# DYNAMICS OF INFORMATION NETWORKS

ANDREI SONTAG [iD],*
TIM ROGERS,* ** AND
CHRISTIAN A YATES,* *** *University of Bath*

## Abstract

We explore a simple model of network dynamics which has previously been applied to the study of information flow in the context of epidemic spreading. A random rooted network is constructed that evolves according to the following rule: at a constant rate, pairs of nodes $(i, j)$ are randomly chosen to interact, with an edge drawn from $i$ to $j$ (and any other out-edge from $i$ deleted) if $j$ is strictly closer to the root with respect to graph distance. We characterise the dynamics of this random network in the limit of large size, showing that it instantaneously forms a tree with long branches that immediately collapse to depth two, then it slowly rearranges itself to a star-like configuration. This curious behaviour has consequences for the study of the epidemic models in which this information network was first proposed.

*Keywords:* Information spread; random recursive trees; preferential attachment

2020 Mathematics Subject Classification: Primary 60J20
Secondary 91D30

## 1. Introduction

There is a growing awareness in the mathematical epidemiology community of the need to model not just the infective status of individuals in a population, but also their knowledge and attitudes. In important early work in this direction, [13] combined a simple model of information spread with standard epidemic dynamics to create a new class of models in which there is feedback between the progress of the disease and altered behaviours as a result of risk awareness. It has been shown that this feedback can fundamentally change the trajectory of an outbreak [12, 13, 15, 16]. Information networks responding to disease and other hazards are of course not unique to humans [23], and the ability to transmit information accurately and efficiently confers significant evolutionary advantages in animal populations. Hence, it is of fundamental interest to understand how information is transmitted in populations and how networks linking information sources are formed.

The literature on the spread of information, such as rumours and opinions, has a wide range of fascinating models and dynamics. General prominent modelling approaches include the use of voter models [8], opinion dynamics [1, 2, 5, 28], rumour spread models [10], first-passage percolation [3, 29], and agent-based models [14]. More specifically, we highlight the use of

stochastic differential equations for representative voters and the microstructure of elections [7], and a point process model based on the assumption that fake news spreads as a two-stage process [19]. The simultaneous transmission of both information and misinformation (or hoaxes) in communities, however, is still under development. The most promising approaches consider contagion processes in networks [24–27] and agent-based models [6, 9, 21].

In this article, we explore in greater depth the paradigmatic model of [13] that has been used to describe the intricate dynamics of awareness spread. This model has been commonly used in conjunction with compartment-based epidemic models, for which analysis is often performed in the limit of large population sizes. For this reason, we are primarily interested in the dynamics of the information network in this same limit. Readers familiar with the literature in probability will recognise the model of [13] as a random recursive tree [18, 20] with a rewiring mechanism, or a preferential attachment tree [4, 11, 17]. As we show in this article, the model exhibits an unusual dynamic in infinite populations: it instantaneously forms long branches, which then immediately collapse to a tree of depth two; it then slowly rearranges to a star-like configuration in a process that takes infinite time.

The article is organised as follows. In Section 2, we introduce the model and key definitions that will be important for expressing our main results in Section 3. In the model, individuals are represented by nodes which are linked whenever information is exchanged between them. One informed node is defined as the root, while the remaining nodes initially have no information. In the infinite population size limit, our theorems show that an information tree containing all nodes forms instantly almost surely, while it takes infinite time until everyone obtains information of the best quality. In this same limit, we show that nodes have depth two almost surely at any time $t > 0$, yet the longest branch ever formed in the tree scales with the logarithm of the population size. This suggests a strange behaviour of the model, where long branches form and collapse in a fraction of time. Proofs of the theorems are provided in Section 4.

## 2. The model and key definitions

Consider a population of identical individuals modelled as nodes in an evolving network. One node is distinguished as the root, to be thought of as the originator of a piece of information. It is convenient to label the nodes by non-negative integers and adopt the convention that node zero is the root. To each pair of nodes $(i, j)$ we associate an independent unit-rate Poisson clock; whenever this clock rings we draw a directed edge from $i$ to $j$ (and delete all other out-edges from $i$) if $j$ is strictly closer to the root. Figure 1 shows snapshots of two simulations of the model, restricted to $N = 100$ and $N = 1000$ nodes.

