Connectivity in bridge-addable graph classes: the McDiarmid-Steger-Welsh conjecture

Guillaume Chapuy* [*] and Guillem Perarnau[†] [**]

[*]IRIF, UMR CNRS 8243, *Université Paris-Diderot, France.*
CRM, UMI CNRS 3457, *Université de Montréal, Canada.*
[**]*School of Computer Science, McGill University, Montréal, QC, Canada.*

April 27, 2017

Abstract

A class of graphs is *bridge-addable* if given a graph G in the class, any graph obtained by adding an edge between two connected components of G is also in the class. We prove a conjecture of McDiarmid, Steger, and Welsh, that says that if $\mathcal{G}_n$ is any class of bridge-addable graphs on n vertices, and G_n is taken uniformly at random from $\mathcal{G}_n$, then G_n is connected with probability at least $e^{-\frac{1}{2}} + o(1)$, when n tends to infinity. This lower bound is asymptotically best possible since it is reached for forests.

Our proof uses a "local double counting" strategy that may be of independent interest, and that enables us to compare the size of two sets of combinatorial objects by solving a related multivariate optimization problem. In our case, the optimization problem deals with partition functions of trees relative to a supermultiplicative functional.

1 Introduction, notation, main result

In this paper, unless otherwise stated, all graphs are finite, simple and with vertex set $\{1, \ldots, n\}$ for some $n \geq 1$. Following [MSW06], we say that a family $\mathcal{G}$ of graphs is *bridge-addable* if the following is true:

> If G is a graph from $\mathcal{G}$, and if e is an edge not in G whose endpoints belong to two different connected components of G, then the graph $G \cup \{e\}$ obtained by adding e to G is in $\mathcal{G}$.

The notion of bridge-addability was motivated by the study of connectivity in random planar graphs (since the class of planar graphs is clearly bridge-addable). Other examples of bridge-addable classes

---

include forests, triangle-free graphs, graphs having a perfect matching, or any minor closed class of graphs whose excluded minors are 2-connected. See [MSW06, ABMR12] for even more examples.

In [MSW05, MSW06], the authors investigate the properties of a graph taken uniformly at random from a bridge-addable class. In particular they show the following:

Proposition 1 ([MSW05]). *For every $\epsilon > 0$, there exists an n_0 such that for every $n \geq n_0$ and any bridge-addable class $\mathcal{G}_n$ of graphs with n vertices, we have*

$$\Pr\left(G_n \text{ is connected}\right) \geq (1 - \epsilon)e^{-1} , \tag{1}$$

where G_n is chosen uniformly at random from $\mathcal{G}_n$.

The authors of [MSW06] conjectured that the constant e^{-1} in (1) can be improved to $e^{-1/2}$. If true, this would be best possible, since it is proved in [Rén59] that if F_n is a uniform random forest on n vertices, then one has when n tends to infinity:

$$\Pr\left(F_n \text{ is connected}\right) \longrightarrow e^{-1/2}.$$

The first partial result towards the proof of the conjecture was obtained by Balister, Bollobás, and Gerke [BBG08] who improved the constant in (1) from e^{-1} to $e^{-0.7983}$. Norin, in an unpublished draft [Nor], improves it to $e^{-2/3}$. Until the present paper these were, as far as we know, the best results under general hypotheses.

Under the (much) stronger hypothesis that the class is also *bridge-alterable* (i.e. that the class is *also* stable by bridge deletion), Addario-Berry, McDiarmid and Reed [ABMR12], and Kang and Panagiotou [KP13] independently improved the constant to the general conjectured value $e^{-1/2}$. Both proofs use the fact that graphs from bridge-alterable classes can be encoded by weighted forests, so that the problem reduces to estimating the connectivity in a random weighted forest. Unfortunately, this encoding fails for general bridge-addable classes, so these proof techniques do not apply to the general case. Note also that many bridge-addable classes are *not* bridge-alterable, for example graphs that admit a perfect matching, graphs that have a component of large size, or graphs in any bridge-addable class that contain a given subgraph (for example, planar graphs containing a path of given length).

In this paper, we prove the general conjecture, *i.e.* we improve the constant in (1) to $e^{-1/2}$ using only the hypothesis of bridge-addability. Our main result is the following:

Theorem 2. *The McDiarmid-Steger-Welsh conjecture is true: For every $\epsilon > 0$, there exists an n_0 such that for every $n \geq n_0$ and any bridge-addable class $\mathcal{G}_n$ of graphs with n vertices we have*

$$\Pr\left(G_n \text{ is connected}\right) \geq (1 - \epsilon)e^{-1/2} ,$$

where G_n is chosen uniformly at random from $\mathcal{G}_n$.

We believe that, beyond the own interest of the conjecture, our proof is interesting for several other reasons. First, the resolution of this problem promotes further this "alternative approach" to random graphs with constraints, and raises the question of finding other class properties, beyond bridge-addability, that give rise to interesting and non-trivial predictable asymptotic behaviour. For example McDiarmid and Scott [MS] have studied *block-classes* (where the class property is stability under taking a two-connected component) and Balister, Bollobás and Gerke [BBG08] have studied the property of 2-*addability*. One can expect that in years to come both more examples and a

greater unification of the techniques will be possible. Second, our proof is based on a combinatorial technique that we call "local double counting" that is interesting on its own and that we believe may be useful in many other situations. We give a quick description of this technique in the next section: roughly speaking, it mixes a classical double counting technique already used in previous works on this topic with a careful analysis of the *local structure* of the objects under study. The idea of taking advantage of the local structure of the graphs to improve the double-counting argument has already been used in the papers [BBG08, Nor] via a *weighted* version of the double-counting setup. As far as we know, the technique we introduce here is of a different nature. In particular, to prevent any confusion, we mention that the word *weight* in this paper is used as in statistical mechanics to refer to the pondering used in the definition of some partition functions that appear in an optimization problem inherited from our double-counting arguments. However here the underlying double-counting is *unweighted* in nature, and the efficiency of our method comes from the fact it is performed *locally* (see Section 2).

Finally, our proof gives more information on random graphs from bridge-addable classes than what is contained in Theorem 2. We are able to quantify the probability of having any small number of connected components:

Theorem 3. *For every $\epsilon > 0$ and for every $k \geq 0$, there exists an n_0 such that for $n \geq n_0$ one has:*

$$\Pr\left(G_n \text{ has at most } k+1 \text{ connected components}\right) \geq \Pr\left(Poisson\left(\frac{1}{2}\right) \leq k\right) - \epsilon.$$

The proof is a simple extension of the proof of Theorem 2 (that corresponds to $k = 0$).

In the rest of the paper, we fix an integer n and a bridge-addable class $\mathcal{G}_n$ such that all graphs in $\mathcal{G}_n$ have n vertices. We let $\mathcal{G}_n^{(i)}$ be the graphs in $\mathcal{G}_n$ having exactly i connected components, and we use the shortcut notations $\mathcal{A}_n := \mathcal{G}_n^{(1)}$ and $\mathcal{B}_n := \mathcal{G}_n^{(2)}$. The main ingredient in the proof of Theorem 2 is the following proposition:

Proposition 4. *For all $\epsilon' > 0$ there exists n_0 such that for $n \geq n_0$ one has:*

$$|\mathcal{B}_n| \leq \left(\frac{1}{2} + \epsilon'\right)|\mathcal{A}_n|. \tag{2}$$

Structure of the paper. We start in Section 2 with a high level discussion on the "local double counting" strategy that we use to prove our main result and that may be of independent interest – the reader only interested in the proofs can skip this discussion. In Section 3, we prove Theorems 2 and 3 admitting Proposition 4. The proof of Proposition 4 occupies the rest of the paper. In Section 4, we define the local parametrization of our graph classes (Section 4.1), we state the main combinatorial double counting argument (Section 4.2), and we use it to obtain a local bound (Corollary 12 in Section 4.3) on a functional of some ratios relating the size of $\mathcal{A}_n$ to the size of certain subsets of $\mathcal{B}_n$. This functional takes the form of a truncated partition function of rooted trees carrying some supermultiplicative weights. In order to use this bound, we study partition functions of trees in Section 5. In Section 5.1, we first study untruncated partition functions, and we relate the rooted and unrooted case via an analogue of the dissymmetry theorem (Lemma 14). In Section 5.2, we transfer the results of Section 5.1 to the setting of truncated partition functions (Proposition 15). Finally, in Section 6, we finish the proof of Proposition 4. In Section 6.1, we show that we can choose a good local partitioning of our graphs classes, that enables us to apply the results of Sections 4 and 5 to conclude the proof in Section 6.2.

2 The "local double counting" approach.

Previous approaches to the problem are based on double counting arguments that, using the addability hypothesis, enable to compare the proportion of graphs having different number of connected components in a class. The basic tool underlying the double counting, which is useful in many situations in combinatorics, is the following: to compare the sizes of two sets $\mathcal{A}$ and $\mathcal{B}$, construct a bipartite graph H on $(\mathcal{A}, \mathcal{B})$ in a way that we can control the degrees of vertices on each side; if all vertices from $\mathcal{A}$ have degree at least $d_{\mathcal{A}}$, and all vertices from $\mathcal{B}$ have degree at most d_B, then $|\mathcal{B}| \leq d_{\mathcal{A}}/d_{\mathcal{B}}|\mathcal{A}|$. In the context of addable classes, the roles of the sets $\mathcal{A}$ and $\mathcal{B}$ are played by $\mathcal{A}_n$ and $\mathcal{B}_n$ with previous notation, and the adjacencies in H are based on the relation of edge deletion. A classical way to strengthen this method is to apply it with an *edge-weighted* bipartite graph structure, for a well chosen notion of weight: this technique is useful in many situations – for example it is this refinement that is used in [BBG08, Nor]. The weights enable to make the double counting more sensitive to the particular structure of the elements of $\mathcal{A}$ and $\mathcal{B}$. Unfortunately, even with the use of edge-weights, approaches that are solely based on a global double counting argument seem to fall short to prove Theorem 2. Our novel approach does not consider edge-weights and exploits the local structure in a different way.

The main feature of our approach, that enables us to reach a tight bound, is that we exploit thoroughly the locality of the adjacencies in H. To this end, we design a "local double counting" strategy, based on several different steps, that we believe deserves to be described at a general level. Indeed it may be useful in many other situations where one wants to compare the size of two combinatorial sets and neither a classical double counting nor a weighted one lead to sharp enough bounds. This is the purpose of this section.

The locality of our approach appears at two different levels. *First*, to each element in $\mathcal{A} \cup \mathcal{B}$ we associate a statistics α, with value in some finite dimensional space $\mathcal{E}$ that we call the parameter space, that is such that if two elements $a \in \mathcal{A}$ and $b \in \mathcal{B}$ are linked by an edge in the bipartite graph H, then their corresponding α-statistics are close in the space $\mathcal{E}$. Moreover, if the statistics α is well chosen, knowing the α-statistics of an element of $\mathcal{A} \cup \mathcal{B}$ allows us to give a more precise bound on its degree in H. For any $\alpha \in \mathcal{E}$ we can then group together, in $\mathcal{A}$ and $\mathcal{B}$, elements whose statistics belongs to a small neighbourhood of α into sets $\mathcal{A}_{[\alpha]}$ and $\mathcal{B}_{[\alpha]}$, and apply the double argument locally to obtain a bound on the "local ratio" $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$.

