# ALMOST ALL TREES ARE ALMOST GRACEFUL

ANNA ADAMASZEK, PETER ALLEN, CODRUŢ GROSU, AND JAN HLADKÝ

ABSTRACT. The Graceful Tree Conjecture of Rosa from 1967 asserts that the vertices of each tree $T$ of order $n$ can be injectively labelled by using the numbers $\{1, 2, \ldots, n\}$ in such a way that the absolute differences induced on the edges are pairwise distinct.

We prove the following relaxation of the conjecture for each $\gamma > 0$ and for all $n > n_0(\gamma)$. Suppose that (i) the maximum degree of $T$ is bounded by $O_\gamma(n/\log n)$, and (ii) the vertex labels are chosen from the set $\{1, 2, \ldots, \lceil (1 + \gamma)n \rceil\}$. Then there is an injective labelling of $V(T)$ such that the absolute differences on the edges are pairwise distinct. In particular, asymptotically almost all trees on $n$ vertices admit such a labelling.

The proof proceeds by showing that a certain very natural randomized algorithm produces a desired labelling with high probability.

## 1. INTRODUCTION

1.1. **Graceful labelling.** Let $G$ be a graph with $n$ vertices and $q$ edges. A *vertex labelling* of $G$ is an assignment of natural numbers to the vertices of $G$, subject to certain conditions. A vertex labelling $f$ is called *graceful* if $f$ is an injection from $V(G)$ to the set $\{1, \ldots, q+1\}$ such that, if each edge $xy \in E(G)$ is assigned the *induced* label $|f(x) - f(y)|$, then the resulting edge labels are distinct. The graph $G$ is called *graceful* if it admits a graceful labelling.

Graceful labellings were first introduced by Rosa [21] under the name of $\beta$-*valuations*. It was Golomb [10] who used the term *graceful* for the first time.

A natural problem associated with graceful labellings is to determine which graphs are graceful. According to an unpublished result of Erdős almost all graphs are not graceful. A version of this argument for a very similar concept of the so-called harmonious labellings, which we introduce in detail in Section 5.3, was later recorded by Graham and Sloane [11]. It was shown by Rosa [21] that a graph with every vertex of even degree and number of edges congruent to 1 or 2 (mod 4) is not graceful. Nevertheless, it appears that many graphs that exhibit some regularity in structure are graceful. For example, paths $P_n$ and wheels $W_n$ are graceful, [21], [8]. A comprehensive survey on the current status of knowledge on graceful graphs can be found in [9].

The most important problem in the area is to determine whether every tree is graceful. Conjectured first by Rosa in 1967, the problem remains wide open.

**Conjecture 1** (Graceful Tree Conjecture). *For any $n$-vertex tree $T$ there exists an injective labelling $\psi : V(T) \to \{1, \ldots, n\}$ that induces pairwise distinct labels on the edges of $T$.*

While Conjecture 1 has attracted a lot of attention, it was proved only for some special classes of trees (paths and caterpillars [21], firecrackers [5], banana trees [22], olive trees [19], trees of diameter at most 7 [23], and several others, see Table 4 in [9]). The related conjecture of Bermond [2] that all lobsters are graceful is still open, and has also been the subject of many papers.

Van Bussel [4] introduced the following relaxation of gracefulness. A map $\psi : V(G) \to [m]$ from a vertex set of a graph is *$m$-graceful* if $\psi$ is injective, and the map $\psi_*$ induced on the edges, $\psi_* : E(G) \to [m-1]$, $\psi_*(xy) := |\psi(x) - \psi(y)|$, is injective as well. Thus, to obtain the original (nonparametric) definition, we take $m = |E(G)| + 1$. If the codomain $[m]$ of an $m$-graceful map is clear from the context, we simply call $\psi$ *graceful*. Van Bussel proved that every tree $T$ admits a $\bigl(2v(T) - 2\bigr)$-graceful labelling.

1.2. **Tree packings.** The motivation for considering graceful labellings comes from the area of graph packings. A collection $G_1, \ldots, G_t$ of graphs *pack* into a graph $H$ if there are embeddings $\psi_1, \ldots, \psi_t$ of $G_1 \ldots, G_t$ into $H$ such that each edge of $H$ is used in at most one embedding. If in addition each edge of $H$ is used in some embedding, we say that $G_1, \ldots, G_t$ *decompose $H$*.

There are several open problems in the area of graph packings. These are special cases of the following general meta-conjecture: if $G_1, \ldots, G_t$ are drawn from a family of 'sparse' graphs, and $H$ is 'dense', then provided some 'simple' necessary conditions are satisfied, $G_1, \ldots, G_t$ pack into $H$.

In particular, we can consider the (sparse) family of trees and the (dense) complete graph. Even here there are several incarnations of the meta-conjecture. The one which will concern us is Ringel's conjecture from 1963.

**Conjecture 2** (Ringel's conjecture, [20]). *For any $(n + 1)$-vertex tree $T$ the complete graph $K_{2n+1}$ can be decomposed into $2n + 1$ edge-disjoint subgraphs isomorphic to $T$.*

This conjecture can be strengthened by requiring the embeddings to have a special structure. Specifically, suppose that the vertices of $K_{2n+1}$ are the integers $0, 1, \ldots, 2n$. For any subgraph $G$ of $K_{2n+1}$ we may define the *cyclic shift* of $G$ as the subgraph $S(G)$ with

$$S(G) = (\{x + 1 \, : \, x \in V(G)\}, \{(x + 1, y + 1) \, : \, (x, y) \in E(G)\})$$

where all addition is performed modulo $2n + 1$.

If $G$ is any graph with $n$ edges, we say that $K_{2n+1}$ can be *cyclically decomposed* into copies of $G$ if there is a subgraph $G' \simeq G$ of $K_{2n+1}$ such that the cyclic shifts $G', S(G'), \ldots, S^{2n}(G')$ are edge-disjoint (and thus form a decomposition of $K_{2n+1}$).

As reported by Rosa [21], the following conjecture is due to Kotzig.

**Conjecture 3** (Ringel–Kotzig conjecture). *For any $(n+1)$-vertex tree $T$ the complete graph $K_{2n+1}$ can be cyclically decomposed into copies of $T$.*

Finally, we can connect this to graceful labellings. If $T$ has a graceful labelling, then $T$ satisfies the Ringel–Kotzig conjecture. Furthermore, if $T$ has an $m$-graceful labelling, then we can cyclically pack $2m-1$ copies of $T$ into $K_{2m-1}$. Specifically, let $\psi : V(T) \to [m]$ be an $m$-graceful labelling of $T$. Then $\psi$ is also an embedding of $T$ into $K_{2m-1}$, since $\psi$ is injective. We claim the cyclic shifts of $\psi$ form a packing of $2m-1$ copies of $T$ into $K_{2m-1}$. Indeed, by symmetry we only have to check that if $uv$ is an edge of $T$, then $\psi(u)\psi(v)$ is not the image of any $u'v' \in E(T)$ under any non-trivial power of the cyclic shift $\psi'$ of $\psi$. Without loss of generality we may assume $1 \le \psi(u) < \psi(v) \le m$. If the range of $\psi'$ contains both $\psi(u)$ and $\psi(v)$ then it must contain the interval from $\psi(u)$ to $\psi(v)$. If $u'v' \in E(T)$ satisfies $\psi'(u')\psi'(v') = \psi(u)\psi(v)$, then we have

$$\left|\psi(u) - \psi(v)\right| = \left|\psi'(u') - \psi'(v')\right| = \left|\psi(u') - \psi(v')\right|,$$

where the final step uses that the range of $\psi'$ contains the interval from $\psi(u)$ to $\psi(v)$. Now since $\psi$ is graceful, we have $u'v' = uv$, and hence $\psi' = \psi$.

Conjecture 2 was wide open until recently when in [3] an approximate version for bounded degree trees was proven.[1] The main result of [3] was extended by Messuti, Rödl, and Schacht [16] to permit almost-spanning bounded degree graphs with limited expansion and later Ferber, Lee, and Mousset [6] allowed for spanning bounded degree graphs with limited expansion. Kim, Kühn, Osthus, and Tyomkyn [15] were able to obtain almost perfect decompositions of a complete graph with an arbitrary family of bounded degree graphs, and using this Joos, Kim, Kühn and Osthus [14] proved the Tree Packing Conjecture and Ringel's Conjecture exactly for arbitrary bounded degree trees. Ferber and Samotij [7] addressed the problem of packing trees with unbounded degrees, in particular packing almost-spanning trees with maximum degree $cn/\log n$ into complete graphs. However the methods used in these papers do not seem not to hint at any approaches for graceful tree labellings. After this paper was made public, Allen, Böttcher, Hladký, and Piguet [1] showed an almost perfect decomposition result for any family of possibly spanning graphs of with maximum degree $cn/\log n$ and constant degeneracy, thus improving (in the setting of complete host graphs) upon [3, 16, 6, 15]. Some ideas used in [1] are inspired by the present work. The strongest result on Conjecture 2 for general trees (with no degree restriction) is a result of Montgomery, Pokrovskiy, and Sudakov [17], who proved that $2\ell+1$ copies of any $(n+1)$-vertex tree $T$ pack into $K_{2\ell+1}$ whenever $\ell \ge n + o(n)$.[2] Actually, such a packing follows immediately by a similar

---

[1]The main result in [3] is more general and allows almost perfect decompositions of a complete graph with an arbitrary family of almost-spanning bounded degree trees, addressing also the Tree Packing Conjecture of Gyarfás.

[2]Theorem 1.3 in [17] contains a slightly weaker statement, but this what its proof actually gives.

reduction as the one given after Conjecture 3 and their main result that such a tree admits a harmonious labelling by a group $\mathbb{Z}_{n+o(n)}$.

1.3. **Our result.** In this paper we prove an approximate version of Conjecture 1 for trees with maximum degree $o\left(\frac{n}{\log n}\right)$. This implies approximate versions of the Ringel–Kotzig and Ringel conjectures.

**Theorem 4.** *For every $\gamma > 0$ there exist $\eta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds for every $n > n_0$. Suppose that $T$ is an $n$-vertex tree and $\Delta(T) \leq \frac{\eta n}{\log n}$. Then there exists a $\lceil(1+\gamma)n\rceil$-graceful labelling $\psi : V(T) \to [\lceil(1+\gamma)n\rceil]$.*

This theorem applies to random trees. More precisely, we work with the set $\mathcal{T}_n$ of all labelled trees on the vertex set $\{1, \ldots, n\}$. By a classical result of Moon [18], a tree selected uniformly at random from the set $\mathcal{T}_n$ has maximum degree $o(\log n)$ with probability tending to 1 as $n$ tends to infinity. In particular Theorem 4 applies to almost all trees.

Our proof of Theorem 4 is an application of the Probabilistic Method, inspired by [3]. More precisely, the proof is an application of the Differential Equations Method (DEM). That is, we run a certain randomized algorithm which sequentially labels the vertices of the $n$-vertex tree $T$, and we prove that with high probability this process produces a $[\lceil(1+\gamma)n\rceil]$-graceful labelling of $T$. As the algorithm progresses with the labelling, the sets of (edge- and vertex-) labels available for future steps keep getting sparser. The key for the analysis of the algorithm is to introduce suitable measures of quasirandomness for these sets, and prove that the sets of available labels evolve in a quasirandom way.

Let us note that DEM has been used extensively in discrete mathematics in particular thanks to the tools developed by Wormald [24]. We do not use Wormald's machinery and it is not clear to us whether that formalism applies in our setting at all. To get a handle on various parameters during the run of the process — this handle provided in other scenarios by DEM itself — we introduce in Section 2 a variant of Hoeffding's bound.

After introducing notation and some preliminary facts in Section 2, we outline the proof of Theorem 4 in Section 3. In Section 3, we also give a detailed description of our labelling algorithm and introduce the key quasirandomness concepts. The missing bits in our proof that the algorithm will produce a labelling required in Theorem 4 are given in Section 4. In Section 5 we suggest various strengthenings of our main result.

## 2. Notation and auxiliary results

For a graph $G$, the *order* of $G$ is the number of vertices of $G$. We write $\Delta(G)$ for the maximum degree of $G$.

We write $a = b \pm \varepsilon$ when we have $a \in [b - \varepsilon, b + \varepsilon]$. Extending this, and in a slight abuse of notation, we write $a \pm \delta = b \pm \varepsilon$ for the inclusion $[a - \delta, a + \delta] \subseteq [b - \varepsilon, b + \varepsilon]$. We write $\log$ for the natural logarithm.

We use $\mathbf{P}[\cdot]$ and $\mathbf{E}[\cdot]$ to denote the probability and the expectation, respectively. All probability spaces considered in this paper are finite. In such a setting, any sigma-algebra is generated by its inclusion-wise minimal nonempty sets, which naturally form a partition of the probability space, and it is convenient to work with the partition rather than the sigma-algebra it generates.

Recall that if $\Omega$ is a finite probability space then a sequence of partitions $(\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n)$ of $\Omega$ is a *filtration* if for each $i \in [n]$, the partition $\mathcal{F}_i$ refines $\mathcal{F}_{i-1}$. Recall that given a function $f : \Omega \to \mathbb{R}$, its *conditional expectation with respect to* $\mathcal{F}_i$, denoted by $\mathbf{E}[f|\mathcal{F}_i]$, is a function $\mathbf{E}[f|\mathcal{F}_i] : \Omega \to \mathbb{R}$ defined by $\mathbf{E}[f|\mathcal{F}_i](\omega) = \mathbf{E}[f|X]$, where $X \in \mathcal{F}_i$ is the cell containing $\omega$. Recall also that in this setting, a function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}_i$-*measurable* if $f$ is constant on each cell of $\mathcal{F}_i$. In this paper, $\Omega$ will be the probability space on which a probabilistic process is defined, the cells of $\mathcal{F}_i$ will be given by all the random choices made up to some time $t_i$ in the process, so choosing $t_i$ increasing in $i$ automatically gives a filtration. In this setting, if a function $f$ is $\mathcal{F}_i$-measurable, that means that its value is fixed by the random choices made in the process up to time $t_i$.

2.1. **Hoeffding's bound.** The following theorem of Hoeffding [12, Theorem 2] gives bounds on sums of independent real-valued bounded random variables.

**Theorem 5.** *Let* $Y_1, \ldots, Y_n$ *be independent random variables with* $0 \le Y_i \le a_i$ *for each* $i \in [n]$. *Let* $X = Y_1 + \cdots + Y_n$, *and let* $\mu = \mathbf{E}X$. *Then we have*

$$\mathbf{P}[X - \mu \ge t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n a_i^2}\right) \quad and$$
$$\mathbf{P}[X - \mu \le -t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n a_i^2}\right).$$

Theorem 5 is one of the most commonly used concentration bounds in probabilistic combinatorics. We shall need an extension of this theorem to non-independent random variables.

**Lemma 6.** *Let* $\Omega$ *be a finite probability space, and* $(\mathcal{F}_0, \ldots, \mathcal{F}_n)$ *a filtration. Suppose that for each* $1 \le i \le n$ *we have a nonnegative real number* $a_i$, *an* $\mathcal{F}_i$-*measurable random variable* $Y_i$ *satisfying* $0 \le Y_i \le a_i$, *nonnegative real numbers* $\mu$ *and* $\nu$, *and an event* $\mathcal{E}$ *with* $\mathbf{P}[\mathcal{E}] > 0$. *Suppose that almost surely, either* $\mathcal{E}$ *does not occur or* $\sum_{i=1}^n \mathbf{E}[Y_i|\mathcal{F}_{i-1}] = \mu \pm \nu$. *Then for each* $t > 0$ *we have*

$$\mathbf{P}\left[\mathcal{E} \text{ and } \left|\sum_{i=1}^n Y_i - \mu\right| \ge \nu + t\right] \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n a_i^2}\right).$$

In applications, $\mathcal{E}$ will be an event that holds with probability close to 1. Then, Lemma 6 controls large deviations of $\sum_{i=1}^n Y_i$.

*Proof.* Given $Y_1, \ldots, Y_n$, we define random variables $U_1, \ldots, U_n$ by $U_i = Y_i$ if $\mathbf{P}[\mathcal{E}|\mathcal{F}_{i-1}] > 0$, and $U_i = 0$ otherwise. Note that $U_i$ is $\mathcal{F}_i$-measurable for each $i$. Furthermore, we claim that

for each $1 \leq k \leq n$ we have almost surely

$$\sum_{i=1}^{k} \mathbf{E}[U_i|\mathcal{F}_{i-1}] \leq \mu + \nu . \tag{1}$$

Indeed, suppose $k$ is minimal such that this statement fails, and let $F$ be a set in $\mathcal{F}_{k-1}$ with $\mathbf{P}[F] > 0$ witnessing its failure. By minimality of $k$, we have $U_k > 0$, so $\mathbf{P}[\mathcal{E}|F] > 0$. Thus with positive probability, $\mathcal{E}$ occurs and $\sum_{i=1}^{n} \mathbf{E}[U_i|\mathcal{F}_{i-1}] \geq \sum_{i=1}^{k} \mathbf{E}[U_i|\mathcal{F}_{i-1}] > \mu + \nu$, contradicting the assumption of the lemma.

Set $Z_0 := 0$ and for $i = 1, \ldots, n$ set $Z_i := \sum_{j=1}^{i}(U_j - \mathbf{E}[U_j|\mathcal{F}_{j-1}])$. It is straightforward to check that $(Z_i)_{i=0}^{n}$ is a martingale which satisfies $|Z_i - Z_{i-1}| = |U_i - \mathbf{E}[U_i|\mathcal{F}_{i-1}]| \leq a_i$. Thus, Azuma's Inequality (see e.g. [13, Theorem 2.25]) gives us

$$\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} a_i^2}\right) \geq \mathbf{P}\left[Z_n - Z_0 \geq t\right] = \mathbf{P}\left[\sum_{i=1}^{n}(U_i - \mathbf{E}[U_i|\mathcal{F}_{i-1}]) \geq t\right]$$

$$\boxed{\text{by (1)}} \quad \geq \mathbf{P}\left[\sum_{i=1}^{n} U_i \geq \mu + \nu + t\right] .$$

If $\mathcal{E}$ occurs then almost surely $Y_i = U_i$ for each $1 \leq i \leq n$. Therefore, we have

$$\mathbf{P}\left[\mathcal{E} \text{ and } \sum_{i=1}^{n} Y_i \geq \mu + \nu + t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} a_i^2}\right).$$

The same argument applied to the random variables $a_i - Y_i$ gives

$$\mathbf{P}\left[\mathcal{E} \text{ and } \sum_{i=1}^{n} Y_i \leq \mu - \nu - t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} a_i^2}\right),$$

and so the lemma statement holds by the union bound.                              $\square$

2.2. **Cutting a tree.** The following lemma (variants of which are well-known) tells us that trees can be easily separated into small components.

**Lemma 7.** *For any $\varepsilon \in (0,1)$, any $n$ such that $\varepsilon n \geq 2 \log n$, and any tree $T$ with $\Delta(T) \leq \frac{\varepsilon^2 n}{4 \log n}$ and $v(T) \leq n$, there exists a set $\mathcal{R}$ of edges of $T$ with $|\mathcal{R}| \leq \varepsilon v(T)$ such that the components of $T - \mathcal{R}$ have order at most $\frac{\varepsilon n}{\log n}$.*

*Proof.* We prove the lemma by induction on $v(T)$. The statement is trivially true for $v(T) \leq \frac{\varepsilon n}{\log n}$, so we may assume $v(T) > \frac{\varepsilon n}{\log n}$. It is enough to show that there exists one edge $uv \in E(T)$ such that one of the two components of $T - uv$ has order between $2\varepsilon^{-1}$ and $\frac{\varepsilon n}{\log n}$, since then the statement follows by applying the induction hypothesis to the other component.

We find the edge $uv$ by the following 'walk' procedure. We start with a leaf vertex $u_1$ and its neighbour $v_1$. Now for each $t \geq 1$ in succession we do the following. If the component of $T - u_t v_t$ containing $v_t$ has order between $2\varepsilon^{-1}$ and $\frac{\varepsilon n}{\log n}$, we stop and return $uv = u_t v_t$. If not, we set $u_{t+1} = v_t$ and $v_{t+1}$ to be the neighbour of $u_{t+1}$ not equal to $u_t$ which maximises the order of the component of $T - u_{t+1} v_{t+1}$ containing $v_{t+1}$.

If at time $t$ the component of $T - u_t v_t$ containing $v_t$ has more than $\frac{\varepsilon n}{\log n}$ vertices, then by averaging, the component of $T - u_{t+1} v_{t+1}$ containing $v_{t+1}$ has at least

$$\tfrac{1}{\Delta(T)}\left(\tfrac{\varepsilon n}{\log n} - 1\right) \geq \tfrac{1}{\Delta(T)} \cdot \tfrac{\varepsilon n}{2 \log n} \geq 2\varepsilon^{-1}$$

vertices, where the first inequality is by choice of $n$ and the second by the bound on $\Delta(T)$. Thus the algorithm finds the desired $uv$. $\qquad\square$

## 3. Setup

Before we embark on the proof of Theorem 4, we outline the main ideas. We write $\tilde{n} = \lceil (1+\gamma)n \rceil$. We wish to label our tree in a random process, at each step labelling one new vertex which has one labelled neighbour (so that the set of labelled vertices is always a subtree), and choosing the new vertex label to be admissible, that is, to avoid re-using either the vertex label or the induced edge label. We keep track of the sets of vertex and induced edge labels which remain available at the $t^{\text{th}}$ step, for which we will write $\mathsf{A}_t$ and $\mathsf{B}_t$ respectively. We will show that $\mathsf{A}_t$ looks like a uniform random subset of $[\tilde{n}]$ of cardinality $\tilde{n} - t$ and that $\mathsf{B}_t$ looks like a uniform random subset of $[\tilde{n} - 1]$ of cardinality $\tilde{n} - t - 1$. The reason why we do this is that if $\mathsf{A}_t$ and $\mathsf{B}_t$ were really uniformly distributed it would be easy to show that at every step there is (with probability very close to 1) an admissible choice for the new label.

Choosing the new labels uniformly from all admissible labels will not lead to random-looking sets $\mathsf{A}_t$ and $\mathsf{B}_t$. Let us illustrate this in the initial situation when, starting with $\mathsf{A}_0 = [\tilde{n}]$ and $\mathsf{B}_0 = [\tilde{n} - 1]$, we label first two neighboring vertices $v_1$ and $v_2$. Then the first created edge label on $v_1 v_2$ will be assigned the smallest possible value of 1 with probability $\frac{2}{\tilde{n}}$, but the largest possible value of $\tilde{n} - 1$ with probability only $\frac{2}{\tilde{n}(\tilde{n}-1)}$, and the probabilities of labels with intermediate values interpolate these two extremes. This effect will persist, so that $\mathsf{B}_t$ will rapidly become non-uniform, being much sparser for small labels than large. That will in turn cause $\mathsf{A}_t$ to evolve in a non-uniform way as well.

But observe that if $uv$ is an edge of $T$, if we choose $\psi(u)$ a uniformly random vertex label, and then choose $\psi(v)$ randomly in a small interval around $\tilde{n} - \psi(u)$ then the distributions of each of $\psi(u)$, $\psi(v)$ and the induced edge label $\left| \psi(u) - \psi(v) \right|$ are close to the uniform distribution. Thus we would like to label our tree such that most edges and vertices are labelled in about this way.

