

A STUDY OF LARGE FRINGE AND NON-FRINGE
SUBTREES IN CONDITIONAL GALTON-WATSON
TREES

XING SHI CAI

April 2016

School of Computer Science
McGill University, Montreal



A thesis submitted to McGill University in partial fulfillment
of the requirements of the degree of Doctor of Philosophy

©Xing Shi Cai 2016

Dedicated to my dear grandpa.

... we all know that behind things that are easily seen there may be years of thinking and/or huge piles of scrap notes that lead nowhere, and one sheet where everything finally worked out nicely.

— Allan Gut [55]

ACKNOWLEDGMENTS

I would like to thank my supervisor Luc Devroye, whom I always look up to but can never be.

I also want to thank my dear friends whom I met at McGill, in particular Laura Eslava, Claude Gravel, Hamed Hatami, Guillem Perarnau, Henning Sulzbach, Liana Yepremyan and Yelena Yuditsky, whose kindness and brilliance I will never forget.

Special thanks go to Cecilia Holmgren for her careful review and helpful comments of this thesis.

ABSTRACT

Random trees are ubiquitous in computer science. One particularly attractive model is the random tree chosen uniformly from a collection of trees. Many of these models are equivalent to the Galton-Watson tree conditional on its size — these trees, in turn, go back to the model proposed by Bienaymé [14], Watson and Galton [94] for the evolution of populations. This thesis deals with some questions about random trees that remained unanswered until now.

In particular, we study the conditions for families of subtrees to exist with high probability (whp) in a Galton-Watson tree of size n . We first give a Poisson approximation of fringe subtree counts, which yields the height of the maximal complete r -ary fringe subtree. Then we determine the maximal K_n such that every tree of size at most K_n appears as a fringe subtree whp. Finally, we study non-fringe subtree counts and determine the height of the maximal complete r -ary non-fringe subtree.

ABRÉGÉ

Les arbres aléatoires sont omniprésents dans la science informatique. Un modèle particulièrement attractif est l'arbre choisi uniformément de manière aléatoire parmi une collection d'arbres. Beaucoup de ces modèles sont équivalents à l'arbre de Galton-Watson conditionnelle sur sa taille — ces arbres, en retour, remontent au modèle proposé par Bienaymé [14], Watson and Galton [94] pour l'évolution des populations. Cette thèse traite des questions sur les arbres aléatoires qui sont restées sans réponse jusqu'à présent.

En particulier, nous étudions les conditions d'existence des familles de sous-arbres avec une forte probabilité dans un arbre de Galton-Watson de la taille n . Nous donnons d'abord une approximation de Poisson du nombre de sous-arbres franges, ce qui donne la hauteur du sous-arbre frange de valence- r complet maximal. Ensuite, nous déterminons le K_n maximal de telle sorte que chaque arbre de taille au plus K_n apparaît comme un sous-arbre frange. Enfin, nous étudions le nombre de sous-arbres non-franges et déterminons la hauteur du sous-arbre non-frange de valence- r complet maximal.

PREFACE

This thesis is a much reworked version of my previous paper of the same title [20], of which my supervisor Luc Devroye is a coauthor. I wrote most of the text, but the mathematical ideas were developed jointly. It also includes a short discussion (Chapter 7) of another paper by us, “The graph structure of a deterministic automaton chosen at random” [21].

During my PhD years, I have also collaborated with Guillem Perarnau Llobet and Bruce Reed in proving an asymptotic version of the acyclic edge colouring conjecture for graphs with large girth [22]. The main idea came from Llobet, but I have made substantial contributions.

When I began my PhD study, I wrote two papers [18, 19] on the Kademia network, which is the most popular searching algorithm for peer-to-peer networks. These are extensions of my Master’s thesis [17].

Here is a full list of my papers:

- X. S. Cai and L. Devroye, “A study of large fringe and non-fringe subtrees in conditional Galton-Watson trees,” *ArXiv e-prints*, 2016. arXiv: [1602.03850](#) [math.PR].
- X. S. Cai and L. Devroye, “The graph structure of a deterministic automaton chosen at random,” *Random Structures and Algorithms*, 2016, To appear. arXiv: [1504.06238](#) [math.PR].
- X. S. Cai, G. Perarnau, B. Reed, and A. B. Watts, “Acyclic edge colourings of graphs with large girth,” *Random Structures and Algorithms*, 2016, To appear. arXiv: [1411.3047](#) [math.CO].
- X. S. Cai and L. Devroye, “The analysis of Kademia for random IDs,” *Internet Mathematics*, vol. 11, no. 6, pp. 572–587, 2015.
- X. S. Cai and L. Devroye, “A probabilistic analysis of Kademia networks,” in *Algorithms and Computation, 24th International Symposium, ISAAC 2013, Hong Kong, China, December 16-18, 2013, Proceedings*, vol. 8283, Heidelberg, Germany: Springer-Verlag, 2013, pp. 711–721.
- X. S. Cai, “A probabilistic analysis of Kademia networks,” Master’s thesis, McGill University, 2012.

CONTENTS

Notation and Symbols	xv
I THE PRELIMINARY	1
1 PROBLEMS AND RESULTS	3
1.1 Fringe and non-fringe subtrees	3
1.2 Conditional Galton-Watson trees	3
1.3 Fringe subtrees of conditional Galton-Watson trees	4
1.4 Non-fringe subtrees of conditional Galton-Watson trees	7
2 CONDITIONAL GALTON-WATSON TREES	9
2.1 Three types of trees	9
2.1.1 Plane trees	9
2.1.2 Galton-Watson trees	10
2.1.3 Simply generated trees	11
2.2 Preorder degree sequences	12
2.2.1 Preorder degree sequences of trees	12
2.2.2 Preorder degree sequences of conditional Galton-Watson trees	13
2.3 Examples of conditional Galton-Watson trees	16
2.3.1 Plane trees	16
2.3.2 Full binary trees	18
2.3.3 Binary trees and d-ary trees	18
2.3.4 Cayley trees	19
2.4 Asymptotic properties of conditional Galton-Watson trees	19
2.4.1 Additive functionals	19
2.4.2 Size-biased Galton-Watson trees	20
2.4.3 Maximum degree	21
2.4.4 Height and width	22
2.4.5 Continuum random tree	22
3 POISSON APPROXIMATION	25
3.1 Two lemmas for Poisson distribution	25
3.2 Stein's method and exchangeable pairs	26
3.3 Construct exchangeable pairs	28
3.4 Lower bounds	30
3.5 Factorial moments	31

II	THE PROOF	33
4	FRINGE SUBTREE COUNTS	35
4.1	Moments of fringe subtree counts	36
4.1.1	Factorial moments	36
4.1.2	Expectation and variance	38
4.2	Proof of Lemma 4.1 and Theorem 1.2	39
4.3	The proof of Theorem 1.3	43
5	FAMILIES OF FRINGE SUBTREES	45
5.1	Coupon collector problem	45
5.2	Complete r -ary fringe subtrees	46
5.2.1	Binary trees	48
5.3	All possible fringe subtrees	48
5.3.1	Computing p_k^{\min}	53
5.3.2	When $L > 0$	56
5.3.3	When $L = 0$	58
6	NON-FRINGE SUBTREES	63
6.1	Intuition and notations	63
6.2	Factorial moments	64
6.2.1	The first factorial moment	64
6.2.2	The second factorial moment	67
6.3	Proof of Theorem 1.4	72
6.4	Complete r -ary non-fringe subtrees	75
6.5	Difference between fringe and non-fringe subtrees	77
III	CONCLUSION	79
7	GALTON-WATSON TREES AND DETERMINISTIC FINITE AUTOMATA CHOSEN AT RANDOM	81
7.1	Graphs and digraphs	81
7.2	Random graphs and digraphs	82
7.3	Random DFA	83
7.4	Our results	84
7.4.1	The one-in-core and the giant	85
7.4.2	Outside the giant	87
7.5	The connection to Galton-Watson trees	88
8	OPEN QUESTIONS	91
8.1	Fringe subtrees	91
8.2	Non-fringe subtrees	91
8.3	Embedded subtrees	92

LIST OF FIGURES

Figure 1	Examples of fringe and non-fringe subtrees	4
Figure 2	Plane trees of at most 4 nodes	9
Figure 3	The preorder degree sequence of a tree and the corresponding Łukasiewicz path	13
Figure 4	Example of $\mathcal{T}_n^{\text{bin}}$	18
Figure 5	Example of d_f	23
Figure 6	Example of $\{T \boxplus T\}$	67
Figure 7	Examples of three cases in \mathcal{B}''	69
Figure 8	Example of T'_1 for T with 7 internal nodes	74
Figure 9	Examples of 2-out digraph with 5 vertices	83
Figure 10	Three layers of $D_{n,k}$	85

LIST OF TABLES

Table 1	Some discrete probability distributions	17
Table 2	Some well-known conditional Galton-Watson trees	17

NOTATION AND SYMBOLS

\mathbb{N}	$\{1, 2, \dots\}$
\mathbb{N}_0	$\{0\} \cup \mathbb{N}$
\mathbb{Z}	the set of all integers
$\llbracket H \rrbracket$	1 if the statement H is true, 0 otherwise
$(x)_r$	$x(x-1)\cdots(x-r+1)$
$\gcd(\mathcal{S})$	The greatest common divisor of integers in \mathcal{S}
$u \vee v$	$\max\{u, v\}$
$u \wedge v$	$\min\{u, v\}$
$(a_i)_{i \geq 1}$	an infinite sequence a_1, a_2, \dots
$(a_i)_{i=1}^n$	a sequence a_1, a_2, \dots, a_n
$\mathbb{E}X$	the expectation of X
$\text{Var}(X)$	the variance of X
$\text{Cov}(X, Y)$	the covariance of X and Y
$\mathbb{E}(X)_r$	the r -th factorial moment of X , i.e., $\mathbb{E}[X(X-1)\cdots(X-r+1)]$
$X \stackrel{\mathcal{L}}{=} Y$	X and Y are identically distributed
$X_n \xrightarrow{P} X$	X_n converges in probability to X
$X_n \xrightarrow{d} X$	X_n converges in distribution to X
$\sigma(X)$	the σ -algebra generated by X
$d_{TV}(X, Y)$	the total variation distance between X and Y
$\text{Po}(\lambda)$	The Poisson distribution with mean λ
$\text{Be}(p)$	Bernoulli (p) distribution
$\text{Bi}(d, p)$	Binomial (d, p) distribution

$\text{Ge}(p)$	Geometric (p) distribution
$\text{Unif}(\mathcal{M})$	Uniform distribution on \mathcal{M}
$\text{N}(0, 1)$	The standard normal distribution
$\text{N}(\mu, \gamma^2)$	The normal distribution with mean μ and variance γ^2
ξ	a non-negative integer-valued random variable with $\mathbb{E}\xi = 1$ and $0 < \sigma^2 := \text{Var}(\xi) < \infty$
p_i	$\mathbb{P}\{\xi = i\}$
$\text{span}(\xi)$	the span of ξ , i.e., $\gcd(\{i \geq 1 : p_i > 0\})$
$\text{supp}(\xi)$	the support of ξ , i.e., $\{i \in \mathbb{N}_0 : p_i > 0\}$
$\text{supp}(\xi, k)$	$\text{supp}(\xi) \cap \{1, \dots, k\}$
\mathfrak{T}	the set of all rooted, ordered and unlabeled trees (plane trees)
\mathfrak{T}_n	the set of all rooted, ordered and unlabeled trees of size n
T	a tree in \mathfrak{T}
$ T $	the size of T , i.e., the number of nodes in T
T_v	the fringe subtree of T rooted at node $v \in T$
$\text{deg}(v)$	the degree of node v
\mathfrak{S}	a set of trees
\mathfrak{S}_n	a sequence of sets of trees
$\mathcal{D}(\mathfrak{S})$	the set of preorder degree sequences of trees in \mathfrak{S}
$\mathcal{D}(\mathfrak{T}_n)$	the set of preorder degree sequences of all trees of size n
\mathcal{R}_n	the set of sequences that are cyclic rotations of sequences in $\mathcal{D}(\mathfrak{T}_n)$
$T' < T$	T has a non-fringe subtree of shape T' at its root
$T' < T_v$	T has a non-fringe subtree of shape T' rooted at v
$(T_n)_{n \geq 1}$	a sequence of trees
\mathcal{T}^{gw}	an unconditional Galton-Watson tree with offspring distribution ξ
$\mathcal{T}_n^{\text{gw}}$	\mathcal{T}^{gw} restricted to the event $ \mathcal{T}^{\text{gw}} = n$

$\mathcal{T}_{n,v}^{gw}$	a fringe subtree of \mathcal{T}_n^{gw} rooted at node $v \in \mathcal{T}_n^{gw}$
$\mathcal{T}_{n,*}^{gw}$	a fringe subtree of \mathcal{T}_n^{gw} rooted at a uniform random node of \mathcal{T}_n^{gw}
\mathcal{T}_n^{bin}	a uniform random binary tree of size n
$N_T(\mathcal{T}_n^{gw})$	the number of fringe subtrees in \mathcal{T}_n^{gw} of the shape T
$\pi(T)$	$\mathbb{P}\{\mathcal{T}^{gw} = T\}$
$N_{\mathfrak{G}}(\mathcal{T}_n^{gw})$	the number of fringe subtrees in \mathcal{T}_n^{gw} that belong to \mathfrak{G}
$\pi(\mathfrak{G})$	$\mathbb{P}\{\mathcal{T}^{gw} \in \mathfrak{G}\}$
$N_T^{nf}(\mathcal{T}_n^{gw})$	the number of non-fringe subtrees of \mathcal{T}_n^{gw} of shape T
$\pi^{nf}(T)$	$\mathbb{P}\{T < \mathcal{T}^{gw}\}$, the probability that \mathcal{T}^{gw} has a non-fringe subtree T at its root
$(\xi_i^n)_{i=1}^n$	$(\xi_1^n, \dots, \xi_n^n)$, the preorder degree sequence of \mathcal{T}_n^{gw}
ξ_1, ξ_2, \dots	iid copies of ξ
S_n	$\sum_{i=1}^n \xi_i$
$(\tilde{\xi}_i^n)_{i=1}^n$	$(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$, a uniform random cyclic rotation of $(\xi_i^n)_{i=1}^n$
$\mathfrak{T}_{\leq k}^+$	the set of all possible trees of size at most k
\mathfrak{T}_k^+	the set of all possible trees of size exactly k
K_n	$\max\{k : \mathfrak{T}_{\leq k}^+ \subseteq \cup_{v \in \mathcal{T}_n^{gw}} \{\mathcal{T}_{n,v}^{gw}\}\}$
T_k^{min}	a tree $T \in \mathfrak{T}_{\leq k}^+$ that minimizes $\mathbb{P}\{\mathcal{T}^{gw} = T\}$
p_k^{min}	$\pi(T_k^{min})$
L_k	$L_k := \min\{p_0 (p_i/p_0)^{1/i} : i \in \text{supp}(\xi, k-1)\}$
L	$\lim_{k \rightarrow \infty} L_k$
\mathfrak{L}	smallest i with $L_{i+1} = L$ if such i exists; ∞ otherwise
T_{k-1}^{star}	the tree of size k with $k-1$ leaves
\vec{d}	a sequence in \mathcal{R}_n , i.e., a rotation of a tree degree sequence of size n
$\text{deg}_i(\vec{d})$	a tree degree sequence that starts at index i in \vec{d}
$T_i((\tilde{\xi}_i^n)_{i=1}^n)$	the fringe subtree of \mathcal{T}_n^{gw} with the preorder degree sequence $\text{deg}_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$

$T_i(\vec{d})$	the tree with the preorder degree sequence $\text{deg}_i(\vec{d})$
$v(T)$	the number of internal nodes in T
$\ell(T)$	the number of leaves in T
$\mathcal{V}(T)$	the set of internal nodes in T
$G(n, m)$	a simple graph of n vertices and m edges chosen uniformly at random
$G(n, p)$	a simple graph of n vertices in which each possible edge appears with probability p independently
$D_{n,k}$	a k -out digraph of n vertices chosen uniformly at random
$D_{n,k}^A$	the condensation DAG of $D_{n,k}$
\mathcal{S}_u	the spectrum of u , i.e., the set of vertices reachable from u
\mathcal{S}_n	the set of vertices in the largest closed scc in $D_{n,k}$
\mathcal{O}_n	the set of vertices in the one-in-core of $D_{n,k}$
DFS	Depth-First-Search
lhs	left hand side
rhs	right hand side
iid	independent and identically distributed
whp	with high probability
DFA	Deterministic Finite Automaton
scc	Strongly Connected Component

Part I

THE PRELIMINARY

CHAPTER 1 PROBLEMS AND RESULTS

We introduce the problems for fringe and non-fringe subtrees in conditional Galton-Watson trees and give some of our main results.

CHAPTER 2 CONDITIONAL GALTON-WATSON TREES

We briefly review some important results on conditional Galton-Watson trees.

CHAPTER 3 POISSON APPROXIMATION

We present Stein's method in the form of exchangeable pairs for Poisson distributions.

PROBLEMS AND RESULTS

1.1 FRINGE AND NON-FRINGE SUBTREES

We study the conditions for families of fringe or non-fringe subtrees to exist with high probability (whp) in a Galton-Watson tree conditional to be of size n . In particular, we want to find the height of the maximal complete r -ary fringe and non-fringe subtrees. We also want to determine the threshold k_n such that all trees of size at most k_n appear as fringe subtrees. In doing so, we extend Janson's [65] result on fringe subtree counts and prove a new concentration theorem for non-fringe subtree counts.

Let \mathfrak{T} be the set of all rooted, ordered, and unlabeled trees, which we call *plane trees*. All trees considered in this thesis belong to \mathfrak{T} unless explicitly stated otherwise. Given a tree $T \in \mathfrak{T}$ and a node $v \in T$, let T_v denote the subtree rooted at v . T_v is usually referred to as a *fringe subtree* of T . If T_v is isomorphic to some tree $T' \in \mathfrak{T}$, then we write $T' = T_v$ and say that T has a *fringe subtree* of shape T' rooted at v , or simply T contains T' as a fringe subtree.

If T_v can be made isomorphic to T' by replacing some or none of its own fringe subtrees with *leaves* (nodes without children), then we write $T' < T_v$ and say that T has a *non-fringe subtree* of shape T' rooted at v , or simply T contains T' as a non-fringe subtree. (Note that $T' = T_v$ implies that $T' < T_v$.) We also write $T' < T$ to denote that T has a non-fringe subtree of shape T' at its root. Figure 1 shows some examples of fringe and non-fringe subtrees.

See Subsection 2.1.1 for more about plane trees.

See page xv for a complete list of notation and symbols.

1.2 CONDITIONAL GALTON-WATSON TREES

Let ξ be a non-negative integer-valued random variable. The Galton-Watson tree $\mathcal{T}^{g\omega}$ with offspring distribution ξ is the random tree generated by starting from the root and independently giving each node a random number of children, where the numbers of children are all distributed as ξ . The conditional Galton-Watson tree $\mathcal{T}_n^{g\omega}$ is $\mathcal{T}^{g\omega}$ restricted to the event $|\mathcal{T}^{g\omega}| = n$, i.e., the event that $\mathcal{T}^{g\omega}$ has n nodes.

In the study of $\mathcal{T}_n^{g\omega}$, the following is usually assumed:

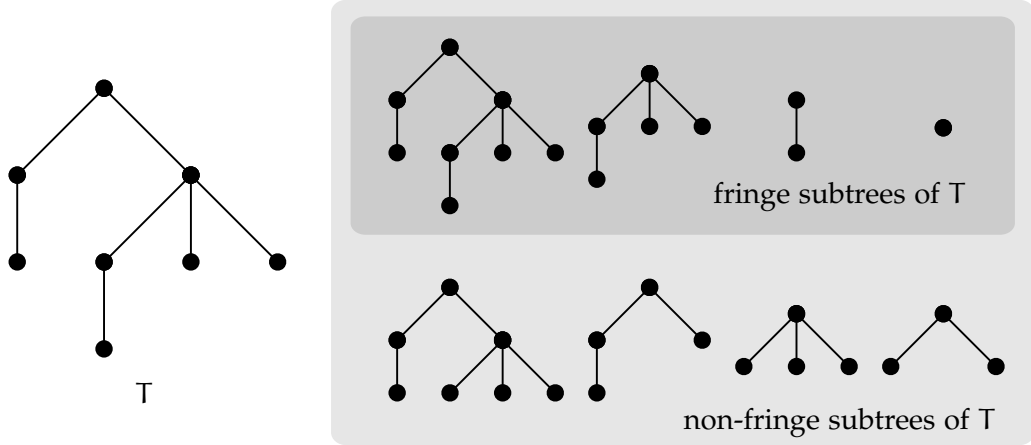


Figure 1: Examples of fringe and non-fringe subtrees

Condition A. Let \mathcal{T}_n^{gw} be a conditional Galton-Watson tree of size n with offspring distribution ξ , such that $\mathbb{E}\xi = 1$ and $0 < \sigma^2 := \text{Var}(\xi) < \infty$. Let \mathcal{T}^{gw} be the corresponding unconditional Galton-Watson tree.

Remark 1.1. Let $\text{span}(\xi) := \gcd\{i \geq 1 : p_i > 0\}$, where $p_i := \mathbb{P}\{\xi = i\}$. In general, there may exist positive integers n such that $\mathbb{P}\{|\mathcal{T}^{gw}| = n\} = 0$ and \mathcal{T}_n^{gw} is not well-defined. But it is easy to show that $\mathbb{P}\{|\mathcal{T}^{gw}| = n\} > 0$ for all $n \geq n_0$ with

$$n - 1 \equiv 0 \pmod{\text{span}(\xi)},$$

where n_0 is a constant depending only on ξ (see [64, cor. 15.6]). Therefore, in this thesis, for all asymptotic results regarding \mathcal{T}_n^{gw} , the limits are always taken as $n \rightarrow \infty$ along the subsequence with $n - 1 \equiv 0 \pmod{\text{span}(\xi)}$. We *do not* emphasize this anymore.

1.3 FRINGE SUBTREES OF CONDITIONAL GALTON-WATSON TREES

Let N_T denote the number of fringe subtrees of \mathcal{T}_n^{gw} of shape T . Extending a result by Aldous [3], Janson [64, thm. 7.12] proved the following:

Theorem 1.1. *Assume Condition A. Let $\mathcal{T}_{n,*}^{gw}$ be a fringe subtree rooted at a uniform random node of \mathcal{T}_n^{gw} . The conditional distribution $\mathcal{L}(\mathcal{T}_{n,*}^{gw} | \mathcal{T}_n^{gw})$ converges in probability to $\mathcal{L}(\mathcal{T}^{gw})$. In other words, for all $T \in \mathfrak{T}$*

$$\frac{N_T(\mathcal{T}_n^{gw})}{n} = \mathbb{P}\{\mathcal{T}_{n,*}^{gw} = T | \mathcal{T}_n^{gw}\} \xrightarrow{p} \mathbb{P}\{\mathcal{T}^{gw} = T\}. \quad (1.1)$$

See also
Subsection 2.4.1.

In [65] Janson further strengthened the above result, proving the asymptotic normality of $N_T(\mathcal{T}_n^{gw})$ by studying additive functionals on \mathcal{T}_n^{gw} .

A natural extension of $N_T(\mathcal{T}_n^{gw})$ is $N_{T_n}(\mathcal{T}_n^{gw})$, where the fixed tree T is replaced by a sequence of trees $(T_n)_{n \geq 1}$. Let $Po(\lambda)$ be a Poisson random variable with mean λ and $N(0, 1)$ be a standard normal random variable. Let $d_{TV}(\cdot, \cdot)$ denote the total variation distance. Let \xrightarrow{d} denote convergence in distribution. Let $\pi(T) := \mathbb{P}\{\mathcal{T}^{gw} = T\}$. We have:

Theorem 1.2. *Assume Condition A. Let $k_n = o(n)$ and $k_n \rightarrow \infty$. Then*

$$\lim_{n \rightarrow \infty} \sup_{T:|T|=k_n} d_{TV}(N_T(\mathcal{T}_n^{gw}), Po(n\pi(T))) = 0. \quad (1.2)$$

Therefore, letting $(T_n)_{n \geq 1}$ be a sequence of trees with $|T_n| = k_n$, we have as $n \rightarrow \infty$:

- (i) If $n\pi(T_n) \rightarrow 0$, then $N_{T_n}(\mathcal{T}_n^{gw}) = 0$ whp.
- (ii) If $n\pi(T_n) \rightarrow \mu \in (0, \infty)$, then $N_{T_n}(\mathcal{T}_n^{gw}) \xrightarrow{d} Po(\mu)$.
- (iii) If $n\pi(T_n) \rightarrow \infty$, then

$$\frac{N_{T_n}(\mathcal{T}_n^{gw}) - n\pi(T_n)}{\sqrt{n\pi(T_n)}} \xrightarrow{d} N(0, 1).$$

(i)-(iii) are simple corollaries of (1.2). See Lemma 3.2.

Remark 1.2. We actually showed that

$$\sup_{T:|T|=k_n} d_{TV}(N_T(\mathcal{T}_n^{gw}), Po(n\pi(T))) = O\left(\sqrt{p_{\max}^k k^{3/2}}\right),$$

where $p_{\max} := \max_{i \geq 0} p_i$.

Let $\llbracket H \rrbracket$ be 1 if a statement H is true and let it be 0 otherwise. Let \mathfrak{T}_k be the set of all trees of size k . For $\mathfrak{S} \subseteq \mathfrak{T}_k$, let

$$\pi(\mathfrak{S}) := \mathbb{P}\{\mathcal{T}^{gw} \in \mathfrak{S}\}, \quad \text{and} \quad N_{\mathfrak{S}}(\mathcal{T}_n^{gw}) := \sum_{v \in \mathcal{T}_n^{gw}} \llbracket \mathcal{T}_{n,v}^{gw} \in \mathfrak{S} \rrbracket.$$

The $\llbracket H \rrbracket$ notation is invented by Iverson [58].

Theorem 1.2 can be generalized as follows:

Theorem 1.3. *Assume Condition A. Let $k_n = o(n)$ and $k_n \rightarrow \infty$. Let \mathfrak{S}_n be a sequence of sets of trees with $\mathfrak{S}_n \subseteq \mathfrak{T}_{k_n}$. We have:*

- (i) If $n\pi(\mathfrak{S}_n) \rightarrow 0$, then $N_{\mathfrak{S}_n}(\mathcal{T}_n^{gw}) = 0$ whp.
- (ii) If $n\pi(\mathfrak{S}_n) \rightarrow \mu \in (0, \infty)$, then $N_{\mathfrak{S}_n}(\mathcal{T}_n^{gw}) \xrightarrow{d} Po(\mu)$.
- (iii) If $n\pi(\mathfrak{S}_n) \rightarrow \infty$, then

$$\frac{N_{\mathfrak{S}_n}(\mathcal{T}_n^{gw}) - n\pi(\mathfrak{S}_n)}{\sqrt{n\pi(\mathfrak{S}_n)}} \xrightarrow{d} N(0, 1).$$

(iv) If $\pi(\mathfrak{S}_n)/\pi(\mathfrak{T}_{k_n}) \rightarrow 0$, then

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(N_{\mathfrak{S}_n}(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(\mathfrak{S}_n))) = 0.$$

The proof of Theorem 1.2 is given in Section 4. It uses many ingredients from previous results on fringe subtrees, especially from Janson [65]. (In particular, Lemma 6.2 of [65] makes the computation of the variance of $N_{\mathfrak{S}_n}(\mathcal{T}_n^{\text{gw}})$ quite easy, which is crucial for the proof.) The key step of the proof is to construct an exchangeable pair for $N_{\mathfrak{T}}(\mathcal{T}_n^{\text{gw}})$. Then we can apply a form of Stein's method [91] called the exchangeable pair [27] to upper bound the convergence of total variation distance.

Chapter 3 explains exchangeable pairs in details.

However, this approach cannot be adapted to prove the convergence of total variation distance for $N_{\mathfrak{S}_n}(\mathcal{T}_n^{\text{gw}})$ in (iv) of Theorem 1.3 without assuming that $\pi(\mathfrak{S}_n)/\pi(\mathfrak{T}_{k_n}) \rightarrow 0$. In particular, it does not work for $N_{\mathfrak{T}_{k_n}}(\mathcal{T}_n^{\text{gw}})$, i.e., the number of fringe subtrees of size k_n . Therefore, to show (i)–(iii) of Theorem 1.3, we instead compute the factorial moments of $N_{\mathfrak{S}_n}(\mathcal{T}_n^{\text{gw}})$ (see Section 4.3).

Binary search trees and recursive trees are also well-studied random tree models (see Drmota [35]). Many authors have found results similar to Theorem 1.2 for these two types of trees, see, e.g., [31, 32, 41, 45, 48]. For recent developments, see Holmgren and Janson [56].

We say that a tree T is *possible* if $\mathbb{P}\{T^{\text{gw}} = T\} > 0$. As an application of Theorems 1.2 and 1.3, we study the following problem — when does $\mathcal{T}_n^{\text{gw}}$ contain all possible trees within a family of trees (possibly depending on n). As shown in Section 5.1, this is essentially a variation of the famous coupon collector problem [40].

In Section 5.2 we solve the above problem for the set of complete r -ary trees. Let $H_{n,r}$ be the maximal integer such that $\mathcal{T}_n^{\text{gw}}$ contains all complete r -ary trees of height at most $H_{n,r}$ as fringe subtrees. Lemma 5.2 shows that $H_{n,r} - \log_r \log n$ converges in probability to an explicit constant.

Let $\mathfrak{T}_{\leq k}^+$ be the set of all possible trees of size at most k . Let K_n be the maximal k such that every tree in $\mathfrak{T}_{\leq k}^+$ appears in $\mathcal{T}_n^{\text{gw}}$ as a fringe subtree. In Section 5.3, we show that, roughly speaking, if the tail of the offspring distribution does not drop off too quickly, $K_n/\log n$ converges in probability to a positive constant. Otherwise, we have $K_n/\log n \xrightarrow{\mathbb{P}} 0$. For example, for a uniform random Cayley tree of size n , we have $K_n \log \log(n)/\log(n) \xrightarrow{\mathbb{P}} 1$. For many well-known Galton-Watson trees, we also give the second order asymptotic term of K_n .

See Example 5.2.

1.4 NON-FRINGE SUBTREES OF CONDITIONAL GALTON-WATSON TREES

Since on average fringe subtrees in $\mathcal{T}_n^{g_w}$ behave like unconditional Galton-Watson trees when n is large, the number of non-fringe subtrees of shape T should be more or less $n\mathbb{P}\{T < \mathcal{T}^{g_w}\}$. Let

$$\pi^{nf}(T) := \mathbb{P}\{T < \mathcal{T}^{g_w}\}, \quad \text{and} \quad N_T^{nf}(\mathcal{T}_n^{g_w}) := \sum_{v \in \mathcal{T}_n^{g_w}} \mathbb{1}[T < \mathcal{T}_{n,v}^{g_w}].$$

We make the above intuition precise in the following theorem:

Theorem 1.4. *Assume Condition A. Let T_n be a sequence of trees with $|T_n| = k_n$ where $k_n \rightarrow \infty$ and $k_n = o(n)$. We have*

(i) *If $n\pi^{nf}(T_n) \rightarrow 0$, then $N_{T_n}^{nf}(\mathcal{T}_n^{g_w}) = 0$ whp.*

(ii) *If $n\pi^{nf}(T_n) \rightarrow \infty$, then*

$$\frac{N_{T_n}^{nf}(\mathcal{T}_n^{g_w})}{n\pi^{nf}(T_n)} \xrightarrow{p} 1.$$

Chyzak, Drmota, Klausner, and Kok [29] studied non-fringe subtrees for various random trees, including *simply generated trees*. They proved that if for all n we have $T_n = T$ where T is fixed, then $N_{T_n}^{nf}(\mathcal{T}_n^{g_w})$ obeys a central limit theorem. However, Theorem 1.4 cannot be simply derived from their result as our T_n depends upon n .

Remark 1.3. It is tempting to try to prove that if $n\pi^{nf}(T_n) \rightarrow \mu \in (0, \infty)$, then we have $N_{T_n}^{nf}(\mathcal{T}_n^{g_w}) \xrightarrow{d} \text{Po}(\mu)$, which is true for fringe subtrees. But unfortunately this is not true in general for non-fringe subtrees. See Lemma 6.14 in Section 6.3. The reason for this is that two non-fringe subtrees can overlap for a large part. Thus the correlation between non-fringe subtrees is stronger than that of fringe subtrees.

CONDITIONAL GALTON-WATSON TREES

In this chapter we review some properties of the conditional Galton-Watson tree \mathcal{T}_n^{gw} . The first three sections are needed for understanding the proof of our main results in [Part ii](#). [Section 2.1](#) introduces three types of trees closely related to \mathcal{T}_n^{gw} . [Section 2.2](#) discusses the preorder degree sequences of \mathcal{T}_n^{gw} . [Section 2.3](#) gives some examples of \mathcal{T}_n^{gw} . In addition, [Section 2.4](#) summarizes a few other important results of \mathcal{T}_n^{gw} including: additive functionals, size-biased Galton-Watson trees, maximum degree, height and width, and continuum random trees.

2.1 THREE TYPES OF TREES

2.1.1 Plane trees

In graph theory, a *tree* is a connected graph without cycles. But trees considered by us are mostly plane trees and we omit the word “plane” when it is clear. Recall that in [Section 1.1](#) plane trees are defined as *rooted, ordered, and unlabeled* trees, which we further elaborate here.

A tree T is *rooted* if it has a unique node marked as the *root*. Nodes that are at (graph) distance x away from the root are called generation x .

For a node $v \in T$, the unique neighbor of v on the path from v to the root is the *parent* of v . Other neighbors of v (if there are any) are the *children* of v . The number of v 's children is its *degree*, which we denote by $\deg(v)$. If a node has no children, then it is a *leaf*, otherwise it is an *internal* node. T is called *ordered* if for each node $v \in T$, the children of v are ordered.

A rooted and ordered tree is *unlabeled* if it does not have labels on nodes. Such a tree is called a plane tree. [Figure 2](#) depicts all plane trees of at most 4 nodes.

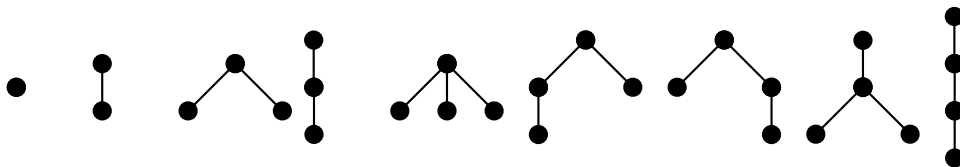


Figure 2: Plane trees of at most 4 nodes

The term *plane* is used because a plane tree can be also seen as a rooted tree together with an embedding on a plane. For a more combinatoric view, see Flajolet and Sedgewick [43, sec. I.5.1].

2.1.2 Galton-Watson trees

“No law, theorem, or
discovery is named
after its
originator”—
Stigler’s law of
eponymy

Galton-Watson trees are named after Watson and Galton for their 1875 paper [94], but the model was first introduced by Bienaymé in 1845 [14]. The authors tried to explain an ostensibly mysterious phenomenon at the time—although the population of a nation kept increasing, a considerable number of family names had disappeared.

Bienaymé studied this problem with a simple probabilistic model: A family starts with a single member, the root, which forms generation 0. Once generation i is born, each of its members reproduces a random number of children according to an offspring distribution ξ , independently from everything else. The newborns form the generation $(i + 1)$ and begin to reproduce generation $(i + 2)$ in the same manner.

This process can repeat forever, an event referred to as *survival*. It can also end when one generation reproduces zero children, an event called *extinction*.

The genealogical tree of such a family is called a Galton-Watson tree with offspring distribution ξ , denoted by $\mathcal{T}^{g\omega}$. Let $|\mathcal{T}^{g\omega}|$ be the size of $\mathcal{T}^{g\omega}$, i.e., the number of nodes in it. Then $|\mathcal{T}^{g\omega}| = \infty$ corresponds to survival and $|\mathcal{T}^{g\omega}| < \infty$ corresponds to extinction. Bienaymé’s surprising result is that even a healthy family in which each member has one child on average is still destined for extinction:

Theorem 2.1. *Let $p_i := \mathbb{P}\{\xi = i\}$. Assume that $p_1 < 1$.*

1. *If $\mathbb{E}\xi \leq 1$, then $\mathbb{P}\{|\mathcal{T}^{g\omega}| < \infty\} = 1$.*
2. *If $\mathbb{E}\xi > 1$, then $\mathbb{P}\{|\mathcal{T}^{g\omega}| < \infty\} \in (0, 1)$.*

The probability $\mathbb{P}\{|\mathcal{T}^{g\omega}| = \infty\}$ is called the *survival* probability of $\mathcal{T}^{g\omega}$. The cases $\mathbb{E}\xi < 1$, $\mathbb{E}\xi = 1$, and $\mathbb{E}\xi > 1$ are called *subcritical*, *critical* and *supercritical* respectively. Usually, only critical Galton-Watson trees are of interest for the reason explained in the next subsection.

See Durrett [37, sec. 5.3.4] for the proof of Theorem 2.1 and the connection between $\mathcal{T}_n^{g\omega}$ and martingales. For a stochastic process perspective of Galton-Watson trees, see Athreya and Ney [12, chap. I].