The *second* way in which our approach is local is that the statistics that we choose is *itself* local. Namely, in our case, the elements of $\mathcal{A} \cup \mathcal{B}$ are graphs, and the finite-dimensional parameter associated to a graph records the number of pendant copies of each tree under a certain size. It is thus a "local observable" – similar to the observables underlying the local limit convergence of [BS01].

In order to make use of this abstract set-up, one needs to be able to work quantitatively with the bounds obtained from this approach: this is done as follows. On the one hand, for each element α in the parameter space we have, provided the previous steps were successful, a relation between each coordinate of α and the ratios $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$. On the other hand, by construction, the statistics α satisfies some simple *global* constraints (in our case, its L^1-norm is smaller than n). Putting all these inequalities together, we obtain some global constraint on a function of the ratio $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$, and the problem of finding the best possible upper bound on the ratio $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$ reduces to an *optimization problem*: namely, how large can the ratio $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$ be, knowing that the global

constraint holds. The maximum of these bounds then leads[1] to a bound on the ratio $|\mathcal{B}|/|\mathcal{A}|$.

In the case addressed in the present paper, this optimization problem takes the somewhat explicit form of optimizing a partition function (or generating function) of unrooted trees, given the constraint that the corresponding partition function of rooted trees is bounded. It is important to note that here, the generating functions we study are *not* the generating functions of the objects in the graph class under study (which is *any* bridge-addable class). It is a generating function of the "local observables" that we have chosen to consider, and that are the same for any graph class. Hence the role of generating functions in the present work is very different from the cases of exactly solvable models, such as random series-parallel or planar graphs [BGKN07, GN09].

Another feature of our method in the present case is that in order to obtain sharper bounds, we also need to partition the set $\mathcal{B}$ into finitely many subsets $\mathcal{B}^U$ (where, in our case, the index U is some unlabeled tree from a fixed finite family). This extra partitioning is not a necessary feature of the "local double counting" strategy we are describing, but it makes it more general. For each U, we then consider the induced bipartite graph on $(\mathcal{A}, \mathcal{B}^U)$, and in each case, we apply the previous ideas to get a bound on the local ratios $|\mathcal{B}^U_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$. We then obtain a bound of the local ratio $|\mathcal{B}_{[\alpha]}|/|\mathcal{A}_{[\alpha]}|$ as a sum of these bounds. As before, an appropriate global constraint on the α-statistics gives rise to an optimization problem, which is now *multidimensional*: how large can the sum of these bounds be, given the global constraint. The number of variables of the problem is thus the number of indexing elements U.

To sum up this discussion, the "local double counting" technique may be useful to bound the ratio of the size of two sets $\mathcal{A}$ and $\mathcal{B}$ in situations where:

0. a classical double counting technique that constructs a bipartite graph on $(\mathcal{A}, \mathcal{B})$ leads to an interesting bound, but not sharp;
1. there is a natural statistics, with value in some multidimensional parameter space (possibly of dimension arbitrarily large), that makes the bipartite graph structure "local"; one expects this statistics to be itself a "local measurement" of the objects under study;
2. one can write a local double counting bound that is more precise than the global one; and
3. there is some global constraint on the statistics that translates into a constraint on the local ratios;

then the method should apply, and one can get in principle a bound (either lower or upper) on the ratios as the solution of an optimization problem. In the case where

4. the set $\mathcal{B}$ may be split into several sets $\mathcal{B}^U$ in order to refine the local bounds,

then the method applies as well, but the optimization problem becomes multidimensional. This will be the case in this paper.

3   Proof of the main results admitting Proposition 4

In this section, we show how to deduce our main results (Theorems 2 and 3) from Proposition 4. The proof of Proposition 4 itself is more complicated, and occupies the remaining sections.

First, the following result shows that the inequality (2) relating $\mathcal{A}_n = \mathcal{G}_n^{(1)}$ and $\mathcal{B}_n = \mathcal{G}_n^{(2)}$ can be "transferred" to a larger number of connected components:

---

[1]The actual situation is a bit more technical. Indeed because the locality of the adjacencies in H is only approximate, there is some overlap between the sets $\mathcal{A}_{[\alpha]}$ to take into account; this is easily handled with an averaging argument (Boxing lemma, Lemma 17).

Proposition 5. *Assume that Proposition 4 is true. Then for all $\epsilon' > 0$, and for each $i_0 \geq 1$, there exists n_0 such that for all $i \leq i_0$ and $n \geq n_0$ one has:*

$$i|\mathcal{G}_n^{(i+1)}| \leq \left(\frac{1}{2} + \epsilon'\right)|\mathcal{G}_n^{(i)}|$$

We note that a similar transfer principle was already used in [ABMR12, Lemma 3.1] in the context of alterable classes and with a different proof specific to that case. A similar argument was used without proof by Norin in his draft [Nor]. We thank Colin McDiarmid for detecting a technical flaw in a previous version of the proof of Proposition 5.

Proof. For every $i \geq 1$ and disjoint sets $V_1, \ldots, V_i$, we write $\Pr(V_1, \ldots, V_i)$ for the probability that for each $j \leq i$, the set V_j induces a connected component in the graph G_n chosen uniformly at random from $\mathcal{G}_n$. Beware that in the following we will use this notation in cases where $V_1, V_2, \ldots, V_i$ is partition of $[n] = \{1, 2, \ldots, n\}$ but also in other cases where it is not.

We consider the following total order on the subsets of $[n]$; for every $V_1, V_2 \subseteq V$ we have $V_1 > V_2$ if $|V_1| > |V_2|$ or $|V_1| = |V_2|$ and the elements of V_1 are smaller in the lexicographical order, than the ones in V_2. We remark that if $V_1 > \cdots > V_i$ is a partition of $[n]$, then $|V_1| \geq \frac{n}{i}$.

Then, with $\uplus$ denoting the disjoint union, we have:

$$\Pr(G \in \mathcal{G}_n^{(i+1)}) = \sum_{\substack{V_1 > \cdots > V_{i+1} \\ V_1 \uplus \cdots \uplus V_{i+1} = [n]}} \Pr(V_1, \ldots, V_{i+1})$$

$$= \frac{1}{i} \sum_{j=1}^{i} \sum_{\substack{V_1 > \cdots > V_{i+1} \\ V_1 \uplus \cdots \uplus V_{i+1} = [n]}} \Pr(V_1, \ldots, V_{i+1})$$

$$= \frac{1}{i} \sum_{\substack{W_1 > \cdots > W_i \\ W_1 \uplus \cdots \uplus W_i = [n]}} \sum_{\substack{W_1 = W_1^1 \uplus W_1^2 \\ W_1^1 > W_2, W_1^1 > W_1^2}} \Pr(W_1^1, W_1^2, W_2, \ldots, W_i),$$

where the last equality is just a change of index, that consists in noting $V_1 = W_1^1$, $V_{j+1} = W_1^2$, $W_1 = V_1 \uplus V_{j+1}$, and in summing first over the set W_1 and then over its partitions into two sets (here $W_2, \ldots, W_i$ denote the remaining sets V_k for $k \notin \{1, j+1\}$). Note the constraint $W_1^1 > W_2$, that comes from the fact that V_1 is the largest set of the partition (for $<$) before the change of index. If we remove this constraint, we only make the sum larger and we get the upper bound:

$$\Pr(G \in \mathcal{G}_n^{(i+1)}) \leq \frac{1}{i} \sum_{\substack{W_1 > \cdots > W_i \\ W_1 \uplus \cdots \uplus W_i = [n]}} \sum_{\substack{W_1 = W_1^1 \uplus W_1^2 \\ W_1^1 > W_1^2}} \Pr(W_1^1, W_1^2, W_2, \ldots, W_i)$$

For any set $W \subseteq V$ we write $G[W]$ for the graph induced by the set of vertices in W. For any graph Z, we denote by $G[W] \cong Z$ the event that W induces a connected component that it is isomorphic to Z. Given a partition $W_1 \uplus \cdots \uplus W_i = V$ such that $W_1 > \cdots > W_i$ and graphs $Z_2, \ldots Z_i$, consider the graph class

$$\mathcal{H}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i) = \{G[W_1]: \ G \in \mathcal{G}_n \text{ and } G[W_j] = Z_j \text{ for } 2 \leq j \leq i\} \, .$$

6

Notice that $\mathcal{H}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i)$ is a bridge-addable class on the set of vertices W_1. Let H be a graph chosen uniformly at random from $\mathcal{H}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i)$. By the remark above, H has order at least $\frac{n}{i+1}$. Let n_0 such that $\frac{n_0}{i_0+1} \geq n_0(\epsilon')$, where $n_0(\epsilon')$ is the constant that appears in Proposition 4. Using this proposition we have

$$\sum_{\substack{W_1 = W_1^1 \cup W_1^2 \\ W_1^1 > W_1^2}} \Pr(W_1^1, W_1^2, W_2, \ldots, W_i) = \sum_{Z_2, \ldots, Z_i} \Pr(H \in \mathcal{H}^{(2)}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i))$$

$$\cdot \Pr(H \in \mathcal{H}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i))$$

$$\leq \sum_{Z_2, \ldots, Z_i} \left(\frac{1}{2} + \epsilon'\right) \Pr(H \in \mathcal{H}^{(1)}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i))$$

$$\cdot \Pr(H \in \mathcal{H}(W_1, W_2 \ldots, W_i; Z_2, \ldots, Z_i))$$

$$= \left(\frac{1}{2} + \epsilon'\right) \Pr(W_1, W_2, \ldots, W_i) .$$

Thus returning to the previous bound we obtain:

$$i \Pr(G \in \mathcal{G}_n^{(i+1)}) \leq \left(\frac{1}{2} + \epsilon'\right) \sum_{\substack{W_1 > \cdots > W_i \\ W_1 \uplus \cdots \uplus W_i = [n]}} \Pr(W_1, W_2, \ldots, W_i)$$

$$= \left(\frac{1}{2} + \epsilon'\right) \Pr(G \in \mathcal{G}_n^{(i)}) . \qquad \square$$

To conclude the proof of Theorem 2, we will also need the following observation from [BBG08]. We include the proof for the sake of completeness.

Lemma 6 ([BBG08]). *For each $i, n \geq 1$ one has:*

$$i|\mathcal{G}_n^{(i+1)}| \leq |\mathcal{G}_n^{(i)}|.$$

Proof. Construct a bipartite graph H on the vertex set $(\mathcal{G}_n^{(i)}, \mathcal{G}_n^{(i+1)})$ by adding an edge between $G_1 \in \mathcal{G}_n^{(i)}$ and $G_2 \in \mathcal{G}_n^{(i+1)}$ if G_2 can be obtained from G_1 by removing an edge. Note that a graph $G_1 \in \mathcal{G}_n^{(i)}$ has degree at most $n - i$ in H, since G_1 has at most $n - i$ cut-edges. Moreover a graph $G_2 \in \mathcal{G}_n^{(i+1)}$ has degree at least $i(n-i)$, by the property of bridge-addability. By counting the edges of H in two different ways, we thus get:

$$i(n - i)|\mathcal{G}_n^{(i+1)}| \leq |E(H)| \leq (n - i)|\mathcal{G}_n^{(i)}|,$$

which concludes the proof. $\qquad \square$

We first prove the main theorem.

