Impossibility of Partial Recovery in the Graph Alignment Problem

Luca Ganassali*, Marc Lelarge†, Laurent Massoulié‡

February 5, 2021

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 NP-hard 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.

1 Introduction

Graph alignment, also known as graph matching, aims at finding a bijective mapping between the vertex sets of two graphs so that the number of adjacency disagreements between the two graphs is minimized. It reduces to the graph isomorphism problem in the noiseless setting where the two graphs can be matched perfectly. The paradigm of graph alignment has found numerous applications across a variety of diverse fields, such as network privacy ([NS08]), computational biology ([SXB08]), computer vision ([CFVS04]), and natural language processing.

Given two graphs with adjacency matrices A and B, the graph matching problem can be viewed as a special case of the quadratic assignment problem (QAP) ([PRW94]):

$$\max_{\Pi} \langle A, \Pi B \Pi^T \rangle \tag{1}$$

where Π ranges over all $n \times n$ permutation matrices, and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product. QAP is NP-hard in general. These hardness results are applicable in the worst case, where the observed graphs are designed by an adversary. In many applications, the graphs can be modeled by random graphs; as such, our focus is not in the worst-case instances, but rather in recovering partially the underlying vertex permutation with high probability.

Correlated Erdős-Rényi model Driven by applications in social networks and biology, a recent line of work ([LFP14, FQM$^+$16, CK17, MX18, CKMP18, DMWX18, CMK18, FMWX19a, GLM19, WXY20, Gan20, FMWX19b, GM20, WXY21]) initiated the statistical analysis of graph matching by assuming that matrices A and B are generated randomly. The simplest such model is the following *correlated Erdős-Rényi model*: we are given two graphs $\mathcal{G}$ and $\mathcal{G}'$ with the same set of nodes $[n]$ and with respectively blue and red edges. The blue and red edges are obtained by sampling uniformly at random:

- with probability $\lambda s/n$ to get two-colored edges;

*INRIA, DI/ENS, PSL Research University, Paris, France. Email: luca.ganassali@inria.fr

†INRIA, DI/ENS, PSL Research University, Paris, France. Email: marc.lelarge@ens.fr

‡MSR-INRIA Joint Centre, INRIA, DI/ENS, PSL Research University, Paris, France. Email: laurent.massoulie@inria.fr

- with probability $\lambda(1-s)/n$ to get a blue (monochromatic) edge;
- with probability $\lambda(1-s)/n$ to get a red (monochromatic) edge;
- with probability $1-\lambda(2-s)/n$ to get a non-edge,

where $\lambda > 0$ and $s \in [0,1]$ are fixed parameters when n will tend to infinity. Hence each $\mathcal{G}$ and $\mathcal{G}'$ is a sparse Erdős-Rényi model with edge probability λ/n. For large values of n, the the fraction of edges of $\mathcal{G}$ (resp. $\mathcal{G}'$) that are shared with $\mathcal{G}'$ (resp. $\mathcal{G}$) is. of order s (see Figure 1).

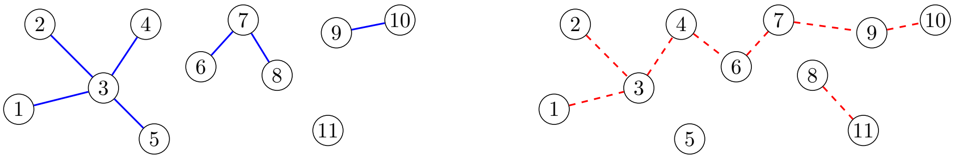


Figure 1: A realization of $(\mathcal{G},\mathcal{G}')$ from the correlated Erdős-Rényi model, with $n=11$, $\lambda=1.9$, and $s=0.7$. For the sake of readability, red edges are always dashed.

We then relabel the vertices of the red graph $\mathcal{G}'$ with an uniform independent permutation $\pi^* \in \mathcal{S}_n$, and we observe $\mathcal{G}$ and $\mathcal{H} := \mathcal{G}'^{\pi^*}$, see Figure 2. Upon observing $\mathcal{G}$ and $\mathcal{H}$, the goal is to recover (or, reconstruct) partially the latent vertex correspondence π^* with probability converging to 1 as $n \to \infty$.

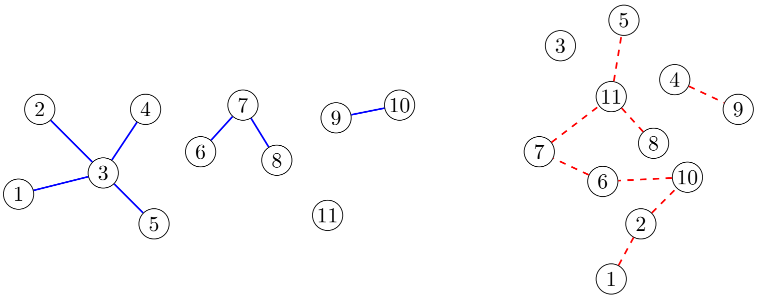


Figure 2: The pair $(\mathcal{G},\mathcal{H})$ corresponding to $(\mathcal{G},\mathcal{G}')$ of Figure 1, after relabeling $\mathcal{G}'$ with the permutation $\pi^* = (6)(1\ 5\ 3\ 11\ 9\ 2\ 8\ 4\ 7\ 10)$.

Partial alignment in the sparse regime We now define our notion of performance. First note that since we are in the sparse regime, even without any noise, i.e. with $s=1$, there is no way to be able to map the $\Theta(n)$ isolated vertices in $\mathcal{G}$ and $\mathcal{H}$ better than chance. Hence, we concentrate on the partial alignment problem where we ask for the best possible fraction of matched vertices between $\mathcal{G}$ and $\mathcal{H}$. More formally, for any $\sigma, \sigma' \in \mathcal{S}_n$, we define *the overlap of σ and σ'* by

$$\mathrm{ov}(\sigma,\sigma') := \sum_{i=1}^{n} \mathbf{1}_{\sigma(i)=\sigma'(i)}. \tag{2}$$

An *estimator* $\hat{\pi}$ is a $\mathcal{S}_n$-valued measurable function of $(\mathcal{G},\mathcal{H})$. Partial alignment thus consists in finding a estimator $\hat{\pi}$ of π^* satisfying $\mathrm{ov}(\hat{\pi},\pi^*) > \alpha n$ with high probability, for some $\alpha > 0$. Let us start by stating a conjecture:

Conjecture.

(*i*) *If $\lambda s \leq 1$, partial reconstruction is impossible, i.e. for any $\alpha > 0$, for all estimator $\hat{\pi}$,*

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

(*ii*) *If $\lambda s > 1$, partial reconstruction is possible (feasible), i.e. there exists $\alpha > 0$ and an estimator $\hat{\pi}$ such that*

$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi},\pi^*) > \alpha n\right) \underset{n\to\infty}{\longrightarrow} 1.$$

Results in the regime with constant mean degree and correlation In this paper, we work in the regime where $\lambda > 0$ and $s \in [0,1]$ are fixed constants. Our results prove part (i) of the conjecture, which had not been previously studied, and give an upper bound on the maximal reachable overlap in case (ii). Let us mention straightaway the related results in our regime that are helpful for our conjecture: [GM20] proves that partial recovery is possible (in polynomial time) in a region $\mathcal{R} := \{(\lambda, s);\ \lambda \in [1, \lambda_0) \text{ and } s \in (s^*(\lambda), 1]\}$ for some function $s^*(\lambda) < 1$, so that interestingly the case $\lambda > \lambda_0$ is left open, nevertheless much in step with (ii). Previous results from [HM20] showed that partial reconstruction was feasible for $\lambda s > C$, with an unspecified constant $C > 20$. At the very time when this paper is being finished, new results from [WXY21] are significantly improving these results, narrowing down the gap for (ii). When translated with our notations, it is shown that partial alignment is possible if $\lambda s \geq 4+\varepsilon$. These results are summed up in a diagram in Figure 3. In particular, our bound is tight and our conjecture is almost solved for the case $s = 1$, with a remaining gap $[\lambda_0, 4]$ being still open.

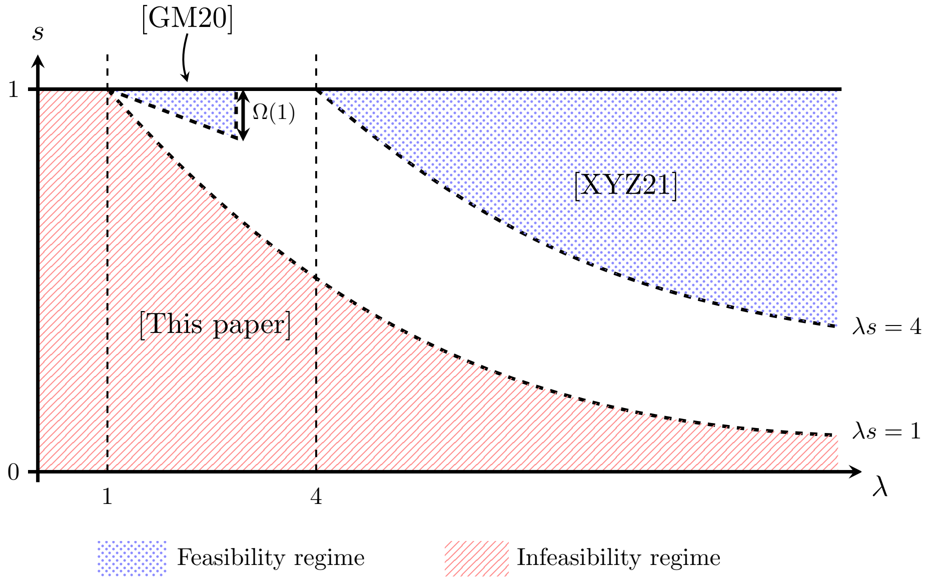


Figure 3: Diagram of the (λ, s) regions where partial reconstruction is known to be impossible (resp. possible), in the sparse regime where λ, s are fixed constants.

