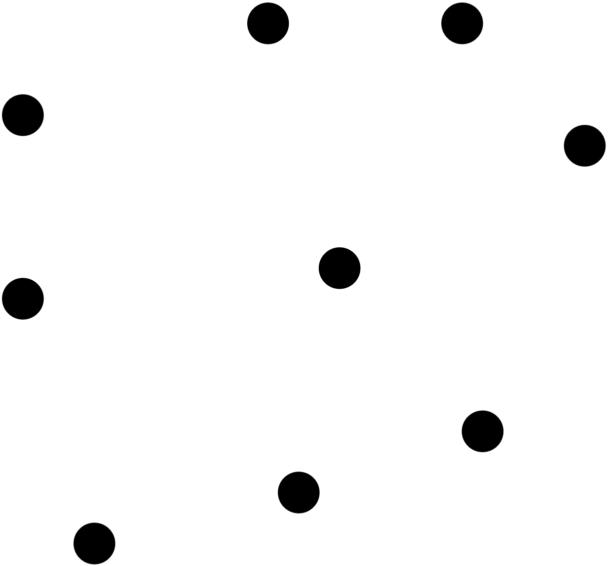
Estimate the matrix of connection probabilities $\boldsymbol{\Theta}_0$

Special case: Stochastic Block Model (SBM)

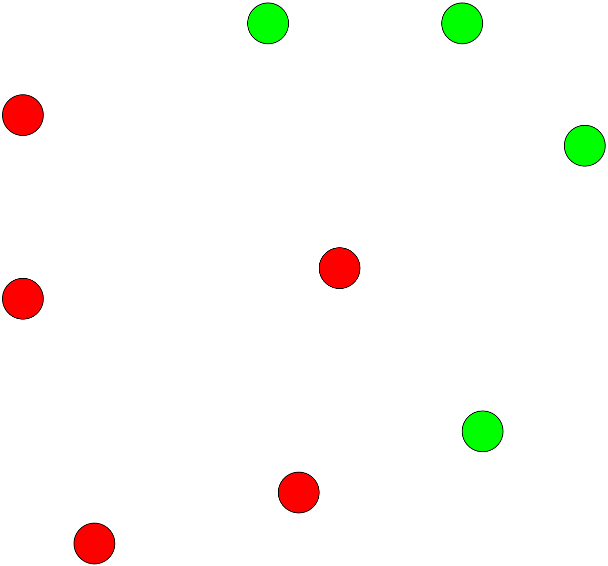
- Parameters:
  - Number of nodes n.
  - Partition of n nodes into k groups $C_1, \dots, C_k$ (communities).
  - Symmetric $k \times k$ matrix $\boldsymbol{Q}$ of inter-community edge probabilities.
- Any two nodes $u \in C_i$ and $v \in C_j$ are connected with probability $\boldsymbol{Q}_{ij}$.
- Degenerate case ($k = 1$): SBM $=$ Erdös-Rényi model.

SBM's are basic approximation units for more complex models.