We will label the vertices of $T$ in the order $v_1, \dots, v_n$ (which we will determine), and in each case we choose an admissible label in a certain short interval $J(v_i)$ within $[\tilde{n}]$. We choose the $J(v_i)$ to be uniformly distributed[3] over $[\tilde{n}]$, and such that for most $v_i$ the intervals $J(v_i)$ and $J(v_j)$, where $v_j$ is the parent of $v_i$ in the ordering, are equally far from and on opposite sides of $\frac{1}{2}(1 + \gamma)n$.

---

[3]A technical complication arises in that the extremes of the interval $[\tilde{n}]$ are not covered as much as the rest; we will ignore this for now.

In reality, we certainly do not choose vertex labels uniformly at random in the intervals; there is a great deal of dependency in the label choices. However, that dependency is sequential: conditioning on the labelling history of the first $t-1$ vertices, we know the distribution of the $t^{\text{th}}$ vertex, and we will be able to show that the average of these distributions is close to uniform (in a sense which we will make precise later). This enables us to apply Lemma 6, which gives us concentration results that imply our quasirandomness properties are maintained. A similar statement applies to the edge labels.

3.1. **Preparation.** We now describe the general setup which we use for proving Theorem 4. We first reduce to the following special case.

**Theorem 8.** *For every $\gamma > 0$ such that $\gamma^{-1}$ is an integer, there exist $\eta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Let $\mu = \left\lceil \exp\left(10^8 \gamma^{-4}\right)\right\rceil^{-1}$, let $\delta_0 = \mu^{2/\mu}$, and let $\delta = \delta_0^{10}$. For every $n > n_0$ divisible by $2\delta^{-1}\gamma^{-1}$, if $T$ is an $n$-vertex tree with $\Delta(T) \leq \frac{\eta n}{\log n}$, there exists a graceful labelling $\psi : V(T) \to [(1+\gamma)n]$.*

(Note that the hierarchy of constants in Theorem 8 is $\gamma > \mu > \delta_0 > \delta > \eta > 0$.)

Before proving this, we show why it implies Theorem 4.

*Proof of Theorem 4.* Given $\gamma > 0$, let $\gamma' = 1/\lceil 2\gamma^{-1}\rceil$. Let $\eta > 0$ and $n_0' \in \mathbb{N}$ be returned by Theorem 8 for input $\gamma'$. Let $\mu = \left\lceil \exp\left(10^8 \gamma'^{-4}\right)\right\rceil^{-1}$, let $\delta_0 = \mu^{2/\mu}$, and let $\delta = \delta_0^{10}$. Let $n_0 \geq n_0'$ be the smallest integer such that

$$(1 + \gamma')\left(n_0 + 2\delta^{-1}\gamma'^{-1}\right) \leq (1+\gamma)n_0$$

and such that $\frac{\eta n}{\log n} \geq 2$. Given $n \geq n_0$, and an $n$-vertex tree $T$ with $\Delta(T) \leq \frac{\eta n}{\log n}$, let $n'$ be the smallest integer which is at least $n$ and divisible by $2\delta^{-1}\gamma'^{-1}$. Let $T'$ be an $n'$-vertex tree obtained by attaching a path with $n' - n$ vertices to a leaf of $T$. Theorem 8 applies, so there is a graceful labelling of $T'$ with $(1 + \gamma')n' \leq (1 + \gamma)n$ labels. The induced labelling of $T$ is also graceful, as desired. $\qquad\square$

We now give the setup we will use to prove Theorem 8.

$\mu,\ \delta_0,\ \delta,\ \varepsilon,$  **Setup 9.** *Given $\gamma > 0$ such that $\gamma^{-1}$ is an integer, we choose $\mu = \left\lceil \exp\left(10^8 \gamma^{-4}\right)\right\rceil^{-1}$. We*
$\eta$          *set $\delta_0 = \mu^{2/\mu}$, $\delta = \delta_0^{10}$, $\varepsilon = \delta^{10}$, and $\eta = \varepsilon^{10}$. Set*

$$n_0 = 2^{10/\eta^2}. \tag{2}$$

*Suppose now that $n > n_0$ divisible by $2\delta_0^{-10}\gamma^{-1}$ is given. Note that since $\mu^{-1}$ is an integer, $\delta_0^{-1}$ is also an integer, so this is possible.*

$\delta_i$        *For each $0 \leq i \leq n$ set $\delta_i = \mu^{(2n-i)/(\mu n)}$. Observe that for $i = 0$, this definition is consistent with the previous definition of $\delta_0$.*

$\ell, m$        *Let $\ell = \frac{1}{2}\delta_0^2 \gamma n$, and $m = \delta_0^2 \ell = \frac{1}{2}\delta_0^4 \gamma n$. Because $n$ is divisible by $2\delta_0^{-4}\gamma^{-1}$, and $\delta_0^{-1}$ is an integer, these two quantities are integers and $m$ divides $\ell$.*

$\tilde{n}$        *Let $\tilde{n} = (1 + \gamma)n$. Because $2\delta_0^{-4}\gamma^{-1}$ divides $n$, in particular $\tilde{n}$ is an integer multiple of $2m$.*
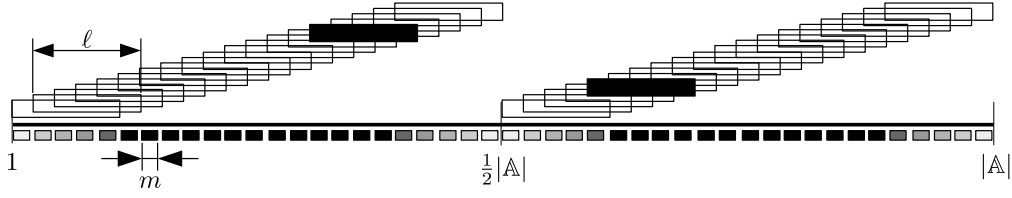
FIGURE 3.1. The intervals $\mathcal{I}_{\mathsf{V}}$ (below the line) and $\mathcal{J}$ (above the line). The shade of grey used for an interval $I \in \mathcal{I}_{\mathsf{V}}$ corresponds to the number of intervals of $\mathcal{J}$ containing $I$. A pair of complementary intervals of $\mathcal{J}$ is highlighted.

Let $\mathbb{A} := [\tilde{n}]$, and $\mathbb{B} := [\tilde{n} - 1]$. We will choose vertex labels from $\mathbb{A}$, and edge labels $\quad \mathbb{A}, \mathbb{B}$ from $\mathbb{B}$.

Let $\mathcal{I}_{\mathsf{V}}$ be the collection of intervals of length $m - 1$ (i.e., size $m$) in $\mathbb{A}$ whose first elements $\quad \mathcal{I}_{\mathsf{V}}$ are in the set
$$\{1, m+1, 2m+1, \ldots, \tilde{n} - m + 1\},$$
and let $\mathcal{I}_{\mathsf{E}}$ be the collection of intervals of length $m - 1$ in $\mathbb{B} \cup \{0\}$ whose first elements are in $\quad \mathcal{I}_{\mathsf{E}}$ the set
$$\{0, m, \ldots, |\mathbb{B}| - m + 1\}.$$

Finally, let $\mathcal{J}$ be the set of intervals of length $\ell - 1$ in $\mathbb{A}$ whose first elements are in the set $\quad \mathcal{J}$
$$\left\{1, m+1, 2m+1, \ldots, \tfrac{\tilde{n}}{2} - m - \ell + 1, \tfrac{\tilde{n}}{2} - \ell + 1, \tfrac{\tilde{n}}{2} + 1, \tfrac{\tilde{n}}{2} + m + 1, \tfrac{\tilde{n}}{2} + 2m + 1, \ldots, \tilde{n} - m - \ell + 1, \tilde{n} - \ell + 1\right\}.$$
(See Figure 3.1.)

For each $J \in \mathcal{J}$, we define the complementary interval $\overline{J} \in \mathcal{J}$ to be the (unique) interval $\quad \overline{J}$ such that the sum of the elements of $J \cup \overline{J}$ is $\ell(\tilde{n} + 1)$. By definition, $J$ and $\overline{J}$ are disjoint, one having only elements less than or equal to $\frac{1}{2}\tilde{n}$ and the other having only elements greater than $\frac{1}{2}\tilde{n}$.

Note that
$$|\mathcal{I}_{\mathsf{V}}| = |\mathcal{I}_{\mathsf{E}}| = \tfrac{\tilde{n}}{m} \quad \text{and} \quad |\mathcal{J}| = \tfrac{\tilde{n}}{m} - 2\left(\tfrac{\ell}{m} - 1\right). \tag{3}$$

In analysing our labelling algorithm, the values $(\delta_i)_{i=1}^n$ will quantify errors in our quasirandomness properties in steps $i = 1, \ldots, n$ (see Claim 3). Our choice of $\delta_i$ ensures that the following holds.

**Fact 10.** *For each $1 \le t \le n$, we have*
$$\sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} \delta \mu^{-1} \delta_{i \delta n} < \tfrac{1}{100} \delta_t. \tag{4}$$

*Proof.* Let us write $a := \mu^{-\delta/\mu}$. Later, we will later need that
$$a \ge 2. \tag{5}$$

We have

$$a = \mu^{-\delta/\mu} = \mu^{-\mu^{\frac{20}{\mu}-1}} .$$

We shall just sketch (5), assuming that $\mu$ is 'very small', i.e., we will look at the limit behaviour of $\mu$ around 0. It is easy (but tedious) to check that these calculations do go through for our $\mu$ (with a lot of room). Recall a basic fact from calculus that $\lim_{x \searrow 0} x^{\frac{1}{x}} = 1$. Similar calculation yield $\lim_{x \searrow 0} x^{\frac{20}{x}-1} = 1$. Hence, $a$ equals $\mu$ taken a power which is close to $-1$, and thus $a$ is very big. In particular (5) holds.

Next, let us bound $\sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} \mu^{-i\delta/\mu}$. Using the formula for geometric sequences, we obtain

$$\sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} \mu^{-i\delta/\mu} = \sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} a^i = \frac{a^{\lceil \frac{t}{\delta n} \rceil + 1} - a}{a-1} \overset{(5)}{\leq} \frac{a^{\lceil \frac{t}{\delta n} \rceil + 1}}{\frac{a}{2}} = 2a^{\lceil \frac{t}{\delta n} \rceil} . \tag{6}$$

Using the definition of $\delta_q$, we can expand the left-hand side as follows,

$$\sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} \delta\mu^{-1}\delta_{i\delta n} = \delta\mu^{-1} \cdot \mu^{2/\mu} \cdot \sum_{i=1}^{\lceil \frac{t}{\delta n} \rceil} \mu^{-i\delta/\mu}$$

$$\overset{(6)}{\leq} \delta\mu^{-1} \cdot \mu^{2/\mu} \cdot 2\mu^{-\lceil t/(\delta n) \rceil \delta/\mu} = 2\delta\mu^{-1} \cdot \delta_{\lceil t/(\delta n) \rceil \delta n}$$

$$\leq 2\delta\mu^{-1} \cdot \mu^{-\delta/\mu} \cdot \delta_t = 2\delta \cdot \mu^{-(\frac{\delta}{\mu}+1)} \cdot \delta_t .$$

Recall that $\delta$ is much smaller than $\mu$. In particular, $\mu^{-(\frac{\delta}{\mu}+1)} < \mu^{-1.1}$. Using the relation between $\delta$ and $\mu$ once again, we get that $2\delta \cdot \mu^{-1.1} < \frac{1}{100}$. Hence, (4) follows. □

The inequality (4) will be important in showing that our error terms do not grow too fast.

3.2. **Assigning intervals of labels to vertices.** Our next step is to give an order on $V(T)$ and for each vertex the promised 'small interval' in $\mathcal{J}$ in which we will eventually choose its label. As mentioned in the outline of the proof, for most edges we will label the two endpoints from complementary small intervals. Specifically, we will do this for all edges not in the set $\mathcal{R}$ given by the following lemma.

**Lemma 11.** *Assume Setup 9. Given an n-vertex tree $T$ with $\Delta(T) \leq \frac{\eta n}{\log n}$, there exists a set $\mathcal{R} \subseteq E(T)$, an ordering $V(T) = \{v_1, \ldots, v_n\}$ of the vertices of $T$, and a collection of intervals $J(v_i) \in \mathcal{J}$ for each $1 \leq i \leq n$ with the following properties.*

*(PRE1) $|\mathcal{R}| \leq \varepsilon n$.*

*(PRE2) For each $i \geq 2$, the vertex $v_i$ has exactly one neighbour in the set $\{v_1, \ldots, v_{i-1}\}$. We call this vertex the parent of $v_i$, and denote it by $v_i^+$.*

*(PRE3) If $v_i v_j \in E(T) \setminus \mathcal{R}$ then $|i - j| \leq \frac{\varepsilon n}{\log n}$.*

*(PRE4) For each interval $S \subseteq [n]$ and each $J \in \mathcal{J}$, we have $\sum_{i \in S} \mathbb{1}_{J(v_i)=J} = \frac{|S|}{|\mathcal{J}|} \pm \delta^2 n$.*

*(PRE5) For each $v_i v_j \in E(T) \setminus \mathcal{R}$ we have $J(v_i) = \overline{J(v_j)}$.*

*Proof.* Given a tree $T$, let $\mathcal{R}$ be the set of edges of $T$ returned by Lemma 7 with input $\varepsilon$. Then we have $|\mathcal{R}| \leq \varepsilon n$, giving (PRE1). Let $T_1, \ldots, T_s$ be the components of $T - \mathcal{R}$.

Let $v_1$ be an arbitrary vertex of $V(T)$. Now for each $i \geq 2$ in turn, we choose $v_i$ to be a vertex of $V(T) \setminus \{v_1, \ldots, v_{i-1}\}$ which has a neighbour in $\{v_1, \ldots, v_{i-1}\}$, if possible choosing a vertex in the same component of $T - \mathcal{R}$ as $v_{i-1}$. (PRE2) follows from this construction.

Suppose now that in the above construction we cannot choose $v_i$ in the same component of $T - \mathcal{R}$ as $v_{i-1}$. Then this is because we already chose all vertices of that component. It follows that each component forms an interval in our ordering on the vertices. In particular, if $v_i v_j \in E(T) \setminus \mathcal{R}$ then $v_i v_j$ is in one component $T_k$ of $T - \mathcal{R}$. Then Lemma 7 tells us that $|i - j| \leq v(T_k) \leq \frac{\varepsilon n}{\log n}$, giving (PRE3).

We now take an arbitrary proper colouring of $T$ with two colours, red and blue. For each component $T_k$ of $T - \mathcal{R}$ we choose independently an interval $J(T_k)$ uniformly at random from $\mathcal{J}$. For each red $v_j \in T_k$ we set $J(v_j) = J(T_k)$, and for each blue $v_j \in T_k$ we set $J(v_j) = \overline{J(T_k)}$. This gives (PRE5) deterministically.

It remains to show that with positive probability we obtain (PRE4). To that end, let $S \subseteq [n]$ be an interval, and $J$ be an element of $\mathcal{J}$. We view $S$ as an interval in $(v_1, \ldots, v_n)$. Consider the intersections of various components $T_k$ with $S$. Since these components form intervals in $(v_1, \ldots, v_n)$, at most two components have non-empty intersection with $S$ and are not contained in $S$. These at most two components contain at most $2\frac{\varepsilon n}{\log n} < \delta^5 n$ vertices. Let $\mathcal{K}$ index the components which are contained in $S$, i.e., $T_k \subseteq S$ for each $k \in \mathcal{K}$. For each $k \in \mathcal{K}$ define $Y_k$ as follows. If $J(T_k) = J$, let $Y_k$ be equal to the number of red vertices in $T_k$; if $\overline{J(T_k)} = J$ let $Y_k$ be equal to the number of blue vertices in $T_k$, and otherwise let $Y_k = 0$.

Because the sets $J(T_k)$ are chosen independently, we get that $Y_k$ are independent random variables, with $0 \leq Y_k \leq v(T_k)$ for each $k \in \mathcal{K}$. We have $\mathbf{E}[Y_k] = \frac{v(T_k)}{|\mathcal{J}|}$ for each $k \in \mathcal{K}$. Because $v(T_k) \leq \frac{\varepsilon n}{\log n}$ for each $k \in \mathcal{K}$, and $\sum_{k \in \mathcal{K}} v(T_k) \leq n$, we have

$$\sum_{k \in \mathcal{K}} v(T_k)^2 \leq \frac{n}{\frac{\varepsilon n}{\log n}} \cdot \left(\frac{\varepsilon n}{\log n}\right)^2 = \frac{\varepsilon n^2}{\log n}.$$

Let $X = \sum_{k \in \mathcal{K}} Y_k$. By Theorem 5 we have

$$\mathbf{P}\left[|X - \mathbf{E}[X]| > \tfrac{1}{4}\delta^4 n\right] \leq 2\exp\left(-\tfrac{1}{8} \cdot \delta^8 n^2 \cdot \tfrac{\log n}{\varepsilon n^2}\right) = 2\exp\left(-\tfrac{1}{8\delta^2}\log n\right) \leq n^{-10},$$

where we used the relation between $\varepsilon$ and $\delta$ from Setup 9.

We see that with probability at least $1 - n^{-10}$ we have

$$X = \frac{|S| \pm \delta^5 n}{|\mathcal{J}|} \pm \tfrac{1}{4}\delta^4 n = \frac{|S|}{|\mathcal{J}|} \pm \tfrac{1}{2}\delta^4 n.$$

Since the number of vertices $v_i \in S$ with $J(v_i) = J$ is $X$, we conclude that for the chosen $S$ and $J$ the probability that (PRE4) fails is at most $n^{-10}$. Taking the union bound over $|\mathcal{J}| \leq n$ choices of $J$ and $\binom{n}{2}$ choices of $S$ we get that with probability at least $1 - n^{-7}$ we have (PRE4), as desired. $\qquad\square$

3.3. **Counting structures, and quasirandom properties.** The phenomenon of quasirandomness is central in discrete mathematics. For example, the celebrated Chung–Graham–Wilson Theorem asserts that if the edge density and the four-cycle density of an $n$-vertex graph $G$ are close to those of an Erdős–Rényi random graph $\mathbb{G}(n, p)$, then $G$ has many other properties as a typical $\mathbb{G}(n, p)$. In the proof of Theorem 8, we need to control the evolution of the sets $\mathsf{A} \subseteq \mathbb{A}$ and $\mathsf{B} \subseteq \mathbb{B}$ of vertex labels and edge labels not used so far during the sequential labeling of $T$. We want to prove that the pair $(\mathsf{A}, \mathsf{B})$ behaves quasirandomly. Thus, in some analogy to the Chung–Graham–Wilson Theorem, we want to come up with quantities control over of which will imply further quasirandomness properties of $(\mathsf{A}, \mathsf{B})$. To this end, we introduce a class $\mathcal{X}$ of 'structures' in Section 3.3.1. We will write $|X(\mathsf{A}, \mathsf{B})|$ for the number of structures of a given type which appear in $(\mathsf{A}, \mathsf{B})$, and the main technical work of this paper will be to show that this number remains close to what one would expect if the two sets were chosen independently at random. In Section 3.3.2 we then explain that these parameters are indeed useful for our graceful labelling. In Section 3.3.3 we then state our main quasirandomness condition and state a useful lemma connected to it.

3.3.1. *Structures.* The key objects that allow us to control quasirandomness are 'structures' defined below.

**Definition 12** (structure). *A structure $X$ is a graph such that*

- *its vertices are labelled with either pairwise distinct elements of $\mathbb{A}$ (we call such vertices* fixed*) or pairwise distinct intervals in $\mathbb{A}$ (we call such vertices* free*), and*
- *its edges are labelled with either pairwise distinct elements of $\mathbb{B}$ (*fixed*) or with distinct choices of special symbols $\mathfrak{e}_1$ or $\mathfrak{e}_2$ (*free*).*

*When dealing with structures, we identify vertices and edges with their labels. So, vertices in $X$ are numbers (if they are fixed) or intervals (if they are free). Likewise, edges in $X$ are numbers or special free symbols $\mathfrak{e}_1$ and $\mathfrak{e}_2$.*

*In any structure $X$, we require that if $u$ and $v$ are fixed vertices, then $uv$ is a fixed edge and we have $uv = |u - v|$, and we require that each free edge has one endpoint fixed and the other free.*

We shall be interested in four groups of structures, denoted by $X_1, \ldots, X_4$. Actual structures in each individual group, say in $X_i$, have the same underlying graph, but differ by labels on vertices and edges. So, let us first describe the graphs underlying these four groups of structures. The graph underlying $X_1$ is a single vertex. The graph underlying $X_2$ and $X_4$ is a path on three vertices. The graph underlying $X_3$ is an edge. Let us now describe the free and fixed vertices and edges of these four groups of structures. The single vertex of $X_1$ is free. Members of $X_2$ have one end-vertex fixed and the two remaining vertices free. The edge connecting the two free vertices is fixed and the other one is free. In $X_3$, one vertex is free and the other is fixed; the edge connecting these two vertices is free. Last, the center of $X_4$ is free,

FIGURE 3.2. Structures for quasirandomness



$$X_1 \llbracket I \rrbracket \qquad X_2 \llbracket a, I, c, I' \rrbracket \qquad X_3 \llbracket a, I \rrbracket \qquad X_4 \llbracket a, a', I \rrbracket$$
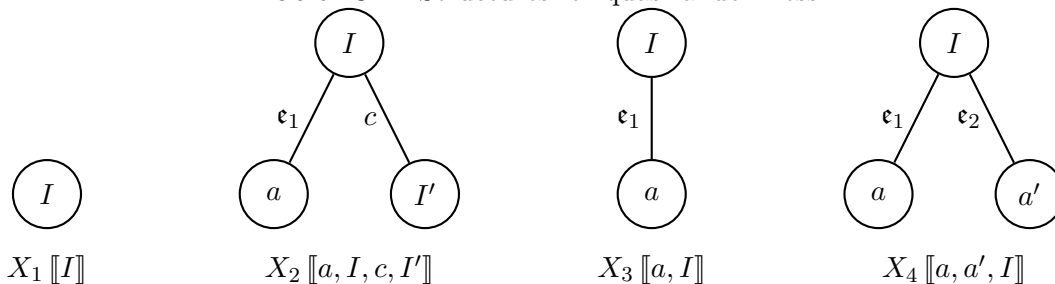
the end-vertices are fixed, and the two edges are free. We then refer to the actual structures in group $X_i$ by writing $X_i \llbracket \cdot \rrbracket$ where the double brackets contains specification of labels of fixed vertices (these are specified by an element of $\mathbb{A}$), free vertices (these are specified by an interval in $\mathbb{A}$), and fixed edges (these are specified by an element of $\mathbb{B}$). Note that free edges are not parametrized in Definition 12, and hence no information regarding them is included the double brackets.

For example, individual structures within the group $X_2$ differ by the actual label on the fixed vertex, the two labels on the free vertices, and the label on the single fixed edge. That is, given $a, a' \in \mathbb{A}$, $c \in \mathbb{B}$ and distinct $I, I' \in \mathcal{I}_V \cup \mathcal{J}$ we write $X_1 \llbracket I \rrbracket$, $X_2 \llbracket a, I, c, I' \rrbracket$, $X_3 \llbracket a, I \rrbracket$ and $X_4 \llbracket a, a', I \rrbracket$ for structures as shown on Figure 3.2.