Main result Let us now mention our main result. An *equivariant estimator* is an estimator $\hat{\pi}$ such that for any $(\mathcal{G}, \mathcal{H})$, we have

$$\hat{\pi}(\mathcal{G}^\sigma, \mathcal{H}) = \sigma^{-1} \circ \hat{\pi}(\mathcal{G}, \mathcal{H}).$$

Note that the restriction to equivariant estimators is very natural, and not restrictive since the maximum a posteriori estimator $\hat{\pi}_{\mathrm{MAP}}$, which is the permutation solving maximization problem (1), can easily be checked to be equivariant. The main result of our paper is as follows:

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

$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi^*) > (c(\lambda s) + \alpha)n\right) \underset{n \to \infty}{\longrightarrow} 0, \tag{3}$$

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

Note that a well-known result (see e.g. [Bol01]) is that $c(\mu)$ is the typical typical fraction of nodes in the largest component of an Erdős-Rényi graph with average degree μ, and that $c(\mu) = 0$ is $\mu \leq 1$, and $c(\mu) \in (0,1)$ whenever $\mu > 1$. Hence, Theorem 1 implies that partial reconstruction is impossible for $\lambda s \leq 1$ with equivariant estimators. Moreover, if $\lambda s > 1$, any equivariant estimator can reach an overlap of at most $c(\lambda s)n + o(n)$. Note that $c(\lambda s)$ is the typical fraction of nodes in the largest component of the intersection graph.

Further related work Graph matching has also been widely studied in the case where the mean degree and correlation are not fixed constants. The model is the same, with adapted notations: the probability for two-colored (resp. monochromatic, non-) edges are now qs (resp. $q(1-s)$, $1-q(2-s)$), with $s \geq q$ and q and s that may depend on n. Note that our study focuses on the sparse setting where $q = \lambda/n$ and s is constant, but lots of interesting results in other regimes are known for partial, almost exact, and exact recovery. We hereafter give them in detail, for completeness.

- For exact recovery (when $\mathrm{ov}(\hat{\pi}, \pi^*) = n$), [CK17] shows that under some additional conditions on the sparsity of the graph, if $\lambda s - \log n \to \infty$ then exact recovery is possible whereas if $\lambda s - \log n \to -\infty$, it is not possible. Very recent work [WXY21] improves this result, showing more precisely that the tight threshold for exact recovery, is at

$$\frac{nq\left(\sqrt{s}-\sqrt{q}\right)^2}{\log n} = 1,$$

  with the only condition that[1] q/s is bounded away from 1.

- For almost exact recovery (i.e. $\mathrm{ov}(\hat{\pi}, \pi^*) \geq (1-\varepsilon)n$ for all $\varepsilon > 0$), in a sparse regime where $q/s = n^{-\Omega(1)}$, [CKMP18] shows that almost exact recovery is possible if and only if $nqs \to \infty$. In a denser regime where $q/s = n^{-o(1)}$, [WXY21] shows that there is a tight threshold exhibiting an "all-or-nothing" phenomenon at

$$\frac{nq\left(s\log\frac{s}{q} - (s-q)\right)}{\log n} = 2,$$

  above which almost exact recovery is possible and below which even partial recovery is impossible.

- For partial recovery, the first investigation made in [HM20] – though rather difficult to translate in our model – showed that $nqs \to 0$ is an impossibility condition, whereas $nqs \geq C$ (with a large, non-explicit constant C), together with some additional sparsity constraints, ensures feasibility. As mentioned, [WXY21] improves these results, showing that in the case $q/s = n^{-\Omega(1)}$, $nqs \geq 4+\varepsilon$ suffices to ensure possibility. In addition, an impossibility condition of the form $nqs \leq 1-\varepsilon$ is also established, but in a denser case, where $q/s = \omega(\log^2 n)$. Note that this last impossibility result does not cover our case, where both the mean degree nq and the correlation parameter s is of order 1.

For the impossibility part, [WXY21] works with the mutual information $I(\pi^*; \mathcal{G}, \mathcal{G}')$, closely related to the minimum mean squared error. They are able to derive an upper bound on the expectation of $\mathrm{ov}(\hat{\pi}, \pi^*)$, for any estimator, which happens to be $o(1)$ when the mean degree in the parent graph of $\mathcal{G}$ and $\mathcal{G}'$ is at least of order $\log^2 n$, but not when λ, s are of order 1. In our result, we do not work directly with the mutual information, but we are considering the posterior distribution of π^*: in simple words, we show that under the assumption $\lambda s < 1$ the posterior distribution puts equal weights on permutations that overlap only on a vanishing fraction of points. This is done by building ad hoc permutations with the probabilistic method.

In this paper, we derive information theoretic results : our proof is not constructive, i.e. not related to a particular algorithm. The search for efficient algorithms is a very active field of research: using spectral methods ([FQM$^+$16, FMWX19b]), degree profiles ([DMWX18]), convex relaxation ([DML17]), etc. Unfortunately, except from [GM20], these algorithms are not known to give a positive fraction of overlap in the regime $\lambda s \geq 1$, hence leaving the question of the tightness of our bound open.

[1] q/s is often referred to as the mean degree in the *parent graph* of $\mathcal{G}, \mathcal{G}'$. Indeed, another common way of generating the two graphs under the correlated Erdős-Rényi model is to consider $\mathcal{F}$ a parent Erdős-Rényi graph of n nodes and mean degree q/s, and perform two independent sub-samplings of $\mathcal{F}$, keeping each edge independently with probability s, forming $\mathcal{G}$ and $\mathcal{G}'$, two correlated Erdős-Rényi graphs of mean degree q.

2 Main results and global intuition

2.1 Some definitions

Throughout the paper, some proposition A_n is said to be true *with high probability* (w.h.p.) if $\mathbb{P}(A_n) \to 1$ when $n \to \infty$.

Finite sets, permutations For all $n > 0$, we define $[n] := \{1, 2, \ldots, n\}$. For any finite set $\mathcal{X}$, we denote by $|\mathcal{X}|$ its cardinal. $\mathcal{S}_{\mathcal{X}}$ is the set of permutations on $\mathcal{X}$. We also denote $\mathcal{S}_k = \mathcal{S}_{[k]}$ for brevity, and we will often identify $\mathcal{S}_k$ to $\mathcal{S}_{\mathcal{X}}$ whenever $|\mathcal{X}| = k$.

Graphs Through all the paper, we will implicitly consider that every graph G of size n has the canonical vertex set $[n]$. We will denote by $E(G)$ its edge set and $e(G)$ its number of edges.

For any pair of graphs (G, G'), both labeled on $[n]$, we denote by $G \vee G'$ (resp. by $G \wedge G'$) the union graph (resp. intersection graph) of G and G'. The symmetric difference of G and G', denoted by $G \triangle G'$, is the subgraph made of edges of $G \vee G'$ that are not in $G \wedge G'$.

In the case where edges are colored (say G is blue and G' is red), these definitions extend to ensure colour preservation: note e.g. that in this case $G \wedge G'$ is simply the subgraph of $G \vee G'$ consisting of two-colored edges (see Figure 4).

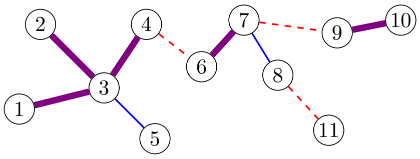


Figure 4: The graph $\mathcal{G} \vee \mathcal{G}'$ with $(\mathcal{G}, \mathcal{G}')$ of Figure 1. For the sake of readability, the two-colored edges of $\mathcal{G} \wedge \mathcal{G}'$ are always drawn thick and purple.

When the pair $(\mathcal{G}, \mathcal{G}')$ is drawn under the correlated Erdős-Rényi model, for all $u, v \in [n]$, we write $u \longleftrightarrow v$ (resp. $u \longleftrightarrow v$) if u and v are connected in $\mathcal{G}$, that is the edge is either blue or two-colored (resp. in $\mathcal{G}'$, either red or two-colored).

For G a graph with vertex set $[n]$ and $\sigma \in \mathcal{S}_n$, we denote by G^σ the *relabeling of G with σ*, which is the graph with same vertex set $[n]$ and edges $(\sigma(u), \sigma(v))$ for all $(u, v) \in E(G)$.

Finally we recall the definition of $c(\mu)$: for all $\mu > 0$, $c(\mu)$ is the greatest non-negative solution to the equation $e^{-\mu x} = 1 - x$. We also recall the fact that for $\mu \leq 1$, $c(\mu) = 0$.

2.2 General intuition on main result

Let us describe the general intuition for our result : recall that we are given $(\mathcal{G}, \mathcal{H})$ drawn under the correlated Erdős-Rényi model with planted relabeling π^*. The idea of the argument for impossibility is to show that, there are w.h.p. lots of permutations that have the same weight for the posterior distribution of π^* given $\mathcal{G}, \mathcal{H}$, and that are far apart. In other words, an informal statement is as follows :

(Informal Statement). *We want to show that there exists lots of relabelings $\mathcal{G}^{\sigma_i}$ of $\mathcal{G}$ such that:*

(i) *There is no way of deciding (statistically) whether the two graphs we observe are $(\mathcal{G}, \mathcal{G}')$ or some $(\mathcal{G}^{\sigma_i}, \mathcal{G}')$.*

(ii) *These relabelings are far apart from each other and small components of $\mathcal{G} \wedge \mathcal{G}'$.*

Let us give a formal version of the previous intuition. First note that for any labeled graphs G, G' on $[n]$:

$$\mathbb{P}(\mathcal{G} = G, \mathcal{G}' = G') = \left(\frac{\lambda s}{n}\right)^{e(G \wedge G')} \left(\frac{\lambda(1-s)}{n}\right)^{e(G \triangle G')} \left(1 - \frac{\lambda(2-s)}{n}\right)^{\binom{n}{2} - e(G \vee G')}.$$

Since

$$e(G \vee G') = e(G) + e(G') - e(G \wedge G') \quad \text{and} \quad e(G \triangle G') = e(G \vee G') - e(G \wedge G'),$$

$\mathbb{P}(\mathcal{G} = G, \mathcal{G}' = G')$ is uniquely determined by $e(G), e(G')$ and $e(G \wedge G')$. In particular, the dependence of the joint distribution in $e(G \wedge G')$ is given by:

$$\mathbb{P}(\mathcal{G} = G, \mathcal{G}' = G') \propto \left[\frac{s(n - \lambda(2-s))}{\lambda(1-s)^2}\right]^{e(G \wedge G')}. \tag{4}$$

In view of (4), preserving the posterior distribution by relabeling a graph $\mathcal{G}$ is simply preserving the number of edges of their intersection graph. We now have a formal rephrasing for our conditions (i) and (ii) above: we encapsulate them in a theorem, which will constitute the bulk of our paper.