2.1.3 Simply generated trees

A conditional Galton-Watson tree of size n , denoted by \mathcal{T}_n^{gw} , is a random tree distributed as \mathcal{T}^{gw} restricted to the event that $|\mathcal{T}^{gw}| = n$. In other words,

Recall $\mathbb{I}[H]$ is the indicator that H is true.

$$\begin{aligned} \mathbb{P}\{\mathcal{T}_n^{gw} = T\} &= \mathbb{P}\{\mathcal{T}^{gw} = T \mid |\mathcal{T}^{gw}| = n\} \\ &= \frac{\mathbb{P}\{[\mathcal{T}^{gw} = T] \cap [|\mathcal{T}^{gw}| = n]\}}{\mathbb{P}\{|\mathcal{T}^{gw}| = n\}} \\ &= \mathbb{I}[|T| = n] \times \frac{\mathbb{P}\{\mathcal{T}^{gw} = T\}}{\mathbb{P}\{|\mathcal{T}^{gw}| = n\}} \\ &= \mathbb{I}[|T| = n] \times \frac{\prod_{v \in T} p_{\deg(v)}}{\sum_{T' \in \mathfrak{T}_n} \prod_{v \in T'} p_{\deg(v)}}, \end{aligned}$$

where \mathfrak{T}_n is the set of all rooted, ordered and unlabeled trees (plane trees) of size n .

The above probability implies another way to define conditional Galton-Watson trees. Let $(w_i)_{i \geq 0}$ be a sequence of non-negative numbers. Let the weight of a tree T be

$$\text{weight}(T) := \prod_{v \in T} w_{\deg(v)}.$$

Meir and Moon [79] call trees with such weights *simply generated trees*.

Let \mathcal{T}_n^{sg} be a random tree of size n such that $\mathbb{P}\{\mathcal{T}_n^{sg} = T\}$ is proportional to $\text{weight}(T)$ if $|T| = n$. In other words,

$$\begin{aligned} \mathbb{P}\{\mathcal{T}_n^{sg} = T\} &= \mathbb{I}[|T| = n] \times \frac{\text{weight}(T)}{\sum_{T' \in \mathfrak{T}_n} \text{weight}(T')} \\ &= \mathbb{I}[|T| = n] \times \frac{\prod_{v \in T} w_{\deg(v)}}{\sum_{T' \in \mathfrak{T}_n} \prod_{v \in T'} w_{\deg(v)}}. \end{aligned}$$

Thus if $\sum_{i \geq 0} w_i = 1$, i.e., $(w_i)_{i \geq 0}$ is a probability distribution, then \mathcal{T}_n^{sg} is nothing but \mathcal{T}_n^{gw} with offspring distribution $p_i = w_i$.

Even if $(w_i)_{i \geq 0}$ is not a probability distribution, the following lemma shows that in many cases, \mathcal{T}_n^{sg} is still equivalent to a conditional Galton-Watson tree.

Lemma 2.1 (Kennedy [68]). *Assume that $w_0 > 0$ and that $w_i > 0$ for some $i \geq 2$. Define*

$$\phi(t) := \sum_{i \geq 0} w_i t^i.$$

If $\phi(\rho) < \infty$ for some $\rho \in (0, \infty]$, then \mathcal{T}_n^{sg} has the same distribution as \mathcal{T}_n^{gw} with offspring distribution

$$p_i = \frac{w_i t^i}{\phi(t)},$$

for all $t \in (0, \rho)$.

This lemma suggests that results of $\mathcal{T}_n^{\text{sg}}$ often also apply to $\mathcal{T}_n^{\text{gw}}$. For example, the theorems in [Chapter 1](#) can also be formulated in terms of $\mathcal{T}_n^{\text{sg}}$. For an encyclopedic survey of simply generated trees, see Janson [64].

2.2 PREORDER DEGREE SEQUENCES

2.2.1 Preorder degree sequences of trees

The *preorder* of nodes in a tree T is the order in which they are visited through the following Depth-First-Search (DFS) procedure:

A stack is a data structure in which data elements are added and removed in First-In-Last-Out order.

1. Let \mathcal{Q} be an empty stack. Mark all nodes *unvisited*.
2. Put the root of T at the head of \mathcal{Q} .
3. Remove a node v at the head of \mathcal{Q} . Mark it as *visited*.
4. Add the children of v at the head of \mathcal{Q} in order of appearance.
5. If \mathcal{Q} is empty, terminate. Otherwise go to step 3.

Let $v_1, \dots, v_{|T|}$ be the nodes of T in preorder. The sequence

$$(\deg(v_1), \deg(v_2), \dots, \deg(v_{|T|}))$$

is called the *preorder degree sequence* of T .

In the above DFS procedure, initially, there is one node in the stack \mathcal{Q} . In each iteration, the size of \mathcal{Q} decreases by one and increases by $\deg(v)$. Since \mathcal{Q} becomes empty only after all nodes have been visited, we have the following well-known lemma (see Janson [64, lem. 15.2] or Flajolet and Sedgewick [43, pp. 74]):

Lemma 2.2. *Let $\mathbb{N} := \{1, 2, \dots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Let $\mathcal{D}(\mathfrak{T}_n) \subseteq \mathbb{N}_0^n$ be the set of preorder degree sequences of all trees of size n . A sequence $(d_1, \dots, d_n) \in \mathcal{D}(\mathfrak{T}_n)$ if and only if it satisfies*

$$\begin{cases} \sum_{i=1}^j d_i \geq j, & (1 \leq j \leq n-1) \\ \sum_{i=1}^n d_i = n-1. \end{cases} \quad (2.1)$$

The path depicted on the right side is called the Łukasiewicz path [43, pp. 75].

Figure 3 gives an example of Lemma 2.2.

We call a sequence in $\cup_{n=1}^{\infty} \mathcal{D}(\mathfrak{T}_n)$ a *tree degree sequence*. Observe that a tree degree sequence cannot be the prefix of another tree degree sequence:

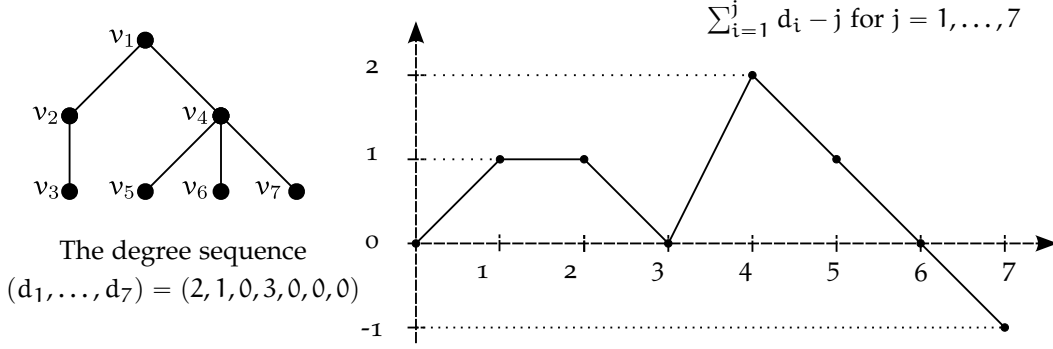


Figure 3: The preorder degree sequence of a tree and the corresponding Łukasiewicz path

Corollary 2.1. *If $(d_1, d_2, \dots, d_n) \in \mathcal{D}(\mathfrak{T}_n)$, then it is impossible that there exists $1 \leq n' < n$ such that $(d_1, d_2, \dots, d_{n'}) \in \mathcal{D}(\mathfrak{T}_{n'})$.*

On the other hand, if the DFS procedure starts from a node v other than the root of T , then the procedure terminates once every node in T_v (the fringe subtree rooted at v) has been visited. This implies the famous *cycle lemma* (see Janson [64, lem. 15.3] or Flajolet and Sedgewick [43, pp. 75]):

Lemma 2.3 (Cycle lemma). *Define*

$$\mathcal{R}_n := \left\{ (r_1, \dots, r_n) \in \mathbb{N}_0^n : \sum_{i=1}^n r_i = n - 1 \right\}.$$

If $(d_1, \dots, d_n) \in \mathcal{R}_n$, then there exists one and only one $j \in \{1, \dots, n\}$ such that

$$(d_{1+j}, d_{2+j}, \dots, d_{n+j}) \in \mathcal{D}(\mathfrak{T}_n),$$

where the indices are taken modulo n .

2.2.2 Preorder degree sequences of conditional Galton-Watson trees

Let $(\xi_i^n)_{i=1}^n := (\xi_1^n, \dots, \xi_n^n)$ denote the preorder degree sequence of $\mathcal{T}_n^{\text{gw}}$. Let $(\tilde{\xi}_i^n)_{i=1}^n := (\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ be a uniform random cyclic rotation of $(\xi_i^n)_{i=1}^n$. Let $(\xi_i)_{i=1}^n := (\xi_1, \dots, \xi_n)$ be independent and identically distributed (iid) copies of ξ . Let $S_n := \sum_{i=1}^n \xi_i$.

The cycle lemma (Lemma 2.3) shows that there is a one-to- n correspondence between $\mathcal{D}(\mathfrak{T}_n)$ and \mathcal{R}_n . This leads to the following well-known connection between $(\tilde{\xi}_i^n)_{i=1}^n$ and (ξ_1, \dots, ξ_n) (see Janson [64, cor. 15.4]):

Lemma 2.4. *We have*

$$\begin{aligned} \mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\} &= \mathbb{P}\{(\xi_1^n, \dots, \xi_n^n) \in \mathcal{D}(\mathfrak{T}_n)\} \\ &= \frac{\mathbb{P}\{(\xi_1, \dots, \xi_n) \in \mathcal{R}_n\}}{n} = \frac{\mathbb{P}\{S_n = n-1\}}{n}. \end{aligned}$$

As a result,

$$(\tilde{\xi}_1^n, \tilde{\xi}_2^n, \dots, \tilde{\xi}_n^n) \stackrel{\mathcal{L}}{=} (\xi_1, \dots, \xi_n \mid S_n = n-1),$$

where $\stackrel{\mathcal{L}}{=}$ denotes “identically distributed” and the right-hand-side (rhs) denotes the sequence (ξ_1, \dots, ξ_n) restricted to the event that $S_n = n-1$.

Recall that $\text{span}(\xi) := \gcd(\{i \geq 1 : p_i > 0\})$, where \gcd denotes the *greatest common divisor*. The following lemma follows from Lemma 2.4 and the *local limit theorem* (see Janson [65, (4.3)] or Kolchin [71]):

Lemma 2.5. *Assume Condition A. We have*

$$\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\} = \frac{1}{n} \mathbb{P}\{S_n = n-1\} \sim \frac{\text{span}(\xi)}{\sqrt{2\pi\sigma^2}} n^{-3/2},$$

as $n \rightarrow \infty$ with $n-1 \equiv 0 \pmod{\text{span}(\xi)}$.

Let $\mathfrak{S} \subseteq \mathfrak{T}$ be a set of trees. Define

$$\pi(\mathbb{T}) := \mathbb{P}\{\mathcal{T}^{\text{gw}} = \mathbb{T}\}, \quad \text{and} \quad \pi(\mathfrak{S}) := \mathbb{P}\{\mathcal{T}^{\text{gw}} \in \mathfrak{S}\}.$$

The following lemma is a special case of [65, lem. 5.1]. We give a short proof for later reference.

Lemma 2.6. *Assume $\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\} > 0$.*

- (i) *Let $\mathbb{T} \in \mathfrak{T}_k$ with $1 \leq k \leq n$. Let $N_{\mathbb{T}}(\mathcal{T}_n^{\text{gw}})$ be the number of fringe subtrees of $\mathcal{T}_n^{\text{gw}}$ in the shape of \mathbb{T} , i.e.,*

$$N_{\mathbb{T}}(\mathcal{T}_n^{\text{gw}}) := \sum_{v \in \mathcal{T}_n^{\text{gw}}} \mathbb{I}[\mathcal{T}_{n,v}^{\text{gw}} = \mathbb{T}].$$

Then

$$\frac{\mathbb{E}[N_{\mathbb{T}}(\mathcal{T}_n^{\text{gw}})]}{n} = \pi(\mathbb{T}) \frac{\mathbb{P}\{S_{n-k} = n-k\}}{\mathbb{P}\{S_n = n-1\}}.$$

- (ii) *Let $\mathfrak{S} \subseteq \mathfrak{T}_k$ be a set of some trees of size k with $1 \leq k \leq n$. Let $N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}})$ be the number of fringe subtrees of $\mathcal{T}_n^{\text{gw}}$ belonging to \mathfrak{S} , i.e.,*

$$N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}}) = \sum_{v \in \mathcal{T}_n^{\text{gw}}} \mathbb{I}[\mathcal{T}_{n,v}^{\text{gw}} \in \mathfrak{S}].$$

Then

$$\frac{\mathbb{E}[N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}})]}{n} = \pi(\mathfrak{S}) \frac{\mathbb{P}\{S_{n-k} = n-k\}}{\mathbb{P}\{S_n = n-1\}}.$$

Proof. Let (d_1, d_2, \dots, d_k) be the preorder degree sequence of T . Recall that $(\xi_1^n, \dots, \xi_n^n)$ is the preorder degree sequence of \mathcal{T}_n^{gw} . Let

$$I_i = \llbracket \xi_i^n = d_1, \xi_{i+1}^n = d_2, \dots, \xi_{i+k-1}^n = d_k \rrbracket,$$

where the indices are taken modulo n .

Note that if $n - k + 1 < i \leq n$, then it is impossible that $I_i = 1$, because the length of the preorder degree sequence of the fringe subtree $\mathcal{T}_{n, v_i}^{gw}$ must be strictly less than k . Therefore, if $I_i > 1$ and $n - k + 1 < i \leq n$, then there exists a $k' < k$ such that $(d_1, d_2, \dots, d_{k'})$ is also a preorder degree sequence, which is impossible by Corollary 2.1.

Therefore for all $1 \leq i \leq n$, $I_i = \llbracket T_{v_i} = T \rrbracket$ and $N_T(\mathcal{T}_n^{gw}) = \sum_{i=1}^n I_i$. Recalling that $(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ is a uniform random rotation of $(\xi_1^n, \dots, \xi_n^n)$ and using Lemma 2.4, we have

$$\begin{aligned} \mathbb{E}[N_T(\mathcal{T}_n^{gw})] &= \mathbb{E} \sum_{i=1}^n I_i = \sum_{i=1}^n \mathbb{P} \left\{ \tilde{\xi}_i^n = d_1, \dots, \tilde{\xi}_{i+k-1}^n = d_k \right\} \\ &= \sum_{i=1}^n \mathbb{P} \left\{ \xi_i = d_1, \dots, \xi_{i+k-1} = d_k \mid S_n = n-1 \right\} \\ &= n \mathbb{P} \left\{ \xi_1 = d_1, \dots, \xi_k = d_k \mid S_n = n-1 \right\} \\ &= n \frac{\mathbb{P} \left\{ [\xi_1 = d_1, \dots, \xi_k = d_k] \cap [S_n = n-1] \right\}}{\mathbb{P} \{S_n = n-1\}}. \end{aligned}$$

Since (d_1, \dots, d_k) is a tree degree sequence, by Lemma 2.2, $\sum_{i=1}^k d_i = k-1$. Therefore, using the mutual independence of ξ_1, \dots, ξ_n , the last expression equals

$$\begin{aligned} &n \frac{\mathbb{P} \left\{ [\xi_1 = d_1, \dots, \xi_k = d_k] \cap [S_{n-k} = n-k] \right\}}{\mathbb{P} \{S_n = n-1\}} \\ &= n \mathbb{P} \{ \mathcal{T}^{gw} = T \} \frac{\mathbb{P} \{S_{n-k} = n-k\}}{\mathbb{P} \{S_n = n-1\}}. \end{aligned}$$

Thus part (i) is proved. Part (ii) follows by summing the equality in (i) over all $T \in \mathfrak{G}$. \square

The following approximations are useful for estimating the expectation and the variance of the fringe subtree counts.

Lemma 2.7 (Lemma 5.2 and 6.2 of Janson [65]). *Assume Condition A and that $\text{span}(\xi) = 1$. We have:*

(i) *Uniformly for all k with $1 \leq k \leq n/2$,*

$$\frac{\mathbb{P} \{S_{n-k} = n-k\}}{\mathbb{P} \{S_n = n-1\}} = 1 + O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right).$$

This is very convenient since we do not need to consider the boundary condition anymore.

Moreover,

$$\frac{\mathbb{P}\{S_{n-1} = n-1\}}{\mathbb{P}\{S_n = n-1\}} = 1 + o\left(n^{-1/2}\right).$$

(ii) Uniformly for all k with $1 \leq k \leq n/4$,

$$\begin{aligned} \frac{\mathbb{P}\{S_{n-2k} = n-2k+1\}}{\mathbb{P}\{S_n = n-1\}} &= \left(\frac{\mathbb{P}\{S_{n-k} = n-k\}}{\mathbb{P}\{S_n = n-1\}} \right)^2 \\ &= -\frac{1}{\sigma^2 n} + o\left(\frac{1}{n}\right) + O\left(\frac{k}{n^{3/2}} + \frac{k^2}{n^2}\right). \end{aligned}$$

Remark 2.1. As shown in the proof of Lemma 2.6, $N_T(\mathcal{T}_n^{gw})$ is equivalent to the number of patterns $d_1, \dots, d_{|T|}$ in the cycle $\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n$. Thus if $\text{span}(\xi) > 1$, we can divide $d_1, \dots, d_{|T|}$ and $\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n$ by $\text{span}(\xi)$ without changing the value of $N_T(\mathcal{T}_n^{gw})$. Therefore, when studying subtree counts, we can always assume that $\text{span}(\xi) = 1$.

Corollary 2.2. For all integers w, z with $w + z \leq n/2$, we have

$$\frac{\mathbb{P}\{S_{n-z} = n-z-w\}}{\mathbb{P}\{S_n = n-1\}} = 1 + O\left(\frac{w+z}{n}\right) + o(n^{-1/2}).$$

Proof. It follows from (i) of Lemma 2.7 that

$$\begin{aligned} &\frac{\mathbb{P}\{S_{n-z} = n-z-w\}}{\mathbb{P}\{S_n = n-1\}} \\ &= \frac{\mathbb{P}\{S_{n-z} = n-z-w\}}{\mathbb{P}\{S_{n-z} = n-z-1\}} \frac{\mathbb{P}\{S_{n-z} = n-z-1\}}{\mathbb{P}\{S_{n-z-1} = n-z-1\}} \frac{\mathbb{P}\{S_{n-z-1} = n-z-1\}}{\mathbb{P}\{S_n = n-1\}} \\ &= \left(1 + O\left(\frac{w}{n-z}\right)\right) \left(1 + o\left((n-z)^{-1/2}\right)\right) \left(1 + O\left(\frac{z}{n}\right)\right) \\ &= 1 + O\left(\frac{w+z}{n}\right) + o(n^{-1/2}). \quad \square \end{aligned}$$

2.3 EXAMPLES OF CONDITIONAL GALTON-WATSON TREES

This section introduces some common examples of \mathcal{T}_n^{gw} . Table 1 summarizes our notations for discrete probability distributions. Table 2 lists conditional Galton-Watson trees with these distributions, which are discussed in the rest of this section.

2.3.1 Plane trees

Assume that $\xi \stackrel{\mathcal{L}}{=} \text{Ge}(1/2)$. Let T be a (plane) tree of size n and (d_1, \dots, d_n) be its preorder degree sequence. By Lemma 2.2 and 2.4, we have

$$\mathbb{P}\{\mathcal{T}_n^{gw} = T\} = \frac{\mathbb{P}\{\mathcal{T}^{gw} = T\}}{\mathbb{P}\{|\mathcal{T}^{gw}| = n\}} = \frac{2^{-\sum_{i=1}^n (d_i+1)}}{\mathbb{P}\{|\mathcal{T}^{gw}| = n\}} = \frac{4^{-(n-1)}}{2\mathbb{P}\{|\mathcal{T}^{gw}| = n\}}.$$

NAME	NOTATION	DEFINITION
Geometric	$\text{Ge}(p)$	$p_i = p(1-p)^i \quad (i \in \mathbb{N}_0)$
Bernoulli	$\text{Be}(p)$	$p_0 = (1-p), p_1 = p$
Uniform	$\text{Unif}(\mathcal{M})$	$p_i = \frac{1}{ \mathcal{M} } \quad (i \in \mathcal{M})$
Binomial	$\text{Bi}(d, p)$	$p_i = \binom{d}{i} (1-p)^{d-i} p^i \quad (i \in \{0, \dots, d\})$
Poisson	$\text{Po}(\lambda)$	$p_i = e^{-\lambda} \lambda^i / i! \quad (i \in \mathbb{N}_0)$

Table 1: Some discrete probability distributions

NAME	OFFSPRING DISTRIBUTION
Plane trees	$\text{Ge}(1/2) \quad p_i = 1/2^{i+1} \quad (i \in \mathbb{N}_0)$
Full binary trees	$2 \times \text{Be}(1/2) \quad p_0 = p_2 = 1/2$
Motzkin trees	$\text{Unif}(\{0, 1, 2\}) \quad p_0 = p_1 = p_2 = 1/3$
Binary trees	$\text{Bi}(2, 1/2) \quad p_0 = p_2 = 1/4, p_1 = 1/2$
d-ary trees	$\text{Bi}(d, 1/d) \quad p_i = \binom{d}{i} (1 - \frac{1}{d})^{d-i} (\frac{1}{d})^i \quad (i \in \{0, \dots, d\})$
Cayley trees	$\text{Po}(1) \quad p_i = e^{-1}/i! \quad (i \in \mathbb{N}_0)$

Table 2: Some well-known conditional Galton-Watson trees

Since this probability does not depend on T , $\mathcal{T}_n^{\text{gw}}$ is uniformly distributed on \mathfrak{T}_n (the set of all plane trees of size n).

In this case, we have $\text{span}(\xi) = 1$ and $\sigma^2 = 2$. So by Lemma 2.3,

$$\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\} \sim \frac{1}{2\sqrt{\pi n^{3/2}}}.$$

Since $\mathcal{T}_n^{\text{gw}}$ has size n , we have

$$1 = \mathbb{P}\{\mathcal{T}_n^{\text{gw}} \in \mathfrak{T}_n\} = \sum_{T \in \mathfrak{T}_n} \mathbb{P}\{\mathcal{T}_n^{\text{gw}} = T\} = |\mathfrak{T}_n| \frac{4^{-(n-1)}}{2\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\}}.$$

Therefore,

$$|\mathfrak{T}_n| = 2 \cdot 4^{n-1} \mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\} \sim \frac{4^{n-1}}{\sqrt{\pi n^{3/2}}}.$$

In fact the exact formula of $|\mathfrak{T}_n|$ is known (see Flajolet and Sedgewick [43, pp. 36]):

Lemma 2.8. *We have*

$$|\mathfrak{T}_n| = \mathfrak{C}_{n-1} := \frac{1}{n} \binom{2n-2}{n-1} \sim \frac{4^{n-1}}{\sqrt{\pi n^{3/2}}},$$

where the numbers \mathfrak{C}_n are called Catalan numbers.

Named after Eugène Catalan for his 1838 paper [24].

2.3.2 Full binary trees

A full binary tree is a tree that contains only degree-2 nodes and leaves. Let $\overline{\mathfrak{B}}_n$ be the set of all full binary trees of size n . Let $T \in \overline{\mathfrak{B}}_n$. Let n_0 and n_2 be the numbers of nodes in T with degree 0 and 2 respectively. By Lemma 2.2, we have

$$n_2 = \frac{n-1}{2}, \quad \text{and} \quad n_0 = \frac{n+1}{2}.$$

Therefore, regardless of the choice of T , we always have

$$\mathbb{P} \{ \mathcal{T}_n^{gw} = T \} = \frac{p_0^{(n+1)/2} p_1^{(n-1)/2}}{\mathbb{P} \{ |\mathcal{T}^{gw}| = n \}}.$$

In other words, if $\text{supp}(\xi) = \{0, 1\}$, e.g., $\xi \stackrel{\mathcal{L}}{=} 2 \text{Ber}(1/2)$, then \mathcal{T}_n^{gw} is uniformly distributed on $\overline{\mathfrak{B}}_n$.

A tree with maximal degree 2 is called a Motzkin tree or a unary-binary tree (see [43, note I.39]). Similar to full binary trees, if $\xi \stackrel{\mathcal{L}}{=} \text{Unif}(\{1, 2, 3\})$, then \mathcal{T}_n^{gw} is uniformly distributed on all Motzkin trees of size n .

2.3.3 Binary trees and d-ary trees

A binary tree is a tree in which each node has two positions, a left one and a right one, where child nodes can attach to. Let $\xi \stackrel{\mathcal{L}}{=} \text{Bi}(2, 1/2)$. Given \mathcal{T}_n^{gw} , for each node $v \in \mathcal{T}_n^{gw}$, we choose $\text{deg}(v)$ of v 's positions uniformly and independently at random. Then we attach v 's children to these positions in the order of these children. Let $\mathcal{T}_n^{\text{bin}}$ be the binary tree constructed in this way. See Figure 4 for an example.

So if $\text{deg}(v) = 2$, then there is no randomness choosing positions for children of v .

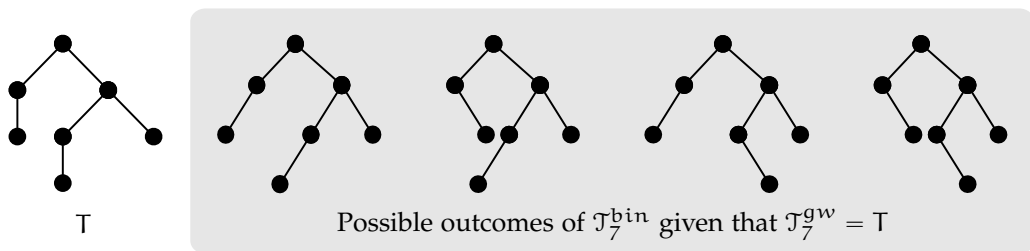


Figure 4: Example of $\mathcal{T}_n^{\text{bin}}$

Recall that \mathfrak{T}_n is the set of all trees of size n .

Let $T \in \mathfrak{T}_n$. Let n_0, n_1, n_2 be the numbers of nodes in T with degree 0, 1, 2 respectively. Let T^{bin} be one possible outcome of $\mathcal{T}_n^{\text{bin}}$ conditioning on that $\mathcal{T}_n^{gw} = T$. Since there are 2^{n_1} possible outcomes of $\mathcal{T}_n^{\text{bin}}$ and they are equally likely, we have

$$\mathbb{P} \{ \mathcal{T}_n^{\text{bin}} = T^{\text{bin}} \mid \mathcal{T}_n^{gw} = T \} = \frac{1}{2^{n_1}}.$$

Therefore,

$$\begin{aligned} \mathbb{P}\{\mathcal{T}_n^{\text{bin}} = T^{\text{bin}}\} &= \mathbb{P}\{\mathcal{T}_n^{\text{bin}} = T^{\text{bin}} \mid \mathcal{T}_n^{\text{gw}} = T\} \frac{\mathbb{P}\{\mathcal{T}^{\text{gw}} = T\}}{\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\}} \\ &= \frac{1}{2^{n_1}} \frac{4^{-n_0} 4^{-n_2} 2^{-n_1}}{\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\}} \\ &= \frac{1}{4^n \mathbb{P}\{|\mathcal{T}^{\text{gw}}| = n\}}. \end{aligned}$$

In other words, $\mathcal{T}_n^{\text{bin}}$ is uniformly distributed on all binary trees of size n . This is well-known from the connections between simply generated trees and Galton-Watson trees (see Example 10.4 of Janson [64]).

Binary trees can be generalized to d -ary trees in which each node has $d \geq 2$ positions where child nodes can attach to. In that case $\xi \stackrel{\mathcal{L}}{=} \text{Bi}(d, 1/d)$. See Example 10.6 of [64].

2.3.4 Cayley trees

A *Cayley tree* T is a *rooted, unordered and labeled tree*. T is labeled means that each node in T has a unique label chosen from $\{1, \dots, |T|\}$. T is unordered means that the child nodes of the same parent are unordered. Thus T is *not* a plane tree.

Let \mathfrak{K}_n be the set of Cayley trees of size n . Let $\xi \stackrel{\mathcal{L}}{=} \text{Poi}(1)$. If we label the nodes of $\mathcal{T}_n^{\text{gw}}$ with $\{1, \dots, n\}$ uniformly at random and forget about the ordering of child nodes, then we get a random Cayley Tree. It is well-known such a random tree is uniformly distributed on \mathfrak{K}_n . See Example 10.2 of Janson [64].

In 1889, Arthur Cayley [25] proved the famous formula bearing his name: $|\mathfrak{K}_n| = n^{n-1}$.

2.4 ASYMPTOTIC PROPERTIES OF CONDITIONAL GALTON-WATSON TREES

This section summarizes some extra asymptotic properties of conditional Galton-Watson trees.

2.4.1 Additive functionals

Let $f : \mathfrak{T} \rightarrow \mathbb{R}$ be a function on trees. Define $F : \mathfrak{T} \rightarrow \mathbb{R}$ by

$$F(T) = \sum_{v \in T} f(T_v),$$

where T_v denotes the fringe subtree of T rooted at v . A function F that can be written in the above form for some f is called an *additive functional*.

If $f(T') = \llbracket T' = T \rrbracket$, then

$$F(\mathcal{T}_n^{\text{gw}}) = \sum_{v \in \mathcal{T}_n^{\text{gw}}} f(\mathcal{T}_{n,v}^{\text{gw}}) = N_T(\mathcal{T}_n^{\text{gw}}).$$

On the other hand, we can write

$$F(\mathcal{T}_n^{\text{gw}}) = \sum_{T \in \mathfrak{T}} f(T) N_T(\mathcal{T}_n^{\text{gw}}).$$

Thus additive functionals and fringe subtree counts are essentially the same thing.

Recall that $\mathcal{T}_{n,*}^{\text{gw}}$ is a fringe subtree rooted at a uniform random node of $\mathcal{T}_n^{\text{gw}}$. Theorem 1.1 naturally extends from fringe subtree counts to additive functionals:

Theorem 2.2 (Janson [65, thm. 1.3]). *Assume Condition A. For a bounded function f on \mathfrak{T} ,*

$$\frac{F(\mathcal{T}_n^{\text{gw}})}{n} = \mathbb{E} [f(\mathcal{T}_{n,*}^{\text{gw}}) \mid \mathcal{T}_n^{\text{gw}}] \xrightarrow{p} \mathbb{E} f(\mathcal{T}^{\text{gw}}). \quad (2.2)$$

In [65], Janson also proved a *central limit theorem* for $F(\mathcal{T}_n^{\text{gw}})$:

Theorem 2.3 (Janson [65, thm. 1.5]). *Assume Condition A. Let $f : \mathfrak{T} \rightarrow \mathbb{R}$ be a function with $\mathbb{E} f(\mathcal{T}^{\text{gw}}) < \infty$. If $\mathbb{E} [f(\mathcal{T}_n^{\text{gw}})^2] \rightarrow 0$, and*

$$\sum_{n=1}^{\infty} \frac{\sqrt{\mathbb{E} [f(\mathcal{T}_n^{\text{gw}})^2]}}{n} < \infty,$$

then

$$\frac{\text{Var} (F(\mathcal{T}_n^{\text{gw}}))}{n} \rightarrow \gamma^2,$$

where γ is a finite constant depending on f and ξ ; moreover

$$\frac{F(\mathcal{T}_n^{\text{gw}}) - n\mathbb{E} [f(\mathcal{T}^{\text{gw}})]}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2),$$

where $N(0, \gamma^2)$ denotes the normal distribution with mean 0 and variance γ^2 .

2.4.2 Size-biased Galton-Watson trees

A random variable $\hat{\xi}$ on \mathbb{N} is said to have ξ -size-biased distribution if

$$\mathbb{P} \{ \hat{\xi} = i \} = i \mathbb{P} \{ \xi = i \} = i p_i, \quad \text{for all } i \in \mathbb{N}.$$

Note that this is indeed a distribution since

$$\sum_{i=1}^{\infty} \mathbb{P} \{ \hat{\xi} = i \} = \sum_{i=1}^{\infty} i p_i = \mathbb{E} \xi = 1.$$

The *size-biased* conditional Galton-Watson tree $\hat{\mathcal{T}}^{g\text{w}}$ is an infinite random tree containing two types of nodes: the *king* and the *common*. A common node reproduces a random number of children which are all common according to ξ independently. A king reproduces a random number of children according to $\hat{\xi}$. Among these children, one is chosen uniformly and independently at random to become the next king, whereas the others become common. The root node is set to be a king. Thus $\hat{\mathcal{T}}^{g\text{w}}$ contains an infinite path consists of kings called the *spine*.

Thus a king has at least one child.

A celebrated result for $\hat{\mathcal{T}}^{g\text{w}}$ is the following:

Theorem 2.4 (Kennedy [68]). *Assume Condition A. Then $\mathcal{T}_n^{g\text{w}}$ converges to $\hat{\mathcal{T}}^{g\text{w}}$ in distribution. Or equivalently, for all fixed $k \in \mathbb{N}$, $\mathcal{T}_n^{g\text{w}}$ truncated at generation k converges in distribution to $\hat{\mathcal{T}}^{g\text{w}}$ truncated at generation k .*

This theorem also extends to simply generated trees. See Theorem 7.1 of Janson [64] for its more complete version.

The size-biased Galton-Watson tree is introduced by Kesten [69]. It has been studied by many authors, e.g., Aldous [5], Aldous and Pitman [8], and Lyons, Pemantle, and Peres [76]. There are also versions of $\hat{\mathcal{T}}^{g\text{w}}$ with a finite spine, see Addario-Berry, Devroye, and Janson [2], Geiger [50], and Sagitov and Serra [88].

2.4.3 Maximum degree

Let $\xi_{\max}^n := \max\{\xi_1^n, \dots, \xi_n^n\}$. Let $\xi_{\max} := \max\{\xi_1, \dots, \xi_n\}$. Lemma 2.4 suggests that ξ_{\max}^n and ξ_{\max} should have similar distributions when n is large. The following theorem makes this precise:

Recall that ξ_1^n, \dots, ξ_n^n is the preorder degree sequence of $\mathcal{T}_n^{g\text{w}}$ and ξ_1, \dots, ξ_n are iid copies of ξ .

Theorem 2.5 (Janson [64, thm. 19.7]). *Assume Condition A. Let $(h_n)_{n \geq 1}$ be a sequence of integers.*

1. *If $n\mathbb{P}\{\xi \geq h_n\} \rightarrow \alpha$ for some $\alpha \in [0, \infty]$, then*

$$\left[\sum_{i=1}^n \mathbb{I}[\xi_i^n \geq h_n] \right] \xrightarrow{d} \text{Po}(\alpha).$$

2. $d_{\text{TV}}(\xi_{\max}^n, \xi_{\max}) \rightarrow 0$.
3. *There exists a sequence $b_n = o(\sqrt{n})$ such that $\xi_{\max}^n \leq b_n$ whp.*

Many authors have studied maximal degrees in $\mathcal{T}_n^{g\text{w}}$, e.g., Kolchin, Sevast'yanov, and Chistyakov [70], Kolchin [71], and Meir and Moon [78]. See Janson [64, sec. 19] for a comprehensive analysis of this problem in the settings of simply generated trees.

2.4.4 *Height and width*

The distance from the root of T to its last generation, which we denote by $H(T)$, is called the *height* of T . The size of the generation that contains the most number of nodes, denoted by $W(T)$, is called the *width* of T .

It is well known that $H(\mathcal{T}_n^{\text{gw}})$ and $W(\mathcal{T}_n^{\text{gw}})$ are both of the order of \sqrt{n} whp. More precisely, we have (see e.g., Kolchin [71], Aldous [5], Chassaing, Marckert, and Yor [26], Janson [61], Drmota [35]):

Theorem 2.6. *Assume Condition A. Then,*

$$\frac{H(\mathcal{T}_n^{\text{gw}})}{\sqrt{n}} \xrightarrow{d} \frac{2}{\sigma} X, \quad \text{and} \quad \frac{W(\mathcal{T}_n^{\text{gw}})}{\sqrt{n}} \xrightarrow{d} \sigma X,$$

where X denotes the distribution of the maximum of a standard Brownian excursion.

Addario-Berry, Devroye, and Janson [2] showed that $W(\mathcal{T}_n^{\text{gw}})$ and $H(\mathcal{T}_n^{\text{gw}})$ have sub-Gaussian tails:

Theorem 2.7. *Assume Condition A. There exist constants C and c depending on ξ , such that for all $x \geq 0$ and $n \geq 1$,*

$$\begin{aligned} \mathbb{P} \{H(\mathcal{T}_n^{\text{gw}}) \geq x\sqrt{n}\} &\leq Ce^{-cx^2}, \\ \mathbb{P} \{W(\mathcal{T}_n^{\text{gw}}) \geq x\sqrt{n}\} &\leq Ce^{-cx^2}. \end{aligned}$$

See also Janson [63] for the joint distribution of $(H(\mathcal{T}_n^{\text{gw}}), W(\mathcal{T}_n^{\text{gw}}))$.

2.4.5 *Continuum random tree*

Let $v_1, \dots, v_{|T|}$ be the vertices of T in preorder. Define

$$W_T(i) = \begin{cases} 0 & \text{if } i = 0, \\ \sum_{j=1}^i (\deg(v_j) - 1) & \text{if } i \in \{1, \dots, n\}. \end{cases}$$

Recall that the plot of this function can be seen as a walk on the discrete lattice $\mathbb{Z} \times \mathbb{Z}$ (see Figure 3), where \mathbb{Z} denotes the set of all integers. The walk starts from $(0, 0)$ and ends at $(n, -1)$. All the points on this walk except the last one stay in the upper-half plane. Given this walk, we can reconstruct T without ambiguity.

