Information-theoretic thresholds for graph alignment.

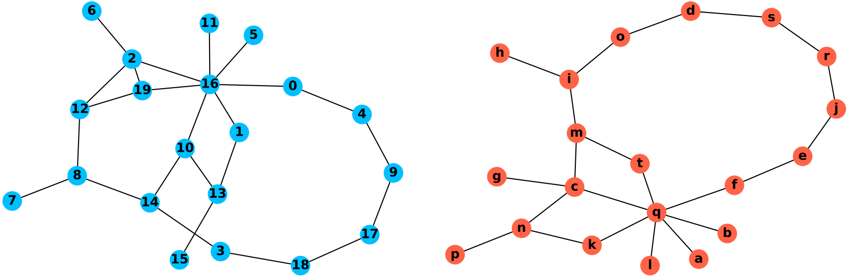
Luca Ganassali

INRIA, Paris

02/04/2020

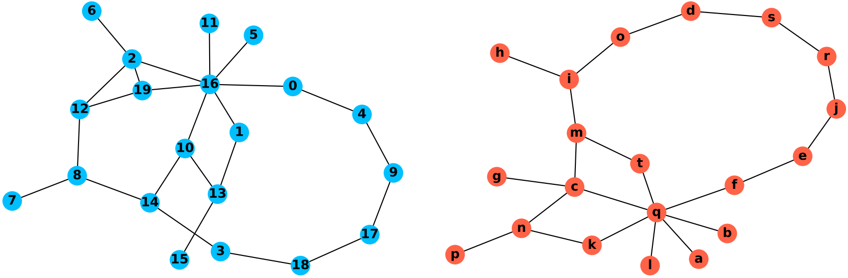
Introduction: graph alignment

Introduction: graph alignment



Two graphs: $G_1 = (V_1, E_1)$ (left) and $G_2 = (V_2, E_2)$ with $|V_1| = |V_2|$.

Introduction: graph alignment



Two graphs: $G_1 = (V_1, E_1)$ (left) and $G_2 = (V_2, E_2)$ with $|V_1| = |V_2|$.
Informal question: can we find a bijective mapping $\Pi : V_1 \to V_2$ such that:

$$G_2 \sim \Pi G_1 \Pi^T.$$

Planted permutation in the correlated Erdős-Rényi model

$ERC(n, p, s)$ model with planted permutation:

- Generate two aligned graphs G_1, G_2' with $V = [n]$, and for all $i, j \in V$, independently,

$$\mathbb{P}\left(i \underset{G_1}{\sim} j, i \underset{G_2'}{\sim} j\right) = ps,$$
$$\mathbb{P}\left(i \underset{G_1}{\sim} j, i \underset{G_2'}{\nsim} j\right) = \mathbb{P}\left(i \underset{G_1}{\nsim} j, i \underset{G_2'}{\sim} j\right) = p(1-s),$$
$$\mathbb{P}\left(i \underset{G_1}{\nsim} j, i \underset{G_2'}{\nsim} j\right) = 1 - p(2-s).$$

- Chose uniformly a random permutation (matrix) Π in S_n and define $G_2 = \Pi G_2' \Pi^T$.

Questions

Consider two models:

- Under $\mathbb{P}$ (planted model), $(G_1, G_2) \sim ERC(n, p, s)$.
- Under $\mathbb{Q}$ (null model), G_1, G_2 are two independent $ER(n, p)$ graphs.

Questions

Consider two models:

- Under $\mathbb{P}$ (planted model), $(G_1, G_2) \sim ERC(n, p, s)$.
- Under $\mathbb{Q}$ (null model), G_1, G_2 are two independent $ER(n, p)$ graphs.

Detection: test $\mathbb{P}$ vs $\mathbb{Q}$.

Questions

Consider two models:

- Under $\mathbb{P}$ (planted model), $(G_1, G_2) \sim ERC(n, p, s)$.
- Under $\mathbb{Q}$ (null model), G_1, G_2 are two independent $ER(n, p)$ graphs.

Detection: test $\mathbb{P}$ vs $\mathbb{Q}$.

'Partial' (Quasi-exact) reconstruction: Under $\mathbb{P}$, find an estimator $\hat{\Pi}$ such that $\sharp \left\{ i, \hat{\Pi}(i) = \Pi(i) \right\} = n - o(n)$ with high probability.

