

Correlation Detection in Trees for Planted Graph Alignment

Luca Ganassali  

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

Laurent Massoulié  

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

Marc Lelarge  

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

Abstract

Motivated by alignment of correlated sparse random graphs, we study a hypothesis problem of deciding whether two random trees are correlated or not. Based on this correlation detection problem, we propose MPAlign, a message-passing algorithm for graph alignment, which we prove to succeed in polynomial time at partial alignment whenever tree detection is feasible. As a result our analysis of tree detection reveals new ranges of parameters for which partial alignment of sparse random graphs is feasible in polynomial time.

We conjecture that the connection between partial graph alignment and tree detection runs deeper, and that the parameter range where tree detection is impossible, which we partially characterize, corresponds to a region where partial graph alignment is hard (not polytime feasible).

2012 ACM Subject Classification Mathematics of computing → Hypothesis testing and confidence interval computation; Mathematics of computing → Max marginal computation; Mathematics of computing → Graph algorithms

Keywords and phrases inference on graphs, hypothesis testing, Erdős-Rényi random graphs

Digital Object Identifier 10.4230/LIPIcs.ITCS.2022.74

Related Version A full version of this paper can be found at: <https://arxiv.org/abs/2107.07623>

Funding This work was partially supported by the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR19-P3IA-0001 (PRAIRIE 3IA Institute).

Acknowledgements The authors would like to thank Guilhem Semerjian for helpful discussions, and the reviewers for relevant comments.

1 Introduction

Graph alignment

Given two graphs $G = (V, E)$ and $G' = (V', E')$ with same node set $V = V' = [n]$, the problem of *graph alignment* consists in identifying a bijective mapping, or *alignment* $\sigma : V \rightarrow V'$ that minimizes

$$\sum_{i,j \in [n]} (\mathbf{1}_{\{i,j\} \in E} - \mathbf{1}_{\{\sigma(i),\sigma(j)\} \in E'})^2,$$

that is the number of disagreements between adjacencies in the two graphs under the alignment σ . This problem reduces to the graph isomorphism problem in the noiseless setting where the two graphs can be matched perfectly (or, are isomorphic). The paradigm of graph alignment has found numerous applications across a variety of diverse fields, such as network privacy [17], computational biology [19], computer vision [4], and natural language processing.



© Luca Ganassali, Laurent Massoulié, and Marc Lelarge;
licensed under Creative Commons License CC-BY 4.0

13th Innovations in Theoretical Computer Science Conference (ITCS 2022).

Editor: Mark Braverman; Article No. 74; pp. 74:1–74:8

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

74:2 Correlation Detection in Trees for Planted Graph Alignment

Given the adjacency matrices A and B of the two graphs, graph alignment can be viewed as an instance of the quadratic assignment problem (QAP) [18]:

$$\arg \max_{\Pi} \langle A, \Pi B \Pi^T \rangle \quad (1)$$

where Π ranges over all $n \times n$ permutation matrices, and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product. QAP is known to be NP-hard in general, as well as some of its approximations [18, 15]. 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 will not be the worst-case instances, as explained hereafter.

Correlated Erdős-Rényi model

A recent thread of research [5, 6, 7, 8, 9, 11, 10] has focused on the study of graph alignment when the two considered graphs are drawn from a generative model under which they are both Erdős-Rényi random graphs. Specifically, for $(\lambda, s) \in \mathbb{R}_+ \times [0, 1]$, the correlated Erdős-Rényi random graph model, denoted $\mathcal{G}(n, q, s)$ with $q = \lambda/n$, consists of two random graphs \mathbf{G}, \mathbf{G}' both with node set $[n]$ generated as follows. Consider an i.i.d. collection $\{(\mathbf{A}_{ij}, \tilde{\mathbf{A}}_{ij})\}_{i < j \in [n]}$ of pairs of correlated Bernoulli random variables with distribution

$$(\mathbf{A}_{ij}, \tilde{\mathbf{A}}_{ij}) = \begin{cases} (1, 1) & \text{with probability } \lambda s/n \\ (1, 0) & \text{with probability } \lambda(1-s)/n \\ (0, 1) & \text{with probability } \lambda(1-s)/n \\ (0, 0) & \text{with probability } 1 - \lambda(2-s)/n. \end{cases} \quad (2)$$