Proof of Theorem 2. Let $\epsilon' = \ln((1 - \epsilon)^{-1} - \epsilon)$ and let $x = \frac{1}{2} + \epsilon'$. Set i_0 to be large enough such that $\frac{1}{i_0!} < \epsilon \cdot e^{1/2}$. Using Proposition 5 recursively with ϵ' and i_0, we have that there exists an n_0 such that for every $i \leq i_0$ and for every $n \geq n_0$

$$|\mathcal{G}_n^{(i+1)}| \leq \frac{x^i}{i!}|\mathcal{A}_n| .$$

Moreover, for every $i > i_0$ we have from Lemma 6:

$$|\mathcal{G}_n^{(i+1)}| \leq \frac{1}{i!}|\mathcal{A}_n| .$$

Using both inequalities we obtain that for any $n \geq n_0$

$$|\mathcal{G}_n| = \sum_{i=1}^{n} |\mathcal{G}_n^{(i)}| \leq \left(\sum_{i=1}^{i_0} \frac{x^{i-1}}{(i-1)!} + \sum_{i=i_0+1}^{n} \frac{1}{i!} \right) |\mathcal{A}_n| \leq \left(e^x + \frac{1}{i_0!} \right) |\mathcal{A}_n|$$

$$\leq \left(e^{\frac{1}{2}+\epsilon'} + \epsilon \cdot e^{1/2} \right) |\mathcal{A}_n| = (1-\epsilon)^{-1} e^{1/2} |\mathcal{A}_n| .$$

The probability that a graph G_n chosen uniformly at random from $\mathcal{G}_n$ is connected is

$$\Pr\left(G_n \text{ is connected} \right) = \frac{|\mathcal{A}_n|}{|\mathcal{G}_n|} \geq (1-\epsilon)e^{-1/2} ,$$

provided that $n \geq n_0$. $\qquad\square$

We conclude by proving the extension of our main result.

Proof of Theorem 3. Let n be a large constant (to be fixed later) and let M_n be the number of components of a graph chosen uniformly at random in $\mathcal{G}_n$ minus one. By Lemma 6, there exists $i_0 \geq 1$ such that

$$\Pr(M_n \geq i_0 + 1) < \epsilon/2 . \qquad (3)$$

We can assume that $k \leq i_0$ since otherwise we are done by (3).

We will use a result on stochastic domination of Poisson distributions given by McDiarmid (see Lemma 3.3 in [McD12]). Given $\alpha > 0$ and $k_0 \geq 1$, if X is a non-negative integer-valued random variable such that for every $k = 0, 1, \ldots, k_0 - 1$ we have

$$\Pr(X = k+1) \leq \frac{\alpha}{k+1} \Pr(X = k) ,$$

then

$$\Pr(k_0 \geq X \geq k) \leq \Pr(\mathrm{Poisson}(\alpha) \geq k) . \qquad (4)$$

By Proposition 5 with the chosen i_0, for every ϵ' there exists an n_0 such that if $n \geq n_0$, the random variable M_n satisfies the hypothesis of Lemma 3.3 in [McD12] with $\alpha = \frac{1}{2} + \epsilon'$. Fix $k_0 = i_0$. Using (3) and (4), we obtain

$$
\begin{aligned}
\Pr(G_n \text{ has at most } k+1 \text{ connected components}) &= \Pr(M_n \leq k) \\
&= 1 - \Pr(i_0 \geq M_n \geq k+1) - \Pr(M_n \geq i_0 + 1) \\
&\geq 1 - \Pr(\mathrm{Poisson}(1/2 + \epsilon') \geq k+1) - \epsilon/2 \\
&= \Pr(\mathrm{Poisson}(1/2 + \epsilon') \leq k) - \epsilon/2 \\
&= \Pr(\mathrm{Poisson}(1/2) \leq k) - \epsilon ,
\end{aligned}
$$

where the last inequality follows provided that ϵ' is small enough with respect to ϵ. $\qquad\square$

8

4 Local double counting and local parameters

We now start the proof of the main technical estimate, namely Proposition 4. The following reduction will be very useful:

Lemma 7 ([BBG08, Lemma 2.1])**.** *Assume that Proposition 4 is true under the additional assumption that all graphs in $\mathcal{A}_n$ and $\mathcal{B}_n$ are forests. Then it is also true without this assumption.*

The proof of this lemma relies on a simple and beautiful argument that consists in splitting the sets $\mathcal{A}_n$ and $\mathcal{B}_n$ into equivalence classes depending on their 2-edge connected blocks, and then choosing a spanning tree arbitrarily in each block. This construction transforms any bridge-addable class of graphs into several bridge-addable classes of forests while preserving the distribution of the number of components, from which the lemma easily follows. We refer to [BBG08] for the full proof.

Thanks to the last lemma, for the rest of the paper, we will make the following assumption:

Assumption: *For all $n \geq 1$, all graphs in $\mathcal{A}_n$ and $\mathcal{B}_n$ are forests.*

4.1 Local parameters and partitions

In order to compare the sizes of $\mathcal{A}_n$ and $\mathcal{B}_n$, we will refine the double counting technique used in the proof of Lemma 6. We will again construct a bipartite graph on the vertex set $(\mathcal{A}_n, \mathcal{B}_n)$ where an edge is placed between $G_1 \in \mathcal{A}_n$ and $G_2 \in \mathcal{B}_n$ if one can be obtained from the other by the deletion of an edge. However, in order to obtain more precise bounds on the degrees of vertices in this bipartite graph, we will partition the sets $\mathcal{A}_n$ and $\mathcal{B}_n$ according to some local parameters of the graphs. Namely, to each graph G we will associate a statistics α^G that records, roughly speaking, the number of pendant copies in G of each tree from some finite family $\mathcal{T}_0$. The vectors α^G will be elements of a space called the parameter space and denoted by $\mathcal{E}$. The set $\mathcal{B}_n$ will be further partitioned according to the isomorphism type of the smallest component, and a special role will be played by subsets where this smallest component belongs to a finite family of trees called $\mathcal{U}_0$. The purpose of this subsection is to set notation and define these partionings.

We write $\mathcal{T}$ for the family of all rooted unlabeled trees and $\mathcal{U}$ for the family of all unrooted unlabeled trees. We also use $\mathcal{T}^\ell$ and $\mathcal{U}^\ell$ for the corresponding sets of *labeled* objects. For every tree $U \in \mathcal{U}$ we note by $\mathrm{Aut}_u(U)$ for the total number of automorphisms of U. For every $T \in \mathcal{T}$ we note by $\mathrm{Aut}_r(T)$ the number of automorphisms of T as a rooted tree (i.e. the number of automorphisms that fix the root of T).

In this section we will fix two finite families $\mathcal{U}_0 \subset \mathcal{U}$ and $\mathcal{T}_0 \subset \mathcal{T}$ such that
- $\mathcal{U}_0$ contains the only unrooted tree of order one (a single vertex).
- $\mathcal{T}_0$ is closed under rooted inclusion; that is, if $T \in \mathcal{T}$ with root at v and $T' \subseteq T$ is a subtree that contains the root, then $T' \in \mathcal{T}$.

We write $t_{\max}$ and $u_{\max}$ for the maximum number of vertices of a tree in $\mathcal{T}_0$ and $\mathcal{U}_0$ respectively.

Following the definition in [ABMR12], for each tree G and each edge e, the *pendant tree* of G in e is denoted by $s(G, e)$ and it is the smallest component of $G - e$, or the component containing vertex 1 if the components have equal size. Since the tree G has $n-1$ edges, every tree on n vertices has exactly $n - 1$ pendant trees.

Define the *parameter space* $\mathcal{E} = [0..(n-1)]^{\mathcal{T}_0}$. To each tree $G \in \mathcal{A}_n$ we associate the vector $\alpha^G \in \mathcal{E}$ such that, $\alpha^G(T)$ is the number of pendant copies of T in G. Precisely, for every rooted tree $T \in \mathcal{T}_0$,

$$\alpha^G(T) := |\{e \in E(G) : \ s(G, e) \cong T\}|,$$

9

where the symbol $\cong$ denotes isomorphisms of *rooted* trees. Then, for each tree $G \in \mathcal{A}_n$, we have that

$$\sum_{T \in \mathcal{T}_0} \alpha^G(T) \leq n - 1. \tag{5}$$

The definition of $\alpha^G(T)$ can be extended to forests as follows. If G is a forest with connected components $G_1 \dots G_i$, where G_1 is the largest component (or, if there is ambiguity, the one containing the smallest vertex among the largest ones), then $\alpha^G(T) = \alpha^{G_1}(T)$ for every rooted tree $T \in \mathcal{T}_0$.

For every $\alpha \in \mathcal{E}$, we use the notation $\mathcal{A}_{n,\alpha}$ to denote the set of graphs $G \in \mathcal{A}_n$ such that $\alpha^G = \alpha$. In other words, $\mathcal{A}_{n,\alpha}$ is the set of graphs in $\mathcal{A}_n$ that have precisely $\alpha(T)$ pendant trees that are isomorphic to T, for each $T \in \mathcal{T}_0$. This partitions $\mathcal{A}_n$ into different sets according to their α-statistics:

$$\mathcal{A}_n = \biguplus_{\alpha \in \mathcal{E}} \mathcal{A}_{n,\alpha} .$$

Note that by (5) one can restrict the previous union to vectors α satisfying $\sum_{T \in \mathcal{T}_0} \alpha(T) \leq n - 1$. For every $\Gamma \subseteq \mathcal{E}$, we also define

$$\mathcal{A}_{n,\Gamma} = \biguplus_{\alpha \in \Gamma} \mathcal{A}_{n,\alpha} .$$

For every unrooted tree $U \in \mathcal{U}$, we let $\mathcal{B}_n^U$ be the subset of $\mathcal{B}_n$ composed by graphs whose smallest component (or the one containing vertex 1 if they have the same size) is isomorphic to U. All the previous definitions for $\mathcal{A}_n$ extend naturally to the sets $\mathcal{B}_n$ and $\mathcal{B}_n^U$. That is, we can partition the set $\mathcal{B}_n$ according to $U \in \mathcal{U}$ and the α-statistics:

$$\mathcal{B}_n = \biguplus_{U \in \mathcal{U}} \mathcal{B}_n^U = \biguplus_{U \in \mathcal{U}} \biguplus_{\alpha \in \mathcal{E}} \mathcal{B}_{n,\alpha}^U .$$

4.2 The main local double counting lemma

In this subsection we construct the promised bipartite graph structure on $(\mathcal{A}_n, \mathcal{B}_n)$, and we analyse *locally* the degrees of this graph. This enables us, for each α in the parameter space $\mathcal{E}$, to compare the number of graphs G in $\mathcal{A}_n$ and $\mathcal{B}_n$ whose statistics α^G is close to α (Corollary 10).

Consider a rooted tree $T \in \mathcal{T}$ and an edge e of T. If we remove e from T, we obtain two connected components: we note $T_- \in \mathcal{T}$ the one containing the root of T, and $U_+ \in \mathcal{U}$ the other one. We let $v_- \in V(T_-)$ and $v_+ \in V(U_+)$ be the two endpoints of e. We emphasize that T_- is considered as a rooted tree (rooted at the root of T), but U_+ is considered as an unrooted tree.

