

**MASTER THESIS**

**First Passage Percolation on  
Inhomogeneous Random Graphs**

**István Kolossváry**

**Supervisors:** Károly Simon

Professor

BME, Institute of Mathematics

Department of Stochastics

Júlia Komjáthy

Research Fellow

BME, Institute of Mathematics

Department of Stochastics

**BME**

**2011**

# Contents

<b>Abstract, Acknowledgements</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
<b>2 The model and the main results</b>	<b>5</b>
2.1 Sketch of proof and structure of the thesis . . . . .	10
<b>3 A branching process</b>	<b>12</b>
3.1 Asymptotics of multi-type BP quantities . . . . .	14
3.2 Proof of Theorem 3.4 . . . . .	16
<b>4 Embedding of BP into IHRG</b>	<b>22</b>
4.1 Marked branching processes . . . . .	23
4.2 Thinning branching processes . . . . .	27
4.3 Connection time . . . . .	31
<b>5 Proof of main results</b>	<b>36</b>
<b>6 Extending to the general model</b>	<b>40</b>
6.1 Approximation of kernels . . . . .	43
6.2 Extension of Theorems 2.3 and 2.4 . . . . .	45
<b>Summary and future plans</b>	<b>48</b>
<b>References</b>	<b>50</b>

## Abstract

This thesis investigates first passage percolation (FPP) on inhomogeneous random graphs (IHRG). The random graph model  $G(n, \kappa)$  we first study is the so-called finite type case of the general model introduced by Bollobás, Janson and Riordan in [6], i.e. each vertex is assigned a type  $s$  from a finite set  $\mathcal{S}$  and two vertices are connected with different probabilities according to their types given by the kernel matrix  $\kappa$ . Each edge is given an independent exponential edge weight with rate 1. Afterwards the general model is also considered. Our main assumption is that the average number of neighbors  $\tilde{\lambda}_n + 1$  of each vertex is independent of its type. We consider the cases where  $\tilde{\lambda}_n \rightarrow \tilde{\lambda}$  is finite or infinite.

The thesis can be considered a generalization of [3] written by Bhamidi, van der Hofstad and Hooghiemstra, where FPP is explored on the Erdős-Rényi random graphs, a special case where  $|\mathcal{S}| = 1$ . We find analogous results for the minimal weight of the path between uniformly chosen vertices in the giant component and for the hopcount, i.e. the number of edges on this minimal weight path.

In the sparse regime, i.e. when  $\tilde{\lambda} < \infty$ , a central limit theorem is proved for the hopcount, with asymptotic mean and variance both equal to  $(\tilde{\lambda} + 1)/\tilde{\lambda} \log n$ . Furthermore, we prove that the minimal weight centered by  $\log n/\tilde{\lambda}$  converges in distribution.

In the dense regime, where  $\tilde{\lambda} = \infty$ , without the edge weights the graph distances between uniformly chosen vertices are  $o(\log n)$ . However with the addition of the exponential edge weights we find that the hopcount  $H_n$  still scales as  $\log n$ , more precisely,  $(H_n - \beta_n \log n)/\sqrt{\log n}$  has a limiting standard normal distribution, where  $\beta_n = (\tilde{\lambda}_n + 1)/\tilde{\lambda}_n$ . The proofs make use of a direct relation between FPP on the IHRG and thinned continuous-time multi-type branching processes.

## Acknowledgements

I thank Károly Simon for his lectures on the topic, which turned my interest towards random graphs even more. I am grateful for the many hours of devoted time put into this research project by Juli Komjáthy.

This research was supported by the grant KTIA-OTKA # CNK 77778, funded by the Hungarian National Development Agency (NFÜ) from a source provided by KTIA.



# 1 Introduction

The study of random graphs has gained increasing attention the past decade due to the fact that large complex networks (such as the World Wide Web) have arisen in different fields of study. To understand the behavior of these networks many different models are introduced even today. Some of them use heuristics while others are mathematically rigorous.

In the 1960's Pál Erdős and Alfréd Rényi were the first to study random graphs. In their seminal model each edge of a graph on  $n$  vertices is present independently with probability  $p$ , these are the Erdős-Rényi (ER) random graphs. Putting  $p = c/n$ , it is a classical result (see [9]) that a giant component emerges (called the supercritical case) if and only if the constant  $c > 1$ . The Erdős-Rényi graphs are homogeneous in the sense that the edges are present with equal probabilities. However many real life networks do not show this property.

In [6] Bollobás, Janson and Riordan introduced a general inhomogeneous random graph model, which also includes the ER random graphs as a special case. The vertices are assigned different types and the edge probabilities depend on these types given by a 'kernel' function. Amongst other results, they characterized the supercritical regime. In this regime they also proved what the typical graph distance is between two randomly chosen vertices of the giant component. A possible generalization is to give the edges different edge weights, which can represent passing time or some cost.

This leads us to the problem of first passage percolation (FPP) in some random environment. Let the environment be some random graph and give each edge a random edge weight, typically independent and identically distributed (i.i.d.) positive random variables. Now think of fluid percolating through the graph from some source at a constant rate. As the environment grows one is interested in the asymptotics of various quantities of the flow.

In [3] Bhamidi, v. d. Hofstad and Hooghiemstra analyzed FPP on the ER random graph with i.i.d. exponential edge weights. They proved that the hopcount, i.e. the number of edges on the shortest-weight path between two randomly chosen vertices in the giant component, follows a central limit theorem. Furthermore, they show convergence in distribution for the weight of the shortest-weight path. Both of these quantities scale as  $\log n$  (see [3, Theorems 2.1 and 2.2]).

The aim of this thesis is to investigate FPP on inhomogeneous random graphs with i.i.d. exponential edge weights. We first consider the so called finite-type case of the general model of [6]. In this setting we show analogous results to that of [3] for the hopcount and the weight of the shortest-weight path. Afterwards we show how these results can be generalized to the original

model of [6]. Our main assumption is that the average number of neighbors of each vertex is independent of its type. We consider the cases when the average degree tends to a finite limit and also when it tends to infinity. Our arguments follow the outline of [3] and [6], however there are new challenges to face. It is standard to couple the exploration process of the graph to a branching process (BP). With the addition of types and exponential edge weights it is not enough to work in discrete time, we rely on the theory of multi-type continuous time branching processes (MCBP, see [1]). Furthermore, because of the types we lose independence in the BP, which leads to the use of more complicated techniques.

## 2 The model and the main results

In this section we define the random graph model  $G(n, \kappa)$  which is the finite-type case of the general model introduced in [6, Example 4.3]. The finite-type case has already been studied by Söderberg [13]. Each vertex is assigned a type from a finite set and the vertices are connected with different probabilities depending on their types given by a kernel matrix. After introducing some standard notation we state the main results concerning the hopcount on  $G(n, \kappa)$  and the weight of the shortest-weight path under different conditions for the average degree.

Let  $\mathcal{S} = \{1, 2, \dots, r\}$  be a finite set equipped with a probability measure  $\mu$ , i.e.  $\mu_i := \mu\{i\} \geq 0$  and  $\mu \cdot \mathbb{1} = 1$ , where  $\mathbb{1} = (1, 1, \dots, 1)^T$ . Each vertex  $i$  is assigned an element from  $\mathcal{S}$ , referred to as its type and denoted by  $t(i)$ . For each  $n$  we have a deterministic or random sample of  $n$  points from  $\mathcal{S}$ . Say  $n_t$  vertices are of type  $t$ , so  $\sum_{t \in \mathcal{S}} n_t = n$ . We assume that the samples are taken in such a way that

$$\frac{n_t}{n} \xrightarrow{\mathbf{P}} \mu_t \quad \text{holds for every } t \in \mathcal{S}. \quad (1)$$

For example if the samples are independent and identically distributed (i.i.d.) with distribution  $\mu$ , then (1) holds by the law of large numbers.

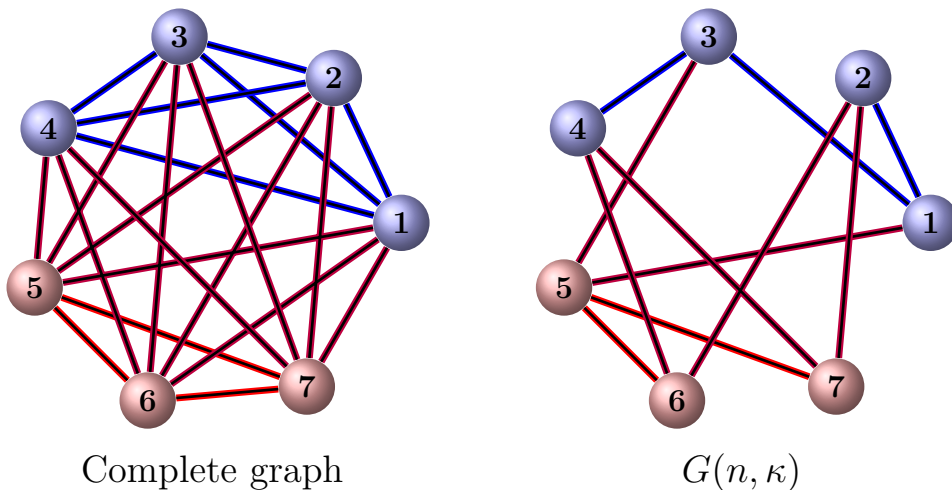
Up until Section 6 the kernel  $\kappa = (\kappa(s, t))_{s, t=1}^r$  will denote a symmetric non-negative  $r \times r$  matrix. The natural interpretation of  $\kappa$  is that it measures the density of edges. More precisely, the random graph  $G(n, \kappa)$  on vertex set  $[n] = \{1, 2, \dots, n\}$  is defined as follows. Two vertices  $i$  and  $j$  (with  $i \neq j$ ) are connected with probability

$$p_{t(i)t(j)} := \min \left\{ \frac{\kappa(t(i), t(j))}{n}, 1 \right\} \quad (2)$$

independently for each possible pair  $i \neq j$ .  $\kappa$  has only finite different values, so for  $n$  large enough ( $n > \max \kappa$ ) the probability simplifies to  $\kappa(t(i), t(j))/n$ . We can see that the Erdős-Rényi random graph is indeed just a very special case of this model. Namely, let  $\mathcal{S}$  consist of a single point and  $\kappa = c$ . Then the probabilities  $p_{ij}$  are all just simply  $c/n$  (for  $n > c$ ).

Finally, independently of all randomness each edge  $e$  is given an edge weight  $E_e$ , which is exponentially distributed with rate 1. During many of our arguments, we will make use of the memoryless property of the exponential distribution.

The following figure shows an example with  $|\mathcal{S}| = 2$  and  $n = 7$ . Start from the complete graph and perform different coin tosses on each edge according to (2) indicated by the various colors, the resulting graph is  $G(n, \kappa)$  (without the edge weights). Here nodes  $\{1, 2, 3, 4\}$  are of type 1 and  $\{5, 6, 7\}$  are of type 2, the edge probabilities  $p_{11} = 2/3 = p_{22}$  and  $p_{12} = p_{21} = 1/2$ .



**Remark 2.1.**

- (a) *The general model in detail will be considered in Section 6. In many cases the results in [6] are first proved for finite-type kernels and then approximation arguments are used for general kernels. This is why it is important to first understand FPP in the finite-type case.*
- (b) *In fact we can assume that  $\mu_i > 0$  for all  $i$ . An argument will be given in Section 6.*
- (c) *We also assume that  $\kappa$  is irreducible. Basically, this means that the vertices of  $G(n, \kappa)$  can't be split into two parts so that the probability of an edge from one part to the other is zero. The formal definition is given in Definition 6.4.*

It will be useful for us to use the following notation:

$$\lambda_{st} := \kappa(s, t)\mu_t, \quad (3)$$

for all  $s, t \in \mathcal{S}$ . Notice that  $\lambda_{st}$  has very literal meaning. For any given vertex of type  $s$  the number of its neighbors of type  $t$  are binomially distributed with parameters  $n_t - \delta_{st}$  and  $\kappa(s, t)/n$  ( $\delta_{st} = 1$  if and only if  $s = t$ ). So from (1) in the limit we get that  $\lambda_{st}$  gives us the average number of type  $t$  neighbors of a type  $s$  vertex. Later this will allow us to approximate with suitable Poisson distributions. From here we construct the matrix

$$A = \begin{pmatrix} \lambda_{11} - 1 & \lambda_{12} & \cdots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} - 1 & \cdots & \lambda_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} - 1 \end{pmatrix}. \quad (4)$$

**Principal Assumption 2.2.** *We will assume that the row sums of the matrix  $A$  are the same and equal to  $\tilde{\lambda} > 0$ .*

This means that in the limit the average degree is independent of the types. This is referred to as the homogeneous case in [6, Example 4.6]. In this case the global behavior of  $G(n, \kappa)$  in the limit is the same as of the ER random graph, but the local behavior can be quite different. The  $\tilde{\lambda} > 0$  condition is necessary and sufficient for a giant component to emerge in  $G(n, \kappa)$  [6, Theorem 3.1].

Let us introduce some standard notation. Let  $\text{Bin}(n, p)$ ,  $\text{Poi}(\lambda)$ ,  $\text{Exp}(\mu)$  respectively denote a binomial, a Poisson and an exponential random variable with the parameters having their usual meaning. Convergence almost surely, in distribution and in probability are denoted by  $\xrightarrow{a.s.}$ ,  $\xrightarrow{d}$ ,  $\xrightarrow{\mathbf{P}}$  respectively. A sequence of events holds with high probability (whp), if it holds with probability tending to 1 as  $n \rightarrow \infty$ . We use the Landau symbols  $O$  and  $o$  with their usual meaning. We say that a sequence of random variables  $X_n$  satisfies  $X_n = O_{\mathbf{P}}(b_n)$  if  $X_n/b_n$  is tight (i.e.  $\forall \varepsilon > 0 \exists M : \mathbf{P}(X_n > Mb_n) < \varepsilon$ ) or  $X_n = o_{\mathbf{P}}(b_n)$  if  $X_n/b_n \xrightarrow{\mathbf{P}} 0$ . Now we turn to the main results.

We investigate the weight and the number of edges on the shortest-weight path between two vertices  $x$  and  $y$ . The following two theorems describe the asymptotic behavior of the weight of the shortest-weight path and the number of edges on this path, the hopcount.

**Theorem 2.3** (CLT for hopcount). *Under Assumption 2.2 the hopcount  $H_n$  between two uniformly chosen vertices, conditioned on being connected,*

satisfies a central limit theorem of the form

$$\frac{H_n - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log n}} \xrightarrow{d} Z,$$

where  $Z$  is a standard normal variable.

Next we consider the asymptotics of the minimal weight path.

**Theorem 2.4** (Limit distribution for minimal weight path). *Under Assumption 2.2 there exists a non-degenerate real valued random variable  $W$ , such that the minimal weight  $W_n$  between two uniformly chosen vertices, conditioned on being connected, satisfies*

$$W_n - \frac{1}{\tilde{\lambda}} \log n \xrightarrow{d} W.$$

The distribution of  $W$  can be determined precisely from multi-type branching process quantities, which we will describe later.

**Remark 2.5** (Joint convergence). *The proof will show that the above convergence hold jointly with independent limits.*

A possible extension is to have  $\kappa$  depend on  $n$ . In this case we have a sequence of matrices  $A_n$ , each satisfying Assumption 2.2 with  $\tilde{\lambda}_n$  as the sum of the rows. Then we can extend our theorems in the following form:

**Corollary 2.6.** *If  $\lim_{n \rightarrow \infty} \tilde{\lambda}_n = \tilde{\lambda} < \infty$  then for the hopcount we have*

$$\frac{H_n - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log n}} \xrightarrow{d} Z,$$

where  $Z$  is a standard normal variable, and for the weight of the FPP

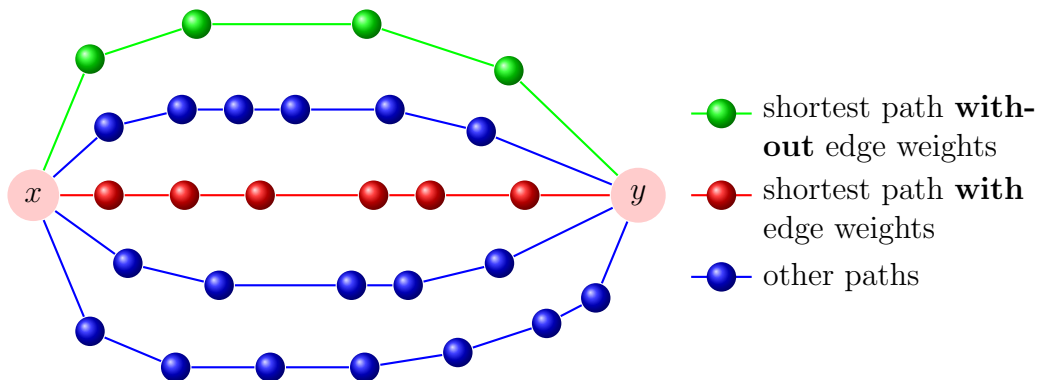
$$W_n - \frac{1}{\tilde{\lambda}} \log n \xrightarrow{d} W.$$

It is interesting to compare these results with its counterpart where there are no edge weights. Then of course  $G_n = W_n$ . Theorem 3.14 of [6] in our context says that in this case the typical distance between two vertices in the giant component still scales as  $\log n$ , but with a different constant:  $1/\log(\tilde{\lambda} + 1)$ . Comparing the constants, we see that

$$\frac{1}{\tilde{\lambda}} < \frac{1}{\log(1 + \tilde{\lambda})} < \frac{\tilde{\lambda} + 1}{\tilde{\lambda}}.$$



This shows that by adding edge weights the structure of the graph changes. Along the shortest-weight path more vertices are visited ( $H_n$ ) than on the path with the least number of vertices. At the same time the weight of the path ( $W_n$ ) becomes smaller than the number of edges on the path with the least number of vertices. The figure below illustrates this, where the red path is the shortest-weight path while the green is the one with the least number of vertices.