Questions

Consider two models:

- Under $\mathbb{P}$ (planted model), $(G_1, G_2) \sim ERC(n, p, s)$.
- Under $\mathbb{Q}$ (null model), G_1, G_2 are two independent $ER(n, p)$ graphs.

Detection: test $\mathbb{P}$ vs $\mathbb{Q}$.

'Partial' (Quasi-exact) reconstruction: Under $\mathbb{P}$, find an estimator $\hat{\Pi}$ such that $\sharp \left\{ i, \hat{\Pi}(i) = \Pi(i) \right\} = n - o(n)$ with high probability.

Exact reconstruction: Under $\mathbb{P}$, find an estimator $\hat{\Pi}$ such that $\hat{\Pi} = \Pi$ with high probability.

IT-threshold for exact reconstruction

IT-threshold for exact reconstruction

We have the following:

Theorem (Cullina, Kiyavash, '18)

(i) Under some mild sparsity constraints ($p \log^2 n \to 0$ and $s \geq p \log^2 n$), if

$$nps - (2) \log n \to \infty,$$

then there exists an estimator $\hat{\Pi}$ that achieves exact reconstruction with high probability.

(ii) If

$$s > p \quad \text{and} \quad nps - \log n \to -\infty,$$

then any estimator $\hat{\Pi}$ verifies $\hat{\Pi} = \Pi$ with probability $o(1)$.

Some notations

For $\pi \in S_n$, define its lifted version $\ell(\pi)$ (bijection on the set of pairs of vertices):

$$\begin{array}{ccccc} \ell(\pi) & : & \binom{[n]}{2} & \to & \binom{[n]}{2} \\ & & \{i,j\} & \mapsto & \{\pi(i),\pi(j)\} \end{array}$$

For any graph G and σ bijection of $\binom{V(G)}{2}$, define $G \circ \sigma$ the 'relabeled graph': it has vertex set $V(G)$ and

$$\forall e \in \binom{V(G)}{2}, \; e \in E(G \circ \sigma) \iff \sigma(e) \in E(G).$$

With these notations we observe $\underline{G_1, G_2 = G_2' \circ \ell(\Pi^{-1})}$ in the *ERC* model.

Some notations

$$p := \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} := \begin{pmatrix} 1 - p(2-s) & p(1-s) \\ p(1-s) & ps \end{pmatrix}.$$

Note that

$$\mathrm{Cov}\left(\mathbf{1}_{e \in G_1} \mathbf{1}_{e \in G_2'}\right) = \det p > 0 \iff s > p.$$

More generally we will assume positive correlation:

$$\frac{p_{00} p_{11}}{p_{01} p_{10}} > 1.$$

For any matrix $m \in \mathcal{M}_2(\mathbb{R})$, we use the notation

$$p^m = p_{00}^{m_{00}} p_{01}^{m_{01}} p_{10}^{m_{10}} p_{11}^{m_{11}}.$$

Some notations

For any graphs g_1 (blue), g_2 (red), define

$$\mu(g_1, g_2) := \begin{pmatrix} \sharp\,\{\text{uncolored edges}\} & \sharp\,\{\text{red edges}\} \\ \sharp\,\{\text{blue edges}\} & \sharp\,\{\text{blue and red edges}\} \end{pmatrix}$$

so that $\boxed{\mathbb{P}(G_1 = g_1, G_2' = g_2) = p^{\mu(g_1, g_2)}}$.

$$\Delta(g_1, g_2) := \sharp\,\{\text{simple-colored edges}\} = \mu(g_1, g_2)_{01} + \mu(g_1, g_2)_{10}.$$

For $\pi \in S_n$,

$$\delta(\pi, g_1, g_2) := \frac{1}{2}\left(\Delta(g_1, g_2 \circ \ell(\pi)) - \Delta(g_1, g_2)\right).$$

$\delta(\pi, g_1, g_2) \leq 0 \iff \pi$ is a better alignment than id for g_1, g_2.