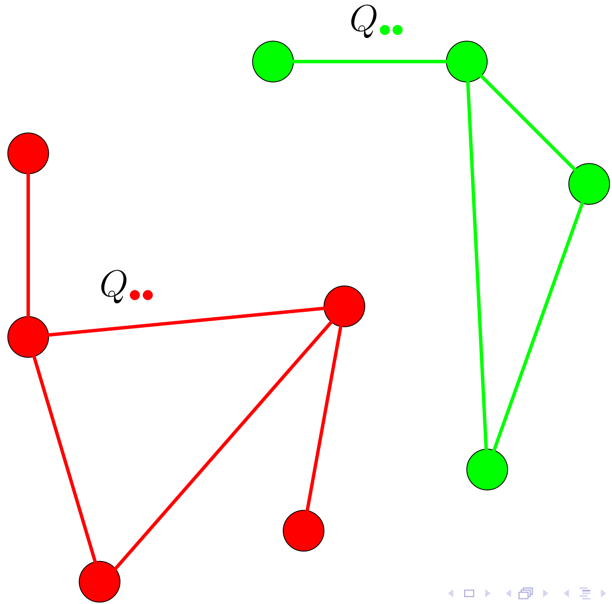
Example: SBM with $k = 2$



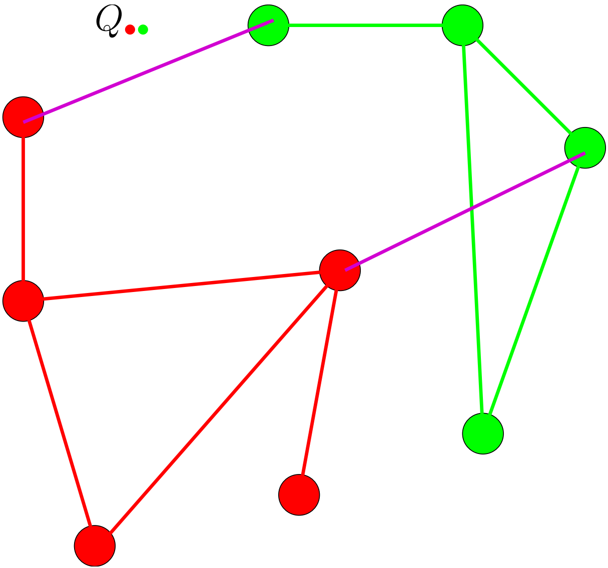
Example: SBM with $k = 2$



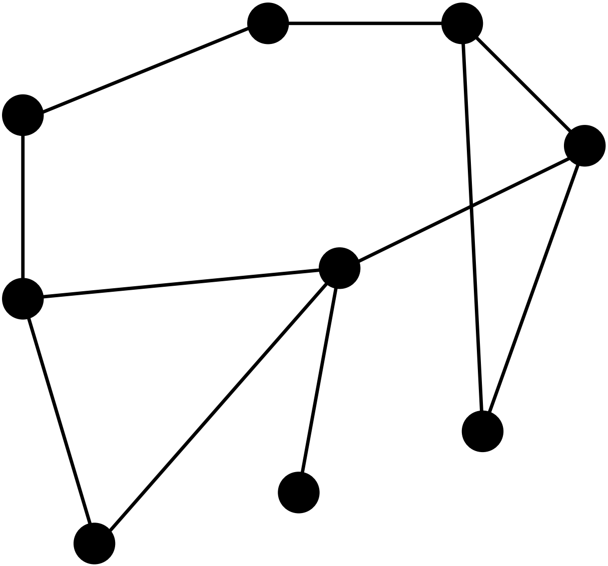
Example: SBM with $k = 2$



Example: SBM with $k = 2$



Example: SBM with $k = 2$



Stochastic Block Model (SBM)

- **Partition** of n nodes into k groups is represented as a mapping

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

- **Connection probabilities**

$$\boldsymbol{\Theta}_{ij} = \boldsymbol{Q}_{z(i)z(j)}, \quad j < i,$$

where $\boldsymbol{Q}$ is a symmetric $k \times k$ matrix of probabilities.

- Equivalent writing:

$$\boldsymbol{\Theta}_{ij} = U_i^T(z)\boldsymbol{Q}U_j(z)$$

where $U_i(z) = (\mathbb{I}_{\{z(i)=1\}}, \dots, \mathbb{I}_{\{z(i)=k\}})^T$ is a binary vector in $\mathbb{R}^k$ with one entry $=1$ and all other entries 0.

Stochastic Block Model (SBM)

- **Connection probabilities**

$$\boldsymbol{\Theta}_{ij} = \boldsymbol{Q}_{z(i)z(j)} = U_i^T(z)\boldsymbol{Q}U_j(z), \quad j < i,$$

where where $U_i(z) = (\mathbb{I}_{\{z(i)=1\}}, \ldots, \mathbb{I}_{\{z(i)=k\}})^T$ is a binary vector in $\mathbb{R}^k$ with one entry $=$1 and all other entries 0.