Similarly as in [3] we also investigate the dense regime, where  $\lim_{n \rightarrow \infty} \tilde{\lambda}_n = \infty$ , i.e. the average degree tends to infinity. In this case whp any two vertices are connected, so the giant component contains  $n(1 - o(1))$  vertices (see [6]). Again comparing with the counterpart without edge weights, the change in the graph structure is even more significant. Without edge weights the graph is ultra small, meaning that graph distances between uniformly chosen vertices are  $o(\log n)$ . However with the addition of edge weights the following theorem states that even the magnitudes do not coincide. We get the same type of behavior as observed in the sparse setting, which means that on the shortest-weight path many more vertices are visited.

**Theorem 2.7** (Dense setting). *Under Assumption 2.2 and if  $\lim_{n \rightarrow \infty} \tilde{\lambda}_n = \infty$ , then we have*

(a)

$$\frac{H_n - \frac{\tilde{\lambda}_n + 1}{\lambda_n} \log n}{\sqrt{\log n}} \xrightarrow{d} Z,$$

where  $Z$  is a standard normal variable and we can substitute  $\frac{\tilde{\lambda}_n + 1}{\lambda_n}$  by 1 if and only if  $\tilde{\lambda}_n / \sqrt{\log n} \rightarrow \infty$ .

(b)

$$\tilde{\lambda}_n W_n - \log n \xrightarrow{d} \tilde{W},$$

where in distribution  $\tilde{W}$  is equal to the sum of independent random variables  $Y_1 + Y_2 - Y_3$ , with  $Y_1$  and  $Y_2$  standard Gumbel random variables and  $Y_3$  another Gumbel random variable with distribution function  $\exp(-e^{-\tilde{\lambda}x})$ .

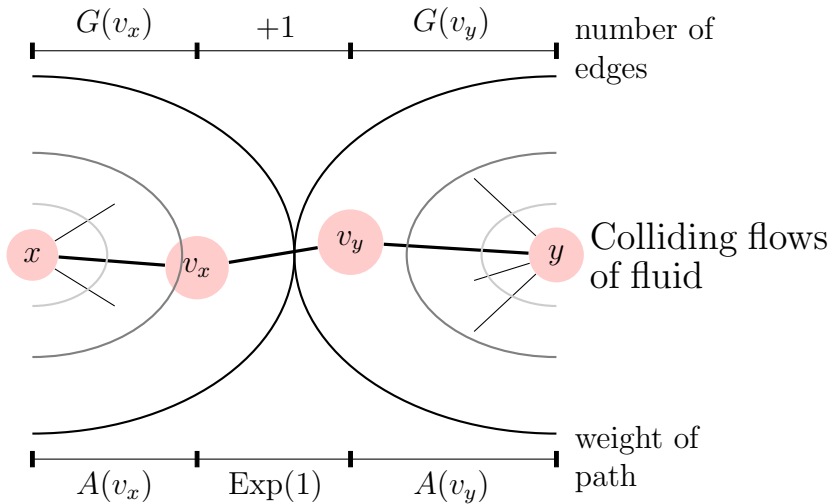
## 2.1 Sketch of proof and structure of the thesis

Roughly, the proof will go as follows. We start exploring the neighborhood of the two vertices  $x$  and  $y$ , respectively. It is standard to relate the exploration process of a random graph to a branching process. Since the number of neighbors of a vertex in the inhomogeneous random graph (IHRG) is binomially distributed it is practical to approximate the exploration by a BP with Poisson offspring, where the parameter is the limit of the product of the parameters of the binomial distribution. Using results from the theory of branching processes we can prove results for the Poisson BP analogous to the main results.

In order to carry through the results to the IHRG, first we need to couple the Poisson BP with the BP coming directly from the IHRG, i.e. where the offspring distribution is the sum of binomial distributions, with parameters depending on the type of the vertex. Secondly, we need to couple the BP to the exploration, so we need to identify the vertices of the BP and the vertices of the IHRG with one another with high probability. To achieve this, marked branching processes are introduced. Marks are assigned to the vertices according to their type in such a way that we can rule out loop and multiple edges in the IHRG. To deal with the problem of cycles in the graph, the notion of thinning is introduced on branching processes. This way we get an embedding of the BP into the IHRG. Perhaps toughest of all is to rigorously catch the idea of colliding the two flows in the IHRG.

To determine the shortest-weight path between two vertices  $x$  and  $y$  in the giant component, think of fluid emanating from these two vertices simultaneously at rate 1. Intuitively it is clear that when these two flows collide, then the shortest-weight path between  $x$  and  $y$  has been found.

The random time of the collision will be referred to as the connection time  $C_{xy}$ . In the figure below, the vertices  $v_x$  and  $v_y$  denote the last wetted vertices (along the shortest path) by the flows from  $x$  and  $y$  respectively before time  $C_{xy}$ .



Let  $G(v_x)$  denote the number of edges between the source  $x$  and vertex  $v_x$  and  $A(v_x)$  the weight of this path. Then for the hopcount  $H_n(x, y)$  and for the weight  $W_n(x, y)$  of the shortest-weight path we get that

$$H_n(x, y) = G(v_x) + G(v_y) + 1,$$

$$W_n(x, y) = A(v_x) + A(v_y) + \text{Exp}(1).$$

Thus it is crucial to determine when the connection actually happens. The difficulty is that we can't let the fluid flow continuously, but only in discrete steps as new vertices are wetted by the flow. Intuitively one could argue that we could also get the shortest path by delaying one of the flows or initiating only one flow. However, we will see that the proof only works when the two flows are about the same size.

Keeping the above in mind here is the outline of the subsequent sections:

- Section 3 could stand alone in its own right. A continuous-time multi-type branching process is introduced that will be related to an exploration process of the giant component of  $G(n, \kappa)$  in Section 4. To keep the paper self-contained, the section is continued with results concerning various quantities of branching processes that will be of use in our arguments. The section is finished with the proof of Theorem 3.4, which is the analog of the main results in the branching process setting.
- In Section 4 the crucial connection is made between the model  $G(n, \kappa)$  of Section 2 and the branching process of Section 3. This is done in a few steps, which are dealt with in separate subsections. We couple the BP of Section 3 to a binomial BP, marked branching processes are

introduced and the notion of thinning is discussed. The final subsection analyzes the connection time rigorously.

- After the preparations of Sections 3 and 4 we prove the main results for the finite-type case model in Section 5.
- Section 6 deals with the general model of [6]. After necessary definitions and a discussion of approximating general kernels, the main results of Section 2, when  $\tilde{\lambda} < \infty$ , are generalized to this setting.

### 3 A branching process

Analogous to the use of the single-type Poisson branching process (BP) in the analysis of the ER random graph, another BP arises naturally in the model  $G(n, \kappa)$ . Namely, when we explore the neighborhood of a vertex  $x$  in the inhomogeneous graph with edge weights, the types of vertices determine the distribution of the number of neighbors, leading to a multi-type BP. We can think of the neighboring vertices being explored at time  $t$  when the distance from  $x$  is less than  $t$ . Thus, the exploration can be considered as a continuous time process. As the edge weights are exponentially distributed, the memoryless property of it implies that it is actually Markovian and can be modeled by a continuous time BP known from the literature (see [1]).

Since the number of type  $t$  neighbors of a vertex of type  $s$  follows the law of  $\text{Bin}(n_t - \delta_{st}, \kappa(s, t)/n)$ , we can well-approximate this by a Poisson distribution with parameter  $\kappa(s, t)\mu_t$ . We will later see that the error arising from this approximation is negligible. Thus, in the branching process when a particle of type  $s \in \mathcal{S}$  dies, it gives birth to a Poisson number of children of type  $t$  with parameter

$$\lambda_{st} := \kappa(s, t)\mu_t,$$

for all  $t \in \mathcal{S}$ . From here we construct the same matrix  $A$  as in (4). In this BP setting, the row sum  $(A\mathbf{1})_s$  gives the expected value of the change in the number of alive individuals after the death of a particle of type  $s$ . We can restate Assumption 2.2 the following way in this setting:

**Assumption 3.1.** *We shall assume that the expected value of the children of a particle doesn't depend on its type, i.e. the sum of each row of  $A$  is equal. Let  $\tilde{\lambda} = (A\mathbf{1})_s > 0$  denote this constant.*

The  $\tilde{\lambda} > 0$  condition guarantees that with strictly positive probability the BP survives, see [1, Chapter V.3]. The survival probability can be determined from the fixpoint equation  $\vec{\rho} = 1 - \exp\{-(A + I)\vec{\rho}\}$ , and assumption 3.1

implies that  $\rho_t = \rho$  for all types  $t$ . For a proof see [6, Chapter 5.]. Making use of the properties of Poisson distributions and ignoring the types, under Assumption 3.1 we can see that the number of children of each particle has a  $\text{Poi}(\lambda + 1)$  distribution. So in distribution this is equal to a single-type branching process with offspring distribution  $\text{Poi}(\tilde{\lambda} + 1)$ . We now formally construct the branching process.

Let  $D_1, D_2, \dots$  be a sequence of non-negative i.i.d. random variables with distribution  $\text{Poi}(\tilde{\lambda} + 1)$  and define the sum  $S_i := \sum_{j=1}^i D_j - (i - 1)$ .

**Construction 3.2.** *1. Start with the root (1st vertex) which dies immediately giving birth to  $D_1$  alive offspring.*

*2. Each alive offspring lives for an  $\text{Exp}(1)$  amount of time, independent of all other randomness involved.*

*3. When the  $i$ -th vertex dies it leaves behind  $D_i$  alive offspring.*

To be precise, conditioned on type  $t(i)$  of the  $i$ -th vertex, each  $D_i$  is the sum of  $r$  different independent Poisson random variables with parameters given by the  $t(i)$ -th row of  $A + I$ . The types are assigned to each new offspring accordingly. In this context we can identify  $S_i$  as the number of alive vertices after the  $i$ -th death. Let  $S_i^{(t)}$  be the number of alive vertices of type  $t$  after the  $i$ -th death. The memoryless property of the exponential distribution and the fact that choosing the minimum out of i.i.d.  $\text{Exp}(1)$  distributions is uniformly distributed implies that looking at the process at the split times of the particles gives the following equivalent description of Construction 3.2:

**Construction 3.3** (Discrete-time reformulation).

*1. At step 1, start with one alive vertex, the root, dying immediately and giving birth to  $D_1$  alive offspring.*

*2. At each step  $i$ , pick one of the alive vertices at random, this vertex dies giving birth to  $D_i$  children.*

Let  $T_1, T_2, \dots$  be the time spacings between the  $i$ -th and  $i + 1$ -th split. Since  $T_i$  conditioned on  $S_i$  is the minimum of  $S_i$  exponential variables, it is exponential with rate  $S_i$ . Denote by  $A_m = T_1 + T_2 + \dots + T_m$  the time of the  $m + 1$ -th split. Thus  $A_m$  expresses the distance of the  $m$ -th closest vertex from the root. We denote the graph distance, called the generation of this vertex from the root by  $G_m$ . By the above mentioned property that the minimum is uniform among the  $S_{m-1}$  alive vertices,  $G_m$  equals in distribution

to the graph distance between the root and a uniformly chosen vertex among all alive vertices at step  $m$ .

Hence  $A_m$  gives the distribution of the weight of the shortest-weight path between the root and the  $m$ -th chosen vertex, while  $G_m$  gives the hopcount on this path. Thus it is of crucial importance for us to understand the behavior of these two quantities. With some non-zero probability it is possible that  $S_i = 0$  for some  $1 \leq i \leq m$ , i.e. the BP dies out. We will always condition on  $S_i > 0$  for all  $1 \leq i \leq m$ , which has strictly positive probability because we constructed a supercritical BP. With this conditioning we have the following asymptotics:

**Theorem 3.4** (Result for BP conditioned on survival). *Conditioned on  $S_i > 0$  for all  $1 \leq i \leq m$ , the following hold for  $m \rightarrow \infty$ :*

$$\frac{G_m - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log m}{\sqrt{\frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log m}} \xrightarrow{d} Z, \quad (5)$$

where  $Z$  is a standard normal variable and

$$A_m - \frac{1}{\tilde{\lambda}} \log m \xrightarrow{a.s.} -\frac{1}{\tilde{\lambda}} \log \left( \frac{1}{\tilde{\lambda} \hat{W}} \right). \quad (6)$$

The distribution of  $\hat{W}$  can be determined from the limit behavior of the BP, see below.

### 3.1 Asymptotics of multi-type BP quantities

In this subsection we quote results from [1, Chapter V] in the context of our branching process. Let us recall our notations.  $S_m = \sum_{j=1}^m D_j - (m-1)$  denotes the number of alive vertices after the  $m$ -th split (the split of the root is the first). Of the  $S_m$  vertices  $S_m^{(t)}$  are of type  $t$ .  $A_m$  denotes the time of the  $m$ -th split. It is also important to keep track of the types of the vertices that split. Let  $N_m^{(t)}$  denote the number of splits of type  $t$  vertices among the first  $m$  splits. We need a further assumption to state theorems about the asymptotics of these quantities.

**Assumption 3.5.** *Assume that the matrix  $A$  is irreducible, meaning that there exists a  $k_0$  s.t.  $A^{k_0}$  has strictly positive entries.*

Then, by the Perron Frobenius theorem we know that  $\tilde{\lambda}$  is the leading eigenvalue of  $A$  and it has associated nonnegative left eigenvector  $\pi = (\pi_1, \pi_2, \dots, \pi_r)$  normalized so that  $\pi \mathbf{1} = 1$ . For  $\text{Poi}(\lambda_{st})$  random variables  $\xi_{st}$

$$\mathbf{E}(\xi_{st} \log \xi_{st}) < \infty \text{ holds for all } s, t \in \mathcal{S}.$$

On the set of non-extinction there must be an infinite number of splits, so  $A_m \rightarrow \infty$  as  $m \rightarrow \infty$ . The following hold:

**Theorem 3.6** (Asymptotics of BP quantities).

(i) On the set of non-extinction as  $m \rightarrow \infty$

$$\lim_{m \rightarrow \infty} \frac{(S_m^{(1)}, \dots, S_m^{(r)})}{\tilde{\lambda} m} = \pi \text{ a.s.}, \quad (7)$$

(ii) and similarly

$$\frac{(N_m^{(1)}, \dots, N_m^{(r)})}{m} \xrightarrow{\text{a.s.}} \pi. \quad (8)$$

(iii)  $m e^{-\tilde{\lambda} A_m} \xrightarrow{\text{a.s.}} \tilde{\lambda}^{-1} W$ .

(iv)  $A_m - (1/\tilde{\lambda}) \log m$  converges to a finite limit  $-\tilde{\lambda}^{-1} \log(\tilde{\lambda}^{-1} W)$  on a set of positive probability and  $\mathbf{P}(W > 0) > 0$ . (a bit stronger than the previous)

The distribution of  $W$  is given by the moment-generating functions  $\varphi_t(x) = \mathbf{E}_t(e^{-xW})$ , that satisfy the equation

$$\varphi_t(x) = \int_0^\infty f^{(t)}(\phi(xe^{-\tilde{\lambda}y})) e^{-y} dy,$$

where  $\phi(x) = (\varphi_1(x), \dots, \varphi_r(x))$  and  $f^{(t)}$  is the  $r$  variable generating function of the offspring distribution of a type  $t$  vertex.  $W$  has an atom at zero of size  $1 - \rho$ , i.e. the extinction probability of the BP. The random variable  $\hat{W}$  in (6) is equal in distribution to  $(W|W > 0)$  (details in [1, Chapter V. 7.5]).

Here we collect two simple consequences of Theorem 3.6. We can write  $S_m = 1 + \sum_{j=1}^m Y_j$ , where  $Y_j^{(t)} = Y_j = D_j - 1$ . ( $Y_j$  doesn't depend on the type because of Assumption 3.1.)

$$\frac{1}{m} S_m = \frac{1}{m} + \frac{1}{m} \sum_{j=1}^m Y_j = \frac{1}{m} + \frac{1}{m} \sum_{t \in \mathcal{S}} \sum_{j=1}^{N_m^{(t)}} Y_j^{(t)} = \sum_{t \in \mathcal{S}} \frac{N_m^{(t)}}{m} \cdot \underbrace{\frac{1}{N_m^{(t)}} \sum_{j=1}^{N_m^{(t)}} Y_j^{(t)}}_{\rightarrow \mathbf{E}(Y_j^{(t)})},$$

by the strong law of large numbers. So as  $m \rightarrow \infty$

$$\frac{1}{m} S_m \xrightarrow{(8)} \sum_{t \in \mathcal{S}} \pi_t \mathbf{E}(Y_j^{(t)}) = \sum_{t \in \mathcal{S}} \pi_t (A\mathbf{1})_t = \pi A\mathbf{1} = \tilde{\lambda} \pi \mathbf{1} = \tilde{\lambda}. \quad (9)$$

Note that here we did not use Assumption 3.1. From here we can see that

$$\frac{S_j^{(t)}}{S_j} = \frac{S_j^{(t)}}{j} \cdot \frac{j}{S_j} \stackrel{(7),(9)}{\rightsquigarrow} \tilde{\lambda} \pi_t \frac{1}{\tilde{\lambda}} = \pi_t. \quad (10)$$

In each case we have convergence almost surely. Estimates can also be given for the magnitude of the error terms in (9) and (10). Depending on the size of the eigenvalue of  $A$  with the second largest real part, one can properly normalize  $S_j - \tilde{\lambda}j$  and  $S_j^{(t)} - \tilde{\lambda}\pi_t j$  such that in distribution it tends to a well-defined normal random variable. For details we refer to [10, Theorems 3.22-24.]. We will later make use of this in some of our arguments.