Consider then a permutation σ^* drawn independently of $\mathbf{A}, \tilde{\mathbf{A}}$ and uniformly at random from the symmetric group \mathcal{S}_n . The two graphs $(\mathbf{G}, \mathbf{G}')$ are then defined by their adjacency matrices \mathbf{A} and \mathbf{A}' such that for all $i < j \in [n]$:

$$\mathbf{A}_{ij} = \mathbf{A}_{ji}, \mathbf{A}'_{ij} = \mathbf{A}'_{ji} = \tilde{\mathbf{A}}_{\sigma^*(i)\sigma^*(j)}.$$

In this setting, the marginal distributions of \mathbf{G} and \mathbf{G}' are identical, namely that of the Erdős-Rényi model $\mathcal{G}(n, q)$ with $q = \lambda/n$.

Planted graph alignment

The previous model is then used to study the mean-case version of graph alignment – namely *planted graph alignment* – consisting in finding an estimator $\hat{\sigma}$ of the planted solution σ^* upon observing \mathbf{G} and \mathbf{G}' . For any subset $\mathcal{C} \subset [n]$, the performance of any one-to-one estimator $\hat{\sigma} : \mathcal{C} \rightarrow [n]$ is now assessed through $\text{ov}(\sigma^*, \hat{\sigma})$, its *overlap* with the unknown permutation σ^* , defined as

$$\text{ov}(\sigma^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) = \sigma^*(i)}. \quad (3)$$

Note that the estimator $\hat{\sigma}$ may not be in \mathcal{S}_n , and only consist in a partial matching. The *error fraction* of $\hat{\sigma}$ with the unknown permutation σ^* is defined as

$$\text{err}(\sigma^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) \neq \sigma^*(i)} = \frac{|\mathcal{C}|}{n} - \text{ov}(\sigma^*, \hat{\sigma}). \quad (4)$$

It is easy to check that the maximum-a-posteriori (MAP) estimator of σ^* given \mathbf{G}, \mathbf{G}' is the solution of the QAP problem (1). However, it is shown [5] that this MAP estimator performs well only in the case where the mean degrees in the graphs are at least $\Omega(\log n)$. Hence in our sparse regime (with constant mean degree) this new measure of performance (3) – which is the one on which we will focus next – differs from that of the non-planted case (1).

A sequence of injective estimators $\{\hat{\sigma}_n\}_n$ – omitting the dependence in n – is said to achieve

- *Exact recovery* if $\mathbb{P}(\hat{\sigma} = \sigma^*) \xrightarrow[n \rightarrow \infty]{} 1$,
- *Almost exact recovery* if $\mathbb{P}(\text{ov}(\sigma^*, \hat{\sigma}) = 1 - o(1)) \xrightarrow[n \rightarrow \infty]{} 1$,
- *Partial recovery* if there exists some $\varepsilon > 0$ such that $\mathbb{P}(\text{ov}(\sigma^*, \hat{\sigma}) > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 1$,
- *One-sided partial recovery* if it achieves partial recovery and $\mathbb{P}(\text{err}(\sigma^*, \hat{\sigma}) = o(1)) \xrightarrow[n \rightarrow \infty]{} 1$.

► **Remark 1.** One-sided partial recovery is by definition at least as hard as partial recovery. From an application standpoint it is more appealing than partial recovery: indeed, it may be of little use to know one has a permutation with 30% of correctly matched nodes if one does not have a clue about which pairs are correctly matched.

This justifies our focus on one-sided recovery, besides the fact that it is more amenable to analysis at this stage. Another main interest of one-sided partial alignment is that its local rephrasing in terms of one-sided tests on trees is the most relevant one, when compared to other types of tests.

