Diameter and Rumour Spreading in Real-World Network Models

by

Abbas Mehrabian

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The so-called 'small-world phenomenon', observed in many real-world networks, is that there is a short path between any two nodes of a network, whose length is much smaller that the network's size, typically growing as a logarithmic function. Several mathematical models have been defined for social networks, the WWW, etc., and this phenomenon translates to proving that such models have a small diameter. In the first part of this thesis, we rigorously analyze the diameters of several random graph classes that are introduced specifically to model complex networks, verifying whether this phenomenon occurs in them.

In Chapter 3 we develop a versatile technique for proving upper bounds for diameters of evolving random graph models, which is based on defining a coupling between these models and variants of random recursive trees. Using this technique we prove, for the first time, logarithmic upper bounds for the diameters of seven well known models. This technique gives unified simple proofs for known results, provides lots of new ones, and will help in proving many of the forthcoming network models are small-world. Perhaps, for any given model, one can come up with an ad hoc argument that the diameter is $O(\log n)$, but it is interesting that a unified technique works for such a wide variety of models, and our first major contribution is introducing such a technique.

In Chapter 4 we estimate the diameter of random Apollonian networks, a class of random planar graphs. We also give lower and upper bounds for the length of their longest paths. In Chapter 5 we study the diameter of another random graph model, called the random surfer Web-graph model. We find logarithmic upper bounds for the diameter, which are almost tight in the special case when the growing graph is a tree. Although the two models are quite different, surprisingly the same engine is used for proving these results, namely the powerful technique of Broutin and Devroye (Large deviations for the weighted height of an extended class of trees, Algorithmica 2006) for analyzing weighted heights of random trees, which we have adapted and applied to the two random graph models. Our second major contribution is demonstrating the flexibility of this technique via providing two significant applications.

In the second part of the thesis, we study rumour spreading in networks. Suppose that initially a node has a piece of information and wants to spread it to all nodes in a network quickly. The problem of designing an efficient protocol performing this task is a fundamental one in distributed computing and has applications in maintenance of replicated databases, broadcasting algorithms, analyzing news propagation is social networks and the spread of viruses on the Internet. Given a rumour spreading protocol, its *spread time* is the time it takes for the rumour to spread in the whole graph.

In Chapter 6 we prove several tight lower and upper bounds for the spread times of two well known randomized rumour spreading protocols, namely the synchronous push&pull protocol and the asynchronous push&pull protocol. In particular, we show the average spread time in both protocols is always at most linear. In Chapter 7 we study the performance of the synchronous push&pull protocol on random k-trees. We show that a.a.s. after a polylogarithmic amount of time, 99 percent of the nodes are informed, but to inform all vertices, a polynomial amount of time is required. Our third majoc contribution is giving analytical proofs for two experimentally verified statements: firstly, the asynchronous push&pull protocol is typically faster than its synchronous variant, and secondly, it takes considerably more time to inform the last 1 percent of the vertices in a social network than the first 99 percent. We hope that our work on the asynchronous push&pull protocol attracts attention to this fascinating model.

Acknowledgements

Firstly I want to thank my supervisor Nick Wormald. It was a great experience to work with Nick. I learnt a lot of things about mathematics and research from him, but most importantly he has been so inspiring for me. After each meeting with him, I found that my energy and desire to proof new results has been doubled. I also want to thank my co-supervisor Joseph Cheriyan, for his support and many helpful advices.

I thank those with whom I collaborated during my PhD, including Hüseyin Acan, Soroush Alamdari, Noga Alon, Andrea Collevecchio, Ehsan Ebrahimzadeh, Linda Farczadi, Jane Gao, Jason Gao, Jeannette Jansen, Marek Lipczak, Dieter Mitsche, Ali Pourmiri, Paweł Prałat, Cristiane Sato, Nick Wormald, and Jonathan Zung.

I thank Alan Frieze, Jim Geelen, Penny Haxell and Srinivasan Keshav for reading a preliminary version of this thesis and their many useful comments.

I thank the C&O department and its chair, Alfred Menezes, for supporting my visits to Monash University in Clayton, and the Monash Mathematics people for their hospitality. My PhD has also been supported by an Ontario Graduate Scholarship and a Vanier Canada Graduate Scholarship.

I took many wonderful courses in my post-graduate studies. I thank all my instructors, and in particular Kasra Alishahi (for stochastic processes), Nikhil Bansal (for discrepancy theory short course), Jim Geelen (for graph theory), Nick Harvey (for randomized algorithms), Anna Lubiw (for graph algorithms), and Peter Winkler (for random walks short course). I thank Costis Georgiou for organizing the C&O algorithms seminars, which were very useful for improving my lecturing skills.

During the PhD, I lived in four different cities, where I found heaps of good friends, whose fellowships have been uplifting and has made my life vivid. I thank all my friends in Waterloo, Clayton, Tehran and Toronto, and in particular, Hüseyin Acan, Sara Ahmadian, Soroush Alamdari, Kasra Alishahi, Yasaman Amin, Farzad Anooshah, Omid Ardakanian, Maryam Ashoori, Safoora Bahrami, Sayeh Bahrololoumi, Keivan Bakhoda, Mohammad Barati, Eduard Biche, Andrea Collevecchio, Hesam Dadafarin, Bahar Davoudi, Ali Dehghan, Morteza Dehghani, Mohammad Derakhshani, Ehsan Ebrahimzadeh, Sepehr Eghbali, Marzieh Fadaee, Niloufar Faghihi, Amir Faraji, Zahra Farajidana, Roozbeh Farhoudi, Jane Gao, Mojtaba Gharibi, Arash Haddadan, Sajjad Hosseini, Zia Jalali, Sahar Karimi, Shima Khoshraftar, Nishad Kothari, Nazanin Mohammadi, Ebrahim Moradi, Shaghayegh Moshtarikhah, Hamoon Mousavi, Nima Mousavi, Ali Nasiri, Maryam Nasirpour, Mohammad Norouzi, Xavier Pérez, Siavash Rahbar, Reza Razavian, Erfan Salavati, Cristiane Sato,

Yasaman Sefidgar, Ali Sharifi, Yousef Sohrabi, Vahid Taslimi, Kamran Tirdad, Mahshid Yassaei, Aynaz Vatankhah, Farshid Zoghalchi, Hassan Zokaei, and Hossein Zokaei.

I also visited several cities, where I enjoyed the hospitality of many people. The list includes Amsterdam (thanks to Marzieh Fadaee and Hamid Ghader), Bolzano (thanks to Andrea Collevecchio and Fabiola Del Greco), Calgary and Banff (thanks to Samaneh Mousavi and Mohammad Khazali), Halifax (thanks to Hoda Chuang, Behrouz Hajisoleimani, Habibeh Naderi, Somayyeh Nasr and Elham Roshanbin), Lausanne (thanks to Pedram Pad), Los Angeles, where I drew most of the figures in this thesis (thanks to Ehsan Ebrahimzadeh), Montreal (thanks to Alireza Mehrabian), Paris (thanks to Ehsan Ebrahimzadeh), Singapore (thanks to Fatima Mohsin and her family), San Jose (thanks to Saeid Farivar and Masoud Farivar), and Vancouver (thanks to Ebrahim Beyrami and Arian Khosravi).

I thank Hania Wormald and Nick Wormald for the parties in Waterloo and Clayton.

I thank Ali Ghodsi and Mohammad Ghaziaskar for the *masnavi-khaani* (Rumi reading) sessions in Waterloo, which have been full of wisdom.

I thank Tohid Ahadifar for his thoughtful life advices.

I thank Ebrahim Moradi and Hamoon Mousavi for being my gym buddies.

I thank the Monash University Outdoors Club for introducing me to the Australian nature, and Ebrahim Moradi, Nima Mousavi, and Farshid Zoghalchi for the two fantastic camping trips in Canada.

I am lucky to have a wonderful family. I thank my lovely grandmother, my caring aunts and uncles, and most importantly, my mother, my father, and my brother, for their love and support.

Finally I thank Josep Diaz for inviting me to the Centre de Recerca Matemàtica in Barcelona, where I am writing these last words.

May God bless us!
A. M.
April 2015

Table of Contents

List of Figures

Chapter 1

Introduction

A point to point communication network can be modelled as an undirected graph: the nodes represent the processors and the links represent communication channels between them. Consider the 'broadcasting problem': a single processor wants to broadcast a piece of information to all other processors in the network. A natural solution is the 'flooding' algorithm: each process sends the piece of information to all of its neighbours. The number of rounds it takes for all processors to learn the information corresponds to the *diameter* of the network graph. In the first part of this thesis (Chapters 3 to 5) we study this parameter in several real-world network models. More details can be found in Section 1.1.

A caveat of the flooding algorithm is that it puts a heavy load on the links. To reduce this load, suppose we enforce the constraint that each processor can send messages to at most one of its neighbours in each round. How quickly can we spread the information in such a setting? In the second part (Chapters 6 and 7) we address this question for two randomized rumour spreading protocols. More details can be found in Section 1.2. Preliminaries appear in Chapter 2, and Chapter 8 articulates the main contributions of the thesis and also contains further research directions. Parts of this thesis are based on joint work with various people and appear in papers, as elaborated in Section 1.3.

1.1 The small-world phenomenon

'Small-world phenomenon' refers to a striking pattern observed in many real-world graphs: most pairs of vertices are connected by a path whose length is considerably smaller than the size of the graph. Travers and Milgram [116] in 1969 conducted an experiment in which

participants were asked to reach a target person by sending a chain letter. The average length of all completed chains was found to be 6.2, an amazingly small number, hence the phrase 'six degrees of separation.' The Webgraph is a directed graph whose vertices are the static web pages,[1] and there is an edge joining two vertices if there is a hyperlink in the first page pointing to the second page. Broder, Kumar, Maghoul, Raghavan, Rajagopalan, Stata, Tomkins, and Wiener [26] in 1999 crawled about 200 million web pages and found that the expected shortest-directed-path distance between two random web pages (when a path exists at all) is 16.18; this figure is 6.83 in the corresponding underlying undirected graph.

Backstrom, Boldi, Rosa, Ugander, and Vigna [8] studied the Facebook graph in May 2011, which had about 721 million vertices. The vertices of this graph are people, and two of them are joined by an edge if they are friends on Facebook. The diameter of the giant component of this graph was found to be 41, and the average distance between reachable pairs was found to be around 4.74. For other examples, see, e.g., Tables 1 and 2 in [4], Table 8.1 in [105] or Table 4 in [95].

Due to the ever growing interest in social networks, the Webgraph, biological networks, etc., in recent years a great deal of research has been built around studying mathematical properties of real world networks (see, e.g., the books [22, 30, 35, 56]). Another fascinating observation on many real-world graphs is that their degree sequences are heavy-tailed and almost obey a *power law*. Namely, for each positive integer k, the fraction of vertices having degree k is almost proportional to $k^{-\beta}$ for some fixed exponent β. As Erdős-Rényi random graphs do not satisfy this property, scholars have defined lots of models recently, aiming at capturing the aforementioned and other properties of real-world graphs (see, e.g., Bonato and Chung [23] or Chakrabarti and Faloutsos [30, Part II] and the references therein). Lots of mathematical models have been defined so far, yet very few rigorously analyzed.

The *diameter* of an undirected graph is the *maximum* shortest-path distance between any two vertices. It is a well known metric quantifying how 'small-world' the graph is; informally speaking, it measures how quickly one can get from one 'end' of the graph to the other. The diameter is related to various processes, e.g. it is within a constant factor of the memory complexity of the depth-first search algorithm. Also, it is a natural lower bound for the mixing time of any random walk ([93, Section 7.1.2]) and the broadcast time of the graph ([76, Section 3]). Another well studied metric is the *average diameter* of a graph, which is the *expected value* of the shortest-path distance between two random vertices. Despite the fact that these are two of the most studied parameters of a network,

[1] A static web page is one that is delivered to the user exactly as stored, in contrast to dynamic web pages which are generated by a web application.

for several models introduced in the literature, the degree sequence is proved to be power-law, but no sublinear upper bound for the diameter or average diameter is known.

In the first part of this thesis, we rigorously analyze the properties of some random graph classes that are introduced specifically to model real-world networks. In particular we study their diameters to verify whether the small-world phenomenon occurs in them. In mathematical terms, this translates into verifying whether these models have small diameter. Roughly speaking, we say a graph is 'small-world' if its diameter has a logarithmic growth. Note that since diameter is an upper bound for the average diameter, all upper bounds we prove for diameter automatically carry over to the average diameter.

In some articles, e.g. [23], when it is said that a graph is small-world, it is also implied that it has a large clustering coefficient, namely that if two vertices have a common neighbour, then they are more likely to be adjacent. In this thesis however, we are not concerned about the clustering coefficients of graphs. Also, we do not discuss the algorithmic aspects of the small-world phenomenon, which are discussed by Kleinberg [85].

All models we consider are probabilistic: due to the complexity of real-world graphs, probabilistic modelling seems inevitable. Consequently, in the analysis we mostly use probabilistic tools (Chapter 2 reviews the required probabilistic preliminaries). Our results hold *asymptotically almost surely (a.a.s.)*, meaning that the probability that they are true approaches 1 as the number of vertices goes to infinity. We study *evolving models* (also called *on-line* or *dynamic* models) only, i.e. the graph changes over time according to pre-defined probabilistic rules, and we are interested in the long-term structure of this evolving graph. Evolving models seem more appropriate in applications in which the network naturally evolves in time, e.g. the Webgraph. All models studied in this thesis are growing models, i.e. in discrete time-steps new vertices and edges appear in the graph, but no deletion occurs.

Some of the models we study generate undirected graphs while others generate directed ones. For directed graphs, since there is no guarantee that there is always a directed path between two given vertices, there is a question of how to define the diameter. We take the approach of [92] and ignore the edge directions when calculating the diameter. In other words, we define the diameter of directed graph as that of its underlying undirected graph.

In this thesis we work with (weakly) connected graphs only, so the diameter and average diameter are well defined. In citing previous work, when we talk about the diameter/average diameter of a disconnected graph, we mean the maximum/average distance between any two vertices in the same connected component.

1.1.1 Previous work

Surprisingly few results are known about the diameters of evolving random graph models, see, e.g., Table 8.2 from the recent monograph [30], or [29, Table III], or the table in [21, p. 162]: each cited table contains a summary of known results on the diameter and other properties of several real-world network models. Chung and Lu [34] defined an evolving (online) and a non-evolving (offline) model. They state that 'The online model is obviously much harder to analyze than the offline model', and hence analyze the former by coupling it with the latter, which had been analyzed before. The difficulty of analyzing evolving models over non-evolving ones arises perhaps from the dependencies between edges in the former models.

Barabási and Albert [9] in 1999 introduced one of the first evolving models for real-world networks, sometimes known as the *preferential attachment model*. Their model can be roughly described as follows (see [20] for a formal definition). We start with a fixed small graph, and in each time-step a new vertex appears and is joined to a fixed number of old vertices, where the probability of joining to each old vertex is proportional to its *degree*. The term 'preferential attachment' here refers to the property that a new vertex prefers attaching to high-degree vertices rather than low-degree ones. This preferential attachment naturally leads to a power-law degree sequence and was perhaps the first explanation for the power-law properties of real-world networks. Their work inspired a lot of research and many models were introduced based on this scheme. A generalization of the Barabási-Albert model is the *linear preference model*, in which the probability of joining the new vertex to a given old vertex is proportional to a linear function of its degree. This evolving model has attracted the most attention; a logarithmic upper bound has been proved for its diameter, and sharper results are known in various special cases, see Remark 3.20 for details. When the new vertex is joined to exactly one vertex in the existing graph (so the resulting evolving graph is always a tree), a general technique based on branching processes is developed by Bhamidi [13], using which he proved the diameter of a variety of preferential attachment trees is a.a.s $\Theta(\log n)$. (In this chapter n always denotes the number of vertices in the graph.)

An important non-evolving model that has been studied extensively is the random graph model with given expected degrees. (A more general version was studied by Bollobás, Janson and Riordan [18] under the name 'inhomogenous random graphs.') Given positive numbers $w_1, w_2, \ldots, w_n$ satisfying $\sum w_k \geq \max\{w_k^2\}$, we build a random graph $G = G(w)$ as follows. The vertices are numbered 1 to n, and each pair $\{i, j\}$ is joined by an edge independently of other pairs and with probability $w_i w_j / \sum w_k$. So, the expected degree of vertex i is w_i; that is, the sequence (w_k) determines the graph's expected degrees. We

can now sample the w_k's according to a power law distribution with exponent β and study the resulting 'power law random graph.' We call this model the *Chung-Lu model* with exponent β. Let $y = \sum w_k^2 / \sum w_k$. In 2001, Chung and Lu [33] introduced this model and proved that for $\beta > 3$, a.a.s. the average diameter is $\Theta(\log n / \log y)$ and the diameter is $\Theta(\log n)$, whereas for $2 < \beta < 3$, a.a.s. the average diameter is $O(\log \log n)$ and the diameter is $\Theta(\log n)$.

Other evolving models whose diameters have been studied in the literature include the Fabrikant-Koutsoupias-Papadimitriou model [11], protean graphs [112], the geometric preferential attachment model [66, 94], and the spatial preferred attachment model [40]. See [18, Section 14] and [117] for collections of results on diameters of non-evolving models.

1.1.2 Chapter 3

There are dozens of research papers in which models for complex networks have been defined and it is shown that these models satisfy the properties of real-world networks, using either experimental analysis or so-called mean-field theory (many fewer works provide rigorous mathematical proofs for their claims). A typical paper focuses on one model or a family of models with some parameter. In Chapter 3 we develop a versatile technique for establishing a.a.s. upper bounds for diameters of a wide range of random graph models. We demonstrate the technique by rigorously proving logarithmic upper bounds for the diameters of several well known models, which are perhaps the most-cited ones among those for which no nontrivial upper bound is known for the diameter.

A somewhat general undirected model is the Cooper-Frieze model, which evolves as follows: in each step, either a new vertex is born and edges are added from it to the existing graph, or edges are added between the existing vertices. The number of added edges is a bounded random variable. One endpoint of each added edge is either the new vertex, or a uniformly random vertex, or a vertex sampled according to the degrees. The other endpoint is either a uniformly random vertex or a vertex sampled according to the degrees. This model contains many other undirected ones as subcases. In Section 3.4 we prove that a.a.s. the Cooper-Frieze model has logarithmic diameter.

Our technique is based on defining a coupling between random graphs and variants of random recursive trees. An important new idea is to use a well known fact to bound the diameter: this fact is that to sample vertices proportional to their degrees, we can choose a random endpoint of a random edge. Many of the models introduced in the literature use this kind of sampling, and we employ this fact to relate them to (variants of) random

recursive trees. A more detailed yet informal description appears in the beginning of Chapter 3.

Our approach gives logarithmic upper bounds that are perhaps not tight, but on the positive side, it is applicable to a broad variety of models, including those incorporating preferential attachment. Another advantage of our technique is simplicity: all proofs in this chapter are elementary and fairly short, and the only probabilistic tools needed are couplings and Chernoff bounds for concentration of binomial random variables. The third advantage of our technique is that the constant factor it gives (hidden in the $O(\log n)$ notion) is typically small: for all the models studied here, the constant is at most 20.

From a wider perspective, it would be appealing to have a mathematical theory for characterizing those evolving random graphs which have logarithmic diameters. Chapter 3 is a fundamental step in building this theory. The technique developed there gives unified simple proofs for known results, provides lots of new ones, and will help in proving many of the forthcoming network models are small-world. We hope this theory will be developed further to cover other network models, e.g. spatial models [80], as well.

The results of this chapter shed light on why the small-world phenomenon is observed in so many real-world graphs. At their core, our arguments are based on the fact that in all considered models, there is a sort of 'rough uniformity' for the (random) destination of each new link. Thus, we may expect that for any growing network in which the endpoints of new links are chosen according to a probability distribution that is 'not too biased,' i.e. does not greatly favour some vertices over others, the diameter grows at most logarithmically. We believe this is the primary reason that most real-world graphs are small-world.

1.1.3 Chapter 4

Despite the great amount of work on models generating graphs with power law degree sequences, a considerably smaller amount of work has focused on generative models for *planar* graphs. In Chapter 4 we study a popular random graph model for generating planar graphs with power law degree sequences, which is defined as follows. Start with a triangle embedded in the plane. In each step, choose a bounded face uniformly at random, add a vertex inside that face and join it to the vertices on the face. After $n-3$ steps, we obtain a (random) triangulated plane graph with n vertices, which is called a *Random Apollonian Network (RAN)*. See Figure 1.1 for an illustration.

In [3] it was shown that a.a.s. the average diameter of a RAN is asymptotic to $\eta_1 \log n$, where $\eta_1 = 6/11 \approx 0.545$. Frieze and Tsourakakis [70] showed that the diameter of a RAN is a.a.s. at most $\eta_2 \log n$, where $\eta_2 \approx 7.081$ is the unique solution greater than 1

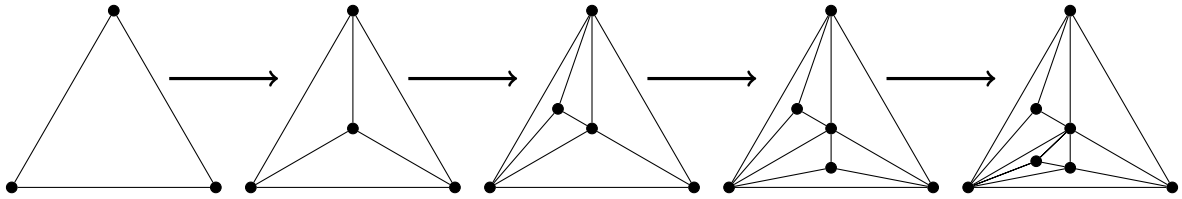


Figure 1.1: the construction of an instance of a RAN with seven vertices

of $\exp(1/x) = 3e/x$, and asked: what is the diameter of a typical RAN? We prove that a.a.s. the diameter of a RAN is asymptotic to $c \log n$, where $c \approx 1.668$ is the solution of an explicit equation.

This result has an interesting story: it was proved independently and simultaneously by three groups who presented it obliviously at the same conference (Random Structures and Algorithms, Poland, August 2013)! The three proofs are quite different. Our proof has appeared in [57] (see [41, Theorem 2] and [87, Theorem 2.2] for the other proofs).

We also study the length of longest simple paths in RANs. Let $\mathcal{L}_m$ be a random variable denoting the number of vertices in a longest path in a RAN with m faces (and $(m+5)/2$ vertices). In [57] we showed there exists a fixed $\delta > 0$ such that $\mathbb{P}\left[\mathcal{L}_m < m/(\log m)^{\delta}\right] \to 1$. Recently, Cooper and Frieze [39] improved this by showing that for every constant $c < 2/3$, we have $\mathbb{P}\left[\mathcal{L}_m \leq m \exp(-\log^c m)\right] \to 1$, and conjectured there exists a fixed $\delta < 1$ such that $\mathbb{P}\left[\mathcal{L}_m \leq m^{\delta}\right] \to 1$. In Chapter 4 we confirm this conjecture by proving that a.a.s. $\mathcal{L}_m < m^{0.99999996}$.

Regarding lower bounds, we show that $\mathcal{L}_m \geq m^{\log 2/\log 3}$, and that $\mathbb{E}\left[\mathcal{L}_m\right] = \Omega\left(n^{0.88}\right)$.

1.1.4 Chapter 5

Recall that the Webgraph is a directed graph whose vertices are the static web pages, and there is an edge joining two vertices if there is a hyperlink in the first page pointing to the second page. As mentioned before, Barabási and Albert [9] introduced one of the first models for the Webgraph, in which the main idea is that a new web page prefers to link to web pages that have been cited more. Pandurangan, Raghavan and Upfal [108] introduced the *PageRank-based selection model* for the Webgraph, in which the main idea is that a new web page prefers to link to web pages that have higher *PageRanks* (see Section 2.6 for the definition of PageRank). Blum, Chan, and Rwebangira [16] introduced a *random-surfer model* for the Webgraph, in which the links of a new web page are chosen by doing independent random walks that start from random web pages and whose lengths

are geometric random variables. Chebolu and Melsted [31] showed that under certain conditions, the previous two models are equivalent. See Chapter 5 for details.

The diameter of the Barabási -Albert model was analyzed by Bollobás and Riordan [20]. Previous work on the PageRank-based selection and random-surfer models has focused on their degree distributions. In Chapter 5 we give logarithmic upper bounds for their diameters. We also give close lower and upper bounds in the special case $d = 1$, namely when the generated graph is (almost) a tree. It turns out that the key parameter in this case is the *height* of the generated random tree. The model is parametrized by $p \in (0, 1)$. We find the a.a.s. asymptotic value of the height for all $p \in [0.21, 1]$, and for $p \in (0, 0.21)$ we provide logarithmic lower and upper bounds.

We remark that although the models studied in this chapter are quite different from RANs, surprisingly the same engine is used for proving these results, namely the technique of Broutin and Devroye [27]. Our two applications of this technique demonstrates its flexibility.

1.2 Randomized rumour spreading

Randomized rumour spreading is an important primitive for information dissemination in networks and has numerous applications in network science, ranging from spreading information in the WWW and Twitter to spreading viruses and diffusion of ideas in human communities. A well studied rumour spreading protocol is the *(synchronous) push&pull protocol*, introduced by Demers, Greene, Hauser, Irish, Larson, Shenker, Sturgis, Swinehart, and Terry [45] and popularized by Karp, Schindelhauer, Shenker, and Vöcking [82]. Suppose that one node in a network is aware of a piece of information, the 'rumour', and wants to spread it to all nodes quickly. The protocol proceeds in rounds. In each round, every *informed* node contacts a random neighbour and sends the rumour to it ('pushes' the rumour), and every *uninformed* node contacts a random neighbour and gets the rumour if the neighbour knows it ('pulls' the rumour).

A point to point communication network can be modelled as an undirected graph: the nodes represent the processors and the links represent communication channels between them. Studying rumour spreading has several applications to distributed computing in such networks, of which we mention just two. The first is in broadcasting algorithms: a single processor wants to broadcast a piece of information to all other processors in the network (see [76] for a survey). There are at least four advantages to the push&pull protocol: it puts much less load on the edges than naive flooding, it is simple (each node makes

a simple local decision in each round; no knowledge of the global topology is needed; no state is maintained), scalable (the protocol is independent of the size of network: it does not grow more complex as the network grows) and robust (the protocol tolerates random node/link failures without the use of error recovery mechanisms, see [62]). A second application comes from the maintenance of databases replicated at many sites, e.g., yellow pages, name servers, or server directories. There are updates injected at various nodes, and these updates must propagate to all nodes in the network. In each round, a processor communicates with a random neighbour and they share any new information, so that eventually all copies of the database converge to the same contents. See [45] for details. Other than the aforementioned applications, rumour spreading protocols have successfully been applied in various contexts such as resource discovery [75], distributed averaging [24], data aggregation [83], and the spread of computer viruses [12]. A discussion of these and other applications can be found in Keshav [84].

In the second part of the thesis we study two randomized rumour spreading protocols, namely the synchronous push&pull protocol and the asynchronous push&pull protocol. We investigate their spread times on various graphs. Given a graph and a starting vertex, the *spread time* of a certain protocol is the time it takes for the rumour to spread in the whole graph, i.e. the time difference between the moment the protocol is initiated and the moment when everyone learns the rumour. This parameter is directly related to the diameter of the graph, the topic of the first part: if the diameter is large, the rumour takes a lot of time to spread. Specifically, the spread time of the synchronous variant is at least the diameter divided by two. In the second part we only consider simple, undirected and connected graphs.

For the synchronous push&pull protocol, it turned out that the spread time is closely related to the *expansion profile* of the graph. Let $\Phi(G)$ and $\alpha(G)$ denote the conductance and the vertex expansion of a graph G, respectively (see Section 7.2 for the definitions). After a series of results by various scholars, Giakkoupis [73, 74] showed the spread time is $O\left(\min\{\Phi(G)^{-1} \cdot \log n, \alpha(G)^{-1} \cdot \log \Delta(G) \cdot \log n\}\right)$. This protocol has recently been used to model news propagation in social networks. Doerr, Fouz, and Friedrich [48] proved an upper bound of $O(\log n)$ for the spread time on Barabási-Albert graphs, and Fountoulakis, Panagiotou, and Sauerwald [69] proved the same upper bound (up to constant factors) for the spread time on Chung-Lu graphs. It is known that a.a.s. Barabási-Albert graphs and Chung-Lu graphs have conductance $\Omega(1)$ (see [36, 103]). So it is not surprising that rumours spread fast on these graphs.

All the above results assumed a synchronized model, i.e. all nodes take action simultaneously at discrete time steps. In many applications and certainly in real-world social networks, this assumption is not very plausible. Boyd, Ghosh, Prabhakar, and Shah [24]

proposed an asynchronous time model with a continuous time line. Each node has its own independent clock that rings at the times of a rate 1 Poisson process. The protocol now specifies for every node what to do when its own clock rings. The rumour spreading problem in the asynchronous time model has so far received less attention. Rumour spreading protocols in this model turn out to be closely related to Richardson's model for the spread of a disease [55] and to first-passage percolation [77] with edges having i.i.d. exponential weights. The main difference is that in rumour spreading protocols each vertex contacts one neighbour at a time. So, for instance in the push protocol, the net communication rate outwards from a vertex is fixed, and hence the rate that the vertex passes the rumour to any one given neighbour is inversely proportional to its degree (the push&pull protocol is a bit more complicated). Hence, the degrees of vertices play a crucial role not seen in Richardson's model or first-passage percolation. However, on regular graphs, the asynchronous push&pull protocol, Richardson's model, and first-passage percolation are essentially the same process, assuming appropriate parameters are chosen. In this sense, Fill and Pemantle [64] and Bollobás and Kohayakawa [19] showed that a.a.s. the spread time of the asynchronous push&pull protocol is $\Theta(\log n)$ on the hypercube graph. Janson [78] and Amini, Draief and Lelarge [5] showed the same results (up to constant factors) for the complete graph and for random regular graphs, respectively. These bounds match the same order of magnitude as in the synchronized case. Doerr, Fouz, Friedrich [50] experimentally compared the spread time in the two time models. They state that 'Our experiments show that the asynchronous model is faster on all graph classes [considered here].' However, a general relationship between the spread times of the two variants has not been proved theoretically.

1.2.1 Chapter 6

In Chapter 6 we answer a fundamental question about the asynchronous push&pull protocol: what are the minimum and maximum spread times on an n-vertex graph? Our proof techniques yield new results on the well studied synchronous version as well. We prove that in either version, the average spread time is at most linear even if only the pull operation is used. As the path graph admits linear spread time, this upper bound is asymptotically best possible, up to the constant factor.

The notion of 'guaranteed spread time' is defined in Chapter 6: this is the smallest number t such that once t units of time have passed, with probability at least $1 - 1/n$ all vertices are informed. We show that in both time models the guaranteed spread time is within a logarithmic factor of the average spread time, so it is $O(n \log n)$. In the asynchronous version, in contrast to the synchronous version, both the average and guaranteed

spread times are $\Omega(\log n)$ on any graph. We give examples of graphs illustrating that these bounds are best possible up to constant factors.

We also compare the spread times of the two versions on the same graph, and prove the first theoretical relationships between their guaranteed spread times. Firstly, in all graphs the guaranteed spread time in the asynchronous version is within an $O(\log n)$ factor of that in the synchronous version. Hence, polylogarithmic upper bounds for the synchronous variant automatically produce similar bounds for the asynchronous one. Next, we find examples of graphs whose asynchronous spread times are logarithmic, but the synchronous versions are polynomially large. Finally, we show for any graph that the ratio of the synchronous spread time to the asynchronous spread time is $O\big(n^{2/3}\big)$.

Previous work on the asynchronous push&pull protocol has focused on special graphs. This thesis is the first systematic study of this protocol on all graphs. We believe this protocol is fascinating and is quite different from its synchronous variant, in the sense that different techniques are required for analyzing it, and the spread times of the two versions can be quite different. Our work makes significant progress on better understanding of this protocol, and will hopefully inspire further research on this problem. Some explicit open problems have been mentioned in Chapter 8.

1.2.2 Chapter 7

In Chapter 7 we study the performance of the synchronous push&pull protocol on two random graph classes. The first one is the class of random k-trees, generated as follows: initially we have a complete graph on k vertices. In every step a new vertex is born, a random k-clique of the current graph is chosen, and the new vertex is joined to all vertices of the k-clique. (A k-clique in a graph is a complete subgraph of size k.) See Figure 1.2 for an illustration with $k = 2$.

We remark that this process is different from the random k-tree process defined by Cooper and Uehara [43] which was further studied in [41].

As in the preferential attachment scheme, the random k-tree process enjoys a 'the rich get richer' effect. Think of the number of k-cliques containing any vertex v as the 'wealth' of v (note that this quantity is linearly related to $\deg(v)$). Then, the probability that the new vertex attaches to v is proportional to the wealth of v, and if this happens, the wealth of v increases by $k - 1$. On the other hand, random k-trees have much larger clustering coefficients than preferential attachment graphs, as all neighbours of each new vertex are joined to each other. It is well-known that real-world networks tend to have large clustering coefficients (see, e.g., [118, Table 1]).

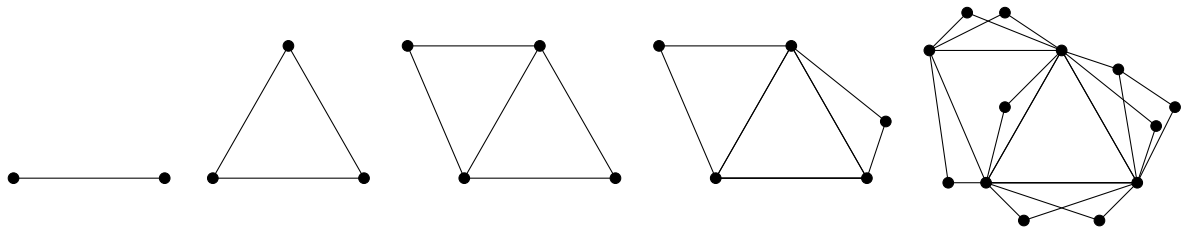


Figure 1.2: the construction of an instance of a random 2-tree with 13 vertices (last few steps are not shown.)

Gao [71] proved that a.a.s. the degree sequence of a random k-tree asymptotically follows a power law distribution with exponent $2 + \frac{1}{k-1}$. In Chapter 3 we show that a.a.s. its diameter is $O(\log n)$. We also show (in Section 7.2) that its clustering coefficient is at least $1/2$, as opposed to Barabási-Albert graphs and Chung-Lu graphs, whose clustering coefficients are $o(1)$ a.a.s. As per these properties, random k-trees may serve as more realistic models for real-world networks.

On the other hand, in Section 7.2 we prove that a.a.s. a random k-tree has conductance $O\left(\log n \cdot n^{-1/k}\right)$ and vertex expansion $O(k/n)$. Therefore we cannot resort to existing results linking the spread time to expansion properties to show rumours spread fast in these graphs. Another interesting structural property of a random k-tree is its *treewidth*. Roughly speaking, the treewidth of a graph measures its similarity to a tree: the treewidth of a tree is 1, and if the treewidth of a graph is small, it is 'tree-like' (see [86] for a comprehensive survey). Gao [72] proved that many random graph models, including Erdős-Rényi random graphs with expected degree $\omega(\log n)$ and preferential attachment graphs with out-degree greater than 11, have treewidth $\Theta(n)$, whereas all random k-trees have treewidth k by construction.

In conclusion, distinguishing features of random k-trees, such as high clustering coefficient, bad expansion (polynomially small conductance) and tree-like structure (small treewidth), inspired us to study randomized rumour spreading on this unexplored random environment.

In Chapter 7 we prove that when $k \geq 2$ is fixed, if initially a random vertex is aware of the rumour, then a.a.s. after $O\left((\log n)^{1+3/k}\right)$ rounds of the synchronous push&pull protocol, the rumour propagates to $n - o(n)$ vertices. Since random k-trees have polynomially small conductance, vertex expansion $O(1/n)$ and constant treewidth, these results demonstrate that push&pull can be efficient even on poorly connected networks.

On the negative side, we prove that a.a.s. the protocol needs at least $\Omega\left(n^{1/(5k)}\right)$ rounds

to inform *all* vertices. Our results imply that after just a polylogarithmic amount of time has passed, 99 percent of the vertices are informed, however, to inform each and every vertex polynomially many rounds are required. This exponential dichotomy between time needed for informing *almost all* and *all* vertices is a striking phenomenon and the main message of this chapter is to present a natural class of random graphs in which this phenomenon can be observed. In fact, in applications such as viral marketing and voting, it is more appealing to inform 99 percent of the vertices very quickly instead of waiting a long time until everyone gets informed. For such applications, our upper bound implies that push&pull can be effective even on poorly connected graphs.

A closely related class of graphs is that of *random k-Apollonian networks*, which generalize random Apollonian networks mentioned in Section 1.1.3. Their construction is very similar to that of random k-trees, with just one difference: if a k-clique is chosen in a certain round, it will never be chosen again. It is known that a.a.s. a random k-Apollonian network exhibits a power law degree distribution and large clustering coefficient [104, 120] and has logarithmic diameter (a proof appears in Section 3.5). Our technique for proving the upper bound for random k-trees successfully carries over to random k-Apollonian networks with no difficulty. We show that when $k \geq 3$ is fixed, if initially a random vertex is aware of the rumour, then a.a.s. after $O\left((\log n)^{1+3/k}\right)$ rounds, the rumour propagates to $n - o(n)$ vertices.

1.3 Summary of main new results

This section contains a list of main new results proved in this thesis. An informal description of our major contributions can be found in Chapter 8.

1. In Chapter 3 we prove that a.a.s. the diameter of graphs generated by each of the following models is $O(\log n)$: the forest fire model (Theorem 3.5), the copying model (Theorem 3.8), the PageRank-based selection model (Theorem 3.12), the Aiello-Chung-Lu models (Theorems 3.18 and 3.27), the generalized linear preference model (Theorem 3.22), directed scale-free graphs (Theorem 3.32), and the Cooper-Frieze model (Theorem 3.34). These results appear in the submitted preprint [98].

2. In Theorem 3.24 we prove for the preferential attachment model with random initial degrees in the case that the initial degrees' distribution has an exponential decay, that, a.a.s. the diameter is $O(\log^3 n)$. This result appears in the submitted preprint [98].

3. In Theorem 4.1 we prove that a.a.s. the diameter of a random Apollonian network is asymptotic to $c \log n$, where $c \approx 1.668$ is the solution of an explicit equation. This result was proved in collaboration with Ebrahimzadeh, Farczadi, Gao, Sato, Wormald, and Zung, and has appeared in [57].

4. In Theorem 4.2 we show that a random Apollonian network always has a cycle and a path of length $n^{0.63}$, and that the expected length of its longest cycles and paths is $\Omega\left(n^{0.88}\right)$. These results were proved in collaboration with Ebrahimzadeh, Farczadi, Gao, Sato, Wormald, and Zung, and have appeared in [57].

5. In Theorem 4.4 we prove that a.a.s. a random Apollonian network does not contain a path of length $n^{0.99999996}$. This result was proved in collaboration with Collevecchio and Wormald, and appears in the submitted preprint [37].

6. In Theorem 5.2 we show that a.a.s. the diameter of the random-surfer Webgraph model with parameters p and d is at most $8e^p(\log n)/p$. In Theorem 5.8 we show the same conclusion holds for the PageRank-based selection Webgraph model. These results were proved in collaboration with Wormald and appear in the submitted manuscript [101], an extended abstract of which has been published [102].

7. In Theorems 5.3 and 5.4, we determine the a.a.s. asymptotic value of the height and diameter of a random surfer tree with parameter p in the regime $p > 0.21$. When $p \leq 0.21$, we provide logarithmic lower and upper bounds. The same conclusions apply to the PageRank-based selection Webgraph model with $\beta = 0$ and $d = 1$. These results were proved in collaboration with Wormald, appear in the submitted manuscript [101], an extended abstract of which has been published [102].

8. In Theorem 6.3 we prove the following bounds for any connected graph G:

$$\begin{aligned} (1-1/n)\,\mathrm{wast}_{\mathsf{a}}(G) \leq\ & \mathrm{gst}_{\mathsf{a}}(G) \ \leq e\,\mathrm{wast}_{\mathsf{a}}(G)\log n\,, \\ \mathrm{wast}_{\mathsf{a}}(G) = \Omega(\log n) \quad & \text{and} \qquad \mathrm{wast}_{\mathsf{a}}(G) = O(n)\,, \\ \mathrm{gst}_{\mathsf{a}}(G) = \Omega(\log n) \quad & \text{and} \quad \mathrm{gst}_{\mathsf{a}}(G) \ = O(n\log n)\,. \end{aligned}$$

   Here, $\mathrm{wast}_{\mathsf{a}}$ and $\mathrm{gst}_{\mathsf{a}}$ denote the 'worst-case average spread time' and 'guaranteed spread time' of the *asynchronous* push&pull protocol, respectively. Moreover, we show these bounds are asymptotically best possible, up to the constant factors. These results were proved in collaboration with Acan, Collevecchio, and Wormald, and appear in the submitted preprint [1].

9. In Theorem 6.4 we show the following hold for any connected graph G:

$$(1-1/n)\,\mathrm{wast}_{\mathsf{s}}(G) \le \ \mathrm{gst}_{\mathsf{s}}(G) \ \le e\,\mathrm{wast}_{\mathsf{s}}(G)\log n\,,$$
$$\mathrm{wast}_{\mathsf{s}}(G) = O(n), \quad \text{and} \qquad \mathrm{gst}_{\mathsf{s}}(G) = O(n\log n)\,.$$

Here, $\mathrm{wast}_{\mathsf{s}}$ and $\mathrm{gst}_{\mathsf{s}}$ denote the 'worst-case average spread time' and 'guaranteed spread time' of the *synchronous* push&pull protocol, respectively. Moreover, we show these bounds are asymptotically best possible, up to the constant factors. These results were proved in collaboration with Acan, Collevecchio, and Wormald, and appear in the submitted preprint [1].

10. In Corollary 6.9 we prove that for any connected graph G we have

$$\frac{\mathrm{gst}_{\mathsf{s}}(G)}{\mathrm{gst}_{\mathsf{a}}(G)} = \Omega(1/\log n) \quad \text{and} \quad \frac{\mathrm{gst}_{\mathsf{s}}(G)}{\mathrm{gst}_{\mathsf{a}}(G)} = O\big(n^{2/3}\big)\,,$$

and the left-hand bound is asymptotically best possible, up to the constant factor. Moreover, we find infinitely many graphs for which this ratio is $\Omega\left(n^{1/3}(\log n)^{-4/3}\right)$. These results were proved in collaboration with Acan, Collevecchio, and Wormald, and appear in the submitted preprint [1].

11. In Theorem 7.3 we show that if initially a random vertex of a random k-tree with fixed $k \ge 2$ knows a rumour, then a.a.s. after $O\left((\log n)^{1+3/k}\right)$ rounds of the synchronous push&pull protocol, $n - o(n)$ vertices will learn the rumour. In Theorem 7.6 we show the same conclusion holds in a random k-Apollonian network with $k \ge 3$. These results were proved in collaboration with Pourmiri, and appear in the submitted manuscript [99], an extended abstract of which has been published [100].

12. In Theorem 7.5 we show that a.a.s. the spread time of the synchronous push&pull protocol on a random k-tree with $k \ge 2$ is at least $n^{1/(5k)}$. This result was proved in collaboration with Pourmiri, appears in the submitted manuscript [99], an extended abstract of which has been published [100].

Chapter 2

Preliminaries

2.1 Notation and terminology

- ✓ All logarithms are in the natural base.
- ✓ $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, and $[n] = \{1, 2, \dots, n\}$.
- ✓ $X \stackrel{d}{=} Y$ means X and Y have the same distribution.
- ✓ $X \stackrel{s}{\leq} Y$ means X is *stochastically smaller than* Y, or X is *stochastically dominated by* Y, that is, for any constant t,

$$\mathbb{P}[X \geq t] \leq \mathbb{P}[Y \geq t] \ .$$

- ✓ A sequence of events $(A_1, A_2, \dots)$ is said to happen *asymptotically almost surely (a.a.s.)* if $\mathbb{P}[A_n]$ approaches 1 as n goes to infinity. In most cases the indexing is implicit and clear from the context and we just say A happens a.a.s. In some cases the events are indexed by a continuous index, which is usually *time*.
- ✓ For two functions $f(n)$ and $g(n)$ we write $f \sim g$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1 \ .$$

✓ For a random variable $X = X(n)$ and a function $f(n)$, we say X is *a.a.s. asymptotic to* $f(n)$ (and write *a.a.s.* $X \sim f(n)$) if for every fixed $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[f(n)(1-\varepsilon) \leq X \leq f(n)(1+\varepsilon)\right] = 1\,.$$

✓ We write *a.a.s.* $X = o\big(f(n)\big)$ if for every constant $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[X \leq \varepsilon f(n)\right] = 1\,.$$

✓ We write *a.a.s.* $X = O\big(f(n)\big)$ if there exists a constant C such that

$$\lim_{n\to\infty} \mathbb{P}\left[X \leq C f(n)\right] = 1\,.$$

✓ $\Delta(G)$ denotes the maximum degree of graph G.

✓ In this thesis graphs and trees are finite, simple and undirected, unless specified otherwise.

✓ The *distance* (also called the *shortest-path distance*) between two vertices is the number of edges in the shortest path connecting them. If the graph is directed, the direction of edges is ignored when calculating the distance.

✓ The *diameter* of a graph is the maximum distance between any two vertices. We will work with (weakly) connected graphs only, so the diameter is always well defined.

✓ The *degree* of a vertex is the number of its incident edges (where a loop is counted twice) and is denoted by $\deg(\cdot)$.

✓ A *leaf* in a graph is a vertex of degree 1.

✓ In some chapters we work with *rooted* graphs/trees, in which there is a designated vertex called the *root*. The *depth* of a vertex is its distance to the root, and the *height* of a graph/tree is the maximum depth of its vertices.

✓ In a rooted tree, a vertex u is a *descendant* of v if the unique path connecting u to the root passes through v. The set of descendants of v is sometimes called its *offspring*. If u is a descendant of v, then v is an *ancestor* of u.

✓ In a rooted tree, vertex p is the *parent* of vertex v if p is the unique neighbour of v that is closest to the root. Note that any non-root vertex has a unique parent.

2.2 Probability distributions

Definition 2.1 (Binomial distribution)**.** Let $\mathrm{Bin}(n,p)$ denote a binomial random variable with parameters n and p; namely for every integer $k \in \{0,1,\dots,n\}$,

$$\mathbb{P}\left[\mathrm{Bin}(n,p)=k\right] = \binom{n}{k} p^k (1-p)^{n-k}\,.$$

Definition 2.2 (Geometric distribution)**.** Let $\mathrm{Geo}(p) \in \{0,1,2,\dots\}$ denote a geometric random variable with parameter p; namely for every $k \in \mathbb{N}_0$, $\mathbb{P}\left[\mathrm{Geo}(p)=k\right] = (1-p)^k p$.

Definition 2.3 (Exponential distribution)**.** Denote by $\mathrm{Exp}(\lambda)$ an exponential random variable with parameter (or rate) λ and mean $1/\lambda$, namely for any $x \geq 0$,

$$\mathbb{P}\left[\mathrm{Exp}(\lambda) \geq x\right] = \exp(-\lambda x)\,.$$

We next define two continuous distributions that appear naturally in studying urn models.

Definition 2.4 (Gamma function)**.** Let $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\mathrm{d}x$.

Definition 2.5 (Beta distribution)**.** Let $\alpha,\beta > 0$. A $\mathrm{Beta}(\alpha,\beta)$ random variable is supported on $(0,1)$ and its density function is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1} \qquad \text{for } x \in (0,1)\,.$$

Definition 2.6 (Dirichlet distribution)**.** Let $\alpha_1,\alpha_2,\dots,\alpha_n$ be positive numbers. The $\mathrm{Dirichlet}(\alpha_1,\alpha_2,\dots\alpha_n)$ distribution has support on the set

$$\left\{(x_1,x_2,\dots,x_n)\colon x_i \geq 0 \text{ for } 1 \leq i \leq n, \text{ and } \sum_{i=1}^n x_i = 1\right\},$$

and its density at point $(x_1,x_2,\dots,x_n)$ equals

$$\frac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{j=1}^n x_j^{\alpha_j-1}.$$

It is known that if the vector $(X_1,X_2,\dots,X_n)$ is distributed as $\mathrm{Dirichlet}(\alpha_1,\alpha_2,\dots,\alpha_n)$, then the marginal distribution of X_i is $\mathrm{Beta}(\alpha_i, \sum_{j\neq i}\alpha_j)$, see, e.g., [81, Section 2.7]. Dirichlet distribution is indeed the multivariate generalization of beta distribution.

2.3 Eggenberger-Pólya urn models

Urn models are important probabilistic objects and naturally appear in the analysis of random (k-)Apollonian networks and random k-trees (Chapters 4 and 6), which are evolving random graphs with reinforcement.

In this section we give some definitions and review some relevant results. See Johnson and Kotz [81] for a general introduction, and Mahmoud [96] for a more recent survey.

Definition 2.7 (Eggenberger-Pólya urn)**.** Start with w_0 white and b_0 black balls in the urn. In every step a ball is drawn from the urn uniformly at random, the ball is returned to the urn, and s balls of the same color are added to the urn. Let $\mathrm{Urn}(w_0, b_0, s, n)$ denote the number of white balls right after n draws.

In Chapter 7 we will consider more complicated urns, see Definition 7.9.

The following result is due to Eggenberger and Pólya [58] (see, e.g., Mahmoud [96, Theorem 5.1.2]).

Theorem 2.8 (limiting distribution in Eggenberger-Pólya urns [58])**.** *For any $\alpha \in [0,1]$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\mathrm{Urn}(a,b,s,n)}{sn} < \alpha\right] = \mathbb{P}\left[\mathrm{Beta}\left(\frac{w}{s}, \frac{b}{s}\right) < \alpha\right] = \frac{\Gamma((w+b)/s)}{\Gamma(w/s)\Gamma(b/s)} \int_0^\alpha x^{\frac{w}{s}-1}(1-x)^{\frac{b}{s}-1}\mathrm{d}x.$$

The following proposition follows from de Finetti's theorem, the fact that distinct draws from an Eggenberger-Pólya urn are exchangeable, and Theorem 2.8; see, e.g., [81, page 181].

Proposition 2.9. *Let $X = \mathrm{Beta}(w/s, b/s)$. Then*

$$\mathrm{Urn}(w,b,s,n) \stackrel{d}{=} w + s \cdot \mathrm{Bin}(n, X)\,.$$

This proposition gives an alternative useful way to generate a random variable distributed as $\mathrm{Urn}(w,b,s,n)$.

Theorem 2.8 can be generalized to get the joint distribution for the proportion of balls of each colour: the following theorem is due to Athreya [7] (see also [81, page 378]).

Theorem 2.10 (limiting joint distribution in multicolour Eggenberger-Pólya urns [7])**.** *Consider an urn with k different colours. Suppose that initially there are $c_i > 0$ balls of*

colour $i \in [k]$. In every step a ball is drawn from the urn uniformly at random, the ball is returned to the urn, and s balls of the same colour are added to the urn. For $i \in [k]$ and $n \in \mathbb{N}$, let $X_{i,n}$ denote the number of balls of colour i after n draws. Then we have the 'weak' convergence

$$\left(\frac{X_{1,n}}{sn}, \frac{X_{2,n}}{sn}, \dots, \frac{X_{k,n}}{sn}\right) \longrightarrow \text{Dirichlet}\left(\frac{c_1}{s}, \frac{c_2}{s}, \dots, \frac{c_k}{s}\right)$$

as $n \to \infty$, which means that for any fixed $A \subseteq \mathbb{R}^k$, as $n \to \infty$ we have

$$\mathbb{P}\left[\left(\frac{X_{1,n}}{sn}, \frac{X_{2,n}}{sn}, \dots, \frac{X_{k,n}}{sn}\right) \in A\right] \to \mathbb{P}\left[\text{Dirichlet}\left(\frac{c_1}{s}, \frac{c_2}{s}, \dots, \frac{c_k}{s}\right) \in A\right] .$$

2.4 Properties of the exponential distribution

Basic properties of the exponential distribution, which have been used in the analysis of the asynchronous push&pull protocol in Chapter 6, are reviewed here. The poissonization technique, which has been used in Chapters 4 and 5 to analyze the height of random trees is also skimmed.

Proposition 2.11. *Let $X_1, \dots, X_k$ be independent exponential random variables with rates $\lambda_1, \dots, \lambda_k$. Then*

$$\min\{X_1, \dots, X_k\} \stackrel{d}{=} \text{Exp}(\lambda_1 + \dots + \lambda_k) .$$

Proof. For any $t \geq 0$,

$$\mathbb{P}\left[\min\{X_1, \dots, X_k\} > t\right] = \prod_{i \in [k]} \mathbb{P}\left[X_i > t\right] = \prod_{i \in [k]} e^{-t\lambda_i} = e^{-t(\sum_{i \in [k]} \lambda_i)} . \quad \blacksquare$$

Proposition 2.12. *Let $k > 0$. If $X \stackrel{d}{=} \text{Exp}(\lambda)$ then $kX \stackrel{d}{=} \text{Exp}(\lambda/k)$.*

Proof. For any $t \geq 0$,

$$\mathbb{P}\left[kX > t\right] = \mathbb{P}\left[X > t/k\right] = e^{-\lambda t/k} . \quad \blacksquare$$

Proposition 2.13. *Let U be a uniform random variable on $(0, 1)$. Then $-\log U_i \stackrel{d}{=} \text{Exp}(1)$.*

Proof. For any $t \geq 0$,

$$\mathbb{P}\left[-\log U_i \geq t\right] = \mathbb{P}\left[U_i \leq e^{-t}\right] = e^{-t} \,. \quad \blacksquare$$

An important property of the exponential distribution is the so-called *memorylessness*.

Proposition 2.14 (memorylessness of the exponential distribution)**.** *Let* $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$ *for an arbitrary* $\lambda > 0$*. For any* $s > t > 0$ *we have*

$$\mathbb{P}\left[X > s | X > t\right] = \mathbb{P}\left[X > s - t\right] \,.$$

Proof.

$$\mathbb{P}\left[X > s | X > t\right] = \frac{\mathbb{P}\left[X > s, X > t\right]}{\mathbb{P}\left[X > t\right]} = \frac{e^{\lambda s}}{e^{\lambda t}} = e^{\lambda s - \lambda t} = \mathbb{P}\left[X > s - t\right] \,. \quad \blacksquare$$

If we think of X as a 'waiting time' for an event to happen, then the memorylessness property says that if we have waited for some time and the event has not happened, the additional time we need to wait is again a random variable with the same distribution, i.e. it does not matter how long we have already waited; the random variable has 'forgotten' how much we have already waited.

Definition 2.15 (Poisson clock)**.** Let $\lambda > 0$ and let $X_1, X_2, \ldots$ be i.i.d. $\mathrm{Exp}(\lambda)$ random variables. A *Poisson clock* with rate (or parameter) λ is a clock that rings at times $X_1, X_1 + X_2, X_1 + X_2 + X_3, \ldots$. The associated *Poisson process* is the list of points $\{X_1, X_1 + X_2, X_1 + X_2 + X_3, \ldots\}$.

Memorylessness of the exponential distribution carries over to the Poisson clock. Suppose we start a Poisson clock at time 0. Then, for any fixed $t > 0$, if we listen to the clock from time t onwards, the clock would work as a Poisson clock started from t. In other words, we can imagine that the clock has been 'restarted' at time t, and it 'forgets' what it has done between 0 and t. This is a 'Markovian' property: the future behaviour is independent of the past! (In fact, it is also independent of the present.)

This is formalized in the following proposition, whose proof is omitted, and can be found in basic textbooks on probability, e.g., Ross [113, Section 5.3.3].

Proposition 2.16 (memorylessness of the Poisson clock)**.** *Let $\mathcal{P} = \{X_1, X_2, \dots\}$ be a Poisson process on $[0, \infty)$, and let τ be a 'stopping time'. That is, to determine if $\tau = t$, we need only know the set $\mathcal{P} \cap [0, t]$. Let N_τ be the smallest index such that $X_{N_\tau} > \tau$. Then*

$$\{X_{N_\tau} - \tau, X_{N_\tau+1} - \tau, X_{N_\tau+2} - \tau, \dots\}$$

is also a Poisson process with the same rate as $\mathcal{P}$.

Two important examples of stopping times are: (1) any fixed time t, and (2) the first time we hear the ith ring, for any fixed $i \in \mathbb{N}$.

Using Proposition 2.16 we can give a short proof for the following proposition.

Proposition 2.17. *Let $X_1, X_2, \dots, X_n$ be i.i.d.* $\mathrm{Exp}(\lambda)$ *random variables. Then we have* $\mathbb{E}[\max\{X_1, \dots, X_n\}] \sim (\log n)/\lambda$.

Proof. Let $Y = \max\{X_1, \dots, X_n\}$. Consider n Poisson clocks with rate λ that start working at time 0. Then Y is the first time that all clocks have rung. By Proposition 2.11, the expected time until we hear a ring is $1/(n\lambda)$. At this point, we can ignore the corresponding clock and wait for the rest of clocks to ring. By memorylessness of the Poisson process, we may assume that the clocks are restarted. Since we have $n-1$ clocks now, the expected waiting time to hear the second ring is $1/\lambda(n-1)$. Continuing similarly and using linearity of expectation, we find

$$\mathbb{E}[Y] = \sum_{i=0}^{n-1} \frac{1}{\lambda(n-i)} \sim (\log n)/\lambda\,. \quad \blacksquare$$

Another nice property of Poisson clocks is that, many Poisson clocks work together as one, in the following sense: assume we have k Poisson clocks with rates $\lambda_1, \lambda_2, \dots, \lambda_k$. We start them at the same time, and put them in a black box so that we cannot see them but can hear them. Whenever any of the clocks ring, we hear it, but we do not know which one has rung. It is known that the collection of rings we hear corresponds to a Poisson process with rate $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$. Moreover, if we hear a ring, then the probability that it was actually the ith clock, equals λ_i/λ (note that, almost surely no two clocks ring at exactly the same time).

Conversely, suppose we have one Poisson clock with rate $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$. Whenever it rings, we independently perform one of k possible actions, action i with probability λ_i/λ. Then, for each $i \in [k]$, the times when action i was performed is itself a Poisson process with rate λ_i. For proofs of all these facts, see, e.g., [113, Section 5.3.4].

2.4.1 Poissonization

Poissonization is a way of 'embedding a discrete-time process into a continuous-time process', which sometimes helps in getting rid of annoying dependencies, and works because of the memorylessness properties of Poisson processes. We illustrate this technique by an example, which has been used in Chapters 4 and 5.

Let $d \in \mathbb{N}$, and suppose that we have a random d-ary tree growing as follows: initially it has only one vertex, the root. In every time-step $i = 1, 2, \ldots$, a leaf is chosen uniformly at random and gives birth to d new children, so the total number of vertices increases by d, and the total number of leaves increases by $d - 1$. We continue this process for a long time, and suppose we are interested in the structure of the tree when it has n vertices.

A major difficulty here is that the branches of the tree are dependent, since the total number of vertices is n. Consider instead a continuous time process that proceeds as follows. At time $t = 0$ the root is born. With every vertex v is associated an independent Poisson clock with rate 1 when it is born. The first time that the clock rings, v dies and gives birth to d new children. Note that there is no dependence between branches here. Moreover, we have the nice property that at any moment you look at the process, the next leaf to die is a uniformly random one! This is because of the memorylessness of the Poisson process: whenever a vertex gives birth to d new children, d new Poisson clocks are created, and the existing Poisson clocks 'forget' how much they have already waited, i.e. we can imagine they are 'restarted.' Hence we have a collection of Poisson clocks with equal rate competing with each other, and so the clock which rings first is a uniformly random one. An inductive argument gives that if we condition on this continuous-time tree to have n vertices, its distribution is exactly like an n-vertex discrete-time tree described above. Also, the independence between branches here makes it much easier to analyze.

A caveat here is that the first model is parametrized by the number of vertices n, whereas the second model is parametrized by time t. How are these two parameters related? This question was answered by Broutin and Devroye [27, Proposition 2].

Proposition 2.18. *Let $N(t)$ denote the number of vertices at time t of the continuous-time d-ary tree process. Then a.a.s. as $t \to \infty$ we have*

$$\log N(t) \sim dt \,.$$

Remark 2.19. Proposition 2 in [27] in fact states a stronger statement, that is,

$$\mathbb{P}\left[\lim_{t\to\infty} \frac{\log N(t)}{t} = d\right] = 1 \,.$$

2.5 Large deviation inequalities

Some standard concentration inequalities are proved here, which are used in various places, in particular in Chapter 5.

Theorem 2.20 (Cramér's Theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d., and let*

$$I(z) = \sup\{zt - \log \mathbb{E}\left[e^{tX_1}\right] : t \in \mathbb{R}\}.$$

For any $a > \mathbb{E}[X_1]$ we have

$$\mathbb{P}[X_1 + \cdots + X_n \geq an] = \exp\left(-I(a)n \pm o(n)\right),$$

and for any $a < \mathbb{E}[X_1]$ we have

$$\mathbb{P}[X_1 + \cdots + X_n \leq an] = \exp\left(-I(a)n \pm o(n)\right) .$$

Proof. See [46, Theorem I.4]. ■

The following three results follow easily from Cramér's Theorem. We provide proofs for completeness.