- **Matrix of connection probabilities** $\boldsymbol{\Theta}_0$ $(n \times n)$ is determined by two parameters: matrix $\boldsymbol{Q}$ $(k \times k)$ and node assignment z.

$$\boldsymbol{\Theta}_0 = U^T(z)\boldsymbol{Q}U(z)$$

with the convention that the diagonal elements are 0, where $U(z)$ is a matrix with columns $U_j(z)$.

Problem 1: Probability Matrix Estimation

Problem 1:

Estimate the matrix of connection probabilities $\mathbf{\Theta}_0$ in the **network sequence model** under the Frobenius loss.

- **Questions:**
  - Fundamental limits of estimation accuracy - minimax rates of convergence on suitable classes of matrices $\mathbf{\Theta}_0$.
  - Model misspecification: Oracle inequalities w.r.t. SBM oracle.
- Two cases regarding the elementwise sup-norm $\|\mathbf{\Theta}\|_\infty$.
  - **Dense graph**: $\|\mathbf{\Theta}\|_\infty$ is fixed.
  - **Sparse graph**: $\rho_n = \|\mathbf{\Theta}\|_\infty$ can be as small as possible.

  Graphs observed in practice are usually sparse graphs.
- Previous work: **Gao, Lu, Zhou (2014)**. Minimax rates for dense graphs.

Graphons

- Real-life networks are in permanent movement and often their size is growing. Thus, it make sense to consider graph limits as $n \to \infty$.
- "Limiting object" independent of the network size n called the **graphon**, introduced by **Lovász and Szegedy (2004)**. Random graph can be viewed as a partial observation of this limiting object.
- Graphons are symmetric measurable functions

$$W : [0,1]^2 \to [0,1].$$

- High level message: every graph limit can be represented by a graphon.

The Graphon Model

- $\xi_1, \dots, \xi_n$ are unobserved (latent) i.i.d. random variables uniformly distributed on $[0, 1]$.
- For $i \neq j$, set
$$\boldsymbol{\Theta}_{ij} = W(\xi_i, \xi_j),$$
and let the diagonal entries $\boldsymbol{\Theta}_{ii} = 0$.
- **Graphon Model**. Conditionally on $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$, the observations $\boldsymbol{A}_{ij}$ for $1 \leq j < i \leq n$ are independent Bernoulli random variables with success probabilities $\boldsymbol{\Theta}_{ij}$.

Remarks

1. Under this model, the observations $\boldsymbol{A}_{ij}$ are not independent.
2. The expected number of edges is $\sim n^2 \Longrightarrow$ **dense** graph.

Sparse Graphon

- Real-life networks usually **sparse** :
  - the number of edges is $o(n^2)$ as $n \to \infty$,
  - the degree of the graph tends to infinity as n grows.
- **Sparse Graphon Model:**

$$\mathbf{\Theta}_{ij} = \rho_n W(\xi_i, \xi_j) \quad , i < j,$$

  where $\rho_n > 0$ such that $\rho_n \to 0$ as $n \to \infty$.
- $\rho_n =$ "expected proportion of non-zero edges".
- Then
  - the number of edges is of the order $O(\rho_n n^2)$,
  - the average degree is of the order $\rho_n n$.
- Sparse graphons are considered by: **Bickel, Chen (2009), Bickel, Chen, Levina 2011),Wolfe, Olhede (2013), Xu et al. (2014)...**

Identifiability of Graphons

- Consider a measure-preserving bijection $\tau : [0, 1] \to [0, 1]$ (with respect to the Lebesgue measure).
- Take any graphon $W(\cdot, \cdot)$.
- The transformed graphon $W(\tau(\cdot), \tau(\cdot))$ induces the same probability measure on $\boldsymbol{A}$ as $W(\cdot, \cdot)$.
- Therefore, one should consider equivalence classes of graphons. Identifiability of graphons makes sense only on a suitably defined quotient space.