Phase diagram

In this sparse regime (where the graphs have constant mean degree λ), it is known [5, 6] that the presence of $\Omega(n)$ isolated vertices in the underlying intersection graph of \mathbf{G} and \mathbf{G}' makes exact and almost exact recovery impossible. The main questions consist then in determining the phase diagram of the model $\mathcal{G}(n, \lambda/n, s)$ for partial alignment (or recovery), namely the range of parameters (λ, s) for which, in the large n limit:

- Any sequence of estimators fails to achieve (one-sided) partial recovery (for any $\varepsilon > 0$). We refer to the corresponding range as the *information-theoretic (IT-)impossible phase*,
- There is a sequence of estimators $\hat{\sigma}$ achieving (one-sided) partial recovery (with some $\varepsilon > 0$), which we refer to as the *IT-feasible phase*,
- There is a sequence of estimators $\hat{\sigma}$ that can be computed in polynomial-time achieving (one-sided) partial recovery (with some $\varepsilon > 0$): the *easy phase*.

An interesting perspective on this problem is provided by research on community detection, or graph clustering, for random graphs drawn according to the stochastic block model. In that setup, above the so-called Kesten-Stigum threshold, polynomial-time algorithms for clustering are known [3, 14, 16], and the consensus among researchers in the field is that no polynomial-time algorithms exist below that threshold. Yet, there is a range of parameters with non-empty interior below the Kesten-Stigum threshold for which exponential-time algorithms are known to succeed at clustering [1]. In other words, for graph clustering, it is believed that there is a non-empty *hard phase*, consisting of the set difference between the IT-feasible phase and the polynomial-time feasible phase.

The picture available to date for partial graph alignment is as follows. Recent work [10] shows that the IT-impossible phase includes the range of parameters $\{(\lambda, s) : \lambda s \leq 1\}$, and Wu et al. [20] have established that the IT-feasible phase includes the range of parameters $\{(\lambda, s) : \lambda s > 4\}$ (condition $\lambda s > C$ for some large C had previously been established in [13]).

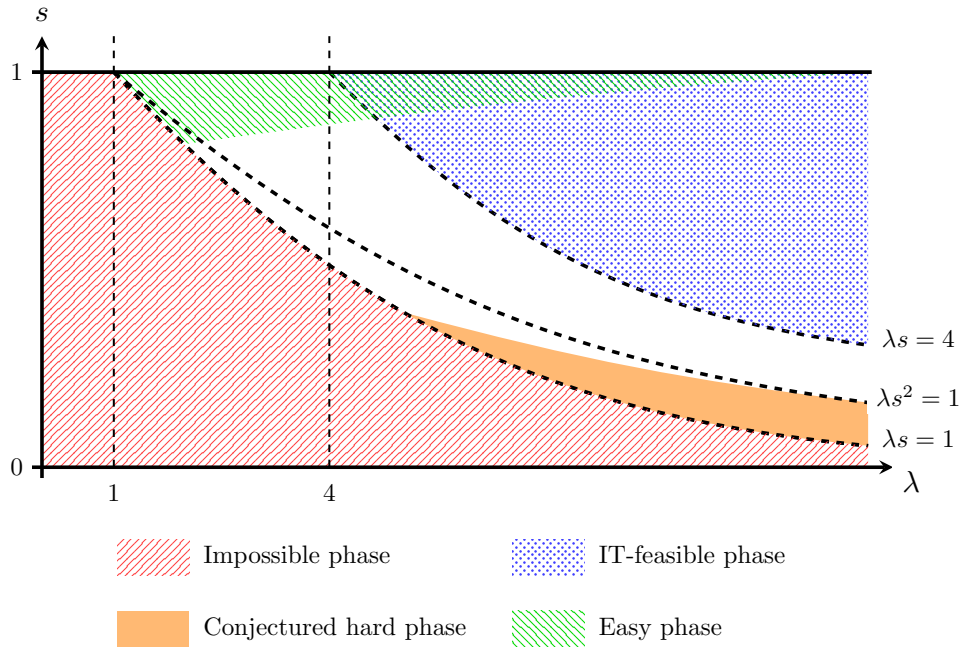


Figure 1 Diagram of the (λ, s) regions where partial recovery is known to be IT-impossible ([10]), IT-feasible ([20]), or easy ([11] and this paper). In the orange region one-sided detectability is impossible in the tree correlation detection problem, and partial graph alignment is conjectured to be hard (this paper).