Lemma 2.21 (Cramér's Theorem for exponential random variables)**.** *Define the function $\Upsilon : (0, \infty) \to \mathbb{R}$ as*

$$\Upsilon(x) = \begin{cases} x - 1 - \log(x) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } 1 < x \, . \end{cases}$$

Let $E_1, E_2, \ldots, E_m$ be independent exponential random variables with parameter 1. For any fixed $x > 0$, as $m \to \infty$ we have

$$\exp\left(-\Upsilon(x)m - o(m)\right) \leq \mathbb{P}[E_1 + E_2 + \cdots + E_m \leq xm] \leq \exp(-\Upsilon(x)m) \, .$$

Proof. We first prove the upper bound. If $x > 1$ then $\exp(-\Upsilon(x)m) = 1$, so we may assume that $0 < x \leq 1$. Let $\theta = 1 - 1/x$. Then we have

$$\mathbb{E}[\exp(\theta E_1)] = \int_0^\infty e^{\theta x} e^{-x} \mathrm{d}x = \frac{1}{1-\theta} \, .$$

Using Markov's inequality and independence we get

$$\begin{aligned}\mathbb{P}\left[E_1 + E_2 + \cdots + E_m \le xm\right] &= \mathbb{P}\left[\exp(\theta E_1 + \cdots + \theta E_m) \ge \exp(\theta xm)\right] \\ &\le \mathbb{E}\left[\exp(\theta E_1 + \cdots + \theta E_m)\right] / \exp(\theta xm) \\ &= \mathbb{E}\left[\exp(\theta E_1)\right] \mathbb{E}\left[\exp(\theta E_2)\right] \ldots \mathbb{E}\left[\exp(\theta E_m)\right] / \exp(\theta xm) \\ &= (1-\theta)^{-m} \exp(-\theta xm) = \exp(-\Upsilon(x)m)\,.\end{aligned}$$

We now prove the lower bound. If $x > 1$, then the result follows from Markov's inequality, so we may assume that $0 < x \le 1$. Let $I(x) = \sup\{\lambda x - \log(\mathbb{E}\left[e^{\lambda E_1}\right]) : \lambda \le 0\}$. Since $\mathbb{E}\left[e^{\lambda E_1}\right] = 1/(1-\lambda)$ for all $\lambda < 1$, the supremum here occurs at $\lambda = 1 - 1/x$, which implies $I(x) = \Upsilon(x)$. Then Cramér's Theorem (Theorem 2.20) gives

$$\mathbb{P}\left[E_1 + E_2 + \cdots + E_m \le xm\right] = \exp(-I(x)m + o(m)) = \exp(-\Upsilon(x)m + o(m))\,,$$

as required. ■

Lemma 2.22. *Define the function $f : (-\infty, 1] \to \mathbb{R}$ as*

$$f(x) = (2-x)^{2-x} p (1-p)^{1-x} (1-x)^{x-1}\,.$$

Let $Z_1, Z_2, \ldots, Z_m$ be independent $1 + \mathrm{Geo}(p)$ random variables, and let $\kappa \ge 1/p$. Then we have

$$\mathbb{P}\left[Z_1 + Z_2 + \cdots + Z_m \ge \kappa m\right] \le f(2-\kappa)^m\,.$$

Proof. Let θ satisfy

$$e^\theta = \frac{\kappa - 1}{\kappa(1-p)}\,.$$

We have

$$\mathbb{E}\left[\exp(\theta Z_1)\right] = \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{\theta k} = \frac{pe^\theta}{1 - e^\theta(1-p)}\,.$$

Thus using Markov's inequality and independence we have

$$\begin{aligned}\mathbb{P}\left[Z_1 + Z_2 + \cdots + Z_m \ge \kappa m\right] &= \mathbb{P}\left[\exp(\theta Z_1 + \cdots + \theta Z_m) \ge \exp(\theta \kappa m)\right] \\ &\le \mathbb{E}\left[\exp(\theta Z_1 + \cdots + \theta Z_m)\right] / \exp(\theta \kappa m) \\ &= \mathbb{E}\left[\exp(\theta Z_1)\right] \mathbb{E}\left[\exp(\theta Z_2)\right] \cdots \mathbb{E}\left[\exp(\theta Z_m)\right] / \exp(\theta \kappa m) \\ &= \left(\frac{pe^{\theta - \theta\kappa}}{1 - e^\theta(1-p)}\right)^m = f(2-\kappa)^m\,.\end{aligned}$$

■

Lemma 2.23. *Let $B_1, \dots, B_n$ be independent Beta$(1/2, 1)$ random variables. For any $\beta > 0$ we have*

$$\mathbb{P}\left(\prod_{i=1}^{n} B_i \le \beta^n\right) \le \left(\frac{e \log(1/\beta)\sqrt{\beta}}{2}\right)^n .$$

Proof. If $\beta \ge e^{-2}$ then the right-hand side is at least 1, so we may assume that $0 < \beta < e^{-2}$. Let $\lambda = -1/2 - 1/\log \beta \in (-1/2, 0)$. We have

$$\mathbf{E}[B_1^\lambda] = \frac{\Gamma(3/2)}{\Gamma(1/2)\Gamma(1)} \int_0^1 x^\lambda x^{-1/2} \mathrm{d}x = \frac{1}{2\lambda + 1} .$$

Hence by Markov's inequality and since the B_i are independent,

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^{n} B_i \le \beta^n\right) &= \mathbb{P}\left(\prod_{i=1}^{n} B_i^\lambda \ge \beta^{\lambda n}\right) \\ &\le \prod_{i=1}^{n} \frac{\mathbb{E}\left[B_1^\lambda\right]}{\beta^\lambda} = \left(\frac{1}{\beta^\lambda(2\lambda+1)}\right)^n = \left(\frac{e \log(1/\beta)\sqrt{\beta}}{2}\right)^n . \qquad \blacksquare\end{aligned}$$

2.6 PageRank

To analyze the diameter of the PageRank-based selection model (in Chapters 3 and 5), we will need a simple result regarding the PageRank distribution, which is proved here. We start by defining the PageRank distribution.

Definition 2.24 (PageRank)**.** Let $p \in (0, 1]$ and let G be a directed graph possibly with loops and multiple edges. *PageRank with restart probability* p is a probability distribution over $V(G)$, which is the stationary distribution of the following random walk. In each step, with probability p we jump to a vertex chosen uniformly at random, and with probability $1 - p$ we walk to a random out-neighbour of the current vertex.

Since $p > 0$ in the above definition, the random walk is aperiodic and irreducible (see, e.g., [113, Section 4] for standard Markov chain definitions). It is known that the stationary distribution of a finite, irreducible, aperiodic Markov chain is unique (see, e.g., [113, Theorem 4.1]) so PageRank is well defined.

It follows from Definition 2.24 that PageRank is the unique probability distribution $\pi_p : V(G) \to [0,1]$ that satisfies

$$\pi_p(v) = \frac{p}{|V(G)|} + (1-p) \sum_{u \in V(G)} \frac{\pi_p(u) \cdot \#(uv)}{\text{out-deg}(u)} , \tag{2.1}$$

where $\#(uv)$ denotes the number of copies of the directed edge uv in the graph (which is zero if there is no edge from u to v), and out-deg(u) denotes the out-degree of u.

PageRank is used as a ranking mechanism in Google [25]. More details and applications can be found in [91].

It will be convenient to have an 'algorithmic' definition for PageRank, namely, a way to sample from the PageRank distribution. This is achieved by the following proposition.

Proposition 2.25. *If we sample a vertex uniformly and perform a simple random walk of length* Geo(p) *starting from the sampled vertex, the last vertex of the walk has PageRank distribution with restart probability p.*

Proof. This was first observed in [31]. Let $\tau \in [0,1]^{V(G)}$ denote the probability distribution of the last vertex, let $\mathcal{P}$ denote the probability transition matrix of the simple random walk on G, and let $\sigma = \left[1/|V(G)|, 1/|V(G)|, \ldots, 1/|V(G)|\right]^T$ be the uniform distribution. Then we have

$$\tau = \sum_{k=0}^{\infty} \left((1-p)^k p\right) \mathcal{P}^k \sigma = p\sigma + (1-p)\mathcal{P} \left(\sum_{k=1}^{\infty} (1-p)^{k-1} p \mathcal{P}^{k-1} \sigma \right) = p\sigma + (1-p)\mathcal{P}\tau .$$

Comparing with (2.1), we find that $\tau = \pi_p$, as required. ∎

2.7 Miscellaneous

Chernoff bounds. Let $X \stackrel{d}{=} \text{Bin}(n,p)$. We refer to the following inequalities, valid for every $\varepsilon \geq 0$, as the lower tail and upper tail Chernoff bounds, respectively. See McDiarmid [97, Theorem 2.3] for proofs.

$$\mathbb{P}\left[X \leq (1-\varepsilon)\mathbb{E}\left[X\right]\right] \leq \exp(-\varepsilon^2 \mathbb{E}\left[X\right]/2) . \tag{2.2}$$

$$\mathbb{P}\left[X \geq (1+\varepsilon)\mathbb{E}\left[X\right]\right] \leq \exp\left(-\frac{\varepsilon^2 \mathbb{E}\left[X\right]}{2+2\varepsilon/3}\right) . \tag{2.3}$$

As noted in [97], these inequalities are true also if X is a sum of arbitrary but independent indicator random variables.

Stirling's approximation. We will use the following bounds for $n!$, known as *Stirling's approximation* (see Feller [63, equation 9.15 in Chapter II]):

$$\sqrt{2\pi n}(n/e)^n < n! < 3\sqrt{n}(n/e)^n \qquad \forall n \in \mathbb{N}\,. \tag{2.4}$$

Galton-Watson branching process. Let Z be an $\mathbb{N}_0$-valued random variable. A Galton-Watson branching process starts with one particle born at time 0. Whenever a particle is born, say at time i, at time $i+1$ it gives birth to a random number of children and dies. The number of children of each particle is distributed as Z and these numbers are independent. An important quantity of interest is the probability of extinction. It is known that if $\mathbb{E}[Z] < 1$ this probability is 1, and if $\mathbb{E}[Z] > 1$, this probability is smaller than 1, see, e.g., [63, Section XII.4].

Couplings. Suppose we are investigating an unknown random object X that somehow resembles a better-known random object Z, and we would like to infer properties of X from properties of Z. The trouble might be that X and Z are defined on disjoint probability spaces. One way is to define a third random object Y (which can be equal to X or Z) and first generate Y, and then generate random objects with the desired distributions of X and Z given Y. We say that X and Z are 'coupled' using a common Y. In this way, X and Z are essentially generated on the same probability space. Of course, this generation should be valid, i.e. if one only looks at X generated this way, she should see 'the real' X, as if it were generated individually. If a valid coupling is built, then properties of Z can perhaps be translated to those of X, achieving our goal.

Chapter 3

Versatile technique for proving upper bounds

In this chapter[1] we present a versatile technique for establishing upper bounds for diameters of random graph models, and demonstrate it by proving logarithmic upper bounds for the diameters of a variety of models, including the following well known ones: the forest fire model [92], the copying model [89], the PageRank-based selection model [108], the Aiello-Chung-Lu models [2], the generalized linear preference model [28], directed scale-free graphs [17], the Cooper-Frieze model [38], and random k-trees [71]. This means that in each of these models, with probability close to 1, for *every* pair (u, v) of vertices there exists a very short (u, v)-path, a path connecting u and v whose length is logarithmic in the number of vertices. These results automatically imply logarithmic upper bounds for average diameters of these models. We also prove polylogarithmic upper bounds for the diameter of the preferential attachment model with random initial degrees [44] in the case that the initial degrees' distribution has an exponential decay. Prior to this work no sublinear upper bound was known even for the average diameter of any of these models. (This claim can quickly be verified by looking at Table 8.2 from the recent monograph [30], or [29, Table III], or the table in [21, p. 162]: each cited table contains a summary of known results on the diameter and other properties of several real-world network models.)

We study *evolving models* only (also called *on-line* or *dynamic* models), i.e. the graph changes over time according to pre-defined probabilistic rules, and we are interested in the long-term structure of this evolving graph. We assume that in discrete time-steps new vertices and edges appear in the graph, but no deletion occurs. The goal is to show that

[1] The contents of this chapter appear in the submitted preprint [98].

the evolving graph at time n has diameter $O(\log n)$ a.a.s. In all models considered here, the Chernoff bounds imply that the number of vertices at time n is $\Theta(n)$ a.a.s., hence we will conclude that a.a.s. the evolving graph has diameter $O(\log n)$ when it has n vertices.

In this chapter we are concerned with upper bounds only and no lower bound for the diameter is proved. However, we believe that for all models considered, at least in the special case when the evolving graph is always a tree, the diameter is $\Theta(\log n)$.

Chung and Lu [34] used couplings with a non-evolving random graph model to prove that the diameter of a certain growth-deletion model is $\Theta(\log n)$. On the one hand, their model is more flexible than the models we consider, as they allow vertex and edge deletions, but on the other hand, their result holds for graphs with at least $\omega(n \log n)$ edges whereas our results cover graphs with $O(n)$ edges, too. Moreover, their proof is quite technical and uses general martingale inequalities. See Remark 3.21 for details.

Our technique and chapter outline

Let us now informally explain our technique. In this chapter when we write a certain graph/tree has a logarithmic diameter/height, we mean its diameter/height has a logarithmic upper bound. An important object in this chapter is a *random recursive tree*, defined as follows: there exists a single node at time 0, and in every time-step $t = 1, 2, \dots$, a new node is born and is joined to a uniformly random node of the current tree. It is known that when this tree has n nodes, a.a.s. its height is $\Theta(\log n)$ [111]. Our technique consists of two main steps: first, we build a *coupling* between our evolving random graph and some variant of a random recursive tree in such a way that the diameter of the graph is not more than four times the height of the tree, and then we prove that a.a.s. the tree has a logarithmic height. The second step is usually straightforward (see Lemma 3.3 for an example) and the tricky part is defining the 'coupled' tree. Let us give some examples.

To distinguish between a vertex of the graph and that of the tree, the latter is referred to as a 'node'. For models studied in Section 3.1, namely the forest fire model [92], the copying model [89], and the PageRank-based selection model [108], the coupled tree is a random recursive tree with weighted edges, which has the same node set as the vertex set of the graph. Let us assume that the initial graph has one vertex, so the tree starts with a single node corresponding to this initial vertex. These models evolve as follows: in every time-step a new vertex, say v, is born and is joined to some random vertices, say $w_1, \dots, w_d$, in the existing graph in such a way that for each j, vertex w_j has a short distance to a uniformly random vertex x_j of the existing graph. In other words, w_j can be obtained by doing a (possibly random) walk with a short length starting from a uniformly

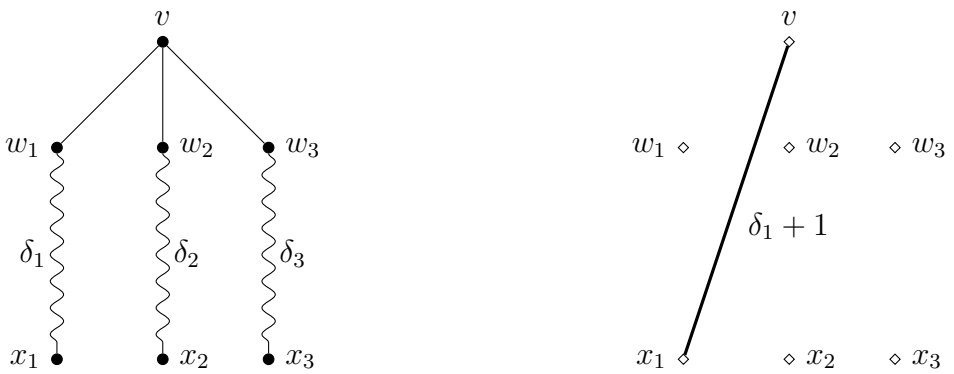


Figure 3.1: in the original graph (left), a new vertex v is born and is joined to some random vertices, say w_1, w_2 and w_3, in such a way that for each j, vertex w_j has distance δ_j to a uniformly random vertex x_j. Each shortest (w_j, x_j)-path has length δ_j and is depicted as a curly line. In the coupled tree (right), a new node v is born and is joined only to node x_1, and the weight of vx_1 is the distance between v and x_1 in the graph, i.e. $\delta_1 + 1$. We do not need the nodes w_2, w_3, x_2, x_3 for the coupling; but they have been depicted for completeness.

random vertex x_j. We let the coupled tree evolve as follows: a new node v is born and is joined to node x_1 in the existing tree, and the weight of the edge vx_1 in the tree is set to be the distance between v and x_1 in the graph, see Figure 3.1. Then by induction, the distance in the graph between the initial vertex and v is at most the weighted distance in the tree between the initial node and node v. Moreover, by construction, the tree evolves as a weighted random recursive tree. Finally, examining the distribution of the weights carefully, we prove that a.a.s. the obtained evolving tree has a logarithmic weighted height.

We remark that in the argument outlined above, we may *ignore* the other neighbours $w_2, \ldots, w_d$ of the new vertex; only the first edge vw_1 is effectively used for bounding the diameter. This is a repeating phenomenon in our arguments. An interesting implication is that we can quickly and locally build a spanning tree with logarithmic diameter as the graph evolves. This might have algorithmic applications.

In Section 3.2 we study models that incorporate preferential attachment. As a simple example, consider the following evolving rule: in every time-step, a vertex is chosen using preferential attachment, i.e. the probability of choosing a specific vertex is proportional to its degree, then a new vertex is born and is joined to the chosen vertex. It is easy to observe that sampling a vertex using preferential attachment can be done by choosing a uniformly random endpoint of a uniformly random edge. Using this sampling procedure,

Figure 3.2: in the original graph (left), an edge e is sampled uniformly, a random endpoint w is chosen, and a new vertex v is born and is joined to w. In the coupled edge-tree (right), a new node vw is born and is joined to a uniformly random node e.

the evolving rule can be re-stated as follows: in every time-step, an edge e is sampled uniformly at random, then a random endpoint w of e is chosen, then a new vertex v is born and is joined to w. One of the main novel ideas in this chapter is introducing *edge trees* and employing them in this context. An *edge* tree is a tree whose nodes correspond to the *edges* of the evolving graph. We couple the evolving graph with an edge tree, and let the edge tree evolve as follows: in the corresponding time-step a new node vw is born and is joined to a uniformly random node e, see Figure 3.2. Clearly, the edge tree indeed grows like a random recursive tree as the graph evolves, hence its height can be easily bounded. Moreover, the constructed coupling implies that the graph's diameter is at most four times the tree's height, so we conclude that a.a.s. the graph has logarithmic diameter. Theorem 3.15 formalizes and generalizes this idea and illustrates the crux of our technique. This theorem states that Model 3.13, a generic model based on the preferential attachment scheme, has logarithmic diameter; the Aiello-Chung-Lu models [2] and the generalized linear preference model [28] are then proved to be special cases of this model.

In the generalized linear preference model, the probability of choosing a specific vertex is proportional to a linear function, say $ax + b$, of the vertex's degree x. Assuming a and b are even positive integers, we handle this by putting a parallel edges corresponding to each edge, and putting $b/2$ loops at each vertex. Then choosing a uniformly random endpoint of a uniformly random edge in the new graph corresponds to sampling according to the linear function of the degrees in the old graph. See Theorem 3.22 for details. At the end of Section 3.2, we also analyze the 'preferential attachment with random initial degrees' model [44], and show that if the initial degrees' distribution has an exponential decay, then a.a.s. the generated graph has a polylogarithmic diameter. This is straightforward to prove using the machinery developed, see Theorem 3.24.

In Section 3.3 we study the 'directed scale-free graphs' [17]. Recall that the diameter of a directed graph is defined as that of the underlying undirected graph (we follow [92] in this regard). When generating a graph using this model, one may sample vertices according

to linear functions of either out-degrees or in-degrees, and the two functions have different constant terms. To cope with this, we introduce 'headless' and 'tailless' edges. These are dummy edges in the graph that do not play any role in connecting the vertices, but they appear in the tree and their job is just to adjust the selection probabilities. Details can be found in Theorem 3.30, which states that a generalized version of directed scale-free graphs has logarithmic diameter.

In Section 3.4 we study the Cooper-Frieze model [38], which is the most general evolving model known to have a power-law degree sequence. In this model, the neighbours of a new vertex can be chosen either according to degrees or uniformly at random. For dealing with this intricacy, we couple with a tree having two types of nodes: some correspond to the vertices, and the others correspond to the edges of the graph. A multi-typed random recursive tree is obtained, in which at every time-step a new node is born and is joined to a node chosen uniformly at random from all nodes of a certain type. We prove that a.a.s this tree has a logarithmic height, and by using the coupling's definition we conclude that a.a.s the Cooper-Frieze model has a logarithmic diameter (see Theorem 3.34). For this model, proving that the tree has logarithmic height is actually the harder step.

Finally, in Section 3.5 we prove logarithmic upper bounds for two further models: random k-trees [71] and random k-Apollonian networks [119]. For the latter, it is already known that a.a.s. the diameter is $O(\log n)$, but our approach gives a shorter proof.

Open problem 3.1. Develop a mathematical theory for characterizing those evolving random graphs which have logarithmic diameters. As a specific question, further develop this technique to cover other network models, e.g. growth-deletion models, accelerated network growth models, and spatial models.

The proof technique we introduce in this chapter could be a fundamental step in building this theory.

Open problem 3.2. Prove nontrivial lower bounds for the diameter of any of the models studied in this chapter.

We include some definitions here. In this chapter, graphs may have parallel edges and loops (note that adding these does not change the diameter). All graphs considered are finite and rooted, i.e. there is a special vertex which is called the root. The *depth* of a vertex is its distance to the root, and the *height* of a graph is the maximum depth of its vertices. Clearly the diameter is at most twice the height, and we always bound the diameter by bounding the height. The depth of vertex v in graph G is denoted by $\text{depth}(v, G)$.

A *growing graph* is a sequence $(G_t)_{t=0}^{\infty}$ of random graphs such that G_t is a subgraph of G_{t+1} for all $t \in \mathbb{N}_0$. We always assume that G_0 has size $O(1)$. A *growing tree* is defined similarly. This sequence can be thought of as a graph 'growing' as time passes, and G_t is the state of the graph at time t. We write informal sentences such as 'at time t, a new vertex is born and is joined to a random vertex of the existing graph,' which formally means 'G_t is obtained from G_{t-1} by adding a new vertex and joining it to a random vertex of G_{t-1}.'

Alternative approach for proving logarithmic diameters

In this section we mention a heuristic for proving logarithmic diameters for certain random graph models, which is due to Cooper, Frieze, and Uehara [41].

Suppose we start with a single vertex v_0 and in step t, where $1 \leq t \leq n$, a new vertex v_t is born and is joined to some of the vertices in the existing graph. Suppose that there exists a fixed $p > 0$ such that for each $1 \leq t \leq n$, in step t vertex v_t is joined to some v_j with $j \leq t/2$ with probability at least p. Then heuristically, we expect the diameter of the resulting graph to be $O(\log n)$: consider the shortest (v_n, v_0)-path $v_n = v_{n_0}, v_{n_1}, v_{n_2}, \ldots, v_0$. Then for each i, with probability at least p, we would have $n_{i+1} \leq n_i/2$. Informally we are 'building' a path to the root, and in each step our 'index' is halved with probability at least p, and so we expect that, with high probability after a logarithmic number of steps our index becomes 0, which means we have reached the root.

To make this rigorous, however, we want the halving to happen with a reasonable probability and *regardless of what we have exposed from the path so far.* In other words, we need that for each $1 \leq t \leq n$, *regardless of any information we may have about steps* $t+1, t+2, \ldots, n$, vertex v_t is joined to some v_j with $j \leq t/2$ with probability at least p. It is not straightforward to verify this condition in the models we consider, especially for those where there are steps in which edges are also added within the existing vertices (handling the revelation of information gets tricky).

Let us elaborate an example that demonstrates it is not easy to see for which models this heuristic gives a rigorous proof. Consider the following evolving random tree, which we call *Model X*. Start with a single vertex v_0. In every time-step $t = 1, 2, 3, \ldots$, a new vertex v_t appears and flips a fair coin: if it is heads, v_t is joined to v_0, otherwise it is joined to the vertex furthest from v_0. In this model, the diameter increases by 1 in each step with probability half, hence it would be around $n/2$ when there are n vertices, which is clearly not logarithmic. However, the model has the property that for each $1 \leq t \leq n$, vertex v_t

is joined to some v_j with $j \leq t/2$ with probability at least $1/2$, so the heuristic suggests that the diameter is logarithmic.

Nevertheless, the required condition can be shown to hold for the plain preferential attachment model and a variation of random k-trees (see [41, Lemma 4]), which leads to a rigorous proof based on this approach for showing these models have logarithmic diameters. This approach tends to give larger constants than that of this chapter. For instance, the upper bound given using the proof in [41, Lemma 4]) for the diameter of a random k-tree is $600 \log n$, whereas in Theorem 3.36 we give an upper bound of $2e \log n + O(1)$.

3.1 Basic technique

The general approach in this chapter is to reduce the problem to proving that a certain variant of a random recursive tree has logarithmic height, and then use a variant of the following lemma, which exemplifies proving such a tree has logarithmic height. We will use a simple inequality. Let $y_1, \dots, y_m$ be positive numbers, and assume $h \in [m]$ and $h > 1$. Then observe that

$$\sum_{1 \leq t_1 < \cdots < t_h \leq m} \left(\prod_{k=1}^{h} y_{t_k} \right) < \frac{1}{h!} \left(\sum_{i=1}^{m} y_i \right)^h . \tag{3.1}$$

Note that the sum in the left-hand-side is over all h-tuples $(t_1, t_2, \dots, t_h)$ satisfying $1 \leq t_1 < \cdots < t_h \leq m$. Let us briefly explain why (3.1) holds. When the right-hand-side is expanded, for each such h-tuple, the product $\prod_{k=1}^{h} y_{t_k}$ is generated precisely $h!$ times, since there are h ways to choose the bracket containing y_{t_1}, $h-1$ ways to choose the bracket containing y_{t_2}, etc. Moreover, in the expansion we have terms such as y_1^h, which do not appear in the left-hand-side, hence the inequality is strict.

The argument in the proof below is inspired by a proof in Frieze and Tsourakakis [70].

Lemma 3.3. *Let $(a_t)_{t \in \mathbb{N}}$ be a sequence of positive integers. Consider a growing tree $(T_t)_{t \in \mathbb{N}_0}$ as follows. T_0 is arbitrary. At each time-step $t \in \mathbb{N}$, a random vector $(W_1, W_2, \dots, W_{a_t}) \in V(T_{t-1})^{a_t}$ is chosen in such a way that for each $i \in [a_t]$ and each $v \in V(T_{t-1})$, the marginal probability $\mathbb{P}[W_i = v]$ equals $|V(T_{t-1})|^{-1}$. In other words, each W_i is a uniformly random node of T_{t-1}; however, the W_i's may be correlated. Then a_t new nodes $v_1, \dots, v_{a_t}$ are born and v_i is joined to W_i for each $i \in [a_t]$. Let $\ell = \ell(n), u = u(n)$ be positive integers such that $\ell \leq a_t \leq u$ for all $t \in [n]$. Then the height of T_n is a.a.s. at most $(u/\ell)e \log n + 2ue + O(1)$.*

Proof. Let $n_0 = |V(T_0)|$. For a given integer $h = h(n)$, let us bound the probability that T_n has a node at depth exactly $h + n_0$. Given a sequence $1 \leq t_1 < t_2 < \cdots <$

$t_h \le n$, the probability that there exists a path $wv_{t_1}v_{t_2}\cdots v_{t_h}$ in T_n with $w \in V(T_0)$ and $v_{t_1}, v_{t_2}, \ldots, v_{t_h} \notin V(T_0)$ such that v_{t_j} is born at time t_j is at most

$$n_0 u^h \prod_{k=1}^{h} \frac{1}{n_0 + \ell \cdot (t_k - 1)} ,$$

since there are n_0 choices for w, at most u^h choices for $(v_{t_1}, \ldots, v_{t_h})$ given $(t_1, \ldots, t_h)$, and for each $k = 1, 2, \ldots, h$, when v_{t_k} is born, there are at least $n_0 + \ell \cdot (t_k - 1)$ nodes available for it to join to. By the union bound and (3.1), the probability that T_n has a node at depth $h + n_0$ is at most

$$\begin{aligned} n_0 u^h \sum_{1 \le t_1 < \cdots < t_h \le n} \left(\prod_{k=1}^{h} \frac{1}{n_0 + \ell \cdot (t_k - 1)} \right) &< \frac{n_0 u^h}{h!} \left(\sum_{j=0}^{n-1} \frac{1}{n_0 + \ell j} \right)^h \\ &< n_0 \left(\frac{ue}{h} \cdot \left(\frac{\log n}{\ell} + 2 \right) \right)^h \Big/ \sqrt{2\pi h} , \end{aligned}$$

where for the second inequality we have used Stirling's approximation (2.4) for $h!$ and the inequality

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} < 1 + \log n .$$

Putting $h \ge (u/\ell)e \log n + 2ue$ makes this probability $o(1)$. Hence a.a.s. the height of T_n is at most $(u/\ell)e \log n + 2ue + n_0$, as required. ■

The first model we study is the *basic forest fire model* of Leskovec, Kleinberg, and Faloutsos [92, Section 4.2.1]. We prove an a.a.s. logarithmic upper bound using an easy application of Lemma 3.3.

Model 3.4. Let $p, q \in [0, 1]$ be arbitrary. Build a growing directed graph as follows. G_0 is an arbitrary weakly connected directed graph. At each time-step $t \in \mathbb{N}$, a new vertex v is born and edges are created from it to the existing graph using the following process.

1. All vertices are marked 'unvisited.' An *ambassador vertex* W is sampled uniformly from the existing graph.
2. Vertex v is joined to W and W is marked as 'visited.'
3. Independently generate two random variables $X = \mathrm{Geo}(p)$ and $Y = \mathrm{Geo}(q)$. Randomly select X unvisited out-neighbours and Y unvisited in-neighbours of W. If not enough unvisited in-neighbours or out-neighbours are available, select as many as possible. Let $W_1, \ldots, W_Z$ denote these vertices.

4. Vertex v is joined to $W_1, \dots, W_Z$, then apply steps 2–4 recursively to each of the vertices $W_1, \dots, W_Z$.

Theorem 3.5. *Consider $(G_t)_{t\in\mathbb{N}_0}$ generated by Model 3.4. A.a.s. for every vertex v of G_n there exists a directed path of length at most $e\log n + O(1)$ connecting v to some vertex of G_0. In particular, a.a.s. the diameter of G_n is at most $2e\log n + O(1)$.*

Remark 3.6. In the original paper [92] where this model was defined, the authors conjecture that the diameter actually decreases as n goes to infinity, however no analytical result is proved.

Proof. We define a growing tree $(T_t)_{t\in\mathbb{N}_0}$ in such a way that T_t is a spanning tree of G_t for all $t \in \mathbb{N}_0$: T_0 is an arbitrary spanning tree of G_0. For every $t \in \mathbb{N}$, if v is the vertex born at time t and w is the corresponding ambassador vertex, then v is joined only to w in T_t. By Lemma 3.3, a.a.s. the height of T_n is at most $e\log n + O(1)$. ■

We next study the *linear growth copying model* of Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins, and Upfal [89, Section 2.1].

Model 3.7. Let $p \in [0,1]$ and $d \in \mathbb{N}$. Build a growing directed graph in which every vertex has out-degree d, and there is a fixed ordering of these d edges. G_0 is an arbitrary weakly connected directed graph with all vertices having out-degree d. In each time-step $t \in \mathbb{N}$ a new vertex v is born and d outgoing edges from v to the existing graph are added, as described below. An ambassador vertex W is sampled uniformly from the existing vertices. For $i \in [d]$, the head of the i-th outgoing edge of v is chosen as follows: with probability p, it is a uniformly random vertex of the existing graph, and with probability $1-p$ it is the head of the i-th outgoing edge of W, in which case we say v has *copied* the i-th outgoing edge of W.

Theorem 3.8. *A.a.s. the diameter of G_n defined in Model 3.7 is at most $4e\log n + O(1)$.*

Proof. We inductively define a growing tree $(T_t)_{t\in\mathbb{N}_0}$ in such a way that the node set of T_t equals the vertex set of G_t for all t. We prove by induction that for each $v \in V(G_t)$, $\operatorname{depth}(v, G_t) \le 2\operatorname{depth}(v, T_t)$. Let T_0 be a breadth-first search tree of G_0, rooted at the root of G_0. For each $t \in \mathbb{N}$, let v be the vertex born at time t, and let w be the corresponding ambassador vertex. We consider two cases:

Case 1. v **copies at least one outgoing edge of** w**.** In this case, we join v to w in T_t. Since v and w have distance 2 in G_t, $\operatorname{depth}(v, G_t) \leq \operatorname{depth}(w, G_t) + 2$, so by the induction hypothesis for w,

$$\operatorname{depth}(v, G_t) \leq \operatorname{depth}(w, G_t) + 2 \leq 2\operatorname{depth}(w, T_t) + 2 = 2\operatorname{depth}(v, T_t)\,,$$

as required.

Case 2. v **does not copy any outgoing edge of** w**.** Let x denote the head of the first outgoing edge of v. In this case, we join v to x in T_t. Using the induction hypothesis for x,

$$\begin{aligned}\operatorname{depth}(v, G_t) \leq \operatorname{depth}(x, G_t) + 1 &\leq 2\operatorname{depth}(x, T_t) + 1 \\ &< 2\operatorname{depth}(x, T_t) + 2 = 2\operatorname{depth}(v, T_t)\,,\end{aligned}$$

as required.

Notice that in either case, node v is joined to a uniformly random node of T_{t-1}. By Lemma 3.3, a.a.s. the height of T_n is at most $e\log n + O(1)$, so a.a.s. the diameter of G_n is at most $4e\log n + O(1)$. ■

3.1.1 Sampling neighbours using PageRank

In this section we study a model in which the neighbours of each new vertex are chosen according to the PageRank distribution, which was defined in Section 2.6.

Model 3.9. Let p_a, p_b, p_c be nonnegative numbers summing to 1, let $q \in (0, 1]$ and $d \in \mathbb{N}$. Build a growing directed graph $(G_t)_{t\in\mathbb{N}_0}$ in which every vertex has out-degree d. G_0 is a weakly connected directed graph with all vertices having out-degree d. In each time-step $t \in \mathbb{N}$, a new vertex is born and d outgoing edges from it to the existing graph are added. The heads of the new edges are chosen independently. For choosing the head of each edge, perform one of the following operations, independently of previous choices.

(a) With probability p_a, the head is a vertex sampled uniformly from the existing graph.

(b) With probability p_b, it is the head of an edge sampled uniformly from the existing graph.

(c) With probability p_c, it is a vertex sampled from the existing graph using the PageRank distribution with restart probability q.

Model 3.9 is defined by Pandurangan, Raghavan, and Upfal [108, Section 2]. They call it the *hybrid selection model*. For the special case $p_b = 0$, which is referred to as the *PageRank-based selection model*, in Chapter 5 we show using a different argument that a.a.s. the diameter is $O(\log n)$. For bounding the diameter of Model 3.9 we will need a lemma.

Lemma 3.10. *There exists a random variable L such that the head of each new edge in Model 3.9 can be obtained by performing a simple random walk of length L in the existing graph starting from a uniformly random vertex. Moreover, $L \overset{s}{\leq} 1 + \mathrm{Geo}(q)$.*

Proof. We claim that

$$L = \begin{cases} 0 & \text{with probability } p_a\,, \\ 1 & \text{with probability } p_b\,, \\ \mathrm{Geo}(q) & \text{with probability } p_c\,. \end{cases}$$

If we sample a vertex uniformly and perform a random walk of length 1, then since all vertices have the same out-degree, the last vertex of the walk is the head of a uniformly random edge. By Proposition 2.25, if we sample a vertex uniformly and perform a random walk of length $\mathrm{Geo}(q)$, the last vertex of the walk has PageRank distribution with restart probability q, and this completes the proof. ■

Definition 3.11 (weighted tree, weighted depth)**.** A *weighted tree* is a tree with nonnegative weights assigned to the edges. The *weighted depth* of a node v is defined as the sum of the weights of the edges connecting v to the root.

Theorem 3.12. *A.a.s. the diameter of G_n defined in Model 3.9 is at most* $18(\log n)/q$.

Proof. We define a growing weighted tree $(T_t)_{t\in\mathbb{N}_0}$ such that for all t, the node set of T_t equals the vertex set of G_t. We prove by induction that the depth of each vertex in G_t is at most its weighted depth in T_t. Let T_0 be a breadth-first search tree of G_0 rooted at the root of G_0, and let all edges of T_0 have unit weights. Assume that when obtaining G_t from G_{t-1}, the heads of the new edges are chosen using the procedure described in Lemma 3.10. For every $t \in \mathbb{N}$, if v is the vertex born at time t, and w and l are the first sampled vertex and length of the first random walk taken, respectively, then v is joined only to w in T_t and the weight of the edge vw is set to $l + 1$. Note that the edge weights are mutually independent. Since the distance between v and w in G_t is at most $l + 1$, by induction the weighted depth of v in T_t is at most the depth of v in G_t. We show that a.a.s. the weighted height of T_n is at most $9 \cdot \log n/q$, and this completes the proof.

By Lemma 3.3, a.a.s. the (unweighted) height of T_n is less than $1.001e\log n$. We prove that any given node at depth at most $1.001e\log n$ of T_n has weighted depth at most $9(\log n)/q$ with probability $1-o(1/n)$, and then the union bound completes the proof. Let v be a node of T_n at depth h, where $h \le 1.001e\log n$. By Lemma 3.10, the weighted depth of v equals the sum of h independent random variables, each stochastically smaller than $2+\mathrm{Geo}(q)$. The probability that the sum of h independent random variables distributed as $2+\mathrm{Geo}(q)$ is greater than $9\log n/q$ equals the probability that the number of heads in a sequence of $9(\log n)/q-h$ independent biased coin flips, each having probability q of being heads, is smaller than h. This probability is

$$\mathbb{P}\left[\mathrm{Bin}\left(\frac{9\log n}{q}-h,q\right)<h\right]. \tag{3.2}$$

Since $h \le 1.001e\log n$,

$$\mu := \left(\frac{9\log n}{q}-h\right)\times q = 9\log n - hq > 6.27\log n\,,$$

and

$$\varepsilon := 1-h/\mu \ge 0.566\,,$$

hence using the lower tail Chernoff bound (2.2) we infer that the probability in (3.2) is less than $\exp\left(-0.566^2\times 6.27(\log n)/2\right) < n^{-1.004}$, as required. ■

3.2 Incorporating preferential attachment: edge trees

In this section we study models incorporating preferential attachment. We first define a model that has a lot of flexibility (Model 3.13) and prove it has logarithmic diameter. Then we reduce Models 3.17, 3.19, and 3.26 to this model.

Model 3.13. Let $(a_t,b_t)_{t\in\mathbb{N}}$ be sequences of nonnegative integers. Consider a growing undirected graph $(G_t)_{t\in\mathbb{N}_0}$ as follows. G_0 is an arbitrary connected graph with at least one edge. At each time-step $t\in\mathbb{N}$, G_t is obtained from G_{t-1} by doing a vertex operation and an edge operation, as defined below.

In a *vertex operation*, if $a_t>0$, a new vertex is born and a_t edges are added in the following manner: Sample an edge uniformly from G_{t-1}, choose one of its endpoints arbitrarily, and join it to the new vertex. For the other a_t-1 new edges, one endpoint is the new vertex, and the other endpoint is arbitrary (can be the new vertex as well). If $a_t=0$ then do nothing in the vertex operation.

Figure 3.3: the right graph is the line graph of the left graph.

In an *edge operation*, sample b_t edges such that each is uniformly random, and choose an arbitrary endpoint of each sampled edge. Then add b_t new edges, joining the chosen vertices to arbitrary vertices of G_{t-1}.

Definition 3.14 (line graph)**.** The *line graph* of a graph H is a simple graph whose vertices are the edges of H, and two edges are adjacent if they have a common endpoint.

See Figure 3.3 for an illustration.

A novel idea in this thesis is introducing edge trees: these are trees coupled with graphs whose nodes correspond to the edges of the graph. The following theorem demonstrates their usage.

Theorem 3.15. *Let $\ell = \ell(n), u = u(n) \in \mathbb{N}$ be such that $\ell \le a_t + b_t \le u$ for every $t \in [n]$. A.a.s. the graph G_n generated by Model 3.13 has diameter at most $4e(u/\ell)\log n + 8eu + O(1)$.*

Proof. We define the *depth* of an edge xy as $1 + \min\{\mathrm{depth}(x), \mathrm{depth}(y)\}$. We inductively define a growing tree $(T_t)_{t\in\mathbb{N}_0}$ such that for all $t \in \mathbb{N}_0$, $V(T_t) = E(G_t) \cup \{\aleph\}$. Here $\aleph$ denotes the root of T_t, which has depth 0. We prove by induction that for all $e \in E(G_t)$, $\mathrm{depth}(e, G_t) \le 2\,\mathrm{depth}(e, T_t)$. Let H be the graph obtained from G_0 by adding an edge labelled $\aleph$ incident to its root. Let T_0 be a breadth-first tree of the line graph of H rooted at $\aleph$. Note that $\mathrm{depth}(\aleph, T_0) = 0$ and $\mathrm{depth}(e, T_0) = \mathrm{depth}(e, G_0)$ for every $e \in E(G_0)$ (see Figure 3.4).

Given T_{t-1}, we define T_t and prove the inductive step. First, consider a vertex operation with $a_t > 0$. Let v be the new vertex, e_1 be the sampled edge, and w_1 be the chosen endpoint of e_1. Notice that $\mathrm{depth}(v, G_t) \le \mathrm{depth}(e_1, G_t) + 1$. In T_t, we join the a_t edges incident with v to e_1 (see Figure 3.5). For any such edge e we have

$$\mathrm{depth}(e, G_t) \le \mathrm{depth}(v, G_t) + 1 \le \mathrm{depth}(e_1, G_t) + 2 \le 2\,\mathrm{depth}(e_1, T_t) + 2 = 2\,\mathrm{depth}(e, T_t),$$

where we have used the inductive hypothesis for e_1 in the third inequality.

Figure 3.4: the graphs G_0 (left) and T_0 (right) in the proof of Theorem 3.15

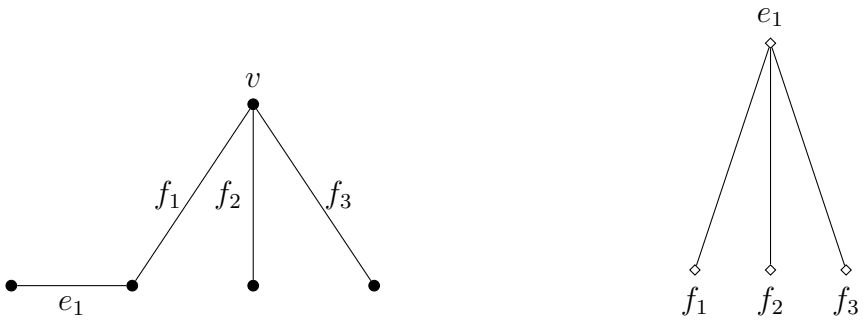


Figure 3.5: in a vertex operation say with $a_t = 3$ (left), a new vertex v is born and a uniformly random edge e_1 is chosen. Then v is joined to an endpoint of e_1 and two arbitrary vertices. In the coupled edge-tree (right), the new edges are joined to e_1.

Figure 3.6: in an edge operation say with $b_t = 3$ (left), three uniformly random edges e_1, e_2, and e_3 are sampled and an endpoint of each of them is joined to an arbitrary vertex. In the coupled edge-tree (right), the new edges w_1x_1, w_2x_2, and w_3x_3 are joined to e_1, e_2, and e_3, respectively.

Second, consider an edge operation. Let $e_1, \dots, e_{b_t}$ be the sampled edges, and let $w_1, w_2, \dots, w_{b_t}$ be the chosen endpoints. For each $j \in [b_t]$, in G_t we join w_j to some vertex of G_{t-1}, say x_j. In T_t, we join the new edge w_jx_j to e_j (see Figure 3.6). We have

$$\begin{aligned}\operatorname{depth}(w_jx_j, G_t) \le \operatorname{depth}(w_j, G_t) + 1 &\le \operatorname{depth}(e_j, G_t) + 1 \\ &\le 2\operatorname{depth}(e_j, T_t) + 1 = 2\operatorname{depth}(w_jx_j, T_t) - 1\,,\end{aligned}$$

where we have used the fact that w_j is an endpoint of e_j for the second inequality, and the inductive hypothesis for e_j for the third inequality.

Hence for all $e \in E(G_t)$, $\operatorname{depth}(e, G_t) \le 2\operatorname{depth}(e, T_t)$. On the other hand, examining the construction of $(T_t)_{t\in\mathbb{N}_0}$ and using Lemma 3.3, we find that a.a.s. the height of T_n is at most $(u/\ell)e\log n + 2ue + O(1)$. This implies that a.a.s. the diameter of G_n is at most $4(u/\ell)e\log n + 8ue + O(1)$. ■

Definition 3.16 (the function ρ_δ)**.** For an undirected graph G and a real number δ, we define the function $\rho_\delta : V(G) \to \mathbb{R}$ as

$$\rho_\delta(v) = \frac{\deg(v) + \delta}{\sum_{u\in V(G)}(\deg(u) + \delta)}\,.$$

Here $\deg(v)$ denotes the degree of vertex v, and a loop is counted twice. Note that if $\delta > -1$ and G has no isolated vertices, then ρ_δ is a probability distribution.

Observe that to sample a vertex using ρ_0, one can sample an edge uniformly and then choose one of its endpoints uniformly. Most of our arguments are based on this crucial fact, and this is the reason for introducing edge trees.

Model 3.17. Let $(x_t)_{t\in\mathbb{N}}$ be a sequence of positive integers, and let $(y_t, z_t)_{t\in\mathbb{N}}$ be sequences of nonnegative integers. Consider a growing undirected graph $(G_t)_{t\in\mathbb{N}_0}$ as follows. G_0 is an arbitrary connected graph with at least one edge. At each time-step $t \in \mathbb{N}$, G_t is obtained from G_{t-1} by performing the following three operations.

1. Sample x_t vertices $N_1, \dots, N_{x_t}$ independently using ρ_0.
2. Sample $2z_t$ vertices $W_1, W_1', W_2, W_2', \dots, W_{z_t}, W_{z_t}'$ independently using ρ_0.
3. Add a new vertex v and add the edges $vN_1, \dots, vN_{x_t}$ and $W_1W_1', \dots, W_{z_t}W_{z_t}'$. Also add y_t loops at v.

Model 3.17 is a generalization of a model defined by Aiello, Chung, and Lu [2, Section 2.1, Model D], which has bounded x_t, y_t, z_t. The following theorem implies that a.a.s. the latter model has diameter $O(\log n)$.

Theorem 3.18. *Let $\ell = \ell(n), u = u(n)$ be positive integers such that $x_t > 0$ and $\ell \le x_t + y_t + z_t \le u$ for all $t \in \mathbb{N}$. A.a.s. the diameter of G_n generated by Model 3.17 is at most $4e(u/\ell)\log n + 8eu + O(1)$.*

Proof. We claim that Model 3.17 is a special case of Model 3.13. Sampling a vertex using ρ_0 corresponds to choosing a random endpoint of a random edge. The three operations of Model 3.17 correspond to applying a vertex operation with $a_t = x_t + y_t$ and an edge operation with $b_t = z_t$. By Theorem 3.15, a.a.s. the diameter of G_n is at most $4e(u/\ell)\log n + 8eu + O(1)$. ■

Next we analyze yet another model by reducing it to Model 3.13.

Model 3.19. Let $\delta \in (-1, \infty)$, $p \in [0, 1]$ and let $(x_t)_{t\in\mathbb{N}}$ be a sequence of positive integers. Consider a growing undirected graph $(G_t)_{t\in\mathbb{N}_0}$ as follows. G_0 is an arbitrary connected graph with at least one edge. At each time-step $t \in \mathbb{N}$, apply exactly one of the following operations: operation (a) with probability p and operation (b) with probability $1 - p$.

(a) Sample x_t vertices independently using ρ_δ, then add a new vertex v and join it to the sampled vertices.

(b) Sample $2x_t$ vertices $W_1, W_1', \dots, W_{x_t}, W_{x_t}'$ independently using ρ_δ, then add the edges $W_1W_1', \dots, W_{x_t}W_{x_t}'$.

Figure 3.7: an example graph G_t (left) and its corresponding $\widehat{G}_t$ (right) with $r = s = 1$

Model 3.19 is a generalization of the *generalized linear preference model* of Bu and Towsley [28], in which x_t is a constant positive integer d for all t. Theorem 3.22 below gives that if δ is rational and nonnegative then a.a.s. the generalized linear preference model has diameter at most $(4 + 2\delta/d)e \log n + O(1)$.

Remark 3.20. Model 3.19 with $p = 1$ and x_t being a constant (independent of t and n) is called the *linear preference model*, whose diameter has been studied extensively. Assume that all vertices of the initial graph have degrees at least d and $x_t = d$ for all t, where $d \in \mathbb{N}$ is fixed. If $d = 1$ and $\delta \geq 0$, Pittel [111] showed the diameter is $\Theta(\log n)$. If $d > 1$ and $\delta \in (-d, 0)$, the diameter is $\Theta(\log \log n)$ as proved by Dommers, van der Hofstad, and Hooghiemstra [51, 117]. If $d > 1$ and $\delta = 0$, the diameter is $\Theta(\log n / \log \log n)$, see Bollobás and Riordan [20]. Finally, if $d > 1$ and $\delta > 0$, the diameter is $\Theta(\log n)$ [51, 117].

Remark 3.21. Chung and Lu [34] studied a variation of Model 3.19 with the following differences: the process is conditioned on generating a graph with no multiple edges or loops; $x_t = d$ for all t, where d may depend on n; there are two additional operations: in the first one, a uniformly random vertex is deleted, and in the second one, x_t uniformly random edges are deleted. They proved that if $d > \log^{1+\Omega(1)} n$, then a.a.s. the evolving graph has diameter $\Theta(\log n)$, where n is the number of vertices.

Theorem 3.22. *Suppose that $\delta = r/s$, where $r \in \mathbb{N}_0$ and $s \in \mathbb{N}$, and suppose that $\ell = \ell(n), u = u(n) \in \mathbb{N}$ are such that $\ell \leq x_t \leq u$ for all t. A.a.s. the diameter of G_n generated by Model 3.19 is at most $4e(u/\ell + \delta/(2\ell)) \log n + O(u)$.*

Proof. For $t \in \mathbb{N}_0$, let $\widehat{G}_t$ be the graph obtained from G_t by copying each edge $2s-1$ times, and adding r loops at each vertex (see Figure 3.7). So $\widehat{G}_t$ has $2s|E(G_t)| + r|V(G_t)|$ edges. Note that the diameters of G_t and $\widehat{G}_t$ are the same. We claim that $(\widehat{G}_t)_{t=0}^{\infty}$ is a special case of the growing graph generated by Model 3.13. First, sampling a vertex of G_{t-1} using ρ_δ corresponds to choosing a random endpoint of a random edge of $\widehat{G}_{t-1}$. Second, applying operation (a) corresponds to applying only a vertex operation with $a_t = 2sx_t + r$ (in which

$2s$ edges are added from the new vertex to each of the sampled endpoints and r loops are added at the new vertex). Third, applying operation (b) corresponds to applying only an edge operation with $b_t = 2sx_t$. Note that, although these b_t edges are not independent, one endpoint of each of them is an endpoint of a uniformly random edge of $\widehat{G}_{t-1}$, so this is indeed a valid edge operation as defined in Model 3.13. By Theorem 3.15, a.a.s the diameter of $\widehat{G}_n$ is at most $4e(u/\ell + \delta/(2\ell)) \log n + (16esu + 8er) + O(1)$, completing the proof. ■

We now analyze the preferential attachment with random initial degrees (PARID) model of Deijfen, van den Esker, van der Hofstad, and Hooghiemstra [44, Section 1.1] by reducing it to Model 3.19.

Model 3.23. Let δ be a constant and let $\{X_t : t \in \mathbb{N}\}$ be a sequence of i.i.d. $\mathbb{N}$-valued random variables such that with probability 1 we have $X_t + \delta > 0$ for all t. Consider a growing undirected graph $(G_t)_{t=0}^{\infty}$ as follows. G_0 is an arbitrary connected graph with at least one edge. At each time-step $t \in \mathbb{N}$, G_t is obtained from G_{t-1} by sampling X_t vertices $N_1, \dots, N_{X_t}$ independently using ρ_δ and adding one new vertex v and X_t new edges $vN_1, \dots, vN_{X_t}$.

Note that X_t is the (random) initial degree of the vertex born at time t. The following theorem implies that if the initial degrees' distribution in the PARID model has an exponential decay (e.g. if it is the Poisson or the geometric distribution), and δ is positive and rational, then a.a.s. the generated graph has a polylogarithmic diameter (the reason is we can take, say, $u = \log^2 n$ in the following theorem).

Theorem 3.24. *Assume that δ is a positive rational number and that $\ell = \ell(n)$ and $u = u(n)$ are positive integers such that $\mathbb{P}[X_1 \notin [\ell, u]] = o(1/n)$. A.a.s. the diameter of G_n generated by Model 3.23 is at most $4e(u/\ell + \delta/(2\ell)) \log n + O(u)$.*

Proof. Since $\mathbb{P}[X_1 \notin [\ell, u]] = o(1/n)$ and the X_i are i.i.d., a.a.s. we have $\ell \le X_t \le u$ for all $t \in [n]$. The rest of the proof is the same as that of Theorem 3.22, where all operations are of type (a). ■

3.2.1 A directed model

In this section we study a directed analogue of Model 3.17, which is also a generalization of a model of Aiello et al. [2]. Sampling probabilities in this model depend on vertices' out-degrees and in-degrees, as defined below.

Definition 3.25 (the functions ρ_δ^{out} and ρ_δ^{in})**.** For a directed graph G and a real number δ, we define the functions $\rho_\delta^{out}, \rho_\delta^{in} : V(G) \to \mathbb{R}$ as

$$\rho_\delta^{out}(v) = \frac{\text{out-deg}(v) + \delta}{\sum_{u \in V(G)} (\text{out-deg}(u) + \delta)}$$

and

$$\rho_\delta^{in}(v) = \frac{\text{in-deg}(v) + \delta}{\sum_{u \in V(G)} (\text{in-deg}(u) + \delta)} .$$

Here out-deg(v) and in-deg(v) denote the out-degree and the in-degree of vertex v, respectively.

Note that in this section we will only work with the case $\delta = 0$. The case $\delta \neq 0$ will be used in Section 3.3 in the definition of Model 3.31.

Model 3.26. Let $\{\chi_t, \gamma_t, z_t, q_t : t \in \mathbb{N}\}$ be sequences of nonnegative integers satisfying $\chi_t + \gamma_t > 0$ for all t. Consider a growing directed graph $(G_t)_{t=0}^\infty$ as follows. G_0 is an arbitrary weakly connected directed graph with at least one edge. At each time-step $t \in \mathbb{N}$, perform the following operations:

1. Sample χ_t vertices $x_1, \dots, x_{\chi_t}$ independently using ρ_0^{out} and γ_t vertices $y_1, \dots, y_{\gamma_t}$ independently using ρ_0^{in}.
2. Sample z_t vertices $w_1, w_2, \dots, w_{z_t}$ independently using ρ_0^{out}, and sample z_t vertices $w'_1, w'_2, \dots, w'_{z_t}$ independently using ρ_0^{in}.
3. Add a new vertex v, and then add the directed edges $w_1 w'_1, \dots, w_{z_t} w'_{z_t}$, $x_1 v, \dots, x_{\chi_t} v$, $v y_1, \dots, v y_{\gamma_t}$. Also add q_t loops at v.

Model 3.26 generalizes [2, Section 2.1, Model C], which has bounded $\chi_t, \gamma_t, z_t, q_t$. The following theorem implies that a.a.s. the diameter of the latter model is $O(\log n)$.

Theorem 3.27. *Let $\ell = \ell(n), u = u(n)$ be positive integers such that $\ell \leq \chi_t + \gamma_t + z_t + q_t \leq u$ for all $t \in \mathbb{N}$. A.a.s. the diameter of G_n generated by Model 3.26 is at most $4e(u/\ell) \log n + 8eu + O(1)$.*

Proof. We claim that the underlying undirected graph of $(G_t)_{t=0}^\infty$ generated by Model 3.26 is a special case of the growing graph generated by Model 3.13. Sampling a vertex using ρ_0^{out} and ρ_0^{in} correspond to choosing the tail and the head of a random edge, respectively. The operations of Model 3.26 correspond to applying a vertex operation with $a_t = \chi_t + \gamma_t + q_t$ and an edge operation with $b_t = z_t$. By Theorem 3.15, a.a.s. the diameter of G_n is at most $4e(u/\ell) \log n + 8eu + O(1)$, as required. ∎

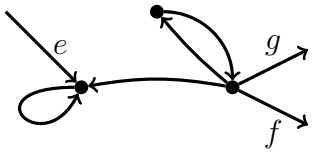


Figure 3.8: a generalized directed graph: edge e is tailless, and edges f and g are headless.

3.3 Directed scale-free graphs: dummy edges

We study two directed models in this section. In contrast to the previous directed model (Model 3.26), in models considered here, the constant term in the definition of attachment probabilities (δ in Model 3.19) can be different for in-degrees and out-degrees. We handle this issue by introducing dummy edges whose role is just to adjust the attachment probabilities (similar to, but more complicated than, what we did in the proof of Theorem 3.22). As in Section 3.2, we first define a general model (Model 3.29) with a lot of flexibility and prove that a.a.s. it has a logarithmic diameter, and then reduce Model 3.31 (which is a generalization of the so-called 'directed scale-free graphs') to that.

Definition 3.28 (Generalized directed graph)**.** In a directed graph, each edge has a tail and a head. A *generalized directed graph* is a directed graph some of whose edges do not have a head or a tail. Edges of such a graph are of three type: *tailless* edges have a head but do not have a tail, *headless* edges have a tail but do not have a head, and *proper* edges have a tail and a head. A *headed* edge is one that is not headless, and a *tailed* edges is one that is not tailless.

See Figure 3.8 for an example of a generalized directed graph.

The following model is a directed analogue of Model 3.13.

Model 3.29. Let $(a_t, b_t, c_t, d_t, \xi_t)_{t\in\mathbb{N}}$ be sequences of nonnegative integers. Consider a growing generalized directed graph $(G_t)_{t=0}^{\infty}$ as follows. G_0 is an arbitrary weakly connected generalized directed graph with at least one edge. At each time-step $t \in \mathbb{N}$, G_t is obtained from G_{t-1} by performing a vertex operation and an edge operation, as defined below.

In a *vertex operation*, if $a_t + b_t > 0$, a new vertex v is born and $a_t + b_t + c_t + d_t$ edges are added in the following manner:

Case 1: If $a_t > 0$, sample a headed edge from G_{t-1} uniformly and add a proper edge from v to its head. Then add $a_t - 1$ new proper edges, tailed at v and headed at arbitrary

vertices of G_{t-1}. Then b_t proper edges are added, tailed at arbitrary vertices of G_{t-1} and headed at v. Then c_t headless edges tailed at v, and d_t tailless edges headed at v are added.

Case 2: If $a_t = 0$, sample a tailed edge from G_{t-1} uniformly and add a proper edge from its tail to v. Then $b_t - 1$ new proper edges are added from arbitrary vertices of G_{t-1} to v. Then c_t headless edges tailed at v, and d_t tailless edges headed at v are added.

If $a_t + b_t = 0$, then do nothing in the vertex operation.

In an *edge operation*, sample ξ_t tailed edges from G_{t-1} such that each is uniformly random, then add ξ_t proper edges, joining the tails of the sampled edges to arbitrary vertices of G_{t-1}.

Theorem 3.30. *Let $\ell = \ell(n), u = u(n) \in \mathbb{N}$ be such that $\ell \le a_t + b_t + \xi_t$ and $a_t + b_t + c_t + d_t + \xi_t \le u$ for every $t \in \mathbb{N}$. A.a.s. the graph G_n generated by Model 3.29 has diameter at most $4e(u/\ell)\log n + 8eu + O(1)$.*

Proof. The argument is similar to that for Theorem 3.15. We define the depth of a headless edge as one plus the depth of its tail, and the depth of a tailless edge as one plus the depth of its head, and the depth of a proper edge uv as $1+\min\{\mathrm{depth}(u), \mathrm{depth}(v)\}$. We inductively define a growing undirected tree $(T_t)_{t=0}^{\infty}$ such that for all $t \in \mathbb{N}_0$, $V(T_t) = E(G_t) \cup \{\aleph\}$. Here $\aleph$ denotes the root of T_t, which has depth 0. We prove by induction that for all $e \in E(G_t)$, $\mathrm{depth}(e, G_t) \le 2\,\mathrm{depth}(e, T_t)$. Let H be the graph obtained from the underlying undirected graph of G_0 by adding an edge labelled $\aleph$ incident to its root. Let T_0 be a breadth-first tree of the line graph of H rooted at $\aleph$. Note that $\mathrm{depth}(\aleph, T_0) = 0$ and $\mathrm{depth}(e, T_0) = \mathrm{depth}(e, G_0)$ for every $e \in E(G_0)$ (recall Figure 3.4).

