Impossibility of Partial Recovery in the Graph Alignment Problem

Luca Ganassali, Marc Lelarge and Laurent Massoulié

34th Annual Conference on Learning Theory, July 2021

INRIA, DI/ENS, PSL Research University, Paris, France

The Graph Alignment problem

Question: Given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$, *what is the best way to match nodes of G with nodes of G'?*

Question: Given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$, *what is the best way to match nodes of G with nodes of G'?*

Minimizing disagreements: Find a bijection $f : V \to V'$ that minimizes

$$\sum_{(i,j)\in V^2} \left(\mathbf{1}_{(i,j)\in E} - \mathbf{1}_{(f(i),f(j))\in E'}\right)^2 ,$$

or, equivalently solve

$$\max_{\Pi} \mathrm{Tr}\left(G\Pi G'\Pi^\top\right) ,$$

where Π runs over all permutation matrices.

The Graph Alignment problem

Question: Given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$, *what is the best way to match nodes of G with nodes of G'?*

Minimizing disagreements: Find a bijection $f : V \to V'$ that minimizes

$$\sum_{(i,j)\in V^2} \left(\mathbf{1}_{(i,j)\in E} - \mathbf{1}_{(f(i),f(j))\in E'}\right)^2 ,$$

or, equivalently solve

$$\max_{\Pi} \mathrm{Tr}\left(G\Pi G'\Pi^\top\right) ,$$

where Π runs over all permutation matrices. $\longleftarrow$ *NP-hard in the worst case*

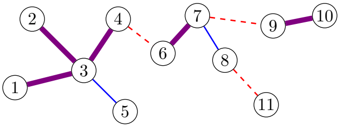
Correlated Erdős-Rényi model $\mathcal{G}(n, q, s)$:

- Draw two graphs $\mathcal{G}, \mathcal{G}'$ with same node set $[n]$, s.t. for all $(i,j) \in \binom{[n]}{2}$:

$$\left(\mathbf{1}_{i \underset{\mathcal{G}}{\sim} j}, \mathbf{1}_{i \underset{\mathcal{G}'}{\sim} j}\right) = \begin{cases} (1,1) & \text{w.p. } qs & \textit{two–coloured} \text{ edge} \\ (0,1),(1,0) & \text{w.p. } q(1-s) & \textit{red} \text{ or } \textit{blue} \text{ edge} \\ (0,0) & \text{w.p. } 1-q(2-s) & \text{non-edge} \end{cases}$$

Sparse setting: $q = \lambda/n \longleftarrow$ *constant mean degree*

$s \in [0,1]$ *is the correlation parameter*



Planted Graph Alignment

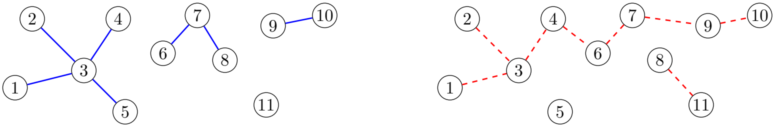
Correlated Erdős-Rényi model $\mathcal{G}(n, q, s)$**:**

- Draw two graphs $\mathcal{G}, \mathcal{G}'$ with same node set $[n]$, s.t. for all $(i, j) \in \binom{[n]}{2}$:

$$\left(\mathbf{1}_{i \underset{\mathcal{G}}{\sim} j}, \mathbf{1}_{i \underset{\mathcal{G}'}{\sim} j}\right) = \begin{cases} (1,1) & \text{w.p. } qs & \text{two}-\text{coloured edge} \\ (0,1),(1,0) & \text{w.p. } q(1-s) & \text{red or blue edge} \\ (0,0) & \text{w.p. } 1-q(2-s) & \text{non-edge} \end{cases}$$

Sparse setting: $q = \lambda/n \longleftarrow$ *constant mean degree*

$s \in [0, 1]$ *is the correlation parameter*



Planted Graph Alignment

Correlated Erdős-Rényi model $\mathcal{G}(n, q, s)$:

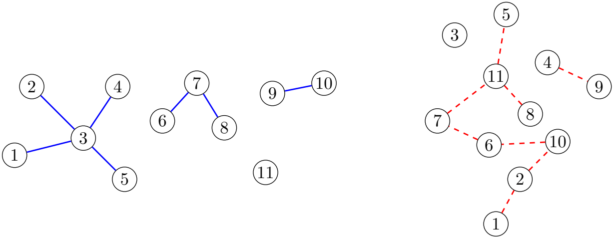
- Draw two graphs $\mathcal{G}, \mathcal{G}'$ with same node set $[n]$, s.t. for all $(i,j) \in \binom{[n]}{2}$:

$$\left(\mathbf{1}_{i \underset{\mathcal{G}}{\sim} j}, \mathbf{1}_{i \underset{\mathcal{G}'}{\sim} j}\right) = \begin{cases} (1,1) & \text{w.p. } qs & \textit{two-coloured} \text{ edge} \\ (0,1),(1,0) & \text{w.p. } q(1-s) & \textit{red} \text{ or } \textit{blue} \text{ edge} \\ (0,0) & \text{w.p. } 1-q(2-s) & \text{non-edge} \end{cases}$$

Sparse setting: $q = \lambda/n \longleftarrow$ *constant mean degree*

$s \in [0,1]$ *is the correlation parameter*

- Relabel the vertices of $\mathcal{G}'$ with a uniform independent permutation π^*: $\mathcal{H} := \mathcal{G}' \circ \pi^*$.



Planted Graph Alignment: state-of-the art

Goal: upon observing $\mathcal{G}$ and $\mathcal{H}$, recover π^* with high probability.

Planted Graph Alignment: state-of-the art

Goal: upon observing $\mathcal{G}$ and $\mathcal{H}$, recover π^* with high probability.

Sparse regime, even with $s = 1$, $\Theta(n)$ isolated vertices $\longleftarrow$ *only partial alignment may be reachable* [Cullina-Kiyavash '16].

Planted Graph Alignment: state-of-the art

Goal: upon observing $\mathcal{G}$ and $\mathcal{H}$, recover π^* with high probability.

Sparse regime, even with $s = 1$, $\Theta(n)$ isolated vertices $\longleftarrow$ *only partial alignment may be reachable* [Cullina-Kiyavash '16].

Notion of performance: for any $\mathcal{S}_n$-valued estimator $\hat{\pi}(\mathcal{G}, \mathcal{H})$, define its overlap with the planted permutation π^*

$$\mathrm{ov}(\hat{\pi}(\mathcal{G}, \mathcal{H}), \pi^*) := \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} \mathbf{1}_{\hat{\pi}(\mathcal{G}^\sigma, \mathcal{H})(i) = \pi^* \circ \sigma^{-1}(i)},$$

Planted Graph Alignment: state-of-the art

Goal: upon observing $\mathcal{G}$ and $\mathcal{H}$, recover π^* with high probability.

Sparse regime, even with $s = 1$, $\Theta(n)$ isolated vertices $\longleftarrow$ *only partial alignment may be reachable* [Cullina-Kiyavash '16].

Notion of performance: for any $\mathcal{S}_n$-valued estimator $\hat{\pi}(\mathcal{G}, \mathcal{H})$, define its overlap with the planted permutation π^*

$$\mathrm{ov}(\hat{\pi}(\mathcal{G}, \mathcal{H}), \pi^*) := \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} \mathbf{1}_{\hat{\pi}(\mathcal{G}^\sigma, \mathcal{H})(i) = \pi^* \circ \sigma^{-1}(i)},$$

Questions: can we hope for $\mathrm{ov}(\hat{\pi}, \pi^*) > \alpha n$ w.h.p. (no computational restrictions)? What is the maximal fraction α?

State-of-the art: in the sparse case where $\lambda > 0$ and $s \in [0, 1]$ are fixed constants: partial recovery is IT-feasible if $\lambda s > 4 + \varepsilon$ [Wu-Xu-Yu '21].

Main result: maximal reachable overlap

Main result: maximal reachable overlap

Theorem 1
For $\lambda > 0$ and $s \in [0, 1]$, we have for any $\alpha > 0$, for any estimator $\hat{\pi}$:

$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi^*) > (c(\lambda s) + \alpha)n\right) \underset{n \to \infty}{\longrightarrow} 0,$$

where $c(\mu)$ is the greatest non-negative solution to the equation $e^{-\mu x} = 1 - x$.

Main result: maximal reachable overlap

Theorem 1
For $\lambda > 0$ and $s \in [0,1]$, we have for any $\alpha > 0$, for any estimator $\hat{\pi}$:

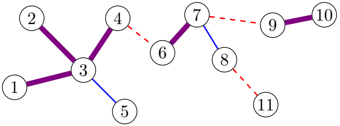
$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi^*) > (c(\lambda s) + \alpha)n\right) \underset{n \to \infty}{\longrightarrow} 0,$$

where $c(\mu)$ is the greatest non-negative solution to the equation $e^{-\mu x} = 1 - x$.

