Probabilistic Analysis of Rumor Spreading Time

Yves Mocquard
IRISA - Université de Rennes 1, France. yves.mocquard@irisa.fr

Bruno Sericola
Inria Rennes - Bretagne Atlantique, France. bruno.sericola@inria.fr

Emmanuelle Anceaume
IRISA - CNRS, Rennes, France. emmanuelle.anceaume@irisa.fr

The context of this work is the well studied dissemination of information in large scale distributed networks through pairwise interactions. This problem, originally called *rumor mongering*, and then *rumor spreading* has mainly been investigated in the synchronous model. This model relies on the assumption that all the nodes of the network act in synchrony, that is, at each round of the protocol, each node is allowed to contact a random neighbor. In this paper, we drop this assumption under the argument that it is not realistic in large scale systems. We thus consider the asynchronous variant, where at random times, nodes successively interact by pairs exchanging their information on the rumor. In a previous paper, we performed a study of the total number of interactions needed for all the nodes of the network to discover the rumor. While most of the existing results involve huge constants that do not allow us to compare different protocols, we provided a thorough analysis of the distribution of this total number of interactions together with its asymptotic behavior. In this paper we extend this discrete-time analysis by solving a conjecture proposed previously and we consider the continuous-time case, where a Poisson process is associated to each node to determine the instants at which interactions occur. The rumor spreading time is thus more realistic since it is the real time needed for all the nodes of the network to discover the rumor. Once again, as most of the existing results involve huge constants, we provide tight bound and equivalent of the complementary distribution of the rumor spreading time. We also give the exact asymptotic behavior of the complementary distribution of the rumor spreading time around its expected value when the number of nodes tends to infinity.

Key words: rumor spreading time, pairwise interactions, Poisson process, Markov chain, analytic performance evaluation

1. Introduction

Randomized rumor spreading is an important mechanism that allows the dissemination of information in large and complex networks through pairwise interactions. This mechanism initially proposed by Demers et al. (1987) for the update of a database replicated at different sites, has then been adopted in many applications ranging from resource discovery as in Harchol-Balter et al. (1999), data-aggregation as in Kempe et al. (2003), complex distributed applications as in Censor-Hillel et al. (2012), or virus propagation in computer networks as in Berger et al. (2005), to mention just a few.

A lot of attention has been devoted to the design and study of randomized rumor spreading algorithms. Initially, some rumor is placed on one of the nodes of a given network, and this rumor is propagated to all the nodes of the network through pairwise interactions between nodes. One of the important questions raised by these protocols is the spreading time, that is time it needs for the rumor to be known by all the nodes of the network.

Several models have been considered to answer this question. The most studied one is the synchronous push-pull model, also called the synchronous random phone call model. This model assumes that all the nodes of the network act in synchrony, which allows the algorithms designed in this model to divide time in synchronized rounds. During each synchronized round, each node i of the network selects at random one of its neighbor j and either sends to j the rumor if i knows it (push operation) or gets the rumor from j if j knows the rumor (pull operation). In the synchronous model, the spreading time of a rumor is defined as the number of synchronous rounds necessary for all the nodes to know the rumor. In one of the first papers dealing with the push operation only, Frieze and Grimmet (1985) proved that when the underlying graph is complete, the ratio of the number of rounds over $\log_2(n)$ converges in probability to $1 + \ln(2)$ when the number n of nodes in the graph tends to infinity.

Further results have been established (see for example Pittel (1987), Karp et al. (2000) and the references therein), the most recent ones resulting from the observation that the rumor spreading time is closely related to the conductance of the graph of the network, see Giakkoupis (2011). Investigations have also been done in different topologies of the network as in Chierichetti et al. (2011), Daum et al. (2016), Fountoulakis and Panagiotou (2013), Panagiotou et al. (2015), in the presence of link or nodes failures as in Feige et al. (1990), in dynamic graphs as in Clementi et al. (2015) and spreading with node expansion as in Giakkoupis (2014).

In distributed networks, and in particular in large scale distributed systems, assuming that all nodes act synchronously is unrealistic. Several authors have recently dropped this assumption by considering an asynchronous model. In the discrete-time case, Acan et al. (2015) study the rumor spreading time for any graph topology. They show that both the average and guaranteed spreading time are $\Omega(n \ln(n))$, where n is the number of nodes in the network. Angluin et al. (2008) analyze the spreading time of a rumor by only considering the push operation (which they call the one-way epidemic operation), and show that with high probability, a rumor injected at some node requires $O(n\ln(n))$ interactions to be spread to all the nodes of the network. This result is interesting, nevertheless the constants arising in the complexity are not determined. In the continuous-time case, Ganesh (2015) considers the propagation of a rumor when there are n independent unit rate Poisson processes, one associated with each node. At a time when there is a jump of the Poisson process associated with node i, this node becomes active, and chooses another node j uniformly at random with which to communicate. Ganesh (2015) analyzes the mean and the variance of the spreading time of the rumor on general graphs and Panagiotou and Speidel (2016) proposes a thorough study for spreading a rumor on particular Erdös-Rényi random graphs.

In the present paper we consider the rumor spreading time in the asynchronous push-pull model for both the discrete and continuous time cases. This model provides minimal assumptions on the computational power of the nodes.

In the discrete-time case, nodes interact by pairs at random and if at least one node possesses the rumor, the other one also gets informed of it. In this case, the spreading time is defined by the number of interactions needed for all the nodes of the network to learn the rumor. In the continuous-time case, as suggested by Ganesh (2015), a Poisson process is associated with each node and at a jump occurrence of Poisson process of a node, this node contacts randomly a neighbor to interact with it as in the discrete-time case, i.e. to get informed of the rumor if one of these two nodes possesses the rumor. The n Poisson processes are suppose to be independent with the same rate.

In Mocquard et al. (2016) we analyzed the rumor spreading time in the discrete-time asynchronous push-pull model. In the present paper we extend the results obtained in Mocquard et al. (2016) in two ways. First, we prove the conjecture formulated therein and second, we deal with the continuous-time asynchronous push-pull model.

The remainder of this paper is organized as follows. Section 2 presents the main results obtained in Mocquard et al. (2016) in the discrete time model needed to solve the continuous-time model. We also prove in this section the conjecture formulated in Mocquard et al. (2016). More precisely, if T_n denotes the total number of interactions needed for all the n nodes to get the rumor then, $\lim_{n\longrightarrow\infty}\mathbb{P}\{T_n > \mathbb{E}(T_n)\} \approx 0.448429663727$, where

$\mathbb{E}(T_n) = (n-1)H_{n-1}$ and H_k is the harmonic series truncated at step k. In Section 3, we consider the continuous time model. A Poisson process is associated with each node and each jump of these independent Poisson processes correspond to an interaction between two different nodes. In this model, the time needed for all the n nodes to get the rumor is denoted by Θ_n. We first give simple expressions of the expected value and variance of Θ_n. Then we give an explicit expression of its distribution and we obtain a simple bound of its complementary distribution which is proved to also be an equivalent of its tail. It is also shown that this bound is much more tight than already known bounds. Finally, we give the limiting distribution of the ratio $\Theta_n/\mathbb{E}(\Theta_n)$ when the number n of nodes tends to infinity. Finally, Section 4 concludes the paper.

2. The discrete time case

We recall in this section the main results obtained in Mocquard et al. (2016) needed to deal with the continuous time case. We also prove the conjecture formulated in Mocquard et al. (2016)

In the discrete time case, the total number of interactions needed so that all the n nodes get the rumor is denoted by T_n. We suppose without any loss of generality that among the n nodes, a single one initially knows the rumour. The case where the number of initial nodes possessing the rumor is greater than one has been considered in Mocquard et al. (2016). A value 0 or 1 is associated with each node. A node with value 1 means that this node knows the rumor and a node with value 0 means that it is not aware of the rumor. For every $t \geq 0$, we denote by $C_t^{(i)}$ the value (0 or 1) of node i at time t. At time 0, all the $C_0^{(i)}$ are equal to 0 except one which is equal to 1 and which corresponds to the node initially knowing the rumor.