For the easy phase, Ganassali and Massoulié [11] have established that it includes the range of parameters $\{(\lambda, s) : \lambda \in [1, \lambda_0], s \in [s(\lambda), 1]\}$ for some parameter $\lambda_0 > 1$ and some function $s(\lambda) : (1, \lambda_0] \rightarrow [0, 1]$. The phase diagram is summed up in Figure 1.

2 Problem description and main contribution

This partial picture leaves open the question of whether, similarly to the case of graph clustering, graph alignment features a hard phase or not. The contribution of the present work can be summarized in three points:

- (1) We investigate a fundamental statistical problem, which to the best of our knowledge had not been previously studied: hypothesis testing for correlation detection in trees. We study the regimes in which the optimal test on trees succeeds or fails in the setting when the trees are correlated Galton-Watson trees (see Theorem 2);
- (2) For this detection problem on trees, the computation of the likelihood ratio can be made recursively on the depth, which yields an optimal message-passing algorithm for this task running in polynomial-time in the number of nodes;
- (3) We then state that the previous detection problem on trees arises naturally from a local point of view in the related problem of one-sided partial recovery for graph alignment. In light of the previous analysis we are able to draw conclusions for our initial problem on graphs; doing so we complete the phase diagram as shown in Figure 1, extending the regime for which one-sided partial alignment is provably feasible in polynomial time, and exhibiting the presence of a conjectured hard phase (see Theorem 4).

Our approach to point (3) follows the way paved by [11]. It essentially relies on an algorithm which lets $\hat{\sigma}(i) = u$ for i such that the local structure of graph \mathbf{G} in the neighborhood of node i is 'close' to the local structure of graph \mathbf{G}' in the neighborhood of node u . As exploited in [11], the neighborhoods to distance d of two nodes i, u in \mathbf{G} and \mathbf{G}' , provided that $u = \sigma^*(i)$, are asymptotically distributed as correlated Galton-Watson branching trees (distribution denoted $\mathbb{P}_{1,d}$). On the other hand, for pairs of nodes (i, u) taken at random in $[n]$, the joint neighborhoods of nodes i and u in \mathbf{G} and \mathbf{G}' respectively, to depth d , are asymptotically distributed as a pair of independent Galton-Watson branching trees (distribution denoted $\mathbb{P}_{0,d}$).

Thus, a fundamental step in our approach is to determine the efficiency of tests for deciding whether a pair of branching trees is drawn from either a product distribution, or a correlated distribution. Previous work [11] relied on tests based on a so-called *tree-matching weight* to measure the similarity between two trees. In the present work we are instead interested in studying the existence of *one-sided tests*, which are tests asymptotically guarantying a vanishing type I error and a non vanishing power. According to the Neyman-Pearson Lemma, optimal one-sided tests are based on the likelihood ratio \mathbf{L}_d of the distributions under the distinct hypotheses $\mathbb{P}_{1,d}$ and $\mathbb{P}_{0,d}$ (trees correlated or not)¹. The mathematical formalization of point (1) here above is the following

► **Theorem 2** (Correlation detection in trees). *Let*

$$\text{KL}_d := \text{KL}(\mathbb{P}_{1,d} \parallel \mathbb{P}_{0,d}) = \mathbb{E}_{1,d} [\log(\mathbf{L}_d)].$$

Then the following propositions are equivalent:

- (i) *There exists a one-sided test for deciding $\mathbb{P}_{0,d}$ versus $\mathbb{P}_{1,d}$,*
- (ii) $\lim_{d \rightarrow \infty} \text{KL}_d = +\infty$ and $\lambda s > 1$,
- (iii) *There exists $(a_d)_d$ such that $a_d \rightarrow \infty$, $\mathbb{P}_{0,d}(\mathbf{L}_d > a_d) \rightarrow 0$ and $\liminf_d \mathbb{P}_{1,d}(\mathbf{L}_d > a_d) > 0$.*
- (iv) *Denoting $\mathbb{P}_0 := \mathbb{P}_{0,\infty}$, the martingale $(\mathbf{L}_d)_d$ (w.r.t. to \mathbb{P}_0) is not uniformly integrable.*
- (v) *with probability² $1 - p_{\text{ext}}(\lambda s) > 0$, \mathbf{L}_d diverges to $+\infty$ with rate $\Omega(\exp(\Omega(1) \times (\lambda s)^d))$.*