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

(ii) $\forall i, j \in [p],\ i \neq j \implies \mathrm{ov}(\sigma_i, \sigma_j) \leq c(\lambda s)n + o(n)$, *where the $o(n)$ is independent of $i, j \in [p]$.*

Let us now explain how Theorem 2 implies our impossibility result via a simple pigeonhole principle.

Proof of Theorem 1. Let us take $\alpha > 0$. We want to control the probability that the overlap between an estimator $\hat{\pi}$ and π^* is greater than $\alpha n + c(\lambda s)n$. Fix $\varepsilon > 0$, and take p large enough so that

$$\alpha \varepsilon p > 2.$$

First note that point (i) together with (4) gives that the joint probability of $(\mathcal{G}, \mathcal{G}', \pi^*)$ is is equal to that of $(\mathcal{G}^{\sigma_i}, \mathcal{G}', \pi^*)$, for all $i \in [p]$. Thus, for all estimator $\hat{\pi}$ depending on $\mathcal{G}, \mathcal{H} = \mathcal{G}'^{\pi^*}$, one has

$$\forall i \in [p],\ \mathrm{ov}\left(\hat{\pi}(\mathcal{G}^{\sigma_i}, \mathcal{H}), \pi^*\right) \overset{(d)}{=} \mathrm{ov}\left(\hat{\pi}(\mathcal{G}, \mathcal{H}), \pi^*\right). \tag{5}$$

Then, since $\hat{\pi}$ is equivariant by assumption,

$$\forall i \in [p],\ \mathrm{ov}\left(\hat{\pi}(\mathcal{G}, \mathcal{H}), \pi^*\right) = \mathrm{ov}\left({\sigma_i}^{-1} \circ \hat{\pi}(\mathcal{G}, \mathcal{H}), \pi^*\right).$$

Let

$$X := \sum_{i \in [p]} \mathbf{1}_{\mathrm{ov}({\sigma_i}^{-1} \circ \hat{\pi}, \pi) > (c(\lambda s) + \alpha)n} = \sum_{i \in [p]} \mathbf{1}_{\mathrm{ov}(\hat{\pi}, \sigma_i \circ \pi) > (c(\lambda s) + \alpha)n}$$

Note that because of point (ii), all overlaps $\mathrm{ov}(\sigma_i \circ \pi, \sigma_j \circ \pi)$ are less than $c(\lambda s)n + o(n)$ for $i \neq j \in [p]$. Thus, there are at least $X \times (\alpha - o(1))n$ distinct points among the node set $[n]$. This gives that one necessarily has

$$X \leq \frac{1}{\alpha - o(1)}. \tag{6}$$

Then, taking the expectation and considering the event on which the set $\{\sigma_i\}_{i \in [p]}$ of Theorem 2 exists – which happens with probability $1 - o(1)$ – gives

$$\mathbb{E}[X] \geq \sum_{i=1}^{p} \mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \sigma_i \circ \pi) > (c(\lambda s) + \alpha)n\right) - p \times o(1) = p \times \mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi) > (c(\lambda s) + \alpha)n\right) - o(1). \tag{7}$$

Hence,

$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi) > (c(\lambda s) + \alpha)n\right) \leq \frac{1}{p(\alpha - o(1))} + o(1). \tag{8}$$

For n large enough, the right-hand side of the last term is less that $\frac{1}{p(\alpha/2)}$, which is less than ε. This proves as desired that for all $\alpha > 0$

$$\mathbb{P}\left(\mathrm{ov}(\hat{\pi}, \pi) > (c(\lambda s) + \alpha)n\right) \underset{n \to \infty}{\longrightarrow} 0. \tag{9}$$

□

We are now left to understand how to build ad hoc permutations verifying points (i) and (ii) of Theorem 2. In order to build these permutations, we are going to relabel the vertices on small tree components of the intersection graph $\mathcal{G} \wedge \mathcal{G}'$. As a first step, we hereafter check that they indeed nearly cover the whole graph, when letting aside the giant component.

2.3 Vertices on small tree components

We briefly recall the definition of the simple Erdős-Rényi model $G(n,p)$: it consist in drawing a (single) graph with node set $[n]$ in which every edge is independently present with probability p. Let us begin with a classical result:

Lemma 2.1 ([Bol01], Corollary 5.8, Theorem 6.11)**.** *Let $G \sim G(n, \mu/n)$ with $\mu > 0$, and $a_n \to \infty$. Then, with high probability, G has a giant component of order $c(\mu)n + o(n)$ and outside the giant component, at least $(1 - c(\mu))n - a_n$ vertices are on tree components.*

We need here a slight adaptation of this result, showing that $(1 - c(\mu))n - o(n)$ vertices are in fact on *small* tree components.

Lemma 2.2. *Let $G \sim G(n, \mu/n)$ with $\mu > 0$, and $K(n) \to \infty$. Then with high probability, $1 - c(\mu))n - o(n)$ vertices are on tree components of size at most $K(n)$.*

Proof. Assume without loss of generality that $K(n) = o(\log n)$. Let $T_>$ be the number of vertices that are on tree components of size $\geq K(n)$. Taking $a_n = o(n)$ in Lemma 2.1, it remains to show that w.h.p., $T_> = o(n)$. This is done easily by bounding very roughly the first moment. Another classical result (see e.g. [JLR00], Theorem 5.4) is that with probability $1 - o(1)$, all tree components are of size $O(\log n)$, which gives

$$\begin{aligned}\frac{\mathbb{E}\left[T_>\right]}{n} &\leq o(1) + \sum_{k=K(n)}^{O(\log n)} \frac{1}{n} \cdot k \cdot \binom{n}{k} k^{k-2} \left(\frac{\mu}{n}\right)^{k-1} \left(1 - \frac{\mu}{n}\right)^{k(n-k)+\binom{k}{2}-k+1} \\ &\leq o(1) + (1 + o(1)) \sum_{k=K(n)}^{O(\log n)} \frac{e^k}{k} \mu^{k-1} e^{-k\mu},\end{aligned}$$

using $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ together with Cayley's formula[2] and the fact that for all indices $K(n) \leq k \leq O(\log n)$ in the sum, $k^2 \leq o(n)$ (uniformly). Now, the series in the right hand term has general terms which is $O\left(e^{-k(\mu - \log \mu + 1)}\right)$, and since $\mu - \log \mu + 1 > 0$ the series converges, which implies that $\mathbb{E}\left[T_> / n\right] = o(1)$. The proof is concluded by Markov's inequality. □

Since in our model $\mathcal{G} \wedge \mathcal{G}'$ is an Erdős-Rényi graph of parameters $(n, \lambda s/n)$, the previous results ensures that all but a vanishing part of the $(1 - c(\mu))n$ vertices outside the giant component are on small (i.e. $\leq K(n)$) tree components of the intersection graph. For the rest of the paper, we will take

$$K(n) = \lfloor \sqrt{\log n} \rfloor.$$

This first step suggests to build the permutations (relabelings) only by looking at $\mathcal{G} \wedge \mathcal{G}'$. Hence, we will first consider the random generation of the intersection graph, then create some permutations σ_i, and finally reveal the monochromatic edges.

The generating process is as follows: since almost all $(1 - c(\mu))n$ vertices are on small trees in $\mathcal{G} \wedge \mathcal{G}'$, we can prove that each small tree up to isomorphism will have a number of occurrences in the intersection graph of order n (this is claimed more precisely in Lemma 3.1). Permuting iteratively these isomorphic trees, we may derange them quite a lot, and each time differently.

In order to prove Theorem 2, we use the *probabilistic method*[3]: we give in the next section a simple detailed stochastic method to build p permutation candidates, and we will next prove that these permutations satisfy conditions (i) and (ii) with positive probability, hence proving the desired existence.

[2] Cayley's formula states that the number of trees on k labeled vertices is k^{k-2}.

[3] The main interest of this widely used method (see [AS16]) is to be non-constructive. Indeed, as detailed in the next Sections, explicitly giving the p permutations considered in Theorem 2 is very cumbersome, because of the extra double edges that may appear (see Section 3.3).

3 Building automorphisms of $\mathcal{G} \wedge \mathcal{G}'$ tree-wise

Through all this section, we work conditionally on the intersection graph $\mathcal{G} \wedge \mathcal{G}'$ (that is the two-colored edges).

3.1 Mathematical formalization

Recall that we fix $K := K(n) = \lfloor \sqrt{\log n} \rfloor$. For all $k \in [K]$, we will denote by $\mathbb{T}_k$ the set of *unlabeled* trees of size k. $\mathbb{T}_k$ can also be viewed as the set of equivalence classes of labeled trees of size k for the isomorphism relation. Note that $\mathbb{T}_k$ is finite and that we can roughly upper bound its size by the number of *labeled* trees of size k which equals k^{k-2}, by Cayley's formula[4].