Given T_{t-1}, we define T_t and prove the inductive step. First, consider a vertex operation, Case 1. Let v be the new vertex and e_1 be the sampled headed edge. Notice that $\mathrm{depth}(v, G_t) \le \mathrm{depth}(e_1, G_t) + 1$. In T_t, we join the $a_t + b_t + c_t + d_t$ new nodes (new edges of G_t) to e_1. For any such edge e we have

$$\mathrm{depth}(e, G_t) \le \mathrm{depth}(v, G_t) + 1 \le \mathrm{depth}(e_1, G_t) + 2 \le 2\,\mathrm{depth}(e_1, T_t) + 2 = 2\,\mathrm{depth}(e, T_t),$$

where we have used the inductive hypothesis for e_1 in the third inequality.

Second, consider a vertex operation, Case 2. Let v be the new vertex and let e_1 be the sampled tailed edge. Notice that $\mathrm{depth}(v, G_t) \le \mathrm{depth}(e_1, G_t) + 1$. In T_t, we join the $b_t + c_t + d_t$ new nodes (new edges of G_t) to e_1. For any such edge e we have

$$\mathrm{depth}(e, G_t) \le \mathrm{depth}(v, G_t) + 1 \le \mathrm{depth}(e_1, G_t) + 2 \le 2\,\mathrm{depth}(e_1, T_t) + 2 = 2\,\mathrm{depth}(e, T_t).$$

Third, consider an edge operation. Let $e_1, \dots, e_{\xi_t}$ be the sampled tailed edges, and denote by $w_1, w_2, \dots, w_{\xi_t}$ their tails. For each $j \in [\xi_t]$, in G_t we join w_j to a vertex of G_{t-1}, say x_j. In T_t, we join the new node $w_j x_j$ to e_j. We have

$$\begin{aligned}\operatorname{depth}(w_j x_j, G_t) \le \operatorname{depth}(w_j, G_t) + 1 &\le \operatorname{depth}(e_j, G_t) + 1 \\ &\le 2\operatorname{depth}(e_j, T_t) + 1 = 2\operatorname{depth}(w_j x_j, T_t) - 1\,,\end{aligned}$$

where we have used the fact that w_j is incident with e_j for the second inequality, and the inductive hypothesis for e_j for the third inequality. Hence for all $e \in E(G_t)$, we have $\operatorname{depth}(e, G_t) \le 2\operatorname{depth}(e, T_t)$, as required. To complete the proof, it suffices to show that a.a.s. the height of T_n is at most $(u/\ell)e \log n + 2ue + O(1)$.

The argument is similar to that for Lemma 3.3. Note that at any time t, graph G_t has at least $|V(T_0)| + \ell t$ proper edges. Let $n_0 = |V(T_0)|$. For a given $h = h(n)$, we bound the probability that T_n has a node at depth exactly $h + n_0$. Given a sequence $1 \le t_1 < \cdots < t_h \le n$, the probability that there exists a path $w v_{t_1} v_{t_2} \dots v_{t_h}$ in T_n with $w \in V(T_0)$ and $v_{t_1}, v_{t_2}, \dots, v_{t_h} \notin V(T_0)$ such that v_{t_j} born at time t_j is at most

$$n_0 u^h \prod_{k=1}^{h} \frac{1}{n_0 + \ell \cdot (t_k - 1)}\,,$$

since there are n_0 choices for w, at most u^h choices for $(v_{t_1}, \dots, v_{t_h})$ given $(t_1, \dots, t_n)$, and for each $k = 1, \dots, h$, when v_{t_k} is born, there are at least $n_0 + \ell \cdot (t_k - 1)$ nodes available for it to join to (corresponding to the proper edges of $G_{t_k - 1}$). By the union bound, the probability that G_n has a node at depth $h + n_0$ is at most

$$\begin{aligned}n_0 u^h \sum_{1 \le t_1 < t_2 < \cdots < t_h \le n} \left(\prod_{k=1}^{h} \frac{1}{n_0 + \ell \cdot (t_k - 1)} \right) &< \frac{n_0 u^h}{h!} \left(\sum_{j=0}^{n-1} \frac{1}{n_0 + \ell j} \right)^h \\ &< n_0 \left(\frac{ue}{h} \cdot \left(\frac{\log n}{\ell} + 2 \right) \right)^h \Big/ \sqrt{2\pi h}\,.\end{aligned}$$

Putting $h \ge (u/\ell)e \log n + 2ue$ makes this probability $o(1)$. Hence a.a.s. the height of T_n is less than $(u/\ell)e \log n + 2ue + O(1)$, as required. ■

The following model is a directed analogue of Model 3.19.

Model 3.31. Let p_a, p_b, p_c be nonnegative numbers summing to 1, and let $\alpha, \beta \in [0, \infty)$. Let $(x_t)_{t \in \mathbb{N}}$ be a sequence of positive integers. Consider a growing directed graph $(G_t)_{t=0}^{\infty}$

Figure 3.9: an example graph G_t (left) and its corresponding $\widehat{G}_t$ (right) with $s = 2, r = 0$, and $q = 1$

as follows. G_0 is an arbitrary weakly connected directed graph. At each time-step $t \in \mathbb{N}$, perform exactly one of the following three operations, with probabilities p_a, p_b, and p_c, respectively.

(a) Sample x_t vertices from the existing graph, independently using ρ_α^{in}. Then add a new vertex and join it to the sampled vertices.

(b) Sample x_t vertices from the existing graph, independently using ρ_β^{out}. Then add a new vertex and join the sampled vertices to it.

(c) Sample x_t vertices $w_1, \dots, w_{x_t}$ independently using the probability distribution ρ_β^{out}, and sample x_t vertices $w'_1, \dots, w'_{x_t}$ independently using ρ_α^{in}. Then add the edges $w_1w'_1, \dots, w_{x_t}w'_{x_t}$.

Model 3.31 is a generalization of *directed scale-free graphs* of Bollobás, Borgs, Chayes, and Riordan [17, Section 2], which has $x_t = 1$ for all t. The following theorem implies that if α and β are rational, then a.a.s. the diameter of the latter model is at most $4e(1 + \alpha + \beta) \log n + O(1)$.

Theorem 3.32. *Suppose that $\alpha = r/s$ and $\beta = q/s$ with $r, q \in \mathbb{N}_0$ and $s \in \mathbb{N}$. Also suppose that $\ell = \ell(n), u = u(n) \in \mathbb{N}$ are such that $\ell \le x_t \le u$ for all t. A.a.s. the diameter of G_n generated by Model 3.31 is at most $4e(u + \alpha + \beta) \log n/\ell + O(u)$.*

Proof. For $t \in \mathbb{N}_0$, let $\widehat{G}_t$ be the generalized directed graph obtained from G_t by copying each edge $s - 1$ times, adding r tailless edge at each vertex, and adding q headless edges at each vertex (see Figure 3.9). So $\widehat{G}_t$ has $s|E(G_t)| + (r + q)|V(G_t)|$ edges. Note that the diameters of G_t and $\widehat{G}_t$ are the same. We claim that $(\widehat{G}_t)_{t=0}^\infty$ is a special case of the growing graph generated by Model 3.29. First, sampling a vertex of G_t using ρ_β^{out} or ρ_α^{in}

correspond to choosing the tail or the head of a uniformly random tailed or headed edge of $\widehat{G}_t$, respectively. Second, applying operation (a) corresponds to applying only a vertex operation with $a_t = sx_t, b_t = 0, c_t = q, d_t = r$. Third, applying operation (b) corresponds to applying only a vertex operation with $a_t = 0, b_t = sx_t, c_t = q, d_t = r$. Fourth, applying operation (c) corresponds to applying only an edge operation with $\xi_t = sx_t$. Note that although the tails of these sx_t edges are not independent, each tail is the tail of a uniformly random tailed edge of $\widehat{G}_t$, hence this is a valid edge operation as defined in Model 3.29. By Theorem 3.30, a.a.s the diameter of $\widehat{G}_n$ is at most $4e(u+\alpha+\beta)\log n/\ell+8e(su+q+r)+O(1)$, completing the proof. ■

3.4 The Cooper-Frieze model: multi-typed edge trees

In this section we study an undirected model that combines uniform and preferential attachment when choosing the neighbours of a new vertex.

Model 3.33. Let $p_a, \dots, p_f$ be nonnegative numbers summing to 1 and satisfying $p_a+p_b > 0$, and let $(x_t)_{t\in\mathbb{N}}$ be a sequence of positive integers. Consider a growing undirected graph $(G_t)_{t=0}^{\infty}$ as follows. G_0 is an arbitrary connected graph. At each time-step $t \in \mathbb{N}$, perform exactly one of the following six operations, with probabilities $p_a, \dots, p_f$ and independently of previous choices.

(a) x_t vertices are sampled uniformly, then a new vertex is born and is joined to the sampled vertices.

(b) x_t vertices are sampled using ρ_0, then a new vertex is born and is joined to the sampled vertices.

(c) x_t+1 vertices are sampled uniformly. Then x_t edges are added joining the first sampled vertex to the others.

(d) A vertex is sampled uniformly and x_t vertices are sampled using ρ_0. Then x_t edges are added joining the first sampled vertex to the others.

(e) A vertex is sampled using ρ_0 and x_t vertices are sampled uniformly. Then x_t edges are added joining the first sampled vertex to the others.

(f) x_t+1 vertices are sampled using ρ_0. Then x_t edges are added joining the first sampled vertex to the others.

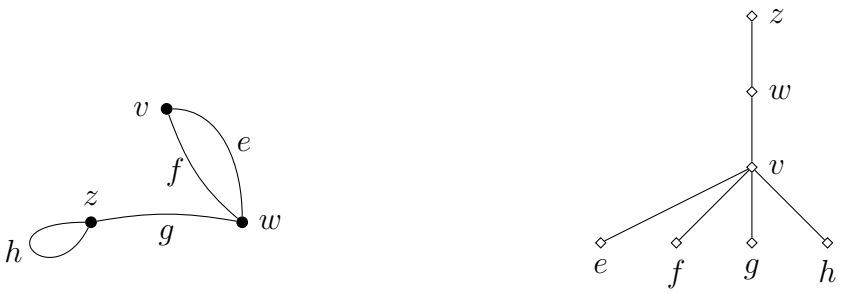


Figure 3.10: the graphs G_0 (left, rooted at z) and T_0 (right) in the proof of Theorem 3.34

Note that each operation increases the number of edges by x_t.

Model 3.33 is a generalization of a model defined by Cooper and Frieze [38, Section 2], in which $x_1, x_2, \dots$ are i.i.d. bounded random variables. The following theorem implies that a.a.s. the diameter of the latter model is $O(\log n)$.

Theorem 3.34. *Let $q = p_a + p_b$ and let $\ell = \ell(n), u = u(n)$ be positive integers such that $\ell \le x_t \le u$ for all t. A.a.s. the diameter of G_n generated by Model 3.33 is at most $4(u/\ell + 11/q)e \log n + 8eu + O(1)$.*

Proof. As before, we define a growing tree whose height multiplied by 2 dominates the height of the graph at any moment, and then we bound the tree's height from above. The main difference with Theorem 3.15 is that in some operations one may sample the *vertices* of the graph uniformly. In a growing tree, when a new vertex v is born and is joined to a vertex w of the existing tree, we say w is the *parent* of v, and that w is *given birth* to v.

We inductively define a growing tree $(T_t)_{t=0}^{\infty}$ such that $V(T_t) = V(G_t) \cup E(G_t)$ for all t, and we prove that $\operatorname{depth}(f, G_t) \le 2 \operatorname{depth}(f, T_t)$ for each vertex or edge f of G_t. A node of T_t is called a *V-node* or an *E-node* if it corresponds to a vertex or an edge of G_t, respectively. We may assume T_0 has been defined (for instance, we can build it by taking a breadth-first search tree of G_0 and joining all the E-nodes to its deepest V-node, see Figure 3.10) and we describe the growth of T_{t-1} to T_t corresponding to each operation.

(a) Let w be the first sampled vertex. In T_t we join all new nodes (corresponding to the new vertex and the new edges in G_t) to w. In this case, a V-node of T_{t-1} has been sampled uniformly and has given birth to one V-node and x_t E-nodes.

(b) For sampling a vertex using ρ_0, we sample a random edge and then choose a random endpoint of it. Let e be the first sampled edge. In T_t we join all new nodes (corresponding to the new vertex and the new edges in G_t) to e. In this case, an E-node of T_{t-1} has been sampled uniformly and is given birth to one V-node and x_t E-nodes.

(c) and (d) Let w be the first sampled vertex. In T_t we join all new nodes (corresponding to the new edges in G_t) to w. In this case, a V-node of T_{t-1} has been sampled uniformly and is given birth to x_t E-nodes.

(e) and (f) For sampling a vertex using ρ_0, we sample a random edge and then choose a random endpoint of it. Let e be the first sampled edge. In T_t we join all new nodes (corresponding to the new edges in G_t) to e. In this case, an E-node of T_{t-1} has been sampled uniformly and is given birth to x_t E-nodes.

Similar to the proofs of Theorems 3.15 and 3.30, an inductive argument gives that $\operatorname{depth}(f, G_t) \le 2\operatorname{depth}(f, T_t)$ for each vertex or edge f of G_t. Hence, showing that a.a.s. the height of T_n is at most $(u/\ell + 11/q)e \log n + 2eu + O(1)$ completes the proof.

For $t \in \mathbb{N}_0$, let $L(t)$ denote the number of V-nodes of T_t. Let $n_0 = |V(T_0)|$ and $m_0 = (9/q)\log n$. Note that $L(t) = n_0 + \operatorname{Bin}(t, q)$. Using the lower tail Chernoff bound (2.2) and the union bound, a.a.s we have $L(t) \ge tq/2$ for all $m_0 \le t \le n$. We condition on an arbitrary vector $(L(0), \dots, L(n)) = (g(0), \dots, g(n))$ for which this event happens.

For a given integer $h = h(n)$, we bound the probability that T_n has a vertex at depth exactly $n_0 + h$. Given a sequence $1 \le t_1 < \cdots < t_h \le n$, the probability that there exists a path $wv_{t_1}v_{t_2}\dots v_{t_h}$ in T_n with $w \in V(T_0)$ and $v_{t_1}, v_{t_2}, \dots, v_{t_h} \notin V(T_0)$ such that v_{t_j} is born at time t_j is at most

$$n_0 \prod_{k=1}^{h} \left(\frac{u}{n_0 + \ell \cdot (t_k - 1)} + \frac{1}{g(t_k - 1)} \right) ,$$

since there n_0 choices for w and for each $k = h, h-1, \dots, 2, 1$, if v_{t_k} wants to choose an E-node as its parent, there are at least $n_0 + \ell \cdot (t_k - 1)$ many E-nodes available for it to join to, and at most u of them were born at time t_{k-1}; and if v_{t_k} wants to choose a V-node as its parent, there are at least $g(t_k - 1)$ many V-nodes available for it to join to, and at most one of them was born at time t_{k-1}. By the union bound and (3.1), the probability that T_n has a vertex at depth $h + n_0$ is at most

$$n_0 \sum_{1 \le t_1 < \cdots < t_h \le n} \left(\prod_{k=1}^{h} \left(\frac{u}{n_0 + \ell \cdot (t_k - 1)} + \frac{1}{g(t_k - 1)} \right) \right) < \frac{n_0}{h!} \left(\sum_{j=0}^{n-1} \frac{u}{n_0 + \ell j} + \sum_{j=0}^{n-1} \frac{1}{g(j)} \right)^h . \tag{3.3}$$

We have

$$\sum_{j=0}^{n-1} \frac{u}{n_0+\ell j} < u + \frac{u}{\ell}\sum_{j=1}^{n-1}\frac{1}{j} < u + (u/\ell)(1+\log n)\,,$$

and

$$\sum_{j=0}^{n-1}\frac{1}{g(j)} = \sum_{j=0}^{m_0-1}\frac{1}{g(j)} + \sum_{j=m_0}^{n-1}\frac{1}{g(j)} \le m_0 + \sum_{j=m_0}^{n-1}\frac{2}{qj} < \frac{11}{q}\,\log n\,.$$

Setting $h \ge (u/\ell + 11/q)e\log n + 2eu$ makes the right hand side of (3.3) become $o(1)$, as required. ■

3.5 Further models

In this section we mention two closely related models for which we can easily prove logarithmic bounds using the technique developed in this chapter. Let $k > 1$ be a positive integer. A *random k-tree*, defined by Gao [71], is built from a k-clique by applying the following operation n times: in every time-step, a k-clique of the existing graph is chosen uniformly at random, a new vertex is born and is joined to all vertices of the chosen k-clique. See Figure 1.2 for an illustration with $k = 2$. (We remark that this process is different from the random k-tree process defined by Cooper and Uehara [43] which was further studied in [41].) *Random k-Apollonian networks* [119] have a similar construction, the only difference being that once a k-clique is chosen in some time-step, it will never be chosen in the future. Cooper, Frieze and Uehara [41, Theorem 2] and independently, Kolossváry, Komjáty and Vágó [87, Theorem 2.2] have recently proved that a.a.s. the diameter of a random k-Apollonian network is asymptotic to $c_k \log n$ for a given constant c_k.

Here we prove that a.a.s. the diameter of a random k-tree is at most $2e\log n + O(1)$, and that a.a.s. the diameter of a random k-Apollonian network is at most $2ek\log n/(k-1) + O(1)$. For the proof for random k-Apollonian networks we need the following variant of Lemma 3.3. For a tree T, denote its set of leaves by $\mathcal{L}(T)$.

Lemma 3.35. *Let $(a_t)_{t\in\mathbb{N}}$ be a sequence of positive integers. Consider a growing tree $(T_t)_{t=0}^{\infty}$ as follows. T_0 is arbitrary. At each time-step $t \in \mathbb{N}$, a random vector $(W_1, W_2, \dots, W_{a_t}) \in \mathcal{L}(T_{t-1})^{a_t}$ is chosen in such a way that for each $i \in [a_t]$ and each $v \in \mathcal{L}(T_{t-1})$, the marginal probability $\mathbb{P}[W_i = v]$ equals $|\mathcal{L}(T_{t-1})|^{-1}$. In other words, each W_i is a uniformly random leaf of T_{t-1}; however, the W_j's may be correlated. Then a_t new nodes $v_1, \dots, v_{a_t}$ are born and v_i is joined to W_i for each $i \in [a_t]$. Let $\ell = \ell(n)$ and $u = u(n)$ be positive*

integers such that $1 < \ell \le a_t \le u$ *for all* $t \in [n]$. *Then the height of* T_n *is a.a.s. at most* $ue \log n/(\ell - 1) + 2ue + O(1)$.

Proof. Let $n_0 = |V(T_0)|$. For a given integer $h = h(n)$, let us bound the probability that T_n has a node at depth exactly $h + n_0$. Given a sequence $1 \le t_1 < t_2 < \cdots < t_h \le n$, the probability that there exists a path $wv_{t_1}v_{t_2}\cdots v_{t_h}$ in T_n with $w \in V(T_0)$ and $v_{t_1}, \ldots, v_{t_n} \notin V(T_0)$ such that v_{t_j} is born at time t_j is at most

$$n_0 u^h \prod_{k=1}^{h} \frac{1}{n_0 + (\ell - 1)(t_k - 1)} ,$$

since for each $k = 1, 2, \ldots, h$, when v_{t_k} is born, there are at least $n_0 + (\ell - 1)(t_k - 1)$ leaves available for it to join to. By the union bound and (3.1), the probability that T_n has a node at depth $h + n_0$ is at most

$$\begin{aligned} n_0 u^h \sum_{1 \le t_1 < \cdots < t_h \le n} \left(\prod_{k=1}^{h} \frac{1}{n_0 + (\ell - 1)(t_k - 1)} \right) &< \frac{n_0 u^h}{h!} \left(\sum_{j=0}^{n-1} \frac{1}{n_0 + (\ell - 1)j} \right)^h \\ &< \left(\frac{ue}{h} \cdot \left(\frac{\log n}{\ell - 1} + 2 \right) \right)^h \Big/ \sqrt{2\pi h} \, . \end{aligned}$$

Putting $h \ge ue \log n/(\ell - 1) + 2ue$ makes this probability $o(1)$. Hence a.a.s. the height of T_n is at most $ue \log n/(\ell - 1) + 2ue + n_0$, as required. ■

Theorem 3.36. *A.a.s. the diameter of an* $(n+k)$*-vertex random* k*-tree is at most* $2e \log n + O(1)$*, and the diameter of an* $(n + k)$*-vertex random* k*-Apollonian network is at most* $2ek \log n/(k - 1) + O(1)$.

Proof. We define the *depth* of a k-clique as the maximum depth of its vertices. Let the first k vertices have depth zero. We couple with a growing tree whose nodes corresponds to the k-cliques of the growing graph. Whenever in the graph a new vertex is born and is joined to the vertices of a k-clique, in the tree the chosen k-clique gives birth to k new children. By induction, the graph's height is always less than or equal to the tree's height.

For the tree corresponding to a random k-tree, in every step a node is chosen uniformly at random and gives birth to k new children, hence a.a.s. its height is bounded by $e \log n + O(1)$ by Lemma 3.3. This gives an a.a.s. upper bound of $2e \log n + O(1)$ for the diameter of the corresponding graph.

For the tree corresponding to a random k-Apollonian network, in every step a leaf is chosen uniformly at random and gives birth to k new children, hence a.a.s. its height is bounded by $ek \log n/(k-1) + O(1)$ by Lemma 3.35. This gives an a.a.s. upper bound of $2ek \log n/(k-1) + O(1)$ for the diameter of the corresponding graph. ■

The proof of Theorem 3.36 indeed implies the following more general results, which will be used in Chapter 7.

Proposition 3.37. *A.a.s. an $(n+k)$-vertex random k-tree has the following property: let $u_h u_{h-1} \cdots u_0$ be any path such that u_i is born later than u_{i-1} for all i; then $h \leq e \log n + O(1)$.*

Proposition 3.38. *A.a.s. an $(n+k)$-vertex random k-Apollonian network has the following property: let $u_h u_{h-1} \cdots u_0$ be any path such that u_i is born later than u_{i-1} for all i; then $h \leq ek \log n/(k-1) + O(1)$.*

Chapter 4

Random Apollonian networks: diameter and longest paths

We have already encountered 'random k-Apollonian networks' in Chapter 3, where we proved an a.a.s. upper bound of $2ek \log n/(k-1) + O(1)$ for their diameter. In this chapter[1] we study random Apollonian networks, which are planar embeddings of random 3-Apollonian networks. Historically, this subclass was defined first; and later it was generalized to higher dimensions. Random Apollonian networks are a popular random graph model for generating planar graphs with power law properties, and can be defined as follows. Start with a triangle embedded in the plane. In each step, choose a bounded face uniformly at random, add a vertex inside that face and join it to the vertices on the face. We call this operation *subdividing* the face. In this chapter, we use the term 'face' to refer to a bounded face, unless specified otherwise. After $n-3$ steps, we have a (random) triangulated plane graph with n vertices and $2n-5$ faces. This is called a *Random Apollonian Network (RAN)* and we study its asymptotic properties, as its number of vertices goes to infinity. The number of edges equals $3n-6$, and hence a RAN is a maximal plane graph. See Figure 4.1 for an illustration.

The term 'Apollonian network' refers to a deterministic version of this process, formed by subdividing all triangles the same number of times, which was first studied in [6, 53]. Andrade, Herrmann, Andrade, and Silva [6] studied power laws in the degree sequences of these networks. Random Apollonian networks were defined by Zhou, Yan and Wang [120]

[1] This chapter is based on joint work with Collevecchio, Ebrahimzadeh, Farczadi, Gao, Sato, Wormald, and Zung. Some of the results therein have already been published [57], and the rest appear in the submitted preprint [37].

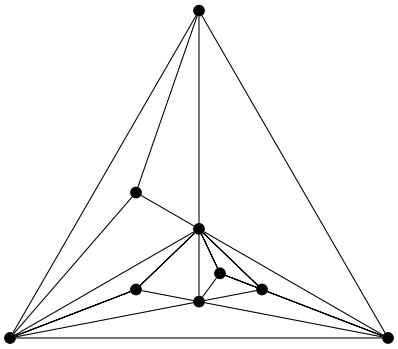


Figure 4.1: a RAN with nine vertices

(see Zhang, Comellas, Fertin and Rong [119] for a generalization to higher dimensions), where it was proved that the diameter of a RAN is a.a.s. $O(\log n)$. It was shown in [120, 104] that RANs exhibit a power law degree distribution. The above proofs were non-rigorous from a mathematical point of view. The average diameter of a RAN was shown to be $\Theta(\log n)$ by Albenque and Marckert [3]. The degree distribution, k largest degrees and k largest eigenvalues of the adjacency matrix (for fixed k) and the diameter were studied by Frieze and Tsourakakis [70].

In this chapter we continue this line of research by studying the asymptotic properties of the longest (simple) paths and cycles in RANs and giving sharp estimates for the diameter of a typical RAN. In Section 4.1 we prove the following result about the diameter.

Theorem 4.1. *A.a.s. the diameter of a RAN on n vertices is asymptotic to $c \log n$, with $c = (1 - \hat{x}^{-1})/\log h(\hat{x}) \approx 1.668$, where*

$$h(x) = \frac{12x^3}{1-2x} - \frac{6x^3}{1-x}\,,$$

and $\hat{x} \approx 0.163$ is the unique solution in the interval $(0.1, 0.2)$ to

$$x(x-1)h'(x) = h(x)\log h(x)\,.$$

We give a high level sketch of the proof of Theorem 4.1. Using some graph theoretic observations, we reduce the problem to estimating the height of a multi-typed random tree. We embed this random tree into a continuous-time branching process using the idea of poissonization (see Section 2.4 for more explanation on this). Broutin and Devroye [27] have developed a powerful technique for analyzing the height of random trees. However,

their theorem requires all vertices in the tree to be of the same type. Using a sneaky coupling we 'sandwich' the height of our tree between the heights of two sequences of complicated random trees. Using [27, Theorem 1] (see also Theorem 4.12) we can estimate the height of these random trees. This theorem gives the height implicitly, as the solution of an equation involving maximizing a certain function, and it takes some real analysis arguments to solve the equation for our trees and giving lower and upper bounds for our original multi-typed tree. More real analysis shows that, fortunately, the lower and upper bounds converge as we refine our coupling, and the limit, which we find as the solution of an implicit equation, is the answer we are seeking.

Let $\mathcal{L}_m$ and $\mathcal{C}_m$ be random variables denoting the number of vertices in a longest path and a longest cycle in a RAN with m faces, respectively. In this chapter we also prove lower bounds for $\mathcal{C}_m$ and $\mathbb{E}[\mathcal{C}_m]$, which immediately yield lower bounds for $\mathcal{L}_m$ and $\mathbb{E}[\mathcal{L}_m]$, noting that $\mathcal{L}_m \geq \mathcal{C}_m$ always.

Theorem 4.2. *For every positive integer m, the following statements are true.*

(a) $\mathcal{L}_m \geq \mathcal{C}_m \geq m^{\log 2/\log 3} + 2$.

(b) $\mathbb{E}[\mathcal{L}_m] \geq \mathbb{E}[\mathcal{C}_m] = \Omega(m^{0.88})$.

Remark 4.3. Using a rather involved argument, Chen and Yu [32, Corollary 3.5] have shown that every 3-connected planar graph on n vertices has a cycle with at least $n^{\log 2/\log 3}$ vertices. By Theorem 4.2(a), every n-vertex random Apollonian network has a cycle with at least $(2n-5)^{\log 2/\log 3}+2$ vertices deterministically, which gives a slightly better result for this subclass, having the same exponent but a larger constant. The example studied in [32, Section 2] shows that the exponent $\log 2/\log 3$ here is the best possible for a deterministic lower bound.

The proof of Theorem 4.2 appears in Section 4.2. The deterministic bound follows from a novel graph theoretic observation and simple induction. For the expected value, we analyze the number of vertices in the triangles of a RAN by modelling the growth of the RAN as a trichromatic Eggenberger-Pólya urn and using the known limiting joint distribution of the proportion of balls of each colour in such an urn (see Theorem 2.10).

Regarding upper bounds, we prove the following result.

Theorem 4.4. *Let $\delta = 1 - 4 \times 10^{-8} = 0.99999996$. A.a.s. as $m \to \infty$ we have $\mathcal{L}_m < m^{\delta}$.*

Open problem 4.5. There is gap between our lower bound in Theorem 4.2 and our upper bound in Theorem 4.4. What is the typical order of magnitude of $\mathcal{L}_m$? Is this variable concentrated around its mean?

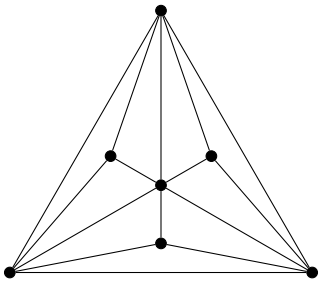


Figure 4.2: the standard 2-subdivision of a triangle

We give a high-level sketch of the proof of Theorem 4.4, which appears in Section 4.3. Using a novel graph theoretic observation, we show it is enough to prove that a.a.s. any *deficient* subtree in an n-vertex random recursive ternary tree has less than n^δ vertices, for some fixed $\delta < 1$. Here, a subtree is called deficient if any vertex has at most eight grandchildren (out of the nine possible). The branch sizes in a random recursive ternary tree are related to the number of balls of different colours in a trichromatic Eggenberger-Pólya urn (see Section 2.3 for the definition). We use de Finetti's theorem to represent these numbers as mixtures of binomial and beta random variables (see Proposition 2.9), allowing us to work in the realm of continuous random variables and consider infinite ternary trees, making things more convenient. A few tricks and sharp concentration bounds for the product of independent beta random variables (Lemma 2.23) complete the proof. The calculations are delicate here, and they have to be so, since we essentially take the union bound over an exponentially large set of objects: the set of vertices at a fixed depth of the tree.

We include some definitions here. Let $\triangle$ be a triangle in a RAN. The *standard 1-subdivision* of $\triangle$ is the set of three triangles obtained from subdividing $\triangle$ once. For $k > 1$, the *standard k-subdivision* of $\triangle$ is the set of triangles obtained from subdividing each triangle in the standard $(k-1)$-subdivision of $\triangle$ exactly once. In Figure 4.2, the standard 2-subdivision of a triangle is illustrated.

Definition 4.6 ($\triangle$-tree of a RAN)**.** Let G be a RAN. We denote the vertices incident with the unbounded face by ν_1, ν_2, ν_3. We define a rooted tree T, called the *$\triangle$-tree* of G, as follows. There is a one to one correspondence between the triangles in G and the nodes of T. For every triangle $\triangle$ in G, we denote its corresponding node in T by $\mathbf{n}^\triangle$. To build T, start with a single root node, which corresponds to the triangle $\nu_1\nu_2\nu_3$ of G. Wherever a triangle $\triangle$ is subdivided into triangles $\triangle_1$, $\triangle_2$, and $\triangle_3$, generate three children $\mathbf{n}^{\triangle_1}$, $\mathbf{n}^{\triangle_2}$, and $\mathbf{n}^{\triangle_3}$ for $\mathbf{n}^\triangle$, and extend the correspondence in the natural manner. Note that this is a ternary tree, with each node having either zero or three children, and has $3n - 8$ nodes

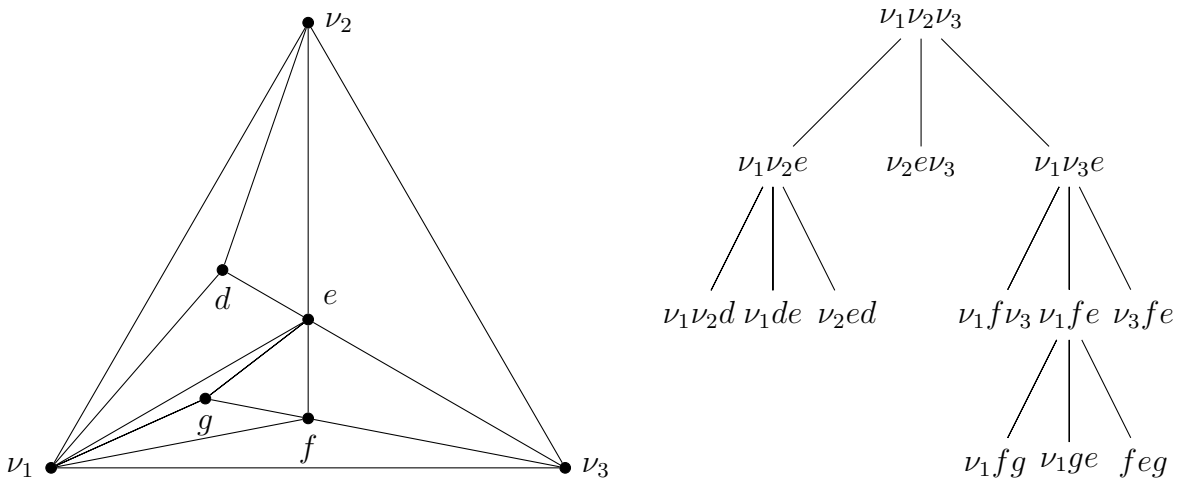


Figure 4.3: a RAN (left) and its corresponding $\triangle$-tree (right)

and $2n-5$ leaves when G has n vertices. We use the term 'nodes' for the vertices of T, so that 'vertices' refer to the vertices of G. Note that the leaves of T correspond to the faces of G. The *depth* of a node $\mathbf{n}^{\triangle}$ is its distance to the root. The set of *grandchildren* of a node is the set of children of its children. See Figure 4.3 for an illustration.

4.1 The diameter

In this section we prove Theorem 4.1. We first prove an important connection with Eggenberger-Pólya urns (defined in Section 2.3).

Proposition 4.7. *Consider a triangle $\triangle$ containing more than one face in a RAN, and let $\triangle_1, \triangle_2, \triangle_3$ be the three triangles in its standard 1-subdivision. After k subdivisions of $\triangle$, the number of faces in $\triangle_i$ is distributed like* $\mathrm{Urn}(1,2,2,k-1)$ *for $i = 1,2,3$.*

Proof. We can analyze the number of faces inside $\triangle_1$ by modelling the process of building the RAN as an Eggenberger-Pólya urn: after the first subdivision of $\triangle$, each of $\triangle_1$, $\triangle_2$, and $\triangle_3$ contains exactly one face. We start with one white ball corresponding to the only face in $\triangle_1$, and two black balls corresponding to the two faces in $\triangle_2$ and $\triangle_3$. In each subsequent step, we choose a face uniformly at random, and subdivide it. If the face is in $\triangle_1$, then the number of faces in $\triangle_1$ increases by 2, and otherwise the number of faces not

in $\triangle_1$ increases by 2. Thus after k subdivisions of $\triangle$, the number of faces in $\triangle_1$ has the same distribution as the number of white balls in an Eggenberger-Pólya urn with $w_0 = 1$, $b_0 = 2$, and reinforcement $s = 2$, after $k - 1$ draws, as required. ■

Throughout this section let G be a RAN with n vertices, and recall that ν_1, ν_2, and ν_3 denote the vertices incident with the unbounded face. For a vertex v of G, let $\tau(v)$ be the minimum distance of v to the boundary, i.e.,

$$\tau(v) = \min\{\mathrm{dist}(v, \nu_1), \mathrm{dist}(v, \nu_2), \mathrm{dist}(v, \nu_3)\} .$$

The *radius* of G is defined as the maximum of $\tau(v)$ over all vertices v.

Lemma 4.8. *Let*

$$h(x) = \frac{12x^3}{1-2x} - \frac{6x^3}{1-x} ,$$

and let $\hat{x}$ be the unique solution in $(0.1, 0.2)$ to

$$x(x-1)h'(x) = h(x)\log h(x) .$$

Finally, let

$$c = \frac{1 - \hat{x}^{-1}}{\log h(\hat{x})} \approx 1.668 .$$

Then the radius of G is a.a.s. asymptotic to $c \log n/2$.

We first show that this lemma implies Theorem 4.1.

Proof of Theorem 4.1. Let $\mathrm{diam}(G)$ denote the diameter of G. Fix arbitrarily small $\varepsilon, \delta > 0$. We show that with probability at least $1 - 2\delta$ we have

$$(1-\varepsilon)c\log n \le \mathrm{diam}(G) \le (1+\varepsilon)c\log n . \tag{4.1}$$

We prove that each of the inequalities in (4.1) holds with probability at least $1 - \delta$ and then apply the union bound. Here and in the following, we assume n is sufficiently large.

For the upper bound, let R be the radius of G. Notice that the distance between any vertex and ν_1 is at most $R + 1$, so $\mathrm{diam}(G) \le 2R + 2$. By Lemma 4.8, with probability at least $1 - \delta$ we have $R \le (1 + \varepsilon/2)c \log n/2$. If this event happens, then $\mathrm{diam}(G) \le (1 + \varepsilon)c \log n$.

We now articulate the proof for the lower bound, which is more involved. Let $\triangle_1$, $\triangle_2$, and $\triangle_3$ be the three triangles in the standard 1-subdivision of the triangle $\nu_1\nu_2\nu_3$, and let n_i be the number of vertices on and inside $\triangle_i$. By Proposition 4.7 we have

$$2n_i - 5 \stackrel{d}{=} \mathrm{Urn}(1, 2, 2, n-4) \quad \text{for } i = 1, 2, 3\,. \tag{4.2}$$

Let M be a positive integer sufficiently large that

$$\mathbb{P}\left[\frac{n_i}{n} < \frac{1}{M}\right] < \delta/6 \quad \text{for } i = 1, 2, 3\,.$$

Such an M exists since (4.2) holds (see Theorem 2.8, the proportional distribution of balls of a given colour in an Eggenberger-Pólya urn). Let A denote the event

$$\min\left\{\frac{n_i}{n} : 1 \le i \le 3\right\} \ge \frac{1}{M}\,.$$

By the union bound, $\mathbb{P}[A] \ge 1 - \delta/2$. We condition on values (n_1, n_2, n_3) such that A happens. Note that we have $\log n_i = \log n - O(1)$ for each i.

For a triangle $\triangle$, $V(\triangle)$ denotes the three vertices of $\triangle$. Note that for $i = 1, 2, 3$, the subgraph induced by vertices on and inside $\triangle_i$ is distributed as a RAN G_i with n_i vertices. Hence by Lemma 4.8 and the union bound, with probability at least $1 - \delta/2$, the radius of each of G_1, G_2 and G_3 is at least $(1-\varepsilon)c\log n/2$. Hence, with probability at least $1 - \delta/2$ there exists $u_1 \in V(G_1)$ with distance at least $(1-\varepsilon)c\log n/2$ to $V(\triangle_1)$, and also there exists $u_2 \in V(G_2)$ with distance at least $(1-\varepsilon)c\log n/2$ to $V(\triangle_2)$. Since any (u_1, u_2)-path must contain a vertex from $V(\triangle_1)$ and $V(\triangle_2)$ (see Figure 4.4), with probability at least $1 - \delta/2$, there exists $u_1, u_2 \in V(G)$ with distance at least $2(1-\varepsilon)c\log n/2$, which implies

$$\mathbb{P}[\mathrm{diam}(G) \ge c(1-\varepsilon)\log n] \ge \mathbb{P}[\mathrm{diam}(G) \ge c(1-\varepsilon)\log n | A]\,\mathbb{P}[A] \ge (1-\delta/2)^2 > 1-\delta. \quad \blacksquare$$

The rest of this section is devoted to the proof of Lemma 4.8. Let T be the $\triangle$-tree of G. We categorize the triangles in G into three types. Let $\triangle$ be a triangle in G with vertex set $\{x, y, z\}$, and assume that $\tau(x) \le \tau(y) \le \tau(z)$. Since z and x are adjacent, we have $\tau(z) \le \tau(x) + 1$. So, $\triangle$ can be categorized to be of one of the following types:

1. if $\tau(x) = \tau(y) = \tau(z)$, then say $\triangle$ is of type 1.
2. If $\tau(x) = \tau(y) < \tau(y) + 1 = \tau(z)$, then say $\triangle$ is of type 2.

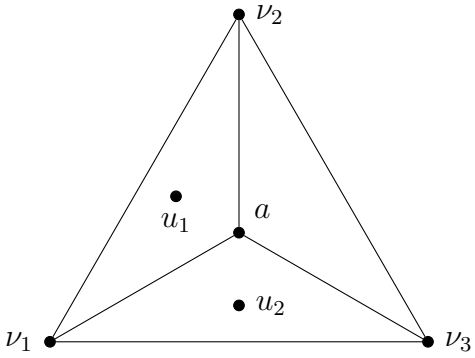


Figure 4.4: any (u_1, u_2)-path must contain a vertex from $\{\nu_1, \nu_2, a\}$ and from $\{a, \nu_1, \nu_3\}$.

3. If $\tau(x) < \tau(x) + 1 = \tau(y) = \tau(z)$, then say $\triangle$ is of type 3.

The type of a node of T is the same as the type of its corresponding triangle. The root of T corresponds to the triangle $\nu_1\nu_2\nu_3$ and the following are easy to observe.

(a) The root is of type 1.

(b) A node of type 1 has three children of type 2.

(c) A node of type 2 has one child of type 2 and two children of type 3.

(d) A node of type 3 has two children of type 3 and one child of type 1.

See Figure 4.5. For a triangle $\triangle$, define $\tau(\triangle)$ to be the minimum of $\tau(u)$ over all $u \in V(\triangle)$. Observe that a node of type 1 or 2 has children with the same τ value as itself; whereas for a node of type 3, the τ value of the type-1 child equals the τ value of the parent plus one, and the two other children have the same τ value as the parent.

Let $\overline{\triangle}$ and $\triangle$ be two triangles of type 1 such that $\mathbf{n}^{\overline{\triangle}}$ is an ancestor of $\mathbf{n}^{\triangle}$ and there is no node of type 1 in the unique path connecting them. Then, the internal vertices of the path connecting $\mathbf{n}^{\overline{\triangle}}$ and $\mathbf{n}^{\triangle}$ consists of a sequence of type-2 nodes and then a sequence of type-3 nodes, hence we have $\tau(\triangle) = \tau(\overline{\triangle}) + 1$ (see Figure 4.5). This determines τ inductively: for every $\mathbf{n}^{\triangle} \in V(T)$, $\tau(\triangle)$ is one less than the number of nodes of type 1 in the path from $\mathbf{n}^{\triangle}$ to the root. We call $\tau(\triangle)$ the *auxiliary depth* of node $\triangle$, and define the *auxiliary height* of a tree T, written $\mathrm{ah}(T)$, to be the maximum auxiliary depth of its nodes. Note that the auxiliary height is always less than or equal to the height. Also, for a

vertex $v \in V(G)$, if $\triangle$ is the triangle that v subdivides, then $\tau(v) = \tau(\triangle)+1$. We augment the tree T by adding specification of the type of each node, and we abuse notation and call the augmented tree the $\triangle$-tree of the RAN. Hence, the radius of the RAN is either $\mathrm{ah}(T)$ or $\mathrm{ah}(T)+1$.

Notice that instead of building T from the RAN G, we can think of the random T as being generated in the following manner: let $n \geq 3$ be a positive integer. Start with a single node as the root of T. So long as the number of nodes is less than $3n-8$, choose a leaf v independently of previous choices and uniformly at random, and add three leaves as children of v. Once the number of nodes becomes $3n-8$, add the information about the types using rules (a)–(d), as follows. Let the root have type 1, and determine the types of other nodes in a top-down manner. For a node of type 1, let its children have type 2. For a node of type 2, select one of the children independently and uniformly at random, let that child have type 2, and let the other two children have type 3. Similarly, for a node of type 3, select one of the children independently of previous choices and uniformly at random, let that child have type 1, and let the other two children have type 3. Henceforth, we will forget about G and focus on finding the auxiliary height of a random tree T generated in this manner.

A major difficulty in analyzing the auxiliary height of the tree generated in the aforementioned manner is that the branches of a node are heavily dependent, as the total number of nodes equals $3n-8$. To remedy this we use the idea of poissonization as explained in Section 2.4.1: we consider another process which has the desired independence and approximates the original process well enough for our purposes. The process, $\widehat{P}$, starts with a single node, the root, which is born at time 0, and is of type 1. From this moment onwards, whenever a node is born (say at time κ), it waits for a random time X, which is distributed exponentially with mean 1, and after time X has passed (namely, at absolute time $\kappa+X$) gives birth to three children, whose types are determined as before (according to the rules (b)–(d)) and dies. Moreover, the lifetime of the nodes are independent. For a nonnegative (possibly random) t, we denote by $\widehat{T}^t$ the random tree obtained by taking a snapshot of this process at time t. By the discussion in Section 2.4.1, for any deterministic $t \geq 0$, the distribution of $\widehat{T}^t$ conditional on $\widehat{T}^t$ having exactly $3n-8$ nodes, is the same as the distribution of T.

Lemma 4.9. *Assume that there exists a constant c such that a.a.s. the auxiliary height of $\widehat{T}^t$ is asymptotic to ct as $t \to \infty$. Then the radius of a RAN with n vertices is a.a.s. asymptotic to $c \log n/2$ as $n \to \infty$.*

Proof. Let $\ell_n = 3n-8$, and let $\varepsilon > 0$ be fixed. For the process $\widehat{P}$, we define three stopping times as follows:

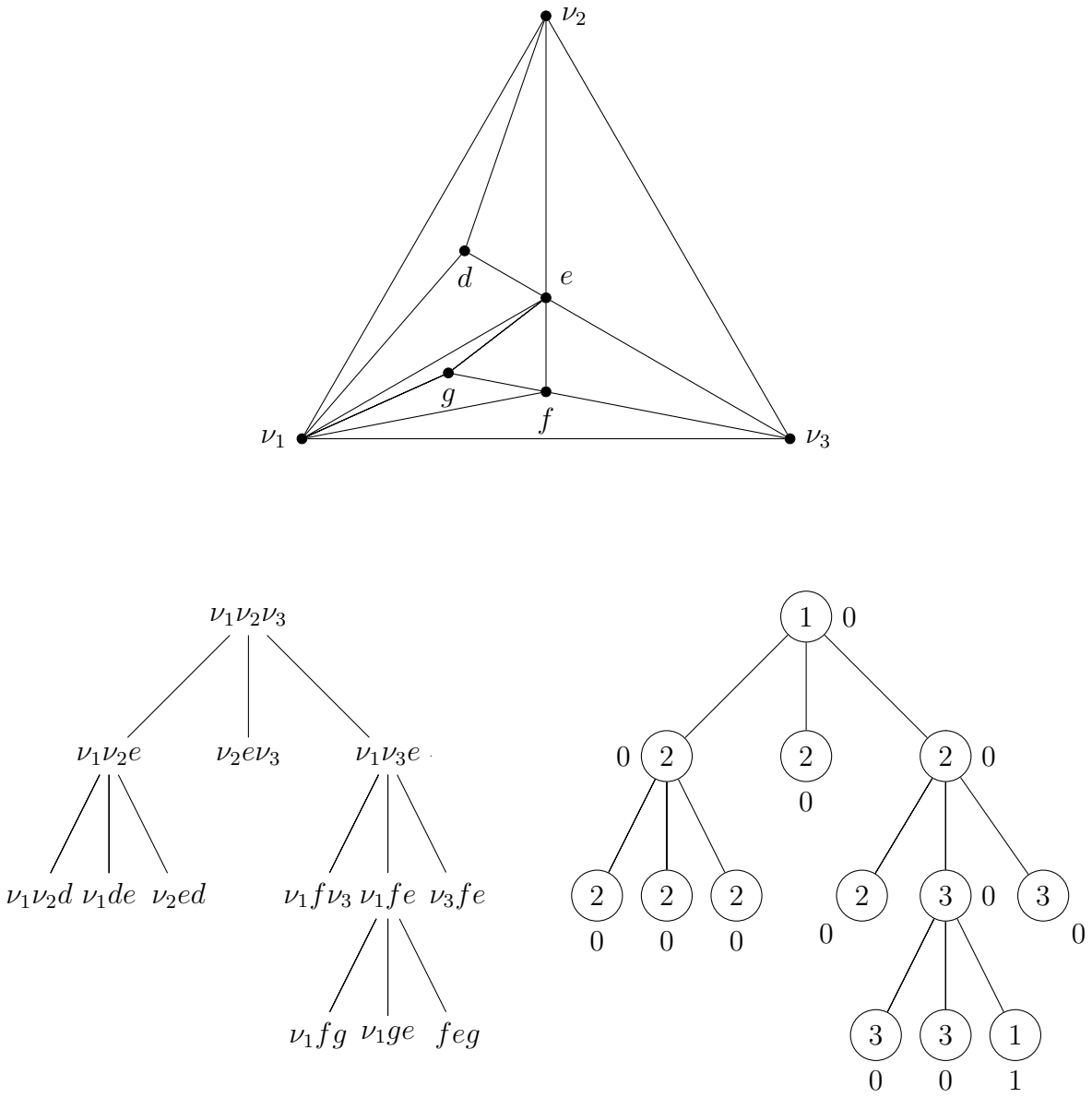


Figure 4.5: a RAN (top) and its corresponding $\triangle$-tree (bottom). In the bottom right part, the type of each node and its τ value are written inside and beside it, respectively.

a_1 is the deterministic time $(1-\varepsilon)\log(\ell_n)/2$.

A_2 is the random time when the evolving tree has exactly ℓ_n nodes.

a_3 is the deterministic time $(1+\varepsilon)\log(\ell_n)/2$.

By Proposition 2.18, a.a.s. as $t \to \infty$ we have

$$\log|V(\widehat{T}^t)| \sim 2t\,,$$

so, as $n \to \infty$, a.a.s.

$$\log|V(\widehat{T}^{a_1})| \sim 2a_1 = (1-\varepsilon)\log(\ell_n)\,,$$

and hence $|V(\widehat{T}^{a_1})| < \ell_n$, which implies $a_1 < A_2$. Symmetrically, it can be proved that a.a.s. as $n \to \infty$ we have $A_2 < a_3$. It follows that a.a.s. as $n \to \infty$

$$\mathrm{ah}\left(\widehat{T}^{a_1}\right) \le \mathrm{ah}\left(\widehat{T}^{A_2}\right) \le \mathrm{ah}\left(\widehat{T}^{a_3}\right)\,.$$

By the assumption, a.a.s. as $n \to \infty$ we have $\mathrm{ah}\left(\widehat{T}^{a_1}\right) \sim (1-\varepsilon)c\log(\ell_n)/2$ and $\mathrm{ah}\left(\widehat{T}^{a_3}\right) \sim (1+\varepsilon)c\log(\ell_n)/2$. On the other hand, as noted above, T has the same distribution as $\widehat{T}^{A_2}$. It follows that a.a.s. as $n \to \infty$

$$1-2\varepsilon \le \frac{2\,\mathrm{ah}(T)}{c\log(\ell_n)} \le 1+2\varepsilon\,.$$

Since ε was arbitrary, the result follows. ■

It will be more convenient to view the process $\widehat{P}$ in the following equivalent way. Let $\widehat{T}$ denote an infinite ternary tree whose nodes have types assigned using rules (a)–(d) and are associated with independent Exp(1) random variables. For convenience, each edge of the tree from a parent to a child is labelled with the random variable associated with the parent, which denotes the age of the parent when the child is born. For every node $u \in V(\widehat{T})$, its *birth time* is defined as the sum of the labels on the edges connecting u to the root, and the birth time of the root is defined to be zero. Given $t \ge 0$, the tree $\widehat{T}^t$ is the subtree induced by nodes with birth time less than or equal to t, and is finite with probability one.

Let $k \ge 3$ be a fixed positive integer. We define two random infinite trees $\underline{T_k}$ and $\overline{T_k}$ as follows. First, we regard $\widehat{T}$ as a tree generated by each node giving birth to exactly three

children with types assigned using (b)–(d), and with an Exp(1) random variable used to label the edges to its children. The tree $\underline{T_k}$ is obtained using the same generation rules as $\widehat{T}$ except that every node of type 2 or 3, whose distance to its closest ancestor of type 1 is equal to k, dies without giving birth to any children. Given $t \geq 0$, the random tree $\underline{T_k^t}$ is, as before, the subtree of $\underline{T_k}$ induced by nodes with birth time less than or equal to t. The tree $\overline{T_k}$ is also generated similarly to $\widehat{T}$, except that for each node u of type 2 (respectively, 3) in $\overline{T_k}$ whose distance to its closest ancestor of type 1 equals k, u has exactly three (respectively, four) children of type 1, and the edges joining u to its children get label 0 instead of random Exp(1) labels. (In the 'evolving tree' interpretation, u immediately gives birth to three or four children of type 1 and dies.) Such a node u is called an *annoying* node. The random tree $\overline{T_k^t}$ is defined as before.

Lemma 4.10. *For every fixed $k \geq 3$, every $t \geq 0$, and every $g = g(t)$, we have*

$$\mathbb{P}\left[\mathrm{ah}\left(\underline{T_k^t}\right) \geq g\right] \leq \mathbb{P}\left[\mathrm{ah}\left(\widehat{T}^t\right) \geq g\right] \leq \mathbb{P}\left[\mathrm{ah}\left(\overline{T_k^t}\right) \geq g\right] .$$

Proof. The left inequality follows from the fact that the random edge labels of $\widehat{T}$ and $\underline{T_k}$ can easily be coupled using a common sequence of independent Exp(1) random variables in such a way that for every $t \geq 0$, the generated $\underline{T_k^t}$ is always a subtree of the generated $\widehat{T}^t$.

For the right inequality, we use a sneaky coupling between the edge labels of $\widehat{T}$ and $\overline{T_k}$. It is enough to choose them using a common sequence of independent Exp(1) random variables $X_1, X_2, \ldots$ and define an injective mapping $f : V(\widehat{T}) \to V\left(\overline{T_k}\right)$ such that for every $u \in V(\widehat{T})$,

(1) the auxiliary depth of $f(u)$ is greater than or equal to the auxiliary depth of u, and

(2) for some I and $J \subseteq I$, the birth time of u equals $\sum_{i \in I} X_i$ and the birth time of $f(u)$ equals $\sum_{j \in J} X_j$.

For annoying nodes, the coupling and the mapping f is shown down to their grandchildren in Figures 4.6 and 4.7. This is easily extended in a natural way to all other nodes of the tree. ■

With a view to proving Lemma 4.8 by appealing to Lemmas 4.9 and 4.10, we will define two sequences $\left(\underline{\rho_k}\right)$ and $\left(\overline{\rho_k}\right)$ such that for each k, a.a.s. the heights of $\overline{T_k^t}$ and $\underline{T_k^t}$ are asymptotic to $\overline{\rho_k} t$ and $\underline{\rho_k} t$, respectively, and also

$$\lim_{k \to \infty} \underline{\rho_k} = \lim_{k \to \infty} \overline{\rho_k} = c \,,$$

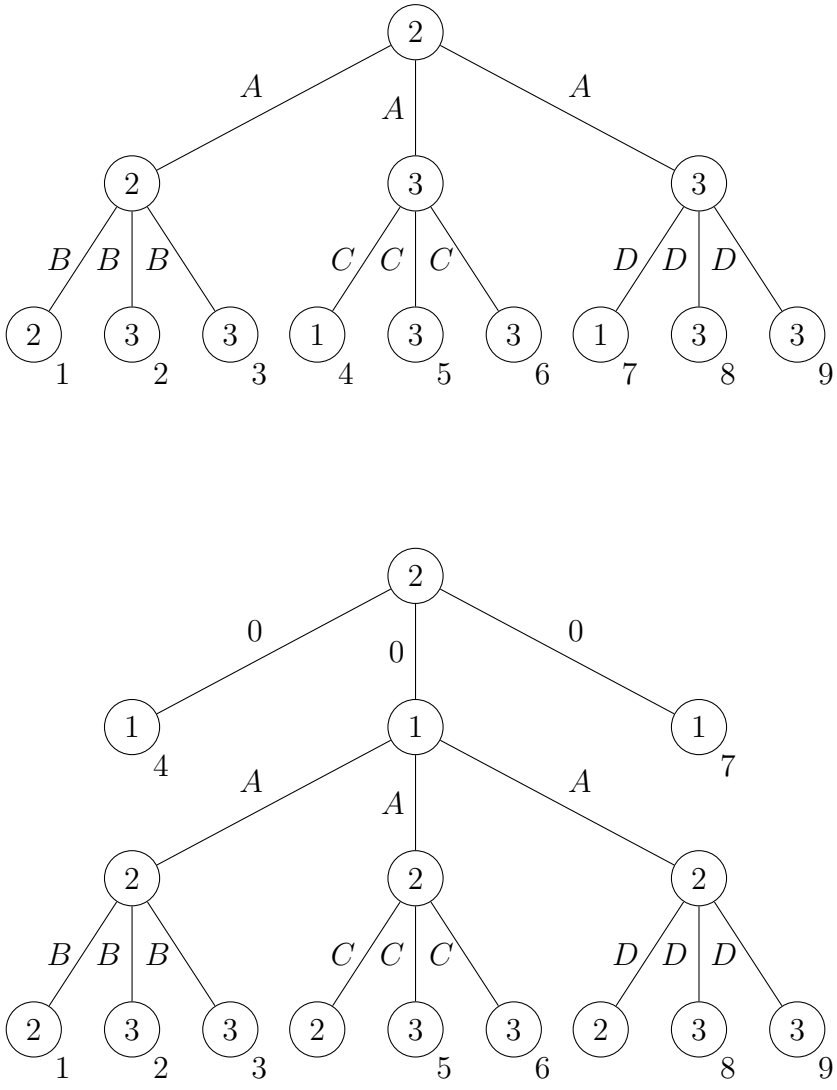


Figure 4.6: illustrating the coupling in Lemma 4.10 for an annoying node of type 2 in $\widehat{T}$. The offspring of the node is shown above and the offspring of the corresponding node in $\overline{T_k}$ is shown below. The type of each node is written inside the node. The coupling of edge labels is defined by the appearance of $A, B, \ldots$ in the two cases. The label 0 is also used in the case of $\overline{T_k}$. The function f is defined by the labels beside the nodes.

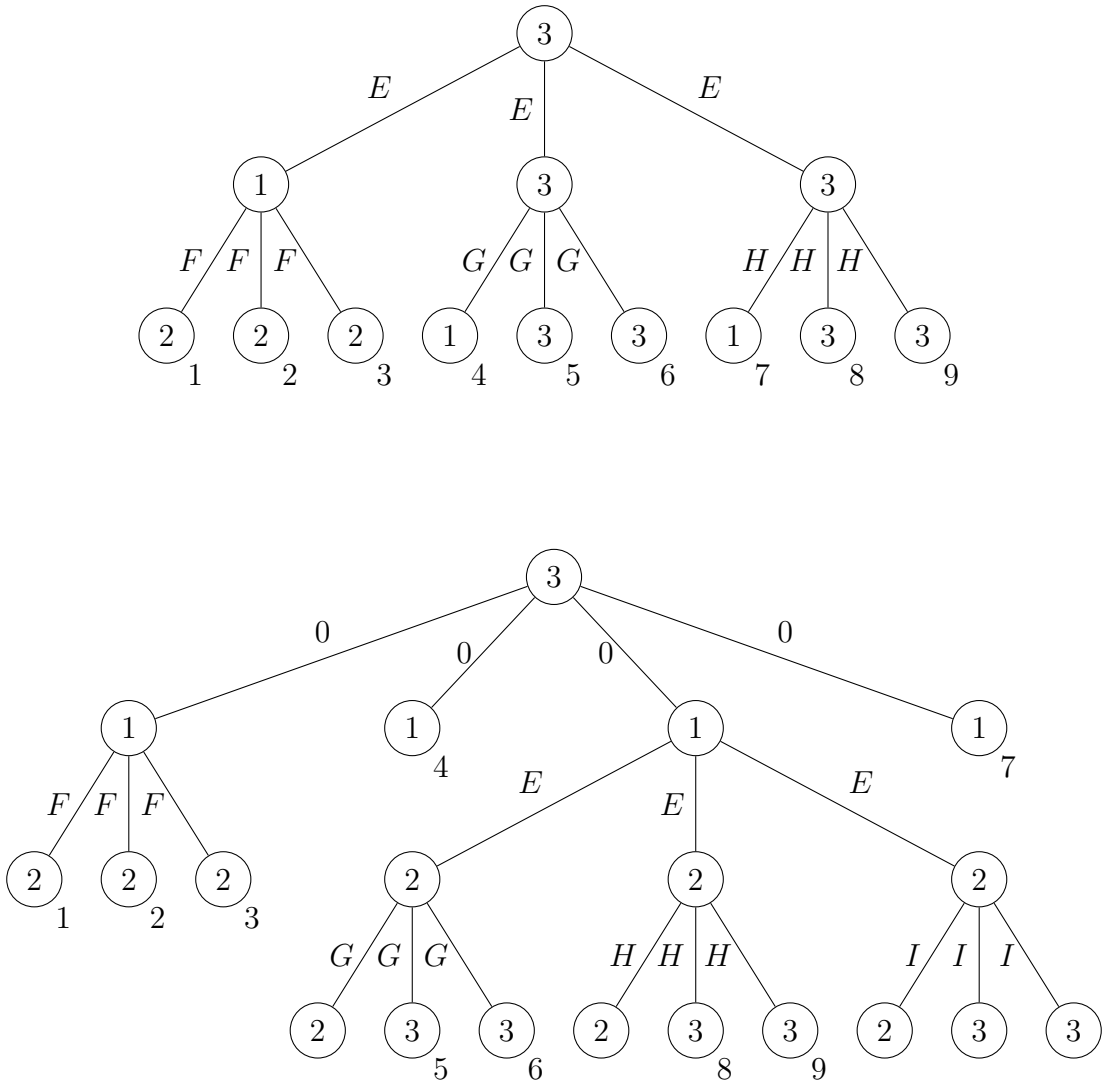


Figure 4.7: the coupling in Lemma 4.10 for an annoying node of type 3 in $\widehat{T}$

where $c \approx 1.668$ is defined in the statement of Lemma 4.8.

