HAL open science

Continuous-Time Stochastic Analysis of Rumor Spreading with Multiple Operations

François Castella, Bruno Sericola, Emmanuelle Anceaume, Yves Mocquard

To cite this version:

François Castella, Bruno Sericola, Emmanuelle Anceaume, Yves Mocquard. Continuous-Time Stochastic Analysis of Rumor Spreading with Multiple Operations. Methodology and Computing in Applied Probability, 2023, 25 (4), pp.82. 10.1007/s11009-023-10058-7 . hal-04255487

HAL Id: hal-04255487

https://cnrs.hal.science/hal-04255487v1

Submitted on 24 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

Continuous-time stochastic analysis of rumor spreading with multiple operations

François Castella[1], Bruno Sericola[2*], Emmanuelle Anceaume[3], Yves Mocquard[2]

[1]University of Rennes, Campus de Beaulieu, Rennes, 35042, France.
[2]Inria, Campus de Beaulieu, Rennes, 35042, France.
[3]CNRS, Campus de Beaulieu, Rennes, 35042, France.

*Corresponding author(s). E-mail(s): bruno.sericola@inria.fr;
Contributing authors: francois.castella@univ-rennes.fr;
emmanuelle.anceaume@irisa.fr; yves.mocquard@irisa.fr;

Abstract

In this paper, we analyze a new asynchronous rumor spreading protocol to deliver a rumor to all the nodes of a large-scale distributed network. This protocol relies on successive pull operations involving $\boldsymbol{k}$ different nodes, with $\boldsymbol{k} \geq \mathbf{2}$, and called $\boldsymbol{k}$-pull operations. Specifically during a $\boldsymbol{k}$-pull operation, an uninformed node $\boldsymbol{a}$ contacts $\boldsymbol{k} - \mathbf{1}$ other nodes at random in the network, and if at least one of them knows the rumor, then node $\boldsymbol{a}$ learns it. We perform a detailed study in continuous-time of the total time $\boldsymbol{\Theta_{k,n}}$ needed for all the $\boldsymbol{n}$ nodes to learn the rumor. These results extend those obtained in a previous paper which dealt with the discrete-time case. We obtain the mean value, the variance and the distribution of $\boldsymbol{\Theta_{k,n}}$ together with their asymptotic behavior when the number of nodes $\boldsymbol{n}$ tends to infinity.

Keywords: Rumor spreading time, $\boldsymbol{k}$-pull protocol, Poisson Process, Markov chain, Asymptotic analysis

1 Introduction

Randomized rumor spreading or gossiping is an important communication mechanism that allows the dissemination of information in large-scale and open networks.

A large-scale and open network comprises a collection of sequential computing entities (*e.g.*, processes, processors, nodes, agents, peers) that join and leave the system at any time, and communicate with one another by exchanging messages. Randomized rumor spreading was initially proposed by Deemers et al. [1] for the update of a database replicated at different sites, and has then been adopted in many applications due to its robustness and simplicity. In contrast to reliable communication broadcasts which must provide agreement on the broadcast value with possibly additional ordering guarantees on the delivery of updates from different sources, a randomized rumor spreading procedure provides reliability only with some probability. A randomized spreading rumor protocol describes the rules required for one or more pieces of information known by an arbitrary node in the network (we call such a node an *informed* node) to be spread to all the nodes of the network. The push and pull protocols are the basic operations used by the nodes to propagate an information over the entire network [1, 2]. With the push operation, an informed node contacts some randomly chosen node in the system, and gives it the rumor while with the pull operation, an uninformed node contacts some random node and asks for the rumor. Note that in both cases the contacted node may already know the rumor or not. The same node can perform both operations according to whether it knows or not the rumor, which corresponds to the push-pull protocol, or performs only one, either a pull or push operation, which corresponds to the pull or push protocols respectively. One of the important questions raised by these protocols is the spreading time, that is the time needed for the rumor to be known by all the nodes of the network.

To answer such a question, one first needs to specify how synchronized nodes are, or in other words whether we suppose that all the nodes of the system act in a synchronous way or not. In the former case, the system model is synchronous while in the latter case it is asynchronous. The most studied one is the synchronous model. This model assumes that all the nodes of the network act synchronously, which allows the algorithms designed in this model to divide time in synchronized rounds. During each synchronized round, each node a of the network selects at random one of its neighbors b and either sends the rumor to b if a knows it (push operation) or gets the rumor from b if b knows the rumor (pull operation). In this model, the spreading time of a rumor is defined as the number of synchronous rounds necessary for all the nodes to know the rumor. When the underlying graph is complete, it has been shown by Frieze [3] that the number of rounds divided by $\log_2(n)$ 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 [4, 5] 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 [6]. Investigations have also been done in different topologies of the network as in [7–10], in the presence of link or nodes failures as in [11], in dynamic graphs as in [12], and in general graphs in terms of vertex expansion [13]. Another alternative consists in letting the nodes make more than one call during the push or pull operations [14]. The authors show that the push-pull protocol takes $O(\log n / \log\log n)$ rounds in expectation if the number of neighbors of a node is chosen independently according to a power law distribution with exponent $\beta \in (2, 3)$.