This is the network model that underlies the early-time and large-population dynamics of the disease awareness model introduced in [13]. In that work, individuals are classified by their information level, with zero representing first-hand information (in the epidemic context, someone currently infected), and generally level $n$ describing people for whom information has travelled $n$ steps to reach. In our model the information level of a node corresponds to its distance from the root. Two extra processes were included in [13] that we do not include here: the fading of information (spontaneous lowering of an individual's information level) and generation of new information (when an infection takes place, the infected individual jumps to information level zero). We have chosen to leave out these elements in order to focus on the dynamics of the information network pertaining to a single initial piece of information. For an analysis of finite-size effects when information fading is included, see our companion work [22].
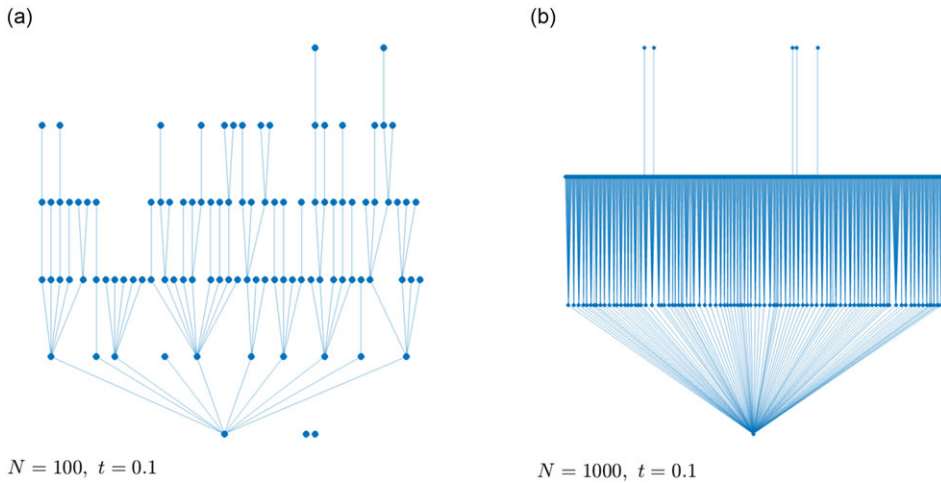
FIGURE 1. Snapshots from two simulations of the information network with different sized populations. (a) Typical structure at $t = 0.1$ for a population of $N = 100$; nodes are spread over several layers and some are yet to attach. (b) Typical structure at $t = 0.1$ for a population of $N = 1000$; nodes are almost entirely condensed into the first two layers.

**Definition 1.** Let $\tau_{ij}$ be the set of (random) times when nodes i and j interacted (whether any edges are redrawn or not). We use $v_i(t)$ to denote the information level of node i at time t. Naturally, $v_0(t) = 0$ for all $t \geq 0$, while for $i > 0$ we have

$$v_i(t) = 1 + \inf\{v_j(s) : j \neq i, s \in \tau_{ij}, s < t\},$$

where we assume the convention $\inf\{\varnothing\} = \infty$.

Many of our results require examination of finite subsets of the population. To that end we introduce, for $i > 1$, the information level of node $i$ when it only has sight of the first $N$ nodes in the network:

$$v_i^N(t) = 1 + \inf\{v_j^N(s) : j \neq i, s \in \tau_{ij}, s < t, j \leq N\},$$

where again $v_0^N(t) = 0$ for all $t \geq 0$.

We note that it is not a priori obvious that the model is well defined in the case of an infinite population. However, the random variables $v_i^N(t)$ are monotonically non-increasing in $N$, which allows us access to several almost sure results characterising the large-$N$ limit.

## 3. Main results

Initially, the network is empty and all nodes are uninformed except the root. The only stationary configuration is the star graph in which every node is attached to the root. Our first results concern the time evolution of the graph in between these limits.

**Theorem 1.** (The graph is a tree.) *Denote by $T_N$ the time until nodes $\{0, \ldots, N-1\}$ are all informed (i.e. are connected to the information network). Then $\mathbb{P}(\lim_{N \to \infty} T_N = 0) = 1$.*

**Theorem 2.** (The graph is never a star.) *Let $T_N^*$ be the first hitting time of the star configuration amongst the first N nodes; that is, $T_N^* = \inf\{t: \max_{i<N} v_i^N(t) = 1\}$. Then $\mathbb{P}(\lim_{N\to\infty} T_N^* = \infty) = 1$.*

The next results characterise the shape of the tree as it evolves. We find that at all positive times the tree has a maximum depth of two, and compute the probability to find a given node at a particular depth.

**Theorem 3.** (The tree has maximum depth two.) *For all $t > 0$,*

$$\mathbb{P}\Big( \lim_{N\to\infty} \max_{i<N}\{v_i^N(t)\} \le 2 \Big) = 1.$$

**Theorem 4.** (Node depth distribution.) *Let $\tau_{i0}$ be the time when node i first interacted with the root. For all $t > 0$, and all nodes $i \neq 0$,*

$$\mathbb{P}\Big( \lim_{N\to\infty} v_i^N(t) = 1 + \mathbf{1}_{\tau_{i0}>t} \Big) = 1,$$

*where, by the definition of the model, $\mathbf{1}_{\tau_{i0}>t}$ is Bernoulli distributed with parameter $e^{-t}$. Furthermore, for all $i \neq 0$, $\mathbb{P}(\lim_{t\to 0^+} v_i(t) = 2) = 1$.*

Hitherto, our theorems have mostly concerned the state of the system at any given time $t$ in the limit of infinite network size. The previous result combined with our next theorem probes the early time limit $t \to 0$, establishing the entrance law of the process in the system with an infinite number of nodes.