Thus the function $W_{\mathcal{T}_n^{\text{gw}}}$ is a random walk that encodes $\mathcal{T}_n^{\text{gw}}$. It is well-known that this walk, suitably rescaled, converges in distribution to a Brownian excursion. (A Brownian motion $(B(t))_{t \geq 0}$ is a real-valued a random process in

which increments have independent Gaussian distributions. It was introduced to model the highly irregular motion of pollen grains in water which was observed by Robert Brown in 1828. A standard Brownian excursion, denoted by $(\epsilon(t))_{t \in [0,1]}$, is $(B(t))_{t \geq 0}$ restricted to the event that $B(t) > 0$ for all $t \in (0, 1)$ and $B(0) = B(1) = 0$. It has applications in many domains, e.g., combinatorics [42, 62], algebra [85], queuing theory [57] and railway traffics [92].

Theorem 2.8 (Marckert and Mokkadem [77]). *Assume Condition A. Let $(\epsilon(t))_{t \in [0,1]}$ be a standard Brownian excursion. We have*

$$\left(\frac{W_{\mathcal{T}_n^{gw}}(\lceil nt \rceil)}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow{d} (\sigma \epsilon(t))_{t \in [0,1]}.$$

This suggests that there exists a tree-like object encoded by a Brownian excursion which is the limit of a rescaled \mathcal{T}_n^{gw} . Indeed, this object is the *continuum random tree* (CRT) introduced by Aldous [4]. Unlike trees we have discussed so far, a CRT is a *real tree*, which is defined as a metric space (\mathcal{M}, d) satisfying that for all $u, v \in \mathcal{M}$:

So a real tree is not a tree!

1. There exists a unique shortest path $[u, v]$ between u and v .
2. $[u, v]$ is the only self-avoiding path containing u and v .

(See Le Gall [73] for more on real trees.)

Let $f : [0, \infty) \rightarrow [0, \infty)$ be a continuous function of bounded support with $f(0) = 0$. We can define a distance function on $[0, \infty)$ according to f :

$$d_f(u, v) = f(u) + f(v) - 2 \inf_{u \wedge v \leq w \leq u \vee v} f(w),$$

where $u \wedge v := \min\{u, v\}$ and $u \vee v := \max\{u, v\}$. See Figure 5 for an example.

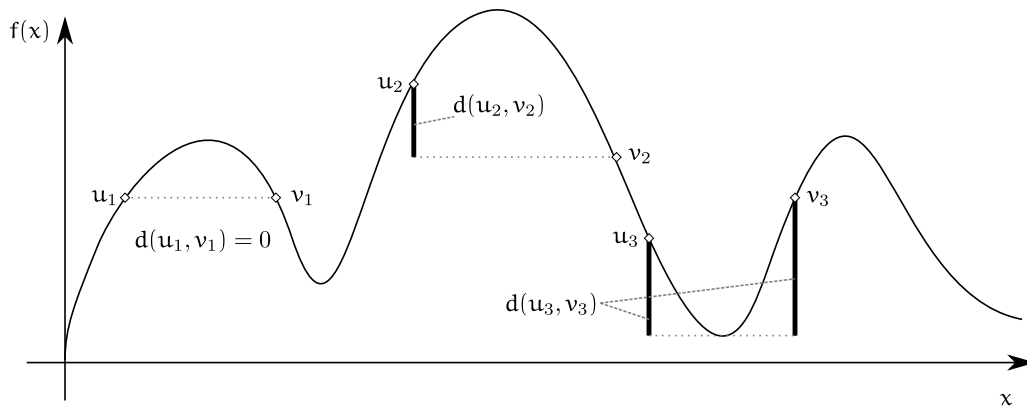


Figure 5: Example of d_f

Let \sim_f be an equivalent class on $[0, \infty)$ such that $u \sim_f v$ if and only if $d_f(u, v) = 0$. Let $\mathcal{M}_f = [0, \infty) / \sim_f$, i.e., the quotient space with respect to \sim_f . Then the metric space $\mathcal{T}_f := (\mathcal{M}_f, d_f)$ is a real tree.

*Though Aldous uses
an equivalent but
different definition.*

Replacing the function f by a standard Brownian excursion ϵ , we have a random real tree \mathcal{T}_ϵ which Aldous [4] calls a CRT. Theorem 2.8 implies that:

Theorem 2.9 (Le Gall [73, thm. 2.5]). *Assume Condition A. We have*

$$\frac{\sigma}{\sqrt{n}} \mathcal{T}_n^{g^w} \xrightarrow{d} \mathcal{T}_\epsilon,$$

where the convergence is in Gromov-Hausdorff distance introduced by Gromov [52].

So Theorem 2.6 (height and width of $\mathcal{T}_n^{g^w}$) can be seen as a result of the above convergence to CRT. See Aldous [4-6], Duquesne [36], Le Gall [73], and Marckert and Mokkadem [77] for more on CRT.

POISSON APPROXIMATION

For most part of this chapter, we introduce tools for estimating the error terms in approximating the sum of indicator random variables with a Poisson distribution. In particular, we review the exchangeable pair method and demonstrate a way to construct such pairs. But the last section also shows how to use factorial moments to prove convergence in distribution to a Poisson or a normal random variable. These results are used for analyzing fringe subtree counts in [Chapter 4](#).

3.1 TWO LEMMAS FOR POISSON DISTRIBUTION

The following lemma is a special case of Roos [86, thm. 1], which applies to mixed Poisson distributions. Barbour, Holst, and Janson proved a similar result using Stein's method [13, thm. 1.C]. We include our proof for its simplicity.

Lemma 3.1. *If $X \stackrel{\mathcal{L}}{=} \text{Po}(\mu)$ and $Y \stackrel{\mathcal{L}}{=} \text{Po}(\nu)$ for $\mu, \nu > 0$, then*

$$d_{\text{TV}}(X, Y) \leq |\sqrt{\mu} - \sqrt{\nu}| = \frac{|\mu - \nu|}{\sqrt{\mu} + \sqrt{\nu}}.$$

Proof. Let $x_i = \mathbb{P}\{X = i\}$ and $y_i = \mathbb{P}\{Y = i\}$. We have

$$\begin{aligned} d_{\text{TV}}(X, Y) &= \frac{1}{2} \sum_{i=1}^{\infty} |x_i - y_i| = \frac{1}{2} \sum_{i=1}^{\infty} |\sqrt{x_i} - \sqrt{y_i}| \times (\sqrt{x_i} + \sqrt{y_i}) \\ &\leq \frac{1}{2} \left(\sum_{i=1}^{\infty} |\sqrt{x_i} - \sqrt{y_i}|^2 \sum_{j=1}^{\infty} (\sqrt{x_j} + \sqrt{y_j})^2 \right)^{1/2} \\ &= \frac{1}{2} \left(\left(2 - \sum_{i=1}^{\infty} 2\sqrt{x_i y_i} \right) \left(2 + \sum_{j=1}^{\infty} 2\sqrt{x_j y_j} \right) \right)^{1/2} \\ &= \left(1 - \left(\sum_{i=1}^{\infty} \sqrt{x_i y_i} \right)^2 \right)^{1/2}, \end{aligned}$$

where the second step uses the Cauchy-Schwarz inequality. An easy calculation shows that

$$\sum_{i=1}^{\infty} \sqrt{x_i y_i} = \sum_{i=1}^{\infty} e^{-\frac{\mu+\nu}{2}} \frac{(\mu\nu)^{i/2}}{i!} = \exp\left(\sqrt{\mu\nu} - \frac{\mu+\nu}{2}\right) = \exp\left(-\frac{(\sqrt{\mu} - \sqrt{\nu})^2}{2}\right).$$

Thus we have

$$\begin{aligned} d_{\text{TV}}(X, Y) &\leq \sqrt{1 - \exp(-(\sqrt{\mu} - \sqrt{\nu})^2)} \\ &\leq \sqrt{(\sqrt{\mu} - \sqrt{\nu})^2} \quad (\text{by } 1 - e^{-x} \leq x) \\ &= |\sqrt{\mu} - \sqrt{\nu}|. \quad \square \end{aligned}$$

In Theorem 1.2, statements (i)–(iii) follow directly from (1.2) by the following lemma, whose simple proof we omit:

Lemma 3.2. *Let X_n be a sequence of integer-valued random variables. Let μ_n be a sequence of non-negative real numbers. Assume that $d_{\text{TV}}(X_n, \text{Po}(\mu_n)) \rightarrow 0$ as $n \rightarrow \infty$. We have:*

- (i) *If $\mu_n \rightarrow 0$, then $X_n = 0$ whp.*
- (ii) *If $\mu_n \rightarrow \mu \in (0, \infty)$, then $X_n \xrightarrow{d} \text{Po}(\mu)$.*
- (iii) *If $\mu_n \rightarrow \infty$, then $(X_n - \mu_n)/\sqrt{\mu_n} \xrightarrow{d} \text{N}(0, 1)$.*

3.2 STEIN'S METHOD AND EXCHANGEABLE PAIRS

Stein's method [91] bounds the errors in approximating the sum of random variables with a normal distribution. It was extended to Poisson distributions by Chen [28] and many others, e.g., Arratia, Goldstein, and Gordon [11] and Barbour, Holst, and Janson [13]. In Chapter 4, we use Stein's method in the form of *exchangeable pairs* to prove Theorem 1.2, which we summarize here.

Let $Z \stackrel{\mathcal{L}}{=} \text{Po}(\lambda)$. Let X be a non-negative integer-valued random variable. For $\mathcal{A} \subseteq \mathbb{N}_0$, define the function $f_{\lambda, \mathcal{A}} : \mathbb{N}_0 \rightarrow \mathbb{R}$ recursively as

$$\begin{cases} f_{\lambda, \mathcal{A}}(0) = 0, \\ f_{\lambda, \mathcal{A}}(i+1) = \frac{1}{\lambda} (if_{\lambda, \mathcal{A}}(i) + \mathbb{I}\{i \in \mathcal{A}\} - \mathbb{P}\{Z \in \mathcal{A}\}). \end{cases}$$

Therefore, for all $i \in \mathbb{N}_0$,

$$\lambda f_{\lambda, \mathcal{A}}(i+1) - if_{\lambda, \mathcal{A}}(i) = \mathbb{I}\{i \in \mathcal{A}\} - \mathbb{P}\{Z \in \mathcal{A}\}.$$

Replacing i with X and taking expectation, we have

$$\mathbb{P}\{X \in \mathcal{A}\} - \mathbb{P}\{Z \in \mathcal{A}\} = \mathbb{E}[\lambda f_{\lambda, \mathcal{A}}(X+1) - X f_{\lambda, \mathcal{A}}(X)].$$

Therefore

$$\begin{aligned} d_{\text{TV}}(X, Z) &= \sup_{\mathcal{A} \subseteq \mathbb{N}_0} |\mathbb{P}\{X \in \mathcal{A}\} - \mathbb{P}\{Z \in \mathcal{A}\}| \\ &= \sup_{\mathcal{A} \subseteq \mathbb{N}_0} |\mathbb{E}[\lambda f_{\lambda, \mathcal{A}}(X+1) - X f_{\lambda, \mathcal{A}}(X)]|. \end{aligned}$$

The function $f_{\lambda, \mathcal{A}}$ has the following nice properties (see Barbour, Holst, and Janson [13, lem. 1.1.1] for a proof):

Lemma 3.3. *Let $f_{\lambda, \mathcal{A}}$ be defined as above. Then for all $\mathcal{A} \subseteq \mathbb{N}_0$, we have*

$$\|f_{\lambda, \mathcal{A}}\| := \max_{i \in \mathbb{N}_0} |f_{\lambda, \mathcal{A}}(i)| \leq \lambda^{-1/2},$$

and

$$\|\Delta f_{\lambda, \mathcal{A}}\| := \max_{i \in \mathbb{N}_0} |f_{\lambda, \mathcal{A}}(i+1) - f_{\lambda, \mathcal{A}}(i)| \leq \lambda^{-1}.$$

A pair of random variables (W, W') is called an *exchangeable pair* if (W, W') has the same joint distribution as (W', W) .

Lemma 3.4 (Ross [87, thm. 4.37]). *Let (W, W') be an exchangeable pair of non-negative integer-valued random variable with $\mathbb{E}W = \lambda$. Let \mathcal{F} be a σ -algebra such that W is \mathcal{F} -measurable. There exists a function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ with*

$$\|f\| \leq \lambda^{-1/2}, \quad \|\Delta f\| \leq \lambda^{-1}, \quad (3.1)$$

such that for all $c \in \mathbb{R}$,

$$\begin{aligned} d_{\text{TV}}(W, \text{Po}(\lambda)) &= \left| \mathbb{E} \left[(\lambda - c \mathbb{P}\{W' = W + 1 | \mathcal{F}\}) f(W + 1) \right] \right. \\ &\quad \left. - \mathbb{E} \left[(W - c \mathbb{P}\{W' = W - 1 | \mathcal{F}\}) f(W) \right] \right|. \end{aligned} \quad (3.2)$$

Example 3.1. Let $(\tilde{I}_i)_{i \geq 1}$ and $(\hat{I}_i)_{i \geq 1}$ be iid $\text{Be}(p)$ random variables. Let $W = \sum_{i=1}^n \tilde{I}_i$. Let $W' = W - \tilde{I}_Z + \hat{I}_Z$ where $Z \stackrel{\mathcal{L}}{=} \text{Unif}(\{1, \dots, n\})$. Thus (W, W') is an exchangeable pair and we have

$$\mathbb{P}\{W' = W + 1 \mid \sigma((\tilde{I}_i)_{i \geq 1})\} = \frac{1}{n} \sum_{i=1}^n (1 - \tilde{I}_i)p,$$

and

$$\mathbb{P}\{W' = W - 1 \mid \sigma((\tilde{I}_i)_{i \geq 1})\} = \frac{1}{n} \sum_{i=1}^n \tilde{I}_i(1 - p).$$

By taking $c = n$, it follows from Lemma 3.1 that there exists $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ satisfying (3.1) with $\lambda = \mathbb{E}W = np$, such that

$$\begin{aligned} &d_{\text{TV}}(W, \text{Po}(\lambda)) \\ &= \left| \mathbb{E} \left[\left(np - \sum_{i=1}^n (1 - \tilde{I}_i)p \right) f(W + 1) \right] - \mathbb{E} \left[\left(W - \sum_{i=1}^n \tilde{I}_i(1 - p) \right) f(W) \right] \right| \\ &= \left| \mathbb{E}[Wp f(W + 1)] - \mathbb{E}[Wp f(W)] \right| \\ &\leq np^2 \|\Delta f\| \leq np^2 \frac{1}{np} = p. \end{aligned}$$

Thus $d_{\text{TV}}(W, \text{Po}(\lambda)) \rightarrow 0$ as $p \rightarrow 0$.

3.3 CONSTRUCT EXCHANGEABLE PAIRS

How we apply the exchangeable pair method in [Section 4.2](#) to subtree counts is of some generality. We summarize it here in a more abstract way.

Let $(\tilde{I}_i, \tilde{J}_i)_{i=1}^n$ be a sequence of pairs of indicator random variables which is permutation invariant, i.e., permuting the sequence does not change its distribution. Let $(\hat{I}_i, \hat{J}_i)_{i=1}^n$ be an independent copy of $(\tilde{I}_i, \tilde{J}_i)_{i=1}^n$. We construct a new sequence $(\bar{I}_i)_{i=1}^n$ as follows. Let Z be a uniform random variable on $\{1, \dots, n\}$. Let

$$\bar{I}_Z = \begin{cases} \hat{I}_Z, & \text{if } \tilde{J}_Z = \hat{J}_Z = 1, \\ \tilde{I}_Z, & \text{otherwise.} \end{cases}$$

Let $\bar{I}_i = \tilde{I}_i$ for $i \in \{1, \dots, n\}$ with $i \neq Z$.

The following lemma gives a sufficient condition for $((\tilde{I}_i)_{i=1}^n, (\bar{I}_i)_{i=1}^n)$ to be an exchangeable pair:

Lemma 3.5. *If I_1 and (I_2, \dots, I_n) are independent conditioning on $[J_1 = 1]$, then $((\tilde{I}_i)_{i=1}^n, (\bar{I}_i)_{i=1}^n)$ is an exchangeable pair.*

Proof. Write event $E = [\tilde{J}_1 = \hat{J}_1 = 1]$. If E does not happen, then $(\bar{I}_i)_{i=1}^n = (\tilde{I}_i)_{i=1}^n$. Let $E' = [Z = 1] \cap E$. Since the sequences that we are considering are all permutation invariant, it suffices to show that for an arbitrary binary sequence $(a_1, \dots, a_n) \in \{0, 1\}^n$, we have

$$\begin{aligned} & \mathbb{P} \left\{ (\tilde{I}_1, \dots, \tilde{I}_n) = (a_1, \dots, a_n), (\bar{I}_1, \dots, \bar{I}_n) = (1 - a_1, a_2, \dots, a_n) \mid E' \right\} \\ &= \mathbb{P} \left\{ (\tilde{I}_1, \dots, \tilde{I}_n) = (a_1, \dots, a_n) \mid E' \right\} \mathbb{P} \left\{ \hat{I}_1 = 1 - a_1 \mid E' \right\} \\ &= \mathbb{P} \left\{ \tilde{I}_1 = a_1 \mid E' \right\} \mathbb{P} \left\{ \tilde{I}_2 = a_2 \dots \tilde{I}_n = a_n \mid E' \right\} \mathbb{P} \left\{ \hat{I}_1 = 1 - a_1 \mid E' \right\} \\ &= \mathbb{P} \left\{ \tilde{I}_1 = 1 - a_1 \mid E' \right\} \mathbb{P} \left\{ \tilde{I}_2 = a_2 \dots \tilde{I}_n = a_n \mid E' \right\} \mathbb{P} \left\{ \hat{I}_1 = a_1 \mid E' \right\} \\ &= \mathbb{P} \left\{ (\tilde{I}_1, \dots, \tilde{I}_n) = (1 - a_1, a_2, \dots, a_n) \mid E' \right\} \mathbb{P} \left\{ \hat{I}_1 = a_1 \mid E' \right\} \\ &= \mathbb{P} \left\{ (\tilde{I}_1, \dots, \tilde{I}_n) = (1 - a_1, a_2, \dots, a_n), (\bar{I}_1, \dots, \bar{I}_n) = (a_1, \dots, a_n) \mid E' \right\}. \quad \square \end{aligned}$$

Let $\Gamma_n = \sum_{i=1}^n \tilde{I}_i$ and $\Psi_n = \sum_{i=1}^n \tilde{J}_i$. Let $\bar{\Gamma}_n = \sum_{i=1}^n \bar{I}_i$. We have:

Lemma 3.6. *Assume that $(\Gamma_n, \bar{\Gamma}_n)$ is an exchangeable pair and that $\tilde{I}_1 = 1$ implies $\tilde{J}_1 = 1$. We have*

$$d_{\text{TV}} \left(\Gamma_n, \text{Po}(n\mathbb{E}\tilde{I}_1) \right) \leq \frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\tilde{J}_1} + \left[\frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\tilde{J}_1} \frac{\text{Var}(\Psi_n)}{\mathbb{E}[\Psi_n]} \right]^{1/2}.$$

Example 3.2. In the case that \tilde{I}_i 's are iid $\text{Be}(p)$ and $\tilde{J}_i = 1$ for all $1 \leq i \leq n$, we have $\Gamma_n \stackrel{\mathcal{L}}{=} \text{Bi}(n, p)$ and $\Psi_n = n$. The above lemma shows that

$$d_{\text{TV}}(\text{Bi}(n, p), \text{Po}(np)) \leq \frac{p}{1} + \left[\frac{p \times 0}{1 \times n} \right]^{1/2} = p,$$

which is the same as the bound given in Example 3.1.

Proof of Lemma 3.6. Let $r_n := \mathbb{E}\tilde{I}_1 = \mathbb{E}\hat{I}_1$ and $q_n := \mathbb{E}\tilde{J}_1 = \mathbb{E}\hat{J}_1$. Let $\mathcal{F} = \sigma((\tilde{I}_i, \tilde{J}_i)_{i=1}^n)$, i.e., the sigma algebra generated by $(\tilde{I}_i, \tilde{J}_i)_{i=1}^n$. Since the event $\bar{\Gamma}_n = \Gamma_n + 1$ happens if and only if $\tilde{J}_Z = \hat{J}_Z = 1$, $\tilde{I}_Z = 0$ and $\hat{I}_Z = 1$, we have

$$\begin{aligned} \mathbb{P}\{\bar{\Gamma}_n = \Gamma_n + 1 | \mathcal{F}\} &= \mathbb{P}\left\{ \bigcup_{i=1}^n \left[Z = i, \tilde{J}_i = 1, \tilde{I}_i = 0, \hat{J}_i = 1, \hat{I}_i = 1 \right] \middle| \mathcal{F} \right\} \\ &= \sum_{i=1}^n \mathbb{P}\{Z = i\} \mathbb{P}\left\{ \tilde{J}_i = 1, \tilde{I}_i = 0 \middle| \mathcal{F} \right\} \mathbb{P}\left\{ \hat{I}_i = 1 \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \tilde{J}_i (1 - \tilde{I}_i) r_n = \frac{r_n}{n} \sum_{i=1}^n (\tilde{J}_i - \tilde{I}_i), \end{aligned}$$

where we use that $[\tilde{I}_i = 1] \subseteq [\tilde{J}_i = 1]$ and $[\hat{I}_i = 1] \subseteq [\hat{J}_i = 1]$.

Let $c = n/\mathbb{E}\tilde{J}_1 = n/q_n$. Note that $\sum_{i=1}^n \mathbb{E}\tilde{I}_i = \mathbb{E}\Gamma_n = nr_n$, and that $\sum_{i=1}^n \mathbb{E}\tilde{J}_i = \mathbb{E}\Psi_n = nq_n$. Thus

$$\begin{aligned} \zeta_1 &:= \mathbb{E}\Gamma_n - c\mathbb{P}\{\bar{\Gamma}_n = \Gamma_n + 1 | \mathcal{F}\} \\ &= \mathbb{E}\Gamma_n - \frac{n}{q_n} \frac{r_n}{n} \sum_{i=1}^n (\tilde{J}_i - \mathbb{E}[\tilde{J}_i] + \mathbb{E}[\tilde{J}_i] - \tilde{I}_i) \\ &= nr_n - \frac{r_n}{q_n} nq_n + \frac{r_n}{q_n} \Gamma_n - \frac{r_n}{q_n} \sum_{i=1}^n (\tilde{J}_i - \mathbb{E}\tilde{J}_i) \\ &= \frac{r_n}{q_n} \Gamma_n + \frac{r_n}{q_n} (\mathbb{E}[\Psi_n] - \Psi_n). \end{aligned}$$

The event $\bar{\Gamma}_n = \Gamma_n - 1$ happens if and only if $\tilde{J}_Z = \hat{J}_Z = 1$, $\tilde{I}_Z = 1$ and $\hat{I}_Z = 0$. Therefore

$$\begin{aligned} \mathbb{P}\{\bar{\Gamma}_n = \Gamma_n - 1 | \mathcal{F}\} &= \mathbb{P}\left\{ \bigcup_{i=1}^n \left[Z = i, \tilde{J}_i = 1, \tilde{I}_i = 1, \hat{J}_i = 1, \hat{I}_i = 0 \right] \middle| \mathcal{F} \right\} \\ &= \sum_{i=1}^n \mathbb{P}\{Z = i\} \mathbb{P}\left\{ \tilde{J}_i = 1, \tilde{I}_i = 1 \middle| \mathcal{F} \right\} \mathbb{P}\left\{ \hat{J}_i = 1, \hat{I}_i = 0 \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \tilde{I}_i \left(\mathbb{P}\left\{ \hat{J}_i = 1 \right\} - \mathbb{P}\left\{ \hat{I}_i = 1 \right\} \right) = \frac{q_n}{n} \Gamma_n - \frac{r_n}{n} \Gamma_n, \end{aligned}$$

where we again use that $[\tilde{I}_i = 1] \subseteq [\tilde{J}_i = 1]$ and $[\hat{I}_i = 1] \subseteq [\hat{J}_i = 1]$. Therefore

$$\zeta_2 := \Gamma_n - c\mathbb{P}\{\bar{\Gamma}_n = \Gamma_n - 1 | \mathcal{F}\} = \Gamma_n - \frac{n}{q_n} \left(\frac{q_n}{n} \Gamma_n - \frac{r_n}{n} \Gamma_n \right) = \frac{r_n}{q_n} \Gamma_n.$$

It follows from Lemma 3.4 that there exists a function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ with $\|\Delta f\| \leq \mathbb{E}[\Gamma_n]^{-1} = (nr_n)^{-1}$ and $\|f\| \leq \mathbb{E}[\Gamma_n]^{-1/2} = (nr_n)^{-1/2}$, such that

$$\begin{aligned}
d_{\text{TV}}(\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) &= |\mathbb{E}[\zeta_1 f(\Gamma_n + 1) - \zeta_2 f(\Gamma_n)]| \\
&= \left| \mathbb{E} \left[\left(\frac{r_n}{q_n} \Gamma_n + \frac{r_n}{q_n} (\mathbb{E}[\Psi_n] - \Psi_n) \right) f(\Gamma_n + 1) - \frac{r_n}{q_n} \Gamma_n f(\Gamma_n) \right] \right| \\
&= \left| \mathbb{E} \left[\frac{r_n}{q_n} \Gamma_n (f(\Gamma_n + 1) - f(\Gamma_n)) + f(\Gamma_n + 1) \frac{r_n}{q_n} (\mathbb{E}[\Psi_n] - \Psi_n) \right] \right| \\
&\leq \frac{r_n}{q_n} \mathbb{E}[\Gamma_n \cdot |f(\Gamma_n + 1) - f(\Gamma_n)|] \\
&\quad + \frac{r_n}{q_n} \mathbb{E}[|f(\Gamma_n + 1)| \cdot |\mathbb{E}[\Psi_n] - \Psi_n|] \\
&\leq \frac{r_n}{q_n} \|\Delta f\| \mathbb{E}[\Gamma_n] + \frac{r_n}{q_n} \|f\| \mathbb{E}[|\mathbb{E}[\Psi_n] - \Psi_n|] \\
&\leq \frac{r_n}{q_n} \frac{1}{nr_n} nr_n + \frac{r_n}{q_n} \frac{1}{\sqrt{nr_n}} (\text{Var}(\Psi_n))^{1/2} \\
&= \frac{r_n}{q_n} + \left[\frac{r_n}{q_n} \frac{\text{Var}(\Psi_n)}{\mathbb{E}\Psi_n} \right]^{1/2}. \quad \square
\end{aligned}$$

3.4 LOWER BOUNDS

While Lemma 3.1 gives us upper bounds of errors in Poisson approximations, it is also useful to know how good these bounds are.

Lemma 3.7 (Barbour, Holst, and Janson [13, thm. 3.A]). *Let W be an integer-valued non-negative random variable with $\mathbb{E}W \geq 1$ and $\text{Var}(W) < \mathbb{E}W$. Let*

$$\Delta = 1 - \frac{\text{Var}(W)}{\mathbb{E}W}.$$

There exists a constant c such that for all such W we have

$$d_{\text{TV}}(W, \text{Po}(\mathbb{E}W)) \geq \frac{c\Delta}{1 + \log \frac{1}{\Delta}} \left\{ 1 \wedge \frac{\mathbb{E}W}{1 + \log \frac{1}{\Delta}} \right\}.$$

Example 3.3. Let $W \stackrel{\mathcal{L}}{=} \text{Bi}(n, p)$. We have $\mathbb{E}W = np$, $\text{Var}(X) = np(1 - p)$ and

$$\Delta = 1 - \frac{\text{Var}(W)}{\mathbb{E}W} = p.$$

Thus by Lemma 3.7,

$$d_{\text{TV}}(W, \text{Po}(W)) \geq \frac{cp}{1 + \log \frac{1}{p}} \left[1 \wedge \frac{np}{1 + \log \frac{1}{p}} \right].$$

Therefore, $d_{\text{TV}}(W, \text{Po}(W))$ does not vanish as long as p stays away from 0.

Note that the condition $\mathbb{E}W > \text{Var}(W)$ in the above lemma is necessary. On the other hand, we have the following lemma for the case $\mathbb{E}W \leq \text{Var}(W)$:

Lemma 3.8 (Barbour, Holst, and Janson [13, thm. 3.B]). *Let W be an integer-valued non-negative random variable with $\mathbb{E}W \geq 1$ and $\mathbb{E}W \leq \text{Var}(W)$. If for some $r > 2$,*

$$\mathbb{E}[|W - \mathbb{E}W|^r] < \infty.$$

Then

$$d_{\text{TV}}(W, \text{Po}(\mathbb{E}W))^{r-2} \geq \frac{(\text{Var}(W) - \mathbb{E}W)^r}{\mathbb{E}[|W - \mathbb{E}W|^r]^2}.$$

3.5 FACTORIAL MOMENTS

Let $(x)_r := x(x-1)\cdots(x-r+1)$. For a random variable X , $\mathbb{E}[(X)_r]$ is called the r -th factorial moment of X . In this section, we introduce a method to prove X converges in distribution to a Poisson or a normal random variable through computing its factorial moments.

Factorial moments is particularly convenient for dealing with sums of indicator variables because of the following lemma (see [93, thm. 2.5]):

Lemma 3.9. *Let $X = \sum_{i=1}^n I_i$ where I_1, \dots, I_n are indicator random variables. Then for all $r \in \mathbb{N}$, we have*

$$\mathbb{E}[(X)_r] = \sum_{i_1, i_2, \dots, i_r} \mathbb{P}\{I_{i_1} = I_{i_2} = \dots = I_{i_r} = 1\},$$

where sum is over all choices of r different indices i_1, \dots, i_r from $\{1, \dots, n\}$.

Convergence in moments does not always imply convergence in distribution. But for Poisson distribution, we have the following well-known result (see [93, thm. 2.4]):

Lemma 3.10. *Let $(X_n)_{n \geq 1}$ be a sequence of integer-valued random variables. If for all fixed $r \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n)_r \rightarrow \lambda^r,$$

with the sum $\lambda \in [0, \infty)$, then $X_n \xrightarrow{d} \text{Po}(\lambda)$.

Remark 3.1. The advantage of using Stein's method is that often we only need to compute the first and second moments. However, there exists an easy formula for computing the factorial moments of the sum of indicator random variables [93, thm. 2.5], which can make the above method more straightforward than Stein's method.

For normal distribution, we have:

Lemma 3.11 (Gao and Wormald [49, thm. 1]). *Assume that $\mu_n \rightarrow \infty$. Let $(X_n)_{n \geq 1}$ be a sequence of integer-valued random variables. If*

$$\sup_{r \leq \sqrt{\mu_n}} \left| \frac{\mathbb{E}(X_n)_r}{\mu_n^r} - 1 \right| \rightarrow 0,$$

then we have

$$\frac{X_n - \mu_n}{\sqrt{\mu_n}} \xrightarrow{d} N(0, 1).$$

Example 3.4. Let $X_n = \sum_{i=1}^n I_i$ where $(I_i)_{i \geq 0}$ are iid $\text{Be}(p_n)$, i.e., $X_n \stackrel{\mathcal{L}}{=} \text{Bi}(n, p_n)$.

Then by Lemma 3.9, we have

$$\mathbb{E}(X_n)_r = \sum_{i_1, \dots, i_r} \mathbb{P}\{I_{i_1} = I_{i_2} = \dots = I_{i_r}\} = (n)_r p_n^r.$$

Thus if $p_n = \lambda/n$ for some constant λ , then $\mathbb{E}(X_n)_r \rightarrow \lambda^r$. It follows from Lemma 3.10, $X_n \xrightarrow{d} \text{Po}(\lambda)$.

On the other hand, if $p_n \rightarrow 0$ and $np_n \rightarrow \infty$, then for all $r \leq \sqrt{np_n}$, we have

$$(np_n)^r \geq \mathbb{E}(X)_r \geq (np_n)^r \left(1 - \frac{r}{n}\right)^r \geq (np_n)^r \left(1 - \frac{r^2}{n}\right) \geq (np_n)^r (1 - p_n),$$

In other words, $\mathbb{E}(X)_r / (np_n)^r \rightarrow 1$. Thus by Lemma 3.11, we have

$$\frac{X_n - np_n}{\sqrt{np_n}} \xrightarrow{d} N(0, 1).$$

Part II

THE PROOF

CHAPTER 4 FRINGE SUBTREE COUNTS

We prove our main results on fringe-subtree counts, i.e., Theorem 1.2 and 1.3.

CHAPTER 5 FAMILIES OF FRINGE SUBTREES

We find the maximal K_n such that every tree of size at most K_n appears as fringe subtree whp. We also determine the height of the maximal complete r -ary fringe subtree.

CHAPTER 6 NON-FRINGE SUBTREES

We prove Theorem 1.4, i.e., the law of large number for non-fringe subtrees.

FRINGE SUBTREE COUNTS

In this chapter we prove Theorem 1.2 and 1.3. For the reader's convenience, we restate Theorem 1.2:

Theorem 1.2. *Assume Condition A. Let $k_n = o(n)$ and $k_n \rightarrow \infty$. Then*

$$\lim_{n \rightarrow \infty} \sup_{T: |T|=k_n} d_{TV}(N_T(\mathcal{T}_n^{gw}), Po(n\pi(T))) = 0. \quad (1.2)$$

Therefore, letting $(T_n)_{n \geq 1}$ be a sequence of trees with $|T_n| = k_n$, we have as $n \rightarrow \infty$:

- (i) If $n\pi(T_n) \rightarrow 0$, then $N_{T_n}(\mathcal{T}_n^{gw}) = 0$ whp.
- (ii) If $n\pi(T_n) \rightarrow \mu \in (0, \infty)$, then $N_{T_n}(\mathcal{T}_n^{gw}) \xrightarrow{d} Po(\mu)$.
- (iii) If $n\pi(T_n) \rightarrow \infty$, then

$$\frac{N_{T_n}(\mathcal{T}_n^{gw}) - n\pi(T_n)}{\sqrt{n\pi(T_n)}} \xrightarrow{d} N(0, 1).$$

Let \mathfrak{G} be a set of trees. Recall that

$$N_{\mathfrak{G}}(\mathcal{T}_n^{gw}) := \sum_{v \in \mathcal{T}_n^{gw}} \mathbb{I}[\mathcal{T}_{n,v}^{gw} \in \mathfrak{G}],$$

i.e., $N_{\mathfrak{G}}(\mathcal{T}_n^{gw})$ is the number of fringe subtrees in \mathcal{T}_n^{gw} whose shapes belong to \mathfrak{G} . Also recall that $\pi(\mathfrak{G}) := \mathbb{P}\{\mathcal{T}^{gw} \in \mathfrak{G}\}$. The major part of this chapter proves the following lemma from which Theorem 1.2 follows immediately:

Lemma 4.1. *Assume Condition A. Let $k = k_n = o(n)$ and $k \rightarrow \infty$. We have*

$$\sup_{\mathfrak{G} \subseteq \mathfrak{T}_k} \frac{d_{TV}(N_{\mathfrak{G}}(\mathcal{T}_n^{gw}), Po(n\pi(\mathfrak{G})))}{\pi(\mathfrak{G})/\pi(\mathfrak{T}_k) + \sqrt{\pi(\mathfrak{G})/\pi(\mathfrak{T}_k)}} \leq 1 + o(k^{-3/2}) + O\left(\frac{k^{1/4}}{\sqrt{n}}\right).$$

The chapter is organized as follows: Section 4.1 estimates the expectation, variance and factorial moments of fringe subtree counts. Section 4.2 applies the exchangeable pair method to finish the proof of Lemma 4.1 as well as Theorem 1.2. In the end, Section 4.3 proves Theorem 1.3.

Remark 4.1. Every lemma in this chapter assumes Condition A. We do not repeat it every time for the sake of brevity.

Recall that $N_T(\mathcal{T}_n^{gw})$ is the number of fringe subtrees of shape T and that $\pi(T) := \mathbb{P}\{\mathcal{T}^{gw} = T\}$.

(i)-(iii) are simple corollaries of (1.2). See Lemma 3.2.

Recall that \mathfrak{T}_k is the set of all trees of size k .

Remark 4.2. If $\pi(T) = 0$, then deterministically $N_T(\mathcal{T}_n^{\text{gw}}) = 0$ and

$$d_{\text{TV}}(N_T(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(T))) = 0.$$

Thus we can assume without loss of generality that

$$\mathbb{P}\{\mathcal{T}^{\text{gw}} \in \mathfrak{T}_{k_n}\} = \mathbb{P}\{|\mathcal{T}^{\text{gw}}| = k_n\} > 0$$

for all n , and that the supremum in (1.2) is taken over all $T \in \mathfrak{T}_{k_n}$ with $\pi(T) > 0$.

4.1 MOMENTS OF FRINGE SUBTREE COUNTS

Let $\mathfrak{S} \in \mathfrak{T}_k$ for some $k \in \mathbb{N}$. Let $\mathcal{D}(\mathfrak{S})$ be the set of preorder degree sequences of trees in \mathfrak{S} . Recall that $(\tilde{\xi}_i^n)_{i=1}^n := (\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ is a uniform random rotation of $(\xi_i^n)_{i=1}^n$. Throughout this chapter, let

$$\tilde{I}_i := \mathbb{1}[(\tilde{\xi}_i^n, \tilde{\xi}_{i+1}^n, \dots, \tilde{\xi}_{i+k-1}^n) \in \mathcal{D}(\mathfrak{S})]. \quad (4.1)$$

See Lemma 2.6 for a similar argument.

