Math. Stat. Learn. 9 (2026), 247–335
DOI 10.4171/MSL/58

Asymmetric graph alignment and the phase transition for asymmetric tree correlation testing

Jakob Maier and Laurent Massoulié

Abstract. Graph alignment – identifying node correspondences between two graphs – is a fundamental problem with applications in network analysis, biology, and privacy research. While substantial progress has been made in aligning correlated Erdős–Rényi graphs under symmetric settings, real-world networks often exhibit asymmetry in both node numbers and edge densities. In this work, we introduce a novel framework for asymmetric correlated Erdős–Rényi graphs, generalizing existing models to account for these asymmetries. We conduct a rigorous theoretical analysis of graph alignment in the sparse regime, where local neighborhoods exhibit tree-like structures. Our approach leverages tree correlation testing as the central tool in our polynomial-time algorithm, MPAlign, which achieves one-sided partial alignment under certain conditions.

A key contribution of our work is characterizing these conditions under which asymmetric tree correlation testing is feasible: If two correlated graphs G and G' have average degrees λs and $\lambda s'$ respectively, where λ is their common density and s, s' are marginal correlation parameters, their tree neighborhoods can be aligned if $ss' > \alpha$, where α denotes Otter's constant and λ is supposed large enough. The feasibility of this tree comparison problem undergoes a sharp phase transition since $ss' \leq \alpha$ implies its impossibility. These new results on tree correlation testing allow us to solve a class of random subgraph isomorphism problems, resolving an open problem in the field.

Contents

Mathematics Subject Classification 2020: 05C80 (primary); 62B10 (secondary).
Keywords: graph alignment, Erdős–Rényi graphs, subgraph isomorphism problem.

1. Introduction

For a pair of simple graphs (G, G') with randomly labeled node sets (V, V'), there are several prominent questions about the joint structure of G and G':

- Are they structurally identical in the sense that there is a bijective function between V and V' mapping edges to edges and non-edges to non-edges? This is the *graph isomorphism problem* (GIP).
- Can G be embedded as a subgraph of G', meaning that there is an injective mapping $V \hookrightarrow V'$ preserving the presence of edges? This is the *subgraph isomorphism problem* (SIP).
- Without any assumption on the graphs, which one-to-one mapping between their nodes maximizes edge overlap? This is the *graph alignment problem* (also called graph matching problem) formulated as an instance of the quadratic assignment problem (QAP): Denoting the adjacency matrices of G and G' as A and A', this problem reads

$$\sigma_{\text{QAP}} \in \operatorname*{arg\,max}_{\substack{\sigma: V \to V', \\ \sigma \text{ injective}}} \sum_{i,j \in V} A_{i,j} A'_{\sigma(i),\sigma(j)}. \tag{QAP}$$

These three problems are increasingly difficult since solving one yields a solution for the ones listed above. While graph isomorphisms can be found in quasi-polynomial time [5], the subgraph isomorphism problem is NP-hard since it can detect a Hamiltonian path by taking G to be the cycle of length $|V'|$. Even more difficult is (QAP), which solves the traveling salesman, k-clique, and subgraph isomorphism problems, among many others [12].

While aligning graphs is NP-hard in the worst case, real-world questions are not necessarily of this complexity. Applications typically consider instances of the graph alignment problem somewhere on a spectrum between hard QAP and easy GIP, potentially allowing for polynomial time algorithms to recover node correspondences.

To model this interpolation between finding simple isomorphisms and solving worst-case QAP, a line of research initiated by [42] has considered G and G' correlated random graphs. This can be seen as an average case analysis of the graph alignment problem, allowing statistical tools to be used. In contrast to the QAP formulation, this approach supposes the existence of a true node correspondence $\sigma_*: V \to V'$ planted into the model. The goal of *planted graph alignment* is to find an estimator $\widehat{\sigma}$ which coincides with σ_* as much as possible.

The existing theory on correlated random graph alignment supposes that G and G' have identical node quantities and the same expected edge numbers. This symmetric framework is restrictive in many applications: Protein interaction graphs [2] have different edge densities across species, brain graphs [47] have different node numbers, and social network analysis requires the retrieval of subgraphs within a big graph [23].

The present paper addresses all of these issues by introducing the novel, more general framework of *asymmetric correlated Erdős–Rényi graphs*, which allows for varying edge densities and differing node numbers. This model can be seen as an average-case interpolation between (QAP) on the one hand, and the subgraph isomorphism problem (SIP) – which is strictly more difficult than the GIP – on the other hand.

In this work, we consider *sparse graphs* with constant average degree, a regime where one cannot recover all node correspondences. We focus on the notion of *one-sided partial alignment* where the goal is to recover a significant portion of node matchings while making a vanishing fraction of errors.

A detailed discussion of the sparse setting is provided in Section 2.1, along with a formal description of the random graph model. For this introductory overview, we give the following definitions:

- *Erdős–Rényi model.* A random graph $G \sim \mathrm{ER}(n, p)$ consists of n nodes where each edge between distinct vertices exists independently with probability p.
- *Parameters.* The correlated asymmetric model is parametrized by a global density level $\lambda > 0$ and correlation parameters $s, s' \in [0, 1]$.
- *Correlated asymmetric model (informal description).* Given an intersection graph $G_* \sim \mathrm{ER}(n_*, \lambda s s'/n_*)$, we add edges and nodes twice independently: Once to obtain $G \sim \mathrm{ER}(n, \lambda s/n)$ and once to get $G' \sim \mathrm{ER}(n', \lambda s'/n')$. The vertex counts n, n' are coupled through n_*, while the graphs G, G' are correlated through G_*.

This enables an informal presentation of our main results.

Theorem 1.1 (Main results – informal statement). *For a pair of sparse asymmetric correlated Erdős–Rényi graphs (G, G') with parameters $\lambda > 0$ and $s, s' \in [0, 1]$ the following statements hold:*

(i) *The polynomial time algorithm* MPALIGN *achieves one-sided partial alignment under the condition that asymmetric tree correlation testing is feasible (see Theorem 3.10).*

(ii) *There is a sharp phase transition for the feasibility of asymmetric tree correlation testing around $\alpha \approx 0.338$ known as Otter's constant: Testing is possible if $ss' > \alpha$ and $\lambda > \lambda_0(s, s')$; it is impossible if $ss' \le \alpha$ (see Theorem 4.10).*

Taking $s < 1$ and $s' = 1$, our model allows G to be a random subgraph of G'. To our knowledge, we contribute the first polynomial-time algorithm to solve this random instance of the SIP. The question about isomorphisms of random graphs is motivated by the work of Babai and Bollobás [6, 9] on efficient *canonical labeling algorithms* that solve (GIP). Extending these results to *random subgraph isomorphisms is an open problem* raised in personal communication with Jiaming Xu.[1] We provide an answer to one instance of that problem.

1.1. Related work

Applications. Research on graph alignment is motivated by a diverse set of applications [15], and has gained popularity through the success in network de-anonymization and consequent privacy breaches during the late 2000s [39, 40]. Researchers in various disciplines have developed graph alignment algorithms for their specific use, such as aligning protein interaction graphs [2, 45], enabling computer vision [7, 16], natural language processing [29], and cell biology [13].

Graph alignment theory. The rigorous theoretical study of graph alignment algorithms was spearheaded by the introduction of correlated random graphs in [42]. The correlated Erdős–Rényi model has since then been thoroughly explored, essentially from two points of view: Information theoretically – *under which parameters is alignment feasible?* [17,18,36,50], and computationally – *which polynomial algorithm can retrieve an alignment?* [4, 22, 24, 37, 51, 52]. All of these papers have in common that they consider either the *dense regime* or the *intermediate regime* where the average node degree across G and G' is growing with the number of nodes. In a notable recent advancement, [21] presents a polynomial-time alignment algorithm for the case with an average degree barely above constant.

Asymmetry. We present one of the first theoretical results on asymmetric graph alignment, but this question has sparked interest in related fields. One example originates from the optimal transport community, where *Unbalanced Gromov–Wasserstein alignment* allows one to compare networks of different sizes [44, 47]. Another example comes from random graph theory and probabilistic subgraph matching [11, 43]. Since our work allows for G to be a subgraph of G', our results enable the matching of such graphs.

To our knowledge, only the authors of [33] conduct a thorough theoretical analysis of graph alignment in the setting with differing node numbers. Their notion of *partially correlated Erdős–Rényi graphs* resembles our correlated graph model but leaves out varying edge densities, and no polynomial matching algorithm is provided.

Sparsity. A branch of research spearheaded by [19, 20, 30] has focused on the *sparse regime*, where the average node degree remains constant. In this setting, Erdős–Rényi

[1] J. Xu, personal communication at MiFODS workshop, MIT (Cambridge, MA, USA), August 19–22, 2019.

graphs are disconnected, making graph matching impossible outside the giant component [26]. However, the local tree-like structure of such graphs has led to intriguing questions about tree correlation testing for graph alignment [25], which were largely resolved in [27, 28]. Complementarily, [37] proposes an algorithm which uses particular rooted-tree substructures called *chandeliers*. This approach yields zero-mismatch partial alignment guarantees around Otter's constant in the sparse regime, while also being applicable to denser regimes.

1.2. Main contributions

We introduce the novel asymmetric correlated Erdős–Rényi model and extend the main results of [27] and [28] to that setting. This extension is non-trivial and yields a substantially stronger result. Notably, the phase transition in (ii) of Theorem 1.1 depends only on the *product* ss'. This simplicity enables a broader interpretation: the density of one graph can compensate for the sparsity of the other, a regime strictly harder than those considered previously. Moreover, local methods, as used here, naturally accommodate asymmetry. This is unlike spectral or first-order optimization approaches, which typically require both graphs to have the same number of nodes.

The non-triviality of this generalization is already apparent in the likelihood ratio diagonalization formula (Theorem 4.7), which differs meaningfully from [28, Theorem 4]: In our case, the functions f are parametrized by (λs) and $(\lambda s')$, not just λ. This asymmetry propagates through the subsequent proofs, requiring considerable care and new work. We therefore make explicit many technical details omitted in [27] and [28] and adjust the presentation for greater transparency. While these generalizations may seem natural in hindsight, executing them rigorously is a main contribution.

To further illustrate how this work goes beyond a direct generalization of [27] and [28], we list some major novelties:

(a) Theorem 3.10 uses concentration of node count measures, distinguishes between matchable and unmatchable nodes, and adjusts the "dangling tree" argument (see Lemma 3.7). None of these are necessary in the symmetric setting.

(b) Lemma 3.11 replaces the 9-line argument in [28, Appendix C.3] with a 7-page proof using graph exploration techniques, correcting an oversight in the earlier treatment of joint graph laws.

(c) Theorem 4.7 and Lemma 4.13 carefully handle convergence of countable sums over unlabeled trees – a crucial subtlety in the asymmetric setting.

Preliminary results appeared in the workshop paper [35]; this document presents a complete and significantly extended version.

1.3. Organization of the paper

Section 2 introduces our asymmetric correlated Erdős–Rényi model in the sparse regime and formalizes the recovery target, namely one-sided partial alignment.

Section 3 then reduces graph alignment to a local comparison of neighborhood trees and presents our polynomial-time algorithm MPALIGN, establishing algorithmic guarantees in Theorem 3.10. This corresponds to point (i) of Theorem 1.1.

Finally, Section 4 studies the underlying *asymmetric tree correlation testing* problem: It derives the feasibility condition, proves the sharp phase transition stated in Theorem 4.10, and thereby identifies the parameter regime in which MPALIGN succeeds. This corresponds to point (ii) of Theorem 1.1.

Section 5 concludes and the Appendix A collects deferred proofs.

1.4. Notations

Let $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ and write $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$. For $k \in \mathbb{N}_{>0}$, define $[k] := \{1, \dots, k\}$ and denote the sum over all injective mappings from $[k]$ to $[n]$ as

$$\sum_{\sigma \colon [k] \hookrightarrow [n]} .$$

For $a, b \in \mathbb{R}$, let $a \wedge b := \min(a, b)$. The complement of a set S is denoted S^c. For a graph G, denote its vertex set $V(G)$ and write $i \in G$ as a shorthand for $i \in V(G)$. Summing over the nodes of a tree t is denoted $\sum_{i \in t}$. The distance between $i, j \in G$ is

$$\mathrm{dist}_G(i, j) := \inf\big\{\mathrm{length}(p) \mid p \text{ is a path in } G \text{ connecting } i \text{ and } j\big\}.$$

Given two probability measures $\mathbb{P}, \mathbb{Q}$ on a probability space $(\Omega, \mathcal{F})$, denote their product measure $\mathbb{P} \otimes \mathbb{Q}$ and their total variation distance

$$d_{\mathrm{TV}}(\mathbb{P}, \mathbb{Q}) = \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

For $n \in \mathbb{N}$, $p \in [0, 1]$, $\mu > 0$, let $\mathrm{Bin}(n, p)$ and $\mathrm{Poi}(\mu)$ denote the binomial and Poisson distributions. If $B \sim \mathrm{Bin}(n, p)$, $X \sim \mathrm{Poi}(\mu)$, write $\pi_\mu(k) := \mathbb{P}(X = k)$ and recall

$$\pi_\mu(k) = e^{-\mu} \frac{\mu^k}{k!}, \quad \mathbb{P}(B = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for all } k \in \mathbb{N}.$$

To describe asymptotic behavior of expressions $f, g \colon \mathbb{N} \to \mathbb{R} \setminus \{0\}$, we use

$$f(N) \in \begin{Bmatrix} \mathcal{O}(g(N)) \\ \Theta(g(N)) \\ o(g(N)) \end{Bmatrix} \iff \begin{cases} f(N)/g(N) \text{ is bounded above,} \\ f(N)/g(N) \text{ is bounded above and below,} \\ \lim_{N\to\infty} f(N)/g(N) = 0. \end{cases}$$

2. Setting and random graph model

In this section, we introduce a novel model of correlated random graphs that allows for varying edge densities and different numbers of nodes. We subsequently introduce the *overlap*, which quantifies agreement with the planted alignment, and the notion of *one-sided alignment*, which captures our recovery objective.

2.1. Asymmetric correlated Erdős–Rényi graphs

The simplest way to describe the law of two asymmetric correlated Erdős–Rényi graphs (G, G') is to algorithmically detail their sampling process.

For this, fix a node number $N \in \mathbb{N}$ and a density parameter $\lambda > 0$. Furthermore, define node correlation parameters $q, q' \in [0,1]$ and edge correlation parameters $r, r' \in (0,1]$. Two correlated random graphs G, G' are sampled as follows:

(1) Let $n_* \sim \mathrm{Bin}(N, qq')$ and sample $G_* \sim \mathrm{ER}(n_*, rr'(\lambda/N))$, which we refer to as the *intersection graph*.

(2) Take $n_+ \sim \mathrm{Bin}(N, q(1-q'))$ and add n_+ isolated nodes to G_*, which now has a total of $n := n_* + n_+$ nodes. Then, independently for each newly added node v and every already present node $w \in [n] \setminus \{v\}$, add an edge connecting v and w with probability $rr'\lambda/N$. This process is called *node augmentation*.

(3) Next, independently for each pair of unconnected nodes $v \neq w \in [n]$, connect them with probability $r(1-r')\lambda/N$. This process is called *edge augmentation*. Call G the graph resulting from steps (2) and (3).

(4) Repeat steps (2) and (3) independently with parameters $n'_+ \sim \mathrm{Bin}(N, (1-q)q')$ and edge augmentation probability $(1-r)r'\lambda/N$. Let $n' := n_* + n'_+$ and call the resulting graph G'.

All four steps are illustrated with an example in Figure 1.

This sampling process yields two asymmetric correlated Erdős–Rényi graphs G and G' whose joint law shall be denoted as $\mathrm{CER}(N, \lambda, q, q', r, r')$. We define their global correlation parameters as

$$s := qr \quad \text{and} \quad s' := q'r',$$

which lets us write $(G, G') \sim \mathrm{CER}(N, \lambda, s, s')$. This notation hides how precisely the graphs are correlated, but we will later see that (s, s') alone determines whether our alignment algorithm works, ignoring the precise value of (q, q', r, r').

Since $\mathbb{E}[n] = qN$ and $\mathbb{E}[n'] = q'N$, the marginal distributions of both graphs are

$$G \sim \mathrm{ER}\Big(n, \frac{\lambda s}{\mathbb{E}[n]}\Big) \quad \text{and} \quad G' \sim \mathrm{ER}\Big(n', \frac{\lambda s'}{\mathbb{E}[n']}\Big) \tag{2.1}$$

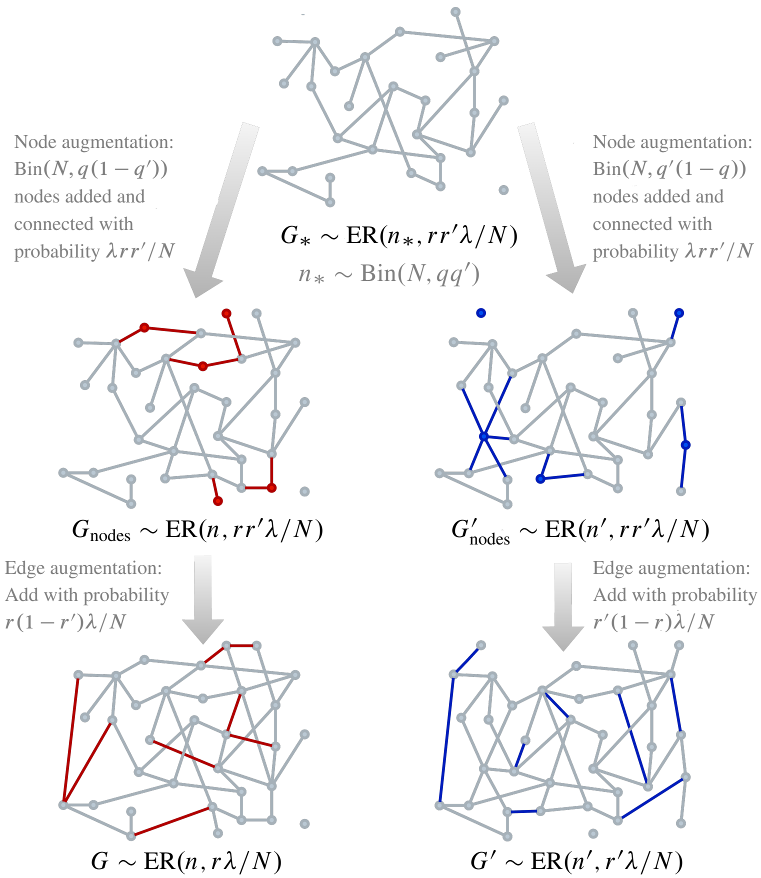

Figure 1. Sampling process of $(G, G') \sim \mathrm{CER}(N, \lambda, q, q', r, r')$. Nodes and edges added during the augmentations are red and blue, respectively.

with expected average degrees $\mathbb{E}[n\lambda s/\mathbb{E}[n]] = \lambda s$ and $\mathbb{E}[n'\lambda s'/\mathbb{E}[n']] = \lambda s'$. Since these average degrees are constant, G and G' are called *sparse* Erdős–Rényi graphs: their adjacency matrices contain a constant number of entries in each column.

Remark 2.1. When $q = q' = 1$, both graphs have the same deterministic node number $n = n' = N$. Therefore, the only difference between G and G' are the average degrees λs and $\lambda s'$. This is what [35] refer to as the *varying edge densities* setting.

In the case where $r = r' = 1$, there may be different node numbers in G and G', but each pair of nodes in both graphs is connected with the same probability $\lambda r/N$. However, unless $q = q'$, their expected average degrees are different, equal to λq and $\lambda q'$, respectively. This is what [35] refer to as the *differing node numbers* setting.

Our definition of $\mathrm{CER}(N, \lambda, q, q', r, r')$ therefore combines both models from [35] into one comprehensive asymmetric Erdős–Rényi model with differing node numbers and varying edge densities at the same time.

Remark 2.2. There are many ways of defining $\mathrm{CER}(N, \lambda, q, q', r, r')$; here, we have chosen to start with an *intersection graph* G_* before performing node and edge *augmentations*. Note that an identically distributed pair of random graphs could equivalently be obtained by starting with a *master graph* $\widehat{G} \sim \mathrm{ER}(N, \lambda/N)$, followed by independent node and edge *deletions*. We refer to [35] for details on the deletion procedure; one can easily verify that both algorithms sample from the same distribution.

2.2. True node correspondences

The planted graph alignment problem consists of retrieving correspondences between the nodes of two graphs. Given a pair $(G, G') \sim \mathrm{CER}(N, \lambda, q, q', r, r')$ from our generative model, there exists a hidden partial mapping between the nodes of G and G', namely via the nodes inherited from G_* that both graphs have in common.

To formalize this, we randomly label the nodes of both G and G' with labels in $[n]$ and $[n']$, respectively (for instance, by choosing the node with number 1 uniformly at random, then picking a second node among the remaining ones, etc.). Then, there are two node subsets $V_* \subset [n]$, $V'_* \subset [n']$ of size n_* and a bijective map $\sigma_*\colon V_* \to V'_*$, which relates all pairs of nodes shared by G and G'.

Remark 2.3. In the setting $q = q' = 1$ where both graphs have n nodes, one has $V_* = V'_* = [n]$ and σ_* is a permutation of $[n]$. If we use the same node labeling for G_* and G, "forgetting" these labels would formally consist of sampling σ_* uniformly among all permutations of $[n]$ and then labeling the nodes of G' as $\sigma_*(1)$, $\sigma_*(2)$, and so on.

Remark 2.4. A key novelty of the $\mathrm{CER}(N, \lambda, q, q', r, r')$ model is step 2 in the sampling process, which allows for different node numbers in both graphs. Omitting this step by picking $q = q' = 1$, one recovers the correlated Erdős–Rényi model typically studied in the literature. Due to the deterministic node numbers in this setting, one can more easily define G and G' via their adjacency matrices A and A': Their correctly permuted entry pairs $(A_{ij}, A'_{\sigma_*(i)\sigma_*(j)})$ are independent tuples of Bernoulli variables with joint distribution

$$\big(A_{ij}, A'_{\sigma_*(i)\sigma_*(j)}\big) = \begin{cases} (1,1) & \text{with probability } (\lambda/n)ss', \\ (1,0) & \text{with probability } (\lambda/n)s(1-s'), \\ (0,1) & \text{with probability } (\lambda/n)s'(1-s), \\ (0,0) & \text{with probability } 1-(\lambda/n)(s+s'-ss'). \end{cases}$$

For $q = q' = 1$, if one additionally supposes $r = r'$, we recover the "standard" correlated Erdős–Rényi model first introduced in [42] and studied extensively, for example, in [27]. In the latter paper and most literature, the parameter λ corresponds to λ/r in our model.

One speaks of *planted graph alignment* since the objective is to recover σ_*, which, by construction, is metaphorically planted into the model, together with the intersection graph G_*. This objective requires us to define an alignment estimator consisting of two sets $\widehat{V} \subset [n]$, $\widehat{V}' \subset [n']$, which ideally contain the nodes of G_*, together with a bijective map $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$ matching nodes in G with nodes in G'. Such an estimator, therefore, consists of the triplet $(\widehat{\sigma}, \widehat{V}, \widehat{V}')$, but we will often just write $\widehat{\sigma}$ to denote all three objects.

We introduce the following performance indicator to measure if such a $\widehat{\sigma}$ is close to the true matching σ_*.

Definition 2.1 (Overlap). For an alignment estimator $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$, define its overlap with the true matching $\sigma_*\colon V_* \to V'_*$ as

$$\mathrm{ov}(\widehat{\sigma}, \sigma_*) := \frac{1}{n_*} \sum_{i \in \widehat{V} \cap V_*} \mathbb{1}_{\widehat{\sigma}(i) = \sigma_*(i)}.$$

Since $n_* = |V_*|$, the overlap measures which proportion of the relevant nodes is correctly matched by $\widehat{\sigma}$. This tells us how well $\widehat{\sigma}$ acts on the true node set but is entirely agnostic to what the estimator does outside of V_*. Consequently, another quantity is required to measure how well $\widehat{\sigma}$ identifies V_* and how many errors it commits in a matching.

Definition 2.2 (Error fraction). For an alignment estimator $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$, one defines its error fraction as

$$\mathrm{err}(\widehat{\sigma}, \sigma_*) := \frac{1}{n} \sum_{i \in \widehat{V}} \mathbb{1}_{i \notin V_* \text{ or } \widehat{\sigma}(i) \neq \sigma_*(i)}.$$

The goal of planted graph alignment is to find a $\widehat{\sigma}$ that can be computed in polynomial time and maximizes the overlap while keeping the error fraction low. For our specific case, we formalize this goal as follows.

2.3. One-sided partial alignment

When seeking to align a pair (G, G') of sparse correlated Erdős–Rényi graphs, it is impossible to recover σ_* completely. The reason for this lies within the structure of G_*, the graph containing all the information about a potential matching between G and G'. Since $G_* \sim \mathrm{ER}(n_*, ss'\lambda/\mathbb{E}[n_*])$, this is a sparse Erdős–Rényi graph with one giant connected component while the remaining graph mainly consists of small trees [9]. This property has been exploited in [26] to prove that a matching between G and G', cannot be done outside this giant component. A more intuitive reason for the impossibility of retrieving σ_* is isolated nodes: Since average node degrees are constant, there is a non-negligible amount of isolated nodes in G_*, which are all indistinguishable.

In the sparse setting, the giant component of G_* contains strictly fewer nodes than the entire graph. Therefore, one can only hope to partially align G with G'. A lower number of matchable nodes also means that every wrongly matched node significantly augments the relative error. This is why authors [19, 27] have focused on alignment estimators that commit a *negligible amount of errors* while *recovering a significant proportion* of the possible graph matching. This is referred to as one-sided partial alignment.

Definition 2.3. For a family of correlated graphs $(G_N, G'_N) \sim \mathrm{CER}(N, \lambda, s, s')$, a sequence of estimators $\widehat{\sigma}_N \colon \widehat{V}_N \to \widehat{V}'_N$ obtained through the same algorithm is said to achieve one-sided partial alignment if there exists $\beta > 0$ such that

$$\mathbb{P}\big(\mathrm{ov}(\sigma_*, \widehat{\sigma}_N) \geq \beta\big) \xrightarrow[N\to\infty]{} 1 \quad \text{and} \quad \mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_N) = o(1)\big) \xrightarrow[N\to\infty]{} 1.$$

We will often omit indexing by N and write $G, G', \widehat{\sigma}$ instead of $G_N, G'_N, \widehat{\sigma}_N$.

Remark 2.5. This definition translates the phrase "$\widehat{\sigma}$ *commits a negligible amount of errors*" into the more precise "*the error fraction of* $\widehat{\sigma}$ *is vanishing with high probability*", which in mathematical terms reads "$\mathbb{P}(\mathrm{err}(\sigma_*, \widehat{\sigma}) = o(1)) \to 1$". More rigorously, this expression signifies that there is a sequence $(a_N)_N$ of $\mathbb{R}_{>0}$ with $a_N \to 0$ and

$$\mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_N) \leq a_N\big) \xrightarrow[N\to\infty]{} 1 \qquad \Longleftrightarrow \qquad \mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_N) > a_N\big) \xrightarrow[N\to\infty]{} 0.$$

We claim that this is equivalent to $\mathrm{err}(\sigma_*, \widehat{\sigma}_N) \to 0$ in probability. To prove the claim, let $\varepsilon > 0$ and take N big enough so that $a_K < \varepsilon$ for all $K > N$. Then, for all such K,

$$\mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_K) > \varepsilon\big) \leq \mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_K) > a_K\big) \xrightarrow[K\to\infty]{} 0.$$

On the other hand, if $\mathrm{err}(\sigma_*, \widehat{\sigma}_N)$ converges in probability, one can take the constant sequence $a_N \equiv \varepsilon$ to obtain the opposite direction immediately. This lets us conclude that

$$\mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}_N) = o(1)\big) \xrightarrow[N\to\infty]{} 1 \qquad \Longleftrightarrow \qquad \mathrm{err}(\sigma_*, \widehat{\sigma}_N) \xrightarrow[N\to\infty]{\text{in } \mathbb{P}} 0.$$

To summarize, sparse Erdős–Rényi graphs admit at best partial alignment, and Definition 2.3 formalizes what we mean by a successful alignment estimator. In the next section, we show that such graphs are locally tree-like, and we leverage this structure to design an efficient alignment algorithm.

3. From graphs to trees and back

Local algorithms are a natural choice for aligning sparse Erdős–Rényi graphs, an intuitive reason being that these graphs have large diameters, and far-apart nodes contain negligible information about each other. Therefore, it instinctively makes sense to align nodes based on their local neighborhoods in the respective graphs.

In this section, we describe the asymptotic structure of these neighborhoods which are Galton–Watson trees. This will allow us to translate the global graph alignment problem to a local tree correlation test. While tree correlation testing deserves a whole section later, we will leave it as a black box here and focus on the algorithm that bridges the gap between graphs and trees, between global structure and local neighborhoods.

Notation for trees. Given $\mu > 0$, the law of a Galton–Watson tree [49] with $\mathrm{Poi}(\mu)$ offspring distribution is denoted as $\mathrm{GW}^{(\mu)}$. Next, for any rooted tree t of potentially infinite size, we denote t_d the *pruning* of t at depth $d \in \mathbb{N}$, i.e., the subtree containing all nodes whose distance to the root is at most d.

For $t \sim \mathrm{GW}^{(\mu)}$, we denote by $\mathrm{GW}_d^{(\mu)}$ the law of the pruned tree t_d. Furthermore, if we consider any fixed tree τ, we denote by $\mathrm{GW}_d^{(\mu)}(\tau)$ the likelihood that the pruning τ_d is of law $\mathrm{GW}_d^{(\mu)}$. Hence, the quantity $\mathrm{GW}_d^{(\mu)}(\tau)$ depends on τ only up to depth d, while ignoring any nodes which are further from the root.

3.1. Tree neighborhoods

As a first step towards understanding neighborhoods in sparse correlated Erdős–Rényi graphs, we describe the asymptotic local structure of a single graph $G \sim \mathrm{ER}(N, \mu/N)$. For a node $i \in G$ and $d \in \mathbb{N}_{>0}$, we denote the neighborhood of i up to depth d as

$$\mathcal{N}_{G,d}(i) = \{j \in V : \mathrm{dist}_G(i, j) \leq d\}.$$

While $\mathcal{N}_{G,d}(i)$ can be seen as a ball of radius d around i, we denote the sphere as

$$\mathcal{S}_{G,d}(i) = \mathcal{N}_{G,d}(i) \setminus \mathcal{N}_{G,d-1}(i) = \{j \in V : \mathrm{dist}_G(i, j) = d\}.$$

The following three lemmata, first presented as [25, Lemmata 2.1, 2.3, and 2.4], describe the behaviors of these neighborhoods.

Lemma 3.1 (Uniform bound on the neighborhood sizes). *Let $G = G_N \sim \mathrm{ER}(N, \mu/N)$ be a family of Erdős–Rényi graphs with $\mu > 1$. Set $d = d_N := \lfloor c \log(N) \rfloor$ for some c fulfilling $c \log(\mu) < 1/2$. Then, for all $\varepsilon > 0$, there exists a constant $C = C(\varepsilon) > 0$ such that*

$$\mathbb{P}\big(\exists i \in V, t \in [d] : |\mathcal{S}_{G,t}(i)| > C \log(N)\mu^t\big) \in \mathcal{O}(N^{-\varepsilon}).$$

Lemma 3.2 (Different neighborhoods are asymptotically independent). *Let G, d, μ, and c be as in Lemma 3.1. Then there exists $\varepsilon > 0$ such that for any pair of nodes $i \neq j \in V$, one has*

$$d_{\mathrm{TV}}\big(\mathrm{Law}\big(\mathcal{N}_{G,d}(i), \mathcal{N}_{G,d}(j)\big), \mathrm{Law}\big(\mathcal{N}_{G,d}(i)\big) \otimes \mathrm{Law}\big(\mathcal{N}_{G,d}(j)\big)\big) \in \mathcal{O}(N^{-\varepsilon}).$$

Consequently, there exists an optimal coupling (see, for instance, [48, Theorem 4.1]*) between $(\mathcal{N}_{G,d}(i), \mathcal{N}_{G,d}(j))$ and $(\mathcal{N}, \mathcal{N}') \sim \mathrm{Law}(\mathcal{N}_{G,d}(i)) \otimes \mathrm{Law}(\mathcal{N}_{G,d}(j))$ fulfilling*

$$\mathbb{P}\big(\big(\mathcal{N}_{G,d}(i), \mathcal{N}_{G,d}(j)\big) \neq (\mathcal{N}, \mathcal{N}')\big) \in o(1).$$

Lemma 3.3 (Neighborhoods look like Galton–Watson trees). *Let G, d, μ, and c be as in Lemma 3.1. Then there exists $\varepsilon > 0$ such that for all nodes $i \in V$,*

$$d_{\mathrm{TV}}\big(\mathcal{N}_{G,d}(i), \mathrm{GW}_d^{(\mu)}\big) \in \mathcal{O}(N^{-\varepsilon}).$$

Asymptotically for $N \to \infty$, Lemma 3.1 allows to control the number of nodes in neighborhoods of size $\mathcal{O}(\log(N))$. Furthermore, Lemma 3.2 states that the neighborhoods of two different fixed nodes behave independently. Finally, Lemma 3.3 affirms that these neighborhoods look as if they were sampled like a Galton–Watson tree with Poisson offspring.

Remark 3.4. Consider the more general case of a graph $G \sim \mathrm{ER}(n, p)$ with random node number $n \sim \mathrm{Bin}(N, q)$ and edge probability $p = \lambda/\mathbb{E}[n]$. Since with high probability, $np = n\lambda/\mathbb{E}[n] \to \lambda$, one can easily adjust the proofs in [25] for the lemmata to remain valid in this random node number case.

We later apply these modified versions of the lemmata to

$$(G, G') \sim \mathrm{CER}(N, \lambda, s, s').$$

To do so, recall from (2.1) that the marginal laws are $G \sim \mathrm{ER}(n, \lambda s/\mathbb{E}[n])$ and $G' \sim \mathrm{ER}(n', \lambda s'/\mathbb{E}[n'])$. The depth-$d$-neighborhoods of all nodes i in G will therefore converge towards $\mathrm{GW}_d^{(\lambda s)}$ and those of $j \in G'$ towards $\mathrm{GW}_d^{(\lambda s')}$. The same reasoning holds for

$$G_* \sim \mathrm{ER}\big(n_*, \lambda r r' q q'/(N q q')\big) = \mathrm{ER}\big(n_*, \lambda s s'/\mathbb{E}[n_*]\big),$$

whose node neighborhoods asymptotically behave as $\mathrm{GW}_d^{(\lambda s s')}$.

3.2. Correlated Galton–Watson trees

From the way we have defined correlated sparse Erdős–Rényi graphs in Section 2.1, it is straightforward to transpose the results about single graph neighborhoods to joint neighborhoods of graph couples. Recall that all information about the correlation of a

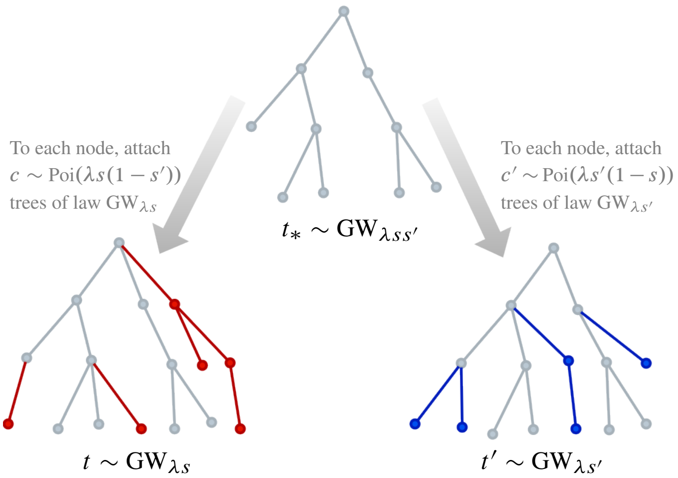

Figure 2. Sampling process of $(t, t') \sim \mathbb{P}^{\mathrm{corr}}$ up to depth 3. The blue and red components are independently sampled and attached to the grey intersection tree.

pair $(G, G') \sim \mathrm{CER}(N, \lambda, s, s')$ is contained in the intersection graph G_*. Translating this to the local level, tree-like neighborhoods of the same node in G and G' are correlated via an *intersection tree* corresponding to the neighborhood in G_*. This leads us to introduce the following model of correlated Galton–Watson trees.

Definition 3.1 (Correlated Galton–Watson trees). For $\lambda > 0$ and $s, s' \in [0, 1]$, define the law of correlated Galton–Watson trees $(t, t') \sim \mathbb{P}^{\mathrm{corr}}$ via the following sampling algorithm:

(1) Sample a rooted tree $t_* \sim \mathrm{GW}^{(\lambda s s')}$, which we refer to as the intersection tree.

(2) Independently for each node i of t_*, sample $c^+(i) \sim \mathrm{Poi}(\lambda s(1-s'))$ and attach $c^+(i)$ additional child nodes to i.

(3) Then, for each child node j added in step 2, independently sample a tree $t^+(j) \sim \mathrm{GW}^{(\lambda s)}$ and attach it, making j its root.

(4) Call the resulting tree t. Then repeat steps 2 and 3 independently with a copy of t_*, but change the laws $c^+(i) \sim \mathrm{Poi}(\lambda s'(1-s))$, $t^+(j) \sim \mathrm{GW}^{(\lambda s')}$, and call the result t'.

Figure 2 illustrates this sampling process. While the procedure yields a pair of potentially infinitely deep trees (t, t'), we will often consider their depth–d-truncated versions (t_d, t'_d) whose law we denote as $\mathbb{P}_d^{\mathrm{corr}}$.

For a pair of trees $(t, t') \sim \mathbb{P}_d^{\mathrm{corr}}$, independently assign random labels to their nodes, like we do for a pair of graphs in Section 2.2. The correlation between t and t'

then relies on the embedded intersection tree t_*. It can be expressed via a mapping $\sigma_*: V_* \to V'_*$, where V_* and V'_* are the node subsets of t and t' corresponding to the nodes of t_*.

The trees from Definition 3.1 are intimately related to sparse correlated Erdős–Rényi graphs. This is formalized in the following lemma, which is an adaptation of [27, Lemma 6.4], and its proof follows directly from Lemma 3.3.

Lemma 3.5. *Given $(G, G') \sim \mathrm{CER}(N, \lambda, s, s')$, let $(i, \sigma_*(i)) \in V_* \times V'_*$ be a pair of aligned nodes in the sense of Section 2.2. If, furthermore, $d = \lfloor c \log(N) \rfloor$ for c fulfilling $2c \log(\lambda \max\{s, s'\}) < 1$, then there exists $\varepsilon > 0$ such that*

$$d_{\mathrm{TV}}\big((\mathcal{N}_{G,d}(i), \mathcal{N}_{G',d}(\sigma_*(i))), \mathbb{P}_d^{\mathrm{corr}}\big) \in \mathcal{O}(N^{-\varepsilon}).$$

Hence, there exists an optimal coupling [48, Theorem 4.1] *between the neighborhood couple $(\mathcal{N}_{G,d}(i), \mathcal{N}_{G',d}(\sigma_*(i)))$ and a pair of trees $(t, t') \sim \mathbb{P}_d^{\mathrm{corr}}$ fulfilling*

$$\mathbb{P}\big((\mathcal{N}_{G,d}(i), \mathcal{N}_{G',d}(\sigma_*(i))) \neq (t, t')\big) \in o(1).$$

Remark 3.6. It may seem surprising that the global correlation parameters s, s' are sufficient in Definition 3.1 to describe the local tree laws, ignoring the fact whether edges or nodes were added during the sampling process of (G, G'). The reason is that adding a new edge or a new node is equivalent in a tree: Both operations result in adding one leaf and one edge. Hence, only the products of edge and node correlation parameters matter, i.e., $s = rq$ and $s' = r'q'$.