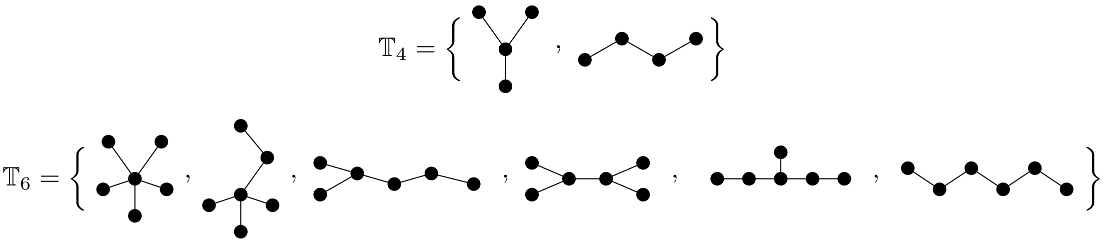

$\mathbb{T}_4 = \{$, $\}$
$\mathbb{T}_6 = \{$, , , , , $\}$

Figure 5: Explicit composition of $\mathbb{T}_4$ (of size 2) and $\mathbb{T}_6$ (of size 6).

For a given tree $\mathbf{T} \in \mathbb{T}_k$, we will denote by $X_{\mathbf{T}}$ the number of distinct connected components of $\mathcal{G} \wedge \mathcal{G}'$ that are isomorphic to $\mathbf{T}$, $H_{\mathbf{T}} := \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{X_{\mathbf{T}}}\}$ the set of the corresponding labeled subgraphs of $\mathcal{G} \wedge \mathcal{G}'$, and $V(H_{\mathbf{T}})$ the set of vertices of $[n]$ that belong to one of the trees in $H_{\mathbf{T}}$.

Our global recursion will be done on the finite set

$$\mathbb{T} := \bigcup_{k=1}^{K} \mathbb{T}_k = \{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_M\}, \tag{10}$$

which we assume to have been ordered increasingly according to tree sizes, for convenience. The global permutation σ is built block-wise by composing permutations $\sigma_{\mathbf{T}}$ for $\mathbf{T} \in \mathbb{T}$ such that each $\sigma_{\mathbf{T}}$ only acts on vertices of $H_{\mathbf{T}}$.

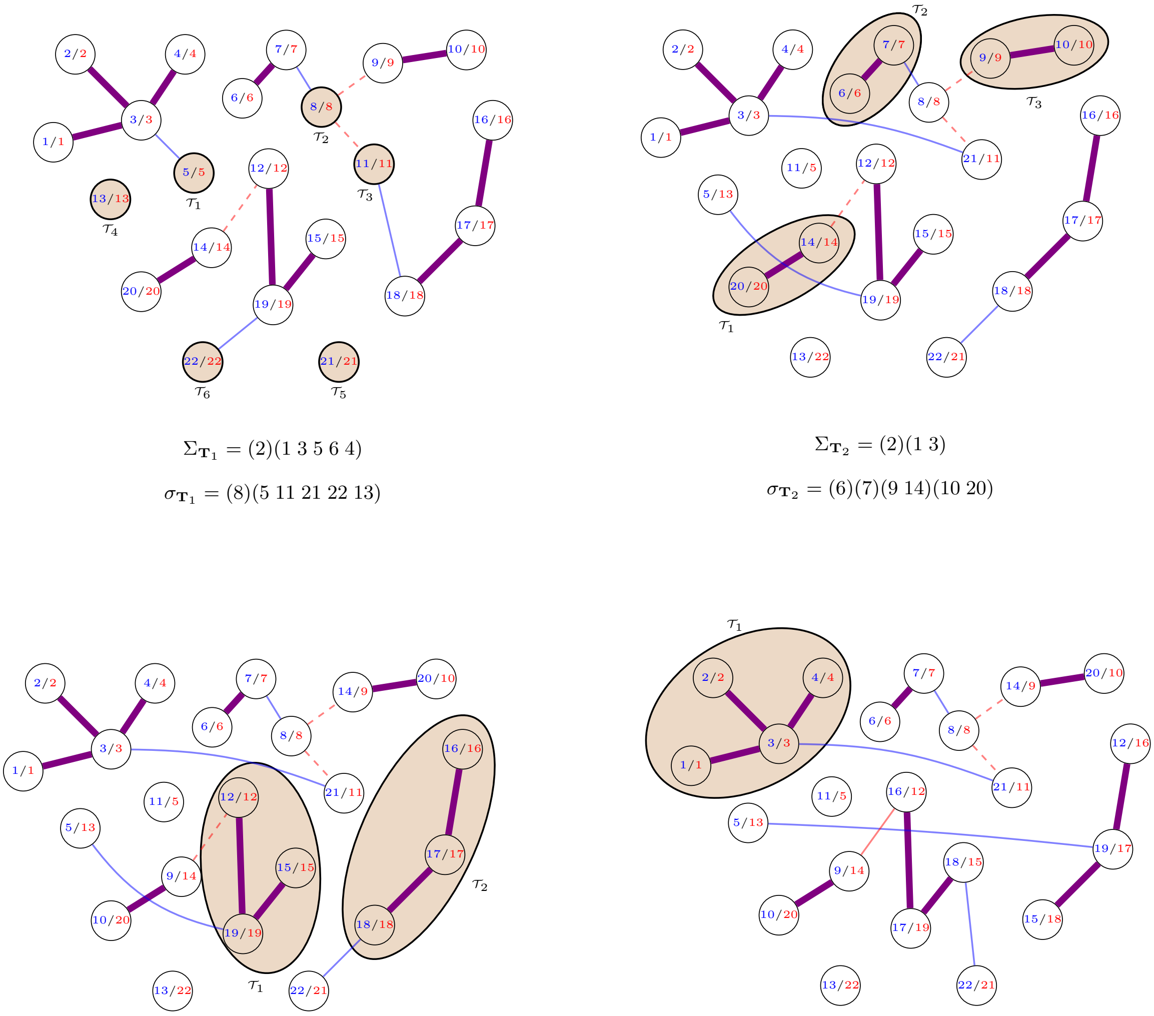
More precisely, for a fixed $\mathbf{T} \in \mathbb{T}$, $\sigma_{\mathbf{T}}$ will consists in permuting the vertices tree by tree, so $\sigma_{\mathbf{T}}$ will be determined by a tree permutation $\Sigma_{\mathbf{T}}$ of size $X_{\mathbf{T}}$. Assume that for all trees $\mathcal{T}_1, \ldots, \mathcal{T}_{X_{\mathbf{T}}}$ isomorphic to $\mathbf{T}$ in $\mathcal{G} \wedge \mathcal{G}'$, we fix some isomorphisms $\psi_1, \ldots, \psi_{X_{\mathbf{T}}}$ such that $\mathcal{T}_i \underset{\psi_i}{\sim} \mathbf{T}$ for all $i \in [X_{\mathbf{T}}]$. More generally we will denote $\mathsf{i}(u)$ the index of the tree that $u \in V(H_{\mathbf{T}})$ belongs to (when there is no ambiguity on $\mathbf{T}$), and $u \simeq u'$ when two vertices of $\mathcal{G} \wedge \mathcal{G}'$ are sent onto the same point of $\mathbf{T}$ by these isomorphisms. Then, the natural definition of the node permutation $\sigma_{\mathbf{T}}$ according to $\Sigma_{\mathbf{T}}$ and these isomorphisms is given by

$$\sigma_{\mathbf{T}} : u \mapsto \begin{cases} \psi^{-1}_{\Sigma_{\mathbf{T}}(\mathsf{i}(u))} \circ \psi_{\mathsf{i}(u)}(u) \;\; (\in \mathcal{T}_{\Sigma_{\mathbf{T}}(\mathsf{i}(u))}) & \text{if } u \in V(H_{\mathbf{T}}), \\ u & \text{if } u \notin V(H_{\mathbf{T}}). \end{cases} \tag{11}$$

Note that by definition, $V(H_{\mathbf{T}})$ is stable by $\sigma_{\mathbf{T}}$, and $\sigma_{\mathbf{T}}$ fixes all nodes in $[n] \setminus V(H_{\mathbf{T}})$.

Recall that M denotes the total size of $\mathbb{T}$ as defined in (10). The recursive construction is as

[4] This upper bound is far from being optimal, but is enough for our use.



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

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

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

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

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

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

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

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

Figure 6: Example of recursive (tree-wise) generation of a permutation with Algorithm 1.

follows :

Algorithm 1: Recursive construction of σ

```
Initialize $\sigma_0 \leftarrow$ id;
for $i = 1$ to $M$ do
    Consider $\mathbf{T} = \mathbf{T}_i$;
    Choose 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 $\Sigma_{\mathbf{T}}$ by (11);
    $\sigma_i \leftarrow \sigma_{\mathbf{T}} \circ \sigma_{i-1}$;
end
return $\sigma = \sigma_M$
```

Note that at the end of the procedure, σ fixes all points that are either on the giant component of the intersection graph, or on a component that is not a tree a size $\leq K(n)$.

Figure 6 gives an example of this random recursive construction (for convenience, $\lambda s < 1$; the true labels are in red, whereas blue labels enables to keep track of the relabeling recursively built on the blue graph).

Through the analysis we will need the following control on $X_{\mathbf{T}}$ for $\mathbf{T} \in \mathbb{T}$:

Lemma 3.1. *Recall that $K(n) = \lfloor \sqrt{\log n} \rfloor$. For all $k \in [K(n)]$, define $f(k) := \frac{(\lambda s)^{k-1} e^{-\lambda s k}}{k!}$. Then, with high probability (on the intersection graph),*

$$\forall k \in [K(n)], \, \forall \mathbf{T} \in \mathbb{T}_k, \, X_{\mathbf{T}} \geq n(1 - o(1)) f(k). \tag{12}$$

The proof of this result is deferred to Appendix B.1.

Remark 3.1. *Note that since $\lambda s e^{-\lambda s} < 1$, $k \mapsto f(k)$ is decreasing with k. Moreover, for $K(n) \leq \sqrt{\log n}$, we have that*

$$f(K(n)) \geq \exp\left(-C\sqrt{\log n} \log \log n\right) \gg n^{-t},$$

for any $t > 0$.

