Link prediction in graphs

Q. Duchemin

Université Paris Est Marne La Vallée

October 2020

Goals of link prediction

- Understand, through a generative model, why different vertices are connected or not.
- Generalise these observations to the rest of the graph.

Motivations

- In social networks [1]
  Shared interests, differences in artistic tastes or political opinion.
- In biological networks [2]
  Interactions between molecules or protein.

---

1. Wasserman, Faust et al., *Social network analysis: Methods and applications*.
2. Madeira et Oliveira, "Biclustering algorithms for biological data analysis: a survey".

A large span of frameworks for link prediction

Supervised / Unsupervised ?

Temporal aspect ?

- Yes : Finding missing links.
- No :
  - Links can be created and destroyed over time.
  - New nodes are entering the graph at each time step. [3]

Topological-based link prediction ? Do we have additional features on the nodes ? [4]

Parametric or Non-parametric model ?

Global or Local [5] method ?

Probabilistic / Geometric model ?

---

3. Dunlavy, Kolda et Acar, "Temporal link prediction using matrix and tensor factorizations".
4. Wang, Satuluri et Parthasarathy, "Local probabilistic models for link prediction".
5. Liben-Nowell et Kleinberg, "The link-prediction problem for social networks".

In this presentation, we will mainly be focused on

Baldin et Berthet, “Optimal link prediction with matrix logistic regression”

which proposed a method which is

- Not temporal
- Supervised
- Global method
- Probabilistic and Parametric

Motivation Adapt usual high-dimensional methods to a model with two covariates (explanatory variables).

Beyond link prediction, this paper allows to study

- Information-Computational gaps.
- General method to establish computational lower bounds.
- Classical statistical and optimization tools : Establishing minimax convergence rate and convex relaxation.

Section 1

Model and Assumptions

We consider a graph $G = ([n], E)$ with adjacency matrix $Y \in \{0,1\}^{n \times n}$ generated from the following **generative model**.

- An explanatory variable $X_i \in \mathbb{R}^d$ is associated to each node $i \in [n]$.
- For some $\Theta_* \in \mathbb{S}^d$,

$$\forall i \in [n],\; Y_{i,i} = 0 \text{ and } \forall i,j \in [n]^2,\; i \neq j,\; Y_{i,j} \sim \mathcal{B}(\pi_{i,j}(\Theta_*)),$$

  where

$$\begin{aligned}\pi_{i,j} : \mathbb{S}^d &\to [0,1] \\ \Theta &\mapsto \mathbb{P}((i,j) \in E) = \sigma(X_i^\top \Theta_* X_j) = \left(1 + \exp(-X_i^\top \Theta_* X_j)\right)^{-1}\end{aligned}$$

Observations

- For all $(i,j) \in \Omega$, $Y_{i,j}$ is observed.
- All the explanatory variables $(X_i)_i$ are known.

Comparaison with other models

Reformulation as a classical logistic regression problem using

$$X_i^\top \Theta_* X_j = \mathrm{Tr}(X_j X_i^\top \Theta_*) = \langle \mathrm{vec}(X_j X_i^\top), \mathrm{vec}(\Theta_*) \rangle.$$

- **Generalised linear model.**
- **Graphon model** with known explanatory variables.
- **Trace regression models.**
  The model is $Y = \mathrm{Tr}(\Theta_*^\top Z) + \epsilon$ with $Z \in \mathbb{R}^{d_1 \times d_2}$ is a matrix of explanatory variables, $\Theta_* \in \mathbb{R}^{d_1 \times d_2}$ is the matrix of regression coefficients, $Y \in \mathbb{R}$ is the response and $\epsilon \in \mathbb{R}$ is the noise.
- **Metric learning.**
  Observations depend on an unknown geometric representation $V_1, \ldots, V_n$ of the variables in a Euclidean space of low dimension. Based on noisy observations of $\langle V_i, V_j \rangle$, we want to recover $(V_i)_i$.
  Taking $X_i = e_i$ and $\Theta_* = V^\top V$ gives $\langle V_i, V_j \rangle = e_i^\top V^\top V e_j = X_i^\top \Theta_* X_j$.