Having described the neighborhoods of correctly matched nodes $(i, \sigma_*(i))$, the following section describes a way of distinguishing these from incorrect matches.

3.3. One-sided hypothesis testing

Given two nodes $i \in G$ and $j \in G'$, there are two cases: Either they should be aligned, i.e., $i \in V_*$ and $\sigma_*(i) = j$, in which case Lemma 3.5 characterizes the joint law $\mathbb{P}^{\mathrm{corr}}$ of their neighborhoods. Or they should not be matched, i.e., $i \neq V_*$ or $\sigma_*(i) \neq j$. Intuitively, if i and j are far apart in G_*, the neighborhood laws of i and j are roughly independent, in the sense that these neighborhoods are approximate samples of the following joint tree law: Let $t \sim \mathrm{GW}^{(\lambda s)}$ and $t' \sim \mathrm{GW}^{(\lambda s')}$ with t, t' independent, and define

$$\mathbb{P}^{\mathrm{ind}} := \mathrm{Law}(t, t') = \mathrm{GW}^{(\lambda s)} \otimes \mathrm{GW}^{(\lambda s')}.$$

Furthermore, for a depth $d \in \mathbb{N}$, we will denote by $\mathbb{P}_d^{\mathrm{ind}}$ the law of the truncated trees (t_d, t'_d).

The approximate independence of the tree neighborhoods of i and j can be established regardless of their distance in G_* by employing the following technique.

Dangling tree trick. Instead of comparing the entire neighborhoods of two nodes, we focus on subsets of these neighborhoods that are *pointing away* from the nodes. Formally, for two adjacent nodes i and i' of G, we define

$$\mathcal{N}_{G,d}(i \to i') := \{j \in \mathcal{N}_{G,d}(i) : \mathrm{dist}_G(j,i') < \mathrm{dist}_G(j,i)\}.$$

If $\mathcal{N}_{G,d}(i)$ is cycle-free, it can be viewed as a tree rooted in i. The subset $\mathcal{N}_{G,d}(i \to i')$ then corresponds to the subtree rooted in i' which has depth $d-1$. Visually speaking, $\mathcal{N}_{G,d}(i \to i')$ describes the neighborhood rooted in i' that is pointing away from i.

Why introduce $\mathcal{N}_{G,d}(i \to i')$? Take $(G,G') \sim \mathrm{CER}(N,\lambda,s,s')$ and recall that their common subgraph G_* is embedded in G and G' via the node sets $V_* \subset V$ and $V'_* \subset V'$, respectively. The following lemma and definition let us bridge the gap between pointing neighborhoods and approximately independent trees.

Lemma 3.7. *For $(G,G') \sim \mathrm{CER}(N,\lambda,s,s')$, let $i \in V$, $j \in V'$ and suppose that*

(i) *either $i \notin V_*$ or $\sigma_*(i) \neq j$;*

(ii) *$\mathcal{N}_{G,2d}(i)$ and $\mathcal{N}_{G',2d}(j)$ are cycle-free;*

(iii) *both nodes have at least three neighbors, named i_1, i_2, i_3 and j_1, j_2, j_3.*

Then, there is at least one pair (i_k, j_k) such that

$$\mathcal{N}_{G,d}(i \to i_k) \cap V_* \quad \text{and} \quad \mathcal{N}_{G,d}(j \to j_k) \cap V'_*$$

have disjoint node sets when embedded in G_*.

Definition 3.2. A pair of trees (t,t') of depth d is said to follow the law $\mathbb{P}^{\mathrm{dis}}_{d-1}$ if they are sampled as $\mathcal{N}_{G,d}(i \to i_k) \cap V_*$ and $\mathcal{N}_{G,d}(j \to j_k) \cap V'_*$ in the above lemma conditioned on (i), (ii), and (iii), and the fact that they have disjoint node sets when embedded in G_*.

In other words, $\mathbb{P}^{\mathrm{dis}}_{d-1}$ is the law of a tree-like neighborhood pair in correlated Erdős–Rényi graphs, conditioned to avoid each other in the intersection graph G_*.

Remark 3.8. This is called the dangling tree trick because Lemma 3.7 guarantees the existence of disjoint subtrees in the neighborhoods $\mathcal{N}_{G,d}(i)$ and $\mathcal{N}_{G',d}(j)$. Since these neighborhoods are assumed to be trees, their pointing subtrees are figuratively *dangling* from their respective roots i and j. The usefulness of this trick stems from the similarity of the laws $\mathbb{P}^{\mathrm{dis}}_d$ and $\mathbb{P}^{\mathrm{ind}}_d$, which we have only hinted at so far and will show later in Lemma 3.11.

Proof of Lemma 3.7. Let τ denote the subgraph of G_* which is spanned by the nodes of $\mathcal{N}_{G,d}(i) \cap V_*$ and rooted in $i_* \in \arg\min_{k \in V_*} d_G(i,k)$. Since $\mathcal{N}_{G,2d}(i)$ is a tree, τ is also cycle-free. Define τ' similarly, as the subgraph with node set $\mathcal{N}_{G',d}(j) \cap V'_*$ rooted in $j_* \in \arg\min_{k \in V_*} d_{G'}(j,k)$.

On the one hand, if τ and τ' do not share any nodes in G_*, they are disjoint subtrees of the same Erdős–Rényi graph. When augmenting both of these subgraphs to become $\mathcal{N}_{G,d}(i)$ and $\mathcal{N}_{G',d}(j)$ during the sampling procedure of (G, G'), they remain disjoint subtrees of those graphs, potentially changing roots from i_* to i and from j_* to j. Since these trees share no nodes, their subtrees do not share any either, and consequently all pairs $(\mathcal{N}_{G,d}(i \to i_m), \mathcal{N}_{G',d}(j \to j_m))$ are of law $\mathbb{P}^{\mathrm{dis}}_{d-1}$.

On the other hand, if τ and τ' share a node in G_*, then there is exactly one path $\mathcal{P}_{i_* \leftrightarrow j_*}$ connecting i_* with j_* and contained in $[\mathcal{N}_{G,d}(i) \cap V_*] \cup [\mathcal{N}_{G',d}(j) \cap V'_*]$: if there was another such path $\mathcal{P}'$, then the union of $\mathcal{P}_{i_* \leftrightarrow j_*}$ and $\mathcal{P}'$ would be a cycle in $\mathcal{N}_{G,2d}(i)$, which is a contradiction to (ii).

Since $\mathcal{N}_{G,d}(i) \cap V_*$ is a tree, all pointing subtrees $\mathcal{N}_{G,d}(i \to i_m) \cap V_*$, $m \in [3]$ are disjoint, and the same holds for $\mathcal{N}_{G,d}(j \to j_m) \cap V'_*$, $m \in [3]$.

Consequently, $\mathcal{P}_{i_* \leftrightarrow j_*}$ can only be contained in two of the three sets

$$\left[\mathcal{N}_{G,d}(i \to i_m) \cap V_*\right] \cup \left[\mathcal{N}_{G',d}(j \to j_m) \cap V'_*\right], \quad m \in [3].$$

In other words, there is $m_* \in [3]$ such that $\mathcal{P}_{i_* \leftrightarrow j_*}$ is disjoint from

$$\left[\mathcal{N}_{G,d}(i \to i_{m_*}) \cap V_*\right] \cup \left[\mathcal{N}_{G',d}(j \to j_{m_*}) \cap V'_*\right].$$

Therefore, $\mathcal{N}_{G,d}(i \to i_{m_*}) \cap V_*$ and $\mathcal{N}_{G',d}(j \to j_{m_*}) \cap V'_*$ share no nodes. Augmenting these two tree-neighborhoods in their respective graphs, one obtains that the couple $\mathcal{N}_{G,d}(i \to i_{m_*}), \mathcal{N}_{G',d}(j \to j_{m_*})$ follows the law $\mathbb{P}^{\mathrm{dis}}_{d-1}$. ∎

The dangling tree trick and the later Lemma 3.11 allow us to consider pairs of independent Galton–Watson trees and compare them to pairs of trees with law $\mathbb{P}^{\mathrm{corr}}_d$.

Given a pair of rooted trees (t, t') and their truncations (t_d, t'_d), our goal is to determine whether they are correlated or independent. For any depth d, this translates to the hypothesis testing problem

$$\begin{aligned} H_0^{(d)} &: (t_d, t'_d) \sim \mathbb{P}^{\mathrm{ind}}_d, && \text{i.e., the trees are independent,} \\ \text{vs. } H_1^{(d)} &: (t_d, t'_d) \sim \mathbb{P}^{\mathrm{corr}}_d, && \text{i.e., the trees are correlated.} \end{aligned} \tag{3.1}$$

In the asymptotic regime where $d \to \infty$, the goal is to find a sequence of tests $\mathcal{T}_d$ to distinguish between $H_0^{(d)}$ and $H_1^{(d)}$ at all depths d. It turns out that the correct performance measure for the test we are looking for is defined below; this definition uses the notation $\mathcal{X}_d :=$ {rooted, unlabeled trees of depth $\leq d$}, and we refer to Section 4.1 for a rigorous introduction of $\mathcal{X}_d$.

Definition 3.3 (One-sided test). We call a function $\mathcal{T}_d \colon \mathcal{X}_d \times \mathcal{X}_d \to \{0, 1\}$ a *test at depth d* and let $\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d = 0)$ denote the probability that a pair of truncated independent trees $(t_d, t'_d) \sim \mathbb{P}^{\mathrm{ind}}_d$ is correctly identified by $\mathcal{T}_d$, which is to say $\mathcal{T}_d(t_d, t'_d) = 0$.

An *asymptotic one-sided test* is then defined as a sequence of functions $(\mathcal{T}_d)_{d\in\mathbb{N}}$ with the following two properties:

- $\mathbb{P}_d^{\mathrm{ind}}(\mathcal{T}_d = 1) \xrightarrow[d\to\infty]{} 0$, i.e., the tests have vanishing type-I error;
- there exists $\beta > 0$ such that $\mathbb{P}_d^{\mathrm{corr}}(\mathcal{T}_d = 1) \geq \beta$ for all but finitely many d, meaning that the tests have significant power β. In other words,

$$\liminf_{d\to\infty} \mathbb{P}_d^{\mathrm{corr}}(\mathcal{T}_d = 1) > 0.$$

The entirety of Section 4 is dedicated to the question of whether one-sided tests exist for different parameter choices of (λ, s, s') and how to define $(\mathcal{T}_d)_{d\in\mathbb{N}}$. For now, we consider those results a black box and suppose that one-sided tests exist, allowing us to define a graph matching algorithm.

3.4. An algorithm to align graphs using tree comparison

Throughout this section, we assume that λ, s, and s' are such that there is a sequence of tests $(\mathcal{T}_d)_{d\in\mathbb{N}}$ for the problem (3.1) which fulfill $\lambda s s' > 1$ and

$$\mathbb{P}_d^{\mathrm{ind}}(\mathcal{T}_d = 1) \leq \exp\big(-C(\lambda s s')^d\big) \quad \text{and} \quad \beta := \liminf_{d\to\infty} \mathbb{P}_d^{\mathrm{corr}}(\mathcal{T}_d = 1) > 0. \tag{3.2}$$

Following our above discussion, the idea for a graph matching algorithm is straightforward: For all pairs of nodes $i \in G$ and $j \in G'$, whose neighborhoods are trees, compare their dangling subtrees using $\mathcal{T}_d$ and decide whether to match i with j. The following algorithm was first introduced in [27]:

Algorithm 1 MPALIGN (message-passing for sparse graph alignment)

Input: Two graphs G, G', a depth parameter d, and a test $\mathcal{T}_d$.
Output: An alignment estimator $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$ where $\widehat{V}, \widehat{V}'$ are node subsets of G, G'.

Initialize $\widehat{V} \leftarrow \emptyset$, $\widehat{V}' \leftarrow \emptyset$ and let $\widehat{\sigma}$ be the empty map,
for $i \in V(G)$ and $j \in V(G')$ **do**
  if neither $\mathcal{N}_{G,2d}(i)$ nor $\mathcal{N}_{G',2d}(j)$ contain cycles, **then**
    if there exist triplets $i_1, i_2, i_3 \in \mathcal{N}_{G,1}(i)$ and $j_1, j_2, j_3 \in \mathcal{N}_{G',1}(j)$ such that

$$\mathcal{T}_{d-1}\big(\mathcal{N}_{G,d}(i \to i_k), \mathcal{N}_{G',d}(j \to j_k)\big) = 1 \text{ for all } k \in \{1,2,3\},$$

then $\widehat{V} \leftarrow \widehat{V} \cup \{i\}$, $\widehat{V}' \leftarrow \widehat{V}' \cup \{j\}$ and set $\widehat{\sigma}(i) = j$.
    end if
  end if
end for
return $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$

Remark 3.9 (Runtime of MPALIGN). We choose $d \in \mathcal{O}(\log(N))$ and show in Proposition 4.3, that $\mathcal{T}_d$ can be computed in polynomial time. Since MPALIGN is essentially a *for*-loop over all $\mathcal{O}(N^2)$ pairs of nodes, this algorithm is therefore polynomial as well. This is the minimum requirement for an efficient graph alignment algorithm.

In the following theorem, we establish that MPALIGN achieves one-sided partial alignment.

Theorem 3.10. *Let $\lambda > 0$ and $s, s' \in [0,1]$ be such that there exist tests $(\mathcal{T}_d)_{d\in\mathbb{N}}$ satisfying the one-sided feasibility conditions in (3.2). Define $(G, G') \sim \mathrm{CER}(N, \lambda, s, s')$ and let $d = \lfloor c \log(N) \rfloor$, where c satisfies $4c \log(\lambda \max\{s, s'\}) < 1$. Then, Algorithm 1 with inputs $(G, G', d, \mathcal{T}_d)$ yields an estimator $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$ which achieves one-sided partial alignment. In other words, there exists $\widetilde{\beta} > 0$ such that*

$$\mathbb{P}\big(\mathrm{ov}(\sigma_*, \widehat{\sigma}) \geq \widetilde{\beta}\big) \xrightarrow[N\to\infty]{} 1, \tag{3.3}$$

$$\mathbb{P}\big(\mathrm{err}(\sigma_*, \widehat{\sigma}) = o(1)\big) \xrightarrow[N\to\infty]{} 1. \tag{3.4}$$

The proof of this theorem builds on the ideas in [27, Section 6], providing additional details and extending the analysis to settings with random node numbers and unequal edge densities.

Although technically involved, the proof offers valuable insights into the interplay between graph- and tree-level structures in the asymmetric setting. This result concludes the chapter and sets the stage for the next section, where we justify the feasibility assumptions in (3.2).

Proof. We start by recalling that the node numbers of G_*, G and G' marginally follow binomial distributions with means

$$\mathbb{E}[n_*] = Nqq, \quad \mathbb{E}[n] = Nq, \quad \mathbb{E}[n'] = Nq'.$$

Throughout this proof, we consider these numbers to be sampled before the remaining sampling process from Definition 3.1. To make this rigorous, define

$$\gamma := \frac{1}{4}\Big(\frac{1}{2} - c \log\big(\lambda \max\{s, s'\}\big)\Big) \quad \text{and} \quad \varepsilon := N^{1/2+\gamma},$$

and condition on the event

$$\mathcal{B} := \{n_* > Nqq' - \varepsilon\} \cap \{n - \varepsilon < Nq < n + \varepsilon\} \cap \{n' - \varepsilon < Nq' < n' + \varepsilon\}.$$

Note that $\gamma > 0$ because $2c \log(\lambda \max\{s, s'\}) \leq 4c \log(\lambda \max\{s, s'\}) < 1$, as assumed in Theorem 3.10. Consequently, using Chernoff's inequality for binomial random variables [14, Theorem 4], the event $\mathcal{B}$ occurs with high probability:

$$\begin{aligned}\mathbb{P}(\mathcal{B}) &\geq 1 - \mathbb{P}\big(n_* - \mathbb{E}[n_*] \leq -\varepsilon\big) - \mathbb{P}\big(|n - \mathbb{E}[n]| \geq \varepsilon\big) - \mathbb{P}\big(|n' - \mathbb{E}[n']| \geq \varepsilon\big) \\ &\geq 1 - e^{-(N^{1+2\gamma})/(2Nqq')} \\ &\qquad - \big(e^{-(N^{1+2\gamma})/(2Nq)} + e^{-(N^{1+2\gamma})/(2Nq+(2/3)N^{1/2+\gamma})}\big) \\ &\qquad - \big(e^{-(N^{1+2\gamma})/(2Nq')} + e^{-(N^{1+2\gamma})/(2Nq'+(2/3)N^{1/2+\gamma})}\big) \\ &= 1 - 3e^{-\Theta(N^{2\gamma})} \xrightarrow[N\to\infty]{} 1. \end{aligned} \tag{3.5}$$

Conditioning on $\mathcal{B}$ will be instrumental to conclude the proofs of both (3.3) and (3.4).

Next, recall that G and G' are randomly labeled augmentations of an intersection graph $G_* \sim \mathrm{ER}(n_*, rr'\lambda/N)$. The nodes of G_* are embedded in G, G', and we identify these embeddings with the node subsets $V_* \subset V, V'_* \subset V'$, respectively. These subsets are related via the true graph alignment $\sigma_*\colon V_* \to V'_*$.

Let $\widehat{\sigma}\colon \widehat{V} \to \widehat{V}'$ be the alignment estimator from Algorithm 1 whose randomness stems from the randomness of (G, G'). For all pairs $i \in V$, $j \in V'$, the *if*-clauses of Algorithm 1 translate to the event

$$\begin{aligned}\mathcal{M}(i,j) &:= \{\text{MPALIGN matches } i \text{ with } j\} \\ &= \big\{\mathcal{N}_{G,2d}(i) \text{ and } \mathcal{N}_{G',2d}(j) \text{ are cycle-free}\big\} \\ &\qquad \cap \big\{i \text{ and } j \text{ have at least three neighbors, } (i_k)_{k=1}^3 \text{ and } (j_k)_{k=1}^3\big\} \\ &\qquad \cap \big\{\mathcal{T}_{d-1}\big(\mathcal{N}_{G,d}(i \to i_k), \mathcal{N}_{G',d}(j \to j_k)\big) = 1 \text{ for } k \in [3]\big\}. \end{aligned} \tag{3.6}$$

For all nodes $i \in V$, define the event of a correct match as

$$\mathcal{M}_*(i) := \{i \in V_*, \widehat{\sigma}(i) = \sigma_*(i)\}.$$

Recalling that $\square^c$ denotes the complement of a set $\square$, this lets us rewrite

$$\mathrm{ov}(\sigma_*, \widehat{\sigma}) = \frac{1}{n_*}\sum_{i\in\widehat{V}} \mathbb{1}_{\mathcal{M}_*(i)} \quad \text{and} \quad \mathrm{err}(\sigma_*, \widehat{\sigma}) = \frac{1}{n}\sum_{i\in\widehat{V}} \mathbb{1}_{\mathcal{M}_*(i)^c}.$$

Proof of (3.3). We start by showing significant overlap. For this, set

$$\widetilde{\beta} := \beta\, \underbrace{\mathbb{P}_d^{\mathrm{corr}}(\text{the intersection tree } t_* \text{ from Definition 3.1 has root degree} \geq 3)}_{=:\alpha},$$

where $\beta > 0$ is the value from (3.2) and α is a strictly positive probability only depending on the parameters λ, s and s'. Hence, $\widetilde{\beta} > 0$. The reason to introduce $\widetilde{\beta}$ is its relationship with $\mathbb{P}(\mathcal{M}_*(i))$: For every $i \in \widehat{V}$, denote by $\mathcal{C}_i$ the event of optimal coupling between the neighborhoods $(\mathcal{N}_{G,2d}(i), \mathcal{N}_{G',2d}(\sigma_*(i)))$ and a tree pair $(t, t') \sim \mathbb{P}_{2d}^{\mathrm{corr}}$.

Applying Lemma 3.5 with parameter $2d$ instead of d and exploiting the hypothesis $4c\log(\lambda\max\{s,s'\})<1$, we get that C_i has probability

$$\mathbb{P}(\mathcal{C}_i)=\mathbb{P}\big((\mathcal{N}_{G,2d}(i),\mathcal{N}_{G',2d}(\sigma_*(i)))=(t,t')\big)=1-o(1).$$

Since $\mathcal{M}_*(i)=\{i\in V_*\}\cap\mathcal{M}(i,\sigma_*(i))$, conditioning on $\mathcal{C}_i$ yields

$$\begin{aligned}\mathbb{P}(\mathcal{M}_*(i))&=\mathbb{P}(\mathcal{C}_i^c,\mathcal{M}_*(i))+\mathbb{P}\big(\mathcal{C}_i,i\in V_*,\mathcal{M}(i,\sigma_*(i))\big)\\&\geq o(1)+\alpha\mathbb{P}^{\mathrm{corr}}_{d-1}(\mathcal{T}_{d-1}=1).\end{aligned}\tag{3.7}$$

The second inequality translates the decomposition of $\mathcal{M}(i,j)$ from (3.6) conditioned on $\mathcal{C}_i$: Under $\mathcal{C}_i$, one has that $\mathcal{N}_{G,d}(i)$ and $\mathcal{N}_{G',d}(\sigma_*(i))$ are cycle-free, the event

$$\{i \text{ and } \sigma_*(i) \text{ have at least three neighbors}\}$$

has probability α and is independent of what happens below the roots, in particular of the event

$$\big\{\mathcal{T}_{d-1}\big(\mathcal{N}_{G,d}(i\to i_k),\mathcal{N}_{G',d}(\sigma_*(i)\to\sigma_*(i)_k)\big)=1 \text{ for } k\in[3]\big\}.$$

The probability of this last event is bounded below by $\mathbb{P}^{\mathrm{corr}}_{d-1}(\mathcal{T}_{d-1}=1)$, the probability that one of these correlated neighborhoods yields a positive test. Independence and this lower bound leads to (3.7).

Next, since $\liminf_{d\to\infty}\mathbb{P}^{\mathrm{corr}}_{d-1}(\mathcal{T}_{d-1}=1)=\beta>0$, one has

$$\mathbb{P}^{\mathrm{corr}}_{d-1}(\mathcal{T}_{d-1}=1)=(1-o(1))\beta,$$

implying

$$\begin{aligned}&\mathbb{P}(\mathcal{M}_*(i))\geq o(1)+\alpha(1-o(1))\beta\\&\implies\tilde{\beta}=\alpha\beta\leq\frac{\mathbb{P}(\mathcal{M}_*(i))-o(1)}{(1-o(1))}\leq(1+o(1))\mathbb{P}(\mathcal{M}_*(i)).\end{aligned}\tag{3.8}$$

Finally, in order to show (3.3), we will prove

$$\mathbb{P}\big(\mathrm{ov}(\sigma_*,\hat{\sigma})<\tilde{\beta}-\delta\big)\xrightarrow[N\to\infty]{}0\quad\text{for all }\delta>0.$$

To do so, we use the conditioning on $\mathcal{B}$ from (3.5), followed by the tower property of conditional expectation, the result from (3.8), and Markov's inequality to obtain the following sequence of upper bounds:

$$\begin{aligned}\mathbb{P}\big(\mathrm{ov}(\sigma_*,\hat{\sigma})<\tilde{\beta}-\delta\big)&\leq\mathbb{P}(\mathcal{B}^c)+\mathbb{P}\big(\mathcal{B},\mathrm{ov}(\sigma_*,\hat{\sigma})<\tilde{\beta}-\delta\big)\\&=o(1)+\mathbb{E}\big[\mathbb{1}_{\mathcal{B}}\mathbb{P}\big(\mathrm{ov}(\sigma_*,\hat{\sigma})<\tilde{\beta}-\delta\mid n_*,n,n'\big)\big]\end{aligned}$$

$$
\begin{aligned}
&= o(1) + \mathbb{E}\Big[\mathbb{1}_{\mathcal{B}}\mathbb{P}\Big(\widetilde{\beta} n_* - \sum_{i\in\widehat{V}} \mathbb{1}_{\mathcal{M}_*(i)} > \delta n_*\Big)\Big] \\
&\le o(1) + \mathbb{E}\Big[\mathbb{1}_{\mathcal{B}}\mathbb{P}\Big((1+o(1))\sum_{i\in\widehat{V}}\big(\mathbb{P}(\mathcal{M}_*(i)) - \mathbb{1}_{\mathcal{M}_*(i)}\big) > \delta n_*\Big)\Big] \\
&\le o(1) + \mathbb{E}\Big[\mathbb{1}_{\mathcal{B}}\frac{1+o(1)}{\delta^2 n_*^2}\big(n \operatorname{Var}\big(\mathbb{1}_{\mathcal{M}_*(1)}\big) + n(n-1)\operatorname{Cov}\big(\mathbb{1}_{\mathcal{M}_*(1)}, \mathbb{1}_{\mathcal{M}_*(2)}\big)\big)\Big].
\end{aligned}
$$

Our goal is now to find upper bounds for the last line: The variance of $\mathbb{1}_{\mathcal{M}_*(1)} \in [0,1]$ is bounded by 1 and the first summand in the expectation vanishes because

$$
\mathbb{E}\Big[\mathbb{1}_{\mathcal{B}}\frac{1+o(1)}{\delta^2 n_*^2} n \operatorname{Var}\big(\mathbb{1}_{\mathcal{M}_*(1)}\big)\Big] \le \frac{1+o(1)}{\delta^2 (Nqq - N^{1/2+\gamma})^2}\big(Nq + N^{1/2+\gamma}\big) \xrightarrow[N\to\infty]{} 0. \tag{3.9}
$$

For two distinct nodes, for instance $i = 1$ and $j = 2$, the term $\operatorname{Cov}(\mathbb{1}_{\mathcal{M}_*(1)}, \mathbb{1}_{\mathcal{M}_*(2)}) \in [-1,1]$ equals 0 if $(\mathcal{N}_{G,d}(i), \mathcal{N}_{G',d}(\sigma_*(i)))$ and $(\mathcal{N}_{G,d}(j), \mathcal{N}_{G',d}(\sigma_*(j)))$ are independent. This independence can be obtained asymptotically by using Lemma 3.2 which leads to $\operatorname{Cov}(\mathbb{1}_{\mathcal{M}_*(1)}, \mathbb{1}_{\mathcal{M}_*(2)}) \in o(1)$. Combining this fact with a similar bound as in (3.9) lets us conclude the proof of (3.3) since

$$
\mathbb{P}\big(\operatorname{ov}(\sigma_*, \widehat{\sigma}) < \widetilde{\beta} - \delta\big) \le o(1) + \mathbb{E}\Big[\mathbb{1}_{\mathcal{B}}\frac{1+o(1)}{\delta^2 n_*^2} n^2 \operatorname{Cov}\big(\mathbb{1}_{\mathcal{M}_*(1)}, \mathbb{1}_{\mathcal{M}_*(2)}\big)\Big] \xrightarrow[N\to\infty]{} 0.
$$

Proof of (3.4). By Remark 2.5, we need to show $\operatorname{err}(\sigma_*, \widehat{\sigma}) \xrightarrow[N\to\infty]{} 0$ in probability. For this, let $\delta > 0$ and use (3.5) to obtain

$$
\mathbb{P}\big(\operatorname{err}(\sigma_*, \widehat{\sigma}) > \delta\big) \le \mathbb{P}\big(\operatorname{err}(\sigma_*, \widehat{\sigma}) > \delta, \mathcal{B}\big) + \mathbb{P}(\mathcal{B}^c) = \mathbb{P}\big(\operatorname{err}(\sigma_*, \widehat{\sigma}) > \delta, \mathcal{B}\big) + o(1).
$$

Furthermore, Markov's inequality yields

$$
\begin{aligned}
\mathbb{P}\Big(\underbrace{\sum_{i\in\widehat{V}} \mathbb{1}_{\mathcal{M}_*(i)^c}}_{=n\times\text{err}} > n\delta, \mathcal{B}\Big) &\le \frac{1}{\delta n}\sum_{i\in V}\mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{B}\big) \\
&\le \frac{1}{\delta}\max_{i\in V}\mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{B}\big).
\end{aligned}
$$

Hence, in order to show $\operatorname{err}(\sigma_*, \widehat{\sigma}) \xrightarrow{\mathbb{P}} 0$ it suffices to check

$$
\mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{B}\big) \xrightarrow{N\to\infty} 0 \quad \text{for all } i \in V.
$$

Fix any $i \in V$ and decompose the event $\mathcal{M}_*(i)^c \cap \{i \in \widehat{V}\}$ of a wrong match by MPALIGN as follows:

$$
\mathcal{M}_*(i)^c \cap \{i \in \widehat{V}\} = \bigcup_{j\in G'\setminus\square} \mathcal{M}(i,j),
$$

where

$$\square = \begin{cases} \emptyset & \text{if } i \in \widehat{V} \setminus V_*, \\ \{\sigma_*(i)\} & \text{if } i \in \widehat{V}, \widehat{\sigma}(i) \neq \sigma_*(i). \end{cases}$$

Next, define the event of exploding neighborhood sizes across G and G' as follows:

$$\begin{aligned} \mathcal{A} := & \left\{\exists i \in V,\, t \in [d] : |\mathcal{S}_{G,t}(i)| > C \log(n)(\lambda s)^t\right\} \\ & \cup \left\{\exists j \in V',\, t' \in [d] : |\mathcal{S}_{G',t'}(j)| > C \log(n')(\lambda s')^t\right\}. \end{aligned}$$

The probability of $\mathcal{A}$ is vanishing according to Lemma 3.1. Consequently,

$$\begin{aligned} \mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{B}\big) &\leq \mathbb{P}(\mathcal{A}) + \mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{A}^c, \mathcal{B}\big) \\ &= o(1) + \mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{A}^c, \mathcal{B}\big), \end{aligned}$$

which leads us to work on the event $\mathcal{M}_*(i)^c \cap \{i \in \widehat{V}\} \cap \mathcal{A}^c \cap \mathcal{B}$ in the sequel. One can decompose this event using the union bound

$$\begin{aligned} & \mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{A}^c, \mathcal{B}\big) \\ & \quad \leq \sum_{j \in G' \setminus \square} \mathbb{P}\big(\mathcal{M}(i,j), \mathcal{A}^c, \mathcal{B}\big) = \sum_{j \in G' \setminus \square} \mathbb{E}\big[\mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{M}(i,j)}\big] \\ & \quad \leq \sum_{j \in G' \setminus \square} \mathbb{E}\big[\mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{T}_{d-1}(\mathcal{N}_{G,d}(i \to i_k), \mathcal{N}_{G',d}(j \to j_k)) = 1 \text{ for some special } k}\big]. \end{aligned}$$

The last inequality uses the dangling tree trick: By "*some special* k", we mean the index k from Lemma 3.7, which lets $\mathcal{N}_{G,d}(i \to i_k) \cap V_*$ and $\mathcal{N}_{G' d}(j \to j_k) \cap V'_*$ be disjoint. From this same lemma, we also know that

$$\big(\mathcal{N}_{G,d}(i \to i_k), \mathcal{N}_{G',d}(j \to j_k)\big) \sim \mathbb{P}^{\text{dis}}_{d-1},$$

which leads to

$$\begin{aligned} \mathbb{P}\big(\mathcal{M}_*(i)^c, i \in \widehat{V}, \mathcal{A}^c, \mathcal{B}\big) &\leq \sum_{j \in G' \setminus \square} \mathbb{E}^{\text{dis}}_{d-1}\big[\mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{T}_{d-1}=1}\big] \\ &\leq n' \mathbb{E}^{\text{dis}}_{d-1}\big[\mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{T}_{d-1}=1}\big]. \end{aligned} \tag{3.10}$$

For two trees t and t', define the likelihood ratio of $\mathbb{P}^{\text{dis}}_{d-1}$ and $\mathbb{P}^{\text{ind}}_{d-1}$ as

$$R_{d-1}(t,t') := \frac{\mathbb{P}^{\text{dis}}_{d-1}(t,t')}{\mathbb{P}^{\text{ind}}_{d-1}(t,t')},$$

and write R_{d-1} without arguments for the random variable if the trees follow some indicated law. This lets us write

$$\begin{aligned} n' \mathbb{E}^{\text{dis}}_{d-1}\big[\mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{T}_{d-1}=1}\big] &= n' \mathbb{E}^{\text{ind}}_{d-1}\big[R_{d-1} \mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}} \mathbb{1}_{\mathcal{T}_{d-1}=1}\big] \\ &\leq n' \sqrt{\mathbb{E}^{\text{ind}}_{d-1}\big[R^2_{d-1} \mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}}\big]} \sqrt{\mathbb{P}^{\text{ind}}_{d-1}(\mathcal{T}_{d-1} = 1)}, \end{aligned} \tag{3.11}$$

where we have used the Cauchy–Schwarz inequality. At this point, we need to *quantify the similarity between* $\mathbb{P}_{d-1}^{\mathrm{dis}}$ *and* $\mathbb{P}_{d-1}^{\mathrm{ind}}$ as announced in Remark 3.8. The following lemma bounds the second moment of their likelihood ratio.

Lemma 3.11. *Under the parameters of Theorem 3.10 and using the above notations, one has*

$$\mathbb{E}_{d-1}^{\mathrm{ind}}\left[R_{d-1}^2 \mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}}\right] \in \mathcal{O}(1).$$

The proof of this lemma, which we defer to Appendix A.1, uses well-established techniques to describe the relationship between graph neighborhoods and branching processes. We refer to [3, p. 161] for a concrete example and mention [10] for an extension to inhomogeneous graphs. The proof presented in Appendix A.1 considerably extends and clarifies its analog in the symmetric case, the proof of [27, Lemma 6.1].

Finally, combining the hypothesis

$$\mathbb{P}_{d-1}^{\mathrm{ind}}(\mathcal{T}_{d-1} = 1) \leq \exp\left(-C(\lambda s s')^{d-1}\right)$$

from (3.2) with Lemma 3.11 and the equations (3.10), (3.11), one obtains

$$\mathbb{P}\left(\mathcal{M}_*(i)^c, i \in \hat{V}, \mathcal{A}^c, \mathcal{B}\right) \leq n' \sqrt{\mathcal{O}(1)} \sqrt{\exp\left(-C(\lambda s s')^{d-1}\right)}.$$

Since $d = \lfloor c \log(N) \rfloor$, one has $n' \leq N < e^{(d+1)/c}$, and hence the right-hand side tends to 0 as we let $d \to \infty$ (and a fortiori for $N \to \infty$). Together with our initial Markov bound on $\mathbb{P}(\mathrm{err} > \delta)$, this lets us conclude that $\mathrm{err}(\sigma_*, \hat{\sigma}) \xrightarrow{\mathbb{P}} 0$, which completes the proof. ∎

4. Tree correlation testing

As the previous section shows, distinguishing between independent and correlated trees is the key building block for sparse graph alignment. Therefore, the statistical feasibility of tree correlation testing in the asymmetric setting is the central question remaining to be answered. This section presents a one-sided test for the following problem which was introduced in (3.1):

$$H_0^{(d)}\colon (t_d, t_d') \sim \mathbb{P}_d^{\mathrm{ind}} \quad \text{vs.} \quad H_1^{(d)}\colon (t_d, t_d') \sim \mathbb{P}_d^{\mathrm{corr}}. \tag{4.1}$$

We show that the existence of such one-sided tests achieves a sharp phase transition at $ss' = \alpha$, where $\alpha \approx 0.3383$. To do so, we present an analysis relying on advantageous properties of the likelihood ratio in the tree correlation testing problem, primarily based on [27, 28]. This section is the theoretical backbone of this paper and presents our novel contribution to tree correlation testing.

We start by proposing a formalism for unlabeled trees before recalling an essential result on the number of such trees. Next, we introduce the likelihood ratio for (4.1) and present two of its key properties. These will then serve as proof engines for the two main results of this section, Theorem 4.9 and Theorem 4.10.

4.1. Unlabeled trees

Since node labels are unknown in the sparse graph alignment problem, local neighborhoods are rooted unlabeled trees. Our goal is to describe the laws of these trees, which requires a rigorous definition of what we mean by the *set of rooted unlabeled trees of depth at most* d, which we denote as $\mathcal{X}_d$.

A first idea is to define $\mathcal{X}_d$ as the set of cycle-free, non-empty graphs t with a distinct root node ρ such that $\operatorname{dist}(\rho, i) \leq d$ for all nodes $i \in t$. For two rooted trees (t, ρ) and (t', ρ'), we consider them equal if there is a graph isomorphism mapping t to t' and ρ to ρ'. The notion of *equality by isomorphism* translates our intuition of the word *unlabeled* since all relabelings of structurally identical rooted trees are treated as a single element of $\mathcal{X}_d$. However, there is an equivalent definition of $\mathcal{X}_d$ which defines the same object but is better suited for some of our analysis. It defines unlabeled trees recursively as follows.

Definition 4.1 (Unlabeled rooted trees). Set $\mathcal{X}_0 := \{\bullet\}$, where $\bullet$ denotes the graph consisting of a single node. For all $d \in \mathbb{N}$, recursively define

$$\mathcal{X}_{d+1} = \Big\{(N_\tau)_{\tau \in \mathcal{X}_d} \in \mathbb{N}^{\mathcal{X}_d} : \sum_{\tau \in \mathcal{X}_d} N_\tau < \infty\Big\}.$$

By this definition, we understand an element $t \in \mathcal{X}_{d+1}$ as a sequence of natural numbers indexed by unlabeled trees of depth at most d, i.e., $t = (N_\tau)_{\tau \in \mathcal{X}_d}$, where only finitely many N_τ are non-zero. For every $\tau \in \mathcal{X}_d$, the number N_τ denotes how many copies of τ are attached to the root of t. This is best understood by looking at Figure 3.

We furthermore set

$$\mathcal{X}_{d+1}^{(n)} := \big\{t = (N_\tau)_{\tau \in \mathcal{X}_d} \in \mathcal{X}_{d+1} : |t| = n\big\},$$

where $|t| = 1 + \sum_{\tau \in \mathcal{X}_d} N_\tau |\tau|$ denotes a tree's node number. In other words, $\mathcal{X}_{d+1}^{(n)}$ contains all unlabeled trees with n nodes and depth bounded by $d + 1$.

It is immediate by induction that $\mathcal{X}_d$ is countable and $\mathcal{X}_d^{(n)}$ is finite for all choices of $d, n \in \mathbb{N}$. Note that a tree of size n can have a maximum depth of $n - 1$, and hence $\mathcal{X}_{n-1}^{(n)}$ describes all unlabeled trees with n nodes. The asymptotic cardinality of $\mathcal{X}_{n-1}^{(n)}$ is explicitly known thanks to the following result due to Otter [41].

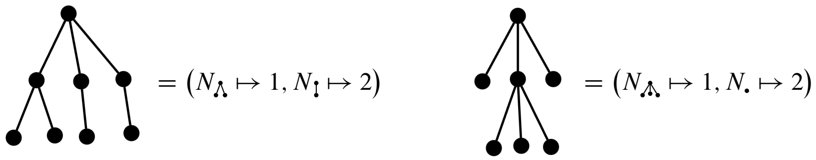

Figure 3. Two examples of unlabeled trees and their representation as sequences of natural numbers indexed by trees.

