TUDOMÁNYOS DIÁKKÖRI DOLGOZAT
SCIENTIFIC STUDENT PAPER

First passage percolation on the Newman-Watts small world model with exponential edge weights

Viktória Vadon

Advisor

Dr. Júlia Komjáthy
postdoctoral research fellow
Eindhoven University of Technology
Stochastic Section

Budapest University of Technology and Economics
2014

Contents

1. Introduction and main results

1.1. **Introduction to random graphs and networks.** Technology and application of arbitrary graph studies reached the phase of turning to the study of large real-life networks. The empirical results are surprising, they found that many real-life networks show some sort of universality by sharing important properties. (For an overview on networks and random graphs see for example [19, 21].) To start with, degree sequences of most of these networks have a power-law tail, i.e., for k large enough, the number of vertices with degree k is proportional to $k^{-\beta}$, for some $\beta \geq 1$. The second commonly shared property is known as "six degrees of separation" for social networks. Based on empirical results, in the graph representing people and their acquaintances, typically two people in the world can be connected with a chain of 6 people. This small world phenomenon, i.e., the diameter being really small compared to the size of the graph, also is a typical property of real-life networks.

Since these complex networks are huge and hence hard to treat globally, local description is important. The local clustering coefficient of a vertex, i.e., the proportion of finished triangles and pairs of edges from the given vertex, was found to be high in most real-life networks. For a social network, this means that two friends of a person are likely to be friends too.

Since the structure and topology of networks affects the performance of various processes on the network, such as the spread of information in biological or human networks, it is an important question for sciences to understand these in more detail. Local structure and neighborhoods depend on the local connection rules of vertices, which are found to be probabilistic. This led to the study of random graphs.

One of the most basic and oldest random graph model is the Erdős-Rényi random graph [20], where there is a fixed number of vertices and each possible edge is added independently with probability p. This model was studied rigorously, however, in some aspects, it does not behave like real-life networks. As a generalization to overcome these aspects, the inhomogeneous graph model was introduced, its most general form befined by Bollobás, Janson and Riordan

[16]. Given the vertices having some types from a (not necessarily finite) type-set, the probability of any edge being added is independent from everything else and depends on the types of the two vertices given by a kernel function κ.

Some random graphs were directly created to model real-life networks, based on the observed properties of the real-life networks under consideration. Preferential attachment models [5, 15] were created to model the time development of internet-related networks. At each time step, a new vertex is added to the network and connects to old vertices independently, with a probability proportional to their degrees. There is a modification that adds new edges in two steps: the first step is the method described above, then we connect the new vertex to the neighbors of its neighbors with a higher probability, to raise more triangles. A possible generalization is a model where vertices and edges can also disappear, or where vertices have some kind of 'fitness' parameter increasing or decreasing the likelihood of getting new edges.

Another different approach is based on the often observed hierarchic structure of networks. Initiated by Barabási, Ravasz and Vicsek [6], hierarchic graph models were created. These models are deterministic, fractal-like graphs with hierarchic structure, where the degree distribution has a power-law tail, and the small world phenomenon is also valid.

When we study processes on networks that move faster than the development of the network itself, we can use static random graph models. (More about those observed processes later.) One commonly used model for this is the configuration model (see in e.g. [14] or [21]). We take n independent copies of a positive, integer-valued random variable to be the degree sequence of the graph and give each vertex as many half edges as its degree is supposed to be. To form edges, we sequentially take a half edge and choose its pair uniformly over the not-yet-used half-edges, (thus we choose a higher-degree vertex with larger probability). We repeat this until all half edges are paired. If the total number of half edges is odd, we delete the last unpaired one. Its popularity comes from the fact that conditioning on the graph being simple leads to uniform distribution among the simple graphs with the given degree sequence.

Another static model is the Watts-Strogatz [33] or Newman-Watts [27] small world model, which we shall introduce later and more rigorously, as it is the center of our attention.

1.2. **First passage percolation.** First passage percolation was first introduced to study spreading dynamics on lattices, in particular on $\mathbb{Z}^d, d \geq 2$. The intuitive idea behind the method is that one imagines water flowing at a constant rate through the (random) medium, the waterfront representing the spread. The model turned out to be able to capture the core idea of several other processes as well, applications include studying graph distances and epidemic spreads.

Note: an epidemic spread is a process on a network. We start the epidemic from a given root (or several ones), which is infected and contagious. The contagious period can be constant, random or infinite. The edge weights correspond to the time needed for the infection to pass through, with finite contagious periods, the neighbors might remain uninfected. It also leads to different models whether we allow vertices to recover. The most simple model, which directly corresponds to the water spread, is when vertices once infected remain so and the contagious period is infinite. Then the epidemic spreads the whole component. We will refer to this easy model later when we write epidemic spread.

Among the first applications of first passage percolation, Janson [22] investigated typical distances and the corresponding hopcount, i.e., the number of edges along the shortest weight path, flooding times as well as diameter on the complete graph. Evidently, without edge weights, all these quantities equal to 1. A strange phenomenon arises though, as adding random exponential edge weights, shortest weight paths drop to order $\log n/n$. Specifically, the shortest weight path, the flooding time and diameter converge to 1, 2, and 3 times $\log n/n$, respectively, and the hopcount jumps up to order $\log n$.

In a sequence of papers (e.g. [10, 11, 12, 32]) van der Hofstad *et al* started to investigate typical distances in random graphs, leading to a large body of first passage percolation on random graphs literature. Their aim was to determine *universality classes* for the random metric spaces that are given

by the shortest path metric when adding some random edge weights to a random graph. They showed (in a sequence of papers) that for random graphs that have no geometry (e.g. the supercritical Erdős-Rényi random graph, the configuration model, or the rank-1 inhomogeneous random graphs), where the degree distribution has finite asymptotic variance, and the edge weights are continuous on $[0, \infty)$, typical distances scale as $\log n$. On the other hand, having power-law degrees with infinite asymptotic variance drastically changes the metric and there are several universality classes. (For instance, for unit edge-weights, the graph is ultra small, i.e., distances scale as $\log \log n$, while for exponential edge-weights, distances are tight [11].) However, for finite means [12](in case of the configuration model) the hopcount remains of order $\log n$.

Recently, Kolossváry and Komjáthy [25] used first passage percolation to determine distances on the inhomogeneous random graph model mentioned above with exponential edge weights, for finite asymptotic mean of degrees. The results are of order $\log n$: the shortest weight path centered with a constant multiple of $\log n$ converges to a non-degenerate random variable, while the hopcount satisfies a central limit theorem with expected value and variance both constant multiples of $\log n$.

In [13] Bhamidi, van der Hofstad and Komjáthy studied epidemic spread on the configuration model with exponential edge weights. They pointed out the connection between first passage times and the epidemic curve, i.e., the asymptotic proportion of infected individuals as a function of time, for a SIR (susceptible-infected-removed) epidemic on the configuration model. After proving our main theorem about typical distances, the same idea yields us the epidemic curve for the Newman-Watts model with exponential edge weights. Earlier, [9] Barbour and Reinert also concerned themselves with epidemic spreads on the Erdős-Rényi random graph and on the configuration model (with bounded degrees), aiming for results that would apply to a wider class of random graphs.

In [2, 18] the competition of two first passage percolations, running on the same graph is investigated. This can be considered a competition between two epidemics, as well as the marketing of two similar products. The results

show that the graph structure has a significant effect on the outcome, as in small worlds, one competitor only gets a negligible part of the vertices, while on regular lattices with linear distances both floods occupy a linear fraction of vertices.

1.3. **The Newman-Watts model.** The Newman-Watts small world model, often referred to as "small world" in short, is one of the first random graphs, created to model real-life networks. It was first introduced by Watts and Strogatz [33], a simplifying modification made by Newman and Watts [27] later. The Watts and Strogatz model starts with a ring of n vertices, each connected to the $k \geq 1$ nearest vertices, then to create shortcuts, each edge is independently rewired to a uniformly chosen other vertex with probability p. This makes the model "tunable" in the sense that the degree of disorder can be manipulated, the graph ranging from completely regular to completely stochastic. Disorder comes with low clustering as well as short distances, while regularity means high clustering, but distances linear in n. However, if p is small but positive, high clustering remains while even comparatively few shortcuts make distances drop significantly, to order $\log n$, hence the small world.

Newman and Watts made the modification that instead of rewiring, all edges of the ring are kept and new shortcut edges are added. The method is similar to the creation of the Erdős-Rényi graph, for each pair of not yet connected vertices, we connect them independently with probability r. It is proved that these models are asymptotically equivalent as n tends to infinity (with the right choice of k, p, r), but the Newman-Watts model is easier to handle since it is always connected. (In this paper, we will pursue our investigations on the Newman-Watts model with $k = 1$.)

The model has been studied from different aspects. Newman *et al* studied distances [28, 29] with simulations and mean-field approximation, as well as the spread of non-deterministic epidemics [26]. The aim was to determine the threshold infection transmitting probability which suffices for a large outbreak of the epidemic.

Aside from the properties of the model that make it similar to real-life networks, the mixing time of a symmetric random walk on the Newman-Watts graph was studied. To define the mixing time, consider the symmetric random walk on the graph as a Markov chain, started with an initial distribution. In each step, the current state of the chain specifies a distribution, that is known to converge to the stationary distribution. The time when the total variation distance of these two drop below 1/4 is called the mixing time. Durrett [19] conjectured that the mixing time is between order of $(\log n)^2$ and $(\log n)^3$, later on Addario-Berry and Lei [1] proved that it is indeed of order $(\log n)^2$.

Barbour and Reinert also investigated typical distances on the Newman-Watts model with first passage percolation. In the first paper [7] they carried through investigations on a continuous circle with circumference L instead of L many vertices, and added $\mathrm{Poi}(L\rho/2)$ many shortcuts at locations chosen according to uniform measure on the circle. Distances are measured by the usual arc measure along the circle, while shortcuts are given length 0. In the second paper [8] they considered a discrete model, with unit edge weights on the cycle as well as the shortcuts, thus the metric was exactly the graph distance. Their results are - not surprisingly - of order $\log n$.

1.4. **Precise model and main results.** We will work on the Newman-Watts small world model [27] with independent random edge weights. We can construct the most convenient version of the model as follows: we take a cycle C_n on n vertices. Then independently for each pair $i, j \in [n] := \{1, 2, \dots, n\}$, where $|i - j| \neq 1 \mod n$, we add the edge (i, j) with probability ρ/n to form shortcuts. Here, ρ is the parameter of the Newman-Watts graph, corresponding to the asymptotic average number of shortcuts from a vertex. Conditioned on the edges of the resulting graph, we assign i.i.d. exponential random variables with parameter 1, in short, Exp(1) edge weights to each edge. We denote the length of edge e by X_e. We write $\mathrm{NW}_n(\rho)$ for a realization of this weighed random graph.

The additional edge weights can represent the time of transmission from one vertex to the other, or the cost along a given edge. We can also think of the graph with the edge weights as a random metric space, where the distance

between any two vertices is the sum of weights along the shortest weight path. Hence, this graph model is a (non-euclidean) random metric space. The distance between two vertices in this metric then can for instance correspond to time of information spread on the network, or can model cost of transmission between the vertices. More precisely, we can model the spread of information from a uniformly chosen vertex, $U \in [n]$, as follows:
We can assume that once a vertex v receives the information at time t, it starts transmitting the information towards all its neighbors at rate 1. Let us denote the vertices that are connected to v by an edge by $\mathrm{H}(v)$, then, for each $w \in \mathrm{H}(v)$, w receives the information at time $t + X_{(v,w)}$. We assume that vertices transmit information only once to each neighbor, even if they receive it more than once, and that there is no time limit for that transmission. (Hence, we can look at this model as a susceptible-infected model for spread of diseases: vertices cannot recover and the infectious period is infinite.)

We say that an event $\mathcal{E}_n$ happens (on $\mathrm{NW}_n(\rho)$) with high probability (w.h.p.) if $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$, that is, the probability that the event holds tends to 1 as the size of the graph tends to infinity. We write shortly Bin and Poi for Binomial and Poisson and distributions, respectively.

In this paper, we investigate typical distances, i.e., the weight of the shortest-weight path between two uniformly chosen vertices U and V. Let Γ_{UV} denote the set of all paths γ in $\mathrm{NW}_n(\rho)$ between U and V. Then the weight of the shortest weight path is defined by

$$\mathcal{P}_n(U,V) := \min_{\gamma\in\Gamma_{UV}} \sum_{e\in\gamma} X_e. \tag{1.1}$$

Theorem 1.1 (Typical distances)**.** *Let U, V be two uniformly chosen vertices in $[n]$. Then, the distance $\mathcal{P}_n(U,V)$ in $\mathrm{NW}_n(\rho)$ with i.i.d $\mathrm{Exp}(1)$ edge weights satisfies w.h.p.*

$$\mathcal{P}_n(U,V) - \frac{1}{\lambda}\log n \xrightarrow{d} -\frac{1}{\lambda}(\log W^U W^V + \Lambda + c),$$

where λ is the largest root of the polynomial $\lambda^2 + (1-\rho)\lambda - 2\rho$, Λ is a standard Gumbel random variable, the random variables W^U, W^V are independent copies of the martingale limit of the multi-type branching process defined below in

Subsection 2.3, and $c = \log(1 - \pi_R^2/2) - \log(\lambda(\lambda+1))$*, furthermore* π *is the stationary type-distribution for the same branching process.*

For the γ path that minimalizes the distance, we call $\mathrm{H}_n(U,V) = |\gamma| = \sum_{e\in\gamma} 1$ the hopcount, i.e., the number of edges along the shortest-weight path.

Theorem 1.2 (Central limit theorem for the hopcount)**.**

$$\frac{\mathrm{H}_n(U,V) - \frac{\lambda+1}{\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}} \xrightarrow{d} X \text{ w.h.p. as } n\to\infty$$

Where X *is a standard normal random variable and* $\xrightarrow{d}$ *means convergence in distribution.*

We consider an epidemic spread on the network as well, with a single source U, all other vertices are susceptible. Once a vertex becomes infected, it remains so, and infects all its neighbors some time later, these transmission times are i.i.d. exponential r.v.'s. (The epidemic occupies the graph in the exact same way the waterflow does, if the edges are pipes with length given by the edge weight.) On $\mathrm{NW}_n(\rho)$, we denote by $\mathrm{I}_n(t,U) := \frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{i \text{ is infected by time } t\}$, the fraction of infected vertices among all vertices at time t, in the epidemic started from the vertex U.

Theorem 1.3 (Epidemic curve)**.** *Define the function* $f(t) = 1 - \mathrm{M}_{W_V}(x(t))$*, where* M_{W_V} *is the moment generating function of* W_V*, and* $x(t) = \left(1 - \frac{1}{2}\pi_R^2\right)\frac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}$*. Then* $\mathrm{I}_n(t + \frac{1}{\lambda}\log n, U)$ *converges to* $f(t + \frac{1}{\lambda}\log W_U)$ *in probability.*

W_U, as in Theorem 1.1, is the martingale limit random variable of the branching process defined in 2.3, started from U.

Remark 1.4. Note that f is a deterministic function with a random shift given by $\frac{1}{\lambda}\log W_U$. This means that throughout the epidemic, the fraction of infected individuals follows a deterministic curve, only the starting time is random. This phenomenon has been observed in real-life epidemics, see e.g. this link or [31]. The time shift depends on W_U, which describes the neighborhood of U. The result corresponds to the heuristics: a bigger value of W_U means the neighborhood of U is "dense", the spread is quick in the initial

stages, hence can infect the whole graph quickly. Indeed, a bigger value of W_U shifts the function more to the left, causing it to take bigger values earlier. It is also just natural that we observe times in the form of $t + \frac{1}{\lambda}\log n$, where t is a real parameter that can take even negative values. As $2t_n = \frac{1}{\lambda}\log n$ is the typical distance, we expect most individuals to be infected in a small time frame around time $2t_n$, hence the fraction of infected individuals is expected to change significantly in the same time frame.

Remark 1.5. To make the distinguishing between shortcut- and cycle-edges easier, let us color the cycle-edges red and the shortcut-edges blue. In this respect, each vertex has two red neighbours, and since there are $\mathrm{Bin}(n-3, \rho/n)$ many blue outgoing edges from the vertex, each vertex has asymptotically $\mathrm{Poi}(\rho)$ many blue neighbours.