### 3.2 Proof of Theorem 3.4

Recall that  $A_m$  gives the distribution of the weight of the shortest-weight path between the root and the  $m$ -th chosen vertex, while  $G_m$  gives the hopcount on this path. We start off by quoting a fundamental result identifying the distributions of  $G_m$  and  $A_m$  as sums of independent random variables. Consider Construction 3.3 throughout the subsection.

**Proposition 3.7** (Distribution of hopcount and FPP). *At step  $m$  choose an alive vertex uniformly at random from all vertices alive at this time. Then*

(a) *the generation of the  $m$ -th chosen vertex satisfies*

$$G_m \stackrel{d}{=} \sum_{i=1}^m I_i, \quad (11)$$

*where  $I_1, I_2, \dots$  are independent Bernoulli random variables with success probability  $\mathbf{P}(I_i = 1 \mid D_1, \dots, D_i) = D_i/S_i$ .*

(b) *the weight of the shortest-weight path between the root and the  $m$ -th chosen vertex satisfies*

$$A_m = T_1 + T_2 + \dots + T_m \stackrel{d}{=} \sum_{i=1}^m E_i/S_i, \quad (12)$$

*where  $E_1, E_2, \dots$  are i.i.d.  $\text{Exp}(1)$  random variables.*

*Proof.* For an analytic proof of (11) see [4] and a new probabilistic proof can be found in [2, Proposition 4.2]. We already argued that  $T_i$  is exponentially distributed with rate  $S_i$ , from here we immediately get part (b), since  $\text{Exp}(n) \stackrel{d}{=} E_1/n$ .  $\square$



From part (a) we see that

$$\mathbf{E}(G_m \mid D_1, \dots, D_m) = \sum_{i=1}^m \frac{D_i}{S_i}, \text{ and}$$

$$\mathbf{Var}(G_m \mid D_1, \dots, D_m) = \sum_{i=1}^m \frac{D_i}{S_i} \left(1 - \frac{D_i}{S_i}\right).$$

Shortly we will see the right magnitude of the two right hand sides. To prove (5) we will show that the standardization

$$\tilde{G}_m = \frac{G_m - \sum_{i=1}^m \frac{D_i}{S_i}}{\sqrt{\sum_{i=1}^m \frac{D_i}{S_i} \left(1 - \frac{D_i}{S_i}\right)}} \quad (13)$$

of  $G_m$  converges in distribution to a normal random variable  $Z$ . This will be done by showing that the Wasserstein distance between the distribution of  $\tilde{G}_m$  and that of  $Z$  goes to 0. This is enough because the Wasserstein metric on the space of probability measures induces the same topology as the topology of weak convergence ([8, Theorem 11.3.3]). The Wasserstein metric on the space of probability measures is defined by

$$\text{Wass}(\mu_1, \mu_2) = \sup \left\{ \left| \int_{\mathbb{R}} g d\mu_1 - \int_{\mathbb{R}} g d\mu_2 \right| : g \text{ is 1-Lipschitz and bounded} \right\}.$$

The following lemma [3, Lemma 3.4] gives an upper bound on the Wasserstein distance of the law of  $\tilde{G}_m$  and of  $Z$ .

**Lemma 3.8.** *Let  $\mu_m$  denote the distribution of  $\tilde{G}_m$  and  $\mu$  the distribution of a standard normal random variable. Then with the abbreviation  $\rho_i = D_i/S_i$*

$$\text{Wass}(\mu_m, \mu) \leq \frac{3}{\sqrt{\sum_{i=1}^m \rho_i(1 - \rho_i)}}.$$

*Proof.* Combining [7, Theorem 3.1 and 3.2] yields

$$\text{Wass}(\mu_m, \mu) \leq 3 \sum_{i=1}^m \mathbf{E}(|\eta_i|^3),$$

where

$$\eta_i = \frac{I_i - \rho_i}{\sqrt{\sum_{i=1}^m \rho_i(1 - \rho_i)}},$$

and where  $I_i$  is a Bernoulli random variable with success probability  $\rho_i$ . From here we obtain that  $\sum_{i=1}^m \mathbf{E}(|\eta_i|^3)$  equals

$$\frac{1}{(\sum_{i=1}^m \rho_i(1 - \rho_i))^{3/2}} \sum_{i=1}^m \rho_i(1 - \rho_i) \cdot \underbrace{(\rho_i^2 + (1 - \rho_i)^2)}_{\leq 1} \leq \frac{1}{\sqrt{\sum_{i=1}^m \rho_i(1 - \rho_i)}}.$$

□

If the process dies out, then the convergence in (13) does not make sense, so we still have to deal with the technical issue of conditioning on survival. Unfortunately, conditioning on survival removes the independence of the increments of  $\{S_j\}_{j=1}^m$ . Note that the process  $\{S_j\}_{j=1}^m$  till extinction is equal in distribution to a random walk on  $\mathbb{Z}$  with independent increments  $D_j - 1 \sim \text{Poi}(\tilde{\lambda} + 1) - 1$ . To overcome the problem of conditioning, we couple the process conditioned on survival to a process conditioned to stay positive only in the first  $\omega_m = o(\log \log m)$  steps and then continue with independent increments. This method will work since after  $\omega_m$  steps the walk has reached height approximately  $\tilde{\lambda}\omega_m$  and from here, whp, will remain positive. The following construction makes this precise. Recall  $S_m = 1 + \sum_{j=1}^m Y_j$ , where  $Y_j = D_j - 1$ .

**Construction 3.9.** Let  $\{S_1^*, \dots, S_{\omega_m}^*\} \sim \{S_1, \dots, S_{\omega_m} | S_i > 0 \forall i < \omega_m\}$ . Furthermore, let  $Y_{\omega_m+1}, Y_{\omega_m+2}, \dots$  be i.i.d. random variables with distribution  $\text{Poi}(\tilde{\lambda} + 1) - 1$ , independent of  $\{S_1^*, \dots, S_{\omega_m}^*\}$ . Define the sequence  $R_1, \dots, R_m$  by

$$R_j := \begin{cases} S_j^*, & j \leq \omega_m \\ S_{\omega_m}^* + \sum_{i=\omega_m+1}^j Y_i, & \omega_m + 1 \leq j \leq m. \end{cases}$$

The following Proposition states that the processes  $\{R_j\}_{j=1}^m$  and  $\{S_j | \forall j : S_j > 0\}_{j=1}^m$  can be coupled whp., i.e. they are equal whp. Furthermore, we can derive good bounds for  $\{R_j\}_{j=1}^m$  from which we can obtain the right magnitude of  $\sum_{i=1}^m \frac{D_i}{S_i}$  and  $\sum_{i=1}^m \frac{D_i}{S_i} \left(1 - \frac{D_i}{S_i}\right)$  in formula (13).

**Proposition 3.10.**

(a)

$$\mathbf{P}(R_1 > 0, \dots, R_m > 0) = 1 - o(1) \text{ as } m \rightarrow \infty,$$

and there exists a coupling such that

$$\mathbf{P}\left(R_j = \hat{S}_j, 1 \leq j \leq m\right) = 1 - o(1) \text{ as } m \rightarrow \infty,$$

where  $\{\hat{S}_1, \dots, \hat{S}_m\} \sim \{S_1, \dots, S_m | S_i > 0 \forall i \leq m\}$ .

(b) There exists a constant  $C > 0$  such that, whp, for all  $j > \log \log m$

$$\tilde{\lambda}j - C\sqrt{j \log j} \leq S_j \leq \tilde{\lambda}j + C\sqrt{j \log j}. \quad (14)$$

(c) Let  $\rho_i = D_i/S_i$ , then

$$\sum_{i=1}^m \rho_i(1 - \rho_i) = \frac{\tilde{\lambda} + 1}{\tilde{\lambda}} \log m + O_{\mathbf{P}}(\log \log m). \quad (15)$$

*Proof of Proposition 3.10, part (a).* With slightly different notation, the proof of part (a) is analogous to the proof in [3, Proposition 3.7], since the increments are  $\text{Poi}(\tilde{\lambda} + 1) - 1$ , so the same random walk argument works. The rough idea is that if the random walk stays positive up to time  $\omega_n$ , then it will be with high probability around its expectation, so from here the event of hitting zero has exponentially small probability.  $\square$

*Proof of Proposition 3.10, part (b).* Define the event

$$E = \left\{ \forall j > \log \log m : \left| S_j - \tilde{\lambda}j \right| < C\sqrt{j \log j} \right\}.$$

Then

$$\begin{aligned} \mathbf{P}(E^C) &= \mathbf{P}\left(\exists j > \log \log m : \left| S_j - \tilde{\lambda}j \right| > C\sqrt{j \log j}\right) \\ &\leq \sum_{j=\log \log m}^{\infty} \mathbf{P}\left(\left| S_j - \tilde{\lambda}j \right| > C\sqrt{j \log j}\right) = \sum_{j=\log \log m}^{\infty} e_j. \end{aligned}$$

To estimate a single error term  $e_j$ , we use the following generalized form of Bernstein's inequality:

**Lemma 3.11** (Bernstein's inequality). *Let  $X_1, \dots, X_m$  be possibly non-independent random variables. Suppose that for all integer  $i > 0$ ,*

$$\begin{aligned} \mathbf{E}(X_i | \mathcal{F}_{i-1}) &= 0 \\ \mathbf{E}(X_i^2 | \mathcal{F}_{i-1}) &\leq R_i \mathbf{E}(X_i^2) \\ \mathbf{E}(X_i^k | \mathcal{F}_{i-1}) &\leq \frac{1}{2} \mathbf{E}(X_i^2 | \mathcal{F}_{i-1}) L^{k-2} k!, \end{aligned}$$

where  $\mathcal{F}_{i-1} = \sigma(X_1, \dots, X_{i-1})$ . Then for  $0 < t \leq \frac{\sqrt{\sum_{i=1}^m R_i \mathbf{E}(X_i^2)}}{2L}$

$$\mathbf{P}\left(\sum_{i=1}^m X_i > 2t \sqrt{\sum_{i=1}^m R_i \mathbf{E}(X_i^2)}\right) \leq \exp(-t^2).$$

*Proof.* Following the standard proof of Bernstein's inequality one can easily see that the statement holds.  $\square$

We apply Lemma 3.11 to the random variables  $X_i := D_i - (\tilde{\lambda} + 1)$ . Thus  $X_i$  are centered Poisson random variables with  $\mathbf{E}(X_i^2) = \tilde{\lambda} + 1$  and  $\sum_{i=1}^m X_i = S_m - m\tilde{\lambda}$ . It can be easily checked that the conditions hold with  $R_i = 1$  and  $L = \sqrt{\tilde{\lambda} + 1}$ . Hence, it follows that

$$\mathbf{P}\left(\left|S_j - \tilde{\lambda}j\right| > 2t\sqrt{(\tilde{\lambda} + 1)j}\right) \leq 2\exp(-t^2), \text{ for } 0 < t \leq \frac{\sqrt{j}}{2}.$$

Plugging in  $t = c\sqrt{\log j}$ , where  $c > 1$ , we get that

$$\begin{aligned} \mathbf{P}(E^C) &\leq \sum_{j=\log \log m}^{\infty} \mathbf{P}\left(\left|S_j - \tilde{\lambda}j\right| > C\sqrt{j \log j}\right) \\ &\leq \sum_{j=\log \log m}^{\infty} \frac{2}{j^c} = \frac{2}{(\log \log m)^{c-1}} + o(1) \xrightarrow{m \rightarrow \infty} 0. \end{aligned} \tag{16}$$

This proves part (b) with  $C = 2c\sqrt{\tilde{\lambda} + 1}$ .  $\square$

*Proof of Proposition 3.10, part (c).* We first claim that  $\sum_{j=1}^m \rho_j^2 = O_{\mathbf{P}}(1)$  i.e. the sequence is tight. Indeed, fix  $\varepsilon > 0$ ,

$$\mathbf{P}\left(\sum_{j=1}^m \frac{D_j^2}{S_j^2} > K\right) \leq \mathbf{P}\left(\sum_{j=1}^m \frac{D_j^2}{S_j^2} > K \mid E\right) + \mathbf{P}(E^C),$$

so by (16) it is enough to show that  $\exists K(\varepsilon)$  such that the first term is  $\leq \varepsilon/2$ . From part (b) and by Markov's inequality we get that

$$\begin{aligned} \mathbf{P}\left(\sum_{j=1}^m \frac{D_j^2}{S_j^2} > K \mid E\right) &\leq \mathbf{P}\left(\sum_{j=1}^m \frac{D_j^2}{(\tilde{\lambda}j - C\sqrt{j \log j})^2} > K\right) \\ &\leq \frac{1}{K} \sum_{j=1}^m \frac{\mathbf{E}(D_j^2)}{(\tilde{\lambda}j - C\sqrt{j \log j})^2} \\ &\leq \frac{1}{K} \underbrace{\frac{2((\tilde{\lambda} + 1)^2 + \tilde{\lambda} + 1)}{\tilde{\lambda}^2}}_{=: c} \underbrace{\sum_{j=1}^m \frac{1}{j^2}}_{\leq \pi^2/6} \leq \frac{\varepsilon}{2}, \end{aligned}$$

if we pick  $K(\varepsilon) \geq c\pi^2/3\varepsilon$ . Hence, we can neglect the term  $\sum_{i=1}^m \rho_i^2$  in (15).

$$\sum_{j=1}^m \rho_j = \sum_{j=1}^m \frac{D_j}{S_j} = \sum_{j \leq \log \log m} \frac{D_j}{S_j} + \sum_{j=\lceil \log \log m \rceil}^m \frac{D_j}{S_j},$$

where the first sum is obviously of order  $O_{\mathbf{P}}(\log \log m)$ . So it is enough to prove that for every  $\varepsilon > 0$  there exists an  $M = M(\varepsilon)$ , such that

$$\mathbf{P}(A) = \mathbf{P}\left(\left|\sum_{j=\lceil \log \log m \rceil}^m \frac{D_j}{S_j} - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log m\right| > M \log \log m\right) \leq \varepsilon.$$

Intersecting with the event  $E$  it is enough to show that  $\mathbf{P}(A|E) \leq \varepsilon/2$ . For shorter formulas we abbreviate  $\log \log m$  with  $x$ . Again applying the bounds in (b)

$$\begin{aligned} \mathbf{P}(A|E) &\leq \mathbf{P}\left(\left|\sum_{j=\lceil \log \log m \rceil}^m \frac{D_j}{\tilde{\lambda}j - \sqrt{j \log j}} - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log m\right| > Mx\right) \\ &\quad + \mathbf{P}\left(\left|\sum_{j=\lceil \log \log m \rceil}^m \frac{D_j}{\tilde{\lambda}j + \sqrt{j \log j}} - \frac{\tilde{\lambda}+1}{\tilde{\lambda}} \log m\right| > Mx\right) \end{aligned} \quad (17)$$

To apply Chebyshev's inequality the variables  $D_j$ -s need to be centered. We will see below that an estimate is needed for the following expression:

$$\begin{aligned} \sum_{j=\log \log m}^m \frac{1}{j - C\sqrt{j \log j}} &= \sum_{j=\log \log m}^m \frac{1}{j \left(1 - C\sqrt{\frac{\log j}{j}}\right)} \\ &= \sum_{j=\log \log m}^m \frac{1}{j} + \sum_{j=\log \log m}^m \sum_{k=1}^{\infty} C^k \frac{(\log j)^{k/2}}{j^{k/2+1}}. \end{aligned}$$

The first sum is  $\log m - \log \log \log m + o(1)$  and using that  $\log j \leq \sqrt{j}$  we show that the second sum is  $o(1)$ :

$$\begin{aligned} \sum_{j=x}^m \sum_{k=1}^{\infty} \frac{(\log j)^{k/2}}{j^{k/2+1}} &\leq \sum_{j=x}^m \sum_{k=1}^{\infty} \frac{j^{k/4}}{j^{k/2+1}} = \sum_{k=1}^{\infty} \sum_{j=x}^m \frac{1}{j^{k/4+1}} \leq \sum_{k=1}^{\infty} \frac{4}{kx^{k/4}} \\ &\leq 4 \sum_{k=1}^{\infty} \left(\frac{1}{x^{1/4}}\right)^k = \frac{(\log \log m)^{-1/4}}{1 - (\log \log m)^{-1/4}} = o(1). \end{aligned}$$

Similarly we get the same estimate for  $\sum_{j=x}^m (j + C\sqrt{j \log j})^{-1}$ . From here

$$\begin{aligned}
& \mathbf{P} \left( \left| \sum_{j=x}^m \frac{D_j}{\tilde{\lambda}j - \sqrt{j \log j}} - \frac{\tilde{\lambda} + 1}{\tilde{\lambda}} \log m \right| > Mx \right) \\
&= \mathbf{P} \left( \left| \sum_{j=x}^m \left( \frac{D_j - (\tilde{\lambda} + 1)}{\tilde{\lambda}j - \sqrt{j \log j}} + \frac{(\tilde{\lambda} + 1)}{\tilde{\lambda}j - \sqrt{j \log j}} \right) - \frac{\tilde{\lambda} + 1}{\tilde{\lambda}} \log m \right| > Mx \right) \\
&\leq \mathbf{P} \left( \left| \sum_{j=x}^m \frac{D_j - (\tilde{\lambda} + 1)}{\tilde{\lambda}j - \sqrt{j \log j}} \right| > Mx - \frac{\tilde{\lambda} + 1}{\tilde{\lambda}} \log \log \log m \right) \\
&\leq \mathbf{P} \left( \left| \sum_{j=\log \log m}^m \frac{D_j - (\tilde{\lambda} + 1)}{\tilde{\lambda}j - \sqrt{j \log j}} \right| > \tilde{M}x \right) \leq \text{Chebysev's ineq.} \\
&\leq \frac{1}{(\tilde{M}x)^2} \sum_{j=\log \log m}^m \frac{\tilde{\lambda} + 1}{(\tilde{\lambda}j - \sqrt{j \log j})^2} \leq \frac{1}{(\tilde{M}x)^2} \frac{\tilde{\lambda} + 1}{\tilde{\lambda}^2 \log \log m},
\end{aligned}$$

which is  $\leq \varepsilon/4$  for large enough  $\tilde{M}$ . We actually see that the proof works not only for  $x = \log \log m$  but for any  $x$  such that  $\log \log \log m = o(x)$ . The other term in (17) can be treated the same way. This concludes the proof of part (c).  $\square$

Let us put all these preparations in the right order to prove Theorem 3.4.