At each discrete instant t, two distinct indexes i and j are successively chosen among the set of nodes $1, \ldots, n$ randomly. We denote by X_t the random variable representing this choice and we suppose that this choice is uniform, i.e we suppose that

$$\mathbb{P}\{X_t = (i,j)\} = \frac{1}{n(n-1)} 1_{\{i \neq j\}}.$$

Once the couple (i,j) is chosen at time $t \geq 1$, we have

$$C_t^{(i)} = C_t^{(j)} = \max\{C_{t-1}^{(i)}, C_{t-1}^{(j)}\} \text{ and } C_t^{(m)} = C_{t-1}^{(m)} \text{ for } m \neq i, j.$$

The random variable T_n, defined by

$$T_n = \inf\{t \geq 0 \mid C_t^{(i)} = 1, \text{ for every } i = 1, \dots, n\},$$

represents the number of interactions needed for all the nodes in the network to know the rumor.

We introduce the discrete-time stochastic process $Y = \{Y_t,\ t \geq 0\}$ with state space $\{1, \dots, n\}$ defined, for all $t \geq 0$, by

$$Y_t = \left| \left\{ i \mid C_t^{(i)} = 1 \right\} \right|.$$

The random variable Y_t represents the number of nodes knowing the rumor at time t. The stochastic process Y is then a homogeneous Markov chain with n states, states $1, \dots, n-1$ being transient and state n absorbing. The random variable T_n can then be written as

$$T_n = \inf\{t \geq 0 \mid Y_t = n\}.$$

It is well-known, see for instance Sericola (2013), that the distribution of T_n is given, for every $k \geq 0$, by

$$\mathbb{P}\{T_n > k\} = \alpha Q^k \mathbb{1}, \tag{1}$$

where α is the row vector containing the initial probabilities of states $1, \dots, n-1$, that is $\alpha_i = \mathbb{P}\{Y_0 = i\} = 1_{\{i=1\}}$, Q is the matrix obtained containing the transition probabilities between transient states, that is, as shown in Mocquard et al. (2016),

$$Q_{i,i} = 1 - \frac{2i(n-i)}{n(n-1)} \text{ for } i = 1, \cdots, n-1 \text{ and } Q_{i,i+1} = \frac{2i(n-i)}{n(n-1)}, \text{ for } i = 1, \cdots, n-2 \tag{2}$$

and $\mathbb{1}$ is the column vector of dimension $n-1$ with all its entries equal to 1.

For $i = 0, \dots, n$, we introduce the notation

$$p_i = \frac{2i(n-i)}{n(n-1)}$$

and we denote by H_k the harmonic series defined by $H_0 = 0$ and $H_k = \sum_{\ell=1}^{k} 1/\ell$, for $k \geq 1$.

If we denote by S_i, for $i = 1, \dots, n-1$, the total time spent by the Markov chain Y in state i, then S_i has a geometric distribution with parameter p_i and we have

$$T_n = \sum_{i=1}^{n-1} S_i.$$

2.1. Analysis of the spreading time

The mean time $\mathbb{E}(T_n)$ needed so that all the nodes get the rumor is then given by

$$\mathbb{E}(T_n) = \alpha(I - Q)^{-1}\mathbb{1}, \tag{3}$$

where I is the identity matrix. Its explicit value has been obtained in Mocquard et al. (2016). It is given, for every $n \geq 1$, by

$$\mathbb{E}(T_n) = (n-1)H_{n-1}. \tag{4}$$

An explicit expression of the distribution of T_n, for $n \geq 2$, has been obtained in the following theorem wich will used to deal with the continuous-time case.

THEOREM 1. *For every $n \geq 1$, $k \geq 0$, we have*

$$\mathbb{P}\{T_n > k\} = \sum_{j=1}^{\lfloor n/2 \rfloor} \left(c_{n-1,j}(1-p_j) + k d_{n-1,j}\right)(1-p_j)^{k-1},$$

where the coefficients $c_{n-1,j}$ and $d_{n-1,j}$, which do not depend on k, are given, for $j = 1, \ldots, n-1$, recursively by

$$c_{1,j} = 1_{\{j=1\}} \text{ and } d_{1,j} = 0$$

and for $i \in \{2, \ldots, n-1\}$ by

$$\begin{aligned}
c_{i,j} &= \frac{p_i c_{i-1,j}}{p_i - p_j} - \frac{p_i d_{i-1,j}}{(p_i - p_j)^2} && \text{for } i \neq j, n-j, \\
d_{i,j} &= \frac{p_i d_{i-1,j}}{p_i - p_j} && \text{for } i \neq j, n-j, \\
c_{i,i} &= 1 - \sum_{j=1, j\neq i}^{\lfloor n/2 \rfloor} c_{i,j} && \text{for } i \leq \lfloor n/2 \rfloor, \\
c_{i,n-i} &= 1 - \sum_{j=1, j\neq n-i}^{\lfloor n/2 \rfloor} c_{i,j} && \text{for } i > \lfloor n/2 \rfloor, \\
d_{i,i} &= p_i c_{i-1,i} && \text{for } i \leq \lfloor n/2 \rfloor, \\
d_{i,n-i} &= p_i c_{i-1,n-i} && \text{for } i > \lfloor n/2 \rfloor.
\end{aligned}$$

Proof. See Mocquard et al. (2016). ∎

2.2. Bounds and asymptotic analysis of the distribution of T_n

The following bound and equivalent of the complementary distribution of T_n will be used in the continuous-time case to obtain similar bound and equivalent.

THEOREM 2. *For all $n \geq 2$ and $k \geq 1$ we have*

$$\mathbb{P}\{T_n > k\} \leq \left(1 + \frac{2k(n-2)^2}{n}\right)\left(1 - \frac{2}{n}\right)^{k-1},$$

$$\mathbb{P}\{T_n > k\} \underset{k \longrightarrow \infty}{\sim} \left(1 + \frac{2k(n-2)^2}{n}\right)\left(1 - \frac{2}{n}\right)^{k-1}.$$

Proof. See Mocquard et al. (2016). ■

Recall that $\mathbb{E}(T_n) = (n-1)H_{n-1}$, where H_k is the harmonic series. We proved in Mocquard et al. (2016) that for all real $c \geq 0$, we have

$$\lim_{n \to \infty} \mathbb{P}\{T_n > c\mathbb{E}(T_n)\} = \begin{cases} 0 & \text{if } c > 1 \\ 1 & \text{if } c < 1. \end{cases} \tag{5}$$

For $c = 1$, this result was formulated in Mocquard et al. (2016) as a conjecture. We are now able to give a proof of it.

THEOREM 3.

$$\lim_{n \longrightarrow \infty} \mathbb{P}\{T_n > \mathbb{E}(T_n)\} = 1 - 2e^{-\gamma} K_1\left(2e^{-\gamma}\right) \approx 0.448429663727.$$

where γ is the Euler's constant given by $\gamma = \lim_{n \longrightarrow \infty}(H_n - \ln(n)) \approx 0.5772156649$ and K_1 is the modified Bessel function of the second kind of order 1 given, for $z > 0$, by

$$K_1(z) = \frac{z}{4}\int_0^{+\infty} t^{-2} e^{-t - z^2/4t} dt.$$

Proof. See Appendix A. ■

Relation (5) shows that for large values of n ($n \longrightarrow \infty$) and for all $\varepsilon > 0$, we have $T_n \leq (1+\varepsilon)\mathbb{E}(T_n)$ with probability 1, $T_n > (1-\varepsilon)\mathbb{E}(T_n)$ with probability 1. Moreover Theorem 3 shows that for large values of n ($n \longrightarrow \infty$), we have $T_n > \mathbb{E}(T_n)$ with probability 0.44843 and thus $T_n = \mathbb{E}(T_n)$ with probability 0.55157.

3. The continuous time case

As in the discrete time case, we suppose without any loss of generality that among the n nodes, a single one initially knows the rumour and a value 0 or 1 is associated with each node. A node with value 1 means that this node knows the rumor and a node with value 0 means that it is not aware of the rumor. For every $t \geq 0$, we denote by $C_t^{(i)}$ the value (0 or 1) of node i at time t. At time 0, all the $C_0^{(i)}$ are equal to 0 except one which is equal to 1 and which corresponds to the node initially knowing the rumor.