3.2 Ensuring that the permutations are 'far apart'

We check in this section that Algorithm 1 generates permutations that will verify condition (ii) of Theorem 2, w.h.p. Let $\sigma_1, \ldots, \sigma_p$ be generated independently with Algorithm 1. We then have the following results:

Lemma 3.2. *With high probability, for all $i \neq j \in [p]$,*

$$\mathrm{ov}(\sigma_i, \sigma_j) = c(\lambda s) n + o(n).$$

This lemma is proved in Appendix B.2. In the sequel we will denote by V_∞ the set of vertices that are on the giant component of $\mathcal{G} \wedge \mathcal{G}'$ (if there is one), and by $V_>$ the vertices of $[n] \setminus V_\infty$ that are *not* on tree components of size $\leq K(n)$. Finally we set $V_{\infty,>} := V_\infty \cup V_>$. Define

$$\mathcal{S}_{in} := \binom{[n] \setminus V_{\infty,>}}{2}, \quad \mathcal{S}_{out} := \binom{[n]}{2} \setminus \left(\binom{V_{\infty,>}}{2} \cap \binom{[n] \setminus V_{\infty,>}}{2} \right), \quad \mathcal{S} := \mathcal{S}_{in} \cup \mathcal{S}_{out}. \tag{13}$$

$\mathcal{S}_{in}$ is the set of edges that have both endpoints outside $V_{\infty,>}$, whereas edges of $\mathcal{S}_{out}$ have exactly one endpoint in $V_{\infty,>}$. We say that an edge $(u,v) \in \binom{[n]}{2}$ is a *common fixed edge* of permutations $\sigma_1, \ldots, \sigma_r$ if

$$\{\sigma_1(u), \sigma_1(v)\} = \ldots = \{\sigma_r(u), \sigma_r(v)\}.$$

For all subset of edges $\mathcal{W} \subseteq \binom{[n]}{2}$, we define

$$F(\mathcal{W}, \sigma_1, \ldots, \sigma_r) := \sum_{e \in \mathcal{W}} \mathbf{1}_{e \text{ is a common fixed edge of } \sigma_1, \ldots, \sigma_r}. \tag{14}$$

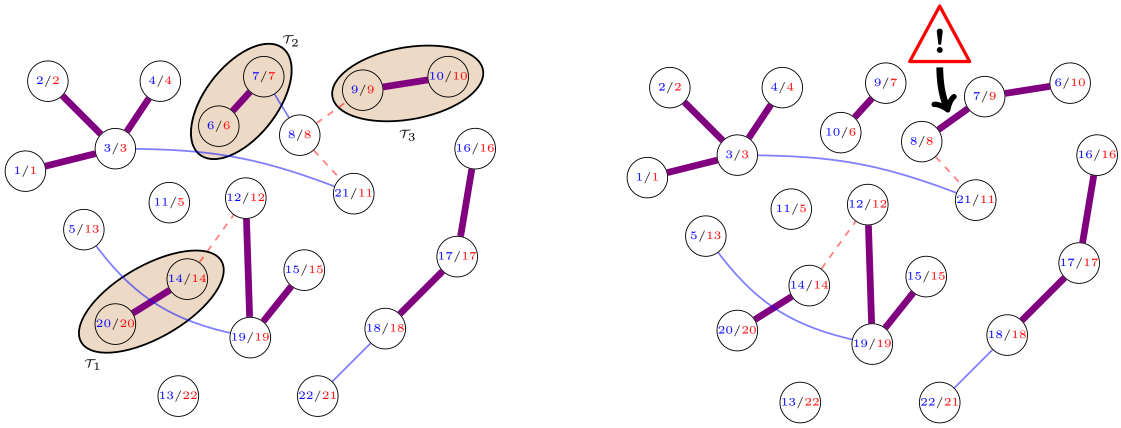
We now state a result – which proof is deferred to B.3 – that will be useful in next section.

Lemma 3.3. *With high probability, we have, for any $t > 0$,*

- *for any $i_1 \neq i_2 \in [p]$,*
$$F(\mathcal{S}, \sigma_{i_1}, \sigma_{i_2}) \leq n^{1+t}, \tag{15}$$
- *for any $i_1, i_2, i_3 \in [p]$ pairwise distinct,*
$$F(\mathcal{S}, \sigma_{i_1}, \sigma_{i_2}, \sigma_{i_3}) \leq n^t, \tag{16}$$
- *for any $r \geq 4$, $i_1, \ldots, i_r \in [p]$ pairwise distinct,*
$$F(\mathcal{S}, \sigma_{i_1}, \ldots, \sigma_{i_r}) = 0. \tag{17}$$

3.3 Emergence of extra double edges

In the example of Figure 6, we can see that the number of two-colored edges in the relabeled union graph $\mathcal{G}^{\sigma_i} \vee \mathcal{G}'$ is constant through time. This property is fundamental for point (i) of Theorem 2. However, depending on the random $\sigma_{\mathbf{T}_i}$ drawn through the process – we recall that they are drawn independently from the monochromatic edges, that are not revealed yet – we may see extra two-colored edges appear (extra double edges hereafter). Figure 7 shows a case in which there is an emergence of an extra double edge in the process.



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

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

Figure 7: Example of the emergence of an extra double edge in Algorithm 1.

Note that the number of two-coloured edges can only be greater or equal to $e(\mathcal{G} \wedge \mathcal{G}')$ through this process, since by definition we are preserving edges of the intersection graph.

The last part of our work is to prove that there is a positive probability that applying independently Alg.1 p times gives p permutations that do not present extra double edges, before using the probabilistic method. This step will require a Poisson approximation, described hereafter.

4 Poisson approximation to avoid extra double edges, proof of Theorem 2.

In this section we introduce n' to be the number of vertices that the permutations actually act on:

$$n' := |[n] \setminus V_{\infty,>}| \sim (1 - c(\lambda s))n \text{ w.h.p.} \tag{18}$$

Then, we assume that we fix a set $\{\sigma_i\}_{i\in[p]}$ of p permutations of $[n']$, verifying :

$$\text{for all } t > 0, \text{for all } m \neq m' \in [p], F(\mathcal{S}, \sigma_m, \sigma_{m'}) \leq n^{1+t}. \tag{H1}$$

$$\text{for all } t > 0, \text{for all } m_1, m_2, m_3 \in [p] \text{ pairwise distinct }, F(\mathcal{S}, \sigma_{m_1}, \sigma_{m_2}, \sigma_{m_3}) \leq n^t. \tag{H2}$$

$$\text{There are no common fixed edge of any } r\text{-tuple in } \{\sigma_i\}_{i\in[p]}. \tag{H3}$$

We will work under the event $\mathcal{E}_{\mathcal{S}}$ on which $n' \sim (1 - c(\lambda s))n$ and $|\mathcal{S}| \sim \binom{n'}{2} \sim n'^2/2 = (1 - c(\lambda s))^2 n^2/2$. It is easy (see e.g. [Bol01]) to show that $\mathcal{E}_{\mathcal{S}}$ is satisfied w.h.p. As explained before, some extra double edges (e.d.e. hereafter) may appear when revealing the non double edges of $\mathcal{S}$ (that is, blue and red edges that are not between vertices of $V_{\infty,>}$). Note that for every edge we have

$$\mathbb{P}\left(u \mathrel{\color{blue}\longleftrightarrow} v \,|\, (u,v) \notin E(\mathcal{G} \wedge \mathcal{G}')\right) = \mathbb{P}\left(u \mathrel{\color{red}\longleftrightarrow} v \,|\, (u,v) \notin E(\mathcal{G} \wedge \mathcal{G}')\right) = \frac{\mathbb{P}\left(u \mathrel{\color{red}\longleftrightarrow} v, (u,v) \notin E(\mathcal{G} \wedge \mathcal{G}')\right)}{\mathbb{P}\left((u,v) \notin E(\mathcal{G} \wedge \mathcal{G}')\right)}$$
$$= \frac{\lambda(1-s)/n}{1-\lambda s/n} \sim \frac{\lambda(1-s)}{n}.$$

For any permutation σ, define the number of created e.d.e. by the relabeling of $\mathcal{G}$ by σ as follows:

$$\Delta(\sigma) := \sum_{\{u,v\} \in \mathcal{S}} \mathbf{1}_{u \mathrel{\color{blue}\longleftrightarrow} v} \mathbf{1}_{\sigma(u) \mathrel{\color{red}\longleftrightarrow} \sigma(v)}. \tag{19}$$

We now present the key result for our analysis, with the notation $n^{\underline{k}}$ for the *falling factorial*

$$n^{\underline{k}} := n(n-1)\cdots(n-k+1).$$

Theorem 3 (Asymptotic Poisson behavior of $\{\Delta(\sigma_i)\}_{i\in[p]}$)**.** *Assume that $\{\sigma_i\}_{i\in[p]}$ verify (H1), (H2) and (H3). Then, for all $\ell_1, \ell_2, \ldots, \ell_p \geq 0$,*

$$\mathbb{E}\left[\Delta(\sigma_1)^{\underline{\ell_1}}\Delta(\sigma_2)^{\underline{\ell_2}}\cdots\Delta(\sigma_p)^{\underline{\ell_p}} \,\middle|\, \mathcal{G} \wedge \mathcal{G}', \mathcal{E}_{\mathcal{S}}\right] \underset{n\to\infty}{\longrightarrow} \left(\frac{\lambda^2(1-s)^2(1-c(\lambda s))^2}{2}\right)^{\ell_1+\ell_2+\ldots+\ell_p}. \tag{20}$$

In other words, conditionally to graph $\mathcal{G} \wedge \mathcal{G}'$ and event $\mathcal{E}_{\mathcal{S}}$, the random variables $\{\Delta(\sigma_i)\}_{i\in[p]}$ are asymptotically distributed as independent Poisson variables of parameter $\frac{\lambda^2(1-s)^2(1-c(\lambda s))^2}{2}$.