*Proof of Theorem 3.4.* We dealt with the technical issue of conditioning on survival in Construction 3.9. In part (c) of Proposition 3.10 we saw that both the conditional expectation and variance of  $G_m$  are indeed equal to  $(\tilde{\lambda} + 1)/\tilde{\lambda} \cdot \log m$ . We argued that it is enough to show that the Wasserstein distance between the standardization of  $G_m$  and that of a standard normal variable tends to zero. According to Lemma 3.8 the Wasserstein distance tends to zero if  $\sum_{j=1}^m \rho_j(1 - \rho_j) \rightarrow \infty$  as  $m \rightarrow \infty$ . But this is immediate from Proposition 3.10 part (a) and (c). Thus (5) is proved.

The almost sure convergence in (6) is basically a restatement of Theorem 3.6 part (iv).  $\square$

**Remark 3.12.** Note that Assumption 3.1 was only needed for the convergence of  $G_m$  and not for  $A_m$ .

## 4 Embedding of BP into IHRG

This section relates the exploration process of the neighborhood of a vertex in the inhomogeneous random graph (IHRG) model  $G(n, \kappa)$  of Section 2 and

the branching process of Section 3 to one another.

This is done by first showing that a multi-type BP with binomially distributed offspring can be coupled to a BP with proper Poisson offspring distribution under Assumption 3.1. As a result Theorem 3.4 also holds for the BP with binomial offspring distribution. To obtain an embedding of the BP into the IHRG, we give the vertices marks to get a continuous-time marked BP (CTMBP) on which we define thinning. This procedure basically deals with the problem of finding the shortest weight path amongst multiple possible paths between any two given vertices.

In Subsection 4.3 we analyze the connection time of the two explorations from vertices  $x$  and  $y$  and establish the relation between the hopcount and shortest weight path in the IHRG and the quantities  $G_m$  and  $A_m$  already studied in the CTMBP.

#### 4.1 Marked branching processes

Let us define a continuous-time multi-type BP with binomially distributed offspring and simultaneously recall the construction of the Poisson counterpart introduced in Section 3. Recall that the set of different types is  $\mathcal{S} = \{1, 2, \dots, r\}$  equipped with a probability measure  $\mu$  and  $n = \sum_{t \in \mathcal{S}} n_t$ , where  $n_t$  satisfies (1) for every  $t \in \mathcal{S}$ . When a particle of type  $s$  dies it gives birth to a random number of offspring  $\eta_{st}$  and  $\xi_{st}$  of type  $t$  in the binomial and Poisson BP respectively, where

$$\eta_{st} \stackrel{d}{=} \text{Bin} \left( n_t - \delta_{st}, \frac{\kappa(s, t)}{n} \right), \text{ and}$$

$$\xi_{st} \stackrel{d}{=} \text{Poi}(\lambda_{st}), \text{ where } \lambda_{st} = \kappa(s, t)\mu_t.$$

Each offspring lives for an exponentially distributed random time with rate one. Recall that  $(n_t - \delta_{st}) \cdot \kappa(s, t)/n \rightarrow \lambda_{st}$ . When a particle of type  $s$  dies its offspring distribution is a sum of independent random variables

$$D_B^{(s)} = \eta_{s1} + \eta_{s2} + \dots + \eta_{sr}, \text{ and}$$

$$D_P^{(s)} = \xi_{s1} + \xi_{s2} + \dots + \xi_{sr} \text{ respectively.}$$

Conditioned on the event that the  $j$ -th split is of a type  $s$  vertex, the distribution of the increment of the process  $D_j^B$  and  $D_j^P$  is equal to  $D_B^{(s)}$  and  $D_P^{(s)}$  respectively. For the number of alive vertices after the  $m$ -th split we write  $S_m^B = D_1^B + \dots + D_m^B - (m - 1)$  and  $S_m^P = D_1^P + \dots + D_m^P - (m - 1)$ . Due to the memoryless property of the exponential distribution, the next

vertex ( $m + 1$ -st) to split is uniformly distributed among the  $S_m^B$  and  $S_m^P$  alive vertices.

Under Assumption 3.1  $D_P^{(s)}$  is equal in distribution to a  $\text{Poi}(\tilde{\lambda} + 1)$  random variable, so in distribution this is indeed the same BP as the one defined in Section 3. We need to make a distinction here with the different types, because shortly we will give the vertices marks according to their type. But let us first show that the results of Theorem 3.4 carry through to the binomial BP.

**Proposition 4.1** (Asymptotics for binomial BP). *Consider the sequence  $D_1^B, D_2^B, \dots$ , defined above. Let  $G_m$  and  $A_m$  be defined as in Proposition 3.7. Then, conditionally on  $S_j^B > 0$  for every  $1 \leq j \leq m$ , the asymptotics for  $G_m$  and  $A_m$  in Theorem 3.4 remain to hold, where the limits are independent.*

*Proof.* It is well known that the random variables  $\eta_{st}$  and  $\xi_{st}$  can be coupled for every  $s, t \in \mathcal{S}$ , by means of a pair of variables  $(\tilde{\eta}_{st}, \tilde{\xi}_{st})$  (with  $\tilde{\eta}_{st} \sim \eta_{st}$  and  $\tilde{\xi}_{st} \sim \xi_{st}$ ) so that for  $n$  large enough ( $|n_t/n - \mu_t| \leq \varepsilon$  holds) we have

$$\mathbf{P}\left(\tilde{\eta}_{st} \neq \tilde{\xi}_{st}\right) \leq n_t \frac{(\kappa(s, t))^2}{n^2} \leq \frac{\mu_t(1 + \varepsilon)(\kappa(s, t))^2}{n} = \frac{\lambda_{st}\kappa(s, t) + o(1)}{n}.$$

Similarly we can couple  $D_B^{(s)}$  and  $D_P^{(s)}$  to obtain a uniform bound for

$$\mathbf{P}\left(\tilde{D}_B^{(s)} \neq \tilde{D}_P^{(s)}\right) \leq \sum_{t \in \mathcal{S}} \frac{\lambda_{st}\kappa(s, t) + o(1)}{n} \leq \frac{(\tilde{\lambda} + 1) \max \kappa + o(1)}{n}.$$

Hence for a single increment of the process

$$\mathbf{P}\left(\tilde{D}_j^B \neq \tilde{D}_j^P\right) = \sum_{s \in \mathcal{S}} \mathbf{P}\left(\tilde{D}_B^{(s)} \neq \tilde{D}_P^{(s)}\right) \mathbf{P}(\text{type-}s \text{ splits}) \leq \frac{\tilde{\lambda} + 1}{n} \max \kappa =: \frac{c}{n}$$

Thus for any  $m = o(n)$ , we can couple the random variables  $D_1^B, \dots, D_m^B$  and  $D_1^P, \dots, D_m^P$ , such that

$$\mathbf{P}\left(\exists j \leq m : \tilde{D}_j^B \neq \tilde{D}_j^P\right) \leq \sum_{j=1}^m \frac{c}{n} = c \frac{m}{n} \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (18)$$

Our aim is to couple  $G_m^B$  and  $G_m^P$ , the generation of the  $m$ -th split in the two processes. So, we show how to couple the Bernoulli random variables  $(I_j^B, I_j^P)$ ,  $1 \leq j \leq m$ , conditioned on  $(\tilde{D}_1^B, \tilde{D}_1^P), \dots, (\tilde{D}_m^B, \tilde{D}_m^P)$  as defined in (11). Let  $U_1, \dots, U_m$  be an i.i.d. sequence of uniform  $(0, 1)$  random variables and define the conditional probabilities

$$\mathbf{P}(I_j^B = 1 | U_j) = \mathbb{1}\left[U_j \leq \frac{\tilde{D}_j^B}{\tilde{S}_j^B}\right], \quad \mathbf{P}(I_j^P = 1 | U_j) = \mathbb{1}\left[U_j \leq \frac{\tilde{D}_j^P}{\tilde{S}_j^P}\right], \quad (19)$$



where  $\tilde{S}_j^B = \tilde{D}_1^B + \dots + \tilde{D}_j^B - (j-1)$ ,  $\tilde{S}_j^P = \tilde{D}_1^P + \dots + \tilde{D}_j^P - (j-1)$  and  $\mathbb{1}[\cdot]$  is the indicator function. Note that by the coupling in (18)

$$\mathbf{P}\left(\tilde{S}_j^B > 0, 1 \leq j \leq m\right) > 0 \text{ for every } m,$$

since the Poisson BP survives with positive probability. The joint distribution of  $(I_j^B, I_j^P)$  is given by (19), namely

$$\begin{aligned} I_j^B = I_j^P = 0, & \text{ if } U_j \leq \min(\tilde{D}_j^B/\tilde{S}_j^B, \tilde{D}_j^P/\tilde{S}_j^P), \\ I_j^B = I_j^P = 1, & \text{ if } U_j \geq \max(\tilde{D}_j^B/\tilde{S}_j^B, \tilde{D}_j^P/\tilde{S}_j^P), \\ I_j^B \neq I_j^P, & \text{ otherwise.} \end{aligned}$$

Let  $G_m^B = I_j^B + \dots + I_m^B$  and  $G_m^P = I_j^P + \dots + I_m^P$ . By the coupling of  $(I_j^B, I_j^P)$  we can see that for  $m = o(n)$

$$\begin{aligned} \mathbf{P}(G_m^B \neq G_m^P) &= \mathbf{P}\left(\sum_{j=1}^m I_j^B \neq \sum_{j=1}^m I_j^P\right) \leq \mathbf{P}(\exists j \leq m : I_j^B \neq I_j^P) \\ &\leq \mathbf{P}(\exists j \leq m : \tilde{D}_j^B \neq \tilde{D}_j^P) \stackrel{(18)}{\leq} c \frac{m}{n} \rightarrow 0. \end{aligned}$$

Combining this with Proposition 3.7 part (a) for  $G_m^P$ , it follows that the asymptotics also hold for  $G_m^B$ . We conclude that the asymptotics for  $A_m^B$  can be proved in a similar manner.  $\square$

The next step is to describe continuous-time marked branching processes (CTMBP). By assigning marks to the individuals in the BP according to their type, we will be able to relate them to the vertices during an exploration of the IHRG. Fix  $n \geq 1$  and denote the set of marks by  $[n] = \{1, 2, \dots, n\}$ . It is important to distinguish the marks according to the types, so  $[n]$  is the disjoint union of the sets of marks  $[n]^{(1)}, \dots, [n]^{(r)}$ , where there are  $n_t$  different marks in  $[n]^{(t)}$ . An individual of type  $s$  will be assigned a mark from  $[n]^{(s)}$ . We consider the same continuous-time BP with binomially distributed offspring defined earlier in the subsection.

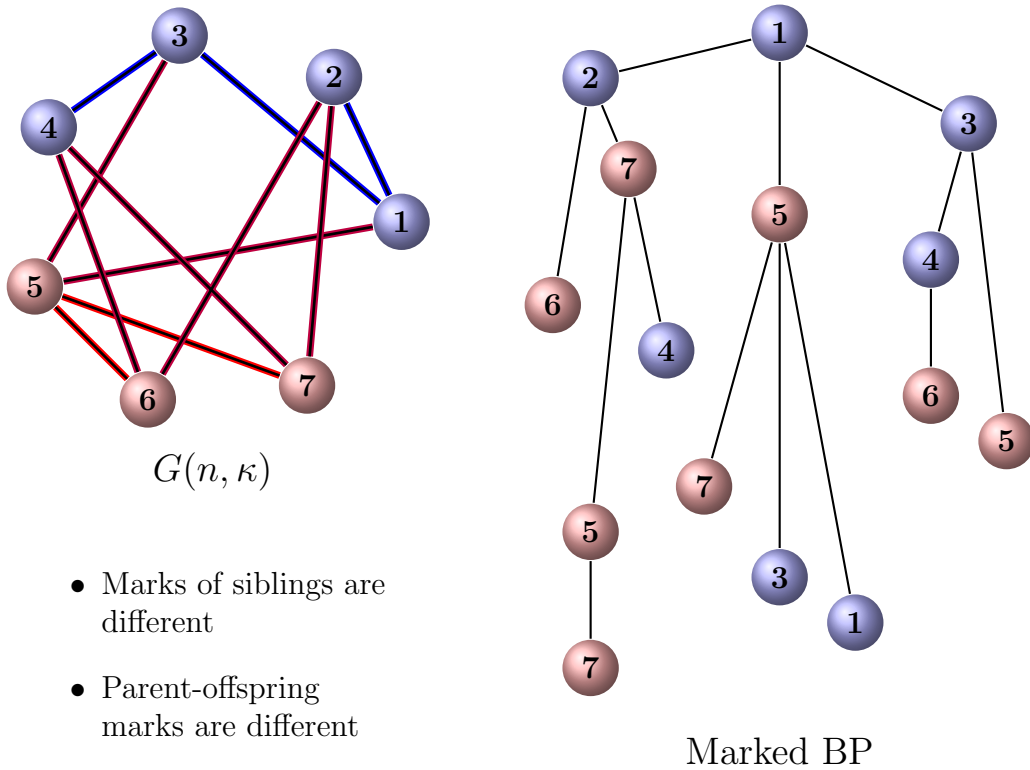
The construction goes as follows. Assume that the root of the BP is of type  $s$ . Assign the mark  $i_0 \in [n]^{(s)}$  to it. The root immediately dies and gives birth to a random number  $\eta_{st}$  of type  $t$  offspring binomially distributed with parameters  $n_t - \delta_{st}$  and  $\kappa(s, t)/n$ . The edges are given Exp(1) edge weights independently of each other. For  $t \neq s$ , the  $\eta_{st}$  "new" individuals of type  $t$  are assigned different marks from  $[n]^{(t)}$  uniformly at random. For  $t = s$ , we choose the marks the same way from  $[n]^{(s)} \setminus i_0$ . From the point of view

of the IHRG the offspring are considered as the neighbors of the vertex that corresponds to the root.

The next individual to split after  $T_1$  time units can be uniformly chosen from the  $S_1^B$  alive vertices (by the memoryless property of the exponential distribution). Assume it is of type  $\tilde{s}$  with mark  $i_1$ . It gives birth to  $\eta_{\tilde{s}t}$  offspring of type  $t$  and choose  $\eta_{\tilde{s}t}$  different marks from  $[n]^{(t)}$  for  $t \neq \tilde{s}$  and  $\eta_{\tilde{s}\tilde{s}}$  different marks from  $[n]^{(\tilde{s})} \setminus i_1$  for  $t = \tilde{s}$ . We update edge weights of the "old" vertices by subtracting  $T_1$  from them or just simply independently draw a new  $\text{Exp}(1)$  edge weight for each alive vertex.

We proceed the same way by induction. The marks of the first  $k$  splits in order are  $i_0, i_1, \dots, i_{k-1}$ . The next vertex to split is uniformly chosen among the  $S_{k-1}^B$  alive vertices at time  $A_k = T_1 + \dots + T_{k-1}$ , and denote its mark by  $i_k$ . It gives birth to offspring analogously as before and assign marks to them as described above.

For a moment let us think in terms of graphs. By assigning the marks this way, we rule out the possibility of loop edges and multiple edges between vertices during an exploration of the IHRG. There still remains the problem of picking the shortest weight path amongst multiple possible paths between any two given vertices, i.e we have to handle cycles.