In the continuous time case, a Poisson process is associated with each node. These n Poisson processes are independent and have the same rate $\lambda > 0$. When the Poisson process associated with node i has a jump, this node chooses another node j randomly, with a given distribution to interact with node i. This is equivalent to consider a single Poisson process with rate $n\lambda$ at the jumps of which two distinct nodes are randomly chosen to interact with a given distribution. Then as in the discrete time case, the two nodes change their value with the maximum value of each node. Again, we want to evaluate the time needed to spread the rumor that is the time needed so that all the nodes get value 1.

We denote by $(\tau_\ell)_{\ell \geq 0}$ the successive jumps of the Poisson process with rate $n\lambda$, with $\tau_0 = 0$. Then once the couple (i, j) is chosen at time τ_ℓ, we have

$$C_t^{(i)} = C_t^{(j)} = \max\{C_{\tau_{\ell-1}}^{(i)}, C_{\tau_{\ell-1}}^{(j)}\} \text{ and } C_t^{(m)} = C_{\tau_{\ell-1}}^{(m)} \text{ for } m \neq i, j \text{ and } t \in [\tau_\ell, \tau_{\ell+1}).$$

For every $\ell \geq 1$, we denote by X_ℓ the random variable representing this choice at time τ_ℓ and we suppose that this choice is uniform, i.e. we suppose that, for all $\ell \geq 1$, we have

$$\mathbb{P}\{X_\ell = (i, j)\} = \frac{1}{n(n-1)} 1_{\{i \neq j\}}.$$

We consider the random variable Θ_n defined by

$$\Theta_n = \inf\{t \geq 0 \mid C_t^{(i)} = 1, \text{ for every } i = 1, \ldots, n\},$$

which represents the time needed for all the nodes in the network to know the rumor.

We introduce the continuous-time stochastic process $Z = \{Z_t,\ t \in \mathbb{R}^+\}$ with state space $\{1, \ldots, n\}$ defined, for all $t \geq 0$, by

$$Z_t = \left| \left\{ i \in \{1, \ldots, n\} \mid C_t^{(i)} = 1 \right\} \right|.$$

The random variable Z_t represents the number of nodes knowing the rumor at time t. The stochastic process Z is then a homogeneous Markov chain with transition rate matrix B. The non zero entries of matrix B are given, for $i=1,\ldots,n$, by

$$\begin{cases} B_{i,i} &= -n\lambda p_i, \\ B_{i,i+1} &= n\lambda p_i, \text{ for } i \neq n. \end{cases}$$

Indeed, when $Z_t = i$, the next node is activated with rate $n\lambda$. In order for process Z to reach state $i+1$ from state i, this activated node, say node ℓ, either possesses the rumor (probability i/n) and the node contacted by ℓ, say m, does not possess the rumor (probability $(n-i)/(n-1)$) or node ℓ does not possess the rumor (probability $(n-i)/n$) and it contacts node m which possesses the rumor (probability $i/(n-1)$). This means that, for $i=1,\ldots,n-1$, the rate $B_{i,i+1}$ is given by

$$B_{i,i+1} = n\lambda \frac{2i(n-i)}{n(n-1)} = n\lambda p_i.$$

The states $1,\ldots,n-1$ of Z are transient and state n is absorbing. The random variable Θ_n can then be written as

$$\Theta_n = \inf\{t \geq 0 \mid Z_t = n\}.$$

It is well-known, see for instance Sericola (2013), that the distribution of Θ_n is given, for every $t \geq 0$, by

$$\mathbb{P}\{\Theta_n > t\} = \alpha e^{Rt} \mathbb{1}, \tag{6}$$

where α is the row vector containing the initial probabilities of states $1,\ldots,n-1$, that is $\alpha_i = \mathbb{P}\{Z_0 = i\} = 1_{\{i=1\}}$, R is the sub-matrix obtained from B by deleting the row and the column corresponding to absorbing state n and $\mathbb{1}$ is the column vector of dimension $n-1$ with all its entries equal to 1. For every $i=1,\ldots,n-1$ we denote by U_i the sojourn time of process Z in state i, that is the time during which the system counts exactly i nodes knowing the rumor. The random variables U_i are independent and exponentially distributed with rate $\mu_i = n\lambda p_i$ and we have

$$\Theta_n = \sum_{i=1}^{n-1} U_i.$$

3.1. Expectation and variance of Θ_n

The expected value and the variance of Θ_n were obtained by Ganesh (2015) in the push model case. We extend these results to the push-pull model in the following two lemmas.

LEMMA 1. *For all $n \geq 2$, we have*

$$\mathbb{E}(\Theta_n) = \frac{(n-1)H_{n-1}}{n\lambda} \text{ and } \mathbb{E}(\Theta_n) \underset{n\longrightarrow\infty}{\sim} \frac{\ln(n)}{\lambda}.$$

Proof. We have

$$\mathbb{E}(\Theta_n) = \sum_{i=1}^{n-1} \mathbb{E}(U_i) = \frac{1}{n\lambda}\sum_{i=1}^{n-1}\frac{1}{p_i} = \frac{1}{n\lambda}\mathbb{E}(T_n) = \frac{(n-1)H_{n-1}}{n\lambda}.$$

The rest of the proof is evident since $H_{n-1} \underset{n\longrightarrow\infty}{\sim} \ln(n)$. ∎

LEMMA 2. *For all $n \geq 2$, we have*

$$\mathrm{Var}(\Theta_n) = \frac{(n-1)^2}{2n^2\lambda^2}\left(\sum_{i=1}^{n-1}\frac{1}{i^2} + \frac{2H_{n-1}}{n}\right) \leq \frac{1}{\lambda^2}\left(\frac{\pi^2}{12} + \frac{H_{n-1}}{n}\right) \text{ and } \lim_{n\longrightarrow\infty}\mathrm{Var}(\Theta_n) = \frac{\pi^2}{12\lambda^2}.$$

Proof. The random variables U_ℓ being independent, we have

$$\begin{aligned}
\mathrm{Var}(\Theta_n) &= \sum_{i=1}^{n-1}\mathrm{Var}(U_i) = \frac{1}{n^2\lambda^2}\sum_{i=1}^{n-1}\frac{1}{p_i^2} = \frac{(n-1)^2}{4\lambda^2}\sum_{i=1}^{n-1}\frac{1}{i^2(n-i)^2} \\
&= \frac{(n-1)^2}{4n^2\lambda^2}\sum_{i=1}^{n-1}\left(\frac{1}{i} + \frac{1}{n-i}\right)^2 = \frac{(n-1)^2}{4n^2\lambda^2}\left(2\sum_{i=1}^{n-1}\frac{1}{i^2} + \frac{4H_{n-1}}{n}\right) \\
&\leq \frac{1}{\lambda^2}\left(\frac{\pi^2}{12} + \frac{H_{n-1}}{n}\right).
\end{aligned}$$

The rest of the proof is evident since $H_{n-1} \underset{n\longrightarrow\infty}{\sim} \ln(n)$. ∎

3.2. Explicit expression of the distribution of Θ_n

The distribution of Θ_n, for $n \geq 2$, which is given by Relation (6) can be easily computed as follows. We make use of the uniformization technique, see for instance Sericola (2013). We introduce the uniformized Markov chain associated with the Markov chain Z which is characterized by its uniformization rate ν and by its transition probability matrix G. The uniformization rate ν must satisfy $\nu \geq \max_{i=1,\dots,n}(-B_{i,i})$ and matrix G is related to the infinitesimal generator R by

$$G = I_n + B/\nu,$$

where I_n denotes the identity matrix of order n. We denote by N_t the number of transitions occurring during the interval $[0,t]$. The process N_t is a Poisson process with rate ν and since $B = -\nu(I_n - G)$, we have $R = -\nu(I_{n-1} - P)$, where P is the sub-matrix obtained from G by deleting the row and the column corresponding to absorbing state n. Relation (6) can then be written as

$$\mathbb{P}\{\Theta_n > t\} = \alpha e^{Rt}\mathbb{1} = \sum_{k=0}^{\infty} e^{-\nu t}\frac{(\nu t)^k}{k!}\alpha P^k \mathbb{1}.$$

It is easily checked that

$$\max_{i=1,\dots,n}(-R_{i,i}) = \max_{i=1,\dots,n}(n\lambda p_i) \leq n\lambda.$$

By taking $\nu = n\lambda$, we get, from Relation (2), $P = Q$ and thus, using (1), this leads to

$$\mathbb{P}\{\Theta_n > t\} = \sum_{k=0}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\mathbb{P}\{T_n > k\} = \sum_{k=0}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\alpha Q^k \mathbb{1}. \tag{7}$$

Using this expression we obtain the following explicit expression of the distribution of Θ_n.