Corollary: Partial recovery is IT-infeasible if $\lambda s \leq 1$.

Intuition: exchanging small tree components

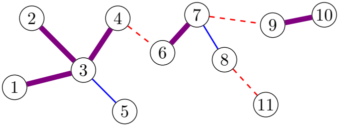
1. Information contained in the **intersection graph** $\mathcal{G} \wedge \mathcal{G}'$:



In our model $\mathcal{G} \wedge \mathcal{G}'$ is an Erdős-Rényi graph: $\mathcal{G} \wedge \mathcal{G}' \sim G(n, \lambda s/n)$.

Intuition: exchanging small tree components

1. Information contained in the **intersection graph** $\mathcal{G} \wedge \mathcal{G}'$:

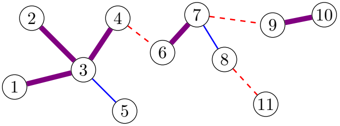


   In our model $\mathcal{G} \wedge \mathcal{G}'$ is an Erdős-Rényi graph: $\mathcal{G} \wedge \mathcal{G}' \sim G(n, \lambda s/n)$.

2. [Erdős, Rényi, Bollobás] typical fraction $c(\lambda s)$ of nodes in the giant component of $\mathcal{G} \wedge \mathcal{G}' \to$ the remaining $(1 - c(\lambda s))n$ nodes are almost all on **small tree components**.

Intuition: exchanging small tree components

1. Information contained in the **intersection graph** $\mathcal{G} \wedge \mathcal{G}'$:



   In our model $\mathcal{G} \wedge \mathcal{G}'$ is an Erdős-Rényi graph: $\mathcal{G} \wedge \mathcal{G}' \sim G(n, \lambda s/n)$.
2. [Erdős, Rényi, Bollobás] typical fraction $c(\lambda s)$ of nodes in the giant component of $\mathcal{G} \wedge \mathcal{G}' \to$ the remaining $(1 - c(\lambda s))n$ nodes are almost all on **small tree components**.
3. For any small tree **T**, a large number of copies of **T** will appear in $\mathcal{G} \wedge \mathcal{G}'$. **Reshuffle them in** $\mathcal{G} \to$ a lot of unnoticed corrupted candidates for $\hat{\pi}$ that are far from π^*.

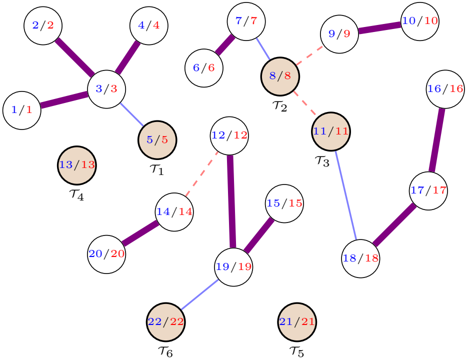
Intuition: exchanging small tree components

Let $\mathbb{T} := \{\mathbf{T}_1, \ldots, \mathbf{T}_M\}$ be the set of all trees (up to isomorphism) of size less than $K := K(n) = \lfloor \log n \rfloor$.

Algorithm 1: Recursive construction of σ

Initialize $\sigma_0 \leftarrow \mathrm{id}$;
for $i = 1$ *to* M **do**
  Consider $\mathbf{T} = \mathbf{T}_i$ and draw uniformly at random the tree permutation $\Sigma_{\mathbf{T}} \in \mathcal{S}_{X_{\mathbf{T}}}$, independently from the past;
  Consider $\sigma_{\mathbf{T}}$ the node permutation associated with $\Sigma_{\mathbf{T}}$;
  $\sigma_i \leftarrow \sigma_{\mathbf{T}} \circ \sigma_{i-1}$;
end
return $\sigma = \sigma_M$

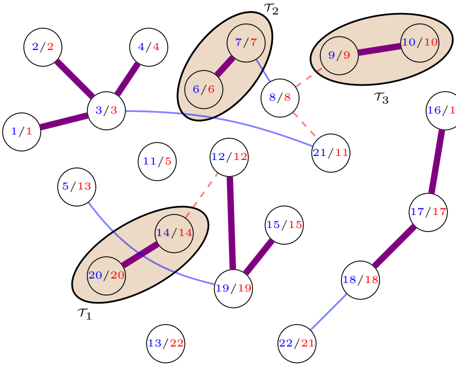
Intuition: exchanging small tree components