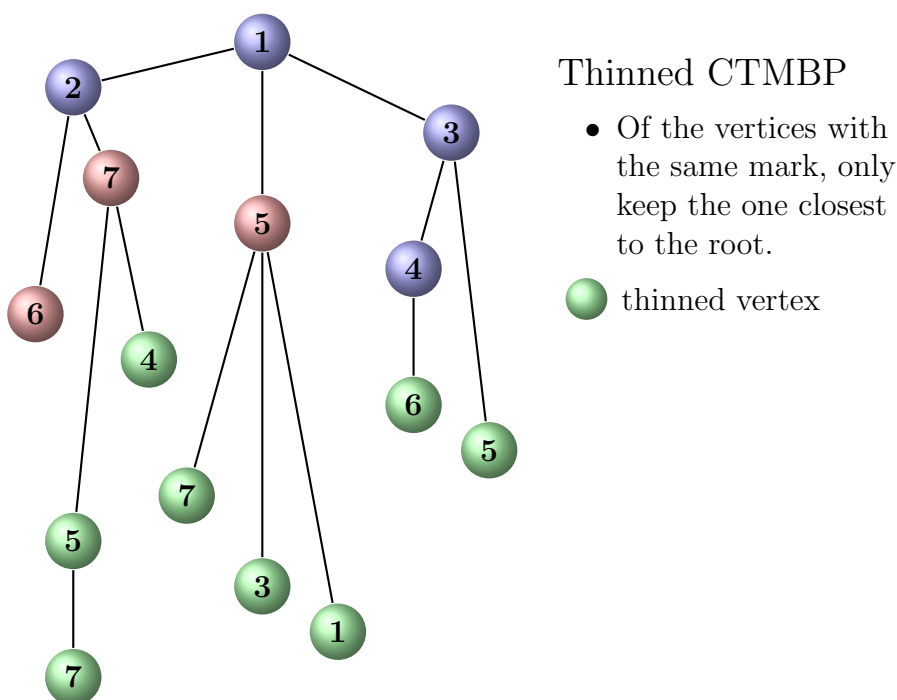
## 4.2 Thinning branching processes

We introduce thinning on branching processes to identify the shortest weight path between any two given vertices in a very natural way. Think of a cycle in a labeled graph with edge weights (assume there are no other cycles within). There are two paths between any two vertices  $x$  and  $y$  on this path, but we are only interested in the shorter one. Let fluid percolate through the graph at a constant rate and keep track of the labels that we reach. From  $x$  the fluid obviously reaches  $y$  along the shorter path first and the other flow reaches  $y$  exactly when the mark of  $y$  reappears on our list of marks already reached. Thus for any shortest weight path between  $x$  and  $y$  the second path will not be used.

In terms of the BP this means that when we reach a mark

$$i_k \in \{i_0, i_1, \dots, i_{k-1}\},$$

i.e. we found a cycle in the exploration of the IHRG, we delete  $i_k$  and the whole subtree starting from it. We call the mark  $i_k$  thinned and the marks of its subtree truncated. As a result we only keep the shortest weight paths between pairs of vertices. We refer to this procedure as thinning the BP.



For each fixed  $n$ , we choose marks from a finite set  $[n]$ , so with probability one at a certain random time all marks will have appeared. This means that the thinned CTMBP dies out, i.e. we found the minimal weight spanning tree

of a component of the IHRG. It is clear that for each  $t \geq 0$ , the set of marks reached by time  $t$  in the thinned CTMBP and the set of vertices reached by time  $t$  in the IHRG are equal in distribution. We successfully embedded the thinned CTMBP into the IHRG, so we immediately get the following:

**Lemma 4.2** (FPP on IHRG is thinned CTMBP). *For any fixed  $n \geq 1$ , consider the thinned CTMBP and IHRG as defined above. Then for any  $i_0 \in [n]$ , the weight  $W_n(i_0, j)$  of the shortest weight path between initial vertex  $i_0$  and any other vertex  $j \in [n]$  in the IHRG is equal in distribution to the weight of the shortest weight path between the root  $i_0$  and  $j$  in the thinned CTMBP. Furthermore, the hopcount between  $i_0$  and  $j$  is also equal in distribution in the IHRG and the thinned CTMBP.*

**Remark 4.3.**

- (a) *If  $i_0$  and  $j$  are not connected then we simply put the weight and hopcount equal to  $+\infty$ .*
- (b) *We did not use that the edge weights are exponentially distributed. So the lemma holds for i.i.d. edge weights with arbitrary continuous distribution supported on  $(0, \infty)$ .*

What remains is to make the intuitive picture of colliding flows of fluid precise. In the sketch of proof we mentioned that we cannot follow the progress of a flow continuously in time. Rather, we will look at the picture at each split time  $A_k$ , which means that we let the fluid percolate on each edge of the CTMBP or equivalently in the exploration process of the IHRG up until time  $A_k$ .

For this we formally introduce the notion of shortest-weight trees  $SWT_k$ , for  $k \geq 1$ . Immediately after time  $A_k$ , the number of individuals that the percolating fluid already reached, the wet individuals (i.e. the ones that split) is  $k + 1$ , whose marks we denoted by  $i_0, i_1, \dots, i_k$ . Let us denote this set by  $M_k^-$  for all  $k \geq 1$ . We also keep track of the marks of the alive vertices,  $M_k^+$ , i.e. the individuals who are the offspring of a wet vertex but the fluid has not reached them yet. Clearly  $|M_k^+| = S_k$ . Let  $SWT_k$  be the collection of the wet and alive marks together with the split times  $A_0, \dots, A_k$ . So  $SWT_0 = (\{i_0\}, A_0 = 0)$ , and

$$SWT_k = (M_k^-, M_k^+, \{A_0, A_1, \dots, A_k\}), \quad k \geq 1. \quad (20)$$

The CTMBP can be uniquely reconstructed from the sequence  $(SWT_k)_{k=1}^\infty$ . Note that  $SWT_k$  contains all the marks in the CTMBP, also the thinned and truncated marks and possibly multiple marks among alive vertices. Later we

will need an upper bound on the expected number of thinned marks (i.e. the number of cycles we find in the exploration process up to and including time  $A_k$ ) together with the truncated marks. Here we give an alternative proof of [3, Lemma 4.5], since we believe the proof in the form presented there requires some correction.

**Lemma 4.4** (Expected number of thinned and truncated marks). *Fix  $k \geq 1$  and denote by  $MT_k^-$  the number of thinned and truncated marks in  $M_k^-$ , and similarly  $AT_k$  stands for the truncated alive marks in  $M_k^+$ . Then*

$$\mathbf{E} (MT_k^- + AT_k) \leq \frac{((\tilde{\lambda} + 1)k + \tilde{\lambda})^2}{\tilde{\lambda}n}. \quad (21)$$

*Proof.* We calculate the size of the subtrees of thinned vertices such that a subtree includes all the wet and alive vertices. Note that this way we might count vertices on multiple occasions, but this upper bound is still sufficient for us. At time  $A_j$  the vertex with mark  $i_j$  splits. Denote the subtree, including wet and alive vertices, of  $i_j$  by  $F_j$ . Then

$$\mathbf{E} (MT_k^- + AT_k) \leq \sum_{j=0}^k \mathbf{E} (|F_j| \cdot \mathbb{1} [i_j \text{ is thinned}]).$$

Recall that  $N_k^{(t)}$  denotes the number of splits of type  $t$  vertices among the first  $k$  splits and there are  $n_t$  different marks that correspond to vertices of type  $t$ . Next we estimate the probability of the event that  $i_j$  is thinned:

$$\begin{aligned} \mathbf{P} (i_j \text{ is thinned}) &= \sum_{t \in \mathcal{S}} \mathbf{P} (i_j \text{ is thinned} \mid \text{type } t) \mathbf{P} (\text{type } t) \\ &= \sum_{t \in \mathcal{S}} \frac{N_j^{(t)}}{n_t} \cdot \frac{S_j^{(t)}}{S_j} \stackrel{(*)}{=} \sum_{t \in \mathcal{S}} \frac{\pi_t^2}{\mu_t} \cdot \frac{j}{n} = \frac{j}{n}, \end{aligned} \quad (22)$$

where for  $(*)$  we used (1), (8) and (10). By [10, Theorems 3.22-24.] it is easy to check that the error term is negligible (after summing over  $j$ ). The last equality is a consequence of Lemma 4.5.

Next step is to determine  $|F_j|$ . By the memoryless property of the process, it is easy to see that the proportion of the subtree of  $i_j$  compared to the size of the BP satisfies

$$\frac{|F_j|}{|F| - j} = \frac{D_{j+1}}{S_{j+1}}, \quad (23)$$

where  $|F| = S_{k+1} + k$  is the number of vertices in the entire BP including wet and alive vertices.

Since  $|F_j|$  and the event  $i_j$  is thinned are independent, we find that putting all of these together with the help of Proposition 3.10 part (b):

$$\begin{aligned}
\mathbf{E}(MT_k^- + AT_k) &\leq \sum_{j=0}^k \mathbf{E}(|F_j|) \mathbf{P}(i_j \text{ is thinned}) \stackrel{(22)}{=} \sum_{j=0}^k \mathbf{E}(|F_j|) \frac{j}{n} \\
&\stackrel{(23)}{\leq} \sum_{j=0}^k \mathbf{E}\left(\frac{D_{j+1}}{S_{j+1}}(S_{k+1} + k)\right) \frac{j}{n} \\
&\leq \sum_{j=1}^k \frac{\mathbf{E}(D_{j+1})}{\tilde{\lambda}(j+1)} \cdot \frac{j+1}{n} (S_{k+1} + k) \\
&\leq \frac{S_{k+1} + k}{\tilde{\lambda}n} \mathbf{E}\left(\sum_{j=1}^k D_{j+1}\right) \leq \frac{((\tilde{\lambda} + 1)k + \tilde{\lambda})^2}{\tilde{\lambda}n} = O\left(\frac{k^2}{n}\right),
\end{aligned}$$

where  $\mathbf{E}\left(\sum_{j=1}^k D_{j+1}\right) = \tilde{\lambda}(k+1) + k$ . □

**Lemma 4.5.** *Under Assumption 3.1 we find that*

$$\pi = \mu,$$

where  $\pi$  is the left eigenvector corresponding to the eigenvalue  $\tilde{\lambda}$  of  $A$  and  $\mu$  is the probability distribution given on  $\mathcal{S}$ .

*Proof of Lemma 4.5.* It is enough to prove that  $\mu A = \tilde{\lambda}\mu$ . Recall that  $A_{st} = \lambda_{st} - \delta_{st}$ , where  $\lambda_{st} = \kappa(s, t)\mu_t$  and  $\kappa$  is a symmetric matrix. Writing out  $(\mu A)_t$ , we find

$$\begin{aligned}
(\mu A)_t &= \sum_{s \in \mathcal{S}} \mu_s A_{st} = \sum_{s \in \mathcal{S}} \mu_s (\kappa(s, t)\mu_t - \delta_{st}) \\
&= \mu_t \sum_{s \in \mathcal{S}} \mu_s \kappa(t, s) - \mu_t = \mu_t \sum_{s \in \mathcal{S}} \lambda_{st} - \delta_{st} = \tilde{\lambda}\mu_t.
\end{aligned}$$

□

We show a similar result to the assertion of Lemma 4.4 for the number of multiple marks among  $M_k^+$ , these are the cycles that will be found in the exploration process after step  $k$ . Let  $MT_k^+$  denote the number of multiple marks in  $M_k^+$ . Instead of proving for general  $k$ , pick  $k = \sqrt{n}$ . We will see in the next subsection why this special case is important.

**Lemma 4.6.** *Let  $a_n = \lceil \sqrt{n} \rceil$ . Then  $MT_{a_n}^+ = O_{\mathbf{P}}(1)$ , i.e. the sequence  $MT_{a_n}^+$  is tight.*

*Proof.* When assigning marks to new vertices of type  $t$  we prescribed some constraints on the set  $[n]^{(t)}$  from which we choose its mark. We can dominate this by repeatedly choosing from  $[n]^{(t)}$  without any constraints. From (7) we know that  $S_{a_n}^{(t)} = \tilde{\lambda}\pi_t a_n(1 + o(1))$ . So let us pick  $\tilde{\lambda}\pi_t a_n(1 + o(1))$  marks from  $[n]^{(t)}$ , always putting back the marks we choose. For  $i \in [n]^{(t)}$  let  $X_i$  be the number of times  $i$  was chosen, thus  $X_i \sim \text{Bin}(\tilde{\lambda}\pi_t a_n(1 + o(1)), 1/n_t)$ . The probability that  $i$  is chosen at least twice is

$$\mathbf{P}(X_i \geq 2) = 1 - \mathbf{P}(X_i = 0) - \mathbf{P}(X_i = 1) = \frac{\tilde{\lambda}^2 \pi_t^2 a_n^2}{2 n_t^2} + O\left(\frac{a_n^3}{n_t^3}\right).$$

Thus for the expected number of multiple marks

$$\mathbf{E}\left(\sum_{i \in [n]^{(t)}} \mathbb{1}[X_i \geq 2]\right) = \sum_{i \in [n]^{(t)}} \mathbf{P}(X_i \geq 2) = \frac{\tilde{\lambda}^2 \pi_t^2}{2 \mu_t} + o(1),$$

independent of  $a_n$ . Thus the expected number of multiple marks is

$$\mathbf{E}(MT_{a_n}^+) = \sum_t \frac{\tilde{\lambda}^2 \pi_t^2}{2 \mu_t} + o(1) = \frac{\tilde{\lambda}^2}{2} + o(1),$$

where Lemma 4.5 was also used. So from a Markov inequality we find that for  $\varepsilon > 0$

$$\mathbf{P}(MT_{a_n}^+ > K) \leq \frac{\tilde{\lambda}^2}{2K} < \varepsilon, \quad \text{if } K = K(\varepsilon) > \frac{\tilde{\lambda}^2}{2\varepsilon}.$$

□

### 4.3 Connection time

It was already mentioned in the sketch of proof that to determine the shortest-weight path between two vertices  $x$  and  $y$ , we let fluid percolate from each one and stop when the two flows collide. For technical reasons we don't let the fluids flow simultaneously from both vertices, rather we let the fluid flow from  $x$  until it reaches some  $a_n$  vertices and then start a flow from  $y$  until the random time  $C_n$ , the connection time, when the two flows collide. Later we will see what the correct choice for  $a_n$  is. To be able to make this precise, we define the connection with the help of the shortest-weight trees defined in (20).

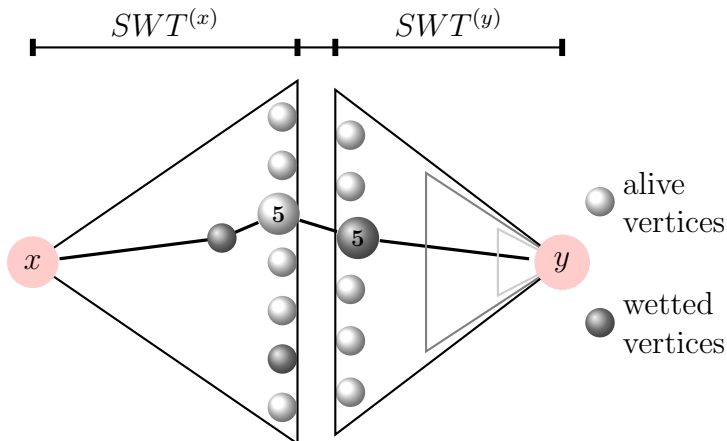
Recall that  $SWT_k^{(x)}$  after the split time  $A_k$  is the collection of wet marks  $M_k^{(x)-}$ , alive marks  $M_k^{(x)+}$  and the split times  $A_0^{(x)}, A_1^{(x)}, \dots, A_k^{(x)}$ . We let the

flow emanating from  $x$  grow for  $a_n$  steps, after which we have wet marks  $M_{a_n}^{(x)-}$  and alive marks  $M_{a_n}^+(x)$ . Then, we start another marked branching process from root  $y$ . Since the flow from  $y$  can only connect to the flow from  $x$  via an alive vertex, a connection is possible only when a mark from  $M_{a_n}^{(x)+}$  appears among the marks in  $M_k^{(y)-}$ . So we have to make a slight modification in the way we assign marks to the vertices in the BP from  $y$ . Everything remains the same except that from the start we leave out the marks  $M_{a_n}^{(x)-}$  from the possible marks  $[n]$ .

From here we define the connection time

$$C_n = \min\{k \geq 0 : M_{a_n}^{(x)+} \cap M_k^{(y)-} \neq \emptyset\}. \quad (24)$$

Note that the final edge still has to be formed. Normally, this would imply that when determining the distribution of the shortest-weight path we would have to add an extra  $\text{Exp}(1)$  random variable for the last edge. However, with some reasoning we will see that the distribution of this last edge is actually the same as letting the fluid flow from  $x$  until the next split. The figure below illustrates the connection time.



Connection time

First we present a series of lemmas and we give proofs afterwards. The next lemma describes the asymptotic distribution of the connection time.

**Lemma 4.7** (Weak convergence of connection time). *Conditioned on the event that both CTMBPs survive, the connection time  $C_n$  satisfies the asymptotics*

$$\frac{C_n a_n}{n} \xrightarrow{d} E, \quad (25)$$

where  $E$  has an exponential distribution with rate  $\tilde{\lambda} \sum_{s \in \mathcal{S}} \pi_s^2 / \mu_s$ , which simplifies to rate  $\tilde{\lambda}$  under Assumption 3.1.



This result supports that the intuitive picture of letting the fluids flow simultaneously or delaying one of them indeed makes no difference in terms of finding the shortest path. However, our next set of arguments work only for a special choice of  $a_n$ .

We want to show that when the connection happens, the probability that a type  $t$  vertex splits in  $SWT^{(y)}$  is the same as the probability that the next vertex to split in  $SWT^{(x)}$  is of type  $t$ . So choosing the mark from  $M_{a_n}^{(x)+}$  at the connection is the same as choosing the next vertex to split in  $SWT^{(x)}$  from the same set. This would allow us to say that the last edge forms simply by letting  $SWT^{(x)}$  flow until the next split  $a_n + 1$ .