THEOREM 4. *For every $n \geq 1$, $t \geq 0$, we have*

$$\mathbb{P}\{\Theta_n > t\} = \sum_{j=1}^{\lfloor n/2 \rfloor} (c_{n-1,j} + n\lambda t d_{n-1,j})\, e^{-n\lambda p_j t},$$

where the coefficients $c_{n-1,j}$ and $d_{n-1,j}$ are given in Theorem 1.

Proof. From Theorem 1, we have for every $n \geq 1$ and $k \geq 0$,

$$\mathbb{P}\{T_n > k\} = \sum_{j=1}^{\lfloor n/2 \rfloor} (c_{n-1,j}(1-p_j) + k d_{n-1,j})\,(1-p_j)^{k-1},$$

where the coefficients $c_{n-1,j}$ and $d_{n-1,j}$ are given in Theorem 1. Using now Relation (7), we obtain

$$\begin{aligned}
\mathbb{P}\{\Theta_n > t\} &= \sum_{k=0}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\left(\sum_{j=1}^{\lfloor n/2 \rfloor} c_{n-1,j}(1-p_j)^k + \sum_{j=1}^{\lfloor n/2 \rfloor} k d_{n-1,j}(1-p_j)^{k-1}\right) \\
&= \sum_{j=1}^{\lfloor n/2 \rfloor} c_{n-1,j} e^{-n\lambda p_j t} + n\lambda t \sum_{j=1}^{\lfloor n/2 \rfloor} d_{n-1,j} e^{-n\lambda p_j t},
\end{aligned}$$

which completes the proof. ■

3.3. Bounds and tail behavior of the distribution of Θ_n

We obtain in this section a very simple bound of the complementary distribution of Θ_n and we show that this bound is also an equivalent of its tail. This bound and equivalent of the quantity $\mathbb{P}\{\Theta_n > t\}$ are derived from Theorem 2.

THEOREM 5. *For all $n \geq 3$ and $t \geq 0$ we have*

$$\mathbb{P}\{\Theta_n > t\} \leq \left[2(n-2)^2\lambda t + \frac{n}{n-2}\right] e^{-2\lambda t},$$

$$\mathbb{P}\{\Theta_n > t\} \underset{t \longrightarrow \infty}{\sim} \left[2(n-2)^2\lambda t + \frac{n}{n-2}\right] e^{-2\lambda t}.$$

Note that for $n=2$, we have $\Theta_2 = U_1$ which is exponentially distributed with rate $\mu_1 = 2\lambda$ and thus $\mathbb{P}\{\Theta_2 > t\} = e^{-2\lambda t}$.

Proof. From Theorem 2, we have for $n \geq 2$ and $k \geq 1$,

$$\mathbb{P}\{T_n > k\} \leq \left(1 + \frac{2k(n-2)^2}{n}\right)\left(1 - \frac{2}{n}\right)^{k-1}.$$

Since $\mathbb{P}\{T_n > 0\} = 1$, this leads to

$$\begin{aligned}
\mathbb{P}\{\Theta_n > t\} &= \sum_{k=0}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\mathbb{P}\{T_n > k\} \\
&\leq e^{-n\lambda t} + \sum_{k=1}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\left(1 + \frac{2k(n-2)^2}{n}\right)\left(1 - \frac{2}{n}\right)^{k-1} \\
&= e^{-n\lambda t} + \sum_{k=1}^{\infty} e^{-n\lambda t}\frac{(n\lambda t)^k}{k!}\left(1 - \frac{2}{n}\right)^{k-1} + 2(n-2)^2\lambda t \sum_{k=1}^{\infty} e^{-n\lambda t}\frac{((n-2)\lambda t)^{k-1}}{(k-1)!} \\
&= e^{-n\lambda t} + \frac{ne^{-n\lambda t}\left(e^{(n-2)\lambda t} - 1\right)}{n-2} + 2(n-2)^2\lambda t e^{-n\lambda t}e^{(n-2)\lambda t} \\
&= \left[2(n-2)^2\lambda t + \frac{n}{n-2}\right] e^{-2\lambda t} - \frac{2}{n-2}e^{-n\lambda t} \\
&\leq \left[2(n-2)^2\lambda t + \frac{n}{n-2}\right] e^{-2\lambda t}.
\end{aligned}$$

which completes the first part of the proof.

On the one hand since $p_1 < p_j$ for $j = 2, \ldots, \lfloor n/2 \rfloor$, we have, from Theorem 1,

$$\mathbb{P}\{T_n > k\} \underset{k \longrightarrow \infty}{\sim} d_{n-1,1}k\left(1 - \frac{2}{n}\right)^{k-1}.$$

On the other hand, from Theorem 2, we have

$$\mathbb{P}\{T_n > k\} \underset{k \longrightarrow \infty}{\sim} \left(1 + \frac{2k(n-2)^2}{n}\right)\left(1 - \frac{2}{n}\right)^{k-1}.$$

These two results imply that

$$d_{n-1,1} = \frac{2(n-2)^2}{n}.$$

In the same way, from Theorem 4, we get

$$\mathbb{P}\{\Theta_n > t\} \underset{t \longrightarrow \infty}{\sim} d_{n-1,1} n\lambda t e^{-n\lambda p_1 t} = 2(n-2)^2 \lambda t e^{-2\lambda t} \underset{t \longrightarrow \infty}{\sim} \left[2(n-2)^2\lambda t + \frac{n}{n-2}\right] e^{-2\lambda t},$$

which completes the proof. ■

We give in the following two different bounds for the quantity $\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\}$, with $c \geq 1$. These bounds will be compared and used to obtain the limiting behaviour of this quantity when the number n of nodes goes to infinity.

Recalling that $\mathbb{E}(\Theta_n) = (n-1)H_{n-1}/(n\lambda)$, a first bound is obtained by an immediate application of Theorem 5.1 of Janson (2014), which leads, for all $n \geq 3$ and for all real number $c \geq 1$, to

$$\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} \leq \frac{1}{c} \exp\left(-\frac{2(n-1)H_{n-1}(c-1-\ln(c))}{n}\right). \tag{8}$$

Note that the right-hand side term is equal to 1 when $c = 1$.

Applying Theorem 5 at point $c\mathbb{E}(\Theta_n)$, we obtain the following second bound.

$$\begin{aligned}
\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} &\leq \left[2(n-2)^2\lambda c\mathbb{E}(\Theta_n) + \frac{n}{n-2}\right] e^{-2\lambda c\mathbb{E}(\Theta_n)} \\
&= \left[\frac{2c(n-2)^2(n-1)H_{n-1}}{n} + \frac{n}{n-2}\right] \exp\left(-\frac{2c(n-1)H_{n-1}}{n}\right).
\end{aligned}$$

From now on we denote this bound by $\varphi(c,n)$ and in the same way, we denote by $\psi(c,n)$ the bound of $\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\}$ obtained in (8). We then have, for $n \geq 3$ and $c \geq 1$,

$$\begin{aligned}
\varphi(c,n) &= \left[\frac{2c(n-2)^2(n-1)H_{n-1}}{n} + \frac{n}{n-2}\right] \exp\left(-\frac{2c(n-1)H_{n-1}}{n}\right), \\
\psi(c,n) &= \frac{1}{c} \exp\left(-\frac{2(n-1)H_{n-1}(c-1-\ln(c))}{n}\right).
\end{aligned}$$

These two bounds are compared in the next theorem.

