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 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.
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.


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 Kademlia 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:

CONTENTS

Notation and Symbols xv

1 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 $T_n^{bin}$  18
Figure 5  Example of $d_f$  23
Figure 6  Example of $\{T \oplus T\}$  67
Figure 7  Examples of three cases in $B''$  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

N \quad \{1,2,\ldots\}

N_0 \quad \{0\} \cup N

Z \quad \text{the set of all integers}

[H] \quad 1 \text{ if the statement } H \text{ is true, } 0 \text{ otherwise}

(x)_r \quad x(x-1)\cdots(x-r+1)

gcd(S) \quad \text{The greatest common divisor of integers in } S

u \lor v \quad \max\{u,v\}

u \land v \quad \min\{u,v\}

(a_i)_{i \geq 1} \quad \text{an infinite sequence } a_1,a_2,\ldots

(a_i)_{i=1}^n \quad \text{a sequence } a_1,a_2,\ldots,a_n

EX \quad \text{the expectation of } X

\text{Var}(X) \quad \text{the variance of } X

\text{Cov}(X,Y) \quad \text{the covariance of } X \text{ and } Y

\text{E}(X)_r \quad \text{the } r\text{-th factorial moment of } X, \text{ i.e., } \mathbb{E}[X(X-1)\cdots(X-r+1)]

X \overset{\text{d}}{=} Y \quad X \text{ and } Y \text{ are identically distributed}

X_n \xrightarrow{p} X \quad X_n \text{ converges in probability to } X

X_n \xrightarrow{d} X \quad X_n \text{ converges in distribution to } X

\sigma(X) \quad \text{the } \sigma\text{-algebra generated by } X

d_{TV}(X,Y) \quad \text{the total variation distance between } X \text{ and } Y

\text{Po} (\lambda) \quad \text{The Poisson distribution with mean } \lambda

\text{Be} (p) \quad \text{Bernoulli (p) distribution}

\text{Bi} (d,p) \quad \text{Binomial (d,p) distribution}
Ge(p) Geometric (p) distribution
Unif(M) Uniform distribution on M
N(0, I) The standard normal distribution
N(\mu, \gamma^2) The normal distribution with mean \mu and variance \gamma^2
\xi, 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 \xi, i.e., gcd\{i \geq 1 : p_i > 0\}
\text{supp}(\xi) the support of \xi, i.e., \{i \in \mathbb{N}_0 : p_i > 0\}
\text{supp}(\xi, k) \text{supp}(\xi) \cap \{1, \ldots, k\}
T the set of all rooted, ordered and unlabeled trees (plane trees)
T_n the set of all rooted, ordered and unlabeled trees of size n
T a tree in 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
deg(v) the degree of node v
\mathcal{G} a set of trees
\mathcal{G}_n a sequence of sets of trees
\mathcal{D}(\mathcal{G}) the set of preorder degree sequences of trees in \mathcal{G}
\mathcal{D}(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}(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 \xi
\mathcal{T}^{gw}_n \mathcal{T}^{gw} restricted to the event |\mathcal{T}^{gw}| = n
\( \tau_{n,v}^{gw} \) a fringe subtree of \( \mathcal{T}_n^{gw} \) rooted at node \( v \in \mathcal{T}_n^{gw} \)

\( \tau_{n,s}^{gw} \) a fringe subtree of \( \mathcal{T}_n^{gw} \) rooted at a uniform random node of \( \mathcal{T}_n^{gw} \)

\( \tau_{n}^{\text{bin}} \) 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 \( T \)

\( \pi(T) \) \( \mathbb{P}\{\mathcal{T}^{gw} = T\} \)

\( N_S(\mathcal{T}_n^{gw}) \) the number of fringe subtrees in \( \mathcal{T}_n^{gw} \) that belong to \( S \)

\( \pi(S) \) \( \mathbb{P}\{\mathcal{T}^{gw} \in S\} \)

\( N_{n_f}(\mathcal{T}_n^{gw}) \) the number of non-fringe subtrees of \( \mathcal{T}_n^{gw} \) of shape \( T \)

\( \pi_{n_f}(T) \) \( \mathbb{P}\{T < \mathcal{T}^{gw}\} \), the probability that \( \mathcal{T}^{gw} \) has a non-fringe subtree \( T \) at its root

\( (\xi^n_i)_{i=1}^n \) \( (\xi^n_1, \ldots, \xi^n_n) \), the preorder degree sequence of \( \mathcal{T}_n^{gw} \)

\( \xi_1, \xi_2, \ldots \) iid copies of \( \xi \)

\( S_n \) \( \sum_{i=1}^n \xi_n \)

\( (\tilde{\xi}^n_i)_{i=1}^n \) \( (\tilde{\xi}^n_1, \ldots, \tilde{\xi}^n_n) \), a uniform random cyclic rotation of \( (\xi^n_i)_{i=1}^n \)

\( \mathcal{T}^+_{\leq k} \) the set of all possible trees of size at most \( k \)

\( \mathcal{T}^+_{= k} \) the set of all possible trees of size exactly \( k \)

\( K_n \) \( \max\{k : \mathcal{T}^+_{\leq k} \subseteq \cup_{v \in \mathcal{T}^{gw}} \{\mathcal{T}_n^{gw}\}\} \)

\( \mathcal{T}^{\text{min}}_k \) a tree \( T \in \mathcal{T}^+_{\leq k} \) that minimizes \( \mathbb{P}\{\mathcal{T}^{gw} = T\} \)

\( p_k^{\text{min}} \) \( \pi(\mathcal{T}^{\text{min}}_k) \)

\( L_k \) \( L_k \coloneqq \min\{p_0 (p_i/p_0)^{1/i} : i \in \text{supp}(\xi, k - 1)\} \)

\( L \) \( \lim_{k \to \infty} L_k \)

\( \mathcal{L} \) smallest \( i \) with \( L_{i+1} = L \) if such \( i \) exists; \( \infty \) otherwise

\( \mathcal{T}^{\text{star}}_{k-1} \) the tree of size \( k \) with \( k - 1 \) leaves

\( \tilde{d} \) a sequence in \( \mathcal{R}_n \), i.e., a rotation of a tree degree sequence of size \( n \)

\( \text{deg}_i(\tilde{d}) \) a tree degree sequence that starts at index \( i \) in \( \tilde{d} \)

\( T_i((\tilde{\xi}^n_i)_{i=1}^n) \) the fringe subtree of \( \mathcal{T}_n^{gw} \) with the preorder degree sequence \( \text{deg}_i(\tilde{\xi}^n_1, \ldots, \tilde{\xi}^n_n) \)
\( T_{i}(\vec{d}) \)  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 \( T \)

\( V(T) \)  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

\( D^A_{n,k} \)  the condensation \( \text{DAG} \) of \( D_{n,k} \)

\( S_u \)  the spectrum of \( u \), i.e., the set of vertices reachable from \( u \)

\( G_n \)  the set of vertices in the largest closed \( \text{scc} \) in \( D_{n,k} \)

\( O_n \)  the set of vertices in the one-in-core of \( D_{n,k} \)

\text{dfs}  Depth-First-Search

\text{lhs}  left hand side

\text{rhs}  right hand side

\text{iid}  independent and identically distributed

\text{whp}  with high probability

\text{DFA}  Deterministic Finite Automaton

\text{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 \( \mathcal{T} \) be the set of all rooted, ordered, and unlabeled trees, which we call plane trees. All trees considered in this thesis belong to \( \mathcal{T} \) unless explicitly stated otherwise. Given a tree \( T \in \mathcal{T} \) 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 \( T' \in \mathcal{T} \), 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 \( T' \prec T_v \) 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 \( T' \prec T_v \).) We also write \( T' \prec T \) 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.

1.2 CONDITIONAL GALTON-WATSON TREES

Let \( \xi \) be a non-negative integer-valued random variable. The Galton-Watson tree \( \mathcal{T}^{gw} \) with offspring distribution \( \xi \) 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 \( \xi \). The conditional Galton-Watson tree \( \mathcal{T}_n^{gw} \) is \( \mathcal{T}^{gw} \) restricted to the event \( |\mathcal{T}^{gw}| = n \), i.e., the event that \( \mathcal{T}^{gw} \) has \( n \) nodes.

In the study of \( \mathcal{T}_n^{gw} \), the following is usually assumed:

See Subsection 2.1.1 for more about plane trees.

See page xv for a complete list of notation and symbols.
Condition A. Let $T_{n}^{gw}$ be a conditional Galton-Watson tree of size $n$ with offspring distribution $\xi$, such that $E\xi = 1$ and $0 < \sigma^2 := \text{Var}(\xi) < \infty$. Let $T^{gw}$ 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 $\mathbb{P}\{|T^{gw}| = n\} = 0$ and $T_{n}^{gw}$ is not well-defined. But it is easy to show that $\mathbb{P}\{|T^{gw}| = n\} > 0$ for all $n \geq n_0$ with

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

where $n_0$ is a constant depending only on $\xi$ (see [64, cor. 15.6]). Therefore, in this thesis, for all asymptotic results regarding $T_{n}^{gw}$, the limits are always taken as $n \to \infty$ 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 $T_{n}^{gw}$ of shape $T$. Extending a result by Aldous [3], Janson [64, thm. 7.12] proved the following:

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

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

In [65] Janson further strengthened the above result, proving the asymptotic normality of $N_T(T_{n}^{gw})$ by studying additive functionals on $T_{n}^{gw}$.
A natural extension of $N_T(T_R^{gw})$ is $N_{T_n}(T_R^{gw})$, where the fixed tree $T$ is replaced by a sequence of trees $(T_n)_{n \geq 1}$. Let $\text{Po}(\lambda)$ be a Poisson random variable with mean $\lambda$ and $N(0, 1)$ be a standard normal random variable. Let $d_{TV}(\cdot, \cdot)$ denote the total variation distance. Let $\pi(T) \mapsto \text{Po}(\pi(T))$. We have:

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

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

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

(i) If $n\pi(T_n) \to 0$, then $N_{T_n}(T_R^{gw}) = 0$ whp.

(ii) If $n\pi(T_n) \to \mu \in (0, \infty)$, then $N_{T_n}(T_R^{gw}) \xrightarrow{d} \text{Po}(\mu)$.

(iii) If $n\pi(T_n) \to \infty$, then

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

**Remark 1.2.** We actually showed that

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

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

Let $[H]$ 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 $\mathfrak{G} \subseteq \mathfrak{T}_k$, let

$$\pi(\mathfrak{G}) := \text{Pr}\{T^{gw} \in \mathfrak{G}\}, \quad \text{and} \quad N_{\mathfrak{G}}(T_R^{gw}) := \sum_{\mathfrak{G}^{gw} \subseteq \mathfrak{G}} [T_R^{gw} \in \mathfrak{G}]_v.$$

Theorem 1.2 can be generalized as follows:

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

(i) If $n\pi(\mathfrak{G}_n) \to 0$, then $N_{\mathfrak{G}_n}(T_R^{gw}) = 0$ whp.

(ii) If $n\pi(\mathfrak{G}_n) \to \mu \in (0, \infty)$, then $N_{\mathfrak{G}_n}(T_R^{gw}) \xrightarrow{d} \text{Po}(\mu)$.

(iii) If $n\pi(\mathfrak{G}_n) \to \infty$, then

$$\frac{N_{\mathfrak{G}_n}(T_R^{gw}) - n\pi(\mathfrak{G}_n)}{\sqrt{n\pi(\mathfrak{G}_n)}} \xrightarrow{d} N(0, 1).$$
(iv) If $\pi(\mathcal{G}_n)/\pi(\mathcal{T}_{k_n}) \to 0$, then

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

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 $N_{\mathcal{G}_n}(\mathcal{T}_n^{gw})$ quite easy, which is crucial for the proof.) The key step of the proof is to construct an exchangeable pair for $N_{\mathcal{T}}(\mathcal{T}_n^{gw})$. 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 $N_{\mathcal{G}_n}(\mathcal{T}_n^{gw})$ in (iv) of Theorem 1.3 without assuming that $\pi(\mathcal{G}_n)/\pi(\mathcal{T}_{k_n}) \to 0$. In particular, it does not work for $N_{\mathcal{T}_{k_n}}(\mathcal{T}_n^{gw})$, 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 $N_{\mathcal{G}_n}(\mathcal{T}_n^{gw})$ (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 $P\{T^{gw} = T\} > 0$. As an application of Theorems 1.2 and 1.3, we study the following problem — when does $T_n^{gw}$ 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 $T_n^{gw}$ 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 $\mathcal{T}_{\leq k}$ be the set of all possible trees of size at most $k$. Let $K_n$ be the maximal $k$ such that every tree in $\mathcal{T}_{\leq k}$ appears in $T_n^{gw}$ 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 $K_n/\log n \xrightarrow{p} 0$. For example, for a uniform random Cayley tree of size $n$, we have $K_n \log \log(n)/\log(n) \xrightarrow{p} 1$. 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 $T_{gw}^n$ behave like unconditional Galton-Watson trees when $n$ is large, the number of non-fringe subtrees of shape $T$ should be more or less $nP\{T < T_{gw}^n\}$. Let

$$\pi^{nf}(T) := P\{T < T_{gw}^n\}, \quad \text{and} \quad N_T^{nf}(T_{gw}^n) := \sum_{v \in T_{gw}^n} [T < T_{gw}^n,v].$$

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 $k_n \to \infty$ and $k_n = o(n)$. We have

(i) If $n\pi^{nf}(T_n) \to 0$, then $N_T^{nf}(T_{gw}^n) = o(1)$ whp.

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

$$\frac{N_T^{nf}(T_{gw}^n)}{n\pi^{nf}(T_n)} \to 1.$$

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 $N_T^{nf}(T_{gw}^n)$ 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 $n\pi^{nf}(T_n) \to \mu \in (0, \infty)$, then we have $N_T^{nf}(T_{gw}^n) \overset{d}{\to} Po(\mu)$, 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.
In this chapter we review some properties of the conditional Galton-Watson tree $T_{gw}^{n}$. 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 $T_{gw}^{n}$. Section 2.2 discusses the preorder degree sequences of $T_{gw}^{n}$. Section 2.3 gives some examples of $T_{gw}^{n}$. In addition, Section 2.4 summarizes a few other important results of $T_{gw}^{n}$ 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 $\text{deg}(v)$. 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

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 \( \xi \), 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 \( \xi \), denoted by \( T_{gw} \). Let \( |T_{gw}| \) be the size of \( T_{gw} \), i.e., the number of nodes in it. Then \( |T_{gw}| = \infty \) corresponds to survival and \( |T_{gw}| < \infty \) 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 \( \mathbb{E}\xi < 1 \), then \( \mathbb{P}(\{ |T_{gw}| < \infty \}) = 1 \).

2. If \( \mathbb{E}\xi > 1 \), then \( \mathbb{P}(\{ |T_{gw}| < \infty \}) \in (0,1) \).

The probability \( \mathbb{P}(\{ |T_{gw}| = \infty \}) \) is called the *survival* probability of \( T_{gw} \). 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 \( T_{gw} \) 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 $T^w_n$, is a random tree distributed as $T^w$ restricted to the event that $|T^w| = n$. In other words,

$$
P \{ T^w_n = T \} = \frac{P \{ T^w = T \ \cap \ |T^w| = n \}}{P \{ |T^w| = n \}}
$$

$$
= \frac{P \{ |T^w| = n \} P \{ T^w = T \}}{P \{ |T^w| = n \}}
$$

$$
= \frac{\sum_{T \in \mathcal{T}_n} \prod_{v \in T} p_{\text{deg}(v)}}{\sum_{T \in \mathcal{T}_n} \prod_{v \in T'} p_{\text{deg}(v)}},
$$

where $\mathcal{T}_n$ 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

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

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

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

$$
P \{ T^*_{n} = T \} = \frac{\text{weight}(T)}{\sum_{T \in \mathcal{T}_n} \text{weight}(T)}
$$

$$
= \frac{\prod_{v \in T} w_{\text{deg}(v)}}{\sum_{T \in \mathcal{T}_n} \prod_{v \in T'} w_{\text{deg}(v)}},
$$

Thus if $\sum_{i \geq 0} w_i = 1$, i.e., $(w_i)_{i \geq 0}$ is a probability distribution, then $T^*_{n}$ is nothing but $T^w_n$ 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, $T^*_{n}$ 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

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

If $\phi(\rho) < \infty$ for some $\rho \in (0, \infty)$, then $T^*_{n}$ has the same distribution as $T^w_n$ with offspring distribution

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

for all $t \in (0, \rho)$.
This lemma suggests that results of $\mathcal{T}_n^g$ often also apply to $\mathcal{T}_n^{gw}$. For example, the theorems in Chapter 1 can also be formulated in terms of $\mathcal{T}_n^g$. 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:

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

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

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

is called the preorder degree sequence of $T$.

In the above DFS procedure, initially, there is one node in the stack $Q$. In each iteration, the size of $Q$ decreases by one and increases by $\deg(v)$. Since $Q$ 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 $\mathbb{N} := \{1, 2, \ldots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Let $\mathcal{D}(\mathcal{T}_n) \subseteq \mathbb{N}_0^n$ be the set of preorder degree sequences of all trees of size $n$. A sequence $(d_1, \ldots, d_n) \in \mathcal{D}(\mathcal{T}_n)$ if and only if it satisfies

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

(2.1)

Figure 3 gives an example of Lemma 2.2.

We call a sequence in $\bigcup_{n=1}^{\infty} \mathcal{D}(\mathcal{T}_n)$ a tree degree sequence. Observe that a tree degree sequence cannot be the prefix of another tree degree sequence:
The degree sequence \( (d_1, \ldots, d_7) = (2, 1, 0, 3, 0, 0, 0) \).

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

**Corollary 2.1.** If \( (d_1, d_2, \ldots, d_n) \in \mathcal{D}(\mathcal{I}_n) \), then it is impossible that there exists \( 1 \leq n' < n \) such that \( (d_1, d_2, \ldots, d_{n'}) \in \mathcal{D}(\mathcal{I}_n) \).

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

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

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

\[
(d_{1+j}, d_{2+j}, \ldots, d_{n+j}) \in \mathcal{D}(\mathcal{I}_n),
\]

where the indices are taken modulo \( n \).

### 2.2.2 Preorder degree sequences of conditional Galton-Watson trees

Let \( (\xi^n_i)_{i=1}^n := (\xi^n_1, \ldots, \xi^n_n) \) denote the preorder degree sequence of \( \mathcal{T}_n^{\text{GW}} \). Let \( (\tilde{\xi}_i^n)_{i=1}^n := (\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n) \) be a uniform random cyclic rotation of \( (\xi^n_i)_{i=1}^n \). Let \( (\bar{\xi}_i)_{i=1}^n := (\xi_1, \ldots, \xi_n) \) be independent and identically distributed (iid) copies of \( \xi \). 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 \( \mathcal{D}(\mathcal{I}_n) \) and \( \mathcal{R}_n \). This leads to the following well-known connection between \( (\tilde{\xi}_i^n)_{i=1}^n \) and \( (\xi_1, \ldots, \xi_n) \) (see Janson [64, cor. 15.4]):
Lemma 2.4. We have
\[
P \{|T_{gw}^n| = n\} = \frac{P \{(\xi_1^n, \ldots, \xi_n^n) \in D(T_n)\}}{n} = \frac{P \{(\xi_1, \ldots, \xi_n) \in I_n\}}{P \{S_n = n - 1\}}.
\]
As a result,
\[(\xi_1^n, \xi_2^n, \ldots, \xi_n^n) \overset{D}{=} (\xi_1, \ldots, \xi_n | S_n = n - 1),\]
where \(\overset{D}{=}\) denotes “identically distributed” and the right-hand-side (rhs) denotes the sequence \((\xi_1, \ldots, \xi_n)\) 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
\[
P \{|T_{gw}^n| = n\} = \frac{1}{n} P \{S_n = n - 1\} \sim \frac{\text{span}(\xi)}{\sqrt{2\pi\sigma^2}} n^{-3/2},
\]
as \(n \to \infty\) with \(n - 1 \equiv 0 \pmod{\text{span}(\xi)}\).

Let \(\mathcal{G} \subseteq \mathcal{T}\) be a set of trees. Define
\[
\pi(T) := P \{T_{gw}^n = T\}, \quad \text{and} \quad \pi(\mathcal{G}) := P \{T_{gw}^n \in \mathcal{G}\}.
\]
The following lemma is a special case of \[65, \text{lem. 5.1}\]. We give a short proof for later reference.