The proof of Theorem 3, based on a fine control of terms of unusually high contribution, is deferred to Appendix A.

4.1 Proof of Theorem 2

Proof. The proof is quite straightforward now. Fixing $p > 0$, Lemma 3.3 gives that (H1), (H2) and (H3) are verified w.h.p. by some $\sigma_1, \ldots, \sigma_p$ generated independently with Algorithm 1. Then, the probability (on the remaining monochrome edges) that the p permutations given satisfy conditions (i) and (ii) of Theorem 2 is equivalent to

$$(1-o(1)) \times \mathbb{P}\left(\mathrm{Poi}\left(\frac{\lambda^2(1-s)^2}{2}\right) = 0\right)^p = (1-o(1))\exp\left(-p\frac{\lambda^2(1-s)^2}{2}\right) > 0, \tag{21}$$

which gives the existence with high probability of a set a permutations of size p satisfying conditions (i) and (ii) of Theorem 2. □

References

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016.

[Bol01] Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

[CFVS04] Donatello Conte, Pasquale Foggia, Mario Vento, and Carlo Sansone. Thirty Years Of Graph Matching In Pattern Recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(3):265–298, 2004.

[CK17] Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs, 2017.

[CKMP18] Daniel Cullina, Negar Kiyavash, Prateek Mittal, and H. Vincent Poor. Partial recovery of Erdős-Rényi graph alignment via k-core alignment. *CoRR*, abs/1809.03553, 2018.

[CMK18] Daniel Cullina, P. Mittal, and N. Kiyavash. Fundamental limits of database alignment. *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 651–655, 2018.

[DML17] Nadav Dym, Haggai Maron, and Yaron Lipman. Ds++: A flexible, scalable and provably tight relaxation for matching problems. *arXiv preprint arXiv:1705.06148*, 2017.

[DMWX18] Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *arXiv e-prints*, page arXiv:1811.07821, Nov 2018.

[FMWX19a] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations I: The gaussian model, 2019.

[FMWX19b] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality, 2019.

[FQM$^+$16] Soheil Feizi, Gerald Quon, Mariana Recamonde Mendoza, Muriel Médard, Manolis Kellis, and Ali Jadbabaie. Spectral alignment of networks. *CoRR*, abs/1602.04181, 2016.

[Gan20] Luca Ganassali. Sharp threshold for alignment of graph databases with gaussian weights, 2020.

[GLM19] L. Ganassali, M. Lelarge, and L. Massoulié. Spectral alignment of correlated Gaussian random matrices. *arXiv e-prints*, page arXiv:1912.00231, November 2019.

[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.

[HM20] Georgina Hall and Laurent Massoulié. Partial Recovery in the Graph Alignment Problem. *arXiv e-prints*, page arXiv:2007.00533, July 2020.

[JLR00] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 2000.

[LFP14] Vince Lyzinski, Donniell E. Fishkind, and Carey E. Priebe. Seeded graph matching for correlated erdos-renyi graphs. *Journal of Machine Learning Research*, 15:3693–3720, 2014.

[MX18] Elchanan Mossel and Jiaming Xu. Seeded graph matching via large neighborhood statistics. *CoRR*, abs/1807.10262, 2018.

[NS08] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125, May 2008.

[PRW94] Panos Pardalos, Franz Rendl, and Henry Wolkowicz. *The Quadratic Assignment Problem: A Survey and Recent Developments*, pages 1–42. 08 1994.

[SXB08] Rohit Singh, Jinbo Xu, and Bonnie Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, 2008.

[WXY20] Yihong Wu, Jiaming Xu, and Sophie H. Yu. Testing correlation of unlabeled random graphs. *arXiv e-prints*, page arXiv:2008.10097, August 2020.

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

A Proof of Theorem 3

Proof of Theorem 3. Let $\ell_1, \ell_2, \dots, \ell_p$ be non negative integers. Recall that conditioned to $\mathcal{G} \wedge \mathcal{G}'$, each edge of $\mathcal{S}$ is independently blue (resp. red) with probability

$$q = q(\lambda, s, n) := \frac{\lambda(1-s)}{n - \lambda s}.$$

Now, let us explain why convergence (20) holds. First recall that for a given $\ell \geq 0$, $\mathbb{E}\left[\Delta(\sigma)^{\underline{\ell}}\right]$ is nothing else but the expected number of (ordered) $p-$tuples of edges $\{u, v\} \in \mathcal{S}$ such that $\mathbf{1}_{u \longleftrightarrow v} \mathbf{1}_{\sigma(u) \longleftrightarrow \sigma(v)} = 1$. Using the notation $\sum^*$ for summation of ordered tuples of edges in $\mathcal{S}$ as well as linearity of expectation, we get:

$$\mathbb{E}\left[\Delta(\sigma_1)^{\underline{\ell_1}} \Delta(\sigma_2)^{\underline{\ell_2}} \cdots \Delta(\sigma_p)^{\underline{\ell_p}}\right] =$$

$$\sum^*_{\substack{\{u_1^{(1)}, v_1^{(1)}\},\\ \{u_2^{(1)}, v_2^{(1)}\},\\ \dots,\\ \{u_{\ell_1}^{(1)}, v_{\ell_1}^{(1)}\}}} \sum^*_{\substack{\{u_1^{(2)}, v_1^{(2)}\},\\ \{u_2^{(2)}, v_2^{(2)}\},\\ \dots,\\ \{u_{\ell_2}^{(2)}, v_{\ell_2}^{(2)}\}}} \cdots \sum^*_{\substack{\{u_1^{(p)}, v_1^{(p)}\},\\ \{u_2^{(p)}, v_2^{(p)}\},\\ \dots,\\ \{u_{\ell_p}^{(p)}, v_{\ell_p}^{(p)}\}}} \mathbb{E}\left[\prod_{m=1}^{p} \prod_{j=1}^{\ell_m} \mathbf{1}_{u_j^{(m)} \longleftrightarrow v_j^{(m)}} \mathbf{1}_{\sigma_m(u_j^{(m)}) \longleftrightarrow \sigma_m(v_j^{(m)})}\right] \tag{22}$$

First observe that the total number of terms N in the previous sum is

$$N := |\mathcal{S}|^{\underline{\ell_1}} \times |\mathcal{S}|^{\underline{\ell_2}} \times \cdots |\mathcal{S}|^{\underline{\ell_p}} \sim \left(\frac{(1 - c(\lambda s))^2 n^2}{2}\right)^{\ell_1 + \ldots + \ell_p},$$

since $|\mathcal{S}| \sim \frac{(1-c(\lambda s))^2 n^2}{2}$ on event $\mathcal{E}_{\mathcal{S}}$.

Lower bound: Observe that the N terms in the sum of eq. (22) are made in general of $2(\ell_1 + \ldots + \ell_p)$ indicator variables, not necessarily distinct. For most of the terms however, all involved edges are distinct, thus independent, and their contribution to the sum is $q^{2(\ell_1 + \ldots + \ell_p)}$.

Whenever a pair of blue (resp. red) indicators are equal, at least one term may be canceled, so the contribution to the expectation is higher than $q^{2(\ell_1 + \ldots + \ell_p)}$.

Whenever a pair of edges that appear in a blue/red pair of indicators are equal, the product of the indicators is necessarily 0 (indeed, an edge in $\mathcal{S}$ cannot be two-colored). These terms, where at least one equality of the form $\left\{u_j^{(m)}, v_j^{(m)}\right\} = \left\{\sigma_{m'}(u_{j'}^{(m')}), \sigma_{m'}(v_{j'}^{(m')})\right\}$ occurs, cover the case where the contribution is strictly less that $q^{2(\ell_1 + \ldots + \ell_p)}$ (it is 0). There are at most

$$\binom{\ell_1 + \ldots + \ell_p}{2} \left(\frac{n^2}{2}\right)^{\ell_1 + \ldots + \ell_p - 1}$$

such terms. Thus

$$\begin{aligned}
\mathbb{E}\left[\Delta(\sigma_1)^{\underline{\ell_1}} \Delta(\sigma_2)^{\underline{\ell_2}} \cdots \Delta(\sigma_p)^{\underline{\ell_p}}\right] &\geq \left(N - \binom{\ell_1 + \ldots + \ell_p}{2} \left(\frac{n^2}{2}\right)^{\ell_1 + \ldots + \ell_p - 1}\right) \times q^{2(\ell_1 + \ldots + \ell_p)} \\
&\sim \left(\frac{(1 - c(\lambda s))^2 n^2}{2}\right)^{\ell_1 + \ldots \ell_p} \times \left(\frac{\lambda(1-s)}{n}\right)^{2(\ell_1 + \ldots + \ell_p)} \\
&\underset{n \to \infty}{\longrightarrow} \left(\frac{\lambda^2 (1-s)^2 (1 - c(\lambda s))^2}{2}\right)^{\ell_1 + \ell_2 + \ldots + \ell_p}.
\end{aligned}$$

Upper bound: The terms that we now want to study are the terms for which the contribution is greater than $q^{2(\ell_1 + \ldots + \ell_p)}$. Looking closely at the general product in (22), an unusual high contribution is the consequence of three possible type of constraints:

(i) constraints of the form $\left\{u_j^{(m)}, v_j^{(m)}\right\} = \left\{u_{j'}^{(m')}, v_{j'}^{(m')}\right\}$: note that since the sums are made of ordered tuples, this equality may happen only for pairs such that $m \neq m'$. Moreover, transitivity of equality implies that a constraint implying some $\left\{u_j^{(m)}, v_j^{(m)}\right\}$ may happen at most once for each $m' \in [p], m' \neq m$ (otherwise we would have a relationship of the from $\left\{u_j'^{(m')}, v_j'^{(m')}\right\} = \left\{u_k'^{(m')}, v_k'^{(m')}\right\}$, which is impossible).

