Distance in the Forest Fire Model
How far are you from Eve?

Varun Kanade
École normale supérieure
varun.kanade@ens.fr

Reut Levi
École normale supérieure
reuti.levi@googlemail.com

Zvi Lotker
Ben Gurion University
zvilo@bgu.ac.il

Frederik Mallmann-Trenn
École normale supérieure,
Simon Fraser University,
mallmann@di.ens.fr

Claire Mathieu
CNRS, École normale supérieure
cmathieu@di.ens.fr

Abstract

Leskovec, Kleinberg and Faloutsos (2005) observed that many social networks exhibit properties such as shrinking (i.e. bounded) diameter, densification, and (power-law) heavy tail degree distributions. To explain these phenomena, they introduced a generative model, called the Forest Fire model, and using simulations showed that this model indeed exhibited these properties; however, proving this rigorously was left as an open problem.

In this paper, we analyse one of these properties, shrinking diameter. We define a restricted version of their model that incorporates the main features that seem to contribute towards this property, and prove that the graphs generated by this model exhibit shrinking distance to the seed graph. We prove that an even simpler model, the random walk model, already exhibits this phenomenon.

1 Introduction

Ten years ago, Leskovec, Kleinberg and Faloutsos introduced the Forest Fire model, a generative model to understand the dynamics of social networks over a long period [22]. They examined real-world networks such as the ArXiv Citation Graph, the Patents Citation Graph, the Autonomous Systems Graph, Affiliation Graphs, the Email Network, the IMDB Actors-to-Movies Network, and a Product Recommendation Network. They observed that these social networks become denser over time. They also made the surprising observation that the effective diameter of the networks "shrinks" over time, instead of growing, as was previously thought. They suggested the *Forest Fire* model as an attempt to explain densification, shrinking diameter, and heavy-tailed distributions of vertex indegrees and outdegrees.

In this model, the evolution initially starts with a fixed seed graph. Time is discrete and at each time t a node u_t arrives, picks a random node, w, in the current graph as its "ambassador" and connects to it. The ambassador is considered *burned* and all other nodes are considered *unburnt*. Node u_t then generates two random numbers x and y and selects x outgoing edges from w and y in-coming edges to w incident to nodes that have not yet been burned. If not enough outgoing or incoming edges are available, u_t selects as many as it can. Let $w_1, w_2, ..., w_{x+y}$ denote the other endpoints of the edges selected. u_t connects to $w_1, w_2, ..., w_{x+y}$, marks them as burned, and then applies the previous step recursively to each w_i. Leskovec *et al.* observed through simulation, that the Forest Fire Model appears to have the shrinking diameter property, but leave open the question of providing a rigorous proof:

> *"Rigorous analysis of the Forest Fire model appears to be quite difficult. However in simula-*

tions we find that [...] we can produce graphs that [...] have diameter that decrease."

This is the starting point of our work. In this paper, we answer this question for a simplified version of the model (Theorem 2.1) and for a variant (Theorem 2.3).

1.1 Related work

There is a extensive variety of models for generating graphs of social networks, each reproducing a subset of properties observed in real-world social networks. The first major line of research considers static graphs, where the number of nodes does not change over the course of time: For example, in *small-world* like models, there is a fixed underlying graph which is augmented by additional links between the vertices. Kleinberg proposed a particular random augmentation of links on the grid and proved that this gives rise to a decentralised greedy algorithms to find short paths among nodes [14]. In a more recent paper, Chaintreau *et al.* proposed a different model, in which similar results are achieved, where the grid is augmented with links generated by random walks on the grid with occasional resets [4].

Other static models focus mainly on reproducing both densification and small diameter simultaneously. One example is the model by Leskovec *et al.* which uses a matrix-operation, namely, the Kronecker product, to generate self-similar graphs recursively [21]. They reproduce a vast number of properties including heavy tails for the in- and out-degree distribution and small diameter. However, the deterministic nature of this model produces unrealistic features. To remedy this drawback, they propose the Stochastic Kronecker Graph (SKG) model which has been very successful and is widely used in simulations. One disadvantage of SKG is that the adjustment of the parameters may have a huge influence on the properties of the resulting graphs. Recently, Seshadhri *et al.* [29] showed that in fact the SKG model bears resemblance to a variant of the Chung and Lu model [6] which generalises classical random graph models. Additionally, Seshadhri *et al.* [28] introduce the Block Two-Level Erdős-Rényi (BTER) model, and demonstrate that it captures observable properties of many real-world social networks.

The second major research line considers graph evolving over time where at each time step new vertices and edges are added to the evolving graph. Barabási *et al.* proposed the so called *preferential attachment model* in which new vertices attach preferentially to vertices with high degree, reproducing the power law distribution over the in-degrees [2]. Building on preferential attachment, Cooper and Frieze propose a model which obeys power law as well as shrinking diameter and densification, unfortunately, it involves many parameters [7]. Recently, Avin *et al.* extended the preferential attachment model to incorporate densification [1]. Krapivsky and Redner investigated the development of random networks as the attachment probability grows [18].

The authors of [15, 19] consider an edge copying evolution in which, on arrival, a new vertex picks an existing node and copies a subset of its neighbours. Another model is the Community Guided Attachment model, in which there is a hierarchical backbone structure that determines the linkage probabilities [22]. Lattanzi and Sivakumar generate random graphs according to an underlying affiliation network: Each node picks a random subset of affiliations and in each affiliation the nodes are connected as a clique (additionally, there is a process of preferential attachment) [20]. They show that this model exhibits shrinking diameter, densification, and a heavy-tailed degree distribution. Moreover, they connected the densification of the network to the non-linearity of the core. The recursive search model proposed by Vazquez is quite similar to the Forest-Fire model [30]. In the recursive search model, vertices are added to the graph one by one; when a new vertex arrives it first connects to a random vertex and then recursively connects to a subset of its unvisited neighbours. The main difference is that in the forest fire model, the average number of neighbours visited out of the current node is constant, where as in the recursive search model this is a constant fraction. Thus, presence of high-degree nodes can make the two models quite different.

In the Random-Surfer Model (RSM), introduced by Blum *et al.* [3], the nodes arrive one by one. Upon arrival, each node performs several random walks from random starting points and connects to the endpoints of the performed walks. Our Random Walk Process (RWP) share resemblance to the RSM. The main difference is that in the RWP a new node connects to all the visited nodes in the random walk (instead of just the endpoint). Chebolu and Melsted [5] proved that the RSM and the PageRank-based selection model, proposed by Pandurangan *et al.* [27], are equivalent and also proved that the expected in-degree of vertices follows a powerlaw distribution. More recently, Mehrabian and Wormald [24] proved logarithmic upper bounds for the diameter in the RSM and the PageRank-based selection model as well as a logarithmic lower bound for a special case where the generated graph is a tree.

The only rigorous work thus far on the forest fire model is by Mehrabian [24] who provide a logarithmic upper bound to the diameter of the forest fire model as well as for other well known models, e.g., the copying model and the PageRank-based selection model.

1.2 Results and techniques

We look at two simplifications of the original Forest Fire model. We focus on understanding the asymptotic directed distance to the seed graph. Although Leskovec *et al.* look at average diameter in the graph (by ignoring the edge orientations), it is also natural to look at the average *directed* distance to the seed graph. Both the models we consider show that the asymptotic average distance to the seed graph is constant.

In the first simplified model, which we call the Forest Fire Process, the evolution initially starts with a fixed seed graph, and time is discrete. At each time t, a node u_t arrives, and is connected to its "ambassador", a node w randomly selected from the nodes already in the network (the direction of the added edge is from u_t to w). One says that the ambassador is *burned*. Node u_t then generates a random number x which is less than or equal to the outdegree of w, and selects x neighbours of w. The process continues recursively while never *burning* the same node twice (see Algorithms 2.1 and 2.2 for more details).[1]

Compared to the model of Leskovec *et al.*, the Forest Fire Process has two differences: it only has forward burning (no burning of edges going backwards); and it has a slightly different burning distribution (binomial versus truncated geometric). The model is parameterised by the average number of neighbours burned, α. When α is larger than some constant, then the average distance to the seed is provably bounded assuming the seed graph is a large enough cycle (Theorem 2.1); when α is less than some other constant, then the average distance to the seed provably grows logarithmically in the size of the graph (Theorem 2.2).

We then look at an even simpler model, which we call the Random Walk Process and which, surprisingly, still exhibits the same phenomenon. Surprising, because the average outdegree is clearly constant, and the model is local, yet the distance to the seed can be bounded. It is parameterised by the average outdegree. For this model our results are more precise: we show a threshold phenomenon. If the average outdegree is less than 4, then the distance to the seed grows logarithmically with the size of the graph; if it's greater than 4, then the distance to the seed is bounded.

How do we prove bounded on the distance to the seed is in the Forest Fire Process? It would be natural to show that the distance to the seed has a drift downwards, but that's not true in general. One challenge in the analysis is the non-monotonicity of the model; augmenting the graph by edges can make it more unlikely for the next arriving vertex to be close to the seed graph. From a mathematical viewpoint, the main novelty of the paper is the definition of an upper bound ϕ on the distance such that, conditioning on any history, after two steps, the expected value of ϕ decreases. Distance is one plus the minimum distance of a neighbour, but in order to have monotonicity, ϕ uses a maximum instead of a minimum.[2] One might think that bounding the distance by the ϕ value—replacing a minimum by a maximum—gives a much too loose approximation but surprisingly it works: in the Forest Fire Process, not just the shortest path but almost all paths to the seed are short. (The intuition for the definition of ϕ is discussed in greater detail in Section 4). Among other techniques, we reduce the analysis of a DAG built on top of a tree to the analysis of a DAG built on top of a line. We make frequent use of coupling arguments, stochastic domination, Galton-Watson branching processes, Hajek's Theorem, and other techniques from discrete probability.

Organisation: In Section 2, we formally define the models we study. In Section 3, we look at simpler versions of these processes and show how they imply properties in the original versions through coupling. In Section 4, we prove the main results for the Forest Fire process. Section 5 proves the main results for the Random Walk process. Finally, we conclude with a discussion. Appendices provide background on Galton-Watson processes (Section A) and other standard results from probability (Section B) that we use.

2 Models and Results

For a directed graph G, and a node $u \in G$, let $\mathrm{outdeg}_G(u)$ denote the outdegree of the node u in G and $\mathcal{N}(u)$ denote the (out)neighbours of u in G. For u, v, let $\mathrm{dist}_G(u, v)$ denote the (directed) distance from u to v in G.

For random variables X and Y we write $X \prec Y$ if X is stochastically dominated by Y, *i.e.*, for all k it holds $\mathbb{P}(X \geq k) \leq \mathbb{P}(Y \geq k)$. We denote by $\mathsf{Bernoulli}(p)$ the Bernoulli distribution with success probability p, by $\mathsf{Bin}(n, p)$ the Binomial distribution, with n independent trials, each having a success probability of p, by $\mathsf{Geom}(p)$ the geometric distribution with success probability p, *i.e.* for $X \sim \mathsf{Geom}(p)$ we have $\mathbb{P}(X = i) = (1-p)^i p$ for

[1] In fact, if one allows the same node to be burnt multiple times, it is rather straightforward to see that one has bounded expected distance to the seed graph as long as the average degree is greater than 1: The process then looks much like a Galton-Watson process with a positive probability of being infinite. Thus, this would imply that some node in the seed graph would be reached by the burning process with constant probability.

[2] In fact, there is some carefully induced non-monotonicity in the definition of ϕ, otherwise, we would not be able to bound its expected value.

$i = 0, 1, \ldots,$ by $\mathsf{Poisson}(\lambda)$ the Poisson distribution with mean λ, and by $\mathsf{Uniform}(b)$ the uniform distribution on the elements $\{0, 1, 2, \ldots, b\}$.

The Forest Fire Process is defined iteratively, starting from a seed graph G_0. Let $G_{t-1} = (V_{t-1}, E_{t-1})$ denote the graph at the end of round $t-1$. In round t, a new node u_t arrives, and chooses a node $u_{t'} \in V_{t-1}$ uniformly at random. We call the node $u_{t'}$ the *ambassador* of the new node u_t. After selecting the ambassador, we *burn* the ambassador, *i.e.*, we add the edge $(u_t, u_{t'})$ to the graph. This then propagates as follows.

First choose a random subset of the edges of G_{t-1} as active edges: every edge (u, v) of G_{t-1} is active independently with probability $\min\{1, \frac{\alpha}{\text{outdeg}(u)}\}$, where α is a parameter of the model. Second, *burn* all vertices of G_{t-1}, reachable from $u_{t'}$ by following directed active edges. Third, add an edge from u_t to every *burnt* vertex. This construction of G_t can be obtained by executing Algorithms 2.1 and 2.2. Although, it is more natural to view burning as a branching process, we describe it as a percolation process in order to avoid the need to define a specific order for the burning process.