Definition 1 (Multiplicities of edges and vertices in rooted or unrooted trees, see Figure 1)**.** Let T, e, T_-, U_+, v_-, v_+ be as above. Then:
- The *rooted multiplicity of the edge e in T*, denoted by $m_T(e)$ is the number of distinct edges in T that are mapped to e by some isomorphism of T (as a rooted tree).
- The *rooted multiplicity of the vertex v_- in T_-*, denoted by $m_{T_-}(v_-)$ is the number of distinct vertices in T_- that are mapped to v_- by some isomorphism of T_- (as a rooted tree, rooted at the root of T).
- The *unrooted multiplicity of v_+ in U_+*, denoted by $n_{U_+}(v_+)$ is the number of distinct vertices in U_+ that are mapped to v_+ by some isomorphism of U_+ (as an unrooted tree).
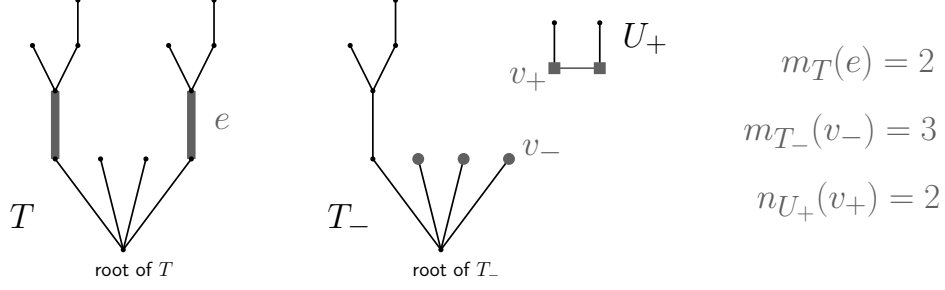
10

Figure 1: Construction of T_- and U_+ from T and e, and the relevant multiplicities.

.

In order to state our main combinatorial lemma, we first need to introduce the concept of boxes. For every $\alpha \in \mathcal{E}$ and $w \geq 1$ we define the *box* $[\alpha]^w \subset \mathcal{E}$ as the parallelepiped:

$$[\alpha]^w := \{\alpha' \in \mathcal{E} : \forall T \in \mathcal{T}_0, \ \alpha(T) \leq \alpha'(T) < \alpha(T) + w\}.$$

The parameters α and w of the box $[\alpha]^w$ will be referred to as its *lower corner* and its *width*. We also define the q-neighbourhood of $[\alpha]^w$ as the set of elements in $\mathcal{E}$ that are at distance at most q from the box. Precisely,

$$[\alpha]_q^w := \{\alpha' \in \mathcal{E} : \forall T \in \mathcal{T}_0, \ \alpha(T) - q \leq \alpha'(T) < \alpha(T) + w + q\}.$$

Note that in many cases $[\alpha]_q^w$ is itself a box, but for the structure of our argument it will be convenient to think of it as a neighbourhood of the box $[\alpha]^w$ in $\mathcal{E}$, hence our different notation.

Here we show the crucial double counting argument that will allow us to compare the sets $\mathcal{A}_n$ and $\mathcal{B}_n$.

Lemma 8 (Local double counting lemma)**.** *There exists a constant $q_* = q_*(\mathcal{U}_0) \leq u_{\max}$ (and independent from n) such that the following is true. Let $T \in \mathcal{T}_0$ and let e be an edge of T. Let T_-, U_+, v_-, v_+ be as above and assume that $T_- \in \mathcal{T}_0$ and $U_+ \in \mathcal{U}_0$.*
Then for every $\alpha \in \mathcal{E}$ and $w \geq 1$ one has:

$$m_T(e) \cdot \big(\alpha(T) + w + q_*\big) \cdot \left|\mathcal{A}_{n,[\alpha]_{q_*}^w}\right| \geq n_{U_+}(v_+) \cdot m_{T_-}(v_-) \cdot \alpha(T_-) \cdot \left|\mathcal{B}_{n,[\alpha]^w}^{U_+}\right| . \tag{6}$$

Proof. The proof of this lemma consists on a simple double counting on the edges of the bipartite graph H with vertex sets $(\mathcal{A}_{n,[\alpha]_{q_*}^w}, \mathcal{B}_{n,[\alpha]^w}^{U_+})$ and where $G_1 \in \mathcal{A}_{n,[\alpha]_{q_*}^w}$ and $G_2 \in \mathcal{B}_{n,[\alpha]^w}^{U_+}$ are joined by an edge if and only if there exists a copy of T in G_1 and an edge e in T such that $G_1 - e = G_2$, $T - e = U_+ \cup T_-$ and $s(G_1, e) = U_+$.

On the one hand, the number of edges in H is $|E(H)| = \sum_{G_2 \in \mathcal{B}_{n,[\alpha]^w}^{U_+}} \deg_H(G_2)$, where $\deg_H(G)$ denotes the degree of G in H. Since $\mathcal{G}_n$ is a bridge-addable class of graphs, we can add any bridge to G_2 and stay in the class. Moreover, if G_1 can be obtained from G_2 by adding a bridge, then G_1 belongs to the $q_* = |U_+| \leq u_{\max}$ neighbourhood of α; that is, gluing a tree of size $|U_+|$ can change the statistics of G_2 by at most $|U_+|$ in each component (for every T, $|\alpha^{G_2}(T) - \alpha^{G_1}(T)| \leq |U_+|$). Then, an easy argument shows that $\deg_H(G_2) = n_{U_+}(v_+)m_{T_-}(v_-)\alpha^{G_2}(T_-)$. Thus,

$$|E(H)| = \sum_{G_2 \in \mathcal{B}_{n,[\alpha]^w}^{U_+}} n_{U_+}(v_+)m_{T_-}(v_-)\alpha^{G_2}(T_-) \geq n_{U_+}(v_+)m_{T_-}(v_-)\alpha(T_-)\left|\mathcal{B}_{n,[\alpha]^w}^{U_+}\right| .$$

11

On the other hand, the number of edges in H is $|E(H)| = \sum_{G_1 \in \mathcal{A}_{n,[\alpha]_{q*}^w}} \deg_H(G_1)$. In this case, it suffices to upper bound the degree of G_1 into the set $\mathcal{B}_{n,[\alpha]^w}^{U_+}$. A simple argument also shows that $\deg_H(G_1) \le m_T(e)\alpha^{G_1}(T)$. Thus,

$$|E(H)| \le \sum_{G_1 \in \mathcal{A}_{n,[\alpha]_{q*}^w}} m_T(e)\alpha^{G_1}(T) \le m_T(e)(\alpha(T) + w + q_*)\left|\mathcal{A}_{n,[\alpha]_{q*}^w}\right| .$$

The lemma follows from the two previous inequalities. $\qquad\square$

Remark 1. The last lemma is also valid in the following degenerate case. Assume that T, viewed as an unrooted tree, is an element of $\mathcal{U}_0$, and let $U_+ = T$ (considered as an unrooted tree). Let conventionally $T_- := \emptyset$. Then (6) holds with the conventions $\mathrm{Aut}_r(\emptyset) = 1$, $\alpha(\emptyset) = n - |T|$, $m_T(e) = 1$. In other words, we have, with v_+ the root of T:

$$\left(\alpha(T) + w + q_*\right) \cdot \left|\mathcal{A}_{n,[\alpha]_{q*}^w}\right| \ge n_T(v_+)(n - |T|) \cdot \left|\mathcal{B}_{n,[\alpha]^w}^T\right| .$$

The proof is similar: we just consider the bipartite graph structure on $(\mathcal{A}_{n,[\alpha]_{q*}^w}, \mathcal{B}_{n,[\alpha]^w}^{U_+})$ defined by the fact that $G_1 \in \mathcal{A}_{n,[\alpha]_{q*}^w}$ and $G_2 \in \mathcal{B}_{n,[\alpha]^w}^T$ are joined by an edge if and only if there exists a copy of T in G_1 and an edge e in G_1 such that $s(G_1, e) = T$. The only thing to note is that for each $G_2 \in \mathcal{B}_{n,[\alpha]^w}^T$, there are $n_T(v_+)(n - |T|)$ ways to add a bridge to G_2 between one of the $n_T(v_+)$ allowed vertices of its connected component isomorphic to T, to any of the $(n - |T|)$ vertices of its other connected component.

Lemma 9. *With the above notation, we have:*

$$\frac{m_T(e)}{\mathrm{Aut}_r(T)} = \frac{m_{T_-}(v_-)n_{U_+}(v_+)}{\mathrm{Aut}_r(T_-)\mathrm{Aut}_u(U_+)}$$

Proof. We will prove the equality by counting the *labeled* rooted trees in $\mathcal{T}^\ell$ with a marked edge e that are isomorphic to the unlabeled rooted tree $T \in \mathcal{T}$. Recall that T and e are such that $T - e = T_- \cup U_+$. On the one hand, there are $|T|!/(\mathrm{Aut}_r(T_-)\mathrm{Aut}_u(U_+))$ different ways to label $T_- \cup U_+$. By definition of $m_{T_-}(v_-)$ and $n_{U_+}(v_+)$, for each of these labellings, there are $m_{T_-}(v_-)n_{U_+}(v_+)$ ways to select an edge e to connect T_- and U_+, such that we obtain a labelling of the tree T with marked edge e (note that all these choices are inequivalent since we work with a labeled structure). On the other hand, there are $|T|!/\mathrm{Aut}_r(T)$ different labellings of the tree T, and by definition each of them gives raise to $m_T(e)$ ways to mark the edge e, by the previous construction. Hence $m_T(e) \cdot |T|!/\mathrm{Aut}_r(T) = m_{T_-}(v_-)n_{U_+}(v_+) \cdot |T|!/(\mathrm{Aut}_r(T_-)\mathrm{Aut}_u(U_+))$, and the lemma is proved. $\quad\square$

The next corollary follows immediately:

Corollary 10. *Let $U \in \mathcal{U}_0$. Then for any T, e, T_-, U_+ as in Definition 1 such that $U_+ \cong U$, and for any $\alpha \in \mathcal{E}$ and $w \ge 1$, one has:*

$$\mathrm{Aut}_u(U) \cdot \left|\mathcal{B}_{n,[\alpha]^w}^U\right| \le \frac{\mathrm{Aut}_r(T)}{\mathrm{Aut}_r(T_-)} \cdot \frac{\alpha(T) + w + q_*}{\alpha(T_-)} \cdot \left|\mathcal{A}_{n,[\alpha]_{q*}^w}\right|,$$

where $q_* = q_*(\mathcal{U}_0)$ is the constant obtained in Lemma 8.

Observe that the last corollary also holds in degenerate case $U_+ = T$, $T_- = \emptyset$, with the notation of Remark 1.

12

4.3 Inductive bounds and tree weights

In this subsection we iterate the bound of Corollary 10 to obtain, for each $\alpha \in \mathcal{E}$ and $T \in \mathcal{T}_0$, a lower bound on $\alpha(T)$ in terms of the ratios $\left|\mathcal{B}^U_{n,[\alpha]^w}\right| / \left|\mathcal{A}_{n,[\alpha]^w_{q_*}}\right|$ for different $U \in \mathcal{U}_0$. We then use inequality (5) to conclude that a certain functional of these ratios is bounded (Corollary 12). This is the main combinatorial step towards proving a bound on the sum of these ratios, which is the quantity revelant to prove Proposition 4 (this will be done in the next section).