In large scale networks, that is in networks involving several thousands of nodes, assuming that all nodes act synchronously is a very strong assumption. Thus several authors, including [15–19], suppose an asynchronous model, that is a model in which nodes asynchronously trigger operations with randomly chosen nodes in the system, either to push, pull or push-pull information. The asynchronous gossip protocol is usually modeled by a continuous-time stochastic (Markovian) process [15–19]. This type of stochastic processes belongs to the death process category, which has many applications in demography, queuing theory, performance engineering, epidemiology, biology and many other distributed applications. For instance, in [20], an analysis of the susceptible-infected model –corresponding to an asynchronous push-pull model– allows us in some cases to explicit the state probabilities by using the Laplace transform on the Kolmogorov forward equation. However, these techniques prove ineffective when the transition rate is non-linear (Laplace transform inversion becomes a tricky exercise). Most of the rumor spreading protocols studied in the asynchronous models rely either on the push-pull operations or on the push operations. Indeed, pushing the information allows us to initiate the rumor very quickly but then struggles to reach the few uninformed nodes. In contrast, the pull algorithm attracted very little attention because this operation was long considered inefficient to spread a rumor within a large scale network [21]. It is actually very useful in systems fighting against message saturation (see for instance [22]). The ineffectiveness of the pull protocol stems from the fact that it takes some time before the rumor reaches a phase of exponential growth.

The objective of this paper is to further develop this line of inquiry by studying the k-pull protocol in the continuous-time case. This protocol counterbalances the slow initiation of pull-based rumour spreading protocols by increasing the chances of learning the rumor with each operation. A local clock following an exponential distribution with rate λ is associated with each uninformed node of the system. Each time the clock of an uniformed node rings, this node contacts $k-1$, with $k \geq 2$ distinct nodes, chosen at random uniformly among the $n-1$ other nodes. If at least one of these contacted nodes knows the rumor, the initiator of the k-pull operation learns the rumor and clears its clock.

Concretely, the k-pull operation is interesting in all situations in which you would like to benefit from multiple concomitant responses to build up your opinion. This is typically the case in fault tolerant distributed applications (including Byzantine fault tolerance, consensus, clock synchronization, leader election), as well as in large distributed applications (e.g., peer-to-peer communication, blockchain) where nodes require sufficiently many responses/votes to cope with the presence of faulty nodes. Other scenario that would take advantage of such an operation is cybersecurity. When investigating cybersecurity incidents, experts typically conduct their investigation in the form of a knowledge graph to explore and discover complex attack paths. For instance, the Defants company [23] has developed a model for representing raw system and network information in the form of a knowledge graph. This graph can contain up to millions of nodes about an incident that took place on a network of several hundreds machines and users. In such a graph, the nodes model the elements of the system: machines, user accounts, session information, files, services and so on. Two nodes are related if an action carried out on one of them has generated the second. These actions

may include decryption, decompression, execution or packet transmission. By relying on the k-pull operation, the expert will be able to compare or audit concomitant responses, giving rise to fast and informed decisions.

The remainder of the paper is organized as follows. In Section 2, we present the asynchronous k-pull protocol and introduce the random variable $\Theta_{k,n}$which represents the total amount of time needed for all the nodes to know the rumor. We prove in Section 3 that the mean number of k-pull operations needed to inform all the n nodes of the system, assuming that a single node initially knows the rumor, that is $\mathbb{E}(\Theta_{k,n})$, is equivalent to $k\ln(n)/(k-1)\lambda$ when the number of nodes n in the system tends to infinity. We also show that the limiting variance of $\Theta_{k,n}$ is equal to $(1+1/(k-1)^2)\pi^2/6\lambda^2$ when n tends infinity. The distribution of the rumor spreading time $\Theta_{k,n}$ is analyzed in Section 4. We provide explicit limiting distributions of $\Theta_{k,n}-\mathbb{E}(\Theta_{k,n})$ and $\Theta_{k,n}-k\ln(n)/(k-1)\lambda$ when n tends to infinity. All these results are illustrated using numerical values. Section 5 concludes the paper.