Coping with the curse of dimensionality

High-dimensional setting $d^2 \gg N = |\Omega|$.

Motivation of the structural assumptions

$$\Theta_* = \sum_{l=1}^{R} \lambda_l u_l u_l^\top$$

The **affinity** $\Sigma_{i,j} := X_i^\top \Theta_* X_j$ between vertices i and j is therefore only a function of the projections of X_i and X_j along the axes u_l i.e

$$\Sigma_{i,j} = \sum_{l=1}^{R} \lambda_l (u_l^\top X_i)(u_l^\top X_j).$$

Prior	Assumption
Only a few of the directions u_l have non-zero impact on the affinity	Θ_* is low-rank
Only few relevant coefficients of X_i and X_j influence the affinity	Sparsity on the u_l $\Leftrightarrow$ Block-sparsity on Θ_*

Block-sparse matrix logistic regression

Notations For any $p, q \in [0, \infty)$ and any $B \in \mathbb{S}^d$,

$$\|B\|_{p,q} = \left\|(\|B_{1,*}\|_p, \ldots, \|B_{d,*}\|_p)\right\|_q ,$$

where $B_{i,*}$ is the i-th row of B.

$$\forall B \in \mathbb{S}^n, \quad \|B\|_{F,\Omega}^2 := \sum_{(i,j)\in\Omega} B_{i,j}^2.$$

For any $k, r \in [d]$ (with $r \leq k$),

$$\mathcal{P}_{k,r}(M) = \left\{\Theta \in \mathbb{S}^d \ : \ \|\Theta\|_{1,1} < M, \ \|\Theta\|_{0,0} \leq k, \ \text{and } \mathrm{rank}(\Theta) \leq r\right\}$$

$\|\cdot\|_{1,1}$ is the element wise l^1 norm on $\mathbb{S}^d$.
$\|\cdot\|_{0,0}$ counts the number of selected variables.

Recovering Θ_* from affinities

Block isometry property

For a matrix $\mathbb{X} \in \mathbb{R}^{d \times n}$ and an integer $s \in [d]$, we define $\Delta_{\Omega,s}(\mathbb{X}) \in (0,1)$ as the smallest positive real such that

$$N(1 - \Delta_{\Omega,s}(\mathbb{X}))\|B\|_F^2 \le \|\mathbb{X}^\top B \mathbb{X}\|_{F,\Omega}^2 \le N(1 + \Delta_{\Omega,s}(\mathbb{X}))\|B\|_F^2,$$

for all matrices $B \in \mathbb{S}^d$ that satisfy the block-sparsity assumption $\|B\|_{0,0} \le s$.

The Block isometry property guarantees that the matrix Θ_* can be recovered from observations of the affinities $\Sigma_{i,j}$.

Block VS Restricted isometry property

Restricted isometry property

For a matrix $A \in \mathbb{R}^{n\times p}$ and an integer $s \in [p]$, $\delta_s(A) \in (0,1)$ is the smallest positive real such that

$$n(1-\delta_s(A))\|v\|_2^2 \le \|Av\|_2^2 \le n(1+\delta_s(A))\|v\|_2^2,$$

for all s-sparse vectors, i.e. satisfying $\|v\|_0 \le s$.
When $p = d^2$, we define $\delta_{\mathcal{B},s}(A)$ as the smallest positive real such that

$$n(1-\delta_{\mathcal{B},s}(A))\|v\|_2^2 \le \|Av\|_2^2 \le n(1+\delta_{\mathcal{B},s}(A))\|v\|_2^2,$$

for all vectors such that $v = \mathrm{vec}(B)$, where B satisfies the block-sparsity assumption $\|B\|_{0,0} \le s$.

For a matrix $\mathbb{X} \in \mathbb{R}^{d\times n}$, let $\mathbb{D}_\Omega \in \mathbb{R}^{N\times d^2}$ be defined row-wise by $\mathbb{D}_\Omega(i,j) = \mathrm{vec}(X_j X_i^\top)$ for all $(i,j) \in \Omega$. It holds that

$$\Delta_{\Omega,s}(\mathbb{X}) = \delta_{\mathcal{B},s}(\mathbb{D}_\Omega).$$

Recovering the affinities from the $\pi_{i,j}$

We do not directly observe the $\Sigma_{i,j}$, but their image through σ. A condition is necessary to ensure that the affinities can be recovered from the observed edges.