Let $T \in \mathcal{T}$ be a rooted tree. A $\mathcal{U}_0$-*admissible decomposition of T* is an increasing sequence $\mathbf{T} = (T_i)_{i \leq \ell}$ of labeled trees

$$T_1 \subset \cdots \subset T_\ell = T$$

for some $\ell \geq 1$ called the *length*, such that $T_1 \in \mathcal{U}_0$ and that for each $2 \leq i \leq \ell$, T_i is obtained by joining T_{i-1} by an edge e_i to some tree $U_i \in \mathcal{U}_0$.

By the choices made at the beginning of Section 4.1 for $\mathcal{U}_0$ (that contains the tree of size one) and $\mathcal{T}_0$ (that it is closed by inclusion), if $T \in \mathcal{T}_0$, then T has at least one $\mathcal{U}_0$-admissible decomposition such that $T_i \in \mathcal{T}_0$ for every $1 \leq i \leq \ell$.

Fix $\alpha \in \mathcal{E}$, $w \geq 1$ and let $q_* = q_*(\mathcal{U}_0)$ be the constant obtained from Lemma 8. Throughout this subsection we will focus on the box $[\alpha]^w$. Let $\Lambda = (\mathbb{R}_+)^{\mathcal{U}_0}$. We define the vector $\mathbf{z}_{n,\alpha} = \mathbf{z}_{n,[\alpha]^w_{q_*}} = (z^U_{n,\alpha})_{U \in \mathcal{U}_0} \in \Lambda$ by:

$$z^U_{n,\alpha} := \mathrm{Aut}_u(U) \frac{|\mathcal{B}^U_{n,[\alpha]^w}|}{|\mathcal{A}_{n,[\alpha]^w_{q_*}}|} \left(1 - \frac{|U|}{n} \right). \tag{7}$$

Observe that if $|\mathcal{B}^U_{n,[\alpha]^w}| > 0$, then, since the class of graphs is bridge-addable, we have $|\mathcal{A}_{n,[\alpha]^w_{q_*}}| > 0$. If $|\mathcal{B}^U_{n,[\alpha]^w}| = 0$, then we set $z^U_{n,[\alpha]^w} := 0$.

For any $\mathbf{z} \in \Lambda$, any $T \in \mathcal{T}$ and any $\mathcal{U}_0$-admissible decomposition of it $\mathbf{T} = (T_i)_{i \leq \ell}$, the *weight* of $\mathbf{T}$ with respect to $\mathbf{z}$ is defined as $\omega(\mathbf{T}, \mathbf{z}) = \prod_{i=1}^{\ell} z^{U_i}$, where $U_i = T_i \setminus T_{i-1}$ as an unrooted tree (here we use the convention $T_0 = \emptyset$).
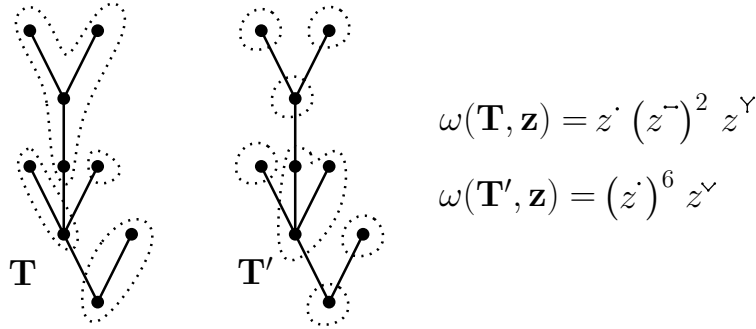

$$\omega(\mathbf{T}, \mathbf{z}) = z^{\cdot} \left(z^{\frown} \right)^2 z^{\curlyvee}$$

$$\omega(\mathbf{T}', \mathbf{z}) = \left(z^{\cdot} \right)^6 z^{\vee}$$

Figure 2: Two $\mathcal{U}_0$-admissible decompositions $\mathbf{T}$ and $\mathbf{T}'$ of the same tree, and the corresponding weights. In this case we assume that the trees $\cdot, \frown, \vee, \curlyvee$ belong to $\mathcal{U}_0$.

Lemma 11. *For any $T \in \mathcal{T}$ and any $\mathcal{U}_0$-admissible decomposition $\mathbf{T}$ of T of length ℓ, one has:*

$$\frac{\alpha(T)}{n} \geq \frac{\omega(\mathbf{T}, \mathbf{z}_{n,\alpha})}{\mathrm{Aut}_r(T)} - \frac{(w + q_*)(2|T|)^{\ell-1}}{n} .$$

Proof. Let $T_1 \subset \cdots \subset T_\ell = T$ be the $\mathcal{U}_0$-admissible decomposition $\mathbf{T}$. We will show the statement using induction on the length of $\mathbf{T}$.

If $\ell = 1$, then T is a rooted copy of U for some unrooted tree $U \in \mathcal{U}_0$. By Remark 1 we have,

$$\frac{\alpha(T)}{n} \geq \frac{\mathrm{Aut}_u(U)\mathrm{Aut}_r(\emptyset)\alpha(\emptyset)\left|\mathcal{B}^U_{n,[\alpha]^w}\right|}{\mathrm{Aut}_r(T)n\left|\mathcal{A}_{n,[\alpha]^w_{q_*}}\right|} - \frac{w+q_*}{n} = \frac{z^U_{n,\alpha}}{\mathrm{Aut}_r(T)} - \frac{w+q_*}{n} = \frac{\omega(\mathbf{T},\mathbf{z}_{n,\alpha})}{\mathrm{Aut}_r(T)} - \frac{w+q_*}{n},$$

where we used the conventions $\mathrm{Aut}_r(\emptyset) = 1$ and $\alpha(\emptyset) = n - |U|$ and the definition of $z^U_{n,\alpha}$.

Let us assume that the inequality is true for every tree $T_- \in \mathcal{T}$ and for every $\mathcal{U}_0$-admissible decomposition of T_- of size at most $\ell - 1$. Let $\mathbf{T}_-$ be the admissible decomposition induced by $\mathbf{T}$ in $T_- = T_{\ell-1}$. Then letting $U = T \setminus T_-$, we have by Lemma 9:

$$\mathrm{Aut}_r(T_-) = \frac{m_{T_-}(v_-)n_U(v_+)}{m_T(e)\mathrm{Aut}_u(U)} \cdot \mathrm{Aut}_r(T) \leq |T|\mathrm{Aut}_r(T),$$

since $n_U(v_+) \leq \mathrm{Aut}_u(U)$ and $m_{T_-}(v_-) \leq |T_-| \leq |T|$.

By using Corollary 10 and the induction hypothesis on T_- we obtain

$$\frac{\alpha(T)}{n} \geq \frac{\mathrm{Aut}_u(U)\mathrm{Aut}_r(T_-)|\mathcal{B}^U_{n,[\alpha]^w}|}{\mathrm{Aut}_r(T)n|\mathcal{A}_{n,[\alpha]^w_{q_*}}|}\alpha(T_-) - \frac{w+q_*}{n}$$

$$= \frac{z^U_{n,\alpha}\mathrm{Aut}_r(T_-)}{\mathrm{Aut}_r(T)} \cdot \frac{\alpha(T_-)}{n-|U|} - \frac{w+q_*}{n}$$

$$\geq \frac{z^U_{n,\alpha}\omega(\mathbf{T}_-,\mathbf{z}_{n,\alpha})}{\mathrm{Aut}_r(T)} - \frac{w+q_*}{n}\left(1 + \frac{\mathrm{Aut}_r(T_-)(2|T|)^{\ell-2}}{\mathrm{Aut}_r(T)}\right)$$

$$\geq \frac{\omega(\mathbf{T},\mathbf{z}_{n,\alpha})}{\mathrm{Aut}_r(T)} - \frac{(w+q_*)(2|T|)^{\ell-1}}{n}.$$

$\square$

Definition 2. For any $\mathbf{z} \in \Lambda$ and any $T \in \mathcal{T}$, we define its *maximum weight* with respect to $\mathbf{z}$, denoted by $\omega(T,\mathbf{z})$, to be the largest weight $\omega(\mathbf{T},\mathbf{z})$, where $\mathbf{T}$ is a $\mathcal{U}_0$-admissible decomposition of T. Note that $\omega(T,\mathbf{z})$ is well defined since each tree T has at least one $\mathcal{U}_0$-admissible decomposition[2].

We introduce the following weighted sum, for $\mathbf{z} \in \Lambda$:

$$Y_{\mathcal{T}_0}(\mathbf{z}) := \sum_{T \in \mathcal{T}_0} \frac{\omega(T,\mathbf{z})}{\mathrm{Aut}_r(T)}.$$

Then we immediately have from the previous lemma:

Corollary 12. *Assume that $\alpha \in \mathcal{E}$ is such that $\sum_{T \in \mathcal{T}_0} \alpha(T) \leq n - 1$. Then one has*

$$Y_{\mathcal{T}_0}(\mathbf{z}_{n,\alpha}) \leq 1 + \frac{C}{n}, \tag{8}$$

where $C = (w+q_*)(2t_{max})^{t_{max}-1}|\mathcal{T}_0|$ is a constant depending only on $\mathcal{T}_0, \mathcal{U}_0$ and w (but not on n).

Proof. This is proved by summing the upper bound of Lemma 11 for the $\mathcal{U}_0$-admissible decomposition that gives the maximum weight and over all $T \in \mathcal{T}_0$. $\square$

---

[2]Here we do a slight abuse of notation by using $\omega(\mathbf{T},\mathbf{z})$ and $\omega(T,\mathbf{z})$ to denote, respectively, the weight of a given decomposition $\mathbf{T}$ and the maximum weight of a decomposition of a given tree T.

5 Partition functions and optimization

In Section 4 we have obtained, for each $\alpha \in \mathcal{E}$, a bound on a functional of the ratios $z_{n,\alpha}^U$ for $U \in \mathcal{U}_0$ (Corollary 12). Note that this functional, namely $Y_{\mathcal{T}_0}(\mathbf{z})$, resembles a truncated version of a partition function[3] of trees weighted by their maximal weight. In this section we are going to use this fact to prove that some other functional of the $z_{n,\alpha}^U$, which we are directly interested in to prove Proposition 4, is bounded.

In this section $\mathcal{U}_0$ and $\epsilon > 0$ are *fixed*.

5.1 Partition functions of rooted and unrooted trees

In this subsection we consider an infinite version of the partition function appearing in the L.H.S. of (8), where the summation on $\mathcal{T}_0$ is replaced by a summation on the set of all rooted trees $\mathcal{T}$. We are going to show that if this partition function is finite, then the unrooted version of this partition function is at most $\frac{1}{2}$ (Lemma 14).

