Impossibility of Partial Recovery in the Graph Alignment Problem

Luca Ganassali†, Laurent Massoulié†,⋆, Marc Lelarge†

†INRIA, DI/ENS, PSL Research University, Paris, France ⋆MSR-Inria Joint Centre

Abstract

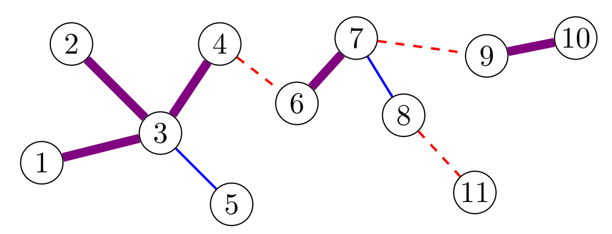
Random graph alignment refers to recovering the underlying vertex correspondence between two random graphs with correlated edges. This can be viewed as an average-case and noisy version of the well-known graph isomorphism problem. For the correlated Erdős-Rényi model, we prove an impossibility result for partial recovery in the sparse regime, with constant average degree and correlation, as well as a general bound on the maximal reachable overlap. Our bound is tight in the noiseless case (the graph isomorphism problem) and we conjecture that it is still tight with noise. Our proof technique relies on a careful application of the probabilistic method to build automorphisms between tree components of a subcritical Erdős-Rényi graph.

Planted graph alignment: correlated Erdős-Rényi model

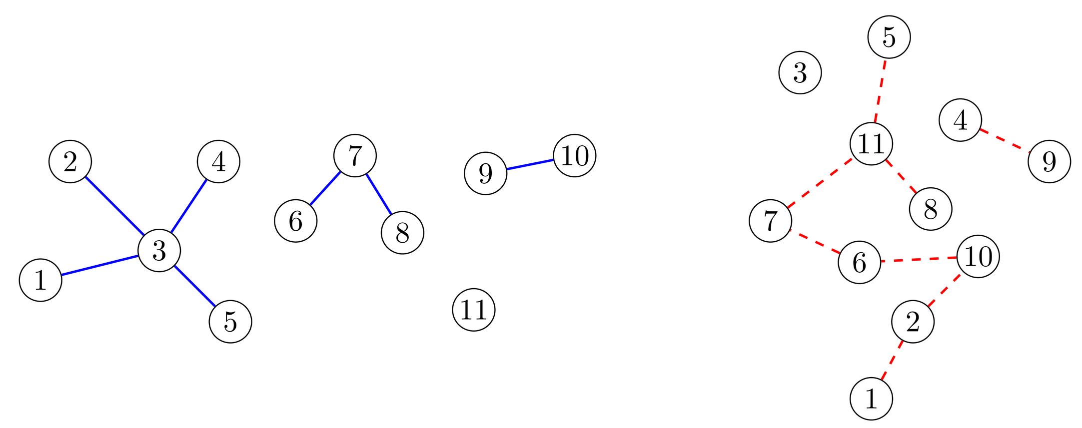
(1) Draw two graphs $\mathcal{G}, \mathcal{G}'$ with same node set $[n]$, s.t. for all $(i,j) \in \binom{[n]}{2}$:

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

Sparse setting: constant mean degree with $q = \lambda/n$, with correlation parameter $s \in [0,1]$.



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



Goal: given $\mathcal{G}, \mathcal{H}$, find an estimator $\hat{\pi}$ that partially recovers π^* w.h.p., that is s.t. $\mathrm{ov}(\hat{\pi}, \pi^*) \geq \alpha n$, for some $\alpha > 0$, where

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

Main Result

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

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

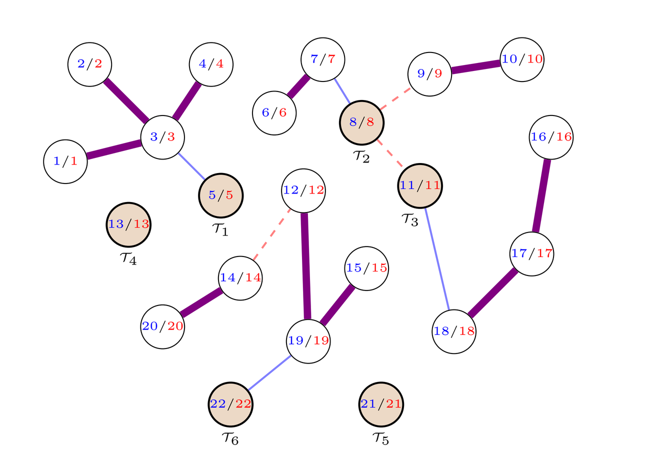
where $c(\mu)$ is the greatest non-negative solution to the equation $e^{-\mu x} = 1 - x$.