Loss function for graphon estimation

- Consider a sparse graphon

$$f(x, y) = \rho_n W(x, y)$$

and its estimator $\tilde{f}(x, y)$.

- The squared error is defined by

$$\delta^2(f, \tilde{f}) := \inf_{\tau \in \mathcal{M}} \int\int_{(0,1)^2} |f(\tau(x), \tau(y)) - \tilde{f}(x, y)|^2 \mathrm{d}x\mathrm{d}y$$

where $\mathcal{M}$ is the set of all measure-preserving bijections $\tau : [0, 1] \to [0, 1]$.

- $\delta(\cdot, \cdot)$ is a metric on the quotient space of graphons: **Lovász (2012).**

Problem 2: Graphon Estimation

Problem 2:

Estimate the sparse graphon function

$$f(x,y) = \rho_n W(x,y)$$

under the $\delta(\cdot,\cdot)$-loss.

- **Question:** Fundamental limits of estimation accuracy - minimax rates of convergence on two classes of graphons:
  - **Step function graphons**: analog of SBM.
  - **Smooth graphons**: W is a smooth function of two variables.
- Previous work: **Wolfe, Olhede (2013)**. Suboptimal upper bounds for smooth graphons.
- Parallel work: **Borgs,Chayes,Smith (2015)**. Upper bounds for step function graphons.

Problem 1: Probability Matrix Estimation

- **General Idea:** Fix an integer $k > 0$ and estimate the matrix $\mathbf{\Theta}_0$ by a block constant matrix with $k \times k$ blocks and such that block size is greater than some given integer n_0. Denote the set of such matrices by
$$\mathrm{SBM}(k, n_0).$$
- **(1) Least squares estimator:** The LS estimator $\widehat{\mathbf{\Theta}}$ of $\mathbf{\Theta}_0$ is a solution of

$$\min_{\mathbf{\Theta} \in \mathrm{SBM}(k,n_0)} \|\boldsymbol{A} - \mathbf{\Theta}\|_F^2.$$

By convention, for all estimators: zero diagonal entries.

Restricted least squares estimator

- **(2) Restricted least squares estimator:** The restricted LS estimator $\widehat{\mathbf{\Theta}}^r$ of $\mathbf{\Theta}_0$ is defined as a solution of

$$\min_{\mathbf{\Theta} \in \mathrm{SBM}(k, \mathbf{1}):\ \|\mathbf{\Theta}\|_\infty \leq \mathbf{r}} \|\boldsymbol{A} - \mathbf{\Theta}\|_F^2.$$

  Here, $r \in (0, 1]$ is a given constant, $\|\mathbf{\Theta}\|_\infty$ is the maximum of components norm.

- Note that for the restricted LS, we allow for any partitions including really unbalanced ones $(n_0 = 1)$.

Oracle inequality for the LS estimator

Theorem

For the network sequence model with $n_0 \geq 2$,

$$\mathbb{E}\left[\frac{1}{n^2}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\|_F^2\right] \leq C \min_{\boldsymbol{\Theta}\in\mathrm{SBM}(k,n_0)} \frac{1}{n^2}\|\boldsymbol{\Theta}_0 - \boldsymbol{\Theta}\|_F^2 + \Delta$$

where

$$\Delta = C\left(\|\boldsymbol{\Theta}_0\|_\infty + \frac{\log(n/n_0)}{n_0}\right)\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right).$$

Balanced partitions: $n_0 = n/k \implies$

$$\frac{\log(n/n_0)}{n_0} = \frac{k\log k}{n}$$

and the term containing n_0 is negligible for $\|\boldsymbol{\Theta}_0\|_\infty > \frac{k\log k}{n}$.