1.4.1. *Comparison to related literature and context.*

Erdős-Rényi graph. The subgraph formed by the blue edges is approximately an Erdős-Rényi graph, with the difference that the cycle always makes $\mathrm{NW}_n(\rho)$ connected, hence there is a single giant component containing all vertices. It is well-known [21] that typical distances in the giant component of the supercritical Erdős Rényi random graph are of order $\log n$, and this is also true [8] for the Newman-Watts model (without the edge weights) for any $\rho > 0$. Hence, adding the cycle-edges makes the graph somewhat similar to the supercritical regime of the Erdős Rényi random graph. On the other hand, the cycle edges add a geometry to the graph, which plays a crucial role throughout the investigation.

Let $\beta = 1/(\rho - 1)$, then with our notations, the typical distance on the Erdős-Rényi graph with $\mathrm{Exp}(1)$ edge weights, conditioned on U and V being connected:

$$\mathcal{P}_n(U, V) - \beta \log n \xrightarrow{d} -\beta \left(\log(\beta W) + \log(\beta W') + \Lambda\right)$$

Where W and W' are independent copies of $(W_0 | W_0 > 0)$, with W_0 being the branching process limiting random variable and Λ is a standard Gumbel random variable. Note that $\rho - 1$ is the expected offspring-1 in the branching process, corresponding to λ, the Malthusian growth parameter in our model.

Hence the difference between typical distances on the Erdős-Rényi random graph model and on the Newman-Watts model is in the constant.

Inhomogeneous random graphs. Recall the inhomogeneous random graph model mentioned above in subsection 1.1. Vertices have types s from a type space S, equipped with a probability measure μ, and conditioned on the types of the vertices, edges are present independently with probabilities that depend on the types. These probabilities are given by a so-called kernel κ. We will see that the exploration process on the $\mathrm{NW}_n(\rho)$ model is similar to the exploration process on inhomogeneous random graphs with exponential edge weights [25] - that is, both of them can be modelled using a continuous time multi-type branching process. It would be natural to guess that typical distances are then the same in these two models for an appropriately chosen κ. It turns out, that this is almost the case: the first order term $\lambda^{-1}\log n$, and the random variables are the same, but the additive constant c in Theorem 1.1 is not: the geometry of the Newman-Watts model modifies how the branching processes can connect to each other, which modifies the constant. Writing their result in the same form as ours, $c_{\mathrm{inhom}} = -\log\left(\lambda(\lambda+1)/\sum_{s\in S}\pi(s)^2/\mu(s)\right)$.

Other results about the Newman-Watts small world model. As mentioned before, Barbour and Reinert investigated typical distances in related models rigorously. In [7] they treated a continuous version of the model. Their result is implicit, but shows the distance is logarithmic function of L:

$$\mathbb{P}(\mathcal{P}_L(U,V) > (\log(L\rho)/2 + x)/\rho) \xrightarrow{d} \int_0^\infty \frac{\mathrm{e}^{-y}\mathrm{d}y}{1 + y(2\mathrm{e}^{2x})}$$

In a subsequent paper [8] they treated the discrete model $\mathrm{NW}_n(\rho(n))$ with unit edge weights. They gave complete characterisation of typical distances in terms of the parameter $\rho(n)$ that might also tend to infinity with n. In particular, they showed that the earlier continuous model is a good approximation only if $\rho(n) \to \rho$: in this case the distances are logarithmic, given by a similar implicit relation.

1.5. **Sketch of the paper.** In what follows, we prove Theorems 1.1, 1.2 and 1.3. The brief idea of the proof is the following: we pick two vertices uniformly, then we start the neighbourhoods of these vertices in the graph in the order that

is the distance from these vertices (Section 2). We show that this procedure w.h.p. results in shortest weight trees (SWT's) from the two vertices that can be coupled to two independent copies of a certain branching process, that will be a continuous time multi-type branching process. We then handle the connection between these two branching processes in Section 3 with the help of a Poisson-approximation. After the detailed analysis of the connection process, the proof of Theorem 1.1 follows by noting that the first connection of the two explored clusters provides a shortest weight path, that can be described uniquely in the Poisson point process. We also heuristically compute the epidemic curve from this, then make it rigorous in 4. Finally we prove the central limit theorem for the hopcount (5).

2. Exploration process

To explore the neighborhood of a vertex, we use a modification of the Dijkstra algorithm. Along the lines of first passage percolation, the intuitive picture is that the edges represent pipes with length given by the edge weight, and water flows through the graph from the source at a unit rate. At any time t, the explored vertices are the wet ones while the actives mean those vertices that compete to become explored (are the end of an edge through which water is currently flowing).

Introduce the following notations: $\mathcal{N}(t)$, $\mathcal{A}(t)$, $\mathcal{U}(t)$ denote the set of explored, active and unexplored vertices at time t, respectively.
$\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$ denotes the remaining lifetime of the currently active vertices, indexed by the vertices; $\mathcal{R}_w(t)$ meaning the remaining lifetime of some vertex $w \in \mathcal{A}(t)$ at time t. We write $\mathrm{N}(t)$, $\mathrm{A}(t)$ for the sizes of explored and active sets at time t. As before, $\mathrm{H}(v)$ denotes the neighbors of a vertex v. Adding the subscript R or B to any quantity corresponds to the same quantity restricted to only the red or blue vertices, respectively. Colors will be given according to the color of the edge we reach the vertex through.

2.1. The exploration algorithm on an arbitrary weighted graph.

Let $i = 1$. The vertex from which we start the exploration process is denoted by

v_1. We color v_1 blue and set the time as $t = T_1 = 0$. Evidently, we take

$$\begin{aligned}\mathcal{N}(0) &= \{v_1\}\\ \mathcal{A}(0) &= \mathrm{H}(v_1)\\ \mathcal{U}(0) &= [n] \setminus (\{v_1\} \cup \mathrm{H}(v_1))\end{aligned}$$

The remaining lifetimes are determined by the edge weights, that is

$$\mathcal{R}_{\{\mathcal{A}(0)\}}(0) = \{\mathcal{R}_w(0) = X_{(v_1,w)} \text{ for all } w \in \mathrm{H}(v_1)\}.$$

We color the active vertices $w \in \mathrm{H}(v_1)$ to have the same color as the edge (v_1, w).

We work with induction from now on. In each step, we increase i by 1. We can construct the continuous time process in steps, namely, at the random times when we explore a new vertex.

Let $\tau_i = \min\left(\mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1})\right)$, the minimum of remaining lifetimes. Then define $T_i := T_{i-1} + \tau_i$, the time when we explore the next vertex. Clearly, the sets do not change until T_i. Formally, for any t s.t. $T_{i-1} \le t < T_i$:

$$\begin{aligned}\mathcal{N}(t) &:= \mathcal{N}(T_{i-1})\\ \mathcal{A}(t) &:= \mathcal{A}(T_{i-1})\\ \mathcal{U}(t) &:= \mathcal{U}(T_{i-1}).\end{aligned}$$

From all the remaining lifetimes, we subtract the time passed: for some $0 \le s \le \tau_i$,

$$\mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1} + s) := \mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1}) - s, \text{ subtracted element-wise.}$$

Note that we define only $\mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(t)$ it for $t = T_i$ too, which will make the included remaining lifetimes continuous. The other sets are defined up to T_i, excluding T_i, which leads to the quantities $\mathrm{A}(t)$ and $\mathrm{N}(t)$ being right-continuous. (So that $\mathrm{A}(T_i)$ means the number of active vertices *after* the i^{th} split.)
At time T_i, the vertex v_i of which the remaining lifetime equals to 0, becomes explored and its neighbors become active. We shall refer to v_i as the i^{th} explored

vertex.

$$\mathcal{N}(T_i) := \mathcal{N}(T_{i-1}) \cup \{v_i\}$$
$$\mathcal{A}(T_i) := \mathcal{A}(T_{i-1}) \setminus \{v_i\} \cup \mathrm{H}(v_i)$$
$$\mathcal{U}(T_i) := \mathcal{U}(T_{i-1}) \setminus \mathrm{H}(v_i)$$

Only the change of the remaining lifetimes set is not trivial.

$$\mathcal{R}_{\{\mathcal{A}(T_i)\}}(T_i) := \mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_i) \setminus \{\mathcal{R}_{v_i}(T_i)\} \cup \{\mathcal{R}_x(T_i) : x \in \mathrm{H}(v_i)\}$$

where $\mathcal{R}_x(T_i) = X_{(v_i,x)}$, the edge weight of (v_i, x), and x also gets the color of (v_i, x).

On an arbitrary connected weighted graph, the exploration process can be continued until all vertices become explored.

Note that this algorithm builds the shortest weight tree SWT from the starting vertex. This tree will be modeled with the branching process.

Remark 2.1. The "set" of actives might contain several occurrences of a vertex, in case at least two neighbors of a vertex are explored already. (The water is approaching the same vertex through several edges at the same time.) We do have to keep them until one of them becomes explored so that we know which one gives the shortest weight path to the vertex from the source. We do not delete other occurrences right away, they will be "thinned" (see later in 2.4.1) when they are supposed to become explored. Hence the coloring might not be well-defined for active vertices, as the colors of the different occurrences can be different, but the coloring is well-defined on explored vertices.

2.2. **Exploration on the weighted Newman-Watts random graph.** Though we consider $\mathrm{NW}_n(\rho)$ as a *realization* of a random graph, to determine the characteristics of the exploration process on $\mathrm{NW}_n(\rho)$, we can not handle it as an arbitrary graph. We have to consider T_i, τ_i, $\mathrm{N}(t)$ and $\mathrm{A}(t)$ as random variables, such as $\mathcal{N}(t)$, $\mathcal{A}(t)$, $\mathcal{U}(t)$ and $\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$ are random sets, and the next explored vertex is chosen randomly.

Lemma 2.2 (Children)**.** *Suppose the vertex v is being explored for the first time (i.e., not "double-explored"). If v is red, one new red and* Binomial$(n-3, \frac{\rho}{n})$ *many new blue active vertices are born. If v is blue, two new red and*

Binomial$(n-4, \frac{\rho}{n})$ *many new blue active vertices are born. The number of new blue active vertices is asymptotically* Poi(ρ) *in both cases.*

Proof. The statement is obvious from Remark 1.5, using the fact that the color of the vertex is determined by the color of the edge it is reached through. □

Lemma 2.3 (Remaining lifetimes)**.** *At any time t, any element of $\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$ is an* Exp(1) *random variable.*

Proof. Since the edges are weighed with i.i.d. exponential random variables, this follows easily from the memoryless property of exponential distribution. □

Corollary 2.4 (Next explored)**.** *The minimum of the remaining lifetimes is unique with probability 1 and the next explored vertex is chosen uniformly over the set of actives.*

Proof. To see the first part, we only have to recall that two out of finite many absolutely continuous random variables are equal with probability zero. The algorithm takes the minimum of independent, identically distributed random variables, which is uniform over the indeces. Since the next explored vertex is the one with the least remaining lifetime, the second part of the statement also holds. □

2.3. **Multi-type branching processes.** We define the following multi-type continuous time branching process that corresponds to the initial stages of the exploration process.

There are two particle types, red (R) and blue (B), and their lifetime is Exp(1), independent from everything else. Particles give birth only once upon their death. They leave behind offspring as in Lemma 2.2: each particle has Poi(ρ) many blue offspring, red particles have one, while blue particles have two red children. Dead and alive particles will correspond to explored and active vertices, respectively. With this wording, for the number of alive particles, we define

Definition 2.5. *We shall write $\mathbf{Z}(t) = (Z_R(t), Z_B(t))$ for the number of alive particles of each type, $Z(t)$ meaning the total number of alive particles. Let*

$\mathrm{N}(t) = \mathrm{N}_R(t) + \mathrm{N}_B(t)$, *where* $\mathrm{N}_q(t)$ *means the number of dead particles of type* $q = R, B$. *We assume the above quantities to be right-continuous. Superscripts* $(R), (B)$ *refer to the process started with a single particle of the given type.*

The exploration process corresponds to the process started with a single blue-type particle, which dies immediately.

2.3.1. *Literature on multi-type branching processes.* Here we restate the necessary theorems from [4] which we will use.

In this subsubsection, n does not refer to the number of vertices in the Newman-Watts graph, but to the number of dead particles in the branching process.

Definition 2.6 (Mean matrix)**.** *Let* $\mathrm{M}(t) := \mathrm{M}_{r,q}(t) = \mathbb{E}[Z_q^{(r)}(t)]$, $(q, r = R, B)$ *the mean matrix, where* $Z_q^{(r)}(t)$ *is as defined above in Definition 2.5.*

It is not hard to see that $\mathrm{M}(t)$ satisfies the semigroup property $\mathrm{M}(t+s) = \mathrm{M}(t)\mathrm{M}(s)$ and the continuity condition $\lim_{t\to 0} \mathrm{M}(t) = \mathrm{I}$, where I denotes the identity matrix. As a result, we have:

Theorem 2.7 (Athreya-Ney)**.** *There exists an infinitesimal generator matrix* Q *so that* $\mathrm{M}(t) = \mathrm{e}^{\mathrm{Q}t}$, *where* $Q_{r,q} = a_r \mathbb{E}[D_q^{(r)}] - \delta_{r,q}$. *Here,* a_r *is the rate of dying for a particle of type* r, *(i.e., the parameter of its exponential lifetime),* D *is the number of offspring with the same sub-end superscript conventions as in Definition 2.5, and* $\delta_{r,q} = \mathbb{1}_{\{r=q\}}$ *(i.e.,* $\delta_{r,q} = 1$ *if and only if* $r = q$*).*

In our case,

$$\mathrm{Q} = \begin{pmatrix} 0 & \rho \\ 2 & \rho - 1 \end{pmatrix}$$

Eigenvalues and eigenvectors of the Q *matrix*. Using the characteristic polynomial,

$$\lambda_{1,2} = \frac{\rho - 1 \pm \sqrt{\rho^2 + 6\rho + 1}}{2}.$$

For $\rho \geq 1$, the maximal eigenvalue is

$$\lambda = \frac{\rho - 1 + \sqrt{\rho^2 + 6\rho + 1}}{2} \tag{2.1}$$

We shall refer to the second eigenvalue by

$$\lambda_2 = \frac{\rho - 1 - \sqrt{\rho^2 + 6\rho + 1}}{2}. \tag{2.2}$$

The corresponding normalised left eigenvector π which satisfies $\pi \mathrm{Q} = \lambda\pi$ gives the stationary type-distribution.

$$\pi = (\pi_R, \pi_B) = \left(\frac{2}{\lambda + 2}, \frac{\lambda}{\lambda + 2}\right) \tag{2.3}$$

We denote the right (column) eigenvector with $\mathbf{u}$ and normalize it so that $\pi\mathbf{u} = 1$. For later use, without computing, we denote by $\mathbf{v}_2$ and $\mathbf{u}_2$ the left (row) and right (column) eigenvector of Q belonging to the eigenvalue λ_2.

A well-known phenomenon, and the most important theorem for our purposes is that the process grows exponentially with rate λ (the so-called Malthusian parameter), more precisely,

Theorem 2.8 (Athreya-Ney)**.**

$$\lim_{t\to\infty} \mathbf{Z}(t)\mathrm{e}^{-\lambda t} = W\pi$$

where W is a nonnegative random variable, the almost sure martingale limit of $W_t := \mathbf{Z}(t)\mathbf{u}\,\mathrm{e}^{-\lambda t}$.