Some notations

Lemma

$$\forall \pi \in S_n, \quad \mu(g_1, g_2 \circ \ell(\pi)) - \mu(g_1, g_2) = \delta(\pi, g_1, g_2) \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

Some notations

Lemma

$$\forall \pi \in S_n, \quad \mu(g_1, g_2 \circ \ell(\pi)) - \mu(g_1, g_2) = \delta(\pi, g_1, g_2) \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

Proof $M : \mu(g_1, g_2 \circ \ell(\pi)) - \mu(g_1, g_2)$. Edge conservation implies

$$M \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T M = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^T.$$

So $M = \lambda \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$, and
$2\lambda = M_{01} + M_{10} = \Delta(g_1, g_2 \circ l(\pi)) - \Delta(g_1, g_2)$. $\square$

Bayesian background

Bayesian background

Bayesian inference tells us that the best estimator is the maximum a posteriori (MAP):

$$\hat{\Pi}_{MAP} \in \arg\max_{\pi} \mathbb{P}\left(\Pi = \pi \middle| G_1, G_2\right).$$

In our case:

Bayesian background

Bayesian inference tells us that the best estimator is the maximum a posteriori (MAP):

$$\hat{\Pi}_{MAP} \in \arg\max_{\pi} \mathbb{P}\left(\Pi = \pi \middle| G_1, G_2\right).$$

In our case:

$$\begin{aligned}
\mathbb{P}\left(\Pi = \pi \middle| G_1 = g_1, G_2 = g_2\right) &\propto \mathbb{P}\left(\Pi = \pi, G_1 = g_1, G_2 = g_2\right) \\
&\propto \mathbb{P}\left(G_1 = g_1, G_2' = g_2 \circ \ell(\pi)\right) \\
&= p^{\mu(g_1, g_2 \circ \ell(\pi))} \\
&\propto \left(\frac{p_{10} p_{01}}{p_{00} p_{11}}\right)^{\delta(\pi, g_1, g_2)} \propto \left(\frac{p_{10} p_{01}}{p_{00} p_{11}}\right)^{\frac{1}{2}\Delta(\pi, g_1, g_2)}.
\end{aligned}$$

Bayesian background

Bayesian inference tells us that the best estimator is the maximum a posteriori (MAP):

$$\hat{\Pi}_{MAP} \in \arg\max_{\pi} \mathbb{P}\left(\Pi = \pi \middle| G_1, G_2\right).$$

In our case:

$$\begin{aligned}
\mathbb{P}\left(\Pi = \pi \middle| G_1 = g_1, G_2 = g_2\right) &\propto \mathbb{P}\left(\Pi = \pi, G_1 = g_1, G_2 = g_2\right) \\
&\propto \mathbb{P}\left(G_1 = g_1, G_2' = g_2 \circ \ell(\pi)\right) \\
&= p^{\mu(g_1, g_2 \circ \ell(\pi))} \\
&\propto \left(\frac{p_{10}p_{01}}{p_{00}p_{11}}\right)^{\delta(\pi, g_1, g_2)} \propto \left(\frac{p_{10}p_{01}}{p_{00}p_{11}}\right)^{\frac{1}{2}\Delta(\pi, g_1, g_2)}.
\end{aligned}$$

So

$$\boxed{\hat{\Pi}_{MAP} = \arg\min_{\pi} \Delta(G_1, G_2 \circ \ell(\pi)) = \arg\min_{\pi} \left\| G_1 - \pi^T G_2 \pi \right\|^2}.$$

Useful bounds

Changing the variable $\pi = \pi \circ \Pi^T$, to decide feasibility, a crucial set is

$$\mathcal{Q} := \{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\} .$$

Useful bounds

Changing the variable $\pi = \pi \circ \Pi^T$, to decide feasibility, a crucial set is

$$\mathcal{Q} := \left\{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\right\}.$$

- $\mathbb{P}\left(\hat{\Pi}_{MAP} \neq \Pi\right) \leq \sum_{\pi \neq \mathrm{id}} \mathbb{P}\left(\pi \in \mathcal{Q}\right)$ (Achievability case)

Useful bounds

Changing the variable $\pi = \pi \circ \Pi^T$, to decide feasibility, a crucial set is

$$\mathcal{Q} := \left\{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\right\}.$$

- $\mathbb{P}\left(\hat{\Pi}_{MAP} \neq \Pi\right) \leq \sum_{\pi \neq \mathrm{id}} \mathbb{P}\left(\pi \in \mathcal{Q}\right)$ (Achievability case)
- $\mathbb{P}\left(\hat{\Pi}_{MAP} = \Pi\right) \leq \mathbb{E}\left[\mathbf{1}_{\mathrm{id} \in \arg\min_\pi \Delta(\pi, G_1, G_2')} / |\mathcal{Q}|\right]$ (Non-feasible case)