$$\Sigma_{\mathbf{T}_1} = (2)(1\ 3\ 5\ 6\ 4)$$

$$\sigma_{\mathbf{T}_1} = (8)(5\ 11\ 21\ 22\ 13)$$

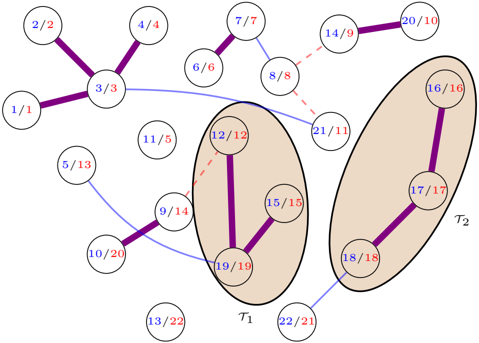
Intuition: exchanging small tree components



$$\Sigma_{\mathbf{T}_2} = (2)(1\ 3)$$

$$\sigma_{\mathbf{T}_2} = (6)(7)(9\ 14)(10\ 20)$$

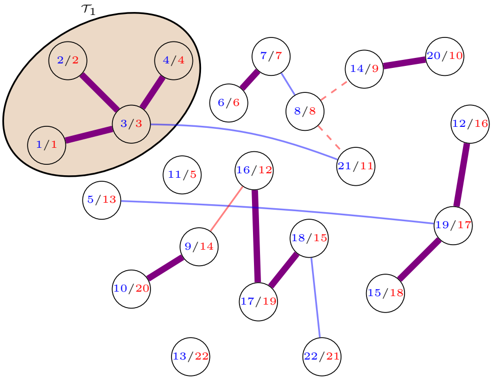
Intuition: exchanging small tree components



$$\Sigma_{\mathbf{T}_3} = (1\ 2)$$

$$\sigma_{\mathbf{T}_3} = (15\ 18)(19\ 17)(12\ 16)$$

Intuition: exchanging small tree components



$$\Sigma_{\mathbf{T}_4} = (1)$$

$$\sigma_{\mathbf{T}_4} = (1)(2)(3)(4)$$

4. Joint distribution for $\mathcal{G}, \mathcal{G}'$:

$$\mathbb{P}(\mathcal{G} = G, \mathcal{G}' = G') = \frac{1}{\mathcal{Z}_{\lambda,s}(G)\mathcal{Z}_{\lambda,s}(G')} \left[\frac{s(n - \lambda(2 - s))}{\lambda(1 - s)^2}\right]^{\left|E(G \wedge G')\right|}$$

4. Joint distribution for $\mathcal{G}, \mathcal{G}'$:

$$\mathbb{P}(\mathcal{G} = G, \mathcal{G}' = G') = \frac{1}{\mathcal{Z}_{\lambda,s}(G)\mathcal{Z}_{\lambda,s}(G')} \left[\frac{s(n - \lambda(2 - s))}{\lambda(1 - s)^2}\right]^{|E(G \wedge G')|}$$

Theorem 2

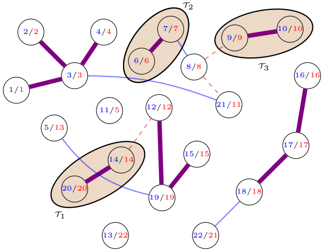
Fix an integer $p > 0$. Consider $(\mathcal{G}, \mathcal{G}')$ drawn under the correlated Erdős-Rényi model. Then, with high probability, there exists $\{\sigma_i\}_{i \in [p]}$ – that depend on the intersection graph $\mathcal{G} \wedge \mathcal{G}'$ – such that

(i) $\forall i \in [p],\ |E(\mathcal{G}^{\sigma_i} \wedge \mathcal{G}')| = |E(\mathcal{G} \wedge \mathcal{G}')|$,

(ii) $\forall i, j \in [p],\ i \neq j \implies \sum_{\ell=1}^{n} \mathbf{1}_{\sigma_i(\ell) = \sigma_j(\ell)} \leq c(\lambda s)n + o(n)$, *where the $o(n)$ is independent of $i, j \in [p]$.*