**Lemma 2.6.** Assume \(P \{|T_{gw}^n| = n\} > 0\).

(i) Let \(T \in \mathcal{T}_k\) with \(1 \leq k \leq n\). Let \(N_T(T_{gw}^n)\) be the number of fringe subtrees of \(T_{gw}^n\) in the shape of \(T\), i.e.,
\[
N_T(T_{gw}^n) := \sum_{v \in T_{gw}^n} [T_{gw}^{n, v} = T].
\]
Then
\[
\frac{E \left[ N_T(T_{gw}^n) \right]}{n} = \pi(T) \frac{P \{S_{n-k} = n - k\}}{P \{S_n = n - 1\}}.
\]

(ii) Let \(\mathcal{G} \subseteq \mathcal{T}_k\) be a set of some trees of size \(k\) with \(1 \leq k \leq n\). Let \(N_{\mathcal{G}}(T_{gw}^n)\) be the number of fringe subtrees of \(T_{gw}^n\) belonging to \(\mathcal{G}\), i.e.,
\[
N_{\mathcal{G}}(T_{gw}^n) = \sum_{v \in T_{gw}^n} [T_{gw}^{n, v} \in \mathcal{G}].
\]
Then
\[
\frac{E \left[ N_{\mathcal{G}}(T_{gw}^n) \right]}{n} = \pi(\mathcal{G}) \frac{P \{S_{n-k} = n - k\}}{P \{S_n = n - 1\}}.
\]
Proof. Let \((d_1, d_2, \ldots, d_k)\) be the preorder degree sequence of \(T\). Recall that \((\xi_1^n, \ldots, \xi_n^n)\) is the preorder degree sequence of \(T_n^{\shuffle w}\). Let

\[
I_i = [\xi_i^n = d_1, \xi_{i+1}^n = d_2, \ldots, \xi_{i+k-1}^n = d_k],
\]

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 \(T_n^{\shuffle w}\) 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 \((d_1, d_2, \ldots, d_{k'})\) is also a preorder degree sequence, which is impossible by Corollary 2.1.

Therefore for all \(1 \leq i \leq n\), \(I_i = [T_{vi} = T]\) and \(N_T(T_n^{\shuffle w}) = \sum_{i=1}^n I_i\). Recalling that \((\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n)\) is a uniform random rotation of \((\xi_1^n, \ldots, \xi_n^n)\) and using Lemma 2.4, we have

\[
\mathbb{E}[N_T(T_n^{\shuffle w})] = \mathbb{E} \sum_{i=1}^n I_i = \sum_{i=1}^n \mathbb{P}\left\{\tilde{\xi}_i^n = d_1, \ldots, \tilde{\xi}_{i+k-1}^n = d_k \mid S_n = n - 1\right\}
\]

\[
= n \mathbb{P}\{\xi_1 = d_1, \ldots, \xi_k = d_k \mid S_n = n - 1\}
\]

\[
= n \frac{\mathbb{P}\{[\xi_1 = d_1, \ldots, \xi_k = d_k] \cap [S_n = n - 1]\}}{\mathbb{P}\{S_n = n - 1\}}.
\]

Since \((d_1, \ldots, d_k)\) is a tree degree sequence, by Lemma 2.2, \(\sum_{i=1}^k d_i = k - 1\).

Therefore, using the mutual independence of \(\xi_1, \ldots, \xi_n\), the last expression equals

\[
\frac{n \mathbb{P}\{[\xi_1 = d_1, \ldots, \xi_k = d_k] \cap [S_{n-k} = n - k]\}}{\mathbb{P}\{S_n = n - 1\}}
\]

\[
= n \mathbb{P}\{T_n^{\shuffle w} = T\} \frac{\mathbb{P}\{S_{n-k} = n - k\}}{\mathbb{P}\{S_n = n - 1\}}.
\]

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

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 \(1 \leq k \leq n/2\),

\[
\frac{\mathbb{P}\{S_{n-k} = n - k\}}{\mathbb{P}\{S_n = n - 1\}} = 1 + O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right).
\]
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,
\[
\frac{\mathbb{P}[S_{n-2k} = n-2k+1]}{\mathbb{P}[S_n = n-1]} = -\frac{1}{\sigma^2 n} + o\left(\frac{1}{n}\right) + O\left(\frac{k}{n^{3/2}} + \frac{k^2}{n^2}\right).
\]

**Remark 2.1.** As shown in the proof of Lemma 2.6, \(N_T(T_n^{gw})\) is equivalent to the number of patterns \(d_1, \ldots, d_{|T|}\) in the cycle \(\hat{\xi}_1^n, \ldots, \hat{\xi}_n^n\). Thus if \(\text{span}(\xi) > 1\), we can divide \(d_1, \ldots, d_{|T|}\) and \(\hat{\xi}_1^n, \ldots, \hat{\xi}_n^n\) by \(\text{span}(\xi)\) without changing the value of \(N_T(T_n^{gw})\). 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
\[
\frac{\mathbb{P}[S_{n-z} = n-z-w]}{\mathbb{P}[S_n = n-1]} = \frac{\mathbb{P}[S_{n-z} = n-z-w]}{\mathbb{P}[S_n = n-1]} \frac{\mathbb{P}[S_{n-z-1} = n-z-1]}{\mathbb{P}[S_{n-1} = n-z-1]} \frac{\mathbb{P}[S_{n-z-1} = n-z-1]}{\mathbb{P}[S_{n-z} = n-z-1]} \frac{\mathbb{P}[S_{n-z} = n-z-1]}{\mathbb{P}[S_n = n-1]} = 1 + O\left(\frac{w}{n-z}\right) + o\left(\frac{(n-z)^{-1/2}}{n}\right)
\]
\[
= \left(1 + O\left(\frac{w}{n-z}\right)\right) \left(1 + o\left(\frac{(n-z)^{-1/2}}{n}\right)\right) \left(1 + O\left(\frac{z}{n}\right)\right) = 1 + O\left(\frac{w+z}{n}\right) + o(n^{-1/2}).
\]

### 2.3 Examples of Conditional Galton-Watson Trees

This section introduces some common examples of \(T_n^{gw}\). 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 \(\xi \overset{D}{=} \text{Ge}(1/2)\). Let \(T\) be a (plane) tree of size \(n\) and \((d_1, \ldots, d_n)\) be its preorder degree sequence. By Lemma 2.2 and 2.4, we have
\[
\mathbb{P}\{T_n^{gw} = T\} = \frac{\mathbb{P}\{T_n^{gw} = T\}}{\mathbb{P}\{|T_n^{gw}| = n\}} = \frac{2^{-\sum_{i=1}^n (d_i+1)}}{\mathbb{P}\{|T_n^{gw}| = n\}} = 4^{-(n-1)} \frac{2^n}{\mathbb{P}\{|T_n^{gw}| = n\}}.
\]
2.3 Examples of Conditional Galton-Watson Trees

<table>
<thead>
<tr>
<th>Name</th>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric</td>
<td>Ge(p)</td>
<td>$p_i = p(1 - p)^i$ (i $\in$ $\mathbb{N}_0$)</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>Be(p)</td>
<td>$p_0 = (1 - p), p_1 = p$</td>
</tr>
<tr>
<td>Uniform</td>
<td>Unif(M)</td>
<td>$p_i = \frac{1}{</td>
</tr>
<tr>
<td>Binomial</td>
<td>Bi(d, p)</td>
<td>$p_i = \binom{d}{i} (1 - p)^{d-i} p^i$ (i $\in$ {0, ..., d})</td>
</tr>
<tr>
<td>Poisson</td>
<td>Po(λ)</td>
<td>$p_i = e^{-\lambda} i^i / i!$ (i $\in$ $\mathbb{N}_0$)</td>
</tr>
</tbody>
</table>

Table 1: Some discrete probability distributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Offspring Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane trees</td>
<td>Ge(1/2)</td>
</tr>
<tr>
<td></td>
<td>$p_i = 1/2^{i+1}$ (i $\in$ $\mathbb{N}_0$)</td>
</tr>
<tr>
<td>Full binary trees</td>
<td>2 × Be(1/2)</td>
</tr>
<tr>
<td></td>
<td>$p_0 = p_2 = 1/2$</td>
</tr>
<tr>
<td>Motzkin trees</td>
<td>Unif({0, 1, 2})</td>
</tr>
<tr>
<td></td>
<td>$p_0 = p_1 = p_2 = 1/3$</td>
</tr>
<tr>
<td>Binary trees</td>
<td>Bi(2, 1/2)</td>
</tr>
<tr>
<td></td>
<td>$p_0 = p_2 = 1/4, p_1 = 1/2$</td>
</tr>
<tr>
<td>d-ary trees</td>
<td>Bi(d, 1/d)</td>
</tr>
<tr>
<td></td>
<td>$p_i = \binom{d}{i} (1 - \frac{1}{d})^{d-i} (\frac{1}{d})^i$ (i $\in$ {0, ..., d})</td>
</tr>
<tr>
<td>Cayley trees</td>
<td>Po(1)</td>
</tr>
<tr>
<td></td>
<td>$p_i = e^{-1} i^i / i!$ (i $\in$ $\mathbb{N}_0$)</td>
</tr>
</tbody>
</table>

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

Since this probability does not depend on $T$, $T_{n}^{gw}$ is uniformly distributed on $\mathcal{T}_n$ (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:

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

Since $T_n^{gw}$ has size $n$, we have

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

Therefore,

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

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

**Lemma 2.8.** We have

$$|\mathcal{T}_n| = C_{n-1} := \frac{1}{n} \binom{2n-2}{n-1} \sim \frac{4^{n-1}}{\sqrt{\pi n^{3/2}}}.$$  

where the numbers $C_n$ 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 \( \mathcal{B}_n \) be the set of all full binary trees of size \( n \). Let \( T \in \mathcal{B}_n \). 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

\[
    \mathbb{P}\{\mathcal{T}_{gw}^n = T\} = \frac{P_0^{(n+1)/2} P_1^{(n-1)/2}}{\mathbb{P}\{|\mathcal{T}_{gw}^n| = n\}}.
\]

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

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 \( \xi \sim \text{Unif}\{1,2,3\} \), then \( \mathcal{T}_{gw}^n \) 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 \( \xi \sim \text{Bi}(2,1/2) \). Given \( \mathcal{T}_{gw}^n \), for each node \( v \in \mathcal{T}_{gw}^n \), 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 \( \mathcal{T}_{bin}^n \) be the binary tree constructed in this way. See Figure 4 for an example.

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

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

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

\[
    \mathbb{P}\{\mathcal{T}_{bin}^n = \mathcal{T}_{bin}^n \mid \mathcal{T}_{gw}^n = T\} = \frac{1}{2^{n_1}}.
\]

![Figure 4: Example of \( \mathcal{T}_{bin}^n \)](image)
Therefore,

\[
P \{ T_n^{\text{bin}} = T^{\text{bin}} \} = P \{ T_n^{\text{bin}} = T^{\text{bin}} \mid T_n^{\text{gw}} = T \} \frac{P \{ T_n^{\text{gw}} = T \}}{P \{ |T_n^{\text{gw}}| = n \}} \\
= \frac{1}{2^n} \frac{4^{-n} 4^{-n} 2^{-n} 1}{P \{ |T_n^{\text{gw}}| = n \}} \\
= \frac{4^n}{P \{ |T_n^{\text{gw}}| = n \}}.
\]

In other words, \( T_n^{\text{bin}} \) 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 \( \xi_t \equiv \text{Bi}(d, 1/d) \). 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 \( \{1, \ldots, |T|\} \). \( T \) is unordered means that the child nodes of the same parent are unordered. Thus \( T \) is not a plane tree.

Let \( \mathcal{R}_n \) be the set of Cayley trees of size \( n \). Let \( \xi_t \equiv \text{Poi}(1) \). If we label the nodes of \( T_n^{\text{gw}} \) with \( \{1, \ldots, n\} \) 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 \( \mathcal{R}_n \). See Example 10.2 of Janson [64].

### 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 \( f : \mathcal{T} \to \mathbb{R} \) be a function on trees. Define \( F : \mathcal{T} \to \mathbb{R} \) 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 \( f(T') = [T' = T] \), then
\[
F(T_n^{gw}) = \sum_{\nu \in T_n^{gw}} f(T_{n,\nu}^{gw}) = N_T(T_n^{gw}).
\]

On the other hand, we can write
\[
F(T_n^{gw}) = \sum_{T \in \mathcal{T}} f(T) N_T(T_n^{gw}).
\]

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

Recall that \( T_{n,s}^{gw} \) is a fringe subtree rooted at a uniform random node of \( T_n^{gw} \).

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 \( \mathcal{T} \),
\[
\frac{F(T_n^{gw})}{n} = \mathbb{E} \left[ f(T_{n,s}^{gw}) \mid T_n^{gw} \right] \Pr f(T^{gw}).
\]

In [65], Janson also proved a central limit theorem for \( F(T_n^{gw}) \):

**Theorem 2.3** (Janson [65, thm. 1.5]). Assume Condition A. Let \( f : \mathcal{T} \to \mathbb{R} \) be a function with \( \mathbb{E} f(T^{gw}) < \infty \). If \( \mathbb{E} \left[ f(T_n^{gw})^2 \right] \to 0 \), and
\[
\sum_{n=1}^{\infty} \sqrt{\mathbb{E} \left[ f(T_n^{gw})^2 \right]} \frac{1}{n} < \infty,
\]
then
\[
\frac{\text{Var} (F(T_n^{gw}))}{n} \to \gamma^2,
\]
where \( \gamma \) is a finite constant depending on \( f \) and \( \xi \); moreover
\[
\frac{F(T_n^{gw}) - n \mathbb{E} \left[ f(T_n^{gw}) \right]}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2),
\]
where \( N(0, \gamma^2) \) denotes the normal distribution with mean 0 and variance \( \gamma^2 \).

### 2.4.2 Size-biased Galton-Watson trees

A random variable \( \hat{\xi} \) on \( \mathbb{N} \) is said to have \( \xi\)-size-biased distribution if
\[
P \{ \hat{\xi} = i \} = iP \{ \xi = i \} = i p_i, \quad \text{for all } i \in \mathbb{N}.
\]

Note that this is indeed a distribution since
\[
\sum_{i=1}^{\infty} P \{ \hat{\xi} = i \} = \sum_{i=1}^{\infty} i p_i = \mathbb{E} \xi = 1.
\]
The size-biased conditional Galton-Watson tree $\hat{T}^{gw}$ 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 $\xi$, 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 $\hat{T}^{gw}$ contains an infinite path consists of kings called the spine.

A celebrated result for $\hat{T}^{gw}$ is the following:

**Theorem 2.4** (Kennedy [68]). Assume Condition A. Then $T^{gw}_n$ converges to $\hat{T}^{gw}$ in distribution. Or equivalently, for all fixed $k \in \mathbb{N}$, $T^{gw}_n$ truncated at generation $k$ converges in distribution to $\hat{T}^{gw}$ 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 $\hat{T}^{gw}$ with a finite spine, see Addario-Berry, Devroye, and Janson [2], Geiger [50], and Sagitov and Serra [88].

2.4.3 Maximum degree

Let $\xi_{\text{max}}^n := \max\{\xi_1^n, \ldots, \xi_n^n\}$. Let $\hat{\xi}_{\text{max}} := \max\{\hat{\xi}_1, \ldots, \hat{\xi}_n\}$. Lemma 2.4 suggests that $\xi_{\text{max}}^n$ and $\hat{\xi}_{\text{max}}$ should have similar distributions when $n$ is large. The following theorem makes this precise:

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

1. If $nP\{\xi \geq h_n\} \to \alpha$ for some $\alpha \in [0, \infty]$, then
   $$\sum_{i=1}^{n} I[\xi_i^n \geq h_n] \overset{d}{\to} \text{Po}(\alpha).$$

2. $d_{TV}(\xi_{\text{max}}^n, \hat{\xi}_{\text{max}}) \to 0$.

3. There exists a sequence $b_n = o(\sqrt{n})$ such that $\xi_{\text{max}}^n \leq b_n$ whp.

Many authors have studied maximal degrees in $T^{gw}_n$, 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 $H(T)$ and $W(T)$ 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,

$$
\frac{H(T)}{\sqrt{n}} \overset{d}{\to} \frac{2}{\sigma} X, \quad \text{and} \quad \frac{W(T)}{\sqrt{n}} \overset{d}{\to} \sigma X,
$$

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

Addario-Berry, Devroye, and Janson [2] showed that $W(T)$ and $H(T)$ have sub-Gaussian tails:

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

$$
\Pr \{ H(T) \geq x \sqrt{n} \} \leq C e^{-cx^2},
$$

$$
\Pr \{ W(T) \geq x \sqrt{n} \} \leq C e^{-cx^2}.
$$

See also Janson [63] for the joint distribution of $(H(T), W(T))$.

2.4.5 Continuum random tree

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

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

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 $W_{T_n^0}$ is a random walk that encodes $T_n^0$. 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 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

$$\left( \frac{W_{\mathcal{T}^{bw}_n}(\lfloor nt \rfloor)}{\sqrt{n}} \right)_{t \in [0,1]} \overset{d}{\to} (\sigma \epsilon(t))_{t \in [0,1]}.$$  

This suggests that there exists a tree-like object encoded by a Brownian excursion which is the limit of a rescaled $\mathcal{T}^{bw}_n$. 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 $u, v \in \mathcal{M}$:

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 $f : [0, \infty) \to [0, \infty)$ 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 $d_f$](image)

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.
Replacing the function $f$ by a standard Brownian excursion $e$, we have a random real tree $T_e$ 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*

$$\frac{\sigma}{\sqrt{n}} T_{gw}^n \xrightarrow{d} T_e,$$

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

So Theorem 2.6 (height and width of $T_{gw}^n$) 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.
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 $X \overset{d}{=} \text{Po}(\mu)$ and $Y \overset{d}{=} \text{Po}(\nu)$ for $\mu, \nu > 0$, then

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

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

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{i=1}^{\infty} |x_i - y_i| = \frac{1}{2} \sum_{i=1}^{\infty} |\sqrt{x_i} - \sqrt{y_i}| \times (\sqrt{x_i} + \sqrt{y_i})$$

$$\leq \frac{1}{2} \left( \sum_{i=1}^{\infty} |\sqrt{x_i} - \sqrt{y_i}|^2 \sum_{j=1}^{\infty} (\sqrt{x_j} + \sqrt{y_j})^2 \right)^{1/2}$$

$$= \frac{1}{2} \left( 2 - \sum_{i=1}^{\infty} 2 \sqrt{x_i y_i} \right) \left( 2 + \sum_{j=1}^{\infty} 2 \sqrt{x_j y_j} \right)^{1/2}$$

$$= \left( 1 - \left( \sum_{i=1}^{\infty} \sqrt{x_i y_i} \right)^2 \right)^{1/2},$$

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

$$\sum_{i=1}^{\infty} \sqrt{x_i y_i} = \sum_{i=1}^{\infty} e^{-\mu/2} \frac{(\mu \nu)^{1/2}}{i!} = \exp \left( \sqrt{\mu \nu} - \frac{\mu + \nu}{2} \right) = \exp \left( -\frac{(\sqrt{\mu} - \sqrt{\nu})^2}{2} \right).$$
Thus we have
\[
\begin{align*}
\text{d}_{TV}(X, Y) &\leq \sqrt{1 - \exp\left(-\left(\sqrt{\mu} - \sqrt{\nu}\right)^2\right)} \\
&\leq \sqrt{(\sqrt{\mu} - \sqrt{\nu})^2} 
\text{ (by } 1 - e^{-x} \leq x) \\
&= |\sqrt{\mu} - \sqrt{\nu}|.
\end{align*}
\]

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 \( \mu_n \) be a sequence of non-negative real numbers. Assume that \( \text{d}_{TV}(X_n, \text{Po}(\mu_n)) \to 0 \) as \( n \to \infty \).

We have:

(i) If \( \mu_n \to 0 \), then \( X_n = 0 \) whp.

(ii) If \( \mu_n \to \mu \in (0, \infty) \), then \( X_n \xrightarrow{d} \text{Po}(\mu) \).