The following definition will be used throughout the rest of the section.

Definition 4.11 (Gamma distribution)**.** For a positive integer s, let Gamma(s) denote the Gamma distribution with mean s, i.e., the distribution of the sum of s independent Exp(1) random variables.

For the rest of this section, asymptotics are with respect to t instead of n, unless otherwise specified. We analyze the heights of $\underline{T_k}$ and $\overline{T_k}$ with the help of a theorem of Broutin and Devroye [27, Theorem 1]. We state here a special case suitable for our purposes, including a trivial correction to the conditions on E.

Theorem 4.12 (see [27])**.** *Let E be a template nonnegative random variable that satisfies $\mathbb{P}[E=0] = 0$ and $\sup\{z : \mathbb{P}[E > z] = 1\} = 0$, and such that $\mathbb{P}[E = z] < 1$ for every $z \in \mathbb{R}$; and for which there exists $\lambda > 0$ such that $\mathbb{E}[\exp(\lambda E)]$ is finite. Let $b > 1$ be a positive integer and let T_∞ be an infinite b-ary tree. Let B be a template random b-vector with each component distributed as E (but not necessarily independent components). For every node u of T_∞, label the edges from u to its children using an independently generated copy of B.*

Given $t \geq 0$, let H_t be the height of the subtree of T_∞ induced by the nodes for which the sum of the labels on their path to the root is at most t. Then, a.a.s. we have $H_t \sim \rho t$, where ρ is the unique solution to

$$\sup\{\lambda/\rho - \log(\mathbb{E}[\exp(\lambda E)]) : \lambda \leq 0\} = \log b\,.$$

For each $i = 2, 3, \ldots$, let $\alpha_i, \beta_i, \gamma_i$ denote the number of nodes of type 1, 2, 3 at depth i of $\widehat{T}$ for which the root is the only node of type 1 in their path to the root. Then rules (a)–(d) for determining node types imply

$$\forall i > 2 \qquad \alpha_i = \gamma_{i-1}, \quad \beta_i = \beta_{i-1}, \quad \gamma_i = 2\beta_{i-1} + 2\gamma_{i-1}\,.$$

These, together with $\alpha_2 = 0$, $\beta_2 = 3$, and $\gamma_2 = 6$ (see Figure 4.8), imply

$$\forall i \geq 2 \qquad \alpha_i = 3 \times 2^{i-1} - 6, \quad \beta_i = 3, \quad \gamma_i = 3 \times 2^i - 6\,. \tag{4.3}$$

Let $\underline{b_k} = \sum_{i=1}^{k} \alpha_i$ and $\overline{b_k} = \sum_{i=1}^{k} \alpha_i + 3\beta_k + 4\gamma_k$.

We define a random infinite tree $\underline{T_k}'$ as follows. The nodes of $\underline{T_k}'$ are the type-1 nodes of $\underline{T_k}$. Let V' denote the set of these nodes. For $u, v \in V'$ such that u is the closest type-1

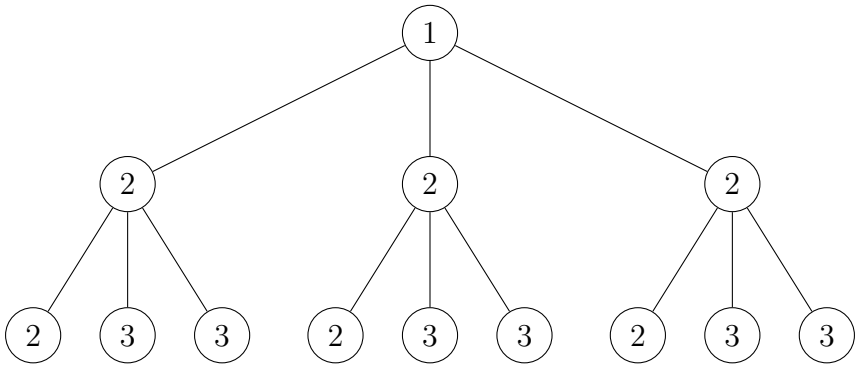


Figure 4.8: we have $\alpha_2 = 0$, $\beta_2 = 3$, and $\gamma_2 = 6$.

ancestor of v in $\underline{T_k}$, there is an edge joining u and v in $\underline{T_k}'$, whose label equals the sum of the labels of the edges in the unique (u, v)-path in $\underline{T_k}$. By construction, for all $t \geq 0$, the height of the subtree of $\underline{T_k}'$ induced by nodes with birth time less than or equal to t equals the auxiliary height of $\underline{T_k}^t$. Let u be a node in $\underline{T_k}'$. Then observe that for each $i = 3, 4, \ldots, k$, u has α_i children whose birth times equal the birth time of u plus a Gamma(i) random variable. In particular, $\underline{T_k}'$ is an infinite $\underline{b_k}$-ary tree.

To apply Theorem 4.12 we need the label of each edge to have the same distribution. For this, we create a random rearrangement of $\underline{T_k}'$. First let $\underline{E_k}$ be the random variable such that for each $3 \leq i \leq k$, with probability $\alpha_i/\underline{b_k}$, $\underline{E_k}$ is distributed as a Gamma(i) random variable. Now, for each node u of $\underline{T_k}'$, starting from the root and in a top-down manner, randomly permute the branches below u. This results in an infinite $\underline{b_k}$-ary tree, every edge of which has a random label distributed as $\underline{E_k}$. Although the labels of edges from a node to its children are dependent, the $\underline{b_k}$-vector of labels of edges from a node to its children is independent of all other edge labels, as required for Theorem 4.12. Let ρ be the solution to

$$\sup\{\lambda/\rho - \log(\mathbb{E}\left[\exp(\lambda \underline{E_k})\right]) : \lambda \leq 0\} = \log \underline{b_k}\,. \tag{4.4}$$

Then by Theorem 4.12, a.a.s. the auxiliary height of $\underline{T_k}^t$, which equals the height of the subtree of $\underline{T_k}'$ induced by nodes with birth time less than or equal to t, is asymptotic to ρt. Notice that we have

$$\mathbb{E}\left[\exp(\lambda\ \mathrm{Exp}(1))\right] = \frac{1}{1-\lambda}\,.$$

So, by the definition of Gamma(s), and since the product of expectation of independent

variables equals the expectation of their product,

$$\mathbb{E}\left[\exp(\lambda \text{ Gamma}(s))\right] = \frac{1}{(1-\lambda)^s}\,.$$

Hence by linearity of expectation,

$$\mathbb{E}\left[\exp(\lambda \underline{E_k})\right] = \sum_{i=3}^{k} \frac{\alpha_i}{\underline{b_k}(1-\lambda)^i}\,. \tag{4.5}$$

One can define a random infinite $\overline{b_k}$-ary tree $\overline{T_k}'$ in a similar way. Let $\overline{E_k}$ be the random variable such that for each $3 \le i \le k-1$, with probability $\alpha_i/\overline{b_k}$, it is distributed as a Gamma(i) random variable, and with probability $(\alpha_k + 3\beta_k + 4\gamma_k)/\overline{b_k}$, it is distributed as a Gamma(k) random variable. Then by a similar argument, a.a.s. the auxiliary height of $\overline{T_k^t}$ is asymptotic to ρt, where ρ is the solution to

$$\sup\{\lambda/\rho - \log(\mathbb{E}\left[\exp(\lambda \overline{E_k})\right]) : \lambda \le 0\} = \log \overline{b_k}\,. \tag{4.6}$$

Moreover, one calculates

$$\mathbb{E}\left[\exp(\lambda \overline{E_k})\right] = \frac{\alpha_k + 3\beta_k + 4\gamma_k}{\overline{b_k}(1-\lambda)^k} + \sum_{i=3}^{k-1} \frac{\alpha_i}{\overline{b_k}(1-\lambda)^i}\,. \tag{4.7}$$

As part of our plan to prove Lemma 4.8, we would like to define $\underline{\rho_k}$ and $\overline{\rho_k}$ in such a way that they are the unique solutions to (4.4) and (4.6), respectively. We first need to establish two analytical lemmas.

For later convenience, we define $\mathcal{F}$ to be the set of positive functions $f : [0.1, 0.2] \to \mathbb{R}$ that are differentiable on $(0.1, 0.2)$, and let $W : \mathcal{F} \to \mathbb{R}^{[0.1,0.2]}$ be the operator defined as

$$Wf(x) = x(x-1)f'(x)/f(x) - \log f(x)\,.$$

Note that Wf is continuous. Define $h \in \mathcal{F}$ as

$$h(x) = \frac{12x^3}{1-2x} - \frac{6x^3}{1-x}\,.$$

Lemma 4.13. *The function Wh has a unique root $\hat{x}$ in $(0.1, 0.2)$.*

Proof. By the definition of $(\alpha_i)_{i\geq 3}$ in (4.3) we have

$$h(x) = \sum_{i\geq 3} \alpha_i x^i \qquad \forall x \in [0.1, 0.2]\,.$$

Since $\alpha_i > 0$ for all $i \geq 3$, we have $h(x) > 0$ and $h'(x) > 0$ for $x \in [0.1, 0.2]$, and hence the derivative of $\log h(x)$ is positive. Moreover, the derivative of $x(x-1)h'(x)/h(x)$ equals $4x(x-1)/(1-2x)^2$, which is negative. Therefore, $Wh(x)$ is a strictly decreasing function on $[0.1, 0.2]$. Numerical calculations give $Wh(0.1) \approx 1.762 > 0$ and $Wh(0.2) \approx -0.831 < 0$. Hence, there is a unique solution to $Wh(x) = 0$ in $(0.1, 0.2)$. ■

Remark 4.14. Numerical calculations give $\hat{x} \approx 0.1629562$.

Define functions $\underline{g_k}, \overline{g_k} \in \mathcal{F}$ as

$$\underline{g_k}(x) = \sum_{i=3}^{k} \alpha_i x^i, \text{ and } \overline{g_k}(x) = (\alpha_k + 3\beta_k + 4\gamma_k)x^k + \sum_{i=3}^{k-1} \alpha_i x^i\,.$$

Note that by (4.5) and (4.7),

$$\underline{b_k}\,\mathbb{E}\left[\exp\left(\lambda \underline{E_k}\right)\right] = \underline{g_k}\left(\frac{1}{1-\lambda}\right), \text{ and } \overline{b_k}\,\mathbb{E}\left[\exp\left(\lambda \overline{E_k}\right)\right] = \overline{g_k}\left(\frac{1}{1-\lambda}\right) \tag{4.8}$$

hold at least when $(1-\lambda)^{-1} \in [0.1, 0.2]$, namely for all $\lambda \in [-9, -4]$.

Lemma 4.15. *Both sequences* $\left(W\underline{g_k}\right)_{k=3}^{\infty}$ *and* $(W\overline{g_k})_{k=3}^{\infty}$ *converge pointwise to* Wh *on* $[0.1, 0.2]$ *as* $k \to \infty$. *Also, there exists a positive integer* k_0 *and sequences* $\left(\underline{x_k}\right)_{k=k_0}^{\infty}$ *and* $(\overline{x_k})_{k=k_0}^{\infty}$ *such that* $W\underline{g_k}\left(\underline{x_k}\right) = W\overline{g_k}\left(\overline{x_k}\right) = 0$ *for all* $k \geq k_0$, *and*

$$\lim_{k\to\infty} \underline{x_k} = \lim_{k\to\infty} \overline{x_k} = \hat{x}\,.$$

Proof. For any $x \in [0.1, 0.2]$, we have

$$\lim_{k\to\infty} \underline{g_k}(x) = h(x), \quad \lim_{k\to\infty} \underline{g_k}'(x) = h'(x), \quad \lim_{k\to\infty} \overline{g_k}(x) = h(x), \quad \lim_{k\to\infty} \overline{g_k}'(x) = h'(x)\,,$$

so the sequences $\left(W\underline{g_k}\right)_{k=3}^{\infty}$ and $(W\overline{g_k})_{k=3}^{\infty}$ converge pointwise to Wh.

Next, we show the existence of a positive integer $\underline{k_0}$ and a sequence $\left(\underline{x_k}\right)_{k=k_0}^{\infty}$ such that $W\underline{g_k}\left(\underline{x_k}\right) = 0$ for all $k \geq k_0$, and

$$\lim_{k\to\infty} \underline{x_k} = \hat{x}\,.$$

The proof for existence of corresponding positive integer $\overline{k_0}$ and the sequence $(\overline{x_k})_{k=k_0}^{\infty}$ is similar, and we may let $k_0 = \max\{\underline{k_0}, \overline{k_0}\}$.

Since $Wh(0.1) > 0$ and $Wh(0.2) < 0$, there exists $\underline{k_0}$ so that for $k \geq \underline{k_0}$, $W\underline{g_k}(0.1) > 0$ and $W\underline{g_k}(0.2) < 0$. Since $W\underline{g_k}$ is continuous for all $k \geq 3$, it has at least one root in $(0.1, 0.2)$. Moreover, since $W\underline{g_k}$ is continuous, the set $\{x : W\underline{g_k}(x) = 0\}$ is a closed set, thus we can choose a root $\underline{x_k}$ closest to $\hat{x}$. We just need to show that $\lim_{k\to\infty} \underline{x_k} = \hat{x}$. Fix an $\varepsilon > 0$. Since $Wh(\hat{x} - \varepsilon) > 0$ and $Wh(\hat{x} + \varepsilon) < 0$, there exists a large enough M such that for all $k \geq M$, $W\underline{g_k}(\hat{x} - \varepsilon) > 0$ and $W\underline{g_k}(\hat{x} + \varepsilon) < 0$. Thus $\underline{x_k} \in (\hat{x} - \varepsilon, \hat{x} + \varepsilon)$. Since ε was arbitrary, we conclude that $\lim_{k\to\infty} \underline{x_k} = \hat{x}$. ■

Let k_0 be as promised by Lemma 4.15 and let $\left(\underline{x_k}\right)_{k=k_0}^{\infty}$ and $(\overline{x_k})_{k=k_0}^{\infty}$ be the sequences given by Lemma 4.15. Define the sequences $\left(\underline{\rho_k}\right)_{k=k_0}^{\infty}$ and $(\overline{\rho_k})_{k=k_0}^{\infty}$ by

$$\underline{\rho_k} = \left(1 - \underline{x_k}^{-1}\right) / \log \underline{g_k}(\underline{x_k}), \quad \overline{\rho_k} = \left(1 - \overline{x_k}^{-1}\right) / \log \overline{g_k}(\overline{x_k}). \tag{4.9}$$

Lemma 4.16. *For every fixed $k \geq k_0$, a.a.s. the heights of $\overline{T_k^t}$ and $\underline{T_k^t}$ are asymptotic to $\overline{\rho_k} t$ and $\underline{\rho_k} t$, respectively.*

Proof. We give the argument for $\overline{T_k^t}$; the argument for $\underline{T_k^t}$ is similar. First of all, we claim that $\log(\mathbb{E}\left[\exp(\lambda \overline{E_k})\right])$ is a strictly convex function of λ over $(-\infty, 0]$. To see this, let $\lambda_1 < \lambda_2 \leq 0$ and let $\theta \in (0, 1)$. Then we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\theta\lambda_1\overline{E_k} + (1-\theta)\lambda_2\overline{E_k}\right)\right] &= \mathbb{E}\left[\left[\exp\left(\lambda_1\overline{E_k}\right)\right]^{\theta}\left[\exp\left(\lambda_2\overline{E_k}\right)\right]^{1-\theta}\right] \\ &< \mathbb{E}\left[\exp\left(\lambda_1\overline{E_k}\right)\right]^{\theta}\mathbb{E}\left[\exp\left(\lambda_2\overline{E_k}\right)\right]^{1-\theta},\end{aligned}$$

where the inequality follows from Hölder's inequality, and is strict as the random variable $\overline{E_k}$ does not have all of its mass concentrated in a single point. Taking logarithms completes the proof of the claim.

It follows that given any value of ρ, $\lambda/\rho - \log(\mathbb{E}\left[\exp(\lambda\overline{E_k})\right])$ is a strictly concave function of $\lambda \in (-\infty, 0]$ and hence attains its supremum at a unique $\lambda \leq 0$.

Now, define

$$\overline{\lambda_k} = 1 - \overline{x_k}^{-1},$$

which is in $(-9, -4)$ as $\overline{x_k} \in (0.1, 0.2)$. Next we will show that

$$\overline{\lambda_k}/\overline{\rho_k} - \log(\mathbb{E}\left[\exp(\overline{\lambda_k}\,\overline{E_k})\right]) = \log \overline{b_k}, \tag{4.10}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\lambda/\overline{\rho_k} - \log(\mathbb{E}\left[\exp(\lambda\overline{E_k})\right])\right]\Big|_{\lambda=\overline{\lambda_k}} = 0, \tag{4.11}$$

which implies that $\overline{\rho_k}$ is the unique solution for (4.6), and thus by Theorem 4.12 and the discussion after it, a.a.s. the height of $\overline{T_k^t}$ is asymptotic to $\overline{\rho_k} t$.

Notice that $\overline{\lambda_k} \in (-9, -4)$, so by (4.8),

$$\overline{b_k}\mathbb{E}\left[\exp(\lambda \overline{E_k})\right] = \overline{g_k}((1-\lambda)^{-1})$$

for λ in a sufficiently small open neighbourhood of $\overline{\lambda_k}$. Taking logarithm of both sides and using (4.9) gives (4.10).

To prove (4.11), note that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\log(\mathbb{E}\left[\exp(\lambda \overline{E_k})\right])\right]\Big|_{\lambda=\overline{\lambda_k}} &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\log \overline{g_k}\left(\left(1-\overline{\lambda_k}\right)^{-1}\right) - \log \overline{b_k}\right]\Big|_{\lambda=\overline{\lambda_k}} \\ &= \frac{\overline{g_k}'\left((1-\overline{\lambda_k})^{-1}\right)}{\left(1-\overline{\lambda_k}\right)^2\, \overline{g_k}\left((1-\overline{\lambda_k})^{-1}\right)} = \overline{x_k}^{\,2}\, \frac{\overline{g_k}'(\overline{x_k})}{\overline{g_k}(\overline{x_k})}\,.\end{aligned}$$

By Lemma 4.15, $W\overline{g_k}(\overline{x_k}) = 0$, i.e.,

$$\overline{x_k}^{\,2}\, \frac{\overline{g_k}'(\overline{x_k})}{\overline{g_k}(\overline{x_k})} = \overline{x_k}^{\,2}\, \frac{\log \overline{g_k}(\overline{x_k})}{\overline{x_k}(\overline{x_k}-1)} = \frac{\log \overline{g_k}(\overline{x_k})}{1-\overline{x_k}^{\,-1}} = \frac{1}{\rho_k}\,,$$

and (4.11) is proved. ■

To complete the proof we will need an analytic lemma, known as Dini's theorem.

Lemma 4.17 (Theorem 7.13 in Rudin [114])**.** *Let E be a compact set. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions continuous on E, which converges to a continuous function f on E. If $f_n(x) \geq f_{n+1}(x)$ for every $n \in \mathbb{N}$ and $x \in E$, then f_n converges to f uniformly on E.*

We now have all the ingredients to prove Lemma 4.8.

Proof of Lemma 4.8. By Lemma 4.9, we just need to show that a.a.s. the auxiliary height of $\widehat{T}^t$ is asymptotic to ct, where

$$c = \frac{1-\hat{x}^{-1}}{\log h(\hat{x})}\,.$$

By Lemma 4.16, a.a.s. the heights of $\overline{T_k^t}$ and $\underline{T_k^t}$ are asymptotic to $\overline{\rho_k} t$ and $\underline{\rho_k} t$, respectively. By Lemma 4.15, $\overline{x_k} \to \hat{x}$ and $\underline{x_k} \to \hat{x}$. Observe that $\left(\underline{g_k}\right)_{k=3}^{\infty}$ and $\left(\overline{g_k}\right)_{k=3}^{\infty}$ converge pointwise to h, and that for every $k \geq 3$ and every $x \in [0.1, 0.2]$, $\underline{g_k}(x) \leq \underline{g_{k+1}}(x)$ and $\overline{g_k}(x) \geq \overline{g_{k+1}}(x)$. Thus by Lemma 4.17, $\left(\underline{g_k}\right)_{k=3}^{\infty}$ and $\left(\overline{g_k}\right)_{k=3}^{\infty}$ converge uniformly to h on $[0.1, 0.2]$.

Hence,

$$\lim_{k\to\infty} \underline{\rho_k} = \lim_{k\to\infty} \frac{1 - \underline{x_k}^{-1}}{\log \underline{g_k}(\underline{x_k})} = \frac{1-\hat{x}^{-1}}{\log h(\hat{x})} = c\,,$$

and

$$\lim_{k\to\infty} \overline{\rho_k} = \lim_{k\to\infty} \frac{1 - \overline{x_k}^{-1}}{\log \overline{g_k}(\overline{x_k})} = \frac{1-\hat{x}^{-1}}{\log h(\hat{x})} = c\,.$$

It follows from Lemma 4.10 that a.a.s. the auxiliary height of $\widehat{T}^t$ is asymptotic to ct, as required. ■

4.2 Lower bounds for longest paths

In this section we prove Theorem 4.2. We first prove part (a), i.e., we give a deterministic lower bound for the length of a longest cycle in a RAN. Recall that $\mathcal{C}_m$ denotes the number of vertices of a longest cycle in a RAN with m faces. Let G be a RAN with m faces, and let v be the unique vertex that is adjacent to ν_1, ν_2, and ν_3 (see Figure 4.9(a)). For $1 \le i \le 3$, let $\triangle_i$ be the triangle with vertex set $\{v, \nu_1, \nu_2, \nu_3\} \setminus \{\nu_i\}$. Define the random variable $\mathcal{L}'_m$ as the largest number L such that for every permutation π on $\{1,2,3\}$, there is a path in G of L edges from $\nu_{\pi(1)}$ to $\nu_{\pi(2)}$ not containing $\nu_{\pi(3)}$. Clearly we have $\mathcal{C}_m \ge \mathcal{L}'_m + 2$.

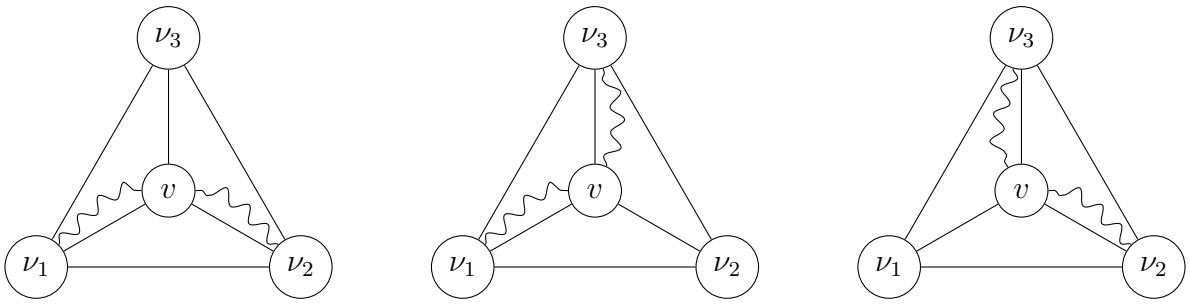
Proof of Theorem 4.2(a). Let $\xi = \log 2/\log 3$. We prove by induction on m that $\mathcal{L}'_m \ge m^\xi$. This is obvious for $m = 1$, so assume that $m > 1$. Let m_i denote the number of faces in $\triangle_i$. Then $m_1 + m_2 + m_3 = m$. By symmetry, we may assume that $m_1 \ge m_2 \ge m_3$. For any given $1 \le i \le 3$, it is easy to find a path avoiding ν_i that connects the other two ν_j's by attaching two appropriate paths in $\triangle_1$ and $\triangle_2$ at vertex v. (See Figures 4.9(a)–(c).) By the induction hypothesis, these paths can be chosen to have lengths at least ${m_1}^\xi$ and ${m_2}^\xi$, respectively. Hence for every permutation π of $\{1,2,3\}$, there is a path from $\nu_{\pi(1)}$ to $\nu_{\pi(2)}$ avoiding $\nu_{\pi(3)}$ with length at least

$${m_1}^\xi + {m_2}^\xi\,. \tag{4.12}$$

It is easily verified that since $m_1 \ge m_2 \ge m_3$ and $m_1 + m_2 + m_3 = m$, the minimum of (4.12) happens when $m_1 = m_2 = m/3$, thus

$$\mathcal{L}'_m \ge {m_1}^\xi + {m_2}^\xi \ge 2\left(\frac{m}{3}\right)^\xi = m^\xi\,,$$

and the proof is complete. ■



(a) Path avoiding ν_3. (b) Path avoiding ν_2. (c) Path avoiding ν_1

Figure 4.9: paths avoiding $\triangle_3$ and one of the ν_i's

Next, we use the same idea to give a larger lower bound for $\mathbb{E}[\mathcal{C}_m]$. Let the random variable X_i denote the number of faces in $\triangle_i$. Then the X_i's have the same distribution and are not independent. As in the proof of Proposition 4.7, we can analyze the evolution of the random vector (X_1, X_2, X_3) by modelling the process of building the RAN as an Eggenberger-Pólya urn: after the first subdivision of $\triangle$, each of $\triangle_1$, $\triangle_2$, and $\triangle_3$ contains exactly one face. We consider an urn with 3 different colours. Start with one ball of each colour, corresponding to the faces in $\triangle_1, \triangle_2$, and $\triangle_3$. In each subsequent step, we choose a face uniformly at random, and subdivide it. For each i, if the face is in $\triangle_i$, then the number of faces in $\triangle_i$ increases by 2. In the urn, we add two balls of colour i. Thus after k subdivisions of $\triangle$, the number of faces in $\triangle_i$ has the same distribution as the number of balls of colour i after $k-1$ draws.

Note that, when the RAN has m faces, the number of subdivisions is $k = (m-1)/2$. It follows from Theorem 2.10 that we have the weak convergence

$$\left(\frac{X_1}{m}, \frac{X_2}{m}, \frac{X_3}{m}\right) \longrightarrow \text{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \tag{4.13}$$

as $m \to \infty$.

We are now ready to prove part (b) of Theorem 4.2.

Proof of Theorem 4.2(b). Let $\zeta = 0.88$. We prove that there exists a constant $\kappa > 0$ such that $\mathbb{E}[\mathcal{L}'_m] \geq \kappa m^\zeta$ holds for all $m \geq 1$. We proceed by induction on m, with the induction base being $m = m_0$, where m_0 is a sufficiently large constant, to be determined later. By choosing κ sufficiently small, we may assume $\mathbb{E}[\mathcal{L}'_m] \geq \kappa m^\zeta$ for all $m \leq m_0$.

For $1 \leq i \leq 3$, let X_i denote the number of faces in $\triangle_i$. Define a permutation σ on $\{1, 2, 3\}$ such that $X_{\sigma(1)} \geq X_{\sigma(2)} \geq X_{\sigma(3)}$, breaking ties randomly. Then σ is a random per-

mutation determined by the X_i and the random choice in the tie-breaking. By symmetry, for every fixed $\sigma' \in S_3$, $\mathbb{P}[\sigma = \sigma'] = 1/6$. As in the proof of part (a), we have

$$\mathcal{L}'_m \geq \mathcal{L}'_{X_{\sigma(1)}} + \mathcal{L}'_{X_{\sigma(2)}}.$$

Taking the expectation on both sides, we have

$$\mathbb{E}[\mathcal{L}'_m] \geq \mathbb{E}\left[\mathcal{L}'_{X_{\sigma(1)}} + \mathcal{L}'_{X_{\sigma(2)}}\right] \geq 6\mathbb{E}\left[(\mathcal{L}'_{X_1} + \mathcal{L}'_{X_2})\mathbb{1}_{X_1>X_2>X_3}\right], \tag{4.14}$$

where the second inequality holds by symmetry and as $\mathbb{P}[\sigma = (1,2,3)] = 1/6$. By the induction hypothesis, for every $x_1, x_2 < m$,

$$\mathbb{E}\left[\mathcal{L}'_{X_1} \mid X_1 = x_1\right] \geq \kappa x_1^\zeta, \text{ and } \mathbb{E}\left[\mathcal{L}'_{X_2} \mid X_2 = x_2\right] \geq \kappa x_2^\zeta.$$

Hence,

$$\mathbb{E}\left[(\mathcal{L}'_{X_1} + \mathcal{L}'_{X_2})\mathbb{1}_{X_1>X_2>X_3}\right] \geq \kappa\mathbb{E}\left[(X_1^\zeta + X_2^\zeta)\mathbb{1}_{X_1>X_2>X_3}\right]. \tag{4.15}$$

Let

$$f(x_1, x_2, x_3) = \frac{1}{2\pi\sqrt{x_1 x_2 x_3}}$$

and

$$S := \{(x_1, x_2, x_3) : x_1, x_2, x_3 \geq 0 \text{ and } x_1 + x_2 + x_3 = 1\}.$$

Note that f and S are the density function and the support of Dirichlet$(1/2, 1/2, 1/2)$, respectively (see Definition 2.6). Since the function

$$(x_1, x_2, x_3) \mapsto (x_1^\zeta + x_2^\zeta)\mathbb{1}_{X_1>X_2>X_3}$$

is bounded and continuous on S, it follows from the convergence (4.13) that

$$\begin{aligned}\mathbb{E}\left[\left(\left(\frac{X_1}{m}\right)^\zeta + \left(\frac{X_2}{m}\right)^\zeta\right)\mathbb{1}_{X_1>X_2>X_3}\right] &\to \int_S (x_1^\zeta + x_2^\zeta)\mathbb{1}_{x_1>x_2>x_3} f(x_1, x_2, x_3)\mathrm{d}(x_1, x_2, x_3) \\ &= \int_{x_1=1/3}^{1} \int_{x_2=(1-x_1)/2}^{\min\{x_1, 1-x_1\}} \frac{x_1^\zeta + x_2^\zeta}{2\pi\sqrt{x_1 x_2(1-x_1-x_2)}}\mathrm{d}x_2\mathrm{d}x_1\end{aligned}$$

as $m \to \infty$ (see, e.g., Billingsley [15, Theorem 29.1 (i)]). By the choice of ζ, we have

$$\int_{x_1=1/3}^{1} \int_{x_2=(1-x_1)/2}^{\min\{x_1, 1-x_1\}} \frac{x_1^\zeta + x_2^\zeta}{2\pi\sqrt{x_1 x_2(1-x_1-x_2)}}\mathrm{d}x_2\mathrm{d}x_1 > 1/6\,.$$

Then, by (4.14) and (4.15),

$$\mathbb{E}[\mathcal{L}'_m] \geq 6\kappa\mathbb{E}\left[(X_1^\zeta + X_2^\zeta)\mathbb{1}_{X_1>X_2>X_3}\right] > \kappa m^\zeta,$$

if we choose m_0 sufficiently large. ■

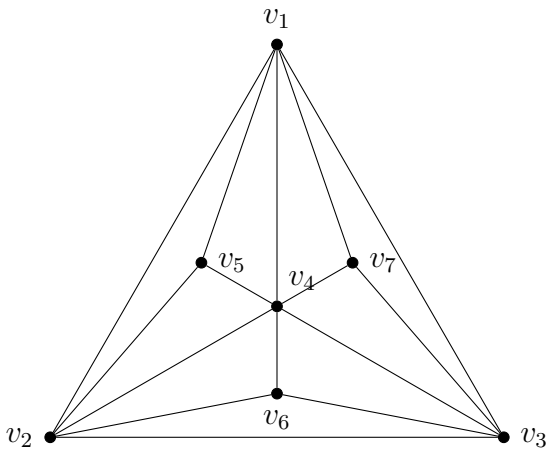


Figure 4.10: a triangle in G corresponding to a node of T with 9 grandchildren. Vertices $v_1, \ldots, v_7$ are the vertices in the boundaries of the triangles corresponding to these grandchildren.

4.3 Upper bounds for longest paths

In this Section we prove Theorem 4.4. Let $\mathcal{T}_t$ denote the $\triangle$-tree of a RAN G with $n = t+3$ vertices and $m = 2t+1$ faces, i.e. a RAN after exactly t subdivisions. We start with a crucial graph theoretic observation.

Recall that the set of *grandchildren* of a node is the set of children of its children, and for a triangle $\triangle$ in G, let $I(\triangle)$ denote the set of vertices of G that are *strictly inside* $\triangle$.

Lemma 4.18. *Let $\mathbf{n}^{\triangle}$ be a node of $\mathcal{T}_t$ with nine grandchildren $\mathbf{n}^{\triangle_1}, \mathbf{n}^{\triangle_2}, \ldots, \mathbf{n}^{\triangle_9}$. Then the vertex set of a path in G intersects at most eight of the $I(\triangle_i)$'s.*

Proof. There are exactly seven vertices in the boundaries of the triangles corresponding to the grandchildren of $\mathbf{n}^{\triangle}$. Let $v_1, \ldots, v_7$ denote such vertices (see Figure 4.10). Let $P = u_1 u_2 \ldots u_p$ be a path in G. When P enters or leaves one of $\triangle_1, \triangle_2, \ldots, \triangle_9$, it must go through a v_i. So P does not contain vertices from more than one triangle between two consecutive occurrences of a v_i. Since P goes through each v_i at most once, the vertices v_i split P up into at most eight sub-paths. Hence P contains vertices from at most eight of the triangles $\triangle_i$. ■

Definition 4.19 (Deficient subtree)**.** A subtree $\mathcal{J}$ of $\mathcal{T}_t$ is called *deficient* if each node of $\mathcal{J}$ has at most eight grandchildren (out of the nine possible).

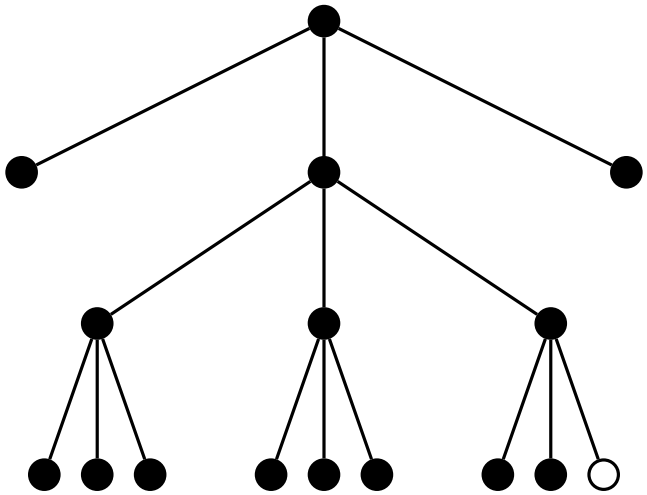


Figure 4.11: a deficient subtree is coloured in black.

See Figure 4.11 for an illustration. In view of Lemma 4.18, we will prove upper bounds for the size of deficient subtrees.

For $\lambda > 1$, define

$$g(\lambda) = \frac{9\lambda}{2(\lambda-1)^{3/2}}\Big(\sqrt{\pi} + \sqrt{\pi}\log\left(\lambda-1\right)/2 + 4/9\Big)\,. \tag{4.16}$$

The main result of this section is the following proposition.

Proposition 4.20. *Let τ, κ, and $\lambda > 2$ be positive constants satisfying*

$$3e\log\tau < 2\sqrt{\tau}\,, \tag{4.17}$$

$$9\,\kappa^{\lambda} g(\lambda) \le 1\,, \tag{4.18}$$

and let d be the largest even integer smaller than $\log t/\log\tau$. Then, a.a.s. as $t\to\infty$, the largest deficient subtree of $\mathcal{T}_t$ has $O\left(8^{d/2} + t\kappa^{-d/2}\right)$ nodes.

We show how this theorem implies Theorem 4.4.

Proof of Theorem 4.4. Set $\lambda = 10^6$, $\kappa = (9g(\lambda))^{-1/\lambda} \approx 1.000000535$, $\tau = 720$, and $\delta = 1 - 4\times 10^{-8}$. Observe that $3e\log\tau < 2\sqrt{\tau}$ and $\delta > \max\{1 - \log(\kappa)/2\log\tau, \log(8)/2\log\tau\}$.

Let P be a path in G and let $R(P)$ denote the set of nodes $\mathbf{n}^{\triangle}$ of $\mathcal{T}_t$ such that $I(\triangle)$ contains some vertex of P. By Lemma 4.18, $R(P)$ induces a deficient subtree of $\mathcal{T}_t$. Hence, using Proposition 4.20 for the second inequality, a.a.s. we have

$$|V(P)| \leq 3 + |R(P)| \leq 3 + O\left(8^{d/2} + t\kappa^{-d/2}\right) < t^{\delta} < (2t+1)^{\delta} ,$$

as required. ■

The rest of this section is devoted to the proof of Proposition 4.20. It will be convenient to view $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_t, \ldots$ as a growing subtree of an infinite ternary tree $\mathcal{T}$. Denote the root of $\mathcal{T}$ by ϱ. For any $v \in V(\mathcal{T})$, the set of nodes of the unique (v, ϱ)-path is denoted by $\pi(v)$.

For any $t \in \mathbb{N}$ and $v \in V(\mathcal{T}_t)$, let $\aleph(v,t)$ denote the number of descendants of v in $\mathcal{T}_t$, including v itself. This is the number of nodes in the 'branch' of $\mathcal{T}_t$ containing v, including v. For any $t \in \mathbb{N}$ and $v \in V(\mathcal{T})$, let

$$\mathrm{Weight}(v,t) = \frac{\aleph(v,t) - 1}{3} \tag{4.19}$$

if $v \in V(\mathcal{T}_t)$, and $\mathrm{Weight}(v,t) = 0$ if $v \notin V(\mathcal{T}_t)$. Note that this is the number of non-leaf nodes in this branch at time t, see Figure 4.12. For $A \subseteq V(\mathcal{T})$ define $\mathrm{Weight}(A,t) = \sum_{v \in A} \mathrm{Weight}(v,t)$.

The main idea of the proof is to show that in a deficient subtree, in every level we lose a certain amount of 'weight', so we cannot gather a lot of nodes. However, working with weights is a bit difficult; using the following lemma, which is based on results about Eggenberger-Pólya urns, we introduce 'masses', which are easier to work with.

Lemma 4.21. *There exist random variables $\{B_v\}_{v \in V(\mathcal{T})}$, such that for any $t \in \mathbb{N}$ and $v \in V(\mathcal{T})$ we have*

$$\mathrm{Weight}(v,t) \overset{s}{\leq} \mathrm{Bin}\left(t, \prod_{\sigma \in \pi(v)} B_\sigma\right) .$$

Moreover, $B_\varrho = 1$ and for all $v \neq \varrho$ we have $B_v \overset{d}{=} \mathrm{Beta}(1/2, 1)$. Also, if u and v are not siblings, then B_u and B_v are independent. Finally, if x, y, z are siblings, then

$$B_x + B_y + B_z = 1 . \tag{4.20}$$

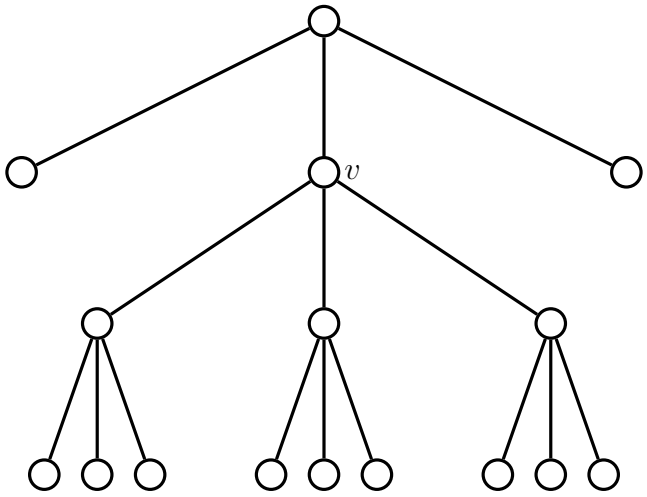


Figure 4.12: we have let $\aleph(v,5) = 13$ and $\mathrm{Weight}(v,5) = 4$.

Proof. Consider a node $v \neq \varrho$ and a positive integer t such that $v \in V(\mathcal{T}_t)$, and let p denote the parent of v. Note that at time t, the number of leaves in the branch at v is $2\mathrm{Weight}(v,t)+1$. Hence, given that at time $t+1$ the weight of p increases, the probability, conditional on the past, that the weight of v increases at the same time, is equal to

$$\frac{2\mathrm{Weight}(v,t)+1}{2\mathrm{Weight}(p,t)+1}.$$

Each time a weight increases, its increment is exactly 1. Thinking of v and its offspring as white balls, and its siblings and their offspring as black balls, the evolution of the numerator of the above expression over time can be modelled using an Eggenberger-Pólya urn, with initial condition $w_0 = 1$, $b_0 = 2$, and reinforcement $s = 2$ (see Section 2.3), that is,

$$2\mathrm{Weight}(v,t)+1 \stackrel{d}{=} \mathrm{Urn}(1,2,2,\mathrm{Weight}(p,t)-1)\,.$$

Let $X \stackrel{d}{=} \mathrm{Beta}(1/2,1)$. Proposition 2.9 states that

$$\mathrm{Urn}(1,2,2,\mathrm{Weight}(p,t)-1) \stackrel{d}{=} 1+2\,\mathrm{Bin}(\mathrm{Weight}(p,t)-1,X)\,.$$

Thus,

$$\mathrm{Weight}(v,t) \stackrel{d}{=} \mathrm{Bin}(\mathrm{Weight}(p,t)-1,X) \stackrel{s}{\leq} \mathrm{Bin}(\mathrm{Weight}(p,t),X)\,.$$

Moreover, the urns corresponding to distinct nodes are mutually independent. Set $B_\varrho = 1$ and for any triple of siblings (x, y, z) we independently assign

$$(B_x, B_y, B_z) \stackrel{d}{=} \text{Dirichlet}(1/2, 1/2, 1/2)\,.$$

By the properties of the Dirichlet distribution (see Definition 2.6) we have $B_x + B_y + B_z = 1$ and each of B_x,B_y, and B_z is distributed as Beta$(1/2, 1)$. Note that Weight$(\varrho, t) = t$. So by induction, Weight(v, t), conditional on $\{B_\sigma\}_{\sigma\in\pi(v)}$, is stochastically smaller than $\text{Bin}\left(t, \prod_{\sigma\in\pi(v)} B_\sigma\right)$, as required. ■

For any node v of $\mathcal{T}$, define

$$\text{Mass}(v) = \prod_{\sigma\in\pi(v)} B_\sigma\,. \tag{4.21}$$

Thanks to Lemma 4.21, we can work with the smooth and somewhat independent masses instead of non-smooth and dependent weights, which makes life easier. This lemma states that the masses approximate the weights well enough for our purposes. Another advantage is that we can now work with the infinite tree $\mathcal{T}$ rather than the finite and unbalanced $\mathcal{T}_t$, so we need not worry about the structure of the tree and which nodes are there by time t etc.

Observation 4.22 (Law of mass conservation)**.** Let $v \in V(\mathcal{T})$ and let x, y, z denotes its children. Then (4.20) implies that $\text{Mass}(v) = \text{Mass}(x) + \text{Mass}(y) + \text{Mass}(z)$.

For a node v of $\mathcal{T}$, we define

$$\Upsilon_v = \min\left\{\frac{\text{Mass}(u)}{\text{Mass}(v)} : u \text{ is a grandchild of } v\right\}\,. \tag{4.22}$$

We will need a technical lemma, whose proof consists of straightforward calculations and appears at the end of the section.

Lemma 4.23. *Let v be a node of $\mathcal{T}$ and let $\lambda > 2$. Then $\mathbb{E}\left[(1-\Upsilon_v)^\lambda\right] < g_1(\lambda) < g(\lambda)$, for some function g_1, where g is defined in (4.16).*

The following lemma, which is the core of our argument, bounds the mass of a deficient subtree. The idea of the proof is that in a deficient subtree, going down every two levels we lose a certain amount of mass.

As in Definition 4.19, we say that a subtree $\mathcal{J}$ of $\mathcal{T}$ is *deficient* if each node of $\mathcal{J}$ has at most eight grandchildren (out of the nine possible).

Denote by $\mathcal{B}_d$ the collection of subsets of $V(\mathcal{T})$ at depth d, with the property that they belong to the same deficient subtree. Note that each element of $\mathcal{B}_{2d}$ has at most 8^d nodes.

Lemma 4.24. *Let $\kappa > 0$ and $\lambda > 2$ be constants satisfying (4.18). A.a.s. as $d \to \infty$ we have*

$$\max_{C \in \mathcal{B}_{2d}} \mathrm{Mass}(C) \leq \kappa^{-d}.$$

Proof. Let $g_1(\lambda)$ be as given by Lemma 4.23. We first prove a claim. For a node v of $V(\mathcal{T})$ at even depth, let $v = v_{2k}, v_{2k-1}, \ldots, v_1, v_0 = \varrho$ be the unique (v, ϱ)-path. We define

$$\rho(v) = \{v_{2k-2}, v_{2k-4}, \ldots, v_2, v_0\}.$$

Claim 4.25. *A.a.s. as $d \to \infty$ the following holds: for all v at depth $2d$ of $\mathcal{T}$ we have*

$$\prod_{\sigma \in \rho(v)} \left(1 - \Upsilon_\sigma\right)^{-1} \geq \kappa^d.$$

Proof of Claim. Let v be an arbitrary node at depth $2d$. Since the Υ_σ are independent for $\sigma \in \rho(v)$ and $\lambda > 2$, Markov's inequality gives

$$\begin{aligned}\mathbb{P}\left[\prod_{\sigma \in \rho(v)} \left(1 - \Upsilon_\sigma\right)^{-1} < \kappa^d\right] &= \mathbb{P}\left[\prod_{\sigma \in \rho(v)} \left(1 - \Upsilon_\sigma\right)^{\lambda} > \kappa^{-\lambda d}\right] \\ &< \mathbb{E}\left[\prod_{\sigma \in \rho(v)} \left(1 - \Upsilon_\sigma\right)^{\lambda}\right] \kappa^{\lambda d} < \left(g_1(\lambda)\kappa^\lambda\right)^d,\end{aligned}$$

where we have used Lemma 4.23 for the last inequality. By (4.18) and since $g_1(\lambda) < g(\lambda)$ we have $9g_1(\lambda)\kappa^\lambda < 1$. Since there are 9^d nodes at depth $2d$ of $\mathcal{T}$, the union bound completes the proof of claim. □

For each node v of $\mathcal{T}$ at even depth, we define its *adjusted mass*, written $\mathrm{AMass}(v)$, as follows. For the root, $\mathrm{AMass}(\varrho) = 1$, and for other nodes v,

$$\mathrm{AMass}(v) = \mathrm{Mass}(v) \times \prod_{\sigma \in \rho(v)} \left(1 - \Upsilon_\sigma\right)^{-1}.$$

For any $A \subseteq V(\mathcal{T})$, let $\mathrm{AMass}(A) = \sum_{v \in A} \mathrm{AMass}(v)$.

Claim 4.26. *For any $d \in \mathbb{N}_0$ and any $C \in \mathcal{B}_{2d}$ we have* $\mathrm{AMass}(C) \le 1$.

Proof of Claim. Let $C \in \mathcal{B}_{2d}$, and let T be a deficient subtree of $\mathcal{T}$ that contains the nodes of C. For any node v of T, denote its set of grandchildren in T by $\mathrm{gc}(v)$. Then by the definition of Υ_v in (4.22) and the Law of mass conservation (Observation 4.22),

$$\mathrm{Mass}(\mathrm{gc}(v)) \le (1 - \Upsilon_v)\mathrm{Mass}(v).$$

Thus $\mathrm{AMass}(\mathrm{gc}(v)) \le \mathrm{AMass}(\{v\})$. Hence, for any $1 \le k \le d$, we have

$$\sum_{v \in V(T), \mathrm{depth}(v)=2k} \mathrm{AMass}(v) \le \sum_{v \in V(T), \mathrm{depth}(v)=2k-2} \mathrm{AMass}(v).$$

Iterating this, we get

$$\mathrm{AMass}(C) = \sum_{v \in C} \mathrm{AMass}(v) \le \mathrm{AMass}(\varrho) = 1. \qquad \square$$

Using Claims 4.25 and 4.26, a.a.s. as $d \to \infty$ we have that for every $C \in \mathcal{B}_{2d}$,

$$1 \ge \mathrm{AMass}(C) = \sum_{v \in C} \left(\mathrm{Mass}(v) \prod_{\sigma \in \rho(v)} (1 - \Upsilon_\sigma)^{-1} \right) \ge \sum_{v \in C} \left(\mathrm{Mass}(v) \kappa^d \right) = \kappa^d \mathrm{Mass}(C)\,,$$

and the lemma follows. ■

We now prove the main result of this section.

Proof of Proposition 4.20. By Lemma 4.21, for any node v and positive integer t we have $\mathrm{Weight}(v,t) \overset{s}{\le} \mathrm{Bin}(t, \mathrm{Mass}(v))$. The upper tail Chernoff bound (2.3) implies

$$\mathbb{P}\left[\mathrm{Weight}(v,t) \ge 3t\mathrm{Mass}(v) \mid \mathrm{Mass}(v) \ge q\right] \le \exp(-tq) \tag{4.23}$$

for any $q > 0$.

Since τ satisfies (4.17), there exists $\tau_1 < \tau$ satisfying

$$3e \log \tau_1 < 2\sqrt{\tau_1}\,. \tag{4.24}$$

Let $\beta = 1/\tau_1$. By Lemma 2.23 (concentration of products of independent beta random variables), for any node v at level d we have

$$\mathbb{P}\left[\mathrm{Mass}(v) < \beta^d\right] = \mathbb{P}\left[\prod_{\sigma\in\pi(v)} B_\sigma < \beta^d\right] \le \left(\frac{e\log(1/\beta)\sqrt{\beta}}{2}\right)^d . \tag{4.25}$$

Note that (4.24) implies that the term in brackets is a constant smaller than 1/3.

Let $\mathcal{A}$ denote the event 'there exists node v at depth d with $\mathrm{Weight}(v,t) \ge 3t\mathrm{Mass}(v)$.' We have

$$\mathbb{P}\left[\mathcal{A}\right] \le 3^d\mathbb{P}\left[\mathrm{Weight}(v,t) \ge 3t\mathrm{Mass}(v)\right] \le 3^d\mathbb{P}\left[\mathrm{Mass}(v) < \beta^d\right] + 3^d\exp\left(-t\beta^d\right) ,$$

where we have used (4.23) for the second inequality. The right-hand-side is $o(1)$: the first term is $o(1)$ since (4.25) holds, and the second one is $o(1)$ as $t^{1/d}\beta \ge \tau\beta$, and $\tau\beta$ is a constant larger than 1. So, a.a.s. as $d\to\infty$, for all nodes v at depth d we have $\mathrm{Weight}(v,t) < 3t\mathrm{Mass}(v)$.

On the other hand, by Lemma 4.24, a.a.s. as $d\to\infty$ we have $\max_{C\in\mathcal{B}_d}\mathrm{Mass}(C) \le \kappa^{-d/2}$. Therefore, a.a.s. as $d\to\infty$ we have

$$\max_{C\in\mathcal{B}_d}\mathrm{Weight}(C,t) < 3t\kappa^{-d/2} . \tag{4.26}$$

Since d is a growing function of t, (4.26) holds also a.a.s. as $t\to\infty$.

Let $\mathcal{J}$ be a deficient subtree of $\mathcal{T}_t$. If $\mathcal{J}$ has k nodes at depths r and $r+1$, then it has at most $8k$ nodes at depths $r+2$ and $r+3$, hence the number of nodes of $\mathcal{J}$ with depth smaller than d is at most

$$4\times(1+8+\cdots+8^{d/2-1}) = 4\times(8^{d/2}-1)/7 < 8^{d/2} .$$

Let C be the set of nodes of $\mathcal{J}$ at depth d. Note that $C\in\mathcal{B}_d$. From the definition of weights in (4.19), the number of nodes of $\mathcal{J}$ with depth at least d is

$$\sum_{v\in C}\aleph(v,t) = 3\mathrm{Weight}(C,t) + |C| \le 9t\kappa^{-d/2} + 8^{d/2} ,$$

where we have used (4.26). Therefore,

$$|V(\mathcal{J})| \le 2\times 8^{d/2} + 9t\kappa^{-d/2} ,$$

as required. ■

Proof of Lemma 4.23. Let $\Upsilon = \Upsilon_v$. By definition of the Mass function in (4.21), for each grandchild u of v, the fraction $\text{Mass}(u)/\text{Mass}(v)$ is a product of two independent $\text{Beta}(1/2, 1)$ random variables. Let B_1 and B_2 be two such variables. The density function of each of B_1 and B_2 is $1/(2\sqrt{x})$ if $x \in (0, 1)$ and 0 elsewhere, hence we have

$$\mathbb{P}\left[B_1 B_2 \le \varepsilon\right] = \int_0^1 \left(\int_0^{\min\{1, \varepsilon/x\}} \frac{1}{2\sqrt{y}} \mathrm{d}y \right) \frac{1}{2\sqrt{x}} \mathrm{d}x = \sqrt{\varepsilon}(1 + \log(1/\varepsilon)/2) \,.$$

Thus

$$\begin{aligned} \mathbb{E}\left[(1-\Upsilon)^\lambda\right] &= \int_0^1 \mathbb{P}\left[(1-\Upsilon)^\lambda \ge x\right] dx \\ &= \int_0^1 \mathbb{P}\left[\Upsilon \le 1 - x^{1/\lambda}\right] dx \,. \end{aligned}$$

Since v has nine grandchildren, by the union bound,

$$\begin{aligned} \mathbb{E}\left[(1-\Upsilon)^\lambda\right] &\le 9 \int_0^1 \mathbb{P}\left[B_1 B_2 \le 1 - x^{1/\lambda}\right] dx \\ &= \frac{9}{2} \int_0^1 \sqrt{1 - x^{1/\lambda}} \, \log\left(\frac{e^2}{1 - x^{1/\lambda}}\right) dx \,. \end{aligned}$$

With the change of variables $y = (\lambda - 1)(1 - x^{1/\lambda})$, we find

$$\begin{aligned} \mathbb{E}\left[(1-\Upsilon)^\lambda\right] &\le \frac{9\lambda}{2(\lambda-1)^{3/2}} \int_0^{\lambda-1} \sqrt{y} \, \log\left(\frac{e^2(\lambda-1)}{y}\right) \left(1 - \frac{y}{\lambda - 1}\right)^{\lambda-1} dy \\ &< \frac{9\lambda}{2(\lambda-1)^{3/2}} \int_0^{\lambda-1} \sqrt{y} \, \log\left(\frac{e^2(\lambda-1)}{y}\right) e^{-y} dy \,. \end{aligned}$$

Define

$$g_1(\lambda) = \frac{9\lambda}{2(\lambda-1)^{3/2}} \int_0^{\lambda-1} \sqrt{y} \, \log\left(\frac{e^2(\lambda-1)}{y}\right) e^{-y} dy \,.$$

Since $\lambda > 2$, we have

$$\int_0^{\lambda-1} \sqrt{y} \, \log\left(e^2(\lambda-1)\right) e^{-y} dy < \int_0^{\infty} \sqrt{y} \, \log\left(e^2(\lambda-1)\right) e^{-y} dy = \log\left(e^2(\lambda-1)\right) \sqrt{\pi}/2 \,,$$

and

$$\int_0^{\lambda-1} \sqrt{y} \, \log\left(1/y\right) e^{-y} dy \le \int_0^{1} \sqrt{y} \, \log\left(1/y\right) e^{-y} dy < \int_0^{1} \sqrt{y} \, \log\left(1/y\right) dy = 4/9 \,,$$

so $\mathbb{E}\left[(1-\Upsilon)^\lambda\right] < g_1(\lambda) < g(\lambda)$, and the proof is complete. ■

Chapter 5

The random-surfer Webgraph model

In this chapter[1] we study the random-surfer Webgraph model, which is similar to the evolving models we studied in Chapter 3: in every step a new vertex arrives and is joined to one or more vertices of the existing graph.

In the preferential attachment model and most of its variations (see, e.g., [9, 51, 54, 88]) the probability that the new vertex attaches to an old vertex v, called the *attraction* of v, is proportional to a deterministic function of the degree of v. In other variations (see, e.g., [14, 61]) the attraction also depends on the so-called 'fitness' of v, which is a random variable generated independently for each vertex and does not depend on the structure of the graph. For analyzing such models when they generate trees, a typical technique is to approximate them with population-dependent branching processes and prove that results on the corresponding branching processes carry over to the original models. For example, Pittel [111] estimated the height of random recursive trees. Bhamidi [13] used this technique to show that the height of a variety of preferential attachment trees is asymptotic to a constant times the logarithm of the number of vertices, where the constant depends on the parameters of the model.

In the random-surfer Webgraph model, however, the attraction of a vertex does not depend only on its degree, but rather on the general structure of the graph, so the branching processes techniques cannot apply directly, and new ideas are needed. The crucial novel idea in our arguments is to reduce the attachment rule to a simple one, with the help of introducing (possibly negative) 'weights' for the edges.

[1] This chapter is based on joint work with Nick Wormald. The results appear in the submitted manuscript [101], an extended abstract of which has been published [102].

Suppose that when a new vertex appears, it builds d new edges to old vertices; to bound the diameter, as we did in Chapter 3, we just employ the first created edge. In the special case $d = 1$, we obtain a random recursive tree with edge weights, and then we adapt a powerful technique developed by Broutin and Devroye [27] (that uses branching processes) to study its weighted height. This technique is based on large deviations, and we have already used a special case of their main result in Chapter 4 (see Theorem 4.12). Their main result [27, Theorem 1] is not applicable here for two reasons. Firstly, the weights of edges on the path from the root to each vertex are not independent, and secondly, the weights can be negative.

We now define the models and state the main results of this chapter. In Section 5.1 we give logarithmic upper bounds for diameters of the random-surfer Webgraph model and the PageRank-based selection Webgraph model in the general case $d \geq 1$. In Sections 5.2–5.4 we focus on the special case $d = 1$ and prove close lower and upper bounds. Section 5.2 contains the main technical contribution of this chapter, where we explain how to transform the random-surfer tree model into one that is easier to analyze. The lower and upper bounds are proved in Sections 5.3 and 5.4, respectively. Concluding remarks appear in Section 5.5.

Definition 5.1 (Random-surfer Webgraph model [16])**.** Let d be a positive integer and let $p \in (0, 1]$. Generate a random directed rooted n-vertex graph, possibly with loops and multiple edges, with all vertices having out-degree d. Start with a single vertex v_0, the root, with d self-loops. At each subsequent step $s \in [n-1]$, a new vertex v_s appears and d edges are created from it to vertices in $\{v_0, v_1, \ldots, v_{s-1}\}$, by doing the following probabilistic procedure d times, independently: choose a vertex u uniformly at random from $\{v_0, v_1, \ldots, v_{s-1}\}$, and a fresh random variable $X = \mathrm{Geo}(p)$; perform a simple random walk of length X starting from u, and join v_s to the last vertex of the walk.

The motivation behind this definition is as follows. Think of the vertex v_s as a new web page that is being set up. Say the owner wants to put d links in her web page. To build each link, she does the following: she goes to a random page. With probability p she likes the page and puts a link to that page. Otherwise, she clicks on a random link on that page, and follows the link to a new page. Again, with probability p she likes the new page and puts a link to that, otherwise clicks on a random link etc., until she finds a desirable page to link to. The geometric random variables correspond to this selection process.

Theorem 5.2. *Let d be a positive integer and let $p \in (0, 1]$. A.a.s. the diameter of the random-surfer Webgraph model with parameters p and d is at most $8e^p(\log n)/p$.*

Recall that the diameter of a directed graph is defined as the diameter of its underlying undirected graph, which is natural here since this model generates directed acyclic graphs (newer vertices always create edges to older vertices).