2 The model

As we will see in this section, the results obtained for the discrete-time model of the k-pull rumor spreading, which has been analyzed in [24], cannot be used to deal with the continuous-time model of the k-pull rumor spreading. Indeed, we need here a much more sophisticated analysis.

We consider a complete network of size n in which each node may be asked for a piece of information (pull event). The algorithm starts with a single node informed of the rumor. A local clock following an exponential distribution with rate λ is associated with each uniformed node of the system. Each time the clock of an uniformed node s rings, this node contacts $k-1$, with $k\geq 2$, distinct nodes, chosen at random uniformly among the $n-1$ other nodes. If at least one of these contacted nodes knows the rumor, node s learns it and clears its clock (*i.e.*, s remains contactable but does not contact other nodes). We suppose that the k-pull operation once triggered is instantaneous *i.e.*, the time for a node to contact $k-1$ other nodes and to receive their response is immediate

Since the clock of an uninformed node rings after a time that is exponentially distributed with rate λ, we naturally introduce the continuous-time Markov chain $Z=\{Z_t,\ t\geq 0\}$, where Z_t represents the number of informed nodes at continuous-time $t\geq 0$. Specifically the transitions of Z occur at successive instants $\tau_0=0$, $\tau_1,\ldots$, where the $\tau_i-\tau_{i-1}$, $i\geq 1$, are independent and exponentially distributed with rate $(n-i)\lambda p_{k,n}(i)$, where

$$p_{k,n}(i)=\begin{cases}1-\dfrac{\binom{n-1-i}{k-1}}{\binom{n-1}{k-1}}=1-\displaystyle\prod_{h=1}^{k-1}\left(1-\frac{i}{n-h}\right) & \text{if } 1\leq i\leq n-k\\[2ex] 1 & \text{if } n-k+1\leq i\leq n-1.\end{cases}\tag{1}$$

Indeed, $p_{k,n}(i)$ is the probability, given that $Z_t = i$, that the set of $k-1$ chosen nodes (*i.e.*, $k-1$ among $n-1$) at the next alarm clock contains at least one node among the i informed nodes. Hence the global clock of the process rings according to an exponential distribution whose rate is proportional to the amount of uniformed nodes. Note that a jump of process Z corresponds to a new informed node.

Observe also that the continuous-time model of the rumor spreading corresponds to a physical time, that is the total amount of time needed for all the n nodes to learn the rumor, while the discrete-time model stands for the total number of k-pull operations needed for all the n nodes to learn the rumor.

We denote by $\Theta_{k,n}$ the random variable defined by

$$\Theta_{k,n} = \inf\{t \geq 0 \mid Z_t = n\}$$

which represents the continuous-time model *i.e.*, the total amount of time needed for all the nodes to know the rumor. The spreading time $\Theta_{k,n}$ can thus be expressed as a sum of independent and exponentially distributed random variables. Specifically, introducing the notation $U_{k,n}(i) = \tau_i - \tau_{i-1}$, for $i \geq 1$ with $\tau_0 = 0$, we have

$$\Theta_{k,n} = \sum_{i=1}^{n-1} U_{k,n}(i), \tag{2}$$

where $U_{k,n}(1), \ldots, U_{k,n}(n-1)$ are independent and $U_{k,n}(i)$ is exponentially distributed with rate $(n-i)\lambda p_{k,n}(i)$.

The authors of [24] used two technical lemmas (Lemma 2 and 3 of [24]) to analyze the asymptotic behavior of the moments and the distribution of the rumor spreading time in the discrete-time case. These two lemmas allow them to provide simple lower and upper bounds for the probabilities $p_{k,n}(i)$. These bounds are then used to get other asymptotically equal bounds of the moments of the discrete-time rumor spreading time which the sum of geometric random variables with parameters $p_{k,n}(i)$. In the continuous-time case, the bounds of the $p_{k,n}(i)$ obtained in [24] do not lead to asymptotically equal bounds of the moments of the continuous-time rumor spreading time which is the sum of exponential random variables with rates $(n-i)\lambda p_{k,n}(i)$. This is due to the multiplicative factor $n-i$ which arises in the rates $(n-i)\lambda p_{k,n}(i)$. We thus need much more refined bound than those obtained in [24].

3 Moments of the rumor spreading time

We analyze in this section the first two moments of the rumor spreading time by using appropriate lower and upper bounds. The following technical lemma is used to obtain their asymptotic behavior. In this lemma, γ is the Euler-Mascheroni constant given by $\gamma \approx 0.5772156649$.