THEOREM 6. *For every $n \geq 5$, there exists a unique $c^* \geq 1$ such that $\varphi(c^*,n) = \psi(c^*,n)$ and we have*

$$\begin{cases} \varphi(c,n) > \psi(c,n) \text{ for all } 1 \leq c < c^* \\ \varphi(c,n) < \psi(c,n) \text{ for all } c > c^*. \end{cases} \tag{9}$$

Furthermore,

$$\lim_{c \longrightarrow \infty} \frac{\varphi(c,n)}{\psi(c,n)} = 0.$$

Proof. Let us introduce the quantities

$$A_n = \frac{(n-1)H_{n-1}}{n}, B_n = 2(n-2)^2 A_n \text{ and } C_n = \frac{n}{n-2}.$$

We then have

$$\frac{\varphi(c,n)}{\psi(c,n)} = \left(B_n c^2 + C_n c\right) e^{-2A_n(1+\ln(c))} = \left(B_n c^{2-2A_n} + C_n c^{1-2A_n}\right) e^{-2A_n}.$$

It is easily checked that the sequence A_n is strictly increasing and that $A_3 = 1$. It follows that for $n \geq 5$, we have $A_n > 1$ and so

$$1 - 2A_n < 2 - 2A_n < 0.$$

This implies that for every $n \geq 5$, the function $\varphi(c,n)/\psi(c,n)$ is strictly decreasing with c on $[1, +\infty)$ and that

$$\lim_{c \longrightarrow \infty} \frac{\varphi(c,n)}{\psi(c,n)} = 0.$$

Consider now the sequences x_n and y_n defined for $n \geq 5$, by

$$x_n = \frac{\varphi(1,n)}{\psi(1,n)} = (B_n + C_n)\, e^{-2A_n} \text{ and } y_n = \frac{2e^{-2}(n-2)^2 A_n}{(n-1)^2}.$$

The sequence A_n being increasing, it is easily checked that sequence y_n is increasing too. Moreover, we have

$$x_n \geq B_n e^{-2(1+\ln(n-1))} = \frac{e^{-2} B_n}{(n-1)^2} = \frac{2e^{-2}(n-2)^2 A_n}{(n-1)^2} = y_n.$$

A simple computation shows that we have $y_{34} > 1$. The sequence y_n being increasing, we obtain $y_n > 1$ for every $n \geq 34$. It follows that we also have $x_n > 1$ for all $n \geq 34$. A numerical computation gives $x_n > 1$ for $n = 5, \ldots, 33$ which means that for all $n \geq 5$, we have $x_n = \varphi(1,n)/\psi(1,n) > 1$. The function $\varphi(c,n)/\psi(c,n)$ being strictly decreasing with c on $[1, +\infty)$, we deduce that there exists a unique solution, called c^*, to the equation $\varphi(c,n)/\psi(c,n) = 1$ and (9) follows. ∎

This theorem shows that our bound $\varphi(c,n)$ is much more tight than the one obtained using the result of Janson (2014), which has been denoted by $\psi(c,n)$, for $c > c^*$, not only because the ratio $\varphi(c,n)/\psi(c,n)$ decreases with c and tends to 0 when c tends to infinity,

Table 1 Values of c^* for different network sizes n.

n	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9
c^*	1.253	1.163	1.128	1.109	1.095	1.085	1.078	1.071	1.066

but also because for every value of n, the value of c^* is very close to 1 as shown in Table 1. Moreover, from Theorem 5, our bound is optimal in the sense that

$$\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} \underset{c \longrightarrow \infty}{\sim} \varphi(c, n).$$

Table 2 and Figure 1 illustrate, for a network composed of $n = 1000$ nodes, the behavior of the bounds $\varphi(c, 1000)$ and $\psi(c, 1000)$, as a function of c, compared to the exact value of complementary distribution function of Θ_{1000} at point $c\mathbb{E}(\Theta_{1000})$, computed using Theorem 4. Table 2 illustrates clearly the result of Theorem 5. Indeed the values of our bound $\varphi(c, 1000)$ are very close to the real value of the complementary distribution function, while the values of $\psi(c, 1000)$ tend to move away from this real value even for small values of c. Note that when $c = 1$ both bounds are useless and the real value $\mathbb{P}\{\Theta_{1000} > \mathbb{E}(\Theta_{1000})\}$ is very close to the limit obtained in Theorem 9 of next section. Figure 1 shows the large gap between the bounds $\varphi(c, 1000)$ and $\psi(c, 1000)$ when c is greater than c^* whose value is $c^* = 1.12819634$. Moreover this large gap increases when n increases since the value of c^* decreases to 1 when n increases, as shown in Table 1.

Table 2 Values of $\mathbb{P}\{\Theta_{1000} > c\mathbb{E}(\Theta_{1000})\}$, $\varphi(c, 1000)$ and $\psi(c, 1000)$ for different values of c.

c	1	1.2	1.4	1.6	1.8	2
$\mathbb{P}\{\Theta_{1000} > c\mathbb{E}(\Theta_{1000})\}$	0.446	0.063	0.005	3.9×10^{-4}	2.6×10^{-5}	1.6×10^{-6}
$\varphi(c, 1000)$	≥ 1	0.288	0.017	9.7×10^{-4}	5.5×10^{-5}	3×10^{-6}
$\psi(c, 1000)$	1.0	0.634	0.276	0.089	0.023	0.005

3.4. Asymptotic analysis of the distribution of Θ_n

We analyze in this section the behavior of the complementary distribution of Θ_n at point $c\mathbb{E}(\Theta_n)$ when the number n of nodes in the network tends to infinity, in function of the value of c.

We prove in the following theorem that the bounds $\varphi(c, n)$ and $\psi(c, n)$, obtained from Theorem 5 and Relation (8) respectively with $t = c\mathbb{E}(T_n)$, both tend to 0 when n goes to infinity.

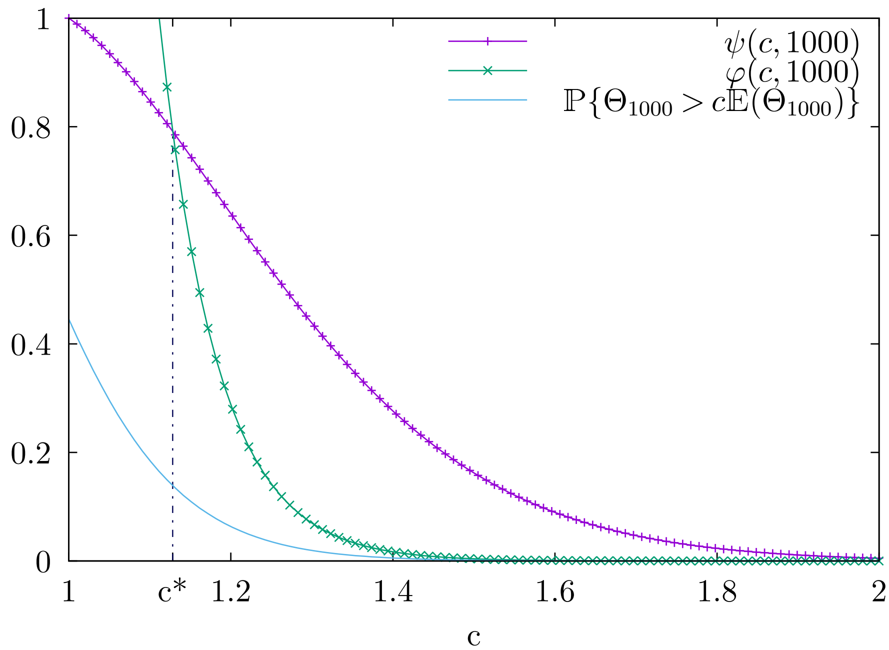

Figure 1 **Bounds $\psi(c,1000)$, $\varphi(c,1000)$ and real value of $\mathbb{P}\{\Theta_{1000} > c\mathbb{E}(\Theta_{1000}\}$ as a function of c. The point at which the bounds are equal is $c^* = 1.12819634$.**

THEOREM 7. *For all real number $c > 1$, we have*

$$\lim_{n \longrightarrow \infty} \varphi(c,n) = 0 \text{ and } \lim_{n \longrightarrow \infty} \psi(c,n) = 0.$$

Proof. It is easily checked that

$$\varphi(c,n) \underset{n \longrightarrow \infty}{\sim} \frac{2cn^2 \ln(n)}{n^{2c}}$$

which tends to 0 when n tends to infinity. Concerning $\psi(c,n)$ we have

$$\psi(c,n) \underset{n \longrightarrow \infty}{\sim} \frac{1}{c} e^{-\ln(n)(c-1-\ln(c))}.$$

For $c > 1$ we have $c - 1 - \ln(c) > 0$ which implies that $\psi(c,n)$ tends to 0 when n tends to infinity. ∎