**Definition 13.** *Let $\mathcal{X}$ be the set of structures of the form $X_1 \llbracket I \rrbracket$, $X_2 \llbracket a, I, c, I' \rrbracket$, $X_3 \llbracket a, I \rrbracket$ and $X_4 \llbracket a, a', I \rrbracket$ where $a, a'$ are distinct elements of $\mathbb{A}$, where $c$ is an element of $\mathbb{B}$, and where $I, I' \in \mathcal{I}_V$ are distinct intervals.*

**Remark 14.** *We shall explain in Section 3.3.2 the essential role structures play in our proof. But at this moment, let us give at least a brief hint behind their definition. Say, we are dealing with labeling of particular three vertices $v_1$, $v_2$, and $v_3$ of our tree, which are in a mutual position as in group $X_2$. Say, for some reason (which we explain in Section 3.3.2), we want to label $v_1$ with $a$, $v_2$ with a label from $I$ and $v_3$ with a label from $I'$. Further, suppose that we want the induced edge label on $v_2 v_3$ to be $c$. Then these requirements are clearly reflected by $X_2 \llbracket a, I, c, I' \rrbracket$. You can also observe that such a situation has exactly 'one degree of freedom', where each degree of freedom represents the possibility of many choices (more than 2) of one label. Indeed, selecting arbitrarily a label $a_2$ for $v_2$ in $I$, we have at most two options, namely $a_2 + c$ and $a_2 - c$ (these numbers need not be in $I'$). In this sense, all structures $X_1 \llbracket I \rrbracket$, $X_2 \llbracket a, I, c, I' \rrbracket$, $X_3 \llbracket a, I \rrbracket$ and $X_4 \llbracket a, a', I \rrbracket$ have exactly one degree of freedom.*

*This perspective also explains why a free edge is not specified by any interval, in contrast to free vertices. Indeed, each free edge in the above structures is incident to a fixed vertex from one side and to a free vertex from another side, and hence specifying an interval for the free vertex already tells us the interval for the induced edge label. The only reason why*

*we introduce the 'special symbols' $\mathfrak{e}_1$ and $\mathfrak{e}_2$ is to distinguish the two edges graph-theoretically (that is, without taking into account different properties their induced edge labels may enjoy).*

*Last, let us explain why Definition 13 we require $I$ and $I'$ to be distinct. Recall that in Lemma 11(PRE5) the intervals assigned to the two endvertices of each edge $T$ (except the special ones in $\mathcal{R}$) are complementary, and hence distinct. $I$ and $I'$ will represent these assigned intervals.*

We can now define define a concept of "one structure following the pattern of another structure" (so, we do not define any notion of "pattern" per se), and of chosen labels.

**Definition 15** (to follow pattern, chosen label). *Suppose that we have a structure $X$. Suppose that $X$ has a free edge $\mathfrak{e}_i$ with fixed endpoint labelled $a$ and free endpoint labelled $I$. We then*
$\mathrm{Diff}(\mathfrak{e}_i; X)$   *write $\mathrm{Diff}(\mathfrak{e}_i; X) := |a - \min(I)|$. We write $\mathrm{Free}_{\mathsf{V}}(X)$ for the set of free vertex labels in $X$,*
$\mathrm{Free}_{\mathsf{V}}(X)$   *and $\mathrm{Free}_{\mathsf{E}}(X)$ for the set of free edge labels in $X$. We write $\mathrm{free}(X)$ for the total number of*
$\mathrm{Free}_{\mathsf{E}}(X)$   *free (vertex- or edge-) labels in $X$.*
$\mathrm{free}(X)$

*Suppose that we are given structures $X$ and $X'$. We say that $X'$ follows the pattern $X$ if $\mathrm{free}(X') = 0$ and there is a graph isomorphism $\varrho$ from $X'$ to $X$ such that*

- *for each vertex $u \in X'$ for which $\varrho(u)$ is fixed in $X$, the labels of $u$ and of $\varrho(u)$ are the same,*
- *for each vertex $u \in X'$ for which $\varrho(u)$ is free in $X$, the label of $u$ is contained in the label of $\varrho(u)$,*
- *for each edge $uv \in X'$ for which $\varrho(uv)$ is a fixed edge in $X$, the labels of $uv$ and of $\varrho(uv)$ are the same.*

*We call the labels of vertices $u \in X'$ for which $\varrho(u)$ is free in $X$,* chosen (vertex) labels. *Likewise, we call the labels of edges $uv \in X'$ for which $\varrho(uv)$ is free in $X$,* chosen (edge)
$\mathrm{Ch}_{\mathsf{V}}(X'; X)$   labels. *We write $\mathrm{Ch}_{\mathsf{V}}(X'; X)$ for the set of chosen vertex labels in $X'$ and $\mathrm{Ch}_{\mathsf{E}}(X'; X)$ for the*
$\mathrm{Ch}_{\mathsf{E}}(X'; X)$   *set of chosen edge labels in $X'$.*
$X(\mathsf{A}, \mathsf{B})$        *Given $\mathsf{A} \subseteq \mathbb{A}$ and $\mathsf{B} \subseteq \mathbb{B}$, and a structure $X$, we write*

$$X(\mathsf{A}, \mathsf{B}) \tag{7}$$

*for the set of all structures following the pattern $X$ whose chosen vertex labels are in $\mathsf{A}$ and whose chosen edge labels are in $\mathsf{B}$.*

Note that the double square brackets use to parametrise the families $X_1, \dots, X_4$ have a different meaning than the parentheses in (7); we can for example write $X_2 \llbracket a, I, c, I' \rrbracket (\mathsf{A}, \mathsf{B})$ for the set of all structures following the pattern $X_2 \llbracket a, I, c, I' \rrbracket$ whose chosen vertex labels are in $\mathsf{A}$ and whose chosen edge label is in $\mathsf{B}$.

3.3.2. *The role of structures $\mathcal{X}$.* At the beginning of Section 3 we outlined the main idea of the proof of Theorem 4 (which is also the main idea of Theorem 8): We proceed labeling the vertices of $T$. During this process, the set $\mathsf{A} \subseteq \mathbb{A}$ of available vertex labels and the set $\mathsf{B} \subseteq \mathbb{B}$ of

available edge labels get sparser. We need to control that the sets $\mathsf{A}$ are $\mathsf{B}$ are spread over the intervals $\mathbb{A}$ and $\mathbb{B}$ in a quasirandom way. Actually, we need to control even the interactions of $\mathsf{A}$ and $\mathsf{B}$; for example if $\mathsf{A}$ consisted of all even numbers and $\mathsf{B}$ of all odd numbers then these sets are very uniformly spread, but it is not possible to label a single new pair of vertices that form an edge of $T$. It turns out that the quantities we need to control for our proof of Theorem 8 are exactly the quantities $|X(\mathsf{A}, \mathsf{B})|$, for each $X \in \mathcal{X}$. In other words, we control the number of structures that follow the pattern $X$ and use elements from $\mathsf{A}$ and $\mathsf{B}$ as the chosen vertex labels and edge labels, respectively. For example, observe that the density $\frac{|\mathsf{A} \cap I|}{|I|}$ of $\mathsf{A}$ on $I \in \mathcal{I}_{\mathsf{V}}$ is equal to $\frac{|X_1[\![I]\!](\mathsf{A}, \mathsf{B})|}{|X_1[\![I]\!](\mathbb{A}, \mathbb{B})|}$. For the other three structures, one should think of $a \in \mathbb{A}$ as the chosen label for some vertex $v_i$ of $T$. Then $|X_3[\![a, I]\!](\mathsf{A}, \mathsf{B})|$ is the number of ways to give a neighbour $v_j$ of $v_i$ a label in $I$ which has not yet been used and which induces an edge label that has not yet been used (which is obviously a useful quantity to control). If one thinks of $a$ as being the chosen label of a vertex $v_i$, then $|X_2[\![a, J(v_j), c, J(v_k)]\!](\mathsf{A}, \mathsf{B})|$ is the number of ways to label a child $v_j$ of $v_i$ and grandchild $v_k$ of $v_i$ within their chosen intervals, not re-using vertex or edge labels used previously, and using the edge label $c$ on $v_j v_k$. The quantity $|X_4[\![a, a', J(v_j)]\!](\mathsf{A}, \mathsf{B})|$ plays a similar rôle, except that we fix the vertex label used for $v_k$ to be $a'$ rather than the edge label for $v_j v_k$.

One of our quasirandomness conditions, (QUASI2) below, states that these quantities are likely to stay close to what one would expect if the sets $\mathsf{A}$ and $\mathsf{B}$ were chosen independently at random; thus $|X_1[\![I]\!](\mathsf{A}, \mathsf{B})|$ will be close to the overall density of $\mathsf{A}$ in $\mathbb{A}$ when $\mathsf{A}$ is generated by running our labelling algorithm for a given time. In addition, we introduce property (QUASI1), which states that the density of $\mathsf{B}$ on each interval $I_{\mathsf{E}} \in \mathcal{I}_{\mathsf{E}}$ is approximately $\frac{|\mathsf{B}|}{|\mathbb{B}|}$.

Since this idea of a structure following the pattern of another is a little complicated, before stating Fact 17 which elaborates on our remark about one degree of freedom from Remark 14, and puts structural and quantitative restrictions on $X(\mathsf{A}, \mathsf{B})$ for any structure $X \in \mathcal{X}$ and sets $\mathsf{A}$ and $\mathsf{B}$, we give a little example.

**Example 16.** *In Figure 3.3 we give an example of structures with a specific label on a specific free vertex or edge. More precisely, we show the only four structures following $X_2[\![10, I, 3, I']\!]$ (for say $I = I' = \{5, 6, \ldots, 15\}$) which induce label $2$ on $\mathfrak{e}_1$. The two leftmost structures are also the only structures following $X_2[\![10, I, 3, I']\!]$ that have label $12$ on $I$. Since $I$ and $I'$ overlap, they cannot be two distinct sets of $\mathcal{I}_{\mathsf{V}}$, and thus $X_2[\![10, I, 3, I']\!] \notin \mathcal{X}$, i.e. this particular structure is not one we will be interested in. We are only interested in $X_2[\![10, I, 3, I']\!] \in \mathcal{X}$, which means $I$ and $I'$ are disjoint. If $10$ is in $I$, we might still have two possible vertex labels in $I$ which induce label $2$ on $\mathfrak{e}_1$ (if $10$ is not in $I$, there will be at most one such choice), but given such a vertex label there is at most one valid choice of vertex label in $I'$.*

We now state the following fact, which generalises the above example.

FIGURE 3.3. The four structures following $X_2 \llbracket 10, I, 3, I' \rrbracket$ (for say $I = I' = \{5, 6, \ldots, 15\}$) which induce label 2 on $\mathfrak{e}_1$. The two leftmost structures are also the only structures following $X_2 \llbracket 10, I, 3, I' \rrbracket$ that have label 12 on $I$.
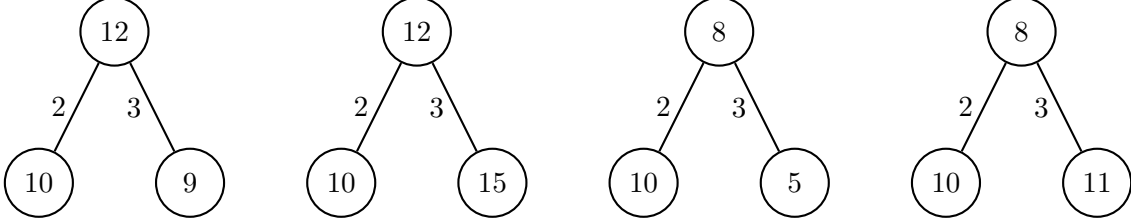


**Fact 17.** *Suppose that $X \in \mathcal{X}$ and sets $\mathsf{A}$ and $\mathsf{B}$ are given.*

(a) *Given any free vertex and a label in $\mathbb{A}$ there is at most one structure in $X(\mathsf{A}, \mathsf{B})$ in which the corresponding vertex has that label. Similarly for any free edge and label in $\mathbb{B}$ there are at most two structures in $X(\mathsf{A}, \mathsf{B})$ in which the corresponding edge has that label.*

(b) *We have $|X(\mathsf{A}, \mathsf{B})| \leq \min_{I \in \mathrm{Free}_\mathsf{V}(X)} |I| = m$.*

*Proof.* We first deal with Part (a), the case of substituting a chosen label, say $a^*$, on a free vertex. The statement is trivial for structures of the groups $X_1$, $X_3$ and $X_4$, because for these, substituting $a^*$ on the originally free vertex fully specifies that structure. For structures $X_2 \llbracket a, I, c, I' \rrbracket$, we recall that Definition 13 requires all elements of $I$ to be smaller than all elements of $I'$ or vice versa. So, if $a^*$ is the chosen vertex label for the smaller of the two intervals, then we know that the bigger of the two must get the label $a^* + c$. If $a^*$ is the chosen vertex label on the bigger of the two intervals, then we know that the smaller of the two must get the label $a^* - c$.

When a chosen edge label, say $c^*$, is substituted in structures $X_2 \llbracket a, I, c, I' \rrbracket$, $X_3 \llbracket a, I \rrbracket$, $X_4 \llbracket a, a', I \rrbracket$ on the free edge $\mathfrak{e}_1$, then in all these cases the free vertex labeled with $I$ must be given a chosen vertex label $a + c^*$ or $a - c^*$. This completely specifies the chosen labels in case of $X_3 \llbracket a, I \rrbracket$, $X_4 \llbracket a, a', I \rrbracket$. In case of $X_2 \llbracket a, I, c, I' \rrbracket$, there is at most one way of choosing a vertex label for the free vertex labeled $I'$ by the argument from the previous paragraph. The free edge $\mathfrak{e}_2$ appears only in structure $X_4 \llbracket a, a', I \rrbracket$, where the argument is completely symmetric.

Part (b) follows from Part (a). □

3.3.3. *The quasirandomness conditions.* As explained in Section 3.3.2, our quasirandomness conditions, given in Definition 19 below, assert that the pair $(\mathsf{A}, \mathsf{B})$ is a quasirandom subset of the pair $(\mathbb{A}, \mathbb{B})$ of density roughly $\frac{|\mathsf{A}|^2}{|\mathbb{A}|^2}$. Here, the quasirandomness condition is expressed via structures from $\mathcal{X}$, that is, by comparing $|X(\mathsf{A}, \mathsf{B})|$ and $|X(\mathbb{A}, \mathbb{B})|$ for $X \in \mathcal{X}$. For this reason, it is convenient to express the quantities $|X(\mathbb{A}, \mathbb{B})|$ explicitly. This is done in the easy lemma below.

**Lemma 18.** *Suppose that $a$ and $a'$ are distinct elements of $\mathbb{A}$, that $c \in \mathbb{B}$, that $I \in \mathcal{I}_\mathsf{V}$, and that $J \in \mathcal{J}$. Then we have the following.*

(a) $\left|X_3 [\![a, J]\!] (\mathbb{A}, \mathbb{B})\right| = \ell \pm 1$,

(b) $\left|X_3 [\![a, I]\!] (\mathbb{A}, \mathbb{B})\right| = m \pm 1$,

(c) $\left|X_4 [\![a, a', J]\!] (\mathbb{A}, \mathbb{B})\right| = \ell \pm 3$,

(d) $\left|X_2 [\![a, J, c, \overline{J}]\!] (\mathbb{A}, \mathbb{B})\right| = \left|\{(b, b') \in J \times \overline{J} : |b - b'| = c\}\right| \pm 3$.

*Proof.* For (a), observe that $X_3 [\![a, J]\!] (\mathbb{A}, \mathbb{B})$ contains all structures that follow the pattern $X_3 [\![a, J]\!]$ and whose single chosen vertex is labelled with an element of $J$. The only exception is when $a \in J$; in that case the structure in which the chosen vertex is labelled with $a$ is not counted (recall that in Definition 12, vertices are labelled with distinct labels). Since $|J| = \ell$, we have that $\left|X_3 [\![a, J]\!] (\mathbb{A}, \mathbb{B})\right| = \ell$ or $\left|X_3 [\![a, J]\!] (\mathbb{A}, \mathbb{B})\right| = \ell - 1$. The proof of (b) is exactly the same.

Similarly, for (c), observe that $X_4 [\![a, a', J]\!] (\mathbb{A}, \mathbb{B})$ contains all structures that follow the pattern $X_4 [\![a, a', J]\!]$ and whose chosen vertex is labelled by any element from $J \setminus \{a, a', \frac{a+a'}{2}\}$ (the reason for excluding $\frac{a+a'}{2}$ is that such a choice would give us the same label on the edges $\mathfrak{e}_1$ and $\mathfrak{e}_2$).

For (d), observe that $X_2 [\![a, J, c, \overline{J}]\!] (\mathbb{A}, \mathbb{B})$ contains all structures that follow the pattern $X_2 [\![a, J, c, \overline{J}]\!]$ and whose label $b$ chosen on the free vertex $J$ and label $b'$ chosen on the free vertex $\overline{J}$ satisfy

(i) $b \in J$, $b' \in \overline{J}$, $|b - b'| = c$, and

(ii) $b \neq a$, $b' \neq a$, $|b - a| \neq |b - b'|$.

The number of pairs $(b, b')$ satisfying (i) is $\left|\{(b, b') \in J \times \overline{J} : |b - b'| = c\}\right|$, and each of the three restrictions in (ii) decreases this number by at most 1. $\qquad\square$

**Definition 19** (Quasirandomness conditions). *A pair of sets* $\mathsf{A} \subseteq \mathbb{A}$ *and* $\mathsf{B} \subseteq \mathbb{B}$ *is* $\alpha$-quasirandom *if we have the following.*

*(QUASI1) For all* $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ *we have* $|I_\mathsf{E} \cap \mathsf{B}| = m \cdot \frac{|\mathsf{A}|}{\tilde{n}} \pm \alpha m$.

*(QUASI2) For all* $X \in \mathcal{X}$ *we have* $|X(\mathsf{A}, \mathsf{B})| = |X(\mathbb{A}, \mathbb{B})| \left(|\mathsf{A}|/\tilde{n}\right)^{\text{free}(X)} \pm \alpha m$.

It may seem strange that we only have $|\mathsf{A}|$ in the estimates, and not $|\mathsf{B}|$, but we will prove that throughout our labelling process, $|\mathsf{A}|$ and $|\mathsf{B}|$ are close enough for the difference to be immaterial. Note that we only insist on counts of structures in $\mathcal{X}$ being preserved, so that $I, I' \in \mathcal{I}_\mathsf{V}$, even though we defined our structures allowing $I, I' \in \mathcal{I}_\mathsf{V} \cup \mathcal{J}$. The following claim lets us deduce the latter (in the cases we need it) from the former.

**Lemma 20.** *Suppose that* $(\mathsf{A}, \mathsf{B})$ *is* $\alpha$-*quasirandom, that* $a$ *and* $a'$ *are distinct elements of* $\mathbb{A}$, *that* $c \in \mathbb{B}$, *and that* $J \in \mathcal{J}$. *If* $\ell \geq 3\alpha^{-1}$, *then we have*

$$\left|X_2 [\![a, J, c, \overline{J}]\!] (\mathsf{A}, \mathsf{B})\right| = \left(|\mathsf{A}|/\tilde{n}\right)^3 \left|\{(b, b') \in J \times \overline{J} : |b - b'| = c\}\right| \pm 3\alpha\ell, \qquad (8)$$

$$\left|X_3 [\![a, J]\!] (\mathsf{A}, \mathsf{B})\right| = \left(|\mathsf{A}|/\tilde{n}\right)^2 \ell \pm 2\alpha\ell \quad and \qquad (9)$$

$$\left|X_4 [\![a, a', J]\!] (\mathsf{A}, \mathsf{B})\right| = \left(|\mathsf{A}|/\tilde{n}\right)^3 \ell \pm 2\alpha\ell. \qquad (10)$$

*Proof.* Let us first establish (9). Let us recall that the interval $J$ is partitioned into $\ell/m$ many intervals $\{I\}_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J}$. Therefore, $X_3 \llbracket a, J \rrbracket (\mathsf{A}, \mathsf{B})$ is partitioned into $\{X_3 \llbracket a, I \rrbracket (\mathsf{A}, \mathsf{B})\}_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J}$. Property (QUASI2) applies to each structure $X_3 \llbracket a, I \rrbracket$ with $I \in \mathcal{I}_\mathsf{V}$ and $I \subseteq J$. Hence,

$$
\begin{aligned}
\left| X_3 \llbracket a, J \rrbracket (\mathsf{A}, \mathsf{B}) \right| &= \sum_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J} \left| X_3 \llbracket a, I \rrbracket (\mathsf{A}, \mathsf{B}) \right| \\
&= (|\mathsf{A}|/\tilde{n})^2 \sum_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J} \left( \left| X_3 \llbracket a, I \rrbracket (\mathbb{A}, \mathbb{B}) \right| \pm \alpha m \right) \\
&= (|\mathsf{A}|/\tilde{n})^2 \left| X_3 \llbracket a, J \rrbracket (\mathbb{A}, \mathbb{B}) \right| \pm \alpha \ell \\
\boxed{\text{by Lemma 18(a)}} \quad &= (|\mathsf{A}|/\tilde{n})^2 \ell \pm 2\alpha \ell \,.
\end{aligned}
$$

The proof of (10) is very similar:

$$
\begin{aligned}
\left| X_4 \llbracket a, a', J \rrbracket (\mathsf{A}, \mathsf{B}) \right| &= \sum_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J} \left| X_4 \llbracket a, a', I \rrbracket (\mathsf{A}, \mathsf{B}) \right| \\
&= (|\mathsf{A}|/\tilde{n})^3 \sum_{I \in \mathcal{I}_\mathsf{V}, I \subseteq J} \left( \left| X_4 \llbracket a, a', I \rrbracket (\mathbb{A}, \mathbb{B}) \right| \pm \alpha m \right) \\
&= (|\mathsf{A}|/\tilde{n})^3 \left| X_4 \llbracket a, a', J \rrbracket (\mathbb{A}, \mathbb{B}) \right| \pm \alpha \ell \\
\boxed{\text{by Lemma 18(c)}} \quad &= (|\mathsf{A}|/\tilde{n})^3 \ell \pm 2\alpha \ell \,.
\end{aligned}
$$

For (8), observe that

$$
\left| X_2 \llbracket a, J, c, \overline{J} \rrbracket (\mathsf{A}, \mathsf{B}) \right| = \sum_{\substack{I, I' \in \mathcal{I}_\mathsf{V} \\ I \subseteq J, I' \subseteq \overline{J}}} \left| X_2 \llbracket a, I, c, I' \rrbracket (\mathsf{A}, \mathsf{B}) \right|
$$

and furthermore in the sum at most $2\ell/m$ terms are non-zero. Summing (QUASI2), and observing that we only need to sum the $\alpha m$ error term $2\ell/m$ times, we obtain

$$
\begin{aligned}
\left| X_2 \llbracket a, J, c, \overline{J} \rrbracket (\mathsf{A}, \mathsf{B}) \right| &= (|\mathsf{A}|/\tilde{n})^3 \left| X_2 \llbracket a, J, c, \overline{J} \rrbracket (\mathbb{A}, \mathbb{B}) \right| \pm 2\alpha \ell \\
\boxed{\text{by Lemma 18(d)}} \quad &= \left| \{ (b, b') \in J \times \overline{J} : |b - b'| = c \} \right| \pm 3\alpha \ell \,.
\end{aligned}
$$

$\square$

### 3.4. The algorithm.

The idea of the labelling algorithm is now straightforward. We will label the vertices in order, choosing at each time $t$ to give $v_t$ a vertex label in $J(v_t)$ which has not previously been used, and which induces an edge label on $v_t v_t^+$ which has not previously been used. Unfortunately, this simple version of the algorithm does not quite maintain the quasirandom properties mentioned above, because different intervals $I \in \mathcal{I}_\mathsf{V}$ are contained in different numbers of intervals $J \in \mathcal{J}$; this is the 'technical complication' mentioned in Footnote 3. Indeed, in Figure 3.1 we saw that the labels around the extremes and the centre of $\mathbb{A}$ are used less frequently than those in the intermediate ranges. To correct this we introduce a distribution $\mathrm{Cor}_\mathsf{V}$ on $\mathcal{I}_\mathsf{V} \cup \{*\}$. At each time we in addition randomly sample

from $\mathrm{Cor_V}$, and either do nothing (if $\mathrm{Cor_V}$ returns $*$) or remove a randomly chosen so far unused vertex label from $I$ (if $\mathrm{Cor_V}$ returns $I$).