Recall that for any rooted tree T, and $\mathbf{z} \in \Lambda$, we defined $\omega(T, \mathbf{z})$ as the maximum weight of an $\mathcal{U}_0$-admissible decomposition of T. For any $\mathbf{z} \in \Lambda$, we let $Y(\mathbf{z}) \in \mathbb{R}_+ \cup \{\infty\}$ be defined by the following infinite sum:

$$Y(\mathbf{z}) := \sum_{T \in \mathcal{T}} \frac{\omega(T, \mathbf{z})}{\mathrm{Aut}_r(T)},$$

where the sum is taken over *all* rooted (unlabeled) trees. Note that, by double-counting, this sum is also equal to the following sum, taken on all rooted *labeled* trees:

$$Y(\mathbf{z}) = \sum_{T \in \mathcal{T}^\ell} \frac{\omega(T, \mathbf{z})}{|T|!}.$$

In words, $Y(\mathbf{z})$ is the exponential partition function of all rooted labeled trees, counted with their maximum weight. We let D be the domain of convergence of this sum:

$$D := \{\mathbf{z} \in \Lambda, Y(\mathbf{z}) < \infty\}.$$

Lemma 13. *D is closed downwards for the product order (i.e. for any $\mathbf{z}, \mathbf{z}' \in \Lambda$ such that $(z')^U \leq z^U$ for every $U \in \mathcal{U}_0$, if $\mathbf{z} \in D$ then $\mathbf{z}' \in D$) and D is bounded.*

Proof. The first assumption is straightforward, so to prove the second one it is enough to see that for each $U \in \mathcal{U}_0$ we have $Y(\mathbf{z}) = \infty$ where $\mathbf{z}$ is zero everywhere except for $z^U = |U|!$. We can construct a labeled tree of size $n|U|$ by attaching successively n copies of U by edges. The number of distinct ways to do that is equal to the number of rooted labeled trees of size n, which is n^{n-1}, times the number of ways to distribute the labels in the different copies of U, which is at least $\frac{1}{n!}\binom{n|U|}{|U|,\ldots,|U|} = \frac{(n|U|)!}{n!(|U|!)^n}$. Of course we do not obtain all trees of size $n|U|$ with this construction, but this is enough to obtain the lower bound:

$$Y(\mathbf{z}) \geq \sum_{n \geq 1} \frac{\frac{1}{n!}\binom{n|U|}{|U|,\ldots,|U|} n^{n-1}}{(n|U|)!} (z^U)^n = \sum_{n \geq 1} \frac{n^{n-1}}{n!(|U|!)^n} (z^U)^n = \sum_{n \geq 1} \frac{n^{n-1}}{n!} \ .$$

---

[3] Here we prefer to use the terminology *partition function* rather than *generating function*, since the second terminology usually refers to *formal* power series. Note that here, since the decomposition of maximal weight underlying the definition of the numbers $\omega(T, \mathbf{z})$ depends on $\mathbf{z}$, the quantity $\omega(T, \mathbf{z})$ is only *piecewise* polynomial in the z^U, $U \in \mathcal{U}_0$, so that $Y_{\mathcal{T}_0}(\mathbf{z})$ is *not* a formal power series in $\mathbf{z}$.

The last sum is divergent, which concludes the proof of the claim. □

We note that $\omega(T, \mathbf{z})$ does not depend on the root of T, so this quantity is well defined for *unrooted* trees $U \in \mathcal{U}$. We can thus introduce the "unrooted version" of the partition function Y:

$$Y^u(\mathbf{z}) := \sum_{U \in \mathcal{U}} \frac{\omega(U, \mathbf{z})}{\mathrm{Aut}_u(U)} = \sum_{U \in \mathcal{U}^\ell} \frac{\omega(U, \mathbf{z})}{|U|!}.$$

Note that $Y^u(\mathbf{z})$ is also given by the following expressions:

$$Y^u(\mathbf{z}) = \sum_{T \in \mathcal{T}^\ell} \frac{\omega(T, \mathbf{z})}{|T| \cdot |T|!} = \sum_{T \in \mathcal{T}} \frac{\omega(T, \mathbf{z})}{|T| \cdot \mathrm{Aut}_r(T)}.$$

It is clear that $Y^u(\mathbf{z}) \leq Y(\mathbf{z})$ and in particular $Y^u(\mathbf{z}) < \infty$ if $\mathbf{z} \in D$.

The following statement, which is a variant in our context of the celebrated *dissymmetry theorem* (see [BLL98]), is where the constant $\frac{1}{2}$ from our main theorem (Theorem 2) appears:

Lemma 14 (Supermultiplicative dissymmetry theorem)**.** *If $\mathbf{z} \in D$, then the rooted series and unrooted series are related by the following inequality:*

$$Y(\mathbf{z}) - Y^u(\mathbf{z}) \geq \frac{1}{2} Y(\mathbf{z})^2. \tag{9}$$

In particular for all $\mathbf{z} \in D$ one has

$$Y^u(\mathbf{z}) \leq \frac{1}{2}. \tag{10}$$

Proof. Let $U \in \mathcal{U}^\ell$ be a labeled unrooted tree. Then the number $e(U) = |U| - 1$ of edges of U and the number $v(U) = |U|$ of vertices of U are related by the equation:

$$v(U) - 1 = e(U) .$$

By multiplying this equality by $\omega(U, \mathbf{z})/|U|!$ and summing over all unrooted labeled trees U, it follows that the quantity $Y^e(\mathbf{z}) := Y(\mathbf{z}) - Y^u(\mathbf{z})$ can be interpreted as the exponential partition function of all labeled trees *with one marked edge*, counted with their maximum weight. Now let U be a labeled tree with a marked edge e. Removing e splits U into two connected components $T_1, T_2 \in \mathcal{T}^\ell$ that are naturally rooted at a vertex, and by definition of the maximum weight we have the supermultiplicativity property:

$$\omega(U, \mathbf{z}) \geq \omega(T_1, \mathbf{z}) \cdot \omega(T_2, \mathbf{z}).$$

Indeed, the right hand side is the weight of the $\mathcal{U}_0$-admissible decomposition of U induced by the decomposition with maximum weight of each of its components, and the weight of this decomposition is a lower bound on the maximum weight. Conversely, given any two rooted labeled trees T_1 and T_2 whose sizes add up to $|U|$, there are $\frac{|U|!}{|T_1|!|T_2|!}$ ways to distribute the labels in $[1..|U|]$ between them to build a labeled tree U of this form, and each tree U with a marked edge is obtained in exactly two ways by this construction. Since all sums are absolutely convergent, we thus get that

$$Y^e(\mathbf{z}) \geq \frac{1}{2} Y(\mathbf{z})^2,$$

which gives (9).

The bound (10) follows since by definition of D, $Y(\mathbf{z})$ is a well defined real number, and since for all $y \in \mathbb{R}$ one has $y - \frac{1}{2}y^2 \leq \frac{1}{2}$. □

16

Remark 2. The partition function $T(x)$ of all rooted trees, which is solution of the equation $T(x) = x \exp(T(x))$, has radius of convergence e^{-1}, and its value at the dominant singularity is $T(e^{-1}) = 1$. Moreover, it is classical that the generating function of *unrooted* trees is given by $T^u(x) = T(x) - \frac{1}{2}T(x)^2$ (see for example [BLL98]). It follows that, at the dominant singularity, one has $T^u(e^{-1}) = \frac{1}{2}$. Note that this implies (9) in the case where $\mathcal{U}_0$ is a singleton, and that this also shows that (9) is tight. We also note that, using classical singularity analysis [FS09], this enables one to reprove the result of Rényi [Rén59] that says that a random forest of size n is connected with probability $e^{-1/2} + o(1)$ when n tends to infinity.

The last partition functions we define are the functionals $Y_{\mathcal{U}_0}^u(\mathbf{z})$ and $\widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z})$, defined by:

$$Y_{\mathcal{U}_0}^u(\mathbf{z}) := \sum_{U \in \mathcal{U}_0} \frac{\omega(U, \mathbf{z})}{\mathrm{Aut}_u(U)} \quad , \quad \widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z}) := \sum_{U \in \mathcal{U}_0} \frac{z^U}{\mathrm{Aut}_u(U)}.$$

Note that the sums are taken over all the elements of $\mathcal{U}_0$, that are considered as *unrooted* trees.

5.2 Optimization

In the last subsection we have shown (Lemma 14) that the fact that $Y(\mathbf{z}) < \infty$ implies that $Y^u(\mathbf{z}) \le \frac{1}{2}$. The goal of this subsection, achieved in the next proposition, is to transfer this result to truncated analogues of these partition functions.

For all $k \ge 1$, define the following truncated version of $Y(\mathbf{z})$:

$$Y_{\le k}(\mathbf{z}) := \sum_{\substack{T \in \mathcal{T} \\ |T| \le k}} \frac{\omega(T, \mathbf{z})}{\mathrm{Aut}_r(T)}.$$

Note that $Y_{\le k}(\mathbf{z})$ is defined by a finite sum, hence it is a well defined real number for *all* $\mathbf{z} \in \Lambda$. We also define $Y_{=k}(\mathbf{z})$ to be the contribution of trees of size exactly k to $Y(\mathbf{z})$:

$$Y_{=k}(\mathbf{z}) := Y_{\le k}(\mathbf{z}) - Y_{\le k-1}(\mathbf{z}).$$

Proposition 15. *There exists a k_*, depending only on ϵ and $\mathcal{U}_0$, such that for every $\mathbf{z} \in \Lambda$ satisfying $Y_{\le k_*}(\mathbf{z}) \le 1.5$, we have*

$$\widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z}) \le \frac{1}{2}(1 + \epsilon).$$

Remark 3. If $k \ge k_*$ then $Y_{\le k_*}(\mathbf{z}) \le Y_{\le k}(\mathbf{z})$. Therefore if necessary the integer k_* can be replaced by any larger value without changing the conclusion of the proposition.

Remark 4. The constant 1.5 in the above proposition could be replaced by any constant larger than 1 (as the proof will show). To keep the notation light we preferred to fix some arbitrary value that is good enough for our proof.

Note that for any $\mathbf{z} \in \Lambda$, if we define $\mathbf{z}_* \in \Lambda$ by the fact that for all $U \in \mathcal{U}_0$ we have $\mathbf{z}_*^U = \omega(U, \mathbf{z})$, then $\omega(T, \mathbf{z}) = \omega(T, \mathbf{z}_*)$ for any tree $T \in \mathcal{T}$ (this is easily seen by considering maximum weight decompositions). Since it is always true that $z^U \le \omega(U, \mathbf{z})$, if follows that replacing $\mathbf{z}$ by $\mathbf{z}_*$ does not change the value of $Y_{\le k}(\mathbf{z})$, while only making $\widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z})$ larger or equal. Therefore

$$\max\{\widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z}) : Y_{\le k}(\mathbf{z}) \le 1.5\} = \max\{\widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z}) : Y_{\le k}(\mathbf{z}) \le 1.5 \text{ and } \forall U \in \mathcal{U}_0, z^U = \omega(U, \mathbf{z})\}$$

$$= \max\{Y_{\mathcal{U}_0}^u(\mathbf{z}) : Y_{\le k}(\mathbf{z}) \le 1.5 \text{ and } \forall U \in \mathcal{U}_0, z^U = \omega(U, \mathbf{z})\}. \quad (11)$$

Now, let us fix a sequence $(\mathbf{z}_k)_{k \geq u_{max}}$, such that for all $k \geq u_{max}$ we have:

$$\mathbf{z}_k \in \{\mathbf{z} : Y_{\leq k}(\mathbf{z}) \leq 1.5 \text{ and } \forall U \in \mathcal{U}_0, z^U = \omega(U, \mathbf{z})\}.$$

Note that the set from where $\mathbf{z}_k$ is selected, is bounded: for $k \geq u_{\max}$, we have $Y_{\leq k}(\mathbf{z}) \geq \omega(T, \mathbf{z})/\mathrm{Aut}_r(T)$ for each rooted tree T that, as an unrooted tree, belongs to $\mathcal{U}_0$. This directly implies that the sequence $\mathbf{z}_k$ is uniformly bounded. We can thus extract an increasing sequence k_i such that the corresponding subsequence converges, and we note:

$$\mathbf{z}_\infty := \lim_{i \to \infty} \mathbf{z}_{k_i}.$$

Our first step in the proof of the proposition is the following lemma.