Useful bounds

Changing the variable $\pi = \pi \circ \Pi^T$, to decide feasibility, a crucial set is

$$\mathcal{Q} := \left\{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\right\}.$$

- $\mathbb{P}\left(\hat{\Pi}_{MAP} \neq \Pi\right) \leq \sum_{\pi \neq \mathrm{id}} \mathbb{P}\left(\pi \in \mathcal{Q}\right)$ (Achievability case)
- $\mathbb{P}\left(\hat{\Pi}_{MAP} = \Pi\right) \leq \mathbb{E}\left[\mathbf{1}_{\mathrm{id} \in \arg\min_\pi \Delta(\pi, G_1, G_2')} / |\mathcal{Q}|\right]$ (Non-feasible case)

Proof for non-feasible case

Proof for non-feasible case

If $nps - \log n \to -\infty$, we can show that

$$\mathcal{Q} := \{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\}$$

is 'too large':

Proof for non-feasible case

If $nps - \log n \to -\infty$, we can show that

$$\mathcal{Q} := \{\pi \in S_n, \, \delta(\pi, G_1, G_2') \leq 0\}$$

is 'too large':

Theorem (Automorphisms of Erdős-Rényi graphs (Bollobas '85))

Let $G \sim ER(n, p)$. If $p \leq \frac{\log n - c_n}{n}$ with $c_n \to \infty$, then there exists some sequence $w(n) \to \infty$ such that

$$\mathbb{P}\left(|\mathrm{Aut}(G)| \leq w(n)\right) \leq \frac{1}{w(n)}.$$

(Proof: $|\mathrm{Aut}(G)| \geq X!$, where X is the number of isolated vertices, then second moment method).

Proof for non-feasible case

If $nps - \log n \to -\infty$, we can show that

$$\mathcal{Q} := \{\pi \in S_n,\, \delta(\pi, G_1, G_2') \leq 0\}$$

is 'too large':

Theorem (Automorphisms of Erdős-Rényi graphs (Bollobas '85))

Let $G \sim ER(n,p)$. If $p \leq \frac{\log n - c_n}{n}$ with $c_n \to \infty$, then there exists some sequence $w(n) \to \infty$ such that

$$\mathbb{P}\left(|\mathrm{Aut}(G)| \leq w(n)\right) \leq \frac{1}{w(n)}.$$

(Proof: $|\mathrm{Aut}(G)| \geq X!$, where X is the number of isolated vertices, then second moment method).

Since $\mathrm{Aut}\left(G_1 \cap G_2'\right) \subset \mathcal{Q}$,

$$\mathbb{P}\left(\hat{\Pi}_{MAP} = \Pi\right) \leq \mathbb{E}\left[1/|\mathcal{Q}|\right] \leq 2/w(n) \to 0.$$

Proof sketch in the achievability case

Proof sketch in the achievability case

Fix $\pi \neq \mathrm{id}$, and bound $\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right)$.

Proof sketch in the achievability case

Fix $\pi \neq \mathrm{id}$, and bound $\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right)$. We note

$$\mathcal{S} := \left\{ e \in \binom{n}{2},\ \ell(\pi)(e) \neq e \right\}$$

Proof sketch in the achievability case

Fix $\pi \neq \mathrm{id}$, and bound $\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right)$. We note

$$\mathcal{S} := \left\{ e \in \binom{n}{2}, \ \ell(\pi)(e) \neq e \right\}$$

Lemma

$$\mathbb{E}\left[\delta(\pi, G_1, G_2')\right] = |\mathcal{S}|(p_{00}p_{11} - p_{01}p_{01}) = |\mathcal{S}| \det p > 0.$$

Proof sketch in the achievability case

Fix $\pi \neq \mathrm{id}$, and bound $\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right)$. We note

$$\mathcal{S} := \left\{ e \in \binom{n}{2}, \; \ell(\pi)(e) \neq e \right\}$$

Lemma

$$\mathbb{E}\left[\delta(\pi, G_1, G_2')\right] = |\mathcal{S}|(p_{00}p_{11} - p_{01}p_{01}) = |\mathcal{S}| \det p > 0.$$