**Theorem 5.** (There is always at least one node with depth one.)

$$\mathbb{P}\Big( \lim_{t\to 0^+} \inf_{i\neq 0}\{v_i(t)\} = 1 \Big) = 1.$$

These results characterise the instantaneous arrangement of the infinite network into a tree of depth two. Amongst the first $N$ nodes, however, we can show that branches of depth $\log N$ are certain to have existed at some point. This implies an interesting explosive behaviour in which arbitrarily long branches form and then immediately collapse.

**Theorem 6.** (Longest branch.) *Consider the length of the longest branch ever formed in the network amongst the first N nodes, $D_N = \max(\{v_i^N(t), t > 0, i \le N\} \setminus \{\infty\})$. This quantity grows logarithmically with N; specifically,*

$$\lim_{N\to\infty} \mathbb{P}\left( \frac{D_N}{\log N} \in [e/2, e] \right) = 1.$$

Together, these results give a complete characterisation of the (somewhat strange) dynamics of the information network: the graph instantaneously forms a tree with long branches which then promptly collapse to depth two; it then takes an infinitely long amount of time to slowly rearrange itself towards a star-like configuration.

## 4. Proofs of the main theorems

In this section we provide the proofs for the main theorems stated in Section 3. We also give two corollaries that relate Theorems 1 and 2 to the variables $v_i^N(t)$.

*Proof of Theorem 1.* For the system with $N$ nodes $\{0, \ldots, N-1\}$, $T_N$ is the sum of $N-1$ independent random variables, $T_N = \sum_{n=1}^{N-1} W_{n,N}$, where we define $W_{n,N}$ as the time between

the $n$th and $(n+1)$th nodes joining the informed group. By definition, $W_{0,N} = 0$, since the initial condition is one node at level 0.

The transition from $n$ informed nodes to $n+1$ occurs when one of the $N-n$ uninformed nodes interacts with one of the $n$ informed nodes in any information level. This transition occurs at a rate $n \cdot (N-n)$. The transition time $W_{n,N}$ is then exponentially distributed with rate $n(N-n)$, mean $(n(N-n))^{-1}$, and variance $(n(N-n))^{-2}$. Hence,

$$\mathbb{E}[T_N] = \sum_{n=1}^{N-1} \mathbb{E}[W_{n,N}] = \sum_{n=1}^{N-1} \frac{1}{n(N-n)} = \frac{1}{N} \sum_{n=1}^{N-1} \left( \frac{1}{n} + \frac{1}{N-n} \right) = \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n} = \frac{2H_{N-1}}{N},$$

$$\mathrm{Var}[T_N] = \sum_{n=1}^{N-1} \mathrm{Var}[W_{n,N}] = \sum_{n=1}^{N-1} \frac{1}{n^2(N-n)^2} = \frac{1}{N^2} \sum_{n=1}^{N-1} \left( \frac{1}{n} + \frac{1}{N-n} \right)^2$$

$$= \frac{2}{N^2} \left( \sum_{n=1}^{N-1} \frac{1}{n^2} \right) + \frac{4}{N^3} \left( \sum_{k=1}^{N-1} \frac{1}{n} \right) = \frac{2}{N^2} \left( \frac{\pi^2}{6} - \psi'(N) \right) + \frac{4}{N^3} H_{N-1},$$

where $H_n = \sum_{k=1}^{n} (1/k)$ is the $n$th harmonic number, and $\psi'(N+1)$ is the first derivative of the digamma function.

We will use the Borel–Cantelli lemma to prove almost sure convergence. Observe that $T_N > 0$ and $T_N > \varepsilon \Rightarrow T_N^2 > \varepsilon^2$. Thus, for all $\varepsilon > 0$,

$$\mathbb{P}(T_N > \varepsilon) = \mathbb{P}(T_N^2 > \varepsilon^2) \leq \frac{\mathbb{E}[T_N^2]}{\varepsilon^2}$$

$$\leq \frac{1}{\varepsilon^2} \left[ \frac{2}{N^2} \left( \frac{\pi^2}{6} - \psi'(N) \right) + \frac{4}{N^3} H_{N-1} + \frac{4H_{N-1}^2}{N^2} \right]$$

by Chebyshev's inequality, where we used $\mathbb{E}[T_N^2] = \mathrm{Var}[T_N] + \mathbb{E}[T_N]^2$.

Fix $\varepsilon > 0$, and consider the set of events $A_N^\varepsilon = \{T_N > \varepsilon\}$. It follows that

$$\sum_{N=1}^{\infty} \mathbb{P}(A_N^\varepsilon) = \sum_{N=1}^{\infty} \mathbb{P}(T_N > \varepsilon) \leq \sum_{N=1}^{\infty} \frac{1}{\varepsilon^2} \left[ \frac{2}{N^2} \left( \frac{\pi^2}{6} - \psi'(N) \right) + \frac{4}{N^3} H_{N-1} + \frac{4H_{N-1}^2}{N^2} \right].$$

It is easy to see that the sum converges. By the Borel–Cantelli lemma, it follows that $\mathbb{P}(\limsup_{N \to \infty} A_N^\varepsilon) = \mathbb{P}(\limsup_{N \to \infty} \{T_N > \varepsilon\}) = 0$.