Lemma 1. *Let g be a Lipschitz function on interval* $[0,1]$.

$$\textit{If } g(0) \neq 0 \textit{ then } \sum_{i=1}^{n} \frac{1}{i} g\left(\frac{i}{n}\right) = g(0)\ln(n) + \gamma g(0) + \int_0^1 \frac{g(x)-g(0)}{x} dx + \varepsilon(n).$$

where $\varepsilon(n)$ is the remainder of the Euler-Mascheroni representation of the harmonic sum in terms of the logarithm which satisfies $\lim_{n\longrightarrow\infty}\varepsilon(n)=0$.

Proof. Using the integral form of the remainder for the Taylor series of function g, we get

$$g(x)-g(0)=xG(x),\ \text{where } G(x)=\int_0^1 g'(tx)dt,$$

function g' being the derivative of g. We then have

$$\sum_{i=1}^{n}\frac{1}{i}g\left(\frac{i}{n}\right)-g(0)\sum_{i=1}^{n}\frac{1}{i}=\sum_{i=1}^{n}\frac{1}{i}\frac{i}{n}G\left(\frac{i}{n}\right)=\frac{1}{n}\sum_{i=1}^{n}G\left(\frac{i}{n}\right)\underset{n\to\infty}{\longrightarrow}\int_0^1 G(x)dx,$$

since the last term is a Riemann sum. We then use the following well-known development of the harmonic sum

$$\sum_{i=1}^{n}\frac{1}{i}=\ln(n)+\gamma+\varepsilon(n),$$

where $\varepsilon(n)$ is such that $\lim_{n\longrightarrow\infty}\varepsilon(n)=0$. □

As usual, for two sequences (u_n) and (v_n) we introduce the notation $u_n \underset{n\longrightarrow\infty}{\sim} v_n$ to mean that $\lim\limits_{n\longrightarrow\infty} u_n/v_n=1$. In particular, from Lemma 1 we get, for any function g Lipschitz on $[0,1]$,

$$\text{If } g(0)\neq 0 \text{ then } \sum_{i=1}^{n}\frac{1}{i}g\left(\frac{i}{n}\right)\underset{n\longrightarrow\infty}{\sim} g(0)\ln(n). \tag{3}$$

3.1 Expected rumor spreading time

Using (1) and (2), the expected rumor spreading time writes

$$\mathbb{E}(\Theta_{k,n})=\frac{1}{\lambda}\sum_{i=1}^{n-1}\frac{1}{(n-i)p_{k,n}(i)}=\frac{1}{\lambda}\sum_{i=1}^{n-k}\frac{1}{(n-i)p_{k,n}(i)}+\frac{1}{\lambda}\sum_{i=1}^{k-1}\frac{1}{i}. \tag{4}$$

Using the fact that $0\leq h\leq k$ in Relation (1), we easily get

$$1-\left(1-\frac{i}{n}\right)^{k-1}\leq p_{k,n}(i)\leq 1-\left(1-\frac{i}{n-k}\right)^{k-1}. \tag{5}$$

Introducing the notation

$$\alpha_n=\sum_{i=1}^{n-k}\frac{1}{(n-i)p_{k,n}(i)},$$

these inequalities lead to

$$\gamma_n\leq\alpha_n\leq\beta_n, \tag{6}$$

where

$$\beta_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)\left[1-\left(1-\frac{i}{n}\right)^{k-1}\right]} \text{ and } \gamma_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)\left[1-\left(1-\frac{i}{n-k}\right)^{k-1}\right]}.$$

We now show that a refined analysis of the terms β_n and γ_n leads to a precise description of the asymptotic behavior of the expected rumor spreading time $\mathbb{E}(\Theta_{k,n})$.

For $k \geq 2$, we introduce the function f_k defined, for all $x \in [0,1]$ by

$$f_k(x) = \frac{1}{1+(1-x)+(1-x)^2+\cdots+(1-x)^{k-2}}.$$

Observe that function f_k is clearly Lipschitz on $[0,1]$. It is frequently used in the remainder of the paper.

Lemma 2. *For all $k \geq 2$, we have*

$$\beta_n = \frac{k\ln(n)}{k-1} + \frac{\gamma k}{k-1} - \sum_{i=1}^{k-1}\frac{1}{i} - \frac{\ln(k-1)}{k-1} + \varepsilon(n),$$

where $\varepsilon(n)$ is such that $\lim_{n\longrightarrow\infty}\varepsilon(n) = 0$.