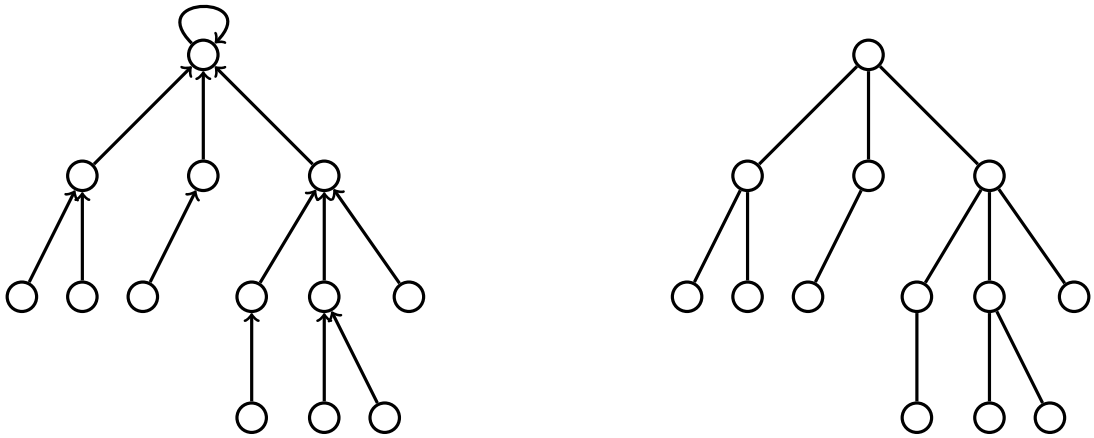


Figure 5.1: a random-surfer Webgraph (left) with $d = 1$ and its corresponding random-surfer tree (right)

Notice that the upper bound in Theorem 5.2 does not depend on d (whereas one would expect that the diameter must decrease asymptotically as d increases). This independence is because in our argument we employ only the first edge created by each new vertex to bound the diameter.

When $d = 1$, we show in Theorem 5.4 below that the diameter is a.a.s. $\Theta(\log n)$. An interesting open problem is to evaluate the asymptotic value of the diameter when $d > 1$. In this regime the diameter might be of a smaller order, e.g. $\Theta(\log n/\log\log n)$, as is the case for the preferential attachment model (see [20, Theorem 1]).

A *random-surfer tree* is an undirected tree obtained from a random-surfer Webgraph with $d = 1$ by deleting the self-loops of the root and ignoring the edge directions. See Figure 5.1 for an illustration.

Theorem 5.3. *For $p \in (0,1)$, let $s = s(p)$ be the unique solution in $(0,1)$ to*

$$s \log\left(\frac{(1-p)(2-s)}{1-s}\right) = 1\,. \tag{5.1}$$

Let $p_0 \approx 0.206$ be the unique solution in $(0,1/2)$ to

$$\log\left(\frac{1-p}{p}\right) = \frac{1-p}{1-2p}\,. \tag{5.2}$$

Define the functions $c_L, c_U : (0,1) \to \mathbb{R}$ as

$$c_L(p) = \exp(1/s)s(2-s)p\,,$$

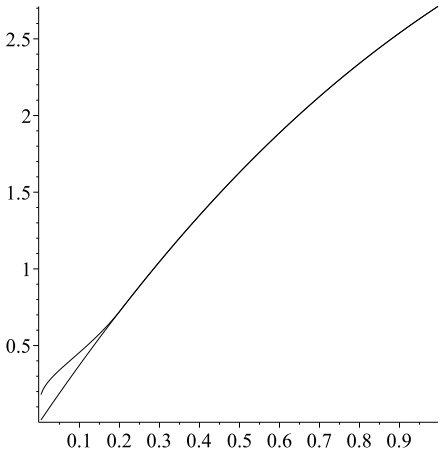


Figure 5.2: the functions c_L and c_U in Theorems 5.3 and 5.4

and

$$c_U(p) = \begin{cases} c_L(p) & \text{if } p_0 \le p < 1 \\ \left(\log\left(\frac{1-p}{p}\right)\right)^{-1} & \text{if } 0 < p < p_0\,. \end{cases}$$

For every fixed $\varepsilon > 0$, a.a.s. the height of the random-surfer tree model with parameter p is between $(c_L(p) - \varepsilon)\log n$ and $(c_U(p) + \varepsilon)\log n$.

The value p_0 and the functions c_L and c_U (plotted in Figure 5.2) are well defined by Lemma 5.21 below. Also, c_L and c_U are continuous, and $\lim_{p\to 0} c_L(p) = \lim_{p\to 0} c_U(p) = 0$ and $\lim_{p\to 1} c_L(p) = e$. We suspect that the gap between our bounds when $p < p_0$ is an artefact of our proof technique, and we do not expect a phase transition in the behaviour of the height at $p = p_0$.

We also prove lower and upper bounds for the diameter, which are close to being tight.

Theorem 5.4. *Let c_L and c_U be defined as in Theorem 5.3. For every fixed $\varepsilon > 0$, a.a.s. the diameter of the random-surfer tree model with parameter $p \in (0,1)$ is between $(2c_L(p) - \varepsilon)\log n$ and $(2c_U(p) + \varepsilon)\log n$.*

Immediately, we have the following corollary.

Corollary 5.5. *Let c_L and p_0 be defined as in Theorem 5.3. For any $p \in [p_0, 1)$, the height of the random-surfer tree model with parameter p is a.a.s. asymptotic to $c_L(p)\log n$ as $n \to \infty$, and its diameter is a.a.s. asymptotic to $2c_L(p)\log n$.*

A natural open problem is to close the gap between the lower and upper bounds in Theorems 5.3 and 5.4 when $p < p_0$. It seems that for solving this problem new ideas are required.

We now define the PageRank-based selection Webgraph model.

Definition 5.6 (PageRank-based selection Webgraph model [108])**.** Let d be a positive integer and let $\beta \in [0, 1]$ and $p \in (0, 1]$. The PageRank-based selection Webgraph model is a random n-vertex directed graph possibly with loops and multiple edges, with all vertices having out-degree d, generated as follows. It starts with a single vertex with d self-loops. At each subsequent step a new vertex appears, chooses d old vertices and attaches to them (where a vertex can be chosen multiple times). These choices are independent and the head of each edge is a uniformly random vertex with probability β, and is a vertex chosen according to the PageRank distribution (with restart probability p) with probability $1-\beta$.

Remark 5.7. The preferential attachment scheme is sometimes justified by stating that vertices with higher degrees are typically more important, so it is more likely that they are linked to in the future. The idea of the above definition is to replace the notion of degree with that of PageRank, which is supposed to be a better measure of importance of a vertex.

Theorem 5.8. *Let d be a positive integer and let $p, \beta \in (0, 1]$. A.a.s. as $n \to \infty$ the diameter of the PageRank-based selection Webgraph model with parameters d, p, and β is at most $8e^p(\log n)/p$.*

Chebolu and Melsted [31, Theorem 1.1] showed the random-surfer Webgraph model is equivalent to the PageRank-based selection Webgraph model with $\beta = 0$ (this fact also follows from Proposition 2.25). Hence Theorems 5.2 follows immediately from Theorem 5.8. Moreover, the conclusions of Theorems 5.3 and 5.4 apply to the PageRank-based selection Webgraph model with $\beta = 0$ and $d = 1$.

In Theorems 5.3 and 5.4 we have assumed that $p < 1$, since the situation for $p = 1$ has been clarified in previous work. Let $p = 1$. Then a random-surfer tree has the same distribution as a random recursive tree, the height of which is a.a.s. asymptotic to $e \log n$ as proved by Pittel [111]. It is not hard to alter his argument to prove the diameter is a.a.s. asymptotic to $2e \log n$. Note that this is consistent with our results, as $\lim_{p\to 1} c_L(p) = e$. For the rest of the chapter, we fix $p \in (0, 1)$.

We give high-level sketches of the proofs. For the general case $d \geq 1$, i.e. Theorems 5.2 and 5.8, we use a coupling with a random recursive tree, and sharp concentration bounds for sums of exponential and geometric random variables (Lemmas 2.21 and 2.22).

For the special case $d = 1$, i.e. Theorems 5.3 and 5.4, our results are sharper and the analysis is much more involved. First, we simplify the attachment procedure by introducing (possibly negative) weights for the edges. Then we embed the tree into a continuous-time branching process using the idea of poissonization. To estimate the height, we would like to apply the result of Broutin and Devroye [27] again, but it is not applicable since the 'weights' are not independent! So we adapt their technique and produce a variant of their theorem which does not require independence, but has a weaker conclusion. For proving this variant we use Galton-Watson branching processes (see Section 2.7). To apply this variant we need a couple of large deviation inequalities for sums of non-independent random variables, and establishing these takes a few pages of calculation. Finally, the answer is given implicitly as a solution of an equation involving maximization of a function. We solve this maximization problem via differentiation and a convexity argument.

We include some definitions here. Define the *depth* of a vertex as the length of a shortest path (ignoring edge directions) connecting the vertex to the root, and the *height* of a graph G, denoted by $\mathrm{ht}(G)$, as the maximum depth of its vertices. Clearly the diameter is at most twice the height. In a weighted tree (a tree whose *edges* are weighted), define the *weight* of a vertex to be the sum of the weights of the edges connecting the vertex to the root, and the *weighted height* of tree T, written $\mathrm{wht}(T)$, to be the maximum weight of its vertices. We view an unweighted tree as a weighted tree with unit edge weights, in which case the weight of a vertex is its depth, and the notion of weighted height is the same as the usual height.

We will need the following two large deviation inequalities, which have been proved in Section 2.5.

Define the function $\Upsilon : (0, \infty) \to \mathbb{R}$ as

$$\Upsilon(x) = \begin{cases} x - 1 - \log(x) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } 1 < x \,. \end{cases} \tag{5.3}$$

Lemma 5.9 (Cramér's Theorem for exponential random variables)**.** *Let $E_1, E_2, \ldots, E_m$ be independent exponential random variables with mean 1. For any fixed $x > 0$, as $m \to \infty$ we have*

$$\exp\left(-\Upsilon(x)m - o(m)\right) \leq \mathbb{P}\left[E_1 + E_2 + \cdots + E_m \leq xm\right] \leq \exp(-\Upsilon(x)m) \,.$$

Define the function $f : (-\infty, 1] \to \mathbb{R}$ as

$$f(x) = (2 - x)^{2-x} p (1 - p)^{1-x} (1 - x)^{x-1} \,. \tag{5.4}$$

Lemma 5.10. *Let $Z_1, Z_2, \ldots, Z_m$ be independent $1 + \mathrm{Geo}(p)$ random variables, and let $\kappa \geq 1/p$. Then we have $\mathbb{P}\left[Z_1 + Z_2 + \cdots + Z_m \geq \kappa m\right] \leq f(2 - \kappa)^m$.*

5.1 Upper bound for the PageRank-based model

In this section we prove Theorem 5.8, which gives an upper bound for the diameter of the PageRank-based selection Webgraph model. Theorem 5.2 follows immediately using [31, Theorem 1.1]. We need a technical lemma, whose proof follows from straightforward calculations.

Lemma 5.11. *Let η, c be positive numbers satisfying $\eta \geq 4e^p/p$ and $c \leq p\eta$. Then we have $-c\Upsilon(1/c) + c\log f(2-\eta/c) < \max\{\eta(1-p)\log(1-p^3), -0.15p\eta\} - 1$.*

Proof. We consider two cases.
Case 1: $c \geq 1$**.** In this case we prove

$$-c\Upsilon(1/c) + c\log f(2-\eta/c) < \eta(1-p)\log(1-p^3) - 1.$$

Notice that we have $1 - c\Upsilon(1/c) = c + c\log(1/c)$, so, using the definition of f and since $\eta(1-p) \leq \eta - c$, the conclusion is implied by

$$c + c\log(1/c) + \eta\log(\eta(1-p)/(\eta-c)) + c\log(p(\eta-c)/((1-p)c)) < (\eta-c)\log(1-p^3)\,.$$

Letting $r = \eta/c$ and since $c > 0$, this statement is equivalent to

$$ep(1-p)^{r-1}r^2(r/(r-1))^{r-1} < \eta(1-p^3)^{r-1}\,.$$

Since $(r/(r-1))^{r-1} < e$, and $1 - p < (1-p^3)e^{-p}$, for this inequality to hold it suffices to have

$$e^{2+p}r^2p\exp(-pr) \leq 4e^p/p \qquad \forall r \in [p^{-1}, \infty)\,,$$

which follows from the fact that $x^2e^{-x} \leq 4e^{-2}$ for all $x \geq 1$.

Case 2: $c < 1$**.** In this case we prove

$$-c\Upsilon(1/c) + c\log f(2-\eta/c) < -0.15p\eta - 1.$$

Since $\Upsilon(1/c) = 0$, this is equivalent to

$$1 + 0.15p\eta + c\log f(2-\eta/c) < 0. \tag{5.5}$$

Note that

$$\left(\frac{\eta/c}{\eta/c-1}\right)^{\eta/c-1} < e\,,$$

so we have

$$\begin{aligned} c \log f(2-\eta/c) &= \log\left((\eta/c)^{\eta} p^{c} (1-p)^{\eta-c} (\eta/c-1)^{c-\eta}\right) \\ &< \log\left((e\eta p/c)^{c}(1-p)^{\eta-c}\right) \leq c\log(e\eta p/c) + cp - p\eta\,, \end{aligned}$$

where we have used $\log(1-p) \leq -p$ in the last inequality. Hence to prove (5.5), since $c > 0$, it suffices to show that

$$\frac{1}{c} + 1 + \log(\eta p/c) + p < 0.85 p\eta/c\,. \tag{5.6}$$

Since $p\eta \geq 4e^{p} \geq 4 > 4c$, we have

$$\begin{aligned} \frac{1}{c} &< 0.25 p\eta/c, \\ 1 + p < \frac{1+p}{c} < \frac{e^{p}}{c} &\leq 0.25 p\eta/c, \\ \log(\eta p/c) &< 0.35 p\eta/c\,, \end{aligned}$$

which imply (5.6). ■

We now describe an alternative way to generate the edge destinations in the PageRank-based selection model. Define the non-negative random variable $\mathcal{L}$ as

$$\mathcal{L} = \mathcal{L}(p,\beta) = \begin{cases} 0 & \text{with probability } \beta\,, \\ \mathrm{Geo}(p) & \text{with probability } 1-\beta\,. \end{cases}$$

Note that $\mathrm{Geo}(p)$ stochastically dominates $\mathcal{L}$. The following lemma is a direct corollary of Proposition 2.25.

Lemma 5.12. *The head of each new edge in the PageRank-based selection model can be obtained by sampling a vertex u uniformly from the existing graph and performing a simple random walk of length $\mathcal{L}$ starting from u.*

We now have the ingredients to prove Theorem 5.8.

Proof of Theorem 5.8. Let $\eta = 4e^{p}/p$. As done several times in Chapter 3, we define an auxiliary tree whose node set equals the vertex set of the graph generated by the PageRank-based selection Webgraph model, and whose weighted height dominates the height of this

graph. Then we show a.a.s. this tree has weighted height at most $\eta \log n$, which completes the proof.

Initially the tree has just one vertex v_0. By Lemma 5.12, the growth of the PageRank-based selection model at each subsequent step $s \in [n-1]$ can be described as follows: a new vertex v_s appears and d edges are created from it to vertices in $\{v_0, v_1, \ldots, v_{s-1}\}$, by doing the following probabilistic procedure d times, independently: choose a vertex u uniformly at random from $\{v_0, v_1, \ldots, v_{s-1}\}$, and a fresh random variable $\mathcal{L}$; perform a simple random walk of length $\mathcal{L}$ starting from u, and join v_s to the last vertex of the walk.

Consider a step s and the first chosen $u \in \{v_0, \ldots, v_{s-1}\}$ and $\mathcal{L}$. In the tree, we join the vertex v_s to u and set the weight of the edge $v_s u$ to be $\mathcal{L} + 1$. Note that the edge weights are mutually independent. Since the (u, v_s)-distance in the graph is at most $\mathcal{L} + 1$, an inductive argument gives that for any vertex v, its weight in the auxiliary tree is greater than or equal to its depth in the graph. Hence, it suffices to show that a.a.s. the weighted height of the auxiliary tree is at most $\eta \log n$. We work with the tree in the rest of the proof.

Let us consider an alternative way to grow the tree, used by Devroye, Fawzi, and Fraiman [47], which results in the same distribution. Let $U_1, U_2, \ldots$ be i.i.d. uniform random variables in $(0, 1)$. Then for each new vertex v_s, we attach it to the vertex $v_{\lfloor sU_s \rfloor}$, which is indeed a vertex uniformly chosen from $\{v_0, \ldots, v_{s-1}\}$.

For convenience, we consider the tree when it has $n+1$ vertices $v_0, v_1, \ldots, v_n$. Let $D(s), W(s)$ denote the depth and the weight of vertex v_s, respectively. We have

$$\begin{aligned}\mathbb{P}\left[\text{wht(auxiliary tree)} > \eta \log n\right] &\leq \sum_{s=1}^{n} \mathbb{P}\left[W(s) > \eta \log n\right] \\ &\leq n\mathbb{P}\left[W(n) > \eta \log n\right] = \sum_{d=1}^{n} \mathcal{A}(d)\,,\end{aligned}$$

where we define

$$\mathcal{A}(d) = n\mathbb{P}\left[D(n) = d\right] \mathbb{P}\left[W(n) > \eta \log n | D(n) = d\right]\,.$$

To complete the proof it is enough to show $\sum_{d=1}^{n} \mathcal{A}(d) = o(1)$.

Let $P(0) = 0$ and for $s = 1, \ldots, n$, let $P(s)$ denote the index of the parent of v_s. We have

$$\mathbb{P}\left[D(n) \geq d\right] = \mathbb{P}\left[D(P(n)) \geq d-1\right] = \cdots = \mathbb{P}\left[D(P^{d-1}(n)) \geq 1\right] = \mathbb{P}\left[P^{d-1}(n) \geq 1\right]\,.$$

Since $P(m) = \lfloor mU_m \rfloor \le mU_m$ for each $0 \le m \le n$ and since the U_i are i.i.d., we have

$$\mathbb{P}\left[P^{d-1}(n) \ge 1\right] \le \mathbb{P}\left[nU_1U_2\dots U_{d-1} \ge 1\right] .$$

Let $E_i = -\log U_i$. Then $E_i \stackrel{d}{=} \mathrm{Exp}(1)$ (see Proposition 2.13), and moreover,

$$\begin{aligned}\mathbb{P}\left[D(n) \ge d\right] &\le \mathbb{P}\left[nU_1U_2\dots U_{d-1} \ge 1\right] \\ &= \mathbb{P}\left[E_1 + \dots + E_{d-1} \le \log n\right] \le \exp\left(-(d-1)\Upsilon\left(\frac{\log n}{d-1}\right)\right) , \end{aligned} \tag{5.7}$$

where we have used Lemma 5.9. The right-hand side is $o(1/n)$ for $d = 1.1e\log n$. Hence to complete the proof we need only show that

$$\mathcal{A}(d) = o(1/\log n) \qquad \forall d \in (0, 1.1e\log n) . \tag{5.8}$$

Fix an arbitrary positive integer $d \in (0, 1.1e\log n)$. The random variable $W(n)$, conditional on $D(n) = d$, is a sum of d i.i.d. $1+\mathcal{L}$ random variables. Since $\mathrm{Geo}(p)$ stochastically dominates $\mathcal{L}$, by Lemma 5.10 and since $\eta > 1.1e/p$, we have

$$\mathbb{P}\left[W(n) > \eta\log n | D(n) = d\right] \le f(2 - \eta\log n/d)^d , \tag{5.9}$$

with f defined in (5.4).

Combining (5.7) and (5.9), we get

$$\mathcal{A}(d) \le \exp\left[\log n - (d-1)\Upsilon\left(\frac{\log n}{d-1}\right) + d\log f(2 - \eta\log n/d)\right] . \tag{5.10}$$

Let $c = d/\log n$ and $c_1 = c - 1/\log n$. Let $\vartheta = \max\{\eta(1-p)\log(1-p^3), -0.15p\eta\}$. Note that ϑ is a negative constant. By Lemma 5.11 and since the function $c\Upsilon(1/c)$ is uniformly continuous on $[0, 1.1e]$, we find that for large enough n,

$$-c_1\Upsilon(1/c_1) + c\log f(2 - \eta/c) < \vartheta/2 - 1 .$$

Together with (5.10), this gives $\mathcal{A}(d) \le \exp(\vartheta\log n/2)$, and (5.8) follows. ■

5.2 Transformations of the random-surfer tree model

In Sections 5.2–5.4 we study the random-surfer tree model. In this section we show how to transform this model three times to obtain a new random tree model, which we analyze in subsequent sections. The first transformation is novel. The second one was perhaps first used by Broutin and Devroye [27], and the third one probably by Pittel [111].

Let us call the random-surfer tree model the *first model*. First, we will replace the attachment rule with a simpler one by introducing *weights* for the edges. In the first model, the edges are unweighted and in every step s a new vertex v_s appears, chooses an old vertex u, and attaches to a vertex on the path connecting u to the root, according to a certain rule. We introduce a second model that is weighted, and such that there is a one to one correspondence between the vertices in the second model and in the first model. For a vertex v in the first model, we denote its corresponding vertex in the second model by $\overline{v}$. In the second model, in every step s a new vertex $\overline{v_s}$ appears, chooses an old vertex $\overline{u}$ and attaches to $\overline{u}$, and the weight $w(\overline{u}\,\overline{v_s})$ of the new edge $\overline{u}\,\overline{v_s}$ is chosen such that the *weight* of $\overline{v_s}$ equals the *depth* of v_s in the first model. Let $w(\overline{u})$ denote the weight of vertex $\overline{u}$. Then it follows from the definition of the random-surfer tree model that

$$w(\overline{u}\,\overline{v_s}) \stackrel{d}{=} \max\{1 - \mathrm{Geo}(p), 1 - w(\overline{u})\}\,. \tag{5.11}$$

The term $\mathrm{Geo}(p)$ is the length of the random walk taken towards the root, and the term $1 - w(\overline{u})$ appears here solely because the weight of $\overline{v_s}$ is at least 1 (in the first model, the depth of v_s is at least 1, since it cannot attach to a vertex higher than the root), and so we must have

$$1 \le w(\overline{v_s}) = w(\overline{u}) + w(\overline{uv_s})\,.$$

Since the depth of v in the first model equals the weight of $\overline{v}$ in the second model, the height of the first model equals the weighted height of the second model (see Figure 5.3).

Remark 5.13. This extra term of $1 - w(\overline{u})$ in (5.11) causes a lot of trouble: it destroys the independence among the edge weights, which means the Broutin-Devroye result does not apply directly, and it is also the reason for the gap between c_L and c_U when $p < p_0$.

We will need to make the degrees of the tree bounded, so we define a third model. In this model, the new vertex can attach just to the leaves. In step s a new vertex v_s appears, chooses a random leaf u and joins to u using an edge with weight distributed as $\max\{1 - \mathrm{Geo}(p), 1 - w(\overline{u})\}$. Simultaneously, a new vertex u' appears and joins to u using an edge with weight 0. Then we have $w(u) = w(u')$ and henceforth u' plays the role of u, i.e. the next vertex wanting to attach to u, but cannot do so because u is no longer a leaf,

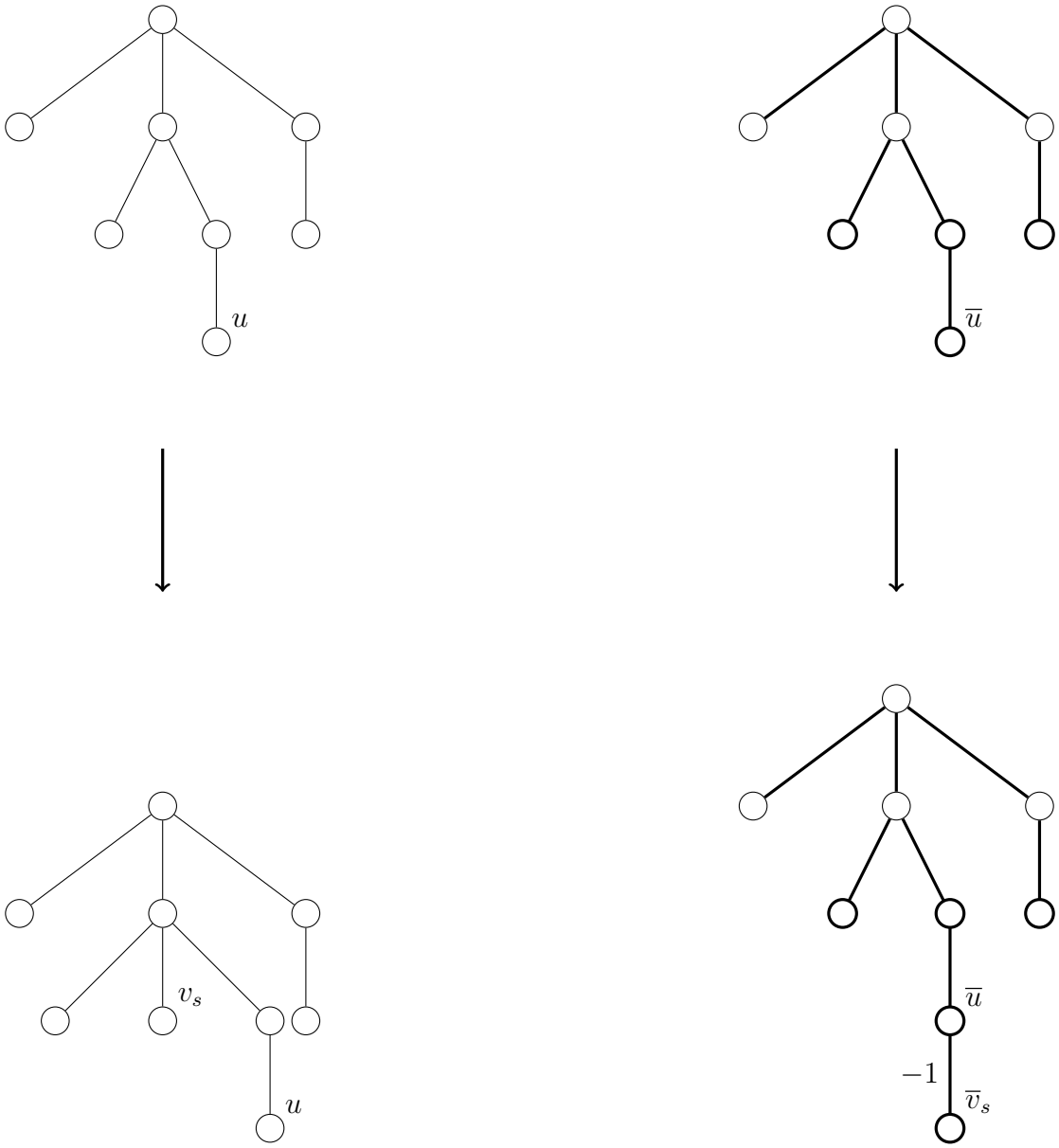


Figure 5.3: a birth in the first model (left) and the corresponding birth in the second model (right) are shown. In the first model, the chosen vertex is u and the length of the random walk is 2. In the second model, the new vertex $\overline{v}_s$ is joined to $\overline{u}$ with weight -1. Note that the depth of v_s in the second model equals the weight of $\overline{v}_s$ in the second model.

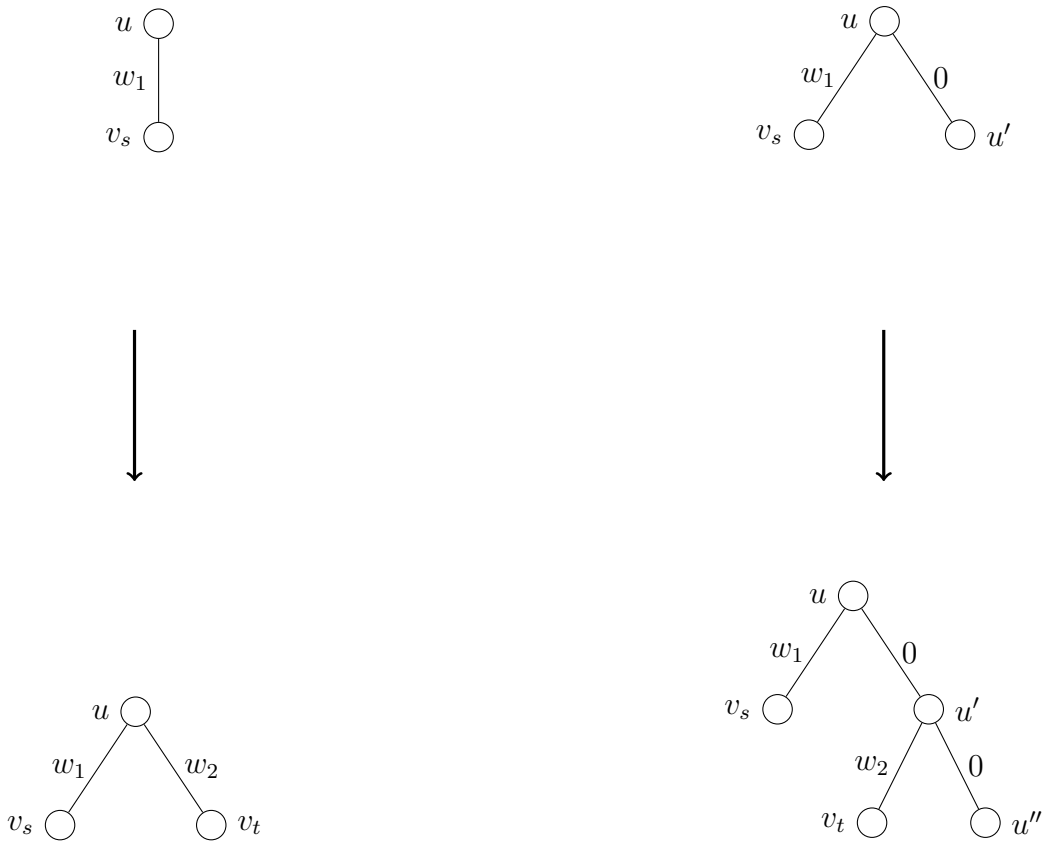


Figure 5.4: two births in the second model (left) and the corresponding births in the third model (right) are shown. In the second model a new vertex v_s is born and is joined to u with weight w_1. In the third model, at the same time a new vertex u' is born and is joined to u with weight 0. Later in the second model a new vertex v_t is born and is joined to u with weight w_2. In the third model v_t is joined to u' with weight w_2 and at the same time a new vertex u'' is born which plays the role of u.

may attach to u' instead, see Figure 5.4. Observe that the weighted height of the third model, when it has $2n-1$ vertices, has the same distribution as the weighted height of the second model with n vertices. In fact the second model may be obtained from the third one by contracting all zero-weight edges. We can thus study the weighted height of the first model by studying it in the third model.

All the above models were defined using discrete time steps. Employing the idea of poissonization explained in Section 2.4.1, we now define a fourth model using the following continuous time branching process, which we call $\mathcal{P}$. At time 0 the root is born. From this moment onwards, whenever a new vertex v is born (say at time κ), it waits for a random time E, which is distributed exponentially with mean 1, and after time E has passed (namely, at absolute time $\kappa+E$) gives birth to two children v_1 and v_2, and dies. The weights of the edges vv_1 and vv_2 are generated as follows: vertex v chooses $i \in \{1,2\}$ independently and uniformly at random. The weight of vv_i is distributed as $\max\{1-\mathrm{Geo}(p), 1-w(v)\}$ and the weight of vv_{3-i} is 0. Given $t \geq 0$, we denote by T_t the random tree obtained by taking a snapshot of this process at time t. By the discussion in in Section 2.4.1, for any stopping time τ, the distribution of T_τ, conditional on T_τ having $2n-1$ vertices, is the same as the distribution of the third model when it has $2n-1$ vertices.

The following lemma, which is very similar to Lemma 4.9, implies that certain results for T_t carry over to results for the random-surfer tree model.

Lemma 5.14. *Assume that there exist constants θ_L, θ_U such that for every fixed $\varepsilon > 0$,*

$$\mathbb{P}\left[\theta_L(1-\varepsilon)t \leq \mathrm{wht}(T_t) \leq \theta_U(1+\varepsilon)t\right] \to 1$$

as $t \to \infty$. Then for every fixed $\varepsilon > 0$, a.a.s. as $n \to \infty$ the height of the random-surfer tree model is between $\theta_L(1-\varepsilon)\log n$ and $\theta_U(1+\varepsilon)\log n$.

Proof. Let $\ell_n = 2n-1$, and let $\varepsilon > 0$ be fixed. For the process $\mathcal{P}$, we define three stopping times as follows:

a_1 is the deterministic time $(1-\varepsilon)\log(\ell_n)$.

A_2 is the random time when the evolving tree has exactly ℓ_n vertices.

a_3 is the deterministic time $(1+\varepsilon)\log(\ell_n)$.

By hypothesis, a.a.s. as $n \to \infty$ we have

$$(1-\varepsilon)\theta_L \log(\ell_n) \leq \mathrm{wht}\left(T_{a_1}\right) \text{ and } \mathrm{wht}\left(T_{a_3}\right) \leq (1+\varepsilon)\theta_U \log(\ell_n)\,. \tag{5.12}$$

By Proposition 2.18, as $t \to \infty$, we have $\log |V(T_t)| \sim t$. This means that, as $n \to \infty$, a.a.s.

$$\log |V(T_{a_1})| \sim a_1 = (1-\varepsilon)\log(\ell_n)\,,$$

and hence $|V(T_{a_1})| < \ell_n$, which implies $a_1 < A_2$. Symmetrically, it can be proved that a.a.s. as $n \to \infty$ we have $A_2 < a_3$. It follows that a.a.s. as $n \to \infty$

$$\mathrm{wht}\,(T_{a_1}) \le \mathrm{wht}\,(T_{A_2}) \le \mathrm{wht}\,(T_{a_3})\ . \tag{5.13}$$

On the other hand, as noted above, T_{A_2} has the same distribution as the third model with $2n-1$ vertices, whose weighted height has the same distribution as that of the random-surfer tree model with n vertices. Chaining (5.12) and (5.13) completes the proof. ■

It will be convenient to define T_t in a static way, which is equivalent to the dynamic definition above.

Definition 5.15 (T_∞, T_t)**.** Let T_∞ denote an infinite binary tree. To every edge e is associated a random vector (E_e, W_e) and to every vertex v a random variable W_v, where the W_e's and W_v's are the *weights*. The law for $\{E_e\}_{e\in E(T)}$ is easy: first with every vertex v we associate independently an Exp(1) random variable, and we let the values of E on the edges joining v to its two children be equal to this variable. In the dynamic interpretation, this random variable denotes the length of life of v. Generation of the weights is done in a top-down manner, where we think, somewhat ironically, of the root as the top vertex. Let the weight of the root be zero. Let v be a vertex whose weight has been determined, and let v_1, v_2 be its two children. Choose $i \in \{1,2\}$ independently and uniformly at random, and then choose $Y = 1 - \mathrm{Geo}(p)$ independently of previous choices. Then let

$$W_{vv_i} = \max\{Y, 1 - W_v\}, \quad W_{v_i} = W_v + W_{vv_i}\,, \tag{5.14}$$

and

$$W_{vv_j} = 0, \quad W_{v_j} = W_v$$

for $j = 3 - i$.

For a vertex v, let $\pi(v)$ denote the set of edges of the unique path connecting v to the root. Note that the weight of any vertex v equals $\sum_{e\in\pi(v)} W_e$. We define the *birth time* of a vertex v, written B_v, as

$$B_v = \sum_{e\in\pi(v)} E_e\,,$$

and the birth time of the root is defined to be zero. Finally, given $t \ge 0$, we define T_t as the subtree of T_∞ induced by vertices with birth time at most t. Note that T_t is finite almost surely.

5.3 Lower bounds for the random-surfer tree model

Here we prove the lower bounds in Theorems 5.3 and 5.4. For this, we consider another infinite binary tree T'_∞ which is very similar to T_∞, except for the generation rules for the weights, which are as follows. Let the weight of the root be zero. Let v be a vertex whose weight has been determined, and let v_1, v_2 be its two children. Choose $i \in \{1, 2\}$ independently and uniformly at random, and choose $Y = 1 - \mathrm{Geo}(p)$ independently of previous choices. Then let

$$W_{vv_i} = Y \text{ and } W_{v_i} = W_v + W_{vv_i} \tag{5.15}$$

and

$$W_{vv_j} = 0 \text{ and } W_{v_j} = W_v$$

for $j = 3 - i$. Comparing (5.15) with (5.14), we find that the weight of every vertex in T'_∞ is stochastically less than or equal to that of its corresponding vertex in T_∞. The tree T'_t is defined as before. Clearly probabilistic lower bounds for $\mathrm{wht}(T'_t)$ are also probabilistic lower bounds for $\mathrm{wht}(T_t)$. Distinct vertices u and v in a tree are called *antipodal* if the unique (u, v)-path in the tree passes through the root.

Lemma 5.16. *Consider the tree T'_∞. Let $\gamma_L : (0, 1) \to \mathbb{R}$ be such that for every $a \in (0, 1)$, each vertex u and each descendant v of u that is m levels deeper,*

$$\mathbb{P}\left[W_v - W_u \geq am\right] \geq \exp(-m\gamma_L(a) - o(m)) \tag{5.16}$$

as $m \to \infty$. Assume that there exist $\alpha^*, \rho^* \in (0, 1)$ with

$$\gamma_L(\alpha^*) + \Upsilon(\rho^*) = \log 2\,. \tag{5.17}$$

Then for every fixed $\varepsilon > 0$, a.a.s. there exist antipodal vertices u, v of T'_t with weights at least $\frac{\alpha^*}{\rho^*}(1 - \varepsilon)t$.

The proof is very similar to the proof of [27, Lemma 4] except a small twist is needed at the end to handle the negative weights.

Proof. Let $c = \frac{\alpha^*}{\rho^*}$, and let $\varepsilon, \delta > 0$ be arbitrary. We prove that with probability at least $1 - \delta$ for all large enough t there exists a pair (u, v) of antipodal vertices of T'_∞ with $\max\{B_u, B_v\} < t$ and $\min\{W_u, W_v\} > (1 - 2\varepsilon)\, ct$.

Let L be a constant positive integer that will be determined later, and let $\alpha = \alpha^*$ and $\rho = \frac{\alpha}{c(1-\epsilon)} > \rho^*$. By (5.17) and since $\rho^* < 1$ and Υ is strictly decreasing on $(0, 1]$, we have

$$\gamma_L(\alpha) + \Upsilon(\rho) < \log 2 \,.$$

Build a Galton-Watson process (see Section 2.7 for the definition) from T'_∞ whose particles are a subset of vertices of T'_∞, as follows. Start with the root as the initial particle of the process. If a given vertex u is a particle of the process, then its potential offspring are its 2^L descendants that are L levels deeper. Moreover, such a descendant v is an offspring of u if and only if $W_v - W_u \geq \alpha L$ and $B_v - B_u \leq \rho L$. As these two events are independent, the expected number of children of u is at least

$$2^L \mathbb{P}\left[W_v - W_u \geq \alpha L\right] \mathbb{P}\left[B_v - B_u \leq \rho L\right] \geq \exp\left[(\log 2 - \gamma_L(\alpha) - \Upsilon(\rho) - o(1))L\right]$$

as $L \to \infty$, by (5.16) and Lemma 5.9 (Cramér's Theorem for exponential random variables). Since we have $\log 2 - \gamma_L(\alpha) - \Upsilon(\rho) > 0$, we may choose L large enough that this expected value is strictly greater than 1. Therefore, this Galton-Watson process survives with probability $q > 0$ (see Section 2.7).

We now boost this probability up to $1 - \delta$, by starting several independent processes, giving more chance that at least one of them survives. Specifically, let b be a constant large enough that

$$(1-q)^{2^{b-1}} < \delta/3 \,.$$

Consider 2^b Galton-Watson processes, which have the vertices at depth b of T'_∞ as their initial particles, and reproduce using the same rule as before. Let a be a constant large enough that

$$2^{b+1}(e^{-a} + (1-p)^{a+2}) < \delta/3 \,,$$

and let A be the event that all edges e in the top b levels of T'_∞ have $E_e \leq a$ and $W_e \geq -a$. Then

$$1 - \mathbb{P}\left[A\right] \leq 2^{b+1}(e^{-a} + (1-p)^{a+2}) < \delta/3 \,.$$

Also, let Q be the event that in each of the two branches of the root, at least one of the 2^{b-1} Galton-Watson processes survives. Then

$$1 - \mathbb{P}\left[Q\right] \leq 2(1-q)^{2^{b-1}} < 2\delta/3 \,,$$

and so with probability at least $1 - \delta$ both A and Q happen.

Assume that both A and Q happen. Let

$$m = \left\lfloor \frac{t(1-\varepsilon)}{\rho L} \right\rfloor$$

and let u and v be particles at generation m of surviving processes in distinct branches of the root. Then u and v are antipodal,

$$\max\{B_u, B_v\} \le ab + m\rho L \le t(1-\varepsilon) + O(1) < t\,,$$

and

$$\min\{W_u, W_v\} \ge -ab + m\alpha L \ge \frac{(1-\varepsilon)\alpha}{\rho}\, t - O(1) > c(1-2\varepsilon)t$$

for t large enough, as required. ■

Let $Y_1, Y_2, \ldots$ be i.i.d. with $Y_i = 1 - \mathrm{Geo}(p)$. Recall the definition of $f : (-\infty, 1] \to \mathbb{R}$ from (5.4):

$$f(x) = (2-x)^{2-x} p (1-p)^{1-x} (1-x)^{x-1}\,.$$

Note that $f(1) = p$ since by convention $0^0 = 1$, and $f\left(2 - p^{-1}\right) = 1$. The following lemma follows by noting that f is positive and the derivative of $\log f$ is $\log\left(\frac{1-x}{(2-x)(1-p)}\right)$.

Lemma 5.17. *The function f is continuous in $(-\infty, 1]$ and differentiable in $(-\infty, 1)$. Moreover, f is increasing on $(-\infty, 2-p^{-1}]$ and decreasing on $[2-p^{-1}, 1]$.*

Lemma 5.18. *(a) There is an absolute constant C such that for any $a \in [2-p^{-1}, 1]$ and any positive integer m we have*

$$\mathbb{P}\left[Y_1 + \cdots + Y_m \ge am\right] \le Cm f(a)^m\,.$$

(b) As $m \to \infty$, uniformly for all $a \in [0,1]$ we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_m \ge am\right] \ge \left[f(a) - o(1)\right]^m\,.$$

(c) If $p \ge 1/2$, then as $m \to \infty$, uniformly for all $a \in [0, 2 - \frac{1}{p}]$ we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_m \ge am\right] \ge \left[1 - o(1)\right]^m\,.$$

Proof. The conclusions are easy to see for $a = 1$, so assume that $a < 1$. First, assume that am is an integer. Consider a sequence of independent biased coin flips, each of which is heads with probability p. A random walker starts from 0, takes one step to the right on seeing heads, and one to the left on seeing tails. Then $Y_1 + \cdots + Y_m$ is the walker's position just after seeing the m-th head. Thus $Y_1 + \cdots + Y_m = am$ if and only if the $(2m - am)$-th

coin comes up heads, and in the first $2m-am$ coin flips we see exactly m heads and $m-am$ tails, so we have

$$\begin{aligned}\mathbb{P}\left[Y_1+\cdots+Y_m=am\right] &= \binom{2m-am-1}{m-1}p^m(1-p)^{m-am}\\ &= \Theta\left(\binom{2m-am}{m}p^m(1-p)^{m-am}\right)\\ &= \Theta\left(f(a)^m/\sqrt{m}\right)\,, \end{aligned} \tag{5.18}$$

where we have used Stirling's approximation (2.4) for the last equality.

(a) Let $a\in[2-p^{-1},1)$, and let C be an absolute constant for the upper bound of Θ in (5.18). Then

$$\begin{aligned}\mathbb{P}\left[Y_1+\cdots+Y_m\ge am\right] &\le m\sup\{\mathbb{P}\left[Y_1+\cdots+Y_m=\alpha m\right]:\alpha\in[a,1]\}\\ &\le C\sqrt{m}\left[\sup\{f(\alpha):\alpha\in[a,1]\}\right]^m\le Cm(f(a))^m\end{aligned}$$

since f is decreasing on $\left[2-\frac{1}{p},1\right]$ by Lemma 5.17 and m is a positive integer.

(b) Assume that $m\to\infty$. Then

$$\mathbb{P}\left[Y_1+\cdots+Y_m\ge am\right]\ge\mathbb{P}\left[Y_1+\cdots+Y_m=\lceil am\rceil\right]=(f(a)-o(1))^m$$

uniformly for all $a\in[0,1)$ by continuity of f.

(c) Assume that $p\ge1/2$ and that $m\to\infty$. Then

$$\begin{aligned}\mathbb{P}\left[Y_1+\cdots+Y_m\ge am\right] &\ge \mathbb{P}\left[Y_1+\cdots+Y_m=\left\lceil\left(2-p^{-1}\right)m\right\rceil\right]\\ &= \left(f\left(2-p^{-1}\right)-o(1)\right)^m=(1-o(1))^m\end{aligned}$$

uniformly for all $a\in[0,2-\frac{1}{p}]$ by continuity of f and since $f(2-p^{-1})=1$. ■

We define a two variable function

$$\Phi(a,s)=p(1-p)(2-s)^2(s-a)-a(1-s)\,, \tag{5.19}$$

and we define a function $\phi:[0,1]\to[0,1]$ as follows: given $a\in[0,1]$, $\phi(a)$ is the unique solution in $[a,1]$ to

$$\Phi(a,\phi(a))=p(1-p)(2-\phi(a))^2(\phi(a)-a)-a(1-\phi(a))=0\,. \tag{5.20}$$

Lemma 5.19(a) below shows that ϕ is well defined. The proof is straightforward.

Lemma 5.19. *(a) Given $a \in [0,1]$, there is a unique solution $s \in [a,1]$ to $\Phi(a,s) = 0$. If $a \in \{0,1\}$ then $\phi(a) = a$. If $a \in (0,1)$ then $0 < a < \phi(a) < 1$.*

(b) If $s = \phi(a)$ then

$$\frac{sf(s)^{a/s}}{a^{a/s}(s-a)^{1-\frac{a}{s}}} = \frac{s}{s-a}\left(\frac{1-s}{(1-p)(2-s)}\right)^a .$$

(c) The function ϕ is increasing on $[0,1]$ and differentiable on $(0,1)$.

(d) The function ϕ is invertible and ϕ^{-1} is increasing. If $s \in \{0,1\}$ then $\phi^{-1}(s) = s$. If $s \in (0,1)$ then $0 < \phi^{-1}(s) < s < 1$.

Proof. (a) The conclusion is clear for $a \in \{0,1\}$, so we may assume that $a \in (0,1)$. Since $\Phi(a,a) < 0$ and $\Phi(a,1) > 0$, there exists at least one $s \in (a,1)$ with $\Phi(a,s) = 0$. We now show that there is a unique such s. Fixing a, since Φ is differentiable with respect to s, it is enough to show that

$$\text{if } \Phi(a,s) = 0, \text{ then } \frac{\partial \Phi}{\partial s} > 0 \tag{5.21}$$

Let $\sigma = p(1-p)$. We have

$$\frac{\partial \Phi}{\partial s} = \sigma(2-s)^2 + a - 2\sigma(2-s)(s-a) .$$

At a point (a,s) with $\Phi(a,s) = 0$, we have

$$\sigma(2-s)^2 = \frac{a(1-s)}{s-a}, \text{ and } \sigma(2-s)(s-a) = \frac{a(1-s)}{2-s} ,$$

so at this point,

$$\frac{\partial \Phi}{\partial s} = a\left(\frac{1}{s-a} + \frac{1}{1-s} - \frac{2}{2-s}\right) ,$$

which is strictly positive because

$$\min\left\{\frac{1}{s-a}, \frac{1}{1-s}\right\} > 1 > \frac{1}{2-s} ,$$

and this proves (5.21).

(b) Plugging the definition of f from (5.4) and using $\Phi(a,s) = 0$ gives this equation.

(c) We first show that ϕ is differentiable and increasing on $(0,1)$. Let $a \in (0,1)$ and let $s = \phi(a)$. We have

$$\frac{\partial \Phi}{\partial a} = s - 1 - p(1-p)(2-s)^2 < 0\,,$$

and $\partial\Phi/\partial s$ is positive as proved in part (a). Hence by the implicit function theorem $\mathrm{d}s/\mathrm{d}a$ exists and is positive, so ϕ is differentiable and increasing on $(0,1)$. Since $\phi(0) = 0$ and $\phi(1) = 1$, ϕ is increasing on $[0,1]$.

(d) Let $s \in [0,1]$. Then $\Phi(0,s)\Phi(s,s) \le 0$ and so there exists at least one $a_0 \in [0,s]$ with $\Phi(a_0,s) = 0$. The function $\Phi(a,s)$ is linear in a and the coefficient of a is non-zero, hence this root a_0 is unique. The function ϕ^{-1} is increasing since ϕ is increasing. The last two statements follow from similar statements proved for ϕ in (a). ■

Next let $\hat{Y}_1, \hat{Y}_2, \dots$ be independent and distributed as follows: for every $i = 1, 2, \dots$ we flip an unbiased coin, if it comes up heads, then $\hat{Y}_i = Y_i$, otherwise $\hat{Y}_i = 0$.

Define the function $g_L : (0,1) \to \mathbb{R}$ as

$$g_L(a) = \begin{cases} 1/2 & \text{if } p > 1/2 \text{ and } 0 < a < 1 - \frac{1}{2p} \\ \frac{\phi(a)-a}{\phi(a)} \left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)} \right)^a & \text{otherwise.} \end{cases}$$

Note that g_L is continuous as $\phi(1 - \frac{1}{2p}) = 2 - 1/p$. The proofs of the following two lemmas are standard and can be skipped on a first reading.

Lemma 5.20. *We have the following large deviation inequality for every fixed $a \in (0,1)$ as $m \to \infty$.*

$$\mathbb{P}\left[\hat{Y}_1 + \cdots + \hat{Y}_m \ge am\right] \ge (2g_L(a) - o(1))^{-m}\,.$$

Proof. We have

$$\mathbb{P}\left[\hat{Y}_1 + \cdots + \hat{Y}_m \ge am\right] = \sum_{k=\lceil am \rceil}^{m} \binom{m}{k} 2^{-m} \times \mathbb{P}\left[Y_1 + \cdots + Y_k \ge am\right]\,,$$

where k denotes the number of $\hat{Y}_i$'s whose value was determined to be equal to Y_i.

If $p > 1/2$ and $0 < a < 1 - \frac{1}{2p}$, then letting $k = \lceil m/2 \rceil$ gives

$$\binom{m}{k} 2^{-m} = \Omega\left(\frac{1}{\sqrt{m}}\right)$$

by Stirling's approximation (2.4), and

$$\mathbb{P}\left[Y_1 + \cdots + Y_k \geq am\right] \geq (1 - o(1))^k$$

by Lemma 5.18(c). This gives

$$\mathbb{P}\left[\hat{Y}_1 + \cdots + \hat{Y}_m \geq am\right] \geq (1 - o(1))^m \,,$$

as required.

Otherwise, let $s = \phi(a)$. Then letting $k = \lceil am/s \rceil$ gives

$$\binom{m}{k} 2^{-m} = \Omega\left(\left[\frac{s(s-a)^{a/s}}{2(s-a)a^{a/s}}\right]^m\right) \Big/ m^2$$

by Stirling's approximation (2.4), and

$$\mathbb{P}\left[Y_1 + \cdots + Y_k \geq am\right] \geq (f(s) - o(1))^k$$

by Lemma 5.18(b) and since f is continuous. Lemma 5.19(b) completes the proof. ■

Lemma 5.21. *(a) There exists a unique solution $p_0 \in (0, 1/2)$ to*

$$\log\left(\frac{1-p}{p}\right) = \frac{1-p}{1-2p} \,.$$

Also, if $p \leq p_0$ then $\log\left(\frac{1-p}{p}\right) \geq \frac{1-p}{1-2p}$.

(b) Given $p \in (0, 1)$, there exists a unique solution $s_0 \in (0, 1)$ to

$$(1-p)(2-s) = \exp(1/s)(1-s) \,.$$

Moreover, if $p > 1/2$ then $s_0 > 2 - p^{-1}$, and if $p_0 < p \leq 1/2$ then $s_0 > \frac{1-2p}{1-p}$.

Proof. (a) The function $r(p) = \log\left(\frac{1-p}{p}\right) - \frac{1-p}{1-2p}$ approaches $+\infty$ when $p \to 0^+$ and approaches $-\infty$ when $p \to \frac{1}{2}^-$. Moreover,

$$r'(p) = \frac{-1}{p(1-p)} - \frac{1}{(1-2p)^2} < 0$$

for $p \in (0, 1/2)$. Hence $r(p)$ has a unique root p_0, and $r(p) \geq 0$ if and only if $p \leq p_0$.

(b) The function

$$\mu(s) = \log(1-p) + \log(2-s) - \log(1-s) - \frac{1}{s}$$

approaches $-\infty$ as $s \to 0^+$, and approaches $+\infty$ as $s \to 1^-$, and its derivative is positive in $(0,1)$, hence it has a unique root s_0 in $(0,1)$. Also we have $\mu(2-p^{-1}) = p/(1-2p)$, which means that if $p > 1/2$ then $s_0 > 2 - p^{-1}$. Moreover, if $p_0 < p \le 1/2$, then by part (a),

$$\mu\left(\frac{1-2p}{1-p}\right) = \log\left(\frac{1-p}{p}\right) - \frac{1-p}{1-2p} = r(p) < 0\,,$$

which means $s_0 > \frac{1-2p}{1-p}$. ■

Lemma 5.22. *Given $\varepsilon > 0$, a.a.s as $t \to \infty$ there exist two antipodal vertices u, v of T'_t with weights at least $c_L(p)(1-\varepsilon)t$. In particular, a.a.s. the weighted height of T'_t is at least $c_L(p)(1-\varepsilon)t$.*

Proof. By Lemma 5.21(b), there is a unique solution $s \in (0,1)$ to

$$(1-p)(2-s) = \exp(1/s)(1-s)\,.$$

By the definition of c_L,

$$c_L = c_L(p) = \exp(1/s)s(2-s)p\,.$$

Lemma 5.20 implies that the assumption (5.16) of Lemma 5.16 holds for the function $\gamma_L(a) = \log(2g_L(a))$. Let $a = \phi^{-1}(s)$ and let $\rho = 1 - \frac{a}{s}$. Since $s \in (0,1)$ we have $0 < a < s < 1$ by Lemma 5.19(d), and thus $\rho \in (0,1)$ as well. Moreover, since $\Phi(a,s) = 0$, we have $c_L = a/\rho$.

We now show that $g_L(a) = \frac{s-a}{s}\exp(a/s)$. This is clear if $p \le 1/2$, so assume that $p > 1/2$. It is easy to verify that $\Phi(1-\frac{1}{2p}, 2-\frac{1}{p}) = 0$. Since $p > 1/2$, by Lemma 5.21(b) we have $s > 2 - \frac{1}{p}$. Since ϕ^{-1} is increasing, we have $a = \phi^{-1}(s) \ge 1 - \frac{1}{2p}$.

From $g_L(a) = \frac{s-a}{s}\exp(a/s)$ we get

$$\log(2g_L(a)) + \rho - 1 - \log(\rho) = \log 2\,,$$

and Lemma 5.16 completes the proof. ■

The lower bound in Theorem 5.3 follows from Lemmas 5.22 and 5.14.

Proof of the lower bound in Theorem 5.4. Fix $\varepsilon > 0$. Let us define the *semi-diameter* of a tree as the maximum weighted distance between any two antipodal vertices. Clearly, semi-diameter is a lower bound for the diameter, so we just need to show a.a.s. as $n \to \infty$ the semi-diameter of the random-surfer model with n vertices is at least $(2c_L(p) - \varepsilon) \log n$. By Lemma 5.22, a.a.s as $t \to \infty$ the semi-diameter of T'_t is at least $(2c_L(p) - \varepsilon)t$. Using an argument similar to the proof of Lemma 5.14 we may conclude that a.a.s. as $n \to \infty$ the semi-diameter of the third model (of Section 5.2) with $2n - 1$ vertices is at least $(2c_L(p) - \varepsilon) \log n$. It is easy to observe that this statement is also true for the random-surfer model with n vertices, and the proof is complete. ■

5.4 Upper bounds for the random-surfer tree model

In this section we prove the upper bounds in Theorems 5.3 and 5.4. As in the previous section, a lot of algebra is involved.

Lemma 5.23. *Let $\gamma_U : [0,1] \to [0,\infty)$ be a continuous function such that for every fixed $a \in [0,1]$ and every vertex v of T_∞ at depth m,*

$$\mathbb{P}\left[\sum_{e \in \pi(v)} W_e > am\right] \leq \exp(-m\gamma_U(a) + o(m)) \tag{5.22}$$

as $m \to \infty$. Define

$$\theta = \sup\left\{\frac{a}{\rho} : \gamma_U(a) + \Upsilon(\rho) = \log 2 : a \in [0,1], \rho \in (0,\infty)\right\}. \tag{5.23}$$

Then for every fixed $\varepsilon > 0$,

$$\mathbb{P}\left[\mathrm{wht}(T_t) > \theta(1+\varepsilon)t\right] \to 0$$

as $t \to \infty$.

The proof is similar to that of [27, Lemma 3], in which the assumption (5.22) is not needed. In fact, in the model studied in [27], the weights $\{W_e : e \in \pi(v)\}$ are mutually independent, and the authors use Cramér's Theorem (Theorem 2.20) to obtain a large deviation inequality for $\sum_{e \in \pi(v)} W_e$, which is similar to (5.22).

Proof. We first prove a claim.

Claim 5.24. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\rho \in \left(0, \frac{1}{\theta(1+\varepsilon)}\right]$,*

$$\Upsilon(\rho) + \gamma_U(\theta(1+\varepsilon)\rho) - \log 2 \geq \delta \,.$$

Proof of Claim. Assume that this is not the case for some $\varepsilon > 0$. This means there exists a sequence $(\rho_i)_{i\in\mathbb{N}}$ such that for all $i \in \mathbb{N}$,

$$\Upsilon(\rho_i) + \gamma_U(\theta(1+\varepsilon)\rho_i) - \log 2 < 1/i \,.$$

Then $(\rho_i)_{i\in\mathbb{N}}$ has a convergent subsequence. Let $\rho^* \in \left[0, \frac{1}{\theta(1+\varepsilon)}\right]$ be the limit. It cannot be the case that $\rho^* = 0$ since $\Upsilon(x) \to \infty$ as $x \to 0$, and γ_U is non-negative. By continuity of Υ and γ_U we have

$$\Upsilon(\rho^*) + \gamma_U(\theta(1+\varepsilon)\rho^*) - \log 2 \leq 0 \,.$$

Since Υ is continuous, decreasing, and attains all values in $[0, \infty)$, we can choose $\rho' \leq \rho^*$ so that

$$\Upsilon(\rho') + \gamma_U(\theta(1+\varepsilon)\rho^*) - \log 2 = 0 \,,$$

But then

$$\frac{\theta(1+\varepsilon)\rho^*}{\rho'} \geq \theta(1+\varepsilon) > \theta \,,$$

contradicting the definition of θ in (5.23). □

Fix $\varepsilon > 0$ and let A_k be the event that there exists a vertex at depth k of T_t with weight larger than $\theta(1+\varepsilon)t$. By the union bound,

$$\mathbb{P}\left[\mathrm{wht}(T_t) > \theta(1+\varepsilon)t\right] \leq \sum_{k=1}^{\infty} \mathbb{P}\left[A_k\right] = \sum_{k>\theta(1+\varepsilon)t} \mathbb{P}\left[A_k\right] \,,$$

as the weights of each edge is at most 1 (see (5.14)).

Let $k > \theta(1+\varepsilon)t$ be a positive integer. A vertex v at depth k of T_∞ is included in T_t and has weight larger than $\theta(1+\varepsilon)t$ if and only if $B_v \leq t$ and $W_v > \theta(1+\varepsilon)t$. These two events are independent by the definition of T_∞. The random variable B_v is distributed as a sum of k independent exponential random variables with mean 1, and so

$$\begin{aligned}\mathbb{P}\left[B_v \leq t, W_v > \theta(1+\varepsilon)t\right] &\leq \exp\left[\left(-\Upsilon(t/k) - \gamma_U\left(\frac{\theta(1+\varepsilon)t}{k}\right) + o(1)\right)k\right] \\ &\leq \exp\left[(-\log 2 - \delta + o(1))k\right] \,,\end{aligned}$$

where we have used Lemma 5.9 and (5.22) for the first inequality, and $\delta > 0$ is the constant provided by the claim. Since there are 2^k vertices at depth k of T_∞, by the union bound

$$\mathbb{P}[A_k] \le 2^k \exp[(-\log 2 - \delta + o(1))k] \le \exp[(-\delta + o(1))k] \ .$$

The $o(1)$ term is less than $\delta/2$ for large enough t, and thus

$$\mathbb{P}[\mathrm{wht}(T_t) > \theta(1+\varepsilon)t] \le \sum_{k>\theta(1+\varepsilon)t} \mathbb{P}[A_k] \le \sum_{k>\theta(1+\varepsilon)t} e^{-k\delta/2} = O\left(e^{-\theta(1+\varepsilon)t\delta/2}\right) = o(1) \ ,$$

as required. ■

Let $Y_1, Y_2, \ldots$ be i.i.d. with $Y_i = 1 - \mathrm{Geo}(p)$, and define random variables $X_1, X_2, \ldots$ as follows:

$$X_1 = \max\{Y_1, 1\} \ ,$$

and for $i \ge 1$,

$$X_{i+1} = \max\{Y_{i+1}, 1 - (X_1 + \cdots + X_i)\} \ .$$

Define the function $h : [0,1] \to \mathbb{R}$ as

$$h(x) = \begin{cases} 1 & \text{if } p \ge \frac{1}{2} \text{ and } 0 \le x \le 2 - \frac{1}{p} \\ \left(\frac{p}{1-p}\right)^x & \text{if } p < \frac{1}{2} \text{ and } 0 \le x \le \frac{1-2p}{1-p} \\ (2-x)^{2-x} p^{1-x} (1-p)^{1-x} (1-x)^{x-1} & \text{otherwise} \ . \end{cases} \tag{5.24}$$

Note that in the third case we have $h(x) = f(x)$, where f is defined in (5.4). It is easy to see that h is continuous. We now prove a lemma whose proof follows from straightforward calculations.