Algorithm 2.1 Forest Fire Process (G_0)

for $t = 1, 2, \ldots$ **do**
  upon arrival of node u_t at time t:
    $\text{amb}(u_t) \leftarrow$ a node chosen u.a.r. from V_{t-1}
    $S \leftarrow \text{Burn}(G_{t-1}, \text{amb}(u_t))$
    $G_t \leftarrow (V_{t-1} \cup \{u_t\}, E_{t-1} \cup \{(u_t, w) : w \in S\})$

Algorithm 2.2 $\text{Burn}(G = (V, E), v)$
// Outputs a subset of V reachable from v

$H \leftarrow \emptyset$
for all $(w, x) \in E$ **do**
  with probability $\min\left\{1, \frac{\alpha}{\text{outdeg}_{G_t}(w)}\right\}$
    $H \leftarrow H \cup \{(w, x)\}$
return $\{x \in V$: there exists a directed path from v to x in $H\}$

We now state our two main results for the forest fire model. The parameters α and the input graph G_0 are fixed and we study the asymptotic properties of the graph G_t. We have not optimised the constants in the theorem statements and expect them to be far from tight.

THEOREM 2.1. *For all $\alpha \geq 100$, if G_0 is a directed cycle with $|G_0| \geq \alpha^{20}$, the Forest Fire Process with parameters α and G_0 has the property of non-increasing distance to G_0, i.e., for every t,*

$$\mathbb{E}[\text{dist}_{G_t}(u, G_0)] = O(1),$$

where the expectation is over a node u, which is picked uniformly at random in G_t, and $\text{dist}(u, G_0)$ is the directed distance.[3]

REMARK 1. *It is not critical that G_0 is a cycle. The main requirement is that once the* burning process *reaches G_0, a large enough constant number of vertices will be burnt. For example, G_0 being an expander, clique, or a strongly connected graph with large girth of with $|G_0|$ large enough suffices. Simulations seem to indicate that G_0 being a single node also result in a similar behaviour.*

THEOREM 2.2. *For all $\alpha \leq 1/(4e)$ and for all G_0, the Forest Fire Process with inputs α and G_0 is such that*

$$\mathbb{E}[\text{dist}_{G_t}(u, G_0)] = \Omega(\log t),$$

using the same notation as above.

The Random Walk Process is defined similarly, parameterised by a number p, $0 < p < 1$. Instead of the burning process, the construction takes a random walk from the starting node, whose length is distributed geometrically. Then, the process adds an edge from u_t to every vertex visited on the walk. We allow the graph to be a directed multi-graph by allowing parallel edges to vertices of G_0. This construction of G_t can be obtained by executing Algorithms 2.3 and 2.4.

Algorithm 2.3 Random Walk Process (G_0)

for $t = 1, 2, \ldots$ **do**
  upon arrival of node u_t at time t:
    $\text{amb}(u_t) \leftarrow$ a node chosen u.a.r. from V_{t-1}
    $S \leftarrow \text{Walk}(G_{t-1}, \text{amb}(u_t))$
    $G_t \leftarrow (V_{t-1} \cup \{u_t\}, E_{t-1} \cup \{(u_t, w) : w \in S\})$

Algorithm 2.4 $\text{Walk}(G, v)$
// Outputs a subset of V reachable from v

$S \leftarrow \{v\}$
with probability $1 - p$
  choose $w \in N(v)$ uniformly at random
  $S \leftarrow S \cup \text{Walk}(G, w)$
return S

Here are our two main results for the random walk model.

THEOREM 2.3. *Let G_0 be a strongly connected graph. The Random Walk Process with parameters $p < 1/3$*

[3] Note that $\text{dist}(v_t, G_0)$, once defined at time t, never changes

and G_0 has the property of non-increasing distance to G_0, i.e., for every t,

$$\mathbb{E}[\mathrm{dist}_{G_t}(u, G_0)] = O(1),$$

using the same notation as above.

THEOREM 2.4. *For all G_0, the Random Walk Process with parameters $p > 1/3$ and G_0, has the property that*

$$\mathbb{E}[\mathrm{dist}_{G_t}(u, G_0)] = \Omega(\log t),$$

using the same notation as above.

3 Relating graph and line processes

3.1 Line processes

To prove the results for the Forest Fire and Random Walk processes, we study related processes: the *Line Fire Process* and the *Line Walk Process*, which we define below. We state two corresponding technical lemmas for each of these Processes; in the next sub-section, we state coupling lemmas to relate the processes and prove the results of Section 2, using the corresponding related results on the line (whose proofs are deferred to later sections), together with the coupling.

When comparing the graph processes, defined in Section 2, (*i.e.* the Forest Fire Process and the Random Walk Process) with the Line Processes (*i.e.* the Line Fire Process and the Line Walk Process), the difference is that while in the graph processes the first step is to select the ambassador at random; in the line processes we skip this step, and force each new node to select the most recently added node as its ambassador, *i.e.*, in the line processes the first step is deterministic and follows the line structure.

Algorithm 3.1 Line Fire Process(L_0) and Line Walk Process(L_0)

for $t = 1, 2, \ldots$ **do**
  upon arrival of node u_t at time t:
    $\mathrm{amb}(u_t) \leftarrow u_{t-1}$
    $S \leftarrow \begin{cases} \mathrm{Burn}(L_{t-1}, \mathrm{amb}(u_t)) & \text{(Line Fire Proc.)} \\ \mathrm{Walk}(L_{t-1}, \mathrm{amb}(u_t)) & \text{(Line Walk Proc.)} \end{cases}$
    $L_t \leftarrow (V_{t-1} \cup \{u_t\}, E_{t-1} \cup \{(u_t, w) : w \in S\})$

Next, we state the relevant lemmas for the line fire process, that are used to prove the above theorems. The proofs of these lemmas are deferred to later sections.

LEMMA 3.1. *Let $\alpha \geq 100$ and let L_0 be a directed cycle, such that $|L_0| \geq \alpha^{20}$. Then the Line Fire Process has the property that*

$$\exists c, \exists \gamma < 1 \text{ s.t. } \forall t \ \forall j \ \mathbb{P}(\mathrm{dist}_{L_t}(v_t, L_0) > j) < c\gamma^j.$$

LEMMA 3.2. *For all L_0, the Line Fire Process with $\alpha \leq \alpha^*$ and L_0, has the property that*

$$\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)] = \Omega(t),$$

Similarly, we state the relevant lemmas for the line walk process.

LEMMA 3.3. *Let L_0 be a strongly connected graph. Then the line walk process with parameters $p < 1/3$ and L_0 has the property that*

$$\exists c, \exists \gamma < 1 \text{ s.t. } \forall t \ \forall j \ \mathbb{P}(\mathrm{dist}_{L_t}(v_t, L_0) > j) < c\gamma^j.$$

LEMMA 3.4. *For all L_0, the line fire process with parameters $p > 1/3$ and L_0 is such that*

$$\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)] = \Omega(t).$$

3.2 The ambassador graph

DEFINITION 1. *In the graph G_t generated by the the forest fire or random walk process, the* ambassador graph *A_t is the sub-graph, consisting of edges $(u, \mathrm{amb}(u))$ for all nodes $u \notin G_0$. These edges are referred to as ambassador edges.*

A_t is a forest of directed trees, rooted at vertices of G_0. First, we observe the following fact.

FACT 3.1. *If (u, v) is an edge of G_t, then there exists an integer k such that $v = \mathrm{amb}^k(u)$, where amb^k denotes k iterative applications of $\mathrm{amb}(\cdot)$.*

To prove our lower bounds, we will use the following bound on the expected distance to the root in the ambassador graph. The following lemma was originally proven in [8, Theorem 10].

LEMMA 3.5. *Let u be a vertex in the ambassador graph A_t chosen uniformly at random. Then*

$$\mathbb{E}[\mathrm{dist}_{A_t}(u, G_0)] = \Theta(\log t)\ .$$

Proof. Let v_k denote the node which arrives at time k. By convention the vertices of G_0 arrive at time 0. Then

$$\mathbb{E}[\mathrm{dist}_{A_t}(u, G_0)] \leq \frac{1}{t} \sum_{k \leq t} \mathbb{E}[\mathrm{dist}_{A_t}(v_k, G_0)]\ . \tag{3.1}$$

For the upper bound, since $\mathrm{dist}_{A_t}(v_k, G_0)$ is at most k in the worst case:

$$\mathbb{E}[\mathrm{dist}_{A_t}(v_k, G_0)] \leq c \log k + k \cdot \mathbb{P}[\mathrm{dist}_{A_t}(v_k, G_0) > c \log k]\ . \tag{3.2}$$

Recall that $\mathrm{dist}_{A_t}(v_k, G_0)$ is the length the path $v_k, \mathrm{amb}(v_k), \mathrm{amb}^2(v_k), \ldots$, until we reach G_0. Let X_i

denote the arrival time of $\text{amb}^{i-1}(v_k)$. We have, by uniform choice of the ambassador of a node:

$$\begin{aligned} X_1 &= k \\ \mathbb{E}(X_i|X_{i-1}) &\leq X_{i-1}/2 \\ X_{\text{dist}_{A_t}(v_k,G_0)} &= 0. \end{aligned}$$

Then, for some constant c by Markov's inequality,

$$\begin{aligned} \mathbb{P}[\text{dist}_{A_t}(v_k, G_0) > c\log k] &= \mathbb{P}[X_{c\log k+1} \geq 1] \\ &\leq \mathbb{E}[X_{c\log k+1}] \leq 1/k. \end{aligned} \tag{3.3}$$

Substituting into Equations (3.2) and (3.1), we obtain $\mathbb{E}[\text{dist}_{A_t}(u, G_0)] = O(\log t)$, as desired.

For the lower bound on the expectation, define Y_i and Z_i to be the indicator variables for the events: on the path from v_k to G_0 there exists a vertex whose arrival time is in $[2^{i-1}, 2^i]$ and $[1, 2^i]$, respectively. Observe that $\mathbb{P}(Y_i = 1|Z_i = 1) = 1/2$. Additionally, $\mathbb{P}(Z_i = 1) = \Omega(1/|G_0|) = \Omega(1)$ for every $i < \log k$. To see this, consider the first time that the path visits a node whose arrival time is at most 2^i. If this node is in G_0 then $Z_i = 0$, otherwise $Z_i = 1$. Thus $\mathbb{E}(Z_i) \geq 2^i/(2^i + |G_0|)$. Hence $\mathbb{P}(Z_i = 1) = \Omega(1)$, for every $i < \log k$. Since $\text{dist}_{A_t}(v_k, G_0)$ is at least the number of intervals that are visited, we obtain for some constant c',

$$\begin{aligned} \mathbb{E}[\text{dist}_{A_t}(v_k, G_0)] &\geq \mathbb{E}\left[\sum_{i\geq 1} Y_i\right] = \sum_{i\geq 1} \mathbb{P}(Y_i = 1) \\ &\geq c' \cdot \log k\,. \end{aligned} \tag{3.4}$$

Recall that k arrived at time k. In the remainder we use (3.4) to show that $\mathbb{E}[\text{dist}_{A_t}(u, G_0)] = \Omega(\log t)$ for u chosen arbitrary at random in A_t. Define $u_{\geq t/2}$ to be a vertex chosen uniformly at random whose arrival time is at least $t/2$, and observe that $\mathbb{E}[\text{dist}_{A_t}(u, G_0)] \geq \frac{1}{3}\mathbb{E}[\text{dist}_{A_t}(u_{\geq t/2}, G_0)]$ for large enough t. Therefore, by Equation (3.4), $\mathbb{E}[\text{dist}_{A_t}(u, G_0)] \geq \frac{c'}{3} \cdot \log(t/2) = \Omega(\log t)$. □

3.3 Coupling

The following lemma shows the relation between the Line Fire and Forest Fire Processes.

Definition 2. *The* level *of a vertex u is its distance to G_0 in the ambassador graph, defined by:*

$$\ell(u) = \begin{cases} 0 & \text{if } u \in G_0 \\ \ell(\text{amb}(u)) + 1 & \text{otherwise.} \end{cases}$$

Lemma 3.6. *Let $t \geq \tau \geq 1$. Consider the Forest Fire Process with seed graph G_0, conditioned on $\ell(u_t) = \tau$. Then the sub-graph of G_t, consisting of G_0 and of all vertices on the path from u_t to G_0 in A_t, and of all edges out of those vertices; has the same distribution as the graph L_τ, with seed graph $L_0 = G_0$. In particular $\text{dist}_{G_t}(u_t, G_0)$ in the Forest Fire Process conditioned on $\ell(u_t) = \tau$ has the same distribution as $\text{dist}_{L_\tau}(v_\tau, G_0)$ in the Line Fire Process. This also holds for the Random Walk Process.*

Proof. The ambassador graph is constructed independently of the Burn process, so we can change the order in which the edges of G_t are constructed, by generating the ambassador graph in a first phase, and then adding the other edges in a second phase. In the first phase, a node only chooses a random ambassador and connects to it. In the second phase, every node invokes the burning process starting with the respective ambassador.

Consider the path in the ambassador graph, going from u_t to G_0, and label its vertices

$$\begin{aligned} &(u_t, \text{amb}(u_t), \text{amb}^2(u_t), \ldots, \text{amb}^{\ell(u_t)}(u_t)) = \\ &\quad = (w_{\ell(u_t)}, \ldots, w_0). \end{aligned}$$

Thus $w_{\ell(u_t)} = u_t$, $w_i = \text{amb}(w_{i+1})$ for $i < \ell(u_t)$, and $w_0 \in G_0$. We claim that the sub-graph induced by $G_0 \cup \{w_0, \ldots, w_{\ell(u_t)}\}$ in the Forest Fire Process, has exactly the same distribution as the graph L_τ, produced by the line fire process with seed G_0, for $\tau = \ell(u_t)$.