Oracle inequality for the LS estimator

Corollary (First oracle inequality)

For the network sequence model with $n_0 \geq 2$*, balanced partition and* $\|\mathbf{\Theta}_0\|_\infty > \frac{k \log k}{n}$,

$$\mathbb{E}\left[\frac{1}{n^2}\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0\|_F^2\right] \leq C_1 \min_{\mathbf{\Theta} \in \mathrm{SBM}(k, n_0)} \frac{1}{n^2}\|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_F^2 + \Delta$$

where

$$\Delta = C_2 \|\mathbf{\Theta}_0\|_\infty \left(\frac{\log k}{n} + \frac{k^2}{n^2}\right).$$

Oracle inequality for the Restricted LS estimator

Theorem (Second oracle inequality)

For the network sequence model with $\|\mathbf{\Theta}_0\|_\infty \le r$, *we have*

$$\mathbb{E}\left[\frac{\|\widehat{\mathbf{\Theta}}^r - \mathbf{\Theta}_0\|_F^2}{n^2}\right] \le \min_{\mathbf{\Theta}\in \mathrm{SBM}(k)} \frac{C_1\|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_F^2}{n^2} + C_2 r\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right)$$

where $\widehat{\mathbf{\Theta}}^r$ *is the restricted LS estimator.*

Probability matrix estimation: sparse SBM

- Given an integer k and any $\rho_n \in (0,1]$, consider the set of all probability matrices corresponding to k-class stochastic block model with connection probability uniformly smaller than ρ_n:

$$\mathcal{T}[k,\rho_n] = \{\boldsymbol{\Theta}_0 \in \mathrm{SBM}(k) : \ \|\boldsymbol{\Theta}_0\|_\infty \leq \rho_n\}.$$

Theorem (Minimax rate for sparse SBM)

For the network sequence model,

$$\inf_{\widehat{\boldsymbol{T}}} \sup_{\boldsymbol{\Theta}_0 \in \mathcal{T}[k,\rho_n]} \mathbb{E}_{\boldsymbol{\Theta}_0}\left[\frac{1}{n^2}\|\widehat{\boldsymbol{T}} - \boldsymbol{\Theta}_0\|_F^2\right] \asymp \min\left(\rho_n\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right), \rho_n^2\right)$$

where $\mathbb{E}_{\boldsymbol{\Theta}_0}$ denotes the expectation with respect to the distribution of $\boldsymbol{A}$ when the underlying probability matrix is $\boldsymbol{\Theta}_0$ and $\inf_{\widehat{\boldsymbol{T}}}$ is the infimum over all estimators.

Probability matrix estimation: SBM

- Previous work: **Gao, Lu, Zhou (2014)** cover the dense case $\rho_n = 1$. The minimax rate over $\mathcal{T}[k, 1]$ is

$$\frac{\log k}{n} + \frac{k^2}{n^2}.$$

Probability matrix estimation for smooth graphons

- Consider now the graphon model.
- **Smoothness:** Assume that the graphon W is in a Hölder ball with smoothness $\alpha > 0$.
- **Approximation of smooth graphon by SBM:** If $\mathbf{\Theta}_0$ is generated by a ρ_n-sparse graphon belonging to a Hölder ball with smoothness α, then
$$\frac{1}{n^2} \min_{\mathbf{\Theta} \in \mathrm{SBM}(k)} \|\mathbf{\Theta}_0 - \mathbf{\Theta}\|_F^2 \leq C \rho_n^2 \left(\frac{1}{k^2}\right)^{\alpha \wedge 1}.$$
- Plugging this into the oracle inequalities we get the bound
$$\rho_n^2 \left(\frac{1}{k^2}\right)^{\alpha \wedge 1} + \rho_n \Big(\frac{\log k}{n} + \frac{k^2}{n^2}\Big)$$
- Remains to minimize over k to achieve optimality.