Lemma 5.25. *There exists an absolute constant C such that for every $a \in [0,1]$ and every positive integer m we have $\mathbb{P}[X_1 + \cdots + X_m > am] \le Cm^2 h(a)^m$.*

Proof. The conclusion is obvious if $p \ge \frac{1}{2}$ and $a \le 2 - \frac{1}{p}$, or if $a = 0$, since in these cases $h(a) = 1$. Also, $\mathbb{P}[X_1 + \cdots + X_m > m] = 0$ so the conclusion is true if $a = 1$, so we may assume that $\max\{0, 2 - \frac{1}{p}\} < a < 1$.

Observe that if $X_1 + \cdots + X_m > am$, there is a subsequence of the form $Y_{m-k+1}, \ldots, Y_m$ whose sum is at least am, and this subsequence contains at least am elements since $Y_i \le 1$ for all i. Hence we have

$$\mathbb{P}[X_1 + \cdots + X_m > am] \le m \max\{\mathbb{P}[Y_1 + \cdots + Y_k \ge am] : k \in [am, m] \cap \mathbb{N}\}$$

as the Y_i's are i.i.d.

For any integer $k \in [am, m]$, by Lemma 5.18(a) we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_k \geq am\right] \leq Ck(f(am/k))^k$$

for an absolute constant C, since $am/k \geq a > 2 - \frac{1}{p}$. Let $r = k/m \in [a, 1]$. So we find that

$$\mathbb{P}\left[X_1 + \cdots + X_m > am\right] \leq Cm^2\big(\sup\{f(a/r)^r : r \in [a,1]\}\big)^m .$$

Let us define

$$\xi(r) = f(a/r)^r = (2r-a)^{2r-a}p^r(1-p)^{r-a}(r-a)^{a-r}r^{-r} .$$

So to complete the proof we just need to show that

$$\sup\{\xi(r) : r \in [a,1]\} \leq h(a) \qquad \forall\, a \in \left(\max\left\{0, 2 - \frac{1}{p}\right\}, 1\right) . \tag{5.25}$$

The function $\xi(r)$ is positive and differentiable for each $a \in (0,1)$, hence the supremum here occurs either at a boundary point or at a point with zero derivative. The derivative of $\log(\xi(r))$ equals

$$\log\left(\frac{p(1-p)(2r-a)^2}{r(r-a)}\right) .$$

Thus $\xi'(r)$ has the same sign as $\overline{\xi}(r) = p(1-p)(2r-a)^2 - r(r-a)$ in $r \in [a,1]$. Notice that $\overline{\xi}(r)$ has two roots

$$r_1 = \frac{ap}{2p-1}, \text{ and } r_2 = \frac{a(1-p)}{1-2p} .$$

We may consider several cases.

Case 0: $p = 1/2$**.** The function $\overline{\xi}$ is positive, so ξ is increasing in $[a,1]$, hence the supremum in (5.25) happens at $r = 1$ and its value is $f(a)$.

Case 1: $p > 1/2$**.** Since $a > 2 - \frac{1}{p}$, we find that $r_1 > 1$ and $r_2 < 0$. Moreover, $\overline{\xi}(a) \geq 0$. Thus $\overline{\xi}$ is non-negative in $[a,1]$, which implies ξ is increasing in $[a,1]$. Thus the supremum in (5.25) happens at $r = 1$ and its value is $f(a)$.

Case 2: $p < 1/2$ **and** $a \leq \frac{1-2p}{1-p}$**.** In this case $r_1 < 0$ and $a \leq r_2 \leq 1$. Since $\overline{\xi}(a) \geq 0$ and $\overline{\xi}(r_1) = \overline{\xi}(r_2) = 0$ and $\overline{\xi}$ is quadratic, the function $\overline{\xi}$ goes from positive to negative at

r_2. Therefore, the function ξ attains its supremum at r_2 and the supremum value in (5.25) equals

$$\xi(r_2) = \left(\frac{p}{1-p}\right)^a .$$

Case 3: $p < 1/2$ and $a > \frac{1-2p}{1-p}$. We find that $r_1 < 0$ and $r_2 > 1$, and $\overline{\xi}(a) \geq 0$, so $\overline{\xi}$ is non-negative in $[a, 1]$, hence ξ is increasing in $[a, 1]$. Thus the supremum in (5.25) happens at $r = 1$ and its value is $f(a)$. This completes the proof of (5.25) and the lemma. ■

Next we define random variables $\hat{X}_1, \hat{X}_2, \ldots$ as follows: for every $i = 1, 2, \ldots$ we flip an independent unbiased coin, if it comes up heads, then $\hat{X}_i = X_i$, otherwise $\hat{X}_i = 0$.

We define the function $g_U : [0, 1] \to \mathbb{R}$ as

$$g_U(a) = \begin{cases} 1/2 & \text{if } p \geq 1/2 \text{ and } 0 \leq a \leq 1 - 1/2p \\ (\frac{1-p}{p})^a/2 & \text{if } p < 1/2 \text{ and } 0 \leq a \leq \frac{1-2p}{2-2p} \\ \frac{1}{p} & \text{if } a = 1 \\ \frac{\phi(a)-a}{\phi(a)} \left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right)^a & \text{otherwise} , \end{cases} \tag{5.26}$$

where ϕ is defined by (5.20). Note that by Lemma 5.19(a), we have $0 < a < \phi(a) < 1$ for $a \in (0, 1)$, so g_U is well defined for all $a \in [0, 1]$. Here is another computational lemma with a somewhat lengthy proof.

Lemma 5.26. *(a) We have the following concentration inequality for every $a \in [0, 1]$ and every positive integer m, where C' is an absolute constant:*

$$\mathbb{P}\left[\hat{X}_1 + \cdots + \hat{X}_m > am\right] \leq C' m^3 (2g_U(a))^{-m} . \tag{5.27}$$

(b) The function g_U is continuously differentiable on $(0, 1)$ and

$$g_U'(a) = \begin{cases} 0 & \text{if } p \geq 1/2 \text{ and } 0 < a \leq 1 - 1/2p \\ \log(\frac{1-p}{p}) g_U(a) & \text{if } p < 1/2 \text{ and } 0 < a \leq \frac{1-2p}{2-2p} \\ \log\left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right) g_U(a) & \text{otherwise} . \end{cases} \tag{5.28}$$

(c) The function $\log g_U(a)$ is increasing and convex. It is strictly increasing when $g_U(a) > 1/2$.

Proof. (a) First, the case $a = 0$ is obvious since $g_U(0) = 1/2$, and the case $a = 1$ is easy since $\mathbb{P}\left[\hat{X}_1 + \cdots + \hat{X}_m > m\right] = 0$. So we may assume that $a \in (0, 1)$.

Letting k of the $\hat{X}_i$'s being equal to X_i and the rest equal to zero, we get

$$\begin{aligned}\mathbb{P}\left[\hat{X}_1 + \cdots + \hat{X}_m > am\right] &= \sum_{k=am}^{m} \binom{m}{k} 2^{-m} \mathbb{P}\left[X_1 + \cdots + X_k > am\right] \\ &\leq m \sup\left\{\binom{m}{rm} 2^{-m} \mathbb{P}\left[X_1 + \cdots + X_{rm} > am\right] : r \in [a, 1]\right\} .\end{aligned}$$

For a given $r \in [a, 1]$, Lemma 5.25 gives

$$\mathbb{P}\left[X_1 + \cdots + X_{rm} > am\right] \leq C(rm)^2 h(a/r)^{rm} \leq Cm^2 h(a/r)^{rm} .$$

Moreover, by Stirling's approximation (2.4), we have

$$\binom{m}{rm} = O\left(\frac{1}{r^{rm}(1-r)^{(1-r)m}}\right) .$$

So, we find that

$$\mathbb{P}\left[\hat{X}_1 + \cdots + \hat{X}_m > am\right] \leq C' m^3 \left(\sup\left\{\frac{h(a/r)^r}{2r^r(1-r)^{1-r}} : r \in [a, 1]\right\}\right)^m .$$

Thus to complete the proof of part (a) we just need to show

$$g_U(a) = \inf\left\{\frac{\zeta - a}{\zeta}\left(\frac{a}{(\zeta - a)h(\zeta)}\right)^{a/\zeta} : \zeta \in [a, 1]\right\} , \tag{5.29}$$

where we have used the change of variable $\zeta = a/r$. For analyzing this infimum we define the two variable function

$$\psi(a, \zeta) = \frac{\zeta - a}{\zeta}\left(\frac{a}{(\zeta - a)h(\zeta)}\right)^{a/\zeta}$$

with domain $\{(a, \zeta) : 0 < a < 1, a \leq \zeta \leq 1\}$, and consider two cases depending on the value of p.

Case 1: $p \geq 1/2$**.** By the definition of h in (5.24) we have

$$\psi(a, \zeta) = \begin{cases} \frac{\zeta - a}{\zeta}\left(\frac{a}{\zeta - a}\right)^{a/\zeta} & \text{if } a \leq \zeta \leq 2 - p^{-1} \\ \frac{\zeta - a}{\zeta}\left(\frac{a}{(\zeta - a)f(\zeta)}\right)^{a/\zeta} & \text{otherwise} , \end{cases}$$

where f is defined in (5.4). Since $f(2-p^{-1})=1$, ψ is continuous here. Let us define $\psi_1(\zeta)=\frac{\zeta-a}{\zeta}\left(\frac{a}{\zeta-a}\right)^{a/\zeta}$ and $\psi_2(\zeta)=\frac{\zeta-a}{\zeta}\left(\frac{a}{(\zeta-a)f(\zeta)}\right)^{a/\zeta}$.

The derivative of $\log\psi_1(\zeta)$ is

$$a\log\left(\frac{\zeta-a}{a}\right)/\zeta^2\,,$$

which is negative for $\zeta<2a$ and positive for $\zeta>2a$. This implies $\psi_1(\zeta)$ is decreasing when $\zeta\le 2a$ and increasing when $\zeta\ge 2a$. So ψ_1 achieves its minimum at $\zeta=2a$, and its minimum value is $1/2$.

The derivative of $\log\psi_2(\zeta)$ is

$$\frac{a}{\zeta^2}\left[\log\left(p(1-p)(2-\zeta)^2(\zeta-a)\right)-\log\left(a(1-\zeta)\right)\right]\,.$$

Comparing with (5.19) we find that this derivative has the same sign as $\Phi(a,\zeta)$. So by Lemma 5.19(a) it vanishes at a unique point $\zeta=\phi(a)$. Also at $\zeta=\phi(a)$ we have $\partial\Phi/\partial\zeta>0$ (see (5.21)), which implies $\Phi(a,\zeta)$ is non-positive when $\zeta\le\phi(a)$ and non-negative when $\zeta\ge\phi(a)$. Thus ψ_2 achieves its minimum at $\phi(a)$, and its minimum value is

$$\psi_2(\phi(a))=\frac{\phi(a)-a}{\phi(a)}\left(\frac{a}{(\phi(a)-a)f(\phi(a))}\right)^{a/\phi(a)}=\frac{\phi(a)-a}{\phi(a)}\left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right)^{a}$$

by Lemma 5.19(b).

We conclude that:

(i) If $2a\le 2-1/p$, then the infimum of ψ occurs at $\zeta=2a$ and its value is $\psi(a,2a)=\psi_1(2a)=1/2$. The reason is that on $[a,2-1/p]$, $\psi=\psi_1$ achieves its minimum at $2a$, and on $[2-1/p,1]$, $\psi=\psi_2$ is increasing since $\Phi(a,2-1/p)\ge 0$.

(ii) If $a\le 2-1/p$ and $2a>2-1/p$, then the infimum occurs at $\zeta=\phi(a)$ and its value is $\frac{\phi(a)-a}{\phi(a)}\left((1-p)(2-\phi(a))/(1-\phi(a))\right)^a$. The reason is that on $[a,2-1/p]$, $\psi=\psi_1$ is decreasing, and on $[2-1/p,1]$, $\psi=\psi_2$ achieves its minimum at $\phi(a)$ since $\Phi(a,2-1/p)\le 0$ and $\Phi(a,1)\ge 0$.

(iii) If $a>2-1/p$, then the infimum occurs at $\zeta=\phi(a)$ and its value is equal to $\frac{\phi(a)-a}{\phi(a)}\left((1-p)(2-\phi(a))/(1-\phi(a))\right)^a$. The reason is that on $[a,1]$, $\psi=\psi_2$ achieves its minimum at $\phi(a)$ since $\Phi(a,a)\le 0$ and $\Phi(a,1)\ge 0$.

Case 2: $p < 1/2$**.** By the definition of h in (5.24) we have

$$\psi(a,\zeta) = \begin{cases} \left(\frac{1-p}{p}\right)^a \frac{\zeta-a}{\zeta} \left(\frac{a}{\zeta-a}\right)^{a/\zeta} & \text{if } a \le \zeta \le \frac{1-2p}{1-p} \\ \frac{\zeta-a}{\zeta} \left(\frac{a}{(\zeta-a)f(\zeta)}\right)^{a/\zeta} & \text{otherwise.} \end{cases}$$

The function ψ is continuous here since

$$f\left(\frac{1-2p}{1-p}\right) = \left(\frac{p}{1-p}\right)^{\frac{1-2p}{1-p}} .$$

Let us define $\psi_3(\zeta) = \left(\frac{1-p}{p}\right)^a \frac{\zeta-a}{\zeta} \left(\frac{a}{\zeta-a}\right)^{a/\zeta}$. Since $\psi_3(\zeta) = \left(\frac{1-p}{p}\right)^a \psi_1(\zeta)$, the function $\psi_3(\zeta)$ is decreasing when $\zeta \le 2a$ and increasing when $\zeta \ge 2a$. So ψ_3 achieves its minimum at $\zeta = 2a$ and its minimum value is $\left(\frac{1-p}{p}\right)^a /2$. We conclude that

(iv) If $a \le 1 - p/(1-p)$ and $2a \le 1 - p/(1-p)$, then the infimum in (5.29) occurs at $\zeta = 2a$ and at this point we have $\psi(a,\zeta) = \left(\frac{1-p}{p}\right)^a /2$. The reason is that on $[a, 1-p/(1-p)]$, $\psi = \psi_3$ achieves its minimum at $2a$, and on $[1-p/(1-p), 1]$, $\psi = \psi_2$ is increasing since $\Phi(a, 1-p/(1-p)) \ge 0$.

(v) If $a \le 1 - p/(1-p)$ and $2a > 1 - p/(1-p)$, then the infimum in (5.29) occurs at $\zeta = \phi(a)$ and its value is equal to $\frac{\phi(a)-a}{\phi(a)} \left((1-p)(2-\phi(a))/(1-\phi(a))\right)^a$. The reason is that on $[a, 1-p/(1-p)]$, $\psi = \psi_3$ is decreasing, and on $[1-p/(1-p), 1]$, $\psi = \psi_2$ achieves its minimum at $\phi(a)$ since $\Phi(a, 1-p/(1-p)) \le 0$ and $\Phi(a,1) \ge 0$.

(vi) If $a > 1 - p/(1-p)$, then the infimum in (5.29) occurs at $\zeta = \phi(a)$ and its value is equal to $\frac{\phi(a)-a}{\phi(a)} \left((1-p)(2-\phi(a))/(1-\phi(a))\right)^a$. The reason is that on $[a,1]$, $\psi = \psi_2$ achieves its minimum at $\phi(a)$ since $\Phi(a,a) \le 0$ and $\Phi(a,1) \ge 0$.

In all cases we proved that $g_U(a)$ actually gives the value of the infimum in (5.29), and this concludes the proof of (5.29) and of part (a).

(b) Consider the definition of g_U in (5.26). The formulae in (5.28) for the cases '$p \ge 1/2$ and $0 < a \le 1 - 1/2p$' and '$p < 1/2$ and $0 < a \le \frac{1-2p}{2-2p}$' follow from definition, so we assume that a is in the 'otherwise' case. We use the equality (5.29). Note that as proved in part (a), the infimum in (5.29) occurs at the point $\zeta = \phi(a)$ that has $\frac{\partial \psi}{\partial \zeta}\Big|_{(a,\phi(a))} = 0$. This

implies for every a_0,

$$\begin{aligned}\frac{\mathrm{d}g_U}{\mathrm{d}a}(a_0) &= \frac{\partial\psi}{\partial a}(a_0,\phi(a_0)) + \frac{\partial\psi}{\partial\zeta}(a_0,\phi(a_0)) \times \frac{\mathrm{d}\phi}{\mathrm{d}a}(a_0) \\ &= \frac{\partial\psi}{\partial a}(a_0,\phi(a_0)) = \frac{\partial}{\partial a}\left[\frac{\zeta-a}{\zeta}\left(\frac{a}{(\zeta-a)f(\zeta)}\right)^{a/\zeta}\right]\Bigg|_{(a_0,\phi(a_0))},\end{aligned}$$

and (5.28) follows from computing this partial derivative and putting $\zeta = \phi(a_0)$.

We next prove the continuity of g_U and its derivative. Note that by Lemma 5.19(a), if $a \in (0,1)$ then $\phi(a) \in (0,1)$. First, g_U is continuous at $a = 1$ since

$$\begin{aligned}\lim_{a\to 1}\frac{\phi(a)-a}{\phi(a)}\left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right)^a &= \lim_{a\to 1}\frac{\phi(a)-a}{\phi(a)}\left(\frac{a}{p(2-\phi(a))(\phi(a)-a)}\right)^a \\ &= \lim_{a\to 1}\frac{(\phi(a)-a)^{1-a}}{\phi(a)}\left(\frac{a}{p(2-\phi(a))}\right)^a = \frac{1}{p}\,.\end{aligned}$$

For $p \geq 1/2$, the only discontinuity for g_U can possibly occur at $b = 1-1/2p$. However at this point we have $\phi(b) = 2b = 2-p^{-1}$ so that $(1-p)(2-\phi(b)) = 1-\phi(b)$. Hence the left and right limits of g_U equal $1/2$, and the left and right limits of g_U' equal 0. Therefore, both g_U and g_U' are continuous at b.

For $p < 1/2$, the only discontinuity for g_U can possibly occur at $c = (1-2p)/(2-2p)$. However at this point $\phi(c) = 2c = (1-2p)/(1-p)$ so that $\frac{(1-p)(2-\phi(c))}{1-\phi(c)} = \frac{1-p}{p}$. Hence the left and right limits of g_U equal $\left(p^{-1}-1\right)^c/2$, and the left and right limits of g_U' equal $\log\left(p^{-1}-1\right)\left(p^{-1}-1\right)^c/2$. Therefore, both g_U and g_U' are continuous at c.

(c) Note that g_U is positive everywhere, so $\log(g_U)$ is (strictly) increasing if and only if g_U is (strictly) increasing. By the formulae for g_U' in part (b), it is easy to see that g_U' is always non-negative, and is positive when $g_U(a) > 1/2$. To show $\log(g_U)$ is convex, we need to show its derivative, i.e. g_U'/g_U is increasing. This also follows from part (b), noting that ϕ is increasing by Lemma 5.19(c). ■

Lemma 5.27. *Let $\omega > 0$ and let $\tau(x) : [0,\omega] \to \mathbb{R}$ be a positive function that is differentiable on $(0,\omega)$ and satisfies*

$$\alpha(x) + \chi(\tau(x)) = 0 \qquad \forall\, x \in [0,\omega] \tag{5.30}$$

for convex functions α, χ, with α increasing and χ decreasing. Assume there exists $x^* \in (0,\omega)$ such that $\tau'(x^*) = \tau(x^*)/x^*$. Then we have

$$\frac{x^*}{\tau(x^*)} \geq \frac{y}{\tau(y)} \tag{5.31}$$

for all $y \in [0, \omega]$.

Proof. We first prove that τ is convex and increasing. Pick $x_1, x_2 \in [0, \omega]$ and $\lambda_1, \lambda_2 \in [0, 1]$ with $\lambda_1 + \lambda_2 = 1$. We need to show that

$$\tau(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 \tau(x_1) + \lambda_2 \tau(x_2)\,. \tag{5.32}$$

We have

$$\begin{aligned}\chi(\lambda_1 \tau(x_1) + \lambda_2 \tau(x_2)) &\leq \lambda_1 \chi(\tau(x_1)) + \lambda_2 \chi(\tau(x_2)) \\ &= -\lambda_1 \alpha(x_1) - \lambda_2 \alpha(x_2) \\ &\leq -\alpha(\lambda_1 x_1 + \lambda_2 x_2) \\ &= \chi(\tau(\lambda_1 x_1 + \lambda_2 x_2))\end{aligned}$$

by convexity of χ, then (5.30), then convexity of α, and then (5.30) again. The equation (5.32) follows since χ is decreasing. Hence τ is convex. Also, τ is increasing since α is increasing and χ is decreasing.

Now, let $y \in [0, \omega]$. We prove (5.31) for $y < x^*$. The proof for $y > x^*$ is similar. By the mean value theorem, there exists $z \in (y, x^*)$ with

$$\tau'(z) = \frac{\tau(x^*) - \tau(y)}{x^* - y}\,.$$

On the other hand, since $z < x^*$ and τ is convex, we have

$$\tau'(z) \leq \tau'(x^*) = \frac{\tau(x^*)}{x^*}\,.$$

The inequality (5.31) follows from these two results. ∎

We are ready to prove the upper bound in Theorem 5.3. The upper bound in Theorem 5.4 follows immediately as in every tree the diameter is at most twice the height.

Proof of the upper bound in Theorem 5.3. Let $c_U = c_U(p)$. By Lemma 5.14 we just need to show that given $\varepsilon > 0$, a.a.s as $t \to \infty$ the weighted height of T_t is at most $(1+\varepsilon)c_U t$. For proving this we use Lemma 5.23. Lemma 5.26 implies that condition (5.22) of Lemma 5.23 holds with $\gamma_U(a) = \log(2g_U(a))$, so we need only show that

$$c_U = \sup\left\{\frac{a}{\rho} : \log(g_U(a)) + \rho - 1 - \log(\rho) = 0 : a \in [0,1], \rho \in (0,\infty)\right\}. \tag{5.33}$$

The function $\rho - 1 - \log(\rho)$ attains all values in $[0, \infty)$ for $\rho \in (0, 1]$. Moreover, it is strictly decreasing for $\rho \in (0, 1]$ and equals 0 for $\rho \in [1, \infty)$. So $\log(g_U(a)) + \rho - 1 - \log(\rho) = 0$ has a unique solution (for ρ) if $0 < g_U(a) < 1$, and no solution if $g_U(a) > 1$. Since $g_U(0) = 1/2$ and $g_U(1) = 1/p$, and the function $g_U(x)$ is continuous and strictly increasing when $g_U(x) > 1/2$, there is a unique x with $g_U(x) = 1$. Denote this point by $a_{\max}$. Define the function $\tau : [0, a_{\max}] \to (0, 1]$ as follows. Let $\tau(a_{\max}) = 1$ and for $x < a_{\max}$ let $\tau(x)$ be the unique number satisfying

$$\log(g_U(x)) + \tau(x) - 1 - \log \tau(x) = 0 \,. \tag{5.34}$$

Hence to prove (5.33) it is enough to show that

$$c_U = \sup \left\{ \frac{x}{\tau(x)} : x \in [0, a_{\max}] \right\} . \tag{5.35}$$

We prove (5.35) using Lemma 5.27. The function $\log(g_U(a))$ is increasing and convex by Lemma 5.26(c), and it is easy to check that the function $\rho - 1 - \log(\rho)$ is decreasing and convex. Moreover, differentiating (5.34) gives

$$\frac{g_U'(x)}{g_U(x)} + \tau'(x) - \frac{\tau'(x)}{\tau(x)} = 0 \,.$$

So by the implicit function theorem τ is differentiable in $x \in (0, a_{\max})$ and

$$\tau'(x) = \frac{\tau(x)}{1 - \tau(x)} \frac{g_U'(x)}{g_U(x)} \,.$$

By Lemma 5.27, we just need to show the existence of $x^* \in (0, a_{\max})$ with

$$c_U = \frac{x^*}{\tau(x^*)} = \frac{1 - \tau(x^*)}{\tau(x^*)} \frac{g_U(x^*)}{g_U'(x^*)} \,. \tag{5.36}$$

We consider two cases. Recall that $p_0 \approx 0.206$ is the solution to

$$\log\left(\frac{1-p}{p}\right) = \frac{1-p}{1-2p} \,,$$

which has a unique solution by Lemma 5.21(a).

Case 1: $0 < p \le p_0$**.** In this case we have

$$c_U = \left(\log\left(\frac{1-p}{p}\right) \right)^{-1} .$$

Let

$$a^* = \left[2 \log\left(\frac{1-p}{p}\right)\right]^{-1} .$$

By Lemma 5.21(a)

$$\log\left(\frac{1-p}{p}\right) \geq \frac{1-p}{1-2p} ,$$

which gives $a^* \leq \frac{1-2p}{2-2p}$, thus

$$g_U(a^*) = \left(\frac{1-p}{p}\right)^{a^*} \Big/ 2 = \exp\left(\frac{1}{2} - \log 2\right) < 1$$

by the definition of g_U in (5.26), and

$$g'_U(a^*) = \log\left(\frac{1-p}{p}\right) g_U(a^*)$$

by Lemma 5.26(b). The definition of τ in (5.34) implies $\tau(a^*) = 1/2$. Moreover,

$$\frac{a^*}{\tau(a^*)} = \left(\log\left(\frac{1-p}{p}\right)\right)^{-1} = \frac{1-\tau(a^*)}{\tau(a^*)} \frac{g_U(a^*)}{g'_U(a^*)} ,$$

which gives (5.36). Finally, since $g_U(a^*) < 1$, we have $a^* \in (0, a_{\max})$, and the proof is complete.

Case 2: $p_0 < p < 1$**.** In this case we have

$$c_U = ps^*(2 - s^*) \exp(1/s^*) ,$$

where $s^* \in (0, 1)$ is the unique solution for

$$s^* \log\left(\frac{(1-p)(2-s^*)}{1-s^*}\right) = 1 . \tag{5.37}$$

Lemma 5.21(b) implies that s^* is well defined. Let $a^* = \phi^{-1}(s^*)$.

We first show that

$$g_U(a^*) = \frac{s^* - a^*}{s^*} \left(\frac{(1-p)(2-s^*)}{1-s^*}\right)^{a^*} . \tag{5.38}$$

If $p > 1/2$, then by Lemma 5.21(b) we have $s^* > 2 - \frac{1}{p}$. It is easy to verify that $\Phi(1 - \frac{1}{2p}, 2 - \frac{1}{p}) = 0$. Since ϕ^{-1} is increasing, we have $a^* = \phi^{-1}(s^*) > 1 - \frac{1}{2p}$, so (5.38) agrees with the definition of g_U in (5.26).

If $p_0 < p \le 1/2$, then by Lemma 5.21(b) we have $s^* > \frac{1-2p}{1-p}$. It is easy to verify that $\Phi(\frac{1-2p}{2-2p}, \frac{1-2p}{1-p}) = 0$. Since ϕ^{-1} is increasing, we have $a^* = \phi^{-1}(s^*) > \frac{1-2p}{2-2p}$, so (5.38) agrees with the definition of g_U in (5.26).

Using (5.37), the equation (5.38) simplifies into

$$g_U(a^*) = \left(1 - \frac{a^*}{s^*}\right) \exp\left(a^*/s^*\right) < \exp(-\frac{a^*}{s^*}) \exp\left(a^*/s^*\right) = 1\,, \tag{5.39}$$

and by Lemma 5.26(b) we have

$$g'_U(a^*) = g_U(a^*) \log\left(\frac{(1-p)(2-s^*)}{1-s^*}\right) = g_U(a^*)/s^*\,.$$

It follows from (5.39) and the definition of τ in (5.34) that $\tau(a^*) = 1 - \frac{a^*}{s^*}$. Using (5.37) and $\Phi(a^*, s^*) = 0$, we get

$$\frac{a^*}{\tau(a^*)} = ps^*(2 - s^*) \exp(1/s^*) = \frac{1 - \tau(a^*)}{\tau(a^*)} \frac{g_U(a^*)}{g'_U(a^*)}\,,$$

which gives (5.36). Finally, since $g_U(a^*) < 1$, we have $a^* \in (0, a_{\max})$, and the proof is complete. ■

5.5 Concluding remarks

There is a common generalization of random recursive trees, preferential attachment trees, and random-surfer trees. Consider i.i.d. random variables $X_1, X_2, \ldots \in \{0, 1, 2, \ldots\}$. Start with a single vertex v_0. At each step s a new vertex v_s appears, chooses a random vertex u in the present graph, and then walks X_s steps from u towards v_0, joining to the last vertex in the walk (if it reaches v_0 before X_s steps, it joins to v_0). Random recursive trees correspond to $X_i = 0$, preferential attachment trees correspond to $X_i = \text{Bernoulli}(1/2)$ (see, e.g., [16, Theorem 3.1]), and random-surfer trees correspond to $X_i = \text{Geo}(p)$. Using the ideas of this chapter, it is possible to obtain lower and upper bounds for the height and the diameter of this general model (similar to Theorems 5.3 and 5.4), provided one

can prove large deviation inequalities (similar to Lemma 5.18) for the sum of X_i's and also large deviation inequalities (similar to Lemma 5.25) for the sum of random variables X'_i, defined as

$$X'_1 = 1, \qquad X'_{i+1} = \max\{1 - X_i, 1 - (X'_1 + \cdots + X'_i)\}\,.$$

Chapter 6

Push&pull protocols

In this chapter[1] we study two randomized rumour spreading protocols which are defined below.

Definition 6.1 (Asynchronous push&pull protocol)**.** Let G be a simple and connected graph, and suppose that an independent Poisson clock of rate 1 is associated with each vertex of G. Suppose that initially, some vertex v of G knows a piece of information, the so-called *rumour*. The rumour spreads in G as follows. Whenever the clock of a vertex x rings, this vertex performs an 'action': it calls a random neighbour y; if x knows the rumour and y does not, then x tells y the rumour (a *push* operation), and if x does not know the rumour and y knows it, y tells x the rumour (a *pull* operation). Note that if both x and y know the rumour or neither of them knows it, then this action is useless. Also, vertices have no memory, hence x may call the same neighbour several consecutive times. The *spread time* of G starting from v, written $\mathrm{ST}_{\mathsf{a}}(G, v)$, is the first time that all vertices of G know the rumour. Note that this is a continuous random variable, with two sources of randomness: the Poisson processes associated with the vertices, and random neighbour-selection of the vertices. The *guaranteed spread time* of G, written $\mathrm{gst}_{\mathsf{a}}(G)$, is the smallest deterministic number t such that for every $v \in V(G)$ we have $\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G, v) > t\right] \leq 1/|V(G)|$. The *worst average spread time* of G, written $\mathrm{wast}_{\mathsf{a}}(G)$, is the smallest deterministic number t such that for every $v \in V(G)$ we have $\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G, v)\right] \leq t$.

Definition 6.2 (Synchronous push&pull protocol)**.** Let G denote a simple and connected graph. Initially some vertex v of G knows the rumour, which spreads in G in a round-robin

[1] This chapter is based on joint work with Acan, Collevecchio, and Wormald. The results therein appear in the submitted preprint [1].

manner: in each round $1, 2, \ldots$, all vertices perform actions simultaneously. That is, each vertex x calls a random neighbour y; if x knows the rumour and y does not, then x tells y the rumour (a *push* operation), and if x does not know the rumour and y knows it, y tells x the rumour (a *pull* operation). Note that this is a synchronous protocol, e.g. a vertex that receives a rumour in a certain round cannot send it on in the same round. The *spread time* of G starting from v, written $\mathrm{ST}_{\mathsf{s}}(G, v)$, is the first time that all vertices of G know the rumour. Note that this is a discrete random variable, with one source of randomness: the random neighbour-selection of the vertices. The *guaranteed spread time* of G, written $\mathrm{gst}_{\mathsf{s}}(G)$, and the *worst average spread time* of G, written $\mathrm{wast}_{\mathsf{s}}(G)$, are defined in an analogous way to the asynchronous case.

In Proposition 6.16, we show that the choice of starting vertex does not affect the spread time as far as we are not concerned with constant factors, and that $\mathrm{wast}_{\mathsf{a}}(G) = \Theta(\mathbb{E}[\mathrm{ST}_{\mathsf{a}}(G, v)])$ and $\mathrm{wast}_{\mathsf{s}}(G) = \Theta(\mathbb{E}[\mathrm{ST}_{\mathsf{s}}(G, v)])$ for any vertex $v \in V(G)$.

In this chapter G always denotes the ground graph which is simple and connected, and n always denotes the number of its vertices, and is assumed to be sufficiently large.

The first main result of this chapter is the following theorem.

Theorem 6.3. *The following hold for any n-vertex graph G.*

$$(1 - 1/n)\,\mathrm{wast}_{\mathsf{a}}(G) \le \mathrm{gst}_{\mathsf{a}}(G) \le e\,\mathrm{wast}_{\mathsf{a}}(G)\log n\,, \tag{6.1}$$

$$\mathrm{wast}_{\mathsf{a}}(G) = \Omega(\log n) \quad \text{and} \quad \mathrm{wast}_{\mathsf{a}}(G) = O(n)\,, \tag{6.2}$$

$$\mathrm{gst}_{\mathsf{a}}(G) = \Omega(\log n) \quad \text{and} \quad \mathrm{gst}_{\mathsf{a}}(G) = O(n \log n)\,. \tag{6.3}$$

Moreover, these bounds are asymptotically best possible, up to the constant factors.

Our proof of the right-hand bound in (6.2) is based on the pull operation only, so this bound applies equally well to the pull protocol.

The arguments for (6.1) and the right-hand bounds in (6.2) and (6.3) can easily be extended to the synchronous variant, giving the following theorem. The bound (6.6) below also follows from [62, Theorem 2.1], but here we also show its tightness.

Theorem 6.4. *The following hold for any n-vertex graph G.*

$$(1 - 1/n)\,\mathrm{wast}_{\mathsf{s}}(G) \le \mathrm{gst}_{\mathsf{s}}(G) \le e\,\mathrm{wast}_{\mathsf{s}}(G)\log n\,, \tag{6.4}$$

$$\mathrm{wast}_{\mathsf{s}}(G) = O(n)\,, \tag{6.5}$$

$$\mathrm{gst}_{\mathsf{s}}(G) = O(n \log n)\,. \tag{6.6}$$

Moreover, these bounds are asymptotically best possible, up to the constant factors.

Open problem 6.5. Find the best possible constants factors in Theorems 6.3 and 6.4.

We next turn to studying the relationship between the asynchronous and synchronous variants on the same graph.

Theorem 6.6. *For any G we have $\mathrm{gst}_{\mathsf{a}}(G) = O\left(\mathrm{gst}_{\mathsf{s}}(G)\log n\right)$, and this bound is best possible, up to the constant factor.*

For all graphs we examined a stronger result holds, which suggests the following conjecture.

Conjecture 6.7. For any n-vertex graph G we have $\mathrm{gst}_{\mathsf{a}}(G) \leq \mathrm{gst}_{\mathsf{s}}(G) + O(\log n)$.

Our last main result is the following theorem, whose proof is somewhat technical, and uses couplings with the sequential rumour spreading protocol (defined on Page 146).

Theorem 6.8. *For any $\alpha \in [0,1)$ we have*

$$\mathrm{gst}_{\mathsf{s}}(G) \leq n^{1-\alpha} + O(\mathrm{gst}_{\mathsf{a}}(G) n^{(1+\alpha)/2}) \,. \tag{6.7}$$

Corollary 6.9. *We have*

$$\frac{\mathrm{gst}_{\mathsf{s}}(G)}{\mathrm{gst}_{\mathsf{a}}(G)} = \Omega(1/\log n) \quad \text{and} \quad \frac{\mathrm{gst}_{\mathsf{s}}(G)}{\mathrm{gst}_{\mathsf{a}}(G)} = O\big(n^{2/3}\big) \,,$$

and the left-hand bound is asymptotically best possible, up to the constant factor. Moreover, there exist infinitely many graphs for which this ratio is $\Omega\left(n^{1/3}(\log n)^{-4/3}\right)$.

Open problem 6.10. What is the maximum possible value of the ratio $\mathrm{gst}_{\mathsf{s}}(G)/\mathrm{gst}_{\mathsf{a}}(G)$ for an n-vertex graph G?

The parameters $\mathrm{wast}_{\mathsf{s}}(G)$ and $\mathrm{wast}_{\mathsf{a}}(G)$ can be approximated easily using the Monte Carlo method: simulate the protocols several times, measuring the spread time of each simulation, and output the average. Another open problem is to design a *deterministic* approximation algorithm for any one of $\mathrm{wast}_{\mathsf{a}}(G)$, $\mathrm{gst}_{\mathsf{a}}(G)$, $\mathrm{wast}_{\mathsf{s}}(G)$ or $\mathrm{gst}_{\mathsf{s}}(G)$.

In this chapter we use standard graph theoretic arguments and well known properties of the exponential distribution and Poisson processes (see Section 2.4 for a review), in particular the memorylessness, and the fact that the union of two Poisson processes is another Poisson process. For proving Theorem 6.8 we define a careful coupling between the synchronous and asynchronous protocols.

We next review some related work. In Section 6.1 we prove some preliminary results. In Section 6.2 we study some examples, which demonstrate tightness of some of the above bounds. Theorems 6.3 and 6.4 are proved in Sections 6.3 and 6.4, respectively. Theorems 6.6 and 6.8 and Corollary 6.9 are proved in Section 6.5.

Related work

A protocol that was studied prior to push&pull is the (synchronous) push protocol, in which the informed nodes push the rumour, but the uninformed ones do nothing. Feige, Peleg, Raghavan and Upfal [62] showed that for any G, a.a.s. the spread time of the push protocol is $\Omega(\log n)$ and $O(\Delta(G)\cdot(\mathrm{diam}(G)+\log n))$. This protocol has been studied on many graph classes such as complete graphs [62, 110], Erdős-Rényi random graphs [62, 67, 106], random regular graphs [10, 68], and hypercube graphs [62]. For most of these classes it turns out that a.a.s. the spread time is $\Theta(\mathrm{diam}(G)+\log n)$, which does not depend on the maximum degree. Interesting connections between the spread time and the cover time/mixing time of the simple random walk on the graph have been proved in [59, 115].

Fountoulakis et al. [69] studied the asynchronous push&pull protocol on Chung-Lu random graphs with exponent between 2 and 3. For these graphs, they showed that a.a.s. after some constant time, $n - o(n)$ nodes are informed. Doerr, Fouz, and Friedrich [49] showed that for the preferential attachment graph (the non-tree case), a.a.s. all but $o(n)$ vertices receive the rumour in time $O\left(\sqrt{\log n}\right)$, but to inform all vertices a.a.s., $\Theta(\log n)$ time is necessary and sufficient. Panagiotou and Speidel [107] studied this protocol on Erdős-Renyi random graphs and proved that if the average degree is $(1+\Omega(1))\log n$, a.a.s. the spread time is $(1+o(1))\log n$.

6.1 Preliminaries

We start by making a few observations.

Observation 6.11. Consider the asynchronous variant. Let uv be an edge. Whenever v's clock rings, it calls u with probability $1/\deg(v)$. Hence, for each vertex v, we can replace v's clock by one Poisson clock for each incident edge, these clocks being independent of all other clocks and having rate $1/\deg(v)$ (see the discussion after Proposition 2.16).

Observation 6.12. Whenever a new vertex is informed, by memorylessness of the Poisson process, we may imagine that all Poisson clocks are restarted (see Proposition 2.16).

The following definition will be used throughout the chapter.

Definition 6.13 (Communication time)**.** For an edge $e = uv$, the *communication time via edge* e, written $T(e)$, is defined as follows. Suppose τ is the first time that one of u and v learns the rumour, and ρ is the first time after τ that one of u and v calls the other one.

Then $T(e) = \rho - \tau$, which is nonnegative. Note that after time ρ, both u and v know the rumour.

Observation 6.14. Let $uv \in E(G)$. In the synchronous version,

$$T(uv) \stackrel{d}{=} 1 + \min\{\mathrm{Geo}(1/\deg(u)), \mathrm{Geo}(1/\deg(v))\} .$$

We get a nicer formula in the asynchronous version.

Proposition 6.15. *Let $uv \in E(G)$. In the asynchronous version,*

$$T(uv) \stackrel{d}{=} \mathrm{Exp}(1/\deg(u) + 1/\deg(v)) . \tag{6.8}$$

Moreover, the random variables $\{T_e\}_{e \in E(G)}$ are mutually independent.

Proof. By Observations 6.11 and 6.12, the $T(e)$'s are mutually independent, and moreover, $T(uv)$ is the minimum of two independent exponential random variables with rates $1/\deg(v)$ and $1/\deg(u)$, and (6.8) follows from Proposition 2.11. ■

We now show that changing the starting vertex affects the spread time by at most a multiplicative factor of 2, and in particular, $\mathrm{wast}_{\mathsf{s}}(G) \le 2\mathbb{E}[\mathrm{ST}_{\mathsf{s}}(G, v)]$ and $\mathrm{wast}_{\mathsf{a}}(G) \le 2\mathbb{E}[\mathrm{ST}_{\mathsf{a}}(G, v)]$ for any vertex v.

Proposition 6.16. *For any two vertices u and v of G we have $\mathrm{ST}_{\mathsf{s}}(G, u) \stackrel{s}{\le} 2\,\mathrm{ST}_{\mathsf{s}}(G, v)$ and also $\mathrm{ST}_{\mathsf{a}}(G, u) \stackrel{s}{\le} 2\,\mathrm{ST}_{\mathsf{a}}(G, v)$.*

Proof. We first consider the synchronous protocol. Let $C(u, v)$ denote the first time that v learns the rumour, assuming initially only u knows it. We claim that

$$C(u, v) \stackrel{d}{=} C(v, u) , \tag{6.9}$$

which would imply

$$\mathrm{ST}_{\mathsf{s}}(G, u) \stackrel{s}{\le} C(u, v) + \mathrm{ST}_{\mathsf{s}}(G, v) \stackrel{d}{=} C(v, u) + \mathrm{ST}_{\mathsf{s}}(G, v) \stackrel{s}{\le} 2\,\mathrm{ST}_{\mathsf{s}}(G, v) .$$

In every round of an execution of the protocol, each vertex contacts a neighbour. We call this an *action*, and the *signature* of this action, is a function $a : V \to V$ mapping each vertex to a neighbour. Hence, m rounds of the protocol can be encoded as $(u, a_1 a_2 \cdots a_m)$, where u is the vertex knowing the rumour initially, and $a_1 a_2 \cdots a_m$ is a sequence of signatures.

Let $I(u, a_1 a_2 \cdots a_m)$ denote the set of informed vertices after m rounds. Note that in each round, the signature of the action taken is a uniformly random one. Hence $\mathbb{P}[C(u,v) \le k]$ equals the proportion of the signature-sequences $a_1 a_2 \cdots a_k$ of length k that satisfy $v \in I(u, a_1 \cdots a_k)$. If $v \in I(u, a_1 \cdots a_k)$, then looking at the (u,v)-path through which v was informed, we see that $u \in I(v, a_k a_{k-1} \cdots a_2 a_1)$. Therefore, $\mathbb{P}[C(u,v) \le k] = \mathbb{P}[C(v,u) \le k]$ for any k, and this proves (6.9).

We now consider the asynchronous protocol. Let $D(u,v)$ denote the first time that v learns the rumour, assuming initially only u knows it. Again, it suffices to prove

$$D(u,v) \stackrel{d}{=} D(v,u)\,. \tag{6.10}$$

By Proposition 6.15, for any edge uv we have $T(uv) \stackrel{d}{=} \mathrm{Exp}(1/\deg(u) + 1/\deg(v))$. Moreover, the variables $\{T(e)\}_{e\in E}$ are mutually independent. We define a collection of mutually independent random variables $\{R(e)\}_{e\in E}$, such that for any edge uv,

$$R(uv) \stackrel{d}{=} \mathrm{Exp}(1/\deg(u) + 1/\deg(v))\,.$$

Let $\mathcal{P}$ denote the set of all (u,v)-paths. Then we have

$$D(u,v) = \min\left\{\sum_{e\in P} T(e) : P \in \mathcal{P}\right\} \stackrel{d}{=} \min\left\{\sum_{e\in P} R(e) : P \in \mathcal{P}\right\}\,.$$

By symmetry, $D(v,u)$ has exactly the same distribution, and (6.10) follows. ■

6.2 Examples

In this section we study some important graphs and bound their spread times, partly for showing tightness of some of the bounds obtained, and partly to serve as an introduction to the behaviour of the protocols.

6.2.1 The complete graph

For the complete graph, K_n, by symmetry what matters at any time is not the actual set of informed vertices, but only the number of vertices that have the rumour. In the asynchronous case, by Proposition 6.15 and Observation 6.11, we can imagine a Poisson

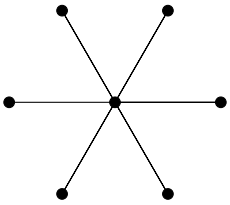


Figure 6.1: the star graph G_7^*

clock for each edge, having rate $2/(n-1)$ and independent of all other clocks. Let $T_1 = 0$ and denote by T_k the first time that there are k informed vertices. We can at this time simply restart all $k(n-k)$ clocks at edges joining informed to uninformed vertices (see Observation 6.12). When the next alarm rings, a new vertex receives the rumour. Thus, $T_{k+1} - T_k$ is distributed as the mimimum of $k(n-k)$ independent exponential random variables each with rate $2/(n-1)$, i.e. as $\mathrm{Exp}(2k(n-k)/(n-1))$. Hence by linearity of expectation,

$$\mathrm{wast}_{\mathsf{a}}(K_n) = \mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(K_n, v)\right] = \mathbb{E}\left[T_n\right] = \mathbb{E}\left[T_1\right] + \sum_{k=1}^{n-1} \mathbb{E}\left[T_{k+1} - T_k\right] = \sum_{k=1}^{n-1} \frac{n-1}{2k(n-k)} \,.$$

We have

$$\sum_{k=1}^{n-1} \frac{n-1}{2k(n-k)} = \left(\frac{n-1}{2n}\right) \sum_{k=1}^{n-1} \left\{\frac{1}{k} + \frac{1}{n-k}\right\} \sim \left(\frac{n-1}{2n}\right) (2\log n) \sim \log n \,,$$

so $\mathrm{wast}_{\mathsf{a}}(K_n) \sim \log n$. In fact, Janson [78, Theorem 1.1(ii)] showed that a.a.s. $\mathrm{ST}_{\mathsf{a}}(K_n, v) \sim \log n$. Moreover, it is implicit in his proof that $\mathrm{gst}_{\mathsf{a}}(K_n) = O(\log n)$, see [78, Theorem 3.2].

For the synchronous version, Karp et al. [82, Theorem 2.1] showed that a.a.s. we have $\mathrm{ST}_{\mathsf{s}}(K_n, v) \sim \log_3 n$. It follows that $\mathrm{wast}_{\mathsf{s}}(K_n) \sim \log_3 n$. It is implicit in their proof that $\mathrm{gst}_{\mathsf{s}}(K_n) = O(\log n)$.

6.2.2 The star

The star G_n^* with n vertices has $n-1$ leaves and a central vertex that is adjacent to every other vertex, see Figure 6.1. Each leaf must communicate with the central vertex through the unique edge connecting them. It is clear that $\mathrm{ST}_{\mathsf{s}}(G_n^*, v) = 1$ if v is the central vertex and $\mathrm{ST}_{\mathsf{s}}(G_n^*, v) = 2$ otherwise.

In the asynchronous case, the spread time is close to the time the last vertex makes its first call. By Proposition 6.15, all communication times are independent and distributed as $\mathrm{Exp}(n/(n-1))$. Let $X_1, \dots, X_{n-1}$ be independent $\mathrm{Exp}(n/(n-1))$ random variables. Then

$$\mathrm{ST}_{\mathsf{a}}(G_n^*, v) \stackrel{d}{=} \begin{cases} \max\{X_1, \dots, X_{n-1}\} & \text{if } v \text{ is the central vertex} \\ X_1 + \max\{X_2, \dots, X_{n-1}\} & \text{if } v \text{ is a leaf.} \end{cases}$$

It follows from Proposition 2.17 that that $\mathrm{wast}_{\mathsf{a}}(G_n^*) \sim \log n$.

Proposition 6.17. *We have* $\mathrm{gst}_{\mathsf{a}}(G_n^*) = \Theta(\log n)$.

Proof. Let v be a leaf. Then

$$\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G_n^*, v) > \log n/2\right] \geq \mathbb{P}\left[X_1 \geq (\log n)/2\right] = \exp\left(-\frac{n \log n}{2(n-1)}\right) > 1/n\,,$$

so $\mathrm{gst}_{\mathsf{a}}(G_n^*) > (\log n)/2$.

For the other direction, since $\max\{X_1, \dots, X_{n-1}\} \stackrel{s}{\leq} X_1 + \max\{X_2, \dots, X_{n-1}\}$, it suffices to show

$$\mathbb{P}\left[X_1 + \max\{X_2, \dots, X_{n-1}\} > 3 \log n\right] < 1/n\,. \tag{6.11}$$

We have

$$\mathbb{P}\left[X_1 > \log n\right] = \exp(-n \log n/(n-1)) < 1/(2n)$$

and

$$\mathbb{P}\left[\max\{X_2, \dots, X_{n-1}\} > 2 \log n\right] \leq n\mathbb{P}\left[X_2 > 2 \log n\right] = \exp\left(\log n - \frac{2n \log n}{n-1}\right) < 1/(2n)\,.$$

Inequality (6.11) follows from the union bound. ∎

This graph gives that the left-hand bounds in (6.1), (6.2), (6.3), (6.4) and Corollary 6.9, and Theorem 6.6, are tight, up to constant factors.

6.2.3 The path

The graph P_n consists just of a path $(v_1, \dots, v_n)$, see Figure 6.2. In this case, the spread times in the synchronous and asynchronous variants are close to each other. We first consider the asynchronous variant. Let e be an edge. By Proposition 6.15, if e connects

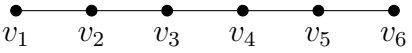


Figure 6.2: the path graph P_6

two internal vertices, then $T(e) \stackrel{d}{=} \text{Exp}(1)$, and otherwise, $T(e) \stackrel{d}{=} \text{Exp}(3/2)$. Thus if the rumour starts from one of the endpoints, say v_1, we have

$$\text{ST}_{\mathsf{a}}(P_n, v_1) \stackrel{d}{=} \sum_{i=1}^{n-1} X_i, \tag{6.12}$$

where X_i's are independent exponential random variables, X_1 and X_{n-1} with rates $3/2$ and the rest with rates 1. It follows that $\mathbb{E}\left[\text{ST}_{\mathsf{a}}(P_n, v_1)\right] = (n-3) + 2(2/3) = n - 5/3$. With similar computations, it is easy to see that this is the worst case, i.e. $\text{wast}_{\mathsf{a}}(P_n) = n - 5/3$. This shows that the right-hand bound in (6.2) is tight, up to the constant factor.

Next we show $\text{gst}_{\mathsf{a}}(P_n) \sim n$. Fix $\varepsilon > 0$. Note that $\sum_{i=2}^{n-2} X_i$ is a sum of i.i.d. random variables, hence by Cramér's Theorem (Theorem 2.20), the probability that it deviates by at least εn from its expected value is $\exp(-\Omega(n))$. Moreover, (6.12) means that $\text{ST}_{\mathsf{a}}(P_n, v_1)$ is $\sum_{i=2}^{n-2} X_i$ plus two exponential random variables with constant rate, and the same statement is true for it as well, so $\text{gst}_{\mathsf{a}}(P_n) \sim n$.

Now consider the synchronous case. Let e be an edge. By Observation 6.14, if e connects two internal vertices, then $T(e) \stackrel{d}{=} 1 + \text{Geo}(3/4)$, and otherwise, $T(e) = 1$. Thus if the rumour starts from one of the endpoints, say v_1, we have

$$\text{ST}_{\mathsf{s}}(P_n, v_1) \stackrel{d}{=} n - 1 + \sum_{i=2}^{n-2} X_i, \tag{6.13}$$

where X_i's are independent $\text{Geo}(3/4)$ random variables. It follows that $\mathbb{E}\left[\text{ST}_{\mathsf{s}}(P_n, v_1)\right] = (4/3)n - 2$. With similar computations, it is easy to see that this is the worst case, i.e. $\text{wast}_{\mathsf{s}}(P_n) = (4/3)n - 2$. This shows that the right-hand bound in (6.5) is tight, up to the constant factor.

Next we show $\text{gst}_{\mathsf{s}}(P_n) \sim n$. Fix $\varepsilon > 0$. Note that $\sum_{i=2}^{n-2} X_i$ is a sum of i.i.d. random variables, hence by Cramér's Theorem (Theorem 2.20), the probability that it deviates by at least εn from its expected value is $\exp(-\Omega(n))$. Moreover, (6.13) means that $\text{ST}_{\mathsf{s}}(P_n, v_1)$ is $\sum_{i=2}^{n-2} X_i$ plus a constant, and the same statement is true for it as well, so $\text{gst}_{\mathsf{s}}(P_n) \sim (4/3)n$.

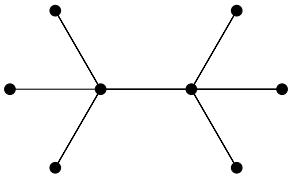


Figure 6.3: the double star graph DS_8

6.2.4 The double star

Consider the tree DS_n consisting of two adjacent vertices of degree $n/2$, referred to as the *star centres*, and $n-2$ leaves, see Figure 6.3. We will show that $\mathrm{gst}_{\mathsf{a}}(DS_n)$ and $\mathrm{gst}_{\mathsf{s}}(DS_n)$ are both $\Theta(n \log n)$, while the average times $\mathrm{wast}_{\mathsf{a}}(DS_n)$ and $\mathrm{wast}_{\mathsf{s}}(DS_n)$ are $\Theta(n)$. This example hence shows tightness of the right-hand bounds in (6.1), (6.3), (6.4) and (6.6) up to constant factors.

The main delay in spreading the rumour in this graph comes from the edge e^* joining the two centres. First, consider the asynchronous case. Here, by Proposition 6.15, $T(e^*) = \mathrm{Exp}(4/n)$. So, the rumour passes from one centre to the other one in $n/4$ time units on average. On the other hand, the leaves learn the rumour in $\Theta(\log n)$ time on average, as in the star graph. Combining the two, we get $\mathrm{wast}_{\mathsf{a}}(DS_n) \sim n/4$.

For the guaranteed spread time, note that if $c < 1/4$ then

$$\mathbb{P}\left[T(e^*) \geq cn \log n\right] = \exp(-n/4 \times cn \log n) \geq 1/n\,.$$

Thus $\mathrm{gst}_{\mathsf{a}}(DS_n) \geq n \log n/4$. Calculations similar to those in Section 6.2.2 show that if $c > 1/4$ then for any vertex v, $\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(DS_n, v) > cn \log n\right] < 1/n$, whence $\mathrm{gst}_{\mathsf{a}}(DS_n) \sim (n \log n)/4$.

In the synchronous case, for any v we have $T(e^*) + 1 \leq \mathrm{ST}_{\mathsf{s}}(DS_n, v) \leq T(e^*) + 2$ and by Observation 6.14,

$$T(e^*) \stackrel{d}{=} 1 + \min\{\mathrm{Geo}(2/n), \mathrm{Geo}(2/n)\} \stackrel{d}{=} 1 + \mathrm{Geo}(4/n - 4/n^2)\,,$$

hence

$$\mathrm{wast}_{\mathsf{s}}(DS_n) = 3 + \mathbb{E}\left[\mathrm{Geo}(4/n - 4/n^2)\right] \sim n/4\,.$$

If $c < 1/4$ then

$$\mathbb{P}\left[3 + \mathrm{Geo}(4/n - 4/n^2) \geq cn \log n\right] \geq (1 - 4/n + 4/n^2)^{cn \log n}\,.$$

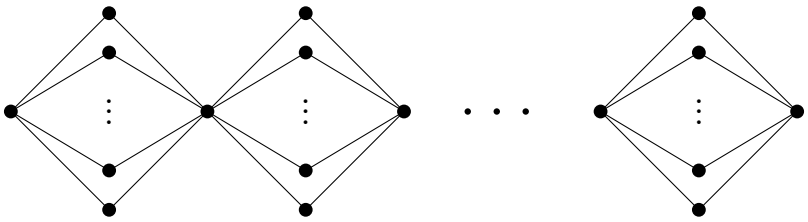


Figure 6.4: a necklace graph, on which the asynchronous push&pull protocol is much quicker than its synchronous variant.

Since $e^{-y} \geq 1 - y \geq e^{-y-y^2}$ for every $y \in [0, 1/4]$,

$$(1 - 4/n + 4/n^2)^{cn \log n} = \exp((-4/n + O(1/n^2))cn \log n) = (1/n)^{4c} e^{o(1)} \geq 1/n \,.$$

While, if $c > 1/4$, then

$$\mathbb{P}\left[3 + \mathrm{Geo}\left(\frac{4}{n} - \frac{4}{n^2}\right) \geq cn \log n\right] = \left(1 - \frac{4}{n} + \frac{4}{n^2}\right)^{cn \log n - 3} \leq \exp(-4c \log n + o(1)) < 1/n,$$

whence $\mathrm{gst_s}(DS_n) \sim (n \log n)/4$.

6.2.5 The necklace graph

Let m and $k \geq 2$ be positive integers, and let G be the necklace graph given in Figure 6.4, where there are m diamonds, each consisting of k edge-disjoint paths of length 2 with the same end vetices, which we call *hubs*. The number of vertices is $n = km + m + 1$. Let us analyze the average spread time.

Consider the asynchronous case first. Proposition 6.15 gives that for each edge e,

$$T(e) \stackrel{d}{=} \mathrm{Exp}(1/2 + 1/k) \stackrel{s}{\leq} \mathrm{Exp}(1/2)$$

and that $\{T(e)\}_{e \in E}$ are independent. Between any two consecutive hubs there are k disjoint paths of length 2, so the communication time between them is stochastically dominated by $Z := \min\{Z_1, \ldots, Z_k\}$, where the Z_i are independent random variables equal in distribution to the sum of two independent $\mathrm{Exp}(1/2)$ random variables.

Lemma 6.18. *We have* $\mathbb{E}[Z] = O(1/\sqrt{k})$.

Proof. For any $t \geq 0$ we have

$$\mathbb{P}[Z > t] = \prod_i \mathbb{P}[Z_i > t] = \mathbb{P}[Z_1 > t]^k \leq \left(1 - \mathbb{P}[\mathrm{Exp}(1/2) \leq t/2]^2\right)^k$$
$$= (2\exp(-t/4) - \exp(-t/2))^k .$$

Using the inequality $2\exp(-t/4) - \exp(-t/2) \leq \exp(-t^2/64)$, valid for all $t \in [0,4]$, we get

$$\mathbb{E}[Z] = \int_0^\infty \mathbb{P}[Z > t]\,\mathrm{d}t \leq \int_0^4 \exp(-kt^2/64)\mathrm{d}t + \int_4^\infty (2\exp(-t/4))^k \mathrm{d}t$$
$$< \int_0^\infty \exp(-kt^2/64)\mathrm{d}t + \frac{2^{k+2}}{ke^k}$$
$$= 8\sqrt{\pi/k} + (2/e)^k 4/k = O(1/\sqrt{k}) . \qquad \blacksquare$$

By Lemma 6.18, the expected time for all the hubs to learn the rumour is $O(mk^{-1/2})$. Once all the hubs learn the rumour, a degree 2 vertex pulls the rumour in Exp(1) time and by Proposition 2.17, the expected value of the maximum of at most km independent Exp(1) variables is $O(\log km)$. So by linearity of expectation,

$$\mathrm{wast}_{\mathsf{a}}(G) = O(\log n + mk^{-1/2}).$$

In the synchronous case, for any G we have $\mathrm{wast}_{\mathsf{s}}(G) \geq \mathrm{diam}(G)$. For this graph, we get $\mathrm{wast}_{\mathsf{s}}(G) \geq 2m$. Choosing $k = \Theta\left((n/\log n)^{2/3}\right)$ and $m = \Theta\left(n^{1/3}(\log n)^{2/3}\right)$ gives

$$\mathrm{wast}_{\mathsf{a}}(G) = O(\log n) \text{ and } \mathrm{wast}_{\mathsf{s}}(G) = \Omega(n^{1/3}(\log n)^{2/3}) .$$

This graph, which has $\mathrm{wast}_{\mathsf{s}}(G)/\mathrm{wast}_{\mathsf{a}}(G) = \Omega\left((n/\log n)^{1/3}\right)$, is the example promised by Corollary 6.9.