Most of the mass of $\mathrm{Cor_V}$ is on $*$. That means that the total number of labels removed during the run will be tiny compared to the 'extra' $\gamma n$ labels we are given by Theorem 8. On the other hand, near the extremes and the centre of $\mathbb{A}$ a substantial proportion of the vertex labels will be removed by $\mathrm{Cor_V}$ without serving as vertex labels. These two contrasting properties are consistent since only a small number of vertex labels are near the extremes and the centre of $\mathbb{A}$.

An analogous complication arises when dealing with edge labels $\mathbb{B}$. Indeed, we can see, for example, that for $\tilde{n} - 1 \in \mathbb{B}$ to appear as an edge label on $uv$, we must have chosen $J(u)$ and $J(v)$ to be the two extreme intervals of $\mathcal{J}$, while for $\frac{n}{2} \in \mathbb{B}$ (or any other edge label not close to 0 or $\tilde{n}$) there are $\frac{2\ell^2}{m^2}$ possible choices of the pair $J(u)$, $J(v)$. To deal with these discrepancies, we introduce a suitable distribution $\mathrm{Cor_E}$ on $\mathcal{I_E} \cup \{*\}$. In analogy with $\mathrm{Cor_V}$, we remove edge vertex labels from intervals $I_\mathsf{E}$ chosen according to $\mathrm{Cor_E}$.

We will see that our labelling algorithm labels $v_i$ more or less uniformly in $J = J(v_i)$. If $v_i v_i^+$ is not in $\mathcal{R}$, then the induced edge label on $v_i v_i^+$ is chosen (approximately) from the distribution in which the probability of choosing $c$ is $\hspace{4cm}$ <span style="color:red">$\mathfrak{Q}(J,c)$</span>

$$\mathfrak{Q}(J,c) := \frac{\left| \left\{ (a,a') \in J \times \overline{J} : |a - a'| = c \right\} \right|}{\ell^2} . \tag{11}$$

For convenience, we define $\mathfrak{Q}(J,0)$ according to the above formula (even though 0 is not an edge label).

We shall need the following simple properties of $\mathfrak{Q}(J,c)$.

**Fact 21.** *(a) For any edge label $c$, there are at most $2/\delta_0^2$ sets $J \in \mathcal{J}$ such that $\mathfrak{Q}(J,c) > 0$.*
*(b) For any edge labels $c$ and $c'$ with $|c - c'| \le 2m$, and any $J \in \mathcal{J}$ we have $\left| \mathfrak{Q}(J,c) - \mathfrak{Q}(J,c') \right| \le \frac{2m}{\ell^2}$.*

*Proof.* Part (a) is obvious. For part (b), we can assume that $J$ precedes $\overline{J}$, and that $c \le c'$. Lets expand the nominators in (11) corresponding to $\mathfrak{Q}(J,c)$ and $\mathfrak{Q}(J,c')$. We see that for all but at most $2m$ many pairs $(a,a') \in J \times \overline{J}$ satisfying $a' - a = c$ we also have that $(a, a' + c' - c) \in J \times \overline{J}$, and conversely, for all but at most $2m$ many pairs $(a,a') \in J \times \overline{J}$ satisfying $a' - a = c'$ we also have that $(a, a' + c - c') \in J \times \overline{J}$. This proves the statement. $\square$

We consider this a 'small error', and use the approximation $\sum_{c \in I_\mathsf{E}} \mathfrak{Q}(J(v_i),c) \approx m\mathfrak{Q}(J(v_i), \min(I_\mathsf{E}))$ for each $I_\mathsf{E} \in \mathcal{I_E}$ in order to simplify the definition of $\mathrm{Cor_E}$ below. Our two correction distributions are then defined by the following formulas, $\hspace{3cm}$ <span style="color:red">$\mathrm{Cor_V}$</span>

<span style="color:red">$\mathrm{Cor_E}$</span>

$$\mathbf{P}[\mathrm{Cor_V} = I] = \frac{1 - \frac{m}{\ell}|\{J \in \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|} \qquad \text{for each } I \in \mathcal{I_V}, \tag{12}$$

$$\mathbf{P}[\mathrm{Cor_E} = I_\mathsf{E}] = \frac{1 - m \sum_{J \in \mathcal{J}} \mathfrak{Q}(J, \min(I_\mathsf{E}))}{|\mathcal{J}|} \qquad \text{for each } I_\mathsf{E} \in \mathcal{I_E}, \tag{13}$$

and $\mathbf{P}[\mathrm{Cor}_\mathsf{V} = *] = 1 - \sum_{I \in \mathcal{I}_\mathsf{V}} \mathbf{P}[\mathrm{Cor}_\mathsf{V} = I]$, and $\mathbf{P}[\mathrm{Cor}_\mathsf{E} = *] = 1 - \sum_{I_\mathsf{E} \in \mathcal{I}_\mathsf{E}} \mathbf{P}[\mathrm{Cor}_\mathsf{E} = I_\mathsf{E}]$. Let us briefly justify that these are really probability distributions, that is, that these formulae are all non-negative. By construction of $\mathcal{I}_\mathsf{V}$ and $\mathcal{J}$, each interval of $\mathcal{I}_\mathsf{V}$ is in at most $\frac{\ell}{m}$ intervals of $\mathcal{J}$, so that the expression in (12) is nonnegative. Similarly, by construction $\sum_{J \in \mathcal{J}} \mathfrak{Q}(J, \min(I_\mathsf{E}))$ is at most $\frac{1}{m}$ for each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, so that (13) is nonnegative. Note that for 'most' $I$ and $I_\mathsf{E}$, actually (12) and (13) evaluate to zero. Now, we have

$$\mathbf{P}[\mathrm{Cor}_\mathsf{V} = *] = 1 - \frac{|\mathcal{I}_\mathsf{V}| - \frac{m}{\ell}|\{(I, J) \in \mathcal{I}_\mathsf{V} \times \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|}.$$

Any $J \in \mathcal{J}$ contains exactly $\frac{\ell}{m}$ intervals $I$ from $\mathcal{I}_\mathsf{V}$. Hence

$$\mathbf{P}[\mathrm{Cor}_\mathsf{V} = *] = \frac{|\mathcal{J}| - |\mathcal{I}_\mathsf{V}| + |\mathcal{J}|}{|\mathcal{J}|} \tag{14}$$

which is nonnegative (and in fact very close to 1) by (3) and (2). Finally, we have

$$\mathbf{P}[\mathrm{Cor}_\mathsf{E} = *] = 1 - \frac{|\mathcal{I}_\mathsf{E}| - m \sum_{I_\mathsf{E} \in \mathcal{I}_\mathsf{E}} \sum_{J \in \mathcal{J}} \mathfrak{Q}(J, \min(I_\mathsf{E}))}{|\mathcal{J}|}.$$

Fix $J \in \mathcal{J}$ and consider the set of differences $\{km : km = |a - a'|, (a, a') \in J \times \overline{J}\}$. This is a set of size $2(\frac{\ell}{m} - 1) + 1$. The largest and the smallest labels in this set can be written as a difference of elements from $J \times \overline{J}$ in exactly $m$ ways each, the second largest and second smallest labels can be written as a difference in $2m$ ways each, and so on. For any $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, $\min(I_\mathsf{E})$ is a multiple of $m$, hence

$$\mathbf{P}[\mathrm{Cor}_\mathsf{E} = *] = \frac{|\mathcal{J}| - |\mathcal{I}_\mathsf{E}| + m|\mathcal{J}| \sum_{j=-\ell/m}^{\ell/m} \frac{\ell - |jm|}{\ell^2}}{|\mathcal{J}|} = \frac{2|\mathcal{J}| - |\mathcal{I}_\mathsf{E}|}{|\mathcal{J}|}, \tag{15}$$

which again by (3) and (2) is very close to 1, and in particular nonnegative.

Given sets $\mathsf{A} \subseteq \mathbb{A}$ and $\mathsf{B} \subseteq \mathbb{B}$, an interval $I \subseteq \mathbb{A}$ and a label $a$, we say that $a' \in \mathbb{A}$ is **admissible label** *admissible for $a$ and $I$ with respect to $\mathsf{A}$ and $\mathsf{B}$* if $a' \in A \cap I$ and $|a' - a| \in C$. We let the set of admissible vertices for $a$ and $I$ with respect to $\mathsf{A}$ and $\mathsf{B}$ be $\mathrm{Adm}(a, I; \mathsf{A}, \mathsf{B})$. Observe that we have

$$|X_3 [\![a, I]\!] (\mathsf{A}, \mathsf{B})| = |\mathrm{Adm}(a, I; \mathsf{A}, \mathsf{B})|. \tag{16}$$

We generate a labelling of $V(T)$ by Algorithm 1. For $t = 1$, at lines 3 and 6 of Algorithm 1 we define

$$\mathrm{Adm}(\psi_0(v_1^+), J(v_1); \mathsf{A}_1, \mathsf{B}_1) := J(v_1) \quad \text{and} \quad \{|\psi_1(v_1^+) - a|\} := \emptyset. \tag{17}$$

**Algorithm 1:** Labeling of $T$.

---

**1** Let $\psi_0 := \emptyset$, let $\mathsf{A}_1 := \mathbb{A}$ and let $\mathsf{B}_1 := \mathbb{B}$ ;

**2 foreach** $t = 1, \ldots, n$ **do**

    **3**    Choose $a \in \mathrm{Adm}\big(\psi_{t-1}(v_t^+), J(v_t); \mathsf{A}_t, \mathsf{B}_t\big)$ uniformly at random     $\triangleright$ may fail,
                                                                           see (17) for $t = 1$;

    **4**    $\psi_t := \psi_{t-1} \cup \{v_t \mapsto a\}$                         $\triangleright$ enhance the partial labelling;

    **5**    $\mathsf{A}_t^r := \mathsf{A}_t \setminus \{a\}$                              $\triangleright$ remove corresponding vertex label;

    **6**    $\mathsf{B}_t^r := \mathsf{B}_t \setminus \{|\psi_t(v_t^+) - a|\}$     $\triangleright$ remove corresponding edge label, see (17) for $t = 1$;

    **7**    Sample $x$ from $\mathrm{Cor}_\mathsf{V}$                        $\triangleright$ correction on vertex labels;

    **8**    **if** $x = I \subseteq \mathcal{I}_\mathsf{V}$ **then**

    **9**         Choose $r_t^v \in I \cap \mathsf{A}_t^r$ uniformly at random                      $\triangleright$ may fail;

   **10**         $\mathsf{A}_{t+1} := \mathsf{A}_t^r \setminus \{r_t^v\}$ ;

   **11**    **end**

   **12**    **else**

   **13**         $\mathsf{A}_{t+1} := \mathsf{A}_t^r, r_t^v := *$ ;

   **14**    **end**

   **15**    Sample $y$ from $\mathrm{Cor}_\mathsf{E}$                      $\triangleright$ correction on edge labels;

   **16**    **if** $y = I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ **then**

   **17**         Choose $r_t^e \in I_\mathsf{E} \cap \mathsf{B}_t^r$ uniformly at random                   $\triangleright$ may fail;

   **18**         $\mathsf{B}_{t+1} := \mathsf{B}_t^r \setminus \{r_t^e\}$ ;

   **19**    **end**

   **20**    **else**

   **21**         $\mathsf{B}_{t+1} := \mathsf{B}_t^r, r_t^e := *$ ;

   **22**    **end**

**23 end**

**24 return** $\psi_n$ ;

---

3.5. **Probabilistic formalities.** To apply Lemma 6, as we will want to do, we need a probability space $\Omega$ and a filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots$. Let $\Omega$ be the set of sequences of length $3n$ over the alphabet $[\tilde{n}] \cup \{*, \mathrm{fail}\}$. We generate a sequence in $\Omega$ from a run of Algorithm 1 by recording, for each time $t = 1, \ldots, n$, the vertex label chosen at line 3, the choice of $r_t^v$ , and the choice of $r_t^e$. In the event that the algorithm fails — which occurs when it requests to sample a uniform element from an empty set in lines 3, 9, or 17 — we record fail at the point when the algorithm fails and in all remaining places of the sequence. We obtain a probability measure on $\Omega$ as the probability that running Algorithm 1 generates a given sequence.

When we use Lemma 6, we will have random variables $Y_1, \ldots, Y_n$ tracking sequential contributions to some parameter. Each $Y_i$ is determined by some initial segment of $\big([\tilde{n}] \cup \{*, \mathrm{fail}\}\big)^{3n}$,

called a *history*, $\mathscr{H}_i(\omega)$ of $\omega \in \Omega$, we have an estimate for $\sum_i \mathbf{E}[Y_i | \mathscr{H}_{i-1}]$, and the lengths of these initial segments are monotone increasing. Since the lengths of the histories are increasing, they generate in the natural way a filtration on $\Omega$, as required for Lemma 6.

In the rest of the paper, we will not need the details of this construction of $\Omega$, but simply the observation that conditioning on some history is equivalent to conditioning on the behaviour of Algorithm 1 up to a given point, and that Lemma 6 applies to random variables of the above type.

## 4. Proof of Theorem 8

4.1. **Technical overview.** Before starting the proof, we give a brief overview of the structure. Ultimately, all we need to do is show that Algorithm 1 runs successfully with positive probability. We will show something rather stronger, namely that in fact with high probability, at each time $t$ in the running of Algorithm 1 the pair $(\mathsf{A}_t, \mathsf{B}_t)$ is $\delta_t$-quasirandom. This is a stronger claim because this quasirandomness in particular asserts that the sets from which labels are chosen at lines 3, 9 and 17 are non-empty.

In turn, to prove $\delta_t$-quasirandomness of $(\mathsf{A}_t, \mathsf{B}_t)$, we consider separately each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ for (QUASI1) and each $X \in \mathcal{X}$ for (QUASI2). We describe our approach for a given $X \in \mathcal{X}$; that for $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ is analogous. We can write $\left|X(\mathsf{A}_t, \mathsf{B}_t)\right|$ as $\left|X(\mathbb{A}, \mathbb{B})\right|$ minus the (random) change caused by the (random) choices of $\psi_1(v_1)$, $r_1^v$, $r_1^e$, $\psi_2(v_2)$, and so on up to $r_{t-1}^e$. Thus what we want to do is estimate the sum of a collection of random variables. These random variables are sequentially dependent, so that we can use Lemma 6 to provide such an estimate. We will see that the probability bounds coming from Lemma 6 are strong enough to simply use the union bound over all choices of $X$ and $t$, completing the proof.

The difficulty in this programme is that in order to apply Lemma 6 we need estimates for the expected changes at each step, conditioned on the history up to that step. In order to obtain these estimates, we need to know that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for earlier $i$. This may seem suspiciously circular: but it is not. To see this, consider the first time $t$ at which quasirandomness fails. This is the first time at which some sum of changes deviates excessively from its expected value. The probability of this event is bounded by Lemma 6 in terms of the sum of the conditional expectations of changes, and those conditional expectations are calculated assuming $\delta_i$-quasirandomness of $(\mathsf{A}_i, \mathsf{B}_i)$ for some values $i < t$, in other words for times $i$ when, because $i < t$ and $t$ is the first time at which quasirandomness fails, we do have this quasirandomness. Lemma 6 then tells us that the event of quasirandomness first failing at time $t$ is unlikely, and sufficiently unlikely that taking a union bound over $t$ we conclude that quasirandomness failing at any time is unlikely.

Let us now discuss how we obtain these sums of conditional expectations. The *removal term* change caused by $r_i^v$ depends only on the sets $\mathsf{A}_i^r$ and $\mathsf{B}_i^r$, and similarly the removal term change caused by $r_i^e$ depends only on $\mathsf{A}_{i+1}$ and $\mathsf{B}_i^r$. It is thus quite easy to estimate the sums of conditional expectations of these changes, which we do in Claim 3. It is rather harder to

estimate the change caused by $\psi_i(v_i)$, which in addition to the sets $\mathsf{A}_i$ and $\mathsf{B}_i$ depends also on the choice of $v_i^+$, which in turn depends on earlier labellings, and so on; ultimately there is some dependence on the entire labelling. Analysing this seems at first hopeless. But in fact there is only significant dependence on $v_i^+$; we will show that, assuming quasirandomness, any choice of $(v_i^+)^+$ leads to the conditional expectation of change when labelling $v_i$ being approximately a quantity, $p_{X,i}$, which we call the *crude estimate*. This quantity $p_{X,i}$ does not depend at all on the labelling process, and thus we can quite easily estimate the sum over $i$ of the $p_{X,i}$'s. Putting this estimate together with the estimated sum of removal terms, which we do in Claim 3, yields the 'correct' value of $\big|X(\mathbb{A}, \mathbb{B})\big| - \big|X(\mathsf{A}_t, \mathsf{B}_t)\big|$. In other words, it is enough to show that the sum of crude estimates corresponds to the actual changes caused by labelling.

This is still not an easy task. We perform it in two steps. First, we argue in Claim 4 that $p_{X,i}$ is approximately the expectation of change caused by labelling $v_i$, conditioned on the history up to immediately before labelling $v_i^+$. We define a *fine estimate* $q_{X,i}$ which corresponds to the expectation of change caused by labelling $v_i$, conditioned on the history up to immediately after labelling $v_i^+$. An application of Lemma 6 then tells us that with high probability the sum of the $p_{X,i}$'s is approximately the sum of the $q_{X,i}$'s. Now, $q_{X,i}$ is still not the conditional expectation we would like to find: some vertices may be labelled in between labelling $v_i^+$ and labelling $v_i$, and these labellings, together with removal terms in the same interval, cause $q_{X,i}$ and the expectation of change caused by labelling $v_i$, conditioned on the history up to immediately before labelling $v_i$, to be different. But provided there are only few such intervening vertices, the difference is small. Our choice of order, using (PRE3), guarantees that for most $i$ there are indeed few such intervening vertices, and we conclude that (deterministically) the sum of the $q_{X,i}$ is close to the sum over $i$ of the expectation of change caused by labelling $v_i$, conditioned on the history up to immediately before labelling $v_i$. This last sum is what we need in order to apply Lemma 6 to estimate the sum of the actual changes caused by labelling the $v_i$, which completes the proof.

In total, then, since the $p_{X,i}$ are quantities independent of the labelling process we do not need to assume anything to estimate their sum in Claim 3. To show that their sum approximates the sum of the $q_{X,i}$'s with high probability, which we do in Claim 4, and to show that the sum of the $q_{X,i}$'s is with high probability close to the sum of the actual changes, we need to assume quasirandomness before the time when we label $v_i$. As this is before time $i$, as discussed this assumption is valid.

### 4.2. Proof of Theorem 8. Given $\gamma > 0$, let constants and sets be as defined in Setup 9. Given an $n$-vertex tree $T$ with $\Delta(T) \leq \frac{\eta n}{\log n}$, Lemma 11 produces an edge set $\mathcal{R}$, an ordering $V(T) = \{v_1, \ldots, v_n\}$ and intervals $J(v_i)$.

In order to apply Lemma 6, we will twice need to use the following upper bound on $\sum_{v \in V(T)} \deg(v)^2$.

$$\sum_{v \in V(T)} \deg(v)^2 \le \Delta(T) \cdot \sum_{v \in V(T)} \deg(v) \le \frac{\eta n}{\log n} \cdot 2e(T) \le \frac{2\eta n^2}{\log n} \,. \tag{18}$$

We run Algorithm 1. We say the algorithm *fails* if at any time a step is not possible: in other words, if the sets from which we should choose uniformly $\psi_t(v_t)$, $r_t^v$, or $r_t^e$, are empty. Then Theorem 8 holds if with positive probability Algorithm 1 does not fail. We will show that with high probability Algorithm 1 maintains the property that $(\mathsf{A}_t, \mathsf{B}_t)$ is $\delta_t$-quasirandom for each $1 \le t \le n$.

For each $1 \le \tau \le n$, let us define the following two events:

$$\mathcal{W}_{<\tau} := \{\text{Algorithm 1 has not failed before time } \tau\} \,,$$
$$\mathcal{U}_\tau := \{\text{we have } |\mathsf{A}_\tau|, |\mathsf{B}_\tau| = \tilde{n} - \tau \pm 10\ell\} \tag{19}$$

**Claim 1.** *We have for the probability of the 'bad event'* $\bigcup_{\tau=1}^n (\mathcal{W}_{<\tau} \setminus \mathcal{U}_\tau)$ *we have*

$$\mathbf{P}\left[\bigcup_{\tau=1}^n (\mathcal{W}_{<\tau} \setminus \mathcal{U}_\tau)\right] \le n^{-1} \,.$$

Rephrasing the claim, with probability at least $1 - n^{-1}$, at each time $1 \le \tau \le n$, it holds that Algorithm 1 has failed before time $\tau$ or we have (19).

*Proof of Claim 1.* Observe that, unless the algorithm fails, in each step after the first (in which no edge label is given to any edge) one vertex label and one edge label is used in the labelling, so $|\mathsf{A}_\tau| \le \tilde{n} - \tau + 1$ and $|\mathsf{B}_\tau| \le |\mathbb{B}| - \tau + 2 = \tilde{n} - \tau + 1$, as needed for the upper-bound.

Let us now turn to the lower-bound. For $t \in [n]$, let $U_t$ be defined as follows:

(a) if Algorithm 1 has not failed until step $t$, let $U_t$ be the indicator that $*$ was not sampled on Line 7 (in step $t$),

(b) if Algorithm 1 has failed before step $t$, let $U_t$ be a Bernoulli random variable with success probability $\frac{2(\ell - m)}{\tilde{n} - 2(\ell - m)}$ (and independent of all other random choices).

For $t \in [n]$, let $W_t$ be defined as follows in the same way, except that in case (a), we use the indicator that $*$ was not sampled on Line 15.

By (3), (14) and (15), we have

$$\mathbf{P}[\mathrm{Cor}_\mathsf{V} \ne *] = \mathbf{P}[\mathrm{Cor}_\mathsf{E} \ne *] = \frac{2(\ell - m)}{\tilde{n} - 2(\ell - m)} \,.$$

Therefore, $U_t$'s and $W_t$'s are independent Bernoulli random variable with success probability $\frac{2(\ell-m)}{\tilde{n}-2(\ell-m)} < \frac{2\ell}{n}$. By Theorem 5, the probability that $\sum_{t=1}^n U_t > 10\ell$ or $\sum_{t=1}^n W_t > 10\ell$ is at most $n^{-1}$. That is, with probability at least $1 - n^{-1}$ we have that Algorithm 1 failed or $|\mathsf{A}_n|, |\mathsf{B}_n| \ge \tilde{n} - n - 10\ell$. Observe that in the case of this good event, we also get $|\mathsf{A}_\tau|, |\mathsf{B}_\tau| \ge \tilde{n} - \tau - 10\ell$ for each $\tau \le n$, no matter at which times the potential non-$*$ samples were sampled. □

The next claim tells us that Algorithm 1 does not fail at line 3 (Claim 2(a)), line 9 (Claim 2(b)), nor at line 17 (Claim 2(c)).