Intuition: $c(\lambda s) \to$ typical fraction of nodes that lie in the giant connected component of the intersection graph (with the correct alignment) $\mathcal{G} \wedge \mathcal{G}'$. Outside this component, all other components in $\mathcal{G} \wedge \mathcal{G}'$ are small trees that cannot be correcly aligned.

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

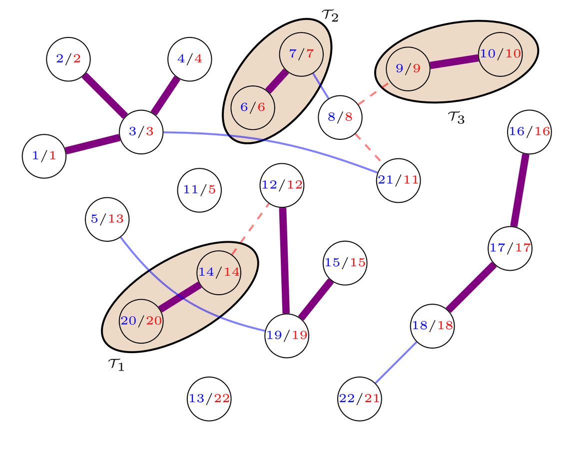
Proof sketch: corrupting the ground truth

Corruption procedure: In $\mathcal{G} \wedge \mathcal{G}'$, for all 'small' tree $\mathbf{T}$, shuffle at random all copies of $\mathbf{T}$ $\to$ blockwise construction of a corrupted version σ of the ground truth s.t. $\mathcal{G}^\sigma \wedge \mathcal{G}'$ 'looks like' $\mathcal{G} \wedge \mathcal{G}'$.



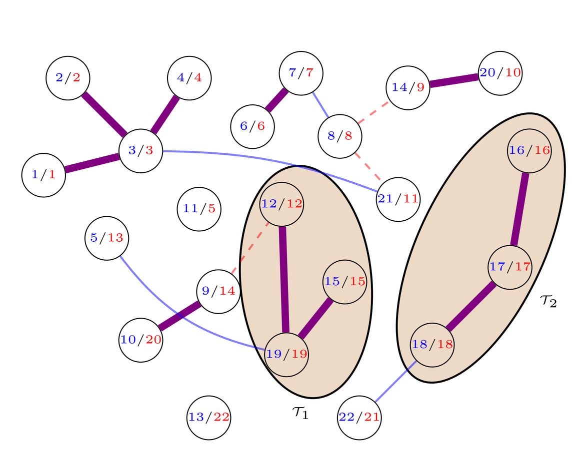
$\Sigma_{\mathbf{T}_1} = (2)(1\ 3\ 5\ 6\ 4)$