Then $N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}}) = \sum_{i=1}^n \tilde{I}_i$. We use \tilde{I}_i to compute the moments of $N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}})$.

4.1.1 Factorial moments

The next lemma can be seen as a generalization of Lemma 2.6:

Lemma 4.2. Let $\mathfrak{S} \subseteq \mathfrak{T}_k$. Let $(\tilde{I}_i)_{i=1}^n$ be as in (4.1). For $k, r \in \mathbb{N}$ with $kr \leq n$, we have

$$\mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} = \dots = \tilde{I}_{k(r-1)+1} = 1\} = \pi(\mathfrak{S})^r \frac{\mathbb{P}\{S_{n-kr} = n - r(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}}$$

Recall that

ξ_1, \dots, ξ_n are iid copies of ξ , and

$$S_n = \sum_{i=1}^n \xi_n.$$

Proof. We have

$$\begin{aligned} & \mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} = \dots = \tilde{I}_{k(r-1)+1} = 1\} \\ &= \mathbb{P}\left\{\bigcap_{j=1}^r \left[(\tilde{\xi}_{k(j-1)+1}^n, \dots, \tilde{\xi}_{kj}^n) \in \mathcal{D}(\mathfrak{S})\right]\right\} \\ &= \mathbb{P}\left\{\bigcap_{j=1}^r \left[(\xi_{k(j-1)+1}, \dots, \xi_{kj}) \in \mathcal{D}(\mathfrak{S})\right] \mid S_n = n - 1\right\} \\ &= \frac{\mathbb{P}\left\{\bigcap_{j=1}^r \left[(\xi_{k(j-1)+1}, \dots, \xi_{kj}) \in \mathcal{D}(\mathfrak{S})\right] \cap [S_n = n - 1]\right\}}{\mathbb{P}\{S_n = n - 1\}} \\ &= \prod_{j=1}^r \mathbb{P}\left\{(\xi_{k(j-1)+1}, \dots, \xi_{kj}) \in \mathcal{D}(\mathfrak{S})\right\} \frac{\mathbb{P}\{S_{n-kr} = (n-1) - r(k-1)\}}{\mathbb{P}\{S_n = n - 1\}} \\ &= \pi(\mathfrak{S})^r \frac{\mathbb{P}\{S_{n-kr} = n - r(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}}, \end{aligned}$$

where the second last equality uses that the total degree of a tree of size k is $k - 1$. \square

Lemma 4.3. Consider n balls labeled $1, \dots, n$ arranged in a cycle. Let $k, r \in \mathbb{N}$ such that $kr \leq n$. The number of choices of r non-overlapping segments of length k from this cycles is

$$n(n - r(k - 1) - 1)_{r-1}.$$

Recall that $(x)_r = \prod_{i=0}^{r-1} (x - i)$.

Proof. We first decide the relative positions of the k segments. This equals the number of ways to arrange r distinct black balls separated by $n - rk$ distinct white balls in a cycle, with the first black ball always occupying a fixed position. There are in total $n - r(k - 1)$ positions. So there are $n - r(k - 1) - 1$ choices of position for the second black ball, $n - r(k - 1) - 2$ choices of position for the third black ball, and so on. In other words, there are $(n - r(k - 1) - 1)_{r-1}$ choices of the relative positions of the k segments. And once this is decided, we have n choices for the starting index of the first segment. So in total there are $n(n - r(k - 1) - 1)_{r-1}$ ways to select these segments. \square

Lemma 4.4. Let $\mathfrak{S} \subseteq \mathfrak{T}_k$. Then for all $r \in \mathbb{N}$ with $kr \leq n$,

$$\mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{gw}))_r = n(n - r(k - 1) - 1)_{r-1} \pi(\mathfrak{S})^r \frac{\mathbb{P}\{S_{n-kr} = n - r(k - 1) - 1\}}{\mathbb{P}\{S_n = n - 1\}}.$$

Proof. Let $(\tilde{I}_i)_{i=1}^n$ be as (4.1). Then $N_{\mathfrak{S}}(\mathcal{T}_n^{gw}) = \sum_{i=1}^n \tilde{I}_i$. By Lemma 3.9, for $rk \leq n$, we have

$$\mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{gw}))_r = \sum_{i_1, \dots, i_r} \mathbb{P}\{\tilde{I}_{i_1} = \dots = \tilde{I}_{i_r} = 1\},$$

where the sum is over all choices of n distinct indices i_1, \dots, i_r from $\{1, \dots, n\}$. Since preorder degree sequences of two fringe subtrees of the same size cannot overlap, if there exist two different indices $s, t \in \{i_1, \dots, i_r\}$ such that

$$\{s, s + 1, \dots, s + k - 1\} \cap \{t, t + 1, \dots, t + k - 1\} \neq \emptyset,$$

where all the numbers are taken modulo n , then $\mathbb{P}\{\tilde{I}_{i_1} = \dots = \tilde{I}_{i_r} = 1\} = 0$. Thus we only need to consider i_1, \dots, i_r without such overlapping indices.

By Lemma 4.3, there are $n(n - r(k - 1) - 1)_{r-1}$ ways to choose such i_1, \dots, i_r . And since $(\tilde{I}_i)_{i=1}^n$ is permutation invariant, for all these choices we have

$$\mathbb{P}\{\tilde{I}_{i_1} = \tilde{I}_{i_2} = \dots = \tilde{I}_{i_r} = 1\} = \mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} \dots = \tilde{I}_{k(r-1)+1} = 1\}.$$

Therefore,

$$\begin{aligned} \mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{gw}))_r &= n(n - r(k - 1) - 1)_{r-1} \mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} \dots = \tilde{I}_{k(r-1)+1} = 1\} \\ &= n(n - r(k - 1) - 1)_{r-1} \pi(\mathfrak{S})^r \frac{\mathbb{P}\{S_{n-kr} = n - r(k - 1) - 1\}}{\mathbb{P}\{S_n = n - 1\}}, \end{aligned}$$

where the last equality follows from Lemma 4.2. \square

Lemma 4.5. *Let $r = r_n = o(\sqrt{n})$ and $k = k_n = o(n/r_n^2)$. We have*

$$\sup_{\mathfrak{G} \subseteq \mathfrak{I}_k} \sup_{s \leq r} \left| \frac{\mathbb{E}(N_{\mathfrak{G}}(\mathcal{T}_n^{\text{gw}}))_s}{(n\pi(\mathfrak{G}))^s} - 1 \right| = o(1).$$

Proof. Since $kr^2 = o(n)$, for $s \leq r$ we have

$$n(n - s(k-1) - 1)_{s-1} = n^s \left(1 + O\left(\frac{r^2 k}{n}\right) \right).$$

And by Corollary 2.2, for $s \leq r$ we have

$$\frac{\mathbb{P}\{S_{n-ks} = n - s(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}} = 1 + O\left(\frac{rk}{n}\right) + o(n^{-1/2}).$$

Therefore, it follows from Lemma 4.4 that for all $s \leq r$,

$$\begin{aligned} \mathbb{E}(N_{\mathfrak{G}}(\mathcal{T}_n^{\text{gw}}))_s &= n(n - r(s-1) - 1)_{s-1} \pi(\mathfrak{G})^s \frac{\mathbb{P}\{S_{n-ks} = n - s(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}} \\ &= (n\pi(\mathfrak{G}))^s \left[1 + O\left(\frac{r^2 k}{n}\right) + o(n^{-1/2}) \right]. \end{aligned} \quad \square$$

4.1.2 Expectation and variance

Lemma 4.6. *Let $k = k_n = o(n)$. Then*

$$\sup_{\mathfrak{G} \subseteq \mathfrak{I}_k} \left| \frac{\mathbb{E}N_{\mathfrak{G}}(\mathcal{T}_n^{\text{gw}})}{n\pi(\mathfrak{G})} - 1 \right| = O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right).$$

Proof. It follows from Lemma 4.5 by taking $r = 1$. □

Lemma 4.7. *Let $(\tilde{I}_i)_{i=1}^n$ be as in (4.1). We have*

$$\begin{aligned} \text{Cov}(\tilde{I}_1, \tilde{I}_{k+1}) &= \mathbb{E}[\tilde{I}_1 \tilde{I}_{k+1}] - \mathbb{E}[\tilde{I}_1] \mathbb{E}[\tilde{I}_{k+1}] \\ &\leq \frac{\mathbb{E}\tilde{I}_1}{n} \left(o(k^{-3/2}) + O\left(\sqrt{\frac{k}{n}}\right) \right). \end{aligned}$$

Proof. It follows from Lemma 2.6 that

$$\mathbb{E}\tilde{I}_1 = \pi(\mathfrak{G}) \frac{\mathbb{P}\{S_{n-k} = n - k\}}{\mathbb{P}\{S_n = n - 1\}}.$$

It follows from Lemma 4.2 by taking $r = 2$ that

$$\mathbb{E}[\tilde{I}_1 \tilde{I}_{k+1}] = \pi(\mathfrak{G})^2 \frac{\mathbb{P}\{S_{n-2k} = n - 2k + 1\}}{\mathbb{P}\{S_n = n - 1\}},$$

By part (ii) of Lemma 2.7 and that $\pi(\mathfrak{S}) \leq \pi(\mathfrak{T}_k) = \Theta(k^{-3/2})$ (Lemma 2.5), we have

$$\begin{aligned}
 & \mathbb{E} \left[\tilde{I}_1 \tilde{I}_{k+1} \right] - \mathbb{E} \left[\tilde{I}_1 \right] \mathbb{E} \left[\tilde{I}_{k+1} \right] \\
 &= \frac{\mathbb{P} \{S_{n-2k} = n - 2k + 1\}}{\mathbb{P} \{S_n = n - 1\}} - \left(\frac{\mathbb{P} \{S_{n-k} = n - k\}}{\mathbb{P} \{S_n = n - 1\}} \right)^2 \\
 &= \pi(\mathfrak{S})^2 \left[-\frac{1}{\sigma^2 n} + o\left(\frac{1}{n}\right) + O\left(\frac{k}{n^{3/2}} + \frac{k^2}{n^2}\right) \right] \\
 &\leq \frac{\pi(\mathfrak{S})}{n} \left[o(\pi(\mathfrak{T}_k)) + O\left(\frac{\pi(\mathfrak{T}_k)k}{\sqrt{n}} + \frac{\pi(\mathfrak{T}_k)k^2}{n}\right) \right] \\
 &= \frac{\pi(\mathfrak{S})}{n} \left[o(k^{-3/2}) + O\left(\frac{k^{-3/2}k}{\sqrt{n}} + \frac{k^{-3/2}k^2}{n}\right) \right] \\
 &= \frac{\mathbb{E}\tilde{I}_1}{n} \left(o(k^{-3/2}) + O\left(\sqrt{\frac{k}{n}}\right) \right). \quad \square
 \end{aligned}$$

Lemma 4.8. *Let $k = k_n = o(n)$. We have*

$$\sup_{\mathfrak{S} \subseteq \mathfrak{T}_k} \frac{\text{Var} (N_{\mathfrak{S}}(\mathcal{T}_n^{gw}))}{\mathbb{E} N_{\mathfrak{S}}(\mathcal{T}_n^{gw})} \leq 1 + o(k^{-3/2}) + O\left(\sqrt{\frac{k}{n}}\right).$$

Proof. Let $(\tilde{I}_i)_{i=1}^n$ be as in (4.1). Then $N_{\mathfrak{S}}(\mathcal{T}_n^{gw}) = \sum_{i=1}^n \tilde{I}_i$. Consider two indices $i \neq j$. If $|i - j| < k$ or $|i + n - j| < k$, then $\mathbb{E}\tilde{I}_i \tilde{I}_j = 0$. This is because the preorder degree sequences of fringe subtrees of size k cannot overlap. So for such i and j we have $\text{Cov}(\tilde{I}_i, \tilde{I}_j) \leq 0$.

On the other hand, if $|i - j| > k$ and $|i + n - j| > k$, then $\text{Cov}(\tilde{I}_i, \tilde{I}_j) = \text{Cov}(\tilde{I}_1, \tilde{I}_{k+1})$ since $(\tilde{\xi}_i^n)_{i=1}^n$ is permutation invariant. Therefore, by Lemma 4.7,

$$\begin{aligned}
 \text{Var} (N_{\mathfrak{S}}(\mathcal{T}_n^{gw})) &= \sum_{1 \leq i \neq j \leq n} \text{Cov}(\tilde{I}_i, \tilde{I}_j) + \sum_{i=1}^n \text{Var}(\tilde{I}_i) \\
 &\leq n^2 \text{Cov}(\tilde{I}_1, \tilde{I}_{k+1}) + n \mathbb{E}[\tilde{I}_1] (1 - \mathbb{E}[\tilde{I}_1]) \\
 &\leq n^2 \frac{\mathbb{E}\tilde{I}_1}{n} \left(o(k^{-3/2}) + O\left(\sqrt{\frac{k}{n}}\right) \right) + n \mathbb{E}[\tilde{I}_1] \\
 &= \left(1 + o(k^{-3/2}) + O\left(\sqrt{\frac{k}{n}}\right) \right) \mathbb{E} [N_{\mathfrak{S}}(\mathcal{T}_n^{gw})]. \quad \square
 \end{aligned}$$

4.2 PROOF OF LEMMA 4.1 AND THEOREM 1.2

In this section we first bound $d_{TV}(N_{\mathfrak{S}}(\mathcal{T}_n^{gw}), \text{Po}(\mathbb{E} N_{\mathfrak{S}}(\mathcal{T}_n^{gw})))$.

Let $\mathfrak{S} \subseteq \mathfrak{T}_k$. Recall that $\mathcal{D}(\mathfrak{S})$ denotes the set of preorder degree sequences of trees in \mathfrak{S} and that \mathfrak{T}_k is the set of all trees of size k . Let

$$\tilde{\mathfrak{J}}_i = \{(\tilde{\xi}_i^n, \tilde{\xi}_{i+1}^n, \dots, \tilde{\xi}_{i+k-1}^n) \in \mathcal{D}(\mathfrak{T}_k)\}, \quad (4.2)$$

where all the indices are modulo n . In other words, $N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}}) = \sum_{i=1}^n \tilde{J}_i$ is the number of fringe subtrees of size k .

The following observation is crucial for proving Lemma 4.1:

Lemma 4.9. *Let $(\tilde{I}_i)_{i=1}^n$ be as in (4.1) and $(\tilde{J}_i)_{i=1}^n$ be as in (4.2). Conditional on $[\tilde{J}_1 = 1]$, \tilde{I}_1 is independent from $\tilde{I}_2, \dots, \tilde{I}_n$.*

Proof. Let $E = [\tilde{J}_1 = 1]$. Since the preorder degree sequences of two fringe subtrees of size k cannot overlap, E implies that $I_j = 0$ for all j such that

$$j \in \{2, \dots, k\} \cap \{n+2-k, \dots, n\}.$$

For j out of this range, I_j is determined by $\tilde{\xi}_{k+1}^n, \dots, \tilde{\xi}_n^n$. Thus it suffices to show that conditional on E , $\tilde{\xi}_1^n, \dots, \tilde{\xi}_k^n$ and $\tilde{\xi}_{k+1}^n, \dots, \tilde{\xi}_n^n$ are independent.

Let $(d_i)_{i=1}^n$ be such that $(\tilde{\xi}_i^n)_{i=1}^n = (d_i)_{i=1}^n$ implies E . In other words,

$$(d_1, \dots, d_k) \in \mathcal{D}(\mathfrak{T}_k), \quad \text{and} \quad \sum_{i=k+1}^n d_i = n - k.$$

Since the only restriction $\tilde{J}_1 = 1$ puts on $\tilde{\xi}_{k+1}^n, \dots, \tilde{\xi}_n^n$ is that they sum up to $n - k$, we have

Recall that ξ_1, \dots, ξ_n are iid copies of ξ .

$$\mathbb{P} \left\{ (\tilde{\xi}_{k+1}^n, \dots, \tilde{\xi}_n^n) = (d_{k+1}, \dots, d_n) \mid E \right\} = \frac{\mathbb{P} \{ (\xi_{k+1}, \dots, \xi_n) = (d_{k+1}, \dots, d_n) \}}{\mathbb{P} \{ S_{n-k} = n - k \}}.$$

On the other hand,

$$\begin{aligned} & \mathbb{P} \left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_k^n) = (d_1, \dots, d_k) \mid E \right\} \\ &= \frac{\mathbb{P} \left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_k^n) = (d_1, \dots, d_k) \right\}}{\mathbb{P} \{ E \}} \\ &= \frac{\mathbb{P} \{ (\xi_1, \dots, \xi_k) = (d_1, \dots, d_k) \} \mathbb{P} \{ S_{n-k} = n - k \}}{\mathbb{P} \{ S_n = n - 1 \}} \frac{1}{\mathbb{P} \{ E \}}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{P} \left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_k^n) = (d_1, \dots, d_k) \mid E \right\} \times \mathbb{P} \left\{ (\tilde{\xi}_{k+1}^n, \dots, \tilde{\xi}_n^n) = (d_{k+1}, \dots, d_n) \mid E \right\} \\ &= \frac{\mathbb{P} \{ (\xi_1, \dots, \xi_n) = (d_1, \dots, d_n) \}}{\mathbb{P} \{ S_n = n - 1 \}} \frac{1}{\mathbb{P} \{ E \}} \\ &= \mathbb{P} \left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n) = (d_1, \dots, d_n) \mid E \right\}. \quad \square \end{aligned}$$

Lemma 4.10. *Let $k = k_n = o(n)$ and $k \rightarrow \infty$. As $n \rightarrow \infty$,*

$$\sup_{\mathfrak{G} \subseteq \mathfrak{T}_k} \frac{d_{\text{TV}}(N_{\mathfrak{G}}(\mathcal{T}_n^{\text{gw}}), \text{Po}(\mathbb{E}N_{\mathfrak{G}}(\mathcal{T}_n^{\text{gw}})))}{\pi(\mathfrak{G})/\pi(\mathfrak{T}_k) + \sqrt{\pi(\mathfrak{G})/\pi(\mathfrak{T}_k)}} \leq 1 + o(k^{-3/2}) + O(\sqrt{k/n}).$$

Proof. Let $(\tilde{I}_i)_{i=1}^n$ and $(\tilde{J}_i)_{i=1}^n$ be as in (4.1) and (4.2). We use the abbreviations $\Gamma_n = N_{\mathcal{T}}(\mathcal{T}_n^{gw})$ and $\Psi_n = N_{\mathfrak{I}_k}(\mathcal{T}_n^{gw})$. In other words $\Gamma_n = \sum_{i=1}^n \tilde{I}_i$ and $\Psi_n = \sum_{i=1}^n \tilde{J}_i$.

Let Z be a uniform random variable on $\{1, \dots, n\}$. We construct a new sequence $(\bar{I}_i)_{i=1}^n$ by letting

$$\bar{I}_Z = \begin{cases} \hat{I}_Z, & \text{if } \tilde{J}_Z = \hat{J}_Z = 1, \\ \tilde{I}_Z, & \text{otherwise,} \end{cases}$$

and $\bar{I}_i = \tilde{I}_i$ for $i \in \{1, \dots, n\}$ with $i \neq Z$.

By Lemma 4.9, conditioning on $\tilde{J}_1 = 1$, \tilde{I}_1 is independent from $\tilde{I}_2, \dots, \tilde{I}_n$. Therefore, it follows from Lemma 3.5 that $((\tilde{I}_i)_{i=1}^n, (\bar{I}_i)_{i=1}^n)$ is an exchangeable pair. Let $\bar{\Gamma}_n = \sum_{i=1}^n \bar{I}_i$. Then $(\Gamma_n, \bar{\Gamma}_n)$ is an exchangeable pair.

Since $\tilde{I}_1 = 1$ implies \tilde{J}_1 , it follows from Lemma 3.6 that

$$d_{TV}(\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) \leq \frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\tilde{J}_1} + \left[\frac{\mathbb{E}\tilde{I}_1 \text{Var}(\Psi_n)}{\mathbb{E}\tilde{J}_1 \mathbb{E}[\Psi_n]} \right]^{1/2}.$$

By Lemma 2.6, we have

$$\frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\tilde{J}_1} = \frac{\mathbb{E}\Gamma_n}{\mathbb{E}\Psi_n} = \frac{\pi(\mathfrak{S}) \frac{\mathbb{P}\{S_{n-1}=n-k\}}{\mathbb{P}\{S_{n-k}=n-k\}}}{\pi(\mathfrak{I}_k) \frac{\mathbb{P}\{S_{n-1}=n-k\}}{\mathbb{P}\{S_{n-k}=n-k\}}} = \frac{\pi(\mathfrak{S})}{\pi(\mathfrak{I}_k)}.$$

And by Lemma 4.8, we have

$$\frac{\text{Var}(\Psi_n)}{\mathbb{E}[\Psi_n]} = 1 + o(k^{-3/2}) + O(\sqrt{k/n}).$$

Therefore

$$\begin{aligned} d_{TV}(\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) &\leq \frac{\pi(\mathfrak{S})}{\pi(\mathfrak{I}_k)} + \left(\frac{\pi(\mathfrak{S})}{\pi(\mathfrak{I}_k)} \left(1 + o(k^{-3/2}) + O(\sqrt{k/n}) \right) \right)^{1/2} \\ &\leq \left(1 + o(k^{-3/2}) + O(\sqrt{k/n}) \right) \left(\frac{\pi(\mathfrak{S})}{\pi(\mathfrak{I}_k)} + \sqrt{\frac{\pi(\mathfrak{S})}{\pi(\mathfrak{I}_k)}} \right). \quad \square \end{aligned}$$

Lemma 4.11. *Let $k = k_n = o(n)$ and $k \rightarrow \infty$. We have*

$$\sup_{\mathfrak{S} \subseteq \mathfrak{I}_k} \frac{d_{TV}(\text{Po}(n\pi(\mathfrak{S})), \text{Po}(\mathbb{E}N_{\mathfrak{S}}(\mathcal{T}_n^{gw})))}{\pi(\mathfrak{S})/\pi(\mathfrak{I}_k) + \sqrt{\pi(\mathfrak{S})/\pi(\mathfrak{I}_k)}} \leq O\left(\frac{k^{1/4}}{\sqrt{n}}\right) + o(k^{-3/4}).$$

Proof. Let $\Gamma_n = N_{\mathfrak{S}}(\mathcal{T}_n^{gw})$. It follows from Lemma 3.1 and Lemma 4.6 that

$$\begin{aligned} d_{TV}(\text{Po}(n\pi(\mathfrak{S})), \text{Po}(\mathbb{E}\Gamma_n)) &\leq \frac{|n\pi(\mathfrak{S}) - \mathbb{E}\Gamma_n|}{\sqrt{n\pi(\mathfrak{S})}} \\ &= \sqrt{n\pi(\mathfrak{S})} \left| 1 - \frac{\mathbb{E}\Gamma_n}{n\pi(\mathfrak{S})} \right| \\ &= \sqrt{n\pi(\mathfrak{S})} \left(O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right) \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{d_{\text{TV}}(\text{Po}(n\pi(\mathfrak{S})), \text{Po}(\mathbb{E}[\Gamma_n]))}{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k) + \sqrt{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k)}} &\leq \frac{d_{\text{TV}}(\text{Po}(n\pi(\mathfrak{S})), \text{Po}(\mathbb{E}[\Gamma_n]))}{\sqrt{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k)}} \\ &\leq \sqrt{n\pi(\mathfrak{T}_k)} \left(O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right) \right) \\ &= O\left(\frac{k^{1/4}}{\sqrt{n}}\right) + o(k^{-3/4}), \end{aligned}$$

where the last step uses $\pi(\mathfrak{T}_k) = \Theta(k^{-3/2})$ from Lemma 2.5. \square

Proof of Lemma 4.1. Let $\Gamma_n = N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}})$. It follows from Lemma 4.11, Lemma 4.10 and the triangle inequality that

$$\begin{aligned} &\frac{d_{\text{TV}}(\Gamma_n, \text{Po}(n\pi(\mathfrak{S})))}{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k) + \sqrt{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k)}} \\ &\leq \frac{d_{\text{TV}}(\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) + d_{\text{TV}}(\text{Po}(\mathbb{E}\Gamma_n), \text{Po}(n\pi(\mathfrak{S})))}{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k) + \sqrt{\pi(\mathfrak{S})/\pi(\mathfrak{T}_k)}} \\ &\leq 1 + o(k^{-3/2}) + O(\sqrt{k/n}) + O\left(\frac{k^{1/4}}{\sqrt{n}}\right) + o(k^{-3/4}) \\ &\leq 1 + o(k^{-3/2}) + O(\sqrt{k/n}). \end{aligned} \quad \square$$

Proof of Theorem 1.2. Let $k = k_n$. Let $p_{\max} := \max_{i \geq 0} p_i$. By the assumptions of Theorem 1.2, $\text{Var}(\xi) > 0$, which implies that ξ is not a constant. Thus $p_{\max} < 1$.

For $T \in \mathfrak{T}_k$, we have

$$\pi(T) := \mathbb{P}\{\mathcal{T}^{\text{gw}} = T\} \leq p_{\max}^k.$$

Therefore, by Lemma 2.5,

$$\frac{\pi(T)}{\pi(\mathfrak{T}_k)} \leq \frac{p_{\max}^k}{\Theta(k^{-3/2})} = O(p_{\max}^k k^{3/2}).$$

It follows from Lemma 4.1 by taking $\mathfrak{S} = \{T\}$ that

$$\begin{aligned} d_{\text{TV}}(N_T(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(T))) &\leq (1 + o(1)) \left(\frac{\pi(T)}{\pi(\mathfrak{T}_k)} + \sqrt{\frac{\pi(T)}{\pi(\mathfrak{T}_k)}} \right) \\ &= O\left(\sqrt{p_{\max}^k k^{3/2}}\right) = o(1). \end{aligned}$$

Statements (i)–(iii) of the theorem follow from Lemma 3.2. \square

Remark 4.3. The proof of Theorem 1.2 uses the condition $k_n \rightarrow \infty$ multiple times. This is a necessary condition. Let $T_n = T$ with a fixed T for all n . Similar to Lemma 4.6 and 4.8, we can show that

$$\mathbb{E}N_{T_n}(\mathcal{T}_n^{\text{gw}}) = (1 + o(1))n\pi(T),$$

and

$$\text{Var}(N_{T_n}(\mathcal{T}_n^{g_w})) \leq (1 + o(1))n\pi(T)(1 - \pi(T)).$$

Therefore,

$$\liminf_{n \rightarrow \infty} \left[1 - \frac{\text{Var}(N_{T_n}(\mathcal{T}_n^{g_w}))}{\mathbb{E}N_{T_n}(\mathcal{T}_n^{g_w})} \right] \geq \pi(T) > 0.$$

So $\text{Var}(N_{T_n}(\mathcal{T}_n^{g_w})) < \mathbb{E}N_{T_n}(\mathcal{T}_n^{g_w})$ for n large and by Lemma 3.7, we have

$$\liminf_{n \rightarrow \infty} d_{TV}(N_{T_n}(\mathcal{T}_n^{g_w}), \text{Po}(\mathbb{E}N_{T_n}(\mathcal{T}_n^{g_w}))) > \frac{c\pi(T)}{1 + \log \frac{1}{\pi(T)}} > 0,$$

for some constant $c > 0$.

4.3 THE PROOF OF THEOREM 1.3

We restate Theorem 1.3 here:

Theorem 1.3. *Assume Condition A. Let $k_n = o(n)$ and $k_n \rightarrow \infty$. Let \mathfrak{S}_n be a sequence of sets of trees with $\mathfrak{S}_n \subseteq \mathfrak{T}_{k_n}$. We have:*

(i) *If $n\pi(\mathfrak{S}_n) \rightarrow 0$, then $N_{\mathfrak{S}_n}(\mathcal{T}_n^{g_w}) = 0$ whp.*

(ii) *If $n\pi(\mathfrak{S}_n) \rightarrow \mu \in (0, \infty)$, then $N_{\mathfrak{S}_n}(\mathcal{T}_n^{g_w}) \xrightarrow{d} \text{Po}(\mu)$.*

(iii) *If $n\pi(\mathfrak{S}_n) \rightarrow \infty$, then*

$$\frac{N_{\mathfrak{S}_n}(\mathcal{T}_n^{g_w}) - n\pi(\mathfrak{S}_n)}{\sqrt{n\pi(\mathfrak{S}_n)}} \xrightarrow{d} N(0, 1).$$

(iv) *If $\pi(\mathfrak{S}_n)/\pi(\mathfrak{T}_{k_n}) \rightarrow 0$, then*

$$\lim_{n \rightarrow \infty} d_{TV}(N_{\mathfrak{S}_n}(\mathcal{T}_n^{g_w}), \text{Po}(n\pi(\mathfrak{S}_n))) = 0.$$

Part (iv) follows immediately from Lemma 4.1. However, to show (i)–(iii), we instead use our estimates of the factorial moments of $N_{\mathfrak{S}_n}(\mathcal{T}_n^{g_w})$.

Proof of (i) and (ii) of Theorem 1.3. Let $k = k_n$ and $\mathfrak{S} = \mathfrak{S}_n$. For a fixed $r \in \mathbb{N}$, since $k = o(n)$, it follows from Lemma 4.5 that,

$$\frac{\mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{g_w}))_r}{(n\pi(\mathfrak{S}))^r} \rightarrow 1.$$

Thus if $n\pi(\mathfrak{S}) \rightarrow \mu \in [0, \infty)$, then $\mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{g_w}))_r \rightarrow \mu^r$. By Lemma 3.10, this implies that

$$N_{\mathfrak{S}}(\mathcal{T}_n^{g_w}) \xrightarrow{d} \text{Po}(\mu).$$

If $\mu = 0$, then $N_{\mathfrak{S}}(\mathcal{T}_n^{g_w}) = 0$ whp since it is integer-valued. \square

Proof of (iii) of Theorem 1.3. Let $k = k_n$ and $\mathfrak{S} = \mathfrak{S}_n$. Since $k \rightarrow \infty$, by Lemma 2.5,

$$\pi(\mathfrak{S}) \leq \pi(\mathfrak{T}_k) = \Theta(k^{-3/2}) = o(1).$$

Thus, letting $r = r_n = \sqrt{n\pi(\mathfrak{S})}$, we have $r = o(\sqrt{n})$. Moreover,

$$\frac{kr^2}{n} = \frac{k n \pi(\mathfrak{S})}{n} \leq k \pi(\mathfrak{T}_k) = k \Theta(k^{-3/2}) = \Theta(k^{-1/2}) = o(1).$$

It follows Lemma 4.5 that

$$\sup_{r \leq \sqrt{n\pi(\mathfrak{S})}} \left| \frac{\mathbb{E}(N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}}))_r}{(n\pi(\mathfrak{S}))^r} - 1 \right| = o(1).$$

By Lemma 3.10, we have

$$\frac{N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}}) - n\pi(\mathfrak{S})}{\sqrt{n\pi(\mathfrak{S})}} \xrightarrow{d} N(0, 1). \quad \square$$

Proof of (iv) of Theorem 1.3. Let $k = k_n$ and $\mathfrak{S} = \mathfrak{S}_n$. By Lemma 4.1, we have

$$d_{\text{TV}}(N_{\mathfrak{S}}(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(\mathfrak{S}))) \leq (1 + o(1)) \left(\frac{\pi(\mathfrak{S})}{\pi(\mathfrak{T}_k)} + \sqrt{\frac{\pi(\mathfrak{S})}{\pi(\mathfrak{T}_k)}} \right) = o(1),$$

where in the last step we use the assumption that $\pi(\mathfrak{S})/\pi(\mathfrak{T}_k) \rightarrow 0$. □

FAMILIES OF FRINGE SUBTREES

In this chapter, we apply [Theorem 1.2](#) and [1.3](#) to study the conditions for $\mathcal{T}_n^{\text{gw}}$ to contain every tree that belongs to a family of trees. [Section 5.1](#) introduces a generalized version of coupon collector problem as our main tool. [Section 5.2](#) determines the height of the largest complete r -ary fringe subtree in $\mathcal{T}_n^{\text{gw}}$. [Section 5.3](#) finds the maximal K_n such that all possible trees of size at most K_n appear in $\mathcal{T}_n^{\text{gw}}$ as fringe subtrees.

5.1 COUPON COLLECTOR PROBLEM

As shown later, our problem is essentially a variation of the famous coupon collector problem—if in every draw we get a coupon of a uniform random type among n types, how many draws do we need to collect all n types of coupons? This problem can be generalized as in the next lemma. For the original problem, see Erdős and Rényi [[40](#)] and Flajolet, Gardy, and Thimonier [[44](#)]. For more about the generalized version defined below, see Neal [[81](#)].

Lemma 5.1 (Generalized coupon collector). *Let X_n be a random variable that takes values in $\{1, \dots, n\}$. Let $p_{n,i} := \mathbb{P}\{X_n = i\}$. Assume that $p_{n,i} > 0$ for all $1 \leq i \leq n$. Let $X_{n,1}, X_{n,2}, \dots$ be i.i.d. copies of X_n . Let*

$$N_n := \inf\{i \geq 1 : |\{X_{n,1}, X_{n,2}, \dots, X_{n,i}\}| = n\}.$$

Let m_n be a positive integer. We have

$$1 - \sum_{i=1}^n (1 - p_{n,i})^{m_n} \leq \mathbb{P}\{N_n \leq m_n\} \leq \frac{1}{\sum_{i=1}^n (1 - p_{n,i})^{m_n}}.$$

Proof. Let $m = m_n$. Let $Z_{n,i} = \mathbb{I}[i \notin \{X_{n,1}, \dots, X_{n,m}\}]$. Then $N_n \leq m$ if and only if $Z_n := \sum_{i=1}^n Z_{n,i} = 0$. So $\mathbb{P}\{N_n \leq m\} = \mathbb{P}\{Z_n = 0\} = 1 - \mathbb{P}\{Z_n \geq 1\}$.

The first inequality of this lemma follows from the following:

$$\mathbb{P}\{Z_n \geq 1\} \leq \mathbb{E}Z_n = \sum_{i=1}^n \mathbb{E}Z_{n,i} = \sum_{i=1}^n \mathbb{P}\{\bigcap_{j=1}^m X_{n,j} \neq i\} = \sum_{i=1}^n (1 - p_{n,i})^m.$$

Recall that $\mathbb{I}[H]$ is the indicator that H is true.

For $1 \leq i \neq j \leq n$, we have

$$\begin{aligned} & \mathbb{E}[Z_{n,i}Z_{n,j}] - \mathbb{E}[Z_{n,i}]\mathbb{E}[Z_{n,j}] \\ &= (1 - p_{n,i} - p_{n,j})^m - (1 - p_{n,i})^m(1 - p_{n,j})^m \\ &\leq (1 - p_{n,i})^m \left[\left(1 - \frac{p_{n,j}}{1 - p_{n,i}}\right)^m - (1 - p_{n,j})^m \right] < 0. \end{aligned}$$

Therefore

$$\begin{aligned} \text{Var}(Z_n) &= \sum_{1 \leq i, j \leq n} \mathbb{E}[Z_{n,i}Z_{n,j}] - \mathbb{E}[Z_{n,i}]\mathbb{E}[Z_{n,j}] \\ &\leq \sum_{1 \leq i \leq n} \left(\mathbb{E}[Z_{n,i}^2] - \mathbb{E}[Z_{n,i}]^2 \right) \leq \mathbb{E}Z_n. \end{aligned}$$

Thus by Chebyshev's inequality, as in the second moment method (see e.g., Alon and Spencer [9, chap. 4]), we have

$$\begin{aligned} \mathbb{P}\{Z_n = 0\} &\leq \mathbb{P}\{|Z_n - \mathbb{E}Z_n| \geq \mathbb{E}Z_n\} \\ &\leq \frac{\text{Var}(Z_n)}{(\mathbb{E}Z_n)^2} \leq \frac{1}{\mathbb{E}Z_n} = \frac{1}{\sum_{i=1}^n (1 - p_{n,i})^m}. \quad \square \end{aligned}$$

5.2 COMPLETE r -ARY FRINGE SUBTREES

Recall that $\pi(T) := \pi(\mathcal{T}^{\text{gw}} = T)$.

Recall that a tree T is called possible if $\pi(T) > 0$. Let $r > 0$ be a fixed integer and h_n be a sequence of positive integers. A simple application of Theorem 1.2 is to find sufficient conditions such that whp every (or not every) possible complete r -ary tree appears in $\mathcal{T}_n^{\text{gw}}$ as fringe subtrees.

Let $\mathfrak{S}_{h,r}$ be the set of all possible complete r -ary trees of height at most h . Let

$$H_{n,r} := \max\{h : \mathcal{T}_n^{\text{gw}} \text{ contains all trees in } \mathfrak{S}_{h,r} \text{ as fringe subtrees}\}.$$

Lemma 5.2. *Assume Condition A and $p_r > 0$ for some $r \geq 2$. Let $h_n \rightarrow \infty$ be a sequence of positive integers. Let*

$$\alpha_r = \log_r \left(\log \frac{1}{p_0} + \frac{1}{r-1} \log \frac{1}{p_r} \right).$$

Let $\omega_n \rightarrow \infty$ be an arbitrary sequence.