Let  $a_n = \lceil \sqrt{n} \rceil$ . Then in Lemma 4.4 and 4.6 we proved that the number of truncated and multiple marks in  $M_{a_n}^{(x)+}$  is of constant order. Thus, in  $SWT^{(x)}$  the mark of the next vertex to split is uniform among the alive marks  $M^{(x)+}$ , where  $|M^{(x)+}| = S$ . The ratio of the alive vertices of type  $t$ ,  $S^{(t)}/S$  tends to  $\pi_t$  (see (10)), which means that the probability that the next vertex to split is of type  $t$  is approximately  $\pi_t$ . For any other choice of  $a_n$  this is not the case. The next lemma shows that when the connection happens the probability that a type  $t$  vertex splits in  $SWT^{(y)}$  is not  $\pi_t$  in general, but under Assumption 3.1 the two probabilities are equal. We believe that the corresponding arguments in [3] require some correction, so the proof given here is new and uses other methods.

**Lemma 4.8.** *In the limit  $n \rightarrow \infty$ , the probability that, when the connection happens, the vertex to split is of type  $t$  satisfies*

$$\mathbf{P}(\text{type } t \text{ splits} \mid \text{connection}) \rightarrow \frac{\pi_t^2 / \mu_t}{\sum_{s \in S} \pi_s^2 / \mu_s}.$$

*Proof of Lemma 4.7.* Let us first show the proof neglecting error terms and then we will make the arguments rigorous. For the probability  $\mathbf{P}(C_n > m)$  we can write

$$\begin{aligned} \mathbf{P}(C_n > m) &= \prod_{j=1}^m \mathbf{P}(C_n > j \mid C_n > j-1) \\ &= \prod_{j=1}^m \mathbf{E}(\mathbf{P}(C_n > i \mid C_n > j-1, \mathcal{F}_{j-1})), \end{aligned} \tag{26}$$

where  $\mathcal{F}_{j-1}$  is the  $\sigma$ -algebra generated by  $SWT_{a_n}^{(x)}$  and  $SWT_{j-1}^{(y)}$ . To calculate  $\mathbf{P}(C_n > j \mid C_n > j-1, \mathcal{F}_{j-1})$  we have to sum over the types in  $SWT^{(y)}$  and find the probability that it does not connect to an alive vertex in  $SWT^{(x)}$  of

the same type. Formally,

$$\mathbf{P}(C_n > j \mid C_n > j-1, \mathcal{F}_{j-1}) = \sum_{t \in \mathcal{S}} \frac{S_{j-1}^{(y)(t)}}{S_{j-1}^{(y)}} \left( 1 - \frac{S_{a_n}^{(x)(t)}}{n_t} \right).$$

Plugging this into (26) with  $m = \frac{nx}{a_n}$  ( $x > 0$ ) yields

$$\begin{aligned} & \mathbf{P}(C_n > xn/a_n) \\ &= \prod_{j=1}^{xn/a_n} \left[ \sum_{t \in \mathcal{S}} \frac{S_{j-1}^{(y)(t)}}{S_{j-1}^{(y)}} \left( 1 - \frac{S_{a_n}^{(x)(t)}}{n_t} \right) \right] = \prod_{j=1}^{xn/a_n} \left[ 1 - \sum_{t \in \mathcal{S}} \frac{S_{j-1}^{(y)(t)}}{S_{j-1}^{(y)}} \cdot \frac{S_{a_n}^{(x)(t)}}{n_t} \right] \\ &\stackrel{(7)}{\approx} \prod_{j=1}^{xn/a_n} \left[ 1 - \sum_{t \in \mathcal{S}} \frac{S_{j-1}^{(y)(t)}}{S_{j-1}^{(y)}} \cdot \frac{\tilde{\lambda} \pi_t a_n}{n_t} \right] \\ &\stackrel{(10)}{\approx} \prod_{j=1}^{xn/a_n} \left[ 1 - \sum_{t \in \mathcal{S}} \frac{a_n}{n} \cdot \frac{\tilde{\lambda} \pi_t^2}{n_t/n} \right] \stackrel{(1)}{\approx} \prod_{j=1}^{xn/a_n} \left[ 1 - \frac{a_n}{n} \sum_{t \in \mathcal{S}} \frac{\tilde{\lambda} \pi_t^2}{\mu_t} \right] \\ &= \exp \left( -\frac{a_n}{n} \tilde{\lambda} \sum_{t \in \mathcal{S}} \frac{\pi_t^2}{\mu_t} \frac{xn}{a_n} + o(1) \right) \rightarrow \exp \left( -\tilde{\lambda} \sum_{t \in \mathcal{S}} \frac{\pi_t^2}{\mu_t} \cdot x \right). \end{aligned}$$

In order to have short formulas we neglected the error terms along the lines but using [10, Theorems 3.22-24.] it is easy to see that the dominating error terms are of order  $n^{Re(\lambda_2)/\tilde{\lambda}-1}$  or  $n^{-3/2}$  depending on whether  $Re(\lambda_2) > \frac{1}{2}\tilde{\lambda}$  or  $Re(\lambda_2) \geq \frac{1}{2}\tilde{\lambda}$ , where  $\lambda_2$  denotes the eigenvalue of  $A$  with the second largest real part. Lemma 4.5 implies that under Assumption 3.1 we have  $\sum_{t \in \mathcal{S}} \pi_t^2/\mu_t = 1$ .  $\square$

**Remark 4.9.** *At this point we can see that the choice  $a_n = \lceil \sqrt{n} \rceil$  implies that the connection from  $y$  happens after exploring vertices of the same order. More precisely, the connection happens after exploring  $\sqrt{n} \cdot E$  vertices, where  $E \stackrel{d}{=} \text{Exp}(\tilde{\lambda})$ . Roughly, this is the same as starting the two flows of fluid at the same time.*

*Proof of Lemma 4.8.* We neglect the error terms along the lines again but note that the very same estimates work as in the proof of Lemma 4.7.

Let us introduce the events  $C := \{\text{connection happens}\}$  and  $t \dagger$  be the event that a type  $t$  vertex splits in  $SWT^{(y)}$  when the connection happens. To form the connection, a vertex of type  $t$  splits among the alive vertices in  $SWT^{(y)}$  and connects to an alive vertex of type  $t$  in  $SWT^{(x)}$  and we have to

sum over the different types. Thus

$$\begin{aligned} \mathbf{P}(C) &= \sum_{t \in \mathcal{S}} \mathbf{P}(C | t \dagger) \mathbf{P}(t \dagger) \\ &= \sum_{t \in \mathcal{S}} \frac{S_{a_n}^{(x)(t)}}{n_t} \cdot \frac{S_{C_n}^{(y)(t)}}{S_{C_n}^{(y)}} = \tilde{\lambda} \sum_{t \in \mathcal{S}} \frac{\pi_t^2}{\mu_t} \cdot \frac{a_n}{n} (1 + o(1)). \end{aligned}$$

Using a simple Bayes' theorem we conclude that

$$\begin{aligned} \mathbf{P}(t \dagger | C) &= \frac{\mathbf{P}(C | t \dagger) \mathbf{P}(t \dagger)}{\mathbf{P}(C)} = \frac{S_{a_n}^{(x)(t)}/n_t \cdot S_{C_n}^{(y)(t)}/S_{C_n}^{(y)}}{\sum_{s \in \mathcal{S}} S_{a_n}^{(x)(s)}/n_s \cdot S_{C_n}^{(y)(s)}/S_{C_n}^{(y)}} \\ &= \frac{\tilde{\lambda} \cdot \pi_t^2/\mu_t \cdot a_n/n}{\tilde{\lambda} \sum_{s \in \mathcal{S}} \pi_s^2/\mu_s \cdot a_n/n} (1 + o(1)) \rightarrow \frac{\pi_t^2/\mu_t}{\sum_{s \in \mathcal{S}} \pi_s^2/\mu_s}. \end{aligned}$$

□

We are now ready to prove the main theorem of this section, which makes the idea outlined in the sketch of proof precise.

**Theorem 4.10** (Connecting hopcount to heights in CTMBP). *Let  $H_n$  and  $W_n$  denote the hopcount and the weight, respectively, of the minimal path between vertices  $x$  and  $y$  in the IHRG  $G(\kappa, n)$ , where we condition on the vertices  $x$  and  $y$  to be in the giant component.*

(a) For  $n \rightarrow \infty$ , whp,

$$H_n \stackrel{d}{=} G_{a_n+1}^{(x)} + G_{C_n}^{(y)}, \quad (27)$$

(b)

$$W_n \stackrel{d}{=} A_{a_n+1}^{(x)} + A_{C_n}^{(y)}, \quad (28)$$

*Proof.* In  $SWT^{(x)}$  the mark to split at time  $A_{a_n}^{(x)}$  is  $i_{a_n}^{(x)}$ , while in  $SWT^{(y)}$  the mark to split at  $A_{C_n}^{(y)}$  is  $i_{C_n}^{(y)}$ . This is the same mark that is alive in  $SWT^{(x)}$  at  $A_{a_n}^{(x)}$ , to make a distinction denote it by  $i_{C_n}^{(x)}$ .

Since  $x$  and  $y$  are connected, whp,  $x$  and  $y$  are contained in the giant component of the IHRG, which implies that both CTMBPs do not die out. The main contribution of Lemma 4.8 and the argument preceding it is that when the connection happens we choose uniformly from the alive vertices in  $SWT^{(x)}$  which in distribution is the same as  $G_{a_n+1}$  and the split time in distribution is  $A_{a_n+1}$ .

From Lemma 4.4 and 4.6 it follows that the probability that the shortest weight path from  $x$  to  $i_{C_n}^{(x)}$  contains a truncated or multiple mark is at most

$$\frac{\mathbf{E}(AT_{a_n}) + MT_{a_n}^+}{a_n} \leq \left( \frac{((\tilde{\lambda} + 1)a_n + \tilde{\lambda})^2}{\tilde{\lambda}n} + O_{\mathbf{P}}(1) \right) \frac{1}{a_n} \rightarrow 0.$$

This shows that whp  $G_{a_n+1}$  gives the hopcount between  $x$  and  $i_{C_n}^{(x)}$ . Similar argument shows that, whp, the hopcount between  $y$  and  $i_{C_n}^{(y)}$  is equal in distribution to  $G_{C_n}^{(y)}$ . This establishes (27).

The argument for part (b) is simple. The minimal weight path is the route  $x \rightarrow i_{C_n}^{(x)} \rightarrow i_{C_n}^{(y)} \rightarrow y$ . The weight on this path is

$$W_n \stackrel{d}{=} W(x \rightarrow i_{C_n}^{(x)}) + W(i_{C_n}^{(x)} \rightarrow i_{C_n}^{(y)}) + W(i_{C_n}^{(y)} \rightarrow y) \stackrel{d}{=} A_{a_n+1}^{(x)} + A_{C_n}^{(y)},$$

where we can simply reallocate the weight  $A_{a_n}^{(x)} \stackrel{d}{=} W(x \rightarrow i_{a_n}^{(x)})$  over to  $x \rightarrow i_{C_n}^{(x)}$  due to our construction and the weight of the last edge  $W(i_{C_n}^{(x)} \rightarrow i_{C_n}^{(y)})$  can be merged into  $A_{a_n+1}^{(x)}$ .  $\square$

## 5 Proof of main results

In this section we prove our main results stated in Section 2.

*Proof of Theorem 2.3.* Observe that, conditioned on  $C_n$  the random variables  $G_{a_n}^{(x)}$  and  $G_{C_n}^{(y)}$  are independent. Introduce the abbreviation  $\beta := (\tilde{\lambda} + 1)/\tilde{\lambda}$  and define the random variables

$$Z_n^{(x)} := \frac{G_{a_n}^{(x)} - \beta \log a_n}{\sqrt{\beta \log a_n}}, \quad Z_{C_n}^{(y)} := \frac{G_{C_n}^{(y)} - \beta \log C_n}{\sqrt{\beta \log C_n}}$$

and denote by  $\xi_1, \xi_2$  two independent standard normal random variables. By Proposition 4.1 we know that  $Z_n^{(x)}$  converges in distribution to  $\xi_1$  and  $Z_{C_n}^{(y)}$  converges to  $\xi_2$  conditioning on the event that  $\{C_n = m_n\}$ , where  $m_n = o(n)$ . Formulating this with the Wasserstein distance (introduced in Subsection 3.2), we find that for any bounded continuous function  $g$  of two real variables

$$\lim_{n \rightarrow \infty} \mathbf{E} \left( g(Z_n^{(x)}, Z_{C_n}^{(y)}) \mid C_n = m_n \right) = \mathbf{E} (g(\xi_1, \xi_2)). \quad (29)$$

Hence by bounded convergence and using (25) and (29),

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{E} \left( g(Z_n^{(x)}, Z_{C_n}^{(y)}) \right) &= \int_0^\infty \lim_{n \rightarrow \infty} \mathbf{E} \left( g(Z_n^{(x)}, Z_{C_n}^{(y)}) \mid C_n = \lceil a_n y \rceil \right) d\mathbf{P}(C_n \leq \lceil a_n y \rceil) \\ &= \int_0^\infty \mathbf{E} (g(\xi_1, \xi_2)) e^{-y} dy = \mathbf{E} (g(\xi_1, \xi_2)), \end{aligned}$$

which shows the joint weak convergence of the pair  $(Z_n^{(x)}, Z_{C_n}^{(y)})$  to  $(\xi_1, \xi_2)$ . By the continuous mapping theorem [5, Theorem 5.1] applied to  $(x, y) \mapsto x + y$  and by Theorem 4.10, we find that

$$\begin{aligned} \frac{H_n - \beta \log n}{\sqrt{\beta \log n}} &= \frac{G_{a_n}^{(x)} - (\beta/2) \log n}{\sqrt{\beta \log n}} + \frac{G_{C_n}^{(y)} - (\beta/2) \log n}{\sqrt{\beta \log n}} \\ &= \frac{G_{a_n}^{(x)} - \beta \log a_n}{\sqrt{\beta \log a_n}} \cdot \frac{\sqrt{\beta \log a_n}}{\sqrt{\beta \log n}} + \frac{G_{C_n}^{(y)} - \beta \log C_n}{\sqrt{\beta \log C_n}} \cdot \frac{\sqrt{\beta \log C_n}}{\sqrt{\beta \log n}} \\ &= \underbrace{Z_n^{(x)} \cdot \sqrt{\frac{\log a_n}{\log n}}}_{\rightarrow 1/\sqrt{2}} + \underbrace{Z_{C_n}^{(y)} \cdot \sqrt{\frac{\log C_n}{\log n}}}_{\rightarrow 1/\sqrt{2}} \xrightarrow{d} \xi_1/\sqrt{2} + \xi_2/\sqrt{2}, \end{aligned}$$

and we note that  $(\xi_1 + \xi_2)/\sqrt{2}$  is again standard normal. When switching from  $(\beta/2) \log n$  to  $\beta \log C_n$ , we suppressed the term  $(\beta \log E)/\sqrt{\log n}$  (where  $E \stackrel{d}{=} \text{Exp}(\tilde{\lambda})$ ), since it is  $o_{\mathbf{P}}(1)$ . Note that the limit is independent of  $C_n$  as well.  $\square$

*Proof of Theorem 2.4.* Similarly to the proof of Theorem 2.3, by Proposition 4.1 for any sequence  $m_n \rightarrow \infty$ ,

$$\left( A_{a_n}^{(x)} - \frac{1}{\tilde{\lambda}} \log a_n, A_{C_n}^{(y)} - \frac{1}{\tilde{\lambda}} \log C_n \mid C_n = m_n \right) \xrightarrow{d} (X_1, X_2), \quad (30)$$

where  $X_1$  and  $X_2$  are two independent copies of  $X = -\frac{1}{\tilde{\lambda}} \log \left( \frac{1}{\tilde{\lambda}} \hat{W} \right)$ . Combining Theorem 4.10, Lemma 4.8 with (30) and the continuous mapping theorem, we obtain

$$W_n - \frac{1}{\tilde{\lambda}} \log n = \underbrace{A_{a_n}^{(x)} - \frac{1}{\tilde{\lambda}} \log a_n}_{\xrightarrow{d} X_1} + \underbrace{A_{C_n}^{(y)} - \frac{1}{\tilde{\lambda}} \log C_n}_{\xrightarrow{d} X_2} + \underbrace{\frac{1}{\tilde{\lambda}} \log \frac{C_n}{a_n}}_{\xrightarrow{d} 1/\tilde{\lambda} \log E},$$

where  $E \stackrel{d}{=} \text{Exp}(\tilde{\lambda})$  is independent of  $(X_1, X_2)$ . Thus the distribution of  $W$  in Theorem 2.4 is the sum  $X_1 + X_2 - 1/\tilde{\lambda} Y$ , where  $Y$  has a Gumbel distribution with distribution function  $\exp(-e^{-\tilde{\lambda}x})$ .  $\square$

*Proof of Corollary 2.6.* It is immediate from the two previous proofs with

$$\frac{\tilde{\lambda}_n + 1}{\tilde{\lambda}_n} \rightarrow \frac{\tilde{\lambda} + 1}{\tilde{\lambda}} \quad \text{and} \quad \frac{1}{\tilde{\lambda}_n} \rightarrow \frac{1}{\tilde{\lambda}}, \quad \text{as } n \rightarrow \infty.$$

$\square$

*Proof of Theorem 2.7.* In the dense setting, where  $\tilde{\lambda} \rightarrow \infty$ , we have a sequence of kernels  $\kappa_n(s, t)$ ,  $n = 1, 2, \dots$  and  $s, t \in \mathcal{S}$ . The type  $t$  neighbors of a type  $s$  vertex have distribution  $\eta_{st}^{(n)} \stackrel{d}{=} \text{Bin}(n_t - \delta_{st}, \kappa_n(s, t)/n)$ . We avoid the coupling with Poisson random variables in Proposition 4.1 by directly giving a proof for the CTMBP where the offspring distribution  $(D_i^{(n)} | \text{type } s \text{ splits})$  is the sum of independent binomial random variables  $\eta_{st}^{(n)}$ ,  $t \in \mathcal{S}$  (for details see Subsection 4.1).