(iii) If \( \mu_n \to \infty \), 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 \( Z \sim \text{Po}(\lambda) \). Let \( X \) be a non-negative integer-valued random variable. For \( A \subseteq \mathbb{N}_0 \), define the function \( f_{\lambda, A} : \mathbb{N}_0 \to \mathbb{R} \) recursively as

\[
\begin{align*}
f_{\lambda, A}(0) &= 0, \\
f_{\lambda, A}(i + 1) &= \frac{1}{\lambda} \left( if_{\lambda, A}(i) + [i \in A] - \mathbb{P}\{Z \in A\}\right).
\end{align*}
\]

Therefore, for all \( i \in \mathbb{N}_0 \),
\[
\lambda f_{\lambda, A}(i + 1) - if_{\lambda, A}(i) = [i \in A] - \mathbb{P}\{Z \in A\}.
\]

Replacing \( i \) with \( X \) and taking expectation, we have
\[
\mathbb{P}\{X \in A\} - \mathbb{P}\{Z \in A\} = \mathbb{E}\left[\lambda f_{\lambda, A}(X + 1) - X f_{\lambda, A}(X)\right].
\]

Therefore
\[
\begin{align*}
\text{d}_{TV}(X, Z) &= \sup_{A \subseteq \mathbb{N}_0} |\mathbb{P}\{X \in A\} - \mathbb{P}\{Z \in A\}| \\
&= \sup_{A \subseteq \mathbb{N}_0} |\mathbb{E}\left[\lambda f_{\lambda, A}(X + 1) - X f_{\lambda, A}(X)\right]|.
\end{align*}
\]
The function \( f_{\lambda,A} \) has the following nice properties (see Barbour, Holst, and Janson [13, lem. 1.1.1] for a proof):

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

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

and

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

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 \( \sigma \)-algebra such that \( W \) is \( \mathcal{F} \)-measurable. There exists a function \( f : \mathbb{N}_0 \to \mathbb{R} \) with

\[
\| f \| \leq \lambda^{-1/2}, \quad \| \Delta f \| \leq \lambda^{-1},
\]

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

\[
d_{TV}(W, \text{Po}(\lambda)) = |\mathbb{E}\left( (\lambda - c\mathbb{P}\{W' = W + 1|\mathcal{F}\}) f(W + 1) - \mathbb{E}\left( (W - c\mathbb{P}\{W' = W - 1|\mathcal{F}\}) f(W)\right)\right)|. \tag{3.1}
\]

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

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

and

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

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

\[
d_{TV}(W, \text{Po}(\lambda))
\]

\[
= \left| \mathbb{E}\left[ (np - \sum_{i=1}^{n} (1 - \tilde{I}_i)p) f(W + 1) \right] - \mathbb{E}\left[ (W - \sum_{i=1}^{n} \hat{I}_i(1 - p)) f(W) \right] \right|
\]

\[
= |\mathbb{E}[Wpf(W + 1)] - \mathbb{E}[Wpf(W)]| \leq np^2 \| \Delta f \| \leq np^2 \frac{1}{np} = p.
\]

Thus \( d_{TV}(W, \text{Po}(\lambda)) \to 0 \) as \( p \to 0 \).
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 \( \{ (\tilde{I}_i, \tilde{J}_i) \}_{i=1}^n \) be an independent copy of \( \{ (\tilde{I}_i, \tilde{J}_i) \}_{i=1}^n \). We construct a new sequence \( \{ I_i \}_{i=1}^n \) as follows. Let \( Z \) be a uniform random variable on \( \{1, \ldots, n\} \). Let

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

Let \( I_i = \tilde{I}_i \) for \( i \in \{1, \ldots, n\} \) with \( i \neq Z \).

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

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

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

\[
P\{ (\tilde{I}_1, \ldots, \tilde{I}_n) = (a_1, \ldots, a_n), (\hat{I}_1, \ldots, \hat{I}_n) = (1 - a_1, a_2, \ldots, a_n) \mid E' \} \\
= P\{ \tilde{I}_1 = a_1 \mid E' \} P\{ \tilde{I}_2 = a_2 \ldots \tilde{I}_n = a_n \mid E' \} P\{ \hat{I}_1 = 1 - a_1 \mid E' \} \\
= P\{ \tilde{I}_1 = 1 - a_1 \mid E' \} P\{ \tilde{I}_2 = a_2 \ldots \tilde{I}_n = a_n \mid E' \} P\{ \hat{I}_1 = a_1 \mid E' \} \\
= P\{ (\tilde{I}_1, \ldots, \tilde{I}_n) = (1 - a_1, a_2, \ldots, a_n) \mid E' \} P\{ \hat{I}_1 = a_1 \mid E' \}.
\]

Let \( \Gamma_n = \sum_{i=1}^n \tilde{I}_i \) and \( \Psi_n = \sum_{i=1}^n \hat{I}_i \). Let \( \hat{\Gamma}_n = \sum_{i=1}^n \tilde{I}_i \). We have:

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

\[
d_{TV} \left( \Gamma_n, Po(n\mathbb{E}\tilde{I}_1) \right) \leq \sqrt{\frac{\mathbb{E}\hat{I}_1}{\mathbb{E}\tilde{I}_1} \left( Var(\Psi_n) \right) \mathbb{E}[\tilde{I}_1]}{\mathbb{E}[\Psi_n]}}.\]
Example 3.2. In the case that $\tilde{I}_i$’s are iid Be($p$) and $\tilde{J}_i = 1$ for all $1 \leq i \leq n$, we have $\Gamma_n \overset{d}{=} Bi(n, p)$ and $\Psi_n = n$. The above lemma shows that

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

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

Proof of Lemma 3.6. Let $r_n := \mathbb{E}{\tilde{I}_1} = \mathbb{E}{\tilde{J}_1}$ and $q_n := \mathbb{E}{\tilde{J}_1} = \mathbb{E}{\tilde{J}_1}$. 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 $\hat{\Gamma}_n = \Gamma_n + 1$ happens if and only if $\hat{J}_Z = \hat{J}_Z = 1, \hat{I}_Z = 0$ and $\hat{I}_Z = 1$, we have

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

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

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

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

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

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

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

$$\zeta_2 := \Gamma_n - c\mathbb{P}\{\hat{\Gamma}_n = \Gamma_n - 1 | \mathcal{F}\} = \Gamma_n - \frac{q_n}{n} \Gamma_n - \frac{r_n}{n} \Gamma_n = \frac{r_n}{q_n} \Gamma_n.$$
It follows from Lemma 3.4 that there exists a function \( f : \mathbb{N}_0 \to \mathbb{R} \) with 
\[
\|\Delta f\| \leq \mathbb{E} [\Gamma_n]^{-1} = (nr_n)^{-1} \quad \text{and} \quad \|f\| \leq \mathbb{E} [\Gamma_n]^{-1/2} = (nr_n)^{-1/2},
\]
such that 
\[
d_{TV} (\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) = \mathbb{E} [\xi_1 f(\Gamma_n + 1) - \xi_2 f(\Gamma_n)]
\]
\[
= \mathbb{E} \left[ \frac{r_n}{q_n} \Gamma_n + \frac{r_n}{q_n} (\mathbb{E} [\Psi_n] - \Psi_n) \right] f(\Gamma_n + 1) - \frac{r_n}{q_n} \mathbb{E} [\Gamma_n] f(\Gamma_n)
\]
\[
= \mathbb{E} \left[ \frac{r_n}{q_n} \Gamma_n (f(\Gamma_n + 1) - f(\Gamma_n)) + f(\Gamma_n + 1) \right] \frac{r_n}{q_n} (\mathbb{E} [\Psi_n] - \Psi_n)
\]
\[
\leq \frac{r_n}{q_n} \mathbb{E} [\Gamma_n \cdot |f(\Gamma_n + 1) - f(\Gamma_n)|]
\]
\[
+ \frac{r_n}{q_n} \mathbb{E} [|f(\Gamma_n + 1) \cdot |\mathbb{E} [\Psi_n] - \Psi_n]]
\]
\[
\leq \frac{r_n}{q_n} \|\Delta f\| \mathbb{E} [\Gamma_n] + \frac{r_n}{q_n} \|f\| \mathbb{E} [\mathbb{E} [\Psi_n] - \Psi_n]
\]
\[
\leq \frac{r_n}{q_n} \frac{1}{nr_n} \mathbb{E} [\Gamma_n] + \frac{r_n}{q_n} \frac{1}{\sqrt{nr_n}} (\text{Var} (\Psi_n))^{1/2}
\]
\[
= \frac{r_n}{q_n} + \left[ \frac{r_n}{q_n} \frac{\text{Var} (\Psi_n)}{\mathbb{E} \Psi_n} \right]^{1/2}.
\]
\[\square\]

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. 3.1]). 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 
\[
d_{TV} (W, \text{Po} (\mathbb{E} W)) \geq \frac{c \Delta}{1 + \log \frac{1}{\Delta}} \left( 1 \wedge \frac{\mathbb{E} W}{1 + \log \frac{1}{\Delta}} \right).
\]

**Example 3.3.** Let \( W \sim \text{Bi}(n, p) \). 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, 
\[
d_{TV} (W, \text{Po} (W)) \geq \frac{cp}{1 + \log \frac{1}{p}} \left[ 1 \wedge \frac{np}{1 + \log \frac{1}{p}} \right].
\]

Therefore, \( d_{TV} (W, \text{Po} (W)) \) 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 $r > 2$,

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

Then

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

### 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 $I_1, \ldots, I_n$ are indicator random variables. Then for all $r \in \mathbb{N}$, we have

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

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

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 $r \in \mathbb{N}$,

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

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

**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 $\mu_n \to \infty$. Let $(X_n)_{n \geq 1}$ be a sequence of integer-valued random variables. If

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

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., $X_n \overset{d}{=} \text{Bi}(n, p_n)$. Then by Lemma 3.9, we have

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

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

On the other hand, if $p_n \to 0$ and $np_n \to \infty$, 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, $\mathbb{E}(X)_r/(np_n)^r \to 1$. 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 \( k_n \to \infty \). Then

\[
\lim_{n \to \infty} \sup_{T : |T| = k_n} d_{TV}(N_T(T_n^{gw}), \text{Po}(n\pi(T))) = 0. \tag{1.2}
\]

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

(i) If \( n\pi(T_n) \to 0 \), then \( N_{T_n}(T_n^{gw}) = o(n) \) whp.

(ii) If \( n\pi(T_n) \to \mu \in (0, \infty) \), then \( N_{T_n}(T_n^{gw}) \xrightarrow{d} \text{Po}(\mu) \).

(iii) If \( n\pi(T_n) \to \infty \), then

\[
\frac{N_{T_n}(T_n^{gw}) - n\pi(T_n)}{\sqrt{n\pi(T_n)}} \xrightarrow{d} N(0,1).
\]

Let \( \mathcal{S} \) be a set of trees. Recall that

\[ N_{\mathcal{S}}(T_n^{gw}) := \sum_{v \in T_n^{gw}} [v \in \mathcal{S}], \]

i.e., \( N_{\mathcal{S}}(T_n^{gw}) \) is the number of fringe subtrees in \( T_n^{gw} \) whose shapes belong to \( \mathcal{S} \). Also recall that \( \pi(\mathcal{S}) := \mathbb{P}(T^{gw} \in \mathcal{S}) \). 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 \( k \to \infty \). We have

\[
\sup_{\mathcal{S} \subseteq \mathcal{T}_k} \frac{d_{TV}(N_{\mathcal{S}}(T_n^{gw}), \text{Po}(n\pi(\mathcal{S})))}{n/o(\pi(\mathcal{S})/\pi(\mathcal{T}_k) + \sqrt{n/O(\pi(\mathcal{S})/\pi(\mathcal{T}_k)})} \leq 1 + o\left(k^{-3/2}\right) + O\left(k^{1/4}/\sqrt{n}\right).
\]

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.
Remark 4.2. If $\pi(T) = 0$, then deterministically $N_T(T_n^w) = 0$ and
\[ d_TV(N_T(T_n^w), Po(n\pi(T))) = 0. \]

Thus we can assume without loss of generality that
\[ \mathbb{P}\{T_n^w \in T_n\} = \mathbb{P}\{|T_n^w| = k_n\} > 0 \]
for all $n$, and that the supremum in (1.2) is taken over all $T \in T_n$ with $\pi(T) > 0$.

4.1 Moments of fringe subtree counts

Let $\mathcal{G} \subseteq T_k$ for some $k \in \mathbb{N}$. Let $\mathcal{D}(\mathcal{G})$ be the set of preorder degree sequences of trees in $\mathcal{G}$. Recall that $(\tilde{\xi}_n^i)_{i=1}^n := (\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n)$ is a uniform random rotation of $(\xi_1^n)_{i=1}^n$. Throughout this chapter, let
\[ \tilde{I}_1 := \| (\tilde{\xi}^n_1, \tilde{\xi}^n_2, \ldots, \tilde{\xi}^n_{n+k-1}) \in \mathcal{D}(\mathcal{G}) \|. \quad (4.1) \]

Then $N_\mathcal{G}(T_n^w) = \sum_{i=1}^n \tilde{I}_i$. We use $\tilde{I}_i$ to compute the moments of $N_\mathcal{G}(T_n^w)$.

4.1.1 Factorial moments

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

Lemma 4.2. Let $\mathcal{G} \subseteq T_k$. Let $(\tilde{I}_i)_{i=1}^n$ be as in (4.1). For $k, r \in \mathbb{N}$ with $kr \leq n$, we have
\[ \mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} \cdots = \tilde{I}_{k(r-1)+1} = 1\} = \pi(\mathcal{G})^r \frac{\mathbb{P}\{S_{n-kr} = n-r(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}} \]

Proof. We have
\[ \mathbb{P}\{\tilde{I}_1 = \tilde{I}_{k+1} \cdots = \tilde{I}_{k(r-1)+1} = 1\} = \mathbb{P}\left\{ \bigcap_{j=1}^r \left( (\tilde{\xi}^n_{k(j-1)+1}, \ldots, \tilde{\xi}^n_{k(j)}) \in \mathcal{D}(\mathcal{G}) \right) \right\} \]
\[ = \mathbb{P}\left\{ \bigcap_{j=1}^r \left( (\tilde{\xi}^n_{k(j-1)+1}, \ldots, \tilde{\xi}^n_{k(j)}) \in \mathcal{D}(\mathcal{G}) \right) \mid S_n = n - 1 \right\} \]
\[ = \mathbb{P}\left\{ \bigcap_{j=1}^r \left( (\tilde{\xi}^n_{k(j-1)+1}, \ldots, \tilde{\xi}^n_{k(j)}) \in \mathcal{D}(\mathcal{G}) \right) \cap [S_n = n - 1] \right\} \]
\[ = \prod_{j=1}^r \mathbb{P}\{ (\tilde{\xi}^n_{k(j-1)+1}, \ldots, \tilde{\xi}^n_{k(j)}) \in \mathcal{D}(\mathcal{G}) \} \frac{\mathbb{P}\{S_{n-kr} = n-r(k-1) - 1\}}{\mathbb{P}\{S_n = n - 1\}}, \]

where the second last equality uses that the total degree of a tree of size $k$ is $k-1$. \qed
Lemma 4.3. Consider \( n \) balls labeled 1, \ldots, \( n \) 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

\[
n(n - r(k - 1) - 1)_{r-1}.
\]

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.

Lemma 4.4. Let \( \mathcal{G} \subseteq \mathbb{T}_k \). Then for all \( r \in \mathbb{N} \) with \( kr \leq n \),

\[
\mathbb{E}(N_{\mathcal{G}}(T_n^{gw}))_r = n(n - r(k - 1) - 1)_{r-1} \mathbb{P}(\mathcal{S}_{n-kr} = n-r(k-1)-1) \mathbb{P}(\mathcal{S}_n = n-1) / \mathbb{P}(\mathcal{S}_n = n).
\]

Proof. Let \( \tilde{I}_i \) be as (4.1). Then \( N_{\mathcal{G}}(T_n^{gw}) = \sum_{i=1}^{n} I_i \). By Lemma 3.9, for \( rk \leq n \), we have

\[
\mathbb{E}(N_{\mathcal{G}}(T_n^{gw}))_r = \sum_{i_1, \ldots, i_r} \mathbb{P}\left\{ \tilde{I}_{i_1} = \cdots = \tilde{I}_{i_r} = 1 \right\},
\]

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

\[
\{s, s+1, \ldots, s+k-1\} \cap \{t, t+1, \ldots, t+k-1\} \neq \emptyset,
\]

where all the numbers are taken modulo \( n \), then \( \mathbb{P}\left\{ \tilde{I}_{i_1} = \cdots = \tilde{I}_{i_r} = 1 \right\} = 0 \).

Thus we only need to consider \( i_1, \ldots, i_r \) without such overlapping indices.

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

\[
\mathbb{P}\left\{ \tilde{I}_{i_1} = \cdots = \tilde{I}_{i_r} = 1 \right\} = \mathbb{P}\left\{ \tilde{I}_1 = \tilde{I}_{k+1} = \cdots = \tilde{I}_{k(r-1)+1} = 1 \right\}.
\]

Therefore,

\[
\mathbb{E}(N_{\mathcal{G}}(T_n^{gw}))_r = n(n - r(k - 1) - 1)_{r-1} \mathbb{P}\left\{ \tilde{I}_1 = \tilde{I}_{k+1} = \cdots = \tilde{I}_{k(r-1)+1} = 1 \right\}
\]

\[
= n(n - r(k - 1) - 1)_{r-1} \mathbb{P}(\mathcal{S}_{n-kr} = n-r(k-1)-1) \mathbb{P}(\mathcal{S}_n = n-1) / \mathbb{P}(\mathcal{S}_n = n-1),
\]

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

$$\sup_{\mathcal{G} \subseteq 2^k} \sup_{s \leq r} \left| \frac{\mathbb{E}(N_{\mathcal{G}}(T_n^{gw}))_s}{n\pi(\mathcal{G})^s} - 1 \right| = o(1).$$

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

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

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

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

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

$$\mathbb{E}(N_{\mathcal{G}}(T_n^{gw}))_s = n(n - r(s - 1) - 1)_{s-1}\pi(\mathcal{G})^s \mathbb{P}\{S_{n-k} = n - s(k - 1) - 1\} \mathbb{P}\{S_n = n - 1\}$$

$$= (n\pi(\mathcal{G}))^s \left[1 + O\left(\frac{r^2k}{n}\right) + o(n^{-1/2})\right].$$

4.1.2 Expectation and variance

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

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

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

$$\text{Cov}(\tilde{I}_1, \tilde{I}_{k+1}) = \mathbb{E}\left[\tilde{I}_1 \tilde{I}_{k+1}\right] - \mathbb{E}[\tilde{I}_1] \mathbb{E}[\tilde{I}_{k+1}]$$

$$\leq \frac{\mathbb{E}\tilde{I}_1}{n} \left(o\left(k^{-3/2}\right) + O\left(\sqrt{\frac{k}{n}}\right)\right).$$

Proof. It follows from Lemma 2.6 that

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

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

$$\mathbb{E}\left[\tilde{I}_1 \tilde{I}_{k+1}\right] = \pi(\mathcal{G})^2 \frac{\mathbb{P}\{S_{n-2k} = n - 2k + 1\}}{\mathbb{P}\{S_n = n - 1\}}.$$
By part (ii) of Lemma 2.7 and that \( \pi(\mathcal{G}) \leq \pi(\mathcal{T}_k) = \Theta(k^{-3/2}) \) (Lemma 2.5), we have

\[
\frac{1}{n} \sum \frac{1}{o(\sqrt{n})} + O \left( \frac{k^{3/2} + k^{-3/2}}{\sqrt{n}} \right) + O \left( \frac{k^{-3/2}}{\sqrt{n}} \right).
\]

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

\[
\sup_{\boldsymbol{\Theta} \in \mathcal{T}_k} \frac{\text{Var} (N_{\boldsymbol{\Theta}}(\mathcal{T}_n^{\mathbb{R}}))}{\mathbb{E} N_{\mathcal{T}_n^{\mathbb{R}}}} \leq 1 + o \left( k^{-3/2} \right) + O \left( \sqrt{\frac{k}{n}} \right).
\]

Proof. Let \( \tilde{I}_i \) be as in (4.1). Then \( N_{\boldsymbol{\Theta}}(\mathcal{T}_n^{\mathbb{R}}) = \sum_{i=1}^n \tilde{I}_i \). 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 \( \text{Cov} (\tilde{I}_i, \tilde{I}_j) = 0 \).

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 \( (\tilde{\theta}_i^n)_{i=1}^n \) is permutation invariant. Therefore, by Lemma 4.7,

\[
\text{Var} (N_{\boldsymbol{\Theta}}(\mathcal{T}_n^{\mathbb{R}})) = \sum_{1 \leq i \neq j \leq n} \text{Cov} (\tilde{I}_i, \tilde{I}_j) + \sum_{i=1}^n \text{Var} (\tilde{I}_i)
\]

\[
\leq n^2 \text{Cov} (\tilde{I}_1, \tilde{I}_{k+1}) + n \mathbb{E} \tilde{I}_1 \left( 1 - \mathbb{E} \tilde{I}_1 \right)
\]

\[
= \frac{n^2}{n} \mathbb{E} \tilde{I}_1 \left( o \left( k^{-3/2} \right) + O \left( \sqrt{k/n} \right) \right) + n \mathbb{E} \tilde{I}_1
\]

\[
= \left( 1 + o \left( k^{-3/2} \right) + O \left( \sqrt{k/n} \right) \right) \mathbb{E} N_{\mathcal{T}_n^{\mathbb{R}}}. \]

4.2 PROOF OF LEMMA 4.1 AND THEOREM 1.2

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

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

\[
\tilde{J}_i = \{(\tilde{\theta}_1^n, \tilde{\theta}_{i+1}^n, \ldots, \tilde{\theta}_{i+k-1}^n) \in \mathcal{D}(\mathcal{T}_k), \}
\]
Recall that $\xi_1, \ldots, \xi_n$ are iid copies of $\xi$.