Identifiability Condition (IC)

There exists a constant $M > 0$ such that for all Θ in the class $\mathcal{P}_{d,d}(M)$ we have $\max_{(i,j)\in\Omega} |X_i^\top \Theta X_j| < M$.

Under (IC),

$$\forall (i,j) \in \Omega, \qquad \inf_{\Theta \in \mathcal{P}_{d,d}(M)} \sigma'(X_i^\top \Theta X_j) \geq \mathcal{L}(M) > 0,$$

where $\mathcal{L}(M) := \sigma'(M) = \sigma(M)(1 - \sigma(M))$.

Preliminaries

Log-likelihood

$$l_Y(\Theta) = -\sum_{(i,j)\in\Omega} \log\left(1 + e^{(2Y_{i,j}-1)X_i^\top \Theta X_j}\right)$$

$-l_Y$ is a convex function of Θ.

Stochastic component of the likelihood

Denoting $l : \Theta \mapsto \mathbb{E}_{\Theta_*}\left[l_Y(\Theta)\right]$, it holds

$$\begin{aligned} l_Y(\Theta) - l(\Theta) &= \sum_{(i,j)\in\Omega} (Y_{i,j} - \pi_{i,j}(\Theta_*))X_i^\top \Theta X_j \\ &= \langle\!\langle \mathcal{E}_\Omega, \mathbb{X}^\top \Theta \mathbb{X} \rangle\!\rangle, \end{aligned}$$

where $\mathcal{E}_\Omega = (Y_{i,j} - \pi_{i,j}(\Theta_*))_{(i,j)\in\Omega}$ with zeros on the complement Ω^c.

$$\begin{aligned} l(\Theta) - l(\Theta_*) &= -\sum_{(i,j)\in\Omega} \mathrm{KL}\left(\pi_{i,j}(\Theta_*), \pi_{i,j}(\Theta)\right) \\ &= -\mathrm{KL}\left(\mathbb{P}_{\Theta_*}, \mathbb{P}_{\Theta}\right). \end{aligned}$$

Section 2

Penalised logistic loss

Penalised logistic loss

$$\hat{\Theta} \in \underset{\Theta \in \mathcal{P}_{d,d}(M)}{\arg\min} \{-l_Y(\Theta) + p(\Theta)\}$$

with a penalty p defined by

$$p(\Theta) = g\left(\mathrm{rank}(\Theta), \|\Theta\|_{0,0}\right) \text{ with } g(R, K) = cKR + cK\log\left(\frac{de}{K}\right),$$

where $c > 0$ is a universal constant and to be specified further.

Non-asymptotic upper bound

Assume the design matrix $\mathbb{X}$ satisfies $\max_{(i,j)\in\Omega}|X_i^\top \Theta_* X_j| < M$ for some $M > 0$ and all Θ_* in a given class. Then

$$\sup_{\Theta_* \in \mathcal{P}_{k,r}(M)} \frac{1}{N}\mathbb{E}\left[\mathrm{KL}(\mathbb{P}_{\Theta_*}, \mathbb{P}_{\hat{\Theta}})\right] \le C_1\left\{\frac{kr}{N} + \frac{k}{N}\log(\frac{de}{k})\right\},$$

where $C_1 > 3c$ is some universal constant for all $k = 1, \dots, d$ and $r = 1, \dots, k$.

Proof (1/2)

Let us recall that $l(\Theta) = \mathbb{E}_{\Theta_*}\left[l_Y(\Theta)\right]$ and that

$$l(\Theta_*) - l(\Theta) = \sum_{(i,j)\in\Omega} \mathrm{KL}\left(\pi_{i,j}(\Theta_*), \pi_{i,j}(\Theta)\right).$$

It suffices to show

$$\sup_{\Theta_*\in\mathcal{P}_{k,r}(M)} \mathbb{P}_{\Theta_*}\left(\underbrace{l(\Theta_*) - l(\hat{\Theta}) + p(\hat{\Theta})}_{:=\tau^2(\hat{\Theta},\Theta_*)} > 2p(\Theta_*) + R_t^2\right) \leq e^{-cR_t}, \quad (*)$$

for any $R_t > 0$ and some numeric constant $c > 0$. Indeed, then taking $R_t^2 = p(\Theta_*)$, it follows that $l(\Theta_*) - l(\hat{\Theta}) \leq 3p(\Theta_*)$ uniformly for all Θ_* in the considered class with probability at least $1 - e^{-c\sqrt{p(\Theta_*)}}$.