Proposition 4.1. *For $\alpha := 0.3383219\ldots$, which is known as Otter's constant, there exists $D > 0$ such that*

$$|\mathcal{X}_{n-1}^{(n)}| \underset{n\to\infty}{\sim} \frac{D}{n^{3/2}} \frac{1}{\alpha^n}.$$

This implies that the power series $\Phi(x) = \sum_{n=1}^{\infty} |\mathcal{X}_{n-1}^{(n)}| x^n$ has radius of convergence $\alpha > 0$ with $\Phi(\alpha) < \infty$.

This proposition plays a key role when examining the likelihood ratio, as we explain in Section 4.3. Before delving into that, we first need to translate the laws of correlated Galton–Watson trees as introduced in Section 3.2 to the new formalism of $\mathcal{X}_{d+1}$.

4.2. Correlated Galton–Watson trees, revisited

Given a pair of trees $(t, t') \sim \mathbb{P}_{d+1}^{\mathrm{corr}}$ represented by their subtree tuples $t = (N_\tau)_{\tau \in \mathcal{X}_d}$ and $t' = (N'_{\tau'})_{\tau' \in \mathcal{X}_d}$ as in Definition 4.1, how can we describe their joint law using only N_τ and $N'_{\tau'}$? The following definition and lemma answer this question.

Definition 4.2. A pair of trees $t = (N_\tau)_\tau$, $t' = (N'_{\tau'})_{\tau'} \in \mathcal{X}_{d+1}$ has joint law $\widetilde{\mathbb{P}_{d+1}^{\mathrm{corr}}}$ if for every $\tau, \tau' \in \mathcal{X}_d$, one has

$$N_\tau = \Delta_\tau + \sum_{\tau' \in \mathcal{X}_d} \Gamma_{\tau,\tau'} \quad \text{and} \quad N'_{\tau'} = \Delta'_{\tau'} + \sum_{\tau \in \mathcal{X}_d} \Gamma_{\tau,\tau'},$$

where the independent random variables Δ, Δ', and Γ follow the laws

$$\begin{aligned} \Delta_\tau &\sim \mathrm{Poi}\big(\lambda s(1-s')\, \mathrm{GW}_d^{(\lambda s)}(\tau)\big), \\ \Delta'_{\tau'} &\sim \mathrm{Poi}\big(\lambda s'(1-s)\, \mathrm{GW}_d^{(\lambda s')}(\tau')\big), \\ \Gamma_{\tau,\tau'} &\sim \mathrm{Poi}\big(\lambda s s' \mathbb{P}_d^{\mathrm{corr}}(\tau, \tau')\big). \end{aligned}$$

Lemma 4.2. *The laws $\mathbb{P}_{d+1}^{\mathrm{corr}}$ from Definition 3.1 and $\widetilde{\mathbb{P}_{d+1}^{\mathrm{corr}}}$ from Definition 4.2 are equal. Hence, we always write $\mathbb{P}_{d+1}^{\mathrm{corr}}$, also when working with the subtree tuple representation.*

Proof. Let $(t, t') \sim \mathbb{P}^{\mathrm{corr}}_{d+1}$ and recall that t, t' are independent augmentations of a common intersection tree $t_* \sim \mathrm{GW}^{(\lambda s s')}$. Every node in t adjacent to the root falls into one of two categories:

(a) the node was present in t_* before the augmentation, in which case the subtree attached to that node is shared between t and t';

(b) the node was not in t_*, in which case the subtree attached to that node is of law $\mathrm{GW}_d^{(\lambda s)}$ and independent from t'.

For category (b), let $c_+ \sim \mathrm{Poi}(\lambda s(1-s'))$ be the number of non-shared children of the root in t. Conditionally on c_+, the attached subtrees are i.i.d. with law $\mathrm{GW}_d^{(\lambda s)}$, meaning that each child carries a mark $\tau \in \mathcal{X}_d$ with probability $\mathrm{GW}_d^{(\lambda s)}(\tau)$. By Poisson thinning for marked Poisson variables, the counts

$$\Delta_\tau := \#\{\text{category (b) root-children in } t \text{ whose subtree equals } \tau\}$$

are independent and satisfy $\Delta_\tau \sim \mathrm{Poi}(\lambda s(1-s')\ \mathrm{GW}_d^{(\lambda s)}(\tau))$ for all $\tau \in \mathcal{X}_d$. The same argument applies to t' yielding $\Delta'_{\tau'}$.

For category (a), let $k \sim \mathrm{Poi}(\lambda s s')$ be the number of shared root-children coming from the intersection tree t_*. Conditionally on k, the pairs of attached subtrees are i.i.d. with joint law $(\tau, \tau') \sim \mathbb{P}^{\mathrm{corr}}_d$. By Poisson thinning, the pair-counts

$$\Gamma_{\tau,\tau'} := \#\{\text{category (a) shared root-children with pair } (\tau, \tau')\}$$

are independent and satisfy $\Gamma_{\tau,\tau'} \sim \mathrm{Poi}(\lambda s s' \mathbb{P}^{\mathrm{corr}}_d(\tau, \tau'))$ for all pairs $(\tau, \tau') \in \mathcal{X}_d^2$. Hence, the total number of category (a) root subtrees equal to τ in t is $\sum_{\tau'} \Gamma_{\tau,\tau'}$. Analogously, $\sum_\tau \Gamma_{\tau,\tau'}$ counts the τ'-subtrees in t'.

Since N_τ and $N'_{\tau'}$ count the number of occurrences of τ and τ' as root subtrees in t and t', respectively, and because every root subtree is either in category (a) or (b), we obtain the expressions for N and N' in Definition 4.2. ∎

The description of $\mathbb{P}^{\mathrm{corr}}_{d+1}$ using subtree tuples was the last necessary ingredient before examining the likelihood ratio of the testing problem (4.1).

4.3. The likelihood ratio and its properties

The Neyman–Pearson lemma strongly suggests using the likelihood ratio when it comes to hypothesis testing. Therefore, we introduce this object in the context of unlabeled trees as a first step.

For two unlabeled trees $\tau, \tau' \in \mathcal{X}_d$ recall that $\tau = \tau'$ means that they are equal up to relabeling. With this in mind, we denote the likelihood of two given trees t, t' being correlated up to depth $d \in \mathbb{N}$ as

$$\mathbb{P}^{\mathrm{corr}}_d(t, t') := \mathbb{P}_{(\tau,\tau') \sim \mathbb{P}^{\mathrm{corr}}}\big((\tau_d, \tau'_d) = (t_d, t'_d)\big) = \mathbb{P}^{\mathrm{corr}}_d\big((\tau, \tau') = (t, t')\big).$$

In the last expression and for the rest of this section, (τ, τ') implicitly denotes the random variable issued from the indicated law ($\mathbb{P}_d^{\mathrm{corr}}$ in this case). Since $\mathbb{P}_d^{\mathrm{corr}}$ only considers trees up to depth d, our short notation $\mathbb{P}_d^{\mathrm{corr}}(t, t')$ more accurately refers to $\mathbb{P}_d^{\mathrm{corr}}(t_d, t'_d)$. We define $\mathbb{P}_d^{\mathrm{ind}}(t, t')$ analogously as the likelihood of (t_d, t'_d) following the distribution $\mathbb{P}_d^{\mathrm{ind}}$.

With these notations, the likelihood ratio for the testing problem (4.1) is

$$L_d(t, t') := \frac{\mathbb{P}_d^{\mathrm{corr}}(t, t')}{\mathbb{P}_d^{\mathrm{ind}}(t, t')}.$$

If (t, t') is a pair of random trees, $L_d(t, t')$ becomes a random variable which we will often denote as L_d when the law of (t, t') is evident from the context.

Recursive property. The reason for studying $L_d(t, t')$ is that it fully determines one-sided testability, as we show later, in "(a) $\Leftrightarrow$ (b)" of Theorem 4.9. Consequently, we will test for tree correlation using the likelihood ratio, making it crucial to find efficient ways of computing it. The following proposition solves this problem by providing a recursive formula for L_d.

Proposition 4.3 (Recursive likelihood ratio formula). *For $t, t' \in \mathcal{X}_d$, let c, c' be the degrees of their respective root nodes. Denote by $t_{[1]}, \ldots, t_{[c]}$ respectively $t'_{[1]}, \ldots, t'_{[c']}$ the subtrees attached to these roots, labeled arbitrarily. Recall that $\sigma\colon [k] \hookrightarrow [c]$ denotes an injective map from $[k]$ to $[c]$ and let $\sum_{\sigma:[k]\hookrightarrow[c]}$ be the sum over all such mappings. Then, for all $d \in \mathbb{N}_{>0}$, the likelihood ratio at depth d can be expressed recursively as*

$$L_d(t, t') = \sum_{k=0}^{c \wedge c'} \psi(k, c, c') \sum_{\substack{\sigma\colon [k]\hookrightarrow[c],\\ \sigma'\colon [k]\hookrightarrow[c']}} \prod_{i=1}^{k} L_{d-1}\big(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big),$$

where

$$\psi(k, c, c') = \exp(\lambda s s') \frac{1}{\lambda^k k!} (1 - s')^{c-k} (1 - s)^{c'-k}.$$

In Appendix A.2, the reader can retrieve the proof of this result.

Remark 4.4. Recall from Remark 3.9 that the graph alignment algorithm MPALIGN is polynomial if tree correlation tests can be computed in polynomial time. Since both c and c' are of order $\mathcal{O}(1)$ in the sparse regime, the recursive formula from Proposition 4.3 enables the desired efficient computation of L_d. With $d \in \mathcal{O}(N)$, this computation is even in $\mathcal{O}(\mathrm{polylog}(N))$.

One can repeatedly apply the recursive likelihood ratio formula to obtain an explicit lower bound for L_{d+k}, which will serve later in the proof of Theorem 4.9.

Proposition 4.5. *Let (t, t') be a pair of rooted trees sharing a common rooted subtree t_*. Denote by $\sigma_*\colon t_* \hookrightarrow t$ and $\sigma'_*\colon t_* \hookrightarrow t'$ the embeddings of t_* into t and t', respectively. For each $i \in t$, let $c_t(i)$ denote its number of child nodes in t, and define $t_{[i]}$ as the subtree rooted at i, containing all its descendants. This notation extends naturally to t' and t_*, and we recall that $(t_*)_d$ denotes the depth-d truncation of t_*. Then, for all $d, k \in \mathbb{N}_{>0}$, the following inequality holds:*

$$\begin{aligned} L_{d+k}(t, t') \geq &\prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big) \\ &\times \prod_{j \in (t_*)_d \setminus (t_*)_{d-1}} L_k\big(t_{[\sigma_*(j)]}, t'_{[\sigma'_*(j)]}\big). \end{aligned}$$

The proof of this proposition is deferred to Appendix A.3.

The representations of L_d from these propositions have their worth outside of computational inquiries, for instance, when proving the next property of L_d.

Martingale property. If we consider the subsequent likelihood ratios $(L_d(t, t'))_d$ as a random process, it is natural to suspect it to be a martingale with respect to $\mathbb{P}^{\text{ind}}$ and the filtration $\mathcal{F}_d = \sigma((t_d, t'_d))$. The martingale properties are automatically true if the random variables in the likelihood ratio have Lebesgue densities [32]. In our case, it can easily be checked by computation: Following the reasoning in [27, Section 2.3], it suffices to verify that the following expression equals 1:

$$\begin{aligned} &\mathbb{E}^{\text{ind}}\left[\sum_{k=0}^{c \wedge c'} \frac{\pi_{\lambda s s'}(k)\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)}{\pi_{\lambda s}(c)\pi_{\lambda s'}(c')}\right] \\ &= \sum_{k=0}^{\infty} \mathbb{E}^{\text{ind}}\left[e^{\lambda s s'} \frac{1}{\lambda^k k!}(1-s')^{c-k}(1-s)^{c'-k} \frac{c!\,c'!}{(c-k)!(c'-k)!}\mathbb{1}_{k \leq c \wedge c'}\right] \\ &= e^{\lambda s s'} \sum_{k=0}^{\infty} \frac{1}{\lambda^k k!} \mathbb{E}^{\text{ind}}\left[(1-s')^{c-k} \frac{c!}{(c-k)!}\mathbb{1}_{k \leq c}\right] \mathbb{E}^{\text{ind}}\left[(1-s)^{c'-k} \frac{c'!}{(c'-k)!}\mathbb{1}_{k \leq c'}\right] \\ &= 1. \end{aligned}$$

The last two equalities follow from the independence of c and c' and their distributions

$$c \sim \text{Poi}(\lambda s) \quad \text{and} \quad c' \sim \text{Poi}(\lambda s').$$

Due to $\mathbb{E}^{\text{ind}}_d[L_d] = 1$ and $(L_d)_d$ is trivially adapted to the filtration, it follows that the likelihood ratio forms a martingale.

Since $L_d \geq 0$, the Martingale Convergence theorem yields an almost sure limit $L_\infty \in L^1$. As a consequence, we obtain a fixed point formula for $\mathbb{E}^{\text{ind}}[L_\infty]$ from the recursive representation of Proposition 4.3: Using that c, c' are independent from $t_{[\sigma(i)]}, t'_{[\sigma'(i)]}$ in Galton–Watson trees, letting $d \to \infty$ and then taking $\mathbb{E}^{\text{ind}}$, we obtain

$$\begin{aligned}\mathbb{E}^{\text{ind}}[L_\infty] &= \mathbb{E}^{\text{ind}}\Bigg[\sum_{k=0}^{c\wedge c'} \psi(k,c,c') \sum_{\substack{\sigma\colon [k]\hookrightarrow[c],\\ \sigma'\colon[k]\hookrightarrow[c']}} \prod_{i=1}^{k} \mathbb{E}^{\text{ind}}[L_\infty]\Bigg] \\ &= \mathbb{E}^{\text{ind}}\Bigg[\sum_{k=0}^{c\wedge c'} \frac{\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)}{\pi_{\lambda s}(c)\,\pi_{\lambda s'}(c')}\;\pi_{\lambda s s'}(k)\mathbb{E}^{\text{ind}}[L_\infty]^k\Bigg] \\ &= \sum_{k=0}^{\infty} \pi_{\lambda s s'}(k)\mathbb{E}^{\text{ind}}[L_\infty]^k,\end{aligned}$$

where the last step follows from a similar computation as above.

Consequently, $\mathbb{E}^{\text{ind}}[L_\infty]$ is a fixed point of the probability generating function of a $\text{Poi}(\lambda s s')$ random variable. This lets us directly link $\mathbb{E}^{\text{ind}}[L_\infty]$ to the theory of Galton–Watson processes with average offspring $\lambda s s'$: Recall that in the supercritical regime of such processes, their extinction probability is equal to the smallest solution to the same fixed point equation [1]. In particular, supposing $\lambda s s' > 1$ and $\mathbb{E}^{\text{ind}}[L_\infty] < 1$ implies that $\mathbb{E}^{\text{ind}}[L_\infty]$ is equal to the extinction probability of a Galton–Watson tree with $\text{Poi}(\lambda s s')$ offspring which we denote by $\mathbb{P}(\text{Ext}(\text{GW}^{(\lambda s s')}))$. This probability is also equal to the relative size of the giant component in a sparse Erdős–Rényi graph [9], illustrating another close link between these graphs and the trees we consider.

Aside from the connection to extinction probabilities of Galton–Watson trees, the condition $\mathbb{E}^{\text{ind}}[L_\infty] < 1$ is equivalent to an additional property of $(L_d)_d$.

Lemma 4.6. *One has $\mathbb{E}^{\text{ind}}[L_\infty] < 1$ if and only if $(L_d)_d$ is not uniformly integrable.*

Proof. Since martingales are uniformly integrable if and only if they converge in L^1, the equivalence to be shown can be reformulated as

$$(L_d)_d \text{ converges in } L^1 \quad\Longleftrightarrow\quad \mathbb{E}^{\text{ind}}[L_\infty] = 1.$$

Assuming L^1-convergence, one has

$$1 = \mathbb{E}^{\text{ind}}[L_d] \to \mathbb{E}^{\text{ind}}[L_\infty],$$

yielding $\mathbb{E}^{\text{ind}}[L_\infty] = 1$.

On the other hand, suppose that

$$\mathbb{E}^{\text{ind}}[L_\infty] = 1 = \lim_{d\to\infty} \mathbb{E}^{\text{ind}}[L_d].$$

Since $L_d \to L_\infty$ almost surely, Scheffé's lemma also implies convergence in L^1, which concludes the proof. ∎

Diagonalization property. With the recursion formula and the martingale property of L_d at hand, we lack a final tool for a more fine-grained analysis of information-theoretic thresholds. The following stunning result generalizes [28, Theorem 4] to the asymmetric case, affirming that L_d can be *diagonalized* over the space of unlabeled trees. We refer to this result as the likelihood ratio diagonalization formula since it resembles the Spectral theorem for matrices, including the orthonormal eigenbasis.

Theorem 4.7. *There exists a set of functions $f_{d,\beta}^{(\mu)}: \mathcal{X}_d \to \mathbb{R}$ with parameters $\mu > 0$, $d \in \mathbb{N}$ and indexed by trees $\beta \in \mathcal{X}_d$ such that for all choices of $s, s' \in [0,1]$, $\lambda > 0$, the following* likelihood ratio diagonalization *formula holds:*

$$\forall t, t' \in \mathcal{X}_d : \quad L_d(t,t') = \sum_{\beta \in \mathcal{X}_d} \sqrt{ss'}^{|\beta|-1} f_{d,\beta}^{(\lambda s)}(t) f_{d,\beta}^{(\lambda s')}(t'). \tag{4.2}$$

Furthermore, the $f_{d,\beta}^{(\mu)}$ have the following properties:

- *Constant value for $\beta = \bullet$, the trivial tree:*

$$\forall \mu > 0,\ \forall t \in \mathcal{X}_d : \quad f_{d,\bullet}^{(\mu)}(t) = 1.$$

- First orthogonality property *(with respect to the Galton–Watson-measure):*

$$\forall \mu > 0,\ \forall \beta, \beta' \in \mathcal{X}_d : \quad \sum_{t \in \mathcal{X}_d} \mathrm{GW}_d^{(\mu)}(t)\, f_{d,\beta}^{(\mu)}(t)\, f_{d,\beta'}^{(\mu)}(t) \;=\; \mathbb{1}_{\beta=\beta'}. \tag{4.3}$$

- Second orthogonality property *(summing over β):*

$$\forall \mu > 0,\ \forall t, t' \in \mathcal{X}_d : \quad \sum_{\beta \in \mathcal{X}_d} f_{d,\beta}^{(\mu)}(t)\, f_{d,\beta}^{(\mu)}(t') \;=\; \frac{\mathbb{1}_{t=t'}}{\mathrm{GW}_d^{(\mu)}(t)}. \tag{4.4}$$

This theorem, particularly the orthogonality properties, is central in our strategy for showing Theorem 4.10 later. Its proof is a computationally heavy induction, which we defer to Appendix A.4.

A first opportunity to see Theorem 4.7 in action is by relating the second moment of the likelihood ratio to the number of unlabeled trees.

Corollary 4.8. *The second moment of the likelihood ratio under the independence assumption can be written as*

$$\mathbb{E}_d^{\mathrm{ind}}[L_d^2] = \sum_{n=1}^{\infty} |\mathcal{X}_d^{(n)}|(ss')^{n-1}.$$

Proof. Using the likelihood ratio diagonalization formula (4.2), we compute

$$\mathbb{E}_d^{\mathrm{ind}}[L_d^2] = \mathbb{E}_d^{\mathrm{ind}}\Bigg[\sum_{\beta,\beta' \in \mathcal{X}_d} \sqrt{ss'}^{|\beta|+|\beta'|-2} f_{d,\beta}^{(\lambda s)}(t)\, f_{d,\beta'}^{(\lambda s)}(t)\, f_{d,\beta}^{(\lambda s')}(t')\, f_{d,\beta'}^{(\lambda s')}(t')\Bigg]$$

$$
\begin{aligned}
&\overset{\text{(a)}}{=} \sum_{\beta,\beta'\in\mathcal{X}_d} \mathbb{E}_d^{\mathrm{ind}}\Big[\sqrt{ss'}^{|\beta|+|\beta'|-2} f_{d,\beta}^{(\lambda s)}(t) f_{d,\beta'}^{(\lambda s)}(t) f_{d,\beta}^{(\lambda s')}(t') f_{d,\beta'}^{(\lambda s')}(t')\Big] \\
&\overset{\text{(b)}}{=} \sum_{\beta,\beta'\in\mathcal{X}_d} \sqrt{ss'}^{|\beta|+|\beta'|-2} \mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta}^{(\lambda s)}(t) f_{d,\beta'}^{(\lambda s)}(t)\big] \mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta}^{(\lambda s')}(t') f_{d,\beta'}^{(\lambda s')}(t')\big] \\
&\overset{\text{(c)}}{=} \sum_{\beta,\beta'\in\mathcal{X}_d} \sqrt{ss'}^{|\beta|+|\beta'|-2} \mathbb{1}_{\beta=\beta'}^2 \\
&= \sum_{\beta\in\mathcal{X}_d} (ss')^{|\beta|-1} \\
&= \sum_{n=1}^{\infty} |\mathcal{X}_d^{(n)}| (ss')^{n-1}.
\end{aligned}
$$

The equation (b) uses independence of t and t' under $\mathbb{P}_d^{\mathrm{ind}}$, while (c) exploits the first orthogonality property (4.3) in the form

$$
\sum_{t\in\mathcal{X}_d} \mathrm{GW}_d^{(\mu)}(t)\, f_{d,\beta}^{(\mu)}(t)\, f_{d,\beta'}^{(\mu)}(t) = \mathbb{E}_{t\sim \mathrm{GW}_d^{(\mu)}}\big[f_{d,\beta}^{(\mu)}(t)\, f_{d,\beta'}^{(\mu)}(t)\big] = \mathbb{1}_{\beta=\beta'}
$$

for parameter choices $\mu = \lambda s$ and $\mu = \lambda s'$. In (a), we use both of these facts to apply Fubini's theorem in order to exchange expectation and summation over the countable set $\mathcal{X}_d \times \mathcal{X}_d$. Using Fubini is possible because $ss' < 1$ and the following application of the Cauchy–Schwarz inequality:

$$
\begin{aligned}
&\mathbb{E}_d^{\mathrm{ind}}\big[|f_{d,\beta}^{(\lambda s)}(t) f_{d,\beta'}^{(\lambda s)}(t) f_{d,\beta}^{(\lambda s')}(t') f_{d,\beta'}^{(\lambda s')}(t')|\big] \\
&\quad = \mathbb{E}_d^{\mathrm{ind}}\big[|f_{d,\beta}^{(\lambda s)}(t) f_{d,\beta'}^{(\lambda s)}(t)|\big] \mathbb{E}_d^{\mathrm{ind}}\big[|f_{d,\beta}^{(\lambda s')}(t') f_{d,\beta'}^{(\lambda s')}(t')|\big] \\
&\quad \le \sqrt{\mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta}^{(\lambda s)}(t)^2\big] \mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta'}^{(\lambda s)}(t)^2\big] \mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta}^{(\lambda s')}(t')^2\big] \mathbb{E}_d^{\mathrm{ind}}\big[f_{d,\beta'}^{(\lambda s')}(t')^2\big]} = 1.
\end{aligned}
$$

This concludes the proof of the corollary. ■

With a full toolbox of results about L_d, we are now ready to present our main theorems about tree correlation testing.

4.4. One-sided testability equivalences

Motivated by the question whether there exist one-sided tests for the tree correlation testing problem (4.1), we start by reformulating it in two steps: First, make the question uniquely about likelihood ratio tests; second, relate the feasibility to more amenable information-theoretic quantities. It turns out that testability is closely linked to the Kullback–Leibler divergence between $\mathbb{P}_d^{\mathrm{corr}}$ and $\mathbb{P}_d^{\mathrm{ind}}$, which is defined as

$$
\mathrm{KL}_d := \mathrm{KL}\big(\mathbb{P}_d^{\mathrm{corr}} \| \mathbb{P}_d^{\mathrm{ind}}\big) = \mathbb{E}_d^{\mathrm{corr}}[\log L_d].
$$

This quantity is *non-decreasing* as a function of the depth d. To see this, let $\mathcal{F}_d$ be the sigma-field containing information up to depth d and apply Jensen's inequality with the convex function $x \mapsto x\log(x)$, as well as the martingale property

$$\mathbb{E}_{d+1}^{\mathrm{ind}}[L_{d+1} \mid \mathcal{F}_d] = L_d,$$

to obtain

$$\begin{aligned}
\mathrm{KL}_{d+1} &= \mathbb{E}_{d+1}^{\mathrm{ind}}\big[\mathbb{E}_{d+1}^{\mathrm{ind}}[L_{d+1}\log(L_{d+1}) \mid \mathcal{F}_d]\big] \\
&\geq \mathbb{E}_{d+1}^{\mathrm{ind}}\big[\mathbb{E}_{d+1}^{\mathrm{ind}}[L_{d+1} \mid \mathcal{F}_d]\log\big(\mathbb{E}_{d+1}^{\mathrm{ind}}[L_{d+1} \mid \mathcal{F}_d]\big)\big] \\
&= \mathbb{E}_{d}^{\mathrm{ind}}[L_d\log(L_d)] = \mathrm{KL}_d\,.
\end{aligned}$$

Consequently, the sequence $(\mathrm{KL}_d)_d$ has a (potentially infinite) limit $\mathrm{KL}_\infty \in [0,\infty]$.

The following theorem achieves the desired reformulation of one-sided testability by providing equivalent conditions that relate KL_d, L_d, and the parameters λ, s, s'.

Theorem 4.9. *In the tree correlation testing problem* (4.1)*, the following are equivalent:*

(a) *there exist one-sided tests $\mathcal{T}_d$ to decide $\mathbb{P}_d^{\mathrm{ind}}$ vs. $\mathbb{P}_d^{\mathrm{corr}}$;*

(b) *there is a sequence of thresholds $\theta_d \to \infty$ such that $\mathbb{P}_d^{\mathrm{ind}}(L_d > \theta_d) \to 0$ and* $\liminf_{d\to\infty}\mathbb{P}_d^{\mathrm{corr}}(L_d > \theta_d) > 0$*;*

(c) *the $\mathbb{P}^{\mathrm{ind}}$-martingale $(L_d)_d$ is not uniformly integrable;*

(d) *$\lambda ss' > 1$ and $KL_\infty = \infty$;*

(e) *$\lambda ss' > 1$ and* $\mathbb{P}^{\mathrm{corr}}(\liminf_d(\lambda ss')^{-d}\log(L_d) \geq C) \geq 1 - \mathbb{P}(\mathrm{Ext}(\mathrm{GW}^{(\lambda ss')}))$ *for a constant $C > 0$ only depending on (λ, s, s') and* $\mathrm{Ext}(\mathrm{GW}^{(\lambda ss')})$ *denoting the extinction event of the Galton–Watson process with offspring* $\mathrm{Poi}(\lambda ss')$.

This result was first presented in the symmetric model in [27, Theorem 1] and was extended to asymmetric correlated graphs in [35, Theorem 1]. Here, we provide a complete and rigorous proof that was partially omitted in [35].

Proof. We start by showing the equivalence (b) $\Leftrightarrow$ (c) followed by (a) $\Leftrightarrow$ (b) and concluding with the circular implications (b) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (b). We defer longer technical passages to Appendix A.5 for better readability.

(b) $\Leftrightarrow$ (c). Given (b), there exists a sequence $\theta_d \to \infty$ such that

$$\liminf_{d\to\infty}\mathbb{P}_d^{\mathrm{corr}}(L_d > \theta_d) > 0.$$

Reformulating this inequality, there exist $\varepsilon > 0$ and $d_0(\varepsilon) \in \mathbb{N}$ such that, for all $d \geq d_0(\varepsilon)$,

$$\mathbb{P}_d^{\mathrm{corr}}(L_d > \theta_d) = \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d > \theta_d}] \geq \varepsilon.$$

This contradicts the definition of uniform integrability of $(L_d)_d$.

For the opposite direction, suppose (c) to be true and invoke Lemma 4.6 to translate it to $\mathbb{E}^{\text{ind}}[L_\infty] < 1$. From this, one can construct a sequence $(\theta_d)_d$ which goes to infinity and satisfies the desired properties, i.e.,

$$\liminf_{d\to\infty} \mathbb{P}^{\text{corr}}(L_d > \theta_d) > 0 \quad \text{and} \quad \mathbb{P}^{\text{ind}}(L_d > \theta_d) \xrightarrow[d\to\infty]{} 0.$$

For details on the construction of $(\theta_d)_d$, we refer the reader to Appendix A.5.

(a) $\Leftrightarrow$ (b). While (a) follows from (b) by taking $\mathcal{T}_d := \mathbb{1}_{L_d > \theta_d}$, the other implication is more involved. Suppose that there exists a sequence of one-sided tests $\mathcal{T}_d$ fulfilling

$$\mathbb{P}^{\text{ind}}(\mathcal{T}_d = 1) =: \alpha_d \to 0 \quad \text{and} \quad \liminf_{d\to\infty} \underbrace{\mathbb{P}^{\text{corr}}(\mathcal{T}_d = 1)}_{=:\beta_d} > 0.$$

The Neyman–Pearson lemma as presented in [31] then yields an optimal test sequence $\widehat{\mathcal{T}}_d$ of the form

$$\widehat{\mathcal{T}}_d = \begin{cases} 1 & \text{if } L_d > \widehat{\theta}_d, \\ \mathbb{1}_{U<\gamma_d} & \text{for } U \sim \text{Unif}([0,1]) \text{ and } \gamma_d \in [0,1] \text{ if } L_d = \widehat{\theta}_d, \\ 0 & \text{if } L_d \le \widehat{\theta}_d, \end{cases}$$

with type-I error equal to that of $\mathcal{T}_d$, i.e.,

$$\alpha_d = \mathbb{P}^{\text{ind}}(\widehat{\mathcal{T}}_d = 1) = \mathbb{P}^{\text{ind}}(L_d > \widehat{\theta}_d) + \gamma_d \mathbb{P}^{\text{ind}}(L_d = \widehat{\theta}_d).$$

We differentiate between two cases depending on whether $\widehat{\theta}_d \to \infty$ or not.

Beginning with the case $\widehat{\theta}_d \not\to \infty$, we show in Appendix A.5 that

$$(\widehat{\theta}_d)_d \text{ is bounded} \implies \mathbb{E}^{\text{ind}}[L_\infty] < 1. \tag{4.5}$$

Since $\mathbb{E}^{\text{ind}}[L_\infty] < 1$ is equivalent to (c), which we have already shown to imply (b), this concludes the first case.

Assuming now that $\widehat{\theta}_d \to \infty$, we set $\theta_d := \widehat{\theta}_d - 1$ and use Markov's inequality to obtain

$$\mathbb{P}_d^{\text{ind}}(L_d > \theta_d) \le \frac{\mathbb{E}_d^{\text{ind}}[L_d]}{\theta_d} = \frac{1}{\widehat{\theta}_d - 1} \xrightarrow[d\to\infty]{} 0.$$

This shows that the likelihood ratio test has vanishing type-I error. It also has significant power since

$$\begin{aligned} \mathbb{P}_d^{\text{corr}}(L_d \ge \theta_d) &\ge \mathbb{P}_d^{\text{corr}}(L_d \ge \widehat{\theta}_d) \\ &\ge \mathbb{P}_d^{\text{corr}}(\mathcal{T}_d = 1) = \beta_d \implies \liminf_{d\to\infty} \mathbb{P}_d^{\text{corr}}(L_d \ge \theta_d) > 0. \end{aligned}$$

Hence, (b) holds true, concluding the equivalence we wanted to show.

(b) $\Rightarrow$ (d). Supposing that (b) is true, we refer to step 5 in the proof of [27, Theorem 1] for a demonstration of $\mathrm{KL}_\infty = \infty$. This proof completely ignores the symmetry assumption; it is therefore independent of s, s' and directly transferable to our asymmetric setting.

In order to show $\lambda s s' > 1$, we assume the contrary, i.e., $\lambda s s' \leq 1$. A standard result from Galton–Watson tree theory (see, for instance, [1]) states that in this case, 1 is the only fixed point in the interval [0, 1] of the probability generating function of $\mathrm{Poi}(\lambda s s')$. As seen in Section 4.3 ("martingale property"), the value $\mathbb{E}^{\mathrm{ind}}[L_\infty]$ fulfills the same fixed point equation, leading to $\mathbb{E}^{\mathrm{ind}}[L_\infty] = 1$. Consequently, one has trivial convergence of the means

$$\mathbb{E}^{\mathrm{ind}}[L_d] = 1 \to 1 = \mathbb{E}^{\mathrm{ind}}[L_\infty],$$

which we combine with the almost sure convergence $L_d \to L_\infty$ and Scheffé's lemma to obtain L^1-convergence of L_d to L_∞. This, in turn, yields that $(L_d)_d$ is a flat martingale, hence uniformly integrable. Since we have already shown (b) $\Leftrightarrow$ (c), this contradicts one-sided testability and lets us conclude that the assumption $\lambda s s' \leq 1$ must be wrong.

(d) $\Rightarrow$ (e). This implication is long and technical, requiring additional results about Galton–Watson trees. We defer it to Appendix A.5.

(e) $\Rightarrow$ (b). The assumption $\lambda s s' > 1$ implies $1 - \mathbb{P}(\mathrm{Ext}(\mathrm{GW}^{(\lambda s s')})) > 0$, and hence

$$\mathbb{P}^{\mathrm{corr}}\big(\liminf_{d\to\infty} (\lambda s s')^{-d} \log(L_d) \geq C\big) > 0.$$

Setting $\theta_d := \exp(C(\lambda s s')^d)$, one has $\theta_d \to \infty$ and Markov's inequality implies vanishing type-I error of the test $\mathbb{1}_{L_d > \theta_d}$:

$$\mathbb{P}^{\mathrm{ind}}(L_d \geq \theta_d) \leq \frac{\mathbb{E}^{\mathrm{ind}}[L_d]}{\theta_d} = \frac{1}{\theta_d} \xrightarrow[d\to\infty]{} 0.$$

To show significant power, we use Fatou's lemma to obtain

$$\begin{aligned}\liminf_{d\to\infty} \mathbb{P}^{\mathrm{corr}}(L_d \geq \theta_d) &\geq \mathbb{P}^{\mathrm{corr}}\big(\liminf_{d\to\infty}\{L_d \geq \theta_d\}\big)\\ &= \mathbb{P}^{\mathrm{corr}}\big(\exists\, D\ \forall d \geq D : L_d \geq \exp\big(C(\lambda s s')^d\big)\big)\\ &= \mathbb{P}^{\mathrm{corr}}\big(\liminf_{d\to\infty} (\lambda s s')^{-d} \log(L_d) \geq C\big) > 0.\end{aligned}$$

This implies (b) and concludes the proof of all equivalences we wanted to show. ∎

4.5. Sharp information-theoretic limits

The previous section translates feasibility of tree correlation testing to concrete statistical properties. Especially (a) $\Leftrightarrow$ (d) of Theorem 4.9 lets us focus on the Kullback–Leibler between $\mathbb{P}_d^{\mathrm{corr}}$ and $\mathbb{P}_d^{\mathrm{ind}}$. The following theorem relates the convergence of $(\mathrm{KL}_d)_d$ directly to the parameters λ, s and s', characterizing a sharp phase transition in asymmetric tree correlation testing around Otter's constant.

Theorem 4.10. *In the tree correlation testing problem* (4.1)*, parametrized by* $s, s' \in (0, 1]$ *and* $\lambda > 0$*, there is a feasibility phase transition around Otter's constant* $\alpha \approx 0.3383$*, which is characterized as follows:*

(i) *if* $ss' \leq \alpha$*, then* $\sup_{d\in\mathbb{N}} \mathrm{KL}_d < \infty$ *and tree correlation testing is impossible;*

(ii) *if* $ss' > \alpha$ *and* $\lambda > \lambda_0$ *for a threshold* $\lambda_0 = \lambda_0(s, s')$ *coming from a Gaussian approximation argument, then* $\mathrm{KL}_d \to \infty$ *and tree correlation testing is possible.*

The conclusions in (i) and (ii) about tree correlation testing are immediate consequences of Theorem 4.9. In particular, one needs to require $\lambda_0 ss' > 1$ for the conclusion in (ii) to be true, adding a requirement on the magnitude of λ_0.

Remark 4.11. For large enough λ, the only parameter determining feasibility of the testing problem is the product ss'. This has the effect that one density parameter s of the first tree can compensate for lower density s' in the second tree. We refer to the concluding Section 5 for the implications of this result on graph alignment.

Remark 4.12 (On the size of λ_0). The sharpness statement in Theorem 4.10 holds only under the additional condition $\lambda > \lambda_0$, which arises from the Gaussian approximation argument later in this section. It is natural to conjecture that $\lambda ss' > 1$ is sufficient for $\mathrm{KL}_d \to \infty$, even more so in the light of Corollary 4.8, which states that $\mathbb{E}^{\mathrm{ind}}[L_d^2]$ only depends on ss', not λ. However, the only link we can establish between KL_d and $\mathbb{E}^{\mathrm{ind}}[L_d^2]$ goes through Gaussian approximation, and we do not have a principled reason to expect such a link to hold beyond that regime. Determining the minimal value of λ_0 therefore remains open.

For the remainder of this section, we prove Theorem 4.10. Showing (i) is straightforward due to our preparatory work, the proof of (ii) requires more involved techniques.

Proof of Theorem 4.10 (i). Let $ss' \leq \alpha$. We start by upper bounding the Kullback–Leibler by using Jensen's inequality:

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}_d^{\mathrm{corr}} \| \mathbb{P}_d^{\mathrm{ind}}\big) &= \mathbb{E}_d^{\mathrm{ind}}\big[L_d \log(L_d)\big] = \mathbb{E}_d^{\mathrm{corr}}\big[\log(L_d)\big] \\ &\leq \log\big(\mathbb{E}_d^{\mathrm{corr}}[L_d]\big) = \log\big(\mathbb{E}_d^{\mathrm{ind}}[L_d^2]\big).\end{aligned}$$

This second moment is characterized in Corollary 4.8. Combining that result with $|\mathcal{X}_d^{(n)}| \le |\mathcal{X}_{n-1}^{(n)}|$ yields

$$\mathbb{E}_d^{\mathrm{ind}}[L_d^2] = \sum_{n=1}^{\infty} |\mathcal{X}_d^{(n)}|(ss')^{n-1} \le (ss')^{-1} \sum_{n=1}^{\infty} |\mathcal{X}_{n-1}^{(n)}|(ss')^n.$$

Due to the asymptotic formula for $|\mathcal{X}_{n-1}^{(n)}|$ from Proposition 4.1, we have

$$\sum_{n=1}^{\infty} |\mathcal{X}_{n-1}^{(n)}|(ss')^n < \infty \quad \iff \quad \sum_{n=1}^{\infty} C n^{-3/2} \frac{(ss')^n}{\alpha^n} < \infty,$$

where finiteness on the right follows from $ss' \le \alpha$. Consequently, $\mathrm{KL}(\mathbb{P}_d^{\mathrm{corr}} \| \mathbb{P}_d^{\mathrm{ind}})$ is bounded above by the logarithm of a finite value, which concludes the proof. ∎

The use of Jensen's inequality in the above proof prevents us from using the same argument to show $\mathrm{KL}(\mathbb{P}_d^{\mathrm{corr}} \| \mathbb{P}_d^{\mathrm{ind}}) = \infty$ in the case where $ss' > \alpha$. This direction requires some additional tools, which we will introduce next.