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

The following observation is crucial for proving Lemma 4.1:

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

**Proof.** Let $E = [\tilde{\jmath}_1 = 1]$. 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 $j \in \{2, \ldots, k\} \cap \{n + 2 - k, \ldots, n\}$.

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

Let $(d_i)_{i=1}^n$ be such that $(\tilde{\xi}_{k+1}^n, \ldots, \tilde{\xi}_n^n) = (d_{k+1}, \ldots, d_n)$ implies $E$. In other words,

$$(d_1, \ldots, d_k) \in D(\Xi_k), \quad \text{and} \quad \sum_{i=k+1}^n d_i = n - k.$$  \(\blacksquare\)

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

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

On the other hand,

$$\mathbb{P}\{(\tilde{\xi}_1^n, \ldots, \tilde{\xi}_k^n) = d_1, \ldots, d_k \mid E\} = \frac{\mathbb{P}\{(\xi_1^n, \ldots, \xi_k^n) = d_1, \ldots, d_k\}}{\mathbb{P}E} \mathbb{P}\{S_n = n - 1\} / \mathbb{P}E.$$  \(\blacksquare\)

Therefore,

$$\mathbb{P}\{(\tilde{\xi}_1^n, \ldots, \tilde{\xi}_k^n) = (d_1, \ldots, d_k) \mid E\} \times \mathbb{P}\{(\tilde{\xi}_{k+1}^n, \ldots, \tilde{\xi}_n^n) = (d_{k+1}, \ldots, d_n) \mid E\} = \mathbb{P}\{(\xi_1^n, \ldots, \xi_n = d_1, \ldots, d_n)\} / \mathbb{P}\{S_n = n - 1\} / \mathbb{P}E = \mathbb{P}\{(\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n) = (d_1, \ldots, d_n) \mid E\}.$$  \(\blacksquare\)

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

$$\sup_{\phi \in \Xi_k} \frac{d_{TV}(N_{\phi}(\mathcal{T}_n^p), \text{Po}(\mathbb{E}N_{\phi}(\mathcal{T}_n^p)))}{\pi(\mathcal{G})/\pi(\Xi_k) + \sqrt{\pi(\mathcal{G})/\pi(\Xi_k)}} \leq 1 + o\left(k^{-3/2}\right) + O\left(\sqrt{k/n}\right).$$
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
\[
\Gamma_n = N_{\mathcal{T}}(\mathcal{J}_{\mathcal{G}}^{\mathcal{W}}) \quad \text{and} \quad \Psi_n = N_{\mathcal{G}}(\mathcal{J}_{\mathcal{G}}^{\mathcal{W}}).
\]
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 \(\{1, \ldots, n\}\). We construct a new sequence \((\tilde{I}_i)_{i=1}^n\) by letting
\[
\tilde{I}_Z = \begin{cases} 
\hat{I}_Z, & \text{if } \hat{Z} = \hat{I}_Z = 1, \\
\hat{I}_Z, & \text{otherwise},
\end{cases}
\]
and \(\tilde{I}_i = \hat{I}_i\) for \(i \in \{1, \ldots, n\}\) with \(i \neq Z\).

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

Since \(\hat{I}_1 = 1\) implies \(\hat{I}_1\), it follows from Lemma 3.6 that
\[
d_{TV}(\Gamma_n, \text{Po}(\mathcal{E}\Gamma_n)) \leq \frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\hat{I}_1} + \left[\frac{\mathbb{E}\tilde{I}_1 \text{Var}(\Psi_n)}{\mathbb{E}\hat{I}_1 \mathbb{E}[\Psi_n]}\right]^{1/2}.
\]

By Lemma 2.6, we have
\[
\frac{\mathbb{E}\tilde{I}_1}{\mathbb{E}\hat{I}_1} = \frac{\mathbb{E}\Gamma_n}{\mathbb{E}\Psi_n} = \frac{\pi(\mathcal{E})}{\pi(\mathcal{G})} = \frac{\pi(\mathcal{E})}{\pi(\mathcal{G})}.
\]
And by Lemma 4.8, we have
\[
\text{Var}(\Psi_n) = 1 + o(k^{-3/2}) + O\left(\sqrt{k/n}\right).
\]
Therefore
\[
d_{TV}(\Gamma_n, \text{Po}(\mathcal{E}\Gamma_n)) \leq \left(\frac{\pi(\mathcal{E})}{\pi(\mathcal{G})}\right) \left(1 + o(k^{-3/2}) + O\left(\sqrt{k/n}\right)\right)^{1/2}
\]
\[
\leq \left(1 + o(k^{-3/2}) + O\left(\sqrt{k/n}\right)\right) \left(\frac{\pi(\mathcal{E})}{\pi(\mathcal{G})}\right) + O\left(\sqrt{k/n}\right).
\]
□

**Lemma 4.11.** Let \(k = k_n = o(n)\) and \(k \to \infty\). We have
\[
\sup_{\Theta \subseteq \mathcal{F}_k} \frac{d_{TV}(\text{Po}(n\pi(\mathcal{E})), \text{Po}(\mathcal{E}\Theta^\mathcal{G}))}{\pi(\Theta)/\pi(\mathcal{G}) + \sqrt{\pi(\Theta)/\pi(\mathcal{G})}} \leq O\left(\frac{k^{1/4}}{\sqrt{n}}\right) + o(k^{-3/4}).
\]

**Proof.** Let \(\Gamma_n = N_{\mathcal{G}}(\mathcal{J}_{\mathcal{G}}^{\mathcal{W}})\). It follows from Lemma 3.1 and Lemma 4.6 that
\[
d_{TV}(\text{Po}(n\pi(\mathcal{E})), \text{Po}(\mathcal{E}\Gamma_n)) \leq \frac{|n\pi(\mathcal{E}) - \mathcal{E}\Gamma_n|}{\sqrt{n\pi(\mathcal{E})}}
\]
\[
= \sqrt{n\pi(\mathcal{E})} \left|1 - \frac{\mathcal{E}\Gamma_n}{n\pi(\mathcal{E})}\right|
\]
\[
= \sqrt{n\pi(\mathcal{E})} \left(O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right)\right).
\]
Therefore,
\[
\frac{d_{TV}(\text{Po}(n\pi(S)), \text{Po}([\Gamma_n]))}{\pi(S)/\pi(\Xi_k) + \sqrt{\pi(S)/\pi(\Xi_k)}} \leq \frac{d_{TV}(\text{Po}(n\pi(S)), \text{Po}([\Gamma_n]))}{\sqrt{\pi(S)/\pi(\Xi_k)}} \\
\leq \sqrt{n\pi(\Xi_k)} \left( O\left(\frac{k}{n}\right) + o\left(n^{-1/2}\right) \right) \\
= O\left(\frac{k^{1/4}}{\sqrt{n}}\right) + o(k^{-3/4}),
\]
where the last step uses \(\pi(\Xi_k) = \Theta(k^{-3/2})\) from Lemma 2.5.

Proof of Lemma 4.1. Let \(\Gamma_n = \text{N}_\Theta(T^w_R)\). It follows from Lemma 4.11, Lemma 4.10 and the triangle inequality that
\[
\frac{d_{TV}(\Gamma_n, \text{Po}(n\pi(S))))}{\pi(S)/\pi(\Xi_k) + \sqrt{\pi(S)/\pi(\Xi_k)}} \\
\leq \frac{d_{TV}(\Gamma_n, \text{Po}(\mathbb{E}\Gamma_n)) + d_{TV}(\text{Po}(\mathbb{E}\Gamma_n), \text{Po}(n\pi(S))))}{\pi(S)/\pi(\Xi_k) + \sqrt{\pi(S)/\pi(\Xi_k)}} \\
\leq 1 + o\left(k^{-3/2}\right) + O\left(\sqrt{k/n}\right) + O\left(k^{1/4}\right) + o(k^{-3/4}) \\
\leq 1 + o\left(k^{-3/2}\right) + O\left(\sqrt{k/n}\right).
\]

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

For \(T \in \Xi_k\), we have
\[
\pi(T) := P\{T^w = T\} \leq p_k.
\]
Therefore, by Lemma 2.5,
\[
\frac{\pi(T)}{\pi(\Xi_k)} \leq \frac{p_k}{\Theta(k^{-3/2})} = O\left(p_k^{-3/2}\right).
\]
It follows from Lemma 4.1 by taking \(S = \{T\}\) that
\[
d_{TV}(N_T(T^w_R), \text{Po}(n\pi(T))) \leq (1 + o(1)) \left(\frac{\pi(T)}{\pi(\Xi_k)} + \sqrt{\frac{\pi(T)}{\pi(\Xi_k)}}\right) \\
= O\left(\sqrt{p_k^{-3/2}}\right) = o(1).
\]
Statements (i)-(iii) of the theorem follow from Lemma 3.2.

Remark 4.3. The proof of Theorem 1.2 uses the condition \(k_n \to \infty\) 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
\[
\text{E}N_{T_n}(T^w_R) = (1 + o(1))n\pi(T),
\]
and
\[ \text{Var}(N_{T_n}(J_n^{gw})) \leq (1 + o(1))n\pi(T)(1 - \pi(T)). \]

Therefore,
\[ \liminf_{n \to \infty} \left[ 1 - \frac{\text{Var}(N_{T_n}(J_n^{gw}))}{\mathbb{E}N_{T_n}(J_n^{gw})} \right] \geq \pi(T) > 0. \]

So \( \text{Var}(N_{T_n}(J_n^{gw})) > \mathbb{E}N_{T_n}(J_n^{gw}) \) for \( n \) large and by Lemma 3.7, we have
\[ \liminf_{n \to \infty} d_{\text{TV}}(N_{T_n}(J_n^{gw}), \text{Po}(\mathbb{E}N_{T_n}(J_n^{gw}))) > \frac{c\pi(T)}{1 + \log \frac{\pi(T)}{\pi(T)}} > 0, \]
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 \( k_n \to \infty \). Let \( \mathcal{S}_n \) be a sequence of sets of trees with \( \mathcal{S}_n \subseteq \mathcal{T}_{k_n} \). We have:

(i) If \( n\pi(\mathcal{S}_n) \to 0 \), then \( N_{\mathcal{S}_n}(J_n^{gw}) = 0 \) whp.

(ii) If \( n\pi(\mathcal{S}_n) \to \mu \in (0, \infty) \), then \( N_{\mathcal{S}_n}(J_n^{gw}) \overset{d}{\to} \text{Po}(\mu) \).

(iii) If \( n\pi(\mathcal{S}_n) \to \infty \), then
\[ \frac{N_{\mathcal{S}_n}(J_n^{gw}) - n\pi(\mathcal{S}_n)}{\sqrt{n\pi(\mathcal{S}_n)}} \overset{d}{\to} N(0, 1). \]

(iv) If \( \pi(\mathcal{S}_n)/n(\mathcal{T}_{k_n}) \to 0 \), then
\[ \lim_{n \to \infty} d_{\text{TV}}(N_{\mathcal{S}_n}(J_n^{gw}), \text{Po}(n\pi(\mathcal{S}_n))) = 0. \]

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

**Proof of (i) and (ii) of Theorem 1.3.** Let \( k = k_n \) and \( \mathcal{S} = \mathcal{S}_n \). For a fixed \( r \in \mathbb{N} \), since \( k = o(n) \), it follows from Lemma 4.5 that,
\[ \frac{\mathbb{E}(N_{\mathcal{S}}(J_n^{gw}))}{(n\pi(\mathcal{S}))^r} \to 1. \]

Thus if \( n\pi(\mathcal{S}) \to \mu \in [0, \infty) \), then \( \mathbb{E}(N_{\mathcal{S}}(J_n^{gw})) \to \mu^r \). By Lemma 3.10, this implies that
\[ N_{\mathcal{S}}(J_n^{gw}) \overset{d}{\to} \text{Po}(\mu). \]

If \( \mu = 0 \), then \( N_{\mathcal{S}}(J_n^{gw}) = 0 \) whp since it is integer-valued. \( \square \)
Proof of (iii) of Theorem 1.3. Let $k = k_n$ and $\mathcal{G} = \mathcal{G}_n$. Since $k \to \infty$, by Lemma 2.5,
$$
\pi(\mathcal{G}) \leq \pi(\mathcal{I}_k) = \Theta(k^{-3/2}) = o(1).
$$
Thus, letting $r = r_n = \sqrt{n \pi(\mathcal{G})}$, we have $r = o(\sqrt{n})$. Moreover,
$$
\frac{kr^2}{n} = \frac{kn \pi(\mathcal{G})}{n} \leq k \pi(\mathcal{I}_k) = k \Theta(k^{-3/2}) = \Theta(k^{-1/2}) = o(1).
$$
It follows Lemma 4.5 that
$$
\sup_{r \in \sqrt{n \pi(\mathcal{G})}} \left| \frac{\mathbb{E}(N_\mathcal{G}(T^{gw}_n))_r}{(n \pi(\mathcal{G}))^r} - 1 \right| = o(1).
$$
By Lemma 3.10, we have
$$
\frac{N_\mathcal{G}(T^{gw}_n) - n \pi(\mathcal{G})}{\sqrt{n \pi(\mathcal{G})}} \xrightarrow{d} N(0, 1).
$$

Proof of (iv) of Theorem 1.3. Let $k = k_n$ and $\mathcal{G} = \mathcal{G}_n$. By Lemma 4.1, we have
$$
\mathrm{d}_{TV}(N_\mathcal{G}(T^{gw}_n), \text{Po}(n \pi(\mathcal{G}))) \leq (1 + o(1)) \left( \frac{\pi(\mathcal{G})}{\pi(\mathcal{I}_k)} + \sqrt{\frac{\pi(\mathcal{G})}{\pi(\mathcal{I}_k)}} \right) = o(1),
$$
where in the last step we use the assumption that $\pi(\mathcal{G})/\pi(\mathcal{I}_k) \to 0$. 

In this chapter, we apply Theorem 1.2 and 1.3 to study the conditions for $T_{gw}^n$ 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 $T_{gw}^n$. Section 5.3 finds the maximal $K_n$ such that all possible trees of size at most $K_n$ appear in $T_{gw}^n$ 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 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 $\{1, \ldots, n\}$. Let $p_{n,i} := P\{X_n = i\}$. Assume that $p_{n,i} > 0$ for all $1 \leq i \leq n$.

Let $X_{n,1}, X_{n,2}, \ldots$ be i.i.d. copies of $X_n$. Let

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

Let $m_n$ be a positive integers. We have

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

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

The first inequality of this lemma follows from the following:

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

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

Therefore

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

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

$$
P \left( Z_n = 0 \right) \leq P \left( \left| Z_n - \mathbb{E} Z_n \right| \geq \mathbb{E} Z_n \right) \\
\leq \frac{\text{Var} \left( Z_n \right)}{(\mathbb{E} Z_n)^2} \leq \frac{1}{\mathbb{E} Z_n} = \frac{1}{\sum_{i=1}^m (1 - p_{n,i})^m}.
$$

### 5.2 Complete $r$-ary fringe subtrees

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 $\mathcal{T}_{r}^{gw}$ as fringe subtrees.

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

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

#### Lemma 5.2

Assume Condition A and $p_r > 0$ for some $r \geq 2$. Let $h_n \to \infty$ 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 $\omega_n \to \infty$ be an arbitrary sequence.

(i) If $h_n \leq \log_r ( \log n - \omega_n ) - \alpha_r$, then whp $\mathcal{T}_{r}^{gw}$ contains all trees in $\mathcal{G}_{h_n,r}$ as fringe subtrees.

(ii) If $h_n \geq \log_r ( \log n + \omega_n ) - \alpha_r$, then whp $\mathcal{T}_{r}^{gw}$ does not contain all trees in $\mathcal{G}_{h_n,r}$ as fringe subtrees.
Also,
\[ H_{n,r} - \log_r \log n \rightarrow -\alpha_r. \]

**Proof.** Let \( h = h_n \). Let \( T_{h}^{r-ary} \) denote the complete \( r \)-ary tree of height \( h \). Note that if \( T_{h}^{r-ary} \) appears in \( T_{h}^{gw} \) as a fringe subtree, then every tree in \( S_{h,r} \) also appears in \( T_{h}^{gw} \) as a fringe subtree. The tree \( T_{h}^{r-ary} \) has
\[ \ell_n = r^h \text{ leaves and } v_n = (r^h - 1)/(r - 1) = (\ell_n - 1)/(r - 1) \text{ internal vertices, which all have degree } r. \]
Thus we have
\[ \pi(T_{h}^{r-ary}) := \mathbb{P} \{ T_{h}^{gw} = T_{h}^{r-ary} \} = p_r^n v_n^\ell_n. \]

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
\[
\log \frac{1}{\pi(T_{h}^{r-ary})} = v_n \log \frac{1}{p_r} + \ell_n \log \frac{1}{p_0} \\
= \ell_n - \frac{1}{r - 1} \log \frac{1}{p_r} + \ell_n \log \frac{1}{p_0} \\
= \ell_n \left( \frac{1}{r - 1} \log \frac{1}{p_r} + \log \frac{1}{p_0} \right) + O(1) \\
\leq \frac{\log n - \omega_n}{r^{\alpha_r}} + O(1) \\
= \log n - \omega_n + O(1).
\]
Thus \( \log(n \pi(T_{h}^{r-ary})) \geq \omega_n - O(1) \rightarrow \infty \), which implies that \( n \pi(T_{h}^{r-ary}) \rightarrow \infty \).

It follows from Theorem 1.2 that \( N_{h}^{r-ary}(T_{h}^{gw}) \overset{p} \rightarrow \infty \). Thus (i) is proved.

Similar computations show that with the assumptions of (ii), \( n \pi(T_{h}^{r-ary}) \rightarrow 0 \), which implies that \( N_{h}^{r-ary}(T_{h}^{gw}) \overset{p} \rightarrow 0 \) by Theorem 1.2. The last statement of the lemma follows directly from (i) and (ii). □

We have a similar result for \( S_{h,1} \), 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 \( T_{h}^{gw} \) contains all trees in \( S_{h_n,1} \) as fringe subtrees.

(ii) If \( h_n \geq (\log n + \omega_n)/\log \frac{1}{p_1} \), then whp \( T_{h}^{gw} \) does not contain all trees in \( S_{h_n,1} \) as fringe subtrees.
Therefore
\[
\frac{H_{n,1}}{\log_{1/p_1(n)}} \to 1.
\]

5.2.1 Binary trees

Recall that in Subsection 2.3.3, we take a \( T_{gw}^n \) with \( \xi \sim \text{Bi}(1/2, 2) \) 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 \( T_{bin}^n \).

We can define fringe subtree for \( T_{bin}^n \) in the same manner as we did for \( T_{gw}^n \). And our analysis of \( T_{gw}^n \) can be easily adapted to \( T_{bin}^n \). For example, let \( T \) be a chain of height \( h \). Let \( T_{bin}^n \) be a binary tree of height \( h \) that has only left children. Then independently each fringe subtree of the shape \( T \) in \( T_{gw}^n \) has probability \( 1/2^h \) to become a fringe subtree of the shape \( T_{bin}^n \). Let \( N_{T_{bin}}(T_{gw}^n) \) be the number of such fringe subtrees in \( T_{bin}^n \). Then by Lemma 4.6, we have
\[
\mathbb{E}[N_{T_{bin}}(T_{bin}^n)] = \frac{1}{2^n} \mathbb{E}[N(T_{gw}^n)] \sim \frac{1}{2^n} n\pi(T) = \frac{n}{2^h} = \frac{n}{4^n}.
\]

Thus with an argument similar to Lemma 5.3, the maximum chain in \( T_{bin}^n \) that contains only left children has height \( \log_4 n + o_p(1) \).