- On $\{\tau^2(\hat{\Theta},\Theta_*) \leq 2p(\Theta_*)\}$, $(*)$ clearly holds.
- On $\{\tau^2(\hat{\Theta},\Theta_*) > 2p(\Theta_*)\}$

$$\langle\!\langle \mathcal{E}_\Omega, \mathbb{X}^\top(\hat{\Theta} - \Theta_*)\mathbb{X}\rangle\!\rangle \geq l(\Theta_*) - l(\hat{\Theta}) + p(\hat{\Theta}) - p(\Theta_*) \geq \frac{1}{2}\tau^2(\hat{\Theta},\Theta_*).$$

Proof (2/2)

Therefore, for any $\Theta_* \in \mathcal{P}_{k,r}(M)$, we have

$$\mathbb{P}_{\Theta_*}\left(\tau^2(\hat{\Theta},\Theta_*) > 2p(\Theta_*) + R_t^2\right) \leq \mathbb{P}_{\Theta_*}\left(\sup_{\tau(\Theta,\Theta_*)\geq R_t} \frac{\langle\langle \mathcal{E}_\Omega, \mathbb{X}^\top(\Theta-\Theta_*)\mathbb{X}\rangle\rangle}{\tau^2(\Theta,\Theta_*)} \geq \frac{1}{2}\right).$$

We apply the **peeling device** : we slice the set $\tau(\Theta,\Theta_*) \geq R_t$ into pieces on which the penalty term $p(\Theta)$ is fixed and the term $l(\Theta_*) - l(\Theta)$ is bounded.

$$\begin{aligned}
&\mathbb{P}_{\Theta_*}\left(\sup_{\tau(\Theta,\Theta_*)\geq R_t} \frac{\langle\langle \mathcal{E}_\Omega, \mathbb{X}^\top(\Theta-\Theta_*)\mathbb{X}\rangle\rangle}{\tau^2(\Theta,\Theta_*)} \geq \frac{1}{2}\right) \\
&\leq \sum_{K=1}^{d}\sum_{R=1}^{K}\sum_{s=1}^{\infty} \mathbb{P}_{\Theta_*}\left(\sup_{\substack{\Theta\,:\,R_t\leq\tau(\Theta,\Theta_*)\leq 2^s R_t \\ \|\Theta\|_{\mathbf{0},\mathbf{0}}=k,\mathrm{rank}(\Theta)=R}} \langle\langle \mathcal{E}_\Omega, \mathbb{X}^\top(\Theta-\Theta_*)\mathbb{X}\rangle\rangle \geq \frac{1}{8}2^{2s}R_t^2\right).
\end{aligned}$$

To end the proof, we apply

- Bousquet's version of Talagrand's inequality.
- Dudley's entropy integral bound.

Going from KL to Σ and Θ.

Let us recall that $l(\Theta) = \mathbb{E}_{\Theta_*}\left[l_Y(\Theta)\right]$ and that

$$l(\Theta_*) - l(\Theta) = \sum_{(i,j)\in\Omega} \mathrm{KL}\left(\pi_{i,j}(\Theta_*), \pi_{i,j}(\Theta)\right).$$

Going from KL to Σ.

Using $\nabla l(\Theta_*) = 0$, it holds using Taylor expansion,

$$\begin{aligned} l(\Theta_*) - l(\hat{\Theta}) &= \frac{1}{2} \sum_{(i,j)\in\Omega} \sigma'\left(X_i^\top \Theta_0 X_j\right) \langle\!\langle X_j X_i^\top, \Theta_* - \hat{\Theta} \rangle\!\rangle^2 \\ &\geq \frac{\mathcal{L}}{2} \sum_{(i,j)\in\Omega} \langle\!\langle X_j X_i^\top, \Theta_* - \hat{\Theta} \rangle\!\rangle^2 \\ &= \frac{\mathcal{L}}{2} \|\mathbb{X}^\top \left(\Theta_* - \hat{\Theta}\right) \mathbb{X}\|_{F,\Omega}^2, \end{aligned}$$

where $\Theta_0 \in [\hat{\Theta}, \Theta_*]$ element-wise.