To prove part (a), concerning the hopcount on the shortest-weight path, it is enough to show that (5) holds with  $(\tilde{\lambda} + 1)/\tilde{\lambda}$  replaced by  $(\tilde{\lambda}_n + 1)/\tilde{\lambda}_n$ , since from here we can argue the same way as we did in the sparse setting. Let  $\xi^{(1)}, \dots, \xi^{(r)}, \xi_1, \xi_2, \dots$  denote independent standard normal random variables. By the CLT we get that

$$(D_i^{(n)} | \text{type } s \text{ splits}) \xrightarrow{d} \sum_{t \in \mathcal{S}} \left( \lambda_{st}^{(n)} + \sqrt{\lambda_{st}^{(n)}} \cdot \xi^{(t)} \right) \stackrel{d}{=} \tilde{\lambda}_n + 1 + \sqrt{\tilde{\lambda}_n + 1} \cdot \xi_i.$$

From here, with the usual notation  $S_i^{(n)} = \sum_{j=1}^i D_j^{(n)} - (i - 1)$ :

$$\begin{aligned} D_i^{(n)} &\xrightarrow{d} \tilde{\lambda}_n + 1 + \sqrt{\tilde{\lambda}_n + 1} \cdot \xi_i, \\ S_i^{(n)} &\xrightarrow{d} i\tilde{\lambda}_n + 1 + \sqrt{i(\tilde{\lambda}_n + 1)} \cdot \tilde{\xi}_i. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{D_i^{(n)}}{S_i^{(n)}} &= \frac{\tilde{\lambda}_n + 1 + \sqrt{\tilde{\lambda}_n + 1} \cdot \xi_i}{i\tilde{\lambda}_n + 1 + \sqrt{i(\tilde{\lambda}_n + 1)} \cdot \tilde{\xi}_i} \\ &= \frac{\tilde{\lambda}_n + 1}{i\tilde{\lambda}_n} \cdot \frac{1 + (\tilde{\lambda}_n + 1)^{-1/2} \cdot \xi_i}{1 + (i\tilde{\lambda}_n)^{-1} + \sqrt{\frac{\tilde{\lambda}_n + 1}{i\tilde{\lambda}_n}} \cdot \tilde{\xi}_i} = \frac{\tilde{\lambda}_n + 1}{i\tilde{\lambda}_n} \cdot \left( 1 + O\left(\frac{1}{\sqrt{\tilde{\lambda}_n + 1}}\right) \right). \end{aligned} \tag{31}$$

Proposition 3.7 states that

$$G_m \stackrel{d}{=} \sum_{i=1}^m I_i, \quad \text{where } \mathbf{P}(I_i = 1 | D_1^{(n)}, \dots, D_m^{(n)}) = D_i^{(n)} / S_i^{(n)},$$

combining this with (31), we find that  $G_m$  satisfies the same asymptotics as

$$\sum_{i=1}^m \hat{I}_i^{(n)}, \quad \text{where } \mathbf{P}(\hat{I}_i^{(n)} = 1) = \frac{\tilde{\lambda}_n + 1}{i\tilde{\lambda}_n},$$

with  $I_1^{(n)}, I_2^{(n)}, \dots, I_m^{(n)}$  independent. Applying Lemma 3.8, we conclude that for any  $m = m_n \rightarrow \infty$ ,

$$\frac{G_m - \beta_n \sum_{i=1}^m 1/i}{\sqrt{\sum_{i=1}^m \beta_n/i (1 - \beta_n/i)}} \xrightarrow{d} Z,$$

where  $Z$  is standard normal and  $\beta_n = (\tilde{\lambda}_n + 1)/\tilde{\lambda}_n$ . Note that the denominator can be replaced with  $\sqrt{\log m}$ , since

$$\sum_{i=1}^m \frac{\beta_n}{i} \cdot \left(1 - \frac{\beta_n}{i}\right) = \beta_n \log m + O(1).$$

The centering constant  $\beta_n \sum_{i=1}^m 1/i$  can be replaced by  $\beta_n \log m$ , which essentially completes the proof of part (a). We can further argue that the centering constant can be replaced by  $\log m$  if and only if  $(\beta_n - 1)\sqrt{\log m} \rightarrow 0$ , or equivalently  $\sqrt{\log n} = o(\tilde{\lambda}_n)$ .

Let us continue with the proof of part (b), concerning the minimal weight. We again use that  $S_i^{(n)}$  concentrates around  $i\tilde{\lambda}_n + 1$ . Proposition 3.7 part (b) states that

$$A_m \stackrel{d}{=} \sum_{i=1}^m E_i/S_i^{(n)} \approx \frac{1}{\tilde{\lambda}} \sum_{i=1}^m E_i/i.$$

Notice that the sequence  $E_m/m, E_{m-1}/m - 1, \dots, E_1/1$  gives in distribution the spacings of the exponential random variables  $E_1, \dots, E_m$ . So the sum  $\sum_{i=1}^m E_i/i$  is equal in distribution to a random variable  $B_m$  that is the maximum of  $m$  independent exponentially distributed random variables with rate 1. The distribution function of  $B_m$  is

$$\mathbf{P}(B_m \leq x) = (1 - e^{-x})^m.$$

From here we find that

$$\begin{aligned} \mathbf{P}\left(\tilde{\lambda}A_{a_n}^{(x)} - \log a_n \leq x\right) &= \mathbf{P}(B_{a_n} \leq x + \log a_n) = (1 - e^{-(x + \log a_n)})^{a_n} \\ &= \exp(-e^{-x} + O(1/a_n)) \rightarrow \exp(-e^{-x}) = \Lambda(x), \end{aligned}$$

where  $\Lambda$  denotes the distribution function of a standard Gumbel random variable. Similarly to the proof of Theorem 2.4, we conclude that

$$(\tilde{\lambda}A_{a_n}^{(x)} - \log a_n, \tilde{\lambda}A_{C_n}^{(y)} - \log C_n) \xrightarrow{d} (Y_1, Y_2),$$

where  $Y_1$  and  $Y_2$  are two independent copies of a standard Gumbel random variable. From Theorem 4.10 we get that by the continuous mapping theorem

$$\tilde{\lambda}W_n - \log n \stackrel{d}{=} \underbrace{\tilde{\lambda}A_{a_n}^{(x)} - \log a_n}_{\xrightarrow{d} Y_1} + \underbrace{\tilde{\lambda}A_{C_n}^{(y)} - \log C_n}_{\xrightarrow{d} Y_2} + \underbrace{\log C_n - \log a_n}_{\xrightarrow{d} -Y_3},$$

where  $Y_3$  is an independent Gumbel random variable with distribution function  $\exp(-e^{-\tilde{\lambda}x})$  by Lemma 4.7.  $\square$

## 6 Extending to the general model

In this section the general model of [6] is described. To extend our results to this setting, following the lines of [6], general kernel functions are approximated by a sequence of finite-type kernels already studied in detail in the previous sections.

In the general setting  $\mathcal{S}$  is a separable metric space equipped with a Borel probability measure  $\mu$ . For each  $n$  we have a deterministic or random sample of  $n$  points  $\mathbf{x}_n = (x_1, \dots, x_n)$  from  $\mathcal{S}$ . We assume that the empirical distribution

$$\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

converges in probability to  $\mu$  as  $n \rightarrow \infty$ , where  $\delta_x$  is the measure consisting of a point mass of weight 1 at  $x$ . This convergence condition is equivalent to the condition

$$\nu_n(S) := \frac{\#\{i : x_i \in S\}}{n} \xrightarrow{\mathbf{P}} \mu(S), \quad (32)$$

for every  $\mu$ -continuity set  $S \subset \mathcal{S}$  (i.e.  $S$  is measurable and there is no mass on the boundary of  $S$ ). This is the analog of condition (1).

The pair  $(\mathcal{S}, \mu)$  is called a ground space and for a sequence  $(\mathbf{x}_n)_{n \geq 1}$  satisfying (32) we say that the triplet  $(\mathcal{S}, \mu, (\mathbf{x}_n)_{n \geq 1})$  defines a vertex space  $\mathcal{V}$ . A (generalized) kernel  $\kappa$  on a ground space is a symmetric non-negative measurable function on  $\mathcal{S} \times \mathcal{S}$ . From here the random graph  $G(n, \kappa)$  is defined the same way as in the finite-type case. For a given kernel  $\kappa$  and vertex space  $\mathcal{V}$ , an edge  $\{ij\}$  (with  $i \neq j$ ) exists with probability

$$p_{ij} := \min \left\{ \frac{\kappa(x_i, x_j)}{n}, 1 \right\}. \quad (33)$$

Independently each edge is given an  $\text{Exp}(1)$  edge weight.

### Remark 6.1.

- (a) Kernel  $\kappa$  is regular finitary [6, Definition 4.4] if  $\mathcal{S}$  has a finite partition into ( $\mu$ -continuity) sets  $S_1, \dots, S_r$  such that  $\kappa$  is constant on each  $S_i \times S_j$  for all  $1 \leq i, j \leq r$ . Clearly, a random graph  $G(n, \kappa)$  generated by a regular finitary kernel has the same distribution as a finite-type graph.



- (b) Note that measure zero sets cannot be ignored the usual way, since changing  $\kappa$  on a set of measure zero may have a significant effect on the graph  $G(n, \kappa)$ .
- (c) This is indeed a very general model. Several examples and relationships to previously studied models can be found in [6, Sections 4 and 16].

The model is too general thus far, so further restrictions are needed for the kernel  $\kappa$ . Some technical conditions are needed and the total number of edges  $e(G)$  need to "behave the right way". These are collected in the following definition.

**Definition 6.2.** A kernel  $\kappa$  is graphical on a vertex space  $(\mathcal{S}, \mu, (\mathbf{x}_n)_{n \geq 1})$  if the following hold:

- (i)  $\kappa$  is continuous almost everywhere (a.e.) on  $\mathcal{S} \times \mathcal{S}$ ,
- (ii)  $\kappa \in L^1(\mathcal{S} \times \mathcal{S}, \mu \times \mu)$ ,
- (iii)

$$\frac{1}{n} \mathbf{E}(e(G(n, \kappa))) \rightarrow \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) d\mu(x) d\mu(y). \quad (34)$$

Condition (iii) says that  $G(n, \kappa)$  has about the right number of edges, since

$$\mathbf{E}(e(G(n, \kappa))) = \mathbf{E} \left( \sum_{i < j} \min \{ \kappa(x_i, x_j) / n, 1 \} \right),$$

which tends to  $n \frac{1}{2} \iint \kappa$  under certain well behaved cases (see [6, Lemma 8.1]). For example, condition (iii) holds whenever  $\kappa$  is bounded and  $\nu$  is a vertex space. Conditions (ii) and (iii) together imply that the (expected) average degree is  $O(1)$ , meaning that the results will only be extended in the sparse setting. Further note that Definition 6.2 was not needed in the finite-type case, since all of the conditions automatically hold in that setting. To use a sequence of approximating kernels we need to let  $\kappa$  depend on  $n$ . The next definition contains the conditions for a sequence of kernels.

**Definition 6.3.** Let  $\nu = (\mathcal{S}, \mu, (\mathbf{x}_n)_{n \geq 1})$  be a vertex space and  $\kappa$  be a kernel on  $\nu$ . A sequence  $\kappa_n$  of kernels on  $(\mathcal{S}, \mu)$  is graphical on  $\nu$  with limit  $\kappa$  if, for a.e.  $(y, z) \in \mathcal{S}^2$ ,

$$y_n \rightarrow y \text{ and } z_n \rightarrow z \text{ imply that } \kappa_n(y_n, z_n) \rightarrow \kappa(y, z),$$

$\kappa$  satisfies conditions (i) and (ii) in Definition 6.2, and

$$\frac{1}{n} \mathbf{E}(e(G(n, \kappa_n))) \rightarrow \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) d\mu(x) d\mu(y). \quad (35)$$

We also need to exclude cases where the vertex set of  $G(n, \kappa)$  can be split into two parts so that the probability of an edge from one part to the other is zero. This is formalized in the following definition, which is the analog of Assumption 3.5 for the matrix  $A$  defined in (4).

**Definition 6.4.** *A kernel  $\kappa$  on a ground space  $(\mathcal{S}, \mu)$  is irreducible if*

$$A \subseteq \mathcal{S} \text{ and } \kappa = 0 \text{ a.e. on } A \times (\mathcal{S} \setminus A) \text{ implies } \mu(A) = 0 \text{ or } \mu(\mathcal{S} \setminus A) = 0.$$

*As a slight modification we say that  $\kappa$  is quasi-irreducible if there is a  $\mu$ -continuity set  $\mathcal{S}' \subseteq \mathcal{S}$  with  $\mu(\mathcal{S}') > 0$  such that the restriction of  $\kappa$  to  $\mathcal{S}' \times \mathcal{S}'$  is irreducible and  $\kappa(x, y) = 0$  if  $x \notin \mathcal{S}'$  or  $y \notin \mathcal{S}'$ .*

Next, we describe the branching process that arises when exploring a component of  $G(n, \kappa)$ . It is a multi-type continuous time branching process with type space  $\mathcal{S}$ , where a particle of type  $x \in \mathcal{S}$ , when it splits, gives birth to a set of offspring distributed as a Poisson process on  $\mathcal{S}$  with intensity measure  $\kappa(x, y)d\mu(y)$ . That is, the number of children with types in a subset  $S \subset \mathcal{S}$  has a Poisson distribution with mean  $\int_S \kappa(x, y)d\mu(y)$ . In the finite-type case if we pick  $S$  to be a single type  $t$ , then we get back the branching process with Poisson offspring distribution with mean  $\kappa(s, t)\mu_t = \lambda_{st}$  studied in detail in Section 3. Each offspring lives for an  $\text{Exp}(1)$  amount of time independently of each other.

We also define an integral operator, whose norm establishes a direct connection with the emergence of a giant component in the random graph  $G(n, \kappa)$  and the survival of the above mentioned branching process. Let  $T_\kappa$  be the integral operator on  $(\mathcal{S}, \mu)$  with kernel  $\kappa$  defined by

$$(T_\kappa f)(x) := \int_{\mathcal{S}} \kappa(x, y)f(y)d\mu(y), \quad (36)$$

for any measurable function  $f$  such that this integral is defined. In the finite-type case this simplifies to a sum, namely

$$T_\kappa f = (A + I)f,$$

where  $f$  is a column vector,  $A$  is the matrix defined in (4) and  $I$  is the identity matrix. The norm of  $T_\kappa$  is

$$\|T_\kappa\| := \sup \{\|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1\} \leq \infty,$$

where  $\|\cdot\|_2$  is the norm of  $L^2(\mathcal{S}, \mu)$ . In the finite-type case easy calculations give us the norm:

$$\begin{aligned} \|T_\kappa\| &= \|A + I\| = \sqrt{\lambda_{\max}((A + I)^*(A + I))} = \sqrt{\lambda_{\max}(A + I)^2} \\ &= \sqrt{(\lambda_{\max}(A + I))^2} = \lambda_{\max}(A) + 1 = \tilde{\lambda} + 1, \end{aligned}$$

where the last equality holds because of Assumption 2.2. Theorem 3.1 of [6] states that a giant component emerges in  $G(n, \kappa_n)$  and its corresponding branching process survives with positive probability if and only if  $\|T_\kappa\| > 1$ . This is why we assumed throughout that  $\tilde{\lambda} > 0$ .

**Remark 6.5.** *Further details and remarks concerning the general model can be found in [6, Section 2].*

It only remains to formulate our main assumption in the general setting. Let  $(\mathcal{S}, \mu)$  be an arbitrary ground space and let  $\kappa$  be such that it satisfies

$$\int_{\mathcal{S}} \kappa(x, y) d\mu(y) = \tilde{\lambda} + 1 \quad \text{for a.e. } x. \quad (37)$$

This means that asymptotically the average degree is independent of the type. We can see that in this case the norm  $\|T_\kappa\|$  is indeed  $\tilde{\lambda} + 1$ . An example for a kernel  $\kappa$  satisfying (37) can be given by taking  $\mathcal{S} = (0, 1]$  (interpreted as the 1 dimensional torus  $\mathbb{T}^1$ ),  $\mu$  as the Lebesgue measure and  $\kappa(x, y) = h(d(x, y))$  for an even function  $h \geq 0$  of period 1, where  $d(x, y)$  is the metric given on  $\mathcal{S}$ . Figures 1 and 2 show the contours of two such examples, where the dark purple strips indicate where the contour is zero.