(ii) constraints of the form $\left\{\sigma_m(u_j^{(m)}), \sigma_m(v_j^{(m)})\right\} = \left\{\sigma_{m'}(u_{j'}^{(m')}), \sigma_{m'}(v_{j'}^{(m')})\right\}$. For the same reasons as in case (i), a constraint implying some $\left\{\sigma_m(u_j^{(m)}), \sigma_m(v_j^{(m)})\right\}$ may happen at most once for each $m' \in [p], m' \neq m$.

(iii) the last case is made of intersection of cases (i) and (ii), i.e. edges satisfying both $\left\{u_j^{(m)}, v_j^{(m)}\right\} = \left\{u_{j'}^{(m')}, v_{j'}^{(m')}\right\}$ and $\left\{\sigma_m(u_j^{(m)}), \sigma_m(v_j^{(m)})\right\} = \left\{\sigma_{m'}(u_{j'}^{(m')}), \sigma_{m'}(v_{j'}^{(m')})\right\}$. This implies in particular that $\left\{u_j^{(m)}, v_j^{(m)}\right\}$ is an common fixed edge for σ_m and $\sigma_{m'}$. By assumption (H3), note that there cannot be a connected path of constraints of the form (iii) of length greater or equal to 3.

Let us now represent these constraints with a dependency graph. Each vertex a the graph represent one edge $\left\{u_j^{(m)}, v_j^{(m)}\right\}$ of the sum, that we will align column-wise according to $m \in [p]$. We put a plain (resp. dashed) edge between two nodes if they are enforced by constraint (i) but not (iii) (resp. (ii) but not (iii)). Finally we draw a thick plain edge between two nodes if they are enforced by constraint (iii).

In view of discussion in points $(i) - (ii) - (iii)$, this dependency graph must be p-partite. Moreover, the subgraph made of plain thick or plain edges (resp. plain thick of dashed edges) only consists in a union of disjoint paths. The thick plain subgraph is only made of isolated edges and paths fo size 3. Finally, transitivity of the equality relationship enables to draw any path in any order: we shall take the left to right order by convention (no backtracking).

We denote by k_1 (resp. k_2) the number of plain (resp. dashed) edges. We also denote track k_3 the number of thick plain isolated edges, and k_4 the number of thick plain isolated paths of length 2. Figure 8 gives an example of such a dependency graph.

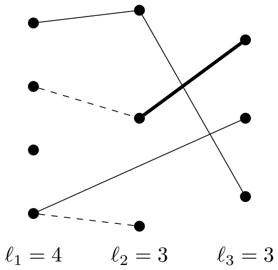


Figure 8: Example of a dependency graph, with $(k_1, k_2, k_3, k_4) = (3, 2, 1, 0)$.

In order to upper bound the contribution due to large terms, we must understand both the expectation of the product of indicators in (22) (this only depends on (k_1, k_2, k_3, k_4)), as well as the number of possible (labeled) dependency graphs with a given (k_1, k_2, k_3, k_4).

First, all plain (resp. dashed) dependency edge makes 1 (resp. 1) indicators disappear in the expectation (for any event $\mathcal{A}$, $\mathbf{1}_{\mathcal{A}}^2 = \mathbf{1}_{\mathcal{A}}$). In the same way, all thick plain isolated edge (resp. thick plain isolated path of length 2) makes 2 (resp. 4) indicators disappear the expectation for a given case with given (k_1, k_2, k_3, k_4) is

$$q^{2(\ell_1 + \ldots + \ell_p) - (k_1 + k_2 + 2k_3 + 4k_4)} \leq C_1 n^{-2(\ell_1 + \ldots + \ell_p) + (k_1 + k_2 + 2k_3 + 4k_4)} \tag{23}$$

where C_1 is a constant depending on $\ell_1, \ldots, \ell_p$, Second, an upper bound for the number of possible (labeled) dependency graphs with a given (k_1, k_2, k_3, k_4) can be established as follows. First, we have $k_1 + k_2 + k_3 + 2k_4$ equalities, leaving at most $\ell_1 + \ldots + \ell_p - (k_1 + k_2 + k_3 + 2k_4)$ degrees of freedom in the choices of the edges. Moreover, we force k_3 of these edges to be common fixed edges between two (distinct) permutations, and k_4 of them to be common fixed edges between three (pairwise distinct) permutations. In view of hypotheses (H1) and (H2), the number of possible (labeled) dependency graphs with a given (k_1, k_2, k_3, k_4) is at most

$$\binom{k_1+k_2+k_3+k_4}{k_3+k_4} |\mathcal{S}|^{\ell_1+\ldots+\ell_p-(k_1+k_2+k_3+2k_4)-k_3-k_4} \times (n^{1+t})^{k_3} \times n^{tk_4}$$
$$\leq C_2 n^{2(\ell_1+\ldots+\ell_p)-2(k_1+k_2)-(3-t)k_3-(6-t)k_4}, \qquad (24)$$

where C_2 is a constant depending on $\ell_1, \ldots, \ell_p$.

Hence, in view of (23) and (24), the total contribution of higher terms is upper bounded by

$$\sum_{s=1}^{\ell_1+\ldots+\ell_p} \sum_{k_1+k_2+k_3+2k_4=s} C_1 C_2 n^{-2(\ell_1+\ldots+\ell_p)+(k_1+k_2+2k_3+4k_4)} n^{2(\ell_1+\ldots+\ell_p)-2(k_1+k_2)-(3-t)k_3-(6-t)k_4}$$
$$\leq C_1 C_2 \sum_{s=1}^{\ell_1+\ldots+\ell_p} \sum_{k_1+k_2+k_3+2k_4=s} n^{-k_1} n^{-k_2} n^{-(1-t)k_3} n^{-(2-t)k_4}$$
$$\leq C_1 C_2 \times (\ell_1+\ldots+\ell_p) \times (\ell_1+\ldots+\ell_p)^{4(\ell_1+\ldots+\ell_p)} \times n^{-(1-t)} \xrightarrow[n\to\infty]{} 0.$$

This last convergence concludes the proof. □

B Proofs of Lemmas

B.1 Proof of Lemma 3.1

Proof. For the control of $X_{\mathbf{T}}$ we follow classical computations made in [Bol01] to establish asymptotic behavior of $X_{\mathbf{T}}$. For our purpose, we only need the two first moments. Assume that $\mathbf{T}$ is of size $k = k(\mathbf{T}) \leq K$, and that its automorphism group has $a = a(\mathbf{T})$ elements. Then, letting $\mu = \lambda s$,

$$\mathbb{E}[X_{\mathbf{T}}] = \binom{n}{k} \times \frac{k!}{a} \times \left(\frac{\mu}{n}\right)^{k-1} \left(1 - \frac{\mu}{n}\right)^{k(n-k)+\binom{k}{2}-k+1}.$$

Indeed, we have $\binom{n}{k}$ choices for the nodes, then $\frac{k!}{a}$ ways of putting the edges. Using $\binom{n}{k} \sim \frac{n^k}{k!}$ and $\left(1-\frac{\mu}{n}\right)^{-k^2+\binom{k}{2}-k+1} \sim 1$ as soon as $k = o(\sqrt{n})$, we get

$$\mathbb{E}[X_{\mathbf{T}}] \sim n\mu^{k-1} e^{-\mu k}/a.$$

We now compute $\mathbb{E}[X_{\mathbf{T}}(X_{\mathbf{T}}-1)]$ by classically counting the number of ordered pairs of distinct isolated tree components of $\mathcal{G} \wedge \mathcal{G}'$ isomorphic to $\mathbf{T}$. This number is then multiplied by the probability of observing these two distinct isolated components. This gives

$$\mathbb{E}[X_{\mathbf{T}}(X_{\mathbf{T}}-1)] = \binom{n}{k}\binom{n-k}{k} \times \left(\frac{k!}{a}\right)^2 \times \left(\frac{\mu}{n}\right)^{2(k-1)} \left(1-\frac{\mu}{n}\right)^{2\left(k(n-2k)+\binom{k}{2}-k+1\right)} \left(1-\frac{\mu}{n}\right)^{k^2}.$$

Here again, $k = o(\sqrt{n})$ gives that

$$\mathbb{E}[X_{\mathbf{T}}(X_{\mathbf{T}}-1)] \sim n^2\mu^{2(k-1)} e^{-2\mu k}/a^2.$$

Denoting $\alpha = \alpha(\mathbf{T}) := n\mu^{k-1}e^{-\mu k}/a(\mathbf{T})$, these computations give that $\mathbb{E}[X_{\mathbf{T}}] \sim \mathrm{Var}(X_{\mathbf{T}}) \sim \alpha(\mathbf{T})$ when $n \to \infty$, uniformly in $k \leq K(n)$ as soon as $K(n) = o(\sqrt{n})$. Let us fix $\varepsilon = \varepsilon(n) > 0$ small enough. Applying Chebyshev's inequality together with the union bound gives

$$\begin{aligned}
\mathbb{P}\left(\exists k \in [K(n)],\, \exists \mathbf{T} \in \mathbb{T},\, X_{\mathbf{T}} \leq (1-\varepsilon)\alpha(\mathbf{T})\right) &\leq \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} \mathbb{P}\left(X_{\mathbf{T}} - \mathbb{E}\left[X_{\mathbf{T}}\right] \leq (1-\varepsilon)\alpha(\mathbf{T}) - \mathbb{E}\left[X_{\mathbf{T}}\right]\right) \\
&\overset{(a)}{\leq} \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} \frac{\mathrm{Var}\left(X_{\mathbf{T}}\right)}{\left((1-\varepsilon)\alpha(\mathbf{T}) - \mathbb{E}\left[X_{\mathbf{T}}\right]\right)^2} \\
&\overset{(b)}{\leq} (1+o(1)) \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} \frac{1}{\varepsilon^2 \alpha(\mathbf{T})} \\
&\overset{(c)}{\leq} (1+o(1)) \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} \frac{1}{\varepsilon^2 n f(k)} \\
&\overset{(d)}{\leq} (1+o(1)) K(n)^{K(n)} \frac{1}{\varepsilon^2 n f(K(n))},
\end{aligned}$$

where

$$f(k) := \frac{\mu^{k-1} e^{-\mu k}}{k!}. \tag{25}$$

We used in (a) that all $(1-\varepsilon)\alpha(\mathbf{T}) - \mathbb{E}\left[X_{\mathbf{T}}\right]$ are negative for n large enough, in (b) uniformity in $k \leq K(n)$, in (c) the lower bound $nf(k)$ for $\alpha(T)$, and finally in (d) that $k \mapsto f(k)$ is decreasing since $\mu e^{-\mu} < 1$.