Proof. For all $k \geq 2$ and $y \in \mathbb{R}$, using the identity

$$1-y^{k-1} = (1-y)\left(1+y+\cdots+y^{k-2}\right),$$

and by definition of function f_k, we have

$$\begin{aligned}
\beta_n &= \sum_{i=1}^{n-k} \frac{f_k\left(\frac{i}{n}\right)}{(n-i)\left[1-\left(1-\frac{i}{n}\right)\right]} = \sum_{i=1}^{n-k} \frac{n}{i(n-i)} f_k\left(\frac{i}{n}\right) \\
&= \sum_{i=1}^{n-k}\left(\frac{1}{i}+\frac{1}{n-i}\right) f_k\left(\frac{i}{n}\right) \qquad (7)\\
&= \sum_{i=1}^{n-k}\frac{1}{i} f_k\left(\frac{i}{n}\right) + \sum_{i=k}^{n-1}\frac{1}{i} f_k\left(1-\frac{i}{n}\right) \\
&= \sum_{i=1}^{n}\frac{1}{i} f_k\left(\frac{i}{n}\right) + \sum_{i=1}^{n}\frac{1}{i} f_k\left(1-\frac{i}{n}\right) \\
&\quad - \sum_{i=n-k+1}^{n}\frac{1}{i} f_k\left(\frac{i}{n}\right) - \sum_{i=1}^{k-1}\frac{1}{i} f_k\left(1-\frac{i}{n}\right) - \frac{1}{n} f_k(0).
\end{aligned}$$

Observing that

$$\lim_{n\longrightarrow\infty}\left[\sum_{i=n-k+1}^{n}\frac{1}{i}f_k\left(\frac{i}{n}\right)+\sum_{i=1}^{k-1}\frac{1}{i}f_k\left(1-\frac{i}{n}\right)+\frac{1}{n}f_k(0)\right]=f_k(1)\sum_{i=1}^{k-1}\frac{1}{i},$$

we obtain

$$\beta_n=\sum_{i=1}^{n}\frac{1}{i}f_k\left(\frac{i}{n}\right)+\sum_{i=1}^{n}\frac{1}{i}f_k\left(1-\frac{i}{n}\right)-f_k(1)\sum_{i=1}^{k-1}\frac{1}{i}+\varepsilon(n),$$

where $\varepsilon(n)$ is such that $\lim_{n\longrightarrow\infty}\varepsilon(n)=0$. Using now twice Lemma 1, once with $g(x)=f_k(x)$ and once again with $g(x)=f_k(1-x)$, we obtain, since function f_k is Lipschitz with $f_k(0)=1/(k-1)\neq 0$ and $f_k(1)=1\neq 0$,

$$\beta_n=\frac{k\ln(n)}{k-1}+\frac{\gamma k}{k-1}-\sum_{i=1}^{k-1}\frac{1}{i}+\int_0^1\left(\frac{f_k(1-x)+f_k(x)}{x}-\frac{k}{(k-1)x}\right)dx+\varepsilon(n).$$

We need to compute the quantity

$$I:=\int_0^1\left(\frac{f_k(1-x)+f_k(x)}{x}-\frac{k}{(k-1)x}\right)dx.$$

Coming back to the definition of f_k, we introduce the polynomial

$$q_k(x)=1+x+\cdots+x^k,$$

so that $f_k(x)=1/q_{k-2}(1-x)$ and $f_k(1-x)=1/q_{k-2}(x)$. On top of that, we observe that $k/(k-1)=1+1/(k-1)=f_k(0)+f_k(1)$. These two observations lead to

$$I=\int_0^1\frac{1}{x}\left(\frac{1}{q_{k-2}(x)}-\frac{1}{q_{k-2}(0)}+\frac{1}{q_{k-2}(1-x)}-\frac{1}{q_{k-2}(1)}\right)dx,$$

and the variable change $x:=1-x$ to deal with the second difference leads to

$$I=\int_0^1\left[\frac{1}{x}\left(\frac{1}{q_{k-2}(x)}-\frac{1}{q_{k-2}(0)}\right)+\frac{1}{1-x}\left(\frac{1}{q_{k-2}(x)}-\frac{1}{q_{k-2}(1)}\right)\right]dx.$$

Now the whole point in order to compute this integral is to factorize x in the first difference $1/q_{k-2}(x)-1/q_{k-2}(0)$, and to factorize $1-x$ in the second difference $1/q_{k-2}(x)-1/q_{k-2}(1)$, so as to remove the apparent singularities and to recover computable quantities.