Theorem 2.9 (Athreya-Ney)**.** *Define T_n, the n^{th} split time, as the time of the n^{th} death in the branching process. (We assume $T_1 = 0$ for the death of the root.) On the event $\{W > 0\}$,*

(i) *For each $q \in (R, B)$, $\lim_{n\to\infty} \mathrm{N}_q(T_n)/\mathrm{N}(T_n) = \lim_{n\to\infty} \mathrm{N}_q(T_n)/n \overset{a.s.}{=} \pi_q$*
(ii) $\lim_{n\to\infty} n\mathrm{e}^{-\lambda T_n} \overset{a.s.}{=} \frac{1}{\lambda}W$

For the number of dead particles,

Corollary 2.10.

$$\mathrm{N}(t)\mathrm{e}^{-\lambda t} \xrightarrow{a.s.} \frac{1}{\lambda}W$$

Proof. From Theorem 2.8 we have $Z(t)\mathrm{e}^{-\lambda t} \xrightarrow{a.s.} W$. Let $S_n = Z(T_n) = Z(T_n+)$, the number of alive particles *after* the n^{th} split. Expressed with the offsprings:

$S_n = 1 + \sum_{k=1}^{n}(D_k - 1)$, D_k meaning the total offspring of v_k, the k^{th} splitting particle. We compute

$$\begin{aligned}\frac{1}{n}S_n &= \frac{1}{n}\left(1 + \sum_{k=1}^{n}(D_k - 1)\right) \\ &= \frac{1}{n} + \frac{1}{n}\mathrm{N}_R(T_n)\,\frac{1}{\mathrm{N}_R(T_n)}\left(\sum_{j=1}^{\mathrm{N}_R(T_n)} D_j^{(R)} - 1\right) \\ &\quad + \frac{1}{n}\mathrm{N}_B(T_n)\,\frac{1}{\mathrm{N}_B(T_n)}\left(\sum_{j'=1}^{\mathrm{N}_B(T_n)} D_{j'}^{(B)} - 1\right).\end{aligned}$$

We use part (i) of Theorem 2.9 and note that the variables inside the sums are i.i.d., hence the Law of Large numbers implies that

$$\lim_{n\to\infty}\frac{1}{n}S_n = \pi_R\left(\mathbb{E}[D^{(R)}] - 1\right) + \pi_B\left(\mathbb{E}[D^{(B)}] - 1\right) = \pi\mathrm{Q}\mathbf{1} = \lambda, \tag{2.4}$$

where $\mathbf{1}$ means the column vector of ones.

Note that $\mathrm{N}(t)$ yields n for the unique n such that $T_n \le t < T_{n+1}$, as well as for such values of t, $Z(t) = Z(T_n) = S_n$ holds, hence

$$S_{\mathrm{N}(t)} = Z(t).$$

Substituting this into Theorem 2.8, we have

$$S_{\mathrm{N}(t)}\mathrm{e}^{-\lambda t} = \mathrm{N}(t)\frac{S_{\mathrm{N}(t)}}{\mathrm{N}(t)}\mathrm{e}^{-\lambda t} \xrightarrow{a.s.} W$$

Since the process is supercritical, $T_n \to \infty$ a.s., thus by (2.4),

$$\lim_{t\to\infty}\frac{S_{\mathrm{N}(t)}}{\mathrm{N}(t)} = \lim_{n\to\infty}\frac{S_{\mathrm{N}(T_n)}}{\mathrm{N}(T_n)} = \lim_{n\to\infty} S_n/n \overset{\text{a.s.}}{=} \lambda$$

follows. Substituting, we have

$$\mathrm{N}(t)\lambda\mathrm{e}^{-\lambda t} \xrightarrow{a.s.} W.$$

□

Throughout the next sections, we develop error bounds on the coupling between the branching process and the exploration process on the graph. For

convenience, we introduce

$$t_n := \frac{1}{2\lambda} \log n \tag{2.5}$$

the times we will observe the branching and exploration processes at, as well as

$$W^{(n)} := e^{-\lambda t_n} Z(t_n), \quad \text{with} \quad W^{(n)} \xrightarrow{a.s.} W, \tag{2.6}$$

the approximations of the martingale limit W at the times t_n.

Note that for large enough n, $W^{(n)} = W(1 + o(1))$, hence we often ignore the superscript when dealing with error terms.

2.4. **Labeling, coupling, error terms.** In this section we relate the branching process to the exploration process on $\mathrm{NW}_n(\rho)$ and calculate the coupling error.

Error bound on coupling the offspring. The branching process is defined with $\mathrm{Poi}(\rho)$ blue offspring distribution, while in the exploration process a vertex actually has $\mathrm{Bin}(n-3, \rho/n)$ (or $\mathrm{Bin}(n-4, \rho/n)$) many blue children, which we have to handle.
Let $Y \sim \mathrm{Poi}(\rho)$ and $X \sim \mathrm{Bin}(n, \rho/n)$. By the usual coupling of Binomial and Poisson random variables, we get $\mathbb{P}(X \neq Y) \leq \frac{\rho^2}{n}$.

First we examine the blue offspring in the case when a red vertex splits. Now, we have to couple a different binomial, $Z \sim \mathrm{Bin}(n-3, \rho/n)$ to Y. Let $V \sim \mathrm{Bin}(3, \rho/n)$, then $Z = X - V$. Then

$$\begin{aligned}
\mathbb{P}(Z \neq Y) &\leq \mathbb{P}(X \neq Y) + \mathbb{P}(V \neq 0) \\
&\leq \frac{\rho^2}{n} + 1 - \mathbb{P}(V = 0) \\
&= \frac{\rho^2}{n} + \frac{3\rho}{n} + o(1/n^2)
\end{aligned}$$

In case a blue vertex splits, $\hat{Z} \sim \mathrm{Bin}(n-4, \rho/n)$ and $\hat{V} \sim \mathrm{Bin}(4, \rho/n)$. Hence, similarly as before, $\mathbb{P}(\hat{V} \neq 0) = 4\frac{\rho}{n} + o(1/n^2)$, thus $\mathbb{P}(\hat{Z} \neq Y) \leq \frac{\rho^2}{n} + \frac{4\rho}{n} + o(1/n^2)$.

Claim 2.11. *The probability that up to k steps, for at least one prticle the blue offspring in the exploration process and the Poisson branching process is different, is at most $k\frac{\rho^2+4\rho}{n}$.*

Proof. In each step, the bound is either $\frac{\rho^2}{n} + \frac{3\rho}{n}$ and $\frac{\rho^2}{n} + \frac{4\rho}{n}$, of which the maximum is $\frac{\rho^2}{n} + \frac{4\rho}{n}$. Hence, by a union bound, multiplying it by the number of steps gives an upper bound. □

Since the red offspring is deterministic, it does not rise an error term.

2.4.1. *Labeling and thinning.* We label the particles of the branching process to correspond to the exploration process. The labeling is as follows:
Note: everything should be interpreted modulo n.

(i) The root is labeled u, the source of the exploration process. u can be U, a uniformly chosen vertex over $[n]$.
(ii) Every other particle gets a label when it is born.
(iii) We distinguish "left type" and "right type" red children. Left type red particles have a left type red child, right type red particles have a right type red child, blue particles have a red child of both types.
(iv) A left type red child of v gets label $v - 1$, a right type red child of v is labeled $v + 1$.
(v) The blue children of v get a set of labels uniformly chosen from $[n]$.

Recall from Remark 2.1 that we might have several occurrences of a vertex (both active and explored) in the exploration process as well as in the branching process, but at most one of them is explored. The labeling fails if two *explored* vertices share the same label.

Lemma 2.12. *The probability that the labeling fails at the i^{th} split is at most $2i/n$.*

Proof. The labeling fails at the i^{th} split if the splitting particle has a label that is already taken by an explored vertex. To estimate the probability of this, we investigate two cases.
When a blue particle splits. Since the label of a blue particle is chosen uniformly over $[n]$, and there are at most $i - 1$ dead labels already, the probability that we choose from this set is $(i - 1)/n$.
When a red particle splits. Since the labeling of red particles is deterministic, we have to investigate them differently. In this part, we have to go back to the

underlying graph model and imagine the exploration process on the cycle. We start with a single blue vertex, then explore its red neighbors, one by one, the interval of red vertices around the original blue one grows, the side is always randomly chosen. Sometimes we find a shortcut, leading to a distant blue vertex, that has a uniform label, and around that we start exploring red vertices in a more deterministic way - the label is given, only the time of exploration is random. This forms a growing set of growing intervals, each having a blue vertex as center, denoted by c_k for the interval I_k. (Actually $c_k = v_{B,k}$, the k^{th} dying blue particle.) Let us write l_k and r_k for the number of explored red vertices to the left and to the right of c_k. (Left and right radius of the k^{th} interval.) Since these intervals are functions of time, we shall write $I_k(T_i)$, $r_k(T_i)$ and $l_k(T_i)$. Recall that $I_k(T_i)$ means the same as $I_k(T_i+)$, since we think of the process as right-continuous.

In this setting, a new red vertex can coincide with an existing vertex if and only if two intervals I and I' grow into each other. It is not hard to see that this can only occur in the following setting:

There exists an $a \in [n]$ such that a vertex with label $a \in I(T_{i-1})$ and at the same time, the verex labeled $a+1 \in I'(T_{i-1})$. Since $a+1 \in \mathrm{H}(a)$, which is explored in $I(T_{i-1})$, $a+1$ is active in $I'(T_{i-1})$. Similarly, a is active in $I'(T_{i-1})$. In case at T_i, I or I' grows on the side towards the other, $a \in I'(T_i)$ or $a+1 \in I(T_i)$. Then either a or $a+1$ is double-explored.

Denote by I^* the interval that grows at time T_i. We want to know the probability of it growing into another I_j interval in exactly this step. Conditioned on c^*, $r^*(T_{i-1}), l^*(T_{i-1}), r_j$ and l_j, there are two possible labels of c_j for this event to happen: on the left side of I^*, at position $c^* - l^*(T_{i-1}) - r_j - 1$ and on the right side, at position $c^* + r^*(T_{i-1}) + l_j + 1$. (Both cases mean the furthest explored red vertices were neighbors at step t_{i-1}.)
Hence, we get that for a single interval I_j,

$$\mathbb{P}(I^*(T_{i-1}) \cap I_j(T_{i-1}) = \varnothing, I^*(T_i) \cap I_j(T_i) \neq \varnothing) = \frac{2}{n}$$

We have to sum this up over all possible intervals $I_j \neq I^*$. Note that there are exactly as many intervals as blue explored vertices (at either T_{i-1} or T_i, since

v_i is red).
Let the event $G_i = \{v_i \text{ is red and its label is already used}\}$.

$$\mathbb{P}(G_i) \leq \sum_{j=1}^{\mathrm{N}_B(T_{i-1})-1} \frac{2}{n} = \frac{2}{n}\left(\mathrm{N}_B(T_{i-1}) - 1\right) \leq \frac{2i}{n},$$

since there are less than i blue explored vertices ($\mathrm{N}_B(T_{i-1}) \leq \mathrm{N}(T_{i-1}) = i - 1$). Note: this proof is not limited to the case when two red dead particles meet (a vertex is double red explored), but also applies when the new red explored vertex coincides with a formerly explored blue one, in case $l_j(T_{i-t}) = 0$ or $r_j(T_{i-t}) = 0$.

Hence, the statement of the lemma follows. □

In case the labeling fails, we can not consider the second (or any later) occurrence of the double-explored vertex and its descendants. (We do not need them, as we already have a shorter path to the given label, then in any paths through the label, we can replace the part from the root to this vertex with the shortest possible and not make the path longer. Also, the offspring of the particle has to be well defined, but in the random construction, another explored occurence would lead to generating a new set of offspring.) We mark this second (or any later) occurrence thinned, and all its descendants ghosts. To carry on with the branching process approximation, we have to bound the proportion of ghosts among the actives.

Ancestral line. We approach the problem of ghost actives with the help of the ancestral line. We define the ancestral line $\mathrm{AL}(v)$ of a vertex v as the chain of particles leading to v from the root, including the root and v itself. Then an alive particle is a ghost if and only if at least one of its ancestors is thinned.
The ancestral line was introduced by Bühler in [17] with the following approach: for each time interval $[T_k, T_{k+1})$ we can allocate a unique particle on the ancestral line that was active in the interval $[T_k, T_{k+1})$. Hence we can look at the ancestral line as a time process too.
Since we do not remove the descendants of thinned particles, only mark them ghosts, there might be particles that are thinned because of ghosts, creating

more ghosts then necessary. This only means that the bound we achieve isn't sharp, however, it suffices.

Bühler introduced the ancestral line to use it in discussing the generation of a randomly chosen alive particle, as well as the degree of relationship of two randomly chosen alive particles. Our aim is different, so we make use only of the indicator decomposition of the generation. For the following observations, we condition on $\{D_i, i = 1, ..., k\}$, where D_i is the total number of offspring of the i^{th} dying particle. Denote by G_k the generation of a uniformly chosen alive particle V after the k^{th} split. Then $G_k = L_1 + L_2 + ... + L_k$, where the indicators L_i are conditionally independent and $L_i = 1$ if and only if the ancestor of V that was alive in the time interval $[T_i, T_{i+1})$ was newborn (born at T_i).

Since the ancestors are chosen uniformly, in the interval $[T_i, T_{i+1})$, D_i many particles are newborn, and S_i many are alive, which yields the probability $\mathbb{P}(L_i = 1) = D_i/S_i$.

Suppose that w_1 and $w_2 \in \mathrm{AL}(v)$, furthermore, w_2 is an offspring of w_1. Then w_2 is newborn at T_i if and only if w_1 is the particle splitting at T_i. Hence, if we define the indicators L_i' by $L_i' = 1$ if and only if the i^{th} splitting particle $v_i \in \mathrm{AL}(v)$, then $\{L_i' = 1\} = \{L_i = 1\}$.

Corollary 2.13. *The probability of the i^{th} dying particle being an ancestor of a V uniformly chosen active:*

$$\mathbb{P}(v_i \in \mathrm{AL}(V)) = \mathbb{P}(L_i' = 1) = \mathbb{P}(L_i = 1) = \frac{D_i}{S_i}.$$

Expected proportion of thinned actives. Let us combine Corollary 2.13 and Lemma 2.12. To be able to do so, we need the following lemma.

Lemma 2.14. *There exists a positive integer-valued random variable K so that K is always finite and for every $i > K$, $S_i = i\lambda(1 + o(1))$.*

We postpone the proof for now.

Lemma 2.15. *Let $\mathcal{A}_G(t) = \{v \in \mathcal{A}(t) : v \text{ is a ghost}\}$ the set of ghost actives at time t and $\mathrm{A}_G(t)$ its size. The proportion $\mathrm{A}_G(t_n + u)/\mathrm{A}(t_n + u)$ tends to 0 in probability as n tends to infinity.*

Remark 2.16. The times we want to observe is $t_n + u$, where $t_n = \frac{1}{2\lambda} \log n$ and u is an arbitrarily small real number (can be negative). The reason is that $\mathrm{e}^{\lambda t_n} = \sqrt{n}$, hence at t_n, the number of active vertices is of order $\sqrt{n}$.

Proof. The proportion $\mathrm{A}_G(t)/\mathrm{A}(t) = \mathbb{P}(V \in \mathcal{A}_G(t))$, where V is uniform over $\mathcal{A}(t)$, i.e., uniformly chosen active individual. We can use Corollary 2.13 for the representation of the ancestral line of $V \in \mathcal{A}(t)$, namely we can write

$$\begin{aligned}\mathbb{P}(V \in \mathcal{A}_G(t)) &= \mathbb{P}(\exists w \in \mathrm{AL}(V) \text{ that is thinned}) \\ &\leq \sum_{i=1}^{\mathrm{N}(t)} \mathbb{P}(v_i \in \mathrm{AL}(V) \text{ and } v_i \text{ is thinned}),\end{aligned}$$

where, as before, v_i means the particle that dies at T_i. Since the labeling is independent of the family tree,

$$\mathbb{P}(V \in \mathcal{A}_G(t)) \leq \sum_{i=1}^{\mathrm{N}(t)} \mathbb{P}\left(v_i \in \mathrm{AL}(V)\right) \cdot \mathbb{P}\left(v_i \text{ is thinned}\right) \leq \sum_{i=1}^{\mathrm{N}(t)} \frac{D_i}{S_i} \frac{2i}{n}. \tag{2.7}$$

According to 2.14, we can split the sum for parts up to K and above. For $i \leq K$, we can use $D_i < S_i$.

$$\begin{aligned}\mathbb{P}(V \in \mathcal{A}_G(t)) &\leq \sum_{i=1}^{K} \frac{2i}{n} + \sum_{i=K+1}^{\mathrm{N}(t)} \frac{D_i}{\lambda(1+o(1))n} \\ &< \frac{K^2}{n} + \frac{\sum_{i=K+1}^{\mathrm{N}(t)} D_i}{\lambda n(1+o(1))} < \frac{K^2}{n} + \frac{\sum_{i=1}^{\mathrm{N}(t)} D_i}{\lambda n(1+o(1))}\end{aligned}$$

Then we make two sums with respect to the color of the splitting particle.

$$\begin{aligned}\sum_{i=1}^{\mathrm{N}(t)} D_i &= \sum_{i=1}^{\mathrm{N}_R(t)} D_i^{(R)} + \sum_{i'=1}^{\mathrm{N}_B(t)} D_{i'}^{(B)} \\ &= \mathrm{N}_R(t) \frac{1}{\mathrm{N}_R(t)} \sum_{i=1}^{\mathrm{N}_R(t)} D_i^{(R)} + \mathrm{N}_B(t) \frac{1}{\mathrm{N}_B(t)} \sum_{i'=1}^{\mathrm{N}_B(t)} D_{i'}^{(B)}\end{aligned}$$

Having averages of i.i.d. random variables, we can apply the Law of Large Numbers, and use that $(\mathrm{N}_R(t)/\mathrm{N}(t), \mathrm{N}_B(t)/\mathrm{N}(t)) \xrightarrow{a.s.} \pi$.

$$\begin{aligned}
&\mathbb{P}(V \in \mathcal{A}_G(t)) \\
&\leq K^2/n + \frac{\mathrm{N}(t)\pi_R\mathbb{E}[D^{(R)}](1+o(1)) + \mathrm{N}(t)\pi_B\mathbb{E}[D^{(B)}](1+o(1))}{\lambda n(1+o(1))} \\
&= K^2/n + \frac{\mathrm{N}(t)\left(\pi_R(\rho+1) + \pi_B(\rho+2)\right)}{\lambda n}(1+o(1))
\end{aligned}$$

Using that π is the left eigenvector of Q with eigenvalue λ, $\pi_R\rho + \pi_B(\rho+1) = \lambda(\pi_R + \pi_B) = \lambda$.
For large enough times, Corollary 2.10 implies $\mathrm{N}(t) = \mathrm{e}^{\lambda t}\frac{1}{\lambda}W(1+o(1))$. Hence

$$\mathrm{A}_G(t)/\mathrm{A}(t) = \mathbb{P}(V \in \mathcal{A}_G(t)) \leq K^2/n + \mathrm{e}^{\lambda t}\frac{\lambda+1}{\lambda^2 n}W(1+o(1)).$$

Thus for the observed time window $t_n + u$, the proportion of the ghost actives

$$\mathrm{A}_G(t_n+u)/\mathrm{A}(t_n+u) \leq K^2/n + \frac{\lambda+1}{\lambda^2}\mathrm{e}^{\lambda u}W\frac{1}{\sqrt{n}}(1+o(1)) \xrightarrow{a.s.} 0.$$

Even though K is a random variable, associated with the deviation of the S_i, it is independent of n, so the convergence remains to hold. □

Now before moving onwards, let us return to the proof of Lemma 2.14. This lemma follows from [3, Theorem 1, Theorem 2]. Here, we restate [3, Theorem 1] in a special case, where each eigenvalue has multiplicity 1. This is still sufficient for our purposes and easier to handle.