Proof:

$$\begin{aligned}
\mathbb{E}\left[\delta(\pi, G_1, G_2')\right] &= \sum_{e} \left(\mathbb{P}\left(e \in G_1, e \in G_2'\right) - \mathbb{P}\left(e \in G_1, \ell(\pi)(e) \in G_2'\right)\right) \\
&= \sum_{e \in \mathcal{S}} \left(p_{11} - (p_{10} + p_{11})(p_{01} + p_{11})\right) \\
&= |\mathcal{S}|(p_{00}p_{11} - p_{01}p_{01}).
\end{aligned}$$

Proof sketch in the achievability case

We now need a 'large-deviation principle' for δ, measured in terms of $|\mathcal{S}|$:

Proof sketch in the achievability case

We now need a 'large-deviation principle' for δ, measured in terms of $|\mathcal{S}|$:

Lemma (Cullina, Kiyavash, '18)

$$\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right) \leq \exp\left(-z|\mathcal{S}|\right),$$

where

$$z = \left(\sqrt{p_{00}p_{11}} - \sqrt{p_{01}p_{10}}\right)^2$$

Proof sketch in the achievability case

We now need a 'large-deviation principle' for δ, measured in terms of $|\mathcal{S}|$:

Lemma (Cullina, Kiyavash, '18)

$$\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right) \leq \exp\left(-z|\mathcal{S}|\right),$$

where

$$z = \left(\sqrt{p_{00}p_{11}} - \sqrt{p_{01}p_{10}}\right)^2$$

(Proof: difficult, based on analysis of the generating function of (δ, μ).)

Proof sketch in the achievability case

We now need a 'large-deviation principle' for δ, measured in terms of $|\mathcal{S}|$:

Lemma (Cullina, Kiyavash, '18)

$$\mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right) \leq \exp\left(-z|\mathcal{S}|\right),$$

where

$$z = \left(\sqrt{p_{00}p_{11}} - \sqrt{p_{01}p_{10}}\right)^2$$

(Proof: difficult, based on analysis of the generating function of (δ, μ).)

Note: if $p \log^2 n \to 0$ and $s \geq p \log^2 n$, then

$$z \sim p_{11} = ps.$$

Proof sketch in the achievability case

A more natural quantity is

$$s := \sharp \{1 \leq i \leq n,\ \pi(i) \neq i\}.$$

Proof sketch in the achievability case

A more natural quantity is

$$s := \sharp \{1 \leq i \leq n,\ \pi(i) \neq i\}.$$

We have

$$\binom{n-s}{2} \leq \binom{n}{2} - |\mathcal{S}| \leq \binom{n-s}{2} + \frac{s}{2},$$

Proof sketch in the achievability case

A more natural quantity is

$$s := \sharp \left\{ 1 \leq i \leq n,\ \pi(i) \neq i \right\}.$$

We have

$$\binom{n-s}{2} \leq \binom{n}{2} - |\mathcal{S}| \leq \binom{n-s}{2} + \frac{s}{2},$$

and in particular,

$$\boxed{|\mathcal{S}| \geq \frac{s(n-2)}{2}}.$$

Proof sketch in the achievability case

A more natural quantity is

$$s := \sharp \left\{ 1 \leq i \leq n,\ \pi(i) \neq i \right\}.$$

We have

$$\binom{n-s}{2} \leq \binom{n}{2} - |\mathcal{S}| \leq \binom{n-s}{2} + \frac{s}{2},$$

and in particular,

$$\boxed{|\mathcal{S}| \geq \frac{s(n-2)}{2}}.$$

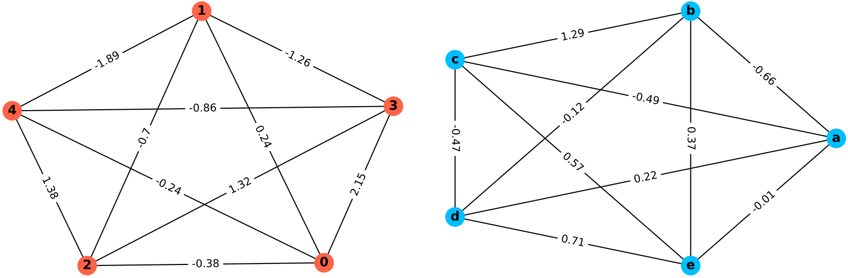
Then,

$$\mathbb{P}\left(\hat{\Pi}_{MAP} \neq \Pi\right) \leq \sum_{t=1}^{n} \sum_{\pi | s=t} \mathbb{P}\left(\delta(\pi, G_1, G_2') \leq 0\right) \leq \sum_{t=1}^{n} n^t \exp\left(-zt(n-2)/2\right)$$