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



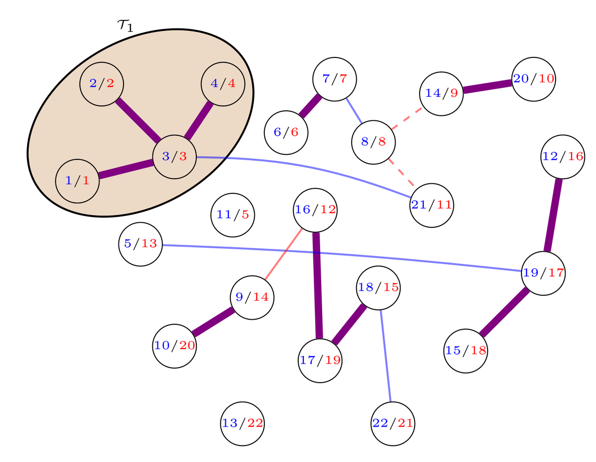
$\Sigma_{\mathbf{T}_2} = (2)(1\ 3)$

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



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

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



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

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

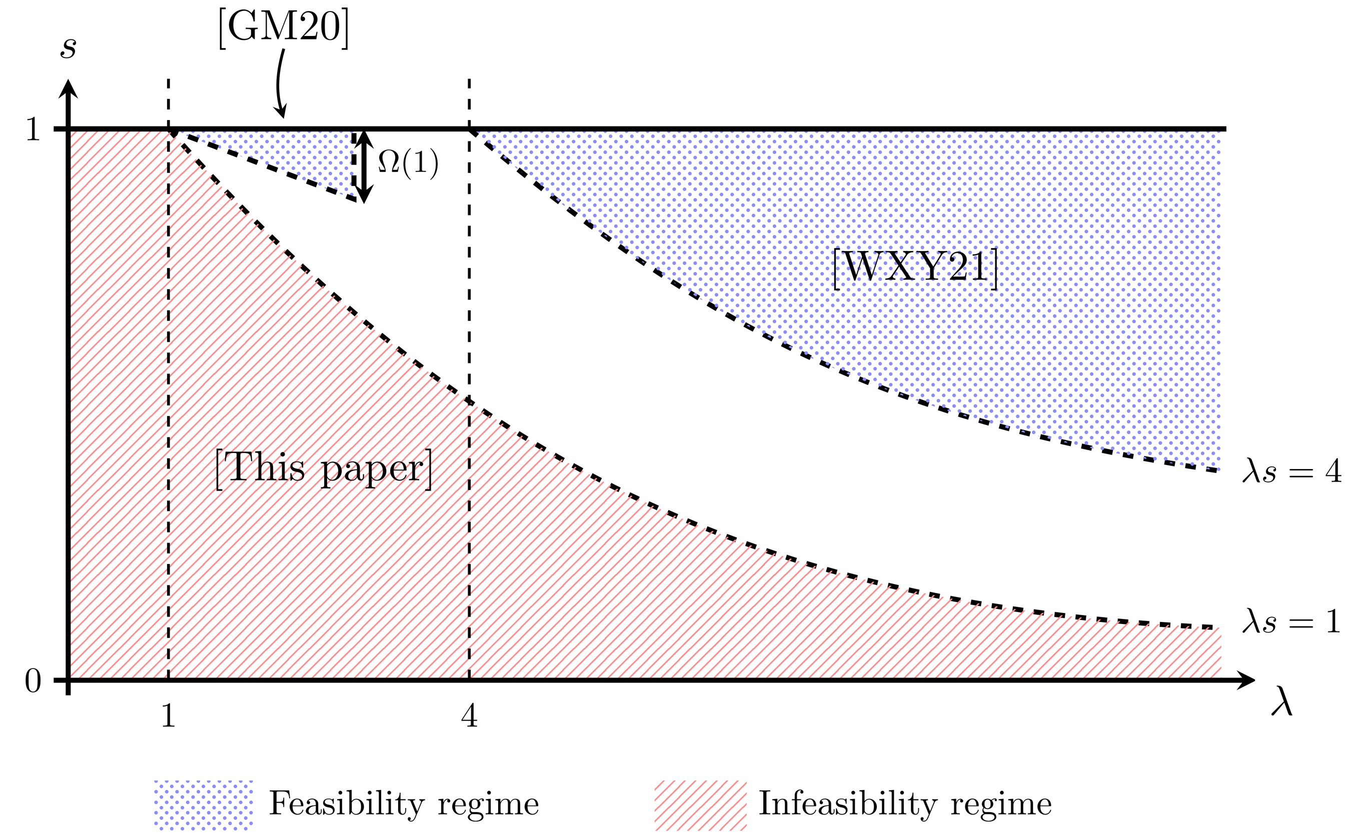
Theorem 2. *Fix an integer $p > 0$. Consider $(\mathcal{G}, \mathcal{G}')$ drawn under the correlated Erdős-Rényi model. Then, with high probability, there exists $\{\sigma_i\}_{i \in [p]}$ – that depend on the intersection graph $\mathcal{G} \wedge \mathcal{G}'$ – such that*

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

(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 $i, j \in [p]$. (far apart corruptions)*

Theorem 1 easily follows from Theorem 2.
Proof of Theorem 2: *(ii)* $\to$ standard random permutation arguments. *(i)* $\to$ Poisson approximation and probabilistic method.

Diagram



Conjecture: $\lambda s = 1$ is the sharp IT-threshold. Existence of hard phase is still open.

Main References

[WXY21] Yihong Wu, Jiaming Xu, and Sophie H. Yu. Settling the sharp reconstruction thresholds of random graph matching, *arxiv preprint*, 2021.

[GM20] Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. volume 125 of *Proceedings of Machine Learning Research*, pages 1633–1665. PMLR, 09–12 Jul 2020.