THEOREM 8. *For all real $c \geq 0$, we have*

$$\lim_{n \to \infty} \mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} = \begin{cases} 0 & \text{if } c > 1 \\ 1 & \text{if } c < 1. \end{cases}$$

Proof. From Theorem 7, both bounds $\varphi(c,n)$ and $\psi(c,n)$ of $\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\}$ tend to 0 when n tends to infinity, for $c > 1$. So using either $\varphi(c,n)$ or $\psi(c,n)$ we deduce that

$$\lim_{n \longrightarrow \infty} \mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} = 0 \text{ for all } c > 1.$$

In the case where $c < 1$, Theorem 5.1 of Janson (2014) leads to

$$\mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} \geq 1 - \exp\left(\frac{-2(n-1)H_{n-1}(c-1-\ln(c))}{n}\right).$$

Since $c - 1 - \ln(c) > 0$ for all $c \in [0,1)$, the right-hand side term of this inequality tends to 1 when $n \longrightarrow \infty$. Thus, $\lim_{n\longrightarrow\infty} \mathbb{P}\{\Theta_n > c\mathbb{E}(\Theta_n)\} = 1$ when $c < 1$. ■

The following theorem considers the case $c = 1$. Note that the result is identical to the one of Theorem 3 in the discrete time case.

THEOREM 9.

$$\lim_{n \longrightarrow \infty} \mathbb{P}\{\Theta_n > \mathbb{E}(\Theta_n)\} = 1 - 2e^{-\gamma} K_1\left(2e^{-\gamma}\right) \approx 0.448429663727.$$

where γ is the Euler's constant given by $\gamma = \lim_{n\longrightarrow\infty}(H_n - \ln(n)) \approx 0.5772156649$ and K_1 is the modified Bessel function of the second kind of order 1 given, for $z > 0$, by

$$K_1(z) = \frac{z}{4}\int_0^{+\infty} t^{-2} e^{-t-z^2/4t} dt.$$

Proof. See Appendix B. ■

Remark. Some possible extensions of this work are the following.

1. We have supposed that the initial number of nodes knowing the rumor is equal to 1. The case where this number is equal to ℓ, with $\ell \geq 2$, has been dealt with in Mocquard et al. (2016) in the discrete time case. This extension to the continuous time case is almost straightforward since it suffices to redefine the random variable Θ_n as $\Theta_n = U_\ell + \cdots U_n$.

2. Instead of considering the total time needed for all the nodes to obtain the rumor, one could be interested in the total time needed for a fixed percentage, say ρ, of the nodes to obtain the rumor. In that case the random variable Θ_n to consider should be redefined as $\Theta_n = U_1 + \cdots U_{\lceil \rho n \rceil}$. Of course this extension could also be combined with the first one above.

3. The instants at which the interactions between nodes occur have been modeled by a Poisson process. This could be generalized by considering, instead of a Poisson process, a Phase-type renewal process which preserves the Markov property and can approximate every point process.

Acknowledgement. We would like to thank Professor Philippe Carmona for his expert advice concerning the proof of Theorem 3.

4. Conclusion

In this paper we have provided a thorough analysis of the rumor spreading time in the asynchronous push-pull model in the continuous time case by completing and extending the results already obtained in the discrete time case. Such a precise analysis is a step towards the design of more complex such as, for instance, the leader election in large distributed systems. Our analysis concerning the tail distribution of the rumor spreading time and its limiting behavior when the number of nodes goes to infinity has never been done in such detail before. It shows that the evaluation of the first moment of the rumor spreading time is far from sufficient to provide a global control of the system.

References

Acan H, Collevecchio A, Mehrabian A, Nick W (2015) On the push&pull protocol for rumour spreading. *Proceedings of the ACM Symposium on Principles of Distributed Systems (PODC)*.

Angluin D, Aspnes J, Eisenstat D (2008) Fast computation by population protocols with a leader. *Distributed Computing* 21(2):183–199.

Berger N, Borgs C, Chayes JT, Saberi A (2005) On the spread of viruses on the internet. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*.

Censor-Hillel K, Haeupler B, Kelner J, Maymounkov P (2012) Global computation in a poorly connected world: Fast rumor spreading with no dependence on conductance. *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*.

Chierichetti F, Lattanzi S, Panconesi A (2011) Rumor spreading in social networks. *Theoretical Computer Science* 412(24):2602–2610.

Clementi A, Crescenzi P, Doerr C, Fraigniaud P, Pasquale F, Silvestri R (2015) Rumor spreading in random evolving graphs. *Random structures and Algorithms* 48(2):290–312.

Daum S, Kuhn F, Maus Y (2016) Rumor spreading with bounded indegree. *Proceedings of the International Colloquium on Structural Information and Communication Complexity (SIROCCO)*.

Demers A, Gealy M, Greene D, Hauser C, Irish W, Larson J, Shenker S, Sturgis H, Swinehart D, Terry D (1987) Epidemic algorithms for replicated datbase maintenance. *Proceedings of the ACM Syposium on Principles of Distributed Systems (PODC).*

Feige U, Peleg D, Raghavan P, Upfal E (1990) Randomized broadcast in networks. *Random Structures and Algorithms* 1(4):447–460.

Fountoulakis N, Panagiotou K (2013) Rumor spreading on random regular graphs and expanders. *Random Structures and Algorithms* 43(2):201–220.

Frieze A, Grimmet G (1985) The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics* 10(1):57–77.

Ganesh AJ (2015) Rumour spreading on graphs. Technical report, URL `https://people.maths.bris.ac.uk/~maajg/teaching/complexnets/rumours.pdf`.

Giakkoupis G (2011) Tight bounds for rumor spreading in graphs of a given conductance. *Proceedings of the International Symposium on Theoretical Aspects of Computer Science (STACS).*

Giakkoupis G (2014) Tight bounds for rumor spreading with vertex expansion. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).*

Gordon L (1989) Bounds for the distribution of the generalized variance. *The Annals of Statistics* 17(4):1684–1692.

Gradshteyn IS, Ryzhik IM (2014) *Table of Integrals, Series, and Products* (Boston, MA, USA: Academic Press), eighth edition.

Harchol-Balter M, Leighton T, Lewin D (1999) Resource discovery in distributed networks. *Proceedings of the ACM Syposium on Principles of Distributed Systems (PODC).*

Janson S (2014) Tail bounds for sums of geometric and exponential variables. Technical report, URL `http://www2.math.uu.se/~svante/papers/sjN14.pdf?`

Karp R, Schindelhauer C, Shenker S, Vocking B (2000) Randomized rumor spreading. *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS).*

Kempe D, Dobra A, Gehrke J (2003) Gossip-based computation of aggregate information. *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS).*

Mocquard Y, Robert S, Sericola B, Anceaume E (2016) Analysis of the propagation time of a rumour in large-scale distributed systems. *Proceedings of the 15th IEEE International Symposium on Network Computing and Applications (NCA)*, 35–42.

Panagiotou K, Perez-Gimenez X, Sauerwald T, Sun H (2015) Randomized rumor spreading: the effect of the network topology. *Combinatorics, Probability and Computing* 24(2):457–479.

Panagiotou K, Speidel L (2016) Asynchronous rumor spreading on random graphs. *Algorithmica* .

Pittel B (1987) On spreading a rumor. *SIAM Journal on Applied Mathematics* 47(1):213–223.

Sericola B (2013) *Markov Chains. Theory, Algorithms and Applications.* Applied stochastic methods series (Wiley).

Stroock DW (2010) *Probability Theory: An Analytic View* (New York, NY, USA: Cambridge University Press), second edition.

Appendix A: Proofs in the discrete time case

We give in this appendix the proof of Theorem 3 which concerns the discrete time case.

In order to prove Theorem 3, we first need a technical Lemma. The random variables S_i are independent and geometrically distributed with parameter $p_i = 2i(n-i)/(n(n-1))$, but since the p_i depend on n, we rename the random variables S_i as $S_{n,i}$ and the parameters p_i as $p_{n,i}$. The spreading time T_n thus writes as $T_n = S_{n,1} + \cdots + S_{n,n-1}$.

We use the notation $X_n \xrightarrow{\mathcal{L}} X$ to express that the sequence of random variables (X_n) converges in distribution (or in law) towards the random variable X when n tends to infinity.