Proof of Theorem 4.10 (ii). Instead of being true for any value of λ, we recall that point (ii) of Theorem 4.10 only holds for all λ surpassing a certain threshold λ_0. This phenomenon has its roots in our proof technique: We will first examine the asymptotic behavior of KL_d when letting $\lambda \to \infty$ and leaving d fixed. This is done via a Gaussian approximation argument, which is introduced in the following lemma.

This lemma will not be precisely enough to show the desired statement, which requires some refined analysis afterward. Nonetheless, it contains the key idea of the proof.

Lemma 4.13 (Gaussian approximation). *Let $d \in \mathbb{N}$ and recall that $\mathcal{X}_{d+1} = \mathbb{N}^{\mathcal{X}_d}$ by the subtree tuple identification. There exists a pair of functions*

$$(y, y') \colon \mathcal{X}_{d+1} \times \mathcal{X}_{d+1} \to \mathbb{R}^{\mathcal{X}_d} \times \mathbb{R}^{\mathcal{X}_d}, \quad (t, t') \mapsto \big(y_\beta(t), y'_\beta(t')\big)_{\beta \in \mathcal{X}_d},$$

with the property that the above map is affine and bijective. The law of the random vector $(y(t), y'(t'))$ shall be denoted either as $\mathcal{L}(y, y')$ in the case $(t, t') \sim \mathbb{P}_{d+1}^{\mathrm{corr}}$, or as $\mathcal{L}(y) \otimes \mathcal{L}(y')$ in the case $(t, t') \sim \mathbb{P}_{d+1}^{\mathrm{ind}}$. One has

$$\mathrm{KL}\big(\mathbb{P}_{d+1}^{\mathrm{corr}} \| \mathbb{P}_{d+1}^{\mathrm{ind}}\big) = \mathrm{KL}\big(\mathcal{L}(y, y') \| \mathcal{L}(y) \otimes \mathcal{L}(y')\big)$$

and that $\mathcal{L}(y, y')$ converges weakly towards the joint law of a pair of random variables

$$\mathcal{L}(y, y') \xrightarrow[\lambda \to \infty]{w} \mathcal{L}(z, z').$$

The pair $(z, z') = ((z_\beta)_{\beta \in \mathcal{X}_d}, (z'_\beta)_{\beta \in \mathcal{X}_d})$ *is a Gaussian vector of infinite dimension with zero mean and covariances defined by*

$$\mathbb{E}[z_\beta z_\gamma] = \mathbb{E}[z'_\beta z'_\gamma] = \mathbb{1}_{\beta=\gamma}, \quad \mathbb{E}[z_\beta z'_\gamma] = \sqrt{ss'}^{|\beta|} \mathbb{1}_{\beta=\gamma} \quad \textit{for all } \beta, \gamma \in \mathcal{X}_d.$$

The proof of this lemma is rather technical and exploits the orthogonality properties from Theorem 4.7. The details are in Appendix A.6, in particular, how the vector (y, y') is defined and how Levy's Continuity theorem implies its weak convergence to (z, z').

For the remainder of this section, let $ss' \geq \alpha$ and recall that our goal is to show

$$\mathrm{KL}_d \xrightarrow[d\to\infty]{} \infty$$

for sufficiently large λ. With Lemma 4.13 at hand, we can take a first step towards this goal and establish a lower bound on KL for large λ.

Lemma 4.14. *For any depth* $d \in \mathbb{N}$, *one has the following asymptotic lower bound on* KL_{d+1}*:*

$$\liminf_{\lambda\to\infty} \mathrm{KL}\big(\mathbb{P}^{\mathrm{corr}}_{d+1} \| \mathbb{P}^{\mathrm{ind}}_{d+1}\big) \geq \sum_{n=1}^{\infty} |\mathcal{X}^{(n)}_{d+1}| (ss')^{n-1}.$$

Proof. The key idea for a lower bound on $\mathrm{KL}(\mathbb{P}^{\mathrm{corr}}_{d+1} \| \mathbb{P}^{\mathrm{ind}}_{d+1})$ is the *lower semi-continuity of the Kullback–Leibler divergence* (see, for instance, [34]): For two sequences of weakly converging random variables

$$X_\lambda \xrightarrow[\lambda\to\infty]{w} X \quad \text{and} \quad Y_\lambda \xrightarrow[\lambda\to\infty]{w} Y,$$

one has

$$\liminf_{\lambda\to\infty} \mathrm{KL}\big(\mathcal{L}(X_n) \| \mathcal{L}(Y_n)\big) \geq \mathrm{KL}\big(\mathcal{L}(X) \| \mathcal{L}(Y)\big).$$

Using Lemma 4.13, we therefore obtain

$$\begin{aligned}\liminf_{\lambda\to\infty} \mathrm{KL}\big(\mathbb{P}^{\mathrm{corr}}_{d+1} \| \mathbb{P}^{\mathrm{ind}}_{d+1}\big) &= \liminf_{\lambda\to\infty} \mathrm{KL}\big(\mathcal{L}(y, y') \| \mathcal{L}(y) \otimes \mathcal{L}(y')\big) \\ &\geq \mathrm{KL}\big(\mathcal{L}(z, z') \| \mathcal{L}(z) \otimes \mathcal{L}(z')\big).\end{aligned}$$

The right-hand side can be explicitly computed after some observations: Recall that the covariance structure of (z, z') is given by

$$\mathbb{E}[z_\beta z_\gamma] = \mathbb{E}[z'_\beta z'_\gamma] = \mathbb{1}_{\beta=\gamma}, \quad \mathbb{E}[z_\beta z'_\gamma] = \sqrt{ss'}^{|\beta|} \mathbb{1}_{\beta=\gamma} \quad \text{for all } \beta, \gamma \in \mathcal{X}_d.$$

This leads us to observe, on the one hand, that an infinite Gaussian vector of law $\mathcal{L}(z) \otimes \mathcal{L}(z')$ has the infinite covariance matrix $\mathrm{diag}(1, 1, 1, \ldots)$. On the other hand, for a vector Z of law $\mathcal{L}(z, z')$, we choose an arbitrary numbering of the trees in $\mathcal{X}_d$, namely $\beta_1, \beta_2, \beta_3, \ldots$ and suppose that the coordinates are ordered as

$$Z = (z_{\beta_1}, z'_{\beta_1}, z_{\beta_2}, z'_{\beta_2}, z_{\beta_3}, z'_{\beta_3}, \ldots).$$

Note that this ordering does not affect the Kullback–Leibler divergence since applying a permutation to the covariance matrix is a similarity transformation.

Observe that the covariance of Z is an infinite block diagonal matrix with the blocks

$$\begin{pmatrix} 1 & \sqrt{ss'}^{|\beta_i|} \\ \sqrt{ss'}^{|\beta_i|} & 1 \end{pmatrix} \quad \text{for } i = 1, 2, 3, \ldots$$

This enables us to compute

$$\begin{aligned} \mathrm{KL}\big(\mathcal{L}(z, z') \| \mathcal{L}(z) \otimes \mathcal{L}(z')\big) &= \mathrm{KL}\big(\mathcal{N}(0, \mathrm{Cov}(Z)) \| \mathcal{N}(0, I)\big) \\ &= -\frac{1}{2} \log\big(\det(\mathrm{Cov}(Z))\big), \end{aligned}$$

where the last equality is a simple computation; see, for instance, [38].

One can deduce that

$$\det(\mathrm{Cov}(Z)) = \prod_{i=1}^{\infty} \big(1 - (ss')^{|\beta_i|}\big)$$

from the block diagonal structure of the covariance matrix. Consequently,

$$\begin{aligned} \liminf_{\lambda \to \infty} \mathrm{KL}\big(\mathbb{P}^{\mathrm{corr}}_{d+1} \| \mathbb{P}^{\mathrm{ind}}_{d+1}\big) &\geq -\frac{1}{2} \log\Bigg(\prod_{\beta \in \mathcal{X}_d} \big(1 - (ss')^{|\beta|}\big)\Bigg) \\ &= \frac{1}{2} \log\Bigg(\prod_{\beta \in \mathcal{X}_d} \frac{1}{1 - (ss')^{|\beta|}}\Bigg). \end{aligned}$$

Since $ss' < 1$ and $\mathcal{X}_d$ is countable, we may develop the geometric series

$$\begin{aligned} \prod_{\beta \in \mathcal{X}_d} \frac{1}{1 - (ss')^{|\beta|}} &= \prod_{\beta \in \mathcal{X}_d} \sum_{\gamma_\beta = 0}^{\infty} (ss')^{\gamma_\beta |\beta|} \\ &= \sum_{(\gamma_\beta)_\beta \in \mathbb{N}^{\mathcal{X}_d}} (ss')^{\sum_\beta \gamma_\beta |\beta|} \\ &= \sum_{\gamma \in \mathcal{X}_{d+1}} (ss')^{|\gamma| - 1}, \end{aligned}$$

where for the last equality we recall that

$$|\gamma| = 1 + \sum_{\beta} \gamma_\beta |\beta|$$

for any $\gamma = (\gamma_\beta)_{\beta \in \mathcal{X}_d} \in \mathcal{X}_{d+1}$. Developing the right-hand expression as in the proof of Corollary 4.8 yields

$$\sum_{\gamma \in \mathcal{X}_{d+1}} (ss')^{|\gamma|-1} = \sum_{n=1}^{\infty} \sum_{\substack{\gamma \in \mathcal{X}_{d+1}, \\ |\gamma|=n}} (ss')^{n-1} = \sum_{n=1}^{\infty} |\mathcal{X}_{d+1}^{(n)}| (ss')^{n-1}.$$

This concludes the proof of the lemma. ∎

Letting $d \to \infty$ in Lemma 4.14 yields

$$\begin{aligned} \lim_{d\to\infty} \liminf_{\lambda\to\infty} \mathrm{KL}\big(\mathbb{P}_{d+1}^{\mathrm{corr}} \| \mathbb{P}_{d+1}^{\mathrm{ind}}\big) &\geq \lim_{d\to\infty} \sum_{n=1}^{\infty} |\mathcal{X}_d^{(n)}| (ss')^{n-1} \\ &= \sum_{n=1}^{\infty} |\mathcal{X}_{n-1}^{(n)}| (ss')^{n-1}, \end{aligned}$$

where the right-hand side diverges due to $ss' \geq \alpha$ and the asymptotics of $|\mathcal{X}_{n-1}^{(n)}|$ from Proposition 4.1. If the limits in d and λ were swapped, this would give the desired results, but we are not allowed to do so. Instead, we need another proof strategy, which we sketch here before providing the details:

(1) From the fact that $\liminf_\lambda \mathrm{KL}(\mathbb{P}_{d+1}^{\mathrm{corr}} \| \mathbb{P}_{d+1}^{\mathrm{ind}})$ diverges for $d \to \infty$, one can derive that there is a depth d_0 and a test $\mathcal{T}_{d_0}$ fulfilling

$$\mathbb{P}_{d_0}^{\mathrm{corr}}(\mathcal{T}_{d_0} = 1) \geq \beta \quad \text{and} \quad \mathbb{P}_{d_0}^{\mathrm{ind}}(\mathcal{T}_{d_0} = 1) \leq \varepsilon. \tag{4.6}$$

(2) This test can then be amplified in the next generation $d_0 + 1$, i.e., there exists a test $\mathcal{T}_{d_0+1}$ based on $\mathcal{T}_{d_0}$ such that for the same constants as before

$$\mathbb{P}_{d_0+1}^{\mathrm{corr}}(\mathcal{T}_{d_0+1} = 1) \geq \beta \quad \text{and} \quad \mathbb{P}_{d_0+1}^{\mathrm{ind}}(\mathcal{T}_{d_0+1} = 1) \leq \frac{\varepsilon}{2}. \tag{4.7}$$

(3) The test amplification strategy is independent of the generation d, meaning we can propagate the bound arbitrarily often. This leads to a sequence of tests $(\mathcal{T}_d)_d$ which share the lower bound $\beta > 0$ on their power and have vanishing type-I error since $\mathbb{P}_d^{\mathrm{ind}}(\mathcal{T}_d = 1) \leq \varepsilon/2^{-(d-d_0)}$. In other words, we obtain a sequence of one-sided tests.

(4) By Theorem 4.9, this is equivalent to the divergence of $\mathrm{KL}(\mathbb{P}_{d+1}^{\mathrm{corr}} \| \mathbb{P}_{d+1}^{\mathrm{ind}})$, which concludes the proof of Lemma 4.13.

While points (3) and (4) of this sketch are rigorous, points (1) and (2) require more detail. This is the content of the following two lemmata, which are proved afterward.

Lemma 4.15 (Point (1) of the proof sketch). *Let $\beta \in (0, 2/15)$ and $\varepsilon > 0$. Then, there exists $\lambda_1 > 0$ and $d_0 \in \mathbb{N}$ such that, for all $\lambda \geq \lambda_1$, there is a test $\mathcal{T}_{d_0} \colon \mathcal{X}_{d_0} \times \mathcal{X}_{d_0} \to \{0, 1\}$ fulfilling*

$$\mathbb{P}^{\mathrm{corr}}_{d_0}(\mathcal{T}_{d_0} = 1) \geq \beta \quad \textit{and} \quad \mathbb{P}^{\mathrm{ind}}_{d_0}(\mathcal{T}_{d_0} = 0) \leq \varepsilon.$$

Lemma 4.16 (Point (2) of the proof sketch). *Given $\beta \in (0, 1)$, there exist constants $\varepsilon > 0$ and $\lambda_0 > 0$ such that the following implication holds: If for all $\lambda \geq \lambda_0$ and any depth $d \in \mathbb{N}$, there exists a test $\mathcal{T}_d \colon \mathcal{X}_d \times \mathcal{X}_d \to \{0, 1\}$ with*

$$\mathbb{P}^{\mathrm{corr}}_{d}(\mathcal{T}_{d} = 1) \geq \beta \quad \textit{and} \quad \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_{d} = 1) \leq \varepsilon,$$

then one can define another test $\mathcal{T}_{d+1} \colon \mathcal{X}_{d+1} \times \mathcal{X}_{d+1} \to \{0, 1\}$ fulfilling the amplified test conditions

$$\mathbb{P}^{\mathrm{corr}}_{d+1}(\mathcal{T}_{d+1} = 1) \geq \beta \quad \textit{and} \quad \mathbb{P}^{\mathrm{ind}}_{d+1}(\mathcal{T}_{d+1} = 1) \leq \frac{\varepsilon}{2}$$

for all $\lambda \geq \lambda_0$.

To see how these results imply points (1) and (2) from the above proof sketch, fix a constant $\beta \in (0, 2/15)$ and let (ε, λ_0) be the corresponding quantities from Lemma 4.16. By Lemma 4.15, there exist λ_1 and d_0 such that (4.6) holds for all $\lambda \geq \lambda_1$. Hence, for $\lambda \geq \max\{\lambda_0, \lambda_1\}$, both lemmata can be applied.

Since (4.6) is nothing but the implication assumption in Lemma 4.16 for $d = d_0$, this lemma generates another test at generation $d_0 + 1$ fulfilling the amplified test conditions presented in (4.7). Point (3) of the proof sketch then iterates the utilization of Lemma 4.16, which concludes the proof of Theorem 4.10 (ii). ∎

All that remains to do now is show both Lemmata 4.15 and 4.16.

The proof of Lemma 4.15 can be copied line-by-line from [28, Appendix B.1], up to notational differences and the fact that the initial condition changes from $s > \sqrt{\alpha}$ to the asymmetric counterpart $ss' > \alpha$. We refer the reader to the original source.

The same cannot be said about the proof of Lemma 4.16, which may be of particular interest since it shows how to obtain an amplified test $\mathcal{T}_{d+1}$ from a lower-depth test $\mathcal{T}_d$.

Proof of Lemma 4.16. Fix a constant $\beta \in (0, 1)$ and a depth d. Next, suppose that there exist $\varepsilon > 0$, $\lambda_0 > 0$ and a test $\mathcal{T}_d \colon \mathcal{X}_d \times \mathcal{X}_d \to \{0, 1\}$ such that for all $\lambda \geq \lambda_0$,

$$\mathbb{P}^{\mathrm{corr}}_{d}(\mathcal{T}_{d} = 1) \geq \beta \quad \text{and} \quad \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_{d} = 1) \leq \varepsilon.$$

To find an amplified test $\mathcal{T}_{d+1}$, we define the value

$$Z(t, t') := \sum_{\substack{(\tau, \tau') \in \mathcal{X}_d^2, \\ \mathcal{T}_d(\tau, \tau') = 1}} \big(N_\tau - \lambda s \, \mathrm{GW}_d^{(\lambda s)}(\tau)\big)\big(N'_{\tau'} - \lambda s' \, \mathrm{GW}_d^{(\lambda s')}(\tau')\big),$$

where $t = (N_\tau)_\tau$ and $t' = (N'_{\tau'})_{\tau'}$ form a pair of trees in $\mathcal{X}_{d+1}$. For some threshold ξ to be determined, we derive the test $\mathcal{T}_{d+1}(t,t') := \mathbb{1}_{Z(t,t')\geq\xi}$, which must fulfill the two requirements

$$\mathbb{P}^{\mathrm{corr}}_{d+1}(\mathcal{T}_{d+1} = 1) = \mathbb{P}^{\mathrm{corr}}_{d+1}\big(Z(t,t') \geq \xi\big) \geq \beta \tag{4.8}$$

and

$$\mathbb{P}^{\mathrm{ind}}_{d+1}(\mathcal{T}_{d+1} = 1) = \mathbb{P}^{\mathrm{ind}}_{d+1}\big(Z(t,t') \geq \xi\big) \leq \frac{\varepsilon}{2}. \tag{4.9}$$

The technical tool for showing these equations is the following lemma, whose proof can be found in Appendix A.7.

Lemma 4.17 (Moments of Z). *The random variable $Z := Z(t,t')$ has the following moments under $\mathbb{P}^{\mathrm{ind}}$ and $\mathbb{P}^{\mathrm{corr}}$:*

(a) $\mathbb{E}^{\mathrm{ind}}_{d+1}[Z] = 0$;
(b) $\mathbb{E}^{\mathrm{corr}}_{d+1}[Z] = \lambda s s' \mathbb{P}^{\mathrm{corr}}_{d}(\mathcal{T}_d = 1)$;
(c) $\mathbb{E}^{\mathrm{ind}}_{d+1}[Z^4] \leq 36(\lambda^2 s s')^2 \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1)^2 + 13\lambda^3 s s' \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1)$;
(d) $\mathrm{Var}^{\mathrm{corr}}_{d+1}[Z] \leq \lambda^2 s s'(1 + s s')\mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1) + \mathbb{E}^{\mathrm{corr}}_{d+1}[Z]$.

To show equation (4.9), we simply combine Markov's inequality with point (c) and the hypothesis $\mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1) \leq \varepsilon$, which yields

$$\begin{aligned}\mathbb{P}^{\mathrm{ind}}_{d+1}\big(Z(t,t') \geq \xi\big) &\leq \frac{1}{\xi^4}\mathbb{E}^{\mathrm{ind}}_{d+1}\big[Z(t,t')^4\big] \\ &\leq \frac{1}{\xi^4}\big[36(\lambda^2 s s')^2 \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1)\varepsilon + 13\lambda^3 s s' \varepsilon\big].\end{aligned}$$

Since $A + B \leq 2\max\{A, B\}$, choosing

$$\xi^4 \geq 2\max\big\{72(\lambda^2 s s')^2 \mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1), 26\lambda^3 s s'\big\}$$

lets us bound the right-hand side by $\varepsilon/2$, which implies $\mathbb{P}^{\mathrm{ind}}_{d+1}(Z(t,t') \geq \xi) \leq \varepsilon/2$. Consequently, setting

$$\xi = \max\big\{4\lambda\sqrt{s s'}\mathbb{P}^{\mathrm{ind}}_{d}(\mathcal{T}_d = 1)^{1/4}, 3\lambda^{3/4}\sqrt[4]{s s'}\big\}$$

implies (4.9).

To show equation (4.8), we first assume $\xi \leq \mathbb{E}^{\mathrm{corr}}_{d+1}[Z]/2$ and compute

$$\begin{aligned}\mathbb{P}^{\mathrm{corr}}_{d+1}(Z \leq \xi) &\leq \mathbb{P}^{\mathrm{corr}}_{d+1}\Big(Z \leq \frac{1}{2}\mathbb{E}^{\mathrm{corr}}_{d+1}[Z]\Big) \\ &\leq \mathbb{P}^{\mathrm{corr}}_{d+1}\Big(|Z - \mathbb{E}^{\mathrm{corr}}_{d+1}[Z]| \geq \frac{1}{2}\mathbb{E}^{\mathrm{corr}}_{d+1}[Z]\Big)\end{aligned}$$

$$
\begin{aligned}
&\leq 4\frac{\mathrm{Var}^{\mathrm{corr}}_{d+1}[Z]}{\mathbb{E}^{\mathrm{corr}}_{d+1}[Z]^2} \\
&\leq 4\frac{\lambda^2 ss'(1+ss')\,\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1)+\lambda ss'\mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d=1)}{\lambda^2 s^2(s')^2\,\mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d=1)^2} \\
&\leq \frac{8\,\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1)}{ss'\beta^2}+\frac{4}{\lambda ss'\beta}.
\end{aligned}
$$

We explain the inequalities one by one: The first one exploits that $\xi \leq \mathbb{E}^{\mathrm{corr}}_{d+1}[Z]/2$ and the second line is true because $0 \leq A \leq B$ implies $|A-2B| \geq B$. Thirdly, we apply Markov's inequality; next, we use points (b) and (d) from Lemma 4.17, and the final inequality follows from $1+ss' \leq 2$ as well as the hypothesis $\mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d=1) \geq \beta$.

Note that one has the implication

$$
\mathbb{P}^{\mathrm{corr}}_{d+1}(Z \leq \xi) \leq 1-\beta \quad\Longrightarrow\quad \mathbb{P}^{\mathrm{corr}}_{d+1}(Z \leq \xi) \geq \beta,
$$

where the latter expression corresponds to (4.8). Hence, all that remains to be done is choosing λ_0 and ε such that for all $\lambda \geq \lambda_0$, the following two equations hold:

$$
\max\left\{4\lambda\sqrt{ss'}\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1)^{1/4}, 3\lambda^{3/4}\sqrt[4]{ss'}\right\} = \xi \leq \frac{1}{2}\mathbb{E}^{\mathrm{corr}}_{d+1}[Z], \tag{4.10}
$$

$$
\frac{8\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1)}{ss'\beta^2}+\frac{4}{\lambda ss'\beta} \leq 1-\beta. \tag{4.11}
$$

Imposing

$$
\lambda \geq \frac{8}{ss'(1-\beta)} \quad\text{and}\quad \mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1) \leq \frac{(1-\beta)ss'\beta^2}{16}
$$

ensures (4.11). Next, by combining

$$
\mathbb{E}^{\mathrm{corr}}_{d+1}[Z] = \lambda ss'\mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d=1) \quad\text{with}\quad \mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d=1) \geq \beta,
$$

one observes that (4.10) is implied by

$$
4\lambda\sqrt{ss'}\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1)^{1/4} \leq \frac{1}{2}\lambda ss'\beta \quad\text{and}\quad 3\lambda^{3/4}\sqrt[4]{ss'} \leq \frac{1}{2}\lambda ss'\beta.
$$

This is equivalent to

$$
\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d=1) \leq \left(\frac{\sqrt{ss'}\beta}{8}\right)^4 \quad\text{and}\quad \frac{6^4}{\beta^4(ss')^3} \leq \lambda.
$$

Finally, we set

$$
\lambda_0 := \max\left\{\frac{6^4}{\beta^4(ss')^3}, \frac{8}{ss'(1-\beta)}\right\} \quad\text{and}\quad \varepsilon := \min\left\{\left(\frac{\sqrt{ss'}\beta}{8}\right)^4, \frac{(1-\beta)ss'\beta^2}{16}\right\}.
$$

Note that both variables only depend on β, s and s'. Consequently, we could have set these values from the start, and all arguments would work without extra assumptions, independently from the depth d. This concludes the proof. ∎

5. Conclusion

In this work, we extend the graph alignment problem to the asymmetric case with varying edge densities and different node numbers. This advances the applicability of graph alignment, but our analysis remains theoretical within the well-studied Erdős–Rényi setting. Section 2 provides a rigorous model definition, adjusted performance metrics, and the idea of one-sided partial alignment in the asymmetric case.

Since our study focuses on sparse graphs, Section 3 translates the problem to correlation testing for local Galton–Watson trees. We generalize prior work to an asymmetric tree model and show that Algorithm 1 achieves alignment on the graph level by comparisons on the tree level.

Finally, Section 4 investigates asymmetric tree correlation testing in more detail. Leveraging advantageous properties of the likelihood ratio, particularly its diagonalization formula, we demonstrate a phase transition in the feasibility of tree correlation testing. The sharp threshold $ss' = \alpha$ provides insights not only for tree correlation testing but also for graph alignment, as summarized in the following corollary.

Corollary 5.1 (Combination of Theorem 3.10 and Theorem 4.10). *For $\lambda > 0$ large enough and $ss' > \alpha$ where α is Otter's constant, Algorithm 1 achieves one-sided recovery of the true alignment σ_*. In particular, for $s = 1$ and $\alpha < s' < 1$, in which case G' is a proper random subgraph of G, one recovers a significant portion of node correspondences while making negligible error.*

This result has new theoretical implications for random subgraph matching. While the general subgraph isomorphism problem is NP-hard, the here presented random version is feasible in polynomial time. This resolves an open problem discussed in personal communication with Jiaming Xu[2] and naturally extends the work of [6, 9].

Open questions. While Corollary 5.1 represents a positive result for asymmetric tree correlation testing, some questions remain:

- We know from Theorem 4.10 that $ss' \leq \alpha$ implies impossibility of tree correlation testing. It remains an open question whether sparse graph alignment is also infeasible in this regime or if a non-local algorithm could function.
- MPALIGN achieves one-sided recovery but its theoretical power is lowered by the dangling tree trick. Disregarding nodes with fewer than three neighbors limits the algorithm's ability to align a significant portion of the giant component. A more refined analysis to extend the algorithm's scope is left for future research.

[2] J. Xu, personal communication at MiFODS workshop, MIT (Cambridge, MA, USA), August 19–22, 2019.

Aside from a refined analysis of the problem at hand, there are several directions in which the concept of asymmetric correlated graphs could enrich the alignment literature. Here are two examples:

- Within the sparse regime, one natural generalization is to study the alignment of correlated configuration models with fixed degree sequences.
- In the denser settings of the correlated Erdős–Rényi model, numerous findings can be generalized to the asymmetric setting. We expect that some extend naturally, while others may require new techniques to accommodate varying node numbers and edge densities.

Many geometric graphs and real-world networks exhibit inherent asymmetry, necessitating alignment techniques beyond the classical symmetric setting. We hope to contribute a step towards deeper understanding of correlated networks beyond the theoretical playground of symmetrical pairs of Erdős–Rényi graphs.

A. Deferred proofs

A.1. Proof of Lemma 3.11

We recall the lemma about the second moment of R_d before moving to its proof. Here, we shift the index from $d-1$ to d compared to Lemma 3.11, which does not impact the result and simplifies notations.

Lemma A.1 (Copy of Lemma 3.11). *Under the parameters of Theorem 3.10 and using the notations from the proof of Theorem 3.10, one has*

$$\mathbb{E}_d^{\mathrm{ind}}\big[R_d^2 \mathbb{1}_{\mathcal{A}^c \cap \mathcal{B}}\big] \in \mathcal{O}(1).$$

Proof. This proof uses vocabulary and ideas from techniques to establish relationships between graph neighborhoods and branching processes; cf. [3, p. 161] and [10].

Let $(G, G') \sim \mathrm{CER}(N, \lambda, s, s')$ and let $i \in V$ as well as $j \in V'$. We describe the sampling process from $\mathbb{P}^{\mathrm{dis}}$ as a simultaneous exploration of G and G', starting at i and j. This process is conditioned on the following hypotheses from Definition 3.2:

(a) either $i \notin V_*$ or $\sigma_*(i) \neq j$;

(b) $\mathcal{N}_{G,2d}(i)$ and $\mathcal{N}_{G',2d}(j)$ are trees;

(c) $\mathcal{N}_{G,d}(i) \cap V_*$ and $\mathcal{N}_{G',d}(j) \cap V'_*$ are disjoint as subgraphs of G_*.

Note that this is not precisely the vocabulary of dangling trees since we are not talking about pointing neighborhoods: Considering $\mathcal{N}_{G,d}(i)$ instead of $\mathcal{N}_{G,d}(i \to i_k)$ simplifies notation but does not influence this proof which could well be written with pointed neighborhoods.

Sampling from $\mathbb{P}^{\text{dis}}$ can be seen as the following graph exploration algorithm: It describes the sampling process of $(\mathcal{N}_{G,d}(i), \mathcal{N}_{G',d}(j))$ conditioned on (a), (b), (c).

Initialization. Initialize $\mathcal{N}_{G,d}(i) := \{i\}$ and $\mathcal{N}_{G',d}(j) := \{j\}$ as the trivial trees consisting only of their roots. These roots are assigned depth 0 in their respective trees. Next, we consider three types of node sets during the exploration process:

- The active sets $\mathcal{A}^*$, $\mathcal{A}$, and $\mathcal{A}'$ which are subsets of V_* (which is identified as V'_* in G'), $V \setminus V_*$, and $V' \setminus V'_*$, respectively. If i is a node of G_*, one initializes $\mathcal{A}^* = \{i\}$, $\mathcal{A} = \emptyset$; else if $i \in V \setminus V_*$, then $\mathcal{A}^* = \emptyset$, $\mathcal{A} = \{i\}$. Similarly, add j to $\mathcal{A}^*$ if contained in G_*, or add it to $\mathcal{A}'$ otherwise.
- The deactivated sets $\mathcal{D}^*$, $\mathcal{D}$, and $\mathcal{D}'$, which are all initialized as $\emptyset$.
- The sets of unvisited nodes $\mathcal{U}^* := V_* \setminus \mathcal{A}^*$, $\mathcal{U} := V \setminus \mathcal{A}$, and $\mathcal{U}' := V'_* \setminus \mathcal{A}'$.

Let A^* denote the cardinality of $\mathcal{A}^*$, D' the cardinality of $\mathcal{D}'$, and so forth for all sets. At all times during the sampling process, one has

$$U^* = n_* - A^* - D^*, \quad U = n_+ - A - D, \quad \text{and} \quad U' = n'_+ - A' - D'.$$

Exploration of the unvisited nodes. While at least one active set ($\mathcal{A}^*$, $\mathcal{A}$, or $\mathcal{A}'$) is non-empty and contains nodes of depth d or less (with respect to the root in their respective graph), execute the following steps:

(1) Arbitrarily pick one node $k \in \mathcal{A}^* \cup \mathcal{A} \cup \mathcal{A}'$ with minimal depth among all such nodes, remove it from its respective active set, and add it to the corresponding deactivated set. For instance, if $k \in \mathcal{A}'$, then move k to $\mathcal{D}'$.

(2) If k is part of $\mathcal{N}_{G,d}(i)$, then sample $\text{Bin}(U^*, \lambda r r'/N)$ child nodes from $\mathcal{U}^*$ and $\text{Bin}(U, \lambda r(1 - r')/N)$ child nodes from $\mathcal{U}$. Remove these child nodes from their respective unvisited node sets and add them to the corresponding active sets.

(3) Else if k is part of $\mathcal{N}_{G',d}(j)$, then sample $\text{Bin}(U^*, \lambda r r'/N)$ child nodes from $\mathcal{U}^*$ and $\text{Bin}(U', \lambda r'(1 - r)/N)$ child nodes from $\mathcal{U}'$. Again, remove these child nodes from their respective unvisited node sets and add them to the corresponding active sets.

(4) Assign correct depths to all new active nodes: If k was of depth d_k, then all children are assigned depth $d_k + 1$.

Computation of the likelihood under $\mathbb{P}_d^{\text{dis}}$. Fix two trees t and t' of depth d and impose an arbitrary breadth-first-search numbering across both trees: Assign the numbers 1 and 2 to the two roots, then number subsequent generations such that nodes of higher depth have strictly higher numbers. Let $K = |t| + |t'|$ be the total number of nodes across both trees and denote by c_κ the number of children of each node $\kappa \in [K]$.

We start by computing $\mathbb{P}_d^{\text{dis}}(t, t')$ which is independent of the arbitrary node numbering $\kappa \in [K]$. For this computation, one goes through the nodes of t and t', computing the likelihood that each child number c_κ equals the binomial distributions from the exploration process.

Note that in steps (2) and (3) of the exploration process, one needs to condition on the fact that edges do not connect active nodes amongst each other. This corresponds to the condition that $\mathcal{N}_{G,d}(i)$ and $\mathcal{N}_{G',d}(j)$ are non-intersecting and cycle-free. It translates to the event $\{\mathrm{Bin}(A^* - 1, \lambda s s') = 0\}$ for nodes of the intersection graph and similarly for other subgraphs. Subtracting the number 1 from A^* compensates for the fact that at the beginning of each iteration, the current node k is still in the active set.

We index the exploration steps by $\kappa \in [K]$ and use that same notation for the evolving set cardinalities $A^*_\kappa, U_\kappa, D'_\kappa$, etc. By "$\kappa \in t$", we denote an iteration over all exploration indices $\kappa \in [K]$ that correspond to nodes of t. The likelihood of (t, t') to be sampled from $\mathbb{P}_d^{\mathrm{dis}}$ can then be computed as follows:

$$\begin{aligned}
\mathbb{P}_d^{\mathrm{dis}}(t,t') &= \prod_{\kappa\in t} \mathbb{P}\Big(\mathrm{Bin}\Big(U^*_\kappa + U_\kappa, \frac{\lambda r}{N}\Big) = c_\kappa\Big)\mathbb{P}\Big(\mathrm{Bin}\Big(A^*_\kappa + A_\kappa - 1, \frac{\lambda r}{N}\Big) = 0\Big) \\
&\quad \times \prod_{\kappa\in t'} \mathbb{P}\Big(\mathrm{Bin}\Big(U^*_\kappa + U'_\kappa, \frac{\lambda r'}{N}\Big) = c_\kappa\Big)\mathbb{P}\Big(\mathrm{Bin}\Big(A^*_\kappa + A'_\kappa - 1, \frac{\lambda r'}{N}\Big) = 0\Big) \\
&= \prod_{\kappa\in t} \binom{U^*_\kappa + U_\kappa}{c_\kappa}\Big(\frac{\lambda r}{N}\Big)^{c_\kappa}\Big(1-\frac{\lambda r}{N}\Big)^{U^*_\kappa + U_\kappa - c_\kappa}\Big(1-\frac{\lambda r}{N}\Big)^{A^*_\kappa + A_\kappa - 1} \\
&\quad \times \prod_{\kappa\in t'} \binom{U^*_\kappa + U'_\kappa}{c_\kappa}\Big(\frac{\lambda r'}{N}\Big)^{c_\kappa}\Big(1-\frac{\lambda r'}{N}\Big)^{U^*_\kappa + U'_\kappa - c_\kappa}\Big(1-\frac{\lambda r'}{N}\Big)^{A^*_\kappa + A'_\kappa - 1} \\
&= \prod_{\kappa\in t} \binom{U^*_\kappa + U_\kappa}{c_\kappa}\Big(\frac{\lambda r}{N}\Big)^{c_\kappa}\Big(1-\frac{\lambda r}{N}\Big)^{U^*_\kappa + U_\kappa + A^*_\kappa + A_\kappa - c_\kappa - 1} \\
&\quad \times \prod_{\kappa\in t'} \binom{U^*_\kappa + U'_\kappa}{c_\kappa}\Big(\frac{\lambda r'}{N}\Big)^{c_\kappa}\Big(1-\frac{\lambda r'}{N}\Big)^{U^*_\kappa + U'_\kappa + A^*_\kappa + A'_\kappa - c_\kappa - 1}.
\end{aligned}$$

To simplify this expression, we track the quantities $A_\kappa, D'_\kappa, U^*_\kappa, \ldots$ throughout the sampling process. The initialization leads to

$$\begin{aligned}
&A^*_1 = \mathbb{1}_{i\in V_*} + \mathbb{1}_{j\in V'_*}, \quad A_1 = \mathbb{1}_{i\notin V_*}, \quad A'_1 = \mathbb{1}_{j\notin V'_*}, \\
&D^*_1 = D_1 = D'_1 = 0.
\end{aligned}$$

These quantities evolve as follows with $\kappa \in [K]$:

$$\begin{aligned}
&A^*_\kappa = A^*_1 - D^*_\kappa + \sum_{\eta\in[\kappa]} c^*_\eta, \quad A_\kappa = A_1 - D_\kappa + \sum_{\eta\in[\kappa]} c^+_\eta, \\
&A'_\kappa = A'_1 - D'_\kappa + \sum_{\eta\in[\kappa]} c'^+_\eta,
\end{aligned}$$

$$D_\kappa^* = \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V_*}, \qquad D_\kappa = \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V\setminus V_*}, \qquad D'_\kappa = \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V'\setminus V'_*},$$
$$U_\kappa^* = n_* - A_\kappa^* - D_\kappa^*, \qquad U_\kappa = n_+ - A_\kappa - D_\kappa, \qquad U'_\kappa = n'_+ - A'_\kappa - D'_\kappa,$$

where $c_\eta = c_\eta^* + c_\eta^+$ denotes the decomposition of children among those in the intersection graph G_* and those in the added nodes $V \setminus V_*$, respectively, $V' \setminus V'_*$. Since the product expansion of $\mathbb{P}_d^{\mathrm{dis}}(t,t')$ is only a function of the sums $A_\kappa^* + A'_\kappa$, $U_\kappa^* + U_\kappa$, $U_\kappa^* + U'_\kappa$ and $A_\kappa^* + A_\kappa$, we compute these sums directly:

$$\begin{aligned}
A_\kappa^* + A_\kappa &= A_1^* + A_1 - D_\kappa^* - D_\kappa + \sum_{\eta\in[\kappa]} (c_\eta^* + c_\eta^+) \\
&= 1 + \mathbb{1}_{j\in V'_*} - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V} + \sum_{\eta\in[\kappa],\eta\in V} c_\eta, \\
A_\kappa^* + A'_\kappa &= A_1^* + A'_1 - D_\kappa^* - D'_\kappa + \sum_{\eta\in[\kappa]} (c_\eta^* + c_\eta'^+) \\
&= 1 + \mathbb{1}_{i\in V_*} - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V'} + \sum_{\eta\in[\kappa],\eta\in V'} c_\eta, \\
U_\kappa^* + U_\kappa &= n_* + n_+ - A_\kappa^* - A_\kappa - D_\kappa^* - D_\kappa \\
&= n - 1 - \mathbb{1}_{j\in V'_*} - \sum_{\eta\in[\kappa],\eta\in V} c_\eta, \\
U_\kappa^* + U'_\kappa &= n_* + n'_+ - A_\kappa^* - A'_\kappa - D_\kappa^* - D'_\kappa \\
&= n' - 1 - \mathbb{1}_{i\in V_*} - \sum_{\eta\in[\kappa],\eta\in V'} c_\eta.
\end{aligned}$$

Introducing the notations $\mathbb{I} := 1 + \mathbb{1}_{j\in V'_*}$ and $\mathbb{I}' := 1 + \mathbb{1}_{i\in V_*}$, this simplifies to

$$\begin{aligned}
U_\kappa^* + U_\kappa &= n - \mathbb{I} - \sum_{\eta\in[\kappa]\cap V} c_\eta, \\
U_\kappa^* + U'_\kappa &= n' - \mathbb{I}' - \sum_{\eta\in[\kappa]\cap V'} c_\eta, \\
U_\kappa^* + U_\kappa + A_\kappa^* + A_\kappa &= n - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V}, \\
U_\kappa^* + U'_\kappa + A_\kappa^* + A'_\kappa &= n' - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V'}.
\end{aligned}$$

Plugging these into the expression for $\mathbb{P}_d^{\mathrm{dis}}(t,t')$ yields

$$\begin{aligned}
\mathbb{P}_d^{\mathrm{dis}}(t,t') = &\prod_{\kappa\in t} \binom{n - \mathbb{I} - \sum_{\eta\in[\kappa]\cap V} c_\eta}{c_\kappa} \left(\frac{\lambda r}{N}\right)^{c_\kappa} \left(1 - \frac{\lambda r}{N}\right)^{n-1-c_\kappa - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V}} \\
&\times \prod_{\kappa\in t'} \binom{n' - \mathbb{I}' - \sum_{\eta\in[\kappa]\cap V'} c_\eta}{c_\kappa} \left(\frac{\lambda r'}{N}\right)^{c_\kappa} \left(1 - \frac{\lambda r'}{N}\right)^{n'-1-c_\kappa - \sum_{\eta\in[\kappa]} \mathbb{1}_{\eta\in V'}}
\end{aligned}$$

$$= \prod_{\kappa \in t} \frac{(n - \mathbb{I} - \sum_{\eta \in [\kappa] \cap V} c_\eta)!}{c_\kappa!(n - \mathbb{I} - \sum_{\eta \in [\kappa-1] \cap V} c_\eta)!} \left(\frac{\lambda s}{Nq}\right)^{c_\kappa} \left(1 - \frac{\lambda s}{Nq}\right)^{n-1-c_\kappa-|[\kappa]\cap V|}$$

$$\times \prod_{\kappa \in t'} \frac{(n' - \mathbb{I}' - \sum_{\eta \in [\kappa] \cap V'} c_\eta)!}{c_\kappa!(n' - \mathbb{I}' - \sum_{\eta \in [\kappa-1] \cap V'} c_\eta)!} \left(\frac{\lambda s'}{Nq'}\right)^{c_\kappa} \left(1 - \frac{\lambda s'}{Nq'}\right)^{n'-1-c_\kappa-|[\kappa]\cap V'|}.$$

Recall that $\mathcal{B} \subset \{n - \varepsilon \le Nq \le n + \varepsilon\} \cap \{n' - \varepsilon \le Nq' \le n' + \varepsilon\}$, where

$$\varepsilon = N^{1/2+\gamma} \quad \text{and} \quad \gamma = \frac{1}{8} - \frac{1}{4} c \log\big(\lambda \max\{s, s'\}\big) > 0.$$

Conditioning on $\mathcal{B}$, one can upper bound $\mathbb{P}_d^{\mathrm{dis}}(t, t')$ as follows:

$$\mathbb{P}_d^{\mathrm{dis}}(t,t')\mathbb{1}_{\mathcal{B}} \le \prod_{\kappa \in t} \frac{(n - \mathbb{I} - \sum_{\eta \in [\kappa] \cap V} c_\eta)!}{c_\kappa!(n - \mathbb{I} - \sum_{\eta \in [\kappa-1] \cap V} c_\eta)!} \frac{(\lambda s)^{c_\kappa}}{(n-\varepsilon)^{c_\kappa}} \left(1 - \frac{\lambda s}{n+\varepsilon}\right)^{n-c_\kappa-|[\kappa]\cap V|}$$

$$\times \prod_{\kappa \in t'} \frac{(n' - \mathbb{I}' - \sum_{\eta \in [\kappa] \cap V'} c_\eta)!}{c_\kappa!(n' - \mathbb{I}' - \sum_{\eta \in [\kappa-1] \cap V'} c_\eta)!} \frac{(\lambda s')^{c_\kappa}}{(n'-\varepsilon)^{c_\kappa}} \left(1 - \frac{\lambda s'}{n'+\varepsilon}\right)^{n'-c_\kappa-|[\kappa]\cap V'|}.$$

In the fraction $(n - \mathbb{I} - \sum_{\eta \in [\kappa] \cap V} c_\eta)!/(n - \mathbb{I} - \sum_{\eta \in [\kappa-1] \cap V} c_\eta)!$, the numerator is equal to the denominator plus c_κ. Hence, using that the numerator is a product of factors bounded by n, one has

$$\frac{(n - \mathbb{I} - \sum_{\eta \in [\kappa] \cap V} c_\eta)!}{(n - \mathbb{I} - \sum_{\eta \in [\kappa-1] \cap V} c_\eta)!} \times \frac{1}{n^{c_\kappa}} \le 1,$$

which still holds for sufficiently large n if one replaces $1/n^{c_\kappa}$ by $1/(n-\varepsilon)^{c_\kappa}$, because $\varepsilon \in o(N)$. Consequently, there exists $N_0 \in \mathbb{N}$ such that for all $N \ge N_0$,

$$\mathbb{P}_d^{\mathrm{dis}}(t,t')\mathbb{1}_{\mathcal{B}} \le \prod_{\kappa \in t} \frac{1}{c_\kappa!} (\lambda s)^{c_\kappa} \left(1 - \frac{\lambda s}{n+\varepsilon}\right)^{n-c_\kappa-|[\kappa]\cap V|} \times \prod_{\kappa \in t'} \frac{1}{c_\kappa!} (\lambda s')^{c_\kappa} \left(1 - \frac{\lambda s'}{n'+\varepsilon}\right)^{n'-c_\kappa-|[\kappa]\cap V'|}. \tag{A.1}$$

For the rest of this proof, we assume $N \ge N_0$ in order for (A.1) to hold.