Theorem 2.17 (Asmussen)**.** *Let $\mathbf{Z}_n$ be a supercritical multitype Galton-Watson process, with dominant eigenvalue λ, the corresponding left and right eigenvector $\mathbf{v}$ and $\mathbf{u}$. For any other eigenvalue ν, $\mathbf{v}_\nu$ and $\mathbf{u}_\nu$ denote the left and right eigenvector belonging to ν.*
Let us decompose an arbitrary vector $\mathbf{a} \in \mathbb{R}^p$ that has the property $\mathbf{v}\cdot\mathbf{a} = 0$ in the basis of the eigenvectors of A and then we define

$$\mu := \sup\{\nu : \mathbf{v}_\nu\mathbf{a} \neq 0\}, \quad \sigma^2 := \lim_{n\to\infty}\frac{|\mathbf{v}|\mathrm{Var}(\mathbf{Z}_n\mathbf{a})}{\lambda^n}$$

If $\mu^2 < \lambda$, then with $C_n = (2\sigma^2\mathbf{Z}_n\mathbf{u}\log n)^{1/2}$

$$\liminf_{n\to\infty}\frac{\mathbf{Z}_n\mathbf{a}}{C_n} = -1 \quad \textit{and} \quad \limsup_{n\to\infty}\frac{\mathbf{Z}_n\mathbf{a}}{C_n} = 1.$$

We also restate [3, Theorem 2] without change.

Theorem 2.18 (Asmussen 2.)**.** *Replacing $n \in \mathbb{N}$ with $t \in [0, \infty)$, Theorem 2.17 remains valid for any supercritical irreducible multi-type Markov branching process.*

Proof of Lemma 2.14. Our aim is to use the previous two theorems for the 2-type branching process that corresponds to the exploration process. Since π and $\mathbf{v}_2$ are linearly independent in our case, there is no such nonzero vector that is orthogonal to both, so for any $\mathbf{a}$ that $\pi\mathbf{a} = 0$, necessarily $\mathbf{v}_2\mathbf{a} \neq 0$, which implies $\mu = \lambda_2$, since there are no more eigenvalues. The eigenvalues of the mean matrix $\mathrm{M}(t)$ are $\mathrm{e}^{\lambda t}$ and $\mathrm{e}^{\lambda_2 t}$. The condition $\mu^2 < \lambda$ in Theorem 2.17 is then equivalent to $2\lambda_2 < \lambda$ which follows from the nonnegativity of ρ, through simple algebraic computations. In our case the asymptotic variance σ^2 is the following:

$$\sigma^2 = \lim_{t\to\infty} \frac{\pi \mathrm{Var}(\mathbf{Z}(t)\mathbf{a})}{\mathrm{e}^{\lambda t}}$$
$$C_t = (2\sigma^2 \mathbf{Z}(t)\mathbf{u} \log t)^{1/2}$$

This implies

$$\limsup_{t\to\infty} \frac{\mathbf{Z}(t)\mathbf{a}}{C_t} = 1 \text{ and } \liminf_{t\to\infty} \frac{\mathbf{Z}(t)\mathbf{a}}{C_t} = -1.$$

Let us apply this theorem for the split time T_i: we get that there is only a finite number of indexes i such that $\left|\frac{\mathbf{Z}(T_i)\mathbf{a}}{C_{T_i}}\right| > 2$. Let the maximum of these indexes be K, a random variable. Since $T_i - 1/\lambda \log i$ has an almost sure limit, T_i is normally of order $\log i$. This implies that C_{T_i} is usually of order $(i \log\log i)^{1/2}$, then by Markov's inequality, C_{T_i} is w.h.p. of at most order $(i \log i)^{1/2}$.

Since $\mathbb{E}[\mathbf{Z}(t)\mathbf{a}] = 0$ if and only if $\pi\mathbf{a} = 0$, we can apply the theorem for the centered version $S_i^c := S_i - \mathbb{E}S_i$. Then for $i > K$, $|S_i^c| \leq C_{T_i}$. The fluctuation is of smaller order then S_i; itself, which means we can indeed write $S_i = i\lambda(1 + o(1))$. For more detail on this, see the proof of [23, Corollary 3.16]. □

2.4.2. *The number of multiple active and active-explored labels.* Recall that in the exploration process, as well as in the branching process there might

be multiple occurrences, see Remark 2.1, as thinning only prevents multiple explored labels.

For the connection process, we want to use that the number of different active labels at T_i is approximately the same as S_i, i.e., there are not many multiple occurrences. For the reasons mentioned above, we still have to deal with labels that are multiply active, or are explored and active at the same time. We will discuss this in the following five cases:

1. A blue active vertex has been already explored.
2. A red active vertex has been already explored.
3. A blue active vertex is also red active.
4. A vertex is double red active.
5. A vertex is double blue active.

We will denote by $p_\alpha(t)$ the probability at time t of a uniform active vertex falling under case $\alpha = 1, \dots, 5$, which is the same as the proportion of vertices falling under case α among all active vertices.

We observe the times $t_n + u$ with $t_n = \frac{1}{2\lambda} \log n$, $u \in \mathbb{R}$, since $\mathbb{E}(\mathrm{A}(t_n)) = W\mathrm{e}^{\lambda t_n} = W\sqrt{n}$, and we shall see that this is the size needed for the two branching processes connect.

Case 1. *Blue active being already explored*. At time t, there are at most $\mathrm{N}(t)$ explored and neither thinned nor ghost labels. Under the condition that the active vertex is blue, since its label is chosen uniformly over $[n]$, the probability that this label has been already explored is $\mathrm{N}(t)/n$. Substitute $\mathrm{N}(t_n + u)$ from Corollary 2.10, then for $t = t_n + u = \frac{1}{2\lambda} \log n + u$,

$$\begin{aligned} &\mathbb{P}(v \text{ is already explored} \mid v \text{ is blue}) = \mathrm{N}(t_n + u)/n \\ &= \left(e^{\lambda u} \frac{1}{\lambda} W(1 + o(1)) \right) / \sqrt{n}. \end{aligned}$$

Thus

$$p_1(t_n + u) \overset{\text{a.s.}}{=} \pi_B \left(e^{\lambda u} \frac{1}{\lambda} W(1 + o(1)) \right) / \sqrt{n} \to 0.$$

Case 2. *Red active being already explored*. We consider this case the same way we considered the thinning of red vertices. In the picture of the growing set of growing intervals on the cycle, a red active is explored if and only if two

intervals are about to grow into each other: the furthest explored red vertices in both interval are neighbors. We call these intervals neighbors. Then, for two neighboring intervals, the active vertices at the end of each interval are explored in the other interval. Let I_i and $I_j, 1 \le i < j \le \mathrm{N}_B(t)$ intervals around $c_i = v_{B,i}$ (i^{th} dying blue) and $c_j = v_{B,j}$ respectively. Conditioned on c_i and all radiuses, there are two possibilities so that I_i and I_j are neighbors: $c_j = c_i + r_i + l_j + 1$ or $c_j = c_i - l_i - r_j - 1$. In other words, conditioned on the position of one of the intervals, there are precisely two positions for the other interval so that our event, that the intervals are neighbors, happens. Thus for each pair of indices the probability of the intervals being neighbors is $2/n$. These are obviously not independent but note that expectation is linear even for sums of dependent random variables. Summing up for all pair of indexes and dividing by the number of all intervals gives the proportion of *case 2* red actives among all red actives.

$$p_2(t) = \frac{\text{number of neighboring intervals}}{\text{number of all intervals}} = \frac{\sum_{1 \le i < j \le \mathrm{N}_B(t)} 2/n}{\mathrm{N}_B(t)}$$
$$= \frac{\binom{\mathrm{N}_B(t)}{2} \cdot 2/n}{\mathrm{N}_B(t)} = (\mathrm{N}_B(t) - 1)2/n$$

For the time window of our interest, substituting $\mathrm{N}(t)$ from Corollary 2.10, and $t = t_n + u$, this means:

$$p_2(t_n + u) \overset{\text{a.s.}}{=} \pi_R(1 + o(1)) \left(\pi_B \sqrt{n}\, \mathrm{e}^{\lambda u} \frac{1}{\lambda} W - 1 \right) 2/n$$
$$= \pi_R \left(\pi_B (1 + o(1)) \mathrm{e}^{\lambda u} \frac{1}{\lambda} W - 1 \right) 2/\sqrt{n} \to 0.$$

Case 3. *Blue active being red active*. Using that the labels of blue vertices are chosen uniformly,

$$\mathbb{P}(v \text{ is red active too} | v \text{ is blue active}) = \mathrm{A}_R(t)/n \overset{\text{a.s.}}{=} \pi_R(1 + o(1))\mathrm{A}(t)/n$$

Hence $p_3(t) \overset{\text{a.s.}}{=} \pi_B \pi_R (1 + o(1))\mathrm{A}(t)/n$.

For the time window of our interest

$$p_3(t_n+u) \overset{\text{a.s.}}{=} \pi_B\pi_R(1+o(1))\sqrt{n}\,\left(e^{\lambda u}\,W(1+o(1))\right)/n$$
$$= \pi_B\pi_R\left(e^{\lambda u}\,W(1+o(1))\right)/\sqrt{n} \to 0.$$

Case 4. *Multiple red active vertices*. This case is very similar to Case 2. A vertex v can be red active twice, if the two intervals that it belongs to are "almost neighbors", that is, both have v as an active vertex on one of their ends. (v is the only vertex separating them.) The only difference is that a multiple red label decreases the number of different labels by 1, not by 2. This reduces the proportion to the half, hence

$$p_4(t_n+u) \overset{\text{a.s.}}{=} \pi_R\left(\pi_B(1+o(1))\mathrm{e}^{\lambda u}\,\frac{1}{\lambda}W-1\right)/\sqrt{n} \to 0.$$

Heuristically, we can say this occurs because a red cannot be double right side active, since that would mean the label on its left is doule-explored, which would have been thinned. (Same for left.)

Case 5. *Multiple blue active vertices*. We will examine this case with an approach different from the others, with an urn scheme. Imagine an urn with n balls, labeled with $[n]$. In each step, we choose one uniformly, note its number, then put it back in the urn. We want to know the number of different labels noted after k steps, or rather the number of those that were chosen more than once.

For any label l, the number of its appearances – denoted by L – in the label sequence has a distribution $L \sim \mathrm{Bin}(k, 1/n)$. The probability that l is chosen twice:

$$\mathbb{P}(L=2) = \binom{k}{2}\left(\frac{1}{n}\right)^2\left(\frac{n-1}{n}\right)^{k-2} < \frac{k^2}{2}\frac{1}{n^2}$$

Here again we need, $k = \mathrm{A}_B(t_n+u)$. Then

$$\mathbb{P}(L=2) \overset{\text{a.s.}}{<} \frac{\pi_B^2}{2}(\sqrt{n})^2\,\mathrm{e}^{2\lambda u}\,W^2/n^2 = \frac{\pi_B^2}{2}\,\mathrm{e}^{2\lambda u}\,W^2/n.$$

For $3 \le i \le k = \mathrm{A}_B(t) = O(\sqrt{n})$, for some constant C_i,

$$\mathbb{P}(L=i) < C_i\frac{k^i}{n^i}.$$

Since $k = O(\sqrt{n})$, there exists a constant C so that

$$\mathbb{P}(L \geq 3) = C\frac{k^3}{n^3} = \hat{C}(W,u)/(\sqrt{n})^3(1+o(1)).$$

Above is the proportion of labels among all labels that are used twice, or more than twice for blue actives. The proportions among the blue actives are:

$$n\,\mathbb{P}(L=2)/\mathrm{A}_B(t_n+u) \overset{\text{a.s.}}{<} \frac{\pi_B}{2}\,\mathrm{e}^{\lambda u}\,W/\sqrt{n}$$
$$n\,\mathbb{P}(L\geq 3)/\mathrm{A}_B(t_n+u) \overset{\text{a.s.}}{<} \tilde{C}(W,u)/(\sqrt{n})^2 = o(1/\sqrt{n}).$$

Hence, we get that the probability of a label being used more than twice is very small (in which case we should subtract it more than once), even smaller order than p_α, so we will neglect it from now on.

The final conclusion is that the proportion of *case 5* actives among all actives is

$$p_5(t_n+u) \overset{\text{a.s.}}{<} \frac{\pi_B^2}{2}\,\mathrm{e}^{\lambda u}\,W/\sqrt{n} \to 0.$$

Corollary 2.19. *Define* $\mathrm{A}_e(t)$ *the effective size of the active set as follows: we subtract from* $\mathrm{A}(t)$ *the number of ghosts, already explored and multiple active labels, to get the number of different labels in* $\mathcal{A}(t)$*. Then*

$$\mathrm{A}_e(t_n+u)/\mathrm{A}(t_n+u) \xrightarrow{\mathbb{P}} 1.$$

Proof. By the previous arguments, clearly we have then

$$\mathrm{A}_e(t_n+u)/\mathrm{A}(t_n+u) \geq 1 - \sum_{\alpha=1}^{5} p_\alpha(t_n+u)$$
$$\geq 1 - \frac{1}{\sqrt{n}}\mathrm{e}^{\lambda u}\,W\left(\frac{\pi_B}{\lambda} + \pi_B\pi_R + \frac{3\pi_R\pi_B}{\lambda} + \frac{\pi_B^2}{2} + \frac{\lambda+1}{\lambda^2}\right)(1+o(1)) - K^2/n,$$

where K is the finite random variable in Lemma 2.14.

(Note that implicitly, both sides are conditioned on $\mathrm{A}(t)$.) □

3. Connection process

Now that we have a good approximation of the shortest weight tree (SWT) started from a vertex, it provides us a method to observe the shortest weight path between two vertices. Let us give a raw sketch of this method before moving into the details. To find the shortest weight path between vertices U

and V, we grow the shortest weight trees from the two vertices (SWT^U and SWT^V) simultaneously. Since the graph is connected, these two trees must intersect after some time. (Given enough time, even one SWT could occupy the whole graph.) The shortest weight path is determined by the first intersection of the *explored* sets.

For technical issues, we choose another way of handling.