To complete the proof, let $\varepsilon_k$, $k \in \mathbb{N}$, be a sequence of decreasing times such that $\varepsilon_k < \varepsilon_{k-1}$, $\varepsilon_k > 0$, and $\lim_{k \to \infty} \varepsilon_k = 0$. Hence, $A_N^{\varepsilon_k} \subset A_N^{\varepsilon_{k+1}}$, and

$$\mathbb{P}\left( \cup_{\varepsilon > 0} \limsup_{N \to \infty} A_N^\varepsilon \right) \leq \sum_{k=1}^{\infty} \mathbb{P}\left( \limsup_{N \to \infty} A_N^{\varepsilon_k} \right) = 0.$$

Consequently, $\mathbb{P}(\lim_{N \to \infty} T_N = 0) = 1$, and the result is proved. $\qquad\square$

The previous proof that the graph is a tree shows that the time until all nodes in the population are aware is almost surely zero. Hence, all nodes, and in particular the maximum of the population, have to be aware almost surely at any time $t > 0$. This suggests the following proposition, which relates the previous result with the maximum information level in the population.

**Proposition 1.** *For all times $t > 0$, $\mathbb{P}(\lim_{N\to\infty}\max_{i<N}\{v_i^N(t)\} < \infty) = 1$.*

In fact, Theorem 3 is a stronger version of this proposition, so we will not prove this weak form here.

*Proof of Theorem* 2. Similarly to the previous case, $\sup_{i<N} v_i^N(t) = 1$ if and only if all nodes have interacted with the root, i.e. the graph is a star. Thus, the time until the star is formed can be written as the sum of the transition times between each node connecting to the root. That is, the time $T_N^*$ until the star forms is again the sum of $N-1$ independent random variables, $T_N^* = \sum_{n=1}^{N-1} V_{n,N}$, where we define $V_{n,N}$ as the time between the $n$th node and the $(n+1)$th connecting to the root. Again, $V_{0,N} = 0$, as there is one node at level 0 from the beginning.

The transition from $n$ to $n+1$ nodes occurs when one of the remaining $N-n$ nodes not connected to the root interacts with the root. This transition has rate $1 \cdot (N-n)$. The transition time $V_{n,N}$ is then an exponentially distributed random variable with rate $(N-n)$, mean $(N-n)^{-1}$, and variance $(N-n)^{-2}$. Consider now the set of random variables $\{\mathcal{V}_n\}_{n=1}^\infty$ such that $\mathcal{V}_n = (1/n)X_n$, where $X_n \sim \mathrm{Exp}(1)$ are independent and identically distributed (i.i.d.). Observe that $\mathcal{V}_n \sim \mathrm{Exp}(n)$. Instead of writing $T_N^*$ as a sum of random variables that depend on $N$, we can use the set $\{\mathcal{V}_n\}_{n=1}^\infty$ to write $T_N^* = \sum_{n=1}^{N-1} \mathcal{V}_n = \sum_{n=1}^{N-1} (1/n)X_n$.

By the monotone convergence theorem, $T_N^* \to T^* \in \mathbb{R} \cup \{\infty\}$.

We will show that $T^* = \infty$ by contradiction. First, consider the random variable

$$S_M = \frac{1}{M-1}\sum_{N=2}^{M} T_N^* = \frac{1}{M-1}\sum_{N=2}^{M}\sum_{n=1}^{N-1}\frac{1}{n}X_n$$

$$= \frac{1}{M-1}\sum_{n=1}^{M-1}(M-n)\frac{X_n}{n}$$

$$= \left(1 + \frac{1}{M-1}\right)T_M^* - \frac{1}{M-1}\sum_{n=1}^{M-1}X_n. \tag{1}$$

Assume $T^*$ is finite. Then, $\lim_{M\to\infty} T_M^*/(M-1) = 0$ almost surely (a.s.). Since $\lim_{M\to\infty} T_M^* \xrightarrow{\text{a.s.}} T^*$, it follows that $\lim_{M\to\infty} S_M \xrightarrow{\text{a.s.}} T^*$. Additionally, by the strong law of large numbers, $\lim_{M\to\infty} (1/(M-1))\sum_{n=1}^{M-1} X_n \xrightarrow{\text{a.s.}} 1$. Taking the limit $M \to \infty$ in (1) leads to $T^* = T^* - 1$, contradicting our assumption that $T^*$ is finite. Thus, $T^* = \infty$ and $T_N^* \to \infty$ almost surely. $\square$

The previous theorem can be written in terms of the $v_i^N(t)$ variables as the following corollary.