Computation of R_d, the likelihood ratio of $\mathbb{P}_d^{\mathrm{dis}}$ and $\mathbb{P}_d^{\mathrm{ind}}$. It is quite simple to compute $\mathbb{P}_d^{\mathrm{ind}}(t,t')$ because of the independence in Galton–Watson trees:

$$\mathbb{P}_d^{\mathrm{ind}}(t,t') = \prod_{\kappa \in t} \mathbb{P}\big(\mathrm{Poi}(\lambda s) = c_\kappa\big) \prod_{\kappa \in t'} \mathbb{P}\big(\mathrm{Poi}(\lambda s') = c_\kappa\big)$$

$$= \prod_{\kappa \in t} e^{-\lambda s} (\lambda s)^{c_\kappa} \frac{1}{c_\kappa!} \prod_{\kappa \in t'} e^{-\lambda s'} (\lambda s')^{c_\kappa} \frac{1}{c_\kappa!}.$$

Comparing this to the upper bound on $\mathbb{P}_d^{\mathrm{dis}}(t,t')$ from (A.1), one obtains

$$\begin{aligned} R_d(t,t')\mathbb{1}_{\mathcal{B}} &= \left(\frac{\mathbb{P}_d^{\mathrm{dis}}(t,t')}{\mathbb{P}_d^{\mathrm{ind}}(t,t')}\right)\mathbb{1}_{\mathcal{B}} \\ &\le \prod_{\kappa\in t} e^{\lambda s}\Big(1-\frac{\lambda s}{n+\varepsilon}\Big)^{n-c_\kappa-|[\kappa]\cap V|} \\ &\quad \times \prod_{\kappa\in t'} e^{\lambda s'}\Big(1-\frac{\lambda s'}{n'+\varepsilon}\Big)^{n'-c_\kappa-|[\kappa]\cap V'|}\mathbb{1}_{\mathcal{B}}. \end{aligned} \tag{A.2}$$

Moving forward, we condition on the event $\mathcal{A}^c$, which we recall to be defined as

$$\begin{aligned} \mathcal{A}^c &= \big\{\forall i\in V, \rho\in[d] : |\mathcal{S}_{G,\rho}(i)| \le C\log(n)(\lambda s)^\rho\big\} \\ &\qquad \cup \big\{\forall j\in V', \rho'\in[d] : |\mathcal{S}_{G',\rho'}(j)| \le C\log(n')(\lambda s')^\rho\big\} \\ &\subset \left\{|t|\le C\log(n)\sum_{\rho=0}^d(\lambda s)^\rho\right\} \cup \left\{|t'|\le C\log(n')\sum_{\rho'=0}^d(\lambda s')^{\rho'}\right\} \\ &\le \big\{|t|\le \widetilde{C}\log(n)(\lambda s)^d\big\}\cup\big\{|t'|\le\widetilde{C}\log(n')(\lambda s')^d\big\}, \end{aligned}$$

where we have introduced the constant

$$\widetilde{C} := C\,\frac{\lambda\max\{s,s'\}}{\lambda\min\{s,s'\}-1} > 0.$$

Under $\mathcal{A}^c$, we have

$$\sum_{\kappa\in t}|[\kappa]\cap V| = 1+2+\cdots+|t| = \frac{|t|(|t|+1)}{2} \le |t|^2 \le \widetilde{C}^2\log(n)^2(\lambda s)^{2d}.$$

Plugging the bounds on $|t|$ and $\sum_\kappa |[\kappa]\cap V|$ into the first product "$\prod_{\kappa\in t}$" into (A.2), we obtain

$$\begin{aligned} &\prod_{\kappa\in t} e^{\lambda s}\Big(1-\frac{\lambda s}{n+\varepsilon}\Big)^{n-c_\kappa-|[\kappa]\cap V|}\mathbb{1}_{\mathcal{A}^c} \\ &\quad = \exp\left[\lambda s|t| + \Big(n|t| - \sum_{\kappa\in t}c_\kappa - \sum_{\kappa\in t}|[\kappa]\cap V|\Big)\log\Big(1-\frac{\lambda s}{n+\varepsilon}\Big)\right]\mathbb{1}_{\mathcal{A}^c} \\ &\quad \overset{(\star)}{\le} \exp\left[\lambda s|t| - \frac{\lambda s}{n+\varepsilon}\Big(n|t| - \sum_{\kappa\in t}c_\kappa - \sum_{\kappa\in t}|[\kappa]\cap V|\Big)\right]\mathbb{1}_{\mathcal{A}^c} \\ &\quad = \exp\left[\Big(1-\frac{n}{n+\varepsilon}\Big)\lambda s|t| + \frac{1}{n+\varepsilon}\lambda s\frac{|t|(|t|+1)}{2}\right]\exp\left[\frac{\lambda s}{n+\varepsilon}\sum_{\kappa\in t}c_\kappa\right]\mathbb{1}_{\mathcal{A}^c} \\ &\quad \le \exp\left[\frac{\varepsilon}{n+\varepsilon}\widetilde{C}\log(n)(\lambda s)^{d+1} + \frac{1}{n+\varepsilon}\widetilde{C}^2\log(n)^2(\lambda s)^{2d+1}\right] \\ &\qquad \times\exp\left[\frac{\lambda s}{n+\varepsilon}\sum_{\kappa\in t}c_\kappa\right], \end{aligned} \tag{A.3}$$

where the inequality $(\star)$ uses $\log(1-x) \le -x$ for $x < 1$. Recalling that $Nq \le n+\varepsilon$ and $\varepsilon = N^{1/2+\gamma}$, we have

$$\frac{\varepsilon}{n+\varepsilon} \le \frac{N^{1/2+\gamma}}{Nq} \quad \text{and} \quad \frac{1}{n+\varepsilon} \le \frac{1}{Nq}.$$

We further bound $\log(n) \le \log(N)$ and use the assumption $d \le c\log(N)$ from Theorem 3.10, which leads to

$$(\lambda s)^d \le (\lambda s)^{c\log(N)} = N^{c\log(\lambda s)} \le N^{c\log(\lambda\max\{s,s'\})} \le N^{1/2-2\gamma}, \tag{A.4}$$

where the last inequality follows from $c\log(\lambda\max\{s,s'\}) \le 1/4 < 1/2$ and

$$\frac{1}{2} - 2\gamma \ge \frac{1}{2} - \frac{1}{4} + \frac{1}{2}c\log\bigl(\lambda\max\{s,s'\}\bigr) > c\log\bigl(\lambda\max\{s,s'\}\bigr).$$

Going back to (A.3), we can thus bound the first term of the last line as follows:

$$\begin{aligned}
&\exp\Bigl[\frac{\varepsilon}{n+\varepsilon}\widetilde{C}\log(n)(\lambda s)^{d+1} + \frac{1}{n+\varepsilon}\widetilde{C}^2\log(n)^2(\lambda s)^{2d+1}\Bigr] \\
&\quad\le \exp\Bigl[\frac{N^{1/2+\gamma}}{Nq}\widetilde{C}\log(N)\lambda s N^{1/2-2\gamma} + \frac{1}{Nq}\widetilde{C}^2\log(N)^2\lambda s N^{1-4\gamma}\Bigr] \\
&\quad= \exp\Bigl[\frac{\widetilde{C}\lambda s}{q}\log(N)N^{-\gamma} + \frac{\widetilde{C}^2\lambda s}{q}\log(N)^2N^{-4\gamma}\Bigr].
\end{aligned}$$

Since $\gamma > 0$, the last line is in $\mathcal{O}(1)$ and we note that its value is deterministic.

The same argument that we started in (A.3) can be used verbatim by replacing s, n, q, t with s', n', q', t' in order to show

$$\begin{aligned}
&\prod_{\kappa\in t'} e^{\lambda s'}\Bigl(1 - \frac{\lambda s'}{n'+\varepsilon}\Bigr)^{n'-c_\kappa-|[\kappa]\cap V'|}\mathbb{1}_{\mathcal{B}\cap\mathcal{A}^c} \\
&\quad\le \exp\Bigl[\frac{\varepsilon}{n'+\varepsilon}\widetilde{C}\log(n')(\lambda s')^{d+1} + \frac{1}{n'+\varepsilon}\widetilde{C}^2\log(n')^2(\lambda s')^{2d+1}\Bigr] \\
&\qquad\times\exp\Bigl[\frac{\lambda s'}{n'+\varepsilon}\sum_{\kappa\in t'} c_\kappa\Bigr] \\
&\quad\le \mathcal{O}(1)\times\exp\Bigl[\frac{\lambda s'}{n'+\varepsilon}\sum_{\kappa\in t'} c_\kappa\Bigr].
\end{aligned}$$

Using the thus obtained bounds on both products "$\prod_{\kappa\in t}$" and "$\prod_{\kappa\in t'}$" in (A.2) yields

$$R_d(t,t')\mathbb{1}_{\mathcal{B}\cap\mathcal{A}^c} \le \mathcal{O}(1)\times\exp\Bigl[\frac{\lambda s}{n+\varepsilon}\sum_{\kappa\in t} c_\kappa\Bigr]\exp\Bigl[\frac{\lambda s'}{n'+\varepsilon}\sum_{\kappa\in t'} c_\kappa\Bigr].$$

Squaring this inequality and using $1/(n+\varepsilon) \le 1/(Nq)$ on $\mathcal{B}$, we get

$$R_d^2(t,t')\mathbb{1}_{\mathcal{B}\cap\mathcal{A}^c} \le \mathcal{O}(1)\times\exp\Bigl[\frac{2\lambda s}{Nq}\sum_{\kappa\in t} c_\kappa\Bigr]\exp\Bigl[\frac{2\lambda s'}{Nq'}\sum_{\kappa\in t'} c_\kappa\Bigr]. \tag{A.5}$$

Computing the expectation. To conclude the proof of Lemma 3.11, one needs to compute the expectation of R_d^2 under $\mathbb{P}_d^{\mathrm{ind}}$. The assumption that $(t,t') \sim \mathbb{P}_d^{\mathrm{ind}}$ is equivalent to $c_\kappa \sim \mathrm{Poi}(\lambda s)$ for $\kappa \in t$ and $c_\kappa \sim \mathrm{Poi}(\lambda s')$ for $\kappa \in t'$ with all c_κ being independent. Hence, taking expectations in (A.5) yields

$$\mathbb{E}_d^{\mathrm{ind}}\big[R_d^2 \mathbb{1}_{\mathcal{B}\cap\mathcal{A}^c}\big] \le \mathcal{O}(1) \times \prod_{\kappa\in t} \mathbb{E}_d^{\mathrm{ind}}\big[e^{(2\lambda s)/(Nq)c_\kappa}\big] \times \prod_{\kappa\in t'} \mathbb{E}_d^{\mathrm{ind}}\big[e^{(2\lambda s')/(Nq')c_\kappa}\big]\mathbb{1}_{\mathcal{A}^c}. \tag{A.6}$$

Since the generating function of c_κ equals $\mu \mapsto \mathbb{E}[e^{\mu c_\kappa}] = \exp(\lambda s(e^\mu - 1))$, we obtain

$$\begin{aligned}\prod_{\kappa\in t} \mathbb{E}_d^{\mathrm{ind}}\big[e^{(2\lambda s)/(Nq)}c_\kappa\big]\mathbb{1}_{\mathcal{A}^c} &= \exp\big(\lambda s\big(e^{(2\lambda s)/(Nq)} - 1\big)\big)^{|t|}\mathbb{1}_{\mathcal{A}^c} \\ &\le \exp\big(\lambda s\big(e^{(2\lambda s)/(Nq)} - 1\big)\widetilde{C}\log(n)(\lambda s)^d\big).\end{aligned}$$

A first order approximation yields $e^{(2\lambda s)/(Nq)} - 1 \in \mathcal{O}(N^{-1})$ and we recall from (A.4) that $(\lambda s)^d \le N^{1/2-2\gamma}$. Hence,

$$\big(e^{(2\lambda s)/(Nq)} - 1\big)\log(n)(\lambda s)^d \le \mathcal{O}\big(N^{-1}\log(N)N^{1/2-2\gamma}\big) \subset o(1),$$

which leads to

$$\exp\big(\lambda s\big(e^{(2\lambda s)/(Nq)} - 1\big)\widetilde{C}\log(n)(\lambda s)^d\big) \in \mathcal{O}(1).$$

Plugging this result into (A.6), we conclude that $d \mapsto \mathbb{E}_d^{\mathrm{ind}}[R_d^2 \mathbb{1}_{\mathcal{A}^c}\mathbb{1}_{\mathcal{B}}]$ is bounded. This finishes the proof. ∎

A.2. Proof of Proposition 4.3

We recall the recursive likelihood ratio property before moving to its proof.

Proposition A.2 (Copy of Proposition 4.3). *For $t, t' \in \mathcal{X}_d$, let c, c' be the degrees of their respective root nodes. Denote by $t_{[1]}, \dots, t_{[c]}$, respectively, $t'_{[1]}, \dots, t'_{[c']}$ the subtrees attached to the roots, labeled randomly. Recall that $\sigma: [k] \hookrightarrow [c]$ denotes an injective map from $[k]$ to $[c]$ and let $\sum_{\sigma:[k]\hookrightarrow[c]}$ be the sum over all such mappings. Then, for all $d \in \mathbb{N}_{>1}$, the likelihood ratio at depth d can be expressed recursively as*

$$L_d(t,t') = \sum_{k=0}^{c\wedge c'} \psi(k,c,c') \sum_{\substack{\sigma: [k]\hookrightarrow[c],\\ \sigma': [k]\hookrightarrow[c']}} \prod_{i=1}^{k} L_{d-1}\big(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big),$$

where

$$\psi(k,c,c') = \exp(\lambda s s')\frac{1}{\lambda^k k!}(1-s')^{c-k}(1-s)^{c'-k}.$$

Proof. For a rooted tree t, denote its root node by ρ_t. The values $c = \deg(\rho_t)$ and $c' = \deg(\rho_{t'})$ represent the number of root children in t and t', respectively. Since t and t' are deterministic, both c and c' are fixed integers.

In the following, (τ, τ') denotes a pair of random trees following the law $\mathbb{P}_d^{\mathrm{corr}}$. Hence, they can be seen as independent augmentations of an intersection tree denoted as τ_* as explained in Definition 3.1. Note that $(\tau, \tau') = (t, t')$ implies the event

$$D := \big\{\deg(\rho_\tau) = c, \deg(\rho'_\tau) = c'\big\},$$

which lets us compute

$$\begin{aligned}
\mathbb{P}_d^{\mathrm{corr}}(t,t') &\overset{\text{(a)}}{=} \sum_{k=0}^{c\wedge c'} \mathbb{P}_d^{\mathrm{corr}}\big((\tau,\tau') = (t,t'), D, \deg(\rho_{\tau_*}) = k\big) \\
&= \sum_{k=0}^{c\wedge c'} \mathbb{P}_d^{\mathrm{corr}}\big(D, \deg(\rho_{\tau_*}) = k\big)\, \underbrace{\mathbb{P}_d^{\mathrm{corr}}\big((\tau,\tau') = (t,t') \mid D, \deg(\rho_{\tau_*}) = k\big)}_{=:\boldsymbol{\xi}} \\
&\overset{\text{(b)}}{=} \sum_{k=0}^{c\wedge c'} \pi_{\lambda s s'}(k)\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)\boldsymbol{\xi}.
\end{aligned}$$

The equality (a) uses $\deg(\rho_{\tau_*}) \le \deg(\rho_\tau) \wedge \deg(\rho_{\tau'})$, while (b) relies on the construction of (τ, τ') as augmentations of τ_*. From the construction of these augmentations, one has

$$\deg(\rho_{\tau_*}) \sim \mathrm{Poi}(\lambda s s'), \deg(\rho_\tau) - \deg(\rho_{\tau_*}) \sim \mathrm{Poi}\big(\lambda s(1-s')\big),$$

and

$$\deg(\rho_{\tau'}) - \deg(\rho_{\tau_*}) \sim \mathrm{Poi}\big(\lambda s'(1-s)\big)$$

with all three variables being independent.

To compute $\boldsymbol{\xi}$, randomly assign the numbers $\{1, \ldots, c\}$ to the child nodes of ρ_t and denote by $t_{[i]} := \mathcal{N}_{t,d}(\rho_t \to i)$ the (unlabeled) subtree rooted in $i \in [c]$, using the notation from Section 3.3. Define $t'_{[1]}, \ldots, t'_{[c']}$ analogously and enumerate the child nodes of τ, τ' in a way that $(\tau_{[1]}, \tau'_{[1]}), \ldots, (\tau_{[k]}, \tau'_{[k]})$ correspond to the root children in τ_* with $k = \deg(\rho_{\tau_*})$. With this notation, the event $(\tau, \tau') = (t, t')$ implies that permutations σ, σ' are correctly mapping the random numberings to the corresponding nodes. Denoting by $\mathfrak{S}_c$ the symmetric group of size c, this leads to

$$\begin{aligned}
\boldsymbol{\xi} = &\sum_{\substack{\sigma\in\mathfrak{S}_c \\ \sigma'\in\mathfrak{S}_{c'}}} \frac{1}{c!c'!} \\
&\times \mathbb{P}_d^{\mathrm{corr}}\left((\tau,\tau') = (t,t') \mid D, \deg(\rho_{\tau_*}) = k, \begin{cases} \tau_i = t_{[\sigma(i)]} & \text{for } i \in [c], \\ \tau'_j = t'_{[\sigma'(j)]} & \text{for } j \in [c'] \end{cases}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\sigma,\sigma'} \frac{1}{c!c'!} \mathbb{P}^{\mathrm{corr}}_{d-1}\big(\big(\tau_{[i]}, \tau'_{[i]}\big) = \big(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big) \text{ for } i \in [k]\big) \\
&\quad \times \mathbb{P}\big(t_{[\sigma(i)]} = \mathrm{GW}^{(\lambda s)}_{d-1} \text{ for } i \in [k+1:c]\big) \\
&\quad \times \mathbb{P}\big(t'_{[\sigma'(i)]} = \mathrm{GW}^{(\lambda s')}_{d-1} \text{ for } i \in [k+1:c']\big) \\
&= \sum_{\sigma,\sigma'} \frac{1}{c!c'!} \bigg(\prod_{i=1}^{k} \mathbb{P}^{\mathrm{corr}}_{d-1}\big(t'_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big)\bigg) \\
&\quad \times \bigg(\prod_{i=k+1}^{c} \mathrm{GW}^{(\lambda s)}_{d-1}\big(t_{[\sigma(i)]}\big)\bigg)\bigg(\prod_{i=k+1}^{c'} \mathrm{GW}^{(\lambda s')}_{d-1}\big(t'_{[\sigma'(i)]}\big)\bigg),
\end{aligned}
$$

where we recall that $\mathrm{GW}^{(\mu)}_{d-1}(t)$ denotes the likelihood of t being a GW-tree of depth $d-1$ and offspring distribution $\mathrm{Poi}(\mu)$.

Under the independence hypothesis, the computation is more immediate:

$$
\begin{aligned}
\mathbb{P}^{\mathrm{ind}}_{d}(t,t') &= \mathrm{GW}^{(\lambda s)}_{d}(t)\ \mathrm{GW}^{(\lambda s')}_{d}(t') \\
&= \pi_{\lambda s}(c)\bigg(\prod_{i=1}^{c} \mathrm{GW}^{(\lambda s)}_{d-1}\big(t_{[i]}\big)\bigg)\pi_{\lambda s'}(c')\bigg(\prod_{i=1}^{c'} \mathrm{GW}^{(\lambda s')}_{d-1}\big(t'_{[i]}\big)\bigg).
\end{aligned}
$$

Putting both results together, we obtain

$$
\begin{aligned}
L_d(t,t') &= \sum_{k=0}^{c\wedge c'} \frac{\pi_{\lambda s s'}(k)\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)}{\pi_{\lambda s}(c)\pi_{\lambda s'}(c')c!c'!} \\
&\quad \times \sum_{\substack{\sigma\in\mathfrak{S}_c,\\ \sigma'\in\mathfrak{S}_{c'}}} \bigg(\prod_{i=1}^{k} \mathbb{P}^{\mathrm{corr}}_{d-1}\big(t'_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big)\bigg) \\
&\quad \times \bigg(\prod_{i=1}^{k} \mathrm{GW}^{(\lambda s)}_{d-1}\big(t_{[\sigma(i)]}\big)\bigg)^{-1} \bigg(\prod_{i=1}^{k} \mathrm{GW}^{(\lambda s')}_{d-1}\big(t'_{[\sigma'(i)]}\big)\bigg)^{-1} \\
&= \sum_{k=0}^{c\wedge c'} \frac{\pi_{\lambda s s'}(k)\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)}{\pi_{\lambda s}(c)\pi_{\lambda s'}(c')c!c'!} \\
&\quad \times \sum_{\substack{\sigma\in\mathfrak{S}_c,\\ \sigma'\in\mathfrak{S}_{c'}}} \prod_{i=1}^{k} L_{d-1}\big(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big) \\
&= \sum_{k=0}^{c\wedge c'} \frac{\psi(k,c,c')}{(c-k)!(c'-k)!} \sum_{\substack{\sigma\in\mathfrak{S}_c,\\ \sigma'\in\mathfrak{S}_{c'}}} \prod_{i=1}^{k} L_{d-1}\big(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\big), \qquad \text{(A.7)}
\end{aligned}
$$

where the term ψ, which is implicitly defined in the last equation, can be simplified to

$$\begin{aligned}\psi(k,c,c') &\overset{\text{def.}}{=} \frac{\pi_{\lambda ss'}(k)\pi_{\lambda s(1-s')}(c-k)\pi_{\lambda s'(1-s)}(c'-k)}{\pi_{\lambda s}(c)\pi_{\lambda s'}(c')}\,\frac{(c-k)!(c'-k)!}{c!\,c'!}\\ &= \exp\bigl(-\lambda ss' - \lambda s(1-s') - \lambda s'(1-s) + \lambda s + \lambda s'\bigr)\\ &\quad\times \frac{1}{k!}\lambda^{-k}(1-s')^{c-k}(1-s)^{c'-k}\\ &= \exp(\lambda ss')\frac{1}{\lambda^k k!}(1-s')^{c-k}(1-s)^{c'-k}.\end{aligned}$$

One can obtain a more compact version the equation (A.7) by realizing that

$$|\mathfrak{S}_c| = (c-k)! \times |\{\sigma\colon [k] \to [c] : \sigma \text{ injective}\}|.$$

Denoting by $\sum_{\sigma:[k]\hookrightarrow[c]}$ the sum over all injective mappings from $[k]$ to $[c]$, we therefore obtain the desired result:

$$L_d(t,t') = \sum_{k=0}^{c\wedge c'} \psi(k,c,c') \sum_{\substack{\sigma\colon [k]\hookrightarrow[c],\\ \sigma'\colon [k]\hookrightarrow[c']}} \prod_{i=1}^{k} L_{d-1}\bigl(t_{[\sigma(i)]}, t'_{[\sigma'(i)]}\bigr). \qquad \blacksquare$$

A.3. Proof of Proposition 4.5

We recall the lower bound on L_{d+k} before moving to its proof.

Proposition A.3 (Copy of Proposition 4.5). *Let (t,t') be a pair of rooted trees sharing a common rooted subtree t_*. Denote by $\sigma_*\colon t_* \hookrightarrow t$ and $\sigma'_*\colon t_* \hookrightarrow t'$ the embeddings of t_* into t and t', respectively.*

For each $i \in t$, let $c_t(i)$ denote its number of child nodes in t, and define $t_{[i]}$ as the subtree rooted at i, containing all its descendants. This notation extends naturally to t' and t_*, and we recall that $(t_*)_d$ denotes the depth-d truncation of t_*. Then, for all $d,k \in \mathbb{N}_{>0}$, the following inequality holds:

$$\begin{aligned}L_{d+k}(t,t') \geq &\prod_{i\in(t_*)_{d-1}} \psi\bigl(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\bigr)\\ &\times \prod_{j\in(t_*)_d\setminus(t_*)_{d-1}} L_k\bigl(t_{[\sigma_*(j)]}, t'_{[\sigma'_*(j)]}\bigr).\end{aligned}$$

Proof. We start by iteratively applying the recursive likelihood formula from Proposition 4.3 to obtain

$$L_d(t,t') = \sum_{\tau\in\mathcal{X}_d} \sum_{\substack{\sigma\colon \tau\hookrightarrow t,\\ \sigma'\colon \tau\hookrightarrow t'}} \prod_{i\in\tau_{d-1}} \psi\bigl(c_\tau(i), c_t(\sigma(i)), c_{t'}(\sigma'(i))\bigr). \tag{A.8}$$

For a rigorous proof of this equation, we refer the reader to [27, Lemma 2.1] whose proof translates word by word to the asymmetric tree correlation setting.

Next, let $t, t' \in \mathcal{X}_{d+k}$ be arbitrary trees that share a common subtree $t_* \in \mathcal{X}_d$ up to depth d. Denote by $\sigma_*\colon t_* \hookrightarrow t$ and $\sigma'_*\colon t_* \hookrightarrow t'$ the respective subtree embeddings. One has the following sequence of computations; each step will be explained right after:

$$
\begin{aligned}
L_{d+k}(t,t') &= \sum_{\tau \in \mathcal{X}_{d+k}} \sum_{\substack{\sigma\colon \tau \hookrightarrow t,\\ \sigma'\colon \tau\hookrightarrow t'}} \prod_{i \in \tau_{d+k-1}} \psi\big(c_\tau(i), c_t(\sigma(i)), c_{t'}(\sigma'(i))\big) \\
&\geq \sum_{\substack{\tau \in \mathcal{X}_{d+k}\\ \tau_d = t_*}} \sum_{\substack{\sigma\colon \tau\hookrightarrow t,\, \sigma = \sigma_* \text{ on } \tau_d\\ \sigma'\colon \tau \hookrightarrow t',\, \sigma' = \sigma'_* \text{ on } \tau_d}} \prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big) \\
&\qquad \times \prod_{j \in (t_*)_d \setminus (t_*)_{d-1}} \prod_{m \in (\tau_{[j]})_{k-1}} \psi\big(c_\tau(m), c_t(\sigma(m)), c_{t'}(\sigma'(m))\big) \\
&= \Bigg(\prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big)\Bigg) \sum_{(\tilde{\tau}_\ell)_\ell \in (\mathcal{X}_k)^{(t_*)_d \setminus (t_*)_{d-1}}} \\
&\quad \sum_{\substack{(\tilde{\sigma}_\ell)_\ell, (\tilde{\sigma}'_\ell)_\ell \colon \tilde{\sigma}_\ell \colon \tilde{\tau}_\ell \hookrightarrow t_{[\sigma_*(\ell)]}\\ \tilde{\sigma}'_\ell\colon \tilde{\tau}_\ell \hookrightarrow t'_{[\sigma'_*(\ell)]}}} \prod_{j \in (t_*)_d \setminus (t_*)_{d-1}} \prod_{m \in (\tilde{\tau}_j)_{k-1}} \psi\big(c_{\tilde{\tau}_j}(m), c_t(\tilde{\sigma}_j(m)), c_{t'}(\tilde{\sigma}'_j(m))\big) \\
&= \Bigg(\prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big)\Bigg) \\
&\qquad \times \prod_{j \in (t_*)_d \setminus (t_*)_{d-1}} \sum_{\tilde{\tau} \in \mathcal{X}_k} \sum_{\substack{\tilde{\sigma}\colon \tilde{\tau} \hookrightarrow t_{[\sigma_*(j)]}\\ \tilde{\sigma}'\colon \tilde{\tau} \hookrightarrow t'_{[\sigma'_*(j)]}}} \prod_{m \in \tilde{\tau}_{k-1}} \psi\big(c_{\tilde{\tau}}(m), c_t(\tilde{\sigma}(m)), c_{t'}(\tilde{\sigma}'(m))\big) \\
&= \prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big) \prod_{j \in (t_*)_d \setminus (t_*)_{d-1}} L_k\big(t_{[\sigma_*(j)]}, t'_{[\sigma'_*(j)]}\big).
\end{aligned}
$$

Here are the explanations for every step of this computation:

- The first equality is an application of (A.8) at depth $d+k$.
- The inequality follows from $\psi \geq 0$ and the omission of some terms in the two sums.
- For the next equality, note that the sum over $\{\tau \in \mathcal{X}_{d+k} \mid \tau_d = t_*\}$ is equivalent to summing over all possible combinations of subtrees $(\tilde{\tau}_\ell)_\ell$ which are elements of $\mathcal{X}_k$ and rooted in their index nodes $\ell \in (t_*)_d \setminus (t_*)_{d-1}$. The injective mappings decompose across these subtrees.
- Next up is a swap of product and sum that simplifies the multinomial expression from before.

- Finally, the last equality is another application of (A.8) to the second product.

The final expression is what we wanted to show. ∎

A.4. Proof of Theorem 4.7

We recall the likelihood ratio diagonalization formula before moving to its proof.

Theorem A.4 (Copy of Theorem 4.7). *There exists a set of functions $f^{(\mu)}_{d,\beta} \colon \mathcal{X}_d \to \mathbb{R}$ with parameters $\mu > 0, d \in \mathbb{N}$ and indexed by trees $\beta \in \mathcal{X}_d$ such that for all choices of $s, s' \in [0,1]$, $\lambda > 0$, the following* likelihood ratio diagonalization *formula holds:*

$$\forall t, t' \in \mathcal{X}_d : \quad L_d(t,t') = \sum_{\beta \in \mathcal{X}_d} \sqrt{ss'}^{|\beta|-1} f^{(\lambda s)}_{d,\beta}(t) f^{(\lambda s')}_{d,\beta}(t'). \tag{A.9}$$

Furthermore, the $f^{(\mu)}_{d,\beta}$ have the following properties:

- *Constant value for $\beta = \bullet$, the trivial tree:*

$$\forall \mu > 0,\ \forall t \in \mathcal{X}_d : \quad f^{(\mu)}_{d,\bullet}(t) = 1. \tag{A.10}$$

- First orthogonality property *(with respect to the Galton–Watson-measure):*

$$\forall \mu > 0,\ \forall \beta, \beta' \in \mathcal{X}_d : \quad \sum_{t \in \mathcal{X}_d} \mathrm{GW}^{(\mu)}_d(t)\, f^{(\mu)}_{d,\beta}(t)\, f^{(\mu)}_{d,\beta'}(t) \;=\; \mathbb{1}_{\beta=\beta'}. \tag{A.11}$$

- Second orthogonality property *(summing over β):*

$$\forall \mu > 0,\ \forall t, t' \in \mathcal{X}_d : \quad \sum_{\beta \in \mathcal{X}_d} f^{(\mu)}_{d,\beta}(t)\, f^{(\mu)}_{d,\beta}(t') \;=\; \frac{\mathbb{1}_{t=t'}}{\mathrm{GW}^{(\mu)}_d(t)}. \tag{A.12}$$

Proof. We prove the theorem by induction on the depth d.

$d = 0$. Since $\mathcal{X}_0 = \{\bullet\}$ only contains the trivial tree, we have

$$\mathbb{P}^{\mathrm{corr}}_0(\bullet,\bullet) = 1, \quad \mathbb{P}^{\mathrm{ind}}_0(\bullet,\bullet) = 1,$$

and therefore $L_0(\bullet,\bullet) = 1$. Setting $f^{(\mu)}_{0,\bullet}(\bullet) := 1$, which naturally yields the property of constant value for the trivial tree, we have

$$\sum_{\beta \in \mathcal{X}_0} \sqrt{ss'}^{|\beta|-1} f^{(\lambda s)}_{0,\beta}(\bullet) f^{(\lambda s')}_{0,\beta}(\bullet) = \sqrt{ss'}^{\,0} f^{(\lambda s)}_{0,\bullet}(\bullet) f^{(\lambda s')}_{d,\bullet}(\bullet) = 1 = L_0(\bullet,\bullet).$$

This proves the likelihood ratio diagonalization formula (A.9) for $d = 0$. Further, note that the first orthogonality property (A.11) is trivial since $\mathcal{X}_0$ only contains one element.

$d \mapsto d+1$. Suppose that the likelihood ratio diagonalization formula (A.9) holds true for $d \in \mathbb{N}_{\geq 0}$. Fixing a pair of trees $(t,t') \in \mathcal{X}_{d+1}$, our goal is to prove that same equation at $d+1$, which can be rewritten as

$$\mathbb{P}^{\text{corr}}_{d+1}(t,t') = \underbrace{\mathrm{GW}^{(\lambda s)}_{d+1}(t)\,\mathrm{GW}^{(\lambda s')}_{d+1}(t')}_{=\mathbb{P}^{\text{ind}}_{d+1}(t,t')} \times \underbrace{\sum_{\beta \in \mathcal{X}_{d+1}} \sqrt{ss'}^{|\beta|-1} f^{(\lambda s)}_{d+1,\beta}(t)\, f^{(\lambda s')}_{d+1,\beta}(t')}_{=L_{d+1}(t,t')}, \tag{A.13}$$

where $f^{\mu}_{d+1,\beta}$ will be defined in the process.

In the following, we identify the fixed tree pair (t,t') with their subtree tuples as in Definition 4.1. This encoding shall be denoted as $t = (N_\tau)_{\tau \in \mathcal{X}_d}$ and $t' = (N'_\tau)_{\tau \in \mathcal{X}_d}$. Since t and t' are finite trees, the maps $\tau \mapsto N_\tau$ and $\tau \mapsto N'_\tau$ are non-zero only for finitely many different subtrees τ. Let $p, p' \in \mathbb{N}$ denote the respective numbers of non-zero coordinates.