5.3 All possible fringe subtrees

Recall that \( T_{\leq k}^+ \) denotes the set of all possible trees of size at most \( k \), i.e.,
\[
T_{\leq k}^+ := \{ T \in T : |T| \leq k, \pi(T) > 0 \}.
\]

Also recall that
\[
K_n := \max\{k : T_{\leq k}^+ \subseteq \bigcup_{v \in T_{gw}^n} \{T_{gw}^n \} \}.
\]

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

**Remark 5.1.** We can assume that \( \mathbb{P}\{|T_{gw}^n| = k_n\} > 0 \) for all \( n \in \mathbb{N} \). Otherwise, let \( k_n' := \max\{i \leq k_n : \mathbb{P}\{|T_{gw}^n| = i\} > 0 \} \). 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 \( T_{gw}^n \stackrel{d}{\to} T_{GW} \) [64, thm. 7.12]. In other words, fringe subtrees on average behave like unconditional Galton-Watson trees. Let \( T_k^{\text{min}} \) be a tree \( T \in T_{\leq k}^+ \) that minimizes \( \pi(T) \). Then \( T_k^{\text{min}} \) is also the least likely tree to appear in \( T_{gw}^n \) as fringe subtree among all trees in \( T_{\leq k}^+ \) when \( n \) is large. So intuitively if
Therefore for

and

implies that

as in (4.2), i.e., \((\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n)\) are the indicators for fringe subtrees of size \(k\). Let \(E\) be the event that \(\tilde{\xi}_i = 1\) if and only if \(i \in \{1, k + 1, \ldots, k(m - 1) + 1\}\). In other words, \(E\) implies that \(T_{n}^{gw}\) contains exactly \(m\) fringe subtrees of size \(k\). It suffices to prove that conditioning on \(E\), we can replace the \(m\) fringe subtrees of size \(k\) by independent copies of \(T_{k}^{gw}\) without changing the distribution of the result.

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

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

Therefore for \(1 \leq i \leq m\), \((d_1, \ldots, d_{k(i-1) + 1}, \ldots, d_{ki})\) is a preorder degree sequence of a tree of size \(k\), which we denote by \(T_i\). Let \((T_{k}^{gw}[i])_{i=1}^{m}\) be \(m\) iid copies of \(T_{k}^{gw}\). Write event

\[ A_i = \left[ T_{k}^{gw}[i] = T_i \right], \]

and

\[ B = \left[ (\tilde{\xi}_{km+1}^n, \ldots, \tilde{\xi}_n^n) = (d_{km+1}, \ldots, d_n) \right]. \]

So we need to show that

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

By Lemma 2.4, the lhs of (5.1) equals

\[
\frac{P \left\{ \left( \tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n \right) = (d_1, \ldots, d_n) \right\} }{P\{E\}} = \frac{P \left\{ \left( \xi_1, \ldots, \xi_n \right) = (d_1, \ldots, d_n) \right\} }{P \{ S_n = n - 1 \} P\{E\}}.
\]
Also by Lemma 2.4, we have

\[ P\{B \cap E\} = \frac{P\{(\xi_{km+1}, \ldots, \xi_n) = (d_{km+1}, \ldots, d_n)\} P\{|T^w| = k\}^m}{P\{S_n = n - 1\}} , \]

and

\[ P\{\bigwedge_{i=1}^m A_i\} = \frac{P\{(\xi_1, \ldots, \xi_{km}) = (d_1, \ldots, d_{km})\}}{P\{|T^w| = k\}^m} . \]

Thus the rhs of (5.1) equals

\[ \frac{P\{B \cap E\}}{P\{E\}} = \frac{P\{(\xi_1, \ldots, \xi_n) = (d_1, \ldots, d_n)\}}{P\{S_n = n - 1\} P\{E\}} . \square \]

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

**Corollary 5.1.** Conditioning on that \( N_{S_k}(T^w_n) = m \), the probability that \( T^w_n \) contains all trees in \( S \subseteq S_k \) equals that \( m \) iid copies of \( T^w_k \) contains all trees in \( S \).

**Lemma 5.5.** Assume Condition A. If \( k_n \to \infty \) and

\[ \frac{np_{k_n}^\text{min}}{k_n} \to \alpha, \]

then whp every possible tree of size \( k_n \) appears in \( T^w_n \).

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

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

\[ P\{A_n \mid B_n(a)\} \geq P\{A_n \mid B_n(b)\} \]

given that \( a \gg b \). Therefore

\[ P\{A_n\} = P\{A_n \cap [N_{S_k}(T^w_n) < y_n]\} + \sum_{i \geq y_n} P\{A_n\mid B_n(i)\} P\{B_n(i)\} \geq P\{A_n\mid B_n(y_n)\} \sum_{i \geq y_n} P\{B_n(i)\} , \]

\[ = P\{A_n\mid B_n(y_n)\} (1 - o(1)) . \]

So it suffices to prove that \( P\{A_n^c\mid B_n(y_n)\} \to 0 \).
By Corollary 5.1 and Lemma 5.1 (the coupon collector), we have
\[
P \{ A_n^c | B_n(y_n) \} \leq \sum_{T \in \mathcal{T}_k} |\pi(T) > 0| \left( 1 - P \left\{ \mathcal{T}_{k}^{gw} = T \right\} \right)^{y_n} \\
\leq \sum_{T \in \mathcal{T}_k} |\pi(T) > 0| \exp \left\{ -y_n P \left\{ \mathcal{T}_{k}^{gw} = T \right\} \right\} \\
\leq \sum_{T \in \mathcal{T}_k} |\pi(T) > 0| \exp \left\{ -\frac{1}{2} n P \{|\mathcal{T}^{g}| = k \} P \{|\mathcal{T}^{gw}| = k \} \right\} \\
\leq O(|\mathcal{T}_k|) \exp \left\{ -np_{k}^{min} \right\}.
\]

By Lemma 2.8, $|\mathcal{T}_k| = O(4^k)$. Since by assumption $np_{k}^{min}/k \to \infty$, we have
\[
P \{ A_n^c | B_n(y_n) \} \leq O(4^k) \exp \left\{ -np_{k}^{min} \right\} = O(1) \exp \{k \log(4) - np_{k}^{min}\} \to 0. \tag*{\square}
\]

**Corollary 5.2.** Assume Condition A. If $k_n \to \infty$ and
\[
np_{k}^{min}/k_n \to \infty,
\]
then whp every tree in $\mathcal{T}_{\leq k_n}^+$ appears in $\mathcal{T}_{k_n}^{g, w}$.

**Proof.** Let $k = k_n$. Let $\mathcal{T}_k^+$ be the set of all possible trees of size exactly $k$, i.e.,
\[
\mathcal{T}_k^+ := \{ T \in \mathcal{T} : |T| = k, \pi(T) > 0 \}.
\]

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

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

**Theorem 5.1.** Assume Condition A. Also assume that as $k \to \infty$,
\[
\log \frac{1}{p_{k}^{min}} \sim \gamma k^\alpha (\log k)^\beta,
\]
where $\alpha \geq 1$, $\beta \geq 0$, $\gamma > 0$ are constants. Let $k_n \to \infty$ be a sequence of positive integers. Let $m = \log n$. Then for all constants $\delta > 0$, we have:
\( (i) \) If
\[
k_n \leq (1 - \delta) \left[ \frac{m}{\gamma (\log m^{1/\alpha})^{\beta}} \right]^{1/\alpha},
\]
then \( \text{whp} \ T_n^{\text{gw}} \) contains all trees in \( \mathcal{T}_{k_n}^{+} \) as fringe subtrees.

\( (ii) \) If
\[
k_n \geq (1 + \delta) \left[ \frac{m}{\gamma (\log m^{1/\alpha})^{\beta}} \right]^{1/\alpha},
\]
then \( \text{whp} \ T_n^{\text{gw}} \) does not contain all trees in \( \mathcal{T}_{k_n}^{+} \) 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 \( p_{k_n}^{\text{min}} \) varies for different offspring distributions. But as mentioned in the introduction, the types of trees that we are interested in all have \( p_{k_n}^{\text{min}} \) 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
\[
\log \frac{1}{p_{k_n}^{\text{min}}} \sim \gamma k^{\alpha} (\log k)^{\beta} \leq (\gamma + o(1)) \frac{(1 - \delta)^{\alpha} m}{\gamma (\log m^{1/\alpha})^{\beta}} (\log m^{1/\alpha})^{\beta} \sim (1 - \delta)^{\alpha} m.
\]

Therefore, recalling \( m = \log n \),
\[
\log np_{k_n}^{\text{min}} = m - \log \frac{1}{p_{k_n}^{\text{min}}} \geq m - (1 + o(1))(1 - \delta)^{\alpha} m = \Omega(m).
\]

It follows that
\[
\log k - \log np_{k_n}^{\text{min}} = O(\log m) - \Omega(m) \to -\infty.
\]

Thus \( np_{k_n}^{\text{min}}/k \to \infty \) and it follows from Corollary 5.2 that \( \text{whp} \ T_n^{\text{gw}} \) contains every tree in \( \mathcal{T}_{k_n}^{+} \) as a fringe subtree.

Similar computations show that (5.3) implies \( np_{k_n}^{\text{min}} \to 0 \). It follows from Theorem 1.2 that \( N_{\text{min}}(T_n^{\text{gw}}) \xrightarrow{p} 0 \). Thus \( \text{whp} \ T_n^{\text{gw}} \) does not contain every tree in \( \mathcal{T}_{k_n}^{+} \) as a fringe subtree.

The rest of this section is organized as follows: In Subsection 5.3.1, we give a general method of finding \( p_{k_n}^{\text{min}} \). 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 ALL POSSIBLE FRINGE SUBTREES

5.3.1 Computing \( p_{k}^{\text{min}} \)

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

\[
\text{supp}(\xi, k) := \text{supp}(\xi) \cap \{1, \ldots, k\}.
\]

Let \( L_{1} := p_{0} \) and for \( k \geq 2 \) let

\[
L_{k} := \min \left\{ p_{0} \left( \frac{p_{k}}{p_{0}} \right)^{1/k} : i \in \text{supp}(\xi, k - 1) \right\}.
\]

Since \( L_{k} \) is non-increasing, \( L := \lim_{k \to \infty} L_{k} \) exists. Equivalently, we have

\[
L := \inf \left\{ p_{0} \left( \frac{p_{k}}{p_{0}} \right)^{1/i} : i \in \text{supp}(\xi) \setminus \{0\} \right\}.
\]

**Theorem 5.2.** Assume Condition A. We have

\[
(p_{k}^{\text{min}})^{1/k} \to L \leq p_{\text{max}} < 1
\]

as \( k \to \infty \), where the limit is taken along the subsequence \( k \) with \( P \{|T^{g_{w}}| = k\} > 0 \).

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

**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 \( P \{|T^{g_{w}}| = k\} > 0 \), there are at least \( k^{-C'_{i}}C''_{i}^{k} \) trees \( T \) of size \( k \) with

\[
0 < \pi(T) \leq C''_{i} \left[ p_{0} \left( \frac{p_{k}}{p_{0}} \right)^{1/i} \right]^{k}.
\]

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

\[
\begin{align*}
y &\equiv x \pmod{i}, \\
y &\equiv 0 \pmod{j}.
\end{align*}
\]

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

\[
\begin{align*}
n_{0} &= k - n_{i} - n_{j}, \\
n_{j} &= \frac{y}{j}, \\
n_{i} &= \frac{k - 1 - y}{i},
\end{align*}
\]
such that \(n_0, n_i, n_j\) are all non-negative integers with
\[
n_0 + n_i + n_j = k, \text{ and } in_i + jn_j = k - 1.
\]

Let \(\Sigma_k(n_0, n_i, n_j)\) 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
\[
|\Sigma_k(n_0, n_i, n_j)| = \frac{1}{k} \binom{k}{n_0, n_i, n_j} = \frac{k!}{k n_0! n_i! n_j!}.
\]

(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
\[
|\Sigma_k(n_0, n_i, n_j)| \geq k^{-C_i'} \left( \frac{1}{i} \right)^{1/i} \left( 1 - \frac{1}{i} \right)^{1-1/i} \left( \frac{1}{1-i} \right)^{1-1/i} := k^{-C_i'} C_i^k.
\]

And for every \(T \in \Sigma_k(n_0, n_i, n_j)\), we have
\[
\pi(T) \leq p_i^n p_0^{-n_0} \leq p_i^{-C_i^*} \left( \frac{p_i}{p_0} \right)^{1/i} \left( \frac{p_0}{p_i} \right)^{(1-1/i)} := C_i'' \left( \frac{p_i}{p_0} \right)^{1/i} \left( \frac{p_0}{p_i} \right)^{(1-1/i)}. \quad \square
\]

**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 \(\omega\) such that the span of \(\text{supp}(\xi, \omega) := \text{supp}(\xi) \cap \{0, \ldots, \omega\}\) 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
\[
n_i = \left\lceil \frac{k - 1 - D}{i} \right\rceil,
\]
such that there exists non-negative integers \(n_0, n_1, \ldots, n_\omega\) 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
\[
\left| n_0 - k \left(1 - \frac{1}{i}\right) \right| \leq D^*, \quad \left| n_i - \frac{k}{i} \right| \leq D^*, \quad \sum_{j=1}^{\omega} n_j \leq D^*.
\]

\((5.4)\)
where \( D^* \) is a constant. From here we can estimate the number of trees \( T \) with \( n_j \) nodes of degree \( j \) for \( j \in \{0, \ldots, \omega \} \cup \{i\} \) and the probability \( \pi(T) \) to finish the proof as before.

If \( i \leq \omega \), similar to the above argument, we can choose
\[
\tau_i = \left\lfloor \frac{k - 1 - D^*}{i} \right\rfloor,
\]
such that there exist non-negative integers \( s_0, \ldots, s_\omega \) satisfying
\[
k = \tau_i + \sum_{j = 1}^{\omega} j s_j, \quad \text{and} \quad k - 1 = \tau_i + \sum_{j = 1}^{\omega} j s_j.
\]
Letting \( n_j = s_j + \lceil j = i \rceil \cdot \tau_i \), we again get (5.4).

**Proof of Theorem 5.2.** Let \( T \) be a tree with \( |T| = t \) and \( \pi(T) > 0 \). So \( T \in \mathfrak{F}^t \). 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
\[
n_0 + n_1 + \ldots + n_{t-1} = t, \quad \text{and} \quad n_1 + 2n_2 + \ldots + (t-1)n_{t-1} = t - 1.
\]
Using the convention that \( 0^0 = 1 \), we have for \( t \geq 2 \)
\[
\pi(T) = p_0^{n_0} p_1^{n_1} \cdots p_{t-1}^{n_{t-1}} \\
= p_0^{n_0 + n_1 + \ldots + n_{t-1}} \left( \frac{p_1}{p_0} \right)^{n_1} \left( \frac{p_2}{p_0} \right)^{n_2} \cdots \left( \frac{p_{t-1}}{p_0} \right)^{n_{t-1}} \\
= p_0^t \prod_{i \in \text{supp}(\xi, t-1)} \left( \frac{p_i}{p_0} \right)^{i n_i} \\
\geq p_0^t \left[ \min_{i \in \text{supp}(\xi, t-1)} \left( \frac{p_i}{p_0} \right)^{i n_i} \right]^\frac{t-1}{t} \sum_{i=1}^{t-1} in_i \\
= p_0 L_t^{-1} \geq p_0 L^{-1}.
\]
Since \( p_{\max}^t \geq \pi(T) \), we have for all \( t \)
\[
p_{\max} \geq p_0^{1/t} L^{1-1/t}.
\]
Therefore \( L \leq p_{\max} < 1 \). Recalling that \( \mathfrak{F}^t \) is the set of possible trees of size \( t \), we have
\[
p_k^{\min} := \min_{1 \leq t \leq k} \min_{T \in \mathfrak{F}^t} \pi(T) \geq \min_{1 \leq t \leq k} p_0 L^{t-1} = p_0 L^{k-1}, \quad (5.5)
\]
where the last equality follows from \( L < 1 \). As a result
\[
\liminf_{k \to \infty} (p_k^{\min})^{1/k} \geq \liminf_{k \to \infty} \left( \frac{p_0}{L} \right)^{1/k} L = L.
\]
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
$$p_k^{\text{min}} \leq P\{T^*_w = T\} \leq C \alpha \left[p_0 \left(\frac{p_\alpha}{p_0}\right)^{1/\alpha}\right]^k \leq C \alpha (L + \varepsilon)^k,$$
where $C_\alpha > 0$ is constant. Thus $\limsup_{k \to \infty} (p_k^{\text{min}})^{1/k} \leq L + \varepsilon$. Since $\varepsilon$ is arbitrary, we have $\limsup_{k \to \infty} (p_k^{\text{min}})^{1/k} \leq L$. \hfill $\Box$

5.3.2 When $L > 0$

If $L > 0$, then by Theorem 5.2,
$$\log \frac{1}{p_k^{\text{min}}} \sim \log \left(\frac{1}{L}\right)^k = \log \left(\frac{1}{L}\right)^k (\log k)^0.$$
Thus we can apply Theorem 5.1 with $\gamma = \log(1/L)$, $\alpha = 1$ and $\beta = 0$ to get
$$\frac{K_n}{\log(n)} \to \frac{1}{\log(1/L)}.$$

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 $\xi \overset{d}{=} 2 \text{Be}(1/2)$ (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 $\xi \overset{d}{=} \text{Bi}(d, 1/d)$ for $d \geq 2$ (d-ary tree), then $L = (d-1)^{d-1}/d^d$.

(iv) If $\xi \overset{d}{=} \text{Ge}(1/2)$ (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 $L = \lim_{k \to \infty} L_k = 1/2$. (ii) and (iii) follow from similar simple calculations.

(iv): For all $i \geq 1$, we have
$$p_0 \left(\frac{p_i}{p_0}\right)^{1/i} = \frac{1}{2} \left(\frac{1}{2i}\right)^{1/i} = 1/4.$$ 
Therefore $L_k = 1/4$ for all $k \geq 1$, and $L = 1/4$. \hfill $\Box$
Define
\[\mathcal{L} = \begin{cases} 
\min\{i \in \mathbb{N} : L_{i+1} = L\} & \text{if } L_j = L \text{ for some } j, \\
\infty & \text{otherwise.}
\end{cases}\]

If \(\mathcal{L} < \infty\), then we call the Galton-Watson tree well-behaved. Examples of such trees include those for which \(\xi\) is bounded, those for which \(\xi\) has a polynomial or slower-than-exponential tail, and the case \(\xi \overset{d}{=} \text{Ge}(1/2)\). 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 \(T^\text{gw}_n\) contains all trees in \(\mathcal{T}_{< k_n}^+\) 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 \(T^\text{gw}_n\) does not contain all trees in \(\mathcal{T}_{< k_n}^+\) as fringe subtrees.

Thus as \(n \to \infty\), we have for well-behaved Galton-Watson trees,

\[\frac{K_n \log(1/L) - \log n}{\log \log n} \to 1.\]

The main idea is that when \(\mathcal{L} < \infty\), there are exponentially many trees of size \(k\) that have small probability to appear as fringe subtrees in \(T^\text{gw}_n\). 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 \(np_{k}^{\min}/k \to \infty\). 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 \(N_{\mathcal{T}_k}(T^\text{gw}_n) \leq y_n := \left\lfloor \frac{3}{2} n \mathbb{P}\{||T^\text{gw}_n|| = k\}\right\rfloor\). Let \(A_n\) be the event that \(T^\text{gw}_n\) contains all possible trees of size \(k\) as fringe subtrees. Let \(B_n(i)\) be the event that \(N_{\mathcal{T}_k}(T^\text{gw}_n) = i\) for some \(i \leq y_n\). Then

\[\mathbb{P}\{A_n\} \leq \mathbb{P}\{N_{\mathcal{T}_k}(T^\text{gw}_n) > y_n\} + \sum_{i \leq y_n} \mathbb{P}\{A_n|B_n(i)\} \mathbb{P}\{B_n(i)\}\]

\[\leq o(1) + \mathbb{P}\{A_n|B_n(y_n)\} \sum_{i \leq y_n} \mathbb{P}\{B_n(i)\},\]

\[\leq o(1) + \mathbb{P}\{A_n|B_n(y_n)\}.\]

Thus it suffices to prove that \(\mathbb{P}\{A_n|B_n(y_n)\} \to 0\).
By Corollary 5.1, we have \( \mathbb{P}\{A_n|B_n(y_n)\} \) equals the probability that \( y_n \) independent copies of \( T_k^{|w} \) do not contain all trees in \( \mathcal{T}^+_k \). It follows from Lemma 5.1 (the coupon collector) that \( \mathbb{P}\{A_n|B_n(y_n)\} \to 0 \) if

\[
\sum_{T \in \mathcal{T}^+_k} (1 - \mathbb{P}\{T_k^{|w} = T\})^{y_n} \to \infty.
\]

By definition of \( \xi \), we have \( L = p_0(p/p_0)^{1/\xi} \). 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 \( \mathcal{T}^+_k \) with

\[
\mathbb{P}\{T_k^{|w} = T\} = \frac{\mathbb{P}\{T_k^{|w} = T\}}{\mathbb{P}\{|T^{|w}| = k\}} \leq C'(p_0(p/p_0)^{1/\xi})^k \frac{C''L^k}{\mathbb{P}\{|T^{|w}| = k\}}.
\]

Therefore

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

See Theorem 5.2 and Lemma 2.5.