**Corollary 1.** *At all times $t > 0$, $\mathbb{P}(\lim_{N\to\infty}\max_{i<N} v_i^N(t) > 1) = 1$.*

The star graph has all nodes, excluding the root, at depth one. Since the maximum is a.s. not at depth one, the tree cannot be a star.

*Proof of Theorem* 3. Taking the system with nodes $\{0, \ldots, N-1\}$, consider first the probability that the maximum information level in the population has depth greater than two. This is equivalent to the probability that at least one node has depth greater than two.

For $i > 0$, we define the variables $\tau_{ij}^1$ as the time when the first meeting between nodes $i$ and $j$ occurs. Note that the $\tau_{ij}^1$ are i.i.d. $\mathrm{Exp}(1)$ random variables. Now, fix $t > 0$ and consider

the events $A_{ij}^t = \{\tau_{j0} < \tau_{ij}^1 < t\}$, i.e. the event that node $j$ exchanged information with the root before interacting with node $i$ for the first time before time $t$. Using the fact that the random variables $\tau_{ij}^1$ and $\tau_{j0}$ are independent, it follows that

$$\mathbb{P}(A_{ij}^t) = \int_0^t \int_0^s e^{-\tau} \times e^{-s} \, d\tau \, ds = \frac{1}{2} e^{-2t} (e^t - 1)^2 = \varepsilon_t > 0. \tag{2}$$

The events $A_{ij}^t$ are mutually independent for fixed $i$ and varying $j$, since meetings between nodes are independent. Furthermore, if event $A_{ij}^t$ occurred, then $v_i^N(t) \leq 2$. Consequently,

$$\mathbb{P}(v_i^N(t) > 2) \leq \prod_{j \neq 0, j \neq i}^{N-1} \mathbb{P}((A_{ij}^t)^c) = (1 - \varepsilon_t)^{N-2}.$$

Let $F_N^t = \{\max_{i < N} v_i^N(t) > 2\}$. Thus,

$$\mathbb{P}(F_N^t) = \mathbb{P}\left(\bigcup_{i=1}^{N-1} \{v_i^N(t) > 2\}\right) \leq \sum_{i=1}^{N-1} (1 - \varepsilon_t)^{N-2} = (N-1)(1 - \varepsilon_t)^{N-2}.$$

Therefore, $\sum_{N=1}^{\infty} \mathbb{P}(F_N^t) = \sum_{N=1}^{\infty} (N-1)(1 - \varepsilon_t)^{N-2} = \varepsilon_t^{-2} < \infty$.

By the Borel–Cantelli lemma, $\mathbb{P}\left(\bigcap_{k=1}^{\infty} \bigcup_{N=k}^{\infty} \{\max_{i \leq N} \{v_i^N(t)\} > 2\}\right) = 0$. Thus, $\mathbb{P}(\lim_{N \to \infty} \max_{i \leq N} \{v_i^N(t)\} \leq 2) = 1$, and Theorem 3 is proved. □

We next prove the following lemma.

**Lemma 1.** *For all $t > 0$, and all nodes $i \neq 0$, $\mathbb{P}(v_i(t) = 1 + \mathbf{1}_{\tau_{i0} > t}) = 1$, where $1 + \mathbf{1}_{\tau_{i0} > t}$ is the Bernoulli random variable determining whether or not the node interacted with the root yet.*

*Proof.* Consider the system with infinite nodes. As before, we define the variables $\tau_{ij}$ as the time when the first meeting between nodes i and j occurs, but this time only for $0 \leq i < j$.

For $i > 0$, we again define the random variables $\mu_i(t) = 1 + \inf_{j > i} \{\mu_j(\tau_{ij}) : \tau_{ij} < t\}$, $\mu_i(0) = \infty$, and $\mu_0(t) = v_0(t) = 0$; in this case, $\mu_i(t)$ is the quality of information of node $i$ when we allow the nodes to meet and exchange information with nodes $j$, $j > i$, only once. Considering again the events $A_{ij}^t = \{\tau_{j0} < \tau_{ij} < t\}$, i.e. the event that node $j$ exchanged information with the root before interacting with node $i$ before time $t$. Recall that the events $A_{ij}^t$ are mutually independent for different $j$. Moreover, for all $t > 0$, $\sum_{j=i+1}^{\infty} \mathbb{P}(A_{ij}^t) = \sum_{j=i+1}^{\infty} \varepsilon_t = \infty$, where $\varepsilon_t$ is the same as in (2). Thus, by the converse Borel–Cantelli lemma, $\mathbb{P}\left(\bigcup_{j > i} A_{ij}^t\right) = 1$ for any $t > 0$. Therefore, at least one of the events $A_{ij}^t$ occurred (in fact, infinitely many of them). Consequently, $\mathbb{P}(v_i(t) \leq 2) \geq \mathbb{P}(\mu_i(t) \leq 2) = 1$. Since $v_i(t) \leq 2 \Rightarrow v_i(t) \in \{1, 2\}$, and $\mathbb{P}(v_i(t) = 1) = 1 - e^{-t}$, the probability that node $i$ interacted with the root, it follows that $\mathbb{P}(v_i(t) = 2) = e^{-t}$. As a final step, note that $v_i(t) = 1 \iff \{\tau_{i0} < t\}$, by definition. Ergo, $1 + \mathbf{1}_{\tau_{i0} > t}$ and $v_i(t)$ are defined in the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and