Definition 3.1 (Collision and connection)**.** *We grow* SWT^U*, the shortest weight tree of* U *till time* $t_n = \frac{1}{2\lambda}\log n$*, and freeze it. Then we grow* SWT^V *till "time"* $t_n + u$*. We say that a collision happens when an active vertex in the shortest weight tree of* U *becomes explored in the process of* V*. If a collision happens at vertex* x *at time* $t_n + u$*, then this determines a path between* U *and* V *with length* $2t_n + u + \mathcal{R}_x^U(t_n)$*, where* $\mathcal{R}_x^U(t_n)$ *is the remaining lifetime of* x *in* SWT^U*, and is an* $\mathrm{Exp}(1)$ *random variable. To find the length of the minimal path – i.e., the connection time – we have to minimise the quantity* $2t_n + u + \mathcal{R}_x^U(t_n)$ *among all collision events.*

Obviously, growing SWT^V after SWT^U results in more thinned vertices in SWT^V, since the labels belonging to explored vertices in SWT^U can not be used again, leading to a bigger thinning probability. Hence, we have to show that the additional number of ghosts caused by this effect is still small. Recall (2.7) and note that the proportion of ghosts depends simultaneously on the thinning probability of the i^{th} explored vertex as well as it being an ancestor of a uniform active vertex. Hence, to determine the additional error term, we have to reconsider the computations in Lemma 2.15 and then show that this term also tends to 0.

Claim 3.2. *If we grow* SWT^V *after* SWT^U *on the same graph* $\mathrm{NW}_n(\rho)$ *and aside from the usual thinning, we have to thin the vertices of* SWT^V *that already occurred in* SWT^U*. Then the effective size of the active set in* SWT^V *for times* $t = t_n + u$ *is asymptotically the same as the size of the active set, that is, the statement of Corollary 2.19 remains valid for* SWT^V *as well.*

Note that it suffices to bound the proportion of ghosts, as the error terms caused by vertices that are multiple active, or active as well as explored in the same SWT, are not increased by the presence of another SWT.

Proof. We once again want an upper bound for the probability of a uniform active being ghost, in a different setting. We use union bound and sum up for all explored vertices the probability of being thinned and being an ancestor of the uniform active at the same time. The arguments with the ancestral line (see 2.4.1) remain valid without any modification, the i^{th} splitting particle is an ancestor of a uniform active with probability D_i/S_i. (The random variables D_i and S_i now mean the number of offspring and alives in SWT^V.) We shall examine the change in the thinning probability.

In case the i^{th} explored vertex is blue, its label is chosen uniformly, thus the probability that this label coincides with a previously chosen label equals $\left(\text{N}^U(t_n)+i-1\right)/n$. In case the i^{th} explored is red, we can use the same idea as before: it has the same label of a vertex explored before if and only if two intervals grow into each other with this step. We now consider the two SWTs together, as one set of intervals. Conditioned on the interval that grows, for any interval the probability that these two grow into each other is $2/n$. The number of intervals is now the total number of blue explored vertices in the two SWTs. Summing it up for all intervals gives the upper bound

$$\sum_{k=1}^{\text{N}_B^U(t_n)+i} 2/n = 2/n\left(\text{N}_B^U(t_n)+i\right) \le 2/n\left(\text{N}^U(t_n)+i\right).$$

Then the common bound for the two colors is $2/n\left(\text{N}^U(t_n)+i\right)$.

For the probability of a uniformly chosen active in SWT^V being a ghost, similarly to (2.7), we have

$$\begin{aligned}\text{A}_G^V(t_n+u)/\text{A}^V(t_n+u) &= \sum_{i=1}^{\text{N}^V(t_n+u)} \frac{D_i}{S_i}\cdot\frac{2}{n}\left(\text{N}^U(t_n)+i\right)\\ &= \sum_{i=1}^{\text{N}^V(t_n+u)} \frac{D_i}{S_i}\cdot\frac{2i}{n} + \sum_{i=1}^{\text{N}^V(t_n+u)} \frac{D_i}{S_i}\cdot\frac{2}{n}\text{N}^U(t_n).\end{aligned} \tag{3.1}$$

We already know from Lemma 2.15 that the first sum tends to 0 as n tends to ∞. Let us investigate the second one.

Using the a.s. finite K value implied by Lemma 2.14, we split the sum again.

$$\Sigma_1 = \frac{2}{n}\mathrm{N}^U(t_n)\sum_{i=1}^{K}\frac{D_i}{S_i} \leq \frac{2}{n}\mathrm{N}^U(t_n)\sum_{i=1}^{K}1 = K\frac{2}{\lambda\sqrt{n}}W_U(1+o(1)) \overset{n\to\infty}{\longrightarrow} 0.$$

For the second one,

$$\Sigma_2 = \frac{2}{n}\mathrm{N}^U(t_n)\sum_{i=K+1}^{\mathrm{N}^V(t_n+u)} D_i/S_i = \frac{2}{n}\mathrm{N}^U(t_n)\sum_{i=K+1}^{\mathrm{N}^V(t_n+u)}\frac{D_i}{i\lambda(1+o(1))}$$

We compute the expected value of the sum $\Sigma_3 := \sum_{i=K+1}^{\mathrm{N}^V(t_n+u)}\frac{D_i}{i\lambda(1+o(1))}$ with tower rule, and using $\mathbb{E}[D_i] = \lambda + 1$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=K+1}^{\mathrm{N}^V(t_n+u)}\frac{D_i}{i\lambda(1+o(1))}\right] &= \mathbb{E}\left(\mathbb{E}\left[\sum_{i=K+1}^{\mathrm{N}^V(t_n+u)}\frac{D_i}{i\lambda(1+o(1))}|\mathrm{N}^V(t_n+u)\right]\right)\\
&= \mathbb{E}\left(\sum_{i=K+1}^{\mathrm{N}^V(t_n+u)}\frac{\mathbb{E}[D_i]}{i\lambda(1+o(1))}\right)\\
&= \frac{\lambda+1}{\lambda}(\mathbb{E}[\log \mathrm{N}^V(t_n+u)] - \log(K+1))(1+o(1))
\end{aligned}$$

Since logarithm is concave, we can use Jensen's inequality to obtain

$$\begin{aligned}
\mathbb{E}[\Sigma_3] &\leq \frac{\lambda+1}{\lambda}(\log \mathbb{E}[\mathrm{N}^V(t_n+u)] - \log(K+1))(1+o(1))\\
&= \frac{\lambda+1}{\lambda}\left(\log\left[\tfrac{1}{\lambda}\sqrt{n}W_V\mathrm{e}^{\lambda u}\right] - \log(K+1)\right)(1+o(1))\\
&= \frac{\lambda+1}{\lambda}\left(\lambda u\tfrac{1}{2}\log n\log\left(\tfrac{1}{\lambda}W_V\right) - \log(K+1)\right)(1+o(1))
\end{aligned}$$

Now we can use Markov's inequality to bound this sum w.h.p.:

$$\mathbb{P}\left(\Sigma_3 \geq \mathbb{E}\Sigma_3\log n\right) \leq \log n \overset{n\to\infty}{\longrightarrow} 0.$$

Then for $\Sigma_2 = \frac{2}{n}\mathrm{N}^U(t_n)\Sigma_3$, the following holds with high probability

$$\begin{aligned}
\Sigma_2 &\leq \frac{2}{n}\mathrm{N}^U(t_n)\log n\mathbb{E}\Sigma_3\\
&\leq \frac{2}{n}\mathrm{N}^U(t_n)\log n\frac{\lambda+1}{\lambda}\left(\lambda u\tfrac{1}{2}\log n\log(\tfrac{1}{\lambda}W_V) - \log(K+1)\right)(1+o(1)).
\end{aligned}$$

Substitute $\mathrm{N}^U(t_N)$ from Corollary 2.10:

$$\begin{aligned}\Sigma_2 &\le \frac{2}{n}(1+o(1))\tfrac{1}{\lambda}\sqrt{n}W_U\log n\\ &\quad\cdot\frac{\lambda+1}{\lambda}\left(\lambda u\tfrac{1}{2}\log n\log\left(\tfrac{1}{\lambda}W_V\right)-\log(K+1)\right)(1+o(1))\\ &= \frac{\log^2 n}{\sqrt{n}}(1+o(1))\left(W_U\frac{(\lambda+1)u\log\left(\frac{1}{\lambda}W_V\right)}{\lambda}(1+o(1))\right)\\ &\quad-\frac{\log n}{\sqrt{n}}(1+o(1))\tfrac{2}{\lambda}W_U\log(K+1)\overset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

And this is what we wanted to prove. □

3.1. **The Poisson point process of collisions.** First we show that with respect to the parameter u in t_n+u, the collision events form a non-homogeneous Poisson point process (PPP) for each pair of colours and we compute the total intensity measure of these PPPs. We get a two dimensional PPP by adding the independent exponential remaining lifetimes $\mathcal{R}_x^U(t_n)$ as second coordinate. (See [30] for this Poisson-approximation.)

It is not hard to see that the collisions indeed form Poisson point processes. Because of the continuous lifetime distribution, the probability that two vertices are explored at the same time in the shortest weight tree of V is 0. Each collision event requires a vertex being explored, thus collision events occur at the same time with probability 0. We will see next that the events are almost independent. From the properties of the exploration process, it is obvious that the probability of a new collision only depends on the sizes of the sets of which we observe the intersection, thus depends on the existence of former collisions only through their effect on the set sizes. By a collision event, the set sizes are reduced at most by 2, the proportion of this change and the set size of order $\sqrt{n}$ (see Corollary 2.19) is asymptotically 0, hence the almost independency.

Given that we have a non-homogenous PPP with intensity measure $\mu(t)$, $t\in\mathbb{R}$, the number of points till time t has distribution $\mathrm{Poi}(M(t))=\mathrm{Poi}(\int_{-\infty}^{t}\mu(s))$, where $M(t)$ is the mean function. Since $\mathbb{E}[\mathrm{Poi}(M(t))]=M(t)$, the derivative of $\mathbb{E}\left[\text{number of points till } t\right]$ gives $\mu(t)$.

To determine the intensity measure of the collision process, we will consider the four collision point processes for each possible pair of colours. Since the labels of blue vertices are chosen uniformly, they can meet any color. Considering the growing set of intervals, we see that red can meet red as well.

Let us introduce the notation for $q, r \in \{R, B\}$

$$\begin{aligned} \mathcal{C}_{q,r}(u) &:= \mathcal{N}_q^V(t_n + u) \cap \mathcal{A}_r^U(t_n), \quad \mathrm{C}_{q,r}(u) := |\mathcal{C}_{q,r}(u)|, \\ \mathcal{C}(u) &:= \bigcup_{q,r \in \{R,B\}} \mathcal{C}_{q,r}(u), \quad \mathrm{C}(u) := |\mathcal{C}(u)| \end{aligned} \tag{3.2}$$

(Note that e.g. $\mathcal{C}_{R,B}(u)$ denotes the set of *red* explored labels in SWT^V that are *blue* active in SWT^U.)

The corresponding intensity measures are denoted by $\mu_{q,r}(u)$ and $\mu(u)$.

By Corollary 2.19, we already proved that the effective size of the active set at times $t_n + u$ for $u \in \mathbb{R}$ is asymptotically the same as the size of the active set, so we use this quantity instead in our computations.

We can handle the 3 cases similarly when blue vertices are involved, hence we investigate the specific parts one by one, then we can make the conclusions at once.

Blue-blue collision. By the definition of the set $\mathcal{C}_{B,B}(u)$, we can write

$$\mathrm{C}_{B,B}(u) = \sum_{x \in \mathcal{N}_B^V(t_n+u)} \mathbb{1}\{x \in \mathcal{A}_B^U(t_n)\}.$$

Note that $\mathbb{E}[\mathbb{1}_x | \mathbb{1}_y]$ is of the same order as $\mathbb{E}[\mathbb{1}_x]$: as the elements of $\mathcal{A}_B^U(t_n)$ are chosen uniformly, the success probability $\mathbb{1}_x$ depends only on the size of the set. The condition $\mathbb{1}_y$ only reduces that size by 1, which is asymptotically neglectable compared to the size of the set. Since the dependence is weak, the usual Poisson approximation yields that $\mathrm{C}_{B,B}(u)$ converges to a Poisson random variable for each $u \in \mathbb{R}$. The events $\{x \in \mathcal{N}_B^V(t_n + u)\}, \{x \in \mathcal{A}_B^U(t_n)\}$ are independent by the labeling procedure for blue labels, Wald's equation yields

$$\mathbb{E}[\mathrm{C}_{B,B}(u)] = \mathbb{E}[\mathrm{N}_B^V(t_n + u)] \cdot \mathbb{P}(x \in \mathcal{A}_B^U(t_n)).$$

Blue-red collision. Similarly to the previous case,

$$\mathrm{C}_{B,R}(u) = \sum_{x \in \mathcal{N}_B^V(t_n+u)} \mathbb{1}\{x \in \mathcal{A}_R^U(t_n)\},$$
$$\mathbb{E}[\mathrm{C}_{B,R}(u)] = \mathbb{E}[\mathrm{N}_B^V(t_n+u)] \cdot \mathbb{P}(x \in \mathcal{A}_R^U(t_n)).$$

Red-blue collision. This time we approach the intersection through indicators of the explored set.

$$\mathcal{C}_{R,B}(u) = \mathcal{N}_R^V(t_n+u) \cap \mathcal{A}_B^U(t_n),$$
$$\mathrm{C}_{R,B}(u) = \sum_{x \in \mathcal{A}_B^U(t_n)} \mathbb{1}\{x \in \mathcal{N}_R^V(t_n+u)\}.$$

Notice that even though the labels of $\mathcal{A}_B^U(t_n)$ were chosen earlier, since this set contains blue vertices, the labels are chosen uniformly over all possible labels. Thus the probability of a blue vertex being included in any set L at any time in the process, is proportional to the size of L. This implies

$$\forall x \in \mathcal{A}_B^U(t_n): \quad \mathbb{P}(x \in \mathcal{N}_R^V(t_n+u)) = \mathrm{N}_R^V(t_n+u)/n.$$

Then, like previously, using Wald's equation we have

$$\mathbb{E}[\mathrm{C}_{R,B}(u)] = \mathbb{E}[\mathcal{A}_B^U(t_n)] \cdot \mathbb{P}(x \in \mathcal{N}_R^V(t_n+u)).$$

Using the asymptotic results for $\mathcal{A}(t), \mathcal{N}(t)$ (from Theorem 2.8 and Corollary 2.10) and the definition of $W^{(n)}$ 2.6, then differentiating with respect to u yields

$$\begin{aligned} \mu_{B,B}(u) &= \pi_B^2 W_V^{(n)} W_U^{(n)} \mathrm{e}^{\lambda u}, \\ \mu_{R,B}(u) &= \pi_B \pi_R W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda u}, \\ \mu_{B,R}(u) &= \pi_B \pi_R W_V^{(n)} W_U^{(n)} \mathrm{e}^{\lambda u}. \end{aligned} \tag{3.3}$$

Red-red collision. We consider this in the setting of the intervals on the cycle. Since we froze the cluster of U, we have SWT^U as a fixed set of intervals, $\{I_k, k = 1, ..., \mathrm{N}_B^U(t_n)\}$, while the set of intervals in SWT^V is growing (and possibly merging) with u $\{J_i(u), i = 1, ..., \mathrm{N}_B^V(t_n+u)\}$. A collision happens when one of the intervals J_i grows into one of the intervals I_k.

Note that here we face a new technical issue. When two intervals $J_i(u)$ and I_k collide at time u, in principle we should stop the evolution of $J_i(u)$, that is, for all $u' > u$ we should have $J_i(u') \equiv J_i(u)$. But, note that this would cause

computational difficulties later, since we would need to condition on all the earlier collisions. Hence, it is easier to do the following as an upper bound on the number of red-red collisions: we let $J_i(u)$ grow further and it might collide with more vertices inside I_k. The error term caused by this is negligible, since this is only part of the thinning of SWT^V imposed by SWT^U, which is negligible in the sense of Claim 3.2.

Then we can easily observe all red-red collisions up to time $t_n + u$. Not just the outmost, but any explored red in SWT^V coinciding with an active red in SWT^U means a (potentially past) collision, as long as the centre of $J_i(u)$ is not inside of I_k (in which case that center should have been thinned).