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

\[
k \log(C) - C' \log(k) + y_n \log \left( 1 - \frac{C''L^k}{\mathbb{P}\{|T^{|w}| = k\}} \right)
\]

\[
\geq \frac{1}{2} k \log(C) - 2y_n \frac{C''L^k}{\mathbb{P}\{|T^{|w}| = k\}}
\]

\[
\geq \frac{1}{2} k \log(C) - 3nC''L^k = \frac{1}{2} k \log(C) - O(nL^k).
\]

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

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

which implies \( \mathbb{P}\{A_n|B_n(y_n)\} \to 0 \).

\[ \square \]

**Remark 5.3.** If \( L > 0 \) and \( \xi = \infty \), then we have \( K_n/\log(n) \overset{p}{\to} 1/\log(1/L) \) 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 \( \xi \) has infinite support and

\[
\liminf_{i \to \infty} p_0 \left( \frac{p_1}{p_0} \right)^{1/i} = 0.
\]
which implies
\[
\limsup_{i \to \infty} \frac{\log(1/p_i)}{i} = \infty,
\]
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 \( f : [0, \infty) \to [0, \infty) \) with \( f(i)/i \uparrow \infty \), then we say that \( \xi \) 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 \( \xi \) 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 \( \mathcal{T}^n_{gw} \) contains all trees in \( \mathcal{T}_{\leq k_n}^+ \) as fringe subtrees.

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

Therefore,
\[
\frac{K_n}{f^{-1}(\log n)} \to 1.
\]

**Proof.** (i): Let \( k = k_n \). Let \( \epsilon > 0 \) be a constant decided later. Since \( \log(1/p_i) \sim f(i) \), there exists an integer \( i(\epsilon) \) such that for all \( i > i(\epsilon) \),
\[
\log p_i \geq -(1 + \epsilon/2)f(i).
\]
Let \( w_i := \log p_{0}(p_{i}/p_{0})^{1/i} \). We have as \( k \to \infty \),
\[
\min_{i(\epsilon) < i < k} \, w_i = \min_{i(\epsilon) < i < k} \left\{ \left(1 - \frac{1}{i}\right) \log(p_0) + \frac{\log(p_i)}{i} \right\} \geq \log(p_0) - \max_{i(\epsilon) < i < k} \left( \frac{1 + \epsilon/2}{i} \right) f(i) \geq \log(p_0) - \frac{(1 + \epsilon/2)f(k - 1)}{k - 1} \to -\infty,
\]
where we use that \( f(i)/i \uparrow \infty \). Since \( \min_{1 < i < k} \, w_i \) is a constant, we have for large \( k \),
\[
\log L_k := \min_{1 < i < k} \, w_i \geq \log(p_0) - \frac{(1 + \epsilon/2)f(k - 1)}{k - 1}.
\]
It follows from (5.5) that for \( k \) large enough,
\[
\log p_k^\min \geq \log(p_0 (k - 1)) \geq \log(p_0) + (k - 1) \log p_0 - (1 + \epsilon/2)f(k - 1) \geq -(1 + \epsilon)f(k - 1),
\]
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 \( \varepsilon \) to be small enough with respect to \( \delta \). Thus \( np_k^{\min} \geq n^{5/2} \). We have \( np_k^{\min}/k \to \infty \). It follows from Corollary 5.2 that \( T_n^{qw} \) 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), \( n\pi(T_{k-1}^{\star}) \to 0 \). Therefore \( T_n^{qw} \) does not contain \( T_{k-1}^{\star} \) whp. \( \Box \)

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

\[
\frac{K_n}{\sqrt{\log(n)}} \frac{p_i}{\sqrt{c'}} \quad \text{as} \quad n \to \infty.
\]

**Example 5.2** (The Cayley trees). Another example is the Galton-Watson tree with offspring distribution \( \xi \), \( \xi \overset{\mathcal{D}}{=} \text{Po}(1) \), 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, \( \xi \) has an \( f \)-super-exponential tail with \( f(i) = i \log(i) \). Since \( f^{-1}(j) \sim j/\log j \), we have

\[
\frac{K_n \log \log n}{\log n} \to 1.
\]

Using (5.5) we can show that if the tail of \( \xi \) 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, \( p_k^{\min} = \pi(T_{k-1}^{\star}) = p_0 (p_0^{-1} - 1/k - 1/(k + 1)) \geq 0 \),

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

Recall that \( \pi(T) := P(T^{qw} = T) \).

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

\[
\log \left(\frac{p_k}{p_0}\right)^{\frac{r_k}{r_i}} - \log \left(\frac{p_{k+1}}{p_0}\right)^{\frac{r_{k+1}}{r_i}} = r_{k+1} - r_k + \log (p_0^{-1}) \left(\frac{1}{k} - \frac{1}{k + 1}\right) \geq r_{k+1} - r_k \geq 0,
\]

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 \( r_k \to \infty \),

\[
p_0 \left(\frac{p_k}{p_0}\right)^{1/k} = p_0^{1-1/k} e^{-r_k} \to 0.
\]
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
\[ p_k^{\text{min}} \geq p_0 L_k^{k-1} = p_0 \left[ p_0 \left( \frac{p_{k-1}}{p_0} \right)^{\frac{1}{k-1}} \right]^{k-1} = p_0^{k-1} p_{k-1} = \pi(T_{k-1}^{\text{star}}). \]

By the definition of \( p_k^{\text{min}} \), we have
\[ p_k^{\text{min}} := \min\{\pi(T) : \|T\| \leq k, \pi(T) > 0\} \leq \pi(T_{k-1}^{\text{star}}). \]

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

In particular, Lemma 5.8 applies to \( \xi \leq P_0(1) \). In this case we have
\[ \log p_k^{\text{min}} = \log(p_0^{k-1} p_{k-1}) = -k \log(k + O(1/k)). \]

It follows from Theorem 5.1 that,
\[ K_n \log \log n \rightarrow_p 1, \]

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

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 \( T_R^{\text{gw}} \) contains all trees in \( \mathcal{Z}_{\leq k_n}^+ \) 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 \( T_R^{\text{gw}} \) does not contain all trees in \( \mathcal{Z}_{\leq k_n}^+ \) as fringe subtrees.
Thus, as \( n \to \infty \),

\[
K_n \frac{\log \log n}{\log n} \overset{p}{\to} \frac{1}{\gamma'}
\]

and more precisely,

\[
\left( \frac{K_n \gamma}{\log n} \left( \log \frac{\log n}{\gamma} - \log \log \frac{\log n}{\gamma} \right) - 1 \right) \times \frac{(\log \log n)^2}{\log \log \log n} \overset{p}{\to} -1.
\]

Proof. Let \( k = k_n \). For (i), we can show that \( np_{k_{\text{min}}} \overset{p}{\to} \infty \). It follows from Corollary 5.2 that whp \( T_{n}^{gw} \) contains all trees of size at most \( k \) as fringe subtrees.

For (ii), similarly we can show that \( np_{k_{\text{min}}} \overset{p}{\to} 0 \). It follows from Theorem 1.2 that whp there is at least one tree in \( \mathcal{T}_{\leq k}^{+} \) that does not appear in \( T_{n}^{gw} \) as fringe subtrees. \( \square \)
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 $T_{gw}^{n_r}$. Finally, Section 6.5 discusses the difference between fringe and non-fringe subtrees.

### 6.1 Intuition and Notations

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

$$N_{n}^{nf}(T_{gw}^{n_r}) := \sum_{u \in T_{gw}^{n_r}} [T < T_{gw}^{n_r}],$$

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

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

$$(\vec{a}_1, 0, \vec{a}_2, 0, \ldots, \vec{a}_\ell, 0) := (a_{1,1}, a_{1,2}, \ldots, a_{1,r(1)}, 0, a_{2,1}, a_{2,2}, \ldots, a_{2,r(2)}, 0, \ldots, a_{\ell,1}, a_{\ell,2}, \ldots, a_{\ell,r(\ell)}, 0)$$  \hspace{1cm} (6.1)

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

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

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

$$(\vec{a}_1, \vec{b}_1, \vec{a}_2, \vec{b}_2, \ldots, \vec{a}_\ell, \vec{b}_\ell)$$  \hspace{1cm} (6.3)
where $\vec{b}_1, \ldots, \vec{b}_t$ are tree degree sequences, i.e., preorder degree sequences of some plane trees. Thus each non-fringe subtree of shape $T$ in $\mathcal{T}_n^\mathrm{gw}$ corresponds to a segment of $(\vec{\xi}_r^n)_{r=1}^n$ of the form of (6.3). If none of the segments overlap with each other, then we can permute them into the form $(\vec{a}_1, \ldots, \vec{a}_\ell, \vec{b}_1, \ldots, \vec{b}_t)$. Since $(\vec{\xi}_r^n)_{r=1}^n$ is permutation invariant, the old and new sequences have the same probability. In other words, $N_1^\mathrm{nf}(\mathcal{T}_n^\mathrm{gw})$ is almost distributed like the number of the patterns $(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_\ell)$ in $(\vec{\xi}_r^n)_{r=1}^n$.

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.,

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

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

$$
deg_i(\vec{d}) := (d_{i_1}, d_{i_1+1}, \ldots, d_{i_1+k_i(\vec{d})-1}) \in \mathcal{D}(\mathfrak{T}_{k_i(\vec{d})}),
$$

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

### 6.2 Factorial Moments

We now prove the exact formulas for the first and second moments of $N_1^\mathrm{nf}(\mathcal{T}_n^\mathrm{gw})$. Throughout this section, let $T$ be a tree with a preorder degree sequence of the form $(\vec{a}_1, \vec{0}, \ldots, \vec{a}_\ell, \vec{0})$ satisfying (6.2). Let $v := v(T)$ and $\ell := \ell(T)$.

#### 6.2.1 The first factorial moment

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

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

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

$$v = t(1) < t(2) < \ldots < t(\ell) < t(\ell+1) \leq n.$$
Thus to reach \( W \) time when the entire walk also a tree degree sequence, i.e., it satisfies (2.1). Moreover the sequence 
\[
(d_1, d_2, \ldots, d_{t(t+1)})
\]
is also a tree degree sequence.

**Proof.** Consider the walk \( W_t \) defined by
\[
W_t = \begin{cases} 
0, & \text{if } t = 0, \\
\sum_{i=1}^{n} d_i - 1, & \text{if } 1 \leq t \leq n.
\end{cases}
\]

Since \( d \in \mathcal{R}_n \),
\[
W_n := \sum_{i=1}^{n} (d_i - 1) = \left( \sum_{i=1}^{n} d_i \right) - n = (n-1) - n = -1.
\]

So this walk starts from 0 and ends at \(-1\) after \( n \) steps. By (6.2),
\[
W_{t(1)} := W_v = \sum_{i=1}^{v} (d_i - 1) = \left( \sum_{s=1}^{\ell} \tilde{d}_s \right) - v = (v + \ell - 1) - v = \ell - 1 \geq 0.
\]

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 \((d_{t(1)+1}, \ldots, d_{t(2)})\) 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 \( 1 \leq i \leq \ell + 1 \). Using the above argument, we also can verify that \( t(3), \ldots, t(\ell + 1) \) are as required.

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

**Lemma 6.2.** Assume that \( \mathbb{P} \{ |T_n^\xi| = n \} > 0 \) and that \( v < n \). We have
\[
\mathbb{E} \left[ \frac{N_n^T (T_n^\xi)}{n} \right] = \frac{\pi_n^T (T) \mathbb{P} \{ S_n - v = n - v - \ell \}}{\mathbb{P} \{ S_n = n - 1 \}}.
\]

Recall that \( \pi_n^T (T) := \mathbb{P} \{ T < T_n^\xi \} \) and \( S_n \) is the sum of \( n \) iid copies of \( \xi \).
Proof. Recall that $T_i(\tilde{x}_1^n, \ldots, \tilde{x}_n^n)$ is the fringe subtree that has the preorder degree sequence $\text{deg}_i(\tilde{x}_1^n, \ldots, \tilde{x}_n^n)$. Let
\[ I_i = \# \{ T < T_i(\tilde{x}_1^n, \ldots, \tilde{x}_n^n) \}. \tag{6.4} \]
Then with an argument similar to Lemma 2.6, $N_{T^n}^{nf}(T^{gw}_n) = \sum_{i=1}^n I_i$. Thus we have
\[ \mathbb{E} \left[ N_{T^n}^{nf}(T^{gw}_n) \right] = \mathbb{E} \left[ \sum_{i=1}^n I_i \right] = n \mathbb{P} \{ I_1 = 1 \}. \]

Recall that $T$ has a preorder degree sequence of the form $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_\ell, 0)$ satisfying (6.2). Let $A \subseteq \mathcal{R}_n$ be the set of sequences such that $(\tilde{c}_r^n)_{r=1}^n \in A$ if and only if $I_1 = 1$. In other words, $\tilde{d} := (d_1, d_2, \ldots, d_n) \in A$ if and only if $\text{deg}_1(\tilde{d}) = (\tilde{a}_1, \tilde{b}_1, \ldots, \tilde{a}_\ell, \tilde{b}_\ell)$ for some $\tilde{b}_1, \ldots, \tilde{b}_\ell$ which are tree degree sequences. By permuting $\text{deg}_1(\tilde{d})$ into $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_\ell, \tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_\ell)$, we get a new sequence $\tilde{d}' := (d_1, d_2', \ldots, d_n') \in A'$ where
\[ A' := \{ (e_1, e_2, \ldots, e_n) \in \mathcal{R}_n : (e_1, e_2, \ldots, e_n) = (d_1, d_2, \ldots, d_n) \}. \]
Such a permutation defines a mapping $f : A \rightarrow A'$.

For every $\tilde{d}' = (d_1', \ldots, d_n') \in A'$, Lemma 6.1 implies that in $\tilde{d}'$ the segment $(\tilde{a}_1, \ldots, \tilde{a}_\ell)$ is followed by $\ell$ tree degree sequences. So we can always reverse the above permutation and find a unique $\tilde{d} \in A$ with $f(\tilde{d}) = \tilde{d}'$. Thus $f$ is a one-to-one mapping. If $\tilde{d}' = f(\tilde{d})$, then $\mathbb{P} \{ (\tilde{c}_r^n)_{r=1}^n = \tilde{d} \} = \mathbb{P} \{ (\tilde{c}_r^n)_{r=1}^n = \tilde{d}' \}$, since $(\tilde{c}_r^n)_{r=1}^n$ is permutation invariant. Therefore we have
\[ \mathbb{P} \{ I_1 = 1 \} = \mathbb{P} \{ (\tilde{c}_r^n)_{r=1}^n \in A \} = \mathbb{P} \{ (\tilde{c}_r^n)_{r=1}^n \in A' \}. \]

Recall that by Lemma 2.4,
\[ (\tilde{c}_r^n)_{r=1}^n \sim (\xi_1, \ldots, \xi_n \mid S_n = n - 1). \]

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

We have
\[ \mathbb{P} \left\{ (\xi_1, \ldots, \xi_v) = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_\ell) \right\} \]
\[ = \mathbb{P} \{ (\xi_1, \xi_2, \ldots, \xi_v) = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_\ell), S_n = n - 1 \} \]
\[ \mathbb{P} \{ S_n = n - 1 \} \]
\[ = \frac{\pi^{nf}(T) \mathbb{P} \{ S_{n-v} = n - v - \ell \}}{\mathbb{P} \{ S_n = n - 1 \}}, \]
where the last step uses $\sum_{s=1}^\ell \tilde{a}_s = v + \ell - 1$ from (6.2). \hfill \Box
6.2.2 The second factorial moment

To compute $\mathbb{E}(N_T^n(J_n^{\omega_T}))$, we enumerate all the cases that $T$ can appear as overlapping non-fringe subtrees by constructing a set of trees $\{T \sqcup T\}$ 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 $V(T)$ be the set of internal nodes of $T$. Then define the set of trees

$$\{T \sqcup T\} = \bigcup_{u \in V(T): T_u < T} \{\text{Splay}(T, u, T)\} \setminus \{T\}.$$ 

Note that $|\{T \sqcup T\}| < v : = |V(T)|$. Also note that given $T' \in \{T \sqcup T\}$ 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 $\{T \sqcup T\}$.

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

Recall that $T$ has $v$ internal nodes and $t$ leaves.

![Figure 6: Example of $\{T \sqcup T\}$](image)

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

$$\mathbb{E}(N_T^n(J_n^{\omega_T})) = \sum_{1 \leq i < j \leq n} \mathbb{E}[I_iI_j] = n \sum_{i=2}^n \mathbb{E}[I_1I_i].$$

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 $P\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\}$ over all $\tilde{d}$ such that the event $(\tilde{\xi}_r^n)_{r=1}^n = \tilde{d}$ implies $I_1I_i = 1$ for some $i$.

**Lemma 6.3.** Let $B$ be the set of pairs $(i, \tilde{d}) \in \{2, \ldots, n\} \times R_n$ such that $T < T_1(\tilde{d})$ and $T < T_1(\tilde{d})$. Then

$$\mathbb{E}(N_T^n(J_n^{\omega_T}))_2 = \sum_{i \geq 2} \mathbb{E}[I_1I_i] = \sum_{(i, \tilde{d}) \in B} P\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\}.$$ 

Recall that $T_1(\tilde{d})$ is the tree with preorder degree sequence $\text{deg}_j(\tilde{d})$. 

Recall that $T_1(\tilde{d})$ is the tree with preorder degree sequence $\text{deg}_j(\tilde{d})$. 

Proof. The event \( I_1 I_k = 1 \) happens if and only if
\[
T < T_1(\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n), \quad \text{and} \quad T < T_1(\tilde{\xi}_1^n, \ldots, \tilde{\xi}_n^n),
\]
i.e., \((i, (\tilde{\xi}_r^n)_{r=1}^n) \in \mathcal{B}\). Therefore
\[
\sum_{i=2}^n \mathbb{E}[I_1 I_i] = \sum_{i=2}^n \sum_{d \in \mathbb{R}_n} \mathbb{P}\left\{(i, \tilde{d}) \in \mathcal{B} \mid (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d}\right\} = \sum_{(i, d) \in \mathcal{B}} \mathbb{P}\left\{(\tilde{\xi}_r^n)_{r=1}^n = \tilde{d}\right\}. \quad \Box
\]

Let \((i, \tilde{d}) \in \mathcal{B}\). If \((\tilde{\xi}_r^n)_{r=1}^n = \tilde{d}\), then \(T < T_1((\tilde{\xi}_r^n)_{r=1}^n)\) and \(T < T_1((\tilde{\xi}_r^n)_{r=1}^n)\). The two fringe subtrees \(T_1((\tilde{\xi}_r^n)_{r=1}^n)\) and \(T_1((\tilde{\xi}_r^n)_{r=1}^n)\) 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 \(T_1((\tilde{\xi}_r^n)_{r=1}^n)\) that correspond to the internal nodes of \(T\) overlap.

To be precise, assume that
\[
\deg_1(\tilde{d}) = (\bar{a}_1, \bar{b}_1, \bar{a}_2, \bar{b}_2, \ldots, \bar{a}_\ell, \bar{b}_\ell),
\]
and
\[
\deg_\ell(\tilde{d}) = (\bar{a}_1, \bar{b}_1, \bar{a}_2, \bar{b}_2, \ldots, \bar{a}_\ell, \bar{b}_\ell),
\]
where \((\bar{a}_1, 0, \ldots, \bar{a}_\ell, 0)\) is the preorder degree sequence of \(T\) which satisfies (6.2) and \(\bar{b}_1, \bar{b}_1, \ldots, \bar{b}_\ell, \bar{b}_\ell\) are tree degree sequences.

Let \(M_1^a(\tilde{d})\) be the set of positions (indices) in \(\tilde{d}\) that are occupied by the \(\bar{a}_1, \ldots, \bar{a}_\ell\) parts of \(\deg_1(\tilde{d})\). Let \(M_1^b(\tilde{d})\) be the set of positions in \(\tilde{d}\) that are occupied by the \(\bar{b}_1, \ldots, \bar{b}_\ell\) parts of \(\deg_1(\tilde{d})\). Let \(M_1(\tilde{d}) = M_1^a(\tilde{d}) \cup M_1^b(\tilde{d})\), i.e., \(M_1(\tilde{d})\) contains the positions occupied by \(\deg_1(\tilde{d})\). Define \(M_1^a(\tilde{d})\), \(M_1^b(\tilde{d})\) and \(M_1(\tilde{d})\) for \(\deg_\ell(\tilde{d})\) accordingly.