Lemma 16. *The point $\mathbf{z}_\infty$ belongs to $\bar{D}$ (the closure of D).*

Proof. We go by contradiction. Suppose $\mathbf{z}_\infty$ does not belong to $\bar{D}$. Then there exists $\delta > 0$ such that $\frac{1}{1+\delta} \circ \mathbf{z}_\infty \notin D$, where we use the notation $\circ$ for the *scaled multiplication* of a vector $\mathbf{z} \in \Lambda$ by a scalar $\lambda \in \mathbb{R}$:

$$(\lambda \circ \mathbf{z})^U = \lambda^{|U|} z^U, \quad U \in \mathcal{U}_0.$$

We thus have $Y(\frac{1}{1+\delta} \circ \mathbf{z}_\infty) = \infty$, *i.e.* $\sum_{\ell \geq 1} Y_{=\ell}(\mathbf{z}_\infty)(1+\delta)^{-\ell} = \infty$. This shows that there exists an infinite sequence $(\ell_j)_{j \geq 1}$ tending to infinity such that for all $j \geq 1$ one has:

$$Y_{=\ell_j}(\mathbf{z}_\infty) \geq (1 + \delta/2)^{\ell_j}.$$

Now we claim that there exists some i_0 such that for $i \geq i_0$ one has, for every rooted tree $T \in \mathcal{T}$:

$$\omega(T, \mathbf{z}_{k_i}) \geq \omega(T, \mathbf{z}_\infty) \left(1 - \frac{\delta}{4}\right)^{|T|}. \tag{12}$$

If we admit this claim, we can conclude the proof as follows. We have, for $i \geq i_0$ and $j \geq 1$:

$$\begin{aligned} Y_{=\ell_j}(\mathbf{z}_{k_i}) &\geq Y_{=\ell_j}(\mathbf{z}_\infty)(1 - \delta/4)^{\ell_j} \\ &\geq ((1 - \delta/4)(1 + \delta/2))^{\ell_j}. \end{aligned}$$

But $(1 - \delta/4)(1 + \delta/2)$ is larger than 1 provided we took δ small enough (and we can do that), so there exists some j such that $\left((1 - \delta/4)(1 + \delta/2)\right)^{\ell_j} > 1.6$, which in turns implies that for $i \geq i_0$, one has $Y_{=\ell_j}(\mathbf{z}_{k_i}) > 1.6$. Now we can choose i large enough so that $k_i \geq \ell_j$, and we get that:

$$Y_{\leq k_i}(\mathbf{z}_{k_i}) \geq Y_{=\ell_j}(\mathbf{z}_{k_i}) > 1.6,$$

which is a contradiction.

So it just remains to prove the claim in (12). We let $I \subset \mathcal{U}_0$ denote the set indexing coordinates of $\mathbf{z}_\infty$ that are equal to zero, *i.e.*:

$$I := \{U \in \mathcal{U}_0, (z_\infty)^U = 0\}.$$

Since $\mathbf{z}_{k_i}$ converges to $\mathbf{z}_\infty$, and since $(\mathbf{z}_\infty)^U \neq 0$ for $U \in \mathcal{U}_0 \setminus I$, each of the ratios $\frac{(z_{k_i})^U}{(z_\infty)^U}$ converge to 1 when i tends to infinity, for $U \in \mathcal{U}_0 \setminus I$. Therefore there exists i_0 such that for $i \geq i_0$, we have for all $U \in \mathcal{U}_0 \setminus I$:

$$\left| \left(\frac{(z_{k_i})^U}{(z_\infty)^U} \right)^{1/|U|} - 1 \right| < \delta/4.$$

18

We can now prove the claim (12). First, if $\omega(T, \mathbf{z}_\infty) = 0$ then the claim is obviously true. Otherwise, consider an $\mathcal{U}_0$-admissible decomposition $\mathbf{T}$ of T that gives rise to the maximum weight $\omega(T, \mathbf{z}_\infty)$. Since $\omega(T, \mathbf{z}_\infty) \neq 0$, the decomposition only uses unrooted trees in $\mathcal{U}_0 \setminus I$. We thus have:

$$\frac{\omega(\mathbf{T}, \mathbf{z}_{k_i})}{\omega(\mathbf{T}, \mathbf{z}_\infty)} = \prod_{U \in \mathcal{U}_0 \setminus I} \left(\frac{(z_{k_i})^U}{(z_\infty)^U} \right)^{\nu(U)}, \tag{13}$$

where $\nu(U)$ is the number of times U is used in the decomposition $\mathbf{T}$. Since $\sum_{U \in \mathcal{U}_0} |U| \nu(U) = |T|$, the ratio (13) is larger than $(1 - \delta/4)^{|T|}$, and the claim follows since $\omega(T, \mathbf{z}_{k_i}) \geq \omega(\mathbf{T}, \mathbf{z}_{k_i})$. $\qquad\square$

We can now prove the proposition:

Proof of Proposition 15. For every $k \geq u_{\max}$, we now choose

$$\mathbf{z}_k \in \underset{\mathbf{z} \in \Lambda}{argmax} \{ Y^u_{\mathcal{U}_0}(\mathbf{z}) : Y_{\leq k}(\mathbf{z}) \leq 1.5 \text{ and } \forall U \in \mathcal{U}_0, \ z^U = \omega(U, \mathbf{z}) \},$$

and we select $\mathbf{z}_\infty$ as before.

From Lemma 16 and the fact that D is closed downwards for the product order, for all $\delta > 0$ we have $\frac{1}{1+\delta} \circ \mathbf{z}_\infty \in D$, where we recycle the notation $\circ$ for the scaled product from the previous proof. From Lemma 14, it follows that $Y^u(\frac{1}{1+\delta} \circ \mathbf{z}_\infty) \leq \frac{1}{2}$. Now, recall that we restricted to $\mathbf{z} \in \Lambda$ such that $z^U = \omega(\mathbf{z}, U)$ for every $U \in \mathcal{U}_0$. This implies that

$$Y^u \left(\frac{1}{1+\delta} \circ \mathbf{z}_\infty \right) \geq Y^u_{\mathcal{U}_0} \left(\frac{1}{1+\delta} \circ \mathbf{z}_\infty \right).$$

Moreover, $Y^u_{\mathcal{U}_0} \left(\frac{1}{1+\delta} \circ \mathbf{z}_\infty \right) \geq (1+\delta)^{-u_{\max}} Y^u_{\mathcal{U}_0}(\mathbf{z}_\infty)$, so $Y^u_{\mathcal{U}_0}(\mathbf{z}_\infty) \leq (1+\delta)^{u_{\max}} \cdot \frac{1}{2}$. Since this is true for any $\delta > 0$, we obtain that $Y^u_{\mathcal{U}_0}(\mathbf{z}_\infty) \leq \frac{1}{2}$. Since $\mathbf{z}_{k_i}$ converges to $\mathbf{z}_\infty$ and $Y^u_{\mathcal{U}_0}(\mathbf{z})$ is clearly continuous, we can choose i large enough and $k_* = k_i$ so that $Y^u_{\mathcal{U}_0}(\mathbf{z}_{k_*}) \leq \frac{1}{2}(1 + \epsilon)$. By the choice of $\mathbf{z}_k$ and by (11), it follows that $\max\{ \widetilde{Y}^u_{\mathcal{U}_0}(\mathbf{z}) : Y_{\leq k_*}(\mathbf{z}) \leq 1.5 \} \leq \frac{1}{2}(1 + \epsilon)$, which implies the proposition. $\qquad\square$

6 Finishing the proof

In this section we conclude the proof of Proposition 4 (hence of the main theorem). The idea of the proof is to combine the main results of Section 4 (Corollary 12) and of Section 5 (Proposition 15) and to apply them to a well chosen set of boxes.

6.1 Boxing lemma

The results of the previous section give us bounds on the variables $\mathbf{z}_{n,\alpha}$ defined by (7), which gives us some control on the ratio of the sizes of the sets $\mathcal{A}_{n,[\alpha]^w_{q_*}}$ and $\mathcal{B}^U_{n,[\alpha]^w}$, where $[\alpha]^w$ is some box inside the parameter space $\mathcal{E}$. In order to use this information in the next subsection, we first show that there exists a partition of the parameter space $\mathcal{E}$ into disjoint boxes $[\beta_i]^w$ such that they are $2q_*$-apart and they capture most of the graphs in $\mathcal{B}^U_n$ for each $U \in \mathcal{U}_0$.

Lemma 17 (Box partitioning lemma). *For all $\epsilon > 0$, $\mathcal{U}_0$ and $\mathcal{T}_0$, there exist q_*, w and n_0 such that for all $n \geq n_0$ the following is true: There exist K and a family of boxes $[\beta_i]^w \subset \mathcal{E}$ of size K such that*

(P₁) *The q_*-neighbourhoods of boxes form a partition of $\mathcal{E}$; i.e.*

$$\mathcal{E} = \biguplus_{i=1}^{K} [\beta_i]_{q_*}^w,$$

(P₂) *Boxes capture a large fraction of each set $\mathcal{B}_n^U$; i.e. for each $U \in \mathcal{U}_0$, we have:*

$$\sum_{i=1}^{K} \sum_{\beta \in [\beta_i]^w} |\mathcal{B}_{n,\beta}^U| \geq (1 - \epsilon)|\mathcal{B}_n^U|.$$

Proof. Let $d = |\mathcal{T}_0|$ and let q_* be chosen as in Lemma 8. Choose $w \geq 0$ and n_0 such that for every $n \geq n_0$

$$1 - \left(1 - \frac{w + 2q_*}{n}\right)^d \left(1 + \frac{2q_*}{w}\right)^{-d} \leq \epsilon|\mathcal{U}_0|^{-1} .$$

Consider the following set

$$\Gamma_0 = \{\beta \in \mathcal{E} : \exists T \in \mathcal{T}_0, \exists j \geq 1, j(w + 2q_*) - 2q_* \leq \beta(T) < j(w + 2q_*)\} .$$

The set Γ_0 can be seen as a subset of hyperplanes of width $2q_*$ equally spaced at distance w in each direction. Observe that $\Gamma_0^c = \mathcal{E} \setminus \Gamma_0$ contains a set of $K = \left(\frac{n}{w+2q_*} - 1\right)^d$ boxes of width w that are $(2q_*)$-apart. However, these boxes might not contain a negligible part of the graphs in $\mathcal{B}_n$.