- (i) If $h_n \leq \log_r(\log n - \omega_n) - \alpha_r$, then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{S}_{h_n,r}$ as fringe subtrees.
- (ii) If $h_n \geq \log_r(\log n + \omega_n) - \alpha_r$, then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{S}_{h_n,r}$ as fringe subtrees.

Also,

$$H_{n,r} - \log_r \log n \xrightarrow{P} -\alpha_r.$$

Proof. Let $h = h_n$. Let $T_h^{r\text{-ary}}$ denote the complete r -ary tree of height h . Note that if $T_h^{r\text{-ary}}$ appears in $\mathcal{T}_n^{\text{gw}}$ as a fringe subtree, then every tree in $\mathfrak{S}_{h,r}$ also appears in $\mathcal{T}_n^{\text{gw}}$ as a fringe subtree. The tree $T_h^{r\text{-ary}}$ has $\ell_n := r^h$ leaves and $v_n := (r^h - 1)/(r - 1) = (\ell_n - 1)/(r - 1)$ internal vertices, which all have degree r . Thus we have

$$\pi(T_h^{r\text{-ary}}) := \mathbb{P}\{\mathcal{T}_n^{\text{gw}} = T_h^{r\text{-ary}}\} = p_r^{v_n} p_0^{\ell_n}.$$

If $h \leq \log_r(\log n - \omega_n) - \alpha_r$, then

$$\ell_n = r^h \leq \frac{\log n - \omega_n}{r^{\alpha_r}}.$$

Therefore

$$\begin{aligned} \log \frac{1}{\pi(T_h^{r\text{-ary}})} &= v_n \log \frac{1}{p_r} + \ell_n \log \frac{1}{p_0} \\ &= \frac{\ell_n - 1}{r - 1} \log \frac{1}{p_r} + \ell_n \log \frac{1}{p_0} \\ &= \ell_n \left(\frac{1}{r - 1} \log \frac{1}{p_r} + \log \frac{1}{p_0} \right) + O(1) \\ &\leq \frac{\log n - \omega_n}{r^{\alpha_r}} r^{\alpha_r} + O(1) \\ &= \log n - \omega_n + O(1). \end{aligned}$$

Thus $\log(n\pi(T_h^{r\text{-ary}})) \geq \omega_n - O(1) \rightarrow \infty$, which implies that $n\pi(T_h^{r\text{-ary}}) \rightarrow \infty$.

It follows from Theorem 1.2 that $N_{T_h^{r\text{-ary}}}(\mathcal{T}_n^{\text{gw}}) \xrightarrow{P} \infty$. Thus (i) is proved.

Similar computations show that with the assumptions of (ii), $n\pi(T_h^{r\text{-ary}}) \rightarrow 0$, which implies that $N_{T_h^{r\text{-ary}}}(\mathcal{T}_n^{\text{gw}}) \xrightarrow{P} 0$ by Theorem 1.2. The last statement of the lemma follows directly from (i) and (ii). \square

We have a similar result for $\mathfrak{S}_{h,1}$, the set of 1-ary trees (chains) of height at most h . The proof is virtually identical to the previous lemma and we leave it to the reader.

Lemma 5.3. *Assume Condition A and $p_1 > 0$. Let $\omega_n \rightarrow \infty$ be an arbitrary sequence. We have:*

- (i) *If $h_n \leq (\log n - \omega_n)/\log \frac{1}{p_1}$, then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{S}_{h_n,1}$ as fringe subtrees.*
- (ii) *If $h_n \geq (\log n + \omega_n)/\log \frac{1}{p_1}$, then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{S}_{h_n,1}$ as fringe subtrees.*

Therefore

$$\frac{H_{n,1}}{\log_{1/p_1}(n)} \xrightarrow{p} 1.$$

5.2.1 Binary trees

Recall that in [Subsection 2.3.3](#), we take a \mathcal{T}_n^{gw} with $\xi \stackrel{\mathcal{L}}{=} \text{Bi}(1/2, 2)$ and attach each degree one child to either the left or the right of its parent uniformly at random. The result is a uniform random binary tree $\mathcal{T}_n^{\text{bin}}$.

We can define fringe subtree for $\mathcal{T}_n^{\text{bin}}$ in the same manner as we did for \mathcal{T}_n^{gw} . And our analysis of \mathcal{T}_n^{gw} can be easily adapted to $\mathcal{T}_n^{\text{bin}}$. For example, let T be a chain of height h . Let T^{bin} be a binary tree of height h that has only left children. Then independently each fringe subtree of the shape T in \mathcal{T}_n^{gw} has probability $1/2^h$ to become a fringe subtree of the shape T^{bin} in $\mathcal{T}_n^{\text{bin}}$. Let $N_{T^{\text{bin}}}(\mathcal{T}_n^{\text{bin}})$ be the number of such fringe subtrees in $\mathcal{T}_n^{\text{bin}}$. Then by [Lemma 4.6](#), we have

$$\mathbb{E}N_{T^{\text{bin}}}(\mathcal{T}_n^{\text{bin}}) = \frac{1}{2^h} \mathbb{E}N_T(\mathcal{T}_n^{gw}) \sim \frac{1}{2^h} n \pi(T) = \frac{1}{2^h} \frac{n}{2^h} = \frac{n}{4^h}.$$

Thus with an argument similar to [Lemma 5.3](#), the maximum chain in $\mathcal{T}_n^{\text{bin}}$ that contains only left children has height $\log_4 n + o_p(1)$.

See [\[59\]](#) for the notation $o_p(1)$.

5.3 ALL POSSIBLE FRINGE SUBTREES

Recall that $\mathfrak{T}_{\leq k}^+$ denotes the set of all possible trees of size at most k , i.e.,

$$\mathfrak{T}_{\leq k}^+ := \{T \in \mathfrak{T} : |T| \leq k, \pi(T) > 0\}.$$

Also recall that

$$K_n := \max\{k : \mathfrak{T}_{\leq k}^+ \subseteq \cup_{v \in \mathcal{T}_n^{gw}} \{\mathcal{T}_{n,v}^{gw}\}\}.$$

We would like to study the growth of K_n with n .

Remark 5.1. We can assume that $\mathbb{P}\{|\mathcal{T}^{gw}| = k_n\} > 0$ for all $n \in \mathbb{N}$. Otherwise, let $k'_n := \max\{i \leq k_n : \mathbb{P}\{|\mathcal{T}^{gw}| = i\} > 0\}$. It is not difficult to show that $k_n - k'_n \leq h = O(1)$ for k_n large. (See [\[64, lem. 12.3\]](#) for details.) Thus this assumption does not change results in this section.

Janson showed that $\mathcal{T}_{n,*}^{gw} \xrightarrow{d} \mathcal{T}^{gw}$ [\[64, thm. 7.12\]](#). In other words, fringe subtrees on average behave like unconditional Galton-Watson trees. Let T_k^{min} be a tree $T \in \mathfrak{T}_{\leq k}^+$ that minimizes $\pi(T)$. Then T_k^{min} is also the least likely tree to appear in \mathcal{T}_n^{gw} as fringe subtree among all trees in $\mathfrak{T}_{\leq k}^+$ when n is large. So intuitively if

T_k^{\min} appears in $\mathcal{T}_n^{\text{gw}}$, then every tree in $\mathfrak{T}_{\leq k}^+$ should appear. And if T_k^{\min} does not appear, then there is at least one tree in $\mathfrak{T}_{\leq k}^+$ that is missing. Therefore, the problem can be reduced to finding

$$p_k^{\min} := \pi(T_k^{\min}) := \mathbb{P}\{\mathcal{T}^{\text{gw}} = T_k^{\min}\} = \min_{T \in \mathfrak{T}_{\leq k}^+} \mathbb{P}\{\mathcal{T}^{\text{gw}} = T\}.$$

The key observation is the following lemma:

Lemma 5.4. *Let $\mathcal{T}_n^{\text{gw}}(\mathfrak{m})$ be $\mathcal{T}_n^{\text{gw}}$ restricted to the event that $N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}}) = \mathfrak{m}$. Let $\mathcal{T}_n^{\text{gw}}(\mathfrak{m})'$ be $\mathcal{T}_n^{\text{gw}}(\mathfrak{m})$ with its fringe subtrees of size k replaced by \mathfrak{m} iid copies of $\mathcal{T}_k^{\text{gw}}$. We have*

$$\mathcal{T}_n^{\text{gw}}(\mathfrak{m}) \stackrel{\mathcal{L}}{=} \mathcal{T}_n^{\text{gw}}(\mathfrak{m})'.$$

Proof. Recall that $(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ is a uniform random rotation of the preorder degree sequence of $\mathcal{T}_n^{\text{gw}}$ and let

$$\tilde{J}_i = \mathbb{I}[(\tilde{\xi}_i^n, \tilde{\xi}_{i+1}^n, \dots, \tilde{\xi}_{i+k-1}^n) \in \mathcal{D}(\mathfrak{T}_k)],$$

as in (4.2), i.e., $(\tilde{J}_i)_{i=1}^n$ are the indicators for fringe subtrees of size k . Let E be the event that $\tilde{J}_i = 1$ if and only if $i \in \{1, k+1, \dots, k(m-1)+1\}$. In other words, E implies that $\mathcal{T}_n^{\text{gw}}$ contains exactly \mathfrak{m} fringe subtrees of size k . It suffices to prove that conditioning on E , we can replace the \mathfrak{m} fringe subtrees of size k by independent copies of $\mathcal{T}_k^{\text{gw}}$ without changing the distribution of the result.

Let $(d_1, \dots, d_n) \in \mathbb{N}_0^n$ be such

$$\left[(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n) = (d_1, \dots, d_n) \right] \in E.$$

Therefore for $1 \leq i \leq \mathfrak{m}$, $(d_{k(i-1)+1}, \dots, d_{ki})$ is a preorder degree sequence of a tree of size k , which we denote by T_i . Let $(\mathcal{T}_k^{\text{gw}}[i])_{i=1}^{\mathfrak{m}}$ be \mathfrak{m} iid copies of $\mathcal{T}_k^{\text{gw}}$. Write event

$$A_i = [\mathcal{T}_k^{\text{gw}}[i] = T_i],$$

and

$$B = \left[(\tilde{\xi}_{k\mathfrak{m}+1}^n, \dots, \tilde{\xi}_n^n) = (d_{k\mathfrak{m}+1}, \dots, d_n) \right].$$

So we need to show that

$$\mathbb{P}\left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n) = (d_1, \dots, d_n) \mid E \right\} = \mathbb{P}\left\{ \left[\bigcap_{i=1}^{\mathfrak{m}} A_i \right] \cap B \mid E \right\}. \quad (5.1)$$

By Lemma 2.4, the lhs of (5.1) equals

$$\frac{\mathbb{P}\left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n) = (d_1, \dots, d_n) \right\}}{\mathbb{P}\{E\}} = \frac{\mathbb{P}\left\{ (\xi_1, \dots, \xi_n) = (d_1, \dots, d_n) \right\}}{\mathbb{P}\{S_n = n-1\} \mathbb{P}\{E\}}.$$

Recall that $N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}})$ is the number of fringe subtrees of size k .

By now this type verification should be routine for the reader.

Recall that ξ_1, \dots, ξ_n are iid copies of ξ and $S_n = \sum_{i=1}^n \xi_i$.

Also by Lemma 2.4, we have

$$\mathbb{P}\{B \cap E\} = \frac{\mathbb{P}\{(\xi_{k_{m+1}}, \dots, \xi_n) = (d_{k_{m+1}}, \dots, d_n)\} \mathbb{P}\{|\mathcal{T}^{gw}| = k\}^m}{\mathbb{P}\{S_n = n-1\}},$$

and

$$\mathbb{P}\{\cap_{i=1}^m A_i\} = \frac{\mathbb{P}\{(\xi_1, \dots, \xi_{k_m}) = (d_1, \dots, d_{k_m})\}}{\mathbb{P}\{|\mathcal{T}^{gw}| = k\}^m}.$$

Thus the rhs of (5.1) equals

$$\mathbb{P}\{\cap_{i=1}^m A_i\} \frac{\mathbb{P}\{B \cap E\}}{\mathbb{P}\{E\}} = \frac{\mathbb{P}\{(\xi_1, \dots, \xi_n) = (d_1, \dots, d_n)\}}{\mathbb{P}\{S_n = n-1\} \mathbb{P}\{E\}}. \quad \square$$

The following corollary connects the existence of a family of subtrees to the coupon collector problem:

Corollary 5.1. *Conditioning on that $N_{\mathfrak{T}_k}(\mathcal{T}_n^{gw}) = m$, the probability that \mathcal{T}_n^{gw} contains all trees in $\mathfrak{S} \subseteq \mathfrak{T}_k$ equals that m iid copies of \mathcal{T}_k^{gw} contains all trees in \mathfrak{S} .*

Lemma 5.5. *Assume Condition A. If $k_n \rightarrow \infty$ and*

$$\frac{np_{k_n}^{\min}}{k_n} \rightarrow \infty,$$

then whp every possible tree of size k_n appears in \mathcal{T}_n^{gw} .

Proof. Let $k = k_n$. Recall that $p_{\max} := \max_{i \geq 0} p_i$ and $p_{\max} < 1$. Therefore $p_k^{\min} \leq p_{\max}^k$. Thus we can assume that $k \leq 2 \log(n) / \log(p_{\max}^{-1})$ when k is large. Otherwise we have $np_k^{\min} \leq n^{-1} \rightarrow 0$, which contradicts the assumption.

Thus by Theorem 1.3, $N_{\mathfrak{T}_k}(\mathcal{T}_n^{gw}) \geq y_n := \lceil \frac{1}{2} n \mathbb{P}\{|\mathcal{T}^{gw}| = k\} \rceil$ whp. Let A_n be the event that \mathcal{T}_n^{gw} contains all possible trees of size k as fringe subtrees. Let $B_n(x)$ be the event that $N_{\mathfrak{T}_k}(\mathcal{T}_n^{gw}) = x$ for given $x \geq y_n$. It follows by Corollary 5.1 that

$$\mathbb{P}\{A_n \mid B_n(a)\} \geq \mathbb{P}\{A_n \mid B_n(b)\}$$

given that $a \geq b$. Therefore

$$\begin{aligned} \mathbb{P}\{A_n\} &= \mathbb{P}\{A_n \cap [N_{\mathfrak{T}_k}(\mathcal{T}_n^{gw}) < y_n]\} + \sum_{i \geq y_n} \mathbb{P}\{A_n \mid B_n(i)\} \mathbb{P}\{B_n(i)\} \\ &\geq \mathbb{P}\{A_n \mid B_n(y_n)\} \sum_{i \geq y_n} \mathbb{P}\{B_n(i)\}, \\ &= \mathbb{P}\{A_n \mid B_n(y_n)\} (1 - o(1)). \end{aligned}$$

So it suffices to prove that $\mathbb{P}\{A_n^c \mid B_n(y_n)\} \rightarrow 0$.

By [Corollary 5.1](#) and [Lemma 5.1](#) (the coupon collector), we have

$$\begin{aligned}
\mathbb{P}\{A_n^c | B_n(y_n)\} &\leq \sum_{T \in \mathfrak{T}_k} \llbracket \pi(T) \geq 0 \rrbracket (1 - \mathbb{P}\{\mathcal{T}_k^{gw} = T\})^{y_n} \\
&\leq \sum_{T \in \mathfrak{T}_k} \llbracket \pi(T) \geq 0 \rrbracket \exp\{-y_n \mathbb{P}\{\mathcal{T}_k^{gw} = T\}\} \\
&\leq \sum_{T \in \mathfrak{T}_k} \llbracket \pi(T) \geq 0 \rrbracket \exp\left\{-\frac{1}{2} n \mathbb{P}\{|\mathcal{T}^{gw}| = k\} \frac{\mathbb{P}\{\mathcal{T}^{gw} = T\}}{\mathbb{P}\{|\mathcal{T}^{gw}| = k\}}\right\} \\
&\leq O(|\mathfrak{T}_k|) \exp\{-np_k^{\min}\}.
\end{aligned}$$

By [Lemma 2.8](#), $|\mathfrak{T}_k| = O(4^k)$. Since by assumption $np_k^{\min}/k \rightarrow \infty$, we have

$$\begin{aligned}
\mathbb{P}\{A_n^c | B_n(y_n)\} &\leq O(4^k) \exp\{-np_k^{\min}\} \\
&= O(1) \exp\{k \log(4) - np_k^{\min}\} \rightarrow 0. \quad \square
\end{aligned}$$

Corollary 5.2. *Assume Condition A. If $k_n \rightarrow \infty$ and*

$$\frac{np_{k_n}^{\min}}{k_n} \rightarrow \infty,$$

then whp every tree in $\mathfrak{T}_{\leq k_n}^+$ appears in \mathcal{T}_n^{gw} .

Proof. Let $k = k_n$. Let \mathfrak{T}_k^+ be the set of all possible trees of size exactly k , i.e.,

$$\mathfrak{T}_k^+ := \{T \in \mathfrak{T} : |T| = k, \pi(T) > 0\}.$$

If $p_1 > 0$, then every tree in \mathfrak{T}_{k-1}^+ is contained in some tree in \mathfrak{T}_k^+ . For example, if $T \in \mathfrak{T}_{k-1}^+$, then we can attach the root of T to a single node and get a new tree $T' \in \mathfrak{T}_k^+$. By [Lemma 5.5](#), whp \mathcal{T}_n^{gw} contains all trees in \mathfrak{T}_k^+ . If this happens, \mathcal{T}_n^{gw} also contains all trees in \mathfrak{T}_{k-1}^+ . It follows that \mathcal{T}_n^{gw} contains all trees in $\mathfrak{T}_{\leq k}^+$ whp.

Otherwise, let r be the smallest positive integer such that $p_r > 0$. Then every tree in \mathfrak{T}_{k-r}^+ is contained in some tree in \mathfrak{T}_k^+ . With union bound, it follows from [Lemma 5.5](#) that whp \mathcal{T}_n^{gw} contains all trees in $\cup_{i=0}^{r-1} \mathfrak{T}_{k-i}^+$. When this happens, \mathcal{T}_n^{gw} contains all trees in $\cup_{i=0}^{r-1} \mathfrak{T}_{k-r-i}^+$. It follows that \mathcal{T}_n^{gw} contains all trees in $\mathfrak{T}_{\leq k}^+$ whp. \square

Theorem 5.1. *Assume Condition A. Also assume that as $k \rightarrow \infty$,*

$$\log \frac{1}{p_k^{\min}} \sim \gamma k^\alpha (\log k)^\beta,$$

where $\alpha \geq 1$, $\beta \geq 0$, $\gamma > 0$ are constants. Let $k_n \rightarrow \infty$ be a sequence of positive integers. Let $m = \log n$. Then for all constants $\delta > 0$, we have:

(i) If

$$k_n \leq (1 - \delta) \left[\frac{m}{\gamma(\log m^{1/\alpha})^\beta} \right]^{1/\alpha}, \quad (5.2)$$

then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.

(ii) If

$$k_n \geq (1 + \delta) \left[\frac{m}{\gamma(\log m^{1/\alpha})^\beta} \right]^{1/\alpha}, \quad (5.3)$$

then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.

As a result,

$$\frac{K_n}{(\log n / (\log \log n)^\beta)^{1/\alpha}} \xrightarrow{p} \left(\frac{\alpha^\beta}{\gamma} \right)^{1/\alpha}.$$

Remark 5.2. The behavior of p_k^{min} varies for different offspring distributions. But as mentioned in the introduction, the types of trees that we are interested in all have p_k^{min} that are covered by Theorem 5.1.

Proof. Let $k = k_n$. Taking a logarithm of (5.2), we have

$$\log k \leq \log(1 - \delta) + \frac{1}{\alpha} \log \frac{m}{\gamma(\log m^{1/\alpha})^\beta} = (1 + o(1)) \log m^{1/\alpha}.$$

Thus

$$\log \frac{1}{p_k^{\text{min}}} \sim \gamma k^\alpha (\log k)^\beta \leq (\gamma + o(1)) \frac{(1 - \delta)^\alpha m}{\gamma(\log m^{1/\alpha})^\beta} (\log m^{1/\alpha})^\beta \sim (1 - \delta)^\alpha m.$$

Therefore, recalling $m = \log n$,

$$\log np_k^{\text{min}} = m - \log \frac{1}{p_k^{\text{min}}} \geq m - (1 + o(1))(1 - \delta)^\alpha m = \Omega(m).$$

It follows that

$$\log k - \log np_k^{\text{min}} = O(\log m) - \Omega(m) \rightarrow -\infty.$$

Thus $np_k^{\text{min}}/k \rightarrow \infty$ and it follows from Corollary 5.2 that whp $\mathcal{T}_n^{\text{gw}}$ contains every tree in $\mathfrak{T}_{\leq k}^+$ as a fringe subtree.

Similar computations show that (5.3) implies $np_k^{\text{min}} \rightarrow 0$. It follows from Theorem 1.2 that $N_{\mathcal{T}_k^{\text{min}}}(\mathcal{T}_n^{\text{gw}}) \xrightarrow{p} 0$. Thus whp $\mathcal{T}_n^{\text{gw}}$ does not contain every tree in $\mathfrak{T}_{\leq k}^+$ as a fringe subtree. \square

The rest of this section is organized as follows: In Subsection 5.3.1, we give a general method of finding p_k^{min} . Then in Subsection 5.3.2 and 5.3.3, we divide offspring distributions in two categories and show that Theorem 5.1 is applicable to all the Galton-Watson trees listed in Table 1.

5.3.1 Computing p_k^{\min}

Recall that $\text{supp}(\xi)$ denotes the support of ξ . Let

$$\text{supp}(\xi, k) := \text{supp}(\xi) \cap \{1, \dots, k\}.$$

Let $L_1 := p_0$ and for $k \geq 2$ let

$$L_k := \min \left\{ p_0 \left(\frac{p_i}{p_0} \right)^{1/i} : i \in \text{supp}(\xi, k-1) \right\}.$$

Since L_k is non-increasing, $L := \lim_{k \rightarrow \infty} L_k$ exists. Equivalently, we have

$$L := \inf \left\{ p_0 \left(\frac{p_i}{p_0} \right)^{1/i} : i \in \text{supp}(\xi) \setminus \{0\} \right\}.$$

Theorem 5.2. *Assume Condition A. We have*

$$(p_k^{\min})^{1/k} \rightarrow L \leq p_{\max} < 1$$

Recall that $p_{\max} := \max_{i \geq 0} p_i < 1$.

as $k \rightarrow \infty$, where the limit is taken along the subsequence k with $\mathbb{P}\{|\mathcal{T}^{g^w}| = k\} > 0$.

In fact, we have a stronger result for the upper bound of $(p_k^{\min})^{1/k}$.

Lemma 5.6. *Assume Condition A. For all fixed i with $p_i > 0$, there exist constants $C_i > 1$ and $C'_i, C''_i, k(i) > 0$ such that for all $k > k(i)$ with $\mathbb{P}\{|\mathcal{T}^{g^w}| = k\} > 0$, there are at least $k^{-C'_i} C_i^k$ trees T of size k with*

$$0 < \pi(T) \leq C''_i \left[p_0 \left(\frac{p_i}{p_0} \right)^{1/i} \right]^k.$$

Recall that $\pi(T) := \mathbb{P}\{\mathcal{T}^{g^w} = T\}$.

Proof of Lemma 5.6 with an extra condition. We give a proof assuming that there exists an integer $j \in \text{supp}(\xi)$ such that i and j are coprime. The general case is not much different and we discuss it afterwards.

Let $x = (k-1) \bmod i$. By the Chinese remainder theorem, there exists a smallest non-negative integer y such that

$$\begin{cases} y \equiv x \pmod{i}, \\ y \equiv 0 \pmod{j}. \end{cases}$$

Note that since x can take at most i different values, y is bounded from above by a constant $k(i)$ depending only on i . Therefore, if $k > k(i) \geq y$, we can choose

$$n_0 = k - n_i - n_j, \quad n_j = \frac{y}{j}, \quad n_i = \frac{k-1-y}{i},$$

such that n_0, n_i, n_j are all non-negative integers with

$$n_0 + n_i + n_j = k, \quad \text{and} \quad in_i + jn_j = k - 1.$$

Let $\mathfrak{T}_k(n_0, n_i, n_j)$ be the set of plane trees of size k that has n_0, n_i and n_j nodes with degree 0, i and j respectively. It is well-known that when the above two conditions hold, we have

$$|\mathfrak{T}_k(n_0, n_i, n_j)| = \frac{1}{k} \binom{k}{n_0, n_i, n_j} = \frac{1}{k} \frac{k!}{n_0! n_i! n_j!}.$$

(See [43, pp. 194].) Since i is a constant and $y \leq k(i)$, there exists a constant C_i^* such that

$$\left| n_0 - k \left(1 - \frac{1}{i} \right) \right| \leq C_i^*, \quad \left| n_i - \frac{k}{i} \right| \leq C_i^*, \quad n_j \leq C_i^*.$$

Using these inequalities and Stirling's approximation (see [43, pp. 407])

$$n! = \Theta \left(\frac{1}{\sqrt{n}} \left(\frac{n}{e} \right)^n \right),$$

it is easy to verify that there exists a constant $C_i' > 0$ such that

$$|\mathfrak{T}_k(n_0, n_i, n_j)| \geq k^{-C_i'} \left(\left(\frac{1}{i} \right)^{1/i} \left(1 - \frac{1}{i} \right)^{1-1/i} \right)^{-k} := k^{-C_i'} C_i^k.$$

And for every $T \in \mathfrak{T}_k(n_0, n_i, n_j)$, we have

$$\pi(T) \leq p_i^{n_i} p_0^{n_0} \leq p_i^{-C_i^*} p_0^{-C_i^*} p_i^{k/i} p_0^{k(1-1/i)} := C_i'' \left[p_0 \left(\frac{p_i}{p_0} \right)^{1/i} \right]^k. \quad \square$$

Proof of Lemma 5.6. Recall that we can assume $\text{span}(\xi) = 1$ (see Remark 2.1). In this case there exists a constant ω such that the span of $\text{supp}(\xi, \omega) := \text{supp}(\xi) \cap \{0, \dots, \omega\}$ is one. Thus by a well-known theorem of Schur (see [43, pp. 258]), there exists a constant D such that for all positive integer $m \geq D$, m can be partitioned into a sum of integers in $\text{supp}(\xi, \omega)$. Thus assuming $i > \omega$, we can choose

$$n_i = \left\lfloor \frac{k-1-D}{i} \right\rfloor,$$

such that there exists non-negative integers $n_0, n_1, \dots, n_\omega$ satisfying

$$k = n_i + \sum_{j=0}^{\omega} n_j, \quad \text{and} \quad k-1 = in_i + \sum_{j=1}^{\omega} jn_j,$$

and

$$\left| n_0 - k \left(1 - \frac{1}{i} \right) \right| \leq D^*, \quad \left| n_i - \frac{k}{i} \right| \leq D^*, \quad \left| \sum_{j=1}^{\omega} n_j \right| \leq D^*, \quad (5.4)$$

where D^* is a constant. From here we can estimate the number of trees T with n_j nodes of degree j for $j \in \{0, \dots, \omega\} \cup \{i\}$ and the probability $\pi(T)$ to finish the proof as before.

If $i \leq \omega$, similar to the above argument, we can choose

$$r_i = \left\lfloor \frac{k-1-D}{i} \right\rfloor,$$

such that there exist non-negative integers s_0, \dots, s_ω satisfying

$$k = r_i + \sum_{j=0}^{\omega} s_j, \quad \text{and} \quad k-1 = ir_i + \sum_{j=1}^{\omega} js_j.$$

Letting $n_j = s_j + \llbracket j = i \rrbracket \cdot r_i$, we again get (5.4). □

Recall that $\llbracket H \rrbracket$ is the indicator that H is true.

Proof of Theorem 5.2. Let T be a tree with $|T| = t$ and $\pi(T) > 0$. So $T \in \mathfrak{T}_{\leq t}^+$. Let n_i be the number of nodes of degree i in T . Note that if $i \notin \text{supp}(\xi)$, then $n_i = 0$. Since by (2.1) the sum of the degrees in a preorder degree sequence equals the size of the tree minus one, we have

$$n_0 + n_1 + \dots + n_{t-1} = t, \quad \text{and} \quad n_1 + 2n_2 + \dots + (t-1)n_{t-1} = t-1.$$

Using the convention that $0^0 = 1$, we have for $t \geq 2$

$$\begin{aligned} \pi(T) &= p_0^{n_0} p_1^{n_1} \dots p_{t-1}^{n_{t-1}} \\ &= p_0^{n_0 + n_1 + \dots + n_{t-1}} \left(\frac{p_1}{p_0}\right)^{n_1} \left(\frac{p_2}{p_0}\right)^{n_2} \dots \left(\frac{p_{t-1}}{p_0}\right)^{n_{t-1}} \\ &= p_0^t \prod_{i \in \text{supp}(\xi, t-1)} \left[\left(\frac{p_i}{p_0}\right)^{\frac{1}{t}} \right]^{in_i} \\ &\geq p_0^t \left[\min_{i \in \text{supp}(\xi, t-1)} \left(\frac{p_i}{p_0}\right)^{\frac{1}{t}} \right]^{\sum_{i=1}^{t-1} in_i} \\ &= p_0 L_t^{t-1} \geq p_0 L^{t-1}. \end{aligned}$$

Since $p_{\max}^t \geq \pi(T)$, we have for all t

$$p_{\max} \geq p_0^{1/t} L^{1-1/t}.$$

Recall that $p_{\max} := \max_{i \geq 0} p_i < 1$.

Therefore $L \leq p_{\max} < 1$. Recalling that \mathfrak{T}_t^+ is the set of possible trees of size t , we have

$$p_k^{\min} := \min_{1 \leq t \leq k} \min_{T \in \mathfrak{T}_t^+} \pi(T) \geq \min_{1 \leq t \leq k} p_0 L^{t-1} = p_0 L^{k-1}, \quad (5.5)$$

where the last equality follows from $L < 1$. As a result

$$\liminf_{k \rightarrow \infty} (p_k^{\min})^{1/k} \geq \liminf_{k \rightarrow \infty} \left(\frac{p_0}{L}\right)^{1/k} L = L.$$

To show the other way, let $\varepsilon > 0$ be a constant, and let

$$\alpha = \min\{i : L_{i+1} \leq L + \varepsilon\}.$$

Therefore $0 < p_0(p_\alpha/p_0)^{1/\alpha} \leq L + \varepsilon$. By Lemma 5.6, there exists at least one tree T of size k such that

$$p_k^{\min} \leq \mathbb{P}\{\mathcal{T}^{\text{gw}} = T\} \leq C_\alpha \left[p_0 \left(\frac{p_\alpha}{p_0} \right)^{\frac{1}{\alpha}} \right]^k \leq C_\alpha (L + \varepsilon)^k,$$

where $C_\alpha > 0$ is constant. Thus $\limsup_{k \rightarrow \infty} (p_k^{\min})^{1/k} \leq L + \varepsilon$. Since ε is arbitrary, we have $\limsup_{k \rightarrow \infty} (p_k^{\min})^{1/k} \leq L$. \square

5.3.2 When $L > 0$

If $L > 0$, then by Theorem 5.2,

$$\log \frac{1}{p_k^{\min}} \sim \log \left(\frac{1}{L} \right) k = \log \left(\frac{1}{L} \right) k (\log k)^0.$$

Thus we can apply Theorem 5.1 with $\gamma = \log(1/L)$, $\alpha = 1$ and $\beta = 0$ to get

$$\frac{K_n}{\log(n)} \xrightarrow{p} \frac{1}{\log(1/L)}.$$

The following lemma computes L for some well-known Galton-Watson trees which we have discussed in Section 2.3.

Lemma 5.7. *We have:*

- (i) If $\xi \stackrel{\mathcal{L}}{=} 2\text{Be}(1/2)$ (full binary tree), then $L = 1/2$.
- (ii) If $p_0 = p_1 = p_2 = 1/3$ (Motzkin tree), then $L = 1/3$.
- (iii) If $\xi \stackrel{\mathcal{L}}{=} \text{Bi}(d, 1/d)$ for $d \geq 2$ (d -ary tree), then $L = (d-1)^{d-1}/d^d$.
- (iv) If $\xi \stackrel{\mathcal{L}}{=} \text{Ge}(1/2)$ (plane tree), then $L = 1/4$.

Proof. (i): If $\xi \sim 2\text{Be}(1/2)$, then $p_0 = 1/2$, $p_2 = 1/2$ and $p_i = 0$ for $i \notin \{0, 2\}$. Thus for $k \geq 3$, we have

$$L_k = \min_{i \in \text{supp}(\xi, k-1)} p_0 \left(\frac{p_i}{p_0} \right)^{1/i} = p_0 \left(\frac{p_2}{p_0} \right)^{1/2} = 1/2.$$

Therefore $L = \lim_{k \rightarrow \infty} L_k = 1/2$. (ii) and (iii) follow from similar simple calculations.

(iv): For all $i \geq 1$, we have

$$p_0 \left(\frac{p_i}{p_0} \right)^{1/i} = \frac{1}{2} \left(\frac{1}{2^i} \right)^{1/i} = 1/4.$$

Therefore $L_k = 1/4$ for all $k \geq 1$, and $L = 1/4$. \square

Define

$$\mathfrak{L} = \begin{cases} \min\{i \in \mathbb{N} : L_{i+1} = L\} & \text{if } L_j = L \text{ for some } j, \\ \infty & \text{otherwise.} \end{cases}$$

If $\mathfrak{L} < \infty$, then we call the Galton-Watson tree *well-behaved*. Examples of such trees include those for which ξ is bounded, those for which ξ has a polynomial or slower-than-exponential tail, and the case $\xi \stackrel{\mathcal{L}}{=} \text{Ge}(1/2)$. Thus the four types of Galton-Watson trees in Lemma 5.7 are all well-behaved. The following theorem gives better thresholds than Theorem 5.1 for such Galton-Watson trees.

Theorem 5.3. *Assume Condition A. Then for all constants $\delta > 0$, we have:*

- (i) *If $k_n \leq (\log n - (1 + \delta) \log \log n) / \log \frac{1}{L}$, then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.*
- (ii) *If $k_n \geq (\log n - (1 - \delta) \log \log n) / \log \frac{1}{L}$ and the Galton-Watson tree is well-behaved, then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.*

Thus as $n \rightarrow \infty$, we have for well-behaved Galton-Watson trees,

$$\frac{k_n \log(1/L) - \log n}{\log \log n} \xrightarrow{p} -1.$$

The main idea is that when $\mathfrak{L} < \infty$, there are exponentially many trees of size k that have small probability to appear as fringe subtrees in $\mathcal{T}_n^{\text{gw}}$. Then we can use Lemma 5.1 (the coupon collector) to find the sufficient condition for one of them to not to appear whp.

Proof. Write $m = \log n$ and $k = k_n$. (i): Using (5.5), it is easy to verify that in this case $np_k^{\min}/k \rightarrow \infty$. Thus (i) follows from Corollary 5.2.

(ii): The proof is similar to the one of Corollary 5.2. As in that proof, we can assume that $k = O(\log n)$. Thus by Theorem 1.3, whp $N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}}) \leq y_n := \lfloor \frac{3}{2}n\mathbb{P}\{|\mathcal{T}^{\text{gw}}| = k\} \rfloor$. Let A_n be the event that $\mathcal{T}_n^{\text{gw}}$ contains all possible trees of size k as fringe subtrees. Let $B_n(i)$ be the event that $N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}}) = i$ for some $i \leq y_n$. Then

$$\begin{aligned} \mathbb{P}\{A_n\} &\leq \mathbb{P}\{N_{\mathfrak{T}_k}(\mathcal{T}_n^{\text{gw}}) > y_n\} + \sum_{i \leq y_n} \mathbb{P}\{A_n | B_n(i)\} \mathbb{P}\{B_n(i)\} \\ &\leq o(1) + \mathbb{P}\{A_n | B_n(y_n)\} \sum_{i \leq y_n} \mathbb{P}\{B_n(i)\}, \\ &\leq o(1) + \mathbb{P}\{A_n | B_n(y_n)\}. \end{aligned}$$

Thus it suffices to prove that $\mathbb{P}\{A_n | B_n(y_n)\} \rightarrow 0$.

By [Corollary 5.1](#), we have $\mathbb{P}\{A_n|B_n(y_n)\}$ equals the probability that y_n independent copies of $\mathcal{T}_k^{g^{w}}$ do not contains all trees in $\mathfrak{T}_{\leq k}^+$. It follows from [Lemma 5.1](#) (the coupon collector) that $\mathbb{P}\{A_n|B_n(y_n)\} \rightarrow 0$ if

$$\sum_{T \in \mathfrak{T}_k^+} (1 - \mathbb{P}\{\mathcal{T}_k^{g^{w}} = T\})^{y_n} \rightarrow \infty.$$

By definition of \mathfrak{L} , we have $L = p_0(p/p_0)^{1/\mathfrak{L}}$. It follows from [Lemma 5.6](#) that there exists constants $C > 1$ and $C', C'' > 0$ such that there are at least $k^{-C'} C^k$ trees T in $\mathfrak{T}_{\leq k}^+$ with

$$\mathbb{P}\{\mathcal{T}_k^{g^{w}} = T\} = \frac{\mathbb{P}\{\mathcal{T}^{g^{w}} = T\}}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}} \leq C'' \frac{(p_0(p/p_0)^{1/\mathfrak{L}})^k}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}} = \frac{C'' L^k}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}}.$$

Therefore

$$\sum_{T \in \mathfrak{T}_k^+} (1 - \mathbb{P}\{\mathcal{T}_k^{g^{w}} = T\})^{y_n} \geq k^{-C'} C^k \left(1 - \frac{C'' L^k}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}}\right)^{y_n}.$$

See [Theorem 5.2](#) and [Lemma 2.5](#).