Going from KL to Θ.

$$\frac{\mathcal{L}}{2} N \left(1 - \Delta_{\Omega,d}\right) \|\Theta_* - \hat{\Theta}\|_F^2 \leq \frac{\mathcal{L}}{2} \|\mathbb{X}^\top \left(\Theta_* - \hat{\Theta}\right) \mathbb{X}\|_{F,\Omega}^2 \leq l(\Theta_*) - l(\hat{\Theta}).$$

Prediction

Measure the prediction error of an estimator $\hat{\Theta}$ by

$$\mathbb{E}\Bigg[\sum_{(i,j)\in\Omega}\big(\pi_{i,j}(\hat{\Theta})-\pi_{i,j}(\Theta_*)\big)^2\Bigg],$$

which is controlled according to the following result using the smoothness of the logistic function σ.

Solving link prediction

Under (IC),

$$\sup_{\Theta_*\in\mathcal{P}_{k,r}(M)}\frac{1}{2N}\mathbb{E}\left[\|\Sigma_*-\hat{\Sigma}\|_{F,\Omega}\right]\le\frac{C_1}{\mathcal{L}(M)}\left(\frac{kr}{N}+\frac{k}{N}\log\left(\frac{de}{k}\right)\right),$$

with $C_1>0$.

Section 3

Performance of the penalised MLE

Low-rank and block-sparse MLE

The rank-constrained maximum likelihood estimators with bounded block size is

$$\hat{\Theta}_{k,r} \in \arg\min_{\Theta \in \mathcal{P}_{k,r}} \{-l_Y(\Theta)\}.$$

Non-asymptotic upper bound on the rate of estimation

Assume the design matrix $\mathbb{X}$ satisfies the block isometry property and $\max_{(i,j)\in\Omega} |X_i^\top \Theta_* X_j| < M$ for some $M > 0$ and all Θ_* in a given class. Then for the maximum likelihood estimator $\hat{\Theta}_{k,r}$,

$$\sup_{\Theta_* \in \mathcal{P}_{k,r}(M)} \mathbb{E}\left[\|\hat{\Theta}_{k,r} - \Theta_*\|_F^2\right] \leq \frac{C_2}{\mathcal{L}(M)(1 - \Delta_{\Omega,2k}(\mathbb{X}))} \left\{\frac{kr}{N} + \frac{k}{N}\log\left(\frac{de}{k}\right)\right\},$$

for all $k = 1, \ldots, d$ and $r = 1, \ldots, k$ and some constant $C_2 > 0$.

Lower bounds

Minimax lower bound

Let the design matrix $\mathbb{X}$ satisfy the block isometry property. Then for estimating $\Theta_* \in \mathcal{P}_{k,r}(M)$ in the matrix logistic regression model, the following lower bound on the rate of estimation holds

$$\inf_{\hat{\Theta}} \sup_{\Theta_* \in \mathcal{P}_{k,r}(M)} \mathbb{E}\left[\|\hat{\Theta} - \Theta_*\|_F^2\right] \geq \frac{C_3}{(1 + \Delta_{\Omega,2k}(\mathbb{X}))}\left(\frac{kr}{N} + \frac{k}{N}\log(\frac{de}{k})\right),$$

where the constant $C_3 > 0$ is independent of d, k, r and the infimum extends overall estimators $\hat{\Theta}$.

The penalised maximum likelihood approach attains the minimax rate of estimation over simultaneously block-sparse and low-rank matrices.

Sparse matrix logistic regression

For any $k, r \in [d]$,

$$\hat{\Theta}_{Lasso} \in \arg\min_{\Theta \in \mathbb{S}^d} \{ -l_Y(\Theta) + \lambda \|\Theta\|_{1,1} \},$$

with $\lambda > 0$ to be chosen further, which is equivalent to the logistic Lasso on $\mathrm{vec}(\Theta)$.