Our strategy for showing (A.13) is to Fourier-invert the characteristic function of $\mathbb{P}^{\text{corr}}_{d+1}$. In the following, we denote by $(r,r') \sim \mathbb{P}^{\text{corr}}_{d+1}$ a random pair of correlated trees, which we also identify with their subtrees tuples $(r,r') = (M_\beta, M'_\beta)_{\beta \in \mathcal{X}_d}$. Note that t, t', N_τ, N'_τ are fixed while r, r', M_τ, M'_τ denote random variables. We define the notation $u \cdot r := \sum_{\tau \in \mathcal{X}_d} u_\tau M_\beta$ for any sequence $u \in \mathbb{R}^{\mathcal{X}_d}$. The characteristic function of $\mathbb{P}^{\text{corr}}_{d+1}$ can then be written as

$$\widehat{\mathbb{P}^{\text{corr}}_{d+1}} \colon \mathbb{R}^{\mathcal{X}_d} \times \mathbb{R}^{\mathcal{X}_d} \to \mathbb{C}, \quad (u,u') \mapsto \mathbb{E}^{\text{corr}}_{d+1}\big[e^{iu\cdot r + iu'\cdot r'}\big].$$

This function is well defined since for almost all $\tau \in \mathcal{X}_d$, we have $(M_\tau, M'_\tau) = (0,0)$, which avoids the problem of infinite sums in the exponent of e. One can easily recover $\mathbb{P}^{\text{corr}}_{d+1}$ from $\widehat{\mathbb{P}^{\text{corr}}_{d+1}}$ using the following Fourier-inversion type formula:

$$\begin{aligned}
\mathbb{P}^{\text{corr}}_{d+1}(t,t') &= \mathbb{E}^{\text{corr}}_{d+1}\big[\mathbb{1}_{r=t} \times \mathbb{1}_{r'=t'}\big] \\
&= \mathbb{E}^{\text{corr}}_{d+1}\big[\mathbb{1}_{\{\forall \tau : N_\tau - M_\tau = 0\}} \times \mathbb{1}_{\{\forall \tau :, N'_\tau - M'_\tau = 0\}}\big] \\
&= \mathbb{E}^{\text{corr}}_{d+1}\left[\frac{1}{(2\pi)^p} \int_{[0,2\pi]^p} e^{iu\cdot(r-t)}\,du \times \frac{1}{(2\pi)^{p'}} \int_{[0,2\pi]^{p'}} e^{iu'\cdot(r'-t')}\,du'\right] \\
&= \frac{1}{(2\pi)^{p+p'}} \int_{[0,2\pi]^p} \int_{[0,2\pi]^{p'}} e^{-iu\cdot t - iu'\cdot t'}\, \widehat{\mathbb{P}^{\text{corr}}_{d+1}}(u,u')\,du\,du'.
\end{aligned} \tag{A.14}$$

Here we have used that the number of non-zero coordinates in $t = (N_\tau)_{\tau \in \mathcal{X}_d}$ and t' equals p and p', respectively. The last equality uses Fubini's theorem together with the fact that

$$|e^{iu\cdot(r-t)} \mathbb{P}^{\text{corr}}_{d+1}(u,u')| \leq 1.$$

To manipulate the term $\widehat{\mathbb{P}^{\mathrm{corr}}_{d+1}}(u,u')$ first recall that $N_\tau = \Delta_\tau + \sum_{\tau'\in\mathcal{X}_d}\Gamma_{\tau,\tau'}$ and $N'_{\tau'} = \Delta'_{\tau'} + \sum_{\tau\in\mathcal{X}_d}\Gamma_{\tau,\tau'}$ with

$$\begin{aligned}\Delta_\tau &\sim \mathrm{Poi}\big(\lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau)\big),\\ \Delta'_{\tau'} &\sim \mathrm{Poi}\big(\lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\big),\\ \Gamma_{\tau,\tau'} &\sim \mathrm{Poi}\big(\lambda ss'\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')\big),\end{aligned}$$

which are all independent Poisson variables. This lets us compute

$$\begin{aligned}&\mathbb{E}\big[e^{iu\cdot r+iu'\cdot r'}\big] = \mathbb{E}\big[e^{iu\cdot\Delta+\sum_{\tau'}u\cdot\Gamma_{:,\tau'}}e^{iu'\cdot\Delta'+\sum_\tau u'\cdot\Gamma_{\tau,:}}\big]\\
&\quad= \prod_{\tau\in\mathcal{X}_d}\mathbb{E}\big[e^{iu_\tau\Delta_\tau}\big]\prod_{\tau\in\mathcal{X}_d}\mathbb{E}\big[e^{iu'_\tau\Delta'_\tau}\big]\prod_{\tau,\tau'\in\mathcal{X}_d}\mathbb{E}\big[e^{i(u_\tau+u'_{\tau'})\Gamma_{\tau,\tau'}}\big]\\
&\quad= \exp\Bigg[\sum_\tau\lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau)(e^{iu_\tau}-1)\\
&\qquad\quad+\sum_{\tau'}\lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s')}(\tau')(e^{iu'_{\tau'}}-1)\\
&\qquad\quad+\sum_{\tau,\tau'}\lambda ss'\,\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')(e^{i(u_\tau+u'_{\tau'})}-1)\Bigg]\\
&\quad\overset{(\star)}{=} \exp\Bigg[\lambda s\sum_\tau\mathrm{GW}_d^{(\lambda s)}(\tau)(e^{iu_\tau}-1)\Bigg]\exp\Bigg[\lambda s'\sum_\tau\mathrm{GW}_d^{(\lambda s')}(\tau)(e^{iu'_\tau}-1)\Bigg]\\
&\qquad\times\exp\Bigg[\lambda ss'\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')(e^{iu_\tau}-1)(e^{iu'_{\tau'}}-1)\Bigg]\\
&\quad= \exp\Bigg[\lambda s\sum_\tau\mathrm{GW}_d^{(\lambda s)}(\tau)(e^{iu_\tau}-1)\Bigg]\exp\Bigg[\lambda s'(\tau)\sum_\tau\mathrm{GW}_d^{(\lambda s')}(e^{iu'_\tau}-1)\Bigg]\\
&\qquad\times\sum_{m=0}^\infty\frac{(\lambda ss')^m}{m!}\Bigg(\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')(e^{iu_\tau}-1)(e^{iu'_{\tau'}}-1)\Bigg)^m.\end{aligned}$$

To obtain the equality $(\star)$, one needs to combine all terms with factor $\lambda ss'$ by using that the marginals of $\mathbb{P}_d^{\mathrm{corr}}$ fulfill

$$\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau') = \sum_\tau\mathrm{GW}_d^{(\lambda s')}(\tau) = \sum_{\tau'}\mathrm{GW}_d^{(\lambda s')}(\tau').$$

To shorten the following explanations, we introduce some notation: Define

$$A_d^{(\mu)}(u) := \exp\Bigg[\mu\sum_\tau\mathrm{GW}_d^{(\mu)}(\tau)(e^{iu_\tau}-1)\Bigg]$$

and

$$B_d(u,u') := \sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')(e^{iu_\tau}-1)(e^{iu'_{\tau'}}-1),$$

which lets us write

$$\widehat{\mathbb{P}^{\mathrm{corr}}_{d+1}}(u,u') = \mathbb{E}\big[e^{iu\cdot r+iu'\cdot r'}\big]$$
$$= A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')\sum_{m=0}^{\infty}\frac{(\lambda ss')^m}{m!}\big(B_d(u,u')\big)^m$$
$$= \sum_{m=0}^{\infty}\frac{(\lambda ss')^m}{m!}\Big(\sqrt[m]{A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')}B_d(u,u')\Big)^m.$$

Note that since $|e^{iu_\tau}-1|\le 2$, one can bound $|B_d(u,u')|\le 4$ as well as

$$\big|A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')\big|\le e^{2\lambda s+2\lambda s'} \implies \Big|\sqrt[m]{A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')}\Big|\le e^{2\lambda s+2\lambda s'}. \quad \text{(A.15)}$$

Consequently, we have the following bound which is uniform in (u,u') and summable in m:

$$\Big|\frac{(\lambda ss')^m}{m!}\Big(\sqrt[m]{A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')}B_d(u,u')\Big)^m\Big|\le\frac{1}{m!}\big(4ss'\lambda e^{2\lambda s+2\lambda s'}\big)^m,$$

which lets us apply Fubini's theorem to obtain

$$\mathbb{P}^{\mathrm{corr}}_{d+1}(t,t')\overset{\text{(A.14)}}{=}\frac{1}{(2\pi)^{p+p'}}\int_{[0,2\pi]^{p+p'}}e^{-iu\cdot t-iu'\cdot t'}$$
$$\times\sum_{m=0}^{\infty}\frac{(\lambda ss')^m}{m!}\Big(\sqrt[m]{A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')}B_d(u,u')\Big)^m\,du\,du'$$
$$=\sum_{m=0}^{\infty}\frac{(\lambda ss')^m}{m!}\frac{1}{(2\pi)^{p+p'}}$$
$$\times\int_{[0,2\pi]^p}\int_{[0,2\pi]^{p'}}e^{-iu\cdot t-iu'\cdot t'}A_d^{(\lambda s)}(u)\,A_d^{(\lambda s')}(u')\big(B_d(u,u')\big)^m\,du\,du'. \quad \text{(A.16)}$$

As a next step, we develop the mth power of the term $B_d(u,u')$. For this, we use the induction hypothesis at depth d under the same form as in (A.13), namely

$$\mathbb{P}^{\mathrm{corr}}_{d}(\tau,\tau') = \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\sum_{\beta\in\mathcal{X}_d}\sqrt{ss'}^{|\beta|-1}f_{d,\beta}^{(\lambda s)}(\tau)f_{d,\beta}^{(\lambda s')}(\tau')$$
$$=\sum_{\beta\in\mathcal{X}_d}\sqrt{ss'}^{|\beta|-1}\underbrace{f_{d,\beta}^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\tau)}_{=:g_{d,\beta}^{(\lambda s)}(\tau)}\underbrace{f_{d,\beta}^{(\lambda s')}(\tau')\,\mathrm{GW}_d^{(\lambda s')}(\tau')}_{=g_{d,\beta}^{(\lambda s')}(\tau')}. \quad \text{(A.17)}$$

In the Fourier-inversion formula (A.14), only finitely many u_τ and u'_τ are non-zero. Consequently, the sum $\sum_{\tau,\tau'}$ appearing in $B_d(u,u')$ is finite and we may exchange this summation with the series $\sum_{\beta\in\mathcal{X}_d}$ from equation (A.17). This lets us compute

$$\begin{aligned}
B_d(u,u') &= \sum_{\beta\in\mathcal{X}_d}\sum_{\tau,\tau'\in\mathcal{X}_d}\sqrt{ss'}^{|\beta|-1} g_{d,\beta}^{(\lambda s)}(\tau)(e^{iu_\tau}-1)g_{d,\beta}^{(\lambda s')}(\tau')(e^{iu'_{\tau'}}-1)\\
&= \sum_{\beta\in\mathcal{X}_d}\sqrt{ss'}^{|\beta|-1}\underbrace{\sum_{\tau\in\mathcal{X}_d} g_{d,\beta}^{(\lambda s)}(\tau)(e^{iu_\tau}-1)}_{=:h_{d,\beta}^{(\lambda s)}(u)}\underbrace{\sum_{\tau'\in\mathcal{X}_d} g_{d,\beta}^{(\lambda s')}(\tau')(e^{iu'_{\tau'}}-1)}_{=h_{d,\beta}^{(\lambda s')}(u')}.
\end{aligned}$$

For the $h_{d,\beta}^{(\mu)}(u)$ defined through the last equation, we have the following bound:

$$\begin{aligned}
\left|h_{d,\beta}^{(\mu)}(u)\right| &\le \sum_{\tau\in\mathcal{X}_d}\left|f_{d,\beta}^{(\mu)}(\tau)\right|\mathrm{GW}_d^{(\mu)}(\tau)|e^{iu_\tau}-1|\\
&\le 2\mathbb{E}_{\tau\sim\mathrm{GW}_d^{(\mu)}}\left[\left|f_{d,\beta}^{(\mu)}(\tau)\right|\right]\\
&\overset{\text{Jensen}}{\le} 2\sqrt{\mathbb{E}_{\tau\sim\mathrm{GW}_d^{(\mu)}}\left[\left(f_{d,\beta}^{(\mu)}(\tau)\right)^2\right]}\overset{(A.11)}{=}2.
\end{aligned}\tag{A.18}$$

Taking the expression we have obtained for $B_d(u,u')$ to the power m and using the multinomial Cauchy product formula, we obtain

$$\begin{aligned}
\left(B_d(u,u')\right)^m &= \sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d},\\ \sum_\beta\gamma_\beta=m}}\frac{m!}{\prod_\beta\gamma_\beta!}\prod_{\beta\in\mathcal{X}_d}\left(\sqrt{ss'}^{-1+|\beta|}\right)^{\gamma_\beta}h_{d,\beta}^{(\lambda s)}(u)^{\gamma_\beta}h_{d,\beta}^{(\lambda s')}(u')^{\gamma_\beta}\\
&= m!\sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d},\\ \sum_\beta\gamma_\beta=m}}\sqrt{ss'}^{-\sum_\beta\gamma_\beta+\sum_\beta\gamma_\beta|\beta|}\prod_{\beta\in\mathcal{X}_d}\frac{1}{\sqrt{\gamma_\beta!}}h_{d,\beta}^{(\lambda s)}(u)\gamma_\beta\frac{1}{\sqrt{\gamma_\beta!}}h_{d,\beta}^{(\lambda s')}(u')^{\gamma_\beta}\\
&= \frac{m!}{(\lambda ss')^m}\sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d}:\\ \sum_\beta\gamma_\beta=m}}\sqrt{ss'}^{\sum_\beta\gamma_\beta|\beta|}\prod_{\beta\in\mathcal{X}_d}\frac{\sqrt{\lambda s}^{\gamma_\beta}}{\sqrt{\gamma_\beta!}}h_{d,\beta}^{(\lambda s)}(u)^{\gamma_\beta}\frac{\sqrt{\lambda s'}^{\gamma_\beta}}{\sqrt{\gamma_\beta!}}h_{d,\beta}^{(\lambda s')}(u')^{\gamma_\beta}.
\end{aligned}$$

Inserting this into equation (A.16) yields

$$\begin{aligned}
&\frac{(\lambda ss')^m}{m!(2\pi)^{p+p'}}\int_{[0,2\pi]^p}\int_{[0,2\pi]^{p'}}e^{-iu\cdot t-iu'\cdot t'}A_d^{(\lambda s)}(u)A_d^{(\lambda s')}(u')\left(B_d(u,u')\right)^m\,du\,du'\\
&= \sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d},\\ \sum_\beta\gamma_\beta=m}}\sqrt{ss'}^{\sum_\beta\gamma_\beta|\beta|}\\
&\quad\times\frac{1}{(2\pi)^p}\int_{[0,2\pi]^p}e^{-iu\cdot t}A_d^{(\lambda s)}(u)\prod_{\beta\in\mathcal{X}_d}\frac{\sqrt{\lambda s}^{\gamma_\beta}}{\sqrt{\gamma_\beta!}}h_{d,\beta}^{(\lambda s)}(u)^{\gamma_\beta}\,du\\
&\quad\times\frac{1}{(2\pi)^{p'}}\int_{[0,2\pi]^{p'}}e^{-iu'\cdot t'}A_d^{(\lambda s')}(u')\prod_{\beta\in\mathcal{X}_d}\frac{\sqrt{\lambda s'}^{\gamma_\beta}}{\sqrt{\gamma_\beta!}}(h_{d,\beta}^{(\lambda s')}(u'))^{\gamma_\beta}\,du',
\end{aligned}\tag{A.19}$$

where swapping the summation in $(\gamma_\beta)_{\beta\in\mathcal{X}_d}$ with the integrals in (u,u') is enabled by Fubini's theorem and the fact that

$$\begin{aligned}
&\sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d}:\\ \sum_\beta \gamma_\beta=m}} \Bigg| A_d^{(\lambda s)}(u) A_d^{(\lambda s')}(u') \sqrt{ss'}^{\sum_\beta \gamma_\beta|\beta|} \\
&\qquad \times \prod_{\beta\in\mathcal{X}_d} \frac{\sqrt{\lambda^2 ss'}^{\gamma_\beta}}{\gamma_\beta!} h_{d,\beta}^{(\lambda s)}(u)^{\gamma_\beta} h_{d,\beta}^{(\lambda s')}(u')^{\gamma_\beta} \Bigg| \\
&\overset{(A.15)}{\leq} \frac{e^{2\lambda(s+s')}}{m!} \sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d}:\\ \sum_\beta \gamma_\beta=m}} \frac{m!}{\prod_{\beta\in\mathcal{X}_d}\gamma_\beta!} \\
&\qquad \times \prod_{\beta\in\mathcal{X}_d} \sqrt{ss'}^{|\beta|} \left| \sqrt{\lambda s} h_{d,\beta}^{(\lambda s)}(u) \sqrt{\lambda s'} h_{d,\beta}^{(\lambda s')}(u') \right|^{\gamma_\beta} \\
&= e^{2\lambda(s+s')} \frac{(\lambda\sqrt{ss'})^m}{m!} \left(\sum_{\beta\in\mathcal{X}_d} \sqrt{ss'}^{|\beta|} \left| h_{d,\beta}^{(\lambda s)}(u) \right| \left| h_{d,\beta}^{(\lambda s')}(u') \right| \right)^m \\
&\overset{(A.18)}{\leq} 4e^{2\lambda(s+s')} \frac{(\lambda\sqrt{ss'})^m}{m!} \sqrt{ss'}^m \left(\sum_{n=1}^{\infty} \left| \{\beta\in\mathcal{X}_d \mid |\beta|=n\} \right| \sqrt{ss'}^{n-1} \right)^m .
\end{aligned}$$

The final expression is finite due to Otter's Proposition 4.1 and $ss' < 1$. It is, furthermore, uniform in u, u' and integrating over $(u,u') \in [0,2\pi]^p \times [0,2\pi]^{p'}$ yields a finite term, which validates the application of Fubini's theorem in (A.19).

Now define

$$g_{d+1,\gamma}^{(\mu)}(u) := \frac{1}{(2\pi)^p} \int_{[0,2\pi]^p} e^{-iu\cdot t} A_d^{(\mu)}(u) \prod_{\beta\in\mathcal{X}_d} \frac{\sqrt{\mu}^{\gamma_\beta}}{\sqrt{\gamma_\beta!}} \left(h_{d,\beta}^{(\mu)}(u)\right)^{\gamma_\beta} du$$

and combine equation (A.19) with the expression for $\mathbb{P}_{d+1}^{\text{corr}}(t,t')$ from (A.16) to obtain

$$\mathbb{P}_{d+1}^{\text{corr}}(t,t') = \sum_{m=1}^{\infty} \sum_{\substack{\gamma=(\gamma_\beta)_{\beta\in\mathcal{X}_d},\\ \sum_\beta \gamma_\beta=m}} \sqrt{ss'}^{\sum_\beta \gamma_\beta|\beta|} g_{d+1,\gamma}^{(\lambda s)}(u) g_{d+1,\gamma}^{(\lambda s')}(u).$$

Note that every tuple $(\gamma_\beta)_{\beta\in\mathcal{X}_d}$ with $\sum_\beta \gamma_\beta = m$ corresponds to a unique tree $\gamma \in \mathcal{X}_{d+1}$ with root degree m by the subtree tuple identification. Every such tree has size

$$|\gamma| = 1 + \sum_{\beta\in\mathcal{X}_d} \gamma_\beta|\beta|.$$

This observation leads to

$$\mathbb{P}^{\mathrm{corr}}_{d+1}(t,t') = \sum_{\gamma\in\mathcal{X}_{d+1}} \sqrt{ss'}^{|\gamma|-1} g^{(\lambda s)}_{d+1,\gamma}(u) g^{(\lambda s')}_{d+1,\gamma}(u). \tag{A.20}$$

This decomposition of $\mathbb{P}^{\mathrm{corr}}_{d+1}(t,t')$ is already similar to our objective in (A.13). Hence, all that remains to do is convert $g^{(\mu)}_{d+1,\gamma}(u)$ into a product between $\mathrm{GW}^{(\mu)}_{d+1}(\tau)$ and $f^{(\mu)}_{d+1,\gamma}(\tau)$, where the function f has the desired orthogonality properties. To do so, we start by developing the definition of $g^{(\mu)}_{d+1,\gamma}(u)$:

$$\begin{aligned}
g^{(\mu)}_{d+1,\gamma}(u) &= \frac{1}{(2\pi)^p}\int_{[0,2\pi]^p} e^{-iu\cdot t} A^{(\mu)}_d(u) \prod_\beta \frac{\sqrt{\mu}^{\gamma_\beta}}{\sqrt{\gamma_\beta !}} \left(h^{(\mu)}_{d,\beta}(u)\right)^{\gamma_\beta} du \\
&= \frac{1}{\prod_\beta \sqrt{\gamma_\beta!}} \frac{1}{(2\pi)^p} \\
\times \int_{[0,2\pi]^p} & e^{-iu\cdot t} e^{\mu\sum_\tau \mathrm{GW}^{\mu}_d(\tau)(e^{iu_\tau}-1)} \prod_\beta \Big(\sqrt{\mu}\sum_{\tau\in\mathcal{X}_d} g^{(\lambda s)}_{d,\beta}(\tau)\,(e^{iu_\tau}-1)\Big)^{\gamma_\beta} du \\
&= \frac{1}{\prod_\beta \sqrt{\gamma_\beta!}} \frac{1}{(2\pi)^p} \\
\times \int_{[0,2\pi]^p} & e^{-iu\cdot t} e^{\mu\sum_\tau \mathrm{GW}^{\mu}_d(\tau)(e^{iu_\tau}-1)} \Big(\prod_\beta \gamma_\beta!\Big)[x^\gamma] e^{\sum_\beta x_\beta \sqrt{\mu}\sum_\tau g^{(\mu)}_{d,\beta}(\tau)\,(e^{iu_\tau}-1)} du.
\end{aligned}$$

In the last step, we have introduced some new notation: $x = (x_\beta)_{\beta\in\mathcal{X}_d}$ denotes a tuple of formal variables, $\gamma = (\gamma_\beta)_{\beta\in\mathcal{X}_d}$ is the tuple integers counting the subtrees of γ. We set $x^\gamma := \prod_\beta x_\beta^{\gamma_\beta}$ and let $[x^\gamma]S$ denote the coefficient of x^γ in a formal series S which contains x as a formal variable. Keeping in mind that we are implicitly dealing with a series, we continue computing $g^{(\mu)}_{d+1,\gamma}(u)$ which is equal to

$$\begin{aligned}
&\sqrt{\prod_\beta \gamma_\beta!}\,\frac{1}{(2\pi)^p} \\
&\quad\times \int_{[0,2\pi]^p} e^{-iu\cdot t}[x^\gamma] e^{\sum_\tau (e^{iu_\tau}-1)\mu\,\mathrm{GW}^{\mu}_d(\tau)}\, e^{\sum_\tau (e^{iu_\tau}-1)\sum_\beta x_\beta\sqrt{\mu} g^{(\mu)}_{d,\beta}(\tau)} du \\
&= \sqrt{\prod_\beta \gamma_\beta!}\,\frac{1}{(2\pi)^p}\int_{[0,2\pi]^p} e^{-iu\cdot t}[x^\gamma] e^{\sum_\tau (e^{iu_\tau}-1)[\mu\,\mathrm{GW}^{(\mu)}_d(\tau)+\sum_\beta x_\beta\sqrt{\mu} g^{(\mu)}_{d,\beta}(\tau)]} du \\
&= \sqrt{\prod_\beta \gamma_\beta!}\,[x^\gamma]\frac{1}{(2\pi)^p}\int_{[0,2\pi]^p} e^{-iu\cdot t} e^{\sum_\tau (e^{iu_\tau}-1)[\mu\,\mathrm{GW}^{(\mu)}_d(\tau)+\sum_\beta x_\beta\sqrt{\mu} g^{(\mu)}_{d,\beta}(\tau)]} du.
\end{aligned}$$

Note that in the last step, we have swapped the integral with $[x^\gamma]$. This requires the use of Fubini's theorem to interchange the exponential series with the integral, which is possible due to

$$\sum_{k=0}^{\infty} \frac{1}{k!}\bigg(\int_{[0,2\pi]^p} \bigg| -iu \cdot t$$

$$+ \mu \sum_{\tau} (e^{iu_\tau} - 1)\Big[\mu \, \mathrm{GW}_d^{(\mu)}(\tau) + \sum_\beta x_\beta \sqrt{\mu} g_{d,\beta}^{(\mu)}(\tau) \Big] \bigg| \, du \bigg)^k$$

$$\leq \sum_{k=0}^{\infty} \frac{1}{k!}\bigg(\prod_{\tau \in \mathcal{X}_d} \int_0^{2\pi} |u_\tau N_\tau| \, du_\tau + 2\mu \bigg| \mu \, \mathrm{GW}_d^{(\mu)}(\tau) + \sum_\beta x_\beta \sqrt{\mu} g_{d,\beta}^{(\mu)}(\tau) \bigg| \bigg)^k$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} (\text{constant independent of } k)^k < \infty.$$

Picking up our computations from above once more and introducing the random variable

$$Z \sim \mathrm{Poi}\bigg(\mu \, \mathrm{GW}_d^{(\mu)}(\tau) + \sum_\beta x_\beta \sqrt{\mu} g_{d,\beta}^{(\mu)}(\tau) \bigg),$$

we obtain

$$g_{d+1,\gamma}^{(\mu)}(u) = \sqrt{\prod_\beta \gamma_\beta !} [x^\gamma] \prod_\tau \frac{1}{(2\pi)^p} \int_0^{2\pi} e^{-iu_\tau N_\tau} \mathbb{E}\big[e^{iu_\tau Z}\big] \, du_\tau$$

$$\overset{(A.14)}{=} \sqrt{\prod_\beta \gamma_\beta !} \, [x^\gamma] \prod_\tau \mathbb{P}_{Z \sim \mathrm{Poi}(\mu \, \mathrm{GW}_d^{(\mu)}(\tau) + \sum_\beta x_\beta \sqrt{\mu} f_{d,\beta}^{(\mu)}(\tau) \, \mathrm{GW}_d^{(\mu)}(\tau))} (Z = N_\tau)$$

$$= \sqrt{\prod_\beta \gamma_\beta !} [x^\gamma] \prod_\tau e^{-\mu \, \mathrm{GW}_d^{(\mu)}(\tau) - \sum_\beta x_\beta \sqrt{\mu} g_{d,\beta}^{(\mu)}(\tau)}$$

$$\times \frac{(\mu \, \mathrm{GW}_d^{(\mu)}(\tau) + \sum_\beta x_\beta \sqrt{\mu} f_{d,\beta}^{(\mu)}(\tau) \, \mathrm{GW}_d^{(\mu)}(\tau))^{N_\tau}}{N_\tau !}$$

$$= e^{-\mu} \sqrt{\prod_\beta \gamma_\beta !} [x^\gamma] e^{-\sum_{\beta,\tau} x_\beta \sqrt{\mu} g_{d,\beta}^{(\mu)}(\tau)} \prod_\tau \frac{(\mu \, \mathrm{GW}_d^{(\mu)}(\tau))^{N_\tau}}{N_\tau !}$$

$$\times \bigg(1 + \sum_\beta x_\beta \frac{1}{\sqrt{\mu}} f_{d,\beta}^{(\mu)}(\tau) \bigg)^{N_\tau}$$

$$= \underbrace{e^{-\sum_\tau \mu \, \mathrm{GW}_d^{(\mu)}(\tau)} \prod_\tau \frac{(\mu \, \mathrm{GW}_d^{(\mu)}(\tau))^{N_\tau}}{N_\tau !}}_{= \mathrm{GW}_{d+1}^{(\mu)}(\tau)} \sqrt{\prod_\beta \gamma_\beta !} [x^\gamma] e^{-\sqrt{\mu} \sum_{\beta,\tau} x_\beta g_{d,\beta}^{(\mu)}(\tau)}$$

$$\times \prod_\tau \bigg(1 + \sum_\beta \frac{x_\beta}{\sqrt{\mu}} f_{d,\beta}^{(\mu)}(\tau) \bigg)^{N_\tau}.$$

Finally, we define our candidate for the depth-$(d+1)$ family of functions $f^{(\mu)}_{d+1,\gamma}$ as

$$f^{(\mu)}_{d+1,\gamma}(\tau) := \sqrt{\prod_\beta \gamma_\beta ! [x^\gamma] e^{-\sqrt{\mu}\sum_{\beta,\tau} x_\beta g^{(\mu)}_{d,\beta}(\tau)} \prod_\tau \Big(1 + \sum_\beta \frac{x_\beta}{\sqrt{\mu}} f^{(\mu)}_{d,\beta}(\tau)\Big)^{N_\tau}}. \tag{A.21}$$

Inserting

$$g^{(\mu)}_{d+1,\gamma}(u) = \mathrm{GW}^{(\mu)}_{d+1}(\tau) f^{(\mu)}_{d+1,\gamma}(\tau)$$

into (A.20) yields the objective defined in (A.13). What remains to show are the properties (A.10), (A.11) and (A.12).

Since $f^{(\mu)}_{d+1,\gamma}$ only depends on the parameter μ instead of the asymmetry parameters s, s', its definition in (A.21) coincides with the one presented in the original paper [28]. Hence, we refer the reader to Steps 2.1 and 2.2 in the proof of [28, Theorem 4] to conclude our proof of Theorem 4.7. ∎

A.5. Proofs of three implications from Theorem 4.9

We recall the theorem about one-sided testability equivalences before proving the three implications from that theorem, which were deferred from Section 4.

Theorem A.5. *In the tree correlation testing problem (4.1), the following are equivalent:*

(a) *there exist one-sided tests $\mathcal{T}_d$ to decide $\mathbb{P}^{\mathrm{ind}}_d$ vs. $\mathbb{P}^{\mathrm{corr}}_d$;*

(b) *there is a sequence of thresholds $\theta_d \to \infty$ such that $\mathbb{P}^{\mathrm{ind}}_d(L_d > \theta_d) \to 0$ and $\liminf_d \mathbb{P}^{\mathrm{corr}}_d(L_d > \theta_d) > 0$;*

(c) *the $\mathbb{P}^{\mathrm{ind}}$-martingale $(L_d)_d$ is not uniformly integrable;*

(d) *$\lambda s s' > 1$ and $KL_\infty = \infty$;*

(e) *$\lambda s s' > 1$ and $\mathbb{P}^{\mathrm{corr}}(\liminf_d (\lambda s s')^{-d} \log(L_d) \geq C) \geq 1 - \mathbb{P}(\mathrm{Ext}(\mathrm{GW}^{(\lambda s s')}))$ for a constant $C > 0$ only depending on (λ, s, s') and $\mathrm{Ext}(\mathrm{GW}^{(\lambda s s')})$ denoting the extinction event of the Galton–Watson process with offspring $\mathrm{Poi}(\lambda s s')$.*

We start with (c) ⇒ (b) before showing (4.5) which is part of (b) ⇒ (a). Finally, we move to the most involved implication, (d) ⇒ (e).

Proof of (c) ⇒ (b) *in Theorem* 4.9. Recall that we assume $\mathbb{E}^{\mathrm{ind}}[L_\infty] < 1$ and want to construct a sequence $\theta_d \to \infty$ fulfilling

$$\liminf_{d\to\infty} \mathbb{P}^{\mathrm{corr}}(L_d > \theta_d) > 0 \quad \text{and} \quad \mathbb{P}^{\mathrm{ind}}(L_d > \theta_d) \xrightarrow[d\to\infty]{} 0.$$

Let θ be any continuity point of $L_\infty \# \mathbb{P}^{\mathrm{ind}}$, the pushforward measure of $\mathbb{P}^{\mathrm{ind}}$ under L_∞. We write

$$\begin{aligned}\mathbb{E}^{\mathrm{ind}}[L_d] &= \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d>\theta}] + \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d\le\theta}] \\ &= \mathbb{P}^{\mathrm{corr}}(L_d > \theta) + \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d\le\theta}].\end{aligned}$$

Since $\mathbb{E}^{\mathrm{ind}}[L_d] = 1$, this implies

$$\liminf_{d\to\infty} \mathbb{P}^{\mathrm{corr}}(L_d > \theta) = 1 - \limsup_{d\to\infty} \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d\le\theta}] \ge 1 - \mathbb{E}^{\mathrm{ind}}[L_\infty], \tag{A.22}$$

where in the last step we have used dominated convergence on $L_d \mathbb{1}_{L_d\le\theta} \le \theta \in L^1$ as well as $L_d \mathbb{1}_{L_d\le\theta} \to L_\infty \mathbb{1}_{L_\infty\le\theta} \le L_\infty$ since θ is a continuity point. Setting $\varepsilon := 1 - \mathbb{E}^{\mathrm{ind}}[L_\infty] > 0$, we can now construct a sequence θ_d with the desired properties.

For continuity points $\theta \ge 0$ of $L_\infty \# \mathbb{P}^{\mathrm{ind}}$ define

$$d(\theta) := \inf\Big\{k \in \mathbb{N} \mid \forall d \ge k : \mathbb{P}^{\mathrm{corr}}(L_d > \theta) > \frac{\varepsilon}{2}\Big\}.$$

Thanks to equation (A.22), this is always a finite quantity which increases with θ since $\mathbb{P}^{\mathrm{corr}}(L_d > \bullet)$ is decreasing in d. We set

$$\widehat{\theta}_d := \sup\big\{\theta \in \mathbb{R}_{\ge 0} \mid \theta \text{ is a continuity point of } L_\infty \# \mathbb{P}^{\mathrm{ind}} \text{ and } d(\theta) \le d\big\}.$$

This yields a sequence of finite real numbers since $\{\theta : d(\theta) \le d\}$ is bounded: Otherwise, there would be $\theta_\ell \to \infty$ with $d(\theta_\ell) \le d$ meaning that $\mathbb{P}^{\mathrm{corr}}(L_d > \theta_\ell) > \varepsilon/2$. This is a contradiction to $L_d \in L^1(\mathbb{P}^{\mathrm{ind}})$, because

$$\mathbb{P}^{\mathrm{corr}}(L_d > \theta_\ell) = \mathbb{E}^{\mathrm{ind}}[L_d \mathbb{1}_{L_d>\theta_\ell}] \xrightarrow{\ell\to\infty} 0. \tag{A.23}$$

The same equation (A.23) also implies $\widehat{\theta}_d \xrightarrow{d\to\infty} \infty$, because arbitrarily high values of $d(\theta)$ allow for arbitrarily high values of θ. Finally, choose the sequence $(\theta_d)_d$ in a way that all θ_d are continuity points of $L_\infty \# \mathbb{P}^{\mathrm{ind}}$ satisfying $\theta_d \in (\widehat{\theta}_d - 1, \widehat{\theta}_d)$. By this construction, we have

$$\liminf_{d\to\infty} \mathbb{P}^{\mathrm{corr}}(L_d > \theta_d) \ge \frac{\varepsilon}{2} > 0 \quad \text{and} \quad \mathbb{P}^{\mathrm{ind}}(L_d > \theta_d) \le \frac{\mathbb{E}^{\mathrm{ind}}[L_d]}{\theta_d} = \frac{1}{\theta_d} \xrightarrow{d\to\infty} 0.$$

This is exactly the desired statement. ■

Proof of equation (4.5) *from* (b) ⇒ (a) *in Theorem* 4.7. Recall that our objective is to show

$$(\widehat{\theta}_d)_d \text{ is bounded} \implies \mathbb{E}^{\mathrm{ind}}[L_\infty] < 1$$

under the assumption $\mathbb{P}^{\mathrm{ind}}(\mathcal{T}_d = 1) =: \alpha_d \to 0$.

If $(\widehat{\theta}_d)_d$ is bounded, there exists $M > 0$ and a converging subsequence $(\widehat{\theta}_{d_k})_k$ such that $\widehat{\theta}_{d_k} < M$ for all k. Since L_k converges to L_∞ in distribution, the Portmanteau theorem applied to the open set (M, ∞) yields

$$\mathbb{P}^{\mathrm{ind}}(L_\infty > M) \leq \liminf_{k\to\infty} \mathbb{P}^{\mathrm{ind}}(L_k > M) \leq \mathbb{P}^{\mathrm{ind}}(L_k > \widehat{\theta}_{d_k}) \xrightarrow{k\to\infty} 0,$$

where the second inequality follows from $\alpha_d \to 0$. Consequently, $\mathbb{P}^{\mathrm{ind}}(L_\infty \leq M) = 1$.

Define the convex function $g(x) = (M - x)_+$ and note that $g(L_d)$ is a submartingale with values in $[0, M]$. For this reason, $d \mapsto \mathbb{E}^{\mathrm{ind}}[g(L_d)]$ is non-decreasing and we can apply the dominated convergence theorem to compute its limit:

$$\lim_{d\to\infty} \mathbb{E}^{\mathrm{ind}}\big[g(L_d)\big] = \mathbb{E}^{\mathrm{ind}}\big[g(L_\infty)\big] = \mathbb{E}^{\mathrm{ind}}\big[(M - L_\infty)_+\big] = M - \mathbb{E}^{\mathrm{ind}}[L_\infty]. \quad \text{(A.24)}$$

Despite the limit random variable being bounded by M, we claim that there exists an index d' for which $\mathbb{P}^{\mathrm{ind}}(L_{d'} > M) > 0$. This is true since assuming the opposite, i.e., $\mathbb{P}^{\mathrm{ind}}(L_d \leq M) = 1$ for all d, implies a contradiction to $\liminf_d \mathbb{P}^{\mathrm{corr}}(\mathcal{T}_d = 1) > 0$:

$$\mathbb{P}^{\mathrm{corr}}(\mathcal{T}_d = 1) = \mathbb{E}^{\mathrm{ind}}\big[L_d \mathbb{1}_{\mathcal{T}_d = 1}\big] \leq M\mathbb{P}^{\mathrm{ind}}(\mathcal{T}_d = 1) \leq M\alpha_d \xrightarrow{d\to\infty} 0.$$

At index d', we therefore have

$$\mathbb{E}^{\mathrm{ind}}\big[g(L_{d'})\big] = \mathbb{E}^{\mathrm{ind}}\big[(M - L_{d'})_+\big] > \mathbb{E}^{\mathrm{ind}}[M - L_{d'}] = M - 1. \quad \text{(A.25)}$$

Combining (A.24) and (A.25) then yields

$$M - 1 < \mathbb{E}^{\mathrm{ind}}\big[g(L_{d'})\big] \leq M - \mathbb{E}^{\mathrm{ind}}[L_\infty] \quad \Longrightarrow \quad \mathbb{E}^{\mathrm{ind}}[L_\infty] < 1,$$

which we wanted to show. ∎

Proof of (d) ⇒ (e) *in Theorem* 4.9. This implication is the most intricate and necessitates introducing additional notation. For a rooted tree τ, we recall that its number of nodes up to depth d is denoted $|\tau_d|$. We further write $\mathcal{G}_d(\tau) := \tau_d \setminus \tau_{d-1}$ for the set of nodes in τ which have exactly depth d. Hence, $|\mathcal{G}_d(\tau_*)|$ is the number of τ_*'s nodes at generation d. Furthermore, we will use Proposition 4.5 and the notation introduced for that statement.

Let $(t, t') \sim \mathbb{P}^{\mathrm{corr}}$ and let t_* be their intersection tree from the sampling process. Call σ_* and σ'_* the injections of t_* into t and t', respectively. The intersection tree's marginal distribution is $t_* \sim \mathrm{GW}_{\lambda ss'}$, and we let $\mu := \lambda ss'$ denote the mean of its offspring distribution. The following lemma describes the asymptotic behavior of $|\mathcal{G}_d(t_*)|$ compared to μ^d.