$$\lesssim \sum_{t=1}^{n} \exp\left(-pst(n-2)/2 + t \log n\right) \leq \boxed{O\left(\exp\left(-psn/2 + \log n\right)\right)}.$$

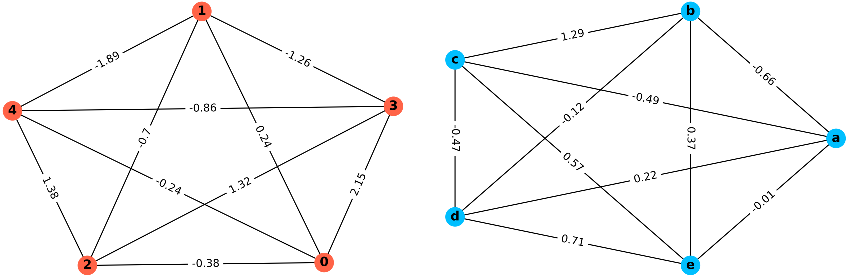
Another setting: weighted graph (or matrix) alignment

Another setting: weighted graph (or matrix) alignment



Two complete graphs G_1, G_2 with correlated weights.

Another setting: weighted graph (or matrix) alignment



Two complete graphs G_1, G_2 with correlated weights.
Informal question: can we find a bijective mapping $\Pi : V_1 \to V_2$ such that:

$$G_2 \sim \Pi G_1 \Pi^T.$$

Standard *GOE* model

Our new planted model $\mathbb{P}$*:*

$$A_{i,j} = A_{j,i} \sim \begin{cases} \frac{1}{\sqrt{n}}\mathcal{N}(0,1) & \text{if } i \neq j, \\ \frac{\sqrt{2}}{\sqrt{n}}\mathcal{N}(0,1) & \text{if } i = j, \end{cases}$$

and H is an independent copy of A. Draw a uniform permutation matrix Π of size $n \times n$, and $B = \Pi^T \left(\sqrt{1-\sigma^2}A + \sigma H\right)\Pi$, where $\sigma = \sigma(N)$ is the noise parameter.

Standard *GOE* model

Our new planted model $\mathbb{P}$*:*

$$A_{i,j} = A_{j,i} \sim \begin{cases} \frac{1}{\sqrt{n}}\mathcal{N}(0,1) & \text{if } i \neq j, \\ \frac{\sqrt{2}}{\sqrt{n}}\mathcal{N}(0,1) & \text{if } i = j, \end{cases}$$

and H is an independent copy of A. Draw a uniform permutation matrix Π of size $n \times n$, and $B = \Pi^T \left(\sqrt{1-\sigma^2}A + \sigma H\right)\Pi$, where $\sigma = \sigma(N)$ is the noise parameter.

Question: What is the IT-threshold for exact recovery?

IT threshold for weighted graph alignment

In this case:

IT threshold for weighted graph alignment

In this case:

$$\mathbb{P}\left(\Pi = \pi \middle| A = a, B = b\right) \propto \mathbb{P}\left(\Pi = \pi, A = a, H = \frac{1}{\sigma}\left(\pi b \pi^T - \sqrt{1-\sigma^2} A\right)\right)$$

$$\propto \exp\left(-n\frac{\mathrm{Tr}(a^2)}{4}\right) \exp\left(-n\frac{\mathrm{Tr}\left(\left(\pi b \pi^T - \sqrt{1-\sigma^2} a\right)^2\right)}{4}\right)$$