Probability matrix estimation for smooth graphons

Denote by $\mathcal{W}(\alpha, \rho_n)$ the class of all sparse graphon models with W in a Hölder ball with smoothness $\alpha > 0$.

Theorem (Estimation rate for smooth sparse graphons)

Let $\rho_n \geq Cn^{-2+\epsilon}$ with an arbitrarily small $\epsilon > 0$. Then

$$\inf_{\widehat{\boldsymbol{T}}} \sup_{\boldsymbol{\Theta}_0 \in \mathcal{W}(\alpha,\rho_n)} \mathbb{E}_{\boldsymbol{\Theta}_0}\left[\frac{1}{n^2}\|\widehat{\boldsymbol{T}} - \boldsymbol{\Theta}_0\|_F^2\right] \asymp \underbrace{\rho_n^{\frac{2+\alpha\wedge 1}{1+\alpha\wedge 1}} n^{-\frac{2(\alpha\wedge 1)}{1+\alpha\wedge 1}}}_{\textbf{nonparametric}} + \underbrace{\frac{\rho_n \log n}{n}}_{\textbf{clustering}}.$$

- Two ingredients of the rates:
  nonparametric rate and **clustering rate** $\frac{\rho_n \log n}{n}$.
- The smoothness index α has an impact on the rate only if $\alpha \in (0, 1)$ and only if the network is not too sparse: $\rho_n \geq Cn^{\alpha-1}(\log n)^{1+\alpha}$.
- For $\alpha \geq 1$ the rate is $\dfrac{\rho_n \log n}{n}$.

From probability matrix estimation to graphon estimation

- To any $n \times n$ probability matrix $\mathbf{\Theta}$ we can associate a graphon.
- Given a $n \times n$ matrix $\mathbf{\Theta}$ with entries in $[0,1]$, define the **empirical graphon** $\widetilde{f}_{\mathbf{\Theta}}$ as the following piecewise constant function:

$$\widetilde{f}_{\mathbf{\Theta}}(x,y) = \mathbf{\Theta}_{\lceil nx \rceil, \lceil ny \rceil}$$

  for all x and y in $(0,1]$.
- This provides a way of deriving an estimator of the graphon function $f(\cdot,\cdot) = \rho_n W(\cdot,\cdot)$ from **any** estimator of the connection probability matrix.
- For example, the empirical graphons associated to the LS estimator and to the restricted LS estimator with threshold r :

$$\widehat{f} = \widetilde{f}_{\widehat{\mathbf{\Theta}}}\,, \qquad \widehat{f}_r = \widetilde{f}_{\widehat{\mathbf{\Theta}}^r}.$$

Agnostic error

- For any estimator $\check{f}$ of the graphon $f = \rho_n W$, we use the loss

$$\delta^2(\check{f}, f) := \inf_{\tau \in \mathcal{M}} \int\int_{(0,1)^2} |f(\tau(x), \tau(y)) - \check{f}(x, y)|^2 \mathrm{d}x\mathrm{d}y$$

  $\mathcal{M}$ is the set of all measure-preserving bijections τ.
- For any estimator $\widehat{\boldsymbol{T}} = \widehat{\boldsymbol{T}}(\boldsymbol{A})$ which is an $n \times n$ matrix with entries in $[0, 1]$:

$$\mathbb{E}\left[\delta^2(\widetilde{f}_{\widehat{\boldsymbol{T}}}, f)\right] \le 2\mathbb{E}\left[\frac{1}{n^2}\|\widehat{\boldsymbol{T}} - \boldsymbol{\Theta}_0\|_F^2\right] + 2\underbrace{\mathbb{E}\left[\delta^2\left(\widetilde{f}_{\boldsymbol{\Theta}_0}, f\right)\right]}_{\textbf{agnostic error}}$$

  (from the triangle inequality). Here, $\widetilde{f}_{\widehat{\boldsymbol{T}}}$ and $\widetilde{f}_{\boldsymbol{\Theta}_0}$ are empirical graphons; $\boldsymbol{\Theta}_0$ is the (**random**) probability matrix associated to W.
- Expectation over the distribution of **unobserved** $\xi_1, \ldots, \xi_n$.