The size of the Γ_0^c satisfies,

$$|\Gamma_0^c| \geq w^d \left(\frac{n}{w + 2q_*} - 1\right)^d \geq \left(1 - \frac{w + 2q_*}{n}\right)^d \left(1 + \frac{2q_*}{w}\right)^{-d} n^d .$$

For the remaining of the proof, let us consider $\Gamma_0 \subseteq \mathbb{Z}^d/\mathcal{E}$. Choose β uniformly at random from $\mathbb{Z}^d/\mathcal{E}$. We write $\Gamma_0 + \beta = \{\gamma + \beta : \gamma \in \Gamma_0\}$ to denote the translation of the set Γ_0 by β. Recall that $|\mathcal{E}| = n^d$. Then

$$|\Gamma_0 + \beta| = |\Gamma_0| \leq \left(1 - \left(1 - \frac{w + 2q_*}{n}\right)^d \left(1 + \frac{2q_*}{w}\right)^{-d}\right) n^d \leq \epsilon|\mathcal{U}_0|^{-1} n^d . \tag{14}$$

We now define the following measure μ over the set $\mathcal{E}$. For each $U \in \mathcal{U}_0$ and $\Gamma \subseteq \mathcal{E}$, let $b_\Gamma^U = \frac{|\mathcal{B}_{n,\Gamma}^U|}{|\mathcal{U}_0||\mathcal{B}_n^U|}$. For every $\Gamma \subseteq \mathcal{E}$, we define

$$\mu(\Gamma) = \sum_{U \in \mathcal{U}_0} b_\Gamma^U .$$

Observe that if $\mu(\Gamma) \leq \epsilon|\mathcal{U}_0|^{-1}$, then, for every $U \in \mathcal{U}_0$

$$|\mathcal{B}_{n,\Gamma}^U| = b_\Gamma^U |\mathcal{U}_0||\mathcal{B}_n^U| \leq \mu(\Gamma)|\mathcal{U}_0||\mathcal{B}_n^U| \leq \epsilon|\mathcal{B}_n^U| .$$

If $\beta \in \mathcal{E}$ is chosen uniformly at random and using (14), we have

$$\mathbb{E}_\beta[\mu(\Gamma_0 + \beta)] = \sum_{\gamma \in \mathcal{E}} \mu(\gamma) \Pr(\gamma \in \Gamma_0 + \beta) = \sum_{\gamma \in \mathcal{E}} \sum_{U \in \mathcal{U}_0} b_\gamma^U \cdot \frac{|\Gamma_0|}{n^d} \leq \epsilon|\mathcal{U}_0|^{-1} ,$$

where we used that $\sum_{\gamma \in \mathcal{E}} \sum_{U \in \mathcal{U}_0} b_\gamma^U = 1$.

Thus, there exists a $\beta_0 \in \mathcal{E}$ such that $\mu(\Gamma_0 + \beta_0) \leq \epsilon |\mathcal{U}_0|^{-1}$. Then, the set $\Gamma_1 = \mathcal{E} \setminus (\Gamma_0 + \beta_0)$ is a set of $(2q_*)$-apart boxes $[\beta_i]^w$ for $i \in [1..K]$ that satisfies

$$\sum_{i=1}^K \sum_{\beta \in [\beta_i]^w} |\mathcal{B}_{n,\beta}^U| = |\mathcal{B}_{n,\Gamma_1}^U| = |\mathcal{B}_n^U| - |\mathcal{B}_{n,(\Gamma_0+\beta_0)}^U| \geq (1-\epsilon)|\mathcal{B}_n^U|\,,$$

for every $U \in \mathcal{U}_0$. $\qquad\square$

6.2   Proof of Proposition 4

Let $\epsilon' > 0$ be fixed, and let $\epsilon := \epsilon'/3$. Let us choose $\mathcal{U}_0$ to be the family of all unrooted unlabeled trees of size at most $u_{\max}$, where u_{max} is chosen such that $(u_{\max})^{-1} < \epsilon$. Now that $\mathcal{U}_0$ is fixed, we can apply Proposition 15, with our current value of ϵ, and we let k_* be the value given by this proposition. The integer k_* depends on ϵ (and also on $\mathcal{U}_0$, that itself depends on ϵ). We now let $\mathcal{T}_0$ be the set of all rooted trees of size at most k_*.

Now that ϵ, $\mathcal{U}_0$, and $\mathcal{T}_0$ are fixed, so is the constant q_* given by Lemma 8. We can then apply Lemma 17 to get some constants w and n_0, such that for every $n \geq n_0$ there exists a family of K boxes $[\beta_i]$ satisfying (P_1) and (P_2). All these constants depend on ϵ (and also on $\mathcal{U}_0$ and $\mathcal{T}_0$, that both also depend on ϵ).

We can now choose $n_1 \geq n_0$ large enough, so that

$$(w + q_*)(2k_*)^{k_*-1}|\mathcal{T}_0| < 0.4n_1. \tag{15}$$

Note that the left-hand side is the quantity C that appears in (8), with $t_{\max} = k_*$. For a further use, we will also assume that $n_1 \geq u_{\max}/\epsilon$.

For $n \geq n_1$, let $[\alpha]^w$ (where $\alpha = \beta_i$, for some $1 \leq i \leq K$) be one of the boxes given by the box partitioning lemma (Lemma 17). Recall the definition of $\mathbf{z}_{n,\alpha} = (z_{n,\alpha}^U)_{U \in \mathcal{U}_0}$ given in (7):

$$z_{n,\alpha}^U := \mathrm{Aut}_u(U) \frac{|\mathcal{B}_{n,[\alpha]^w}^U|}{|\mathcal{A}_{n,[\alpha]_{q_*}^w}|} \left(1 - \frac{|U|}{n}\right).$$

Then we have by Corollary 12 and the bound (15) that:

$$Y_{\leq k_*}(\mathbf{z}_{n,\alpha}) = \sum_{T \in \mathcal{T}_0} \frac{\omega(T, \mathbf{z}_{n,\alpha})}{\mathrm{Aut}_r(T)} \leq 1.4\,.$$

From Proposition 15, this implies that:

$$\sum_{U \in \mathcal{U}_0} \frac{|\mathcal{B}_{n,[\alpha]^w}^U|}{|\mathcal{A}_{n,[\alpha]_{q_*}^w}|} \left(1 - \frac{|U|}{n}\right) = \sum_{U \in \mathcal{U}_0} \frac{z_{n,\alpha}^U}{\mathrm{Aut}_u(U)} = \widetilde{Y}_{\mathcal{U}_0}^u(\mathbf{z}_{n,\alpha}) \leq \frac{1}{2}(1+\epsilon)\,.$$

Since $|U|/n \leq u_{\max}/n < \epsilon$ for all $U \in \mathcal{U}_0$, we deduce that we have:

$$\sum_{U \in \mathcal{U}_0} \left|\mathcal{B}_{n,[\alpha]^w}^U\right| \leq \frac{1}{2} \left|\mathcal{A}_{n,[\alpha]_{q_*}^w}\right| (1+\epsilon)(1-\epsilon)^{-1}. \tag{16}$$

21

We now sum the last inequality over all the boxes $[\beta_i]^w$ provided by Lemma 17. We obtain:

$$\sum_{U \in \mathcal{U}_0} (1 - \epsilon)|\mathcal{B}_n^U| \leq \sum_{U \in \mathcal{U}_0} \sum_{i=1}^{K} |\mathcal{B}_{n,[\beta_i]^w}^U| \leq \frac{1}{2}(1+\epsilon)(1-\epsilon)^{-1} \sum_{i=1}^{K} |\mathcal{A}_{n,[\beta_i]_{q_*}^w}| \leq \frac{1}{2}|\mathcal{A}_n|(1+\epsilon)(1-\epsilon)^{-1}.$$

Here: the central inequality is the summation of the previous bound; the leftmost inequality comes from Property (P_2) of Lemma 17 (boxes capture most of the mass of the sets $\mathcal{B}_n^U$); the rightmost inequality comes from Property (P_1) of Lemma 17 (boxes are $(2q_*)$-apart, so the sets $[\beta_i]_{q_*}^w$ are disjoint). We have just proved that for $n \geq n_1$,

$$\sum_{U \in \mathcal{U}_0} |\mathcal{B}_n^U| \leq \frac{1}{2}(1+\epsilon)(1-\epsilon)^{-2}|\mathcal{A}_n|.$$

Again, using the same simple double counting argument as we used in the proof of Lemma 6, we have:

$$u_{\max}(n - u_{\max}) \sum_{U \in \mathcal{U} \setminus \mathcal{U}_0} |\mathcal{B}_n^U| \leq n|\mathcal{A}_n|.$$

By assumption, $(u_{\max})^{-1} < \epsilon$, so $\frac{n}{u_{\max}(n-u_{\max})} = \frac{1}{u_{\max}(1-u_{\max}/n)} \leq 2\epsilon$ if n is large enough. Therefore, for n large enough, we have:

$$\sum_{U \in \mathcal{U} \setminus \mathcal{U}_0} |\mathcal{B}_n^U| \leq 2\epsilon|\mathcal{A}_n|. \tag{17}$$

Putting all bounds together, we obtain:

$$|\mathcal{B}_n| = \sum_{U \in \mathcal{U}} |\mathcal{B}_n^U| = \sum_{U \in \mathcal{U}_0} |\mathcal{B}_n^U| + \sum_{U \in \mathcal{U} \setminus \mathcal{U}_0} |\mathcal{B}_n^U| \leq \left((1+\epsilon)(1-\epsilon)^{-2} + 2\epsilon\right) \frac{1}{2}|\mathcal{A}_n|.$$

Now, when ϵ' is small enough and $\epsilon := \epsilon'/3$, we have $(1+\epsilon)(1-\epsilon)^{-2} + 2\epsilon \leq (1 + 2\epsilon')$. We thus get:

$$|\mathcal{B}_n| \leq \left(\frac{1}{2} + \epsilon'\right)|\mathcal{A}_n|,$$

which concludes the proof of Proposition 4.

Acknowledgements

References

[ABMR12] L. Addario-Berry, C. McDiarmid, and B. Reed. Connectivity for bridge-addable monotone graph classes. *Combin. Probab. Comput.*, 21(6):803–815, 2012.

[BBG08] P. Balister, B. Bollobás, and S. Gerke. Connectivity of addable graph classes. *J. Combin. Theory Ser. B*, 98(3):577–584, 2008.

[BGKN07] M. Bodirsky, O. Giménez, M. Kang, and M. Noy. Enumeration and limit laws for series-parallel graphs. *European J. Combin.*, 28(8):2091–2105, 2007.

[BLL98] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures*, volume 67 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1998. Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.

[BS01] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[FS09] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[GN09] O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs. *J. Amer. Math. Soc.*, 22(2):309–329, 2009.

[KP13] M. Kang and K. Panagiotou. On the connectivity of random graphs from addable classes. *J. Combin. Theory Ser. B*, 103(2):306–312, 2013.

[McD12] C. McDiarmid. Connectivity for random graphs from a weighted bridge-addable class. *Electron. J. Combin.*, 19(4):Paper 53, 20, 2012.

[MS] C. McDiarmid and A. Scott. Random graphs from a block class. *European Journal of Combinatorics, to appear.*

[MSW05] C. McDiarmid, A. Steger, and D.J. A. Welsh. Random planar graphs. *J. Combin. Theory Ser. B*, 93(2):187–205, 2005.

[MSW06] C. McDiarmid, A. Steger, and D.J. A. Welsh. Random graphs from planar and other addable classes. In *Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, pages 231–246. Springer, Berlin, 2006.

[Nor] S. Norin. Connectivity of addable classes of forests. available at the author's webpage.

[Rén59] A. Rényi. Some remarks on the theory of trees. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 4:73–85, 1959.