To prove this, we couple the burning decisions of w_i and v_i. When $i = 0$, both graphs are G_0. Assume by induction that the sub-graphs induced by $G_0 \cup \{w_0, \ldots, w_{i-1}\}$, in the Forest Fire Process and in the graph L_{i-1} in the Line Fire Process, are identically distributed, hence coupled. Then the Burn process, starting at w_i, can clearly also be coupled with the Burn process of vertex v_i, to give the desired result.

The same proof also works for the random walk process. □

3.4 Proofs of the Graph results from the Line results

Proof. [Proof of Theorems 2.1 and 2.3] Since $\text{dist}(u_t, G_0)$, once defined at time t, never changes, it suffices to show that $\mathbb{E}[\text{dist}_{G_t}(u_t, G_0)] = O(1)$.

$$\begin{aligned} \mathbb{E}[\text{dist}_{G_t}(u_t, G_0)] &= \sum_{b=1}^{t} \mathbb{P}(\text{dist}_{G_t}(u_t, G_0) > b) \\ &= \sum_{b=1}^{t}\sum_{\tau=1}^{t} \mathbb{P}(\text{dist}_{G_t}(u_t, G_0) > b|\ell(u_t) = \tau) \cdot \mathbb{P}(\ell(u_t) = \tau), \end{aligned}$$

where $\ell(u_t)$ is distributed according to the construction of the ambassador tree in the first phase. From Lemma 3.6, $\mathbb{P}(\text{dist}_{G_t}(u_t, G_0) > b|\ell(u_t) = \tau) =$

$\mathbb{P}(\text{dist}_{L_\tau}(v_\tau, G_0) > b)$. From Lemma 3.1 for the Line Fire Process (Lemma 3.3 for the Line Walk Process), $\mathbb{P}(\text{dist}_{L_\tau}(v_\tau, G_0) > b) \le c\gamma^b$. Thus

$$\mathbb{E}[\text{dist}_{G_t}(u_t, G_0)] \le \sum_{b=1}^{t}\sum_{\tau=1}^{t} c\gamma^b \cdot \mathbb{P}(\ell(u_t) = \tau)$$
$$\le \sum_{b=1}^{t} c\gamma^b \le \frac{c}{1-\gamma} = O(1).$$

□

Proof. [Proofs of Theorems 2.2 and 2.4] As in the beginning of the proof of Theorem 2.1, using Lemma 3.6 and exchanging the order of summations, we can write

$$\begin{aligned}
&\mathbb{E}[\text{dist}_{G_t}(u_t, G_0)] = \\
&= \sum_{b=1}^{t}\sum_{\tau=1}^{t} \mathbb{P}(\text{dist}_{L_\tau}(v_\tau, G_0) > b) \cdot \mathbb{P}(\ell(u_t) = \tau) \\
&= \sum_{\tau=1}^{t} \mathbb{E}[\text{dist}_{L_\tau}(u_\tau, G_0)] \cdot \mathbb{P}(\ell(u_t) = \tau) \\
&\ge \sum_{\tau=1}^{t} c\tau \cdot \mathbb{P}(\ell(u_t) = \tau) = c \cdot \mathbb{E}[\ell(u_t)] = \Omega(\log t),
\end{aligned}$$

where for the Line Fire Process we used Lemma 3.2 and Lemma 3.5. For the Line Walk Process, we use Lemmas 3.4 and Lemma 3.5 instead. Thus $\mathbb{E}[\text{dist}_{G_t}(u_{t'}, G_0)] = \Omega(\log t)$ for every $t' \ge t/2$, hence we obtain the desired result.

□

4 Analysis of the Line Fire Process

Throughout Sections 4.1, 4.2, 4.3, and 4.4 we assume that $\alpha \ge 100$, L_0 is a directed cycle with $|L_0| \ge \alpha^{20}$. In order to prove Lemma 3.1, we define a function ϕ, such that for all t, $\text{dist}_{L_t}(v_t) \le \phi(v_t)$, and which is more amenable to analysis. Let $\delta = \alpha^{20}$.

- $\phi(v) = 0$ if $v \in L_0$.
- $\phi(v) = 1 + \max_{v' \in \mathcal{N}(v)}\{\phi(v')\}$ if $\text{outdeg}(v) < \delta$.
- $\phi(v) = \max\left\{\phi(\text{amb}(v)) - 2, 1 + \max_{\substack{v' \in \mathcal{N}(v) \\ v' \ne \text{amb}(v)}} \phi(v')\right\}$ otherwise

4.1 High Level Proof Overview

We first give some intuition about the definition of ϕ. We would like to argue that no matter what happens up to time t, $\text{dist}(v_{t+1}, L_0)$ is less than $\text{dist}(v_t, L_0)$ in expectation. This does not seem to be possible when using distance directly; we can construct graphs where this is not true. However, these graphs are unlikely to arise under the Line Fire Process. Analysing ϕ instead gets around this issue. In fact, we show that $\phi(v_{2t+2}) - \phi(v_{2t})$ has negative expectation, no matter what the history up to time $2t$. A low value $\phi(v_t)$ implies that not only is there one short path from v_t to L_0, but most paths from v_t to L_0 are short. However, note that not all paths are short, in particular the path $v_t, v_{t-1}, \ldots, v_0$. Furthermore, while it is true for most nodes, it is not necessarily true that all nodes are well connected to the seed graph. Note that the definition of ϕ makes a special case for the ambassador when the degree is large. For an edge (v, u) if $u \ne \text{amb}(v)$, $\phi(u) < \phi(v)$. We will call edges $(v, \text{amb}(u))$ *ambassador edges*.

We start from an arbitrary history (and hence an arbitrary graph) at time $2t$. (See Figure 1a: the nodes are arranged by their ϕ value, ambassador edges are marked red and may point upwards, *i.e.*, an increase in ϕ-value, all other edges point strictly downward, *i.e.*, a decrease in ϕ-value.) The good event at time $2t+1$ involves two things: (i) the degree of v_{2t+1} is at least δ (ii) All neighbours v of v_{2t+1}, except possibly the ambassador, are such that $\phi(v) \le \phi(v_{2t}) - 2$. We give a high-level idea why this is likely (formal proof in Lemma 4.5). The burning process stops at any node only with probability $\approx (1 - \alpha/d)^d \approx e^{-\alpha}$, thus it is quite likely that at least δ nodes are burnt starting at v_{2t}. For the second part, at the very first stage, *i.e.*, neighbours of v_{2t} that are burnt, almost all neighbours (except possibly the ambassador if the v_{2t} has out-degree at least δ), will cause a drop in ϕ-value of at least 1. Subsequently, if we look at any path in the burning process, every edge traversed implies that the ϕ value dropped by at least 1, except if the edge was an ambassador (red) edge at a high-degree node, where it may increase by 2 (see definition of ϕ). A large fraction of such red edges are not likely to appear on any path (ambassador edges of low-degree nodes are not a problem, by definition of ϕ). The edges burnt when v_{2t+1} arrives are shown in Figure 1b.

Given that the good event at time $2t+1$ happens, the good event at time $2t+2$ again involves two things: (i) the degree of v_{2t+2} is at least δ (ii) *all* neighbours v of v_{2t+2} satisfy $\phi(v) \le \phi(v_{2t}) - 2$. First, it is easily checked that if the good event happens, indeed it is the case that $\phi(v_{2t+2}) \le \phi(v_{2t}) - 1$, *i.e.*, a decrease. The first part of the good event is similar to the previous case. For the second part, we again consider the first step of the burning process, *i.e.*, the burnt neighbours of v_{2t+1}. Since most neighbours of v_{2t+1} have ϕ-value at most $\phi(v_{2t}) - 2$, with high probability *all* burnt neighbours

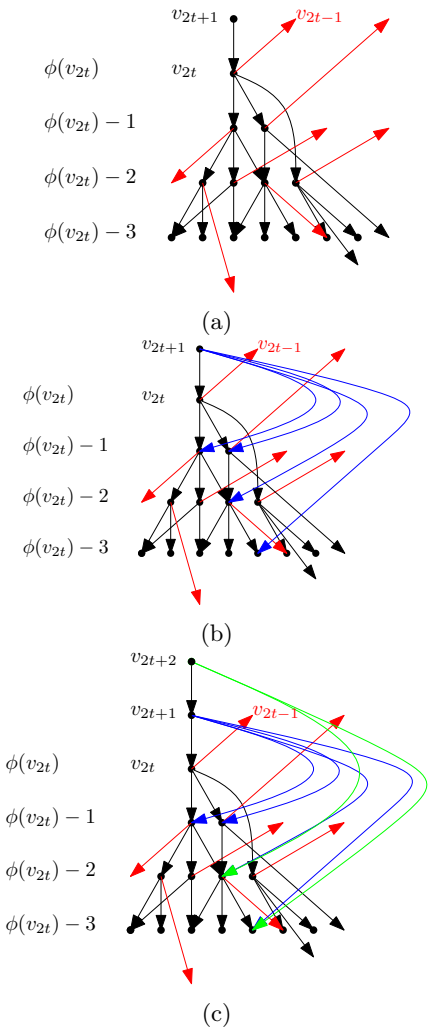

Figure 1: The red edges are ambassador edges. The blue edges are the neighbours of v_{2t+1} and the green edges are the neighbours of v_{2t+2}. When v_{2t+1} arrives it is likely to only have very few edges to nodes with ϕ-values $\geq \phi(v_{2t}) - 1$. When v_{2t} arrives it is likely not to have any neighbours (except of the ambassador) with ϕ-values $\leq \phi(v_{2t}) - 2$.

will satisfy this. Further down in the burn process, it is unlikely that ϕ-value increases as argued earlier, since a large fraction of ambassador (red) edges would have to be followed which is unlikely (See Figure 1c).

Formally we can show that given any history up to time $2t$, $\phi(v_{2t+2}) - \phi(v_{2t})$ has a sub-exponential tail and negative expectation, which implies by Hajek's theorem (Theorem 4.1) that $\mathbb{E}[\phi(v_{2t})] = O(1)$.

4.2 Proof of Lemma 3.1

We now formalise the high-level ideas presented in the previous section. We begin by proving that ϕ indeed dominates distance.

FACT 4.1. *If v arrives at time t, then*

$$\text{dist}_{L_t}(v, L_0) \leq \phi(v).$$

Proof.

$$\text{dist}_{L_t}(v, L_0) = \begin{cases} 0 \text{ if } v \in L_0 \\ 1 + \min\limits_{v' \in \mathcal{N}(v)} \{\text{dist}_{L_t}(v', L_0)\} \text{ otherwise} \end{cases}$$

$$\leq \begin{cases} 0 & \text{if } v \in L_0 \\ 1 + \phi(\text{amb}(v)) & \text{if outdeg}(v) = 1 \\ 1 + \min\limits_{v' \in \mathcal{N}(v) \setminus \{\text{amb}(v)\}} \{\phi(v')\} & \text{otherwise} \end{cases}$$

Comparing the right hand side to the definition of ϕ, and noting that in the graph L_t all (directed) edges point to vertices that arrived earlier (*i.e.*, for any edge $(v_t, v_{t'})$, $t > t'$), the result follows by induction on t. □

The proof of Lemma 3.1 relies on Hajek's theorem (Theorem 4.1), which we now state in a slightly simplified form.

THEOREM 4.1. (HAJEK'S THEOREM [13]) *Let $(Y_t)_{t \geq 0}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Assume the following two conditions hold:*

(i) (Majorisation) There exists a random variable Z and a constant $\lambda > 0$, such that $\mathbb{E}[e^{\lambda Z}]$ is finite, and $(|Y_{t+1} - Y_t| \big| \mathcal{F}_t) \prec Z$ for all $t \geq 0$; and

(ii) (Negative bias) There exist $a, \varepsilon_0 > 0$, such for all t we have $\mathbb{E}[Y_{t+1} - Y_t \mid \mathcal{F}_t, Y_t > a] \leq -\varepsilon_0$

Then there exist $\eta, C > 0$ such that for all b and t we have $\mathbb{P}(Y_t \geq b \mid \mathcal{F}_0) \leq Ce^{-\eta b}$.

Let $(\mathcal{F}_t)_{t \geq 0}$ denote the history of random choices up to time t for the line fire process. We state two lemmas that prove that $(\phi(v_{2t+2}) - \phi(v_{2t}) \mid \mathcal{F}_{2t})$ indeed satisfy the conditions of Hajek's theorem. The proofs of these lemmas appear in subsequent subsections.