**Claim 2.** *Suppose that* $(\mathsf{A}_t, \mathsf{B}_t)$ *is* $\delta_t$-*quasirandom and* (19) *holds. Then we have*

(a) $\left| X_3 \llbracket a, I \rrbracket (\mathsf{A}_t, \mathsf{B}_t) \right| > 0$,

(b) $\left| X_1 \llbracket I \rrbracket (\mathsf{A}_t, \mathsf{B}_t) \right| > 0$, *and*

(c) $|I_\mathsf{E} \cap \mathsf{B}_t^r| > 0$.

*Proof.* Let us first prove (a). By choice of $\delta_t$, for each $a \in \mathbb{A}$ and $I \in \mathcal{I}_\mathsf{V}$, using (QUASI2) we have

$$\left| X_3 \llbracket a, I \rrbracket (\mathsf{A}_t, \mathsf{B}_t) \right| = \left| X_3 \llbracket a, I \rrbracket (\mathbb{A}, \mathbb{B}) \right| \cdot \left( \frac{|\mathsf{A}_t|}{\tilde{n}} \right)^2 \pm \delta_t m$$

$$\boxed{\text{by Lemma 18(b) and (19)}} \quad \geq m \cdot \left( \frac{\tilde{n} - n - 10\ell}{\tilde{n}} \right)^2 - \delta_t m$$

$$\boxed{\text{Setup 9, and } \delta_t \leq \delta_n = \mu^{1/\mu}} \quad \geq \left( \gamma - 10\delta_0^2 - \mu^{1/\mu} \right) m > 0.$$

Mutatis mutandis we obtain (b).

Last, let us turn to (c). We have

$$|I_\mathsf{E} \cap \mathsf{B}_t^r| = |I_\mathsf{E} \cap \mathsf{B}_t| \pm 1 \overset{\text{(QUASI1)}}{=} m \cdot \frac{|\mathsf{A}_t|}{\tilde{n}} \pm \delta_t m \pm 1 \overset{\text{(19)}}{\geq} m \cdot \frac{\tilde{n} - n - 10\ell}{\tilde{n}} - \delta_t m - 1 > 0.$$

$\square$

Thus, in order to prove Theorem 8, it is enough to show that with high probability $(\mathsf{A}_t, \mathsf{B}_t)$ is $\delta_t$-quasirandom for each $1 \leq t \leq n$. We now embark upon proving this.

Since the vertex labels $r_t^v$ and $r_t^e$ are chosen uniformly at random within intervals of respectively $\mathcal{I}_\mathsf{V}$ and $\mathcal{I}_\mathsf{E}$, it is quite easy to analyse their effect on (QUASI1) and (QUASI2). It is rather harder to analyse the effect of the edge and vertex labels used at step $t$, since these are not chosen uniformly. However, the idea one should have in mind is that this choice is 'close to uniform' in a sense we will make precise later, and thus it is useful to write down 'crude estimates' for the effect of the vertex and edge labels used at step $t$ in the labelling which pretends these choices are really uniform. Specifically, the following estimates correspond (more or less) to the expected effect if $\phi_t(v_t)$ were chosen uniformly from the unused vertex labels in $J(v_t)$, if the edge label $\left| \phi_t(v_t) - \phi_t(v_t^+) \right|$ were chosen to be $c \in \mathsf{B}_t$ with probability proportional to $\mathfrak{Q}(J(v_t), c)$, independently, and if (QUASI1) and (QUASI2) held with zero error at time $t$. Of course all these assumptions are false, but we will see that 'on average' they hold, which is enough for our proof.

For $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ and $1 \leq t \leq n$, we define

$$\begin{aligned} \textit{the crude estimate} \quad & p_{I_\mathsf{E}, t} := m \cdot \mathfrak{Q}(J(v_t), \min(I_\mathsf{E})) \quad \text{and} \\ \textit{the removal term} \quad & r_{I_\mathsf{E}, t} := \mathbb{1}_{r_t^e \in I_\mathsf{E}}. \end{aligned} \tag{20}$$

$p_{I_\mathsf{E}, t}$

$r_{I_\mathsf{E}, t}$

The crude estimate is an idealised version of the expected number of edge labels in $I_\mathsf{E}$ that are used at time $t$ (equivalently, to the probability that at time $t$ we use an edge label in the interval $I_\mathsf{E}$). The removal term is the indicator of the event that at time $t$ we remove an edge label from $I_\mathsf{E}$. For $X \in \mathcal{X}$ we define similar terms, with the same intent. Again, the crude estimate $p_{X,t}$ is an estimate for the expected change $\big|X(\mathsf{A}_t, \mathsf{B}_t)\big| - \big|X(\mathsf{A}_t^r, \mathsf{B}_t^r)\big|$, and the removal term $r_{X,t}$ is the actual change $\big|X(\mathsf{A}_t^r, \mathsf{B}_t^r)\big| - \big|X(\mathsf{A}_{t+1}, \mathsf{B}_{t+1})\big|$. For the latter, recall that $\mathsf{A}_t^r$ and $\mathsf{B}_t^r$ are the available vertex and edge labels, respectively, at time $t$ after removing the vertex label and edge label used in labelling $T$. So,

<span style="color:red">$p_{X,t}$<br>$r_{X,t}$</span>

$$p_{X,t} := \frac{|X(\mathbb{A},\mathbb{B})|(\tilde{n}-t)^{\mathrm{free}(X)-1}}{\tilde{n}^{\mathrm{free}(X)-1}} \Big( \sum_{I \in \mathrm{Free}_\mathsf{V}(X)} \frac{\mathbb{1}_{I \subseteq J(v_t)}}{\ell} + \sum_{\mathfrak{e} \in \mathrm{Free}_\mathsf{E}(X)} \mathfrak{Q}(J(v_t), \mathrm{Diff}(\mathfrak{e}; X)) \Big), \text{ and}$$

$$r_{X,t} := \Big| \big\{ X' \in X(\mathsf{A}_t^r, \mathsf{B}_t^r) : r_t^v \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } r_t^e \in \mathrm{Ch}_\mathsf{E}(X'; X) \big\} \Big|. \tag{21}$$

Note that the crude estimates $p_{I_\mathsf{E},t}$ and $p_{X,t}$ are determined before the algorithm starts, and so are their corresponding partial sums $\sum_{i=1}^{t} p_{I_\mathsf{E},i}$ and $\sum_{i=1}^{t} p_{X,i}$. By contrast, the partial sums of the removal terms, $\sum_{i=1}^{t} r_{I_\mathsf{E},i}$ and $\sum_{i=1}^{t} r_{X,i}$, are sums of random variables which in principle depend upon all of the random choices we make throughout the labelling. However, if we assume that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \leq i \leq t-1$ then we can obtain good bounds on these partial sums which hold with high probability by considering only the choice of the $r_i^e$ and $r_i^v$.

We estimate the partial sums $\sum_{i=1}^{t} r_{I_\mathsf{E},i}$ and $\sum_{i=1}^{t} r_{X,i}$ together. The reason is that eventually we will be able to show that (for example) $\sum_{i=1}^{t} p_{I_\mathsf{E},i}$ is with high probability a good estimate for the number of edge labels in $I_\mathsf{E}$ used in the labelling up to time $t$, and it follows that $\sum_{i=1}^{t}(p_{I_\mathsf{E},i} + r_{I_\mathsf{E},i})$ is a good estimate for $|\mathbb{B} \cap I_\mathsf{E}| - |\mathsf{B}_t \cap I_\mathsf{E}|$, which is what we want to know in order to verify (QUASI1). Recall that in the introduction we mentioned that our proof can be seen as an application of the Differential Equations Method. This claim is where we (implicitly) verify that the crude estimates we chose actually correspond to solutions to some differential equations: one should understand the right hand sides of (22) and (23) as (what we expect for) the differences $|\mathbb{B} \cap I_\mathsf{E}| - |\mathsf{B}_t \cap I_\mathsf{E}|$ and $|X(\mathbb{A},\mathbb{B})| - |X(\mathsf{A}_t, \mathsf{B}_t)|$, respectively.

**Claim 3.** *With probability at least $1 - 2n^{-1}$, for each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, each $X \in \mathcal{X}$, and each $1 \leq t \leq n$, if $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \leq i < t$ we have*

$$\sum_{i=1}^{t}(p_{I_\mathsf{E},i} + r_{I_\mathsf{E},i}) = \frac{t}{|\mathcal{J}|} \pm \tfrac{1}{4}\delta_t m \ , \ and \tag{22}$$

$$\sum_{i=1}^{t}(p_{X,i} + r_{X,i}) = |X(\mathbb{A},\mathbb{B})| \frac{\tilde{n}^{\mathrm{free}(X)} - (\tilde{n}-t)^{\mathrm{free}(X)}}{m|\mathcal{J}|\tilde{n}^{\mathrm{free}(X)-1}} \pm \tfrac{1}{4}\delta_t m\,. \tag{23}$$

*Proof of Claim 3.* We have

$$\sum_{i=1}^{t} p_{I_E,i} = \sum_{i=1}^{t} \sum_{J \in \mathcal{J}} \mathbb{1}_{J(v_i)=J} \cdot p_{I_E,i}$$

$$\boxed{\text{by (20)}} \quad = \sum_{i=1}^{t} \sum_{J \in \mathcal{J}} \mathbb{1}_{J(v_i)=J} \cdot m\mathfrak{Q}(J, \min(I_E))$$

$$\boxed{\text{by (PRE4), with } S = [t]} \quad = \left(\tfrac{t}{|\mathcal{J}|} \pm \delta^2 n\right) \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_E)). \tag{24}$$

Let us now turn to the quantity $\sum_{i=1}^{t} r_{I_E,i}$. We have

$$\mathbf{E}\left[\sum_{i=1}^{t} r_{I_E,i}\right] \overset{(20)}{=} \mathbf{E}\left[\sum_{i=1}^{t} \mathbb{1}_{r_i^e \in I_E}\right]$$

$$\boxed{\text{by the way } r_i^e \text{ is chosen on lines 15 and 17 of Algorithm 1}} \quad = t \cdot \mathbf{P}[\mathrm{Cor}_E = I_E]$$

$$\boxed{\text{by (13)}} \quad = t \cdot \frac{1 - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_E))}{|\mathcal{J}|}.$$

The events $\{r_i^e \in I_E\}_{i=1}^{t}$ are independent, and thus Theorem 5 gives us that

$$\sum_{i=1}^{t} r_{I_E,i} = t \cdot \frac{1 - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_E))}{|\mathcal{J}|} \pm \delta^2 n$$

with probability at least $1 - n^{-10}$. Putting this together with (24), we get that (22) holds for that $t$ and that $I_E$ with probability at least $1 - n^{-10}$. Taking the union bound over all choices of $t$ and $I_E \in \mathcal{I}_E$, we see that with probability at least $1 - n^{-2}$, (22) holds for all $t$ and $I_E$ as desired.

For (23) we need to be a little more careful: the quantity $p_{X,i}$ depends on $i$ as well as $J(v_i)$, and (with similar effect) the quantity $r_{X,i}$ depends on $|X(A_i^r, B_i^r)|$ and $|A_i^r \cap I|$ and $|B_i^r \cap I_E|$ for each $I \in \mathcal{I}_V$ and $I_E \in \mathcal{I}_E$ as well as the outcomes of $\mathrm{Cor}_V$ and $\mathrm{Cor}_E$. We divide the interval $[t]$ into intervals $S_1, \ldots, S_{\lceil t/\delta n \rceil}$, all except possibly the last consisting of $\delta n$ elements. Note $S_j$ that by assumption $\delta n$ is an integer which divides $n$. The point of doing this is that any time $i \in S_j$, $|X(A_i^r, B_i^r)|$, $|A_i^r \cap I|$ and $|B_i^r \cap I_E|$ are up to a small error constant on any $S_j$.

For $i \in S_j$, since $\tilde{n} - j\delta n \geq \gamma n$, we have

$$\tilde{n} - i \leq \tilde{n} - j\delta n + \delta n \leq (\tilde{n} - j\delta n)\left(1 + \tfrac{\delta}{\gamma}\right).$$

Thus, for any $1 \leq s \leq 3$,

$$(\tilde{n} - i)^s \leq (\tilde{n} - j\delta n)^s \left(1 + \tfrac{\delta}{\gamma}\right)^s \leq \left(1 + 8\delta\gamma^{-1}\right)(\tilde{n} - j\delta n)^s,$$

and hence $(\tilde{n} - i)^s = \left(1 \pm 8\delta\gamma^{-1}\right)(\tilde{n} - j\delta n)^s$. Using this and (PRE4), for each $X \in \mathcal{X}$ we have

$$
\sum_{i \in S_j} p_{X,i} = \left(1 \pm \tfrac{8\delta}{\gamma}\right) \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\text{free}(X)-1}}{\tilde{n}^{\text{free}(X)-1}} \sum_{I \in \text{Free}_{\mathsf{V}}(X)} \left(\tfrac{|S_j|}{|\mathcal{J}|} \pm \delta^2 n\right) \frac{|\{J \in \mathcal{J} : I \subseteq J\}|}{\ell}
$$
$$
+ \left(1 \pm \tfrac{8\delta}{\gamma}\right) \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\text{free}(X)-1}}{\tilde{n}^{\text{free}(X)-1}} \sum_{\mathfrak{e} \in \text{Free}_{\mathsf{E}}(X)} \sum_{J \in \mathcal{J}} \left(\tfrac{|S_j|}{|\mathcal{J}|} \pm \delta^2 n\right) \mathfrak{Q}(J, \text{Diff}(\mathfrak{e}; X)) \tag{25}
$$

$$
= \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\text{free}(X)-1}}{\tilde{n}^{\text{free}(X)-1}} \sum_{I \in \text{Free}_{\mathsf{V}}(X)} \frac{|S_j| \cdot |\{J \in \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|\ell}
$$
$$
+ \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\text{free}(X)-1}}{\tilde{n}^{\text{free}(X)-1}} \sum_{\mathfrak{e} \in \text{Free}_{\mathsf{E}}(X)} \sum_{J \in \mathcal{J}} \frac{|S_j|}{|\mathcal{J}|} \mathfrak{Q}(J, \text{Diff}(\mathfrak{e}; X)) \pm 100\delta^2 n\gamma^{-1} . \tag{26}
$$

Let us hint where the value of the final error term $\pm 100\delta^2 n\gamma^{-1}$ in (26) comes from. There are two error terms in (25). To bound the error introduced by the term $\pm\frac{8\delta}{\gamma}$, we use that

$$
\frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\text{free}(X)-1}}{\tilde{n}^{\text{free}(X)-1}} \overset{\text{F17(b)}}{\leq} m ,
$$
$$
\sum_{I \in \text{Free}_{\mathsf{V}}(X)} \left(\tfrac{|S_j|}{|\mathcal{J}|} \pm \delta^2 n\right) \cdot \frac{|\{J \in \mathcal{J} : I \subseteq J\}|}{\ell} \leq 2\left(\tfrac{\delta n}{2\frac{n}{m}} + \delta^2 n\right) \cdot \frac{\delta_0^{-2}}{\ell} \leq 5\delta , \text{ and}
$$
$$
\sum_{\mathfrak{e} \in \text{Free}_{\mathsf{E}}(X)} \sum_{J \in \mathcal{J}} \left(\tfrac{|S_j|}{|\mathcal{J}|} \pm \delta^2 n\right) \cdot \mathfrak{Q}(J, \text{Diff}(\mathfrak{e}; X)) \overset{\text{F21(a)}}{\leq} 2 \cdot 2/\delta_0^2 \cdot \left(\tfrac{\delta n}{2\frac{n}{m}} + \delta^2 n\right) \cdot \frac{1}{\ell} \leq 5\delta .
$$

The error coming from the term $\pm\delta^2 n$ can be bounded similarly.

We now estimate $\sum_{i \in S_j} r_{X,i}$. We will use Lemma 6 to do this. To that end, for each $i = 0, \ldots, n-1$, let $\mathscr{H}_i$ be the history up to and including the choice of $\psi_{i+1}(v_{i+1})$, and let $\mathscr{H}_n$ be the complete history. Hence, the difference between $\mathscr{H}_{n-1}$ and $\mathscr{H}_n$ is only in the information about the choice of $r_n^v$ and $r_n^e$. Let $\mathcal{E}_t$ be the event that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \leq i < t$ from which we subtract the event $\bigcup_{\tau=1}^n (\mathcal{W}_{<\tau} \setminus \mathcal{U}_\tau)$ from Claim 1.[4] Suppose now that $\mathcal{E}_t$ occurs. That is in the calculations below, we shall work with an arbitrary conditional subspace $\mathscr{H}_i$, for some $i < t$, but only with such that $\mathscr{H}_i \cap \mathcal{E}_t$ has positive probability.

Since $(\mathsf{A}_i^r, \mathsf{B}_i^r)$ differs by one vertex and one edge label[5] from $(\mathsf{A}_i, \mathsf{B}_i)$, and for any given $X \in \mathcal{X}$ these two labels meet at most three $X' \in X(\mathsf{A}_i, \mathsf{B}_i)$, we have $\left|X(\mathsf{A}_i^r, \mathsf{B}_i^r)\right| = \left|X(\mathsf{A}_i, \mathsf{B}_i)\right| \pm 3$. Furthermore, $(\mathsf{A}_{i+1}, \mathsf{B}_i^r)$ differs by at most one vertex label from $(\mathsf{A}_i^r, \mathsf{B}_i^r)$, and this vertex label meets at most one $X' \in X(\mathsf{A}_i^r, \mathsf{B}_i^r)$. Thus, using (QUASI2) to estimate $\left|X(\mathsf{A}_i, \mathsf{B}_i)\right|$, for

---

[4]We emphasize that the event $\bigcup_{\tau=1}^n (\mathcal{W}_{<\tau} \setminus \mathcal{U}_\tau)$ involves conditions even on times $\tau > t$.

[5]with the only exception $i = 1$ when we have $\mathsf{B}_i^r = \mathsf{B}_i$

all $i \in S_j$ we have

$$
\left| X(\mathsf{A}_{i+1}, \mathsf{B}_i^r) \right|, \left| X(\mathsf{A}_i^r, \mathsf{B}_i^r) \right| = \frac{|X(\mathbb{A}, \mathbb{B})| \cdot |\mathsf{A}_i|^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \pm \delta_i m \pm 4
$$

$$
\overset{(19)}{=} \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - i \pm 10\ell)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \pm \delta_i m \pm 4 \tag{27}
$$

$$
= \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \pm 2\delta_i m \; .
$$

Given $I \in \mathcal{I}_\mathsf{V}$, by (QUASI2) with the structure $X_1 \llbracket I \rrbracket$, we have

$$
|\mathsf{A}_i^r \cap I| = m \cdot \frac{\tilde{n} - i \pm 10\ell}{\tilde{n}} \pm \delta_i m = m \cdot \frac{\tilde{n} - j\delta n}{\tilde{n}} \pm 2\delta_i m \; . \tag{28}
$$

Using this together with (27), since $r_i^v$ is chosen uniformly in $\mathsf{A}_i^r \cap I$ for an interval $I \in \mathcal{I}_\mathsf{V}$ drawn from $\mathrm{Cor}_\mathsf{V}$, we have

$$
\mathbf{E}\left[ \left| \left\{ X' \in X(\mathsf{A}_i^r, \mathsf{B}_i^r) : r_i^v \in \mathrm{Ch}_\mathsf{V}(X'; X) \right\} \right| \,\middle|\, \mathscr{H}_{i-1} \right]
$$

$$
\boxed{\text{by (27) and (28)}} \quad = \sum_{I \in \mathrm{Free}_\mathsf{V}(X)} \mathbf{P}[\mathrm{Cor}_\mathsf{V} = I] \frac{\frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \pm 2\delta_i m}{m \frac{\tilde{n} - j\delta n}{\tilde{n}} \pm 2\delta_i m}
$$

$$
\boxed{\text{by (12)}} \quad = \left( \sum_{I \in \mathrm{Free}_\mathsf{V}(X)} \frac{1 - \frac{m}{\ell} |\{J \in \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|} \right) \cdot \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X) - 1}}{m \tilde{n}^{\mathrm{free}(X) - 1}} \pm \frac{10\delta_i}{\gamma |\mathcal{J}|} \; .
$$

Observe that, since the free vertex labels of $X \in \mathcal{X}$ are distinct members of $\mathcal{I}_\mathsf{V}$, they are disjoint and hence

$$
\text{any given vertex label is in } \mathrm{Ch}_\mathsf{V}(X'; X) \text{ for at most one } X' \in X(\mathbb{A}, \mathbb{B}). \tag{29}
$$

Let us now fix an interval $S_j$. We apply Lemma 6, with $|S_j|$ many random variables $\left| \left\{ X' \in X(\mathsf{A}_i^r, \mathsf{B}_i^r) : r_i^v \in \mathrm{Ch}_\mathsf{V}(X'; X) \right\} \right|$ for $i \in S_j$ and with the event $\mathcal{E}_t$. Observe that these random variables are upper-bounded by 1 by (29). Then Lemma 6 states that if $\mathcal{E}_t$ occurs, then with probability at least $1 - 2\exp\left( -\frac{2\delta^4 n^2}{n} \right) \geq 1 - n^{-10}$, we have

$$
\sum_{i \in S_j} \left| \left\{ X' \in X(\mathsf{A}_i^r, \mathsf{B}_i^r) : r_i^v \in \mathrm{Ch}_\mathsf{V}(X'; X) \right\} \right|
$$

$$
= |S_j| \left( \sum_{I \in \mathrm{Free}_\mathsf{V}(X)} \frac{1 - \frac{m}{\ell} |\{J \in \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|} \right) \cdot \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X) - 1}}{m \tilde{n}^{\mathrm{free}(X) - 1}} \pm \frac{10\delta n \delta_{j\delta n}}{\gamma |\mathcal{J}|} \pm \delta^2 n \tag{30}
$$

$$
= |S_j| \left( \sum_{I \in \mathrm{Free}_\mathsf{V}(X)} \frac{1 - \frac{m}{\ell} |\{J \in \mathcal{J} : I \subseteq J\}|}{|\mathcal{J}|} \right) \cdot \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X) - 1}}{m \tilde{n}^{\mathrm{free}(X) - 1}} \pm \frac{20\delta n \delta_{j\delta n}}{\gamma |\mathcal{J}|} \; .
$$

We now turn to estimating the effects of the $r_i^e$. Let us fix an arbitrary history $\mathscr{H}_{i-1} \subseteq \mathcal{E}_t$ which leads to a given set $\mathsf{B}_i^r$. By (QUASI1), for each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ we have $|\mathsf{B}_i^r \cap I_\mathsf{E}| = m \frac{\tilde{n} - j\delta n}{\tilde{n}} \pm 2\delta_i m$. Given any set $L \subseteq \mathsf{B}_i^r$ of edge labels with $\max(L) - \min(L) \leq m$, the set $L$ is contained in two consecutive intervals of $\mathcal{I}_\mathsf{E}$. Let these be $I_\mathsf{E}^{(1)}$ and $I_\mathsf{E}^{(2)}$, and let $L_1$ and $L_2$ be the corresponding subsets of $L$. Suppose that $c \in I_\mathsf{E}^{(1)} \cup I_\mathsf{E}^{(2)}$. Then Fact 21(a) tells us that