$$\mathbb{P}(\{\omega \in \Omega : v_i(t)[\omega] = 1 + \mathbf{1}_{\tau_{i0} > t}[\omega]\}) = 1 \implies \mathbb{P}(v_i(t) = 1 + \mathbf{1}_{\tau_{i0} > t}) = 1. \quad \square$$

*Proof of Theorem 4.* To complete the proof of Theorem 4, note that $\lim_{N \to \infty} v_i^N(t) = v_i(t)$ by Definition 1. Thus, from Lemma 1, $\mathbb{P}(\lim_{N \to \infty} v_i^N(t) = 1 + \mathbf{1}_{\tau_{i0} > t}) = 1$, and since $\mathbb{P}(\tau_{i0} > 0) = 1$ given $\tau_{i0} \sim \text{Exp}(1)$, it follows that $\mathbb{P}(\lim_{t \to 0^+} v_i(t) = 2) = 1$, and the theorem is proved. □

*Proof of Theorem* 5. Define the random variables $X_n = \inf_{i \neq 0, i \leq n}\{v_i(n^{-\alpha})\}$ and $Y_n = \inf_{i \neq 0}\{v_i(n^{-\alpha})\}$ for fixed $\alpha \in (0, 1)$, i.e. $Y_n$ is the infimum information level when considering all nodes in the infinite tree at time $n^{-\alpha}$, while $X_n$ is the infimum information level of the subset of nodes $\{1, \ldots, n\}$ of the infinite tree.

Note that, almost surely, $Y_n \leq X_n$ for all $n$, and $1 \leq Y_n$. Hence,

$$1 \leq \lim_{t \to 0^+} \inf_{i \neq 0}\{v_i(t)\} = \lim_{n \to \infty} Y_n \leq \lim_{n \to \infty} X_n.$$

We want to show that $X_n \xrightarrow{\text{a.s.}} 1$ as $n \to \infty$. By definition, $\inf_{i \neq 0, i \leq n}\{v_i(t)\} > 1 \Leftrightarrow$ for all $i \neq 0$, $i \leq n$, $v_i(t) > 1$, i.e. not a single node has interacted with the root before time $t$. As we know, the probability that a given node did not interact with the root before time $t$ is $\mathrm{e}^{-t}$. Thus, $\mathbb{P}(X_n > 1) = \mathrm{e}^{-n^{1-\alpha}}$, and $\sum_{n=1}^{\infty} \mathbb{P}(X_n > 1) = \sum_{n=1}^{\infty} \mathrm{e}^{-n^{1-\alpha}} < \infty$.

By the Borel–Cantelli lemma,

$$\mathbb{P}\left(\lim_{n \to \infty} X_n > 1\right) = 0 \Longrightarrow \mathbb{P}\left(\lim_{n \to \infty} X_n = 1\right) = 1,$$

because $\{X_n > 1\}^{\mathrm{c}} = \{X_n = 1\}$. Therefore,

$$\mathbb{P}\left(\lim_{n \to \infty} Y_n \leq 1\right) = \mathbb{P}\left(\lim_{n \to \infty} Y_n = 1\right) = \mathbb{P}\left(\lim_{t \to 0^+} \inf_{i \neq 0}\{v_i(t)\} = 1\right) = 1. \qquad \square$$

*Proof of Theorem* 6. In our model, the tree is frequently reshaped by reattachment events that move nodes closer to the root. Moreover, the rate and consequences of these events depends on the configuration of the tree. It will turn out, however, that long branches form before any reattachment event has occurred.

Recall that $\tau_{ij}$ denotes the set of times that nodes $i$ and $j$ are in contact, and the information level of node $i$ at time $t$ in the finite-$N$ model is defined as

$$v_i^N(t) = 1 + \inf\{v_j^N(s) \colon j \neq i, s < t, j \leq N, s \in \tau_{ij}\}.$$

Introduce the restricted set of times $\widetilde{\tau}_{ij} = \tau_{ij} \setminus (\inf\{t \colon v_i^N(t) < \infty\}, \infty)$ in which all contact events after the initial attachment of $i$ are discarded. These times define a coupled instance of the model in which reattachment is ignored via

$$\widetilde{v}_i^N(t) = 1 + \inf\{\widetilde{v}_j^N(s) \colon j \neq i, s < t, j \leq N, s \in \widetilde{\tau}_{ij}\}.$$

Since contact events can only lower the information level of a node, we immediately have the bound

$$v_i^N(t) \leq \widetilde{v}_i^N(t). \tag{3}$$

Additionally, the two processes are identical up to the random time $t_N^{\star}$ of the first reattachment event. To control the maximal informational level in the original model, we consider the structure of the coupled process at time $t_N^{\star}$ and in its end state.