LEMMA 4.1. (MAJORIZATION) *Let Z be the random variable taking values over all even integers greater than or equal to 4, defined by : $\mathbb{P}(Z = 2i) = 3^{-i-1}$ for $i \geq 3$, and $\mathbb{P}(Z = 4) = 1 - \sum_{i\geq 3}\mathbb{P}(Z = 2i)$. Then $(|\phi(v_{2t+2}) - \phi(v_{2t})|\big|\mathcal{F}_{2t}) \prec Z$.*

LEMMA 4.2. (NEGATIVE BIAS) *There a exists constants $\varepsilon_0 > 0$ such that for every t we have,*

$$\mathbb{E}\left[\phi(v_{2t+2}) - \phi(v_{2t})\big|\mathcal{F}_{2t},\ \phi(v_{2t}) > 2\right] \leq -\varepsilon_0$$

Proof. [Proof of Lemma 3.1] Let $Y_t = \phi(v_{2t})$. Whenever $\lambda < \ln(3)/2$, we have $\mathbb{E}[e^{\lambda Z}]$ is finite for the random variable Z defined in the statement of Lemma 4.1. Thus the sequence $(Y_t)_{t\geq 0}$ with respect to the filtration $(\mathcal{F}_{2t})_{t\geq 0}$ satisfies the two conditions of Theorem 4.1 by Lemmas 4.1 and 4.2, hence Lemma 3.1 follows. □

4.3 Proofs of Lemmas 4.1 and 4.2

We first prove the following side lemma.

LEMMA 4.3. $\mathbb{P}(\text{outdeg}_{L_t}(v_t) < \delta \mid \mathcal{F}_{t-1}) \leq \frac{1}{2\alpha}$.

Proof. We show by induction that the set returned by $\text{Burn}(L_{t-1}, v_{t-1})$ has size at least j with probability at least $1 - (j-1)\cdot e^{-\alpha}$, for $1 \leq j \leq \delta$. Note that since $\delta = \alpha^{20}$, for α large enough, $(\delta - 1)\cdot e^{-\alpha} \leq \frac{1}{2\alpha}$, as required.

To prove the statement by induction, we invoke the principle of deferred decisions of random choices for the activation of the edges. We will also make a slightly stronger induction hypothesis: There is a path $w_0, w_1, \ldots, w_{j-1}$, with $v_{t-1} = w_0$, such that $\text{Burn}(L_{t-1}, v_{t-1})$ will return all of $\{w_0, w_1, \ldots, w_{j-1}\}$. Furthermore, all the edges (w_i, w_{i+1}) are activated and the out-edges of $w_0, \ldots w_{j-2}$ are the only edges for which the random choices have already been made. Clearly for $j = 0$, the statement is true, because v_{t-1} will always be returned. Suppose the statement is true for $j < \delta$. We will prove the corresponding statement for $j+1$. We now look at the activation choices for the out-edges of w_{j-1}. The probability that at least one out-edge is activated is at least $1 - e^{-\alpha}$ (since $\mathbb{P}(\mathsf{Bin}(d, \min\{1, \frac{\alpha}{d}\}) = 0) \leq e^{-\alpha}$). Furthermore, since $j < \delta$, this out-edge cannot be to any of the vertices $\{w_0, \ldots, w_{j-2}\}$, as there is only one directed cycle in L_{t-1} and this has length δ. Thus, by induction and union bound, we have the required result. □

In our proofs, it is useful to rephrase the process $\text{Burn}(G, \alpha)$ defined in Algorithm 2.2 as a tree process, rather than a percolation process. We define $\text{BurnBFS}(G, v)$ in Algorithm 4.1. We assume that vertices have a natural order in the graph, for examples for graphs evolving in time, the vertices are ordered according to their time of arrival. Thus, when indexing a set we assume that the vertices are indexed in this order.

Algorithm 4.1 $\text{BurnBFS}(G, v)$

$M_0 := \{v\}$
for $i = 1, 2, \ldots$ **do**
  $M_i := \emptyset$
  for all $w \in M_{i-1}$ **do**
    for all edges (w, x) **do**
      activate edge (w, x) with probability $\min\{1, \frac{\alpha}{\text{outdeg}_G(w)}\}$
      if (w, x) is activated and $x \notin \bigcup_{j\leq i} M_j$ **then**
        add x to M_i
        set parent$(x) := w$
return $\bigcup_{j\geq 0} M_j$

First, we note that if the burning decisions made for activation of edges (w, x) in Algorithm 4.1 are coupled with those made in Algorithm 2.2, the set of vertices returned by the two processes is exactly the same. Thus, this is indeed another view of the burning process. The burning process BurnBFS produces a tree T with vertices $\cup_{j\geq 0} M_j$, and edges (w, x) for which the *if* condition in Algorithm 4.1 was satisfied. (We remark that if H_v is the induced sub-graph of H (defined in Algorithm 2.2) consisting of all nodes reachable from v, then T is simply the unique BFS tree of H_v starting at v using the order on the vertices.)

In the present section, we will fix some graph G (say some L_t produced by the line fire process), and look at calls made to BurnBFS with this graph as input. Thus, the only source of randomness is the activation decision of the edges. The following is a key technical lemma, whose proof we defer to the next subsection.

LEMMA 4.4. *Fix some i, execute $\text{BurnBFS}(G, v)$ until M_i is created (if at all). Let $u \in M_i$ be a fixed vertex and let $\mathcal{H}$ denote the history of activation decisions made by algorithm BurnBFS thus far. Let T_u be the (random) sub-tree of T with root u and let $k \geq 0$, then*

$$\mathbb{P}(\exists u' \in T_u \setminus \{u\}, \phi(u') \geq \phi(u) + k \mid \mathcal{H}) \leq \frac{3^{-k}}{12\alpha^4}.$$

We now prove the majorisation Lemma.

Proof. [Proof of Lemma 4.1] Observe that Z can equivalently be defined by $\mathbb{P}(Z \geq i) = 3^{-\lceil i/2\rceil}/2$ if $i \geq 5$ and $\mathbb{P}(Z \geq i) = 1$ if $i \leq 4$.

Fix the history $\mathcal{F}_{2t}$ up to time $2t$ and let

$$\begin{aligned}\Delta &= \phi(v_{2t+2}) - \phi(v_{2t}) \\ &= (\phi(v_{2t+2}) - \phi(\mathrm{amb}(v_{2t+2}))) \\ &+ (\phi(v_{2t+1}) - \phi(\mathrm{amb}(v_{2t+1}))).\end{aligned}$$

If $\Delta \geq i \geq 4$ then at least one of the two expressions on the right hand side exceeds $\lceil i/2 \rceil \geq 2$, so $\mathbb{P}(\Delta \geq i \mid \mathcal{F}_{2t})$ is less than or equal to

$$\begin{aligned}&\mathbb{P}(\phi(v_{2t+1}) - \phi(\mathrm{amb}(v_{2t+1})) \geq \lceil i/2 \rceil \mid \mathcal{F}_{2t}) \\ &\quad + \mathbb{P}(\phi(v_{2t+2}) - \phi(\mathrm{amb}(v_{2t+2})) \geq \lceil i/2 \rceil \mid \mathcal{F}_{2t+1}).\end{aligned}$$

Thus we need an upper bound on $\mathbb{P}(\phi(v) - \phi(\mathrm{amb}(v)) \geq j \mid \mathcal{F})$ for $j \geq 2$, where $\mathcal{F}$ is the history right before the arrival of v.

Consider the process $\mathrm{BurnBFS}(G, \mathrm{amb}(v))$ with output $T_{\mathrm{amb}(v)}$ executed to construct $\mathcal{N}(v)$. By definition of ϕ, the sub-tree $T_{\mathrm{amb}(v)}$ needs to contain a node with ϕ-value at least $\phi(\mathrm{amb}(v)) + j - 1$. We use Lemma 4.4 by setting $u = \mathrm{amb}(v)$, $k = \lceil i/2 \rceil - 1$, and can therefore write:

(4.5)

$$\mathbb{P}(\Delta \geq i) \leq 2\frac{3^{-(\lceil i/2 \rceil - 1)}}{36\alpha^4} = 3^{-\lceil i/2 \rceil}\tfrac{1}{6\alpha^4} \leq \mathbb{P}(Z \geq i),$$

thus $\Delta \prec Z$. On the other hand, by definition of ϕ, $\phi(v_{2t+2}) - \phi(v_{2t}) \geq -4$, so $-\Delta \prec Z$. Thus $|\Delta| \prec Z$, which completes the proof of Lemma 4.1. □

To prove the Lemma 4.2 (negative bias), we need to analyse the process over two consecutive steps. We start from an arbitrary history $\mathcal{F}_{2t}$. We first establish some properties that after one step hold with high probability (w.r.t. α).

Lemma 4.5. *Fix $\mathcal{F}_{2t}$ and consider the arrival of v_{2t+1}. Let $z = \phi(v_{2t})$. Then, with probability at least $1 - 1/\alpha$, the following holds: $\mathcal{N}(v_{2t+1})$ contains exactly one node with value z (namely, v_{2t}), at most 6α nodes with value $z-1$, and all other nodes have value at most $z-2$. Moreover,* $\mathrm{outdeg}(v_{2t+1}) \geq \delta$.

Proof. Consider the process $\mathrm{BurnBFS}(G, v_{2t})$ executed to construct $\mathcal{N}(v_{2t+1})$. By Chernoff bounds (Lemma B.2), with probability at least $1 - 2^{-6\alpha}$ we have $|M_1| \leq 6\alpha$. Assume this holds. What's the out-degree of v_{2t}? If it is less than δ, then, by definition of ϕ, all nodes of $\mathcal{N}(v_{2t})$, and in particular all elements of M_1, have ϕ-value at most $z-1$. If it is greater than or equal to δ, then by definition of ϕ all but one nodes of $\mathcal{N}(v_{2t})$ have value less than or equal to $z-1$, and with high probability $1 - 6\alpha/\delta$ the set M_1 does not contain the single exceptional node of $\mathcal{N}(v_{2t})$, and, assuming this holds, all elements of M_1 have ϕ-value less than or equal to $z-1$. Then, Lemma 4.4 (with $k=0$) applied to all sub-trees rooted at nodes of M_1 shows that with probability at least $1 - 6\alpha/(12\alpha^4)$, all other nodes visited by $\mathrm{BurnBFS}(G, v_{2t})$ have ϕ-value less than or equal to $z-2$. Assume this holds.

Moreover, by Lemma 4.3, with probability at least $1 - 1/(2\alpha)$, we have $\mathrm{outdeg}(v_{2t+1}) > \delta$. Assume this holds.

Assuming all those high-probability events hold, $\mathcal{N}(v_{2t+1})$ satisfies all the statements of the lemma. The probability that one of the assumptions we made along the way fails to be realised is, by union bound, at most

$$2^{-6\alpha} + \frac{6\alpha}{\delta} + \frac{6\alpha}{12\alpha^4} + \frac{1}{2\alpha} \leq \frac{1}{\alpha}.$$

□

Proof. [Proof of Lemma 4.2] Fix $\mathcal{F}_{2t}$ and consider the arrival of v_{2t+1}. With probability at least $1 - 1/\alpha$ the situation described in Lemma 4.5 happens. Assume that to be the case, and consider the arrival of v_{2t+2}. Consider the process $\mathrm{BurnBFS}(G, v_{2t+1})$ executed to construct $\mathcal{N}(v_{2t+2})$. Once again, by Chernoff bounds (Lemma B.2), with probability at least $1 - 2^{-6\alpha}$ we have $|M_1| \leq 6\alpha$. Assume this holds. By Lemma 4.5 we know that v_{2t+1} has at least δ neighbours, of which only $6\alpha+1$ may have ϕ-value greater than or equal to $z-1$. With probability at least $1 - 6\alpha(6\alpha+1)/\delta$, none of the nodes of M_1 are in that set, and therefore all nodes of M_1 have ϕ-value less than or equal to $z-2$. Then, Lemma 4.4 (for $k=1$) applied to all sub-trees rooted at nodes of M_1 shows that with probability at least $1 - 6\alpha/(36\alpha^4)$, all other nodes visited by $\mathrm{BurnBFS}(G, v_{2t+1})$ have ϕ-value less than or equal to $z-2$. Assume that holds.

Now, what's the out-degree of v_{2t+2}? By Lemma 4.3, with probability at least $1 - 1/(2\alpha)$, we have $\mathrm{outdeg}(v_{2t+2}) > \delta$. Assume this holds. Then by definition of ϕ and since $\phi(v_{2t}) > 2$, we obtain $\phi(v_{2t+2}) \leq z - 1$, and thus

$$\phi(v_{2t+2}) \leq \phi(v_{2t}) - 1.$$

The probability that one of the assumptions we made along the way fails to be realised is at most

$$\frac{1}{\alpha} + 2^{-6\alpha} + \frac{6\alpha(6\alpha+1)}{\delta} + \frac{6\alpha}{36\alpha^4} + \frac{1}{2\alpha} \leq \frac{2}{\alpha}.$$

To recap, if we let $\Delta = \phi(v_{2t+2}) - \phi(v_{2t})$, we have just proved that

$$\mathbb{P}[\Delta \leq -1] \geq 1 - \frac{2}{\alpha}.$$

To compute the expectation (implicitly conditioning on

$\mathcal{F}_{2t}$), we now write

$$\mathbb{E}[\Delta] \leq \sum_{k\geq 4} k \cdot \mathbb{P}(\Delta = k) + 3 \cdot \mathbb{P}(0 \leq \Delta \leq 3) - \mathbb{P}(\Delta \leq -1)$$

$$\leq \sum_{k\geq 4} \mathbb{P}(\Delta \geq k) + 3 \cdot \frac{2}{\alpha} - \left(1 - \frac{2}{\alpha}\right). \tag{4.6}$$

The first term on the right hand side can be bounded using Equation (4.5):

$$\sum_{k\geq 4} \mathbb{P}(\Delta \geq k) \leq 2 \sum_{k\geq 4} 3^{-\lceil k/2 \rceil} \frac{1}{\alpha^4} \leq \frac{1}{\alpha^2}.$$

We finally obtain

$$\mathbb{E}[\Delta \mid \mathcal{F}_{2t}] \leq -1 + \frac{8}{\alpha} + \frac{1}{\alpha^2} < 0,$$

hence the negative bias. □

4.4 Proof of Lemma 4.4

Proof. The main idea of the proof is to couple the tree process defined by BurnBFS with a Galton-Watson Process. Let i be as in the statement of the lemma, and suppose that the sets $M_0, \ldots, M_i$ have already been fixed by the activation decisions in BurnBFS. We look at $u \in M_i$, the designated vertex in the statement of the lemma.