$\mathfrak{Q}(J, \min(I_{\mathsf{E}}^{(1)})) = \mathfrak{Q}(J, c) = \mathfrak{Q}(J, \min(I_{\mathsf{E}}^{(2)})) = 0$ for all but at most $2\frac{\ell}{m}$ choices of $J \in \mathcal{J}$. By Fact 21(b) the three quantities never differ by more than $\frac{2m}{\ell^2}$. Thus we have

$$\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_{\mathsf{E}}^{(g)})) = \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, c) \pm 8\frac{m}{\ell}$$

for each $g = 1, 2$. Using this, we have for an arbitrary history $\mathscr{H}_{i-1} \subseteq \mathcal{E}_t$ which leads to the given set $\mathsf{B}_i^r$, that

$$\mathbf{P}[r_i^e \in L | \mathscr{H}_{i-1}] = \frac{1 - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_{\mathsf{E}}^{(1)}))}{|\mathcal{J}|} \cdot \frac{|L_1|}{m\frac{\tilde{n}-j\delta n}{\tilde{n}} \pm 2\delta_i m}$$

$$+ \frac{1 - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \min(I_{\mathsf{E}}^{(2)}))}{|\mathcal{J}|} \cdot \frac{|L_2|}{m\frac{\tilde{n}-j\delta n}{\tilde{n}} \pm 2\delta_i m}$$

$$= \frac{1 \pm 8\frac{m}{\ell} - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, c)}{|\mathcal{J}|} \cdot \frac{|L|}{m\frac{\tilde{n}-j\delta n}{\tilde{n}} \pm 2\delta_i m}$$

$$= \frac{\left(1 - \sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, c)\right)|L|}{|\mathcal{J}|m\frac{\tilde{n}-j\delta n}{\tilde{n}}}\left(1 \pm 10\gamma^{-1}\delta_i\right) \pm \frac{16|L|}{\ell|\mathcal{J}|\gamma},$$

where $c \in I_{\mathsf{E}}^{(1)} \cup I_{\mathsf{E}}^{(2)}$ is arbitrary. Recall that $\mathsf{A}_{i+1} = \mathsf{A}_i^r \setminus \{r_i^v\}$. Using the above calculation and (27), we obtain that when $i \in S_j$,

$$\mathbf{E}\left[\left|\left\{X' \in X(\mathsf{A}_{i+1}, \mathsf{B}_i^r) : r_i^e \in \mathrm{Ch}_{\mathsf{E}}(X'; X)\right\}\right|\middle|\mathscr{H}_i, r_i^v\right]$$

$$= \left(\frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-j\delta n)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \pm 2\delta_i m\right) \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \left(\frac{(1-\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \mathrm{Diff}(\mathfrak{e}; X)))}{|\mathcal{J}|m\frac{\tilde{n}-j\delta n}{\tilde{n}}}\left(1 \pm 10\gamma^{-1}\delta_i\right) \pm \frac{16}{\ell|\mathcal{J}|\gamma}\right)$$

$$= \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-j\delta n)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)}} \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \frac{(1-\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \mathrm{Diff}(\mathfrak{e}; X))}{|\mathcal{J}|m\frac{\tilde{n}-j\delta n}{\tilde{n}}} \pm \frac{40m}{\ell|\mathcal{J}|\gamma} \pm \frac{50\delta_i}{\gamma|\mathcal{J}|}$$

$$= \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-j\delta n)^{\mathrm{free}(X)-1}}{\tilde{n}^{\mathrm{free}(X)-1}} \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \frac{(1-\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \mathrm{Diff}(\mathfrak{e}; X))}{|\mathcal{J}|m} \pm \frac{60\delta_i}{\gamma|\mathcal{J}|}.$$

Now any given edge label is in at most four $X' \in X(\mathsf{A}_{i+1}, \mathsf{B}_i^r)$, and $\mathscr{H}_i, r_i^v$ is a history, so by Lemma 6, if $\mathcal{E}_t$ occurs then with probability at least $1 - 2\exp\left(-\frac{2\delta^4 n^2}{16n}\right) \geq 1 - n^{-10}$ we have

$$\sum_{i \in S_j} \left|\left\{X' \in X(\mathsf{A}_i^r, \mathsf{B}_i^r) : r_i^e \in \mathrm{Ch}_{\mathsf{E}}(X'; X)\right\}\right|$$

$$= |S_j|\frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-j\delta n)^{\mathrm{free}(X)-1}}{\tilde{n}^{\mathrm{free}(X)-1}}\left(\sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \frac{(1-\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \mathrm{Diff}(\mathfrak{e}; X)))}{|\mathcal{J}|m}\right) \pm \frac{60\delta n\delta_{j\delta n}}{\gamma|\mathcal{J}|} \pm \delta^2 n \qquad (31)$$

$$= |S_j|\frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-j\delta n)^{\mathrm{free}(X)-1}}{\tilde{n}^{\mathrm{free}(X)-1}}\left(\sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \frac{(1-\sum_{J \in \mathcal{J}} m\mathfrak{Q}(J, \mathrm{Diff}(\mathfrak{e}; X)))}{|\mathcal{J}|m}\right) \pm \frac{70\delta n\delta_{j\delta n}}{\gamma|\mathcal{J}|}.$$

Putting together (26), (30) and (31), with probability at least $1 - n^{-9}$ we have that if $\mathcal{E}_t$ occurs then

$$\sum_{i \in S_j} (p_{X,i} + r_{X,i}) = \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - j\delta n)^{\mathrm{free}(X)-1}}{\tilde{n}^{\mathrm{free}(X)-1}} |S_j| \cdot \frac{\mathrm{free}(X)}{m|\mathcal{J}|} \pm \frac{200\delta n \delta_{j\delta n}}{\gamma|\mathcal{J}|} . \tag{32}$$

By the union bound over all $X \in \mathcal{X}$, all $1 \le t \le n$ and all $j$, we see that with probability at least $1 - n^{-2}$, if $\mathcal{E}_t$ occurs then the equation (32) holds for all $X \in \mathcal{X}$, all times $t$ and all sets $S_j$. Now one part of $\mathcal{E}_t$ is the assumption that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \le i < t$, and the other part is the good event of Claim 1. The latter event occurs with probability at least $1 - n^{-1}$ by Claim 1, so that with probability at least $1 - 2n^{-1}$ the following holds. Whenever $t$ is such that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for $1 \le i < t$, we have (32) for each $X \in \mathcal{X}$ and $j$.

Suppose now that, for some $t$, we have (32) for each $X \in \mathcal{X}$ and $j$. To complete the proof of the claim, we need to show that putting together these partial sums on short intervals, we obtain the desired (23). Here we are implicitly verifying that we have a solution to a certain first-order differential equation (which we do not write down as we do not need to know it), and consequently an integral naturally appears.

We have

$$\int_{x=0}^{t} (\tilde{n} - x - \delta n)^{\mathrm{free}(X)-1} dx \le \sum_{j=1}^{\lceil \frac{t}{\delta n} \rceil} (\tilde{n} - j\delta n)^{\mathrm{free}(X)-1} |S_j| \le \int_{x=0}^{t} (\tilde{n} - x)^{\mathrm{free}(X)-1} dx .$$

Plugging this into (32), we get

$$\sum_{i=1}^{t} (p_{X,i} + r_{X,i}) = \int_{x=0}^{t} \frac{|X(\mathbb{A}, \mathbb{B})|(\tilde{n} - x)^{\mathrm{free}(X)-1} \mathrm{free}(X)}{\tilde{n}^{\mathrm{free}(X)-1} m|\mathcal{J}|} dx \pm \frac{\mathrm{free}(X)\delta n}{|\mathcal{J}|} \pm \sum_{j=1}^{\lceil \frac{t}{\delta n} \rceil} \frac{200\delta n \delta_{j\delta n}}{\gamma|\mathcal{J}|}$$

$$= |X(\mathbb{A}, \mathbb{B})| \frac{\tilde{n}^{\mathrm{free}(X)} - (\tilde{n} - t)^{\mathrm{free}(X)}}{\tilde{n}^{\mathrm{free}(X)-1} m|\mathcal{J}|} \pm \tfrac{1}{4}\delta_t m ,$$

as desired, where the final line follows from the choice of $\delta$, from (4), and since $\frac{n}{|\mathcal{J}|} < 2m$. $\quad\square$

We would now like to argue that $\sum_{i=1}^{t} p_{I_\mathsf{E},i}$ is a good estimate for the number of edge labels in $I_\mathsf{E}$ used in the labelling process up to time $t$. However, we are not able to do this in one step. Instead, we define a *fine estimate* $q_{I_\mathsf{E},i}$, which plays the same rôle as $p_{I_\mathsf{E},i}$ except that we condition on the behaviour of Algorithm 1 up to and including the time $h$ at which we label $v_h = v_i^+$. We will see that at least $\sum_{i=1}^{t} p_{I_\mathsf{E},i}$ is a good estimate for $\sum_{i=1}^{t} q_{I_\mathsf{E},i}$. We define similarly a fine estimate $q_{X,i}$ corresponding to $p_{X,i}$. We write down formulae valid for $i \ge 2$, when $v_h = v_i^+$ exists.

$q_{I_\mathsf{E},i}$

$q_{X,i}$

$$q_{I_\mathsf{E},i} := \frac{\left|\left\{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h) : |a - \psi_h(v_h)| \in I_\mathsf{E}\right\}\right|\tilde{n}^2}{|\mathsf{A}_h|^2\ell}$$

$$q_{X,i} := \sum_{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h)} \frac{\left|\left\{X' \in X(\mathsf{A}_h, \mathsf{B}_h) : a \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } |a - \psi_h(v_h)| \in \mathrm{Ch}_\mathsf{E}(X'; X)\right\}\right|}{|\mathsf{A}_h|^2\ell\tilde{n}^{-2}} \quad (33)$$

Additionally, we define $q_{I_\mathsf{E},1} = q_{X,1} = 0$.

Observe that $\frac{|\mathsf{A}_h||\mathsf{B}_h|\ell}{\tilde{n}|\mathbb{B}|}$ is, by (QUASI2) with the structure $X_3 \llbracket \psi_h(v_h), J(v_i) \rrbracket$, a good estimate for the size of the set of vertex labels from which we will label $v_i$, ignoring any changes that may occur between time $h$ and the time $i$ when we label $v_i$. Thus $q_{I_\mathsf{E},i}$ is a good estimate for the expectation of labelling $v_i$ in such a way as to use an edge label in $I_\mathsf{E}$, if we ignore any changes that might occur between times $h$ and $i$. We will see that it is usually reasonable to ignore such changes. Similarly, $q_{X,i}$ is a good estimate for the expected number of structures following the pattern $X$ whose chosen labels contain either the vertex or edge label used at time $i$, ignoring any changes between times $h$ and $i$, and estimating the size of the set of vertex labels from which we label $v_i$ by $\frac{|\mathsf{A}_h||\mathsf{B}_h|\ell}{\tilde{n}|\mathbb{B}|}$.

We now show that the partial sums of the crude estimates are, with high probability, good estimates for the partial sums of the fine estimates. There are two parts to this. First, we will argue that if $v_i v_i^+ \notin \mathcal{R}$, then we have $\mathbf{E}[q_{I_\mathsf{E},i}|\psi_{h-1}] \approx p_{I_\mathsf{E},i}$, and similarly for the $q_{X,i}$. In other words, $p_{I_\mathsf{E},i}$ is a good estimate for the expectation of $q_{I_\mathsf{E},i}$ conditioned on the labelling history up to the time immediately before labelling $v_i^+$, whatever that history might be (as long as it maintains quasirandomness). This is the main combinatorial work in our proof. Second, we observe that the effect of the remaining terms where $v_i v_i^+ \in \mathcal{R}$ is small, simply because $\mathcal{R}$ is small, and apply Lemma 6 to argue that the sum of conditional expectations is with high probability close to the partial sum $\sum_{i=1}^t q_{I_\mathsf{E},i}$, and similarly for the $q_{X,i}$.

**Claim 4.** *With probability at least $1 - 4n^{-1}$ the following holds. For each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, each $X \in \mathcal{X}$, each $1 \le t \le n$, and each $1 \le k \le \left\lceil \frac{t}{\delta n} \right\rceil$, if $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \le i < t$, we have*

$$\sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} q_{I_\mathsf{E},i} = \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \pm \frac{2000\delta_{k\delta n}\delta m}{\gamma^4} \ , \ and \quad (34)$$

$$\sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} q_{X,i} = \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{X,i} \pm \frac{10^6 \delta m \delta_{k\delta n}}{\gamma^7} \ . \quad (35)$$

*Proof of Claim 4.* For this proof, for each $0 \le i \le n$, let $\mathscr{H}_i$ denote the history up to, but not including, the labelling of $v_{i+1}$. Thus $\mathscr{H}_0$ is the empty history.

We begin with (34). Given $I_\mathsf{E}$ and $1 \le t \le n$, and $1 \le k \le \left\lceil \frac{t}{\delta n} \right\rceil$, we define random variables for each $1 \le h < t$ by

$$Y_h = \sum_{i:h<i\le t} q_{I_\mathsf{E},i} \cdot \mathbb{1}_{v_i v_h \in E(T)} \cdot \mathbb{1}_{(k-1)\delta n+1 \le i \le \max(t,k\delta n)} \cdot$$

We have

$$\sum_{h=1}^{t-1} Y_h = \sum_{h=1}^{t-1} \sum_{i:h<i\leq t} q_{I_{\mathsf{E}},i} \cdot \mathbb{1}_{v_i v_h \in E(T)} \cdot \mathbb{1}_{(k-1)\delta n+1 \leq i \leq \max(t,k\delta n)} \cdot \tag{36}$$

Let us now fix any $i$ and look at the coefficient of $q_{I_{\mathsf{E}},i}$ on the right-hand side of (36). Firstly, the coefficient is never more than 1 since there is at most one $h$ which makes the indicator $\mathbb{1}_{v_i v_h \in E(T)}$ non-zero, namely that corresponding to the parent of $i$ (c.f. (PRE2)). Secondly, the coefficient is zero outside the range $[(k-1)\delta n + 1, \max(t, k\delta n)]$ due to the indicator $\mathbb{1}_{(k-1)\delta n+1 \leq i \leq \max(t,k\delta n)}$. On the other hand, if these two conditions are fulfilled, then the coefficient of $q_{I_{\mathsf{E}},i}$ is 1. Indeed, in that case we can always find a parent $v_h$ with $h < i$. (Note that in our setting, $i$ is always more than 1, and so $v_i$ always has a parent.) We conclude that

$$\sum_{h=1}^{t-1} Y_h = \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} q_{I_{\mathsf{E}},i} \ . \tag{37}$$

We will apply Lemma 6 to estimate the sum on the left-hand side. As in the proof of Claim 3, we let $\mathcal{E}_t$ be the event that $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \leq i < t$, and the good event of Claim 1 holds. In particular, for each $h$ we have $|\mathsf{A}_h|, |\mathsf{B}_h| \geq \frac{\gamma}{2}\tilde{n}$.

We need to show that, assuming $\mathcal{E}_t$, we can give good bounds on $\sum_{h=1}^{t-1} \mathbf{E}[Y_h | \mathscr{H}_{h-1}]$. In turn, to obtain such bounds it is enough to show that $\mathbf{E}[q_{I_{\mathsf{E}},i} | \mathscr{H}_{h-1}] \approx p_{I_{\mathsf{E}},i}$ for each $1 < h < i \leq t$ with $v_i v_h \in E(T) \setminus \mathcal{R}$. The terms with $h = 1$ or $v_i v_h \in \mathcal{R}$ contribute at most $\frac{\eta n}{\log n} + \varepsilon n$ to the sum by assumption on $\Delta(T)$ and by (PRE1), which is small enough to ignore.

Suppose we have $1 < h < i \leq t$, with $v_h = v_i^+$ and $v_i v_h \in E(T) \setminus \mathcal{R}$. We say that $(a, a')$ is an *admissible pair* if $a \in J(v_h) \cap \mathsf{A}_h$ and $a' \in \overline{J(v_i)} \cap \mathsf{A}_h$, and $|\psi_{h-1}(v_h^+) - a|, |a' - a| \in \mathsf{B}_h$ <span style="color:red">admissible pair</span> are distinct. Note that since $J(v_h) = \overline{J(v_i)}$ by (PRE5), $a$ and $a'$ are distinct. It follows that $\psi_{h-1} \cup \{v_h \mapsto a, v_i \mapsto a'\}$ is a graceful labelling of $T[v_1, \ldots, v_h, v_i]$.

Note that

$$\mathbf{E}[q_{I_{\mathsf{E}},i} | \mathscr{H}_{h-1}] = \mathbf{E}_{a \sim \mathsf{UNIFORM}(\mathrm{Adm}(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h))} \left[ \mathbf{E}\left[ q_{I_{\mathsf{E}},i} | \mathscr{H}_{h-1}, \psi_h(v_h) = a \right] \right],$$

because we choose $\psi_h(v_h)$ uniformly. Thus, by definition of $q_{I_{\mathsf{E}},i}$, we have

$$\frac{|\mathsf{A}_h|^2 \ell}{\tilde{n}^2} \mathbf{E}[q_{I_{\mathsf{E}},i} | \mathscr{H}_{h-1}] = \frac{\sum_{(a,a') \text{ admissible}} \mathbb{1}_{|a-a'| \in I_{\mathsf{E}} \cap \mathsf{B}_h}}{\left| \mathrm{Adm}(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h) \right|}$$

$$\overset{(16)}{=} \frac{\sum_{(a,a') \text{ admissible}} \mathbb{1}_{|a-a'| \in I_{\mathsf{E}} \cap \mathsf{B}_h}}{\left| X_3 \left[\!\left[ \psi_{h-1}(v_h^+), J(v_h) \right]\!\right] (\mathsf{A}_h, \mathsf{B}_h) \right|}$$

$$= \frac{\sum_{c \in I_{\mathsf{E}} \cap \mathsf{B}_h} \left| X_2 \left[\!\left[ \psi_{h-1}(v_h^+), \overline{J(v_i)}, c, J(v_i) \right]\!\right] (\mathsf{A}_h, \mathsf{B}_h) \right|}{\left| X_3 \left[\!\left[ \psi_{h-1}(v_h^+), J(v_h) \right]\!\right] (\mathsf{A}_h, \mathsf{B}_h) \right|} \ .$$

Therefore,

$$\mathbf{E}[q_{I_{\mathsf{E}},i} | \mathscr{H}_{h-1}] = \frac{\tilde{n}^2 \sum_{c \in I_{\mathsf{E}} \cap \mathsf{B}_h} |X_2 \left[\!\left[ \psi_{h-1}(v_h^+), \overline{J(v_i)}, c, J(v_i) \right]\!\right] (\mathsf{A}_h, \mathsf{B}_h)|}{|\mathsf{A}_h|^2 \ell |X_3 \left[\!\left[ \psi_{h-1}(v_h^+), J(v_h) \right]\!\right] (\mathsf{A}_h, \mathsf{B}_h)|} \ .$$

Because we assume $\mathcal{E}_t$ and $h < t$, so $(\mathsf{A}_h, \mathsf{B}_h)$ is $\delta_h$-quasirandom and we can use Lemma 20 to estimate both the $X_2$-term in the numerator and the $X_3$-term in the denumerator. We obtain

$$\mathbf{E}[q_{I_\mathsf{E},i}|\mathscr{H}_{h-1}] \overset{(8),(9)}{=} \frac{\tilde{n}^2 \sum_{c \in I_\mathsf{E} \cap \mathsf{B}_h} \left( \ell^2 \mathfrak{Q}(J(v_i), c)|\mathsf{A}_h|^3 \tilde{n}^{-3} \pm 3\delta_h \ell \right)}{|\mathsf{A}_h|^2 \ell \left( \ell |\mathsf{A}_h|^2 \tilde{n}^{-2} \pm 2\delta_h \ell \right)}$$

$$= \frac{\tilde{n} \sum_{c \in I_\mathsf{E} \cap \mathsf{B}_h} \mathfrak{Q}(J(v_i), c)}{|\mathsf{A}_h|} \pm \frac{50\delta_h m}{\ell \gamma^4}$$

$$\boxed{\text{by Fact 21(b)}} \quad = \frac{\tilde{n}|I_\mathsf{E} \cap \mathsf{B}_h| \mathfrak{Q}(J(v_i), \min(I_\mathsf{E}))}{|\mathsf{A}_h|} \pm \frac{60\delta_h m}{\ell \gamma^4} \, .$$

By (QUASI1) we have $|I_\mathsf{E} \cap \mathsf{B}_h| = m\frac{|\mathsf{A}_h|}{\tilde{n}} \pm \delta_h m$, so that

$$\mathbf{E}[q_{I_\mathsf{E},i}|\mathscr{H}_{h-1}] = m\mathfrak{Q}(J(v_i), \min(I_\mathsf{E})) \pm \tfrac{70\delta_h m}{\gamma^4 \ell} = p_{I_\mathsf{E},i} \pm \tfrac{70\delta_i m}{\gamma^4 \ell} \, , \tag{38}$$

where we use $\delta_h < \delta_i$ since $h < i$.

For most values of $i$ we actually have a stronger estimate than (38). If $\mathfrak{Q}(J(v_i), c) = 0$ for all $c \in I_\mathsf{E}$, then $p_{I_\mathsf{E},i} = q_{I_\mathsf{E},i} = 0$ by (20) and (33). That is, in this situation we have

$$\mathbf{E}[q_{I_\mathsf{E},i}|\mathscr{H}_{h-1}] = p_{I_\mathsf{E},i} \, . \tag{39}$$

For any given $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, by Fact 21(a) there are at most $2\frac{\ell}{m}$ sets $J \in \mathcal{J}$ such that $\mathfrak{Q}(J, c) \neq 0$ for some $c \in I_\mathsf{E}$, so by (PRE4), and choice of $\delta$, for each $1 \leq k \leq \delta^{-1}$, the number of $i \in \{(k-1)\delta n + 1, \ldots, k\delta n\}$ such that $\mathfrak{Q}(J(v_i), c) \neq 0$ for some $c \in I_\mathsf{E}$ is at most $2\frac{\ell}{m} \cdot \frac{2\delta n}{|\mathcal{J}|}$.