Then for any pair I_k and $J_i(u)$ the probability that they had already made a collision by time $t_n + u$ is

$$\mathbb{P}(J_i(u) \cap I_k \neq \varnothing) = (l_i^V(u) + r_i^V(u))/n, \tag{3.4}$$

where $l_i^V(u)$ and $r_i^V(u)$ mean the left and right radius of $J_i(u)$, similarly as in paragraph 2.4.1. To see this, condition on I_k and all radiuses included. Then the left hand side active red of I_k coinciding with any of the $r_i^V(u)$ explored right hand side reds of $J_i(u)$ means a (potentially past) collision and each determines the position of c_i^V, the centre of $J_i(u)$. Thus the position of c_i^V being among these $r_i^V(u)$ many positions for on the left of I_k means a collision between I_k and $J_i(u)$. Similarly, there are $l_i^V(u)$ many good positions on the right. Using that c_i^V is chosen uniformly over $[n]$ – since it is a blue vertex – yields the right hand side of (3.4).

Then the expected number of red-red collisions up to time $t_n + u$:

$$\begin{aligned}\mathbb{E}[\mathrm{C}_{R,R}(u)] &= \mathbb{E}\left[\sum_{k=1}^{\mathrm{N}_B^U(t_n)} \sum_{i=1}^{\mathrm{N}_B^V(t_n+u)} \mathbb{1}\{I_k \cap J_i(u) \neq \varnothing\}\right] \\ &= \sum_{k=1}^{\mathrm{N}_B^U(t_n)} \sum_{i=1}^{\mathrm{N}_B^V(t_n+u)} \mathbb{P}(I_k \text{ and } J_i(u) \text{ have made a collision})\end{aligned}$$

Substituting, we have

$$\begin{aligned}\mathbb{E}[\mathrm{C}_{R,R}(u)] &= \sum_{k=1}^{\mathrm{N}_B^U(t_n)} \sum_{i=1}^{\mathrm{N}_B^V(t_n+u)} (l_i^V(u) + r_i^V(u))/n \\ &= \frac{\mathrm{N}_B^U(t_n)}{n} \cdot \left[\sum_{i=1}^{\mathrm{N}_B^V(t_n+u)} (l_i^V(u) + r_i^V(u)) \right].\end{aligned}$$

Notice that $\mathrm{A}_R^U(t_n) = 2\mathrm{N}_B^U(t_n)$, since there are two red actives in each I_k. Also the radius of an interval is the number of explored reds belonging to it, hence the sum of all radiuses yields the number of all explored reds.

$$\begin{aligned}\mathbb{E}[\mathrm{C}_{R,R}(u)] &= \frac{1}{2}\mathrm{A}_R^U(t_n)\frac{1}{n} \cdot \mathrm{N}_R^V(t_n+u) \\ &\approx \frac{1}{2}\pi_R W_U^{(n)}\sqrt{n}\frac{1}{n}\pi_R\frac{1}{\lambda}W_V^{(n)}\sqrt{n}\mathrm{e}^{\lambda u} \\ &= \frac{1}{2}\pi_R^2\frac{1}{\lambda}W_U^{(n)}W_V^{(n)}\mathrm{e}^{\lambda u}\end{aligned}$$

Then differentiation yields

$$\mu_{R,R}(u) = \frac{1}{2}\pi_R^2 W_U^{(n)}W_V^{(n)}\mathrm{e}^{\lambda u} \tag{3.5}$$

Heuristically, this factor $\frac{1}{2}$, that doesn't correspond to the other 3 cases, arises because right side reds can only meet left side reds and vice versa, hence only 2 types of red-red collision can occur out of the possible 4. This can be made rigorous: for the half of reds that are left side reds in an interval I_k, the center $J_i(u)$ has to be within $r_i(u)$ steps on the cycle from the leftmost red in I_k, but not included in I_k, for I_k and $J_i(u)$ to determine a collision. For the right side reds, we can say the same with $l_i(u)$. In total, we have half times the sum of right radiuses and one half times the sum of left radiuses in the intervals $J_i(u)$, which is one half times the number of explored reds in SWT^V at time $t_n + u$.

Theorem 3.3 (Total intensity measure of the collision PPP)**.** *The total intensity measure of the collision Poisson point process is*

$$\mu(u) = W_U^{(n)}W_V^{(n)}\mathrm{e}^{\lambda u}\left(\pi_B^2 + 2\pi_B\pi_R + \frac{1}{2}\pi_R^2\right) = W_U^{(n)}W_V^{(n)}\mathrm{e}^{\lambda u}\left(1 - \frac{1}{2}\pi_R^2\right)$$

Proof. We discuss that the processes are asymptotically independent. In the collision process, we actually consider intersections of sets (explored sets of both colors in both SWT's). Dependencies arise from two phenomena: dependence of the set sizes (explored sets in the same SWT), and given the set sizes, the "competition" for colliding vertices. (E.g. for the intersections $\mathcal{N}_B^V \cap \mathcal{N}_B^U$ and $\mathcal{N}_B^V \cap \mathcal{N}_R^U$.)

Among these sets, only the pairs N_B^U, N_R^U and N_B^V, N_R^V are related, and even these only depend weakly: they can gain a new element when an active vertex becomes explored. Then the offspring of the explored vertex appears in the active set, that might change the distribution of the color of the next explored vertex, however for the times we observe, the active set is of order $\sqrt{n}$, compared to this, the change of proportion that constant many new actives cause is asymptotically neglectable. (Several steps can't ruin it either, by the almost sure convergence to the stationary type-distribution.)

Given the sizes of the explored sets (of both colors in both SWT's), each collision event in one of the processes reduces by 1 the number of possible colliding vertices for each other process, which is neglectable compared to the set sizes of order $\sqrt{n}$.

Then by the asymptotical independency, summing up the equations in (3.3) and (3.5) yields the total intensity measure. □

3.2. **Proof of Main Theorem.** It is well known [30] that given a one-dimensional (non-homogenous) Poisson point process, adding identically distributed random variables independent from each other and the PPP, as second coordinates, we get a two-dimensional non-homogeneous Poisson point process.

Then the points $(U_i, Y_i), i = 1, 2, ...$ form a two-dimensional PPP, where the coordinates U_i are the points in the collision PPP with intensity measure $\mu(u)$ and the coordinates Y_i are the remaining lifetimes, i.e., i.i.d. Exp(1) random variables, independent of the U_i-s. The density of the joint intensity measure is $f(u, y) = \mu(u) \cdot \mathrm{e}^{-y}$. Define $\nu(A) = \int_A f(u, y)\mathrm{d}y\mathrm{d}u$, then the number of points in A has distribution $\mathrm{Poi}(\nu(A))$.

To get the shortest path between U and V, recall from Definition 3.1 that we have to minimize the sum of time passed and the remaining lifetimes

over the collision events. Mathematically, we want to minimize the quantity $U_i + Y_i$ over all points (U_i, Y_i) of the PPP defined above. Note that event $\{\min_j U_j + Y_j \geq z\}$ is equivalent to the event that there is no point in the infinite triangle $\Delta(z) = \{(x, y) : y > 0, x + y < z\}$ in the two-dimensional PPP defined above. We calculate

$$\begin{aligned}\nu(\Delta(z)) &= \int_{-\infty}^{z} \int_{0}^{z-u} \mu(u) \cdot \mathrm{e}^{-y} \mathrm{d}y \mathrm{d}u \\ &= \int_{-\infty}^{z} W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda u} \left(1 - \frac{1}{2}\pi_R^2\right) (1 - \mathrm{e}^{-(z-u)}) \mathrm{d}u \\ &= W_U^{(n)} W_V^{(n)} \left(1 - \frac{1}{2}\pi_R^2\right) \frac{1}{\lambda(\lambda+1)} \mathrm{e}^{\lambda z}.\end{aligned}$$

For short, we denote by

$$\mathcal{W}^{(n)} = W_U^{(n)} W_V^{(n)} \left(1 - \frac{1}{2}\pi_R^2\right) \frac{1}{\lambda(\lambda+1)}. \tag{3.6}$$

As well as $\mathcal{W} = W_U W_V \left(1 - \frac{1}{2}\pi_R^2\right) \frac{1}{\lambda(\lambda+1)}$. Then we can reformulate

$$\nu(\Delta(z)) = \mathcal{W}^{(n)} \mathrm{e}^{\lambda u} \tag{3.7}$$

Turn our attention back to $\mathcal{P}_n(U, V)$, the shortest weight path between U and V. By the previous argument, we conclude

$$\begin{aligned}&\mathbb{P}(\mathcal{P}_n(U, V) \geq z + 2t_n | \mathcal{W}^{(n)}) \\ &= \mathbb{P}(\mathrm{Poi}(\nu(\Delta(z))) = 0 | \mathcal{W}^{(n)}) = \exp\{-\nu(\Delta(z))\}\end{aligned} \tag{3.8}$$

Rearranging the left hand side and substituting the computed value of $\nu(\Delta(z))$, we get

$$\mathbb{P}(2t_n - \mathcal{P}_n(U, V) \leq -z | \mathcal{W}^{(n)}) = \exp\{-\mathcal{W}^{(n)} \mathrm{e}^{\lambda z}\}$$

We rearrange the right hand side first.

$$\begin{aligned}\exp\{-\mathcal{W}^{(n)} \mathrm{e}^{\lambda z}\} &= \exp\{-\exp\{\lambda z + \log \mathcal{W}^{(n)}\}\} \\ &= \exp\{-\exp\{-\left(-\lambda z - \log \mathcal{W}^{(n)}\right)\}\}\end{aligned}$$

We rearrange the original left hand side too, to get something similar to the exponential instead of $-z$ in the probability. First, note that we can write $<$

instead of $\leq$ since the involved distributions are continuous, and we use the definition of t_n.

$$\begin{aligned}
&\mathbb{P}(2t_n - \mathcal{P}_n(U,V) \leq -z) = \mathbb{P}\left(2\frac{1}{2\lambda}\log n - \mathcal{P}_n(U,V) < -z\right) \\
&= \mathbb{P}\left(\log n - \lambda\mathcal{P}_n(U,V) < -\lambda z\right) \\
&= \mathbb{P}\left(\log n - \lambda\mathcal{P}_n(U,V) - \log \mathcal{W}^{(n)} < -\lambda z - \log \mathcal{W}^{(n)} | \mathcal{W}^{(n)}\right)
\end{aligned}$$

Comparing these results,

$$\begin{aligned}
&\mathbb{P}\left(\log n - \lambda\mathcal{P}_n(U,V) - \log \mathcal{W}^{(n)} < -\lambda z - \log \mathcal{W}^{(n)} | \mathcal{W}^{(n)}\right) \\
&= \exp\{-\exp\{-\left(-\lambda z - \log \mathcal{W}^{(n)}\right)\}\}
\end{aligned}$$

We can recognize the cumulative distribution function of a standard Gumbel random variable Λ, $\mathbb{P}(\Lambda < x) = \exp(-\exp(-x))$, with $x = -\lambda z - \log \mathcal{W}^{(n)}$. That implies

$$(\log n - \lambda\mathcal{P}_N(U,V) - \log \mathcal{W}^{(n)}) | \mathcal{W}^{(n)} \xrightarrow{d} \Lambda,$$

where Λ has standard Gumbel distribution. Rearranging and substituting $\mathcal{W}^{(n)}$ from (3.6), and using that the martingales $(W_U^{(n)}, W_V^{(n)}) \xrightarrow{a.s.} (W_U, W_V)$,

$$\begin{aligned}
&\mathcal{P}_n(U,V) - \frac{1}{\lambda}\log n \xrightarrow{d} -\frac{1}{\lambda}\Lambda - \frac{1}{\lambda}\log \mathcal{W} \\
&= -\frac{1}{\lambda}\Lambda - \frac{1}{\lambda}\log(W_U W_V) - \frac{1}{\lambda}\log\left(1 - \frac{1}{2}\pi_R^2\right) + \frac{1}{\lambda}\log\left(\lambda(\lambda+1)\right)
\end{aligned}$$

4. Epidemic curve

Recall the definition of the epidemic curve function from 1.4. The discussion of the epidemic curve will consist of three parts: first, we heuristically find the function by computing the expected value of $\mathrm{I}_n(t,U)$. Then we prove the convergence by bounding the second moment. Finally, we give a characterization of M_{W_V}, the moment generating function of the random variable W_V, that determines the epidemic curve function f.

4.1. **Heuristics, expected value.** First, suppose U and the martingale $W_U^{(n)}$ describing its neighborhood are given. Then about the fraction of infected individuals, we can say

$$\begin{aligned}\mathbb{E}\left[\mathrm{I}_n(t,U)\,\middle|W_U^{(n)}\right] &= \frac{1}{n}\sum_{w\in[n]}\mathbb{E}\left[\mathbb{1}(w \text{ is infected by time t})\,\middle|W_U^{(n)}\right]\\ &= \frac{1}{n}\sum_{w\in[n]}\mathbb{P}\left(w \text{ is infected by time t}\,\middle|W_U^{(n)}\right).\end{aligned}$$

We take the average of the probability of being infected over all vertices, this is the same as the probability that a uniform vertex is infected. Then

$$\begin{aligned}\mathbb{E}\left[\mathrm{I}_n(t,U)\,\middle|W_U^{(n)}\right] &= \mathbb{P}\left(V \text{ uniform vertex is infected by time t}\,\middle|W_U^{(n)}\right)\\ &= \mathbb{P}\left(\mathcal{P}_n(U,V)\le t\,\middle|W_U^{(n)}\right).\end{aligned}$$

The distribution of $\mathcal{P}_n(U,V)$ is known conditioned on W_U and W_V. Since only U is given and V is our variable, we use total law of expectation with respect to V. Finally we have

$$\mathbb{E}\left[\mathrm{I}_n(t,U)|W_U^{(n)}\right] = \mathbb{E}\left[\mathbb{P}\left(\mathcal{P}_n(U,V)\le t|W_V^{(n)},W_U^{(n)}\right)\,\middle|W_U^{(n)}\right]$$

We already know from 3.2 that

$$\begin{aligned}&\mathbb{P}\left(\mathcal{P}_n(U,V)\ge z+2t_n|\,W_U^{(n)},W_V^{(n)}\right)\\ &= \mathbb{P}\left(\mathcal{P}_n(U,V)\ge z+\tfrac{1}{\lambda}\log n\big|\,W_U^{(n)},W_V^{(n)}\right)\\ &= \exp\left\{-W_U^{(n)}W_V^{(n)}\left(1-\tfrac{1}{2}\pi_R^2\right)\tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda z}\right\}.\end{aligned}\tag{4.1}$$

We use the transform $z=t-\frac{1}{\lambda}\log W_U^{(n)}$ and take the complement event. (Note that $\mathcal{P}_n(U,V)=t$ has 0 probability.)

$$\begin{aligned}&\mathbb{P}\left(\mathcal{P}_n(U,V)\le t-\tfrac{1}{\lambda}\log W_u^{(n)}+\tfrac{1}{\lambda}\log n\,\middle|W_V^{(n)},W_U^{(n)}\right)\\ &= 1-\exp\left\{-W_U^{(n)}W_V^{(n)}\left(1-\tfrac{1}{2}\pi_R^2\right)\frac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\frac{1}{W_U^{(n)}}\right\}\\ &= 1-\exp\left\{-W_V^{(n)}\left(1-\tfrac{1}{2}\pi_R^2\right)\frac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\right\}\end{aligned}$$

Then substituting this into (4.1),

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{I}_n(t-\tfrac{1}{\lambda}\log W_U^{(n)}+\tfrac{1}{\lambda}\log n,U)|W_U^{(n)}\right]\\
&=\mathbb{E}\left[1-\exp\left\{-W_V^{(n)}\left(1-\tfrac{1}{2}\pi_R^2\right)\tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\right\}\middle| W_U^{(n)}\right]\\
&=1-\mathbb{E}\exp\left\{-W_V^{(n)}\left(1-\tfrac{1}{2}\pi_R^2\right)\tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\right\}.
\end{aligned}$$

We recognize that the second term is the moment generating function of $W_V^{(n)}$, $\mathrm{M}_{W_V^{(n)}}(x)$, at $x(t)=-\left(1-\frac{1}{2}\pi_R^2\right)\frac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}$.
Then $\mathbb{E}\left[\mathrm{I}_n(t-\frac{1}{\lambda}\log W_U^{(n)}+\frac{1}{\lambda}\log n,U)\right]=1-M_{W_V^{(n)}}(x(t))$, which is equivalent to

$$\mathbb{E}[\mathrm{I}_n(t+\tfrac{1}{\lambda}\log n,U)]=1-M_{W_V^{(n)}}\left(x(t+\tfrac{1}{\lambda}\log W_U^{(n)})\right).$$

Note that $W_U^{(n)}$ converges to W_U almost surely, which implies convergence in probability of their moment generating functions. Then it's natural to conjecture the epidemic curve function to be $f(t)=1-M_{W_V}(x(t))$.