6.3 Extremal spread times for the asynchronous push&pull protocol

In this section we prove Theorem 6.3.

6.3.1 Proof of (6.1) and its tightness

For a given $t \geq 0$, consider the protocol which is the same as push&pull except that, if the rumour has not spread to all vertices by time t, then the new process reinitializes. Coupling the new process with push&pull, we obtain for any $k \in \mathbb{N}_0$ that

$$\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > kt\right] \leq \mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > t\right]^k. \tag{6.14}$$

and

$$\mathbb{P}\left[\mathrm{ST}_{\mathsf{s}}(G,v) > kt\right] \leq \mathbb{P}\left[\mathrm{ST}_{\mathsf{s}}(G,v) > t\right]^k. \tag{6.15}$$

Combining (6.14) with

$$\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > e\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G,v)\right]\right] < 1/e,$$

which comes directly from Markov's inequality, we obtain

$$\mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > e\log n\,\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G,v)\right]\right] < 1/n.$$

Since $\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G,v)\right] \leq \mathrm{wast}_{\mathsf{a}}(G)$ for all v, this gives the right-hand inequality in (6.1) directly from the definition of $\mathrm{gst}_{\mathsf{a}}$. This inequality is tight up to the constant factor, as the double star has $\mathrm{wast}_{\mathsf{a}}(DS_n) = \Theta(n)$ and $\mathrm{gst}_{\mathsf{a}}(DS_n) = \Theta(n\log n)$ (see Section 6.2.4).

To prove the left-hand inequality, let $\tau = \mathrm{gst}_{\mathsf{a}}(G)$ and let v be a vertex such that $\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G,v)\right] = \mathrm{wast}_{\mathsf{a}}(G)$. Then

$$\mathrm{wast}_{\mathsf{a}}(G) = \int_0^\infty \mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > t\right] dt = \sum_{i\in\mathbb{N}_0} \int_{i\tau}^{(i+1)\tau} \mathbb{P}\left[\mathrm{ST}_{\mathsf{a}}(G,v) > t\right] dt \leq \sum_{i\in\mathbb{N}_0} \frac{\tau}{n^i}$$

by (6.14) with $t = \tau$. Hence $\mathrm{wast}_{\mathsf{a}}(G) \leq \tau/(1-1/n)$. This inequality is tight up to a constant factor, as the star has $\mathrm{wast}_{\mathsf{a}}(G_n^*) = \Theta(\mathrm{gst}_{\mathsf{a}}(G_n^*)) = \Theta(\log n)$ (see Section 6.2.2).

6.3.2 Proof of the right-hand bound in (6.2) and its tightness

We will actually prove this using pull operations only. Indeed we will show $\mathrm{wast}_{\mathsf{a}}^{\mathsf{pull}}(G) \leq 4n$, where the superscript pull means the pull protocol only. Since the path has $\mathrm{wast}_{\mathsf{a}}^{\mathsf{pull}}(P_n) \geq \mathrm{wast}_{\mathsf{a}}(P_n) = \Theta(n)$ (see Section 6.2.3), this bound would be tight up to the constant factor.

The proof is by induction: we prove that when there are precisely m uninformed vertices, just b of which have informed neighbours (we call these b vertices the *boundary*

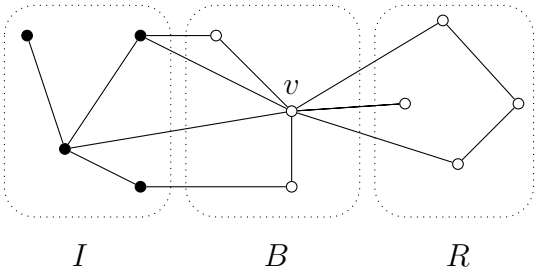


Figure 6.5: first case in the proof of the right-hand bound in (6.2): there exists a boundary vertex v with $3 = d_R(v) \geq d_B(v) = 2$ (informed vertices are black).

vertices), the expected remaining time for the rumour to reach all vertices is at most $4m - 2b$. The inductive step is proved as follows. Let I denote the set of informed vertices, B the set of boundary vertices, and R the set of the remaining vertices. Let $|B| = b$ and $|B| + |R| = m$. Let $d(v)$ denote the degree of v in G and, for a set S of vertices, let $d_S(v)$ denote the number of neighbours of v in S. We consider two cases.

Firstly, suppose that there exists a boundary vertex v with $d_R(v) \geq d_B(v)$ (see Figure 6.5). We can for the next step ignore all calls from vertices other than v, so the process is forced to wait until v is informed before any other vertices. This clearly gives an upper bound on the spread time. The expected time taken for v to pull the rumour from vertices in I is

$$\frac{d(v)}{d_I(v)} = \frac{d_I(v) + d_R(v) + d_B(v)}{d_I(v)} \leq 1 + \frac{2d_R(v)}{d_I(v)} \leq 1 + 2d_R(v).$$

Once v is informed, the number of uninformed vertices decreases by 1, and the number of boundary vertices increases by $d_R(v) - 1$. The inductive hypothesis concludes this case since

$$1 + 2d_R(v) + 4(m-1) - 2(b + d_R(v) - 1) < 4m - 2b.$$

Otherwise, if there is no such v, then any boundary vertex v has a 'pulling rate' of

$$\frac{d_I(v)}{d_I(v) + d_R(v) + d_B(v)} \geq \frac{1}{1 + d_R(v) + d_B(v)} \geq \frac{1}{2d_B(v)} \geq \frac{1}{2b}.$$

Since there are b boundary vertices, together they have a pulling rate of at least $1/2$ (see, e.g., Proposition 2.11), so the expected time until a boundary vertex is informed is at most 2. Once this happens, m decreases by 1 and b either does not decrease or decreases by at most 1, and the inductive hypothesis concludes the proof.

6.3.3 Proof of the left-hand bound in (6.2) and its tightness

In this section we show for any vertex v_0 of a graph G we have $\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G, v_0)\right] = \Omega(\log n)$. This is tight as the star has $\mathrm{wast}_{\mathsf{a}}(G_n^*) = O(\log n)$ (see Section 6.2.2). We give an argument for an equivalent protocol, defined below.

Definition 6.19 (Two-clock-per-edge protocol)**.** On every edge place two Poisson clocks, one near each end vertex. All clocks are independent. On an edge joining vertices u and v, the clocks both have rate $\deg(u)^{-1} + \deg(v)^{-1}$. Note that this is the rate of calls along that edge, combined, from u and v (see Proposition 6.15). At any time that the clock near u on an edge uv rings, and v knows the rumour but u does not, the rumour is passed to u.

Proposition 6.20. *The two-clock-per-edge protocol is precisely equivalent to the asynchronous push&pull protocol.*

Proof. Consider an arbitrary moment during the execution of the two-clock-per-edge protocol. Let I denote the set of informed vertices. For any edge uv with $u \in I$ and $v \notin I$, the rate of calls along uv is $\deg(u)^{-1}+\deg(v)^{-1}$. Moreover, the edges act independently. So, the behaviour of the protocol at this moment is exactly the same as that of the asynchronous push&pull protocol. Hence, the two protocols are equivalent. ■

In view of Proposition 6.20, we may work with the two-clock-per-edge protocol instead. Let X_v be the time taken for the first clock located near v to ring. Then X_v is distributed as $\mathrm{Exp}(f(v))$ where $f(v) = 1 + \sum \deg(u)^{-1}$, the sum being over all neighbours u of v. Hence, $\sum f(v) = 2n$.

On the other hand, for a vertex $v \neq v_0$ to learn the rumour, at least one of clocks located near v must ring. Thus

$$\max\{X_v : v \in V(G)\backslash\{v_0\}\} \overset{s}{\leq} \mathrm{ST}_{\mathsf{a}}(G, v_0)\,.$$

Let $X = \max\{X_v : v \in V(G)\backslash\{v_0\}\}$. Hence to prove $\mathbb{E}\left[\mathrm{ST}_{\mathsf{a}}(G, v_0)\right] = \Omega(\log n)$ it suffices to show $\mathbb{E}\left[X\right] = \Omega(\log n)$.

Let $\tau = \log(n-1)/3$ and $A = V(G) \setminus \{v_0\}$, Then we have

$$\begin{aligned}
\mathbb{P}[X < \tau] &= \prod_{v \in A} (X_v < \tau) \\
&= \prod_{v \in A} (1 - e^{-\tau f(v)}) \\
&\leq \exp\left(- \sum_v e^{-\tau f(v)} \right) \\
&\leq \exp\left(-(n-1) e^{-\tau \sum_v f(v)/(n-1)} \right) \\
&\leq \exp\left(-(n-1) e^{-3\tau} \right) = e^{-1} .
\end{aligned}$$

Here the first inequality follows from $1 - x \leq e^{-x}$, the second from the arithmetic-geometric mean inequality, and the last one from $2n = \sum_v f(v) \leq 3(n-1)$ which holds for $n \geq 3$. Consequently, we have

$$\mathbb{E}[X] \geq \mathbb{P}[X \geq \tau] \, \tau \geq (1 - e^{-1}) \log(n-1)/3 = \Omega(\log n) ,$$

as required.

6.3.4 Proof of (6.3) and its tightness

The bounds in (6.3) follow immediately from (6.1) and (6.2). The left-hand bound is tight as the star has $\mathrm{gst}_{\mathsf{a}}(G_n^*) = \Theta(\log n)$ (see Section 6.2.2), and the right-hand bound is tight as the double star has $\mathrm{gst}_{\mathsf{a}}(DS_n) = \Theta(n \log n)$ (see Section 6.2.4).

6.4 Extremal spread times for the synchronous push&pull protocol

In this section we prove Theorem 6.4. The bound $\mathrm{gst}_{\mathsf{s}}(G) = O(n \log n)$ is a direct consequence of bounds (6.5) and (6.4). This bound is tight as the double star has guaranteed spread time $\Theta(n \log n)$ (see Section 6.2.4). The proof of (6.4) and its tightness are exactly the same as that for (6.1). Below we will prove bound (6.5), i.e. $\mathrm{wast}_{\mathsf{s}}(G) = O(n)$, which would be tight, up to the constant factor, as the path has diameter $n-1$ and hence $\mathrm{wast}_{\mathsf{s}}(P_n) \geq n - 1$.

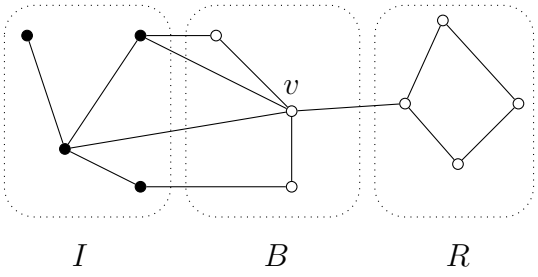


Figure 6.6: second case in the proof of (6.5): for all boundary vertices v we have $d_R(v) < d_B(v)$ (informed vertices are black).

The proof for (6.5) is similar to the one for the right-hand bound in (6.2) given in Section 6.3.2. Let $\alpha = \sqrt{e}/(\sqrt{e}-1)$. We consider the pull operation only, and will prove inductively that when there are m uninformed vertices and b boundary vertices, the expected remaining time for the rumour to reach all vertices is at most $(2+\alpha)m - 2b$, and it follows that $\mathrm{wast}_{\mathsf{s}}(G) = O(n)$. The inductive step is proved as follows. Let I denote the set of informed vertices, B the set of boundary vertices, and R the set of the remaining vertices. Let $|B| = b$ and $|B| + |R| = m$. Let $d(v)$ denote the degree of v in G and, for a set S of vertices, let $d_S(v)$ denote the number of neighbours of v in S. Consider two cases.

Firstly, suppose that there is a vertex $v \in B$ such that $d_R(v) \geq d_B(v)$. In this case, for the next step, we ignore all calls from vertices other than v and wait until v is informed before any other uninformed vertex. This gives an upper bound on the spread time. The expected time taken for v to pull the rumour from vertices in I is

$$1 + \mathbb{E}\left[\mathrm{Geo}\left(\frac{d_I(v)}{d(v)}\right)\right] = \frac{d(v)}{d_I(v)} = \frac{d_I(v) + d_R(v) + d_B(v)}{d_I(v)} \leq 2d_R(v) + 1.$$

Once v is informed, the number of uninformed vertices decreases by 1 and the number of boundary vertices increases by $d_R(v) - 1$. By the inductive hypothesis the expected time for the spread of the rumour is at most

$$2d_R(v) + 1 + (\alpha+2)(m-1) - 2(b + d_R(v) - 1) < (\alpha+2)m - 2b.$$

Next consider the case that $d_R(v) < d_B(v)$ for all $v \in B$ (see Figure 6.6). For each boundary vertex v we have

$$\frac{d_I(v)}{d(v)} = \frac{d_I(v)}{d_I(v) + d_R(v) + d_B(v)} \geq \frac{1}{1 + d_R(v) + d_B(v)} \geq \frac{1}{2d_B(v)} \geq \frac{1}{2b}.$$

Let X denote the time taken until the next vertex is informed. Then we have

$$X = 1 + \min\{X_1, X_2, \ldots, X_b\}\,,$$

where the X_i's are geometric random variables with parameters at least $1/2b$, and correspond to the waiting times of the boundary vertices, and they are independent since we are considering pull operations only. Thus we have

$$\mathbb{E}\left[X-1\right] = \sum_{t\in\mathbb{N}} \mathbb{P}\left[X-1 \geq t\right] = \sum_{t\in\mathbb{N}} \prod_{i\in[b]} \mathbb{P}\left[X_i \geq t\right] \leq \sum_{t\in\mathbb{N}} \left(1-\frac{1}{2b}\right)^{tb} \leq \sum_{t\in\mathbb{N}} e^{-t/2} = \alpha - 1\,,$$

so a boundary vertex learns the rumour after at most α units of time on average, at which time the number of boundary vertices either does not decrease or decreases by 1. By inductive hypothesis again, the average spread time is at most

$$\alpha + (\alpha+2)(m-1) - 2(b-1) = (\alpha+2)m - 2b,$$

which completes the proof.

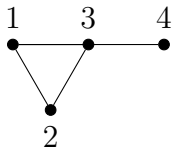
6.5 Comparison of the two protocols

Assuming Theorems 6.6 and 6.8, in this section we prove Corollary 6.9. The left-hand bound follows from Theorem 6.6; it is tight, up to the constant factor, as the star has $\mathrm{gst}_{\mathsf{a}}(G_n^*) = \Theta(\log n)$ and $\mathrm{gst}_{\mathsf{s}}(G_n^*) = 2$ (see Section 6.2.2). The right-hand bound follows from Theorem 6.8 by choosing $\alpha = 1/3$. A graph G was given in Section 6.2.5 having $\mathrm{wast}_{\mathsf{s}}(G)/\mathrm{wast}_{\mathsf{a}}(G) = \Omega\left((n/\log n)^{1/3}\right)$. Using (6.1) and (6.4), we get $\mathrm{gst}_{\mathsf{s}}(G)/\mathrm{gst}_{\mathsf{a}}(G) = \Omega\left(n^{1/3}(\log n)^{-4/3}\right)$ for this G.

In the following subsections we prove Theorems 6.6 and 6.8.

6.5.1 The lower bound

In this section we prove Theorem 6.6. Let G be an n-vertex graph and let s denote the vertex starting the rumour. We give a coupling between the two versions. Consider a 'collection of calling lists for vertices': for every vertex u, we have an infinite list of vertices, each entry of which is a uniformly random neighbour of u, chosen independently from other entries, see Figure 6.7. The coupling is built by using the same collection of calling lists



calling list for vertex 1:	2	2	3	2	3	$\cdots$
calling list for vertex 2:	1	3	3	1	3	$\cdots$
calling list for vertex 3:	2	2	1	4	1	$\cdots$
calling list for vertex 4:	3	3	3	3	3	$\cdots$

Figure 6.7: a collection of calling lists for vertices

for the two versions of the push&pull protocol. Note that $\mathrm{ST}_{\mathsf{s}}(G,s)$ is determined by this collection, but to determine $\mathrm{ST}_{\mathsf{a}}(G,s)$ we also need to know the Poisson processes associated with the vertices.

Let B denote the event '$\mathrm{ST}_{\mathsf{s}}(G,s) \le 2\,\mathrm{gst}_{\mathsf{s}}(G)$', which depends on the calling lists only. Inequality (6.15) gives $\mathbb{P}[B^c] \le 1/n^2$. Partition the time interval $[0, 2\,\mathrm{gst}_{\mathsf{s}}(G) \times 4\log n)$ into subintervals $[0, 4\log n)$, $[4\log n, 8\log n)$, etc. Consider a 'decelerated' variant of the asynchronous push&pull protocol in which each vertex makes a call the first time its clock rings in each subinterval (if it does), but ignores later clock rings in that subinterval (if any). The spread time in this protocol is stochastically larger than that in the asynchronous push&pull protocol, so without loss of generality we may and will work with the decelerated variant. Let A denote the event 'during each of these $2\,\mathrm{gst}_{\mathsf{s}}(G)$ subintervals, all clocks ring at least once.' If A happens, then an inductive argument gives that for any $1 \le k \le 2\,\mathrm{gst}_{\mathsf{s}}(G)$, the set of informed vertices in the decelerated variant at time $4k\log n$ contains the set of informed vertices after k rounds of the synchronous version. Hence, if both A and B happen, then we would have

$$\mathrm{ST}_{\mathsf{a}}(G,s) \le (4\log n)\,\mathrm{ST}_{\mathsf{s}}(G,s) \le (8\log n)\,\mathrm{gst}_{\mathsf{s}}(G)\,.$$

Hence to complete the proof, we need only show that $\mathbb{P}[A^c] \le 1/n - 1/n^2$.

Let I denote a given subinterval. In the asynchronous version, the clock of any given vertex rings with probability at least $1 - n^{-4}$ during I. By the union bound, all clocks ring at least once during I, with probability at least $1 - n^{-3}$. The number of subintervals in the definition of A is $2\,\mathrm{gst}_{\mathsf{s}}(G)$, which is $O(n\log n)$ by (6.6). By the union bound again, $\mathbb{P}[A^c] = O(\log n/n^2)$, as required.

Theorem 6.6 is tight, up to the constant factor, as the star has $\mathrm{gst}_{\mathsf{a}}(G_n^*) = \Theta(\log n)$ and $\mathrm{gst}_{\mathsf{s}}(G_n^*) = 2$ (see Section 6.2.2).

6.5.2 The upper bound

In this section we prove Theorem 6.8. We first sketch the proof. The main ingredients in the proof are a coupling between the two protocols, and sharp concentration bounds. Consider the asynchronous version. List the vertices in the order their clocks ring. The list ends once all the vertices are informed. Now consider the natural coupling between the two protocols, the synchronous actions follow the same ordering as in the list. We partition the list into blocks according to a certain rule in such a way that the blocks have the following property: the synchronous protocol in each round will inform a superset of the set of vertices informed by the asynchronous variant in any single block. For example, if we require that in each block each vertex communicates with the others at most once, then we would have this property. However, in order to get our bound, we need to use a more delicate rule for building the blocks. To conclude, we find an upper bound for the number of blocks, which coincides with the right-hand side of (6.7).

We now give the details. Fix $\alpha \in [0, 1)$. We want to prove

$$\mathrm{gst}_{\mathsf{s}}(G) \leq n^{1-\alpha} + 64\mathrm{gst}_{\mathsf{a}}(G)n^{(1+\alpha)/2} \,. \tag{6.16}$$

Let us fix an arbitrary starting vertex. Let $B_1, B_2, \ldots$ be an i.i.d. sequence of vertices, where B_i is a uniformly random vertex of G. For each i, let W_i be a uniformly random neighbour of B_i, chosen independently of all other choices. Hence, $W_1, W_2, \ldots$ is also an i.i.d. sequence of vertices (not necessarily having uniform distribution). We define a coupling between the two protocols by using the two sequences $(B_i)_{i\in\mathbb{N}}$ and $(W_i)_{i\in\mathbb{N}}$.

To define the coupled asynchronous scenario, we also need to know the ringing times of the clocks. Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. exponentials with rate n (and mean $1/n$), and let this sequence be independent of $(B_i)_{i\in\mathbb{N}}$ and $(W_i)_{i\in\mathbb{N}}$. Then the coupled asynchronous scenario proceeds as follows: at time Z_1 the clock of vertex B_1 rings and it contacts W_1, then at time $Z_1 + Z_2$ the clock of B_2 rings and it contacts W_2, and so on.

We now define a third rumour spreading scenario, which corresponds to the so-called *sequential protocol* [115]. This protocol works as the asynchronous one except we put $Z_i = 1$ for all i. Hence, the scenario only depends on the sequences $(B_i)_{i\in\mathbb{N}}$ and $(W_i)_{i\in\mathbb{N}}$. Let N denote the first time that this protocol has informed all the vertices. Note that $N \geq n - 1$. Observe that, in the asynchronous scenario, all vertices are informed right after N clocks have rung, and the spread time is $\sum_{i=1}^{N} Z_i$.

The following lemma relates $\mathrm{gst}_{\mathsf{a}}(G)$ and N.

Lemma 6.21. *Define the event*

$$\mathcal{A} := \{N \leq 4n\,\mathrm{gst}_{\mathsf{a}}(G)\} \,.$$

Then we have $\mathbb{P}[\mathcal{A}] \geq 1 - O(1/n^2)$.

Proof. As the spread time of the asynchronous scenario is $\sum_{i=1}^{N} Z_i$, by definition of $\mathrm{gst}_{\mathsf{a}}$ we have $\mathbb{P}\left[\sum_{i=1}^{N} Z_i > \mathrm{gst}_{\mathsf{a}}(G)\right] \leq 1/n$. By (6.14) we have

$$\mathbb{P}\left[\sum_{i=1}^{N} Z_i > 2\,\mathrm{gst}_{\mathsf{a}}(G)\right] \leq 1/n^2\,.$$

So we need only show that

$$\mathbb{P}\left[N > 2\sum_{i=1}^{N} nZ_i\right] = O(1/n^2)\,.$$

Since Z_i's are i.i.d. exponentials with rate n, the random variables nZ_i are i.i.d. exponentials with rate 1 (see, e.g., Proposition 2.12), so for any fixed t, Lemma 2.21 (concentration of the sum of i.i.d. exponential random variables) gives

$$\mathbb{P}\left[\sum_{i=1}^{t} nZ_i < t/2\right] = e^{-ct}$$

for some positive constant c. Since $N \geq n-1$, we have

$$\mathbb{P}\left[N > 2\sum_{i=1}^{N} nZ_i\right] \leq \sum_{t=n-1}^{\infty} \mathbb{P}\left[N > 2\sum_{i=1}^{N} nZ_i \,\middle|\, N = t\right] = \sum_{t=n-1}^{\infty} e^{-ct} = \exp(-\Omega(n))\,,$$

as required. ∎

To define the coupled synchronous scenario, we need some definitions. For each vertex v, let $\pi(v)$ denote the probability that $W_j = v$. Recall that this probability does not depend on j. Call a vertex v *special* if $\pi(v) > n^{\alpha-1}$. Note that since $\sum \pi(v) = 1$, there are less than $n^{1-\alpha}$ special vertices.

We partition the list $B_1, W_1, B_2, W_2, \ldots$ into infinitely many finite *blocks* as follows. The first block is of the form

$$B_1, W_1, B_2, W_2, \ldots, B_j, W_j\,,$$

with j as large as possible, subject to the following conditions:

1. We have $B_i \notin \{B_1, W_1, \ldots, B_{i-1}, W_{i-1}\}$ for all $1 < i \leq j$.

2. If $W_i \in \{B_1, W_1, \ldots, B_{i-1}, W_{i-1}\}$ for some $1 < i \leq j$, then W_i is special.

Note that we choose the block to be as long as possible, hence we stop at W_j only if B_{j+1} already appears in $B_1, \ldots, W_j$, or W_{j+1} is non-special and it appears in $B_1, \ldots, W_j$, or both. If we have stopped at W_j, then a new block is started from B_{j+1}, and this process is iterated forever to define all the blocks. Note that each block has an even number of elements.

Let $S_1, S_2, \ldots$ denote the sizes of the blocks, and let N_b be the smallest number such that

$$S_1 + S_2 + \cdots + S_{N_b} \geq 2N \, .$$

The following lemma relates the spread time of the synchronous protocol and N_b.

Lemma 6.22. *The spread time of the synchronous push&pull protocol is stochastically smaller than* $N_b + n^{1-\alpha}$.

Proof. In this proof we only consider the finite list $B_1, W_1, \ldots, B_N, W_N$, which is partitioned into blocks as discussed before. We further split the blocks into smaller ones according to the following rule. Let v be a special vertex and assume that in the sequential scenario, it is informed exactly at time i. So, either $B_i = v$ or $W_i = v$. If the elements $B_i, W_i, B_{i+1}, W_{i+1}$ are contained in the same block, then we split this block at this point, putting everything up to B_i, W_i in one block and B_{i+1}, W_{i+1} and everything after in the other one. Since the number of splits equals the number of special vertices, and there are less than $n^{1-\alpha}$ special vertices, the new total number of blocks is less than $N_b + n^{1-\alpha}$. We work with these refined blocks for the rest of the proof.

We couple with a modified version of the synchronous push&pull protocol, which we call the *lazy* scenario. We define the coupled lazy scenario inductively using the blocks. Assume that the kth block is

$$B_i, W_i, B_{i+1}, W_{i+1}, \ldots, B_j, W_j \, .$$

Then in the kth round of the lazy scenario, vertex B_i contacts W_i, vertex B_{i+1} contacts W_{i+1} and so on, up until vertex B_j contacts W_j (all these communications happen at the same time). Moreover, a vertex that does not appear in this block, does not perform any action in the kth round. It is clear that stochastic upper bounds for the spread time of this lazy scenario carries over to the synchronous push&pull scenario.

To complete the proof we will show that the set of vertices informed by the lazy scenario after k rounds equals the set of vertices informed by the sequential scenario right after time $(S_1 + \cdots + S_k)/2$. (The factor of 2 appears here because a block with r communications has length $2r$.) The proof proceeds by induction. Assume that the kth block is

$$B_i, W_i, B_{i+1}, W_{i+1}, \ldots, B_j, W_j \,.$$

If no repetition happens in this block at all, then it is clear that the lazy scenario in one round informs every vertex which the sequential one informs during times $i, i+1, \ldots, j$. Notice the possible problem if a repetition happens: if during this block, x contacts y and tells her the rumour for the first time, and z also contacts y and asks her the rumour, then in the sequential scenario both y and z will learn the rumour by time j, whereas in the lazy scenario this is not the case because these operations happen at exactly the same time. However, if v is a repeated vertex in this block, then v is a special vertex, and moreover by the secondary splitting of the blocks, we know that it cannot be the case that v is informed in this block for the first time and appears again later in the block. Hence, no 'informing path' of length greater than one can appear in this block, and the proof is complete. ■

Let $k = 64\,\mathrm{gst}_{\mathsf{a}}(G)n^{(1+\alpha)/2}$. The following lemma bounds N_b.

Lemma 6.23. *Define the event*

$$\mathcal{B} := \{S_1 + \cdots + S_k \geq 8\,\mathrm{gst}_{\mathsf{a}}(G)n\} \,.$$

Then we have $\mathbb{P}[\mathcal{B}] \geq 1 - O(1/n^2)$.

Before proving this lemma, let us see why it concludes the proof of Theorem 6.8. By Lemmas 6.21 and 6.23 and the union bound, with probability at least $1 - 1/n$ both events $\mathcal{A}$ and $\mathcal{B}$ happen. Assume this is the case. Then we have

$$S_1 + \cdots + S_k \geq 8\,\mathrm{gst}_{\mathsf{a}}(G)n \geq 2N \,,$$

which means $N_b \leq k$ by the definition of N_b. Together with Lemma 6.22, this implies that with probability at least $1 - 1/n$, the spread time of the synchronous push&pull protocol is at most $k + n^{1-\alpha}$, which gives (6.16).

Proof of Lemma 6.23. Let $\ell = n^{(1-\alpha)/2}/4$. We first show that

$$\mathbb{P}[S_1 > 2\ell] \geq 1/2 \,. \tag{6.17}$$

Let j be arbitrary. We compute the conditional probability of $\{S_1 \geq 2j+2\}$ given that $\{S_1 \geq 2j\}$. On the event $\{S_1 \geq 2j\}$, the conditional probability that B_{j+1} is a repetition of a vertex already in the block $B_1, W_1, B_2, W_2, \ldots, B_j, W_j$ is $2j/n$. The probability that W_{j+1} is a repetition of a non-special vertex in the block is bounded above by $2jn^{\alpha-1}$, since there are at most $2j$ distinct vertices in the block so far, and W_{j+1} is a given non-special vertex with probability at most $n^{\alpha-1}$. So, we have

$$\mathbb{P}\left[S_1 \geq 2j+2 | S_1 \geq 2j\right] \geq 1 - 4jn^{\alpha-1} \geq \exp(-8jn^{\alpha-1}) \,.$$

Consequently,

$$\mathbb{P}\left[S_1 > 2\ell\right] \geq \prod_{j=1}^{\ell} \exp(-8jn^{\alpha-1}) = \exp\left(-4n^{\alpha-1}\ell(\ell+1)\right) \geq 1/2$$

by the choice of ℓ, so (6.17) holds.

Observe that the block sizes $S_1, S_2, \ldots$ are i.i.d., and each of them is at least 2ℓ with probability at least $1/2$. So we have

$$\mathbb{P}\left[S_1 + \cdots + S_k \leq k\ell/2\right] \leq \mathbb{P}\left[\mathrm{Bin}(k, 1/2) \leq k/4\right] \leq \exp(-k/16) = O(1/n^2) \,,$$

where for the second inequality, we have used the lower tail Chernoff bound (2.2). ■

Chapter 7

Synchronous push&pull protocol on two random graphs

In this chapter[1] we study the synchronous push&pull protocol on random k-trees and random k-Apollonian networks, both of which we have already encountered in Section 3.5. We recall the definition of random k-trees.

Definition 7.1 (Random k-tree process [71])**.** Let k be a positive integer. Build a sequence $G(0), G(1), \ldots$ of random graphs as follows. The graph $G(0)$ is just a clique on k vertices. For each $1 \leq t \leq n$, $G(t)$ is obtained from $G(t-1)$ as follows: a k-clique of $G(t-1)$ is chosen uniformly at random, a new vertex is born and is joined to all vertices of the chosen k-clique. The graph $G(n)$ is called a *random k-tree* on $n+k$ vertices. See Figure 7.1 for an illustration.

Remark 7.2. A *k-tree* is defined in a similar manner (see, e.g., [86, Definition 2.1.8]), the only difference being that in each step we choose an *arbitrary* k-clique instead of a uniformly random one.

Sometimes it is convenient to view this as a 'random graph evolving in time.' In this interpretation, in every round $1, 2, \ldots$, a new vertex is 'born' and is added to the evolving graph, and $G(t)$ denotes the graph at the end of round t. Observe that $G(t)$ has $k+t$ many vertices and $kt+1$ many k-cliques. Also, if $t > 0$, then each vertex in $G(t)$ has degree at least k.

[1] This chapter is based on joint work with Ali Pourmiri. The results therein appear in the submitted manuscript [99], an extended abstract of which has been published [100].

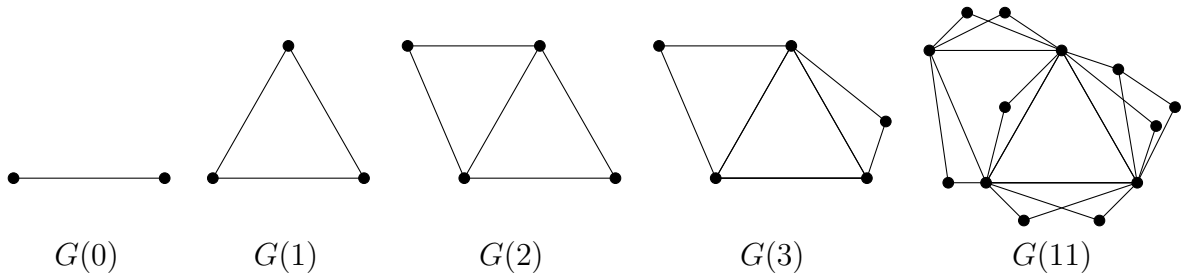


Figure 7.1: an instance of a random 2-tree process

As in the preferential attachment scheme, the random k-tree process enjoys a 'the rich get richer' effect. Think of the number of k-cliques containing any vertex v as the 'wealth' of v (note that this quantity is linearly related to $\deg(v)$). Then, the probability that the new vertex attaches to v is proportional to the wealth of v, and if this happens, the wealth of v increases by $k-1$.

The first result of this chapter is the following theorem.

Theorem 7.3. *Let $k \geq 2$ be fixed and let $f(n) = o(\log\log n)$ be an arbitrary function going to infinity with n. If initially a random vertex of an $(n+k)$-vertex random k-tree knows a rumour, then a.a.s. after $O\left((\log n)^{1+\frac{2}{k}} \cdot \log\log n \cdot f(n)^{\frac{3}{k}}\right)$ rounds of the synchronous push&pull protocol, $n - o(n)$ vertices will know the rumour.*

Remark 7.4. There are two sources of randomness here: the structure of the graph, and the actions of the protocol.

We give a high-level sketch of the proof of Theorem 7.3. Let G be a random k-tree. Let $m = o(n)$ be a suitably chosen parameter, and note that $G(m)$ is a subgraph of G. Consider the connected components of $G - G(m)$. Most vertices born later than round m have relatively small degree, so most these components have a small maximum degree (and logarithmic diameter) thus the rumour spreads quickly inside each of them. A vertex $v \in V(G(m))$ typically has a large degree, but this means there is a high chance that v has a neighbour x with small degree, which quickly receives the rumour from v and spreads it (or vice versa). We build an almost-spanning tree T of $G(m)$ with logarithmic height, such that for every edge uv of T, one of u and v have a small degree, or u and v have a common neighbour with a small degree. Either of these situations mean the rumour is exchanged quickly between u and v. This tree T then works as a 'highway system' to spread the rumour within vertices of $G(m)$ and from them to the components of $G - G(m)$. The main

novelty in this proof is how the almost-spanning tree is built and used. The degrees in a random k-tree, and the sizes of the components of $G - G(m)$ are naturally related to the number of balls of a given colour in certain triangular urn models. In our proofs, we exploit these connections and apply some recent results on triangular urn models from the work of Flajolet, Dumas and Puyhaubert [65] and Janson [79].

Our second main result is the following theorem, which gives a polynomial lower bound for the spread time.

Theorem 7.5. *Let $k \geq 2$ be fixed and let $f(n) = o(\log \log n)$ be an arbitrary function going to infinity with n. Suppose that initially one vertex in the random k-tree, $G(n)$, knows the rumour. Then, a.a.s. the synchronous push&pull protocol needs at least $n^{(k-1)/(k^2+k-1)} f(n)^{-3}$ rounds to inform all vertices of $G(n)$.*

We give a high-level sketch of the proof of Theorem 7.5. A *barrier* in a graph is a subset D of edges of size $O(1)$, whose deletion disconnects the graph. If both endpoints of every edge of a barrier D have large degrees, then the protocol needs a large time to pass the rumour through D. For proving Theorem 7.5, we prove a random k-tree has a barrier a.a.s. The main novelty in this proof is introducing and using the notion of a barrier.

It is instructive to contrast Theorems 7.3 and 7.5. The former implies that if you want to inform almost all the vertices, then you just need to wait for a polylogarithmic number of rounds. The latter implies that, however, if you want to inform each and every vertex, then you have to wait for polynomially many rounds. This is a striking phenomenon and the main message of this chapter is to present a natural class of random graphs in which this phenomenon can be observed. In fact, in applications such as viral marketing and voting, it is more appealing to inform 99 percent of the vertices very quickly instead of waiting a long time until everyone gets informed. For such applications, Theorem 7.3 implies that the synchronous push&pull protocol can be effective even on poorly connected graphs.

It is worth mentioning that bounds for the number of rounds to inform *almost all* vertices have already appeared in the literature, see for instance [49, 69]. In particular, for Chung-Lu graphs with exponent in $(2, 3)$, it is shown [69] that a.a.s. after $O(\log \log n)$ rounds the rumour spreads in $n - o(n)$ vertices, but to inform *all* vertices of the giant component, $\Theta(\log n)$ rounds are needed. This result also shows a great difference between the two cases, however in both cases the required time is quite small.

Informally speaking, a random k-tree looks like the graph in Figure 7.2: there is a big chunk B, containing almost all vertices, in which the rumour spreads quickly, and there are small pieces here and there that are loosely connected to B and it takes a lot of time for the rumour to pass from a piece to B or vice versa.

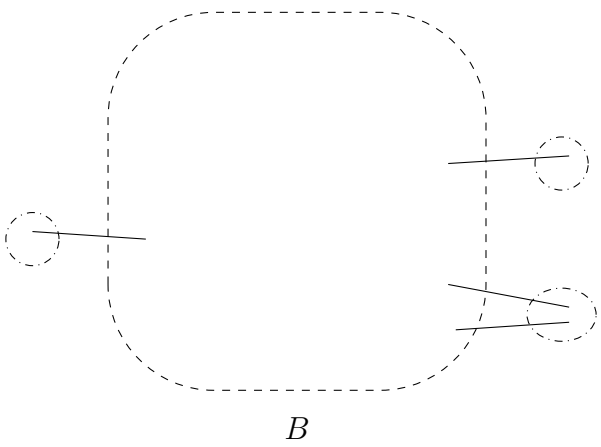


Figure 7.2: informally speaking, a random k-tree looks like this graph: there is a big chunk B, containing almost all vertices, in which the rumour spreads quickly, and there are small pieces here and there that are loosely connected to B and it takes a lot of time for the rumour to pass from a piece to B or vice versa.

A closely related class of graphs is the class of *random k-Apollonian networks*, which was also discussed in Section 3.5. Recall that their construction is very similar to that of random k-trees, with just one difference: if a k-clique is chosen in a certain round, it will never be chosen again. Our third main result is the following theorem.

Theorem 7.6. *Let $k \geq 3$ be fixed and let $f(n) = o(\log\log n)$ be an arbitrary function going to infinity with n. Assume that initially a random vertex of an $(n+k)$-vertex random k-Apollonian network knows a rumour. Then, a.a.s. after*

$$O\left((\log n)^{(k^2-3)/(k-1)^2} \cdot \log\log n \cdot f(n)^{2/k}\right)$$

rounds of the synchronous push&pull protocol, at least $n - o(n)$ vertices will know the rumour.

The proof of Theorem 7.6 is along the lines of that of Theorem 7.3, although there are several differences. Note that we have $(k^2-3)/(k-1)^2 < 1+2/k$, so our upper bound for random k-Apollonian networks is slightly stronger than that for random k-trees.

Unfortunately, our technique for proving Theorem 7.5 does not extend to random k-Apollonian networks, although we believe that a.a.s. we need a polynomial number of rounds to inform all vertices in a random k-Apollonian network as well. We leave this as a conjecture.

For the rest of the chapter, k is a constant greater than 1, and the asymptotics are for n going to infinity. For brevity, the synchronous push&pull protocol will be simply called the 'protocol.' Several times in this chapter we use urn models to analyze the vertices' degrees and the number of vertices in certain parts of a random k-tree. The connections with urn models are built in Section 7.1. In Section 7.2 we study basic properties of random k-trees, demonstrating their similarities with real-world networks. Theorems 7.3, 7.5, and 7.6 are proved in Sections 7.3, 7.4, and 7.5, respectively.

7.1 Connections with urn models

In this section we build some connections between random k-trees and urn models (see Section 2.3). First we recall a definition from that section.

Definition 7.7 (Eggenberger-Pólya urn)**.** Start with W_0 white and B_0 black balls in an urn. In every step a ball is drawn from the urn uniformly at random, the ball is returned to the urn, and k balls of the same color are added to the urn. Then $\mathrm{Urn}(W_0, B_0, k, n)$ denotes the number of white balls right after n draws.

Proposition 7.8. *Let* $X \stackrel{d}{=} \mathrm{Urn}(a, b, k, n)$ *and* $w = a + b$*. Then*

$$\mathbb{E}\left[X^2\right] = \left(a + \frac{a}{w}\, kn\right)^2 + \frac{abk^2n(kn+w)}{w^2(w+k)}$$

and for any $c \geq (a+b)/k$ *we have*

$$\mathbb{P}\left[X = a\right] \leq \left(\frac{c}{c+n}\right)^{a/k}.$$

Proof. The first statement follows from the following well known formulae for the expected value and the variance of X (see [96, Corollary 5.1.1] for instance):

$$\mathbb{E}\left[X\right] = a + \frac{a}{w}\, kn\,, \quad \mathrm{variance}(X) = \frac{abk^2n(kn+w)}{w^2(w+k)}\,.$$

For the second statement, we have

$$\begin{aligned}
\mathbb{P}[X=a] &= \frac{b}{a+b}\cdot\frac{b+k}{a+b+k}\cdot\cdots\frac{b+(n-1)k}{a+b+(n-1)k} \\
&= \prod_{i=0}^{n-1}\left(1-\frac{a}{a+b+ik}\right) \\
&\le \prod_{i=0}^{n-1}\left(1-\frac{a}{ck+ik}\right) \\
&\le \exp\left(-\sum_{i=0}^{n-1}\frac{a}{ck+ik}\right) \\
&= \left\{\exp\left(\sum_{i=0}^{n-1}\frac{1}{c+i}\right)\right\}^{-a/k} \le \left\{\exp\left(\int_{x=c}^{c+n}\frac{\mathrm{d}x}{x}\right)\right\}^{-a/k} = \left(\frac{c}{c+n}\right)^{a/k}. \quad \blacksquare
\end{aligned}$$

In this chapter we consider a generalized version of the Eggenberger-Pólya urn, which is known as a 'triangular urn.'

Definition 7.9 (Triangular urn)**.** Let α, γ, δ be nonnegative integers. We start with W_0 white and B_0 black balls in an urn. In every step a ball is drawn from the urn uniformly at random and returned to the urn. Additionally, if the ball is white, then δ white balls and γ black balls are returned to the urn; otherwise, i.e. if the ball is black, then α black balls (and no white ball) are returned to the urn. Let $\mathrm{Urn}\left(W_0, B_0, \begin{bmatrix}\delta & \gamma \\ 0 & \alpha\end{bmatrix}, n\right)$ denote the number of white balls right after n draws.

Note that Eggenberger-Pólya urns correspond to the matrix $\begin{bmatrix}s & 0 \\ 0 & s\end{bmatrix}$.

We will need two inequalities regarding the Gamma function.

Proposition 7.10. *(a) Let m and k be positive integers. Then we have $\Gamma(m+1/k) \le \Gamma(m)m^{1/k}$.*

(b) Let $a > 0$ and $b \ge 1$. Then $\Gamma(a)a^b \le \Gamma(a+b)$.

Proof. (a) Laforgia [90, inequality (2.2)] proved

$$\frac{\Gamma(x+1)}{\Gamma(x+s)} \ge x^{1-s} \qquad \forall x > 0, s \in [0,1]\,. \tag{7.1}$$

Letting $x = m$ and $s = 1/k$ we get

$$\frac{m\Gamma(m)}{\Gamma(m+1/k)} = \frac{\Gamma(m+1)}{\Gamma(m+1/k)} \geq m^{1-1/k} ,$$

and the conclusion follows.

(b) We use induction on $\lfloor b \rfloor$. First, if $\lfloor b \rfloor = 1$, then using (7.1) with $x = a + b - 1$ and $s = 2 - b$ gives

$$\frac{\Gamma(a+b)}{a\Gamma(a)} = \frac{\Gamma(a+b)}{\Gamma(a+1)} \geq (a+b-1)^{b-1} \geq a^{b-1} ,$$

hence $\Gamma(a+b) \geq \Gamma(a)a^b$. For $\lfloor b \rfloor \geq 2$, by inductive hypothesis we have

$$\Gamma(a+b) = (a+b-1)\Gamma(a+b-1) \geq (a+b-1)\Gamma(a)a^{b-1} \geq \Gamma(a)a^b ,$$

as required. ■

The following proposition follows from known results.

Proposition 7.11. *Let* $X \stackrel{d}{=} \mathrm{Urn}\left(W_0, B_0, \begin{bmatrix} \delta & \gamma \\ 0 & \alpha \end{bmatrix}, n\right)$ *and let* r *be a positive integer. If* $\gamma, \delta > 0$, $\alpha = \gamma + \delta$, *and* $r\delta \geq \alpha$, *then we have*

$$\mathbb{E}\left[X^r\right] \leq \left(\frac{\alpha n}{W_0 + B_0}\right)^{r\delta/\alpha} \prod_{i=0}^{r-1} (W_0 + i\delta) + O\left(n^{(r-1)\delta/\alpha}\right) .$$

Proof. [65, Proposition 15] states

$$\mathbb{E}\left[X^r\right] = n^{r\delta/\alpha}\delta^r \frac{\Gamma(W_0/\delta + r)\Gamma((W_0+B_0)/\alpha)}{\Gamma(W_0/\delta)\Gamma((W_0+B_0+r\delta)/\alpha)} + O\left(n^{(r-1)\delta/\alpha}\right) ,$$

Note that

$$\frac{\Gamma(W_0/\delta + r)}{\Gamma(W_0/\delta)} = \prod_{i=0}^{r-1} (i + W_0/\delta) ,$$

and the inequality

$$\frac{\Gamma((W_0+B_0)/\alpha)}{\Gamma((W_0+B_0+r\delta)/\alpha)} \leq ((W_0+B_0)/\alpha)^{-r\delta/\alpha}$$

follows from Proposition 7.10(b), since $r\delta \geq \alpha$. ■

Proposition 7.12. *Suppose that in $G(j)$ vertex x has $A > 0$ neighbours, and is contained in B many k-cliques. Conditional on this, the degree of x in $G(n+j)$ is distributed as*

$$A + \left(\mathrm{Urn}\left(B, kj+1-B, \begin{bmatrix} k-1 & 1 \\ 0 & k \end{bmatrix}, n \right) - B \right) \Big/ (k-1)\,.$$

Proof. We claim that the total number of k-cliques containing x in $G(n+j)$ is distributed as $\mathrm{Urn}\left(B, kj+1-B, \begin{bmatrix} k-1 & 1 \\ 0 & k \end{bmatrix}, n \right)$. At the end of round j, there are B many k-cliques containing x, and $kj+1-B$ many k-cliques not containing x. In each subsequent round $j+1, \ldots, j+n$, a random k-clique is chosen and k new k-cliques are created. If the chosen k-clique contains x, then $k-1$ new k-cliques containing x are created, and 1 new k-clique not containing x is created. Otherwise, i.e. if the chosen k-clique does not contain x, then no new k-cliques containing x is created, and k new k-cliques not containing x are created, and the claim follows.

Hence the number of k-cliques that are created in rounds $j+1, \ldots, j+n$ and contain x is distributed as $\mathrm{Urn}\left(B, kj+1-B, \begin{bmatrix} k-1 & 1 \\ 0 & k \end{bmatrix}, n \right) - B$, and the proof follows by noting that every new neighbour of x creates $k-1$ new k-cliques containing x. ■

In view of Proposition 7.12, we set

$$\mathcal{K} := \begin{bmatrix} k-1 & 1 \\ 0 & k \end{bmatrix}$$

throughout this chapter.

Combining Propositions 7.11 and 7.12 we obtain the following lemma.

Lemma 7.13. *Let $1 \le j \le n$ and let q be a positive integer. Let x denote the vertex born in round j of the random k-tree process. Conditional on any $G(j)$, the probability that x has degree greater than $k + q(n/j)^{(k-1)/k}$ in $G(n)$ is $O\left(q\sqrt{q}\exp(-q)\right)$.*

Proof. Let $X = \mathrm{Urn}\,(k, kj-k+1, \mathcal{K}, n-j)$. By Proposition 7.12, $\deg(x)$ is distributed as $k + (X-k)/(k-1)$. By Proposition 7.11,

$$\mathbb{E}\left[X^q\right] \le (1+o(1)) \left(\frac{k(n-j)}{kj+1} \right)^{\frac{q(k-1)}{k}} \prod_{i=0}^{q-1} (k + i(k-1)) \le \left(\frac{n}{j} \right)^{\frac{q(k-1)}{k}} (k-1)^q (q+1)!\,.$$

Thus,

$$\begin{aligned}\mathbb{P}\left[\deg(x) > k + q(n/j)^{(k-1)/k}\right] &= \mathbb{P}\left[X - k > q(k-1)(n/j)^{(k-1)/k}\right] \\ &\leq \frac{\mathbb{E}\left[X^q\right]}{(q(k-1)(n/j)^{(k-1)/k})^q} \\ &\leq (q+1)!q^{-q} = O\left(q\sqrt{q}\exp(-q)\right) ,\end{aligned}$$

by Stirling's approximation (2.4), as required. ■

7.2 Basic properties of random k-trees

We have already seen that a random k-tree a.a.s. has a logarithmic diameter (see Theorem 3.36). In this section we prove that random k-trees exhibit another important property observed in real-world networks: a large clustering coefficient. We also prove that random k-trees do not expand well, confirming our claim in the Introduction that random k-trees are poorly connected graphs and thus existing techniques do not apply.

Definition 7.14 (Clustering coefficient)**.** The *clustering coefficient* of a graph G, written $cc(G)$, is defined as

$$cc(G) = \frac{1}{|V(G)|}\sum_{u\in V(G)}\frac{|\langle N(u)\rangle|}{\binom{\deg(u)}{2}} ,$$

where $|\langle N(u)\rangle|$ denotes the number of edges xy such that both x and y are adjacent to u.

Let $G(0), G(1), \ldots$ be defined as in Definition 7.1.

Proposition 7.15. *For every positive integer n, deterministically, $cc(G(n)) > 1/2$.*

Proof. Let u be a vertex of $G = G(n)$. When u is born, we have $|\langle N(u)\rangle| = \binom{k}{2}$. Whenever a new vertex is joined to u, $|\langle N(u)\rangle|$ increases by $k-1$. Hence we have $|\langle N(u)\rangle| = (k-1)(\deg(u) - k/2)$. Since $\deg(u) \geq k$ we get

$$\frac{|\langle N(u)\rangle|}{\binom{\deg(u)}{2}} \geq \frac{k}{\deg(u)} .$$

By the Cauchy-Schwarz inequality we have

$$\left(\sum_{u\in V(G)}\deg(u)\right)\left(\sum_{u\in V(G)}1/\deg(u)\right) \geq |V(G)|^2 = (n+k)^2 ,$$

hence

$$cc(G) \geq \frac{1}{|V(G)|} \sum_{u \in V(G)} \frac{k}{\deg(u)} \geq \frac{k}{n+k} \cdot \frac{(n+k)^2}{2|E(G)|} > \frac{1}{2} \,. \qquad \blacksquare$$

Definition 7.16 (Vertex expansion)**.** The *vertex expansion, vertex isoperimetric number* of a graph G (also known as the *vertex isoperimetric number* of G), written $\alpha(G)$, is defined as

$$\alpha(G) = \min \left\{ \frac{|\partial S|}{|S|} : S \subseteq V(G), 0 < |S| \leq |V(G)|/2 \right\} ,$$

where ∂S denotes the set of vertices in $V(G) \setminus S$ that have a neighbour in S.

Definition 7.17 (Conductance)**.** The *conductance* of a graph G, written $\Phi(G)$, is defined as

$$\Phi(G) = \min \left\{ \frac{e(S, V(G) \setminus S)}{\texttt{vol}(S)} : S \subseteq V(G), 0 < \texttt{vol}(S) \leq \texttt{vol}(V(G))/2 \right\} ,$$

where $e(S, V(G) \setminus S)$ denotes the number of edges between S and $V(G) \setminus S$, and $\texttt{vol}(S) = \sum_{u \in S} \deg(u)$ for every $S \subseteq V(G)$.

Proposition 7.18. *Deterministically $G(n)$ has vertex expansion $O(k/n)$, and a.a.s. its conductance is $O\left(\log n \cdot n^{-1/k}\right)$.*

Proof. Let $G = G(n)$. Since G is a k-tree and by definition has treewidth k, by [86, Lemma 5.3.1] there exists a partition (A, B, C) of $V(G)$ such that

1. $|C| = k + 1$,
2. $(n-1)/3 \leq |A| \leq 2(n-1)/3$ and $(n-1)/3 \leq |B| \leq 2(n-1)/3$, and
3. there is no edge between A and B.

At least one of A and B, say A, has size less than $(n+k)/2$. Then

$$\alpha(G) \leq \frac{|\partial A|}{|A|} \leq \frac{k+1}{(n-1)/3} = O(k/n) \,.$$

At least one of A and B, say B, has volume less than $\texttt{vol}(G)/2$. Then since all vertices in G have degrees at least k,

$$\Phi(G) \leq \frac{e(B, A \cup C)}{\texttt{vol}(B)} \leq \frac{e(B, C)}{k|B|} \leq \frac{(k+1)\Delta(G)}{k(n-1)/3} = O(\Delta(G)/n) \,.$$

Hence to prove $\Phi(G) = O\left(\log n \cdot n^{-1/k}\right)$ it suffices to show that a.a.s. we have

$$\Delta(G) \leq k + (2\log n)n^{1-1/k}\,. \tag{7.2}$$

Let $q = \lfloor 2\log n \rfloor$ and let x be a vertex born in one of the rounds $1, 2, \ldots, n$. By Lemma 7.13,

$$\mathbb{P}\left[\deg(x) > k + qn^{1-1/k}\right] = O(q\sqrt{q}\exp(-q)) = o(1/n)\,.$$

An argument similar to the proof of Lemma 7.13 shows that the probability that a vertex in $G(0)$ has degree greater than $k + qn^{1-1/k}$ is $o(1/n)$ as well. A union bound over all vertices gives(7.2), as required. ■

7.3 Upper bound for random k-trees

In this section we prove Theorem 7.3. Once we have the following lemma, our problem reduces to proving a structural result for random k-trees.

Lemma 7.19. *Let χ and τ be fixed positive integers. Let G be an n-vertex graph and let $\Sigma \subseteq V(G)$ with $|\Sigma| = n - o(n)$ be such that for every pair of vertices $u, v \in \Sigma$ there exists a (u,v)-path $uu_1u_2\cdots u_{l-1}v$ such that $l \leq \chi$ and for every $0 \leq i \leq l-1$ we have $\min\{\deg(u_i), \deg(u_{i+1})\} \leq \tau$ (where we define $u_0 = u$ and $u_l = v$). If a random vertex in G knows a rumour, then a.a.s. after $6\tau(\chi + \log n)$ rounds of the protocol, at least $n - o(n)$ vertices will know the rumour.*

Proof. The proof is along the lines of that of [62, Theorem 2.2]. We show that given any $u, v \in \Sigma$, if u knows the rumour then with probability at least $1 - o\left(n^{-2}\right)$ after $6\tau(\chi + \log n)$ rounds v will know the rumour. The lemma follows by using the union bound and noting that a random vertex lies in Σ a.a.s. Consider the (u,v)-path $uu_1u_2\cdots u_{l-1}v$ promised by the hypothesis. We bound from below the probability that the rumour is passed through this path.

For every $0 \leq i \leq l-1$, the number of rounds taken for the rumour to pass from u_i to u_{i+1} is one plus a geometric random variable with success probability at least $1/\tau$ (if $\deg(u_i) \leq \tau$, this is the number of rounds needed for u_i to push the rumour along the edge, and if $\deg(u_{i+1}) \leq \tau$, this is the number of rounds needed for u_{i+1} to pull the rumour along the edge). The random variables corresponding to distinct edges are mutually independent. Hence the probability that the rumour is not passed in $6\tau(\chi + \log n)$ rounds is at most the probability that the number of heads in a sequence of $6\tau(\chi + \log n)$ independent biased

coin flips, each having probability $1/\tau$ of being heads, is less than l. Let X denote the number of heads in such a sequence. Then using the lower tail Chernoff bound (2.2) and noting that $\mathbb{E}[X] = 6(\chi + \log n)$ we get

$$\mathbb{P}[X < l] \leq \mathbb{P}[X \leq \mathbb{E}[X]/6] \leq \exp(-(5/6)^2 \mathbb{E}[X]/2) \leq \exp(-(5/6)^2(6\log n)/2)\,,$$

which is $o(n^{-2})$, as required. ■

Let $f(n) = o(\log\log n)$ be an arbitrary function going to infinity with n, and let

$$m = \left\lceil \frac{n}{f(n)^{3/(k-1)}(\log n)^{2/(k-1)}} \right\rceil . \tag{7.3}$$

Also let $q = \lceil 4\log\log n \rceil$ and let

$$\tau = 2k + q(n/m)^{1-1/k} . \tag{7.4}$$

By Theorem 3.36, a.a.s. a random k-tree on $n+k$ vertices has diameter $O(\log n)$. Theorem 7.3 thus follows from Lemma 7.19 and the following structural result, which we prove in the rest of this section.

Lemma 7.20. *Let G be an $(n+k)$-vertex random k-tree. A.a.s. there exists $\Sigma \subseteq V(G)$ satisfying the conditions of Lemma 7.19 with τ defined in (7.4) and $\chi = O(\log n + \mathrm{diam}(G))$.*

For the rest of this section, G is an $(n+k)$-vertex random k-tree, and whenever we talk about the degree of a vertex, we mean its degree in G, unless specified otherwise. Recall from Definition 7.1 that $G = G(n)$, where $G(0), G(1), \ldots$ is the random k-tree process. Consider the graph $G(m)$, which has $k+m$ vertices and $mk+1$ many k-cliques. For an edge e of $G(m)$, let $N(e)$ denote the number of k-cliques of $G(m)$ containing e. We define a spanning forest F of $G(m)$ as follows: for every $1 \leq t \leq m$, if the vertex x born in round t is joined to the k-clique C, then in F, x is joined to a vertex $u \in V(C)$ such that

$$N(xu) = \max_{v \in V(C)} N(xv)\,.$$

Note that F consists of k trees and the k vertices of $G(0)$ lie in distinct trees. Think of these trees as rooted at these vertices. The tree obtained from F by merging these k vertices is the 'highway system' described in the sketch of the proof of Theorem 7.3. Roughly speaking, the proof has three parts: first, we show that this tree has a small height (Lemma 7.21); second, we show that each edge in this tree quickly exchanges the

rumour with a reasonably large probability (Lemma 7.23); and finally we show that almost all vertices in $G - G(m)$ have quick access to and from F (Lemma 7.24).

Let LOG denote the event 'each tree in F has height $O(\log n)$.' Proposition 3.37 states that a.a.s. an $(n+k)$-vertex random k-tree has the following property: let $u_h u_{h-1} \cdots u_0$ be any path such that u_i is born later than u_{i-1} for all i; then $h \le e \log n + O(1)$. The following lemma is an immediate corollary of this proposition.

Lemma 7.21. *Asymptotically almost surely* LOG *happens.*

We prove Lemma 7.20 conditional on the event LOG. In fact, we prove it for any $G(m)$ that satisfies LOG. Let G_1 be an arbitrary instance of $G(m)$ that satisfies LOG. So, G_1 and F are fixed in the following, and all randomness refers to rounds $m+1, \dots, n$. The following deterministic lemma will be used in the proof of Lemma 7.23.

Lemma 7.22. *Assume that $xy \in E(F)$ and x is born later than y. If the degree of x in G_1 is greater than $2k-2$, then $N(xy) \ge (k^2-k)/2$.*

Proof. Assume that x is joined to $u_1, \dots, u_k$ when it is born, and that $v_1, v_2, \dots, v_{k-1}, \dots$ are the neighbours of x that are born later than x, in the order of birth. Let Ψ denote the number of pairs (u_j, C), where $1 \le j \le k$, and C is a k-clique in G_1 containing the edge xu_j. Consider the round in which vertex x is born and is joined to $u_1, \dots, u_k$. For every $j \in \{1, \dots, k\}$, the vertex u_j is contained in $k-1$ new k-cliques, so in this round Ψ increases by $k(k-1)$. For each $i \in \{1, \dots, k-1\}$, consider the round in which vertex v_i is born. This vertex is joined to x and $k-1$ neighbours of x. At this round x has neighbour set $\{u_1, \dots, u_k, v_1, \dots, v_{i-1}\}$. Thus at least $k-i$ of the u_j's are joined to v_i in this round. Each vertex u_j that is joined to v_i in this round is contained in $k-2$ new k-cliques that contain x as well, so in this round Ψ increases by at least $(k-i)(k-2)$. Consequently, we have

$$\Psi \ge k(k-1) + \sum_{i=1}^{k-1} (k-i)(k-2) = k^2(k-1)/2 \,.$$

By the pigeonhole principle, there exists some $\ell \in \{1, \dots, k\}$ such that the edge xu_ℓ is contained in at least $(k^2-k)/2$ many k-cliques, and this completes the proof. ∎

A vertex of G is called *modern* if it is born later than the end of round m, and is called *traditional* otherwise. In other words, vertices of G_1 are traditional and vertices of $G - G_1$ are modern. We say edge $uv \in E(G)$ is *fast* if at least one of the following is true: $\deg(u) \le \tau$, or $\deg(v) \le \tau$, or u and v have a common neighbour w with $\deg(w) \le \tau$. For

an edge $uv \in E(F)$, let $p_S(uv)$ denote the probability that uv is not fast (slow), and let p_S denote the maximum of p_S over all edges of F. Recall that G_1 and F are fixed, so these probabilities are well defined.