$\mathcal{D}_h$     We want to sum up (38) and (39). To this end, for $h = 2, \ldots, t-1$, set

$$\mathcal{D}_h = \{i \in \mathbb{N} \, : \, (k-1)\delta n + 1 \leq i \leq \max(t, k\delta n), v_i^+ = v_h, v_i v_h \notin \mathcal{R}\} \, . \tag{40}$$

Observe that $\dot{\bigcup}_{h=2}^{t-1} \mathcal{D}_h = \{i \in \mathbb{N} \, : \, (k-1)\delta n + 1 \leq i \leq \max(t, k\delta n), v_i^+ \neq v_1, v_i v_i^+ \notin \mathcal{R}\}$. Thus, we have

$$\sum_{h=2}^{t-1} \sum_{i \in \mathcal{D}_h} \mathbf{E}[q_{I_\mathsf{E},i}|\mathscr{H}_{h-1}] \overset{(38),(39)}{=} \left( \sum_{\substack{i=(k-1)\delta n + 1 \\ v_i^+ \neq v_1, v_i v_i^+ \notin \mathcal{R}}}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \right) \pm \frac{4\ell\delta n}{m|\mathcal{J}|} \cdot \frac{70\delta_{k\delta n} m}{\ell \gamma^4} \, , \tag{41}$$

since $i \leq k\delta n$ and thus $\delta_i \leq \delta_{k\delta n}$. Taking into account the at most $\varepsilon n$ values of $i$ with $v_i v_i^+ \in \mathcal{R}$ and at most $\frac{\eta n}{\log n}$ values of $i$ with $v_i^+ = v_1$, each of which terms contributes an error

of at most 1, and using (3), we have

$$\sum_{h=1}^{t-1} \mathbf{E}[Y_h | \mathscr{H}_{h-1}] = \sum_{h=1}^{t-1} \sum_{i=(k-1)\delta n+1, \, v_i^+=v_h}^{\max(t,k\delta n)} \mathbf{E}[q_{I_\mathsf{E},i} | \mathscr{H}_{h-1}]$$

$$= \Big( \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \Big) \pm \frac{600\delta_{k\delta n}\delta m}{\gamma^4} \pm \varepsilon n \pm \frac{\eta n}{\log n}$$

$$= \Big( \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \Big) \pm \frac{1000\delta_{k\delta n}\delta m}{\gamma^4} .$$

We are now in a position to apply Lemma 6, with the random variables $(Y_h)_{h=1}^{t-1}$ satisfying $0 \leq Y_h \leq \deg_T(v_h)$ for each $h$, and with the event $\mathcal{E}_t$. By (18) we have $\sum_{h=1}^{t-1} \deg_T(v_h)^2 \leq \frac{2\eta n^2}{\log n}$, so applying Lemma 6 we conclude that with probability at least $1 - \exp\big(-\frac{2\delta^2 n^2 \log n}{2\eta n^2}\big) > 1 - n^{-10}$, if $\mathcal{E}_t$ holds, we have

$$\sum_{h=1}^{t-1} Y_h \overset{(37)}{=} \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} q_{I_\mathsf{E},i} = \Big( \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \Big) \pm \frac{1000\delta_{k\delta n}\delta m}{\gamma^4} \pm \delta n$$

$$= \Big( \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{I_\mathsf{E},i} \Big) \pm \frac{2000\delta_{k\delta n}\delta m}{\gamma^4} .$$

Taking the union bound over $1 \leq t \leq n$ and $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ and $k$, we conclude that with probability at least $1 - n^{-2}$, if $\mathcal{E}_t$ holds then (34) holds for each $1 \leq t \leq n$, each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$, and each $k$. Recall that the good event of Claim 1 holds with probability at least $1 - n^{-1}$. Thus the following event holds with probability at least $1 - 2n^{-1}$. For each $1 \leq t \leq n$, if $(\mathsf{A}_i, \mathsf{B}_i)$ is $\delta_i$-quasirandom for each $1 \leq i < t$, then (34) holds for each $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$ and each $k$.

The proof that with high probability (35) holds follows the same idea, although the combinatorial manipulations are a little more involved. Now, given $X \in \mathcal{X}$ and $1 \leq t \leq n$, we define                                                                                        $Y_h$

$$Y_h = \sum_{h<i\leq t} q_{X,i} \cdot \mathbb{1}_{v_i v_h \in E(T)} \cdot \mathbb{1}_{(k-1)\delta n+1 \leq i \leq \max(t,k\delta n)} \cdot$$

and again the critical point is to show that, assuming $\mathcal{E}_t$, for each $1 < h < i \leq t$ such that $v_i v_h \in E(T) \setminus \mathcal{R}$, we have $\mathbf{E}[q_{X,i} | \mathscr{H}_{h-1}] \approx p_{X,i}$.

As before, given a choice of $1 < h < i \leq t$ with $v_h = v_i^+$ such that $v_i v_h \notin \mathcal{R}$, we say that $(a, a')$ is an *admissible pair* if $a \in J(v_h) \cap \mathsf{A}_h$ and $a' \in J(v_i) \cap \mathsf{A}_h$, and $\big|\psi_{h-1}(v_h^+) - a\big|, |a' - a| \in \mathsf{B}_h$ are distinct.                    admissible pair

By (33), and since $\psi_h(v_h)$ is chosen uniformly from $\mathrm{Adm}\big(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h\big)$, we have

$$
\frac{|\mathsf{A}_h|^2 \ell \mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] \cdot \big|\mathrm{Adm}\big(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h\big)\big|}{\tilde{n}^2}
$$

$$
= \sum_{(a,a') \text{ admissible}} \sum_{X' \in X(\mathsf{A}_h, \mathsf{B}_h)} \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X) \text{ or } |a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)} \tag{42}
$$

$$
= \sum_{X' \in X(\mathsf{A}_h, \mathsf{B}_h)} \sum_{(a,a') \text{ admissible}} \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X) \text{ or } |a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)}
$$

$$
= \sum_{X' \in X(\mathsf{A}_h, \mathsf{B}_h)} \sum_{(a,a') \text{ admissible}} \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X)} + \mathbb{1}_{|a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)} - \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X)} \mathbb{1}_{|a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)} \,,
$$

where the last equality holds since vertex labels of $X'$ are by definition pairwise distinct, and edge labels of $X'$ are by definition pairwise distinct.

For a given $X' \in X(\mathsf{A}_h, \mathsf{B}_h)$, we have

$$
\sum_{(a,a') \text{ admissible}} \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X)} = \sum_{a' \in \mathrm{Ch}_\mathsf{V}(X';X) \cap J(v_i)} \Big|X_4 \big[\!\big[\psi_{h-1}(v_h^+), a', J(v_h)\big]\!\big] (\mathsf{A}_h, \mathsf{B}_h)\Big|
$$

$$
\overset{(10)}{=} \sum_{a' \in \mathrm{Ch}_\mathsf{V}(X';X) \cap J(v_i)} \big(\tfrac{|\mathsf{A}_h|^3 \ell}{\tilde{n}^3} \pm 2\delta_h \ell\big) \tag{43}
$$

$$
= \sum_{I \in \mathrm{Free}_\mathsf{V} X} \mathbb{1}_{I \subseteq J(v_i)} \big(\tfrac{|\mathsf{A}_h|^3 \ell}{\tilde{n}^3} \pm 2\delta_h \ell\big) \,,
$$

where we can apply Lemma 20 since $(\mathsf{A}_h, \mathsf{B}_h)$ is by assumption $\delta_h$-quasirandom. Similarly, we have

$$
\sum_{(a,a') \text{ admissible}} \mathbb{1}_{|a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)} = \sum_{c \in \mathrm{Ch}_\mathsf{E}(X';X)} \Big|X_2 \big[\!\big[\psi_{h-1}(v_h^+), J(v_h), c, J(v_i)\big]\!\big] (\mathsf{A}_h, \mathsf{B}_h)\Big|
$$

$$
\overset{(8)}{=} \sum_{c \in \mathrm{Ch}_\mathsf{E}(X';X)} \big(\tfrac{|\mathsf{A}_h|^3 \ell^2 \mathfrak{Q}(J(v_h),c)}{\tilde{n}^3} \pm 3\delta_h \ell\big) \tag{44}
$$

$$
\boxed{\text{by Fact 21(b)}} \;\; = \sum_{\mathfrak{e} \in \mathrm{Free}_\mathsf{E}(X)} \big(\tfrac{|\mathsf{A}_h|^3 \ell^2 \mathfrak{Q}(J(v_h),\mathrm{Diff}(\mathfrak{e};X))}{\tilde{n}^3} \pm 4\delta_h \ell\big) \,,
$$

where the final line follows since $\mathrm{Diff}(\mathfrak{e}; X)$ is within $m$ of the edge label chosen for $\mathfrak{e}$ in any $X'$ following the pattern $X$. Finally, since $\mathrm{Ch}_\mathsf{E}(X'; X)$ is a set of size at most 2, for each $a' \in \mathrm{Ch}_\mathsf{V}(X'; X)$ there are at most 4 choices of $a$ such that $|a - a'| \in \mathrm{Ch}_\mathsf{E}(X'; X)$. Since $\mathrm{Ch}_\mathsf{V}(X'; X)$ is a set of size at most 2, in total there are at most 8 pairs $(a, a')$ such that $a' \in \mathrm{Ch}_\mathsf{V}(X'; X)$ and $|a - a'| \in \mathrm{Ch}_\mathsf{E}(X'; X)$. We therefore have

$$
\sum_{(a,a') \text{ admissible}} \mathbb{1}_{a' \in \mathrm{Ch}_\mathsf{V}(X';X)} \mathbb{1}_{|a-a'| \in \mathrm{Ch}_\mathsf{E}(X';X)} \leq 8 \,. \tag{45}
$$

Observe that the final expression in each of (43), (44) and (45) is independent of $X'$. Thus, substituting into (42), we have

$$
\begin{aligned}
&\frac{|\mathsf{A}_h|^2 \ell \mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] \cdot \big|\mathrm{Adm}\big(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h\big)\big|}{\tilde{n}^2} \\
={}& \big|X(\mathsf{A}_h, \mathsf{B}_h)\big| \sum_{I \in \mathrm{Free}_{\mathsf{V}} X} \mathbb{1}_{I \subseteq J(v_i)}\Big(\frac{|\mathsf{A}_h|^3 \ell}{\tilde{n}^3} \pm 2\delta_h \ell\Big) \\
&+ \big|X(\mathsf{A}_h, \mathsf{B}_h)\big| \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \Big(\frac{|\mathsf{A}_h|^3 \ell^2 \mathfrak{Q}(J(v_h), \mathrm{Diff}(\mathfrak{e};X))}{\tilde{n}^3} \pm 4\delta_h \ell\Big) \pm 8m \\
\overset{\mathrm{F17(b)}}{=}{}& \frac{|\mathsf{A}_h|^3 \ell^2 \big|X(\mathsf{A}_h, \mathsf{B}_h)\big|}{\tilde{n}^3}\Big(\sum_{I \in \mathrm{Free}_{\mathsf{V}}(X)} \frac{\mathbb{1}_{I \subseteq J(v_i)}}{\ell} + \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \mathfrak{Q}(J(v_i), \mathrm{Diff}(\mathfrak{e};X))\Big) \pm 20\delta_h \ell m \\
\overset{\mathrm{(QUASI2)}}{=}{}& \frac{|\mathsf{A}_h|^{3+\mathrm{free}(X)} \ell^2 \big|X(\mathbb{A}, \mathbb{B})\big|}{\tilde{n}^{3+\mathrm{free}(X)}}\Big(\sum_{I \in \mathrm{Free}_{\mathsf{V}}(X)} \frac{\mathbb{1}_{I \subseteq J(v_i)}}{\ell} + \sum_{\mathfrak{e} \in \mathrm{Free}_{\mathsf{E}}(X)} \mathfrak{Q}(J(v_i), \mathrm{Diff}(\mathfrak{e};X))\Big) \pm 40\delta_h \ell m \\
\overset{(21)}{=}{}& \frac{|\mathsf{A}_h|^{3+\mathrm{free}(X)} \ell^2 \big|X(\mathbb{A}, \mathbb{B})\big|}{\tilde{n}^{3+\mathrm{free}(X)}}\Big(\frac{\tilde{n}^{\mathrm{free}(X)-1} p_{X,i}}{|X(\mathbb{A}, \mathbb{B})|(\tilde{n}-i)^{\mathrm{free}(X)-1}}\Big) \pm 40\delta_h \ell m \\
={}& \frac{|\mathsf{A}_h|^{3+\mathrm{free}(X)} \ell^2 p_{X,i}}{\tilde{n}^4 (\tilde{n}-h)^{\mathrm{free}(X)-1}} \pm 50\delta_h \ell m\,.
\end{aligned}
$$

$$(46)$$

where for the second equality we use the fact that $\mathfrak{Q}(J(v_i), \mathrm{Diff}(\mathfrak{e};X)) = \mathfrak{Q}(J(v_h), \mathrm{Diff}(\mathfrak{e};X))$ since $J(v_i) = \overline{J(v_h)}$ by (PRE5), for the third we use the assumption that $(\mathsf{A}_h, \mathsf{B}_h)$ is $\delta_h$-quasirandom, and for the last line we use the fact $i = h \pm \frac{\varepsilon n}{\log n}$, which holds by (PRE3) since $v_i v_h \notin \mathcal{R}$.

Now, since $(\mathsf{A}_h, \mathsf{B}_h)$ is $\delta_h$-quasirandom, by (9) and since $|\mathsf{A}_h| \geq \frac{\gamma}{2}\tilde{n}$ by (19), we have

$$
\big|\mathrm{Adm}\big(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h\big)\big| = (|\mathsf{A}_h|^2/\tilde{n}^2)\ell \pm 2\delta_h \ell = \big(1 \pm \tfrac{8\delta_h}{\gamma^2}\big)|\mathsf{A}_h|^2 \tilde{n}^{-2}\ell\,. \tag{47}
$$

We can rewrite (46) as

$$
\begin{aligned}
\mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] &= \frac{\frac{|\mathsf{A}_h|^{1+\mathrm{free}(X)} \ell p_{X,i}}{\tilde{n}^2(\tilde{n}-h)^{\mathrm{free}(X)-1}} \pm \frac{50\delta_h m \tilde{n}^2}{|\mathsf{A}_h|^2}}{\big|\mathrm{Adm}\big(\psi_{h-1}(v_h^+), J(v_h); \mathsf{A}_h, \mathsf{B}_h\big)\big|} \\
\boxed{\text{substituting (47) and using that } |\mathsf{A}_h| \geq \tfrac{\gamma}{2}\tilde{n}} \quad &= \frac{|\mathsf{A}_h|^{\mathrm{free}(X)-1} p_{X,i}}{(\tilde{n}-h)^{\mathrm{free}(X)-1}}\big(1 \pm \tfrac{16\delta_h}{\gamma^2}\big) \pm \frac{2000\delta_h m}{\gamma^7 \ell}\,.
\end{aligned}
$$

By (19), we have $|\mathsf{A}_h| = \tilde{n} - h \pm 10\ell$. Furthermore, by (21) we have $p_{X,i} \le 4m/\ell$. Finally, we have free$(X) \le 3$. We thus get

$$\mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] = \frac{(\tilde{n} - h \pm 10\ell)^{\text{free}(X)-1}p_{X,i}}{(\tilde{n} - h)^{\text{free}(X)-1}}\left(1 \pm \tfrac{16\delta_h}{\gamma^2}\right) \pm \frac{2000\delta_h m}{\gamma^7 \ell}$$

$$= p_{X,i}\left(1 \pm \tfrac{16\delta_h}{\gamma^2}\right)\left(1 \pm \tfrac{20\ell}{\gamma n}\right)^2 \pm \frac{2000\delta_h m}{\gamma^7 \ell} \tag{48}$$

$$= p_{X,i} \pm \frac{100\delta_h m}{\gamma^2 \ell} \pm \frac{800m}{\gamma n} \pm \frac{2000\delta_h m}{\gamma^7 \ell} = p_{X,i} \pm \tfrac{3000\delta_i m}{\gamma^7 \ell} \, .$$

As before, for most values of $i$ we obtain a stronger estimate. If $\mathfrak{Q}(J(v_i), c) = 0$ for each $c \in \text{Ch}_\mathsf{E}(X'; X)$ and each $X' \in X(\mathbb{A}, \mathbb{B})$, and $I \cap J(v_i) = \emptyset$ for each $I \in \text{Free}_\mathsf{V}(X)$, then $p_{X,i} = q_{X,i} = 0$ by (21) and (33). In this situation, we can write

$$\mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] = p_{X,i} \, . \tag{49}$$

Since $X$ has at most two free edge labels, for each of which an edge label can be chosen in an interval of length $m$, there are at most $4\frac{\ell}{m}$ sets $J \in \mathcal{J}$ such that $\mathfrak{Q}(J, c) \ne 0$ for some $c \in \text{Ch}_\mathsf{E}(X'; X)$ and $X' \in X(\mathbb{A}, \mathbb{B})$. Since $X$ has at most two free vertex labels, there are at most $2\frac{\ell}{m}$ sets $J \in \mathcal{J}$ such that $I \subseteq J$ for some $I \in \text{Free}_\mathsf{V}(X)$. Putting this together, for all but at most $6\frac{\ell}{m}$ sets $J \in \mathcal{J}$, if $J(v_i) = J$ we have (49). For each $k \ge 1$, and each interval $i \in \{(k-1)\delta n + 1, \dots, k\delta n\}$, by (PRE4), for all but at most $\frac{2\delta n}{|\mathcal{J}|} \cdot 6\frac{\ell}{m}$ choices of $i$ we have (49). Thus, using the notation from (40), after summing up (48) and (49) we obtain

$$\sum_{h=2}^{t-1} \sum_{i \in \mathcal{D}_h} \mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}] = \Big( \sum_{\substack{i=(k-1)\delta n+1 \\ v_i^+ \ne v_1, v_i v_i^+ \notin \mathcal{R}}}^{\max(t,k\delta n)} p_{X,i} \Big) \pm \frac{12\ell\delta n}{m|\mathcal{J}|} \cdot \frac{3000\delta_{k\delta n}m}{\ell\gamma^7} \, ,$$

where (as in (41)) since $i \le k\delta n$ we have $\delta_i \le \delta_{k\delta n}$. Since $0 \le q_{X,i}, p_{X,i} \le 6$ for each $i$, taking into account the at most $\varepsilon n$ values of $i$ with $v_i v_i^+ \in \mathcal{R}$ and at most $\frac{\eta n}{\log n}$ values with $v_i^+ = v_1$, we have

$$\sum_{h=1}^{t-1} \mathbf{E}[Y_h|\mathscr{H}_{h-1}] = \sum_{h=1}^{t-1} \sum_{i=(k-1)\delta n+1 \, , \, v_i^+ = v_h}^{\max(t,k\delta n)} \mathbf{E}[q_{X,i}|\mathscr{H}_{h-1}]$$

$$\stackrel{(3)}{=} \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{X,i} \pm \frac{10^5 \delta m \delta_{k\delta n}}{\gamma^7} \pm 6\varepsilon n \pm 6\frac{\eta n}{\log n} \, .$$

Finally, since $0 \leq Y_h \leq 6 \deg_T(v_h)$, by Lemma 6 and (18), with probability at least $1 - n^{-10}$, if $\mathcal{E}_t$ holds then we have

$$\sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} q_{X,i} = \sum_{h=1}^{t-1} Y_h = \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{X,i} \pm \frac{2 \cdot 10^5 \delta m \delta_{k\delta n}}{\gamma^7} \pm \delta^2 n$$

$$= \sum_{i=(k-1)\delta n+1}^{\max(t,k\delta n)} p_{X,i} \pm \frac{10^6 \delta m \delta_{k\delta n}}{\gamma^7} .$$

Taking the union bound over the choices of $t$ and $X$ and $k$, we see that with probability at least $1 - n^{-8}$, for each $t$ such that $\mathcal{E}_t$ holds, we have (35) for all $X \in \mathcal{X}$ and $k$. Since the good event of Claim 1 holds with probability at least $1 - n^{-1}$, we conclude that the statement holds with probability at least $1 - 4n^{-1}$, as desired. □

We are now in a position to complete the proof of Theorem 8 by showing that with high probability $(\mathsf{A}_t, \mathsf{B}_t)$ is $\delta_t$-quasirandom for each $0 \leq t \leq n-1$. Suppose that the good events of Claims 1, 3 and 4 hold; this event, which we denote by $\mathcal{E}$, has probability at least $1 - 7n^{-1}$. Let $\mathscr{H}_0$ be the empty history, and $\mathscr{H}_i$ denote the history up to and including the labelling of $v_i$ for each $1 \leq i \leq n$.

The proof that $(\mathsf{A}_t, \mathsf{B}_t)$ is $\delta_t$-quasirandom goes by induction on $t$, with the base case $t = 0$ being trivial. Observe that $(\mathsf{A}_1, \mathsf{B}_1) = (\mathbb{A}, \mathbb{B})$ is by definition $\delta_1$-quasirandom. Now suppose that $(\mathsf{A}_i, \mathsf{B}_i)$ are $\delta_i$-quasirandom for each $1 \leq i \leq t$.