Theorem

Assume the design matrix $\mathbb{X}$ satisfies the block isometry property and $\max_{(i,j)\in\Omega} |X_i^\top \Theta X_j| < M$ for some $M > 0$ and all Θ_* in a given class. Then for $\lambda = C_4 \sqrt{\log d}$, where $C_4 > 0$ is an appropriate universal constant,

$$\sup_{\Theta_* \in \mathcal{P}_{k,r}(M)} \mathbb{E}\left[\|\hat{\Theta}_{Lasso} - \Theta_*\|_F^2 \right] \le \frac{C_5}{\mathcal{L}(M)(1 - \Delta_{\Omega,2k}(\mathbb{X}))} \frac{k^2}{N} \log d,$$

for all $k = 1, \ldots, d$ and $r = 1, \ldots, k$ and some universal constant $C_5 > 0$.

Section 4

Computational lower bounds

The Planted Clique problem

Computational lower bound It is the fastest rate of estimation attained by a (randomised) polynomial-time algorithm in the worst-case scenario.

Idea : Detecting a subspace of $\mathcal{P}_{k,r}$ can be computationally as hard as solving the dense subgraph detection problem.

The Planted Clique problem

- $G(n, 1/2)$ is the distribution of Erdos Renyi graphs
- $G(n, 1/2, k, q)$ is the distribution of graphs constructed by
  - first picking k vertices independently at random.
  - connecting all edges in-between with probability $q \in (1/2, 1]$.
  - then joining each remaining pair of distinct vertices by an edge independently at random with probability 1/2.

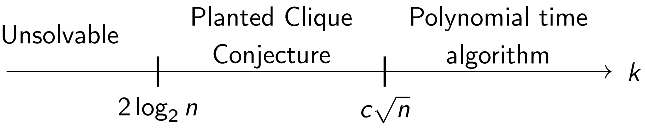
Planted clique problem refers to the hypothesis testing problem of

$$\mathbb{H}_0 : A \sim G(n, 1/2) \quad \mathbb{H}_1 : A \sim G(n, 1/2, k, 1).$$

based on observing a random graph A drawn from either $G(n, 1/2)$ or $G(n, 1/2, k, 1)$.

Dense subgraph detection

Planted Clique problem



Extension to the dense subgraph detection problem

$$\mathbb{H}_0 : A \sim G(n, 1/2) \text{ vs } \mathbb{H}_1 : A \sim G(n, 1/2, k, q), \quad q \in (1/2, 1].$$

The dense subgraph detection conjecture (DSD Conjecture)

For any sequence $k = k_n$ such that $k \le n^{\beta}$ for some $0 < \beta < 1/2$, and any $q \in (1/2, 1]$, there is no (randomised) polynomial-time algorithm that can correctly identify the dense subgraph with probability tending to 1 as $n \to \infty$, i.e. for any sequence of (randomised) polynomial-time tests $(\psi_n : \mathbb{G}_n \to \{0, 1\})_n$, we have

$$\liminf_{n\to\infty} \{\mathbb{P}_0(\psi_n(A) = 1) + \mathbb{P}_1(\psi_n(A) = 0)\} \ge 1/3.$$

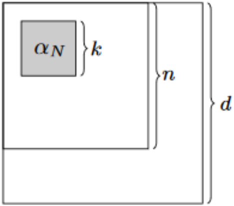
Reduction to the dense subgraph detection problem

Sampling from $G(n, 1/2) \Leftrightarrow \Theta_0 = 0 \in \mathbb{R}^{d \times d}$

Sampling from $G(n, 1/2, k, q)$ **?**

We consider

- $N = |\Omega| = \binom{n}{2}$.
- $\forall i \in [n]$, $X_i = N^{1/4} e_i \in \mathbb{R}^d$.
- $\alpha_N = \frac{\alpha}{\sqrt{N}}$ for some $\alpha > 0$.