Bounding the Agnostic Error

- **Step function graphons:** Let $\mathcal{W}[k]$ be the set of all k-step graphons, i.e., the subset of graphons W such that for some $k \times k$ symmetric matrix $\boldsymbol{Q}$ and some $\phi : [0,1] \to [k]$,

$$W(x,y) = \boldsymbol{Q}_{\phi(x),\phi(y)} \quad \text{for all } x, y \in [0,1] \ .$$

  Let $\boldsymbol{\Theta}_0$ be the probability matrix associated to W and $f = \rho_n W$. Then the agnostic error satisfies:

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

Bound for the δ-risk of step-function graphon

Theorem

Consider the ρ_n-sparse step-function graphon model: $W \in \mathcal{W}[k]$. If $\rho_n \le r$, the restricted LS empirical graphon estimator $\widehat{f}_r$ satisfies

$$\mathbb{E}\left[\delta^2\left(\widehat{f}_r, f\right)\right] \le C\left[r\left(\frac{k^2}{n^2} + \frac{\log(k)}{n}\right) + \rho_n^2\sqrt{\frac{k}{n}}\right].$$

Bound for the δ-risk of step-function graphon

- Assume $r \asymp \rho_n$. The achievable rate:

$$\min\left(\rho_n\left(\frac{k^2}{n^2}+\frac{\log(k)}{n}\right)+\rho_n^2\sqrt{\frac{k}{n}},\ \rho_n^2\right).$$

- Three zones:
  (i) **Weakly sparse graphs:** $\rho_n \geq \frac{\log(k)}{\sqrt{kn}} \vee \left(\frac{k}{n}\right)^{3/2}$. The bound is $\rho_n^2\sqrt{k/n}$ $\Longrightarrow$ the agnostic error dominates. **Optimal.**
  (ii) **Moderately sparse graphs:** $\frac{\log(k)}{n} \vee \left(\frac{k}{n}\right)^2 \leq \rho_n \leq \frac{\log(k)}{\sqrt{kn}} \vee \left(\frac{k}{n}\right)^{3/2}$. The bound is $\rho_n\left(\frac{k^2}{n^2}+\frac{\log(k)}{n}\right)$ $\Longrightarrow$ the probability matrix estimation error dominates. **Optimal up to logs:** Minimax lower bound $\rho_n\left(\frac{k^2}{n^2}+\frac{1}{n}\right)$.
  (iii) **Highly sparse graphs:** $\rho_n \leq \frac{\log(k)}{n} \vee \left(\frac{k}{n}\right)^2$. The bound of the theorem is suboptimal. The optimal rate ρ_n^2 is achieved by the estimator $\tilde{f} \equiv 0$.

Bound for the δ-risk of smooth graphon

Lemma

For the α-smooth ρ_n-sparse graphon model the agnostic error is bounded as

$$\mathbb{E}\left[\delta^2(\widetilde{f}_{\boldsymbol{\Theta}_0}, f)\right] \le C\frac{\rho_n^2}{n^{\alpha\wedge 1}}.$$

Theorem

Consider the α-smooth ρ_n-sparse graphon model. Assume that $r \ge \rho_n \ge Cn^{-2}$. Then the restricted LS empirical graphon estimator $\widehat{f}_r$ satisfies

$$\mathbb{E}\left[\delta^2\left(\widehat{f}_r, f_0\right)\right] \le C\left\{r^{\frac{2+\alpha\wedge 1}{1+\alpha\wedge 1}} n^{-\frac{2(\alpha\wedge 1)}{1+\alpha\wedge 1}} + \frac{r\log n}{n} + \frac{\rho_n^2}{n^{\alpha\wedge 1}}\right\}.$$