Since $L < 1$ and $\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\} = \Theta(k^{-3/2})$, we have $L^k/\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\} = o(1)$. Thus for k large enough, the logarithm of the above is

$$\begin{aligned} & k \log(C) - C' \log(k) + y_n \log \left(1 - \frac{C'' L^k}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}}\right) \\ & \geq \frac{1}{2} k \log(C) - 2y_n \frac{C'' L^k}{\mathbb{P}\{|\mathcal{T}^{g^{w}}| = k\}} \\ & \geq \frac{1}{2} k \log(C) - 3nC'' L^k = \frac{1}{2} k \log(C) - O(nL^k). \end{aligned}$$

By our assumptions, $k = \Omega(\log n)$ and $L^k \leq (\log n)^{1-\delta}/n$. Since $C > 1$, we have

$$\frac{1}{2} k \log(C) - O(nL^k) \geq \Omega(\log n) - O\left(n \frac{(\log n)^{1-\delta}}{n}\right) \rightarrow \infty,$$

which implies $\mathbb{P}\{A_n|B_n(y_n)\} \rightarrow 0$. \square

Remark 5.3. If $L > 0$ and $\mathfrak{L} = \infty$, then we have $K_n/\log(n) \xrightarrow{P} 1/\log(1/L)$ by [Theorem 5.1](#). But the second order term of K_n is sensitive to small modifications of the offspring distribution, which makes it more difficult to analyze without additional information of the distribution.

5.3.3 When $L = 0$

It is clear that $L = 0$ if and only if ξ has infinite support and

$$\liminf_{i \rightarrow \infty} p_0 \left(\frac{p_i}{p_0}\right)^{1/i} = 0,$$

which implies

$$\limsup_{i \rightarrow \infty} \frac{\log(1/p_i)}{i} = \infty,$$

along the subsequence with $p_i > 0$. If in addition we have $p_i > 0$ for all $i \geq 0$ and $\log(1/p_i) \sim f(i)$ for some $f : [0, \infty) \rightarrow [0, \infty)$ with $f(i)/i \uparrow \infty$, then we say that ξ has an *f-super-exponential tail*. We have the following threshold for Galton-Watson trees with such a property.

Theorem 5.4. *Assume Condition A and that ξ has an f-super-exponential tail. Let f^{-1} denote the inverse of f . Then for all constants $\delta > 0$, we have*

- (i) *If $k_n \leq f^{-1}((1 - \delta) \log n) + 1$, then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.*
- (ii) *If $k_n \geq f^{-1}((1 + \delta) \log n) + 1$, then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.*

Therefore,

$$\frac{K_n}{f^{-1}(\log n)} \xrightarrow{p} 1.$$

Proof. (i): Let $k = k_n$. Let $\varepsilon > 0$ be a constant decided later. Since $\log(1/p_i) \sim f(i)$, there exists an integer $i(\varepsilon)$ such that for all $i > i(\varepsilon)$,

$$\log p_i \geq -(1 + \varepsilon/2)f(i).$$

Let $w_i := \log p_0 (p_i/p_0)^{1/i}$. We have as $k \rightarrow \infty$,

$$\begin{aligned} \min_{i(\varepsilon) < i < k} w_i &= \min_{i(\varepsilon) < i < k} \left\{ \left(1 - \frac{1}{i}\right) \log(p_0) + \frac{\log p_i}{i} \right\} \\ &\geq \log(p_0) - \max_{i(\varepsilon) < i < k} \frac{(1 + \varepsilon/2)f(i)}{i} \\ &= \log(p_0) - \frac{(1 + \varepsilon/2)f(k-1)}{k-1} \rightarrow -\infty, \end{aligned}$$

where we use that $f(i)/i \uparrow \infty$. Since $\min_{1 \leq i \leq i(\varepsilon)} w_i$ is a constant, we have for large k ,

$$\log L_k := \min_{1 \leq i < k} w_i \geq \log(p_0) - \frac{(1 + \varepsilon/2)f(k-1)}{k-1}.$$

It follows from (5.5) that for k large enough,

$$\begin{aligned} \log p_k^{\min} &\geq \log(p_0 L_k^{k-1}) \\ &\geq \log(p_0) + (k-1) \log p_0 - (1 + \varepsilon/2)f(k-1). \\ &\geq -(1 + \varepsilon)f(k-1), \end{aligned}$$

where the last step uses $f(k)/k \uparrow \infty$.

The assumption $k - 1 \leq f^{-1}((1 - \delta) \log n)$ implies that $f(k - 1) \leq (1 - \delta) \log n$ and $k = O(\log n)$. Thus

$$\log p_k^{\min} \geq -(1 + \varepsilon)(1 - \delta) \log n \geq -(1 - \delta/2) \log n,$$

if we choose ε to be small enough with respect to δ . Thus $np_k^{\min} \geq n^{\delta/2}$. We have $np_k^{\min}/k \rightarrow \infty$. It follows from [Corollary 5.2](#) that $\mathcal{T}_n^{\text{gw}}$ contains all possible trees of size at most k as fringe subtree whp.

(ii): Let T_{k-1}^{star} be the tree in which one node has degree $k - 1$ and all other nodes are leaves. Computations similar to above show that by the assumption of (ii), $n\pi(T_{k-1}^{\text{star}}) \rightarrow 0$. Therefore $\mathcal{T}_n^{\text{gw}}$ does not contain T_{k-1}^{star} whp. \square

Example 5.1 (The discrete Gaussian distribution). If $p_i = ce^{-c'i^2}$ for some appropriate positive normalization constants c and c' , then ξ has an $c'i^2$ -super-exponential tail. Thus [Theorem 5.4](#) applies and we have

$$\frac{K_n}{\sqrt{\log(n)}} \xrightarrow{p} \frac{1}{\sqrt{c'}},$$

as $n \rightarrow \infty$.

Example 5.2 (The Cayley trees). Another example is the Galton-Watson tree with offspring distribution $\xi \stackrel{\mathcal{L}}{=} \text{Po}(1)$, i.e., the Cayley tree. It has $p_i = e^{-1}/i!$ and $\log(1/p_i) \sim i \log(i)$ by Stirling's approximation. In other words, ξ has an f -super-exponential tail with $f(i) = i \log(i)$. Since $f^{-1}(j) \sim j/\log j$, we have

$$\frac{K_n \log \log n}{\log n} \xrightarrow{p} 1. \quad (5.6)$$

Using (5.5) we can show that if the tail of ξ drops fast enough then the least possible tree of size at most k is T_{k-1}^{star} .

Lemma 5.8. *Assume Condition A. If $p_i > 0$ for all $i \geq 0$ and $p_i^{1/i} \downarrow 0$, then for k large enough, $p_k^{\min} = \pi(T_{k-1}^{\text{star}}) = p_0^{k-1} p_{k-1}$.*

Recall that $\pi(T) := \mathbb{P}\{\mathcal{T}^{\text{gw}} = T\}$.

Proof. Let $r_i = \log(1/p_i^{1/i})$. We have for k large enough,

$$\begin{aligned} \log \left(\frac{p_k}{p_0} \right)^{\frac{1}{k}} - \log \left(\frac{p_{k+1}}{p_0} \right)^{\frac{1}{k+1}} &= r_{k+1} - r_k + \log(p_0^{-1}) \left(\frac{1}{k} - \frac{1}{k+1} \right) \\ &\geq r_{k+1} - r_k \geq 0, \end{aligned}$$

since r_k is eventually non-decreasing. In other words, $(p_k/p_0)^{1/k}$ is eventually non-increasing. On the other hand, since $r_k \rightarrow \infty$,

$$p_0 \left(\frac{p_k}{p_0} \right)^{1/k} = p_0^{1-1/k} e^{-r_k} \rightarrow 0.$$

Therefore, for k large enough, we have

$$L_k := \min_{1 \leq i < k} p_0 \left(\frac{p_i}{p_0} \right)^{1/i} = p_0 \left(\frac{p_{k-1}}{p_0} \right)^{\frac{1}{k-1}}.$$

By (5.5), we have

$$p_k^{\min} \geq p_0 L_k^{k-1} = p_0 \left[p_0 \left(\frac{p_{k-1}}{p_0} \right)^{\frac{1}{k-1}} \right]^{k-1} = p_0^{k-1} p_{k-1} = \pi(T_{k-1}^{\text{star}}).$$

By the definition of p_k^{\min} , we have

$$p_k^{\min} := \min\{\pi(T) : |T| \leq k, \pi(T) > 0\} \leq \pi(T_{k-1}^{\text{star}}).$$

Therefore $p_k^{\min} = \pi(T_{k-1}^{\text{star}}) = p_0^{k-1} p_{k-1}$. \square

In particular, Lemma 5.8 applies to $\xi \stackrel{\mathcal{L}}{=} \text{Po}(1)$. In this case we have

$$\log p_k^{\min} = \log(p_0^{k-1} p_{k-1}) = -k \log k(1 + O(1/k)).$$

It follows from Theorem 5.1 that,

$$K_n \frac{\log \log n}{\log n} \xrightarrow{p} 1,$$

as $n \rightarrow \infty$. This matches (5.6) given by Theorem 5.4. However, we can be more precise by applying the following theorem with $\gamma = 1$.

Theorem 5.5. *Assume Condition A. Also assume that*

$$\log(1/p_i^{\min}) = \gamma(\log i)(i + O(1)),$$

where $\gamma > 0$ is a constant. Define

$$m = \log n, \quad m_1 = \log(m/\gamma), \quad m_2 = \log m_1.$$

Let $k_n \rightarrow \infty$ be a sequence of positive integers. Then for all constants $\delta > 0$, we have:

(i) If

$$k_n \leq \frac{m}{\gamma(m_1 - m_2)} \left(1 - (1 + \delta) \frac{m_2}{m_1(m_1 - m_2)} \right),$$

then whp $\mathcal{T}_n^{\text{gw}}$ contains all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.

(ii) If

$$k_n \geq \frac{m}{\gamma(m_1 - m_2)} \left(1 - (1 - \delta) \frac{m_2}{m_1(m_1 - m_2)} \right),$$

then whp $\mathcal{T}_n^{\text{gw}}$ does not contain all trees in $\mathfrak{T}_{\leq k_n}^+$ as fringe subtrees.

Thus, as $n \rightarrow \infty$,

$$k_n \frac{\log \log n}{\log n} \xrightarrow{p} \frac{1}{\gamma},$$

and more precisely,

$$\left(\frac{k_n \gamma}{\log n} \left(\log \frac{\log n}{\gamma} - \log \log \frac{\log n}{\gamma} \right) - 1 \right) \times \frac{(\log \log n)^2}{\log \log \log n} \xrightarrow{p} -1.$$

Proof. Let $k = k_n$. For (i), we can show that $np_k^{\text{min}}/k \rightarrow \infty$. It follows from [Corollary 5.2](#) that whp $\mathcal{T}_n^{\text{gw}}$ contains all trees of size at most k as fringe subtrees.

For (ii), similarly we can show that $np_k^{\text{min}} \rightarrow 0$. It follows from [Theorem 1.2](#) that whp there is at least one tree in $\mathfrak{T}_{\leq k}^+$ that does not appear in $\mathcal{T}_n^{\text{gw}}$ as fringe subtrees. \square

NON-FRINGE SUBTREES

In this chapter we prove Theorem 1.4, the concentration of non-fringe subtree counts in conditional Galton-Watson trees. Section 6.1 gives an intuition of the proof. Section 6.2 calculates the first and second moments of non-fringe subtree counts. Section 6.3 proves Theorem 1.4. As an application of this theorem, Section 6.4 studies the height of the maximal complete r -ary non-fringe subtree in $\mathcal{T}_n^{\text{gw}}$. Finally, Section 6.5 discusses the difference between fringe and non-fringe subtrees.

6.1 INTUITION AND NOTATIONS

Given a tree T , let $\nu(T)$ be the number of its internal nodes and let $\ell(T)$ be the number of its leaves. Recall that

$$N_T^{\text{nf}}(\mathcal{T}_n^{\text{gw}}) := \sum_{u \in \mathcal{T}_n^{\text{gw}}} \mathbb{1}[T < \mathcal{T}_{n,u}^{\text{gw}}],$$

where $T < T'$ denotes that T' has a non-fringe subtree of shape T at its root.

To simplify the notation, write $\nu := \nu(T)$ and $\ell := \ell(T)$. By Lemma 2.2, T has a preorder degree sequence which consists of ℓ sequences of positive integers (some of which may be empty) separated by ℓ zeros. More precisely, it has the form

$$\begin{aligned} (\vec{a}_1, 0, \vec{a}_2, 0, \dots, \vec{a}_\ell, 0) := & (a_{1,1}, a_{1,2}, \dots, a_{1,r(1)}, 0, \\ & a_{2,1}, a_{2,2}, \dots, a_{2,r(2)}, 0, \\ & \dots, \\ & a_{\ell,1}, a_{\ell,2}, \dots, a_{\ell,r(\ell)}, 0) \end{aligned} \quad (6.1)$$

for non-negative integers $r(1), r(2), \dots, r(\ell)$ and that

$$\sum_{s=1}^{\ell} r(s) = \nu, \quad a_{s,t} \in \mathbb{N}, \quad \sum_{s=1}^{\ell} \vec{a}_s := \sum_{s=1}^{\ell} \sum_{t=1}^{r(s)} a_{s,t} = \nu + \ell - 1. \quad (6.2)$$

Therefore, if $T < T'$, then T' has a preorder degree sequence of the form

$$(\vec{a}_1, \vec{b}_1, \vec{a}_2, \vec{b}_2, \dots, \vec{a}_\ell, \vec{b}_\ell) \quad (6.3)$$

$\mathbb{1}[H]$ is the indicator that H is true.

We will refer to this form many times.

Recall that $(\tilde{\xi}_r^n)_{r=1}^n$ is a uniform random rotation of the preorder degree sequence of $\mathcal{T}_n^{\text{gw}}$.

where $\vec{b}_1, \dots, \vec{b}_\ell$ are *tree degree sequences*, i.e., preorder degree sequences of some plane trees. Thus each non-fringe subtree of shape T in $\mathcal{T}_n^{\text{gw}}$ corresponds to a segment of $(\tilde{\xi}_r^n)_{r=1}^n$ of the form of (6.3). If none of the segments overlap with each other, then we can permute them into the form $(\vec{a}_1, \dots, \vec{a}_\ell, \vec{b}_1, \dots, \vec{b}_\ell)$. Since $(\tilde{\xi}_r^n)_{r=1}^n$ is permutation invariant, the old and new sequences have the same probability. In other words, $N_T^{\text{nf}}(\mathcal{T}_n^{\text{gw}})$ is almost distributed like the number of the patterns $(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_\ell)$ in $(\tilde{\xi}_r^n)_{r=1}^n$.

The problem with this argument is that non-fringe subtrees can overlap. But as shown later in this section, under the assumptions of Theorem 1.4, the effect of such overlaps is negligible.

Recall that we use $\mathcal{D}(\mathfrak{T}_n)$ to denote the set of preorder degree sequences of trees of size n . Let \mathcal{R}_n be the set of sequences that are cyclic rotations of sequences in $\mathcal{D}(\mathfrak{T}_n)$, i.e.,

See also Lemma 2.3.

$$\mathcal{R}_n := \left\{ (r_1, \dots, r_n) \in \mathbb{N}_0^n : \sum_{i=1}^n r_i = n - 1 \right\}.$$

Given $\vec{d} := (d_1, \dots, d_n) \in \mathcal{R}_n$, let $k_i(\vec{d}) \geq 1$ be the unique integer such that

$$\text{deg}_i(\vec{d}) := (d_i, d_{i+1}, \dots, d_{i+k_i(\vec{d})-1}) \in \mathcal{D}(\mathfrak{T}_{k_i(\vec{d})}),$$

where the indices are all modulo n . Lemma 2.2 guarantees that such $\text{deg}_i(\vec{d})$ exists and is unambiguous. Let $T_i(\vec{d})$ be the tree with the preorder degree sequence $\text{deg}_i(\vec{d})$. Let $T_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ be the fringe subtree of $\mathcal{T}_n^{\text{gw}}$ with the preorder degree sequence $\text{deg}_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$.

6.2 FACTORIAL MOMENTS

We now prove the exact formulas for the first and second moments of $N_T^{\text{nf}}(\mathcal{T}_n^{\text{gw}})$. Throughout this section, let T be a tree with a preorder degree sequence of the form $(\vec{a}_1, 0, \dots, \vec{a}_\ell, 0)$ satisfying (6.2). Let $v := v(T)$ and $\ell := \ell(T)$.

In other words, T has v internal nodes and ℓ leaves.

6.2.1 The first factorial moment

Lemma 6.1. *Let $\vec{d} := (d_1, d_2, \dots, d_n) \in \mathcal{R}_n$ such that*

$$(d_1, \dots, d_v) = (\vec{a}_1, \dots, \vec{a}_\ell).$$

Then there exist $t(1), t(2), \dots, t(\ell + 1)$ such that

$$v = t(1) < t(2) < \dots < t(\ell) < t(\ell + 1) \leq n$$

and

$$(d_{t(i)+1}, d_{t(i)+2}, \dots, d_{t(i+1)})$$

is a tree degree sequence (satisfying (2.1)), for all $1 \leq i \leq \ell$. Moreover the sequence

$$(d_1, d_2, \dots, d_{t(\ell+1)})$$

is also a tree degree sequence.

Proof. Consider the walk W_t defined by

$$W_t = \begin{cases} 0, & \text{if } t = 0, \\ \sum_{i=1}^t (d_i - 1), & \text{if } 1 \leq t \leq n. \end{cases}$$

Since $\vec{d} \in \mathcal{R}_n$,

$$W_n := \sum_{i=1}^n (d_i - 1) = \left(\sum_{i=1}^n d_i \right) - n = (n - 1) - n = -1.$$

So this walk starts from 0 and ends at -1 after n steps. By (6.2),

$$W_{t(1)} := W_v = \sum_{i=1}^v (d_i - 1) = \left(\sum_{s=1}^{\ell} \vec{d}_s \right) - v = (v + \ell - 1) - v = \ell - 1 \geq 0.$$

A finite tree has at least one leaf. So $\ell \geq 1$.

Thus to reach -1 , the walk has to reach $\ell - 2$ first. Let $t(2)$ be the first time this happens after step v . Then

$$\sum_{t(1) < i \leq t(2)} (d_i - 1) = W_{t(2)} - W_{t(1)} = (\ell - 2) - (\ell - 1) = -1,$$

and

$$\sum_{t(1) < i \leq j} (d_i - 1) = W_j - W_{t(1)} > (\ell - 2) - (\ell - 1) = -1, \quad t(0) < j < t(2).$$

So $(d_{t(1)+1}, \dots, d_{t(2)})$ satisfies (2.1), i.e., it is a tree degree sequence.

Let $t(i)$ be the first time that W_t reaches $\ell - i$ after step v for $3 \leq i \leq \ell + 1$. Using the above argument, we also can verify that $t(3), \dots, t(\ell + 1)$ are as required.

Since by (6.2), $\vec{d}_1, \dots, \vec{d}_\ell$ contains only positive integers, $t(\ell + 1)$ is also the first time when the entire walk W_t reaches -1 . So the sequence $d_1, d_2, \dots, d_{t(\ell+1)}$ is also a tree degree sequence, i.e., it satisfies (2.1). \square

Lemma 6.2. Assume that $\mathbb{P}\{|\mathcal{T}^{gw}| = n\} > 0$ and that $v < n$. We have

$$\frac{\mathbb{E} [N_T^{nf}(\mathcal{T}_n^{gw})]}{n} = \pi^{nf}(T) \frac{\mathbb{P}\{S_{n-v} = n - v - \ell\}}{\mathbb{P}\{S_n = n - 1\}}.$$

Recall that $\pi^{nf}(T) := \mathbb{P}\{T < \mathcal{T}^{gw}\}$ and S_n is the sum of n iid copies of ξ .

Proof. Recall that $T_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$ is the fringe subtree that has the preorder degree sequence $\deg_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)$. Let

$$I_i = \llbracket T < T_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n) \rrbracket. \quad (6.4)$$

Then with an argument similar to [Lemma 2.6](#), $N_{\uparrow}^{\text{nf}}(\mathcal{T}_n^{\text{gw}}) = \sum_{i=1}^n I_i$. Thus we have

$$\mathbb{E} [N_{\uparrow}^{\text{nf}}(\mathcal{T}_n^{\text{gw}})] = \mathbb{E} \left[\sum_{i=1}^n I_i \right] = n\mathbb{P} \{I_1 = 1\}.$$

Recall that T has a preorder degree sequence of the form $(\vec{a}_1, 0, \dots, \vec{a}_\ell, 0)$ satisfying [\(6.2\)](#). Let $\mathcal{A} \subseteq \mathcal{R}_n$ be the set of sequences such that $(\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{A}$ if and only if $I_1 = 1$. In other words, $\vec{d} := (d_1, d_2, \dots, d_n) \in \mathcal{A}$ if and only if $\deg_1(\vec{d}) = (\vec{a}_1, \vec{b}_1, \dots, \vec{a}_\ell, \vec{b}_\ell)$ for some $\vec{b}_1, \dots, \vec{b}_\ell$ which are tree degree sequences. By permuting $\deg_1(\vec{d})$ into $(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_\ell, \vec{b}_1, \vec{b}_2, \dots, \vec{b}_\ell)$, we get a new sequence $\vec{d}' := (d'_1, d'_2, \dots, d'_n) \in \mathcal{A}'$ where

$$\mathcal{A}' := \{(e_1, e_2, \dots, e_n) \in \mathcal{R}_n : (e_1, e_2, \dots, e_n) = (\vec{a}_1, \vec{a}_2, \dots, \vec{a}_\ell)\}.$$

Such a permutation defines a mapping $f : \mathcal{A} \rightarrow \mathcal{A}'$.

For every $\vec{d}' = (d'_1, \dots, d'_n) \in \mathcal{A}'$, [Lemma 6.1](#) implies that in \vec{d}' the segment $(\vec{a}_1, \dots, \vec{a}_\ell)$ is followed by ℓ tree degree sequences. So we can always reverse the above permutation and find a unique $\vec{d} \in \mathcal{A}$ with $f(\vec{d}) = \vec{d}'$. Thus f is a one-to-one mapping. If $\vec{d}' = f(\vec{d})$, then $\mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d}' \right\}$, since $(\tilde{\xi}_r^n)_{r=1}^n$ is permutation invariant. Therefore we have

$$\mathbb{P} \{I_1 = 1\} = \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{A} \right\} = \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{A}' \right\}.$$

ξ_1, \dots, ξ_n are iid
copies of ξ .

Recall that by [Lemma 2.4](#),

$$(\tilde{\xi}_r^n)_{r=1}^n \sim (\xi_1, \dots, \xi_n \mid S_n = n - 1).$$

Recall that
 $\pi^{\text{nf}}(T) :=$

$\mathbb{P} \{T < \mathcal{T}^{\text{gw}}\}.$

We have

$$\begin{aligned} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{A}' \right\} &= \mathbb{P} \left\{ (\tilde{\xi}_1^n, \tilde{\xi}_2^n, \dots, \tilde{\xi}_v^n) = (\vec{a}_1, \vec{a}_2, \dots, \vec{a}_\ell) \right\} \\ &= \frac{\mathbb{P} \left\{ (\xi_1, \xi_2, \dots, \xi_v) = (\vec{a}_1, \vec{a}_2, \dots, \vec{a}_\ell), S_n = n - 1 \right\}}{\mathbb{P} \{S_n = n - 1\}} \\ &= \pi^{\text{nf}}(T) \frac{\mathbb{P} \{S_{n-v} = n - v - \ell\}}{\mathbb{P} \{S_n = n - 1\}}, \end{aligned}$$

where the last step uses $\sum_{s=1}^{\ell} \vec{a}_s = v + \ell - 1$ from [\(6.2\)](#). □

6.2.2 The second factorial moment

To compute $\mathbb{E}(N_T^{nf}(\mathcal{T}_n^{gw}))_2$, we enumerate all the cases that T can appear as overlapping non-fringe subtrees by constructing a set of trees $\{T \boxplus T\}$ as follows. For trees T, S and node $u \in T$, let $T' = \text{Splay}(T, u, S)$ denote tree T with subtree T_u replaced by S . Thus $T'_u = S$. Let $\mathcal{V}(T)$ be the set of internal nodes of T . Then define the set of trees

$$\{T \boxplus T\} = \left[\bigcup_{u \in \mathcal{V}(T): T_u < T} \{\text{Splay}(T, u, T)\} \right] \setminus \{T\}.$$

Note that $|\{T \boxplus T\}| < v := |\mathcal{V}(T)|$. Also note that given $T' \in \{T \boxplus T\}$ we can always find a unique node u in T such that $T' = \text{Splay}(T, u, T)$. See Figure 6 for an example of $\{T \boxplus T\}$.

Recall that T_u is the fringe subtree of T rooted at $u \in T$.

Recall that T has v internal nodes and ℓ leaves.

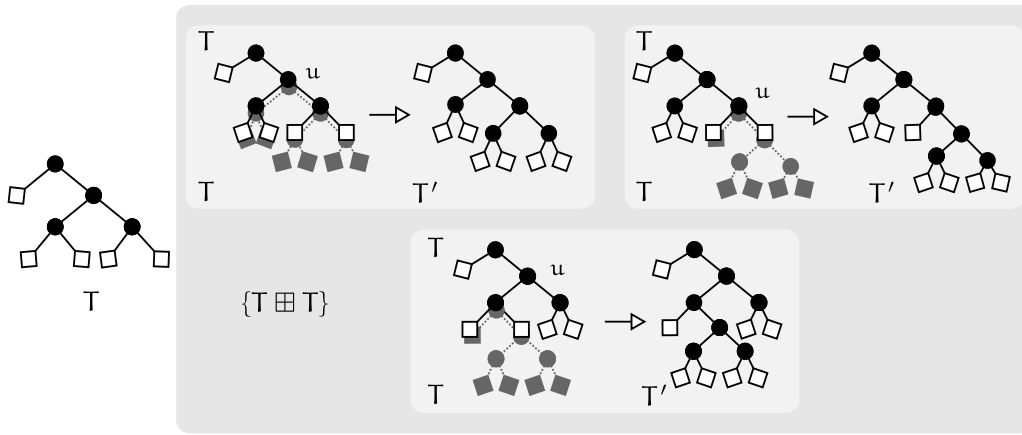


Figure 6: Example of $\{T \boxplus T\}$

Let I_i be defined as in (6.4) (in Lemma 6.2). Since I_1, \dots, I_n are indicator random variables and permutation invariant, by Lemma 3.9, we have

$$\mathbb{E}(N_T^{nf}(\mathcal{T}_n^{gw}))_2 = \sum_{1 \leq i \neq j \leq n} \mathbb{E}[I_i I_j] = n \sum_{i=2}^n \mathbb{E}[I_1 I_i].$$

However, since preorder degree sequences of non-fringe subtrees can overlap, computing this sum is not as simple as in the case of fringe subtrees. Instead we sum up $\mathbb{P}\left\{(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}\right\}$ over all \vec{d} such that the event $(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}$ implies $I_1 I_i = 1$ for some i .

Lemma 6.3. Let \mathcal{B} be the set of pairs $(i, \vec{d}) \in \{2, \dots, n\} \times \mathcal{R}_n$ such that $T < T_1(\vec{d})$ and $T < T_i(\vec{d})$. Then

$$\frac{\mathbb{E}(N_T^{nf}(\mathcal{T}_n^{gw}))_2}{n} = \sum_{i \geq 2} \mathbb{E}[I_1 I_i] = \sum_{(i, \vec{d}) \in \mathcal{B}} \mathbb{P}\left\{(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}\right\}.$$

Recall that $T_j(\vec{d})$ is the tree with preorder degree sequence $\text{deg}_j(\vec{d})$.

Proof. The event $I_1 I_i = 1$ happens if and only if

$$T < T_1(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n), \quad \text{and} \quad T < T_i(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n),$$

i.e., $(i, (\tilde{\xi}_r^n)_{r=1}^n) \in \mathcal{B}$. Therefore

$$\begin{aligned} \sum_{i=2}^n \mathbb{E} [I_1 I_i] &= \sum_{i=2}^n \sum_{\vec{d} \in \mathcal{R}_n} \mathbb{1}[(i, \vec{d}) \in \mathcal{B}] \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\ &= \sum_{(i, \vec{d}) \in \mathcal{B}} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\}. \end{aligned} \quad \square$$

Let $(i, \vec{d}) \in \mathcal{B}$. If $(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}$, then $T < T_1((\tilde{\xi}_r^n)_{r=1}^n)$ and $T < T_i((\tilde{\xi}_r^n)_{r=1}^n)$. The two fringe subtrees $T_1((\tilde{\xi}_r^n)_{r=1}^n)$ and $T_i((\tilde{\xi}_r^n)_{r=1}^n)$ may or may not overlap. But what is more important is whether the parts of $T_1((\tilde{\xi}_r^n)_{r=1}^n)$ and $T_i((\tilde{\xi}_r^n)_{r=1}^n)$ that correspond to the internal nodes of T overlap.

To be precise, assume that

$$\text{deg}_1(\vec{d}) = (\vec{a}_1, \vec{b}_1, \vec{a}_2, \vec{b}_2, \dots, \vec{a}_\ell, \vec{b}_\ell),$$

and

$$\text{deg}_i(\vec{d}) = (\vec{a}_1, \vec{b}'_1, \vec{a}_2, \vec{b}'_2, \dots, \vec{a}_\ell, \vec{b}'_\ell),$$

where $(\vec{a}_1, 0, \dots, \vec{a}_\ell, 0)$ is the preorder degree sequence of T which satisfies (6.2) and $\vec{b}_1, \vec{b}'_1, \dots, \vec{b}_\ell, \vec{b}'_\ell$ are tree degree sequences.

Let $\mathcal{M}_1^a(\vec{d})$ be the set of positions (indices) in \vec{d} that are occupied by the $\vec{a}_1, \dots, \vec{a}_\ell$ parts of $\text{deg}_1(\vec{d})$. Let $\mathcal{M}_1^b(\vec{d})$ be the set of positions in \vec{d} that are occupied by the $\vec{b}_1, \dots, \vec{b}_\ell$ parts of $\text{deg}_1(\vec{d})$. Let $\mathcal{M}_1(\vec{d}) = \mathcal{M}_1^a(\vec{d}) \cup \mathcal{M}_1^b(\vec{d})$, i.e., $\mathcal{M}_1(\vec{d})$ contains the positions occupied by $\text{deg}_1(\vec{d})$. Define $\mathcal{M}_i^a(\vec{d})$, $\mathcal{M}_i^b(\vec{d})$ and $\mathcal{M}_i(\vec{d})$ for $\text{deg}_i(\vec{d})$ accordingly.

To make the above definitions more formal, we can let

$$\mathcal{M}_{1,s}^a(\vec{d}) := \left\{ j \bmod n : 0 < j - \sum_{s=1}^{t-1} (|\vec{a}_s| + |\vec{b}_s|) \leq |\vec{a}_t| \right\},$$

and

$$\mathcal{M}_{1,s}^b(\vec{d}) := \left\{ j \bmod n : |\vec{a}_t| < j - \sum_{s=1}^{t-1} (|\vec{a}_s| + |\vec{b}_s|) \leq |\vec{a}_t| + |\vec{b}_t| \right\},$$

where $|\vec{a}_s|$ and $|\vec{b}_s|$ denote the length of \vec{a}_s and \vec{b}_s respectively. We can also define $\mathcal{M}_{i,s}^a(\vec{d})$ and $\mathcal{M}_{i,s}^b(\vec{d})$ accordingly. Then

$$\mathcal{M}_1^a(\vec{d}) := \bigcup_{s=1}^{\ell} \mathcal{M}_{1,s}^a(\vec{d}), \quad \text{and} \quad \mathcal{M}_1^b(\vec{d}) := \bigcup_{s=1}^{\ell} \mathcal{M}_{1,s}^b(\vec{d}).$$

We consider two cases — $\mathcal{M}_1^a(\vec{d})$ and $\mathcal{M}_i^a(\vec{d})$ overlap or not. Let

$$\mathcal{B}' = \left\{ (i, \vec{d}) \in \mathcal{B} : \mathcal{M}_1^a(\vec{d}) \cap \mathcal{M}_i^a(\vec{d}) = \emptyset \right\}.$$

Let $\mathcal{B}'' := \mathcal{B} \setminus \mathcal{B}'$. We first deal the easier case $(i, \vec{d}) \in \mathcal{B}''$.

Lemma 6.4. *Assume that $\mathbb{P}\{|\mathcal{T}^{gw}| = n\} > 0$ and $v = v(T) \leq n/2$. We have*

$$\sum_{(i, \vec{d}) \in \mathcal{B}''} \mathbb{P}\left\{(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}\right\} = 2 \sum_{T' \in \{T \boxplus T\}} \pi^{nf}(T') \frac{\mathbb{P}\{S_{n-v(T')} = n - v(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}}.$$

Proof. Let $(i, \vec{d}) \in \mathcal{B}''$. Since $\mathcal{M}_1^a(\vec{d}) \cap \mathcal{M}_i^a(\vec{d}) \neq \emptyset$, when $(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}$, either $T_i((\tilde{\xi}_r^n)_{r=1}^n)$ is a fringe subtree of $T_1((\tilde{\xi}_r^n)_{r=1}^n)$ rooted at a node that corresponds to an internal node of T (regarding that $T < T_1((\tilde{\xi}_r^n)_{r=1}^n)$), or vice versa. Thus there exists a $T' \in \{T \boxplus T\}$ such that either $T' < T_1((\tilde{\xi}_r^n)_{r=1}^n)$ or $T' < T_i((\tilde{\xi}_r^n)_{r=1}^n)$. By symmetry of these two events, we have

$$\begin{aligned} & \sum_{(i, \vec{d}) \in \mathcal{B}''} \mathbb{P}\left\{(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}\right\} \\ &= 2 \sum_{T' \in \{T \boxplus T\}} \mathbb{P}\left\{T' < T_1(\tilde{\xi}_1^n, \dots, \tilde{\xi}_n^n)\right\} \\ &= 2 \sum_{T' \in \{T \boxplus T\}} \pi^{nf}(T') \frac{\mathbb{P}\{S_{n-v(T')} = n - v(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}}, \end{aligned} \quad (6.5)$$

where the last step follows from Lemma 6.2. \square

If $(i, \vec{d}) \in \mathcal{B}'$ and $(\tilde{\xi}_r^n)_{r=1}^n = \vec{d}$, we have three cases as shown by Figure 7:

- (i) $T_1((\tilde{\xi}_r^n)_{r=1}^n)$ and $T_i((\tilde{\xi}_r^n)_{r=1}^n)$ do not overlap.
- (ii) $T_1((\tilde{\xi}_r^n)_{r=1}^n)$ contains $T_i((\tilde{\xi}_r^n)_{r=1}^n)$, and the root of the latter does not correspond to an internal node of T (regarding that $T < T_1((\tilde{\xi}_r^n)_{r=1}^n)$).
- (iii) The reverse of case (ii).

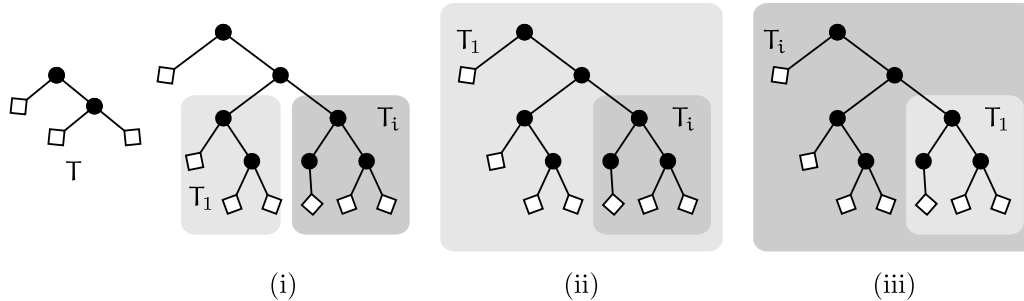


Figure 7: Examples of three cases in \mathcal{B}'' , with $T_1 = T_1((\tilde{\xi}_r^n)_{r=1}^n)$ and $T_i = T_i((\tilde{\xi}_r^n)_{r=1}^n)$

We formally state the three cases as follows:

Recall that $v(T)$ and $\ell(T)$ are the numbers of internal nodes and leaves of T .

Recall that $T < T'$ denotes that T' has a non-fringe subtree of shape T at its root.

Lemma 6.5. *If $(i, \vec{d}) \in \mathcal{B}'$, then one of the following must be true:*

- (i) $\mathcal{M}_1(\vec{d}) \cap \mathcal{M}_i(\vec{d}) = \emptyset$.
- (ii) $\mathcal{M}_i(\vec{d}) \subseteq \mathcal{M}_{1,s}^b(\vec{d})$ for some $1 \leq s \leq \ell$.
- (iii) $\mathcal{M}_1(\vec{d}) \subseteq \mathcal{M}_{i,s}^b(\vec{d})$ for some $1 \leq s \leq \ell$.

It turns out we can unify the three cases:

The indices are
modulo n .