$\mathcal{G}_k^{\alpha_N} \ni \Theta =$ 

$$\mathcal{G}_k^{\alpha_N} := \left\{ \Theta \in \mathcal{P}_{k,1}(M) \;:\; \exists S \in \mathcal{S}_k([n]) \text{ s.t. } \Theta_{i,j} = \begin{cases} \alpha_N & \text{if } i,j \in S \\ 0 & \text{otherwise.} \end{cases} \right\}.$$

Then for any $\Theta \in \mathcal{G}_k^{\alpha_N}$,

$$\mathbb{P}\left((i,j) \in E | X_i, X_j\right) = \left(1 + e^{-X_i^\top \Theta X_j}\right)^{-1} = \begin{cases} \left(1 + e^{-\alpha}\right)^{-1} & \text{if } \Theta_{i,j} = \alpha_N \\ 1/2 & \text{otherwise.} \end{cases}.$$

The testing problem

$$\mathbb{H}_0 : Y \sim \mathbb{P}_{\Theta_0} \quad \text{vs} \quad \mathbb{H}_1 : Y \sim \mathbb{P}_{\Theta}, \Theta \in \mathcal{G}_k^{\alpha_N}.$$

$$\Leftrightarrow$$

The dense subgraph detection problem with $q = (1 + e^{-\alpha})^{-1}$.

Computational lower bound of order k^2/N

Let $\mathcal{F}_k$ be any class of matrices containing $\mathcal{G}_k^{\alpha_N} \cup \{\Theta_0\}$. Let $c > 0$ and $f(k, d, N)$ with $f(k, d, N) \leq ck^2/N$ for $k = k_n < n^{\beta}, 0 < \beta < 1/2$ and a sequence $d = d_n$, for all $n > m_0 \in \mathbb{N}$.

If (DSD Conjecture) holds, for some design $\mathbb{X}$ that fulfils the block isometry property, for any estimator $\hat{\Theta}$, computable in polynomial time, there exists a sequence $(k, d, N) = (k_n, d_n, N)$, such that

$$\frac{1}{f(k, d, N)} \sup_{\Theta_* \in \mathcal{F}_k} \mathbb{E}\left[\|\hat{\Theta} - \Theta_*\|_F^2\right] \to +\infty.$$

Let's sketch the proof !

Proof of the computational lower bound (1/2)

We here provide a proof of the computational lower bound on the prediction error. Assume that there exists a hypothetical estimator $\hat{\Theta}$ computable in polynomial time such that

$$\limsup_{n\to\infty} \frac{1}{f(k,d,N)} \sup_{\Theta_* \in \mathcal{F}_k} \frac{1}{N} \mathbb{E}\left[\|\mathbb{X}^\top(\hat{\Theta} - \Theta_*)\mathbb{X}\|_{F,\Omega}^2\right] \leq b < \infty,$$

for all sequences $(k,d,N) = (k_n, d_n, N)$ and a constant b. Then by Markov's inequality, we have

$$\frac{1}{N}\|\mathbb{X}^\top(\hat{\Theta} - \Theta_*)\mathbb{X}\|_{F,W}^2 \leq uf(k,d,N),$$

for some numeric constant $u > 0$ with probability $1 - b/u$ for all $\Theta_* \in \mathcal{F}_k$. Following the reduction scheme, we consider the design vectors $X_i = N^{1/4} e_i$, $i = 1, \ldots, n$ and the subset of edges Ω, such that

$$\frac{1}{N}\|\mathbb{X}^\top(\hat{\Theta} - \Theta_*)\mathbb{X}\|_{F,\Omega}^2 = \sum_{(i,j)\in\Omega} (\hat{\Theta}_{i,j} - (\Theta_*)_{i,j})^2 = \|\hat{\Theta} - \Theta_*\|_{F,\Omega}^2,$$

for any $\Theta_* \in \mathcal{G}_{\alpha_N k}$.