Now suppose $1 < i \leq t$ is such that $v_i v_i^+ \notin \mathcal{R}$. By (PRE3), $v_h = v_i^+$ satisfies $h \geq i - \frac{\varepsilon n}{\log n}$, and thus $(\mathsf{A}_h, \mathsf{B}_h)$ and $(\mathsf{A}_i, \mathsf{B}_i)$ differ by at most $\frac{2\varepsilon n}{\log n}$ vertex labels and at most $\frac{2\varepsilon n}{\log n}$ edge labels. Thus we have

$$|\mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i)| = |\mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h)| \pm \tfrac{4\varepsilon n}{\log n} , \tag{50}$$

and for any $I_\mathsf{E} \in \mathcal{I}_\mathsf{E}$,

$$\begin{aligned} &\left|\left\{ a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i) : |a - \psi_h(v_h)| \in I_\mathsf{E} \right\}\right| \\ =& \left|\left\{ a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h) : |a - \psi_h(v_h)| \in I_\mathsf{E} \right\}\right| \pm \tfrac{4\varepsilon n}{\log n} . \end{aligned} \tag{51}$$

Therefore,

$$\begin{aligned} \left|\mathrm{Adm}\big(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i\big)\right| &\overset{(50)}{=} \left|\mathrm{Adm}\big(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h\big)\right| \pm \tfrac{4\varepsilon n}{\log n} \\ \boxed{\text{by (16) and (9)}} \quad &= \ell\big(|\mathsf{A}_h|/\tilde{n}\big)^2 \pm 2\delta_h \ell \pm \tfrac{4\varepsilon n}{\log n} \\ \boxed{\text{we have } \delta_h < \delta_i \text{ since } h < i} \quad &= \ell |\mathsf{A}_h|^2 \tilde{n}^{-2} \pm 3\delta_i \ell . \end{aligned} \tag{52}$$

Thus we have

$$
\mathbf{P}\big[|\psi_i(v_i) - \psi_i(v_i^+)| \in I_{\mathsf{E}}\big|\mathscr{H}_{i-1}\big]
$$

$$
= \frac{\big|\{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i) : |a - \psi_h(v_h)| \in I_{\mathsf{E}}\}\big|}{\big|\mathrm{Adm}\big(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i\big)\big|} \tag{53}
$$

$$
\boxed{\text{by (52) and (51)}} \quad = \frac{\big|\{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h) : |a - \psi_h(v_h)| \in I_{\mathsf{E}}\}\big| \pm \frac{4\varepsilon n}{\log n}}{\ell|\mathsf{A}_h|^2 \tilde{n}^{-2} \pm 3\delta_i\ell} \,.
$$

We now need to make some effort to transform the error $\pm 2\delta_i\ell$ from the denominator and the term $\frac{4\varepsilon n}{\log n}$ from the nominator in a way that will be convenient later. Let us do some preparations first. Let us write $w_1 := |\{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h) : |a - \psi_h(v_h)| \in I_{\mathsf{E}}\}|$, $w_1' := w_1 \pm \frac{4\varepsilon n}{\log n}$ and $w_2 := \ell|\mathsf{A}_h|^2\tilde{n}^{-2}$. Recall that $|\mathsf{A}_h| \geq \gamma\tilde{n}$, and so

$$
w_2 \geq \gamma^2\ell \,. \tag{54}
$$

We have

$$
\frac{\delta_i\ell}{w_2} \overset{(54)}{\leq} \frac{\delta_i}{\gamma^2} < 0.01 \,. \tag{55}
$$

We have

$$
\frac{w_1'}{w_2 \pm 3\delta_i\ell} = \frac{w_1}{w_2} \cdot \frac{1}{1 \pm \frac{3\delta_i\ell}{w_2}} \pm \frac{\frac{4\varepsilon n}{\log n}}{w_2 - 3\delta_i\ell}
$$

$$
\boxed{\text{using the fact } \frac{1}{1\pm x} = 1 \pm 2x \text{ valid for } |x| < 0.2, \text{ c.f. (55)}} \quad = \frac{w_1}{w_2} \cdot \left(1 \pm \frac{6\delta_i\ell}{w_2}\right) \pm \frac{4\varepsilon n}{0.97 w_2 \log n} \tag{56}
$$

$$
\boxed{\text{using (54) and (55)}} \quad = \frac{w_1}{w_2} \cdot \left(1 \pm \frac{6\delta_i}{\gamma^2}\right) \pm \frac{12\varepsilon n}{\gamma^2\ell\log n} \,.
$$

We can thus continue (53) as follows,

$$
\mathbf{P}\big[|\psi_i(v_i) - \psi_i(v_i^+)| \in I_{\mathsf{E}}\big|\mathscr{H}_{i-1}\big]
$$

$$
= \frac{\big|\{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h) : |a - \psi_h(v_h)| \in I_{\mathsf{E}}\}\big| \pm \frac{4\varepsilon n}{\log n}}{\ell|\mathsf{A}_h|^2\tilde{n}^{-2} \pm 3\delta_i\ell} \tag{57}
$$

$$
= \frac{w_1'}{w_2 \pm 3\delta_i\ell} \overset{(56)}{=} \frac{w_1}{w_2} \cdot \left(1 \pm \frac{6\delta_i\ell}{w_2}\right) \pm \frac{12\varepsilon n}{\gamma^2\ell\log n}
$$

$$
\boxed{\text{by (33)}} \quad = q_{I_{\mathsf{E}},i}\left(1 \pm \frac{6\delta_i}{\gamma^2}\right) \pm \frac{12\varepsilon n}{\gamma^2\ell\log n} \,.
$$

We now argue that a similar equation for $q_{X,i}$ holds. Given $X \in \mathcal{X}$, we have

$$
\sum_{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i)} \big|\{X' \in X(\mathsf{A}_i, \mathsf{B}_i) : a \in \mathrm{Ch}_{\mathsf{V}}(X'; X) \text{ or } |a - \psi_h(v_h)| \in \mathrm{Ch}_{\mathsf{E}}(X'; X)\}\big|
$$

$$
= \sum_{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h)} \big|\{X' \in X(\mathsf{A}_h, \mathsf{B}_h) : a \in \mathrm{Ch}_{\mathsf{V}}(X'; X) \text{ or } |a - \psi_h(v_h)| \in \mathrm{Ch}_{\mathsf{E}}(X'; X)\}\big| \pm 24 \cdot \frac{2\varepsilon n}{\log n}
$$

and putting this together with (52) we obtain

$$\mathbf{E}\Big[\big|\{X' \in X(\mathsf{A}_i, \mathsf{B}_i) : \psi_i(v_i) \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } |\psi_i(v_i) - \psi_h(v_h)| \in \mathrm{Ch}_\mathsf{E}(X'; X)\}\big|\,\Big|\mathscr{H}_{i-1}\Big]$$

$$= \frac{\displaystyle\sum_{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i)} \big|\{X' \in X(\mathsf{A}_i, \mathsf{B}_i) : a \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } |a - \psi_h(v_h)| \in \mathrm{Ch}_\mathsf{E}(X'; X)\}\big|}{\big|\mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_i, \mathsf{B}_i)\big|}$$

$$= \frac{\displaystyle\sum_{a \in \mathrm{Adm}(\psi_h(v_h), J(v_i); \mathsf{A}_h, \mathsf{B}_h)} \big|\{X' \in X(\mathsf{A}_h, \mathsf{B}_h) : a \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } |a - \psi_h(v_h)| \in \mathrm{Ch}_\mathsf{E}(X'; X)\}\big| \pm \frac{50\varepsilon n}{\log n}}{\ell|\mathsf{A}_h|^2 \tilde{n}^{-2} \pm 3\delta_i\ell}\,.$$

Now, dealing with the error terms as in (56), we obtain

$$\mathbf{E}\Big[\big|\{X' \in X(\mathsf{A}_i, \mathsf{B}_i) : \psi_i(v_i) \in \mathrm{Ch}_\mathsf{V}(X'; X) \text{ or } |\psi_i(v_i) - \psi_h(v_h)| \in \mathrm{Ch}_\mathsf{E}(X'; X)\}\big|\,\Big|\mathscr{H}_{i-1}\Big]$$

$$\stackrel{(33)}{=} q_{X,i}\big(1 \pm \tfrac{50\delta_i}{\gamma^2}\big) \pm \tfrac{400\varepsilon n}{\gamma^2\ell\log n}\,. \tag{58}$$

We are finally in position to estimate $\sum_{i=2}^{t} \mathbf{P}\big[|\psi_i(v_i) - \psi_i(v_i^+)| \in I_\mathsf{E}\big|\mathscr{H}_{i-1}\big]$, which is a key quantity in order to verify (QUASI1). Putting together Claim 4 and (57), we have

$$\sum_{i=2}^{t} \mathbf{P}\big[|\psi_i(v_i) - \psi_i(v_i^+)| \in I_\mathsf{E}\big|\mathscr{H}_{i-1}\big] = \sum_{i=1}^{t}\Big(q_{I_\mathsf{E},i}\big(1 \pm \tfrac{6\delta_i}{\gamma^2}\big) \pm \tfrac{12\varepsilon n}{\gamma^2\ell\log n}\Big)$$

$$= \sum_{k=1}^{\lceil\frac{t}{\delta n}\rceil}\Big(\big(1 \pm \tfrac{6\delta_{k\delta n}}{\gamma^2}\big)\sum_{i=(k-1)\delta n+1}^{\max(t, k\delta n)} q_{I_\mathsf{E},i}\Big) \pm \tfrac{12\varepsilon n^2}{\gamma^2\ell\log n}$$

$$\boxed{\text{by (34)}} \quad = \sum_{k=1}^{\lceil\frac{t}{\delta n}\rceil}\Big(\big(1 \pm \tfrac{6\delta_{k\delta n}}{\gamma^2}\big)\sum_{i=(k-1)\delta n+1}^{\max(t, k\delta n)} p_{I_\mathsf{E},i} \pm \tfrac{4000\delta_{k\delta n}\delta m}{\gamma^4}\Big) \pm 50\delta m$$

$$\boxed{\text{by (22)}} \quad = \sum_{i=1}^{t} p_{I_\mathsf{E},i} \pm \sum_{k=1}^{\lceil\frac{t}{\delta n}\rceil}\Big(\tfrac{6\delta_{k\delta n}}{\gamma^2}\cdot\tfrac{2\delta n}{|\mathscr{J}|} + \tfrac{4000\delta_{k\delta n}\delta m}{\gamma^4}\Big) \pm 50\delta m$$

$$\boxed{\text{by (3)}} \quad = \sum_{i=1}^{t} p_{I_\mathsf{E},i} \pm \sum_{k=1}^{\lceil\frac{t}{\delta n}\rceil}\Big(\tfrac{24\delta_{k\delta n}\delta m}{\gamma^2} + \tfrac{4000\delta_{k\delta n}\delta m}{\gamma^4}\Big) \pm 50\delta m$$

$$\boxed{\text{by (4), recall } \mu^{-1} \gg \gamma^{-4} \text{ by Setup 9}} \quad = \sum_{i=1}^{t} p_{I_\mathsf{E},i} \pm \tfrac{1}{4}\delta_t m\,.$$

$$\tag{59}$$

We apply Lemma 6 for random variables $\big(\mathbb{1}_{|\psi_i(v_i)-\psi_i(v_i^+)|\in I_\mathsf{E}}\big)_{i=2}^{t}$ and the event $\mathcal{E}$. Let us go through the assumption of Lemma 6. Histories naturally generate a filtration, as explained in Section 3.5. Obviously, our random variables are bounded from above by 1. Last, (59) calculates the expectation needed for the lemma. Lemma 6 tells us that with probability at

least $1 - n^{-10}$ if $\mathcal{E}$ occurs we have

$$\sum_{i=2}^{t} \mathbb{1}_{|\psi_i(v_i)-\psi_i(v_i^+)|\in I_{\mathsf{E}}} = \sum_{i=1}^{t} p_{I_{\mathsf{E}},i} \pm \tfrac{1}{4}\delta_t m \pm \delta n \,,$$

and, using Claim 3, we conclude that

$$|\mathsf{B}_{t+1} \cap I_{\mathsf{E}}| = |I_{\mathsf{E}}| - \sum_{i=2}^{t} \mathbb{1}_{|\psi_i(v_i)-\psi_i(v_i^+)|\in I_{\mathsf{E}}} - \sum_{i=1}^{t} r_{I_{\mathsf{E}},i} = |I_{\mathsf{E}}| - \tfrac{t}{|\mathcal{J}|} \pm \tfrac{3}{4}\delta_t m \overset{(19)}{=} \frac{m|\mathsf{A}_{t+1}|}{\tilde{n}} \pm \delta_{t+1} m$$

as required for (QUASI1). Taking the union bound over $I_{\mathsf{E}}$, with probability at least $1 - n^{-9}$, if $\mathcal{E}$ occurs we have (QUASI1) (with parameter $\delta_{t+1}$) for $(\mathsf{A}_{t+1}, \mathsf{B}_{t+1})$.

By a similar argument, but using (58) in place of (57), with probability at least $1 - n^{-5}$, if $\mathcal{E}$ occurs we have (QUASI2) (with parameter $\delta_{t+1}$) for $(\mathsf{A}_{t+1}, \mathsf{B}_{t+1})$. Putting these together, we see that if $\mathcal{E}$ occurs then with high probability we fulfilled Definition 19. More precisely, the following event has probability at least $1 - n^{-4}$. Either $\mathcal{E}$ does not occur, or $(\mathsf{A}_{t+1}, \mathsf{B}_{t+1})$ is $\delta_{t+1}$-quasirandom (and therefore Algorithm 1 does not fail at time $t+1$).

Taking the union bound over $1 \leq t \leq n$, we conclude by induction that with probability at least $1 - n^{-3}$ the following event occurs. Either $\mathcal{E}$ dos not occur, or $(\mathsf{A}_{t+1}, \mathsf{B}_{t+1})$ is $\delta_{t+1}$-quasirandom for each $1 \leq t \leq n$ (and so Algorithm 1 does not fail at any time).

Finally, as we noted above, $\mathcal{E}$ occurs with probability at least $1 - 7n^{-1}$, so we conclude that Algorithm 1 succeeds with probability at least $1 - 8n^{-1} > 0$. When it succeeds, the resulting $\psi_n$ is the desired graceful labelling of $T$ with $(1 + \gamma)n$ labels. $\qquad\square$

## 5. Concluding remarks

5.1. **Improvements on Theorem 4.** It would be desirable to remove the degree constraint of Theorem 4 and prove that all sufficiently large trees have approximate graceful labellings. Inspection of our proof reveals that the log-factor in our degree bound is required only in order to have polynomially small probabilities in various places, which in turn we need to apply the union bound. If we could somehow do without this, we would otherwise require only $\Delta(T) \leq \eta n$. In our proof, we take the union bound over all times $t$ and intervals $S \subseteq [t]$, and over all structures $\mathcal{X}$ (and over a bounded number of other choices). The former can be avoided: it suffices to establish quasirandomness for $t$ a multiple of $\varepsilon n$, and similarly to discretise the choice of $S$ in (PRE4). The latter cannot so easily be avoided: we need to ensure that (for example) at each time $t$ there are admissible vertices for labelling each $v_t$, which requires $X_3 \llbracket \psi_{t-1}(v_t^+), J(v_t) \rrbracket (\mathsf{A}_t, \mathsf{B}_t)$ to be non-empty. We cannot afford occasional failures here, which is what we would expect if the probability of its being empty were a small constant, rather than polynomially small. Nevertheless, it is possible that with more care, following a strategy similar to that presented here one can handle all trees with maximum degree $\varepsilon n$, and perhaps even all trees.

It would be very interesting to obtain genuine graceful labellings for a large class of trees. Perhaps the approach here can be put together with the Absorbing Method for this purpose. We are currently investigating this possibility.

5.2. **Bipartite graceful labelling.** We believe our result can be extended to the notion introduced as $\alpha$-valuations by Rosa [21], and now commonly referred to as *bipartite graceful labellings*. Namely, we say that a labelling $\psi$ of a bipartite graph is *bipartite graceful* if it is graceful and all vertex labels in one colour class $V_1$ are smaller than in the other $V_2$. However, already Rosa found out that the concept is too restrictive in that the bipartite version of Conjecture 1 does not hold (for example, a complete ternary tree of depth 2 and with 13 vertices does not have a bipartite graceful labelling). However, it seems likely that small modifications to our proof can be used to show that the trees as in Theorem 4 have approximate bipartite graceful labellings. That is, the labelling $\psi$ in Theorem 4 can be taken such that $\psi(V_1) \subseteq \{1, \ldots, |V_1| + \frac{\gamma n}{2}\}$ and $\psi(V_2) \subseteq \{|V_1| + \frac{\gamma n}{2} + 1, \ldots, n + \gamma n\}$.

Let us briefly sketch the required modifications. By a similar reduction as that to Theorem 8, we can assume that $|V_1|, |V_2| \geq \gamma n$ in addition to any divisibility properties we need. We need to redefine $J$ and 'complementary interval' to replace the 'midpoint' $\frac{1+\gamma}{2}n$ in our proof with $|V_1| + \frac{\gamma n}{2}$. We then need to alter our procedure of picking the intervals $J(v_i)$, insisting that $J(v_i)$ is always below the midpoint for $v_i \in V_1$, and always above for $v_i \in V_2$, but otherwise following the same random procedure. We will also need to alter the distributions $\mathrm{Cor_V}$ and $\mathrm{Cor_E}$ appropriately. Finally, we apply Algorithm 1, and claim that it succeeds with high probability; doing so it returns an approximate bipartite graceful labelling.

In order to prove this, it is necessary to change (PRE4), taking into account the ratio of vertices in $V_1$ and in $V_2$ on the interval $S$. Similarly, it is necessary to change the quasirandomness definition (QUASI2), again taking into account the ratio of vertices in $V_1$ and $V_2$ up to time $t$. We expect that these modifications to our proof suffice, but these promise significant extra technical complexity, and we have not checked the details.

In light of this sketch, we do not think Theorem 4 should be seen as strong evidence in favour of the Graceful Tree Conjecture. The above sketch would be similarly strong evidence in favour of the statement that all trees admit a bipartite graceful labelling: which is false.

5.3. **Harmonious labellings.** The concept of *harmonious labellings* is the same as of graceful labellings, except that the formula $|f(x) - f(y)|$ defining the label induced on the edge $xy$ is replaced by $(f(x) + f(y)) \mod q$, where $q$ is the number of edges of the graph in question. This concept was introduced by Graham and Sloane [11] who also put forward the counterpart to Conjecture 1.

**Conjecture 22** (Harmonious Tree Conjecture)**.** *For any n-vertex tree $T$ there exists an injective labelling $\psi : V(T) \to [n]$ such that the values*

$$\Big(\psi(u) + \psi(v) \mod (n-1)\Big)_{uv \in E(T)} \tag{60}$$

*are pairwise distinct.*

Actually, Conjecture 22 can be generalized to Abelian groups. In that setting, the conjecture says that given an $n$-vertex tree $T$ and an Abelian group $\Gamma$ of order $n$ there exists an injective labelling $\psi : V(T) \to \Gamma$ such that the values $\psi(u) + \psi(v)$ on the edges $uv \in E(T)$ are pairwise distinct. Firstly, note that the original Conjecture 22 corresponds to $\Gamma = \mathbb{Z}_n$. Secondly, observe that if the generalized conjecture holds for all trees and all Abelian groups of the order exactly as of the tree, then it holds also for all trees and all Abelian groups of the order which is at least the order of the tree.

Conjecture 22 is, too, open. The strongest result by far, obtained very recently by Montgomery, Pokrovskiy, and Sudakov [17], is an asymptotic solution of the group-theoretic version of Conjecture 22.

**Theorem 23.** *For every $\gamma > 0$ there exists $n_0 \in \mathbb{N}$ such for every $n > n_0$, every $n$-vertex tree $T$ and every Abelian group $\Gamma$ of order at least $(1+\gamma)n$, there exists a map $\psi : V(T) \to \Gamma$ such that the values $\psi(u) + \psi(v)$ on the edges $uv \in E(T)$ are pairwise distinct.*

Theorem 23 is a quick consequence of results on containment of rainbow trees, which are the main focus of [17]. The most notable feature of Theorem 23, compared to our Theorem 4 as well as the tree packing results mentioned in Section 1.2, is that there is no upper bound on the maximum degree of $T$.

Since the current paper appeared before [17], and since the methods used in [17] are very different, we would like to comment how the tools we introduced here may be used to obtain a counterpart of Theorem 4 for harmonious labellings. The bound on the maximum degree of the tree $T$ would stay $\frac{\eta n}{\log n}$. The relaxation compared to Conjecture 22 would amount to $\psi$ mapping to $[\tilde{n}]$, $\tilde{n} := \lceil (1+\gamma)n \rceil$, and modulus in (60) being $\tilde{n} - 1$.

In order for our analysis of Algorithm 1 to work, we need the property that the marginal distributions of vertex labels and edge labels are close to uniform throughout the whole process. We obtained this by our careful choice of the sets $J(v)$ for $v \in V(T)$. Once we have this, the remaining analysis does not essentially require gracefulness.

In order to modify our method to work for harmonious labellings we would choose $\mathcal{J}$ as follows. We consider $[\tilde{n}]$ with the natural cyclic order, and let $\mathcal{J}$ be the collection of intervals of length $\ell - 1$ in this order starting at $1, m+1, \ldots, \tilde{n} - m + 1$. We would not need to define 'complementary interval'. Then, in Lemma 11, we would simply choose $J(v)$ independently and uniformly at random from $\mathcal{J}$ for each $v \in V(T)$. It is now obvious that if each $v$ were labelled uniformly in $J(v)$ then the result is a uniform distribution of vertex labels, and easy to check that the distribution of edge labels is also uniform. We expect that from this point one can simply follow the algorithm and analysis given, making the obvious small changes to obtain a harmonious rather than graceful labelling. However we have not checked the details.

Note that in our approach we cut the tree $T$ into small subtrees by removing the edges $\mathcal{R}$. This is not used only to assign the intervals $J(v)$, but also to guarantee that most vertices

are labelled shortly after their parent is labelled. The former property is not required for harmonious labelling, but the latter property is still required for the analysis.

## 6. ACKNOWLEDGMENT

## REFERENCES

1. P. Allen, J. Böttcher, J. Hladký, and D. Piguet, *Packing degenerate graphs*, Adv. Math. **354** (2019), 106739.
2. J. C. Bermond, *Graceful graphs, radio antennae and French windmills*, Graph Theory and Combinatorics, Pitman, London, 1979, pp. 18–37.
3. J. Böttcher, J. Hladký, D. Piguet, and A. Taraz, *An approximate version of the tree packing conjecture*, Israel J. Math. **211** (2016), no. 1, 391–446.
4. F. Van Bussel, *Relaxed graceful labellings of trees*, Electronic J. Combin. **9** (2002), 12p.
5. W. C. Chen, H. I. Lü, and Y. N. Yeh, *Operations of interlaced trees and graceful trees*, Southeast Asian Bull. Math. **21** (1997), 337–348.
6. A. Ferber, C. Lee, and F. Mousset, *Packing spanning graphs from separable families*, Israel J. Math. **219** (2017), no. 2, 959–982.
7. A. Ferber and W. Samotij, *Packing trees of unbounded degrees in random graphs*, J. Lond. Math. Soc. (2) **99** (2019), no. 3, 653–677.
8. R. Frucht, *Graceful numbering of wheels and related graphs*, Annals New York Academy of Sciences **319** (1979), 219–229.
9. J.A. Gallian, *A dynamic survey of graph labeling*, Electronic J. Combin. (2014), 384p.
10. S. W. Golomb, *How to number a graph*, Graph Theory and Computing (R. C. Read, ed.), Academic Press, 1972, pp. 23–37.
11. R. L. Graham and N. J. A. Sloane, *On additive bases and harmonious graphs*, SIAM Journal on Algebraic and Discrete Methods **1** (1980), no. 4, 382–404.
12. W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.
13. S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
14. F. Joos, J. Kim, D. Kühn, and D. Osthus, *Optimal packings of bounded degree trees*, arXiv:1606.03953, to appear in Journal of the EMS.
15. J. Kim, D. Kühn, D. Osthus, and M. Tyomkyn, *A blow-up lemma for approximate decompositions*, Trans. Amer. Math. Soc. **371** (2019), no. 7, 4655–4742.
16. S. Messuti, V. Rödl, and M. Schacht, *Packing minor-closed families of graphs into complete graphs*, J. Combin. Theory Ser. B **119** (2016), 245–265.

17. R. Montgomery, A. Pokrovskiy, and B. Sudakov, *Embedding rainbow trees with applications to graph labelling and decomposition*, arXiv:1803.03316, to appear in Journal of the EMS.

18. J. W. Moon, *On the maximum degree in a random tree*, Michigan Math. J. **15** (1968), no. 4, 429–432.

19. A. M. Pastel and H. Raynaud, *Numerotation gracieuse des oliviers*, Colloq. Grenoble, Publications Université de Grenoble, 1978, pp. 218–223.

20. G. Ringel, *Problem 25*, Theory of Graphs and its Applications (Proc. Sympos. Smolenice 1963), Nakl. CSAV, Praha, 1964, p. 162.

21. A. Rosa, *On certain valuations of the vertices of a graph*, Theory of Graphs (Internat. Symposium, Rome, July 1966), Gordon and Breach, N. Y. and Dunod Paris, 1967, pp. 349–355.

22. G. Sethuraman and J. Jesintha, *All banana trees are graceful*, Advances Appl. Disc. Math. **4** (2009), 53–64.

23. T.-M. Wang, C.-C. Yang, L.-H. Hsu, and E. Cheng, *Infinitely many equivalent versions of the graceful tree conjecture*, Appl. Anal. Discrete Math. **9** (2015), 1–12.

24. N. C. Wormald, *Differential equations for random processes and random graphs*, Ann. Appl. Probab. **5** (1995), no. 4, 1217–1235.

*Affiliation at time of submission:* Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark.

*Email address*: `a.m.adamaszek@gmail.com`

Department of Mathematics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

*Email address*: `p.d.allen@lse.ac.uk`

*Affiliation at time of submission:* Freie Universität Berlin, Germany.

*Email address*: `grosu.codrut@gmail.com`

Institute of Mathematics of the Czech Academy of Sciences, Žitná 25, Praha.

*Email address*: `hladky@math.cas.cz`