To make the above definitions more formal, we can let
\[
M^a_{1,s}(\tilde{d}) := \left\{ j \mod n : 0 < j - \sum_{s=1}^{\ell-1} (|\bar{a}_s| + |\bar{b}_s|) \leq |\bar{a}_s| \right\},
\]
and
\[
M^b_{1,s}(\tilde{d}) := \left\{ j \mod n : |\bar{a}_s| < j - \sum_{s=1}^{\ell-1} (|\bar{a}_s| + |\bar{b}_s|) \leq |\bar{a}_s| + |\bar{b}_s| \right\},
\]
where \(|\bar{a}_s|\) and \(|\bar{b}_s|\) denote the length of \(\bar{a}_s\) and \(\bar{b}_s\) respectively. We can also define \(M^a_{1,s}(\tilde{d})\) and \(M^b_{1,s}(\tilde{d})\) accordingly. Then
\[
M^a_1(\tilde{d}) := \bigcup_{s=1}^\ell M^a_{1,s}(\tilde{d}), \quad \text{and} \quad M^b_1(\tilde{d}) := \bigcup_{s=1}^\ell M^b_{1,s}(\tilde{d}).
\]
We consider two cases — $M^f_t(\tilde{d})$ and $M^f_t(d)$ overlap or not. Let

$$B' = \left\{ (i, \tilde{d}) \in B : M^f_t(\tilde{d}) \cap M^f_t(d) = \emptyset \right\}.$$ 

Let $B'' := B \setminus B'$. We first deal the easier case $(i, \tilde{d}) \in B''$.

**Lemma 6.4.** Assume that $\mathbb{P}\{|T^\#| = n\} > 0$ and $\nu = \nu(T) \leq n/2$. We have

$$\sum_{(i, \tilde{d}) \in B''} \mathbb{P}\{(\hat{\xi}^n_r)^n_{r=1} = \tilde{d}\} = 2 \sum_{T' \in \{T \sqcup T\}} \pi^{n_f}(T') \frac{\mathbb{P}\{S_{n - \nu(T')} = n - \nu(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}}.$$

**Proof.** Let $(i, \tilde{d}) \in B''$. Since $M^f_t(\tilde{d}) \cap M^f_t(d) \neq \emptyset$, when $(\hat{\xi}^n_r)^n_{r=1} = \tilde{d}$, either $T_1((\hat{\xi}^n_r)^n_{r=1})$ is a fringe subtree of $T_1((\hat{\xi}^n_r)^n_{r=1})$ rooted at a node that corresponds to an internal node of $T$ (regarding that $T \prec T_1((\hat{\xi}^n_r)^n_{r=1})$), or vice versa. Thus there exists a $T' \in \{T \sqcup T\}$ such that either $T' \prec T_1((\hat{\xi}^n_r)^n_{r=1})$ or $T' \prec T_1((\hat{\xi}^n_r)^n_{r=1})$. By symmetry of these two events, we have

$$\sum_{(i, \tilde{d}) \in B''} \mathbb{P}\{(\hat{\xi}^n_r)^n_{r=1} = \tilde{d}\} = 2 \sum_{T' \in \{T \sqcup T\}} \mathbb{P}\{T' \prec T_1(\hat{\xi}^n_1, \ldots, \hat{\xi}^n_n)\} = 2 \sum_{T' \in \{T \sqcup T\}} \pi^{n_f}(T') \frac{\mathbb{P}\{S_{n - \nu(T')} = n - \nu(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}},$$

where the last step follows from Lemma 6.2.

If $(i, \tilde{d}) \in B'$ and $(\hat{\xi}^n_r)^n_{r=1} = \tilde{d}$, we have three cases as shown by Figure 7:

(i) $T_1((\hat{\xi}^n_r)^n_{r=1})$ and $T_i((\hat{\xi}^n_r)^n_{r=1})$ do not overlap.

(ii) $T_i((\hat{\xi}^n_r)^n_{r=1})$ contains $T_i((\hat{\xi}^n_r)^n_{r=1})$, and the root of the latter does not correspond to an internal node of $T$ (regarding that $T \prec T_1((\hat{\xi}^n_r)^n_{r=1})$).

(iii) The reverse of case (ii).

Figure 7: Examples of three cases in $B''$, with $T_1 = T_1((\hat{\xi}^n_r)^n_{r=1})$ and $T_i = T_i((\hat{\xi}^n_r)^n_{r=1})$

We formally state the three cases as follows:

Recall that $\nu(T)$ and $\ell(T)$ are the numbers of internal nodes and leaves of $T$.

Recall that $T \prec T'$ 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) \(M_i(\vec{d}) \cap M_i(\vec{d}) = \emptyset\).

(ii) \(M_i(\vec{d}) \subseteq M_{i,s}^b(\vec{d})\) for some \(1 \leq s \leq \ell\).

(iii) \(M_i(\vec{d}) \subseteq M_{i,s}^b(\vec{d})\) 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 \((s, (e_1, \ldots, e_n)) \in \{v + 1, \ldots, n\} \times \mathbb{R}_n\) be a pair such that

\[(e_1, e_2, \ldots, e_v) = (\vec{a}_1, \ldots, \vec{a}_\ell), \quad \text{and} \quad (e_s, e_{s+1}, \ldots, e_{s+v-1}) = (\vec{a}_1, \ldots, \vec{a}_\ell)\].

Let \(\mathcal{B}^*\) be the set of such pairs. We have

\[
\sum_{(i, \vec{d}) \in \mathcal{B}'} \mathbb{P}\left\{ (\vec{c}_r)_{r=1}^n = \vec{d} \right\} = \sum_{(i, \vec{e}) \in \mathcal{B}^*} \mathbb{P}\left\{ (\vec{c}_r)_{r=1}^n = \vec{e} \right\}.
\]

Proof. Let \((i, \vec{d}) \in \mathcal{B}'\) with \(\vec{d} = (d_1, \ldots, d_n)\). Arrange \(d_1, \ldots, d_n\) in a cycle. Paint the segment \(\deg_1(\vec{d})\) red and the segment \(\deg_i(\vec{d})\) blue.

For case (i) and (ii) of Lemma 6.5, we first permute the red segment from \((\vec{a}_1, \vec{b}_1, \ldots, \vec{a}_\ell, \vec{b}_\ell)\) to \((\vec{a}_1, \ldots, \vec{a}_\ell, \vec{b}_1, \ldots, \vec{b}_\ell)\). Then we permute the blue segment of from \((\vec{a}_1, \vec{b}_1, \ldots, \vec{a}_\ell, \vec{b}_\ell)\) to \((\vec{a}_1, \ldots, \vec{a}_\ell, \vec{b}_1, \ldots, \vec{b}_\ell)\). 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 \((s, (e_1, \ldots, e_n)) \in \mathcal{B}'\). So this permutation defines a mapping \(f : \mathcal{B}' \to \mathcal{B}^*\).

Given \((s, \vec{e}) = (s, (e_1, \ldots, e_n)) \in \mathcal{B}\), by Lemma 6.1, we can find \(\ell\) consecutive tree degree sequences \(\vec{g}_1, \ldots, \vec{g}_\ell\) after \(e_1, \ldots, e_v\) and \(\ell\) consecutive tree degree sequences \(\vec{g}_1', \ldots, \vec{g}_\ell'\) after \(e_s, \ldots, e_{s+v-1}\). Also by Lemma 6.1,

\[
\deg_1(\vec{e}) = (e_1, \ldots, e_v, \vec{g}_1, \ldots, \vec{g}_\ell)
\]

and

\[
\deg_s(\vec{e}) = (e_s, \ldots, e_{s+v-1}, \vec{g}_1', \ldots, \vec{g}_\ell')
\]

are both tree degree sequences. If the two sequences do not overlap, then we are in case (i). If \(\deg_1(\vec{e})\) contains \(\deg_s(\vec{e})\), 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

\[
P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{e} \right\}.
\]

The Lemma follows by summing over all pairs in \(\mathcal{B}'\).

Lemma 6.7. Assume that \(P \{ |\mathcal{G}^9| = n \} > 0\) and \(\nu = \nu(T) \leq n/2\). We have

\[
\sum_{(i, \vec{d}) \in \mathcal{B}'} P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = (n - 2\nu + 1) \pi^n f(T)^2 \frac{P\{S_{n-2\nu} = n + 1 - 2(\nu + \ell)\}}{P\{S_n = n - 1\}}.
\]  

(6.6)

Proof. Given \((i, (d_1, \ldots, d_n)) \in \mathcal{B}^*\), we can move the segment \((d_1, \ldots, d_{i+\nu-1})\) to the position \(\nu + 1\) to get a new sequence \((d'_1, \ldots, d'_n) \in \mathcal{C}\), where

\[
\mathcal{C} := \{(e_1, \ldots, e_n) \in \mathcal{R}_n : (e_1, \ldots, e_{2\nu}) = (\vec{a}_1, \ldots, \vec{a}_\ell, \vec{a}_{\ell+1}, \ldots, \vec{a}_n)\}.
\]

Since there are \(n - 2\nu + 1\) possible values of \(i\), this permutation gives us a \((n - 2\nu + 1)\)-to-one mapping \(h : \mathcal{B}^* \to \mathcal{C}\). Since \((\tilde{\xi}_r^n)_{r=1}^n\) is permutation invariant, if \(\vec{d}' = h((i, \vec{d}))\), then

\[
P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d}' \right\}.
\]

Therefore

\[
\sum_{(i, \vec{d}) \in \mathcal{B}^*} P \left\{ (\tilde{\xi}_r^n)_{r=1}^n = \vec{d} \right\} = (n - 2\nu + 1) P \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\}.
\]

We obtain as usual

\[
P \left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\} = P \left\{ (\tilde{\xi}_1^n, \ldots, \tilde{\xi}_{2\nu}^n) = (\vec{a}_1, \ldots, \vec{a}_\ell, \vec{a}_{\ell+1}, \ldots, \vec{a}_n) \right\}
\]

\[
= P \left\{ (\xi_1, \ldots, \xi_{2\nu}) = (\vec{a}_1, \ldots, \vec{a}_\ell, \vec{a}_{\ell+1}, \ldots, \vec{a}_n) \mid S_n = n - 1 \right\}
\]

\[
= P \left\{ (\xi_1, \ldots, \xi_{2\nu}) = (\vec{a}_1, \ldots, \vec{a}_\ell, \vec{a}_{\ell+1}, \ldots, \vec{a}_n) \right\}
\]

\[
\times \frac{P \{S_{n-2\nu} = (n - 1) - 2(\nu + \ell - 1)\}}{P \{S_n = n - 1\}}
\]

\[
= \pi^n f(T)^2 \frac{P \{S_{n-2\nu} = n + 1 - 2(\nu + \ell)\}}{P \{S_n = n - 1\}}.
\]
It follows from Lemma 6.6 that
\[
\sum_{(i,d) \in \mathbb{B}'} \mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\} \\
= \sum_{(i,d) \in \mathbb{B}'} \mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\} \\
= (n - 2v + 1)\mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n \in \mathcal{C} \right\} \\
= (n - 2v + 1)\pi^{n}\mathbb{P}\left\{ S_{n-2v} = n + 1 - 2(v+\ell) \right\} \\
\mathbb{P}\left\{ S_n = n - 1 \right\} .
\]
\[
\square
\]
Putting everything together, we have:

**Lemma 6.8.** Assume that \( \mathbb{P}\{ |\mathcal{T}_R^a| = n \} > 0 \) and \( v = v(T) \leq n/2 \). We have
\[
\mathbb{E}(N_T^{n}(\mathcal{T}_R^a))_2 = n(n - 2v + 1)\pi^{n}\mathbb{P}\left\{ S_{n-2v} = n + 1 - 2(v+\ell) \right\} \\
\mathbb{P}\left\{ S_n = n - 1 \right\} \\
+ 2n \sum_{T' \in \{T \mid T\}} \pi^{n}(T')\mathbb{P}\left\{ S_{n-v(T')} = n - v(T') - \ell(T') \right\} \\
\mathbb{P}\left\{ S_n = n - 1 \right\} .
\]

**Proof.** The lemma follows by combining Lemma 6.3, 6.4, and 6.7:
\[
\mathbb{E}(N_T^{n}(\mathcal{T}_R^a))_2 = \sum_{(i,d) \in \mathbb{B}} \mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\} \\
= \sum_{(i,d) \in \mathbb{B}'} \mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\} + \sum_{(i,d) \in \mathbb{B}''} \mathbb{P}\left\{ (\tilde{\xi}_r^n)_{r=1}^n = \tilde{d} \right\} \\
= (n - 2v + 1)\pi^{n}\mathbb{P}\left\{ S_{n-2v} = n + 1 - 2(v+\ell) \right\} \\
\mathbb{P}\left\{ S_n = n - 1 \right\} \\
+ 2 \sum_{T' \in \{T \mid T\}} \pi^{n}(T')\mathbb{P}\left\{ S_{n-v(T')} = n - v(T') - \ell(T') \right\} \\
\mathbb{P}\left\{ S_n = n - 1 \right\} .
\]
\[
\square
\]

### 6.3 Proof of Theorem 1.4

Let \((T_n)_{n \geq 1}\) be a sequence of trees. Let \(v_n := v(T_n)\) and \(\ell_n := \ell(T_n)\). This section proves Theorem 1.4. We restate it here:

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

(i) If \(n\pi^{n}(T_n) \to 0\), then \(N_T^{n}(\mathcal{T}_R^a) = 0 \text{ whp.}\)

(ii) If \(n\pi^{n}(T_n) \to \infty\), then
\[
\frac{N_T^{n}(\mathcal{T}_R^a)}{n\pi^{n}(T_n)} \to 1.
\]
Part (i) of the Theorem follows from Lemma 6.9

**Lemma 6.9.** Assume Condition A. If $|T_n| = o(n)$ and $\pi^{nf}(T_n) > 0$ for all $n \in \mathbb{N}$, then

$$\frac{\mathbb{E}N_{T_n}^{nf}(T_n^{gw})}{n\pi^{nf}(T_n)} \to 1.$$  

Thus if $n\pi^{nf}(T_n) \to 0$, then $N_{T_n}^{nf}(T_n^{gw}) \xrightarrow{P} 0$.

**Proof.** $|T_n| = o(n)$ implies that $v_n = o(n)$ and $\ell_n = o(n)$. Therefore it follows from Corollary 2.2 and Lemma 6.2 that

$$\frac{\mathbb{E}N_{T_n}^{nf}(T_n^{gw})}{n\pi^{nf}(T_n)} = \frac{\mathbb{P}\{S_n - v_n = n - v_n - \ell_n\}}{\mathbb{P}\{S_n = n - 1\}} \to 1.$$  

**Lemma 6.10.** Assume Condition A. If $|T_n| = o(n)$ and $n\pi^{nf}(T_n) \to \infty$, then

$$\frac{\mathbb{E}(N_{T_n}^{nf}(T_n^{gw}))_2}{(n\pi^{nf}(T_n))^2} \to 1.$$  

**Proof.** Let $v = v_n$, $\ell = \ell_n$ and $T = T_n$. Since $|T_n| = v + \ell$, we have $v = o(n)$ and $\ell = o(n)$. If $T' \in \{\bar{T} \uplus T\}$, then $v(T') < 2v = o(n)$ and $\ell(T') < 2\ell = o(n)$. Therefore, it follows from Corollary 2.2 and Lemma 6.8 that

$$\mathbb{E}(N_{T_n}^{nf}(T_n^{gw}))_2 = n(n - 2v + 1)\pi^{nf}(T)\frac{\mathbb{P}\{S_n - 2v = n + 1 - 2(v + \ell)\}}{\mathbb{P}\{S_n = n - 1\}}$$

$$+ 2n \sum_{T' \in \{\bar{T} \uplus T\}} \pi^{nf}(T')\frac{\mathbb{P}\{S_n - v(T') = n - v(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}}$$

$$= (1 + o(1))(n\pi^{nf}(T))^2 + O(n) \sum_{T' \in \{\bar{T} \uplus T\}} \pi^{nf}(T').$$

Thus it suffices to show that

$$n \sum_{T' \in \{\bar{T} \uplus T\}} \pi^{nf}(T') = o(n\pi^{nf}(T))^2.$$  

Consider the superset $\mathcal{H}$ of $\{\bar{T} \uplus T\}$ that contains trees which can be obtained by replacing a proper non-leaf subtree of $T$ with another copy of $T$. (We do not restrict where this replacement can happen as in the definition of $\{\bar{T} \uplus T\}$.) Note that $|\mathcal{H}| = v - 1$, since $T$ has $v$ internal nodes and one of them is the root.

If $T' \in \mathcal{H}$, then $T'$ contains $T$ as a fringe subtree. Thus $\pi^{nf}(T') \leq \pi^{nf}(T)$. In the case that $v$ is bounded, we have

$$n \sum_{T' \in \mathcal{H}} \pi^{nf}(T') \leq n v \pi^{nf}(T) = O(n\pi^{nf}(T)) = o(n\pi^{nf}(T))^2.$$  

Thus we can assume that $v \to \infty$. 

For $T' \in \mathcal{H}$, if $T'$ has at least $3v/2$ internal nodes, call $T'$ \textit{big}, otherwise call it \textit{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 $\pi_{nf}(T') \leq \pi_{nf}(T)\frac{v}{2}$. Since $v \to \infty$ and $p_{max} < 1$, $v\frac{v}{2} = o(1)$. Using that $|\mathcal{H}| < v$, we have

$$n \sum_{T' \in \mathcal{H}_b} \pi_{nf}(T') \leq n v\pi_{nf}(T)\frac{v}{2} = o(n\pi_{nf}(T)).$$

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 $T_1'$ 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 $T_1'$, see Figure 8.

![Figure 8: Example of $T_1'$ for $T$ with 7 internal nodes](image)

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 $T_1'$ has at least $i$ internal nodes besides these in the copy of $T$ that replaced $T_{i,j}$, we have $\pi_{nf}(T_1') \leq \pi_{nf}(T)\frac{i}{v}$. In summary, we have

$$n \sum_{T' \in \mathcal{H}_s} \pi_{nf}(T') \leq n \sum_{i=1}^{v} \pi_{nf}(T_1') \leq n \sum_{i=1}^{v} \pi_{nf}(T)\frac{i}{v} \leq O(n\pi_{nf}(T)).$$

Therefore,

$$n \sum_{T' \in \mathcal{H}_s} \pi_{nf}(T') \leq n \sum_{T' \in \mathcal{H}_b} \pi_{nf}(T') + n \sum_{T' \in \mathcal{H}_s} \pi_{nf}(T') = o(n\pi_{nf}(T)) + O(n\pi_{nf}(T)) = o(n\pi_{nf}(T))^2.$$

Part (ii) of Theorem 1.4 is contained in the following lemma:

**Lemma 6.11.** Assume Condition A. If $|T_n| = o(n)$ and $n\pi_{nf}(T_n) \to \infty$, then

$$\frac{\text{Var} \left( N_{T_n}^{nf}(\mathcal{J}^w_{T_n}) \right)}{(\mathbb{E} N_{T_n}^{nf}(\mathcal{J}^w_{T_n}))^2} \to 0.$$

Therefore,

$$\frac{N_{T_n}^{nf}(\mathcal{J}^w_{T_n})}{n\pi_{nf}(T_n)} \to 1.$$
6.4 Complete r-Ary Non-Fringe Subtrees

Proof. Let \( X_n = N_n^r(T_n^g) \). By Lemma 6.9,

\[
\mathbb{E}X_n \sim n\pi^{n_f}(T_n) \to \infty.
\]

By Lemma 6.10, we have \( \mathbb{E}(X_n)_2 = (1 + o(1))(n\pi^{n_f}(T_n))^2 \). Therefore,

\[
\begin{align*}
\text{Var}(X_n) &= \mathbb{E}[(X_n)_2] + \mathbb{E}[X_n] - (\mathbb{E}[X_n])^2 \\
&= (1 + o(1))(n\pi^{n_f}(T_n))^2 \\
&\quad + (1 + o(1))(n\pi^{n_f}(T_n)) - (1 + o(1))(n\pi^{n_f}(T_n))^2 \\
&= o(n\pi^{n_f}(T_n))^2 + o(\mathbb{E}[X_n]^2).
\end{align*}
\]

Thus by Chebyshev’s inequality, for all \( \epsilon > 0 \)

\[
\mathbb{P}\left\{ \left| \frac{X_n}{\mathbb{E}X_n} - 1 \right| > \epsilon \right\} \leq \frac{\text{Var}(X_n)}{\epsilon^2\mathbb{E}[X_n]^2} \to 0.
\]

In other words, \( X_n/\mathbb{E}[X_n] \xrightarrow{p} 1. \) \( \Box \)

6.4 Complete r-Ary Non-Fringe Subtrees

Let \( T_{h}^{r, \text{ary}} \) denote the complete r-ary tree of height \( h \). Let \( \bar{H}_{n,r} \) be the maximal \( h \) such that \( T_{h}^{r, \text{ary}} \) appears as a non-fringe subtree in \( T_n^g \).