Lemma 3. *Let $(Z_i)_{i\geq 1}$ be a sequence of i.i.d. random variables exponentially distributed with rate 1 and let W be defined by*

$$W = \sum_{i=1}^{\infty} \frac{Z_i - 1}{2i}.$$

We then have

$$\frac{T_n - \mathbb{E}(T_n)}{n} \xrightarrow{\mathcal{L}} W^{(1)} + W^{(2)}$$

where W^1 and W^2 are i.i.d. with the same distribution as W.

Proof. For each fixed i, we have $\lim_{n\longrightarrow\infty} p_{n,i} = 0$. It follows that for every $x \geq 0$, we have

$$\mathbb{P}\{p_{n,i}S_{n,i} > x\} = \mathbb{P}\{S_{n,i} > x/p_{n,i}\} = (1-p_{n,i})^{\lfloor x/p_{n,i} \rfloor},$$

which tends to e^{-x} when n tends to infinity, since $p_{n,i}$ tends to 0. If Z_i is a random variable exponentially distributed with rate 1, we have shown that $p_{n,i}S_{n,i} \xrightarrow{\mathcal{L}} Z_i$. Moreover since the $(S_{n,i})_{i=1,\dots,n-1}$ are independent, the $(Z_i)_{i\geq 1}$ are also independent.

Observing now that for each fixed i, we have $\lim_{n\longrightarrow\infty} np_{n,i} = 2i$ and defining $R_{n,i} = S_{n,i} - \mathbb{E}(S_{n,i})$ we obtain, since $\mathbb{E}(S_{n,i}) = 1/p_{n,i}$,

$$\frac{R_{n,i}}{n} = \frac{S_{n,i} - \mathbb{E}(S_{n,i})}{n} = \frac{p_{n,i}S_{n,i} - 1}{np_{n,i}} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{2i}. \tag{10}$$

Suppose that $n = 2k+1$. We then have

$$T_{2k+1} - \mathbb{E}(T_{2k+1}) = \frac{1}{2k+1}\left(\sum_{i=1}^{k} R_{2k+1,i} + \sum_{i=1}^{k} R_{2k+1,2k+1-i}\right) = V_k + \overline{V}_k, \tag{11}$$

where

$$V_k = \frac{1}{2k+1}\sum_{i=1}^{k} R_{2k+1,i} \text{ and } \overline{V}_k = \frac{1}{2k+1}\sum_{i=1}^{k} R_{2k+1,2k+1-i}.$$

The random variables V_k and $\overline{V}_k$ are independent and they also have the same distribution. Indeed, since $p_{n,i} = p_{n,n-i}$ the variables $R_{n,i}$ and $R_{n,n-i}$ have the same distribution.

The rest of the proof consists in checking the hypothesis of the principle of accompanying laws of Theorem 3.1.14 of Stroock (2010). We introduce the notation

$$W_{m,k} = \frac{1}{2k+1} \sum_{i=1}^{m-1} R_{2k+1,i}.$$

Using the fact that $\mathbb{E}(R_{n,i}) = 0$ and that the $R_{n,i}$ are independent, we have

$$\begin{aligned}
\mathbb{E}((V_k - W_{m,k})^2) &= \mathbb{E}\left(\left[\frac{1}{2k+1}\sum_{i=m}^{k} R_{2k+1,i}\right]^2\right) = \mathrm{Var}\left(\frac{1}{2k+1}\sum_{i=m}^{k} R_{2k+1,i}\right) \\
&= \frac{1}{(2k+1)^2}\sum_{i=m}^{k} \mathrm{Var}(R_{2k+1,i}) = \frac{1}{(2k+1)^2}\sum_{i=m}^{k} \mathrm{Var}(S_{2k+1,i}) \\
&= \frac{1}{(2k+1)^2}\sum_{i=m}^{k} \frac{1-p_{2k+1,i}}{p_{2k+1,i}^2} \leq \frac{1}{(2k+1)^2}\sum_{i=m}^{k} \frac{1}{p_{2k+1,i}^2}.
\end{aligned}$$

Recalling that $p_{2k+1,i} = 2i(2k+1-i)/(2k(2k+1))$, we obtain

$$\mathbb{E}((V_k - W_{m,k})^2) \leq k^2 \sum_{i=m}^{k} \frac{1}{i^2(2k+1-i)^2}.$$

In this sum we have $2k+1-i \geq k$. This leads to

$$\mathbb{E}((V_k - W_{m,k})^2) \leq \sum_{i=m}^{k} \frac{1}{i^2}.$$

We then have

$$\lim_{m\longrightarrow\infty} \limsup_{k\longrightarrow\infty} \mathbb{E}((V_k - W_{m,k})^2) \leq \lim_{m\longrightarrow\infty} \sum_{i=m}^{\infty} \frac{1}{i^2} = 0.$$

Using now the Markov inequality, we obtain, for all $\varepsilon > 0$,

$$\mathbb{P}\{|V_k - W_{m,k}| \geq \varepsilon\} = \mathbb{P}\{(V_k - W_{m,k})^2 \geq \varepsilon^2\} \leq \frac{\mathbb{E}((V_k - W_{m,k})^2)}{\varepsilon^2}.$$

Putting together these results, we have shown that for all $\varepsilon > 0$, we have

$$\lim_{m\longrightarrow\infty} \limsup_{k\longrightarrow\infty} \mathbb{P}\{|V_k - W_{m,k}| \geq \varepsilon\} = 0 \tag{12}$$

Let us introduce the notation

$$W_m = \sum_{i=1}^{m-1} \frac{Z_i - 1}{2i}.$$

Using (10) and the fact that the $R_{n,i}$ are independent, we have

$$W_{m,k} \xrightarrow{\mathcal{L}} W_m \text{ as } k \longrightarrow \infty. \tag{13}$$

The hypothesis of the principle of accompanying laws of Theorem 3.1.14 of Stroock (2010) are properties (10) and (13). We can thus conclude that

$$V_k \xrightarrow{\mathcal{L}} W \text{ as } k \longrightarrow \infty.$$

Similarly, we have

$$\overline{V}_k \xrightarrow{\mathcal{L}} W \text{ as } k \longrightarrow \infty.$$

This means, from Relation (11), that

$$\frac{T_{2k+1}-\mathbb{E}(T_{2k+1})}{2k+1} \xrightarrow{\mathcal{L}} W^{(1)}+W^{(2)},$$

where $W^{(1)}$ and $W^{(2)}$ are i.i.d. and distributed as W.

The same reasoning applies in the case where $n=2k$. ■

We are now ready to prove Theorem 3.

Theorem 3

$$\lim_{n\longrightarrow\infty}\mathbb{P}\{T_n>\mathbb{E}(T_n)\}=1-2e^{-\gamma}K_1\left(2e^{-\gamma}\right)\approx 0.448429663727.$$

where γ is the Euler-Mascheroni constant given by $\gamma=\lim_{n\longrightarrow\infty}(H_n-\ln(n))\approx 0.5772156649$ and K_1 is the modified Bessel function of the second kind of order 1 given, for $z>0$, by

$$K_1(z)=\frac{z}{4}\int_0^{+\infty}t^{-2}e^{-t-z^2/4t}dt.$$

Proof. Louis Gordon has proved in Gordon (1989) that

$$-\gamma+\sum_{i=1}^{+\infty}\frac{1-Z_i}{i}\overset{\mathcal{L}}{=}\ln Z_1,$$

where (Z_i) are i.i.d. exponential with rate 1 and γ is the Euler-Mascheroni constant. Thus, by definition of W in Lemma 3, we have

$$W\overset{\mathcal{L}}{=}-\frac{\gamma+\ln Z_1}{2}.$$

Introducing $W^{(1)}\overset{\mathcal{L}}{=}-(\gamma+\ln Z_1)/2$ and $W^{(2)}\overset{\mathcal{L}}{=}-(\gamma+\ln Z_2)/2$, we obtain from Lemma 3,

$$\begin{aligned}
\lim_{n\longrightarrow\infty}\mathbb{P}\left\{T_n>\mathbb{E}(T_n)\right\} &= \mathbb{P}\left\{W^{(1)}+W^{(2)}>0\right\}\\
&= \mathbb{P}\left\{-2\gamma-\ln\left(Z_1Z_2\right)>0\right\}\\
&= \mathbb{P}\left\{Z_1Z_2<e^{-2\gamma}\right\}\\
&= \int_0^{\infty}\left(1-\exp(-e^{-2\gamma}/t)\right)e^{-t}dt\\
&= 1-\int_0^{\infty}\exp(-t-e^{-2\gamma}/t)dt.
\end{aligned}$$