The network constructed by the process ignoring reattachment is an instance of a random recursive tree [18, 20]. Write $\widetilde{D}_M$ for the length of the longest branch of this tree when it has $M$ nodes. [20] proved that $\widetilde{D}_M / \log M \xrightarrow{\text{a.s.}} \mathrm{e}$ as $M \to \infty$. This result translates immediately to an almost sure upper bound on the original model via (3). Specifically, $D_N < \widetilde{D}_N$ implies

$$\mathbb{P}\left(\lim_{N \to \infty} \frac{D_N}{\log N} \leq \mathrm{e}\right) = 1,$$

which in turn implies the convergence in probability upper bound in the statement of Theorem 6.

For the lower bound on $D_N$, we compute the size $M_N^\star$ of the random recursive tree at time $t_N^\star$. It turns out that $M_N^\star$ is of order $\sqrt{N}$.

First, note that the rate of attachment events when the informed tree has size $n$ is $n(N - n)$, while the rate of reattachment events is smaller than $n(n - 1)$. Therefore, the probability $q_n$ that an attachment event occurs before a reattachment event when the tree has size $n$ is bounded from below by

$$q_n \geq \frac{n(N - n)}{n(N - n) + n(n - 1)} = \frac{N - n}{N - 1}.$$

It is in fact neater to study $M_{N+1}^\star$, since we may compute

$$\mathbb{P}(M_{N+1}^\star \geq m) \geq \prod_{n=1}^{m} q_n = \frac{N!}{(N - m)! N^m}.$$

Using the well-known factorial bounds

$$\sqrt{2\pi N}\left(\frac{N}{e}\right)^N e^{1/(12N+1)} < N! < \sqrt{2\pi N}\left(\frac{N}{e}\right)^N e^{1/12N},$$

it follows that

$$\mathbb{P}(M_{N+1}^\star \geq m) > \frac{\sqrt{N} N^N \exp\{-N + (1/(12N + 1))\}}{\sqrt{N - m}(N - m)^{(N-m)} \exp\{-N + m + (1/(12(N - m)))\} N^m},$$

$$= \left(\frac{N}{N - m}\right)^{N-m+1/2} \exp\{-m + (1/(12N + 1)) - (1/(12(N - m)))\},$$

$$\geq \exp\left\{-\frac{m^2}{N} + \frac{m}{2N} + \frac{1}{12N + 1} - \frac{1}{12(N - m)}\right\},$$

where we have used the inequality $\log(N - m) \leq \log N - (m/N)$.

To understand the scaling law of $M_N$ we consider $m = N^\alpha$, where $0 < \alpha \leq 1$. From the above, we obtain

$$\mathbb{P}(M_{N+1}^\star \geq N^\alpha) \geq \exp\left\{-N^{2\alpha-1} + \frac{1}{2}N^{\alpha-1} + \frac{1}{12N + 1} - \frac{1}{12N(1 - N^{\alpha-1})}\right\}.$$

Thus, for all $\alpha < \frac{1}{2}$, $\lim_{N\to\infty} \mathbb{P}(M_{N+1}^* \geq N^\alpha) = 1$, and in particular

$$\lim_{N\to\infty} \mathbb{P}(2 \log M_{N+1}^* \geq \log N) = 1.$$

To end the proof of Theorem 6, we note that $D_N \geq \widetilde{D}_{M_N^\star}$, where it is known that almost surely $\widetilde{D}_{M_N^\star} / \log M_N^\star \to e$ as $M_N^\star \to \infty$. So, for any $\epsilon > 0$ we may choose $N$ large enough that

$$\mathbb{P}(\{2 D_N > (1 - \epsilon)e(2 \log M_N^\star)\} \cap \{2 \log M_N^\star \geq \log N\}) > 1 - \epsilon.$$

Sending $\epsilon \to 0$ and passing to large $N$, we obtain

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{D_N}{\log N} \geq \frac{e}{2}\right) = 1. \qquad \square$$

## Acknowledgements

## Funding information

## Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

## References

[1] ACEMOĞLU, D. AND OZDAGLAR, A. (2010). Opinion dynamics and learning in social networks. *Dynamic Games Appl.* **1**, 3–49.

[2] ACEMOĞLU, D., COMO, G., FAGNANI, F. AND OZDAGLAR, A. (2013). Opinion fluctuations and disagreement in social networks. *Math. Operat. Res.* **38**, 1–27.

[3] AUFFINGER, A., DAMRON, M. AND HANSON, J. (2017). *50 Years of First-Passage Percolation* (Univ. Lect. Ser. 68). American Mathematical Society, Providence, RI.