Sharp Oracle Inequality by Aggregation

- Let $\mathbf{\Theta}_{ij}$ be connection probabilities in a network. Assignment of n nodes to k classes: $z : \{1, ..., n\} \to \{1, ..., k\}$.
- **Stochastic block model (SBM)**:

$$\mathbf{\Theta}_{ij} = \boldsymbol{Q}_{z(i),z(j)} = U_i^T(z)\boldsymbol{Q}U_j(z)$$

  where $\boldsymbol{Q}$ is a symmetric $k \times k$ matrix of probabilities and $U_i(z) = (\mathbb{I}_{\{z(i)=1\}}, \dots, \mathbb{I}_{\{z(i)=k\}})^T$.
- **Our aim**: when $\mathbf{\Theta}_{ij}$ are **arbitrary** and we observe

$$Y_{ij} = \mathbf{\Theta}_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \ iid,$$

  find an estimator that is as close as possible to the best approximation of $\mathbf{\Theta}_{ij}$ by a SBM.
  $\Longrightarrow$ Oracle inequality with SBM oracle.

Sharp Oracle Inequality by Aggregation

- We have $M = k^n$ possible assignments z: $z^1, ..., z^M$.
- To each fixed z we associate the LS estimator: averaging of Y_{ij} over the blocks defined by this assignment.
- Thus, we get linear estimators $\hat{\mathbf{\Theta}}^1, ..., \hat{\mathbf{\Theta}}^M$ of matrix $\mathbf{\Theta}$.
- For each $\hat{\mathbf{\Theta}}^m$, bias-variance decomposition yields

$$\mathbb{E}\|\hat{\mathbf{\Theta}}^m - \mathbf{\Theta}_0\|_F^2 \leq \min_{\boldsymbol{Q}} \|U^T(z^m)\boldsymbol{Q}U(z^m) - \mathbf{\Theta}_0\|_F^2 + \sigma^2 k^2$$

  where $\|\cdot\|_F$ is the Frobenius norm.
- We aggregate these estimators using the Exponential Weighting procedure. The aggregate $\hat{\mathbf{\Theta}}^{EW}$ satisfies

$$\mathbb{E}\frac{1}{n^2}\|\hat{\mathbf{\Theta}}^{EW} - \mathbf{\Theta}_0\|_F^2 \leq \min_m \mathbb{E}\frac{1}{n^2}\|\hat{\mathbf{\Theta}}^m - \mathbf{\Theta}_0\|_F^2 + 4\sigma^2 \frac{\log M}{n^2}.$$

Sharp Oracle Inequality by Aggregation

- Combining the two inequalities and $M = k^n$ we get

$$
\begin{aligned}
\mathbb{E}\frac{1}{n^2}\|\hat{\mathbf{\Theta}}^{EW} - \mathbf{\Theta}_0\|_F^2 &\le \underbrace{\min_{\boldsymbol{Q},z}\frac{1}{n^2}\|U^T(z)\boldsymbol{Q}U(z) - \mathbf{\Theta}_0\|_F^2}_{\textbf{oracle error}} \\
&\quad + \sigma^2\left(\frac{k^2}{n^2} + \frac{4\log M}{n^2}\right) \\
&\le \textbf{oracle error} + C\left(\frac{k^2}{n^2} + \frac{\log k}{n}\right).
\end{aligned}
$$

- The same oracle error as before:

$$
\min_{\boldsymbol{Q},z}\frac{1}{n^2}\|U^T(z)\boldsymbol{Q}U(z) - \mathbf{\Theta}_0\|_F^2 = \min_{\mathbf{\Theta}\in \mathrm{SBM}(k)}\|\mathbf{\Theta} - \mathbf{\Theta}_0\|_F^2.
$$

- Inequality with leading constant 1.
- The rate $\frac{k^2}{n^2} + \frac{\log k}{n}$ is minimax optimal for SBM.