Lemma 7.23. *We have $p_S = o(1/(f(n)\log n))$.*

Proof. Let $xy \in E(F)$ be arbitrary. By symmetry we may assume that x is born later than y. By Lemma 7.22, at least one of the following is true: vertex x has less than $2k-1$ neighbours in G_1, or $N(xy) \geq (k^2-k)/2$. So we may consider two cases.

- Case 1: vertex x has less than $2k-1$ neighbours in G_1. Assume that x has A neighbours in G_1 and lies in B many k-cliques in G_1. Then we have $A \leq 2k-2$ and

$$B \leq k + (k-1)(A-k) \leq k(k-1)\,.$$

  Let $X \stackrel{d}{=} \mathrm{Urn}\,(B, km+1-B, \mathcal{K}, n-m)$. Then by Proposition 7.12 the degree of x is distributed as $A + (X-B)/(k-1)$. By Proposition 7.11,

$$\begin{aligned}
\mathbb{E}\,[X^q] &\leq (1+o(1))\left(\frac{k(n-m)}{km+1}\right)^{\frac{q(k-1)}{k}} \prod_{i=0}^{q-1}(B+i(k-1)) \\
&\leq (1+o(1))\left(\frac{n}{m}\right)^{\frac{q(k-1)}{k}} (k-1)^q \prod_{i=0}^{q-1}(k+i) \\
&\leq (k-1)^q (k+q)! \left(\frac{n}{m}\right)^{\frac{q(k-1)}{k}}\,,
\end{aligned}$$

  where we have used $B \leq k(k-1)$ for the second inequality. Therefore,

$$\begin{aligned}
p_S(xy) &\leq \mathbb{P}\left[\deg(x) > 2k + q(n/m)^{\frac{k-1}{k}}\right] \\
&\leq \mathbb{P}\left[X \geq (k-1)q(n/m)^{\frac{k-1}{k}}\right] \\
&\leq \frac{\mathbb{E}\,[X^q]}{(k-1)^q q^q (n/m)^{\frac{q(k-1)}{k}}} \leq \frac{(k+q)!}{q^q} = o\left(\frac{1}{f(n)\log n}\right)\,,
\end{aligned}$$

  since $q \geq 4\log\log n$, k is fixed, and $f(n) = o(\log\log n)$.

- Case 2: $N(xy) \geq (k^2-k)/2$. In this case we bound from below the probability that there exists a modern vertex w that is adjacent to x and y and has degree at most τ. We first bound from above the probability that x and y have no modern common neighbours. For this to happen, none of the k-cliques containing x and y must be chosen in rounds $m+1,\dots,n$. This probability equals $\mathbb{P}\left[\mathrm{Urn}(N(xy), mk+1-N(xy), k, n-m) = N(xy)\right]$. Since $N(xy) \geq (k^2-k)/2$, by Proposition 7.8 we have

$$\mathbb{P}\left[\mathrm{Urn}(N(xy), mk+1-N(xy), k, n-m) = N(xy)\right] \leq \left(\frac{m+1}{n+1}\right)^{\frac{k-1}{2}},$$

which is $o\left(1/(f(n)\log n)\right)$ by the definition of m in (7.3).

Now, assume that x and y do have a modern common neighbour w. If there are multiple such vertices, choose the one that is born first. Since w appears later than round m, by Lemma 7.13,

$$p_S(xy) \leq \mathbb{P}\left[\deg(w) > k + q(n/m)^{(k-1)/k}\right] = O\left(q\sqrt{q}\exp(-q)\right) = o\left(\frac{1}{f(n)\log n}\right). \quad \blacksquare$$

Enumerate the k-cliques of G_1 as $C_1,\dots,C_{mk+1}$. Then choose $r_1 \in C_1,\dots,r_{mk+1} \in C_{mk+1}$ arbitrarily, and call them the *representative vertices*. Starting from G_1, when modern vertices are born in rounds $m+1,\dots,n$ until G is formed, every clique C_i 'grows' to a random k-tree with a random number of vertices, which is a subgraph of G. Enumerate these subgraphs as $H_1,\dots,H_{mk+1}$, and call them the *pieces*. More formally, $H_1,\dots,H_{mk+1}$ are induced subgraphs of G such that a vertex v is in $V(H_j)$ if and only if every path connecting v to a traditional vertex intersects $V(C_j)$. In particular, $V(C_j) \subseteq V(H_j)$ for all $j \in \{1,\dots,mk+1\}$. Note that the H_j's may intersect, as a traditional vertex may lie in more than one C_j, however every modern vertex lies in a unique piece. See Figure 7.3 for an illustration.

A traditional vertex is called *nice* if it is connected to some vertex in $G(0)$ via a path of fast edges. Since F has height $O(\log n)$ and each edge of F is fast with probability at least $1-p_S$, the probability that a given traditional vertex is *not* nice is $O(p_S \log n)$ by the union bound. A piece H_j is called *nice* if all its modern vertices have degrees at most τ, and the vertex r_j is nice. A modern vertex is called *nice* if it lies in a nice piece. A vertex/piece is called *bad* if it is not nice.

Lemma 7.24. *The expected number of bad vertices is $o(n)$.*

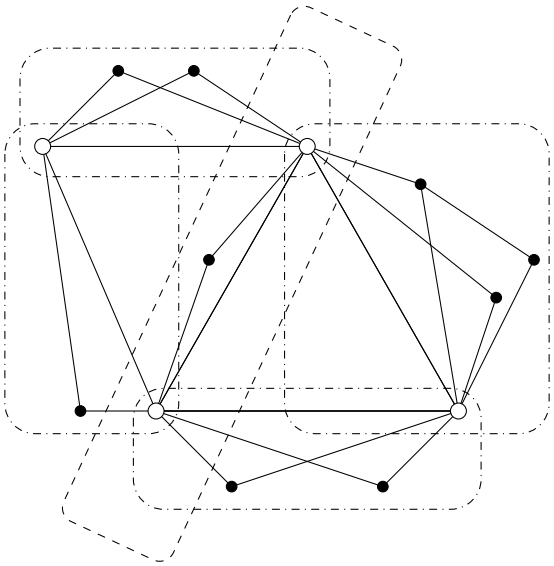


Figure 7.3: illustration of the pieces with $k = 2$, $m = 2$, and $n = 11$. Traditional vertices are white and modern vertices are black. There are five pieces, illustrated by the five rectangles.

Proof. The total number of traditional vertices is $k + m = o(n)$ so we may just ignore them in the calculations below. Let $\eta = nf(n)/m = o(\log^3 n)$. Say piece H_j is *sparse* if $|V(H_j)| \leq \eta + k$. We first bound the expected number of modern vertices in non-sparse pieces. Observe that the number of modern vertices in a given piece is distributed as $X = (\mathrm{Urn}(1, km, k, n - m) - 1)/k$. Using Proposition 7.8 we get $\mathbb{E}[X^2] \leq 2kn^2/m^2$. By the second moment method, for every $t > 0$ we have

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X^2]}{t^2} \leq \frac{2kn^2}{m^2t^2}\,.$$

The expected number of modern vertices in non-sparse pieces is thus at most

$$\begin{aligned}(km+1)\sum_{i=0}^{\infty}(2^{i+1}\eta)\mathbb{P}\left[2^i\eta < X \leq 2^{i+1}\eta\right] &\leq \sum_{i=0}^{\infty}(2^{i+1}\eta)(km+1)\frac{2kn^2}{m^2\eta^2 2^{2i}} \\ &\leq O\left(\frac{n^2}{m\eta}\right)\sum_{i=0}^{\infty}2^{-i} = O\left(\frac{n^2}{m\eta}\right)\,,\end{aligned}$$

which is $o(n)$ by the definition of η.

We now bound from above the expected number of modern vertices in sparse bad pieces. Indeed, we find an upper bound for the expected number of bad pieces, and multiply by η. A piece H_j can be bad in two ways:

(1) the representative vertex r_j is bad: the probability of this is $O(p_S \log n)$. Therefore, the expected number of pieces that are bad due to this reason is $O(mkp_S \log n)$, which is $o(n/\eta)$ by Lemma 7.23.

(2) there exists a modern vertex in H_j with degree greater than τ: the probability that a given modern vertex has degree greater than τ is $O\left(q\sqrt{q}\exp(-q)\right)$ by Lemma 7.13. So the average number of modern vertices with degree greater than τ is $O\left(nq\sqrt{q}\exp(-q)\right)$. Since every modern vertex lies in a unique piece, the expected number of pieces that are bad because of this reason is bounded by $O\left(nq\sqrt{q}\exp(-q)\right) = o(n/\log^3 n)$.

So the expected number of bad pieces is $o(n/\eta + n/\log^3 n)$, and the expected number of modern vertices in sparse bad pieces is $o(n + \eta n/\log^3 n) = o(n)$. ■

We now prove Lemma 7.20, which concludes the proof of Theorem 7.3.

Proof of Lemma 7.20. Let Σ denote the set of nice modern vertices. By Lemma 7.24 and using Markov's inequality, we have $|\Sigma| = n - o(n)$ a.a.s. Let $\{a_1, \ldots, a_k\}$ denote the vertex set of $G(0)$. Using an argument similar to the proof of Lemma 7.23, it can be proved that given $1 \leq i < j \leq k$, the probability that edge $a_i a_j$ is not fast is $o(1)$. Since the total number of such edges is a constant, a.a.s. all such edges are fast. Let u and v be nice modern vertices, and let r_u and r_v be the representative vertices of the pieces containing them, respectively. Since the piece containing u is nice, there exists a (u, r_u)-path whose vertices except possibly r_u all have degrees at most τ. The length of this path is at most $\operatorname{diam}(G)$. Since r_u is nice, for some $1 \leq i \leq n$ there exists an (r_u, a_i)-path in F consisting of fast edges. Appending these paths gives a (u, a_i)-path with length at most $\operatorname{diam}(G) + O(\log n)$ such that for every pair of consecutive vertices in this path, one of them has degree at most τ. Similarly, for some $1 \leq j \leq n$ there exists a (v, a_j)-path of length $O(\log n + \operatorname{diam}(G))$, such that one of every pair of consecutive vertices in this path has degree at most τ. Since the edge $a_i a_j$ is fast a.a.s., we can build a (u, v)-path of length $O(\log n + \operatorname{diam}(G))$ of the type required by Lemma 7.19, and this completes the proof. ■

7.4 Lower bound for random k-trees

In this section we prove Theorem 7.5.

Definition 7.25 (Barrier)**.** A pair $\{C_1, C_2\}$ of disjoint k-cliques in a connected graph is called a *D-barrier* if

(i) the set of edges between C_1 and C_2 is a cut-set, i.e. deleting them disconnects the graph, and

(ii) the degree of each vertex in $V(C_1) \cup V(C_2)$ is at least D.

Proposition 7.26. *If graph G has a D-barrier, then for any vertex v and any $a \in \mathbb{N}$ we have $\mathbb{P}[\mathrm{ST}_{\mathsf{s}}(G, v) \leq a] \leq 2k^2 a/D$.*

Proof. Let $\{C_1, C_2\}$ be a D-barrier in G. Assume that initially, the set of informed vertices is precisely C_1, and let T_{C_1,C_2} denote the first time that a vertex in C_2 learns the rumour. Then we have $T_{C_1,C_2} \overset{s}{\leq} \mathrm{ST}_{\mathsf{s}}(G, v)$ for any vertex v. Any given vertex in C_1 pushes the rumour to C_2 with probability at most k/D, and any given vertex in C_2 pulls the rumour with probability at most k/D, hence for any $a \in \mathbb{N}$ by the union bound we have

$$\mathbb{P}[\mathrm{ST}_{\mathsf{s}}(G, v) \leq a] \leq \mathbb{P}[T_{C_1,C_2} \leq a] \leq a \times 2k \times 2k/D\,,$$

as required. ■

For proving the next lemma we need a proposition which follows from known results.

Proposition 7.27. *There exists $\delta > 0$ such that for all large n we have*

$$\mathbb{P}\left[\mathrm{Urn}(1, 1, \mathcal{K}, n-k) < \delta n^{(k-1)/k}\right] < 1/(2k+1)\,.$$

Proof. Let $Y \overset{d}{=} \mathrm{Urn}(1, 1, \mathcal{K}, n-k)$. Janson [79, Theorem 1.3(v)] proved that as $n \to \infty$, $n^{-(k-1)/k} Y$ converges to a positive random variable W. Let $\delta > 0$ be small enough that $\mathbb{P}[W < \delta] < 1/(3k)$. Then for large enough n,

$$\mathbb{P}\left[Y < \delta n^{(k-1)/k}\right] = \mathbb{P}\left[n^{-(k-1)/k} Y < \delta\right] < 1/(2k+1)\,,$$

as required. ■

Lemma 7.28. *The graph $G(n)$ has an $\Omega(n^{1-1/k})$-barrier with probability $\Omega(n^{1/k-k})$.*

Proof. Let $u_1, \dots, u_k$ be the vertices of $G(0)$, and let $v_1, \dots, v_k$ be the vertices of $G(k) - G(0)$ in the order of appearance. We define two events. Event A is that for every $1 \le i \le k$, when v_i appears, it attaches to $v_1, v_2, \dots, v_{i-1}, u_i, u_{i+1}, \dots, u_k$; and for each $1 \le i, j \le k$, u_i and v_j have no common neighbour in $G(n) - G(k)$. Event B is that all vertices of $G(k)$ have degree $\Omega(n^{(k-1)/k})$ in $G(n)$. Note that if A and B both happen, then the pair $\{u_1 u_2 \dots u_k, v_1 v_2 \dots v_k\}$ is an $\Omega(n^{(k-1)/k})$-barrier in $G(n)$. To prove the lemma we will show that $\mathbb{P}[A] = \Omega(n^{1/k-k})$ and $\mathbb{P}[B|A] = \Omega(1)$.

For A to happen, first, the vertices $v_1, \dots, v_k$ must choose the specified k-cliques, which happens with constant probability. Moreover, the vertices appearing after round k must not choose any of the k^2-1 many k-cliques that contain both u_i's and v_j's. Since $1-y \ge e^{-y-y^2}$ for every $y \in [0, 1/4]$,

$$\begin{aligned}
\mathbb{P}[A] &= \Omega(\mathbb{P}\left[\mathrm{Urn}(k^2-1, 2, k, n-k) = k^2-1\right]) \\
&= \Omega\left(\prod_{i=0}^{n-k-1}\left(\frac{2+ik}{k^2+1+ik}\right)\right) \\
&\ge \Omega\left(\prod_{i=0}^{4k-1}\left(\frac{2+ik}{k^2+1+ik}\right)\prod_{i=4k}^{n-k-1}\left(1-\frac{k^2-1}{ik}\right)\right) \\
&\ge \Omega\left(\exp\left(-\sum_{i=4k}^{n-k-1}\left\{\frac{k^2-1}{ik}+\left(\frac{k^2-1}{ik}\right)^2\right\}\right)\right)
\end{aligned}$$

which is $\Omega(n^{1/k-k})$ since

$$\sum_{i=4k}^{n-k-1}\frac{k^2-1}{ik} \le (k-1/k)\log n + O(1) \text{ and } \sum_{i=4k}^{n-k-1}\left(\frac{k^2-1}{ik}\right)^2 = O(1)\,.$$

Conditional on A and using an argument similar to that in the proof of Proposition 7.12, the degree of each of $u_1, \dots, u_k, v_1, \dots, v_k$ in $G(n)$ is at least $k+(\mathrm{Urn}(1,1,\mathcal{K},n-k)-1)/(k-1)$. By Proposition 7.27 there exists $\delta > 0$ such that

$$\mathbb{P}\left[\mathrm{Urn}(1,1,\mathcal{K},n-k) < \delta n^{(k-1)/k}\right] < 1/(2k+1)\,.$$

By the union bound, the probability that all vertices $u_1, \dots, u_k, v_1, \dots, v_k$ have degrees at least $\delta n^{(k-1)/k}/(k-1)$ is at least $1/(2k+1)$, hence $\mathbb{P}[B|A] \ge 1/(2k+1) = \Omega(1)$. ■

Let $f(n) = o(\log\log n)$ be any function going to infinity with n, and let

$$m = \left\lceil f(n) n^{1-k/(k^2+k-1)} \right\rceil\,. \tag{7.5}$$

Note that the value of m is different from that in Section 7.3, although its role is somewhat similar. Consider the random k-tree process up to round m. Enumerate the k-cliques of $G(m)$ as $C_1, \dots, C_{mk+1}$. Starting from $G(m)$, when new vertices are born in rounds $m+1, \dots, n$ until $G = G(n)$ is formed, every clique C_i 'grows' to a random k-tree with a random number of vertices, which is a subgraph of G. Enumerate these subgraphs as $H_1, \dots, H_{mk+1}$, and call them the *pieces* (see Figure 7.3). We say a piece is *moderate* if its number of vertices is between $n/(mf(n))$ and $nf(n)/m$. Note that the number of vertices in a piece has expected value $\Theta(n/m)$. The following lemma is proved by showing this random variable does not deviate much from its expected value.

Lemma 7.29. *Asymptotically almost surely there are $o(m)$ non-moderate pieces.*

Proof. We prove the first piece, H_1, is moderate a.a.s. By symmetry, this would imply that the average number of non-moderate pieces is $o(m)$. By Markov's inequality, this gives that a.a.s. there are $o(m)$ non-moderate pieces. Let X denote the number of vertices of H_1. Note that X is distributed as $k + \mathrm{Urn}(1, km, k, n-m)$; so its expected value is $k + \frac{n-m}{1+km} = \Theta(n/m)$. By Markov's inequality, $\mathbb{P}[X > nf(n)/m] = o(1)$.

For bounding $\mathbb{P}[X < n/(mf(n))]$, we use an alternative way to generate the random variable $\mathrm{Urn}(1, km, k, n-m)$. Let $Z \stackrel{d}{=} \mathrm{Beta}(1/k, m)$. Proposition 2.9 states that

$$\mathrm{Urn}(1, km, k, n-m) \stackrel{d}{=} 1 + k\,\mathrm{Bin}(n-m, Z)\,,$$

which gives

$$X - k \stackrel{d}{=} 1 + k\,\mathrm{Bin}(n-m, Z)\,.$$

Note that

$$\begin{aligned}\mathbb{P}[Z < 3/(mf(n))] &= \frac{\Gamma(m+1/k)}{\Gamma(m)\Gamma(1/k)} \int_0^{3/(mf(n))} x^{1/k-1}(1-x)^{m-1}dx \\ &\le \frac{m^{1/k}}{\Gamma(1/k)} \int_0^{3/(mf(n))} x^{1/k-1}dx = \frac{3^{1/k}k}{\Gamma(1/k)f(n)^{1/k}} = o(1)\,,\end{aligned}$$

where we have used the fact $\Gamma(m+1/k) \le \Gamma(m)m^{1/k}$, see Proposition 7.10(a). On the other hand, the lower tail Chernoff bound (2.2) gives

$$\begin{aligned}\mathbb{P}[X < n/(mf(n)) | Z \ge 3/(mf(n))] &\le \mathbb{P}[\mathrm{Bin}(n-m, 3/(mf(n))) < n/(kmf(n))] \\ &\le \exp(-3(n-m)/(8mf(n))) = o(1)\,,\end{aligned}$$

thus $\mathbb{P}[X < n/(mf(n))] = o(1)$. ∎

Proof of Theorem 7.5. Consider an alternative way to generate $G(n)$ from $G(m)$: first, we determine how many vertices each piece has, and then we expose the structure of the pieces. Let Y denote the number of moderate pieces. By Lemma 7.29 we have $Y = \Omega(m)$ a.a.s. We prove the theorem conditional on $Y = y$, where $y = \Omega(m)$ is otherwise arbitrary. Note that after the sizes of the pieces are exposed, what happens inside each piece in rounds $m+1, \dots, n$ is mutually independent from other pieces. Let H be a moderate piece with n_1 vertices. By Lemma 7.28, the probability that H has an $\Omega(n_1^{1-1/k})$-barrier is $\Omega(n_1^{1/k-k})$. Since $n/(mf(n)) \le n_1 \le nf(n)/m$, the probability that H has a $\Omega((n/(mf(n))^{1-1/k})$-barrier is $\Omega((nf(n)/m)^{1/k-k})$. Since there are $y = \Omega(m)$ moderate pieces in total, the probability that no moderate piece has an $\Omega\left((n/(mf(n)))^{1-1/k}\right)$-barrier is at most

$$(1 - \Omega((nf(n)/m)^{1/k-k}))^y \le \exp(-\Omega(f(n))) = o(1)\,,$$

by the definition of m in (7.5). So, a.a.s. there exists an $\Omega\left(n^{(k-1)/(k^2+k-1)} f(n)^{-2}\right)$-barrier in $G(n)$. By Proposition 7.26, a.a.s. the spread time is $\Omega\left(n^{(k-1)/(k^2+k-1)} f(n)^{-3}\right)$, as required. ■

7.5 Upper bound for random k-Apollonian networks

In this section we analyze the protocol on random k-Apollonian networks and prove Theorem 7.6. The proof follows the line of that for random k-trees. First, we recall the definition of the random k-Apollonian process.

Definition 7.30 (Random k-Apollonian process)**.** Let k be a positive integer. Build a sequence $A(0)$, $A(1)$, ... of random graphs as follows. The graph $A(0)$ is just a clique on k vertices. This k-clique is marked as *active*. For each $1 \le t \le n$, $A(t)$ is obtained from $A(t-1)$ as follows: an active k-clique of $A(t-1)$ is chosen uniformly at random, a new vertex is born and is joined to all vertices of the chosen k-clique. The chosen k-clique is marked as *non-active*, and all the new k-cliques are marked as active in $A(t)$. The graph $A(n)$ is called a *random k-Apollonian network (k-RAN)* on $n+k$ vertices.

We first prove a counterpart of Proposition 7.12 and Lemma 7.13 for k-RANs.

Proposition 7.31. *Suppose that in $A(j)$ vertex x has $D > 0$ neighbours, and is contained in B active k-cliques. Conditional on this, the degree of x in $A(n+j)$ is distributed as*

$$D + \left(\mathrm{Urn}\left(B, (k-1)j + 1 - B, \begin{bmatrix} k-2 & 1 \\ 0 & k-1 \end{bmatrix}, n\right) - B\right) \Big/ (k-2)\,.$$

Proof. We claim that the number of active k-cliques containing x in $G(n+j)$ is distributed as $\mathrm{Urn}\left(B, (k-1)j+1-B, \begin{bmatrix} k-2 & 1 \\ 0 & k-1 \end{bmatrix}, n\right)$. At the end of round j, there are B active k-cliques containing x, and $(k-1)j+1-B$ many k-cliques not containing x. In each subsequent round $j+1, \ldots, j+n$, a random active k-clique is chosen and k new active k-cliques are created, and one active k-clique is deactivated. If the chosen active k-clique contains x, then $k-2$ new active k-cliques containing x are created, and 1 new active k-clique not containing x is created. Otherwise, just $k-1$ new active k-cliques not containing x are created, and the claim follows.

Hence the number of active k-cliques that are created in rounds $j+1, \ldots, j+n$ and contain x is distributed as $\mathrm{Urn}\left(B, (k-1)j+1-B, \begin{bmatrix} k-2 & 1 \\ 0 & k-1 \end{bmatrix}, n\right) - B$, and the conclusion follows by noting that every new neighbour of x creates $k-2$ new active k-cliques containing x. ■

Lemma 7.32. *Let $1 \le j \le n$ and let q be a positive integer. Let x denote the vertex born in round j. Conditional on any $A(j)$, the probability that x has degree greater than $k + q(n/j)^{(k-2)/(k-1)}$ in $A(n)$ is $O\left(q\sqrt{q}\exp(-q)\right)$.*

Proof. Let $X \stackrel{d}{=} \mathrm{Urn}\left(k, (k-1)(j-1), \begin{bmatrix} k-2 & 1 \\ 0 & k-1 \end{bmatrix}, n-j\right)$. Proposition 7.31 implies that the degree of x in $A(n)$ is distributed as $k + (X-k)/(k-2)$. By Proposition 7.11,

$$\mathbb{E}\left[X^q\right] \le (1+o(1)) \left(\frac{(k-1)(n-j)}{(k-1)j+1}\right)^{\frac{q(k-2)}{k-1}} \prod_{i=0}^{q-1} (k+i(k-2)) \le \left(\frac{n}{j}\right)^{\frac{q(k-2)}{k-1}} (k-2)^q (q+1)!\,.$$

Thus,

$$\begin{aligned} \mathbb{P}\left[\deg(x) > k + q(n/j)^{(k-2)/(k-1)}\right] &= \mathbb{P}\left[X - k > (k-2)q(n/j)^{(k-2)/(k-1)}\right] \\ &\le \frac{\mathbb{E}\left[X^q\right]}{\left((k-2)q(n/j)^{(k-2)/(k-1)}\right)^q} \\ &\le (q+1)!q^{-q} = O\left(q\sqrt{q}\exp(-q)\right)\,. \end{aligned}$$

■

Fix $k > 2$ and let $f(n) = o(\log\log n)$ be an arbitrary function going to infinity with n, and let

$$m = \left\lceil \frac{n}{(\log n)^{2/(k-1)} f(n)^{(2k-2)/(k^2-2k)}} \right\rceil\,. \tag{7.6}$$

Finally, let $q = \lceil 4 \log\log n \rceil$ and let

$$\tau = 2k + q(n/m)^{(k-2)/(k-1)} \,. \tag{7.7}$$

By Theorem 3.36, a.a.s. a k-RAN on $n+k$ vertices has diameter $O(\log n)$. Theorem 7.6 thus follows from Lemma 7.19 and the following structural result, which we prove in the rest of this section.

Lemma 7.33. *Let A be an $(n+k)$-vertex k-RAN. A.a.s. there exists $\Sigma \subseteq V(A)$ satisfying the conditions of Lemma 7.19 with τ defined in (7.7) and $\chi = O(\log n + \mathrm{diam}(A))$.*

The proof of Lemma 7.33 is along the lines of that of Lemma 7.20. For the rest of this section, $A = A(n)$ is an $(n+k)$-vertex k-RAN, and whenever we talk about the degree of a vertex, we mean its degree in A, unless specified otherwise. Consider the graph $A(m)$, which has $k+m$ vertices and $m(k-1)+1$ active k-cliques. For any edge e of $A(m)$, let $N^*(e)$ denote the number of active k-cliques of $A(m)$ containing e. Note that, since $k > 2$, for each edge e, the number of active k-cliques containing e does not decrease as the k-RAN evolves. We define a spanning forest F of $A(m)$ as follows: at round 0, F has k isolated vertices, i.e. the vertices of $A(0)$; then for every $1 \le t \le m$, if the vertex x born in round t is joined to the k-clique C, then in F, x is joined to a vertex $u \in V(C)$ such that

$$N^*(xu) = \max_{v \in V(C)} N^*(xv) \,.$$

Note that F consists of k trees and the k vertices of $A(0)$ lie in distinct trees. Let LOG denote the event 'each tree in F has height $O(\log n)$.'

Proposition 3.38 states that a.a.s. an $(n+k)$-vertex random k-Apollonian network has the following property: let $u_h u_{h-1} \cdots u_0$ be any path such that u_i is born later than u_{i-1} for all i; then $h \le ek \log n/(k-1) + O(1)$. The following lemma is an immediate corollary of this proposition.

Lemma 7.34. *Asymptotically almost surely* LOG *happens.*

We prove Lemma 7.33 conditional on the event LOG. In fact, we prove it for any $A(m)$ that satisfies LOG. Let A_1 be an arbitrary instance of $A(m)$ that satisfies LOG. So, A_1 and F are fixed in the following, and all randomness refers to rounds $m+1, \ldots, n$. The following deterministic lemma will be used in the proof of Lemma 7.36.

Lemma 7.35. *Assume that $xy \in E(F)$ and x is born later than y. If the degree of x in A_1 is at least $2k-1$, then $N^*(xy) \ge (k-1)^2/2$.*

Proof. Assume that x is joined to $u_1, \dots, u_k$ when it is born, and that $v_1, v_2, \dots, v_{k-1}, \dots$ are the neighbours of x that are born later than x, in the order of birth. Let Ψ denote the number of pairs (u_j, C), where C is an active k-clique in A_1 with $xu_j \in E(C)$. Consider the round in which vertex x is born and is joined to $u_1, \dots, u_k$. For every $j \in \{1, \dots, k\}$, the edge xu_j is contained in $k-1$ new active k-cliques, so in this round Ψ increases by $k(k-1)$. For each $i \in \{1, \dots, k-1\}$, consider the round in which vertex v_i is born. At least $k-i$ of the u_j's are joined to v_i in this round. Each vertex u_j that is joined to v_i in this round is contained in $k-2$ new k-cliques that contain x, and one k-clique containing u_j is deactivated. Hence in this round Ψ increases by at least $(k-i)(k-3)$. Consequently, right after v_{k-1} is born, we have

$$\Psi \geq k(k-1) + \sum_{i=1}^{k-1} (k-i)(k-3) = (k-1)^2 k/2 \,.$$

By the pigeonhole principle, there exists some $\ell \in \{1, \dots, k\}$ such that the edge xu_ℓ is contained in at least $(k-1)^2/2$ active k-cliques, and this completes the proof, as the number of active k-cliques containing xu_ℓ will not decrease later. ■

A vertex of A is called *modern* if it is born later than the end of round m, and is called *traditional* otherwise. In other words, vertices of A_1 are traditional and vertices of $A - A_1$ are modern. We say edge $uv \in E(A)$ is *fast* if at least one of the following is true: $\deg(u) \leq \tau$, or $\deg(v) \leq \tau$, or u and v have a common neighbour w with $\deg(w) \leq \tau$. For an edge $uv \in E(F)$, let $p_S(uv)$ denote the probability that uv is not fast (slow), and let p_S denote the maximum of p_S over all edges of F. These probabilities are well defined since we have fixed A_1 and F.

Lemma 7.36. *We have* $p_S = o(1/(f(n) \log n))$.

Proof. The proof is similar to that of Lemma 7.23. Let $xy \in E(F)$ be arbitrary. By symmetry we may assume that x is born later than y. By Lemma 7.35, at least one of the following is true: vertex x has less than $2k-1$ neighbours in A_1, or $N^*(xy) \geq (k-1)^2/2$. So we may consider two cases.

- Case 1: vertex x has less than $2k-1$ neighbours in A_1. Suppose that x has D neighbours in A_1 and lies in B active k-cliques in A_1. Then $D \leq 2k-2$ and

$$B \leq k + (k-2)(D-k) \leq k(k-2) \,.$$

Let

$$X \stackrel{d}{=} \mathrm{Urn}\left(B, (k-1)m+1-B, \begin{bmatrix} k-2 & 1 \\ 0 & k-1 \end{bmatrix}, n-m\right) .$$

Then by Proposition 7.31, the degree of x is distributed as $D + (X-B)/(k-2)$. By Proposition 7.11,

$$\begin{aligned}\mathbb{E}\left[X^q\right] &\le (1+o(1))\left(\frac{(k-1)(n-m)}{(k-1)m+1}\right)^{\frac{q(k-2)}{k-1}} \prod_{i=0}^{q-1}(B+i(k-2)) \\ &= O\left(\left(\frac{n}{m}\right)^{\frac{q(k-2)}{k-1}}(k-2)^q(k+q)!\right) ,\end{aligned}$$

where we have used $B \le k(k-2)$. Therefore,

$$\begin{aligned}p_S(xy) &\le \mathbb{P}\left[\deg(x) > 2k + q(n/m)^{\frac{k-2}{k-1}}\right] \\ &\le \mathbb{P}\left[X \ge (k-2)q(n/m)^{\frac{k-2}{k-1}}\right] \\ &\le \frac{\mathbb{E}\left[X^q\right]}{(k-2)^q q^q (n/m)^{\frac{q(k-2)}{k-1}}} = O\left(\frac{(k+q)!}{q^q}\right) = o\left(\frac{1}{f(n)\log n}\right) ,\end{aligned}$$

since $q \ge 4\log\log n$, $f(n) = o(\log\log n)$, and k is fixed.

- Case 2: $N^*(xy) \ge (k-1)^2/2$. In this case we bound from below the probability that there exists a modern vertex w that is adjacent to x and y and has degree at most τ. We first bound from above the probability that x and y have no modern common neighbours. For this to happen, none of the k-cliques containing x and y must be chosen in rounds $m+1, \dots, n$. This probability equals

$$p := \mathbb{P}\left[\mathrm{Urn}(N^*(xy), m(k-1)+1-N^*(xy), k-1, n-m) = N^*(xy)\right] .$$

Since $N^*(xy) \ge (k-1)^2/2$, by Proposition 7.8 we have

$$p \le \left(\frac{m+1}{n}\right)^{(k-1)/2} = o\left(\frac{1}{f(n)\log n}\right) ,$$

by the definition of m in (7.6).

Now, assume that x and y do have a modern common neighbour w. If there are multiple such vertices, choose the one that is born first. Since w appears later than round m, by Lemma 7.32,

$$p_S(xy) \leq \mathbb{P}\left[\deg(w) > k + q(n/m)^{(k-2)/(k-1)}\right] = O\left(q\sqrt{q}\exp(-q)\right) = o\left(\frac{1}{f(n)\log n}\right). \qquad \blacksquare$$

Enumerate the k-cliques of A_1 as C_1, C_2, ..., and $C_{m(k-1)+1}$. Then choose $r_1 \in C_1, \ldots, r_{m(k-1)+1} \in C_{m(k-1)+1}$ arbitrarily, and call them the *representative vertices*. Starting from A_1, when modern vertices are born in rounds $m+1, \ldots, n$ until A is formed, every clique C_i 'grows' to a k-RAN with a random number of vertices, which is a subgraph of A. Enumerate these subgraphs as $H_1, \ldots, H_{m(k-1)+1}$, and call them the *pieces*. More formally, $H_1, \ldots, H_{m(k-1)+1}$ are induced subgraphs of A such that a vertex v is in $V(H_j)$ if and only if every path connecting v to a traditional vertex intersects $V(C_j)$.

A traditional vertex is called *nice* if it is connected to some vertex in $A(0)$ via a path of fast edges. Since F has height $O(\log n)$ and each edge of F is fast with probability at least $1 - p_S$, the probability that a given traditional vertex is *not* nice is $O(p_S \log n)$ by the union bound. A piece H_j is called *nice* if all its modern vertices have degrees at most τ, and the vertex r_j is nice. A modern vertex is called *nice* if it lies in a nice piece. A vertex/piece is called *bad* if it is not nice.

Lemma 7.37. *The expected number of bad vertices is $o(n)$.*

Proof. The proof is very similar to that of Lemma 7.24, except we use Lemmas 7.32 and 7.36 instead of Lemmas 7.13 and 7.23, respectively. $\blacksquare$

The proof of Lemma 7.33 is exactly the same as that of Lemma 7.20, except we use Lemmas 7.36 and 7.37 instead of Lemmas 7.23 and 7.24, respectively. This concludes the proof of Theorem 7.6.

Chapter 8

Conclusions and open problems

In this chapter we first explain the major contributions of the thesis, and then mention some research directions and open problems.

In Chapter 3 we presented a versatile technique for proving logarithmic upper bounds for diameters of certain evolving random graph models. This technique gives unified simple proofs for known results, provides lots of new ones, and will help in proving many of the forthcoming network models are small-world. Perhaps, for any given model, one can come up with an ad hoc argument that the diameter is $O(\log n)$, but it is interesting that a unified technique works for such a wide variety of models, and our first major contribution is introducing such a technique.

In Chapters 4 and 5 we estimated the diameter of two random graph models. Although the two models are quite different, surprisingly the same engine is used for proving these results, namely the powerful technique of Broutin and Devroye [27] for analyzing weighted heights of random trees, which we have adapted and applied to the two random graph models. Our second major contribution is demonstrating the flexibility of this technique via providing two significant applications.

Our third major contribution appears in Chapters 6 and 7, where we gave analytical proofs for two experimentally verified statements: firstly, the asynchronous push&pull protocol is typically faster than its synchronous variant (see, e.g., [50, Figures 4 and 5]), and secondly, it takes considerably more time to inform the last 1 percent of the vertices in a social network than the first 99 percent (see, e.g., [109, Figure 1]). We hope that our work on the asynchronous push&pull protocol attracts attention to this fascinating model.

Let us now turn to open problems.

Open problem 8.1 (Page 33)**.** Develop a mathematical theory for characterizing those evolving random graphs which have logarithmic diameters.

It would be interesting to have several meta-theorems so that whenever a new model is proposed, it can be quickly determined whether the diameter is logarithmic. This is a rather ambitious project, nevertheless very beneficial for network science. The proof technique we introduced in Chapter 3 could be a fundamental step in building this theory. One can try to further develop this technique to cover other network models, e.g. growth-deletion models [34, 42], accelerated network growth models [52], and spatial models [80].

In this thesis we only considered growth models: vertices and edges are never deleted from the graph. In most real-world networks, however, deletions exist. Growth-deletion models are harder to analyze than growth-only models, and we are aware of only two papers [34, 42] that consider vertex/edge deletions. A specific obstacle in bounding the diameters of growth-deletion models is that, the evolving graph may get disconnected and then get connected again, so the diameter could become undefined during the generation process.

In accelerated network growth models, the number of edges in the graph is superlinear. As time passes, the number of edges added in each step increases, see [52]. The techniques of Chapter 3 are probably applicable and give some upper bounds for the diameter.

In spatial models, vertices are embedded in a metric space, and link formation depends on the relative position of vertices in the space. Consider a social network for example. Each person has a vector of attributes (age, location, occupation, etc.) and two individuals that are 'close' in the underlying metric space are more likely to be friends. Spatial models have gained a lot of interest in recent years, see [80] for a survey. It is not straightforward to prove logarithmic upper bounds for diameters of spatial models, but proving general results would be very interesting.

Open problem 8.2. Consider the following growing random tree model: start from a single node, and suppose in every step a new node is born and is joined to a random node of the existing tree, sampled according to some probability distribution. If this distribution is the uniform distribution, then we obtain a random recursive tree, and it was proved by Pittel [111], and also follows from Lemma 3.3, that a.a.s. the height is $\Theta(\log n)$

For which distributions is the diameter logarithmic? It is reasonable to conjecture that, if the distribution is 'close enough' to uniform (e.g. according to the Kullback-Leibler divergence or the Shannon entropy) then the height would still be $\Theta(\log n)$. It would be nice to prove such a theorem. If the distribution is 'far' from uniform, then not much can be said: the height could be 1 if each new node is attached to the root, or it could be $n-1$ if each new node is attached to the farthest vertex from the root.

Open problem 8.3 (Page 33)**.** Prove nontrivial lower bounds for the diameter of any of the models studied in Chapter 3. Several logarithmic upper bounds were proved in Chapter 3, but it seems completely new ideas are required for proving lower bounds, and the author is not aware of any general approach.

Open problem 8.4 (Page 60)**.** What is the typical order of magnitude of $\mathcal{L}_m$, the length of a longest path in a RAN? Is this variable concentrated around its mean? In Theorem 4.2 we showed $\mathcal{L}_m > m^{\log 2/\log 3}$ and $\mathbb{E}[\mathcal{L}_m] = \Omega(m^{0.88})$ and in Theorem 4.4 we showed that a.a.s. we have $\mathcal{L}_m < m^{0.99999996}$. What is the correct answer? This question, which seems to be difficult, is interesting from the mathematical point of view, but the author is not aware whether it has any applications.

Open problem 8.5 (Page 92)**.** In Theorem 5.4 we showed that the diameter of the random-surfer Webgraph is a.a.s. $\Theta(\log n)$ when each vertex has out-degree $d = 1$. What is the order of magnitude of the diameter when $d > 1$?

The random-surfer Webgraph has similarities with the preferential attachment model (see Section 5.5), for which the diameter is of order $\Theta(\log n/\log\log n)$ when $d > 1$ (by [20, Theorem 1]). However, if we change the preferential attachment rule slightly so that each vertex v is chosen with probability proportional to $\deg(v) + \delta$ for some $\delta > 0$, the diameter becomes $\Theta(\log n)$ (by [51, Theorems 1.3 and 1.4]), but if $\delta \in (-d, 0)$ the diameter is $\Theta(\log\log n)$ (by [51, Theorems 1.6 and 1.7]). Given these, it is not easy to guess what the answer should be for the random-surfer Webgraph model. The answer might indeed depend on the value of p, e.g. in [31, Theorem 1.3] it is shown that a phase transition occurs in the root's degree when p passes $1/2$.

Open problem 8.6 (Page 94)**.** What are the asymptotic values of the height and diameter of the random-surfer tree when $p < p_0$? There is a gap between our lower and upper bounds in Theorems 5.3 and 5.4 (see Figure 5.2). It seems one cannot estimate the diameter in this regime by adapting the technique of Broutin and Devroye.

Open problem 8.7 (Page 128)**.** Find the best possible constant factors in Theorems 6.3 and 6.4. These results provide the extremal spread times for the synchronous and asynchronous push&pull protocols.

We conjecture that the path graph has the maximum average spread time, and the double star has the maximum guaranteed spread time. For the asynchronous variant, we conjecture that the complete graph is the fastest graph. We have proved these conjectures up to constant factors, and it would be interesting to either disprove them, or prove them up to $1 + o(1)$ factors; but the author has no idea how to do so.

Open problem 8.8 (Page 129)**.** In Theorem 6.6 we proved that any n-vertex graph G has $\mathrm{gst}_{\mathsf{a}}(G) = O\left(\mathrm{gst}_{\mathsf{s}}(G) \log n\right)$. For all graphs we examined a stronger result holds, which suggests that indeed for any n-vertex graph G we have $\mathrm{gst}_{\mathsf{a}}(G) \leq \mathrm{gst}_{\mathsf{s}}(G) + O(\log n)$. Does this bound hold?

It might be possible to prove $\mathrm{gst}_{\mathsf{a}}(G) \leq \max\{2\,\mathrm{gst}_{\mathsf{s}}(G), O(\log n)\}$ by considering the path through which the rumour passes (for each vertex), and using concentration of sums of independent exponentials (along the lines of [60, Theorem 6]). The author has tried this approach but failed to produce a valid coupling. Still, there might exist a sneaky way to make it work.

Open problem 8.9 (Page 129)**.** In Corollary 6.9 we proved that for any n-vertex G,

$$\frac{\mathrm{gst}_{\mathsf{s}}(G)}{\mathrm{gst}_{\mathsf{a}}(G)} = O\big(n^{2/3}\big)\,,$$

and that there exist infinitely many graphs for which this ratio is $\Omega\left(n^{1/3}(\log n)^{-4/3}\right)$. What is the maximum possible value of the ratio $\mathrm{gst}_{\mathsf{s}}(G)/\,\mathrm{gst}_{\mathsf{a}}(G)$ for an n-vertex graph G? There is a big gap in the exponent here, and it would be great to find the right answer.

It seems improving the upper bound might be easier than improving the lower bound: we have not used much 'graph theory' in proving the upper bound, in particular the proof is valid for graphs with multiple edges as well. On the other hand, the author cannot think of any graph better than the necklace graph, in terms of having a large gap between asynchronous and synchronous spread times.

Open problem 8.10 (Page 129)**.** The parameters $\mathrm{wast}_{\mathsf{s}}(G)$ and $\mathrm{wast}_{\mathsf{a}}(G)$ can be approximated easily using the Monte Carlo method: simulate the rumour spreading protocols several times, measuring the spread time of each simulation, and output the average. An open problem is to design a *deterministic* approximation algorithm for any one of $\mathrm{wast}_{\mathsf{a}}(G)$, $\mathrm{gst}_{\mathsf{a}}(G)$, $\mathrm{wast}_{\mathsf{s}}(G)$ or $\mathrm{gst}_{\mathsf{s}}(G)$.

Computers cannot produce real randomness, hence it is important to know whether a randomized algorithm can be turned into a deterministic one. This is called *derandomization*. Often devising a deterministic algorithm requires exploiting the problem's structure, hence a deterministic algorithm typically provides a better understanding of the problem in hand.

Open problem 8.11. In some practical applications, performing both push and pull is expensive, so only push operations or pull operations can be performed in any given step. It can be observed that push operations are more effective in the beginning of spread, and

pull operations become more effective later on. Consider the following protocol: up until some time, only push operations are performed, and after that time, only pull operations are performed. When is the best time to make the transition in this protocol?

Open problem 8.12 (Page 154)**.** In Theorem 7.5 we showed that a.a.s. the spread time of the synchronous push&pull protocol on a random k-tree is polynomially large. Our technique does not extend to random k-Apollonian networks, although we believe that a.a.s. we need a polynomial number of rounds to inform all vertices in a random k-Apollonian network as well. For establishing this result, perhaps one needs to define a new notion of 'barrier,' which would be useful in proving lower bounds for spread times.

References

[1] H. Acan, A. Collevecchio, A. Mehrabian, and N. Wormald. On the push&pull protocol for rumour spreading. *arXiv*, 1411.0948 [cs.DC], 2014. submitted, 19 pages.

[2] W. Aiello, F. Chung, and L. Lu. Random evolution in massive graphs. In *42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 510–519. IEEE Computer Soc., Los Alamitos, CA, 2001.

[3] M. Albenque and J.-F. Marckert. Some families of increasing planar maps. *Electron. J. Probab.*, 13:no. 56, 1624–1671, 2008.

[4] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, 74(1):47–97, 2002.

[5] H. Amini, M. Draief, and M. Lelarge. Flooding in weighted sparse random graphs. *SIAM J. Discrete Math.*, 27(1):1–26, 2013.

[6] J. S. Andrade, H. J. Herrmann, R. F. S. Andrade, and L. R. da Silva. Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs. *Phys. Rev. Lett.*, 94:018702, Jan 2005.

[7] K. B. Athreya. On a characteristic property of Polya's urn. *Studia Sci. Math. Hungar.*, 4:31–35, 1969.

[8] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 33–42, New York, NY, USA, 2012. ACM.

[9] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[10] P. Berenbrink, R. Elsässer, and T. Friedetzky. Efficient randomised broadcasting in random regular networks with applications in peer-to-peer systems. In *Proc. 27th Symp. Principles of Distributed Computing (PODC)*, pages 155–164, 2008.

[11] N. Berger, B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Degree distribution of the FKP network model. *Theoret. Comput. Sci.*, 379(3):306–316, 2007.

[12] N. Berger, C. Borgs, J.T. Chayes, and A. Saberi. On the spread of viruses on the Internet. In *Proc. 16th Symp. Discrete Algorithms (SODA)*, pages 301–310, 2005.

[13] S. Bhamidi. Universal techniques to analyze preferential attachment trees: global and local analysis. preprint, available at http://www.unc.edu/~bhamidi/, 2007.

[14] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436–442, 2001.

[15] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[16] A. Blum, T.-H. H. Chan, and M. R. Rwebangira. A random-surfer web-graph model. In *Proceedings of the 8th Workshop on Algorithm Engineering and Experiments and the 3rd Workshop on Analytic Algorithmics and Combinatorics*, ALENEX/ANALCO '06, pages 238–246, 2006.

[17] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 132–139, New York, 2003. ACM.

[18] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[19] B. Bollobás and Y. Kohayakawa. On Richardson's model on the hypercube. In *Combinatorics, geometry and probability (Cambridge, 1993)*, pages 129–137. Cambridge Univ. Press, Cambridge, 1997.

[20] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, January 2004.

[21] A. Bonato. A survey of models of the web graph. In *Combinatorial and algorithmic aspects of networking*, volume 3405 of *Lecture Notes in Comput. Sci.*, pages 159–172. Springer, Berlin, 2005.

[22] A. Bonato. *A course on the web graph*, volume 89 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[23] A. Bonato and F. Chung. Complex networks. In J. L. Gross, J. Yellen, and P. Zhang, editors, *Handbook of Graph Theory*, chapter 12.1, pages 1456–1476. Chapman & Hall/CRC, 2nd edition, 2013.

[24] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[25] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(17):107 – 117, 1998. Proceedings of the Seventh International World Wide Web Conference.

[26] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Comput. Netw.*, 33(1-6):309–320, June 2000.

[27] N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46(3-4):271–297, 2006.

[28] T. Bu and D. Towsley. On distinguishing between internet power law topology generators. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 638–647 vol.2, 2002.

[29] D. Chakrabarti and C. Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Computing Surveys*, 38(1), 2006. Article 2.

[30] D. Chakrabarti and C. Faloutsos. *Graph Mining: Laws, Tools, and Case Studies*. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, 2012.

[31] P. Chebolu and P. Melsted. Pagerank and the random surfer model. In *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 1010–1018, Philadelphia, PA, USA, 2008.

[32] G. Chen and X. Yu. Long cycles in 3-connected graphs. *J. Combin. Theory Ser. B*, 86:80–99, 2002.

[33] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2003.

[34] F. Chung and L. Lu. Coupling online and offline analyses for random power law graphs. *Internet Math.*, 1(4):409–461, 2004.

[35] F. Chung and L. Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.

[36] F. Chung, L. Lu, and V. Vu. The spectra of random graphs with given expected degrees. *Internet Mathematics*, 1(3):257–275, 2003.

[37] A. Collevecchio, A. Mehrabian, and N. Wormald. Longest paths in random Apollonian networks and largest r-ary subtrees of random d-ary recursive trees. *arXiv*, 1404.2425 [math.PR], 2014. submitted, 14 pages.

[38] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures Algorithms*, 22(3):311–335, 2003.

[39] C. Cooper and A. Frieze. Long paths in random Apollonian networks. *arXiv*, 1403.1472v1 [math.PR], 2014.

[40] C. Cooper, A. Frieze, and P. Prałat. Some typical properties of the spatial preferred attachment model. *Internet Mathematics*, 10(1-2):116–136, 2014.

[41] C. Cooper, A. Frieze, and R. Uehara. The height of random k-trees and related branching processes. *Random Structures Algorithms*, 45(4):675–702, 2014.

[42] C. Cooper, A. Frieze, and J. Vera. Random deletion in a scale-free random graph process. *Internet Math.*, 1(4):463–483, 2004.

[43] C. Cooper and R. Uehara. Scale free properties of random k-trees. *Mathematics in Computer Science*, 3(4):489–496, 2010.

[44] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, 47(1):41–72, 2009.

[45] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *Proc. 6th Symp. Principles of Distributed Computing (PODC)*, pages 1–12, 1987.

[46] F. den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[47] L. Devroye, O. Fawzi, and N. Fraiman. Depth properties of scaled attachment random recursive trees. *Random Structures & Algorithms*, 41(1):66–98, 2012.

[48] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *Proc. 43th Symp. Theory of Computing (STOC)*, pages 21–30, 2011.

[49] B. Doerr, M. Fouz, and T. Friedrich. Asynchronous rumor spreading in preferential attachment graphs. In *Proc. 13th Scandinavian Workshop Algorithm Theory (SWAT)*, pages 307–315, 2012.

[50] B. Doerr, M. Fouz, and T. Friedrich. Experimental analysis of rumor spreading in social networks. In *Design and analysis of algorithms*, volume 7659 of *Lecture Notes in Comput. Sci.*, pages 159–173. Springer, Heidelberg, 2012.

[51] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment models. *Journal of Statistical Physics*, 139(1):72–107, 2010.

[52] S. N. Dorogovtsev and J. F. F. Mendes. Accelerated growth of networks. In *Handbook of graphs and networks*, pages 318–341. Wiley-VCH, Weinheim, 2003.

[53] J. P. K. Doye and C. P. Massen. Self-similar disk packings as model spatial scale-free networks. *Phys. Rev. E (3)*, 71(1):016128, 12, 2005.

[54] E. Drinea, A. Frieze, and M. Mitzenmacher. Balls and bins models with feedback. In *Proceedings of the 13th annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '02, pages 308–315, Philadelphia, PA, USA, 2002.

[55] R. Durrett. Stochastic growth models: recent results and open problems. In *Mathematical approaches to problems in resource management and epidemiology (Ithaca, NY, 1987)*, volume 81 of *Lecture Notes in Biomath.*, pages 308–312. Springer, Berlin, 1989.

[56] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[57] E. Ebrahimzadeh, L. Farczadi, P. Gao, A. Mehrabian, C. M. Sato, N. Wormald, and J. Zung. On longest paths and diameter in random Apollonian networks. *Random Structures Algorithms*, 45(4):703–725, 2014.

[58] F. Eggenberger and G. Pólya. Über die statistik verketteter vorgänge. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.

[59] R. Elsässer and T. Sauerwald. Cover time and broadcast time. In *STACS 2009: 26th International Symposium on Theoretical Aspects of Computer Science*, volume 3 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 373–384. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2009.

[60] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoret. Comput. Sci.*, 410(36):3414–3427, 2009.

[61] G. Ergün and G. J. Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1–2):261–272, 2002.

[62] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Struct. Algorithms*, 1(4):447–460, 1990.

[63] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[64] J. A. Fill and R. Pemantle. Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *Ann. Appl. Probab.*, 3(2):593–629, 1993.

[65] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Discrete Math. Theor. Comput. Sci. Proc., AG, pages 59–118. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006.

[66] A. Flaxman, A. Frieze, and J. Vera. A geometric preferential attachment model of networks. II. *Internet Math.*, 4(1):87–111, 2007.

[67] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable broadcasting in random networks and the effect of density. In *Proc. 29th IEEE Conf. Computer Communications (INFOCOM)*, pages 2552–2560, 2010.

[68] N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and expanders. In *Proc. 14th Intl. Workshop on Randomization and Comput. (RANDOM)*, pages 560–573, 2010.

[69] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *Proc. 23th Symp. Discrete Algorithms (SODA)*, pages 1642–1660, 2012.

[70] A. Frieze and C. E. Tsourakakis. Some properties of random Apollonian networks. *Internet Mathematics*, 10(1-2):162–187, 2014.

[71] Y. Gao. The degree distribution of random k-trees. *Theor. Comput. Sci.*, 410(8-10):688–695, 2009.

[72] Y. Gao. Treewidth of Erdős-Rényi random graphs, random intersection graphs, and scale-free random graphs. *Discrete Applied Mathematics*, 160(4-5):566–578, 2012.

[73] G. Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *28th International Symposium on Theoretical Aspects of Computer Science (STACS 2011)*, volume 9, pages 57–68, 2011.

[74] G. Giakkoupis. Tight bounds for rumor spreading with vertex expansion. In *Proc. 25th Symp. Discrete Algorithms (SODA)*, pages 801–815, 2014.

[75] M. Harchol-Balter, F. Thomson Leighton, and D. Lewin. Resource discovery in distributed networks. In *Proc. 18th Symp. Principles of Distributed Computing (PODC)*, pages 229–237, 1999.

[76] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18(4):319–349, 1988.

[77] C. D. Howard. Models of first-passage percolation. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 125–173. Springer, Berlin, 2004.

[78] S. Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.*, 8(4):347–361, 1999.

[79] S. Janson. Limit theorems for triangular urn schemes. *Probab. Theory Related Fields*, 134(3):417–452, 2006.

[80] J. Janssen. Spatial models for virtual networks. In *Programs, proofs, processes*, volume 6158 of *Lecture Notes in Comput. Sci.*, pages 201–210. Springer, Berlin, 2010.

[81] N. L. Johnson and S. Kotz. *Urn models and their application*. John Wiley & Sons, New York-London-Sydney, 1977. An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.

[82] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized Rumor Spreading. In *Proc. 41st Symp. Foundations of Computer Science (FOCS)*, pages 565–574, 2000.

[83] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proc. 44th Symp. Foundations of Computer Science (FOCS)*, pages 482–491, 2003.

[84] S. Keshav. Efficient and decentralized computation of approximate global state. *SIGCOMM Comput. Commun. Rev.*, 36(1):69–74, January 2006.

[85] J. Kleinberg. Complex networks and decentralized search algorithms. In *International Congress of Mathematicians. Vol. III*, pages 1019–1044. Eur. Math. Soc., Zürich, 2006.

[86] T. Kloks. *Treewidth: Computations and Approximations*. Springer-Verlag, 1994.

[87] I. Kolossváry, J. Komjáty, and L. Vágó. Degrees and distances in random and evolving Apollonian networks. *arXiv*, arXiv:1310.3864v1 [math.PR], 2013.

[88] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63:066123, May 2001.

[89] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *41st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 57–65. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.

[90] A. Laforgia. Further inequalities for the gamma function. *Mathematics of Computation*, 42(166):597–600, 1984.

[91] A. N. Langville and C. D. Meyer. Deeper inside PageRank. *Internet Math.*, 1(3):335–380, 2004.

[92] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), March 2007. Article 2.

[93] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009.

[94] A. Li and P. Peng. The small-community phenomenon in networks. *Math. Structures Comput. Sci.*, 22(3):373–407, 2012.

[95] G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida. New kid on the block: Exploring the Google+ social graph. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, IMC '12, pages 159–170, New York, NY, USA, 2012. ACM.

[96] H. M. Mahmoud. Pólya urn models and connections to random trees: A review. *Journal of the Iranian Statistical Society*, 2(1):53–114, 2003.

[97] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[98] A. Mehrabian. Justifying the small-world phenomenon via random recursive trees. *arXiv*, 1410.6397 [cs.DM], 2014. submitted, 30 pages.

[99] A. Mehrabian and A. Pourmiri. Randomized rumor spreading in poorly connected small-world networks. *arXiv*, 1410.8175 [cs.SI], 2014. submitted, 28 pages.

[100] A. Mehrabian and A. Pourmiri. Randomized rumor spreading in poorly connected small-world networks. In F. Kuhn, editor, *Distributed Computing (DISC '14)*, volume 8784 of *Lecture Notes in Computer Science*, pages 346–360. Springer Berlin Heidelberg, 2014.

[101] A. Mehrabian and N. Wormald. It's a small world for random surfers. *arXiv*, 1404.7189v1 [cs.DM], 2014. submitted, 37 pages.

[102] A. Mehrabian and N. Wormald. It's a small world for random surfers. In K. Jansen, J. D. P. Rolim, N. R. Devanur, and C. Moore, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM '14)*, volume 28 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 857–871. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014.

[103] M. Mihail, C. H. Papadimitriou, and A. Saberi. On certain connectivity properties of the internet topology. In *Proc. 44thSymp. Foundations of Computer Science (FOCS)*, pages 28–35, 2003.

[104] M. Mungan. Comment on "Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs". *Phys. Rev. Lett.*, 106:029802, Jan 2011.

[105] M. Newman. *Networks: An Introduction*. Oxford University Press Inc., New York, USA, 2010.

[106] K. Panagiotou, X. Pérez-Giménez, T. Sauerwald, and H. Sun. Randomized rumour spreading: The effect of the network topology. *Combinatorics, Probability and Computing*, 24:457–479, March 2015.

[107] K. Panagiotou and L. Speidel. Asynchronous rumor spreading on random graphs. In Leizhen Cai, Siu-Wing Cheng, and Tak-Wah Lam, editors, *Algorithms and Computation*, volume 8283 of *Lecture Notes in Computer Science*, pages 424–434. Springer Berlin Heidelberg, 2013.

[108] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. *Internet Mathematics*, 3(1):1–20, 2006. Conference version in COCOON 2002.

[109] C. Patsonakis and M. Roussopoulos. Asynchronous rumour spreading in social and signed topologies. *arXiv*, 1310.6119v3 [cs.SI], 2015.

[110] B. Pittel. On spreading a rumor. *SIAM J. Appl. Math.*, 47(1):213–223, 1987.

[111] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures & Algorithms*, 5(2):337–347, 1994.

[112] P. Prałat. The diameter of protean graphs. *Discrete Math.*, 308(15):3399–3406, 2008.

[113] S. M. Ross. *Introduction to probability models*. Academic Press, Inc., Boston, MA, fifth edition, 1993.

[114] W. Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Company, Inc., New York-Toronto-London, 1953.

[115] T. Sauerwald. On mixing and edge expansion properties in randomized broadcasting. *Algorithmica*, 56(1):51–88, 2010.

[116] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):pp. 425–443, 1969.

[117] R. van der Hofstad. Random graphs and complex networks. Lecture notes in preparation, available at http://www.win.tue.nl/~rhofstad/, October 2014.

[118] D. J. Watts and D. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[119] Z. Zhang, F. Comellas, G. Fertin, and L. Rong. High-dimensional Apollonian networks. *J. Phys. A*, 39(8):1811–1818, 2006.

[120] T. Zhou, G. Yan, and B.-H. Wang. Maximal planar networks with large clustering coefficient and power-law degree distribution. *Phys. Rev. E*, 71:046141, Apr 2005.

Index