Let w be some vertex in T, the tree generated by BurnBFS, and say $w \in M_{k-1}$. We are interested in understanding the random variable that is the number of children of w in T. Let M_k^w denote the set M_k right after the activation decisions for edges of vertices in M_{k-1} that are before w in the ordering are completed. Let $S = \{x \in \mathcal{N}(w) \mid x \notin \cup_{j<k} M_j \text{ and } x \notin M_k^w\}$ be the random variable (depending on the choices made while determining $M_0, \ldots, M_{k-1}, M_k^w \setminus \{w\}$), that is the set of potential children of w. Let $p = \min\{1, \frac{\alpha}{\text{outdeg}(w)}\}$. Let B be obtained by adding each $x \in S$ to B with probability p. Thus, $Z_w = |B| \sim \mathsf{Bin}(|S|, p)$ and B corresponds to the activated edges that lead to nodes not already in $\cup_{j<k} M_j \cup M_k^w$. We define R_w to be a random variable: if $\text{outdeg}(w) < \delta$, $R = 0$; otherwise $R = 1$ if and only if $\text{amb}(w) \in B$. We will call the edge $(w, \text{amb}(w))$ in T red if $R_w = 1$. We are interested in the random variables (Z_w, R_w) (note that they are dependent on random choices made earlier in the process defined above; however, to minimise cumbersome notation we will not make this explicit).

We will now define a branching process that is completely independent of the line fire process. It is a Galton-Watson process, with some designated red edges. Let Z', R' be random variables where $Z' \sim 1 + \lceil e\alpha \rceil + \mathsf{Poisson}(e\alpha)$ and $R' \sim \mathsf{Bernoulli}(\alpha/\delta)$. Z' is the random variable that defines the offspring distribution of the Galton-Watson process, and if $R' = 1$, the edge between the node and its "first" child is marked red. We will show that this process stochastically dominates the branching process resulting from a call to Burn, in a particular technical sense.

We have the following claim:

Claim 1. *Let w be some node in M_{k-1}, and let S, p, B be as defined above. Let (Z_w, R_w) be the random variables defined above for the burn process. Let Z', R' be as used to define the independent Galton-Watson process. Then, whenever $\alpha \leq \delta$, there exists a coupling of the random variables such that $Z_w < Z'$ and $R_w \leq R'$.*

Proof. We look at three cases.

1. If $\text{outdeg}(w) < \delta$, then $R_w = 0 \leq R'$. So we only need to define a coupling so that $Z_w < Z'$
2. If $\text{outdeg}(w) \geq \delta$, but $\text{amb}(w) \notin S$ (that is $\text{amb}(w)$ is in some M_j for $j \leq k$ already when the activation decisions for out-edges of w were made), then $R_w = 0$, and again we just need to define a coupling so that $Z_w < Z'$.
3. Finally, when $\text{outdeg}(w) \geq \delta$ and $\text{amb}(w) \in S$, we can couple as follows: note that $Z_w = R_w + \tilde{Z}_w$, where $R_w \sim \mathsf{Bernoulli}(p)$ and $\tilde{Z}_w \sim \mathsf{Bin}(|S| - 1, p)$ (when $\text{outdeg}(w) \geq \delta$, $p = \alpha/\text{outdeg}(w)$, as long as $\delta \geq \alpha$). Note that (Z_w, R_w) have the exact same joint distribution as defined above, since effectively we are making the choice of whether or not $\text{amb}(w)$ should be included in B independently of the other elements. Since $\alpha/\text{outdeg}(w) \leq \alpha/\delta$, it is clear that we can couple R_w and R' so that $R_w \leq R'$. Thus, again it remains only to show a coupling such that $Z_w < Z'$.

For all $\alpha \geq 1$, it follows that $\mathsf{Bin}(n, p)$ is stochastically dominated by $\mathsf{Poisson}(e\alpha)$ whenever $p \leq \alpha/n$ and $n \geq e\alpha$ (see *e.g.* [17]). When, $n < e\alpha$, clearly $\mathsf{Bin}(n, p)$ is stochastically dominated by $1 + \lceil e\alpha \rceil$. The additional 1 in the definition of Z' takes care of the strict inequality made in the claim. This completes the proof. □

Let $\mathcal{T}'$ denote the (possibly infinite) Galton-Watson tree with offspring distribution Z' and some edges marked "red" as defined above. We define a coupling between the (random) sub-tree generated by the Burn process T_u (rooted at $u \in M_i$) and $\mathcal{T}'$ inductively below: This results in an injective map σ from $V(T_u)$ to $V(\mathcal{T}')$, where $V(T)$ denotes the vertices in tree T. Let ρ denote the root of $\mathcal{T}'$, then σ is defined as follows

(through coupling and induction on $\text{dist}_{T_u}(u,w)$). Note that u is the only vertex with $\text{dist}_{T_u}(u,u)=0$.

1. $\sigma(u)=\rho$
2. Suppose all $w \in T_u$ with $\text{dist}_{T_u}(u,w) \le \Delta$ are mapped under σ to some vertices in $\mathcal{T}'$. We look at the time when activation decisions for some w such that $\text{dist}_{T_u}(u,w)=\Delta$ are made. For each such w, we apply the coupling defined in Claim 1. Let (Z_w, R_w) be the corresponding random variables and let (Z', R') be the independent instantiation of the random variables denoting the children of $\sigma(w)$ in $\mathcal{T}'$. By the coupling, we have $Z_w < Z'$ and $R_w \le R'$. If $R_w = 1$, we set $\sigma(\text{amb}(w))$ to be the "red" (first) child of $\sigma(w)$. The remaining $Z_w - 1$ children of w can be mapped to the subsequent $Z_w - 1$ children of $\sigma(w)$, which is possible by the coupling. If $R_w = 0$, all Z_w children of w are mapped to the non-"red" children of $\sigma(w)$, which again is possible since $Z' > Z_w$. This defines the map σ for all vertices w, such that $\text{dist}_{T_u}(u,w) = \Delta+1$.

We observe that the map σ satisfies the following properties by definition:

1. If (w,x) is an edge in T_u, then $(\sigma(w),\sigma(x))$ is an edge in $\mathcal{T}'$, and furthermore the edge $(\sigma(w),\sigma(x))$ points away from the root.
2. If (w,x) is coloured red in T_u, then $(\sigma(w),\sigma(x))$ is coloured red in $\mathcal{T}'$.

Finally, we define a function $\phi' : V(\mathcal{T}') \to \mathbb{Z}$, on the nodes of the tree $\mathcal{T}'$ as follows:

1. $\phi'(\rho) = \phi(u)$
2. For w', let $\text{parent}(w')$ denote the parent of w' in $\mathcal{T}'$. Then,

$$\phi'(w') = \begin{cases} \phi'(\text{parent}(w')) + 2 & \text{if } (\text{parent}(w'), w') \\ & \text{is red} \\ \phi'(\text{parent}(w')) - 1 & \text{otherwise} \end{cases}$$

We check the following fact:

Claim 2. *For every $w \in V(T_u)$, $\phi(w) \le \phi'(\sigma(w))$*

Proof. The proof is based on induction on $\text{dist}_{T_u}(u,w)$. Clearly, when $\text{dist}_{T_u}(u,w)=0$, it must be the case that $w=u$, and we have $\phi'(\sigma(u)) = \phi'(\rho) = \phi(u)$. Suppose, this holds for all w such that $\text{dist}_{T_u}(u,w) \le \Delta$. Consider an edge (w,x) in T_u, such that $\text{dist}_{T_u}(u,x) = \Delta+1$. Then, we consider two cases:

- If (w,x) is coloured red, $x = \text{amb}(w)$. Also, in this case, the edge $(\sigma(w),\sigma(x))$ in $\mathcal{T}'$ is also coloured red. Hence by definition $\phi'(\sigma(x)) = \phi'(\sigma(w)) + 2 \ge \phi(w)+2$. On the other hand, by definition $\phi(x) \le \phi(w)+2$. (Note that (w,x) is red, implies that $\text{outdeg}(w) \ge \delta$ and $x = \text{amb}(w)$; thus, by definition of ϕ, $\phi(w) \ge \phi(x) - 2$.) Hence, we have $\phi(x) \le \phi'(\sigma(x))$.
- On the other hand, if (w,x) is not red, we have that $\phi'(\sigma(x)) = \phi'(\sigma(w)) - 1 \ge \phi(w) - 1$. Also by definition of ϕ, we know that $\phi(w) \ge \phi(x)+1$ for all $x \in \mathcal{N}(w) \setminus \{\text{amb}(w)\}$. This completes the proof.

□

Using Claim 2 we have, $\mathbb{P}(\exists u' \in T_u \setminus \{u\}, \phi(u') \ge \phi(u)+k \mid \mathcal{H}) \le \mathbb{P}(\exists v' \in \mathcal{T}' \setminus \{\rho\}, \phi'(v') \ge \phi'(\rho)+k)$. Thus, it only remains to analyse ϕ' on $\mathcal{T}'$ and prove the required bound. Let $\alpha' = (1 + \lceil e\alpha \rceil + e\alpha)$; for $i \ge 0$, let $b_i = (k+1)(6\alpha')^i$ and let N_i be the random variable denoting the number of nodes of $\mathcal{T}'$ at distance i from the root ρ. We have the following:

(4.7)
$$\begin{aligned}&\mathbb{P}(\exists v' \in \mathcal{T}' \setminus \{\rho\}, \phi'(v') \ge \phi'(\rho)+k) \le \\ &\quad \le \mathbb{P}(\exists v' \in T' \setminus \{\rho\}, \phi'(v') \ge \phi'(\rho)+k \mid \forall i, N_i \le b_i) \\ &\quad + \mathbb{P}(\exists i : N_i \ge b_i)\end{aligned}$$

We bound the two terms of Eq. (4.7) separately. To bound the first term, we use a union bound:

(4.8)
$$\begin{aligned}&\mathbb{P}(\exists v' \in \mathcal{T}' \setminus \{\rho\}, \phi'(v') \ge \phi'(\rho)+k \mid \forall i, N_i \le b_i) \le \\ &\quad \le \sum_{j \ge 1} b_j \max_{v'} \mathbb{P}(\phi'(v') \ge \phi'(\rho)+k \mid \text{dist}(\rho,v') = j)\end{aligned}$$

We now bound for an arbitrary (directed) path $\mathcal{P}$, with vertices $v_0 = \rho, v_1, \ldots, v_j = v'$ in $\mathcal{T}'$, the probability that $\phi'(v') \ge \phi'(\rho)+k$. Note that as we go down the tree $\mathcal{T}'$, the value of ϕ' only decreases, except on red edges. Observe by definition of the tree, that the number of children of any node, distributed according to $Z' \sim 1 + \lceil e\alpha \rceil + \mathsf{Poisson}(e\alpha)$ is independent of whether or not the first node is coloured red. Therefore, by assuming that every edge along the path can potentially be *red*, we are only increasing the probability that for some node v', $\phi'(v') \ge \phi'(\rho)+k$. Note that the probability that any edge out of a node is red, denoted by p_r is at most α/δ (for the first child of a node, which always exists since the number of children is at least $1 + \lceil e\alpha \rceil$, the probability is p_r, for the remaining it is 0). Let r denote the number of red edges in the path $v_0 = \rho, v_1, \ldots, v_j = v'$, then the number

of non-red edges is $j-r$. Thus, by definition of ϕ', $\phi'(v') = \phi'(\rho) + 3r - j$, and hence for $\phi'(v') \geq \phi(\rho)+k$ to be true, it must be the case that $r \geq \lceil \frac{j+k}{3} \rceil$. Thus, we have

$$\begin{aligned}
&\mathbb{P}(\text{At least } \left\lceil \tfrac{j+k}{3} \right\rceil \text{ edges in } \mathcal{P} \text{ are red}) \leq \\
&\qquad \leq \sum_{b=\left\lceil \frac{j+k}{3} \right\rceil}^{j} \binom{j}{b} p_r^b (1-p_r)^{(j-b)} \\
&\qquad \leq \sum_{b\geq\left\lceil \frac{j+k}{3} \right\rceil} \left(\frac{ej}{b}\right)^b p_r^b \\
&\qquad \leq \sum_{b\geq\left\lceil \frac{j+k}{3} \right\rceil} (3e\cdot p_r)^b \\
&\qquad \leq (3e\cdot p_r)^{\frac{j+k}{3}} \cdot \tfrac{1}{1-3ep_r}
\end{aligned}$$