Lemma 6.6. *Let $(s, (e_1, \dots, e_n)) \in \{v+1, \dots, n\} \times \mathcal{R}_n$ be a pair such that*

$$(e_1, e_2, \dots, e_v) = (\vec{a}_1, \dots, \vec{a}_\ell), \quad \text{and} \quad (e_s, e_{s+1}, \dots, e_{s+v-1}) = (\vec{a}_1, \dots, \vec{a}_\ell).$$

Let \mathcal{B}^* be the set of such pairs. We have

$$\sum_{(i, \vec{d}) \in \mathcal{B}'} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = \sum_{(i, \vec{e}) \in \mathcal{B}^*} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{e} \right\}.$$

Proof. Let $(i, \vec{d}) \in \mathcal{B}'$ with $\vec{d} = (d_1, \dots, d_n)$. Arrange d_1, \dots, d_n in a cycle. Paint the segment $\text{deg}_1(\vec{d})$ red and the segment $\text{deg}_i(\vec{d})$ blue.

For case (i) and (ii) of Lemma 6.5, we first permute the red segment from $(\vec{a}_1, \vec{b}_1, \dots, \vec{a}_\ell, \vec{b}_\ell)$ to $(\vec{a}_1, \dots, \vec{a}_\ell, \vec{b}_1, \dots, \vec{b}_\ell)$. Then we permute the blue segment of from $(\vec{a}_1, \vec{b}'_1, \dots, \vec{a}_\ell, \vec{b}'_\ell)$ to $(\vec{a}_1, \dots, \vec{a}_\ell, \vec{b}'_1, \dots, \vec{b}'_\ell)$. It is clear this can be done in case (i). For case (ii), the first step of the permutation moves the blue segment but does not change its contents. So we can carry out the second step without problem.

In case (iii), we reverse the order of the two steps. After this the starting position of the red segment may have changed. We rotate the new sequence such that the red segment still starts from position 1. In all three cases, we get a new pair $(s, (e_1, \dots, e_n)) \in \mathcal{B}^*$. So this permutation defines a mapping $f: \mathcal{B}' \rightarrow \mathcal{B}^*$.

Given $(s, \vec{e}) = (s, (e_1, \dots, e_n)) \in \mathcal{B}$, by Lemma 6.1, we can find ℓ consecutive tree degree sequences $\vec{g}_1, \dots, \vec{g}_\ell$ after e_1, \dots, e_v and ℓ consecutive tree degree sequences $\vec{g}'_1, \dots, \vec{g}'_\ell$ after e_s, \dots, e_{s+v-1} . Also by Lemma 6.1,

$$\text{deg}_1(\vec{e}) = (e_1, \dots, e_v, \vec{g}_1, \dots, \vec{g}_\ell)$$

and

$$\text{deg}_s(\vec{e}) = (e_s, \dots, e_{s+v-1}, \vec{g}'_1, \dots, \vec{g}'_\ell)$$

are both tree degree sequences. If the the two sequences do not overlap, then we are in case (i). If $\text{deg}_1(\vec{e})$ contains $\text{deg}_s(\vec{e})$, then we are in case (ii), otherwise we are in case (iii).

In other words, we can reverse the permutation and get a unique $(i, \vec{d}) \in \mathcal{B}'$ such that $f((i, \vec{d})) = (s, \vec{e})$. Thus f is a one-to-one mapping. And since $(\tilde{\xi}_r^n)_{r=1}^n$ is permutation invariant, we have

$$\mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{e} \right\}.$$

The Lemma follows by summing over all pairs in \mathcal{B}' . \square

Lemma 6.7. *Assume that $\mathbb{P} \{ |\mathcal{T}^{gw}| = n \} > 0$ and $v = v(T) \leq n/2$. We have*

$$\begin{aligned} & \sum_{(i, \vec{d}) \in \mathcal{B}'} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\ &= (n - 2v + 1) \pi^{nf}(T)^2 \frac{\mathbb{P} \{ S_{n-2v} = n + 1 - 2(v + \ell) \}}{\mathbb{P} \{ S_n = n - 1 \}}. \end{aligned} \quad (6.6)$$

Proof. Given $(i, (d_1, \dots, d_n)) \in \mathcal{B}^*$, we can move the segment (d_i, \dots, d_{i+v-1}) to the position $v + 1$ to get a new sequence $(d'_1, \dots, d'_n) \in \mathcal{C}$, where

$$\mathcal{C} := \{ (e_1, \dots, e_n) \in \mathcal{R}_n : (e_1, \dots, e_{2v}) = (\vec{a}_1, \dots, \vec{a}_\ell, \vec{a}_1, \dots, \vec{a}_\ell) \}.$$

Since there are $n - 2v + 1$ possible values of i , this permutation gives us a $(n - 2v + 1)$ -to-one mapping $h : \mathcal{B}^* \rightarrow \mathcal{C}$. Since $(\tilde{\xi}_r^n)_{r=1}^n$ is permutation invariant, if $\vec{d}' = h((i, \vec{d}))$, then

$$\mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d}' \right\}.$$

Therefore

$$\sum_{(i, \vec{d}) \in \mathcal{B}^*} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = (n - 2v + 1) \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\}.$$

We obtain as usual

$$\begin{aligned} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\} &= \mathbb{P} \left\{ (\tilde{\xi}_1^n, \dots, \tilde{\xi}_{2v}^n) = (\vec{a}_1, \dots, \vec{a}_\ell, \vec{a}_1, \dots, \vec{a}_\ell) \right\} \\ &= \mathbb{P} \left\{ (\xi_1, \dots, \xi_{2v}) = (\vec{a}_1, \dots, \vec{a}_\ell, \vec{a}_1, \dots, \vec{a}_\ell) \mid S_n = n - 1 \right\} \\ &= \mathbb{P} \left\{ (\xi_1, \dots, \xi_{2v}) = (\vec{a}_1, \dots, \vec{a}_\ell, \vec{a}_1, \dots, \vec{a}_\ell) \right\} \\ &\quad \times \frac{\mathbb{P} \{ S_{n-2v} = (n - 1) - 2(v + \ell - 1) \}}{\mathbb{P} \{ S_n = n - 1 \}} \\ &= \pi^{nf}(T)^2 \frac{\mathbb{P} \{ S_{n-2v} = n + 1 - 2(v + \ell) \}}{\mathbb{P} \{ S_n = n - 1 \}}. \end{aligned}$$

It follows from [Lemma 6.6](#) that

$$\begin{aligned}
& \sum_{(i, \vec{d}) \in \mathcal{B}'} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\
&= \sum_{(i, \vec{d}) \in \mathcal{B}^*} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\
&= (n - 2\nu + 1) \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\} \\
&= (n - 2\nu + 1) \pi^{\text{nf}}(\mathbb{T})^2 \frac{\mathbb{P} \{S_{n-2\nu} = n + 1 - 2(\nu + \ell)\}}{\mathbb{P} \{S_n = n - 1\}}. \quad \square
\end{aligned}$$

Putting everything together, we have:

Lemma 6.8. *Assume that $\mathbb{P} \{|\mathcal{J}^{\text{gw}}| = n\} > 0$ and $\nu = \nu(\mathbb{T}) \leq n/2$. We have*

$$\begin{aligned}
\mathbb{E}(N_{\mathbb{T}}^{\text{nf}}(\mathcal{J}_n^{\text{gw}}))_2 &= n(n - 2\nu + 1) \pi^{\text{nf}}(\mathbb{T})^2 \frac{\mathbb{P} \{S_{n-2\nu} = n + 1 - 2(\nu + \ell)\}}{\mathbb{P} \{S_n = n - 1\}} \\
&\quad + 2n \sum_{\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}} \pi^{\text{nf}}(\mathbb{T}') \frac{\mathbb{P} \{S_{n-\nu(\mathbb{T}')} = n - \nu(\mathbb{T}') - \ell(\mathbb{T}')\}}{\mathbb{P} \{S_n = n - 1\}}.
\end{aligned}$$

Proof. The lemma follows by combining [Lemma 6.3](#), [6.4](#), and [6.7](#):

$$\begin{aligned}
\frac{\mathbb{E}(N_{\mathbb{T}}^{\text{nf}}(\mathcal{J}_n^{\text{gw}}))_2}{n} &= \sum_{(i, \vec{d}) \in \mathcal{B}} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\
&= \sum_{(i, \vec{d}) \in \mathcal{B}'} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} + \sum_{(i, \vec{d}) \in \mathcal{B}''} \mathbb{P} \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} \\
&= (n - 2\nu + 1) \pi^{\text{nf}}(\mathbb{T})^2 \frac{\mathbb{P} \{S_{n-2\nu} = n + 1 - 2(\nu + \ell)\}}{\mathbb{P} \{S_n = n - 1\}} \\
&\quad + 2 \sum_{\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}} \pi^{\text{nf}}(\mathbb{T}') \frac{\mathbb{P} \{S_{n-\nu(\mathbb{T}')} = n - \nu(\mathbb{T}') - \ell(\mathbb{T}')\}}{\mathbb{P} \{S_n = n - 1\}}. \quad \square
\end{aligned}$$

6.3 PROOF OF THEOREM [1.4](#)

Recall that $\nu(\mathbb{T})$ and $\ell(\mathbb{T})$ are the numbers of internal nodes and leaves of \mathbb{T} .

Let $(\mathbb{T}_n)_{n \geq 1}$ be a sequence of trees. Let $\nu_n := \nu(\mathbb{T}_n)$ and $\ell_n := \ell(\mathbb{T}_n)$. This section proves [Theorem 1.4](#). We restate it here:

Theorem 1.4. *Assume [Condition A](#). Let \mathbb{T}_n be a sequence of trees with $|\mathbb{T}_n| = k_n$ where $k_n \rightarrow \infty$ and $k_n = o(n)$. We have*

(i) *If $n\pi^{\text{nf}}(\mathbb{T}_n) \rightarrow 0$, then $N_{\mathbb{T}_n}^{\text{nf}}(\mathcal{J}_n^{\text{gw}}) = 0$ whp.*

(ii) *If $n\pi^{\text{nf}}(\mathbb{T}_n) \rightarrow \infty$, then*

$$\frac{N_{\mathbb{T}_n}^{\text{nf}}(\mathcal{J}_n^{\text{gw}})}{n\pi^{\text{nf}}(\mathbb{T}_n)} \xrightarrow{\mathbb{P}} 1.$$

Part (i) of the Theorem follows from [Lemma 6.9](#)

Lemma 6.9. *Assume Condition A. If $|\mathbb{T}_n| = o(n)$ and $\pi^{nf}(\mathbb{T}_n) > 0$ for all $n \in \mathbb{N}$, then*

$$\frac{\mathbb{E}N_{\mathbb{T}_n}^{nf}(\mathcal{J}_n^{gw})}{n\pi^{nf}(\mathbb{T}_n)} \rightarrow 1.$$

Thus if $n\pi^{nf}(\mathbb{T}_n) \rightarrow 0$, then $N_{\mathbb{T}_n}^{nf}(\mathcal{J}_n^{gw}) \xrightarrow{P} 0$.

Proof. $|\mathbb{T}_n| = o(n)$ implies that $v_n = o(n)$ and $\ell_n = o(n)$. Therefore it follows from [Corollary 2.2](#) and [Lemma 6.2](#) that

$$\frac{\mathbb{E}N_{\mathbb{T}_n}^{nf}(\mathcal{J}_n^{gw})}{n\pi^{nf}(\mathbb{T}_n)} = \frac{\mathbb{P}\{S_{n-v_n} = n - v_n - \ell_n\}}{\mathbb{P}\{S_n = n - 1\}} \rightarrow 1. \quad \square$$

Lemma 6.10. *Assume Condition A. If $|\mathbb{T}_n| = o(n)$ and $n\pi^{nf}(\mathbb{T}_n) \rightarrow \infty$, then*

$$\frac{\mathbb{E}(N_{\mathbb{T}_n}^{nf}(\mathcal{J}_n^{gw}))_2}{(n\pi^{nf}(\mathbb{T}_n))^2} \rightarrow 1.$$

Proof. Let $v = v_n$, $\ell = \ell_n$ and $\mathbb{T} = \mathbb{T}_n$. Since $|\mathbb{T}_n| = v + \ell$, we have $v = o(n)$ and $\ell = o(n)$. If $\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}$, then $v(\mathbb{T}') < 2v = o(n)$ and $\ell(\mathbb{T}') < 2\ell = o(n)$. Therefore, it follows from [Corollary 2.2](#) and [Lemma 6.8](#) that

$$\begin{aligned} \mathbb{E}(N_{\mathbb{T}}^{nf}(\mathcal{J}_n^{gw}))_2 &= n(n - 2v + 1)\pi^{nf}(\mathbb{T})^2 \frac{\mathbb{P}\{S_{n-2v} = n + 1 - 2(v + \ell)\}}{\mathbb{P}\{S_n = n - 1\}} \\ &\quad + 2n \sum_{\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}} \pi^{nf}(\mathbb{T}') \frac{\mathbb{P}\{S_{n-v(\mathbb{T}')} = n - v(\mathbb{T}') - \ell(\mathbb{T}')\}}{\mathbb{P}\{S_n = n - 1\}} \\ &= (1 + o(1))(n\pi^{nf}(\mathbb{T}))^2 + O(n) \sum_{\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}} \pi^{nf}(\mathbb{T}'). \end{aligned}$$

Thus it suffices to show that

$$n \sum_{\mathbb{T}' \in \{\mathbb{T} \boxplus \mathbb{T}\}} \pi^{nf}(\mathbb{T}') = o(n\pi^{nf}(\mathbb{T}))^2.$$

Consider the superset \mathcal{H} of $\{\mathbb{T} \boxplus \mathbb{T}\}$ that contains trees which can be obtained by replacing a proper non-leaf subtree of \mathbb{T} with another copy of \mathbb{T} . (We do not restrict where this replacement can happen as in the definition of $\{\mathbb{T} \boxplus \mathbb{T}\}$.) Note that $|\mathcal{H}| = v - 1$, since \mathbb{T} has v internal nodes and one of them is the root.

If $\mathbb{T}' \in \mathcal{H}$, then \mathbb{T}' contains \mathbb{T} as a fringe subtree. Thus $\pi^{nf}(\mathbb{T}') \leq \pi^{nf}(\mathbb{T})$. In the case that v is bounded, we have

$$n \sum_{\mathbb{T}' \in \mathcal{H}} \pi^{nf}(\mathbb{T}') \leq nv\pi^{nf}(\mathbb{T}) = O(n\pi^{nf}(\mathbb{T})) = o(n\pi^{nf}(\mathbb{T}))^2.$$

Thus we can assume that $v \rightarrow \infty$.

For $T' \in \mathcal{H}$, if T' has at least $3v/2$ internal nodes, call T' *big*, otherwise call it *small*. Let \mathcal{H}_b and \mathcal{H}_s be the sets of big and small trees in \mathcal{H} respectively.

If $T' \in \mathcal{H}_b$, then besides internal nodes that correspond to internal nodes of T , T' contains at least $v/2$ extra internal nodes. So we have $\pi^{nf}(T') \leq \pi^{nf}(T)p_{\max}^{v/2}$. Since $v \rightarrow \infty$ and $p_{\max} < 1$, $vp_{\max}^{v/2} = o(1)$. Using that $|\mathcal{H}| < v$, we have

Recall that $p_{\max} := \max_{i \geq 0} p_i < 1$.

$$n \sum_{T' \in \mathcal{H}_b} \pi^{nf}(T') \leq nv\pi^{nf}(T)p_{\max}^{v/2} = o(n\pi^{nf}(T)).$$

Let $T_{i,j}$ be a fringe subtree in T whose root is at depth i and is the j -th node of this level. If replacing $T_{i,j}$ with a copy of T makes a new tree T'_i that has strictly less than $3v/2$ internal nodes, then $T_{i,j}$ must contain more than $v/2$ internal nodes. Therefore, for each i , there is at most one possible such j . For an example of T'_i , see Figure 8.

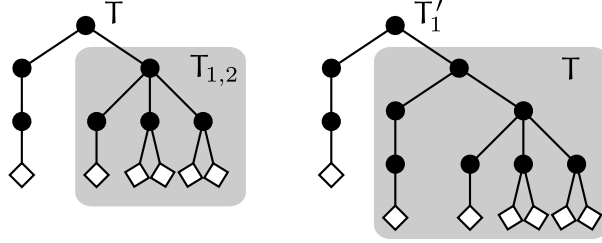


Figure 8: Example of T'_i for T with 7 internal nodes

As T has v internal nodes, there are at most $v - 1$ possible i that can make $T_{i,j}$ a proper and non-leaf subtree. Since T'_i has at least i internal nodes besides these in the copy of T that replaced $T_{i,j}$, we have $\pi^{nf}(T'_i) \leq \pi^{nf}(T)p_{\max}^i$. In summary, we have

We can assume T'_i always exists since we only need an upper bound.

$$n \sum_{T' \in \mathcal{H}_s} \pi^{nf}(T') \leq n \sum_{i=1}^v \pi^{nf}(T'_i) \leq n \sum_{i=1}^v \pi^{nf}(T)p_{\max}^i \leq O(n\pi^{nf}(T)).$$

Therefore,

$$\begin{aligned} n \sum_{T' \in \{T \boxplus T\}} \pi^{nf}(T') &\leq n \sum_{T' \in \mathcal{H}_b} \pi^{nf}(T') + n \sum_{T' \in \mathcal{H}_s} \pi^{nf}(T') \\ &= o(n\pi^{nf}(T)) + O(n\pi^{nf}(T)) = o(n\pi^{nf}(T))^2. \quad \square \end{aligned}$$

Part (ii) of Theorem 1.4 is contained in the following lemma:

Lemma 6.11. *Assume Condition A. If $|T_n| = o(n)$ and $n\pi^{nf}(T_n) \rightarrow \infty$, then*

$$\frac{\text{Var}\left(\mathbb{N}_{T_n}^{nf}(\mathcal{J}_n^{gw})\right)}{(\mathbb{E}\mathbb{N}_{T_n}^{nf}(\mathcal{J}_n^{gw}))^2} \rightarrow 0.$$

Therefore,

$$\frac{\mathbb{N}_{T_n}^{nf}(\mathcal{J}_n^{gw})}{n\pi^{nf}(T_n)} \xrightarrow{p} 1.$$

Proof. Let $X_n = N_{T_n}^{nf}(\mathcal{T}_n^{gw})$. By Lemma 6.9,

$$\mathbb{E}X_n \sim n\pi^{nf}(T_n) \rightarrow \infty.$$

By Lemma 6.10, we have $\mathbb{E}(X_n)_2 = (1 + o(1))(n\pi^{nf}(T_n))^2$. Therefore,

$$\begin{aligned} \text{Var}(X_n) &= \mathbb{E}[(X_n)_2] + \mathbb{E}[X_n] - (\mathbb{E}[X_n])^2 \\ &= (1 + o(1))(n\pi^{nf}(T_n))^2 \\ &\quad + (1 + o(1))(n\pi^{nf}(T_n)) - (1 + o(1))(n\pi^{nf}(T_n))^2 \\ &= o(n\pi^{nf}(T_n))^2 + o(\mathbb{E}[X_n]^2). \end{aligned}$$

Thus by Chebyshev's inequality, for all $\varepsilon > 0$

$$\mathbb{P}\left\{\left|\frac{X_n}{\mathbb{E}X_n} - 1\right| > \varepsilon\right\} \leq \frac{\text{Var}(X_n)}{\varepsilon^2 \mathbb{E}[X_n]^2} \rightarrow 0.$$

In other words, $X_n/\mathbb{E}[X_n] \xrightarrow{p} 1$. □

6.4 COMPLETE r -ARY NON-FRINGE SUBTREES

Let T_h^{r-ary} denote the complete r -ary tree of height h . Let $\bar{H}_{n,r}$ be the maximal h such that T_h^{r-ary} appears as a non-fringe subtree in \mathcal{T}_n^{gw} .

Lemma 6.12. *Assume Condition A and $p_r > 0$ for some $r \geq 2$. Let $h_n \rightarrow \infty$ be a sequence of positive integers. Let*

$$\alpha_r = \log_r \left(\frac{1}{r-1} \log \frac{1}{p_r} \right).$$

Let $\omega_n \rightarrow \infty$ be an arbitrary sequence. We have:

- (i) *If $h_n \leq \log_r(\log n - \omega_n) - \alpha_r$, then whp \mathcal{T}_n^{gw} contains T_h^{r-ary} as non-fringe subtrees.*
- (ii) *If $h_n \geq \log_r(\log n + \omega_n) - \alpha_r$, then whp \mathcal{T}_n^{gw} does not contain T_h^{r-ary} as non-fringe subtrees.*

As a result,

$$\bar{H}_{n,r} - \log_r \log n \xrightarrow{p} -\alpha_r.$$

Proof. Let $h = h_n$. The tree T_h^{r-ary} has

$$v_n = \frac{r^h - 1}{r - 1} = \frac{r^h}{r - 1} + O(1)$$

internal vertices, which all have degree r . Since $\pi^{nf}(\mathbb{T}_h^{r\text{-ary}}) = p_r^{v_n}$, we have

$$\begin{aligned} \log \frac{1}{\pi^{nf}(\mathbb{T}_h^{r\text{-ary}})} &= v_n \log \frac{1}{p_r} \\ &= \frac{r^h}{r-1} \log \frac{1}{p_r} + O(1) \\ &= r^h r^{\alpha_r} + O(1) \\ &\leq \frac{\log n - \omega_n}{r^{\alpha_r}} r^{\alpha_r} + O(1) \\ &= \log n - \omega_n + O(1), \end{aligned}$$

where the inequality comes from the assumption of h . Thus

$$n\pi^{nf}(\mathbb{T}_h^{r\text{-ary}}) \geq e^{\omega_n - O(1)} \rightarrow \infty.$$

It follows from Theorem 1.4 that $N_{\mathbb{T}_h^{r\text{-ary}}}^{nf}(\mathcal{T}_n^{gw}) \geq 1$ whp. Thus (i) is proved.

Similar computations show that for (ii) $n\pi^{nf}(\mathbb{T}_h^{r\text{-ary}}) \rightarrow 0$, which implies that $N_{\mathbb{T}_h^{r\text{-ary}}}^{nf}(\mathcal{T}_n^{gw}) = 0$ whp by Theorem 1.4. The last statement of the lemma follows directly from (i) and (ii). \square

Lemma 6.13. *Assume Condition A and $p_1 > 0$. Let $\omega_n \rightarrow \infty$ be an arbitrary sequence.*

We have:

- (i) *If $h_n \leq (\log n - \omega_n)/\log \frac{1}{p_1}$, then whp \mathcal{T}_n^{gw} contains $\mathbb{T}_h^{1\text{-ary}}$ as non-fringe subtrees.*
- (ii) *If $h_n \geq (\log n + \omega_n)/\log \frac{1}{p_1}$, then whp \mathcal{T}_n^{gw} does not contain $\mathbb{T}_h^{1\text{-ary}}$ as non-fringe subtrees.*

Therefore

$$\frac{\bar{H}_{n,1}}{\log_{1/p_1}(n)} \xrightarrow{p} 1.$$

Proof. For (i), we have

$$\pi^{nf}(\mathbb{T}_h^{1\text{-ary}}) = p_1^h \geq \frac{e^{\omega_n}}{n}.$$

So $n\pi^{nf}(\mathbb{T}_h^{1\text{-ary}}) \rightarrow \infty$. It follows from Theorem 1.4 that $N_{\mathbb{T}_h^{1\text{-ary}}}^{nf} \geq 1$ whp. For (ii), we have $n\pi^{nf}(\mathbb{T}_h^{1\text{-ary}}) \rightarrow 0$, so it follows from Theorem 1.4 that $N_{\mathbb{T}_h^{1\text{-ary}}}^{nf} = 0$ whp. \square

Example 6.1 (The binary tree). Recall that when $p_0 = p_2 = 1/4$ and $p_1 = 1/2$, \mathcal{T}_n^{gw} is equivalent to a uniform random binary tree of size n . It follows from Lemma 6.13 that $\bar{H}_{n,1}/\log_2 n \xrightarrow{p} 1$. This result was previously proved by Devroye, Flajolet, Hurtado, Noy, and Steiger [33].

See Subsection 2.3.3.

6.5 DIFFERENCE BETWEEN FRINGE AND NON-FRINGE SUBTREES

In Theorem 1.4, we have shown that if

$$n\pi(\mathbb{T}_n) \rightarrow \mu \in [0, \infty),$$

then

$$d_{TV}(\mathbb{N}_{\mathbb{T}_n}(\mathcal{T}_n^{gw}), \text{Po}(\mu)) \rightarrow 0.$$

This is not true in general for non-fringe subtrees, as shown by the following lemma.

Lemma 6.14. *Assume Condition A. Let $L_{h(n)}$ be a chain (complete 1-ary tree) of height $h(n)$. Let $X_n = N_{L_{h(n)}}^{nf}(\mathcal{T}_n^{gw})$. If*

$$n\pi^{nf}(L_{h(n)}) := n\mathbb{P}\{L_{h(n)} < \mathcal{T}^{gw}\} \rightarrow \mu \in (0, \infty),$$

then

$$\mathbb{E}X_n \rightarrow \mu, \quad \text{Var}(X_n) \rightarrow \mu \frac{1+p_1}{1-p_1}, \quad \mathbb{E}(X_n - \mathbb{E}X_n)^3 \rightarrow \mu \frac{3p_1^2 + 2p_1 + 1}{(1-p_1)^2}.$$

As a result, $\liminf_{n \rightarrow \infty} d_{TV}(X_n, \text{Po}(\mu)) > 0$.

Proof. Let $h = h(n)$. Since $n\pi^{nf}(L_h) = np_1^h \rightarrow \mu \in (0, \infty)$, we have $h = \log_{1/p_1} n + O(1)$. L_h has h internal nodes and one leaf. Thus it follows from Lemma 6.2 and 2.7 that

$$\mathbb{E}X_n = n\pi^{nf}(L_h) \frac{\mathbb{P}\{S_{n-h} = n-h-1\}}{\mathbb{P}\{S_n = n-1\}} \rightarrow \mu.$$

Since $\{L_h \boxplus L_h\} = \{L_{h+i} : 1 \leq i \leq h-1\}$, by Lemma 2.7,

$$\begin{aligned} \zeta_1 &:= 2n \sum_{T' \in \{L_h \boxplus L_h\}} \pi^{nf}(T') \frac{\mathbb{P}\{S_{n-v(T')} = n-v(T') - \ell(T')\}}{\mathbb{P}\{S_n = n-1\}} \\ &= 2n \sum_{i=1}^{h-1} \pi^{nf}(L_{h+i}) \frac{\mathbb{P}\{S_{n-h-i} = n-h-i-1\}}{\mathbb{P}\{S_n = n-1\}} \\ &= (1+o(1))2n \sum_{i=1}^{h-1} p_1^{h+i} = (1+o(1))2np_1^h \sum_{i=1}^{h-1} p_1^i \rightarrow 2\mu \frac{p_1}{1-p_1}. \end{aligned}$$

We also have by Lemma 2.7,

$$\zeta_2 := n(n-2h)\pi^{nf}(L_h)^2 \frac{\mathbb{P}\{S_{n-2h} = n+1-2(h+1)\}}{\mathbb{P}\{S_n = n-1\}} \rightarrow \mu^2.$$

Therefore, it follows from Lemma 6.8 that

$$\mathbb{E}(X_n)_2 = \zeta_1 + \zeta_2 \rightarrow 2\mu \frac{p_1}{1-p_1} + \mu^2.$$

Recall that $\pi(\mathbb{T}) := \mathbb{P}\{\mathcal{T}^{gw} = \mathbb{T}\}$ and $N_{\mathbb{T}}(\mathcal{T}_n^{gw})$ is the number of fringe subtrees of shape \mathbb{T} in \mathcal{T}_n^{gw} .

Thus

$$\text{Var}(X_n) = \mathbb{E}[(X_n)_2] + \mathbb{E}[X_n] - \mathbb{E}[X_n]^2 \rightarrow \mu \frac{1+p_1}{1-p_1}.$$

So we have

$$\frac{\text{Var}(X_n)}{\mathbb{E}[X_n]} \rightarrow \frac{1+p_1}{1-p_1} > 1.$$

With an argument similar to Lemma 6.8, we can compute $\mathbb{E}[(X_n)_3]$, which yields

$$\mathbb{E}[(X_n - \mathbb{E}X_n)^3] \rightarrow \mu \frac{3p_1^2 + 2p_1 + 1}{(1-p_1)^2}.$$

Since

$$\begin{aligned} \mathbb{E}[|X_n - \mathbb{E}X_n|^3] &= 2\mathbb{E}[|X_n - \mathbb{E}X_n|^3 \times \mathbb{1}_{\{X_n < \mathbb{E}X_n\}}] + \mathbb{E}[(X_n - \mathbb{E}X_n)^3] \\ &\leq O(1) + \mathbb{E}[(X_n - \mathbb{E}X_n)^3], \end{aligned}$$

the above limit implies that

$$C := \limsup_{n \rightarrow \infty} \mathbb{E}[|X_n - \mathbb{E}X_n|^3] < \infty.$$

It follows from Lemma 3.8 that

$$d_{\text{TV}}(X_n, \text{Po}(\mathbb{E}X_n)) = \left(\frac{\text{Var}(X_n) - \mathbb{E}X_n}{\mathbb{E}(|X_n - \mathbb{E}X_n|^3)^{2/3}} \right)^3.$$

Therefore

$$\liminf_{n \rightarrow \infty} d_{\text{TV}}(X_n, \text{Po}(\mathbb{E}X_n)) \geq \frac{1}{C^2} \left(\frac{1+p_1}{1-p_1} \right)^3 > 0.$$

Since $\mathbb{E}X_n \rightarrow \mu$, we also have $\liminf_{n \rightarrow \infty} d_{\text{TV}}(X_n, \text{Po}(\mu)) > 0$. □

Part III

CONCLUSION

CHAPTER 7 GALTON-WATSON TREES AND DETERMINISTIC FINITE AUTOMATA CHOSEN AT RANDOM

We introduce our results on random k -out digraph which inspired the topic of this thesis.

CHAPTER 8 OPEN QUESTIONS

We summarize some open questions about subtrees in large conditional Galton-Watson trees.

GALTON-WATSON TREES AND DETERMINISTIC FINITE AUTOMATA CHOSEN AT RANDOM

The main results proved in this thesis ([Theorem 1.2](#), [1.3](#) and [1.4](#)) are essentially consequences of the fact that fringe subtrees in a large conditional Galton-Watson tree behave more or less like independent unconditional Galton-Watson trees. In this chapter, we briefly discuss the Deterministic Finite Automaton (DFA) chosen at random, i.e., the k -out digraph model, in which we can also find many tree structures that resemble iid copies of Galton-Watson trees. This has in fact motivated the research topic of this thesis.

This chapter is based on Cai and Devroye [[21](#)], and is organized as follows: [Section 7.1](#) defines some basic concepts of graphs and digraphs. [Section 7.2](#) introduces some well-known random graph and digraph models. [Section 7.3](#) discusses the history of the k -out digraph model. [Section 7.4](#) presents our results on k -out digraphs. [Section 7.5](#) explains the connection between random k -out digraphs and Galton-Watson trees and how it motivated this thesis.

“The greatest tool in the hands of the probabilist is independence”—Luc Devroye

7.1 GRAPHS AND DIGRAPHS

A *graph* is a pair $(\mathcal{V}, \mathcal{E})$ where \mathcal{V} is a set of vertices and \mathcal{E} is a multiset of *unordered* pairs of vertices in \mathcal{V} , which are called *edges*. For an edge $e = (u, v)$, u and v are called the *endpoints* of e . If $u = v$, then e is called a *self-loop*. The number of edges u belongs to is called the degree of u (counting each self-loop twice). If self-loops and multiple edges between two vertices are prohibited, then the graph is called a *simple* graph. A list of edges $(u_1, v_1), (u_2, v_2), \dots, (u_\ell, v_\ell)$ is called a *path* from u_1 to v_ℓ if $v_i = u_{i+1}$ for all $1 \leq i < \ell$. A path that does not visit any vertex twice is called *simple*. A graph is *connected* if there is path between every pair of vertices.

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph. Let $\mathcal{V}' \subseteq \mathcal{V}$. The *induced subgraph* $G[\mathcal{V}']$ is the graph $(\mathcal{V}', \mathcal{E}')$, where an edge e belongs to \mathcal{E}' if and only if $e \in \mathcal{E}$ and both of its endpoints are in \mathcal{V}' . $G[\mathcal{V}']$ is a *component* of G if it is connected and adding any vertex to \mathcal{V}' will make it disconnected.

A *directed graph* or *digraph* $(\mathcal{V}, \mathcal{A})$ is a pair where \mathcal{V} is a set of vertices and \mathcal{A} is a multiset of *ordered* pairs of vertices in \mathcal{V} , which are called *directed edges* or

arcs. For an arc $a = \{u, v\} \in \mathcal{A}$, u is its *starting vertex* and v is its *end vertex*. The number of arcs of which u is the starting (end) vertex is the *out-degree* (*in-degree*) of u . A list of arcs $\{u_1, v_1\}, \{u_2, v_2\}, \dots, \{u_\ell, v_\ell\}$ is called a *directed path* from u_1 to v_ℓ if $v_i = u_{i+1}$ for all $1 \leq i < \ell$. A digraph is *strongly connected* if there is a directed path from every vertex to all other vertices.

Let $D = (\mathcal{V}, \mathcal{A})$ be a digraph. Let $\mathcal{V}' \subseteq \mathcal{V}$. The *induced sub-digraph* $D[\mathcal{V}']$ is the digraph $(\mathcal{V}', \mathcal{A}')$ where an arc a belongs to \mathcal{A}' if and only if $a \in \mathcal{A}$ and both of its starting and end vertices are in \mathcal{V}' . $D[\mathcal{V}']$ is called a *strongly connected component* (scc) of D if it is a strongly connected and adding any vertex to \mathcal{V}' will make it not strongly connected.

For more basics of graph theory, see Diestel [34, chap. 1].

7.2 RANDOM GRAPHS AND DIGRAPHS

With the rapid growth of the internet, networks containing millions of nodes have become common. These include the World Wide Web, social networks, and Peer-to-peer networks. Among computer scientists and mathematicians, there is great interest in analyzing such networks using graphs as models. But for large and complex networks, it is difficult to describe their graph structure deterministically. So random graphs, i.e., graphs constructed according to some probabilistic rules, are often used instead.

One of the earliest random graph models is the $G(n, m)$, that is, a simple graph of n vertices and m edges chosen uniformly at random from all such graphs. In 1960, Erdős and Rényi [39] proved some surprising results of $G(n, m)$, which started the field of random graphs. A similar model $G(n, p)$ was introduced by Gilbert [51]. It is defined as a simple graph of n vertices in which each possible edge appears with probability p independently.

Another example which is of particular interest in recent years is the configuration model introduced by Bollobás [16]. In this model, each vertex is assigned a deterministic number of half-edges. Then pairs of half-edges are connected uniformly at random to form edges. It has been proved a very useful tool for analyzing many other random graphs. For example, Molloy and Reed [80] studied the conditions for the configuration model to have a component of linear size and their results can be also applied to $G(n, m)$ and $G(n, p)$.

The counterpart of $G(n, m)$ and $G(n, p)$ for digraphs, denoted by $D(n, m)$ and $D(n, p)$, were studied by Łuczak [75] and Karp [66] respectively. There is also

a directed version of the configuration model which was introduced by Cooper and Frieze [30].

There are many classic textbooks on random graphs. See, e.g., Bollobás [15], Janson, Łuczak, and Rucinski [60], van der Hofstad [93], and Frieze and Karoński [47].

7.3 RANDOM DFA

The *deterministic finite automaton* (DFA) is widely used in computational complexity theory. Formally, a DFA is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$, where Q is a finite set called the set of states, Σ is a finite set called the alphabet, $\delta : Q \times \Sigma \rightarrow Q$ is the transition function, $q_0 \in Q$ is the start state, and $F \subseteq Q$ is the set of accept states.

If q_0 and F are ignored, a DFA with n states and a alphabet of size k can be seen as a digraph with vertex set $\{1, \dots, n\}$ in which each vertex has k out-arcs labeled by $1, \dots, k$, which we call a k -out digraph. Note that such a digraph can have self-loops and multiple arcs from one vertex to another. For an introduction to DFA and its applications, see Sipser [90]. See Figure 9 for examples of k -out digraphs.

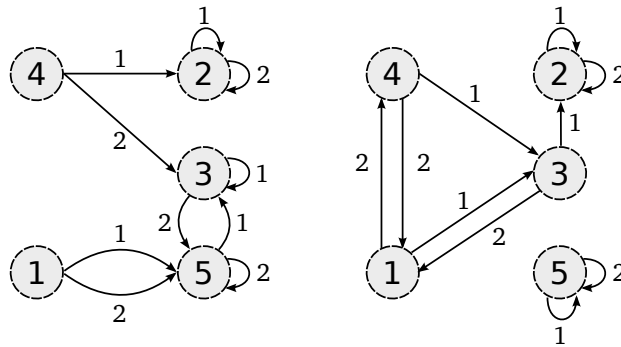


Figure 9: Examples of 2-out digraph with 5 vertices

Let $D_{n,k}$ denote a digraph chosen uniformly at random from all k -out digraphs of n vertices. Equivalently $D_{n,k}$ is a random k -out digraph of n vertices with the endpoints of its kn arcs chosen independently and uniformly at random.