Lemma A.6. *For $t_* \sim \mathrm{GW}_{\lambda ss'}$, set $w_d := |\mathcal{G}_d(t_*)|/\mu^d$. Then, $(w_d)_d$ is a positive martingale whose almost sure limit w satisfies*

$$\mathbb{P}(w > 0) = \mathbb{P}\big(\mathrm{Ext}^c(t_*)\big).$$

For a proof of this standard result on branching processes, we refer the reader to [1, Lemma 2.13].

As a consequence of Lemma A.6, the martingale $\mu^{-d}\,|\mathcal{G}_d(t_*)|$ converges almost surely towards a random variable w.

From now on, we condition on the event $\mathrm{Ext}^c(t_*) = \{t_* \text{ survives}\}$. This is possible since $\mu > 1$ implies $\mathbb{P}(\mathrm{Ext}^c(t_*)) > 0$. By Lemma A.6, we can therefore assume $w > 0$.

Using Proposition 4.5, for all $d, k \in \mathbb{N}$, we have that

$$L_{d+k}(t, t') \geq \underbrace{\prod_{i \in (t_*)_{d-1}} \psi\big(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\big)}_{=: A_d} \times \underbrace{\prod_{j \in \mathcal{G}_d(t_*)} L_k\big(t_{[\sigma_*(j)]}, t'_{[\sigma'_*(j)]}\big)}_{=: B_{d,k}}.$$

We individually examine A_d and $B_{d,k}$, starting with the latter. Taking the logarithm, we get

$$\log(B_{d,k}) = |\mathcal{G}_d(t_*)| \underbrace{\frac{1}{|\mathcal{G}_d(t_*)|} \sum_{i \in \mathcal{G}_d(t_*)} \log\Big(L_k\big(t_{[\sigma_*(i)]}, t'_{[\sigma'_*(i)]}\big)\Big)}_{=: a_d}.$$

For $i \neq j$, the random variables $(t_{[\sigma_*(i)]}, t'_{[\sigma'_*(i)]})$ and $(t_{[\sigma_*(j)]}, t'_{[\sigma'_*(j)]})$ are independent since they represent disjoint branches of a GW-tree. They both follow the same law $\mathbb{P}^{\mathrm{corr}}$, and the logarithms of their likelihoods are in $L^1(\mathbb{P}^{\mathrm{corr}})$. Therefore, the law of large numbers is applicable, yielding $a_d \to \mathbb{E}^{\mathrm{corr}}[\log L_k]$ almost surely. In other words, on an event of probability 1, there exists a sequence $0 < \varepsilon_d \to 0$ such that for all d,

$$\varepsilon_d \geq \big|\mathbb{E}^{\mathrm{corr}}\big[\log(L_k)\big] - a_d\big|, \quad \text{that is, } a_d \geq \big|\mathbb{E}^{\mathrm{corr}}\big[\log(L_k)\big]\big| - \varepsilon_d. \tag{A.26}$$

Similarly, since $|\mathcal{G}_d(t_*)|\mu^{-d} \to w$ almost surely, there is a sequence $0 < \delta_d \to 0$ with

$$\delta_d \geq \big||\mathcal{G}_d(t_*)|\mu^{-d} - w\big|, \quad \text{implying } |\mathcal{G}_d(t_*)| \geq w\mu^d - \delta_d\mu^d \text{ for all } d. \tag{A.27}$$

Noting that $\mathbb{E}^{\mathrm{corr}}[\log(L_k)] = \mathrm{KL}_k$, we can combine (A.26) with (A.27) to obtain

$$\begin{aligned} B_{d,k} = \exp\big\{|\mathcal{G}_d(t_*)|a_d\big\} &\geq \exp\big\{|\mathcal{G}_d(t_*)|(\mathrm{KL}_k - \varepsilon_d)\big\} \\ &\geq \exp\big\{(w\mu^d - \delta_d\mu^d)(\mathrm{KL}_k - \varepsilon_d)\big\} \\ &\geq \exp\big\{\mathrm{KL}_k\, w\mu^d - \mu^d(\delta_d\,\mathrm{KL}_k + \varepsilon_d w)\big\}. \end{aligned}$$

Let n be large enough so that $\mathrm{KL}_k > 1$ and fix an element ω of the probability space Ω for which all sequences converge. Choosing d sufficiently large, one has

$$\delta_d(\omega) \le \frac{1}{4} w(\omega) \quad \text{and} \quad \varepsilon_d(\omega) \le \frac{1}{4}\,\mathrm{KL}_k,$$

yielding

$$\mu^d \delta_d(\omega)\,\mathrm{KL}_k + \mu^d \varepsilon_d(\omega) w(\omega) \le \frac{1}{2}\,\mathrm{KL}_k\, w(\omega)\mu^d.$$

This implies that on an event $\mathcal{B}$ of probability 1,

$$B_{d,k} \ge \exp\left\{\frac{1}{2}\,\mathrm{KL}_k\, w\mu^d\right\} \quad \text{for } d \text{ large enough.} \tag{A.28}$$

Turning to A_d now, we start by introducing the notation

$$F_i := \log\bigl(\psi\bigl(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\bigr)\bigr).$$

As before, we take the logarithm to rewrite

$$\log(A_d) = \sum_{i \in (t_*)_{d-1}} F_i = \sum_{g=0}^{d-1} \sum_{i \in \mathcal{G}_g(t_*)} F_i.$$

One has the following lemma.

Lemma A.7. *The random variables $(F_i)_{i \in (t_*)_d}$ are identically distributed with finite mean and finite variance.*

Proof of Lemma A.7. We start by noting that the law of F_i does not depend on i since

$$\begin{aligned}
F_i &= \log\bigl(\psi\bigl(c_{t_*}(i), c_t(\sigma_*(i)), c_{t'}(\sigma'_*(i))\bigr)\bigr) \\
&= \log\Bigl(e^{\lambda s s'} \lambda^{-c_{t_*}(i)} \frac{1}{c_{t_*}(i)!} (1-s')^{c_t(\sigma_*(i)) - c_{t_*}(i)} (1-s)^{c_{t'}(\sigma'_*(i)) - c_{t_*}(i)}\Bigr) \\
&= \lambda s s' - c_{t_*}(i)\log(\lambda) - \log(c_{t_*}(i)!) \\
&\qquad + \bigl(c_t(\sigma_*(i)) - c_{t_*}(i)\bigr)\log(1-s') + \bigl(c_{t'}(\sigma'_*(i)) - c_{t_*}(i)\bigr)\log(1-s) \\
&\overset{\text{Law}}{=} \lambda s s' - \bigl[X\log(\lambda) + \log(X!)\bigr] + Y\log(1-s') + Y'\log(1-s) =: F,
\end{aligned}$$

where $X \sim \mathrm{Poi}(\lambda s s')$, $Y \sim \mathrm{Poi}(\lambda s(1-s'))$, and $Y' \sim \mathrm{Poi}(\lambda s'(1-s))$ are independent Poisson variables. This follows from the definition of (λ, s, s')-augmentations.

To show that F has finite second moment, define the independent random variables

$$W := \lambda s s' + Y\log(1-s') + Y'\log(1-s), \quad Z := X\log(\lambda) + \log(X!),$$

so that $F = W - Z$. Since Y and Y' are Poisson variables, one has $\mathbb{E}[W^2] < \infty$ and it remains to check that $\mathbb{E}[Z^2] < \infty$. For $x \in \mathbb{N}_{\geq 1}$, one has

$$\log(x!) \leq \log(x^x) = x \log x \leq x^2,$$

which leads to

$$\begin{aligned}\mathbb{E}[Z^2] &= \log(\lambda)^2 \mathbb{E}[X^2] + 2\log(\lambda)\mathbb{E}\big[X \log(X!)\big] + \mathbb{E}\big[\log(X!)^2\big] \\ &\in \mathcal{O}\big(\mathbb{E}[X^2] + \mathbb{E}[X^3] + \mathbb{E}[X^4]\big).\end{aligned}$$

Since the Poisson distribution has moments of all orders, this expression is finite. Thus $\mathbb{E}[Z^2] < \infty$, leading to $\mathbb{E}[F^2] < \infty$. In particular, F has finite variance and mean. ■

According to Lemma A.7, all F_i are identically distributed and we write $f := \mathbb{E}[F]$ for their common mean. Let $\mathcal{G}_g$ be a shorthand for $\mathcal{G}_g(t_*)$ and set

$$\widehat{S}_d := \sum_{g=0}^{d-1} \sum_{i \in \mathcal{G}_g} (F_i - f) = \log(A_d) - f \sum_{g=0}^{d-1} |\mathcal{G}_g|.$$

We define the sigma-fields $\mathcal{F}_d := \sigma(F_i : i \in \mathcal{G}_g, g \leq d-1)$ and make the following claim:

$$\mathbb{E}[\widehat{S}_d^2] = \mathbb{E}[\widehat{S}_{d-1}^2] + \mathrm{Var}(F)\, \mu^{d-1}. \tag{A.29}$$

To show this equation, use the tower rule of conditional expectation to compute

$$\begin{aligned}\mathbb{E}[\widehat{S}_d^2] &= \mathbb{E}\big[\mathbb{E}\big[\widehat{S}_d^2 \mid \mathcal{F}_{d-1}\big]\big] \\ &= \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg(\widehat{S}_{d-1} + \sum_{i \in \mathcal{G}_{d-1}} (F_i - f)\Bigg)^2 \mid \mathcal{F}_{d-1}\Bigg]\Bigg] \\ &= \mathbb{E}[\widehat{S}_{d-1}^2] + \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg(\sum_{i \in \mathcal{G}_{d-1}} (F_i - f)\Bigg)^2 \mid \mathcal{F}_{d-1}\Bigg]\Bigg] \\ &\quad + 2\mathbb{E}\Bigg[\widehat{S}_{d-1}\mathbb{E}\Bigg[\sum_{i \in \mathcal{G}_{d-1}} (F_i - f) \mid \mathcal{F}_{d-1}\Bigg]\Bigg].\end{aligned}$$

Conditionally on $\mathcal{F}_{d-1}$, the random variables $(F_i)_{i \in \mathcal{G}_{d-1}}$ are independent of common law F, because the descendants of distinct nodes in generation $d-1$ are generated independently in the Galton–Watson process. Therefore,

$$\mathbb{E}\big[F_i - f \mid \mathcal{F}_{d-1}\big] = 0 \quad \text{and} \quad \mathbb{E}\big[(F_i - f)^2 \mid \mathcal{F}_{d-1}\big] = \mathrm{Var}(F) \quad \text{for all } i \in \mathcal{G}_{d-1}.$$

The zero-mean and independence of $(F_i - f), (F_j - f)$ for $i \neq j$ implies that mixed terms vanish when we develop the following square:

$$\mathbb{E}\Bigg[\Bigg(\sum_{i \in \mathcal{G}_{d-1}} (F_i - f)\Bigg)^2 \mid \mathcal{F}_{d-1}\Bigg] = \sum_{i \in \mathcal{G}_{d-1}} \mathrm{Var}(F) = \mathrm{Var}(F)|\mathcal{G}_{d-1}|.$$

This proves (A.29), which we apply recursively to obtain

$$\mathbb{E}[\widehat{S}_d^2] = \sum_{g=1}^{d} \mathrm{Var}(F)\mu^{g-1} = \mathrm{Var}(F)\frac{\mu^d - 1}{\mu - 1}. \tag{A.30}$$

This allows us to establish

$$\mathbb{P}\big(|\widehat{S}_d| \geq d\mu^{d/2}\big) \leq \frac{\mathbb{E}[\widehat{S}_d^2]}{d^2\mu^d} \leq \frac{1}{d^2}\mathrm{Var}(F)\frac{1-\mu^{-d}}{\mu - 1} \leq \frac{C}{d^2},$$

where we have used Markov's inequality, equation (A.30), Lemma A.7, and the fact that $1 - \mu^{-d} \leq 1$. The constant C may depend on μ and $\mathrm{Var}(F)$ but is independent of d. Consequently, the probabilities of the events $\{|\widehat{S}_d| \geq d\mu^{d/2}\}$ are summable, and by Borel–Cantelli there almost surely exists an $D \in \mathbb{N}$ such that

$$\left|\log(A_d) - f\sum_{g=0}^{d-1}|\mathcal{G}_g|\right| < d\mu^{d/2} \quad \text{for all } d \geq D. \tag{A.31}$$

Recall that Lemma A.6 states that, almost surely,

$$\mu^{-g}|\mathcal{G}_g| \xrightarrow[g\to\infty]{} w, \quad \text{with } w(\omega) > 0 \text{ for almost all } \omega \in \mathrm{Ext}^c(t_*).$$

Fix $\delta \in (0,1)$. Then for almost every $\omega \in \mathrm{Ext}^c(t_*)$, there exists $D'(\omega) \in \mathbb{N}$ such that

$$(1-\delta)w(\omega)\mu^g \leq |\mathcal{G}_g|(\omega) \leq (1+\delta)w(\omega)\mu^g \quad \text{for all } g \geq D'(\omega). \tag{A.32}$$

Next, derive a lower bound on the drift term $f\sum_{g=0}^{d-1}|\mathcal{G}_g|(\omega)$: If $f \geq 0$, then trivially

$$f\sum_{g=0}^{d-1}|\mathcal{G}_g|(\omega) \geq 0.$$

If $f < 0$, then using the upper bound in (A.32) for all $d \geq D'(\omega)$,

$$\begin{aligned} f\sum_{g=0}^{d-1}|\mathcal{G}_g|(\omega) &\geq f\sum_{g=0}^{D'(\omega)-1}|\mathcal{G}_g|(\omega) + f(1+\delta)w(\omega)\sum_{g=D'(\omega)}^{d-1}\mu^g \\ &\geq -C_1(\omega) - \frac{|f|(1+\delta)}{\mu-1}w(\omega)\mu^d \end{aligned}$$

for a finite constant $C_1(\omega) \geq 0$. Hence, for all sufficiently large d,

$$f\sum_{g=0}^{d-1}|\mathcal{G}_g|(\omega) \geq -\frac{|f|(1+\delta)}{\mu-1}w(\omega)\mu^d - C_1(\omega).$$

Combining this with (A.31), we obtain that, for almost all $\omega \in \mathrm{Ext}^c(t_*)$, there exists a finite $D''(\omega)$ such that for all $d \geq D''(\omega)$,

$$\log(A_d)(\omega) \geq -\frac{|f|(1+\delta)}{\mu-1}\, w(\omega)\mu^d - C_1(\omega) - d\mu^{d/2}. \tag{A.33}$$

Since $w(\omega) > 0$, by increasing $D''(\omega)$ if needed, we may ensure that

$$C_1(\omega) + d\mu^{d/2} \leq w(\omega)\mu^d \quad \text{for } d \geq D''(\omega),$$

which, if plugged into (A.33), leads to

$$\log(A_d)(\omega) \geq -\left(\frac{|f|(1+\delta)}{\mu-1} + 1\right) w(\omega)\mu^d \quad \text{for } d \geq D''(\omega). \tag{A.34}$$

Let $\mathcal{A}$ denote the event where (A.34) holds, which has probability

$$\mathbb{P}(\mathcal{A}) \geq \mathbb{P}\big(\mathrm{Ext}^c(t_*)\big).$$

For every $\omega \in \mathcal{A}$ and sufficiently large d,

$$\log(A_d)(\omega) \geq -C_A w(\omega)\mu^d, \quad \text{where } C_A := \frac{|f|(1+\delta)}{\mu-1} + 1.$$

Since C_A is deterministic, we can choose k large enough so that $\mathrm{KL}_k/2 > C_A$. Combining (A.34) with (A.28), we obtain that on the event $\mathcal{A} \cap \mathcal{B}$,

$$\log\big(L_{d+k}(t,t')\big) \geq \log(A_d) + \log(B_{d,k}) \geq \Big(-C_A + \frac{1}{2}\mathrm{KL}_k\Big) w\mu^d$$

for all sufficiently large d. Since $\mathbb{P}(\mathcal{B}) = 1$, we conclude

$$\begin{aligned}\mathbb{P}^{\mathrm{corr}}\Big(\liminf_{d\to\infty} \mu^{-(d+k)} \log(L_{d+k}) \geq \mu^{-k}\Big(-C_A + \frac{1}{2}\mathrm{KL}_k\Big)\Big) &\geq \mathbb{P}(\mathcal{A} \cap \mathcal{B}) \\ &\geq \mathbb{P}\big(\mathrm{Ext}^c(t_*)\big).\end{aligned}$$

The result (e) from Theorem A.5 then follows by setting

$$C := \mu^{-k}\Big(C' + \frac{1}{2}\mathrm{KL}_k\Big) > 0$$

and observing that

$$\mathbb{P}\big(\mathrm{Ext}^c(t_*)\big) = 1 - \mathbb{P}\big(\mathrm{Ext}(\mathrm{GW}_{\lambda s s'})\big).$$

As a final remark, the constant C can be computed from μ, $\mathbb{E}[F]$, k, and KL_k, which all solely depend on λ, s, and s'. ■

A.6. Proof of Lemma 4.13

We start by recalling Lemma 4.13 on Gaussian approximation, which is instrumental for proving (ii) in Theorem 4.10.

Lemma A.8 (Copy of Lemma 4.13). *Let $d \in \mathbb{N}$ and recall that $\mathcal{X}_{d+1} = \mathbb{N}^{\mathcal{X}_d}$ by the subtree tuple identification. There exists a pair of functions*

$$(y, y') \colon \mathcal{X}_{d+1} \times \mathcal{X}_{d+1} \to \mathbb{R}^{\mathcal{X}_d} \times \mathbb{R}^{\mathcal{X}_d}, \quad (t, t') \mapsto \big(y_\beta(t), y'_\beta(t')\big)_{\beta \in \mathcal{X}_d}$$

with the property that the above map is affine and bijective. The law of the random vector $(y(t), y'(t')))$ shall be denoted as $\mathcal{L}(y, y')$ if $(t, t') \sim \mathbb{P}^{\mathrm{corr}}_{d+1}$, and $\mathcal{L}(y) \otimes \mathcal{L}(y')$ in the case where $(t, t') \sim \mathbb{P}^{\mathrm{ind}}_{d+1}$.

One has

$$\mathrm{KL}\big(\mathbb{P}^{\mathrm{corr}}_{d+1} \| \mathbb{P}^{\mathrm{ind}}_{d+1}\big) = \mathrm{KL}\big(\mathcal{L}(y, y') \| \mathcal{L}(y) \otimes \mathcal{L}(y')\big)$$

and that $\mathcal{L}(y, y')$ converges weakly towards the joint law of a pair of random variables

$$\mathcal{L}(y, y') \xrightarrow[\lambda \to \infty]{w} \mathcal{L}(z, z').$$

The pair $(z, z') = ((z_\beta)_{\beta \in \mathcal{X}_d}, (z'_\beta)_{\beta \in \mathcal{X}_d})$ is a Gaussian vector of infinite dimension with zero mean and covariances defined by

$$\mathbb{E}[z_\beta z_\gamma] = \mathbb{E}[z'_\beta z'_\gamma] = \mathbb{1}_{\beta = \gamma}, \quad \mathbb{E}[z_\beta z'_\gamma] = \sqrt{ss'}^{|\beta|} \mathbb{1}_{\beta = \gamma} \quad \text{for all } \beta, \gamma \in \mathcal{X}_d.$$

Proof. The idea for proving the Gaussian approximation lemma is to transform the (integer-valued) subtree tuple representation of a tree t into a closely related tuple of real random variables $y(t)$ with zero mean and normalized variance. Showing weak convergence will then use Levy's Continuity theorem, applied to infinite-dimensional random vectors.

We begin by letting $t = (N_\tau)_{\tau \in \mathcal{X}_d}$, $t' = (N'_\tau)_{\tau \in \mathcal{X}_d}$ be two trees. For all $\beta \in \mathcal{X}_d$, set

$$y_\beta(t) := \frac{1}{\sqrt{\lambda s}} \sum_{\tau \in \mathcal{X}_d} f^{(\lambda s)}_{d,\beta}(\tau)\big(N_\tau - \lambda s\, \mathrm{GW}^{(\lambda s)}_d(\tau)\big),$$
$$y'_\beta(t') := \frac{1}{\sqrt{\lambda s'}} \sum_{\tau \in \mathcal{X}_d} f^{(\lambda s')}_{d,\beta}(\tau)\big(N'_\tau - \lambda s'\, \mathrm{GW}^{(\lambda s')}_d(\tau)\big).$$

The function $y := (y_\beta)_{\beta \in \mathcal{X}_d} \colon \mathcal{X}_{d+1} \to \mathbb{R}^{\mathcal{X}_d}$ is clearly affine from the definition. It is furthermore bijective with its inverse map given by $t(y) = (N_\tau(y))_{\tau \in \mathcal{X}_d}$, where

$$N_\tau(y) = \lambda s\, \mathrm{GW}^{(\lambda s)}_d(\tau) + \sqrt{\lambda s}\, \mathrm{GW}^{(\lambda s)}_d(\tau) \sum_{\beta \in \mathcal{X}_d} y_\beta f^{(\lambda s)}_{d,\beta}(\tau).$$

This is indeed the inverse because

$$\begin{aligned}
\sum_{\beta\in\mathcal{X}_d} y_\beta f_{d,\beta}^{(\lambda s)}(\tau) &= \frac{1}{\sqrt{\lambda s}} \sum_{\beta\in\mathcal{X}_d}\sum_{\hat\tau\in\mathcal{X}_d} f_{d,\beta}^{(\lambda s)}(\hat\tau)\big(N_{\hat\tau} - \lambda s\,\mathrm{GW}_d^{(\lambda s)}(\hat\tau)\big) f_{d,\beta}^{(\lambda s)}(\tau)\\
&= \frac{1}{\sqrt{\lambda s}} \sum_{\hat\tau\in\mathcal{X}_d}\big(N_{\hat\tau} - \lambda s\,\mathrm{GW}_d^{(\lambda s)}(\hat\tau)\big) \sum_{\beta\in\mathcal{X}_d} f_{d,\beta}^{(\lambda s)}(\hat\tau) f_{d,\beta}^{(\lambda s)}(\tau)\\
&\overset{(4.4)}{=} \frac{1}{\sqrt{\lambda s}} \sum_{\hat\tau\in\mathcal{X}_d}\big(N_{\hat\tau} - \lambda s\,\mathrm{GW}_d^{(\lambda s)}(\hat\tau)\big)\frac{\mathbb{1}_{\tau=\hat\tau}}{\mathrm{GW}_d^{(\lambda s)}(\tau)}\\
&= \frac{1}{\sqrt{\lambda s}}\left(\frac{N_\tau}{\mathrm{GW}_d^{(\lambda s)}} - \lambda s\right).
\end{aligned}$$

The Kullback–Leibler divergence is invariant to affine, bijective transformations (see, for instance, [46]), and therefore

$$\mathrm{KL}\big(\mathbb{P}_{d+1}^{\mathrm{corr}}\|\mathbb{P}_{d+1}^{\mathrm{ind}}\big) = \mathrm{KL}\big(\mathcal{L}(y,y')\|\mathcal{L}(y)\otimes\mathcal{L}(y')\big),$$

where we recall that $\mathcal{L}(y,y')$ is the law of $(y(t),y'(t'))$ if $(t,t')\sim\mathbb{P}_{d+1}^{\mathrm{corr}}$, and $\mathcal{L}(y)\otimes\mathcal{L}(y')$ is the law of $(y(t),y'(t'))$ if $(t,t')\sim\mathbb{P}_{d+1}^{\mathrm{ind}}$. These notations do not mention λ and d, but we emphasize that y and y' depend on these parameters.

Next, we will show that a random vector $(y,y')\sim\mathcal{L}(y,y')$ converges in distribution for $\lambda\to\infty$. For this, let

$$(z,z') = \big((z_\beta)_{\beta\in\mathcal{X}_d},(z'_\beta)_{\beta\in\mathcal{X}_d}\big)$$

be an infinite-dimensional Gaussian vector with zero expectation and the following covariance structure:

$$\mathbb{E}[z_\beta z_\gamma] = \mathbb{E}[z'_\beta z'_\gamma] = \mathbb{1}_{\beta=\gamma},\quad \mathbb{E}[z_\beta z'_\gamma] = \sqrt{ss'}^{|\beta|}\mathbb{1}_{\beta=\gamma}\quad \text{for all } \beta,\gamma\in\mathcal{X}_d.$$

To show

$$(y,y')\xrightarrow[\lambda\to\infty]{(d)}(z,z'),$$

we start with some clarifications on weak convergence in metric spaces. Note that both (y,y') and (z,z') are elements of $\mathbb{R}^{\mathcal{X}_d}\times\mathbb{R}^{\mathcal{X}_d}$, the space of real-valued functions indexed by trees. According to [8, Example 1.2], this space can be made into a complete separable metric space by endowing it with the distance

$$\begin{aligned}
\delta\big((y,y'),(z,z')\big) := \frac{1}{2}\sum_{\beta\in\mathcal{X}_d} 3^{-|\beta|}\min\big(|y_\beta - z_\beta|,1\big)\\
+\frac{1}{2}\sum_{\beta\in\mathcal{X}_d} 3^{-|\beta|}\min\big(|y'_\beta - z'_\beta|,1\big).
\end{aligned}$$

This metric is well defined since, for any $(y, y'), (z, z')$, one has

$$\delta\big((y, y'), (z, z')\big) \leq \sum_{\beta} 3^{-|\beta|} = \sum_{n=1}^{\infty} |\mathcal{X}_d^{(n)}| 3^{-n} \leq \Phi\Big(\frac{1}{3}\Big) \leq \Phi(\alpha) < \infty,$$

where finiteness follows from Proposition 4.1.

Since both (y, y') and (z, z') are random variables with values in the metric space of real-valued sequences $(\mathbb{R}^{\mathcal{X}_d} \times \mathbb{R}^{\mathcal{X}_d}, \delta)$, we know from [8, Example 2.4] that the sequences with finitely many non-zero elements form a *convergence-determining class*. This means that to establish the desired convergence in distribution, it suffices to check $(y, y') \to (z, z')$ for all finite-dimensional subsequences.

To do so, Levy's Continuity theorem allows us to consider (logarithms of) characteristic functions: For any two sequences $v = (v_\beta)_{\beta \in \mathcal{X}_d}$, $v' = (v'_\beta)_{\beta \in \mathcal{X}_d} \in \mathbb{R}^{\mathcal{X}_d}$ with only finitely many non-zero coordinates, our goal is to show

$$\log\big(\mathbb{E}\big[e^{iv\cdot y + iv'\cdot y'}\big]\big) \xrightarrow[\lambda\to\infty]{} \log\big(\mathbb{E}\big[e^{iv\cdot z + iv'\cdot z'}\big]\big),$$

where we use the notation $v \cdot y := \sum_\beta v_\beta y_\beta$.

The vector (z, z') being Gaussian, we can immediately compute the right-hand side:

$$\begin{aligned}
\log\big(\mathbb{E}\big[e^{iv\cdot z + iv'\cdot z'}\big]\big) &= -\frac{1}{2}(v, v')^\top \operatorname{Cov}(z, z')(v, v')^\top \\
&= -\frac{1}{2}(v_{\beta_1}, v'_{\beta_1}, v_{\beta_2}, \dots) \operatorname{Cov}(Z)\, (v_{\beta_1}, v'_{\beta_1}, v_{\beta_2}, \dots)^\top \\
&= -\frac{1}{2} \sum_{\beta \in \mathcal{X}_d} (v_\beta)^2 + (v'_\beta)^2 + 2\sqrt{ss'}^{|\beta|} v_\beta v'_\beta.
\end{aligned}$$

The left-hand side is more involved. Recall that

$$N_\tau = \Delta_\tau + \sum_{\tau' \in \mathcal{X}_d} \Gamma_{\tau,\tau'} \quad \text{and} \quad N'_{\tau'} = \Delta'_{\tau'} + \sum_{\tau \in \mathcal{X}_d} \Gamma_{\tau,\tau'}$$

with

$$\begin{aligned}
\Delta_\tau &\sim \operatorname{Poi}\big(\lambda s(1-s')\, \mathrm{GW}_d^{(\lambda s)}(\tau)\big), \\
\Delta'_{\tau'} &\sim \operatorname{Poi}\big(\lambda s'(1-s)\ \mathrm{GW}_d^{(\lambda s')}(\tau')\big), \\
\Gamma_{\tau,\tau'} &\sim \operatorname{Poi}\big(\lambda ss'\, \mathbb{P}_d^{\mathrm{corr}}(\tau, \tau')\big).
\end{aligned}$$

Using the notations

$$v \cdot f_{d,\bullet}^{(\mu)}(\tau) := \sum_{\beta \in \mathcal{X}_d} v_\beta f_{d,\beta}^{(\mu)}(\tau) \quad \text{and} \quad \Gamma_{\tau,\bullet} = \sum_{\tau'} \Gamma_{\tau,\tau'},$$

compute

$$
\begin{aligned}
e^{iv\cdot y+iv'\cdot y'} &= \exp\Big[\frac{i}{\sqrt{\lambda s}}\sum_{\tau\in\mathcal{X}_d} v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\big(\Delta_\tau+\Gamma_{\tau,\bullet}-\lambda s\,\mathrm{GW}_d^{(\lambda s)}(\tau)\big)\Big]\\
&\qquad\times\exp\Big[\frac{i}{\sqrt{\lambda s'}}\sum_{\tau'\in\mathcal{X}_d} v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau')\big(\Delta'_{\tau'}+\Gamma_{\bullet,\tau'}-\lambda s'\,\mathrm{GW}_d^{(\lambda s')}(\tau')\big)\Big]\\
&= \exp\Big[-i\sqrt{\lambda s}\sum_{\tau} v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\tau)\\
&\qquad\qquad -i\sqrt{\lambda s'}\sum_{\tau'} v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\Big]\\
&\qquad\times\prod_{\tau,\tau'}\exp\Big(\frac{i}{\sqrt{\lambda s}}v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)+\frac{i}{\sqrt{\lambda s'}}v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau')\Big)^{\Gamma_{\tau,\tau'}}\\
&\qquad\times\prod_{\tau}\exp\Big(\frac{i}{\sqrt{\lambda s}}v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\Big)^{\Delta_\tau}\prod_{\tau'}\exp\Big(\frac{i}{\sqrt{\lambda s'}}v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau')\Big)^{\Delta'_{\tau'}}.
\end{aligned}
$$

Since Δ_τ, $\Delta'_{\tau'}$, and $\Gamma_{\tau,\tau'}$ are independent Poisson variables, taking the expectation and then the logarithm on both sides yields

$$
\begin{aligned}
\log\mathbb{E}\big[e^{iv\cdot y+iv'\cdot y'}\big] &= -i\sqrt{\lambda s}\sum_{\tau} v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\tau)\\
&\quad -i\sqrt{\lambda s'}\sum_{\tau'} v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\\
&\quad +\sum_{\tau,\tau'}\lambda ss'\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')\Big(\exp\Big[\frac{i}{\sqrt{\lambda s}}v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)+\frac{i}{\sqrt{\lambda s'}}v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau)\Big]-1\Big)\\
&\quad +\sum_{\tau}\lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau)\Big(\exp\Big[\frac{i}{\sqrt{\lambda s}}v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\Big]-1\Big)\\
&\quad +\sum_{\tau'}\lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\Big(\exp\Big[\frac{i}{\sqrt{\lambda s'}}v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau)\Big]-1\Big).
\end{aligned}
$$

The next step consists in developing all the exponentials as power series. Writing

$$
w(\tau) := v\cdot f_{d,\bullet}^{(\lambda s)}(\tau)\quad\text{and}\quad w'(\tau') := v'\cdot f_{d,\bullet}^{(\lambda s')}(\tau'),
$$

one has

$$
\begin{aligned}
\log\mathbb{E}\big[e^{iv\cdot y+iv'\cdot y'}\big] &= -i\sqrt{\lambda s}\sum_{\tau} w(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\tau)-i\sqrt{\lambda s'}\sum_{\tau'} w'(\tau')\,\mathrm{GW}_d^{(\lambda s')}(\tau')\\
&\quad +\sum_{\tau,\tau'}\lambda ss'\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')\Big(\frac{iw(\tau)}{\sqrt{\lambda s}}+\frac{iw'(\tau)}{\sqrt{\lambda s'}}\\
&\qquad -\frac{w(\tau)^2}{2\lambda s}-\frac{w'(\tau)^2}{2\lambda s'}-\frac{w(\tau)w'(\tau')}{\lambda\sqrt{ss'}}+\mathcal{O}\big(\lambda^{-3/2}\big)\Big)
\end{aligned}
$$

$$+\sum_{\tau}\lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau)\Big(\frac{iw(\tau)}{\sqrt{\lambda s}}-\frac{w(\tau)^2}{2\lambda s}+\mathcal{O}\big(\lambda^{-3/2}\big)\Big)$$
$$+\sum_{\tau'}\lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\Big(\frac{iw'(\tau')}{\sqrt{\lambda s'}}-\frac{w'(\tau')^2}{2\lambda s'}+\mathcal{O}\big(\lambda^{-3/2}\big)\Big).$$

To understand how the right-hand side behaves for $\lambda\to\infty$, we start with the terms in $\mathcal{O}(\lambda^{-3/2})$. Each of these terms is multiplied by λ, leaving them still of order $\lambda^{-1/2}$ which converges to 0. In other words, these terms vanish in the limit.

We move on to extracting all the terms of order $\sqrt{\lambda}$ which cancel each other out:

$$\begin{aligned}&-i\sqrt{\lambda s}\sum_{\tau}w(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\tau)-i\sqrt{\lambda s'}\sum_{\tau'}w'(\tau')\,\mathrm{GW}_d^{(\lambda s')}(\tau')\\ &\quad+is'\sqrt{\lambda s}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w(\tau)+is\sqrt{\lambda s'}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w'(\tau')\\ &\quad+i(1-s')\sqrt{\lambda s}\sum_{\tau}\mathrm{GW}_d^{(\lambda s)}(\tau)w(\tau)+i(1-s)\sqrt{\lambda s'}\sum_{\tau'}\mathrm{GW}_d^{(\lambda s')}(\tau')w'(\tau')=0.\end{aligned}$$

For this computation, we have used the identity

$$\sum_{\tau}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')=\mathrm{GW}_d^{(\lambda s')}(\tau'),$$

as well as

$$\sum_{\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')=\mathrm{GW}_d^{(\lambda s)}(\tau).$$

Next, the terms of order $\lambda^0=1$ are

$$\begin{aligned}&-\frac{s'}{2}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w(\tau)^2-\frac{s}{2}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w'(\tau)^2\\ &\qquad-\sqrt{ss'}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w(\tau)w'(\tau')\\ &\qquad-\frac{1-s'}{2}\sum_{\tau}\mathrm{GW}_d^{(\lambda s)}(\tau)w(\tau)^2-\frac{1-s}{2}\sum_{\tau'}\mathrm{GW}_d^{(\lambda s')}(\tau')w'(\tau')^2\\ &\quad=-\frac{1}{2}\sum_{\tau}\mathrm{GW}_d^{(\lambda s)}(\tau)w(\tau)^2-\frac{1}{2}\sum_{\tau'}\mathrm{GW}_d^{(\lambda s')}(\tau')w'(\tau')^2\\ &\qquad-\sqrt{ss'}\sum_{\tau,\tau'}\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')w(\tau)w'(\tau')\\ &\quad\overset{(\star)}{=}-\frac{1}{2}\sum_{\beta}(v_\beta)^2-\frac{1}{2}\sum_{\beta}(v'_\beta)^2-\sum_{\beta}v_\beta v'_\beta\sqrt{ss'}^{|\beta|-1}.\end{aligned}$$

The last expression equals $\log \mathbb{E}[e^{iv\cdot z+iv'\cdot z'}]$, which will conclude the proof. All that remains to show is, therefore, the equality ($\star$), which requires some additional computations. We start by using the first orthogonality property of $f_{d,\beta}^{\lambda s}$ (equation (4.3)) to compute

$$\begin{aligned}\sum_{\tau} \mathrm{GW}_d^{(\lambda s)}(\tau) w(\tau)^2 &= \sum_{\tau} \mathrm{GW}_d^{(\lambda s)}(\tau) \Big(\sum_{\beta} v_\beta f_{d,\beta}^{(\lambda s)}(\tau)\Big)^2 \\ &= \sum_{\beta,\gamma} v_\beta v_\gamma \sum_{\tau} \mathrm{GW}_d^{(\lambda s)}(\tau) f_{d,\beta}^{(\lambda s)}(\tau) f_{d,\gamma}^{(\lambda s)}(\tau) \\ &= \sum_{\beta,\gamma} v_\beta v_\gamma \mathbb{1}_{\beta=\gamma} = \sum_{\beta} (v_\beta)^2.\end{aligned}$$

Similarly,

$$\sum_{\tau'} \mathrm{GW}_d^{(\lambda s')}(\tau') w'(\tau')^2 = \sum_{\beta} (v'_\beta)^2.$$

Next, equation (A.9) under the form

$$\mathbb{P}_d^{\mathrm{corr}}(\tau,\tau') = \mathrm{GW}_d^{(\lambda s)}(\tau)\, \mathrm{GW}_d^{(\lambda s')}(\tau') \sum_{\delta \in \mathcal{X}_d} \sqrt{ss'}^{\,|\delta|-1} f_{d,\delta}^{(\lambda s)}(\tau) f_{d,\delta}^{(\lambda s')}(\tau')$$

implies the following:

$$\begin{aligned}\sum_{\tau,\tau'} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau') w(\tau) w'(\tau') &= \sum_{\tau,\tau'} \mathrm{GW}_d^{(\lambda s)}(\tau)\, \mathrm{GW}_d^{(\lambda s')}(\tau') \\ &\quad \times \sum_{\delta} \sqrt{ss'}^{\,|\delta|-1} f_{d,\delta}^{(\lambda s)}(\tau) f_{d,\delta}^{(\lambda s')}(\tau') \sum_{\beta,\gamma} v_\beta f_{d,\beta}^{(\lambda s)}(\tau) v'_\gamma f_{d,\gamma}^{(\lambda s')}(\tau') \\ &= \sum_{\beta,\gamma,\delta} v_\beta v'_\gamma \sqrt{ss'}^{\,|\delta|-1} \Big(\sum_{\tau} \mathrm{GW}_d^{(\lambda s)}(\tau) f_{d,\delta}^{(\lambda s)}(\tau) f_{d,\beta}^{(\lambda s)}(\tau)\Big) \\ &\quad \times \Big(\sum_{\tau'} \mathrm{GW}_d^{(\lambda s')}(\tau') f_{d,\delta}^{(\lambda s')}(\tau') f_{d,\gamma}^{(\lambda s')}(\tau')\Big) \\ &= \sum_{\beta,\gamma,\delta} v_\beta v'_\gamma \sqrt{ss'}^{\,|\delta|-1} \mathbb{1}_{\delta=\beta} \mathbb{1}_{\delta=\gamma} = \frac{1}{\sqrt{ss'}} \sum_{\beta} v_\beta v'_\beta \sqrt{ss'}^{\,|\beta|}.\end{aligned}$$

Putting everything together, the equality ($\star$) in our earlier computation is now immediate. This finishes the proof of

$$\log \mathbb{E}\big[e^{iv\cdot y+iv'\cdot y'}\big] \xrightarrow[\lambda\to\infty]{} \log \mathbb{E}\big[e^{iv\cdot z+iv'\cdot z'}\big],$$

which implies that

$$\mathcal{L}(y,y') \xrightarrow[\lambda\to\infty]{w} \mathcal{L}(z,z'). \qquad \blacksquare$$

A.7. Proof of Lemma 4.17

We recall the technical lemma about the moments of Z before providing proof of it.