IT threshold for weighted graph alignment

In this case:

$$\mathbb{P}\left(\Pi = \pi \middle| A = a, B = b\right) \propto \mathbb{P}\left(\Pi = \pi, A = a, H = \frac{1}{\sigma}\left(\pi b \pi^T - \sqrt{1-\sigma^2} A\right)\right)$$

$$\propto \exp\left(-n\frac{\mathrm{Tr}(a^2)}{4}\right) \exp\left(-n\frac{\mathrm{Tr}\left(\left(\pi b \pi^T - \sqrt{1-\sigma^2} a\right)^2\right)}{4}\right)$$

So

$$\boxed{\hat{\Pi}_{MAP} = \arg\min_{\pi} \underbrace{\mathrm{Tr}\left(\left(\pi B \pi^T - \sqrt{1-\sigma^2} A\right)^2\right)}_{=:C(\pi,A,H)}}.$$

IT threshold for weighted graph alignment

Make the variable change $\tau = \pi \Pi^T$ and define:

$$\delta(\tau, A, B) = C(\pi, A, B) - C(\Pi, A, B).$$

IT threshold for weighted graph alignment

Make the variable change $\tau = \pi \Pi^T$ and define:

$$\delta(\tau, A, B) = C(\pi, A, B) - C(\Pi, A, B).$$

We have

$$\begin{aligned}
&\delta(\tau, A, B) \\
&= (1-\sigma^2) \underbrace{\mathrm{Tr}\left((\tau A \tau^T - A)^2\right)}_{=:Q(A)_\tau} - 2\sigma\sqrt{1-\sigma^2}\underbrace{\mathrm{Tr}\left(\tau H \tau^T (\tau A \tau^T - A)\right)}_{=:L(A,H)_\tau \overset{(d)}{=} \mathcal{N}\left(0, \frac{1}{n} Q(A)_\tau\right)}
\end{aligned}$$

IT threshold for weighted graph alignment

Defining as before

$$s := \sharp \{1 \leq i \leq n,\ \pi(i) \neq i\},$$

further analysis shows that $Q(A)_\tau = \mathbb{P}\left(\delta(\tau, A, B) \leq 0\right)$ is of order $\Theta(1) \times s$.

IT threshold for weighted graph alignment

Defining as before

$$s := \sharp \{1 \leq i \leq n,\ \pi(i) \neq i\},$$

further analysis shows that $Q(A)_\tau = \mathbb{P}\left(\delta(\tau, A, B) \leq 0\right)$ is of order $\Theta(1) \times s$.

$$\begin{aligned}
&\mathbb{P}\left(\delta(\tau, A, B) \leq 0\right) \\
&= \mathbb{P}\left((1-\sigma^2)\mathcal{N}\left(1, \frac{4\sigma^2}{ns\Theta(1)(1-\sigma^2)}\right) \leq 0\right) \leq \exp\left(-C\frac{1-\sigma^2}{\sigma^2}ns\right).
\end{aligned}$$

IT threshold for weighted graph alignment

Defining as before

$$s := \sharp \left\{ 1 \leq i \leq n,\ \pi(i) \neq i \right\},$$

further analysis shows that $Q(A)_\tau = \mathbb{P}\left(\delta(\tau, A, B) \leq 0\right)$ is of order $\Theta(1) \times s$.

$$\begin{aligned} & \mathbb{P}\left(\delta(\tau, A, B) \leq 0\right) \\ & = \mathbb{P}\left((1-\sigma^2)\mathcal{N}\left(1, \frac{4\sigma^2}{ns\Theta(1)(1-\sigma^2)}\right) \leq 0 \right) \leq \exp\left(-C\frac{1-\sigma^2}{\sigma^2} ns \right). \end{aligned}$$

So we need to compare $C\frac{1-\sigma^2}{\sigma^2}n$ to $\log n$, i.e. σ to $\frac{1}{\left(1+\frac{\log n}{n}\right)^{1/2}} \sim 1 - \frac{\log n}{2n}$.

IT threshold for weighted graph alignment

IT threshold for weighted graph alignment

Theorem (Work in progress...)

There exist $0 < c < C$ such that:

(*i*) *If*

$$1 - \sigma \geq C \frac{\log n}{n},$$

then there exists an estimator $\hat{\Pi}$ that achieves exact reconstruction with high probability.

(*ii*) *If*

$$1 - \sigma \leq c \frac{\log n}{n},$$

then any estimator $\hat{\Pi}$ verifies $\hat{\Pi} = \Pi$ with probability $o(1)$.

Thank you!