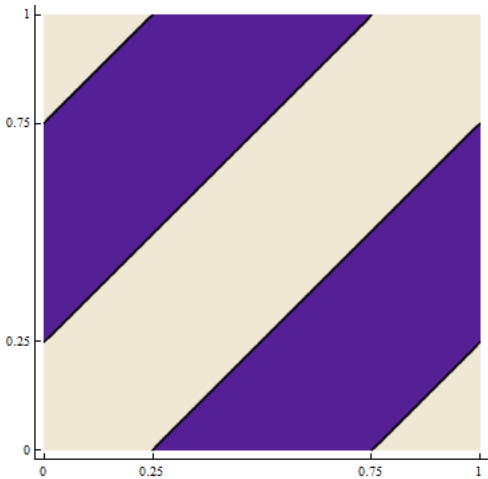


Figure 1:  $h(z) = \mathbb{1} [|z| < 0.25]$

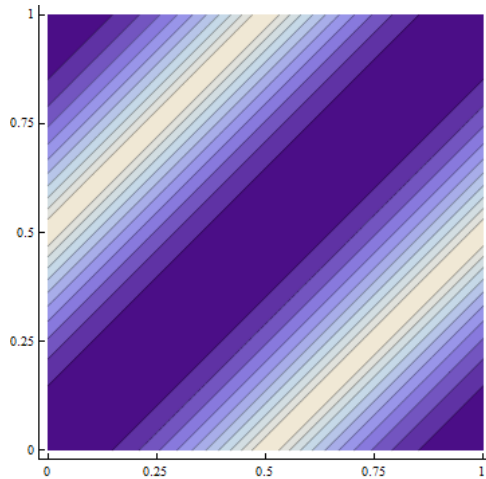


Figure 2:  $h(z) = z^2$

## 6.1 Approximation of kernels

As mentioned before, the extension of the main results to the general model will be done by constructing a sequence of partitions of  $\mathcal{S}$  for which a sequence

of lower and upper approximations of the limiting kernel  $\kappa$  are defined, each satisfying (37). The approximating kernels are regular finitary kernels, so the results of the previous sections apply to them (the regular finitary case and the finite-type case differ only in notation). The idea of using approximating kernels comes from [6, Section 7].

In the regular finitary case we may assume that the type space  $\mathcal{S}$  is finite. Furthermore, we can assume that  $\mu_s > 0$  for every  $s \in \mathcal{S}$ , however an argument is needed. We cannot simply ignore such types, since measure zero sets can alter  $G(n, \kappa)$  significantly. We can argue as follows.

Suppose that  $\mu_s = 0$  for some  $s \in \mathcal{S}$ . Start by redefining the kernel  $\kappa'(s, t) = \kappa'(t, s) := \max \kappa$  for every  $t \in \mathcal{S}$  and leaving it alone otherwise. Then define a new probability measure  $\mu'$  by shifting some small mass  $\eta$  over to  $s$  from the other types. Clearly  $\mu'_t > 0$  for every  $t \in \mathcal{S}$ . Possibly the types of some of the vertices changes, so change them correspondingly. This way we obtain a vertex space  $\nu' = (\mathcal{S}, \mu', (\mathbf{x}'_n)_{n \geq 1})$  with kernel  $\kappa'$ . It is not hard to see that we can couple  $G(n, \kappa)$  and  $G'(n, \kappa')$  so that  $G(n, \kappa) \subseteq G'(n, \kappa')$ . Finally, letting  $\eta \rightarrow 0$ , the norm of  $T_{\kappa'}$  with respect to  $\mu'$  tends to the norm of  $T_{\kappa'}$  with respect to  $\mu$ , which is equal to the norm of  $T_{\kappa}$ , since  $\kappa = \kappa'$  a.e. Iterating for other measure-0 types, we can see that it suffices to consider cases where  $\mu_t > 0 \forall t \in \mathcal{S}$ .

We continue with the definitions of the approximating kernels. Given a sequence of finite partitions  $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$ ,  $m \geq 1$ , of  $\mathcal{S}$  and an  $x \in \mathcal{S}$ , we define  $i_m(x)$  as the element of  $\mathcal{P}_m$  in which  $x$  falls, formally  $x \in A_{m, i_m(x)}$ . As usual,  $\text{diam}(A)$  denotes  $\sup\{d(x, y) : x, y \in A\}$  for  $A \subset \mathcal{S}$ , where  $d$  is the metric on our metric space  $\mathcal{S}$ . Lemma 7.1 of [6] states that for any ground space  $(\mathcal{S}, \mu)$  there exists a sequence of finite partitions of  $\mathcal{S}$  such that

- 1) each  $A_{mi}$  is a  $\mu$ -continuity set,
- 2) for each  $m$ ,  $\mathcal{P}_{m+1}$  refines  $\mathcal{P}_m$ ,
- 3) for a.e.  $x \in \mathcal{S}$ ,  $\text{diam}(A_{m, i_m(x)}) \rightarrow 0$ , as  $m \rightarrow \infty$ .

For such a sequence of partitions we can define a sequence of approximations of  $\kappa$  by taking its average on each  $A_{mi} \times A_{mj}$ :

$$\bar{\kappa}_m(x, y) := \frac{1}{\mu(A_{m, i_m(x)}) \cdot \mu(A_{m, i_m(y)})} \iint_{A_{m, i_m(x)} \times A_{m, i_m(y)} \kappa(s, t) d\mu(s) d\mu(t). \quad (38)$$

If  $\kappa$  is continuous a.e. then property 3) implies that  $\bar{\kappa}_m(x, y) \rightarrow \kappa(x, y)$  for a.e. every  $(x, y) \in \mathcal{S}^2$ . To be able to apply our theorems for finite-type kernels we need to guarantee that Assumption 2.2 holds for all  $\bar{\kappa}_m$ . Thus, considering

$\bar{\kappa}_m$  as a finite-type kernel with respect to the partition  $(A_{m1}, \dots, A_{mM_m})$ , the row sums of  $\bar{\kappa}_m$  weighted by  $\mu$  must be equal to some constant  $c_m$ . In fact we can show that this holds with  $c_m \equiv \tilde{\lambda} + 1$ :

$$\begin{aligned} \sum_j \bar{\kappa}_m(A_{mi} \times A_{mj}) \mu(A_{mj}) &\stackrel{(38)}{=} \sum_j \frac{1}{\mu(A_{mi})} \iint_{A_{mi} \times A_{mj}} \kappa(s, t) d\mu(s) d\mu(t) \\ &= \frac{1}{\mu(A_{mi})} \int_{A_{mi}} \left( \sum_j \int_{A_{mj}} \kappa(t, s) d\mu(s) \right) d\mu(t) \\ &\stackrel{(37)}{=} \frac{1}{\mu(A_{mi})} \int_{A_{mi}} (\tilde{\lambda} + 1) d\mu(t) = \tilde{\lambda} + 1. \end{aligned}$$

Note that we did not make any use of the properties of the partition or the ground space, we only used assumption (37) and the fact that  $\kappa$  is symmetric.

**Remark 6.6.** Originally in [6], lower and upper approximations were defined for  $\kappa$  by taking the infimum or supremum of  $\kappa$  on each  $A_{mi} \times A_{mj}$ :

$$\begin{aligned} \kappa_m^-(x, y) &:= \inf \{ \kappa(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)} \}, \\ \kappa_m^+(x, y) &:= \sup \{ \kappa(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)} \}. \end{aligned}$$

Assume that  $\kappa$  is bounded, otherwise  $\kappa_m^+$  may be  $+\infty$ . In general, an appropriate partition sequence does not necessarily define homogeneous  $\kappa_m^-$  and  $\kappa_m^+$  approximating sequences. However, if  $\mathcal{S} = \mathbb{T}^n$  with its usual metric and  $\mu$  Lebesgue measure, then the partition by dyadic rationals in each coordinate leads to homogeneous approximating kernels.

## 6.2 Extension of Theorems 2.3 and 2.4

We have made all the necessary preparations to extend the results of Theorems 2.3 and 2.4 to the general model described in this section. Recall that  $W_n$  denotes the weight of the shortest-weight path between two uniformly chosen vertices in the giant component, while  $H_n$  denotes the hopcount (number of edges) on this path.

**Theorem 6.7** (Hopcount in general setting). *Let  $(\mathcal{S}, \mu)$  be an arbitrary ground space and  $\kappa$  be a uniformly continuous, quasi-irreducible, graphical kernel on  $(\mathcal{S}, \mu)$  that satisfies  $\sup \kappa(x, y) < \infty$  and*

$$\int_{\mathcal{S}} \kappa(x, y) d\mu(y) = \tilde{\lambda} + 1 < \infty \quad \text{for a.e. } x \in \mathcal{S}.$$

Then the hopcount  $H_n$  between two uniformly chosen vertices, conditioned on being connected, satisfies a central limit theorem of the form

$$\frac{H_n - \frac{\tilde{\lambda}+1}{\lambda} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\lambda} \log n}} \xrightarrow{d} Z,$$

where  $Z$  is a standard normal variable.

*Proof.* Let  $(\mathcal{S}, \mu)$  be an arbitrary ground space and kernel  $\kappa$  satisfies the conditions of the theorem. These define the sequence of random graphs  $(G(n, \kappa))_{n \geq 1}$ . Take any sequence of finite partitions  $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$ ,  $m \geq 1$ , that satisfy properties 1), 2) and 3) described in Subsection 6.1. For each  $m$ , consider the finite type approximating kernel  $\bar{\kappa}_m$  defined in (38), with ground space  $(\mathcal{S}_m, \mu)$  (where  $|\mathcal{S}_m| = M_m$ ). As a result we obtain the sequence  $(G(n, \bar{\kappa}_m))_{n, m \geq 1}$ . Note that in the proofs for finite type kernels none of the estimates depend on  $\mu_t$  or the cardinality of  $\mathcal{S}_m$ , so all the error terms are uniform. The condition  $\sup \kappa(x, y) < \infty$  is necessary because it is used in the proof of Proposition 4.1.

To prove the results we let  $n$  and  $m$  tend to  $\infty$  simultaneously in a carefully chosen way. For fixed  $m$ , Theorem 2.3 states that

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \frac{H_n^m - \frac{\tilde{\lambda}+1}{\lambda} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\lambda} \log n}} < x \right) = \Phi(x),$$

where  $\Phi(x)$  is the distribution function of a standard normal random variable. On the other hand, letting  $m$  depend on  $n$ , from the error estimate given in Wasserstein metric in Lemma 3.8 it is easy to see that

$$\left| \mathbf{P} \left( \frac{H_n^{m(n)} - \frac{\tilde{\lambda}+1}{\lambda} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\lambda} \log n}} < x \right) - \Phi(x) \right| \leq \frac{C(\tilde{\lambda})}{\sqrt{\log n}}, \quad (39)$$

where  $C(\tilde{\lambda})$  is a  $\tilde{\lambda}$ -dependent constant.

To be able to couple the graphs  $G(n, \kappa)$  and  $G(n, \bar{\kappa}_m)$ , we need a relation between  $\kappa(x, y)$  and  $\bar{\kappa}_m(x, y)$ . Since  $\kappa$  is uniformly continuous, for  $(u, v) \in A_{m, i_m(x)} \times A_{m, i_m(y)}$ :

$$\bar{\kappa}_m(u, v) \leq \kappa(x, y)(1 \pm \varepsilon_m) \quad \text{if } |u - x| < \delta_m \text{ and } |v - y| < \delta_m,$$

where  $\text{diam}(A_{m, i_m(x)}) < \delta_m$  and  $\text{diam}(A_{m, i_m(y)}) < \delta_m$ . Abbreviate  $A_{m, i_m(x)}$  by  $A_{mx}$ . In  $G(n, \kappa)$  the probability that there exists an edge between two

vertices of types  $x$  and  $y$  is  $\kappa(x, y)/n$ . While in  $G(n, \bar{\kappa}_m)$ , between types  $A_{mx}$  and  $A_{my}$  this probability is  $\bar{\kappa}_m(x, y)/n$ , which we can now replace by  $\kappa(x, y)(1 \pm \varepsilon_m)$ . Thus

$$\mathbf{P}(\mathbb{1}[\{x, y\} \in e(G(n, \kappa))] \neq \mathbb{1}[\{A_{mx}, A_{my}\} \in e(G(n, \bar{\kappa}_m))]) \leq \frac{2\varepsilon_m}{n}.$$

Summing over all possible edges, we find for the edge sets that

$$\mathbf{P}(e(G(n, \kappa)) \neq e(G(n, \bar{\kappa}_m))) \leq \frac{2n^2\varepsilon_m}{2n} = n\varepsilon_m.$$

Now fix  $m$ . The  $\delta_m$  and uniform continuity of  $\kappa$  defines  $\varepsilon_m$ . Let

$$N_{0,m} := \max \left\{ N : \varepsilon_m N \sqrt{\log N} \leq 1 \right\}.$$

Clearly for all  $n < N_{0,m}$ , the coupling between  $G(n, \bar{\kappa}_m)$  and  $G(n, \kappa)$  fails only with probability less than  $1/\sqrt{\log n}$ . On the other hand, for an  $n$  we can pick a large enough  $m$  which approximates  $G(n, \kappa)$  well, namely

$$m(n) := \inf \{ m : N_0(m) > n \}.$$

Now we can couple  $G(n, \kappa)$  to  $G(n, \bar{\kappa}_{m(n)})$  so that the coupling fails with probability less than  $1/\sqrt{\log n}$ . Under the coupling, also for the hopcount we have

$$\mathbf{P}(H_n \neq H_n^{m(n)}) \leq 1/\sqrt{\log n} = o(1).$$

Combining this error bound with the one in (39) we obtain that

$$\begin{aligned} & \mathbf{P} \left( \frac{H_n - \frac{\tilde{\lambda}+1}{\lambda} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\lambda} \log n}} < x \right) \\ &= \mathbf{P} \left( \frac{H_n^{m(n)} - \frac{\tilde{\lambda}+1}{\lambda} \log n}{\sqrt{\frac{\tilde{\lambda}+1}{\lambda} \log n}} < x \mid H_n = H_n^{m(n)} \right) (1 - o(1)) + \mathbf{P}(H_n \neq H_n^{m(n)}) \\ &= \Phi(x) \left( 1 - O(\sqrt{\log n}^{-1}) \right) + C(\tilde{\lambda})/\sqrt{\log n} + 1/\sqrt{\log n}. \end{aligned}$$

Finally letting  $n \rightarrow \infty$  (thus  $m(n) \rightarrow \infty$  also), we obtain the desired result.  $\square$

We conjecture that the assertion of Theorem 2.4 also goes through to the general setting. The main difficulty is that the limiting random variable  $W$  depends on the cardinality of the partition, whereas for the hopcount the limit is always a standard normal random variable regardless of the number of types.

**Conjecture 6.8.** *Using the assumptions of Theorem 6.7, there exists a non-degenerate real valued random variable  $W$ , such that the minimal weight  $W_n$  satisfies*

$$W_n - \frac{1}{\lambda} \log n \xrightarrow{d} W,$$

where the distribution of  $W$  can be determined precisely from the behavior of the multi-type branching process that arises when exploring a component of  $G(n, \kappa)$  (defined earlier in the section).

## Summary and future plans

The aim of the thesis was to study first passage percolation on inhomogeneous random graphs. For supercritical graphs convergence in distribution was shown for the weight of the shortest-weight path and for the hopcount between two uniformly chosen vertices in the giant component under the assumption that the average degree of a vertex did not depend on its type. First, a finite-type model was considered. The proof consisted of three main parts: 1) proving analogous results for an appropriate branching process, 2) embedding of the thinned BP into an exploration process of the random graph, 3) rigorous analysis of the connection time. Second, with the additional ingredient of approximating general kernels the results were extended to the general model.

Future plans include the elimination of the constraint imposed on the row sum of  $A$  in Assumption 2.2, i.e. letting the average number of neighbors of a vertex depend on its type. This assumption played a vital role in some of our arguments. Without it

- the Poisson BP introduced in Section 3 can not be considered as a single-type BP,
- the ratio of a thinned subtree in Lemma 4.4 does not satisfy the simple relation in (23),
- the distribution of the next vertex to split in  $SWT^{(x)}$  is not the same as the distribution of the vertex that splits in  $SWT^{(y)}$  when the connection happens.

In this direction we already have some partial results. Another possible direction is to generalize the exponential edge weights. It seems reasonable to make an effort with phase-type distributions [11] since on the one hand they are a generalization of exponential distributions, thus inherit some nice



properties. On the other hand the set of phase-type distributions is dense in the set of non-negative distributions [12], so it could be possible to generalize to arbitrary edge weights this way.

## References

- [1] ATHREYA, K. AND NEY, P. (2004). *Branching Processes*. Dover Publications.
- [2] BHAMIDI, S., VAN DER HOFSTAD, R., AND HOOGHIMSTRA, G. (2010). First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.* 20(5), 1907–1965.
- [3] BHAMIDI, S., VAN DER HOFSTAD, R., AND HOOGHIMSTRA, G. (2011). First passage percolation on the erdős-rényi random graph. *Combinatorics, Probability and Computing* 20, 683–707.
- [4] BÜHLER, W. (1971). Generations and degree of relationship in supercritical markov branching processes. *Probability Theory and Related Fields* 18(2), 141–152.
- [5] BILLINGSLEY, P. (1999). *Convergence of probability measures*, second edition ed. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.
- [6] BOLLOBÁS, B., JANSON, S., AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Struct Algor* 31, 3–122.
- [7] CHEN, L. AND SHAO, Q. (2005). *Stein’s method for normal approximation*. Lecture notes series IMS, National University Singapore. 1–59.
- [8] DUDLEY, R. (2002). *Real Analysis and Probability*. Cambridge University Press.
- [9] ERDŐS, P. AND RÉNYI, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5, 17–61.
- [10] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized pólya urns. *Appl* 110, 177–245.
- [11] LATOUCHE, G. AND RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM, Philadelphia.

- [12] NELSON, R. (1995). *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag.
- [13] SÖDERBERG, B. (2002). General formalism for inhomogeneous random graphs. *Phys. Rev. E* 66, 066121.