Lemma A.9 (Copy of Lemma 4.17). *The random variable* $Z := Z(t,t')$ *has the following moments under* $\mathbb{P}^{\mathrm{ind}}$ *and* $\mathbb{P}^{\mathrm{ind}}$*:*

(a) $\mathbb{E}^{\mathrm{ind}}_{d+1}[Z] = 0$;

(b) $\mathbb{E}^{\mathrm{corr}}_{d+1}[Z] = \lambda s s' \mathbb{P}^{\mathrm{corr}}_d(\mathcal{T}_d = 1)$;

(c) $\mathbb{E}^{\mathrm{ind}}_{d+1}[Z^4] \le 36(\lambda^2 s s')^2 \mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d = 1)^2 + 13\lambda^3 s s' \mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d = 1)$;

(d) $\mathrm{Var}^{\mathrm{corr}}_{d+1}[Z] \le \lambda^2 s s'(1+ss')\mathbb{P}^{\mathrm{ind}}_d(\mathcal{T}_d = 1) + \mathbb{E}^{\mathrm{corr}}_{d+1}[Z]$.

Proof. We start by introducing centered versions of the Poisson variables N_τ and $N'_{\tau'}$, namely

$$\widetilde{N}_\tau := N_\tau - \lambda s \,\mathrm{GW}_d^{(\lambda s)}(\tau) \quad \text{and} \quad \widetilde{N}'_{\tau'} := N'_{\tau'} - \lambda s' \,\mathrm{GW}_d^{(\lambda s')}(\tau'),$$

which lets us write

$$Z = Z(t,t') = \sum_{(\tau,\tau') :\, \mathcal{T}_d(\tau,\tau')=1} \widetilde{N}_\tau \widetilde{N}'_{\tau'}.$$

What follows are the proofs of points (a), (b), (c), and (d) in their alphabetic order.

Proof of (a). This is immediate from the independence of $(N_\tau)_\tau$ and $(N'_{\tau'})_{\tau'}$ under $\mathbb{P}^{\mathrm{ind}}_{d+1}$ and the centeredness of $\widetilde{N}_\tau$ and $\widetilde{N}'_{\tau'}$:

$$\mathbb{E}^{\mathrm{ind}}_{d+1}[Z] = \sum_{\substack{(\tau,\tau')\in\mathcal{X}_d^2,\\ \mathcal{T}_d(\tau,\tau')=1}} \mathbb{E}^{\mathrm{ind}}_{d+1}[\widetilde{N}_\tau]\mathbb{E}^{\mathrm{ind}}_{d+1}[\widetilde{N}'_{\tau'}] = 0. \qquad \blacksquare$$

Proof of (b). Recall that if $((N_\tau)_\tau, (N'_{\tau'})_{\tau'}) \sim \mathbb{P}^{\mathrm{corr}}_{d+1}$, then we can write

$$N_\tau = \Delta_\tau + \sum_{\tau'\in\mathcal{X}_d} \Gamma_{\tau,\tau'} \quad \text{and} \quad N'_{\tau'} = \Delta'_{\tau'} + \sum_{\tau\in\mathcal{X}_d} \Gamma_{\tau,\tau'},$$

where

$$\begin{aligned}
\Delta_\tau &\sim \mathrm{Poi}\big(\lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau)\big),\\
\Delta'_{\tau'} &\sim \mathrm{Poi}\big(\lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s')}(\tau')\big),\\
\Gamma_{\tau,\tau'} &\sim \mathrm{Poi}\big(\lambda s s' \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau')\big)
\end{aligned}$$

are independent Poisson variables. For each one of these variables, we define their centered versions

$$\widetilde{\Delta}_\tau := \Delta_\tau - \lambda s(1-s')\,\mathrm{GW}_d^{(\lambda s)}(\tau),$$

$$\widetilde{\Delta}'_{\tau'} := \Delta'_{\tau'} - \lambda s'(1-s)\,\mathrm{GW}_d^{(\lambda s)}(\tau),$$
$$\widetilde{\Gamma}_{\tau,\tau'} := \Gamma_{\tau,\tau'} - \lambda s s' \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau'),$$

which fulfill $\widetilde{N}_\tau = \widetilde{\Delta}_\tau + \sum_{\tau' \in \mathcal{X}_d} \widetilde{\Gamma}_{\tau,\tau'}$ and $\widetilde{N}'_{\tau'} = \widetilde{\Delta}'_{\tau'} + \sum_{\tau \in \mathcal{X}_d} \widetilde{\Gamma}_{\tau,\tau'}$, because of the marginals

$$\sum_{\tau'} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau') = \mathrm{GW}_d^{(\lambda s)}(\tau) \quad \text{and} \quad \sum_{\tau'} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau') = \mathrm{GW}_d^{(\lambda s')}(\tau').$$

Exploiting the independence of Δ_τ, $\Delta'_{\tau'}$, and $\Gamma_{\tau,\tau'}$, we conclude the proof of (b) with the following computation:

$$\begin{aligned}
\mathbb{E}_{d+1}^{\mathrm{corr}}[Z] &= \sum_{\substack{(\tau,\tau')\in\mathcal{X}_d^2,\\ \mathcal{T}_d(\tau,\tau')=1}} \mathbb{E}_{d+1}^{\mathrm{corr}}\big[\widetilde{N}_\tau \widetilde{N}'_{\tau'}\big] \\
&= \sum_{\substack{(\tau,\tau')\in\mathcal{X}_d^2,\\ \mathcal{T}_d(\tau,\tau')=1}} \mathbb{E}_{d+1}^{\mathrm{corr}}\big[(\widetilde{\Gamma}_{\tau,\tau'})^2\big] = \lambda s s' \mathbb{P}_d^{\mathrm{corr}}\big(\mathcal{T}_d(\tau,\tau')=1\big). \qquad \blacksquare
\end{aligned}$$

Proof of (c). We introduce the shortcut notation $\mathcal{S} := \{(\tau,\tau') \in \mathcal{X}_d^2 \mid \mathcal{T}_d(\tau,\tau') = 1\}$ and develop the fourth moment of Z as

$$\begin{aligned}
\mathbb{E}_{d+1}^{\mathrm{ind}}[Z^4] &= \mathbb{E}_{d+1}^{\mathrm{ind}}\Bigg[\bigg(\sum_{(\tau,\tau')\in\mathcal{S}} \widetilde{N}_\tau \widetilde{N}'_{\tau'}\bigg)^4\Bigg] \\
&= \sum_{\substack{(\tau_i,\tau'_i)\in\mathcal{S},\\ i\in\{1,\dots,4\}}} \mathbb{E}_{d+1}^{\mathrm{ind}}\Bigg[\prod_{i=1}^4 \widetilde{N}_{\tau_i}\Bigg] \mathbb{E}_{d+1}^{\mathrm{ind}}\Bigg[\prod_{i=1}^4 \widetilde{N}'_{\tau'_i}\Bigg]. \qquad \text{(A.35)}
\end{aligned}$$

Since $\widetilde{N}_\tau$ is independent from $\widetilde{N}_\theta$ for $\tau \neq \theta$ and both random variables are centered, we have that $\mathbb{E}_{d+1}^{\mathrm{ind}}[\prod_{i=1}^4 \widetilde{N}_{\tau_i}] = 0$ in all but the two following cases:

- all four τ_i are equal, i.e., $|\{\tau_1,\tau_2,\tau_3,\tau_4\}| = 1$;
- the τ_i are pairwise equal but not all the same, i.e., $|\{\tau_1,\tau_2,\tau_3,\tau_4\}| = 2$.

The identical observation is true for $\mathbb{E}_{d+1}^{\mathrm{ind}}[\prod_{i=1}^4 \widetilde{N}'_{\tau'_i}]$. Let

$$\mathcal{S}(m,r) := \big\{(\tau_i,\tau'_i)_{i=1}^4 \in \mathcal{S}^4 \mid |\{\tau_1,\tau_2,\tau_3,\tau_4\}| = m \text{ and } |\{\tau'_1,\tau'_2,\tau'_3,\tau'_4\}| = r\big\},$$

and denote $\Sigma(r,m)$ the sum from (A.35) restricted to summands in $\mathcal{S}(r,m)$ instead of $\mathcal{S}$. This lets us write

$$\mathbb{E}_{d+1}^{\mathrm{ind}}[Z^4] = \Sigma(1,1) + \Sigma(1,2) + \Sigma(2,1) + \Sigma(2,2),$$

and we are left with examining the terms $\Sigma(r,m)$ one by one.

Recall that for a Poisson variable $X \sim \mathrm{Poi}(\mu)$, the fourth moment of its centered version $\widetilde{X} = X - \mu$ is equal to $\mathbb{E}[(\widetilde{X})^4] = 3\mu^2 + \mu$. This lets us compute

$$\begin{aligned}
\Sigma(1,1) &= \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{ind}}_{d+1}\big[(\widetilde{N}_\tau)^4\big]\mathbb{E}^{\mathrm{ind}}_{d+1}\big[(\widetilde{N}'_{\tau'})^4\big] \\
&= \sum_{(\tau,\tau')\in\mathcal{S}} \big(3(\lambda s)^2\,\mathrm{GW}^{(\lambda s)}_d(\tau)^2 + \lambda s\,\mathrm{GW}^{(\lambda s)}_d(\tau)\big) \\
&\qquad \times \big(3(\lambda s')^2\,\mathrm{GW}^{(\lambda s')}_d(\tau')^2 + \lambda s'\,\mathrm{GW}^{(\lambda s')}_d(\tau')\big) \\
&\le \lambda^2 s s' \sum_{(\tau,\tau')\in\mathcal{S}} 9\lambda^2 s s'\,\mathrm{GW}^{(\lambda s)}_d(\tau)^2\,\mathrm{GW}^{(\lambda s')}_d(\tau')^2 \\
&\qquad + (3\lambda s + 3\lambda s' + 1)\,\mathrm{GW}^{(\lambda s)}_d(\tau)\,\mathrm{GW}^{(\lambda s')}_d(\tau') \\
&\le 9(\lambda^2 s s')^2 \mathbb{P}^{\mathrm{ind}}_d(\mathcal{S})^2 + 7\lambda^3 s s' \mathbb{P}^{\mathrm{ind}}_d(\mathcal{S}),
\end{aligned}$$

where we have used $\mathrm{GW}^{(\mu)}_d(\tau) \le 1$ in the first inequality. The second inequality relies on

$$\sum_{i=0}^{\infty} p_i^2 \le \Big(\sum_i p_i\Big)^2$$

for $p_i \ge 0$, as well as $\lambda \ge 1$ and

$$\mathbb{P}^{\mathrm{ind}}_d(\mathcal{S}) = \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}^{\mathrm{ind}}_d(\tau,\tau') = \sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}^{(\lambda s)}_d(\tau)\,\mathrm{GW}^{(\lambda s')}_d(\tau').$$

Next, we compute

$$\begin{aligned}
\Sigma(1,2) &= \sum_{\tau\in\mathcal{X}_d} \mathbb{E}^{\mathrm{ind}}_{d+1}\big[(\widetilde{N}_\tau)^4\big]\binom{4}{2} \sum_{\substack{\tau' : (\tau,\tau')\in\mathcal{S}\\ \theta' : (\tau,\theta')\in\mathcal{S}}} \mathbb{E}^{\mathrm{ind}}_{d+1}\big[(\widetilde{N}'_{\tau'})^2\big]\mathbb{E}^{\mathrm{ind}}_{d+1}\big[(\widetilde{N}'_{\theta'})^2\big]\underbrace{\mathbb{1}_{\tau'\neq\theta'}}_{\le 1} \\
&\le 3\sum_{\tau\in\mathcal{X}_d} \big(3(\lambda s)^2\,\mathrm{GW}^{(\lambda s)}_d(\tau)^2 + \lambda s\,\mathrm{GW}^{(\lambda s)}_d(\tau)\big) \\
&\qquad \times \sum_{\substack{\tau' : (\tau,\tau')\in\mathcal{S}\\ \theta' : (\tau,\theta')\in\mathcal{S}}} (\lambda s')^2\,\mathrm{GW}^{(\lambda s')}_d(\tau')\,\mathrm{GW}^{(\lambda s')}_d(\theta') \\
&= 9(\lambda^2 s s')^2 \sum_{\tau} \mathrm{GW}^{(\lambda s)}_d(\tau)^2 \sum_{\substack{\tau' : (\tau,\tau')\in\mathcal{S}\\ \theta' : (\tau,\theta')\in\mathcal{S}}} \mathrm{GW}^{(\lambda s')}_d(\tau')\,\mathrm{GW}^{(\lambda s')}_d(\theta') \\
&\qquad + 3\lambda^3 s (s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}^{(\lambda s)}_d(\tau)\,\mathrm{GW}^{(\lambda s')}_d(\tau') \sum_{\theta' : (\tau,\theta')\in\mathcal{S}} \mathrm{GW}^{(\lambda s')}_d(\theta').
\end{aligned}$$

Using

$$\sum_i p_i^2 \le \Big(\sum_i p_i\Big)$$

again, we obtain

$$
\begin{aligned}
&\sum_{\tau} \mathrm{GW}_d^{(\lambda s)}(\tau)^2 \sum_{\substack{\tau' \colon (\tau,\tau')\in\mathcal{S}\\ \theta' \colon (\tau,\theta')\in\mathcal{S}}} \mathrm{GW}_d^{(\lambda s')}(\tau')\,\mathrm{GW}_d^{(\lambda s')}(\theta') \\
&\quad= \sum_{\tau} \sum_{\tau' \colon (\tau,\tau')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau') \sum_{\theta' \colon (\tau,\theta')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\theta') \\
&\quad= \sum_{\tau} \Bigg(\sum_{\tau' \colon (\tau,\tau')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau') \Bigg)^2 \\
&\quad\le \Bigg(\sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau') \Bigg)^2 = \mathbb{P}_d^{\mathrm{ind}}(\mathcal{S})^2.
\end{aligned}
$$

Plugging this into our earlier expression for $\Sigma(1,2)$ and using $s' \le 1$ as well as the inequality

$$
\sum_{\theta' \colon (\tau,\theta')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s')}(\theta') \le 1,
$$

one obtains

$$
\Sigma(1,2) \le 9(\lambda^2 s s')^2 \mathbb{P}_d^{\mathrm{ind}}(\mathcal{S})^2 + 3\lambda^3 s s' \mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}).
$$

Since this last term is invariant to swapping s and s', an analog computation yields the same upper bound for $\Sigma(2,1)$.

The final term to examine is $\Sigma(2,2)$. Since every combination of τ_i's we sum over fulfills $\{\tau_i \mid i \in [4]\} = \{\tau, \theta\}$ for $\tau \ne \theta$, there are $\binom{4}{2} = 3$ choices for (τ,θ). The same thing holds for the summation over τ_i'. Consequently,

$$
\begin{aligned}
\Sigma(2,2) &= \binom{4}{2} \sum_{\substack{\tau,\theta\in\mathcal{X}_d,\\ \tau\ne\theta}} \mathbb{E}_{d+1}^{\mathrm{ind}}\big[(\widetilde{N}_\tau)^2\big]\mathbb{E}_{d+1}^{\mathrm{ind}}\big[(\widetilde{N}_\theta)^2\big] \\
&\qquad \times \binom{4}{2} \sum_{\substack{\tau' \colon (\tau,\tau')\in\mathcal{S}\\ \theta' \colon (\tau,\theta')\in\mathcal{S},\\ \tau'\ne\theta'}} \mathbb{E}_{d+1}^{\mathrm{ind}}\big[(\widetilde{N}'_{\tau'})^2\big]\mathbb{E}_{d+1}^{\mathrm{ind}}\big[(\widetilde{N}'_{\theta'})^2\big] \\
&\le 9 \sum_{\tau,\theta\in\mathcal{X}_d} (\lambda s)^2\,\mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s)}(\theta) \\
&\qquad \times \sum_{\substack{\tau' \colon (\tau,\tau')\in\mathcal{S}\\ \theta' \colon (\tau,\theta')\in\mathcal{S}}} (\lambda s')^2\,\mathrm{GW}_d^{(\lambda s')}(\tau')\,\mathrm{GW}_d^{(\lambda s')}(\theta') \\
&= 9(\lambda^2 s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s)}(\tau)\,\mathrm{GW}_d^{(\lambda s')}(\tau') \sum_{(\theta,\theta')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s')}(\theta')\,\mathrm{GW}_d^{(\lambda s)}(\theta) \\
&= 9(\lambda^2 s s')^2 \mathbb{P}_d^{\mathrm{ind}}(\mathcal{S})^2.
\end{aligned}
$$

Putting all things together, we obtain the desired inequality:

$$\mathbb{E}^{\mathrm{ind}}_{d+1}[Z^4] = \sum_{r,m=1}^{2} \Sigma(r,m) \le 36(\lambda^2 s s')^2\, \mathbb{P}^{\mathrm{ind}}_d(\mathcal{S})^2 + 13\lambda^3 s s' \mathbb{P}^{\mathrm{ind}}_d(\mathcal{S}). \qquad \blacksquare$$

Proof of (d). Reusing the notations from the proof of (b), we can write

$$\begin{aligned}\mathbb{E}^{\mathrm{corr}}_{d+1}[Z^2] = \sum_{\substack{(\tau,\tau')\in\mathcal{S},\\ (\theta,\theta')\in\mathcal{S}}} \mathbb{E}^{\mathrm{corr}}_{d+1}\Bigg[&\Big(\widetilde{\Delta}_\tau + \sum_{\beta'} \widetilde{\Gamma}_{\tau,\beta'}\Big)\Big(\widetilde{\Delta}'_{\tau'} + \sum_{\beta} \widetilde{\Gamma}_{\beta,\tau'}\Big)\\ &\times \Big(\widetilde{\Delta}_\theta + \sum_{\gamma'} \widetilde{\Gamma}_{\theta,\gamma'}\Big)\Big(\widetilde{\Delta}'_{\theta'} + \sum_{\gamma} \widetilde{\Gamma}_{\gamma,\theta'}\Big)\Bigg].\end{aligned}$$

The product inside the expectation expands into a polynomial in the variables $(\widetilde{\Delta}, \widetilde{\Delta}', \Gamma)$ indexed accordingly by trees. Since these random variables are independent and centered, the only monomes not being canceled by the expectation are of the following forms:

- $\widetilde{\Delta}_\tau^2(\widetilde{\Delta}'_{\tau'})^2$; denote the sum over these monomes as $\Sigma(2,2,0)$,
- $\widetilde{\Delta}_\tau^2(\sum_\beta \widetilde{\Gamma}_{\beta,\tau'})^2$; denote the sum over these monomes as $\Sigma(2,0,2)$,
- $(\widetilde{\Delta}'_{\tau'})^2(\sum_{\gamma'} \widetilde{\Gamma}_{\tau,\gamma'})^2$; denote the sum over these monomes as $\Sigma(0,2,2)$,
- $(\sum_\beta \widetilde{\Gamma}_{\beta,\tau'})^2(\sum_{\gamma'} \widetilde{\Gamma}_{\tau,\gamma'})^2$; denote the sum over these monomes as $\Sigma(0,0,4)$.

As in the proof of (c), we compute an upper bound of these terms one by one, starting with

$$\begin{aligned}\Sigma(2,2,0) &= \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Delta}_\tau^2]\mathbb{E}^{\mathrm{corr}}_{d+1}\big[(\widetilde{\Delta}'_{\tau'})^2\big]\\ &= \lambda^2 s s'(1-s)(1-s') \underbrace{\sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}^{(\lambda s)}_d(\tau)\ \mathrm{GW}^{(\lambda s')}_d(\tau')}_{=\mathbb{P}^{\mathrm{ind}}_d(\mathcal{S})}.\end{aligned}$$

Next,

$$\Sigma(2,0,2) = \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Delta}_\tau^2]\mathbb{E}^{\mathrm{corr}}_{d+1}\Bigg[\Big(\sum_\beta \widetilde{\Gamma}_{\beta,\tau'}\Big)^2\Bigg] = \lambda^2 s^2 s'(1-s')\, \mathbb{P}^{\mathrm{ind}}_d(\mathcal{S}),$$

and similarly

$$\Sigma(0,2,2) = \lambda^2 (s')^2 s(1-s)\mathbb{P}^{\mathrm{ind}}_d(\mathcal{S}).$$

The more involved term is

$$\Sigma(0,0,4) = \sum_{\substack{(\tau,\tau')\in\mathcal{S},\\ (\theta,\theta')\in\mathcal{S}}} \sum_{\substack{\beta,\beta'\in\mathcal{X}_d,\\ \gamma,\gamma'\in\mathcal{X}_d}} \mathbb{E}^{\mathrm{corr}}_{d+1}\big[\widetilde{\Gamma}_{\tau,\beta'}\widetilde{\Gamma}_{\beta,\tau'}\widetilde{\Gamma}_{\theta,\gamma'}\widetilde{\Gamma}_{\gamma,\theta'}\big]$$

$$
\begin{aligned}
&= \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^4_{\tau,\tau'}] + \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\tau,\tau'}] \sum_{\substack{(\theta,\theta')\in\mathcal{S},\\(\theta,\theta')\neq(\tau,\tau')}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\theta,\theta'}] \\
&\quad + \sum_{\substack{(\tau,\tau')\in\mathcal{S}\\ \beta,\beta'\in\mathcal{X}_d,\\ (\beta,\tau')\neq(\tau,\beta')}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\tau,\beta'}]\mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\beta,\tau'}] \\
&\quad + \sum_{\substack{(\tau,\tau')\in\mathcal{S},\\ (\theta,\theta')\in\mathcal{S},\\ (\tau,\theta')\neq(\theta,\tau')}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\tau,\theta'}]\mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\theta,\tau'}].
\end{aligned}
$$

Recalling $\Gamma_{\tau,\tau'} \sim \mathrm{Poi}(\lambda s s' \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau'))$, we compute an upper bound for the summands one by one. First, one has

$$
\begin{aligned}
\sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^4_{\tau,\tau'}] &= \sum_{(\tau,\tau')\in\mathcal{S}} 3(\lambda s s')^2 \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau')^2 + \lambda s s'\, \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau') \\
&= (\lambda s s')\mathbb{P}^{\mathrm{corr}}_d(\mathcal{S}) + 3(\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau')^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\tau,\tau'}] \sum_{\substack{(\theta,\theta')\in\mathcal{S},\\(\theta,\theta')\neq(\tau,\tau')}} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\theta,\theta'}] \\
&\qquad = (\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau') \sum_{\substack{(\theta,\theta')\in\mathcal{S},\\(\theta,\theta')\neq(\tau,\tau')}} \mathbb{P}^{\mathrm{corr}}_d(\theta,\theta') \\
&\qquad = (\lambda s s')^2 \left(\mathbb{P}^{\mathrm{corr}}_d(\mathcal{S})^2 - \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}^{\mathrm{corr}}_d(\tau,\tau')^2 \right).
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta,\beta'\in\mathcal{X}_d} \mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\tau,\beta'}]\mathbb{E}^{\mathrm{corr}}_{d+1}[\widetilde{\Gamma}^2_{\beta,\tau'}]\, \mathbb{1}_{(\beta,\tau')\neq(\tau,\beta')} \\
&\quad = (\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta\in\mathcal{X}_d} \mathbb{P}^{\mathrm{corr}}_d(\beta,\tau') \sum_{\beta'\in\mathcal{X}_d} \mathbb{P}^{\mathrm{corr}}_d(\tau,\beta')\mathbb{1}_{(\beta,\tau')\neq(\tau,\beta')} \\
&\quad = (\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta\in\mathcal{X}_d} \mathbb{P}^{\mathrm{corr}}_d(\beta,\tau') \left(-\mathbb{P}^{\mathrm{corr}}_d(\beta,\tau') + \sum_{\beta'} \mathbb{P}^{\mathrm{corr}}_d(\tau,\beta') \right) \\
&\quad = -(\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta\in\mathcal{X}_d} \mathbb{P}^{\mathrm{corr}}_d(\beta,\tau')^2 \\
&\qquad + (\lambda s s')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta\in\mathcal{X}_d} \mathbb{P}^{\mathrm{corr}}_d(\beta,\tau')\, \mathrm{GW}^{(\lambda s')}_d(\tau')
\end{aligned}
$$

$$\leq -(\lambda ss')^2 \sum_{(\beta,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\beta,\tau')^2 + (\lambda ss')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathrm{GW}_d^{(\lambda s')}(\tau')\,{}_d^{(\lambda s)}(\tau)$$
$$= (\lambda ss')^2\Bigg(\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) - \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')^2\Bigg),$$

where we have used

$$\sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\beta\in\mathcal{X}_d} \mathbb{P}_d^{\mathrm{corr}}(\beta,\tau')^2 \geq \sum_{(\beta,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\beta,\tau')^2.$$

This inequality can be recycled for the last summand: Using $\mathcal{S} \subset \mathcal{X}_d^2$ for a first upper bound, followed by the above inequality where β is replaced by θ, one obtains

$$\sum_{(\tau,\tau')\in\mathcal{S}} \sum_{(\theta,\theta')\in\mathcal{S}} \mathbb{E}_{d+1}^{\mathrm{corr}}[\widetilde{\Gamma}_{\tau,\theta'}^2]\mathbb{E}_{d+1}^{\mathrm{corr}}[\widetilde{\Gamma}_{\theta,\tau'}^2]\mathbb{1}_{(\tau,\theta')\neq(\theta,\tau')}$$
$$\leq \sum_{(\tau,\tau')\in\mathcal{S}} \sum_{\theta,\theta'\in\mathcal{X}_d} \mathbb{E}_{d+1}^{\mathrm{corr}}[\widetilde{\Gamma}_{\tau,\theta'}^2]\mathbb{E}_{d+1}^{\mathrm{corr}}[\widetilde{\Gamma}_{\theta,\tau'}^2]\mathbb{1}_{(\theta,\tau')\neq(\tau,\theta')}$$
$$\leq (\lambda ss')^2\Bigg(\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) - \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')^2\Bigg).$$

Putting all the information on the summands together, we obtain

$$\begin{aligned}\Sigma(0,0,4) \leq{}& 3(\lambda ss')^2 \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')^2 \\ &+ (\lambda ss')^2\Bigg(\mathbb{P}_d^{\mathrm{corr}}(\mathcal{S})^2 - \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')^2\Bigg) \\ &+ 2(\lambda ss')^2\Bigg(\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) - \sum_{(\tau,\tau')\in\mathcal{S}} \mathbb{P}_d^{\mathrm{corr}}(\tau,\tau')^2\Bigg) + (\lambda ss')\mathbb{P}_d^{\mathrm{corr}}(\mathcal{S}) \\ ={}& (\lambda ss')^2\mathbb{P}_d^{\mathrm{corr}}(\mathcal{S})^2 + 2(\lambda ss')^2\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) + (\lambda ss')\mathbb{P}_d^{\mathrm{corr}}(\mathcal{S}) \\ ={}& 2(\lambda ss')^2\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) + \mathbb{E}_{d+1}^{\mathrm{corr}}[Z]^2 + \mathbb{E}_{d+1}^{\mathrm{corr}}[Z],\end{aligned}$$

where the last equality follows from (b). This lets us conclude

$$\begin{aligned}\mathbb{E}_{d+1}^{\mathrm{corr}}[Z^2] &= \Sigma(2,2,0) + \Sigma(2,0,2) + \Sigma(0,2,2) + \Sigma(0,0,4) \\ &\leq \lambda^2 ss'(1+ss')\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) + \mathbb{E}_{d+1}^{\mathrm{corr}}[Z]^2 + \mathbb{E}_{d+1}^{\mathrm{corr}}[Z],\end{aligned}$$

where the factor multiplying $\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S})$ is obtained from

$$\begin{aligned}&\lambda^2 ss'(1-s)(1-s') + \lambda^2(s')^2 s(1-s) \\ &\quad + \lambda^2 s^2 s'(1-s') + 2\lambda^2 s^2(s')^2 = \lambda^2 ss'(1+ss'),\end{aligned}$$

and (d) follows from

$$\mathrm{Var}_{d+1}^{\mathrm{corr}}[Z] = \mathbb{E}_{d+1}^{\mathrm{corr}}[Z^2] - \mathbb{E}_{d+1}^{\mathrm{corr}}[Z]^2 \leq \lambda^2 ss'(1+ss')\mathbb{P}_d^{\mathrm{ind}}(\mathcal{S}) + \mathbb{E}_{d+1}^{\mathrm{corr}}[Z]. \qquad \blacksquare$$

This completes the proof of Lemma 4.17. ■

Acknowledgments. We thank Luca Ganassali for insightful and enriching discussions, which helped us clarify the most intricate details. We also express our gratitude to the two anonymous reviewers for their very detailed and knowledgeable feedback.

Funding. This work was partially supported by the French government through the France 2030 program, managed by the French National Research Agency (ANR), under grant ANR-23-IACL-0008, PR[AI]RIE-PSAI, Paris School of Artificial Intelligence.

References

[1] R. Abraham and J.-F. Delmas, An introduction to Galton–Watson trees and their local limits. [v1] 2015, [v3] 2024, arXiv:1506.05571v3

[2] A. E. Aladağ and C. Erten, SPINAL: Scalable protein interaction network alignment. *Bioinformatics* **29** (2013), no. 7, 917–924

[3] N. Alon and J. H. Spencer, *The probabilistic method*. Fourth edn., Wiley Ser. Discrete Math. Optim., John Wiley, Hoboken, NJ, 2016 Zbl 1333.05001 MR 3524748

[4] E. Araya and H. Tyagi, Graph matching via convex relaxation to the simplex. *Found. Data Sci.* **7** (2025), no. 2, 464–501 Zbl 1574.62131 MR 4853283

[5] L. Babai, Group, graphs, algorithms: The graph isomorphism problem. In *Proceedings of the International Congress of Mathematicians. Rio de Janeiro, 2018. Vol. IV. Invited lectures*, pp. 3319–3336, World Scientific, Hackensack, NJ, 2018 Zbl 1490.68116 MR 3966534

[6] L. Babai, P. Erdős, and S. M. Selkow, Random graph isomorphism. *SIAM J. Comput.* **9** (1980), no. 3, 628–635 Zbl 0454.05038 MR 0584517

[7] A. Berg, T. Berg, and J. Malik, Shape matching and object recognition using low distortion correspondences. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'05. Vol. 1*, pp. 26–33, IEEE, 2005

[8] P. Billingsley, *Convergence of probability measures*. Second edn., Wiley Ser. Probab. Stat., John Wiley, New York, NY, 1999 Zbl 0944.60003 MR 1700749

[9] B. Bollobás, *Random graphs*. Second edn., Fifth print edition, Cambridge Stud. Adv. Math. 73, Cambridge University Press, Cambridge, 2008 Zbl 0979.05003 MR 1864966

[10] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** (2007), no. 1, 3–122 Zbl 1123.05083 MR 2337396

[11] M. Brocheler, A. Pugliese, and V. Subrahmanian, Probabilistic subgraph matching on huge social networks. In *2011 international conference on advances in social networks analysis and mining*, pp. 271–278, IEEE, 2011

[12] E. Çela, *The quadratic assignment problem*. Comb. Optim. 1, Kluwer Academic, Dordrecht, 1998 Zbl 0909.90226 MR 1490831

[13] S. Chen, S. Jiang, Z. Ma, G. P. Nolan, and B. Zhu, One-way matching of datasets with low rank signals. 2022, arXiv:2204.13858v2

[14] F. Chung and L. Lu, Concentration inequalities and martingale inequalities: A survey. *Internet Math.* **3** (2006), no. 1, 79–127 Zbl 1111.60010 MR 2283885

[15] D. Conte, P. Foggia, C. Sansone, and M. Vento, Thirty years of graph matching in pattern recognition. *nt. J. Pattern Recognit. Artif. Intell.* **18** (2004), no. 03, 265–298

[16] T. Cour, P. Srinivasan, and J. Shi, Balanced graph matching. In *Advances in Neural Information Processing Systems 19*, pp. 313–320, MIT Press, Cambridge, MA, 2007

[17] D. Cullina and N. Kiyavash, Improved achievability and converse bounds for Erdős–Rényi graph matching. *ACM SIGMETRICS Performance Evaluation Review* **44** (2016), no. 1, 63–72

[18] D. Cullina and N. Kiyavash, Exact alignment recovery for correlated Erdős–Rényi graphs. [v1] 2017, [v2] 2018, arXiv:1711.06783v2

[19] D. Cullina, N. Kiyavash, P. Mittal, and H. V. Poor, Partial recovery of Erdős–Rényi graph alignment via k-core alignment. *ACM SIGMETRICS Performance Evaluation Review* **48** (2020), no. 1, 99–100

[20] J. Ding and H. Du, Matching recovery threshold for correlated random graphs. *Ann. Statist.* **51** (2023), no. 4, 1718–1743 Zbl 1533.05244 MR 4658574

[21] J. Ding and Z. Li, A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. [v1] 2023, [v2] 2024, arXiv:2306.00266v2

[22] J. Ding, Z. Ma, Y. Wu, and J. Xu, Efficient random graph matching via degree profiles. *Probab. Theory Related Fields* **179** (2021), no. 1-2, 29–115 Zbl 1460.05171 MR 4221654

[23] W. Fan, Graph pattern matching revised for social network analysis. In *Proceedings of the 15th International Conference on Database Theory*, pp. 8–21, Association for Computing Machinery, NY, New York, 2011 Zbl 1359.68007 MR 3476899

[24] Z. Fan, C. Mao, Y. Wu, and J. Xu, Spectral graph matching and regularized quadratic relaxations. II. *Found. Comput. Math.* **23** (2023), no. 5, 1567–1617 Zbl 1522.90092 MR 4649431

[25] L. Ganassali and L. Massoulié, From tree matching to sparse graph alignment. In *Proceedings of Thirty Third Conference on Learning Theory*, pp. 1633–1665, Proceedings of Machine Learning Research 125, PMLR, 2020

[26] L. Ganassali, L. Massoulié, and M. Lelarge, Impossibility of partial recovery in the graph alignment problem. In *Proceedings of Thirty Fourth Conference on Learning Theory*, pp. 2080–2102, Proceedings of Machine Learning Research 134, PMLR, 2021

[27] L. Ganassali, L. Massoulié, and M. Lelarge, Correlation detection in trees for planted graph alignment. *Ann. Appl. Probab.* **34** (2024), no. 3, 2799–2843

[28] L. Ganassali, L. Massoulié, and G. Semerjian, Statistical limits of correlation detection in trees. *Ann. Appl. Probab.* **34** (2024), no. 4, 3701–3734 Zbl 1548.05297 MR 4783028

[29] A. Haghighi, A. Y. Ng, and C. D. Manning, Robust textual inference via graph matching. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 387–394, Association for Computational Linguistics, 2005

[30] G. Hall and L. Massoulié, Partial recovery in the graph alignment problem. *Oper. Res.* **71** (2023), no. 1, 259–272 Zbl 1540.05122 MR 4585566

[31] M. Hallin, Neyman–Pearson lemma. In *Encyclopedia of Environmetrics*, John Wiley, 2006

[32] L. P. Hansen, Likelihood processes, 2019, http://larspeterhansen.org/wp-content/uploads/2019/02/Chapter-8.pdf, visited on 29 April 2026

[33] D. Huang, X. Song, and P. Yang, Information-theoretic thresholds for the alignments of partially correlated graphs. *IEEE Trans. Inform. Theory* **71** (2025), no. 12, 9674–9697 Zbl 08173783 MR 4999707

[34] D. J. C. MacKay, *Information theory, inference and learning algorithms*. Cambridge University Press, New York, 2003 Zbl 1055.94001 MR 2012999

[35] J. Maier and L. Massoulié, Asymmetric tree correlation testing for graph alignment. In *2023 IEEE Information Theory Workshop. ITW*, pp. 503–508, IEEE, 2023

[36] C. Mao, M. Rudelson, and K. Tikhomirov, Exact matching of random graphs with constant correlation. *Probab. Theory Related Fields* **186** (2023), no. 1-2, 327–389 Zbl 1512.05361 MR 4586222

[37] C. Mao, Y. Wu, J. Xu, and S. H. Yu, Random graph matching at Otter's threshold via counting chandeliers. In *STOC'23. Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pp. 1345–1356, Association for Computing Machinery, New York, 2023 Zbl 07844676 MR 4617472

[38] Math StackExchange, Derivation of mutual information's closed-form analytical solution, 2024, https://math.stackexchange.com/questions/3914899/, visited on 29 April 2026

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

[40] A. Narayanan and V. Shmatikov, De-anonymizing social networks. In *2009 30th IEEE Symposium on Security and Privacy*, pp. 173–187, IEEE, 2009

[41] R. Otter, The number of trees. *Ann. of Math. (2)* **49** (1948), 583–599 Zbl 0032.12601 MR 0025715

[42] P. Pedarsani and M. Grossglauser, On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1235–1243, Association for Computing Machinery, 2011

[43] C. Schellewald and C. Schnörr, Probabilistic subgraph matching based on convex relaxation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition. EMMCVPR 2005*, pp. 171–186, Lect. Notes Comput. Sci. 3757, Springer, Berlin, 2005

[44] T. Séjourné, F.-X. Vialard, and G. Peyré, The unbalanced Gromov Wasserstein distance: Conic formulation and relaxation. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, pp. 8766–8779, Curran Associates, Red Hook, NY, 2021

[45] R. Singh, J. Xu, and B. Berger, Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. U.S.A.* **105** (2008), no. 35, 12763–12768

[46] J. Soch, Proof: Invariance of the Kullback–Leibler divergence under parameter transformation. StatProofBook, 2020, https://statproofbook.github.io/P/kl-inv.html, visited on 29 April 2026

[47] A. Thual, Q. H. Tran, T. Zemskova, N. Courty, R. Flamary, S. Dehaene, and B. Thirion, Aligning individual brains with fused unbalanced Gromov Wasserstein. In *Advances in Neural Information Processing Systems 35* pp. 21792–21804, Curran Associates, Red Hook, NY, 2022

[48] C. Villani, *Optimal transport: Old and new*. Grundlehren Math. Wiss. 338, Springer, Berlin, 2009 Zbl 1156.53003 MR 2459454

[49] H. W. Watson and F. Galton, On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland* **4** (1875), 138–144

[50] Y. Wu, J. Xu, and S. H. Yu, Settling the sharp reconstruction thresholds of random graph matching. *IEEE Trans. Inform. Theory* **68** (2022), no. 8, 5391–5417 Zbl 1505.60015 MR 4476403

[51] L. Yartseva and M. Grossglauser, On the performance of percolation graph matching. In *Proceedings of the First ACM Conference on Online Social Networks*, pp. 119–130, Association for Computing Machinery, New York, NY, 2013

[52] M. Zaslavskiy, F. Bach, and J.-P. Vert, A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31** (2008), no. 12, 2227–2242

Received 22 July 2025; revised 14 January 2026.

Jakob Maier
INRIA Paris, ENS - PSL Research University, 48, rue Barrault, 75013 Paris, France;
jakob.maier@inria.fr
Author IDs: ORCID 0000-0002-7697-8457

Laurent Massoulié
INRIA Paris, ENS - PSL Research University, 48, rue Barrault, 75013 Paris, France;
laurent.massoulie@inria.fr
Author IDs: zbMATH massoulie.laurent MR 354412 ORCID 0000-0001-7263-0069