4.2. **Rigorous proof.** The heuristics above showed that the expected value of $\mathrm{I}_n(t,U)$ indeed converges in probability to the defined f function at the given place. We prove the theorem by showing that the variation of $\mathrm{I}_n(t,U)$ converges to 0, then Chebyshev's inequality yields that $\mathrm{I}_n(t,U)$ converges to its expected value in probability.

Denote by $\mathbb{1}_i=\mathbb{1}\{i \text{ is infected}\}$. Let us calculate

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_i\right)\\
&=\frac{1}{n^2}\sum_{i\in[n]}\mathrm{Var}\mathbb{1}_i+\frac{2}{n^2}\sum_{i<j\in[n]}\mathrm{Cov}[\mathbb{1}_i,\mathbb{1}_j]
\end{aligned}$$

Since $\mathbb{1}_i$ is an indicator, $\mathrm{Var}\mathbb{1}_i\le 1$, hence $\frac{1}{n^2}\sum_{i\in[n]}\mathrm{Var}\mathbb{1}_i\le\frac{1}{n}$.

$\mathrm{Cov}[\mathbb{1}_i,\mathbb{1}_j]$ $= \mathbb{E}[\mathbb{1}_i\mathbb{1}$

$=\mathbb{P}(i \text{ and } j \text{ are both are infected}|W_U^{(n)})-\mathbb{P}(i \text{ is infected}|W_U^{(n)})\mathbb{P}(j \text{ is infected}|W_U^{(n)})$

Imagine now three exploration processes on NW_n, one from U, one from i and one from j. We want a connection simultaneously between SWT^U and SWT^i

as well as between SWT^U and SWT^j. Recall the coupling between the graph and the branching process, as well as the thinning inside one tree and between the trees, all have error terms of order $1/\log n$. This also applies if there are three SWT's, the error terms increase by constant multiples. These connection processes between SWT^U and the other two are related only through the intersection of SWT^i and SWT^j, which is of order $1/\log n$, which means they are asymptotically independent. Then $\mathbb{P}(i \text{ and } j \text{ are both are infected}|W_U^{(n)})$ is close to $\mathbb{P}(i \text{ is infected}|W_U^{(n)})\mathbb{P}(j \text{ is infected}|W_U^{(n)})$, the error term, which is the covariance, is of order $1/\log n$.

This works if i and j are fairly apart, say $(i-j) \bmod n > \log n$. The number of "bad pairs", which are close is $n \log n/2$, compared to the number of all pairs $\binom{n}{2}$, the fraction goes to 0. Even for these, the covariance is bounded by 1. Then the sum divided by n^2 goes to 0.

With that, we have bounded the variance by a term that goes to 0, which finishes the proof.

4.3. **Characterization of the epidemic curve function.** We start with the well-known recursion formula for the number of alive particles in the branching process arising from the branching property:

$$\begin{aligned}\mathbf{Z}^{(B)}(t) = &\sum_{i=1}^{D_R^{(B)}} \mathbb{1}\{X_i > t\} + \mathbb{1}\{X_i < t\}\mathbf{Z}^{(R)}(t - X_i) \\ &+ \sum_{j=1}^{D_B^{(B)}} \mathbb{1}\{X_j > t\} + \mathbb{1}\{X_j < t\}\mathbf{Z}^{(B)}(t - X_i),\end{aligned}$$

and similarly for the split of a red particle. Note that the branching property guarantees that all random variables on the right hand side are independent.

(Recall that $\mathbf{Z}^{(q)}(t)$ means the vector of alive red and blue particles at time t in the process started with a single particle of type q, $D_r^{(q)}$ means the number of r-type children of a particle with type q, and X_i is the Exp(1) lifetime of a particle.)

Multiplying both sides with $e^{-\lambda t}$ we get

$$\begin{aligned}\mathrm{e}^{-\lambda t}\mathbf{Z}^{(B)}(t) = &\sum_{i=1}^{D_R^{(B)}} \mathrm{e}^{-\lambda t}\mathbb{1}\{X_i > t\} + \mathbb{1}\{X_i < t\}\mathrm{e}^{-\lambda X_i}\mathrm{e}^{-\lambda(t-X_i)}\mathbf{Z}^{(R)}(t-X_i)\\ &+ \sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda t}\mathbb{1}\{X_j > t\} + \mathbb{1}\{X_j < t\}\mathrm{e}^{-\lambda X_j}\mathrm{e}^{-\lambda(t-X_j)}\mathbf{Z}^{(B)}(t-X_i).\end{aligned}$$

We then let t go to infinity. The left hand side converges to $W^{(B)}$. $\mathbb{1}\{X_i > t\}$ is bounded, hence multiplied with $\mathrm{e}^{-\lambda t}$, it goes to 0. $\mathbb{1}\{X_i < t\}$ goes to 1, as Exp(1) is always finite. $\mathrm{e}^{-\lambda(t-X_j)}\mathbf{Z}^{(B)}(t-X_i)$, since we subtract the same X_i, converges to $W_i^{(B)}$, which has the same distribution as $W^{(B)}$. The other terms can be handles similarly, altogether we have

$$W^{(B)} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{D_R^{(B)}} \mathrm{e}^{-\lambda X_i} W_i^{(R)} + \sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda X_j} W_j^{(B)},$$

where $W_i^{(R)}$ are independent copies of $W^{(R)} = \lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z^{(R)}(t)$, and $W_j^{(B)}$ are independent copies of $W^{(B)} \stackrel{\mathrm{d}}{=} W_V$. Denote the moment generating functions of $W^{(B)}$ and $W^{(R)}$ respectively $M_{W^{(B)}}$, $M_{W^{(R)}}$.

To obtain functional equations for the moment generation functions, we apply the function $x \mapsto \mathbb{E}[\mathrm{e}^{\vartheta x}]$ to both sides. Recall that a red particle has one red child and Poi(ρ) many blue, while a blue has two red children and Poi(ρ) many blue. From the independency, guaranteed by the branching property, expectation factorizes.

$$\mathbb{E}\mathrm{e}^{\vartheta W^{(B)}} = \left(\mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda X_i} W^{(R)}\}\right]\right)^2 \cdot \mathbb{E}\left(\exp\left\{\vartheta \textstyle\sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda X_j} W_j^{(B)}\right\}\right)$$

We use law of total expectation with respect to X_i to compute

$$\begin{aligned}J^{(R)} :&= \mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda X_i} W^{(R)}\}\right] = \int_0^\infty \mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda x} W^{(R)}\}\right] \mathrm{e}^{-x}\mathrm{d}x\\ &= \int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x\end{aligned}$$

Let $J^{(B)}$ defined similarly with $M_{W^{(B)}}$.

We apply tower rule for the last term:

(4.2)
$$\mathbb{E}\left(\exp\left\{\vartheta\sum_{j=1}^{D_B^{(B)}}\mathrm{e}^{-\lambda X_j}W_j^{(B)}\right\}\right)=\mathbb{E}\left(\mathbb{E}\left[\exp\left\{\vartheta\sum_{j=1}^{D_B^{(B)}}\mathrm{e}^{-\lambda X_j}W_j^{(B)}\right\}\middle|\, D_B^{(B)}\right]\right)$$

The moment generating function of a non-random sum of independent random variables is the product of the moment generating functions.

$$\mathbb{E}\left[\exp\left\{\vartheta\sum_{j=1}^{D_B^{(B)}}\mathrm{e}^{-\lambda X_j}W_j^{(B)}\right\}\middle|\, D_B^{(B)}\right]=\prod_{j=1}^{D_B^{(B)}}\mathbb{E}\left[\exp\{\vartheta\mathrm{e}^{-\lambda X_j}W_j^{(B)}\}\right]=\prod_{j=1}^{D_B^{(B)}}J^{(B)}$$
$$=\left(J^{(B)}\right)^{D_B^{(B)}}.$$

Substitute into (4.2):

$$\mathbb{E}\left[\exp\left\{\vartheta\sum_{j=1}^{D_B^{(B)}}\mathrm{e}^{-\lambda X_j}W_j^{(B)}\right\}\right]=\mathbb{E}\left[\left(J^{(B)}\right)^{D_B^{(B)}}\right].$$

Above is the probability generating function of $D_B^{(B)}\overset{\mathrm{d}}{=}\mathrm{Poi}(\rho)$, hence

$$\mathbb{E}\left[\exp\left\{\vartheta\sum_{j=1}^{D_B^{(B)}}\mathrm{e}^{-\lambda X_j}W_j^{(B)}\right\}\right]=\exp\left\{\rho\left(J^{(B)}-1\right)\right\}$$

We can rewrite this as

$$\begin{aligned}J^{(B)}-1&=\int_0^\infty M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x-1\\&=\int_0^\infty M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x-\int_0^\infty\mathrm{e}^{-x}\mathrm{d}x\\&=\int_0^\infty\left(M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})-1\right)\mathrm{e}^{-x}\mathrm{d}x\end{aligned}$$

We can now conclude

$$M_{W^{(B)}}(\vartheta)$$
$$=\left(\int_0^\infty M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x\right)^2\cdot\exp\left\{\rho\left(\int_0^\infty\left(M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})-1\right)\mathrm{e}^{-x}\mathrm{d}x\right)\right\}$$

Similarly for $M_W^{(R)}$, using that $D_R^{(R)}=1$ and $D_B^{(R)}\overset{\mathrm{d}}{=}\mathrm{Poi}(\rho)$,

$$M_{W^{(R)}}(\vartheta)$$
$$=\left(\int_0^\infty M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x\right)\cdot\exp\left\{\rho\left(\int_0^\infty\left(M_{W^{(R)}}(\vartheta\mathrm{e}^{-\lambda x})-1\right)\mathrm{e}^{-x}\mathrm{d}x\right)\right\}$$

The moment generating functions are described by these functional equations, and according to [4], there exist proper moment generating functions satisfying these functional equations.

5. Central limit theorem for hopcount

Recall that the hopcount $\mathrm{H}_n(U,V)$ is the number of edges along the shortest weight path between U and V, and Theorem 1.2 claims it follows normal distribution with mean and variance both $\frac{\lambda+1}{\lambda}\log n$.

We consider the shortest weight path in two parts: the paths from each U and V to the vertex where the connection happens, denote it by Y. These paths are disjoint with the exception of Y, hence if suffices to determine their lengths. Denote by $G^{(U)}(Y)$ the generation of Y in SWT^U, similarly for V. Then the required steps from the root U to Y is exactly $G^{(U)}(Y)$.

Claim 5.1. *The choice of Y is asymptotically independent in the two* SWT*'s.*

Proof. Conditioned on Y being the connecting vertex, it is uniformly chosen over the active set of SWT^V. That determines its label, and it determines which particle is chosen in SWT^U *through the label*. Since the labeling is independent of the structure of the family tree, aside from the thinning, the choice of Y in SWT^U is independent from its choice in SWT^V. We already bounded the fraction of ghost particles (those who have one of their ancestors thinned) by a term that goes to 0 in 2.15, hence asymptotic independence holds. □

With these notations, $\mathrm{H}_n(U,V) = G^{(U)}(Y) + G^{(V)}(Y)$, and the two terms are independent. We reformulate the theorem using these terms:

$$\begin{aligned}\frac{\mathrm{H}_n(U,V) - \frac{\lambda+1}{\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}} &= \frac{G^{(U)}(Y) + G^{(V)}(Y) - \frac{\lambda+1}{\lambda}\log n}{\sqrt{\frac{\lambda+1}{la}\log n}} \\ &= \frac{G^{(U)}(Y) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}} + \frac{G^{(V)}(Y) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}}\end{aligned}$$

Considering that the terms are independent, it suffices to show that both $\frac{G^{(U)}(Y) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}}$ and $\frac{G^{(V)}(Y) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{\lambda}\log n}}$ has normal distribution with mean 0 and

variance $\frac{1}{2}$, which is equivalent to $\frac{G^{(U)}(Y)-\frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$ and $\frac{G^{(V)}(Y)-\frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$ having standard normal distribution.

5.1. **Generation of connecting vertex in** SWT^V**.** Since SWT^V grows, the connecting vertex is a splitting particle at some splitting time T_k, which is always chosen uniformly over the active set. Then we can recall Bühler's decomposition, that we already used in 2.4.1, for Y's generation as $G_k = \sum -i = 1^k \mathbb{1}_i$, where conditioned on the offspring variables D_i, the indicators are independent and have success probability $\mathbb{P}(\mathbb{1}_i = 1) = \frac{D_i}{S_i}$.

In our case the number of splits is a random variable. We run SWT^V till time $t_n + u$ for some u that is given with the coordinates of the connection point in the two-dimensional PPP (see 3.2), hence there are $\mathrm{N}(t_n + u)$ many explored vertices. We intentionally chose t_n such that $\mathrm{N}(t_n)$ would be of order $\sqrt{n}$, and u is a bounded random variable. (This can be proven by letting $u \to \infty$ in (3.7): the measure of the infinite triangle goes to infinity, hence the probability that it contains no point, that is, the connection does not happen up to $t_n + u$, goes to 0.) Hence we will write $C\sqrt{n}$ instead of $\mathrm{N}(t_n + u)$. C is then a random variable and might depend on n, but is bounded w.h.p.

We decompose

$$\frac{\sum_{i=1}^{C\sqrt{n}} \mathbb{1}_i - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = \frac{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i} - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$
$$+ \frac{\sum_{i=1}^{C\sqrt{n}} \mathbb{1}_i - \sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}}{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}} \cdot \frac{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

Denote by

$$B_1 = \frac{\sum_{i=1}^{C\sqrt{n}} \mathbb{1}_i - \sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}}{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}}$$

$$B_2 = \frac{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

$$B_+ = \frac{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i} - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

Our aim is to show that Lindeberg's CLT is applicable for B_1, while B_2 converges to 1, and B_3 converges to 0.

5.1.1. *Term B_1.* We have a sum of independent indicators, hence it is indeed the expectation and deviation that we subtract and divide with. We only need to check Lindeberg's condition for the central limit theorem to hold. IN this case, Lindeberg's condition is

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \sum_{i=1}^{C\sqrt{n}} \mathbb{E}\left[\mathbb{1}_i^2 \cdot \mathbb{1}\left\{|\mathbb{1}_i - D_i/S_i| > \varepsilon \sum_{i=1}^{c\sqrt{n}} D_i/S_i\,(1 - D_i/S_i)\right\}\right] < \infty$$

Since $\mathbb{1}_i$ is a Bernoulli random variable, $|\mathbb{1}_i - D_i/S_i| \leq 1$. Then $\sum_{i=1}^{C\sqrt{n}} D_i/S_i\,(1 - D_i/S_i) \to \infty$ would ensure Lindeberg's condition, since $\forall\varepsilon\exists n_\varepsilon$ such that $\varepsilon \cdot \sum_{i=1}^{c\sqrt{n_\varepsilon}} D_i/S_i\,(1 - D_i/S_i) > 1$, then we have a 0 sum for all $n \geq n_\varepsilon$, thus the limit is $0 < \infty$.