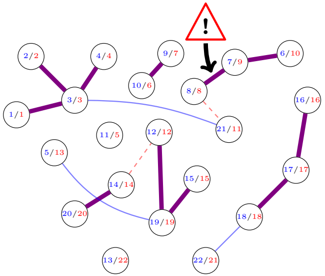
- Point (*ii*) **(all permutations are far apart)**: *lower bound for $X_{\mathbf{T}}$*, number of components isomorphic to $\mathbf{T}$ in $\mathcal{G} \wedge \mathcal{G}'$ holding uniformly w.h.p. + *standard random permutation arguments*.
- Point (*i*) **(same number of double edges)**: Define $\mathcal{S}$: $\{u, v\} \in \mathcal{S}$ if edge $\{u, v\}$ is not double. Control of the *number of extra double edges*:

$$\Delta(\sigma) := \sum_{\{u,v\} \in \mathcal{S}} \mathbf{1}_{u \longleftrightarrow v} \mathbf{1}_{\sigma(u) \longleftrightarrow \sigma(v)}.$$



$\Sigma_{\mathbf{T}_2} = (1)(2\ 3)$

$\sigma_{\mathbf{T}_2} = (14)(20)(7\ 9)(6\ 10)$



Proof sketch for Theorem 2

- Point (*ii*) **(all permutations are far apart)**: *lower bound for* $X_{\mathbf{T}}$, number of components isomorphic to **T** in $\mathcal{G} \wedge \mathcal{G}'$ holding uniformly w.h.p. + *standard random permutation arguments.*
- Point (*i*) **(same number of double edges)**: Define $\mathcal{S}$: $\{u,v\} \in \mathcal{S}$ if edge $\{u,v\}$ is not double. Control of the *number of extra double edges*:

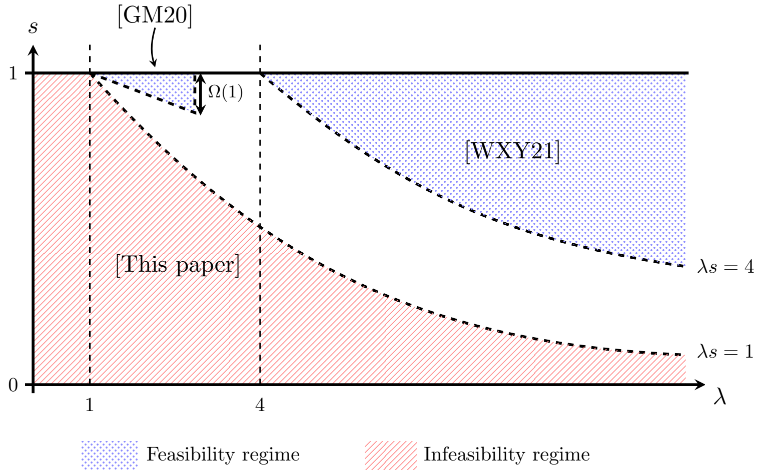
$$\Delta(\sigma) := \sum_{\{u,v\} \in \mathcal{S}} \mathbf{1}_{u \longleftrightarrow v} \mathbf{1}_{\sigma(u) \longleftrightarrow \sigma(v)}.$$

Conditionally to the two-coloured edges and to σ, $\Delta(\sigma)$ is *asymptotically Poisson* with parameter

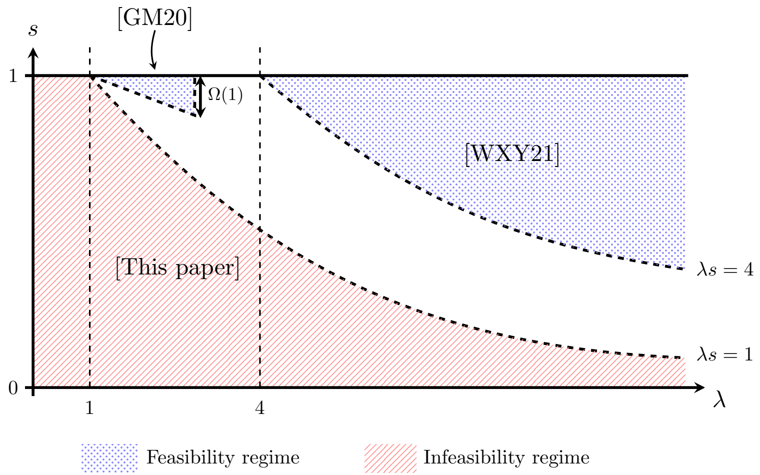
$$\frac{(1-c(\lambda s)n)^2}{2} \times \frac{\lambda^2(1-s)^2}{n^2} = \frac{\lambda^2(1-s)^2(1-c(\lambda s))^2}{2}.$$

$\longrightarrow$ *probabilistic method.*

Conclusion: diagram for partial recovery



Conclusion: diagram for partial recovery



Conjecture: *$\lambda s = 1$ is the sharp threshold.*

Thank you!