[4] BARABÁSI, A. (2016). *Network Science*. Cambridge University Press.

[5] BAUMGAERTNER, B. O., FETROS, P. A., KRONE, S. M. AND TYSON, R. C. (2018). Spatial opinion dynamics and the effects of two types of mixing. *Physical Review E* **98**, 022310.

[6] BRAINARD, J., HUNTER, P. AND HALL, I. (2020). An agent-based model about the effects of fake news on a norovirus outbreak. *Revue d'Épidémiologie et de Santé Publique* **68**, 99–107.

[7] BRODY, D. C. AND MEIER, D. M. (2022). Mathematical models for fake news. In *Financial Informatics*, eds D. BRODY, L. HUGHSTON, and A. MACRINA. WORLD SCIENTIFIC, Singapore, pp. 405–423.

[8] CASTELLANO, C., FORTUNATO, S. AND LORETO, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646.

[9] CISNEROS-VELARDE, P., OLIVEIRA, D. F. M. AND CHAN, K. S. (2019). Spread and control of misinformation with heterogeneous agents. In *Complex Networks X*, eds S. P. CORNELIUS, C. GRANELL MARTORELL, J. GÓMEZ-GARDEÑES, and B. GONÇALVES. SPRINGER, CHAM, pp. 75–83.

[10] DAVIS, J. T. *et al.* (2020). Phase transitions in information spreading on structured populations. *Nature Phys.* **16**, 590–596.

[11] DURRETT, R. (2006). *Random Graph Dynamics*. Cambridge University Press.

[12] FUNK, S., GILAD, E. AND JANSEN, V. (2010). Endemic disease, awareness, and local behavioural response. *J. Theoret. Biol.* **264**, 501–509.

[13] FUNK, S., GILAD, E., WATKINS, C. AND JANSEN, V. A. A. (2009). The spread of awareness and its impact on epidemic outbreaks. *Proc. Nat. Acad. Sci.* 106, 6872–6877.

[14] HALBACH, P. *et al.* (2020). Investigating key factors for social network evolution and opinion dynamics in an agent-based simulation. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Communication, Organization and Work*, ed V. G. DUFFY. SPRINGER, CHAM, pp. 20–39.

[15] JUHER, D., KISS, I. Z. AND SALDAÑA, J. (2015). Analysis of an epidemic model with awareness decay on regular random networks. *J. Theoret. Biol.* **365**, 457–468.

[16] KISS, I. Z., CASSELL, J., RECKER, M. AND SIMON, P. L. (2010). The impact of information transmission on epidemic outbreaks. *Math. Biosci.* **225**, 1–10.

[17] LYON, M. R. AND MAHMOUD, H. M. (2020). Trees grown under young-age preferential attachment. *J. Appl. Prob.* **57**, 911–927.

[18] MAHMOUD, H. M., SMYTHE, R. T. AND SZYMAŃSKI, J. (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms* **4**, 151–176.

[19] MURAYAMA, T., WAKAMIYA, S., ARAMAKI, E. AND KOBAYASHI, R. (2021). Modeling the spread of fake news on twitter. *PLoS ONE* **16**, e0250419.

[20] PITTEL, B. (1994). Note on the heights of random recursive trees and random *m*-ary search trees. *Random Structures Algorithms* **5**, 337–347.

[21] ROSS, B. *et al.* (2019). Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *Europ. J. Inf. Sys.* **28**, 394–412.

[22] SONTAG, A., ROGERS, T. AND YATES, C. A. (2023). Stochastic drift in discrete waves of nonlocally interacting particles. *Phys. Rev. E* **107**, 014128.

[23] STOCKMAIER, S. *et al.* (2021). Infectious diseases and social distancing in nature. Science 371, eabc8881.

[24] TAMBUSCIO, M., OLIVEIRA, D. F. M., CIAMPAGLIA, G. L. AND RUFFO, G. (2018). Network segregation in a model of misinformation and fact-checking. *J. Comp. Social Sci.* **1**, 261–275.

[25] TAMBUSCIO, M. AND RUFFO, G. (2019). Fact-checking strategies to limit urban legends spreading in a segregated society. *Appl. Network Sci.* **4**, 116.

[26] TAMBUSCIO, M., RUFFO, G., FLAMMINI, A. AND MENCZER, F. (2015). Fact-checking effect on viral hoaxes. In *Proc. 24th Int. Conf. World Wide Web*. ACM, New York.

[27] TÖRNBERG, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE* **13**, e0203958.

[28] VAZQUEZ, F., KRAPIVSKY, P. L. AND REDNER, S. (2003). Constrained opinion dynamics: Freezing and slow evolution. *J. Phys. A* **36**, L61–L68.

[29] XIE, J. *et al.* (2021). Detecting and modelling real percolation and phase transitions of information on social media. *Nat. Hum. Behav.* **5**, 1161–1168.