We give a lower bound for the sum of variances.

$$\sigma_n^2 = \sum_{i=1}^{C\sqrt{n}} D_i/S_i\,(1 - D_i/S_i) = \sum_{i=1}^{C\sqrt{n}} D_i/S_i - \sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2.$$

Recall (future citation: lemma about S_i), and split the sum according to the random variable K. Each vertex has at least least on red child, $D_i \geq 1$. Then

$$\sum_{i=1}^{C\sqrt{n}} D_i/S_i = \sum_{i=1}^{K} D_i/S_i + \sum_{i=K+1}^{C\sqrt{n}} D_i/S_i$$
$$\geq \sum_{i=1}^{K} D_i/S_i + \sum_{i=K+1}^{C\sqrt{n}} \frac{1}{i\lambda(1+o_i(1))}.$$

Since K is a.s. finite, the first term is finite. By Janson (.....), $o_i(1) \leq i^{-1/2}$, we have $i(1+o_i(1)) \leq i + \sqrt{i} < 2i$, which implies that $\frac{1}{i(1+o_i(1))} \geq \frac{1}{2i}$, and thus the second term tends to infinity with order at least $\log n$. Observe now $\sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2$, that is the second term in σ_n^2. Since $D_i = Poi(\rho) + 1 + \mathbb{1}\{i^{\text{th}} \text{ explored is blue}\}$, and the terms are independent, $\text{Var}D_i \leq \rho+0+1 < \infty$, then D_i has finite second moment which we denote by M_2.

$$\sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2 = \sum_{i=1}^{K} D_i^2/S_i^2 + \sum_{i=K+1}^{C\sqrt{n}} D_i^2/S_i^2$$
$$= \sum_{i=1}^{K} D_i^2/S_i^2 + \sum_{i=K+1}^{C\sqrt{n}} D_i^2/\left(i^2\lambda^2(1+o(1))^2\right).$$

The first term is a.s. finite. For the second one, we apply Markov's inequality, also using that the expectation is linear.

$$\mathbb{P}\left[\sum_{i=K+1}^{C\sqrt{n}} D_i^2/\left(i^2\lambda^2(1+o(1))^2\right) \geq \sum_{i=K+1}^{C\sqrt{n}} M_2/\left(i^2\lambda^2(1+o(1))^2\right) \cdot \log\log n\right]$$
$$\leq 1/\log\log n,$$

which tends to 0 as n does to infinity. Hence with high probability,

$$\sum_{i=K+1}^{C\sqrt{n}} D_i^2/\left(i^2\lambda^2(1+o(1))^2\right) \leq \frac{\pi^2}{6}\frac{1}{\lambda^2} M_2 \log\log n.$$

This means we subtract something which is w.h.p a much smaller term, then σ_n^2 goes to infinity w.h.p.

5.1.2. *Term B_2.*

$$\frac{\sum_{i=1}^{C\sqrt{n}} D_i/S_i(1-D_i/S_i)}{\frac{\lambda+1}{2\lambda}\log n}$$
$$= \frac{\sum_{i=1}^{K} D_i/S_i(1-D_i/S_i)}{\frac{\lambda+1}{2\lambda}\log n} + \frac{\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i}{\frac{\lambda+1}{2\lambda}\log n} - \frac{\sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2}{\frac{\lambda+1}{2\lambda}\log n}.$$

The first fraction tends to 0, as the numerator is a.s. finite and the denominator tends to infinity. We showed in the previous subsubsection that the nominator of the third fraction is w.h.p. of order $\log\log n$, which devided by $\log n$ goes to 0. We have yet to show that the second term goes to 1. Let $\mathcal{F}_n = \sigma(D_1, ..., D_n)$ the filtration generated by the random variables D_i. Then

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i}{\frac{\lambda+1}{2\lambda}\log n} = \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{i\lambda(1+o(1))} - \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}}{\frac{\lambda+1}{2\lambda}\log n} + \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}}{\frac{\lambda+1}{2\lambda}\log n}.$$

For $\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{i\lambda(1+o(1))} - \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}$, we want to use Chebyshev's inequality. We compute the second moment, which is the same as the variance. The expected value, by tower rule, equals to 0.

$$\mathbb{E}\left[\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{i\lambda(1+o(1))} - \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}\right)^2\right]$$
$$= \sum_{i,j=K+1}^{C\sqrt{n}} \frac{\mathbb{E}\left[(D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})(D_j - \mathbb{E}[D_j|\mathcal{F}_{j-1})]\right]}{\lambda^2 ij(i+o(1))}$$

In case $j < i$, by tower rule,

$$\mathbb{E}\left[(D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})(D_j - \mathbb{E}[D_j|\mathcal{F}_{j-1})]\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[(D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})(D_j - \mathbb{E}[D_j|\mathcal{F}_{j-1})]|\mathcal{F}_j\right]\right]$$

$(D_j - \mathbb{E}[D_j|\mathcal{F}_{j-1})$ is measurable with respect to $\mathcal{F}_j$, then we have

$$\mathbb{E}\left[(D_j - \mathbb{E}[D_j|\mathcal{F}_{j-1})\mathbb{E}\left[D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]|\mathcal{F}_j\right]\right].$$

This equals to 0, since $j < i$ implies $\mathcal{F}_j \subseteq \mathcal{F}_{i-1}$ and we can use tower rule again:

$$\mathbb{E}\left[D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]|\mathcal{F}_j\right] = \mathbb{E}[D_i|\mathcal{F}_j] - \mathbb{E}\left[\mathbb{E}[D_i|\mathcal{F}_{i-1}]|\mathcal{F}_j\right]$$
$$= \mathbb{E}[D_i|\mathcal{F}_j] - \mathbb{E}[D_i|\mathcal{F}_j] = 0.$$

Then the sum is only the sum of conditional variances of D_i's, which is simply the sum of variances, as the D_i's are independent. We earlier gave the bound $\mathrm{Var}D_i \leq \rho + 1$.

$$\mathbb{E}\left[\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{i\lambda(1+o(1))} - \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}\right)^2\right] = \sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})^2]}{\lambda^2 i^2(1+o(1)))}$$
$$\leq \frac{\pi^2}{6}\frac{1}{\lambda^2}(\rho+1)(i+o(1)).$$

Then Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\sum_{i=K+1}^{C\sqrt{n}} (D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})\right| \geq \log\log n \cdot \frac{\pi^2}{6}\frac{1}{\lambda^2}(\rho+1)(1+o(1))\right) \tag{5.1}$$
$$\leq \frac{1}{(\log\log n)^2} \overset{n\to\infty}{\longrightarrow} 0.$$

Which means $\sum_{i=K+1}^{C\sqrt{n}}(D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1})$ is w.h.p. $o(\log n)$ and tends to 0 devided by $\log n$.

Now to show that $\frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}}{\frac{\lambda+1}{2\lambda}\log n}$ tends to 1, we compute

$$\mathbb{E}[D_i|F_{i-1}] = \mathbb{E}D_i^B \frac{S_{i-1}^B}{S_{i-1}} + \mathbb{E}D_i^R \frac{S_{i-1}^R}{S_{i-1}}$$

It is known that $\left(\frac{S_{i-1}^R}{S_{i-1}}, \frac{S_{i-1}^B}{S_{i-1}}\right)$ converges to (π_R, π_B) and by Janson (.....) the error term is $o_i(1) = i^{-1/2}$. Then

$$\mathbb{E}[D_i|F_{i-1}] = (\rho+2)\pi_B(1+o_i(1)) + (\rho+1)\pi_R(1+o_i(1))$$
$$= (\lambda+1)(1+o_i(1))$$

Subsituting, we get

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}}{\frac{\lambda+1}{2\lambda}\log n} = \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{(\lambda+1)(1+o_i(1))}{i\lambda(1+o(1))}}{\frac{\lambda+1}{2\lambda}\log n}$$
$$= \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{(1+o_i(1))}{i}}{\frac{1}{2}\log n} = \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{1}{i}}{\frac{1}{2}\log n} + \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{o_i(1)}{i}}{\frac{1}{2}\log n}$$

Where, introducing a small constant error term δ from the integral approximation,

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{1}{i}}{\frac{1}{2}\log n} = \frac{\log(C\sqrt{n}) - \log(K+1) + \delta}{\frac{1}{2}\log n}$$
$$= 1 + \frac{\log C - \log(K+1) + \delta}{\frac{1}{2}\log n} \overset{n\to\infty}{\longrightarrow} 1.$$

Since $o_i(1)$ is of order $i^{-1/2}$, we can say for a small $\varepsilon < \frac{1}{2}$ that $o_i(1) \leq i^{-\frac{1}{2}+\varepsilon}$. Then $\sum_{i=K+1}^{C\sqrt{n}} \frac{o_i(1)}{i} \leq \sum_{i=0}^{\infty} i^{-3/2+\varepsilon}$ is summable and finite, devided by $\log n$ it tends to 0.

5.1.3. *Term B_3.* As usual, we split the sum:

$$\frac{\left(\sum_{i=1}^{C\sqrt{n}} D_i/S_i\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = \frac{\sum_{i=1}^{K} D_i/S_i}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} + \frac{\left(\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

Since $D_i/S_i \leq 1$, the first fraction is something finite over $\sqrt{\log n}$ and tends to 0.

$$\frac{\left(\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = \frac{\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{i\lambda(1+o(1))}\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$
$$= \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} + \frac{\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

The nominator $\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}$ was previously proven in (5.1) to be w.h.p of order $\log\log n$, hence devided by $\sqrt{\log n}$ it goes to 0.

We also calculated $\mathbb{E}[D_i|\mathcal{F}_{i-1}] = (\lambda+1)(1+o_i(1))$. Then

$$\frac{\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(1))}\right) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = \frac{\lambda+1}{\lambda}\frac{\left(\sum_{i=K+1}^{C\sqrt{n}}(1+o_i(1))/i\right) - \frac{1}{2}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$
$$= \frac{\lambda+1}{\lambda}\left(\frac{\sum_{i=K+1}^{C\sqrt{n}} 1/i - \frac{1}{2}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} + \frac{\sum_{i=K+1}^{C\sqrt{n}} o_i(1)/i}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}\right)$$

The $o_i(1)$ terms are the same as before, hence devided by i they are summable, the second term tends to 0. Using integral approximation and introducing the

small constant error term δ,

$$\left(\sum_{i=K+1}^{C\sqrt{n}} 1/i\right) - \frac{1}{2}\log n = \big(\delta + \log(C\sqrt{n}) - \log(K+1)\big) - \frac{1}{2}\log n$$
$$= \delta + \log C - \log(K+1)$$

which is indeed $o(\sqrt{\log n})$.

5.2. **Generation of connecting vertex in** SWT^U**.** For the frozen side, we have to use a different approach, since conditioned on that the connection happens, the choice of Y is *not* uniform over the actives. One can calculate (we will omit this) by law of total probability, splitting by the type of connection (blue-blue, red-blue, blue-red, left red-right red, right red-left red), that any active red label in SWT^U is chosen with asymptotic probability $\frac{1}{\mathrm{A}^{(U)}(t_n)} \frac{1-\frac{\pi_R}{2}}{1-\frac{\pi_R^2}{2}}$, while any blue label is chosen with asymptotical probability $\frac{1}{\mathrm{A}^{(U)}(t_n)} \frac{1}{1-\frac{\pi_R^2}{2}}$. (We replaced the fractions $\frac{\mathrm{A}_R^{(V)}(t_n+u)}{\mathrm{A}^{(V)}(t_n+u)}$ by their limit, π_R.) However,

Claim 5.2. *Conditioned on the connecting vertex having a blue label in* SWT^U*, asymtotically it is a uniformly chosen particle with a blue label. Same holds for red: if the connecting vertex has red label, it is uniform over the red active particles.*

Note the difference between being a red label and a red particle: there can be multiple active particles with the same label.

Proof. Obviously, when we calculate the probability of a blue label being chosen, we get the same probability for any blue label. (We do not know, hence cannot use anything specific.) It yields that the label is uniform over the possible blue labels. Now consider different actives with the same label V_i, suppose there is m_i many of them. We call the number m_i the multiplicity of the label. Out of these m_i many particles, the connecting vertex is the one with the least remaining lifetime. The remaining lifetimes are independent Exp(1) random variables, hence their minimum is uniformly chosen, meaning given the label, each instance is chosen with probability $\frac{1}{m_i}$. Then a blue particle which has

multiplicity m_i has probability $\frac{1}{m_i}\frac{1}{\mathrm{A}^{(U)}(t_n)}\frac{1}{1-\frac{\pi_R^2}{2}}$ to be the connecting vertex. We claim that asymptotically, this leads to a uniform distribution. Corollary 2.19 implies that the fraction of multiple actives tends to 0 (at time $t_n + u$, as n goes to infinity). Hence if we pick an active vertex, it has multiplicity 1 w.h.p., including red and blue instances. (This also implies that asymptotically, the label has a well defined color and if it is blue, its possible red instances will not add up to its probability.) Hence asymptotically, we have a uniform distribution.
The argument is very similar in the case of red. □

We define the color sets $\mathcal{CS}_1 = \{R\}$ and $\mathcal{CS}_2 = \{B\}$. For a color set $\mathcal{CS}$, let $\mathcal{A}_{\mathcal{CS}} = \cup_{q\in\mathcal{CS}}\mathcal{A}_q$ the set of actives with any color from the color set. The previous claim can be reformalized like this: for both of these sets $\mathcal{CS}_1$ and $\mathcal{CS}_2$, conditioned on a particle being in $\mathcal{A}_{\mathcal{CS}_i}$, the particle is asymptotically uniform over $\mathcal{A}_{\mathcal{CS}_i}$. Then we can apply Khralamov's [24] results. He proves that if an active particle is uniform over $\mathcal{A}_{\mathcal{CS}}$ for some color set $\mathcal{CS}$, the generation of that particle admits a central limit theorem. Also the mean and variance in this CLT does *not* depend on the given color set, implying it is the same for all such color sets.

Note that when we discussed $G^{(V)}(Y)$, we already proved a CLT for the same branching process, when the particle is uniformly chosen over $\mathcal{A}_{\{R,B\}}$, with mean and variance both $\frac{\lambda+1}{2\lambda}\log n$. Which means that whether Y is red or blue in SWT^U, its generation $G^{(U)}(Y)$ admits a central limit theorem with mean and variance $\frac{\lambda+1}{2\lambda}\log n$. This completes the proof.

References

[1] L. Addario-Berry and T. Lei. The mixing time of the newman: Watts small world. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 1661–1668. SIAM, 2012.

[2] T. Antunović, Y. Dekel, E. Mossel, and Y. Peres. Competing first passage percolation on random regular graphs. *ArXiv e-prints*, Sept. 2011.

[3] S. Asmussen. Almost sure behavior of linear functionals of supercritical branching processes. *Transactions of the American Mathematical Society*, 231(1):233–248, 1977.

[4] K. Athreya and P. Ney. *Branching Processes*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer Berlin Heidelberg, 1972.

[5] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[6] A.-L. Barabási, E. Ravasz, and T. Vicsek. Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 299(3–4):559 – 564, 2001.

[7] A. D. Barbour and G. Reinert. Small worlds. *Random Structures & Algorithms*, 19(1):54–74, 2001.

[8] A. D. Barbour and G. Reinert. Discrete small world networks. *eprint arXiv:cond-mat/0304020*, Apr. 2003.

[9] A. D. Barbour and G. Reinert. Approximating the epidemic curve. *Electron. J. Probab*, 18(54):1–30, 2013.

[10] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on the erdős–rényi random graph. *Combinatorics, Probability and Computing*, 20:683–707, 9 2011.

[11] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. Universality for first passage percolation on sparse random graphs. *ArXiv e-prints*, Oct. 2012.

[12] S. Bhamidi, R. van der Hofstad, G. Hooghiemstra, et al. First passage percolation on random graphs with finite mean degrees. *The Annals of Applied Probability*, 20(5):1907–1965, 2010.

[13] S. Bhamidi, R. van der Hofstad, and J. Komjathy. The front of the epidemic spread and first passage percolation. *ArXiv e-prints*, Jan. 2014.

[14] B. Bollobás. *Random Graphs*, volume 73. Cambridge University Press, 2001.

[15] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[16] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[17] W. J. Bühler et al. The distribution of generations and other aspects of the family structure of branching processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*. The Regents of the University of California, 1972.

[18] M. Deijfen and R. van der Hofstad. The winner takes it all. *ArXiv e-prints*, June 2013.

[19] R. Durrett. *Random graph dynamics*, volume 20. Cambridge university press, 2007.

[20] P. Erdos and A. Rényi. {On the evolution of random graphs}. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[21] R. v. d. Hofstad. Random graphs and complex networks. lecture notes.

[22] S. Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8:347–361, 7 1999.

[23] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.

[24] B. Kharlamov. The numbers of generations in a branching process with an arbitrary set of particle types. *Theory of Probability & Its Applications*, 14(3):432–449, 1969.

[25] I. Kolossváry and J. Komjáthy. First Passage Percolation on Inhomogeneous Random Graphs. *ArXiv e-prints*, Jan. 2012.

[26] C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678–5682, May 2000.

[27] M. Newman and D. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4–6):341 – 346, 1999.

[28] M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84:3201–3204, Apr 2000.

[29] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Phys. Rev. E*, 60:7332–7342, Dec 1999.

[30] S. I. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

[31] M. Torok, A. Nelson, L. Alexander, G. C. Mejia, and P. D. MacDonald. Epidemic curves ahead. *Focus on Field Epidemiology*, 1, 2004.

[32] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. *Distances in random graphs with finite mean and infinite variance degrees*. EURANDOM, 2005.

[33] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.