Concerning the first difference, we observe that

$$q_{k-2}(0)-q_{k-2}(x)=-x-x^2-\cdots-x^{k-2}=-xq_{k-3}(x),$$

which gives

$$\frac{1}{x}\left(\frac{1}{q_{k-2}(x)}-\frac{1}{q_{k-2}(0)}\right)=-\frac{q_{k-3}(x)}{q_{k-2}(0)q_{k-2}(x)}=-\frac{q_{k-3}(x)}{q_{k-2}(x)}.$$

The other difference requires slightly more attention. We claim that the following formula holds

$$q_{k-2}(1)-q_{k-2}(x)=(1-x)\left[(k-2)q_{k-3}(x)-xq'_{k-3}(x)\right]. \tag{8}$$

Admitting the result for the time being, we obtain

$$\begin{aligned}\frac{1}{1-x}\left(\frac{1}{q_{k-2}(x)}-\frac{1}{q_{k-2}(1)}\right)&=\frac{(k-2)q_{k-3}(x)-xq'_{k-3}(x)}{q_{k-2}(1)q_{k-2}(x)}\\&=\frac{(k-2)q_{k-3}(x)-xq'_{k-3}(x)}{(k-1)q_{k-2}(x)}.\end{aligned}$$

Eventually, this provides

$$\begin{aligned}I&=\int_0^1\left[-\frac{q_{k-3}(x)}{q_{k-2}(x)}+\frac{(k-2)q_{k-3}(x)-xq'_{k-3}(x)}{(k-1)q_{k-2}(x)}\right]dx\\&=-\frac{1}{k-1}\int_0^1\frac{q_{k-3}(x)+xq'_{k-3}(x)}{q_{k-2}(x)}dx\\&=-\frac{1}{k-1}\int_0^1\frac{q_{k-3}(x)+xq'_{k-3}(x)}{xq_{k-3}(x)+1}dx,\end{aligned}$$

where the last equality comes from the observation $q_{k-2}(x)=xq_{k-3}(x)+1$. As a final result, we recover, since $(xq_{k-3}(x)+1)'=xq'_{k-3}(x)+q_{k-3}(x)$, the value

$$I=-\frac{\ln\left(p_{k-3}(1)\right)}{k-1}=-\frac{\ln\left(k-1\right)}{k-1}.$$

There remains to prove formula (8). The formula can easily be proved using a recursion procedure. Alternatively, one may write, using the fact that $1-x^j=(1-x)\sum_{\ell=0}^{j-1}x^\ell$, the relations

$$\begin{aligned}q_{k-2}(1)-q_{k-2}(x)&=\sum_{j=1}^{k-2}\left(1-x^j\right)=(1-x)\sum_{j=1}^{k-2}\sum_{\ell=0}^{j-1}x^\ell=(1-x)\sum_{\ell=0}^{k-3}\sum_{j=\ell+1}^{k-2}x^\ell\\&=(1-x)\sum_{\ell=0}^{k-3}(k-2-\ell)x^\ell=(1-x)\left[(k-2)q_{k-3}(x)-xq'_{k-3}(x)\right],\end{aligned}$$

which completes the proof. □

We consider now the term γ_n.

Lemma 3. *For all $k \geq 2$, we have*

$$\lim_{n \longrightarrow \infty} (\gamma_n - \beta_n) = 0.$$

Proof. By definition of function f_k, we obtain easily

$$\gamma_n = \sum_{i=1}^{n-k} \frac{f_k\left(\frac{i}{n-k}\right)}{(n-i)\left[1-\left(1-\frac{i}{n-k}\right)\right]} = (n-k)\sum_{i=1}^{n-k} \frac{1}{i(n-i)} f_k\left(\frac{i}{n-k}\right)$$
$$= \frac{n-k}{n}\sum_{i=1}^{n-k}\left(\frac{1}{i}+\frac{1}{n-i}\right) f_k\left(\frac{i}{n-k}\right) \tag{9}$$

The function f_k being Lipschitz on interval $[0,1]$, defining $C_k = \max_{z\in[0,1]} |f_k'(z)|$ where f_k' is the derivative of f_k, we get, using Relations (7) and (9),

$$\left|\gamma_n - \frac{n-k}{n}\beta_n\right| \leq \frac{n-k}{n}\sum_{i=1}^{n-k}\left(\frac{1}{i}+\frac{1}{n-i}\right)\left|f_k\left(\frac{i}{n-k}\right) - f_k\left(\frac{i}{n}\right)\right|$$
$$\leq C_k \frac{n-k}{n}\sum_{i=1}^{n-k}\left(\frac{1}{i}+\frac{1}{n-i}\right)\frac{ik}{n(n-k)}$$
$$\leq C_k \frac{k}{n}\left(\sum_{i=1}^{n-k}\frac{1}{i} + \sum_{i=k}^{n-1}\frac{1}{i}\right).$$