Let u be the function defined on $(0,+\infty)$ by $u(t)=\exp(-t-e^{-2\gamma}/t)$. We easily get

$$\lim_{t\longrightarrow 0^+}u(t)=0 \text{ and } \lim_{t\longrightarrow\infty}u(t)=0,$$

which implies that

$$\int_0^{\infty}u'(t)dt=0. \tag{14}$$

The derivative u' of u is given by

$$\begin{aligned}
u'(t) &= \left(-1+e^{-2\gamma}t^{-2}\right)u(t)\\
&= -u(t)+e^{-2\gamma}u(t)t^{-2}
\end{aligned} \tag{15}$$

Integrating (15) over $(0,+\infty)$ and using (14), we obtain

$$\int_0^{\infty}u(t)dt=e^{-2\gamma}\int_0^{\infty}u(t)t^{-2}dx.$$

By definition of function u, this leads to

$$\lim_{n\longrightarrow\infty} \mathbb{P}\{T_n > \mathbb{E}(T_n)\} = 1 - e^{-2\gamma}\int_0^\infty t^{-2}\exp(-t - e^{-2\gamma}/t)dx$$
$$= 1 - 2e^{-\gamma}K_1\left(2e^{-\gamma}\right) \approx 0.448429663727,$$

where K_1 is the well-known modified Bessel function of the second kind of order 1, see for instance expression 8.432.6 Gradshteyn and Ryzhik (2014). ■

This theorem is as expected similar to the one obtained in the discrete time case.

Appendix B: Proofs in the continuous time case

We give in this appendix the proof of Theorem 9 which concerns the continuous time case.

In order to prove Theorem 9, we first need, as in the discrete time case, the following technical Lemma. The random variables U_i are independent and exponentially distributed with rate $\mu_i = 2\lambda i(n-i)/(n-1)$, but since the μ_i depend on n, we rename the random variables U_i as $U_{n,i}$ and the parameters μ_i as $\mu_{n,i}$. The spreading time Θ_n thus writes as $\Theta_n = U_{n,1} + \cdots + U_{n,n-1}$.

Lemma 4. *Let $(Z_i)_{i\geq 1}$ be a sequence of i.i.d. random variables exponentially distributed with rate 1 and let W be defined by*

$$W = \sum_{i=1}^{\infty} \frac{Z_i - 1}{2\lambda i}.$$

We then have

$$\Theta_n - \mathbb{E}(\Theta_n) \xrightarrow{\mathcal{L}} W^{(1)} + W^{(2)}$$

where $W^{(1)}$ and $W^{(2)}$ are i.i.d. with the same distribution as W.

Proof. For all $n \geq 2$, $i = 1, \ldots, n-1$ and $x \geq 0$, we have

$$\mathbb{P}\{\mu_{n,i}U_{n,i} > x\} = \mathbb{P}\{U_{n,i} > x/\mu_{n,i}\} = e^{-x}.$$

Thus if Z_i is a random variable exponentially distributed with rate 1, we have $\mu_{n,i}U_{n,i} \stackrel{\mathcal{L}}{=} Z_i$. Moreover since the $(U_{n,i})_{i=1,\ldots,n-1}$ are independent, the $(Z_i)_{i\geq 1}$ are also independent.

Observing now that for each fixed i, we have $\lim_{n\longrightarrow\infty} \mu_{n,i} = 2\lambda i$ and defining $R_{n,i} = U_{n,i} - \mathbb{E}(U_{n,i})$ we obtain, since $\mathbb{E}(U_{n,i}) = 1/\mu_{n,i}$,

$$R_{n,i} = U_{n,i} - \mathbb{E}(U_{n,i}) = \frac{\mu_{n,i}S_{n,i} - 1}{\mu_{n,i}} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{2\lambda i}. \tag{16}$$

Suppose that $n = 2k+1$. Defining

$$V_k = \sum_{i=1}^{k} R_{2k+1,i} \text{ and } \overline{V}_k = \sum_{i=1}^{k} R_{2k+1,2k+1-i},$$

we have

$$\Theta_{2k+1} - \mathbb{E}(\Theta_{2k+1}) = V_k + \overline{V}_k. \tag{17}$$

The random variables V_k and $\overline{V}_k$ are independent and they also have the same distribution. Indeed, since $\mu_{n,i} = \mu_{n,n-i}$ the variables $R_{n,i}$ and $R_{n,n-i}$ have the same distribution.

As in the discrete time case, the rest of the proof consists in checking the hypothesis of the principle of accompanying laws of Theorem 3.1.14 of Stroock (2010). We introduce the notation

$$W_{m,k} = \sum_{i=1}^{m-1} R_{2k+1,i}.$$

Using the fact that $\mathbb{E}(R_{n,i}) = 0$ and that the $R_{n,i}$ are independent, we have

$$\begin{aligned}\mathbb{E}((V_k - W_{m,k})^2) &= \mathbb{E}\left(\left[\sum_{i=m}^{k} R_{2k+1,i}\right]^2\right) = \mathrm{Var}\left(\sum_{i=m}^{k} R_{2k+1,i}\right) \\ &= \sum_{i=m}^{k} \mathrm{Var}(R_{2k+1,i}) = \sum_{i=m}^{k} \mathrm{Var}(U_{2k+1,i}) = \sum_{i=m}^{k} \frac{1}{\mu_{2k+1,i}^2}.\end{aligned}$$

Recalling that $\mu_{2k+1,i} = 2\lambda i(2k+1-i)/(2k)$, we obtain

$$\mathbb{E}((V_k - W_{m,k})^2) = \frac{k^2}{\lambda^2} \sum_{i=m}^{k} \frac{1}{i^2(2k+1-i)^2}.$$

In this sum we have $2k+1-i \geq k$. This leads to

$$\mathbb{E}((V_k - W_{m,k})^2) \leq \frac{1}{\lambda^2} \sum_{i=m}^{k} \frac{1}{i^2}.$$

We then have

$$\lim_{m \longrightarrow \infty} \limsup_{k \longrightarrow \infty} \mathbb{E}((V_k - W_{m,k})^2) \leq \frac{1}{\lambda^2} \lim_{m \longrightarrow \infty} \sum_{i=m}^{\infty} \frac{1}{i^2} = 0.$$

Introducing the random variable

$$W_m = \sum_{i=1}^{m-1} \frac{Z_i - 1}{2\lambda i},$$

the rest of the proof is exactly as in the discrete time case. ■

We are now ready to prove Theorem 9.

Theorem 9

$$\lim_{n \longrightarrow \infty} \mathbb{P}\{\Theta_n > \mathbb{E}(T_n)\} = 1 - 2e^{-\gamma} K_1\left(2e^{-\gamma}\right) \approx 0.448429663727.$$

Proof. Louis Gordon has proved in Gordon (1989) that

$$-\gamma + \sum_{i=1}^{+\infty} \frac{1 - Z_i}{i} \overset{\mathcal{L}}{=} \ln Z_1,$$

where (Z_i) are i.i.d. exponential with rate 1 and γ is the Euler-Mascheroni constant. Thus, by definition of W in Lemma 4, we have

$$W \overset{\mathcal{L}}{=} -\frac{\gamma + \ln Z_1}{2\lambda}.$$

Introducing $W^{(1)} \overset{\mathcal{L}}{=} -(\gamma + \ln Z_1)/2\lambda$ and $W^{(2)} \overset{\mathcal{L}}{=} -(\gamma + \ln Z_2)/2\lambda$, we get from Lemma 4,

$$\lim_{n \longrightarrow \infty} \mathbb{P}\left\{T_n > \mathbb{E}(T_n)\right\} = \mathbb{P}\left\{W^{(1)} + W^{(2)} > 0\right\} = \mathbb{P}\left\{-2\gamma - \ln\left(Z_1 Z_2\right) > 0\right\}.$$

The rest of the proof is similar to that of Theorem 3. ■