Substituting this bound in Eq. (4.8), we get

$$\begin{aligned}
&\mathbb{P}(\exists v' \in \mathcal{T}' \setminus \{\rho\}, \phi'(v') \geq \phi'(\rho)+k \mid \forall i, N_i \leq b_i) \leq \\
&\quad \leq \sum_{j\geq 1} b_j \tfrac{1}{1-3ep_r} \left((3ep_r)^{1/3}\right)^{j+k} \\
&\quad \leq (k+1)\frac{1}{1-3ep_r}\left((3ep_r)^{1/3}\right)^k \sum_{j\geq 1}\left(6\alpha'(3ep_r)^{1/3}\right)^j \\
&\quad \leq (k+1)\frac{1}{1-3ep_r}\left((3ep_r)^{1/3}\right)^k \cdot \frac{6\alpha'(3ep_r)^{1/3}}{1-6\alpha'(3ep_r)^{1/3}}
\end{aligned} \tag{4.9}$$

Now, we analyse the second term of Eq. (4.7). Let X_i denote the event that $N_i \geq b_i$. Thus, we are interested in bounding $\mathbb{P}(\exists i, X_i)$. Observe, that:

$$\mathbb{P}(\exists i, X_i) \leq \sum_{i\geq 1} \mathbb{P}(X_i \mid \neg X_{i-1}).$$

We observe that N_i is a sum of N_{i-1} independent copies of $Z' \sim 1 + \lceil e\alpha \rceil + \mathsf{Poisson}(e\alpha)$. Lemma B.3 proves that,

$$\mathbb{P}(X_i \mid \neg X_{i-1}) \leq 2^{-6eab_{i-1}}$$

Thus, we have:

$$\begin{aligned}
\mathbb{P}(\exists i, X_i) &\leq \sum_{i\geq 1} 2^{-6eab_{i-1}} \\
&\leq 2\cdot 2^{-6eab_0} = 2\cdot 2^{-(k+1)(36e\alpha)}
\end{aligned} \tag{4.10}$$

Substituting (4.9) and (4.10) in Eq. (4.7), we get that whenever $\alpha \geq 100$, $\delta \geq \alpha^{20}$, it is the case that $p_r \leq \alpha^{-19}$, which concludes the proof of the lemma. □

4.5 Lower bound: Proof of Lemma 3.2

Proof. As mentioned earlier, the neighbours of a node v_t in the Line Fire process can be represented by the vertices of a tree T rooted at v_{t-1}, in which every node v appears at most once. Furthermore, the number of edges percolated by v is $X_v \sim \mathsf{Bin}\left(|\mathcal{N}(v)|, \min\left\{1, \frac{\alpha}{|\mathcal{N}(v)|}\right\}\right) = \mathsf{Bin}\left(|\mathcal{N}(u)|, \frac{\alpha}{|\mathcal{N}(v)|}\right)$.

We define a process P in which, at the arrival of v_t, node v_{t-1} percolates $Y_{v_{t-1}} \sim \mathsf{Poisson}(e\alpha)$ outgoing edges u.a.r.. Whenever an edge (v_1, v_2) is percolated, v_2 percolates $Y_{v_2} \sim \mathsf{Poisson}(e\alpha)$ of it's outgoing edges u.a.r.. Let T'_{v_t} be the resulting tree. v_t connects then to all nodes of T'_{v_t} (once). Note that a node can be several times in T'_{v_t} and we assume that every time it is added, it chooses $\mathsf{Poisson}(e\alpha)$ u.a.r. to percolate independent of former choices. Since $Pr(X_v \geq k) \leq Pr(Y_v \geq k)$ for $k \geq 1$, we can couple the trees T_v and T'_{v_t} such that if $v \in T$, then $v \in T'_{v_t}$. Let $\Delta_t = \mathrm{dist}_{L_t}(v_t, L_0) - \mathrm{dist}_{L_{t-1}}(v_{t-1}, L_0)$. We have $-\infty < \Delta_t \leq 1$. The distance of a node v to L_0 in L_t equals the number of nodes on the shortest path plus one. Hence, we obtain a crude bound on $\mathbb{P}(\Delta_t = -k)$ by bounding $\mathbb{P}(|T'_{v_t}| \geq k+1)$. As we will argue in the following, $\mathbb{P}(|T'| \geq k+1)$ has an exponential tail distribution. Observe, that T'_{v_t} is GW-tree with offspring distribution $\mathsf{Poisson}(e\alpha)$. We have

$$\mathbb{P}(\Delta_t = -k | \mathcal{F}_{t-1}) \leq \mathbb{P}(|T'_{v_t}| \geq k+1) \leq e^{-(k+1)},$$

where the last inequality follows from Lemma 1.9 of [9]. Hence,

$$\begin{aligned}
\mathbb{E}[\Delta_t | \mathcal{F}_{t-1}] &\geq 1\cdot \mathbb{P}(X_{v_{t-1}} = 0) + \sum_{k\geq 1}(-k)\mathbb{P}(\Delta_t = -k | \mathcal{F}_{t-1}) \\
&\geq \min_{d\geq 1}\{(1-\alpha/d)^d\} + \sum_{k\geq 1} -ke^{-(k+1)} \\
&\geq 1 - \alpha - 0.4 \geq 0.5.
\end{aligned}$$

Hence, $\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)|\mathcal{F}_{t-1}] \geq \mathrm{dist}_{L_{t-1}}(v_{t-1}, L_0) + 1/2$. We have,

$$\begin{aligned}
\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)] &= \mathbb{E}[\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)|\mathcal{F}_{t-1}]] \\
&\geq \mathbb{E}[\mathrm{dist}_{L_{t-1}}(v_{t-1}, L_0)] + 1/2.
\end{aligned}$$

Hence, by repeating this iteratively, we get $\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)] = \Omega(t)$, which yields the claim. □

5 Analysis of the Line Walk Process

In this section we analyse the Line Walk Process. Upon arrival, v_t performs a random walk on L_{t-1} starting at v_{t-1} and stopping at any node with probability

$p > 0$. After the random walk stops, it connects to all visited nodes. Hence, the length of the random walk is distributed as $\mathsf{Geom}(p)$,[4] and hence the degree of v_t is distributed as $1 + \mathsf{Geom}(p)$ (it always connects to the ambassador).

Our proof strategy at a high-level is to understand the neighbourhood of v_t, in particular, we are interested in understanding the *oldest* neighbour of v_t. In this section, we use the term *length* of an edge to denote the temporal difference between the time of arrival of it's endpoints. We characterise these lengths in terms of Galton-Watson trees.

For simplicity, we consider L_0 to be a single node, denoted by v_0, with a self-loop; however, all the proofs in this section are valid for every L_0 which is a strongly connected graph as long as we allow the graphs L_t to be multi-graphs. Note that since the self loop of L_0 is the only directed cycle, all multiple edges have to point to v_0.

5.1 Characterising Edges of v_t through Galton-Watson trees

To analyse the lengths of the out-edges of v_t, it is useful to apply the principle of deferred decisions for the random choices made in constructing the graph L_t. We start the construction process backward starting at time t, only adding edges as we "need" them. Note that the edges (v_k, v_{k-1}) for $k = 1, \ldots, t$ are all deterministically added in the line walk process, as is the edge (v_0, v_0). Let $X_{v_t} \sim \mathsf{Geom}(p)$ denote the length of the random walk performed upon arrival of vertex v_t, starting at v_{t-1}. We index the edges of v_t, so that the *zeroth* edge is the ambassador edge (v_t, v_{t-1}), the first edge is to the first vertex encountered in the random walk after v_{t-1} and so on. In order to determine the first edge traversed by the random walk (assuming $X_{v_t} > 0$), we will have to determine some edges of v_{t-1}, and continue to do so recursively, making only the random choices that are needed. It turns out that the random choices we make can be succinctly modelled as a Galton Watson tree as we describe next.

Let $\mathsf{HalfGeom}(p)$ denote the distribution $\mathsf{Uniform}(X)$ where $X \sim \mathsf{Geom}(p)$.[5] Consider a Galton-Watson tree with offspring distribution $\mathsf{HalfGeom}(p)$. Let $\mathcal{T}$ be a tree generated by this process. We define how we can interpret $\mathcal{T}$ to reveal the first out-edge of v_t. Note that the index of the edge out of v_{t-1} we follow is exactly distributed as $\mathsf{HalfGeom}(p)$, and to determine this edge, we go back further and apply the same process. This is what gives rise to the Galton-Watson Process.

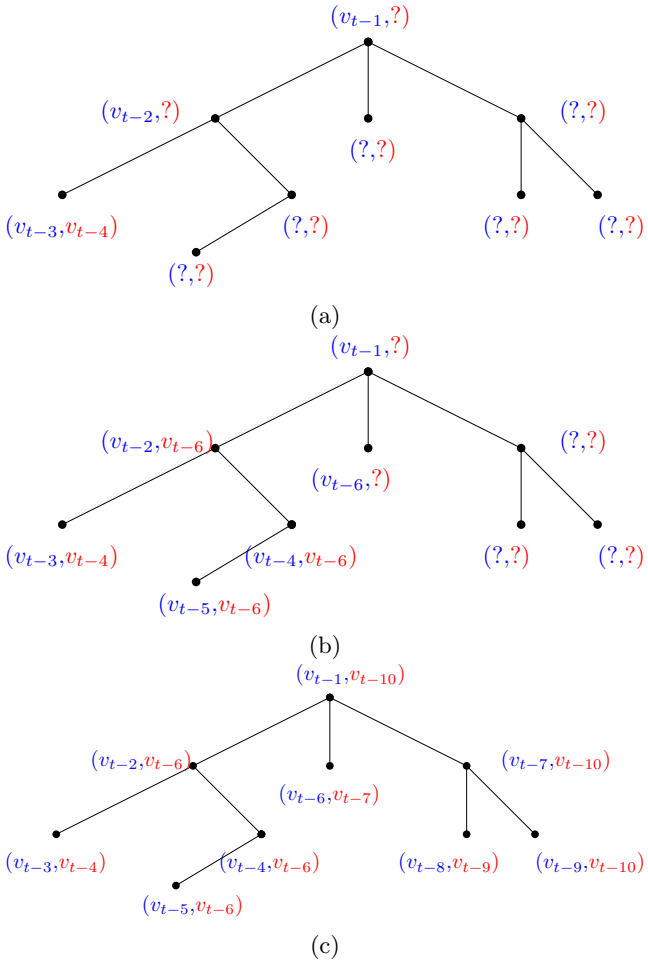

Figure 2: Interpretation of a Galton-Watson tree to determine edges of L_t

Before defining this mapping formally it would be helpful to go through an example. Figure 2a shows a (random) Galton-Watson tree $\mathcal{T}$. We perform a depth-first exploration of $\mathcal{T}$ assigning labels to the nodes along the way. The number of children at the root was drawn from $\mathsf{HalfGeom}(p)$, say it was 3. This indicates that we follow the *third* edge out of v_{t-1}. Thus, we will label the root with the end points of this edge, however for now we don't know the destination so temporarily it's label is $(v_{t-1}, ?)$. To determine its destination we need to simulate the random walk performed when v_{t-1} arrived, which is done at the three children of the root. The random walk started at $v_{t-2} = \mathrm{amb}(v_{t-1})$. The number of children of the first (leftmost) child of the root, also distributed as $\mathsf{HalfGeom}(p)$, say it was 2, indicates that

[4] We consider $\mathsf{Geom}(p)$ to represent the number of failures when success probability is p; thus it can take values in the set of non-negative integers.

[5] Recall that $\mathsf{Uniform}(X)$ chooses a number uniformly at random in $\{0, 1, \ldots, X\}$ so is well defined even when $X = 0$.

the first edge of v_{t-1} was decided by following the second edge of v_{t-2}. Temporarily, the label of this node is set to $(v_{t-2}, ?)$. In Figure 2a, the first child of the node labelled $(v_{t-2}, ?)$ has no children, this indicates that the *zeroth* edge out of v_{t-3}, *i.e.*, the edge (v_{t-3}, v_{t-4}) was followed when v_{t-2} arrived and started a random walk at v_{t-3}. This indicates that the first edge of v_{t-2}, is (v_{t-2}, v_{t-4}).

Then, the second child of the node labelled $(v_{t-2}, ?)$ in Figure 2a determines the second step of the random walk when v_{t-2} arrived. After the first step, the node reached was v_{t-4}, so the second step was on some edge out of v_{t-4}, which is indicated by analysing the subtree. Continuing, this way, we notice that the first edge of v_{t-1} was (v_{t-1}, v_{t-6}). The second edge is determined by analysing the subtree at the second child of the root (see Figure 2b). And finally, continuing the depth-first traversal, we can determine that the third edge of v_{t-1} was (v_{t-1}, v_{t-10}), and hence the first edge of v_t would be (v_t, v_{t-10}) (Figure 2c). It is no coincidence that the length of the edge (v_{t-1}, v_{t-10}), *i.e.*, $(t-1)-(t-10) = 9$ is the number of nodes in the tree. This can be verified by the following observation: Whenever a node is visited for the first time in the DFS traversal, if $v_{t'}$ was the last blue label assigned, the current node receives the blue label $v_{t'-1}$. Note that in case the tree is infinite or has more than $t-1$ nodes, at some point of time the label $(v_0, ?)$ is assigned. But we know that all subsequent edges must point to v_0 (see Figure 3). Basically, we interpret the self loop as infinitely many self-loops, so all of the *zeroth*, *first*, *etc.* edges of v_0 go to v_0, so there is no need to explore the subtrees. Thus, nodes may have multiple edges to v_0; in Figure 3 v_5 has at least two edges to v_0.