This bound tends to 0 when n tends to infinity. Using Lemma 2, we have $\lim_{n\longrightarrow\infty} \beta_n/n = 0$. It follows that

$$\lim_{n\longrightarrow\infty} |\gamma_n - \beta_n| \leq \lim_{n\longrightarrow\infty}\left|\gamma_n - \frac{n-k}{n}\beta_n\right| + k\lim_{n\longrightarrow\infty}\frac{\beta_n}{n} = 0,$$

which completes the proof. □

The following theorem gives the asymptotic behavior of the expected rumor spreading time. It will be used in Corollary 12 to get an asymptotic behavior of the distribution of $\Theta_{k,n}$.

Theorem 4. *For all $k \geq 2$, we have*

$$\mathbb{E}(\Theta_{k,n}) = \frac{1}{\lambda}\left(\frac{k\ln(n)}{k-1} + \frac{\gamma k}{k-1} - \frac{\ln(k-1)}{k-1}\right) + \varepsilon(n),$$

where $\varepsilon(n)$ is such that $\lim_{n\longrightarrow\infty}\varepsilon(n) = 0$.

k	$\mathbb{E}\left(\Theta_{k,10^3}\right)$	$F_{k,10^3}$	$\mathbb{E}\left(\Theta_{k,10^4}\right)$	$F_{k,10^4}$	$\mathbb{E}\left(\Theta_{k,10^5}\right)$	$F_{k,10^5}$
2	14.95397	14.96994	19.57305	19.57511	24.18003	24.18028
3	10.87364	10.88088	14.33386	14.33476	17.78853	17.78864
4	9.60801	9.613757	12.68314	12.68387	15.75380	15.75399
5	9.00458	9.009640	11.88722	11.88787	14.76602	14.76610
6	8.65542	8.660076	11.42258	11.42318	14.18621	14.18628
7	8.42946	8.433840	11.11962	11.12019	13.80647	13.80654
8	8.27209	8.276265	10.90724	10.90779	13.53925	13.53932
9	8.15665	8.160662	10.75054	10.75107	13.34141	13.34148
10	8.06861	8.072498	10.63041	10.63093	13.18929	13.18935

Table 1 Values of $\mathbb{E}\left(\Theta_{k,n}\right)$ and $F_{k,n}$ for $k = 2, \ldots, 10$, $\lambda = 1$ and $n = 10^3, 10^4, 10^5$.

Proof. Relations (4) and (6) give

$$\frac{\gamma_n}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{k-1}\frac{1}{i} \leq \mathbb{E}(\Theta_{k,n}) \leq \frac{\beta_n}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{k-1}\frac{1}{i}.$$

Using Lemma 2 and Lemma 3, we easily get the desired result. □

In particular, from Theorem 4 we get,

$$\mathbb{E}(\Theta_{k,n}) \underset{n\longrightarrow\infty}{\sim} \frac{k\ln(n)}{(k-1)\lambda}.$$

We illustrate our results through Table 1, which gives the expected rumor spreading time $\mathbb{E}(\Theta_{k,n})$ when $\lambda = 1$, for different values of k and n and the approximation given by Theorem 4. For that purpose, we introduce the notation

$$F_{k,n} = \frac{1}{\lambda}\left(\frac{k\ln(n)}{k-1} + \frac{\gamma k}{k-1} - \frac{\ln(k-1)}{k-1}\right).$$

We observe that the asymptotic value $F_{k,n}$ is very close to $\mathbb{E}\left(\Theta_{k,n}\right)$. Observe also, as expected, that $\mathbb{E}\left(\Theta_{k,n}\right)$ increases with n and decreases with k. Moreover, for example when $k = 5$, we see that if the local clock of each uninformed node rings at an expected frequence of 1 per unit of time (i.e. $\lambda = 1$) then, when the number n of nodes is equal to 10^5, the expected rumor spreading time is equal to 14.76602 units of time, which is quite small regarding the large number of nodes.

3.2 Variance of the rumor spreading time

The following lemma is needed to obtain the limiting value of the variance of $\Theta_{k,n}$ when n tends to infinity.