► **Remark 3.** This Theorem gives general necessary and sufficient conditions for the existence of a one-sided test in the tree correlation detection problem. In this full version of this paper [12], several more explicit results, giving sufficient conditions in terms of λ and s for points (ii) or (iii) to be verified or to fail, are stated as corollaries and contribute to the phase diagram in Figure 1.

We also mention that condition (v) will be used in the design of MPAlign (Algorithm 1) choosing an appropriate threshold that will guarantee for the method to output both a substantial part of the underlying permutation and a vanishing number of mismatches.

► **Theorem 4** (Consequences for one-sided partial graph alignment). *For given (λ, s) , if one-sided correlation detection is feasible, i.e. any of the conditions in Theorem 2 holds, then one-sided partial alignment in the correlated Erdős-Rényi model $\mathcal{G}(n, \lambda/n, s)$ is achieved in polynomial time by our algorithm MPAlign (Algorithm 1).*

¹ This guarantees that whenever the test based on tree-matching weight in [11] succeeds, the optimal test studied in this paper also succeeds. On this point, Theorem 3, Section 3 in the full version [12] gives more insight on the extension of the regime established in [11] where partial alignment is feasible (for small λ and s close to 1).

² This probability $1 - p_{\text{ext}}(\lambda s)$ is defined as the probability that a Galton-Watson tree of offspring $\text{Poi}(\lambda s)$ survives.

► **Conjecture.** *We conjecture that if one-sided correlation detection in trees fails, i.e. none of the equivalent conditions in Theorem 2, then no polynomial-time algorithm achieves (one-sided) partial recovery. In view of Theorem 5 of Section 5 of the full version of the paper [12], which guarantees existence of a non-empty parameter region where one-sided tree detection fails while partial graph alignment can be done in non-polynomial time, our conjecture would imply the hard phase to be non-empty.*

3 The MPAlign algorithm

Intuition: extending the tree correlation detection problem

Let $(\mathbf{G}, \mathbf{G}') \sim \mathcal{G}(n, q = \lambda/n, s)$, with underlying alignment σ^* . In order to distinguish matched pairs of nodes (i, u) , we consider their neighborhoods $\mathcal{N}_{d, \mathbf{G}}(i)$ and $\mathcal{N}_{d, \mathbf{G}'}(u)$ at a given depth d : these neighborhoods are close to Galton-Watson trees. In the case where the two vertices are actual matches, i.e. $u = \sigma^*(i)$, we are exactly in the setting of our tree correlation detection problem under \mathbb{P}_1 : point (v) of in Theorem 2 shows that there exists a threshold β_d such that with probability at least $1 - p_{\text{ext}}(\lambda s) > 0$,

$$L_d(i, u) := L_d(\mathcal{N}_{d, \mathbf{G}}(i), \mathcal{N}_{d, \mathbf{G}'}(u)) > \beta_d,$$

when $d \rightarrow \infty$. Point (v) of Theorem 2 shows that this threshold β_d can be e.g. taken to be $\exp(n^\gamma)$ for some $\gamma \in (0, c \log(\lambda s))$.