**Lemma 6.12.** Assume Condition \( A \) and \( p_r > 0 \) for some \( r \geq 2 \). Let \( h_n \to \infty \) be a sequence of positive integers. Let

\[
\alpha_r = \log_r \left( \frac{1}{r-1} \log \frac{1}{p_r} \right).
\]

Let \( \omega_n \to \infty \) be an arbitrary sequence. We have:

(i) If \( h_n \leq \log_r(\log n - \omega_n) - \alpha_r \), then whp \( T_n^g \) contains \( T_{h}^{r, \text{ary}} \) as non-fringe subtrees.

(ii) If \( h_n \geq \log_r(\log n + \omega_n) - \alpha_r \), then whp \( T_n^g \) does not contain \( T_{h}^{r, \text{ary}} \) 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 \( T_{h}^{r, \text{ary}} \) has

\[
\nu_n = \frac{r^h - 1}{r-1} = \frac{r^h}{r-1} + O(1)
\]
internal vertices, which all have degree $r$. Since $\pi^{nf}(T_h^{r-ary}) = p_r^n v_n$, we have

$$\log \frac{1}{\pi^{nf}(T_h^{r-ary})} = v_n \log \frac{1}{p_r}$$

$$= \frac{r^h}{r-1} \log \frac{1}{p_r} + O(1)$$

$$= r^h \alpha_r + O(1)$$

$$\leq \log \frac{n - \omega_n}{r^\alpha_r} + O(1)$$

$$= \log n - \omega_n + O(1),$$

where the inequality comes from the assumption of $h$. Thus

$$n\pi^{nf}(T_h^{r-ary}) \geq e^{\omega_n - O(1)} \to \infty.$$ 

It follows from Theorem 1.4 that $N_{T_h^{r-ary}}^{nf}(T_h^{qw}) \geq 1$ whp. Thus (i) is proved.

Similar computations show that for (ii) $n\pi^{nf}(T_h^{r-ary}) \to 0$, which implies that $N_{T_h^{r-ary}}^{nf}(T_h^{qw}) = 0$ whp by Theorem 1.4. The last statement of the lemma follows directly from (i) and (ii).

**Lemma 6.13.** Assume Condition A and $p_1 > 0$. Let $\omega_n \to \infty$ be an arbitrary sequence. We have:

(i) If $h_n \leq (\log n - \omega_n)/\log \frac{1}{p_r}$, then whp $T_h^{qw}$ contains $T_h^{1-ary}$ as non-fringe subtrees.

(ii) If $h_n \geq (\log n + \omega_n)/\log \frac{1}{p_1}$, then whp $T_h^{qw}$ does not contain $T_h^{1-ary}$ as non-fringe subtrees.

Therefore

$$\lim \frac{\tilde{H}_{n,1}}{\log_{1/p_1}(n)} \to 1.$$ 

**Proof.** For (i), we have

$$\pi^{nf}(T_h^{1-ary}) = p_1^n \geq e^{\omega_n}/n.$$ 

So $n\pi^{nf}(T_h^{1-ary}) \to \infty$. It follows from Theorem 1.4 that $N_{T_h^{1-ary}}^{nf} \geq 1$ whp. For (ii), we have $n\pi^{nf}(T_h^{1-ary}) \to 0$, so it follows from Theorem 1.4 that $N_{T_h^{1-ary}}^{nf} = 0$ whp.

**Example 6.1** (The binary tree). Recall that when $p_0 = p_2 = 1/4$ and $p_1 = 1/2$, $T_h^{qw}$ is equivalent to a uniform random binary tree of size $n$. It follows from Lemma 6.13 that $\tilde{H}_{n,1}/\log_2 n \to 1$. This result was previously proved by Devroye, Flajolet, Hurtado, Noy, and Steiger [33].
6.5 DIFFERENCE BETWEEN FRINGE AND NON-FRINGE SUBTREES

In Theorem 1.4, we have shown that if
\[ n\pi(T_n) \to \mu \in [0, \infty), \]
then
\[ d_{TV}(N_{T_n}(\mathcal{T}^w_n), \text{Po} (\mu)) \to 0. \]
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 \( X_n = N_{L_{h(n)}}^{\text{nfr}}(\mathcal{T}^w_n) \). If
\[ n\pi^{\text{nfr}}(L_{h(n)}) := n\mathbb{P}\{L_{h(n)} < \mathcal{T}^w_n\} \to \mu \in (0, \infty), \]
then
\[ \mathbb{E}X_n \to \mu, \quad \text{Var}(X_n) \to \mu \frac{1 + p_1}{1 - p_1}, \quad \mathbb{E}(X_n - \mathbb{E}X_n)^2 \to \mu \frac{3p_1^2 + 2p_1 + 1}{(1 - p_1)^2}. \]

As a result, \( \liminf_{n \to \infty} d_{TV}(X_n, \text{Po} (\mu)) > 0. \)

**Proof.** Let \( h = h(n) \). Since \( n\pi^{\text{nfr}}(L_h) = np_1^h \to \mu \in (0, \infty) \), 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
\[ \mathbb{E}X_n = n\pi^{\text{nfr}}(L_h) \frac{\mathbb{P}\{S_{n-h} = n - h - 1\}}{\mathbb{P}\{S_n = n - 1\}} \to \mu. \]

Since \( \{L_h \uplus L_h\} = \{L_{h+1} : 1 \leq i \leq h - 1\} \), by Lemma 2.7,
\[ \zeta_1 := 2n \sum_{T' \in \{L_h \uplus L_h\}} \pi^{\text{nfr}}(T') \frac{\mathbb{P}\{S_{n-v(T')} = n - v(T') - \ell(T')\}}{\mathbb{P}\{S_n = n - 1\}} \]
\[ = 2n \sum_{i=1}^{h-1} \pi^{\text{nfr}}(L_{h+i}) \frac{\mathbb{P}\{S_{n-h-i} = n - h - i - 1\}}{\mathbb{P}\{S_n = n - 1\}} \]
\[ = (1 + o(1))2n \sum_{i=1}^{h-1} p_1^{h+i} = (1 + o(1))2np_1^h \sum_{i=1}^{h-1} p_1^i \to 2\mu \frac{p_1}{1 - p_1}. \]

We also have by Lemma 2.7,
\[ \zeta_2 := n(n - 2h)\pi^{\text{nfr}}(L_h)^2 \frac{\mathbb{P}\{S_{n-2h} = n + 1 - 2(h + 1)\}}{\mathbb{P}\{S_n = n - 1\}} \to \mu^2. \]

Therefore, it follows from Lemma 6.8 that
\[ \mathbb{E}(X_n)_2 = \zeta_1 + \zeta_2 \to 2\mu \frac{p_1}{1 - p_1} + \mu^2. \]
Thus
\[
\text{Var}(X_n) = \mathbb{E}[(X_n)_2] + \mathbb{E}[X_n] - \mathbb{E}[X_n]^2 \to \mu \frac{1 + p_1}{1 - p_1}.
\]
So we have
\[
\frac{\text{Var}(X_n)}{\mathbb{E}[X_n]} \to \frac{1 + p_1}{1 - p_1} > 1.
\]
With an argument similar to Lemma 6.8, we can compute \(\mathbb{E}[(X_n)_3]\), which yields
\[
\mathbb{E}[(X_n - \mathbb{E}X_n)^3] \to \mu \frac{3p_1^3 + 2p_1 + 1}{(1 - p_1)^2}.
\]
Since
\[
\mathbb{E}[|X_n - \mathbb{E}X_n|^3] = 2\mathbb{E}[|X_n - \mathbb{E}X_n|^3 \times [X_n < \mathbb{E}X_n]] + \mathbb{E}[(X_n - \mathbb{E}X_n)^3]
\]
\[
\leq O(1) + \mathbb{E}[(X_n - \mathbb{E}X_n)^3],
\]
the above limit implies that
\[
C := \limsup_{n \to \infty} \mathbb{E}[|X_n - \mathbb{E}X_n|^3] < \infty.
\]
It follows from Lemma 3.8 that
\[
d_{TV}(X_n, \text{Po}(\mathbb{E}X_n)) = \left(\frac{\text{Var}(X_n) - \mathbb{E}X_n}{\mathbb{E}[|X_n - \mathbb{E}X_n|^3]^{2/3}}\right)^3.
\]
Therefore
\[
\liminf_{n \to \infty} d_{TV}(X_n, \text{Po}(\mathbb{E}X_n)) \geq \frac{1}{C^2} \left(\frac{1 + p_1}{1 - p_1}\right)^3 > 0.
\]
Since \(\mathbb{E}X_n \to \mu\), we also have \(\liminf_{n \to \infty} d_{TV}(X_n, \text{Po}(\mu)) > 0\).
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.
The main results proved in this thesis \((\text{Theorem } 1.2, 1.3 \text{ and } 1.4)\) 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: \(\text{Section } 7.1\) defines some basic concepts of graphs and digraphs. \(\text{Section } 7.2\) introduces some well-known random graph and digraph models. \(\text{Section } 7.3\) discusses the history of the \(k\)-out digraph model. \(\text{Section } 7.4\) presents our results on \(k\)-out digraphs. \(\text{Section } 7.5\) explains the connection between random \(k\)-out digraphs and Galton-Watson trees and how it motivated this thesis.

### 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 \((u_1, v_1), (u_2, v_2), \ldots, (u_\ell, v_\ell)\) is called a path from \(u_1\) to \(v_\ell\) 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 a 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
For an arc \( a = (u, v) \in A \), \( u \) 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 \( \{u_1, v_1\}, \{u_2, v_2\}, \ldots, \{u_t, v_t\} \) is called a directed path from \( u_1 \) to \( v_t \) if \( v_i = u_{i+1} \) for all \( 1 \leq i < t \). A digraph is strongly connected if there is a directed path from every vertex to all other vertices.

Let \( D = (V, A) \) be a digraph. Let \( V' \subseteq V \). The induced sub-digraph \( D[V'] \) is the digraph \( (V', A') \) where an arc \( a \) belongs to \( A' \) if and only if \( a \in A \) and both of its starting and end vertices are in \( V' \). \( D[V'] \) is called a strongly connected component (scc) of \( D \) if it is a strongly connected and adding any vertex to \( V' \) 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, \(\Sigma\) is a finite set called the alphabet, \(\delta : Q \times \Sigma \to Q\) 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 \(\{1, \ldots, n\}\) in which each vertex has \(k\) out-arcs labeled by \(1, \ldots, k\), 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](image)

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 \(\{1, \ldots, n\}\) 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 \( \nu_k \) and \( \sigma_k \) are explicitly defined constants.

Given a set of vertices \( S \subseteq \{1, \ldots, n\} \), call \( S \) closed if there does not exist an arc with starting vertex in \( S \) and end vertex in \( S^c := \{1, \ldots, n\}\setminus S \). Let \( \mathcal{S}_n \) be the largest closed set of vertices such that \( D_{n,k}[S] \) 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 \( S_1 = \{1, \ldots, n\} \) 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.)

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].
### 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 $G^c_n := \{1, \ldots, n\} \setminus G_n$, i.e., $G^c_n$ is the set of vertices that are outside the giant. The structure of $D_{n,k}^A$ depends on $D_{n,k}[G^c_n]$, the digraph induced by $G^c_n$. Our analysis shows that in $D_{n,k}[G^c_n]$ 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 $G^c_n$ 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 $O_n$ denote the set of vertices in the one-in-core of $D_{n,k}$.

Note that $G_n \subseteq O_n$ since a SCC induces a sub-digraph with each vertex having in-degree at least one. Also note that cycles cannot exist outside $O_n$, for otherwise they contradict the maximality of $O_n$. Now assume that every vertex can reach $G_n$, which happens whp by Grusho [54]. Then $D_{n,k}$ can be divided into three layers: the center is $G_n$; then comes $O_n \setminus G_n$, which consists of cycles outside $G_n$ and paths from these cycles to $G_n$; the outermost is $O^c_n := \{1, \ldots, n\} \setminus O_n$, which is acyclic.

![Diagram of three layers](image)

Figure 10: Three layers of $D_{n,k}$: the giant $G_n$; the one-in-core $O_n$; and the whole graph.
Since there cannot be many vertices in cycles outside the giant, the middle layer \( \mathcal{G}_n \) 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 \( S \) with \( |S| = s \) such that:

(i) \( D_{n,k}[S] \), the sub-digraph induced by \( S \), has minimum in-degree one (surjective) and there are no arcs going from \( S \) to \( S^c \) (closed), which we refer to as \( S \) being a \( k \)-surjection (since \( D_{n,k}[S] \) is equivalent to a surjective function from \( \{1, \ldots, ks\} \) to \( \{1, \ldots, s\} \));

(ii) \( D_{n,k}[S^c] \) 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 \( S \), (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,

\[
\mathbb{P} \{ |\mathcal{O}_n| = s \} = \sum_{S \subseteq \{1, \ldots, n\} : |S| = s} \mathbb{P} \{ \mathcal{O}_n = S \}.
\]

Thus we can compute the characteristic function of \( (|\mathcal{O}_n| - \nu_k n)/\sqrt{n} \) and prove the following theorem:

**Theorem 7.1** (Cai and Devroye [21, thm. 1]). We have as \( n \to \infty \),

\[
\frac{|\mathcal{O}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),
\]

\[
\frac{|\mathcal{G}_n| - \nu_k n}{\sigma_k \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),
\]

\[
\max_{u \in \{1, \ldots, n\}} |\mathcal{O}_u| - \nu_k n \xrightarrow{d} \mathcal{N}(0, 1),
\]

where \( \nu_k \) and \( \sigma_k \) are constants defined by

\[
\nu_k := \frac{\tau_k}{k},
\]

\[
\sigma_k^2 := \frac{\tau_k}{ke^{\tau_k}(1 - ke^{-\tau_k})},
\]

and \( \tau_k \) 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 \( \mathbb{P} \{ |S_1| = s \} \), 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 \(|\Omega_n| - |\mathcal{G}_n| = O_p(1)\) \cite[lem. 9]{21}. 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 \cite[thm. 1.3]{84} 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 \cite{82} and Pittel \cite{83}).

### 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 \cite[thm. 2]{21}). We have:

1. Let \(L_n\) be the length of the longest cycle in \(D_{n,k}[\mathcal{G}_n^c]\). Then \(L_n = O_p(1)\).

2. Let \(C_n\) be the number of cycles in \(D_{n,k}[\mathcal{G}_n^c]\). Then
   \[
   C_n \xrightarrow{d} \text{Po} \left( \log \frac{1}{1 - ke^{-\tau_k}} \right).
   \]

3. Let \(C_{n,\ell}\) be the number of cycles of length \(\ell\) 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 \cite[thm. 3]{21}). 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

1. \(\mathbb{P} \left\{ \bigcup_{x \in S'_x} \text{arc}(D_{n,k}[\mathcal{G}_n^c]) - |S'_x| \geq 1 \right\} = o(1)\), 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.

2. Let \(Y_n := \max_{x \in S'_x} |S'_x|\). Let \(\lambda_k := (k - \tau_k) \left( \frac{\tau_k}{k - 1} \right)^{k-1}\). Then
   \[
   \frac{Y_n}{\log n} \xrightarrow{p} \frac{1}{\log(1/\lambda_k)}.
   \]

3. 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
   \[
   W_n := \max_{x \in S'} \min_{y \in S} \text{dist}(x, y),
   \]
i.e., $W_n$ is the maximum distance from a vertex in $G^c_n$ to $G_n$. 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 $D_{n,k}[G^c_n]$. Then

$$\frac{M_n}{\log n} \xrightarrow{p} \frac{1}{\log(e^{\tau_k}/k)}.$$ 

(v) Let $X_n := \max_{x \in G_n^c} \max_{y \in S_k} \text{dist}(x,y)$. 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:

1. Let $Q$ be an empty queue. Mark all vertices unvisited.
2. Put an arbitrary vertex $w$ in $Q$.
3. Remove a vertex $u$ at the head of $Q$. Mark it as visited.
4. Add at the end of $Q$ the starting vertices of the in-arcs of $u$, except those vertices that have already been visited.
5. If $Q$ is empty, terminate. Otherwise go to step 3.

In this way, every vertex that can reach vertex $w$ is visited in the end.

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 $\nu_k$. This is why $|G_n| \approx \nu_k n$.

Remark 7.3. It is well-known that for $k > 1$, $|C_{\text{max}}^n|$—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, $(|C_{\text{max}}^n| - \nu_k n)/\sigma_k \sqrt{n}$ 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 $\text{Po}(k)$. 

**Compare with the DFS procedure for Galton-Watson trees described in Subsection 2.2.1.**

**Bi(d, p) denotes the binomial (d, p) distribution.**

**Recall that $G_n$ is set of vertices in the largest closed SCC in $D_{n,k}$.**
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:

1. Let \( Q \) 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 \( Q \).
3. Remove a vertex \( u \) at the head of \( Q \). Mark it as visited.
4. Add at the end of \( Q \) 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 \( Q \) 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 \( \nu_k \) 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 \( Q \) if and only if: (i) one or more of the \( k \) arcs staring 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 \( T_n^{gw} \), fringe subtrees behave like independent copies of unconditional Galton-Watson trees. Therefore, it is natural to guess that \( H_{n,k} \), the height of the maximal complete \( k \)-ary non-fringe subtree of \( T_n^{gw} \), should also be \( (1 + o_p(1)) \log_k \log n \). The intuition turns out to be correct, recall that \( T_n^{gw} \) 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 $T_n^{gw}$ 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
\[ \frac{\pi(S_n)}{\pi(T_{k_n})} \to 0, \tag{8.1} \]
then
\[ d_{TV}(N_{\pi T_{k_n}}(T_{R^w}), \text{Po}(n\pi(S_n))) = o(1). \]
The condition (8.1) may not be necessary. For example, if \( S_n = T_{k_n} \), then \( \pi(S_n)/\pi(T_{k_n}) = 1 \) and the above result does not apply. However, by Lemma 4.6 and 4.7
\[ \frac{\text{Var}(N_{T_{k_n}}(T_{R^w}))}{\mathbb{E}N_{T_{k_n}}(T_{R^w})} \sim \frac{n\pi(T_{k_n})}{n\pi(S_n)} = 1. \]
In other words, we cannot use Lemma 3.7 or 3.8 to prove
\[ \lim \inf d_{TV}(N_{\pi T_{k_n}}(T_{R^w}), \text{Po}(n\pi(T_{k_n}))) > 0. \]
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
\[ n\pi^{n}f(T_n) := n\mathbb{P}\{T_n < T_{R^w}\} \to \infty, \]
then
\[ \frac{N_{T_{n}}^{n}f(\mathcal{T}_{R^w})}{n\pi^{n}f(T_n)} \sim 1. \]
We believe that under the same assumption it is also possible to prove
\[ \frac{N_{T_{n}}^{n}f(\mathcal{T}_{R^w}) - n\pi^{n}f(T_n)}{\sqrt{n\pi^{n}f(T_n)}} \to N(0, 1). \]
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 $\mathcal{S}_{k_n}$ 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 $\mathcal{S}_{k_n}$ 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 $T_{gw}^{qw}$ 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 $T_{gw}^{qw}$. This can be seen as a generalization of the height of $T_{gw}^{qw}$, since that $T_{gw}^{qw}$ has height $h$ is equivalent to that the maximal embedded one-ary subtree of $T_{gw}^{qw}$ has height $h$. 

for the width and height of conditioned Galton-Watson trees,” The Annals


tic Analysis (Durham, 1990), ser. London Mathematical Society Lecture
pp. 23–70.


mappings,” Random Structures and Algorithms, vol. 5, no. 4, pp. 487–512,
1994.

Watson processes,” Annales de l’Institut Henri Poincaré Probabilités et Statis-

2008.

2015.

approximations: The Chen-Stein method,” The Annals of Probability, vol. 17,


[82] X. Pérez-Giménez and N. Wormald, “Asymptotic enumeration of strongly
connected digraphs by vertices and edges,” Random Structures and Algo-

[83] B. Pittel, “Counting strongly-connected, moderately sparse directed graphs,”

[84] B. Pittel and D. Poole, “Asymptotic distribution of the numbers of ver-
tices and arcs of the giant strong component in sparse random digraphs,”


[86] B. Roos, “Improvements in the Poisson approximation of mixed Poisson
distributions,” Journal of Statistical Planning and Inference, vol. 113, no. 2,


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.


[91] C. Stein, “A bound for the error in the normal approximation to the dis-
btribution 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,” Jour-
nal of Applied Mathematics and Stochastic Analysis, vol. 4, no. 1, pp. 1–27,