Taking now e.g. $\varepsilon = n^{-1/4}$, the last fact to check to establish the Lemma is that $K^K/f(K) = o(n^{1/2})$ when $K = K(n) = \log^{1/2}(n)$:

$$\begin{aligned}
K^K/f(K) &= K^K K! (1/\mu)^{K-1} e^{\mu K} \\
&\leq \exp\left(2K \log K + (\log(1/\mu) + \mu) K\right) \\
&= \exp\left(\log^{1/2}(n) \log\log n + (\log(1/\mu) + \mu) \log^{1/2}(n)\right) = o(n^{1/2}).
\end{aligned}$$

□

B.2 Proof of Lemma 3.2

Proof. Denote $T_\infty := |V_\infty|$ and $T_> := |V_>|$. First notice that for any permutations σ_i, σ_j with $i \neq j$ generated with Algorithm 1, we have the following equality:

$$\mathrm{ov}(\sigma_i, \sigma_j) = T_\infty + T_> + \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} k \cdot \mathrm{ov}(\Sigma_{\mathbf{T}}^{(i)}, \Sigma_{\mathbf{T}}^{(j)}), \tag{26}$$

where $\Sigma_{\mathbf{T}}^{(i)}$ (resp. $\Sigma_{\mathbf{T}}^{(j)}$) is the tree permutation associated with $\mathbf{T}$ in σ_i (resp. in σ_j). We know that $T_\infty = c(\lambda s)n + o(n)$ w.h.p. and by Lemma 2.2, $T_> = o(n)$ w.h.p.

Define

$$\mathrm{ov}'(\sigma_i, \sigma_j) := \sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} k \cdot \mathrm{ov}(\Sigma_{\mathbf{T}}^{(i)}, \Sigma_{\mathbf{T}}^{(j)}), \tag{27}$$

the second term in (26). We dominate $\mathrm{ov}'(\sigma_i, \sigma_j)$ as follows:

Lemma B.1. *If $X = \mathrm{ov}(\Sigma_{\mathbf{T}}^{(i)}, \Sigma_{\mathbf{T}}^{(j)})$,, then for all $t \in \mathbb{R}$,*

$$\mathbb{E}\left[e^{tX}\right] \leq \exp(e^t). \tag{28}$$

Proof.

$$\mathbb{E}\left[e^{tX}\right] = \sum_{m\geq 0} e^{tm} \mathbb{P}(X \geq m).$$

Noting that $\mathbb{P}(X \geq m) \leq \mathbb{E}\left[\binom{X}{m}\right]$ and that

$$\mathbb{E}\left[\binom{X}{m}\right] = \frac{1}{m!}\mathbb{E}\left[X(X-1)\dots(X-m+1)\right]$$
$$= \frac{1}{m!}k(k-1)\dots(k-m+1)\frac{(k-m)!}{k!} = \frac{1}{m!}$$

gives

$$\mathbb{E}\left[e^{tX}\right] \leq \sum_{m\geq 0} \frac{e^{tm}}{m!} \leq \exp(e^t).$$

□

Using independence of the X variables, Equation (28) of Lemma B.1 give that for all $t \in \mathbb{R}$,

$$\mathbb{E}\left[e^{t\cdot \mathrm{ov}'(\sigma_i,\sigma_j)}\right] \leq \prod_{k=1}^{K(n)} \prod_{\mathbf{T}\in\mathbb{T}_k} \exp(e^{tk}) \leq \exp\left(e^{tK(n)}K(n)^{K(n)+1}\right). \tag{29}$$

Now, we use the classical Chernoff bound, for positive t,

$$\mathbb{P}\left(\mathrm{ov}'(\sigma_i,\sigma_j) \geq n^\alpha\right) \leq \exp\left(-tn^\alpha + e^{tK(n)}K(n)^{K(n)+1}\right)$$
$$\leq \exp\left(-\frac{n^\alpha}{K(n)}\left[\log\left(\frac{n^{1-\alpha}}{K(n)^{K(n)+2}}\right) - 1\right]\right),$$

taking $t = \frac{1}{K(n)}\log\left(\frac{n^\alpha}{K(n)^{K(n)+2}}\right)$. The right hand side tend to 0 for any $\alpha \in (0,1)$, and a simple use of the union bound ends the proof. □

B.3 Proof of Lemma 3.3

Proof. Fix $t > 0$. We use a standard first moment method. We will use the results of Lemmas 2.2 and 3.1, conditioning on the event $\mathcal{A}$ where the corresponding results hold. Since $\mathbb{P}(\mathcal{A}) = 1 - o(1)$, this conditioning is legitimate for our purpose.

Step 1. Let us first control the term $F(\mathcal{S}_{out}, \sigma_{i_1}, \dots, \sigma_{i_r})$: edges of $\mathcal{S}_{out}$ are made of exactly one vertex in $V_{\infty,>}$. There are at most n^2 such edges, and the probability for a given edge of $\mathcal{S}_{out}$ being a common fixed edge of $\sigma_{i_1}, \dots, \sigma_{i_r}$ is $\frac{1}{X_{\mathbf{T}}^{r-1}}$, which can be upper-bounded on $\mathcal{A}$ by $(nf(K(n)))^{1-r} \leq n^{1-r+t/2}$ by Remark 3.1. Edges of $\mathcal{S}_{out}$ thus have a contribution in $\mathbb{E}\left[F(\sigma_{i_1}, \dots, \sigma_{i_r})|\mathcal{A}\right]$ of at most $n^{3-r+t/2}$.

Step 2. In the edges appearing in $F(\sigma_{i_1}, \dots, \sigma_{i_r})$, we consider three cases:

(i) edges of **Intra**: these are edges made with two vertices in the same tree $\mathcal{T} \sim \mathbf{T} \in \mathbb{T}$. On event $\mathcal{A}$, there are at most

$$\sum_{k=1}^{K(n)} \sum_{\mathbf{T}\in\mathbb{T}_k} X_{\mathbf{T}}k^2 \leq nK(n)$$

such edges. The probability for a given edge of **Intra** made of vertices of $\mathbf{T} \in \mathbb{T}$ being a common fixed edge of $\sigma_{i_1}, \dots, \sigma_{i_r}$ is $\frac{1}{X_{\mathbf{T}}^{r-1}}$, which can be upper-bounded by $(nf(K(n)))^{1-r} \leq n^{1-r+t/2}$. Edges of **Intra** thus have a contribution in $\mathbb{E}\left[F(\sigma_{i_1}, \dots, \sigma_{i_r})|\mathcal{A}\right]$ of at most $n^{2-r+t/2}$.

(ii) edges of **Inter**$_1$: these are edges made with two vertices u, v in different trees $\mathcal{T} \neq \mathcal{T}'$ (but that may be $\sim$ to the same $\mathbf{T} \in \mathbb{T}$), and verifying $u \not\sim v$. There are at most n^2 such edges. Since $u \not\sim v$, there are only one possibility to map two edges of **Inter**$_1$. The probability for a given edge of **Inter**$_1$ made of vertices of $\mathcal{T} \sim \mathbf{T}, \mathcal{T}' \in \mathbf{T}'$ being a common fixed edge is $\frac{1}{(X_{\mathbf{T}}(X_{\mathbf{T}}-1))^{r-1}}$, and edges of **Inter**$_1$ thus have a contribution in the expectation of at most $n^{4-2r+t/2}$.

(iii) edges of **Inter**$_2$: these are edges similar to case (ii), except that their endpoints belong necessarily to isomorphic trees, and verifying $u \simeq v$. There are at most n^2 such edges. Since $u \simeq v$, there are two ways to map two edges of **Inter**$_2$. The probability for a given edge of **Inter**$_2$ made of vertices of $\mathcal{T}, \mathcal{T}' \sim \mathbf{T}$ being a common fixed edge is time $\left(\frac{2}{X_{\mathbf{T}}(X_{\mathbf{T}}-1)}\right)^{r-1}$, and edges of **Inter**$_2$ thus have a contribution in the expectation of at most $n^{4-2r+t/2}$.

Step 3. The first two steps show that $\mathbb{E}\left[F(\sigma_{i_1}, \dots, \sigma_{i_r})|\mathcal{A}\right] \leq Cn^{3-r+t/2}$ for all $t > 0$. Summing over all possible r-tuples of permutations, Markov inequality yields

$$\begin{aligned}\mathbb{P}\left(\exists r \geq 4,\, \exists \sigma_{i_1}, \dots, \sigma_{i_r} \text{ pairwise distinct, } F(\mathcal{S}, \sigma_{i_1}, \dots, \sigma_{i_r}) \geq 1\right) &\leq o(1) + \sum_{r=4}^{\infty} p^r C n^{3-r+t/2} \\ &\leq C p^4 n^{t/2-1} \to 0,\end{aligned}$$

for t small enough, and

$$\mathbb{P}\left(\exists \sigma_{i_1}, \sigma_{i_2}, \sigma_{i_3} \text{ pairwise distinct, } F(\mathcal{S}, \sigma_{i_1}, \sigma_{i_2}, \sigma_{i_3}) \geq n^t\right) \leq o(1) + p^3 \times Cn^{-t/2} \to 0,$$

and

$$\mathbb{P}\left(\exists \sigma_{i_1} \neq \sigma_{i_2}, F(\mathcal{S}, \sigma_{i_1}, \sigma_{i_2}) \geq n^{1+t}\right) \leq o(1) + p^2 \times Cn^{-t/2} \to 0.$$

□