Online Supplement for
Probabilistic Analysis of Rumor Spreading Time

Yves Mocquard[*],
Université de Rennes 1 / IRISA,
yves.mocquard@irisa.fr

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

Emmanuelle Anceaume
CNRS/IRISA,
emmanuelle.anceaume@irisa.fr

July 2, 2018

Abstract

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.

Keywords Rumor spreading time, Pairwise interactions, Poisson process, Markov chain, Analytic performance evaluation

1 Organisation of the paper

This paper is online supplement for the INFORMS Journal on Computing paper, entitled "Probabilistic Analysis of Rumor Spreading Time", by Yves Mocquard, Bruno Sericola and Emmanuelle Anceaume [25].

This online supplement is organized as follows. After a short introduction, this paper provides proofs of Theorems 3 and 9 that appear in [25]. Specifically, we provide explicit expressions of the total number

---

[*] This work was partially funded by the French ANR project SocioPlug (ANR-13-INFR-0003), and by the DeSceNt project granted by the Labex CominLabs excellence laboratory (ANR-10-LABX-07-01).

of interactions needed for all the n nodes to get a rumor in both discrete and continuous time of the asynchronous push-pull model, model which provides minimal assumptions on the computational power of the nodes.

2 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 [11] for the update of a database replicated at different sites, has then been adopted in many applications ranging from resource discovery as in [20], data-aggregation as in [22], complex distributed applications as in [4], or virus propagation in computer networks as in [3], 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, [14] 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 [30, 21] 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 [16]. Investigations have also been done in different topologies of the network as in [5, 10, 13, 27], in the presence of link or nodes failures as in [12], in dynamic graphs as in [6] and spreading with node expansion as in [17].

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, [1] 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. [2] 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, [15] 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. [15] analyzes the mean and the variance of the spreading time of the rumor on general graphs and [28] proposes a thorough study for spreading a rumor on particular Erdös-Rényi random graphs. In [9] the authors propose a different model in which, in addition to spreaders and ignorants, is introduced the notion of stiflers. A stifler learns the rumor but does not propagate it. A stifler results from the interaction between two spreaders, or between a spreader and a stifler. These authors have conjectured that the number of stiflers is asymptotically normal with mean and variance linear in n, where n is the size of the system. This conjecture has been proved in [29]. This model has been generalized by [23] where the authors assume moreover that each spreader ceases to propagate the rumour right after being involved in a random number of stifling experiences. Under a general initial configuration they establish the asymptotic behaviour of the ultimate proportion of ignorants as the population size grows to infinity. In [7], the authors propose a model in which spreaders have a random emission capital that decreases at each emission. They study the proportion of ignorants that receive the information before the emission capital of all the spreaders is exhausted, as well as the exhaustion time. This work is extended Erdös-Rényi random graphs in [8].

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 [15], 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 supposed to be independent with the same rate.

3 The discrete time case

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 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 rumor. The case where the number of initial nodes possessing the rumor is greater than one has been considered in [24]. 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

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

The random variable T_n, defined by

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

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 $\{1, \ldots, n\}$ defined, for all $t \geq 0$, by

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

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 $1, \ldots, n-1$ 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 [31], 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 [24],

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

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

For $i \in \{0, \dots, n\}$, 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 $i \in \{1, \dots, n-1\}$, 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

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

3.1 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 [24]. It is given, for every $n \geq 1$, by

$$\mathbb{E}(T_n) = (n-1)H_{n-1} \underset{n \longrightarrow \infty}{\sim} n \ln(n). \tag{4}$$

In the same way, the explicit value of the variance $\text{Var}(T_n)$ can be found in [24]. It is given by

$$\text{Var}(T_n) = \frac{(n-1)^2}{2} \sum_{\ell=1}^{n-1} \frac{1}{\ell^2} - \frac{n-1}{n} H_{n-1} \underset{n \longrightarrow \infty}{\sim} \frac{\pi^2 n^2}{12}.$$

An explicit expression of the distribution of T_n, for $n \geq 2$, has been obtained in the following theorem which will be 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 $j \in \{1, \dots, n-1\}$, recursively by

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

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

$$\left\{\begin{array}{rcll} c_{i,j} & = & \dfrac{p_i c_{i-1,j}}{p_i - p_j} - \dfrac{p_i d_{i-1,j}}{(p_i - p_j)^2} & \textit{for} \quad i \neq j, n-j, \\ d_{i,j} & = & \dfrac{p_i d_{i-1,j}}{p_i - p_j} & \textit{for} \quad i \neq j, n-j, \\ c_{i,i} & = & 1 - \displaystyle\sum_{j=1, j\neq i}^{\lfloor n/2 \rfloor} c_{i,j} & \textit{for} \quad i \leq \lfloor n/2 \rfloor, \\ c_{i,n-i} & = & 1 - \displaystyle\sum_{j=1, j\neq n-i}^{\lfloor n/2 \rfloor} c_{i,j} & \textit{for} \quad i > \lfloor n/2 \rfloor, \\ d_{i,i} & = & p_i c_{i-1,i} & \textit{for} \quad i \leq \lfloor n/2 \rfloor, \\ d_{i,n-i} & = & p_i c_{i-1,n-i} & \textit{for} \quad i > \lfloor n/2 \rfloor. \end{array}\right. \tag{5}$$