At the same time, when nodes u and $\sigma^*(i)$ are distinct and sufficiently far away, we can argue that we are also – with high probability – in the setting of the tree correlation detection problem under \mathbb{P}_0 : since $\mathbb{E}_0[\mathbf{L}_d] = 1$, Markov’s inequality shows that with high probability when $d \rightarrow \infty$,

$$L_d(i, u) \leq \beta_d.$$

Message-passing for the likelihood ratios

As mentioned earlier, the likelihood ratios can be computed efficiently on a graph, giving the exact expression for a *message-passing* procedure, assuming that all neighborhoods are locally tree-like at depth d . We define *oriented likelihood ratios*: for any $i, j \in V(\mathbf{G})$ and $u, v \in V(\mathbf{G}')$, we write $L_d(i \leftarrow j, u \leftarrow v)$ for the likelihood ratio at depth d of two trees, the first one (resp. second one) being rooted at i in \mathbf{G} (resp. u in \mathbf{G}') where the edge $\{i, j\}$ (resp. $\{u, v\}$), if initially present, has been deleted. These oriented likelihood ratios satisfy the following recursion:

$$L_d(i \leftarrow j, u \leftarrow v) = \sum_{k=0}^{d_i \wedge d'_u - 1} \psi(k, d_i - 1, d'_u - 1) \sum_{\substack{\sigma \in \mathcal{S}([k], \mathcal{N}_{\mathbf{G}}(i) \setminus \{j\}) \\ \sigma' \in \mathcal{S}([k], \mathcal{N}_{\mathbf{G}'}(u) \setminus \{v\})}} \prod_{\ell=1}^k L_{d-1}(\sigma(\ell) \leftarrow i, \sigma'(\ell) \leftarrow u), \quad (5)$$

where d_i (resp. d'_u) is the degree of i in \mathbf{G} (resp. of u in \mathbf{G}'), $\mathcal{N}_{\mathbf{G}}(i)$ (resp. $\mathcal{N}_{\mathbf{G}'}(u)$) is the set of neighbors of i in \mathbf{G} (resp. of u in \mathbf{G}'), $\mathcal{S}(A, B)$ is the set of injective mappings from A to B , and

$$\psi(k, c, c') := e^{\lambda s} \times \frac{s^k \bar{s}^{c+c'-2k}}{\lambda^k k!}.$$

The likelihood ratio at depth d between i and u is obtained by computing

$$L_d(i, u) = \sum_{k=0}^{d_i \wedge d'_u} \psi(k, d_i, d'_u) \sum_{\substack{\sigma \in \mathcal{S}([k], \mathcal{N}_{\mathbf{G}}(i)) \\ \sigma' \in \mathcal{S}([k], \mathcal{N}_{\mathbf{G}'}(u))}} \prod_{\ell=1}^k L_{d-1}(\sigma(\ell) \leftarrow i, \sigma'(\ell) \leftarrow u). \quad (6)$$

A natural idea is then to compute for each pair (i, u) the likelihood ratio $L_d(i, u)$ with d large enough (typically scaled in $\Theta(\log n)$ where n is the number of vertices in \mathbf{G} and \mathbf{G}') and to compare it to β_d to decide whether i in \mathbf{G} is matched to u in \mathbf{G}' ³.

Algorithm description

Our algorithm is as follows:

■ **Algorithm 1** MPAlign: Message-passing algorithm for sparse graph alignment.

Input: Two graphs G and G' of size n , average degree λ , depth d , threshold parameter β

Output: A set of pairs $\mathcal{M} \subset V(G) \times V(G')$.

$\mathcal{M} \leftarrow \emptyset$

Compute $L_d(i \leftarrow j, u \leftarrow v)$ for all $\{i, j\} \in E$ and $\{u, v\} \in E'$ thanks to (5)

for $(i, u) \in V(G) \times V(G')$ **do**

if $\mathcal{N}_G(i, d)$ and $\mathcal{N}_{G'}(u, d)$ contain no cycle, and
 $\exists \{j_1, j_2, j_3\} \subset \mathcal{N}_G(i), \exists \{v_1, v_2, v_3\} \subset \mathcal{N}_{G'}(u)$ such that
 $L_{d-1}(j_t \leftarrow i, v_t \leftarrow u) > \beta$ for all $t \in \{1, 2, 3\}$ **then**
 | $\mathcal{M} \leftarrow \mathcal{M} \cup \{(i, u)\}$

end

end

return \mathcal{M}

► **Remark 5.** Each iteration in (5) requires the computation of a matrix of size $O(n^2)$, each entry of which can be computed in time $O((d_{\max}!)^2)$ with d_{\max} the maximum degree in G and G' . Under the correlated Erdős-Rényi model, $d_{\max} = O\left(\frac{\log n}{\log \log n}\right)$ [2], so that $d_{\max}!$ is polynomial in n . Each iteration is thus polynomial in n and since d is taken order $\Theta(\log(n))$, the algorithm runs in polynomial time.