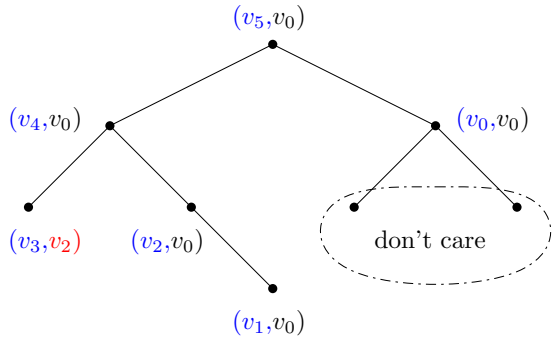

Figure 3: When the tree is larger than $t-1$, multiple edges are added to v_0

Remark 2. *This is why allowing multiple edges is important; it allows us to determine the degree of a node without observing the rest of the graph, i.e., the length of the random walk can be determined without knowing which edges are being traversed, since if we hit v_0, we continue to stay at v_0 for the required number of steps.*

We now define formally how the tree $\mathcal{T}$ can be interpreted as a set of edges that are added to L_t, when starting the construction process backward from time t. (Also, we will not determine random choices that are not needed to determine the first out-edge of v_t.) To avoid confusion between the nodes in the graph L_t and the nodes in the Galton Watson Tree, we will consistently use w to denote nodes in the Galton Watson tree. For a node $w \in \mathcal{T}$, let parent(w) denote the parent of w in the Galton Watson tree $\mathcal{T}$. Let children(w) denote an ordered list of children of w. If w is the root, parent(w) is undefined, and if w is a leaf, children(w) is an empty list. We describe how the nodes of the tree $\mathcal{T}$ will be viewed as a set of edges that are added to L_t. Thus each node $w \in \mathcal{T}$ will have labels source(w) and dest(w), which are vertices in L_t. We perform a DFS traversal on $\mathcal{T}$, with the property that when a node is visited for the first time during the traversal, a value for source(w) is set, and when it is visited for the last time, then the value for dest(w) is set. Initially, for all nodes w, source(w) =? and dest(w) =?. A special case in the traversal is that if source$(w) = v_0$, then no further exploration going down the tree is performed. If a node $w \in \mathcal{T}$ has $i \sim \mathsf{HalfGeom}(p)$ children, this is indicates that the edge $(\text{source}(w), \text{dest}(w))$, was the result of a random walk that followed the edges $(\text{source}(w_1), \text{dest}(w_1)), \ldots, (\text{source}(w_i), \text{dest}(w_i))$, where $w_1, \ldots, w_i$ are the children of w in $\mathcal{T}$. If w has 0 children, the edge is $(\text{source}(w), \text{amb}(\text{source}(w)))$. Formally, the traversal by the following rules (recall that we are determining the edges of v_t, by starting a random walk at v_{t-1}):

1. Let $w = \rho$ be the root node. Set source$(w) = v_{t-1}$
2. At w, if source$(w) = v_0$, set dest$(w) = v_0$, go to parent(w) if it exists; otherwise stop.
3. If w has no children and source$(w) \neq v_0$, set dest$(w) =$ amb(source(w)). Go to parent(w), if it exists, otherwise stop.
4. If all children of w have both labels set, set dest$(w) =$ dest(w'), where w' is the *last* child of w. Go to parent(w), if it exists, otherwise stop.
5. If w has some children with both labels assigned and remaining children with no labels assigned, let w' be the last child of w for which both labels were assigned and $\tilde{w}$ the first child with no labels assigned. Let source$(\tilde{w}) =$ dest(w'), go to $\tilde{w}$.

6. If no child of w has yet been visited and $\text{source}(w) \neq v_0$, then let w' denote the first child of w. Set $\text{source}(w') = \text{amb}(\text{source}(w))$. Go to w'.

At the end of the this process, all the edges $(\text{source}(w), \text{dest}(w))$ for every node $w \in \mathcal{T}$ are added to the graph L_t. There may be nodes $w \in \mathcal{T}$ with $\text{dest}(w) = \text{amb}(\text{source}(w))$, these are edges of the form (v_k, v_{k-1}) that were already present, and so are not added. The same is the case when $\text{source}(w) = \text{dest}(w) = v_0$; in this case, the subtree of w has no labels and is ignored (see Figure 3). This simply means that the seed graph L_0 was hit while trying to add an edge out of v_t and this means that (v_t, v_0) will be an edge in the graph. In particular, note that if $\mathcal{T}$ is infinite this will always be the case. Note that if $(v_t, v_{t'})$ was the first edge out of v_t, with $t' \neq 0$, the second edge (if $X_{v_t} > 1$) is determined by the same process: again $\mathcal{T}$ is generated by the Galton-Watson process with offspring distribution $\mathsf{HalfGeom}(p)$, the starting label at the root is now $(v_{t'}, ?)$ (see Figure 4).

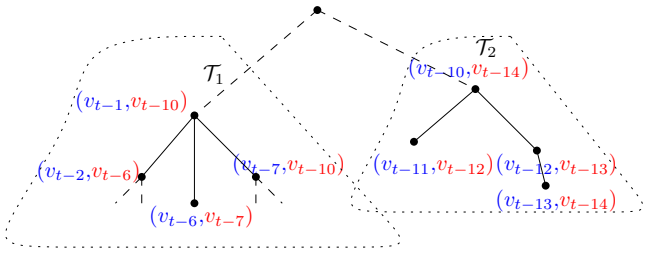

Figure 4: Illustration of how multiple edges of v_t are determined using independent Galton-Watson tree processes with offspring distribution $\mathsf{HalfGeom}(p)$. The tree $\mathcal{T}_1$ is the same as the tree in Figure 2

Thus, we can summarise the above explanation by the following fact.

FACT 5.1. *Fix v_t and let $X_{v_t} \sim \mathsf{Geom}(p)$, $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{X_{v_t}}$ be i.i.d. Galton Watson trees with offspring distribution $\mathsf{HalfGeom}(p)$). Then*

- *The neighbours of v_t are distributed as*

$$v_{t-1}, v_{\min\{0, t-1-\sum_{j=1}^{1}|\mathcal{T}_j|\}}, \ldots, v_{\min\{0, t-1-\sum_{j=1}^{X_{v_t}}|\mathcal{T}_j|\}}$$

- *Let $\tau = 1 + \sum_{j=1}^{X_{v_t}} |\mathcal{T}_j|$, and let $L_t[\tau]$ be the subgraph of induced by by the vertices $v_0, \ldots, v_\tau$. Then,*

$$\mathbb{P}\left(L_t[\tau] \mid X_{v_t}, \mathcal{T}_1, \ldots, \mathcal{T}_{X_{v_t}}, \tau < t\right) = \mathbb{P}(L_\tau),$$

where $\mathbb{P}(L_\tau)$ is the probability of generating a specific graph under the line fire process.

5.2 Proofs of Lemmas 3.3 and 3.4

Let $\mathcal{T}$ be a Galton-Watson tree with offspring distribution $\mathsf{HalfGeom}(p)$. It is well-known that if the expectation of the offspring distribution is strictly larger than 1, then with constant probability, $\mathcal{T}$ is infinite. The following lemma bounds the probability for this event in terms of p.

LEMMA 5.1. *Let $\mathcal{T}$ be a Galton-Watson tree with offspring distribution $\mathsf{HalfGeom}(p)$. $\mathcal{T}$ is infinite with probability at least $1 - \frac{3(1-3p)}{p+5}$ for every $0 < p < 1/3$.*

Proof. Let $\hat{X}_n$ denote the number of nodes of the n-th generation of $\mathcal{T}$. Let $\mu = \mathbb{E}(\hat{X}_1)$. Consider X_1 to be drawn in the following way: first we draw $Z \sim \mathsf{Geom}(p)$, and then draw $\hat{X}_1$ uniformly at random from $\{0, \ldots, Z\}$. We have $\mathbb{E}[\hat{X}_1 | Z = k] = \sum_{i=0}^{k} i/(k+1) = k/2$. Hence, it is implied that $\mu = \mathbb{E}[Z]/2 = (1-p)/(2p)$. Similarly, $\mathbb{E}[\hat{X}_1^2 | Z = k] = \sum_{i=0}^{k} i^2/(k+1) = k(2k+1)/6$, implying that the variance of $\hat{X}_1$ is

$$\begin{aligned}
\mathbb{V}(\hat{X}_1) &= \mathbb{E}[\hat{X}_1^2] - \mathbb{E}[\hat{X}_1]^2 \\
&= \tfrac{2\mathbb{E}[Z^2]+\mathbb{E}[Z]}{6} - E[\hat{X}_1]^2 \\
&= \tfrac{(1-p)(4-p)}{6p^2} - \tfrac{(1-p)^2}{4p^2} \\
&= \tfrac{(1-p)(p+5)}{12p^2},
\end{aligned}$$

where the third equality follows from the facts that $\mathbb{E}[Z^2] = (1-p)(2-p)/p^2$ and $\mathbb{E}[Z] = (1-p)/p$.

For $\mu > 1$ it holds that $\mathbb{E}[X_n] = \mu^n$ and $\mathbb{V}(X_n) \leq \frac{\mathbb{V}(B_v)}{\mu(\mu-1)} \cdot \mu^{2n}$ [11]. By Chebyshev's inequality $\mathbb{P}(|X_n - \mu^n| \geq \mu^n) \leq \frac{\mu(\mu-1)}{\mathbb{V}(B_v)} = \frac{3(1-3p)}{p+5}$. Hence $\mathcal{T}$ is infinite with probability at least $1 - \frac{3(1-3p)}{p+5}$ which yields the claim. □

Proof. [Proof of Lemma 3.3] The probability at the arrival of a node v_t to create at least one tree $\mathcal{T}$ is at least $(1-p)$. Hence the property that the neighbourhood of v_t is infinite is at least $q \geq (1-p)\left(1 - \frac{3(1-3p)}{p+5}\right) > 0$, by Lemma 5.1. If $\mathcal{T}$ is infinite, then we must have hit a node of L_0 and thus $\text{dist}_{L_t}(v_t, L_0) = 1$.

If not, let $(v_t, v_{t'})$ be the longest edge of v_t, with $t' = t - \left(1 + \sum_{j=1}^{X_{v_t}} |\mathcal{T}_j|\right) > 0$. By Fact 5.1, conditioned on $X_{v_t}, \mathcal{T}_1, \ldots, \mathcal{T}_{X_{v_t}}$, the subgraph of L_t induced by vertices $v_0, \ldots, v_{t-t'}$ follows the same distribution as $L_{t-t'}$. Hence, similarly as before, Lemma 5.1 gives that the probability for $v_{t-t'}$ to be connected to L_0 is at least q. Repeating this iteratively gives that the probability for v_t to have a distance of at least j is bounded by $(1-q)^{j-1}$. The claim follows since $q > 0$ is a constant. □

We now investigate the setting where $p > 1/3$.

LEMMA 5.2. *There exists a constant $c = c(p) > 0$ such that $\mathbb{P}(|\mathcal{T}| > n) \leq (1-c(3-\frac{1}{p}))^n$ for every $1/3 < p \leq 1$.*

We are ready to prove Lemma 3.4.

Proof. [Proof of Lemma 3.4] Let $\mathcal{T}_1, \ldots, \mathcal{T}_{X_{v_t}}$ be the Galton-Watson trees created at the arrival of v_t, with $X_{v_t} \sim \mathsf{Geom}(p)$. Pick $Y \sim \mathsf{Uniform}(X_{v_t})$. Hence, $(1 + |\mathcal{T}_1| + \cdots + |\mathcal{T}_Y|)$ and $(1 + |\mathcal{T}_{Y+1}| + \cdots + |\mathcal{T}_{X_{v_t}}|)$ follow the same distribution as $|\mathcal{T}|$, where $\mathcal{T}$is a Galton-Watson tree with offspring distribution $\mathsf{HalfGeom}(p)$.

Therefore, by Lemma 5.2, we get

$$\begin{aligned}\mathbb{P}(1 + |\mathcal{T}_1| + \cdots + |\mathcal{T}_{X_{v_t}}| > n) &\leq 2\mathbb{P}\left(1 + |\mathcal{T}| > n/2\right) \\ &\leq 2\left(1 - c(3 - \tfrac{1}{p})\right)^{n/2-1} . \qquad (5.11)\end{aligned}$$

Let a be a constant that will be determined later. For every v_i define m_i to be $(1 + |\mathcal{T}_1| + \cdots + |\mathcal{T}_{X_{v_t}}|)$ if $(1 + |\mathcal{T}_1| + \cdots + |\mathcal{T}_{X_{v_t}}|) \geq a$ and 0 otherwise. Observe that

$$\mathrm{dist}_{L_t}(v_t, L_0) \geq \left(t - \sum_{i=1}^{t} m_i\right)/a \, . \qquad (5.12)$$

Let $\gamma = (1 - c(3 - \frac{1}{p}))^{1/2} < 1$. By Equation (5.11),

$$\mathbb{E}\left[\sum_{i=1}^{t} m_i\right] \leq \tfrac{2t}{\gamma} \cdot \sum_{j \geq a} j\gamma^j = \tfrac{2t}{\gamma} \cdot \tfrac{\gamma^a(a(1-\gamma)+\gamma)}{(\gamma-1)^2} < t/2 \, , \qquad (5.13)$$

where the last inequality holds for sufficiently large a. We deduce from Equations (5.12) and (5.13) that $\mathbb{E}[\mathrm{dist}_{L_t}(v_t, L_0)] = \Omega(t)$, as desired.