Proof. See [24]. ■

3.2 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 [24]. ■

Recall that $\mathbb{E}(T_n) = (n-1)H_{n-1}$, where H_k is the harmonic series. We proved in [24] that for all real $c \geq 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\{T_n > c\mathbb{E}(T_n)\} = \left\{\begin{array}{ll} 0 & \text{if } c > 1 \\ 1 & \text{if } c < 1. \end{array}\right. \tag{6}$$

For $c = 1$, this result was formulated in [24] as a conjecture. We are now able to give a proof of it.

In order to prove Theorem 3 of [25], 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 $(S_{n,i})_{i\in\{1,\ldots,n-1\}}$ are independent, the $(Z_i)_{i\geq 1}$ are also independent.

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

$$\frac{R_{n,i}}{n} = \frac{S_{n,i} - \mathbb{E}(S_{n,i})}{n} = \frac{p_{n,i}S_{n,i} - 1}{np_{n,i}} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{2i}. \tag{7}$$

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

$$\frac{T_{2k+1} - \mathbb{E}(T_{2k+1})}{2k+1} = \frac{1}{2k+1}\left(\sum_{i=1}^{k} R_{2k+1,i} + \sum_{i=1}^{k} R_{2k+1,2k+1-i}\right) = V_k + \overline{V}_k, \tag{8}$$

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 [32]. 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

$$\lim_{m\longrightarrow\infty}\limsup_{k\longrightarrow\infty}\mathbb{P}\{|V_k-W_{m,k}|\geq\varepsilon\}=0 \tag{9}$$

Let us introduce the notation

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

Using (7) and the fact that the $R_{n,i}$ are independent, we have

$$W_{m,k}\xrightarrow{\mathcal{L}}W_m \text{ as } k\longrightarrow\infty. \tag{10}$$

The hypothesis of the principle of accompanying laws of Theorem 3.1.14 of [32] are properties (7) and (10). 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 (8), 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 of [25].

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 [18] 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,

$$\begin{aligned}\lim_{n \longrightarrow \infty} \mathbb{P}\left\{T_n > \mathbb{E}(T_n)\right\} &= \mathbb{P}\left\{W^{(1)} + W^{(2)} > 0\right\} \\ &= \mathbb{P}\left\{-2\gamma - \ln\left(Z_1 Z_2\right) > 0\right\} \\ &= \mathbb{P}\left\{Z_1 Z_2 < e^{-2\gamma}\right\} \\ &= \int_0^{\infty} \left(1 - \exp(-e^{-2\gamma}/t)\right) e^{-t} dt \\ &= 1 - \int_0^{\infty} \exp(-t - e^{-2\gamma}/t) dt.\end{aligned}$$

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

$$\int_0^{\infty} u'(t) dt = 0. \tag{11}$$

The derivative u' of u is given by

$$\begin{aligned}u'(t) &= \left(-1 + e^{-2\gamma} t^{-2}\right) u(t) \\ &= -u(t) + e^{-2\gamma} u(t) t^{-2}\end{aligned} \tag{12}$$

Integrating (12) over $(0, +\infty)$ and using (11), 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

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

where K_1 is the well-known modified Bessel function of the second kind of order 1, see for instance expression 8.432.6 of [19]. ■

4 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 rumor 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, as suggested by [15], 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

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

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

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

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 $\{1,\dots,n\}$ defined, for all $t \geq 0$, by

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

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 $i \in \{1,\dots,n\}$, 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 $i \in \{1,\dots,n-1\}$, 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 $1,\dots,n-1$ 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 [31], that the distribution of Θ_n is given, for every $t \geq 0$, by

$$\mathbb{P}\{\Theta_n > t\} = \alpha e^{Rt} \mathbb{1}, \tag{13}$$

where α is the row vector containing the initial probabilities of states $1,\dots,n-1$, 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 $i \in \{1,\dots,n-1\}$ 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

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

Since the μ_i depend on n, we rename the random variables U_i as $U_{n,i}$ and the parameters μ_i as $\mu_{n,i}$.

In order to prove Theorem 9 of [25], we first need, as in the discrete time case, the following technical lemmas The spreading time Θ_n thus writes as $\Theta_n = U_{n,1} + \cdots + U_{n,n-1}$. Note that this result has also been obtained by [26] using a quite different proof.

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 $n \geq 2$, $i \in \{1, \ldots, n-1\}$ 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 $\mu_{n,i} U_{n,i} \overset{\mathcal{L}}{=} Z_i$. Moreover since the $(U_{n,i})_{i\in\{1,\ldots,n-1\}}$ 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 $\mathbb{E}(U_{n,i}) = 1/\mu_{n,i}$,

$$R_{n,i} = U_{n,i} - \mathbb{E}(U_{n,i}) = \frac{\mu_{n,i} S_{n,i} - 1}{\mu_{n,i}} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{2\lambda i}. \tag{14}$$

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

$$\Theta_{2k+1} - \mathbb{E}(\Theta_{2k+1}) = V_k + \overline{V}_k. \tag{15}$$

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 [32]. 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 of [25].