The authors recall that a full version of this paper can be found at:

<https://arxiv.org/abs/2107.07623>.

³ As already noted in [11], this first strategy produces in fact many false positive matches, e.g. it might match i with v if (i, j) is an edge of G and $(v, \sigma^*(j))$ is an edge of G' . To fix this, we use the *dangling trees trick*, already introduced in [11] (and improved here by consideration of three rather than two dangling trees): instead of just looking at their neighborhoods, we look for the downstream trees from distinct neighbors of i in \mathbf{G} and of u in \mathbf{G}' . The trick is now to match i with u if and only if there exists three distinct neighbors j_1, j_2, j_3 of i in \mathbf{G} (resp. v_1, v_2, v_3 of u in \mathbf{G}') such that all three of the likelihood ratios $L_{d-1}(j_t \leftarrow i, v_t \leftarrow u)$ for $t \in \{1, 2, 3\}$ are larger than β (see Algorithm 1). The proof of Theorem 7 in the full version [12] explains how this trick avoids false positives and why three dangling trees is a good choice.

References

- 1 Jessica E. Banks, C. Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. *ArXiv*, abs/1607.01760, 2016.
- 2 Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001. doi:10.1017/CB09780511814068.
- 3 Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *The Annals of Probability*, 46(1):1–71, 2018. doi:10.1214/16-AOP1142.
- 4 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. doi:10.1142/S0218001404003228.
- 5 Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs, 2017. arXiv:1711.06783.
- 6 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. arXiv:1809.03553.
- 7 Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *arXiv e-prints*, page arXiv:1811.07821, November 2018. arXiv:1811.07821.
- 8 Osman Emre Dai, Daniel Cullina, and Negar Kiyavash. Database Alignment with Gaussian Features. *arXiv e-prints*, page arXiv:1903.01422, March 2019. arXiv:1903.01422.
- 9 Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality, 2019. arXiv:1907.08883.
- 10 Luca Ganassali, Marc Lelarge, and Laurent Massoulié. Impossibility of partial recovery in the graph alignment problem, 2021. arXiv:2102.02685.
- 11 Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Machine Learning Research*, volume 125, pages 1633–1665. PMLR, 09–12 July 2020. URL: <http://proceedings.mlr.press/v125/ganassali20a.html>.
- 12 Luca Ganassali, Laurent Massoulié, and Marc Lelarge. Correlation detection in trees for partial graph alignment. *arXiv e-prints*, page arXiv:2107.07623, July 2021. arXiv:2107.07623.
- 13 Georgina Hall and Laurent Massoulié. Partial Recovery in the Graph Alignment Problem. *arXiv e-prints*, page arXiv:2007.00533, July 2020. arXiv:2007.00533.
- 14 Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013. doi:10.1073/pnas.1312486110.
- 15 Konstantin Makarychev, Rajsekar Manokaran, and Maxim Sviridenko. Maximum quadratic assignment problem: Reduction from maximum label cover and lp-based approximation algorithm. *CoRR*, abs/1403.7721, 2014. arXiv:1403.7721.
- 16 Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016. doi:10.1214/15-AAP1145.
- 17 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. doi:10.1109/SP.2008.33.
- 18 Panos Pardalos, Franz Rendl, and Henry Wolkowicz. *The Quadratic Assignment Problem: A Survey and Recent Developments*, pages 1–42. DIMACS/AMS, August 1994. doi:10.1090/dimacs/016/01.
- 19 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. doi:10.1073/pnas.0806627105.
- 20 Yihong Wu, Jiaming Xu, and Sophie H. Yu. Settling the sharp reconstruction thresholds of random graph matching. *ArXiv*, abs/2102.00082, 2021.