□

Proof. [Proof of Lemma 5.2] Consider a one-by-one exploration process of $\mathcal{T}$, e.g., a breadth-first exploration. We maintain a queue of active nodes of $\mathcal{T}$. Initially we insert the root to the queue and at each time step i we remove one node from the queue and insert $\xi_i \sim \mathsf{HalfGeom}(p)$ nodes to the queue. Let A_i denote the number of nodes in the queue after time step i. Thus $A_0 = 1$ and for every $i > 0$, $A_i = A_{i-1} - 1 + \xi_i$. Observe that

$$\begin{aligned}\mathbb{P}\left(|\mathcal{T}_p| > n\right) &= \mathbb{P}\left(A_1, \ldots, A_n > 0\right) && (5.14)\\ &\leq \mathbb{P}\left(A_n > 0\right) && (5.15)\\ &\leq \mathbb{P}\left(A_0 + \sum_{i=1}^{n} \xi_i > n\right) && (5.16)\\ &= \mathbb{P}\left(\sum_{i=1}^{n} \xi_i \geq n\right) , && (5.17)\end{aligned}$$

where the inequality follows from Markov's inequality For every $t \geq 0$ it holds that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{k} \xi_i \geq k\right) &= \mathbb{P}\left(e^{t\sum_{i=1}^{k}\xi_i} \geq e^{tk}\right) \\ &\leq e^{-tk}\mathbb{E}\left[e^{t\sum_{i=1}^{k}\xi_i}\right] \\ &= \left(\tfrac{\mathbb{E}[e^{t\xi}]}{e^t}\right)^k ,\end{aligned}$$

where the inequality follows from Markov's inequality. We obtain

$$\begin{aligned}e^{-t}\mathbb{E}[e^{t\xi}] &= e^{-t} \cdot \sum_{i=0}^{\infty} \mathbb{P}(\xi = i) \cdot e^{ti} \\ &= e^{-t} \cdot \sum_{i=0}^{\infty}\sum_{k=i}^{\infty} \tfrac{(1-p)^k p}{k+1} \cdot e^{ti} \\ &= e^{-t} \cdot \sum_{k=0}^{\infty}\sum_{i=0}^{k} \tfrac{(1-p)^k p}{k+1} \cdot e^{ti} \\ &= e^{-t} \cdot \sum_{k=0}^{\infty} \tfrac{(1-p)^k p}{k+1} \cdot \sum_{i=0}^{k} e^{ti} \\ &= e^{-t} \cdot \sum_{k=0}^{\infty} \tfrac{(1-p)^k p}{k+1} \cdot \tfrac{e^{t(k+1)}-1}{e^t-1} \\ &= \tfrac{p}{1-p} \cdot \tfrac{1}{e^t(e^t-1)} \cdot \sum_{k=1}^{\infty} \tfrac{(1-p)^k(e^{tk}-1)}{k} \\ &= -\tfrac{p}{1-p} \cdot \tfrac{1}{e^t(e^t-1)} \cdot \left(\ln((1-(1-p)e^t) - \ln p\right)\end{aligned}$$

where in the last equality we used the facts that $(1-p)e^t \in (0,1)$ and $p \in (0,1)$ and that $\sum_{i=1}^{\infty} x^i/i = -\ln(1-x)$ when $|x| < 1$. Let $\gamma = e^t - 1$. We obtain:

$$\begin{aligned}e^{-t}\mathbb{E}[e^{t\xi}] &= -\tfrac{p}{1-p} \cdot \tfrac{1}{\gamma(1+\gamma)} \cdot \ln\left(\tfrac{1-(1-p)e^t}{p}\right) \\ &= -\tfrac{p}{1-p} \cdot \tfrac{1}{\gamma(1+\gamma)} \cdot \ln\left(1 - \tfrac{(1-p)\gamma}{p}\right) \\ &= \tfrac{1}{1+\gamma} \cdot \sum_{i=1}^{\infty} \left(\tfrac{(1-p)\gamma}{p}\right)^{i-1} \cdot \tfrac{1}{i} \\ &= 1 - \left(1 - \tfrac{1-p}{2p}\right) \cdot \tfrac{\gamma}{1+\gamma} + O(\gamma^2) \\ &= 1 - \tfrac{1}{2}\left(3 - \tfrac{1}{p}\right) \cdot \tfrac{\gamma}{1+\gamma} + O(\gamma^2) \, .\end{aligned}$$

Since $p > 1/3$, there exists t and an absolute constant $c > 0$ such that $e^{-t}\mathbb{E}[e^{t\xi}] \leq 1 - c(3 - \frac{1}{p})$. We deduce $\mathbb{P}(|\mathcal{T}| > n) \leq (1 - c(3 - \frac{1}{p}))^n$.

□

6 Discussion

The Forest Fire model was proposed by Leskovec *et al.* to explain several properties of social networks, *shrinking diameter* being an important one, in addition to

densification and heavy-tailed degree distributions [22]. As the graphs generated are directed, we focused on distance to the seed graph, rather than diameter as the property of interest. This work shows that in a restricted version of the Forest Fire model, we can prove that this distance remains bounded, even as the graph size increases, albeit with some conditions on the seed graph. We show that qualitatively similar results can also be obtained in an even more basic model, the random walk model. Our upper and lower bounds can be used to give bounds on the undirected diameter, with loss of constant factors.

There are several natural open questions. The obvious one, is whether one can remove the conditions on the seed graph, or the requirement of multiple-edges in the random walk case. Our simulation results seem to suggest that starting with a single node as a seed graph should also result in similar behaviour. The next is whether one can address densification. Without backward burning, it is clear that the out-degree of any vertex in G_t can be at most logarithmic in t. This follows from the fact that the edges have to be on directed paths in the ambassador tree. Thus, we cannot expect the average edge density to be more than logarithmic in the number of nodes. Also, for this reason the out-degrees cannot have a heavy tail. In simulations, the in-degrees did exhibit power law behaviour.

References

[1] C. Avin, Z. Lotker, Y. Nahum, and D. Peleg. Core size and densification in preferential attachment networks. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015*, pages 492–503, 2015.

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

[3] A. Blum, T. Chan, and M. Rwebangira. A random-surfer web-graph model. In *Proceedings of the Third Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2006*, pages 238–246, 2006.

[4] A. Chaintreau, P. Fraigniaud, and E. Lebhar. Networks become navigable as nodes move and forget. In *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008*, pages 133–144, 2008.

[5] P. Chebolu and P. Melsted. Pagerank and the random surfer model. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008*, pages 1010–1018. Society for Industrial and Applied Mathematics, 2008.

[6] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2003.

[7] C. Cooper and A. Frieze. A general model of web graphs. *Random Struct. Algorithms*, 22(3):311–335, 2003.

[8] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24(3):277–298, 1987.

[9] M. Draief and L. Massoulié. *Epidemics and Rumours in Complex Networks*. Cambridge University Press, New York, NY, USA, 1st edition, 2010.

[10] J. Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *Journal of Applied Probability*, 36(2):pp. 301–309, 1999.

[11] J. Gravner. Branching Processes. https://www.math.ucdavis.edu/~gravner/MAT135B/materials/ch14.pdf. Online, accessed 7-July-2015.

[12] T. Gutierrez-Bunster, U. Stege, A. Thomo, and J. Taylor. How do biological networks differ from social networks? (an experimental study). In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014, 2014*, pages 744–751, 2014.

[13] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):pp. 502–525, 1982.

[14] J. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14, NIPS 2001*, pages 431–438, 2001.

[15] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The web as a graph: Measurements, models, and methods. In *Computing and Combinatorics*, volume 1627 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, 1999.

[16] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The web as a graph: Measurements, models, and methods. In *COCOON*, pages 1–17, 1999.

[17] A. Klenke and L. Mattner. Stochastic ordering of classical discrete distributions. *Adv. in Appl. Probab.*, 2(42):392–410, 2010.

[18] P. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63:066123, May 2001.

[19] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proceedings. 41st Annual Symposium on Foundations of Computer Science. 2000*, pages 57–65, 2000.

[20] S. Lattanzi and D. Sivakumar. Affiliation networks. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009*, pages 427–434, 2009.

[21] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *Knowledge Discovery in Databases: PKDD 2005*, pages 133–145, 2005.

[22] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.

[23] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, in preparation,

version available at `http://pages.iu.edu/~rdlyons/`, 2015.

[24] A. Mehrabian and N. Wormald. It's a small world for random surfers. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014*, pages 857–871, 2014.

[25] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[26] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[27] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. *Internet Mathematics*, 3(1):1–20, 2006.

[28] C. Seshadhri, T. Kolda, and A. Pinar. Community structure and scale-free collections of erdős-rényi graphs. *Phys. Rev. E*, 85:056109, May 2012.

[29] C. Seshadhri, A. Pinar, and T. Kolda. The similarity between stochastic kronecker and chung-lu graph models. In *Proceedings of the Twelfth SIAM International Conference on Data Mining*, pages 1071–1082, 2012.

[30] A. Vazquez. Disordered networks generated by recursive searches. *EPL (Europhysics Letters)*, 54(4):430, 2001.

A Background: Galton-Watson processes

The analysis uses reductions to Galton-Watson branching processes. A Galton-Watson process is a stochastic process $\{X_n\}$ which evolves according to the recurrence formula $X_0 = 1$ and $X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$, where $\{\xi_j^{(n)} : n, j \in \mathbb{N}\}$ is a set of i.i.d. natural number-valued random variables. The interpretation is as follows: the process builds a random tree. X_n can be thought of as the number of descendants of the root in the n^{th} generation, and $\xi_j^{(n)}$ can be thought of as the number of children (in generation $n+1$) of the j^{th} of these (n^{th} generation) descendants. The recurrence relation states that the number of descendants in the $(n+1)^{st}$ generation is the sum, over all n^{th} generation descendants, of the number of children of that descendant. For more information, see [23].

B Auxiliary lemmas

LEMMA B.1. *Let $X_n \sim Bin(n, \alpha/n)$ for all n. Let $Y \sim Poi(e\alpha)$. Let $\alpha \leq 1/e$. For all n and $k \geq 1$ we have $Pr(X_n \geq k) \leq Pr(Y \geq k)$.*

Proof. Fix an arbitrary n and $k \geq 1$. $Pr(X_n = k) \leq \binom{n}{k}(\alpha/n)^k \leq \alpha^k/k! \leq \frac{\alpha^k}{k!}\frac{e}{e^{e\alpha}} \leq \frac{(e\alpha)^k}{k!}\frac{1}{e^{e\alpha}} = Pr(Y = k)$. □

LEMMA B.2. (CHERNOFF BOUND [26]) *Let $X = \sum_i X_i$ be the sum of 0/1 independent random variables. Let $R \geq 6E[X]$. Then $Pr(X \geq R) \leq 2^{-R}$.*

LEMMA B.3. (POISSON TAIL BOUND) *Let $\alpha \geq 1$. Let $\alpha' = 1 + \lceil e\alpha \rceil + e\alpha$. Let N_i be sum of N_{i-1} independent $(1 + \lceil e\alpha \rceil + \mathsf{Poisson}(e\alpha))$-distributed random variables. Let $b_i = (6\alpha')^i(k+1)$. Then, $\mathbb{P}(N_i \geq b_i | N_{i-1} \leq b_{i-1}) \leq 2^{-6e\alpha b_{i-1}}$.*

Proof. Due to the independence of the random variables, we have $N_{i-1} \cdot \mathsf{Poisson}(e\alpha) = \mathsf{Poisson}(e\alpha N_{i-1})$. In the following we fix $N_{i-1} = n$ for $n \leq b_{i-1}$. We derive, by using the definition b_i and by applying the Poisson tail bound given in [25],

$$
\begin{aligned}
&\mathbb{P}(N_i \geq b_i | n \leq b_{i-1}) \leq \\
&\leq \mathbb{P}(\mathsf{Poisson}(e\alpha n) + n(1 + \lceil e\alpha \rceil) \geq b_i | n \leq b_{i-1}) \\
&\leq \mathbb{P}(\mathsf{Poisson}(e\alpha n) \geq (6\alpha')b_{i-1} - b_{i-1}(1 + \lceil e\alpha \rceil) | n \leq b_{i-1}) \\
&\leq \mathbb{P}(\mathsf{Poisson}(e\alpha n) \geq 6e\alpha b_{i-1} | n \leq b_{i-1}) \\
&\leq \frac{e^{-e\alpha n}(e^2\alpha n)^{6e\alpha b_{i-1}}}{(6e\alpha b_{i-1})^{6e\alpha b_{i-1}}} \\
&\leq \frac{(e^2\alpha b_{i-1})^{6e\alpha b_{i-1}}}{(6e\alpha b_{i-1})^{6e\alpha b_{i-1}}} \\
&\leq 2^{-6e\alpha b_{i-1}}.
\end{aligned}
$$

□