Proof of the computational lower bound (2/2)

Thus, in order to separate the hypotheses

$$\mathbb{H}_0 : Y \sim \mathbb{P}_0 \text{ vs. } \mathbb{H}_1 : Y \sim \mathbb{P}_\Theta,\ \Theta \in \mathcal{G}_k^{\alpha_N},$$

it is natural to employ the following test

$$\psi(Y) = \mathbb{1}\left\{\|\hat{\Theta}\|_{F,\Omega} \geq \tau_{d,k}(u)\right\},$$

where $\tau_{d,k}^2(u) = uf(k,d,N)$. The type I error of this test is controlled automatically : $\mathbb{P}_0(\psi = 1) \leq b/u$. For the type II error,

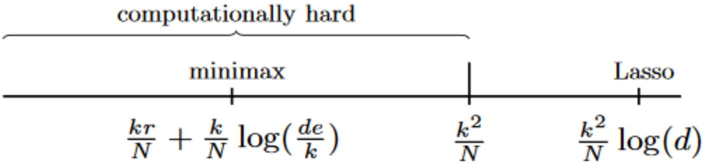
$$\begin{aligned}\sup_{\Theta \in \mathcal{G}_k^{\alpha_N}} \mathbb{P}_\Theta(\psi = 0) &= \sup_{\Theta \in \mathcal{G}_k^{\alpha_N}} \mathbb{P}_\Theta(\|\hat{\Theta}\|_{F,\Omega} < \tau_{d,k}(u)) \\ &\leq \sup_{\Theta \in \mathcal{G}_k^{\alpha_N}} \mathbb{P}_\Theta(\|\hat{\Theta} - \Theta\|_{F,\Omega} > \|\Theta\|_{F,\Omega} - \tau_{d,k}(u)) \\ &\leq b/u,\end{aligned}$$

provided $k^2\alpha_N^2 \geq 4\tau_{d,k}^2(u) = 4uf(k,d,N)$ which holds if $\alpha^2 \geq 4uc$. Hence,

$$\limsup_{n\to\infty} \left\{\mathbb{P}_0(\psi(Y) = 1) + \sup_{\Theta \in \mathcal{G}_k^{\alpha_N}} \mathbb{P}_\Theta(\psi(Y) = 0)\right\} \leq 2b/u < 1/3.$$

Summary

Statistical and computational trade-off in high-dimensional estimation.



The computational gap is most noticeable for the matrices of rank 1.

Conclusion

- The matrix logistic regression model is very natural to study the connection between **statistical accuracy** and **computational efficiency.**
- Block-sparsity is a **limiting model selection criterion** for polynomial-time estimation in the logistic regression model.
- With a larger parameter space, while the statistical rates might be worse, they might be closer to those that are computationally achievable.
- The logistic regression is also a representative of a **large class of generalised linear models.**

Information-Computational gap : the example of the SBM

Let $n, k \in \mathbb{N}^*$, $p = (p_1, \ldots, p_k)$ a probability vector and $W \in S_k([0,1])$.

Definition : SBM

(X, G) is drawn under $SBM(n, p, W)$ if :

- $X_u \sim p, \; \forall u \in [n]$.
- $G_{u,v} \sim \mathcal{B}(W(X_u, X_v)), \; , v \in [n], \; u \neq v$.

Definition : Agreement

Let $x, y \in [k]^n$.

$$A(x, y) = \max_{\pi \in S_k} \frac{1}{n} \sum_{u=1}^{n} \mathbb{1}_{x_u = \pi(y_u)}.$$

Exact Recovery : $\mathbb{P}(A(X, \hat{X}) = 1) = 1 - o(1)$.

$\longrightarrow$ No IC gap.

Information-Computational gap : the example of the SBM

Definition : Weak recovery

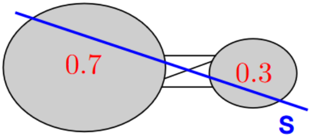
WR requires us to separate at least two communities.

Weak recovery is solvable in $SBM(n, p, W)$ if :

$$\exists \epsilon > 0, i, j \in [k], \text{ and an algo returning a partition } (S, S^c) \text{ of } [n] \text{ s.t.}$$

$$\mathbb{P}\left(\frac{|\Omega_i \cap S|}{|\Omega_i|} - \frac{|\Omega_j \cap S|}{|\Omega_j|} \geq \epsilon \right) = 1 - o(1),$$

where $|\Omega_i| = \{u \in [n] \ : \ X_u = i\}$.



$\longrightarrow$ No IC gap for $k = 2$ and the conjecture from Decelle and al. states that there is an IT gap for $k \geq 3$.