When $k = 1$, $D_{n,k}$ is equivalent to a uniform random mapping from $\{1, \dots, n\}$ to itself, which has been well studied by Kolchin [71], Flajolet and Odlyzko [46], and Aldous and Pitman [7]. In $D_{n,1}$, the largest strongly connected component (scc) has expected size $\Theta(\sqrt{n})$, and so does the size of the longest cycle. However, as shown later, for $k \geq 2$, the largest scc has expected size $\Theta(n)$.

From now on we assume that $k \geq 2$. Let \mathcal{S}_u (the *spectrum* of u) be the set of vertices in $D_{n,k}$ that are reachable from vertex u , including u itself. In 1973 Grusho [54] proved that $(|\mathcal{S}_1| - \nu_k n) / \sigma_k \sqrt{n}$ converges in distribution to the standard normal distribution, where ν_k and σ_k are explicitly defined constants.

Given a set of vertices $\mathcal{S} \subseteq \{1, \dots, n\}$, call \mathcal{S} *closed* if there does not exist an arc with starting vertex in \mathcal{S} and end vertex in $\mathcal{S}^c := \{1, \dots, n\} \setminus \mathcal{S}$. Let \mathcal{G}_n be the largest closed set of vertices such that $D_{n,k}[\mathcal{S}]$ is a scc. (To avoid ambiguity, if the largest closed scc is not unique, let \mathcal{G}_n be the vertex set of the largest closed scc that contains the smallest vertex-label.) We call \mathcal{G}_n the *giant*.

Grusho also proved that $|\mathcal{G}_n|$ has the same limit distribution as $|\mathcal{S}_1|$ by showing that whp \mathcal{G}_n is reachable from all vertices and that $|\mathcal{S}_1| - |\mathcal{G}_n| = o_p(\sqrt{n})$ (see [59] for the notation). His proof relies on a result by Sevast'yanov [89] which approximates the exploration of \mathcal{S}_1 with a Gaussian process.

In 2012 Carayol and Nicaud [23] proved a local limit theorem for $|\mathcal{S}_1|$ by analyzing the limit behavior of the probability that $|\mathcal{S}_1| = s$ for an s close to $\nu_k n$. Their proof depends on a theorem by Korshunov [72] which says that conditioned on every vertex having in-degree at least one, the probability that $\mathcal{S}_1 = \{1, \dots, n\}$ tends to some constant. Carayol and Nicaud derived a simple and explicit formula of this constant from their theorem. (The same formula is also proved by Lebensztayn [74] using Lagrange series.)

Both [89] and [72] are not available in English though.

Lately the simple random walk (SRW) on $D_{n,k}$ has gained some attention for its applications in machine learning. Addario-Berry, Balle, and Perarnau [1] studied the stationary distribution of the SRW by analyzing the distances in $D_{n,k}$. They proved that the diameter and the typical distance, rescaled by $\log n$, converge in probability to explicit constants. Angluin and Chen [10] studied the rate of the convergence to the stationary distribution of the SRW. They also suggested an algorithm for learning a uniformly random DFA under Kearns' statistical query model [67].

7.4 OUR RESULTS

This section summarizes our own results for k -out digraphs from Cai and Devroye [21].

7.4.1 The one-in-core and the giant

A digraph can be uniquely decomposed into sccs which form a directed acyclic graph (DAG) through a process called condensation that contracts every scc into a single vertex while keeping all the arcs between them (see [53, chap. 3.2]). Let $D_{n,k}^A$ denote the condensation DAG of $D_{n,k}$.

Let $\mathcal{G}_n^c := \{1, \dots, n\} \setminus \mathcal{G}_n$, i.e., \mathcal{G}_n^c is the set of vertices that are outside the giant. The structure of $D_{n,k}^A$ depends on $D_{n,k}[\mathcal{G}_n^c]$, the digraph induced by \mathcal{G}_n^c . Our analysis shows that in $D_{n,k}[\mathcal{G}_n^c]$ the total number of cycles and the number of cycles of a fixed length both converge to Poisson distributions with constant means. So the number of cycles and the length of the longest cycle are both $O_p(1)$ (see [59] for the notation). Furthermore, these cycles are vertex-disjoint whp. Therefore, almost every vertex in \mathcal{G}_n^c is a scc itself and $D_{n,k}^A$ is very much like $D_{n,k}$ with the giant contracted into a single vertex.

The d -core of an undirected graph is the maximum induced subgraph in which all vertices have degree at least d . Similarly the d -in-core of a digraph can be defined as the maximum induced sub-digraph in which all vertices have in-degree at least d . Let \mathcal{O}_n denote the set of vertices in the one-in-core of $D_{n,k}$.

Note that $\mathcal{G}_n \subseteq \mathcal{O}_n$ since a scc induces a sub-digraph with each vertex having in-degree at least one. Also note that cycles cannot exist outside \mathcal{O}_n , for otherwise they contradict the maximality of \mathcal{O}_n . Now assume that every vertex can reach \mathcal{G}_n , which happens whp by Grusho [54]. Then $D_{n,k}$ can be divided into three layers: the center is \mathcal{G}_n ; then comes $\mathcal{O}_n \setminus \mathcal{G}_n$, which consists of cycles outside \mathcal{G}_n and paths from these cycles to \mathcal{G}_n ; the outermost is $\mathcal{O}_n^c := \{1, \dots, n\} \setminus \mathcal{O}_n$, which is acyclic.

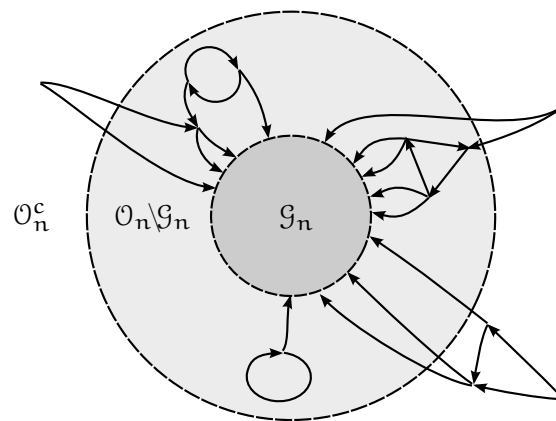


Figure 10: Three layers of $D_{n,k}$: the giant \mathcal{G}_n ; the one-in-core \mathcal{O}_n ; and the whole graph.

Since there cannot be many vertices in cycles outside the giant, the middle layer $\mathcal{O}_n \setminus \mathcal{G}_n$ must be very “thin”. Thus if we can prove $(|\mathcal{O}_n| - \nu_k n) / \sqrt{n}$ converges to a normal distribution, then we can also prove it for $|\mathcal{G}_n|$. The event $|\mathcal{O}_n| = s$ happens if and only if there is a set of vertices \mathcal{S} with $|\mathcal{S}| = s$ such that:

- (i) $D_{n,k}[\mathcal{S}]$, the sub-digraph induced by \mathcal{S} , has minimum in-degree one (*surjective*) and there are no arcs going from \mathcal{S} to \mathcal{S}^c (*closed*), which we refer to as \mathcal{S} being a *k-surjection* (since $D_{n,k}[\mathcal{S}]$ is equivalent to a surjective function from $\{1, \dots, ks\}$ to $\{1, \dots, s\}$);
- (ii) $D_{n,k}[\mathcal{S}^c]$ is acyclic.

The probability of (i) can be computed by counting the number of surjective functions. And we are able to show that the probability of (ii) converges to a constant. Note that for a fixed set \mathcal{S} , (i) and (ii) are independent because they depend on the endpoints of two disjoint sets of arcs. Thus we can get the limit of $\mathbb{P}\{|\mathcal{O}_n| = s\}$. Since the one-in-core of a digraph is unique,

$$\mathbb{P}\{|\mathcal{O}_n| = s\} = \sum_{\mathcal{S} \subseteq \{1, \dots, n\}; |\mathcal{S}|=s} \mathbb{P}\{\mathcal{O}_n = \mathcal{S}\}.$$

Thus we can compute the characteristic function of $(|\mathcal{O}_n| - \nu_k n) / \sqrt{n}$ and prove the following theorem:

Recall that $N(0, 1)$
is the standard
normal distribution.

Theorem 7.1 (Cai and Devroye [21, thm. 1]). *We have as $n \rightarrow \infty$,*

$$\begin{aligned} \frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} &\xrightarrow{d} N(0, 1), \\ \frac{|\mathcal{G}_n| - \nu_k n}{\sigma_k \sqrt{n}} &\xrightarrow{d} N(0, 1), \\ \frac{\max_{u \in \{1, \dots, n\}} |\mathcal{S}_u| - \nu_k n}{\sigma_k \sqrt{n}} &\xrightarrow{d} N(0, 1), \end{aligned}$$

where ν_k and σ_k are constants defined by

$$\nu_k := \frac{\tau_k}{k}, \quad \sigma_k^2 := \frac{\tau_k}{k e^{\tau_k} (1 - k e^{-\tau_k})},$$

and τ_k is the unique positive solution of $1 - \tau_k/k - e^{-\tau_k} = 0$.

Remark 7.1. Although our formula for $\mathbb{P}\{|\mathcal{O}_n| = s\}$ resembles Carayol and Nicaud’s formula for $\mathbb{P}\{|\mathcal{S}_1| = s\}$, we actually prove the result from scratch without relying on previous work. Since we are able to derive explicit expressions of all the constants in our formula, the computation of the characteristic function becomes quite simple. Furthermore, to our knowledge this is the first self-contained proof.

Remark 7.2. We also showed that $|\mathcal{O}_n| - |\mathcal{G}_n| = O_p(1)$ [21, lem. 9]. The intuition is that a digraph with minimal in-degree and out-degree at least one is likely to have a large scc. This phenomenon is also observed in $D(n, p)$. Pittel and Poole [84, thm. 1.3] showed that in $D(n, p)$ the $(1, 1)$ -core—the maximal induced subdigraph in which each vertex has in-degree and out-degree at least one—differs from the largest scc in size by at most $O((\log n)^8)$, whp. This intuition is also used for studying the asymptotic counts of strongly connected digraphs (see Pérez-Giménez and Wormald [82] and Pittel [83]).

Recall that $D(n, p)$ is the directed version of $G(n, p)$.

7.4.2 Outside the giant

We also studied the part of $D_{n,k}$ outside the giant, which determines the structure of $D_{n,k}^A$ and supports the proof of Theorem 7.1. Our results are summarized in two theorems:

Theorem 7.2 (Cai and Devroye [21, thm. 2]). *We have:*

(i) *Let L_n be the length of the longest cycle in $D_{n,k}[\mathcal{G}_n^c]$. Then $L_n = O_p(1)$.*

(ii) *Let C_n be the number of cycles in $D_{n,k}[\mathcal{G}_n^c]$. Then*

$$C_n \xrightarrow{d} \text{Po} \left(\log \frac{1}{1 - ke^{-\tau_k}} \right).$$

Recall that $\text{Po}(\lambda)$ is the Poisson distribution with mean λ .

(iii) *Let $C_{n,\ell}$ be the number of cycles of length ℓ in $D_{n,k}[\mathcal{G}_n^c]$. Then for all fixed $\ell \geq 1$,*

$$C_{n,\ell} \xrightarrow{d} \text{Po} \left(\frac{(ke^{-\tau_k})^\ell}{\ell} \right).$$

Theorem 7.3 (Cai and Devroye [21, thm. 3]). *Let $S'_x := S_x \cap \mathcal{G}_n^c$, i.e., S'_x is the spectrum of x in $D_{n,k}[\mathcal{G}_n^c]$. Then*

(i) $\mathbb{P} \left\{ \bigcup_{x \in \mathcal{G}_n^c} [\text{arc}(D_{n,k}[S'_x]) - |S'_x| \geq 1] \right\} = o(1)$, where $\text{arc}(\cdot)$ denotes the number of arcs in a digraph. In other words, whp every spectrum in $D_{n,k}[\mathcal{G}_n^c]$ is a tree or a tree plus an extra arc.

(ii) Let $Y_n := \max_{x \in \mathcal{G}_n^c} |S'_x|$. Let $\lambda_k := (k - \tau_k) \left(\frac{\tau_k}{k-1} \right)^{k-1}$. Then

$$\frac{Y_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/\lambda_k)}.$$

(iii) Let $\text{dist}(x, y)$ be the distance from x to y , i.e., the length of the shortest directed path from x to y . Let $\text{dist}(x, S) := \min_{y \in S} \text{dist}(x, y)$. Let

$$W_n := \max_{x \in \mathcal{G}_n^c} \min_{y \in \mathcal{G}_n} \text{dist}(x, y),$$

i.e., W_n is the maximum distance from a vertex in \mathcal{G}_n^c to \mathcal{G}_n . Then

$$\frac{W_n}{\log_k \log n} \xrightarrow{p} 1.$$

(iv) Let M_n be the length of the longest simple directed path in $D_{n,k}[\mathcal{G}_n^c]$. Then

$$\frac{M_n}{\log n} \xrightarrow{p} \frac{1}{\log(e^{\tau_k}/k)}.$$

(v) Let $X_n := \max_{x \in \mathcal{G}_n^c} \max_{y \in \mathcal{S}'_x} \text{dist}(x, y)$. Then

$$\frac{X_n}{\log n} \xrightarrow{p} \frac{1}{\log(e^{\tau_k}/k)}.$$

7.5 THE CONNECTION TO GALTON-WATSON TREES

A technique often used in studying random graphs (or digraphs) is to approximate a graph exploration process with a Galton-Watson tree. In the case of $D_{n,k}$, consider the following backward exploration:

Compare with the DFS procedure for Galton-Watson trees described in [Subsection 2.2.1](#).

1. Let \mathcal{Q} be an empty queue. Mark all vertices *unvisited*.
2. Put an arbitrary vertex w in \mathcal{Q} .
3. Remove a vertex u at the head of \mathcal{Q} . Mark it as *visited*.
4. Add at the end of \mathcal{Q} the starting vertices of the in-arcs of u , except those vertices that have already been visited.
5. If \mathcal{Q} is empty, terminate. Otherwise go to step 3.

In this way, every vertex that can reach vertex w is visited in the end.

$\text{Bi}(d, p)$ denotes the binomial (d, p) distribution.

Recall that \mathcal{G}_n is set of vertices in the largest closed scc in $D_{n,k}$.

In $D_{n,k}$, the in-degree of a vertex is a $\text{Bi}(nk, 1/n)$ random variable, which converges in distribution to $\text{Po}(k)$. Thus the above exploration process is approximately a DFS traverse of a Galton-Watson tree with offspring distribution $\text{Po}(k)$. So the probability that vertex w is in a large scc should be close to the probability that the Galton-Watson tree survives, which is in fact exactly ν_k . This is why $|\mathcal{G}_n| = (\nu_k + o_p(1))n$.

Remark 7.3. It is well-known that for $k > 1$, $|\mathcal{C}_{\max}^n|$ —the size of the largest component in $G(n, m = nk/2)$ —is also $(\nu_k + o_p(1))n$ (see Erdős and Rényi [39]). Moreover, $(|\mathcal{C}_{\max}^n| - \nu_k n)/\sigma_k \sqrt{n}$ also converges in distribution to a standard normal distribution (see, e.g., Durrett [38]). This is because in $G(n, m = nk/2)$ the degree of a vertex is approximately $\text{Po}(k)$.

The proof of [Theorem 7.3](#) is also based on the use of Galton-Watson trees for approximating graph exploration processes. For example, part (iii) of the theorem says that W_n , the maximal distance from a vertex in \mathcal{G}_n^c to \mathcal{G}_n , is $(1 + o_p(1)) \log_k \log n$. To see this, consider the modified version of the above exploration process:

$$\mathcal{G}_n^c := \{1, \dots, n\} \setminus \mathcal{G}_n.$$

1. Let \mathcal{Q} be an empty queue. Mark all vertices in \mathcal{G}_n^c *unvisited*.
2. Put an arbitrary vertex $w \in \mathcal{G}_n^c$ in \mathcal{Q} .
3. Remove a vertex u at the head of \mathcal{Q} . Mark it as *visited*.
4. Add at the end of \mathcal{Q} the end vertices of the out-arcs of u , except those vertices that have already been visited and those in \mathcal{G}_n .
5. If \mathcal{Q} is empty, terminate. Otherwise go to step 3.

In other words, we go forward instead of backward, and we stay in \mathcal{G}_n^c . For a vertex $u \in \mathcal{G}_n^c$, each of its k out-arcs has probability about ν_k to end at a vertex in \mathcal{G}_n . So this forward exploration process is approximately a Galton-Watson tree with offspring distribution $\text{Bi}(k, \nu_k)$.

When an unvisited vertex u is visited, less than k new vertices is added to the queue \mathcal{Q} if and only if: (i) one or more of the k arcs starting from u end at \mathcal{G}_n or (ii) at vertices that have already been visited. We can show the second situation is unlikely. So the event that u is of distance h away from \mathcal{G}_n has approximately the same probability of the event that the Galton-Watson tree has only degree- k nodes in the first h generations, i.e., it has a complete k -ary tree of height h as a non-fringe subtree at its root.

Since $|\mathcal{G}_n^c| = (1 - \nu_k + o_p(1))n$, W_n should thus be close to the maximal height of complete k -ary trees that appear as non-fringe subtrees at the roots of $\Theta(n)$ independent Galton-Watson trees. Let W'_n denote the latter. We can use the standard first and second moment method to show

$$W'_n = (1 + o_p(1)) \log_k \log n.$$

This is indeed how we proved W_n is also of this value.

The above intuitive explanation of W_n inspired us to look for similar situations where a random object contains many Galton-Watson-tree-like sub-objects. And it is well-known that in a large $\mathcal{T}_n^{\text{gw}}$, fringe subtrees behave like independent copies of unconditional Galton-Watson trees. Therefore, it is natural to guess that $\bar{H}_{n,k}$, the height of the maximal complete k -ary non-fringe subtree of $\mathcal{T}_n^{\text{gw}}$, should also be $(1 + o_p(1)) \log_k \log n$. The intuition turns out to be correct,

Recall that $\mathcal{T}_n^{\text{gw}}$ is a Galton-Watson tree conditioning on having size n .

as shown by [Lemma 6.12](#). This is how our investigation of fringe and non-fringe subtrees in $\mathcal{T}_n^{\text{gw}}$ begun, which eventually lead to this thesis.

OPEN QUESTIONS

In this chapter we summarize some questions that remain open.

8.1 FRINGE SUBTREES

Part (iv) of Theorem 1.3 shows that if

$$\frac{\pi(\mathfrak{S}_n)}{\pi(\mathfrak{T}_{k_n})} \rightarrow 0, \quad (8.1)$$

then

$$d_{\text{TV}}(\mathbb{N}_{\mathfrak{S}_n}(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(\mathfrak{S}_n))) = o(1).$$

The condition (8.1) may not be necessary. For example, if $\mathfrak{S}_n = \mathfrak{T}_{k_n}$, then $\pi(\mathfrak{S}_n)/\pi(\mathfrak{T}_{k_n}) = 1$ and the above result does not apply. However, by Lemma 4.6 and 4.7

$$\frac{\text{Var}(\mathbb{N}_{\mathfrak{T}_{k_n}}(\mathcal{T}_n^{\text{gw}}))}{\mathbb{E}\mathbb{N}_{\mathfrak{T}_{k_n}}(\mathcal{T}_n^{\text{gw}})} \sim \frac{n\pi(\mathfrak{T}_{k_n})}{n\pi(\mathfrak{T}_{k_n})} = 1.$$

In other words, we cannot use Lemma 3.7 or 3.8 to prove

$$\liminf d_{\text{TV}}(\mathbb{N}_{\mathfrak{T}_{k_n}}(\mathcal{T}_n^{\text{gw}}), \text{Po}(n\pi(\mathfrak{T}_{k_n}))) > 0.$$

Yet to remove the requirement of (8.1), it seems that a very different method is needed.

8.2 NON-FRINGE SUBTREES

Theorem 1.4 shows that if

$$n\pi^{\text{nf}}(\mathbb{T}_n) := n\mathbb{P}\{\mathbb{T}_n < \mathcal{T}^{\text{gw}}\} \rightarrow \infty,$$

then

$$\frac{\mathbb{N}_{\mathbb{T}_n}^{\text{nf}}(\mathcal{T}_n^{\text{gw}})}{n\pi^{\text{nf}}(\mathbb{T}_n)} \xrightarrow{p} 1.$$

We believe that under the same assumption it is also possible to prove

$$\frac{\mathbb{N}_{\mathbb{T}_n}^{\text{nf}}(\mathcal{T}_n^{\text{gw}}) - n\pi^{\text{nf}}(\mathbb{T}_n)}{\sqrt{n\pi^{\text{nf}}(\mathbb{T}_n)}} \xrightarrow{d} \mathbb{N}(0, 1),$$

Recall that $\pi(\mathfrak{S}) := \pi(\mathcal{T}^{\text{gw}} \in \mathfrak{S})$ and \mathfrak{T}_{k_n} is the set of plane trees of size k_n .

Recall that $\mathbb{T} < \mathbb{T}'$ denotes that \mathbb{T}' has a non-fringe subtree of shape \mathbb{T} at its root.

Recall that $\mathbb{N}(0, 1)$ is the standard normal distribution.

using the method of moments as we did for (iii) of Theorem 1.3.

Theorem 1.3 generalizes Theorem 1.2 by considering the number of fringe subtrees whose shapes belong to a set of trees \mathfrak{S}_{k_n} instead of being a single tree T_n . It may be possible to generalize Theorem 1.4 in a similar way, i.e., we consider the non-fringe subtrees whose shapes belong to a set of trees \mathfrak{S}_{k_n} instead of being a single tree T_n .

Another problem which may be of interest is to get a non-fringe version of Theorem 5.1, i.e., what are the sufficient conditions for all (or not all) trees of size at most k_n to appear in \mathcal{T}_n^{gw} as non-fringe subtrees?

8.3 EMBEDDED SUBTREES

Let T be a tree and v be a node of T . Recall that T_v denotes the fringe subtree rooted at v . If by removing some or none the subtrees of T_v , we can make it isomorphic to another tree T' , then we say that T contains an *embedded subtree* of the shape T' at v . A more challenging problem is to determine the height of the maximal complete r -ary embedded subtree in \mathcal{T}_n^{gw} . This can be seen as a generalization of the height of \mathcal{T}_n^{gw} , since that \mathcal{T}_n^{gw} has height h is equivalent to that the maximal embedded one-ary subtree of \mathcal{T}_n^{gw} has height h .

BIBLIOGRAPHY

- [1] L. Addario-Berry, B. Balle, and G. Perarnau, "Diameter and stationary distribution of random r -out digraphs," *ArXiv e-prints*, 2015. arXiv: [1504.06840 \[math.PR\]](#).
- [2] L. Addario-Berry, L. Devroye, and S. Janson, "Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees," *The Annals of Probability*, vol. 41, no. 2, pp. 1072–1087, 2013.
- [3] D. Aldous, "Asymptotic fringe distributions for general families of random trees," *The Annals of Applied Probability*, vol. 1, no. 2, pp. 228–266, 1991.
- [4] D. Aldous, "The continuum random tree. I," *The Annals of Probability*, vol. 19, no. 1, pp. 1–28, 1991.
- [5] D. Aldous, "The continuum random tree II: An overview," in *Stochastic Analysis (Durham, 1990)*, ser. London Mathematical Society Lecture Note Series, vol. 167, Cambridge, UK: Cambridge University Press, 1991, pp. 23–70.
- [6] D. Aldous, "The continuum random tree. III," *The Annals of Probability*, vol. 21, no. 1, pp. 248–289, 1993.
- [7] D. J. Aldous and J. Pitman, "Brownian bridge asymptotics for random mappings," *Random Structures and Algorithms*, vol. 5, no. 4, pp. 487–512, 1994.
- [8] D. Aldous and J. Pitman, "Tree-valued Markov chains derived from Galton-Watson processes," *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, vol. 34, no. 5, pp. 637–686, 1998.
- [9] N. Alon and J. H. Spencer, *The Probabilistic Method*, 3rd. New York: Wiley, 2008.
- [10] D. Angluin and D. Chen, "Random walks on random uni-regular graphs," 2015.
- [11] R. Arratia, L. Goldstein, and L. Gordon, "Two moments suffice for Poisson approximations: The Chen-Stein method," *The Annals of Probability*, vol. 17, no. 1, pp. 9–25, 1989.

- [12] K. B. Athreya and P. E. Ney, *Branching Processes*, ser. Die Grundlehren der mathematischen Wissenschaften. Berlin: Springer-Verlag, 1972.
- [13] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*. New York: Oxford University Press, 1993, pp. 511–511.
- [14] I. J. Bienaymé, “De la loi de multiplication et de la durée des familles,” *Société Philomatique Paris*, 1845, Reprinted in Kendall (1975).
- [15] B. Bollobás, *Random Graphs*, ser. Cambridge Studies in Advanced Mathematics. Cambridge, UK: Cambridge University Press, 2001.
- [16] B. Bollobás, “A probabilistic proof of an asymptotic formula for the number of labelled regular graphs,” *European Journal of Combinatorics*, vol. 1, no. 4, pp. 311–316, 1980.
- [17] X. S. Cai, “A probabilistic analysis of Kademlia networks,” Master’s thesis, McGill University, 2012.
- [18] X. S. Cai and L. Devroye, “A probabilistic analysis of Kademlia networks,” in *Algorithms and Computation, 24th International Symposium, ISAAC 2013, Hong Kong, China, December 16-18, 2013, Proceedings*, vol. 8283, Heidelberg, Germany: Springer-Verlag, 2013, pp. 711–721.
- [19] X. S. Cai and L. Devroye, “The analysis of Kademlia for random IDs,” *Internet Mathematics*, vol. 11, no. 6, pp. 572–587, 2015.
- [20] X. S. Cai and L. Devroye, “A study of large fringe and non-fringe subtrees in conditional Galton-Watson trees,” *ArXiv e-prints*, 2016. arXiv: [1602.03850 \[math.PR\]](#).
- [21] X. S. Cai and L. Devroye, “The graph structure of a deterministic automaton chosen at random,” *Random Structures and Algorithms*, 2016, To appear. arXiv: [1504.06238 \[math.PR\]](#).
- [22] X. S. Cai, G. Perarnau, B. Reed, and A. B. Watts, “Acyclic edge colourings of graphs with large girth,” *Random Structures and Algorithms*, 2016, To appear. arXiv: [1411.3047 \[math.CO\]](#).
- [23] A. Carayol and C. Nicaud, “Distribution of the number of accessible states in a random deterministic automaton,” in *29th International Symposium on Theoretical Aspects of Computer Science*, ser. Leibniz International Proceedings in Informatics, vol. 14, Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012, pp. 194–205.
- [24] E. Catalan, “Note sur une équation aux différences finies.,” *Journal de mathématiques pures et appliquées*, pp. 508–516, 1838.

- [25] A. Cayley, "A theorem on trees," *Quarterly Journal of Mathematics*, vol. 23, pp. 376–378, 1889.
- [26] P. Chassaing, J.-F. Marckert, and M. Yor, "The height and width of simple trees," in *Mathematics and Computer Science (Versailles, 2000)*, ser. Trends in Mathematics, Basel: Birkhäuser, 2000, pp. 17–30.
- [27] S. Chatterjee, P. Diaconis, and E. Meckes, "Exchangeable pairs and Poisson approximation," *Probability Surveys*, vol. 2, pp. 64–106, 2005.
- [28] L. H. Y. Chen, "Poisson approximation for dependent trials," *The Annals of Probability*, vol. 3, no. 3, pp. 534–545, 1975.
- [29] F. Chyzak, M. Drmota, T. Klausner, and G. Kok, "The distribution of patterns in random trees," *Combinatorics, Probability and Computing*, vol. 17, no. 01, pp. 21–59, 2008.
- [30] C. Cooper and A. Frieze, "The size of the largest strongly connected component of a random digraph with a given degree sequence," *Combinatorics, Probability and Computing*, vol. 13, pp. 319–337, 03 2004.
- [31] L. Devroye, "Limit laws for local counters in random binary search trees," *Random Structures and Algorithms*, vol. 2, no. 3, pp. 303–315, 1991.
- [32] L. Devroye, "Limit laws for sums of functions of subtrees of random binary search trees," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 152–171, 2002.
- [33] L. Devroye, P. Flajolet, F. Hurtado, M. Noy, and W. Steiger, "Properties of random triangulations and trees," *Discrete and Computational Geometry*, vol. 22, no. 1, pp. 105–117, 1999.
- [34] R. Diestel, *Graph Theory*, 3rd, ser. Graduate Texts in Mathematics. Berlin: Springer-Verlag, 2005.
- [35] M. Drmota, *Random Trees*. Vienna: Springer, 2009.
- [36] T. Duquesne, "A limit theorem for the contour process of conditioned Galton-Watson trees," *The Annals of Probability*, vol. 31, no. 2, pp. 996–1027, 2003.
- [37] R. Durrett, *Probability: Theory and Examples*, 4th, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press, 2010.
- [38] R. Durrett, *Random Graph Dynamics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press, 2007.

- [39] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.
- [40] P. Erdős and A. Rényi, "On a classical problem of probability theory," *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, vol. 6, pp. 215–220, 1961.
- [41] Q. Feng, H. M. Mahmoud, and A. Panholzer, "Phase changes in subtree varieties in random recursive and binary search trees," *SIAM Journal on Discrete Mathematics*, vol. 22, no. 1, pp. 160–184, 2008.
- [42] P. Flajolet and G. Louchard, "Analytic variations on the Airy distribution," *Algorithmica*, vol. 31, no. 3, pp. 361–377, 2001.
- [43] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge, UK: Cambridge University Press, 2009.
- [44] P. Flajolet, D. Gardy, and L. Thimonier, "Birthday paradox, coupon collectors, caching algorithms and self-organizing search," *Discrete Applied Mathematics*, vol. 39, no. 3, pp. 207–229, 1992.
- [45] P. Flajolet, X. Gourdon, and C. Martínez, "Patterns in random binary search trees," *Random Structures and Algorithms*, vol. 11, no. 3, pp. 223–244, 1997.
- [46] P. Flajolet and A. Odlyzko, "Random mapping statistics," in *Advances in Cryptology—EUROCRYPT 89*, ser. Lecture Notes in Computer Science, vol. 434, Berlin, Germany: Springer-Verlag, 1990, pp. 329–354.
- [47] A. Frieze and M. Karoński, *Introduction to Random Graphs*. Cambridge, UK: Cambridge University Press, 2015.
- [48] M. Fuchs, "Subtree sizes in recursive trees and binary search trees: Berry-Esseen bounds and Poisson approximations," *Combinatorics, Probability and Computing*, vol. 17, no. 5, pp. 661–680, 2008.
- [49] Z. Gao and C. N. Wormald, "Asymptotic normality determined by high moments, and submap counts of random maps," *Probability Theory and Related Fields*, vol. 130, no. 3, pp. 368–376, 2004.
- [50] J. Geiger, "Elementary new proofs of classical limit theorems for Galton-Watson processes," *Journal of Applied Probability*, pp. 301–309, 1999.
- [51] E. N. Gilbert, "Random graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

- [52] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, ser. Modern Birkhäuser Classics. Boston: Birkhäuser, 2007.
- [53] J. Gross and J. Yellen, *Handbook of Graph Theory*, ser. Discrete Mathematics and Its Applications. Boca Raton: CRC Press, 2004.
- [54] A. A. Grusho, "Limit distributions of certain characteristics of random automaton graphs," *Mathematical Notes of the Academy of Sciences of the USSR*, vol. 14, no. 1, pp. 633–637, 1973.
- [55] A. Gut, *Probability: A Graduate Course*. New York: Springer-Verlag, 2013.
- [56] C. Holmgren and S. Janson, "Limit laws for functions of fringe trees for binary search trees and random recursive trees," *Electronic Journal of Probability*, vol. 20, p. 51, 2015.
- [57] D. L. Iglehart, "Functional central limit theorems for random walks conditioned to stay positive," *The Annals of Probability*, vol. 2, no. 4, pp. 608–619, 1974.
- [58] K. E. Iverson, *A Programming Language*. Wiley, 1962.
- [59] S. Janson, "Probability asymptotics: notes on notation," *ArXiv e-prints*, 2011. arXiv: [1108.3924](https://arxiv.org/abs/1108.3924) [math.PR].
- [60] S. Janson, T. Łuczak, and A. Rucinski, *Random Graphs*, ser. Wiley Series in Discrete Mathematics and Optimization. New York: Wiley, 2011.
- [61] S. Janson, "Random cutting and records in deterministic and random trees," *Random Structures and Algorithms*, vol. 29, no. 2, pp. 139–179, 2006.
- [62] S. Janson, "Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas," *Probability Surveys*, vol. 4, pp. 80–145, 2007.
- [63] S. Janson, "On the asymptotic joint distribution of height and width in random trees," *Studia Scientiarum Mathematicarum Hungarica*, vol. 45, no. 4, pp. 451–467, 2008.
- [64] S. Janson, "Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation," *Probability Surveys*, vol. 9, pp. 103–252, 2012.
- [65] S. Janson, "Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees," *Random Structures and Algorithms*, vol. 48, no. 1, pp. 57–101, 2016.
- [66] R. M. Karp, "The transitive closure of a random digraph," *Random Structures and Algorithms*, vol. 1, no. 1, pp. 73–93, 1990.

- [67] M. Kearns, "Efficient noise-tolerant learning from statistical queries," *Journal of the ACM*, vol. 45, no. 6, pp. 983–1006, 1998.
- [68] D. P. Kennedy, "The Galton-Watson process conditioned on the total progeny," *Journal of Applied Probability*, pp. 800–806, 1975.
- [69] H. Kesten, "Subdiffusive behavior of random walk on a random cluster," *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, vol. 22, no. 4, pp. 425–487, 1986.
- [70] V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov, *Random Allocations*. Washington, D.C.: Winston, 1978.
- [71] V. F. Kolchin, *Random Mappings*, ser. Translations Series in Mathematics and Engineering. New York: Optimization Software, 1986.
- [72] A. Korshunov, "The number of initially connected automata," *Problemy Kibernetiki*, vol. 34, pp. 5–82, 1978.
- [73] J.-F. Le Gall, "Random trees and applications," *Probability Surveys*, vol. 2, pp. 245–311, 2005.
- [74] E. Lebensztayn, "On the asymptotic enumeration of accessible automata," *Discrete Mathematics and Theoretical Computer Science*, vol. 12, no. 3, pp. 75–79, 2010.
- [75] T. Łuczak, "The phase transition in the evolution of random digraphs," *Journal of Graph Theory*, vol. 14, no. 2, pp. 217–223, 1990.
- [76] R. Lyons, R. Pemantle, and Y. Peres, "Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes," *The Annals of Probability*, vol. 23, no. 3, pp. 1125–1138, 1995.
- [77] J.-F. Marckert and A. Mokkadem, "The depth first processes of Galton-Watson trees converge to the same Brownian excursion," *The Annals of Probability*, vol. 31, no. 3, pp. 1655–1678, 2003.
- [78] A. Meir and J. W. Moon, "On nodes of large out-degree in random trees," *Congressus Numerantium*, vol. 82, pp. 3–13, 1991.
- [79] A. Meir and J. W. Moon, "On the altitude of nodes in random trees," *Canadian Journal of Mathematics*, vol. 30, no. 1978, pp. 997–1015, 1978.
- [80] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures and Algorithms*, vol. 6, no. 2–3, pp. 161–180, 1995.
- [81] P. Neal, "The generalised coupon collector problem," *Journal of Applied Probability*, vol. 45, no. 3, pp. 621–629, 2008.

- [82] X. Pérez-Giménez and N. Wormald, "Asymptotic enumeration of strongly connected digraphs by vertices and edges," *Random Structures and Algorithms*, vol. 43, no. 1, pp. 80–114, 2013.
- [83] B. Pittel, "Counting strongly-connected, moderately sparse directed graphs," *Random Structures and Algorithms*, vol. 43, no. 1, pp. 49–79, 2013.
- [84] B. Pittel and D. Poole, "Asymptotic distribution of the numbers of vertices and arcs of the giant strong component in sparse random digraphs," *Random Structures and Algorithms*, 2015, To appear.
- [85] M. Reineke, "Cohomology of noncommutative Hilbert schemes," *Algebras and Representation Theory*, vol. 8, no. 4, pp. 541–561, 2005.
- [86] B. Roos, "Improvements in the Poisson approximation of mixed Poisson distributions," *Journal of Statistical Planning and Inference*, vol. 113, no. 2, pp. 467–483, 2003.
- [87] N. Ross, "Fundamentals of Stein's method," *Probability Surveys*, vol. 8, pp. 210–293, 2011.
- [88] S. Sagitov and M. C. Serra, "Multitype Bienaymé-Galton-Watson processes escaping extinction," *Advances in Applied Probability*, vol. 41, no. 1, pp. 225–246, 2009.
- [89] B. A. Sevast'yanov, "Convergence to Gaussian and Poisson processes of the distribution of the number of empty cells in the classical problem of pellets," *Teoriya Veroyatnostei i ee Primeneniya*, vol. 12, no. 1, pp. 144–153, 1967.
- [90] M. Sipser, *Introduction to the Theory of Computation*. Independence, KY, USA: Cengage Learning, 2012.
- [91] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, Berkeley, CA, USA: University of California Press, 1972, pp. 583–602.
- [92] L. Takács, "On a probability problem connected with railway traffic," *Journal of Applied Mathematics and Stochastic Analysis*, vol. 4, no. 1, pp. 1–27, 1991.
- [93] R. van der Hofstad, *Random Graphs and Complex Networks*. 2014, vol. 1.

- [94] H. W. Watson and F. Galton, "On the probability of the extinction of families," *The Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 4, pp. 138–144, 1875.