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 [18] 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. ■

References

[1] Huseyin Acan, Andrea Collevecchio, Abbas Mehrabian, and Wormald Nick. On the push&pull protocol for rumour spreading. In *Proceedings of the ACM Symposium on Principles of Distributed Systems (PODC)*, 2015.

[2] Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(2):183–199, 2008.

[3] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2005.

[4] Keren Censor-Hillel, Bernhard Haeupler, Jonathan Kelner, and Petar Maymounkov. Global computation in a poorly connected world: Fast rumor spreading with no dependence on conductance. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, 2012.

[5] Flavio Chierichetti, Silvio Lattanzi, and Alessandro Panconesi. Rumor spreading in social networks. *Theoretical Computer Science*, 412(24):2602–2610, 2011.

[6] Andrea Clementi, Pierluigi Crescenzi, Carola Doerr, Pierre Fraigniaud, Francesco Pasquale, and Riccardo Silvestri. Rumor spreading in random evolving graphs. *Random structures and Algorithms*, 48(2):290–312, 2015.

[7] F. Comets, F. Delarue, and R. Schott. Information transmission under random emission constraints. *Environmental Modelling & Software*, 23(6):973–1009, 2014.

[8] F. Comets, C. Gallesco, S. Popov, and M. Vachkovskaia. Constrained information transmission on Erdös-Rényi graphs. *Markov Processes and Related Fields*, 22:111–138, 2016.

[9] D.J. Daley and David George Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965.

[10] Sebastian Daum, Fabian Kuhn, and Yannic Maus. Rumor spreading with bounded indegree. In *Proceedings of the International Colloquium on Structural Information and Communication Complexity (SIROCCO)*, 2016.

[11] Alan Demers, Mark Gealy, Dan Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard Sturgis, Dand Swinehart, and Doug Terry. Epidemic algorithms for replicated datbase maintenance. In *Proceedings of the ACM Syposium on Principles of Distributed Systems (PODC)*, 1987.

[12] Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.

[13] Nicolaos Fountoulakis and Konstantinos Panagiotou. Rumor spreading on random regular graphs and expanders. *Random Structures and Algorithms*, 43(2):201–220, 2013.

[14] Alan Frieze and Geoffrey Grimmet. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10(1):57–77, 1985.

[15] A. J. Ganesh. Rumour spreading on graphs. Technical report, 2015.

[16] George Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *Proceedings of the International Symposium on Theoretical Aspects of Computer Science (STACS)*, 2011.

[17] George Giakkoupis. Tight bounds for rumor spreading with vertex expansion. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.

[18] Louis Gordon. Bounds for the distribution of the generalized variance. *The Annals of Statistics*, 17(4):1684–1692, 1989.

[19] Izrail S. Gradshteyn and Iosif M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, Boston, MA, USA, eighth edition, 2014.

[20] Mor Harchol-Balter, Tom Leighton, and Daniel Lewin. Resource discovery in distributed networks. In *Proceedings of the ACM Syposium on Principles of Distributed Systems (PODC)*, 1999.

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

[22] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003.

[23] Elcio Lebensztayn, Anderson F.P. Machado, and Pablo Martin Rodrguez. On the behaviour of a rumour process with random stifling. *Environmental Modelling & Software*, 26:517–522, 2011.

[24] Yves Mocquard, Samantha Robert, Bruno Sericola, and Emmanuelle Anceaume. Analysis of the propagation time of a rumour in large-scale distributed systems. In *Proceedings of the 15th IEEE International Symposium on Network Computing and Applications (NCA)*, 2016.

[25] Yves Mocquard, Bruno Sericola, and Emmanuelle Anceaume. Probabilistic analysis of rumor spreading time. *INFORMS Journal on Computing*, 2018. Under press.

[26] Stanislav Molchanov and Joseph M. Whitmeyer. Two Markov Models of the Spread of Rumors. *Journal of Mathematical Sociology*, 34:157–166, 2010.

[27] Konstantinos Panagiotou, Xavier Perez-Gimenez, Thomas Sauerwald, and Hé Sun. Randomized rumor spreading: the effect of the network topology. *Combinatorics, Probability and Computing*, 24(2):457–479, 2015.

[28] Konstantinos Panagiotou and Leo Speidel. Asynchronous rumor spreading on random graphs. *Algorithmica*, 78(3):968–989, 2017.

[29] B. Pittel. On a daley-kendall model of random rumours. *Journal of Applied Probability*, 27(1):14–27, 1990.

[30] Boris Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[31] Bruno Sericola. *Markov Chains. Theory, Algorithms and Applications.* Applied stochastic methods series. Wiley, 2013.

[32] Daniel W. Stroock. *Probability Theory: An Analytic View.* Cambridge University Press, second edition, 2010.