Lemma 5. *Let g be a Lipschitz function on interval* $[0, 1]$.

$$\textit{If } g(0) \neq 0 \textit{ then } \lim_{n\longrightarrow\infty}\sum_{i=1}^{n}\frac{1}{i^2}g\left(\frac{i}{n}\right) = g(0)\frac{\pi^2}{6}.$$

Proof. Function g being Lipschitz on interval $[0,1]$, for all $x, y \in [0,1]$, we have

$$|g(x) - g(y)| \leq C|x - y|, \text{ where } C = \max_{z \in [0,1]} |g'(z)|$$

and g' is the derivative of function g. We then have, by taking $x = i/n$ and $y = 0$,

$$\left| \sum_{i=1}^{n} \frac{1}{i^2} g\left(\frac{i}{n}\right) - g(0) \sum_{i=1}^{n} \frac{1}{i^2} \right| \leq \sum_{i=1}^{n} \frac{1}{i^2} \left| g\left(\frac{i}{n}\right) - g(0) \right| \leq C \frac{1}{n} \sum_{i=1}^{n} \frac{1}{i} \underset{n \longrightarrow \infty}{\sim} C \frac{\ln(n)}{n}.$$

This means that

$$\lim_{n \longrightarrow \infty} \sum_{i=1}^{n} \frac{1}{i^2} g\left(\frac{i}{n}\right) = g(0) \sum_{i=1}^{\infty} \frac{1}{i^2} = g(0) \frac{\pi^2}{6},$$

which completes the proof. □

Using (1) and (2), the variance of the rumor spreading time writes

$$\mathbb{V}(\Theta_{k,n}) = \frac{1}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{(n-i)^2 p_{k,n}^2(i)} = \frac{1}{\lambda^2} \sum_{i=1}^{n-k} \frac{1}{(n-i) p_{k,n}^2(i)} + \frac{1}{\lambda^2} \sum_{i=1}^{k-1} \frac{1}{i^2}. \tag{10}$$

As we did for the expected rumor spreading time and introducing the notation

$$\delta_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)^2 p_{k,n}^2(i)},$$

we obtain

$$\eta_n \leq \delta_n \leq \zeta_n, \tag{11}$$

where

$$\zeta_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)^2 \left[1 - \left(1 - \frac{i}{n} \right)^{k-1} \right]^2}, \; \eta_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)^2 \left[1 - \left(1 - \frac{i}{n-k} \right)^{k-1} \right]^2}.$$

In the following two lemma, we obtain the limiting value of both ζ_n and η_n when n tends to infinity.

Lemma 6. *For all $k \geq 2$, we have*

$$\lim_{n \longrightarrow \infty} \zeta_n = \frac{1}{(k-1)^2} \frac{\pi^2}{6} + \sum_{i=k}^{\infty} \frac{1}{i^2}.$$

Proof. By definition of function f_k, we have

$$\zeta_n = \sum_{i=1}^{n-k} \frac{f_k^2\left(\frac{i}{n}\right)}{(n-i)^2\left[1-\left(1-\frac{i}{n}\right)\right]^2} = \sum_{i=1}^{n-k} \frac{n^2}{i^2(n-i)^2} f_k^2\left(\frac{i}{n}\right). \tag{12}$$

Observing that

$$\frac{n^2}{i^2(n-i)^2} = \frac{1}{(n-i)^2} + \frac{1}{i^2} + \frac{2}{n}\left(\frac{1}{i} + \frac{1}{n-i}\right), \tag{13}$$

we have

$$\zeta_n = \sum_{i=1}^{n-k} \frac{1}{(n-i)^2} f_k^2\left(\frac{i}{n}\right) + \sum_{i=1}^{n-k} \frac{1}{i^2} f_k^2\left(\frac{i}{n}\right) + \frac{2}{n}\sum_{i=1}^{n-k}\left(\frac{1}{i} + \frac{1}{n-i}\right) f_k^2\left(\frac{i}{n}\right).$$

We denote respectively by $\zeta_{n,1}$, $\zeta_{n,2}$ and $\zeta_{n,3}$ these three sums. Concerning $\zeta_{n,3}$, we have

$$\sum_{i=1}^{n-k}\left(\frac{1}{i} + \frac{1}{n-i}\right) f_k^2\left(\frac{i}{n}\right) = \sum_{i=1}^{n-k} \frac{1}{i} f_k^2\left(\frac{i}{n}\right) + \sum_{i=k}^{n-1} \frac{1}{i} f_k^2\left(1-\frac{i}{n}\right).$$

Applying twice Relation (3), once with $g(x) = f_k^2(x)$ and once again with $g(x) = f_k^2(1-x)$, we obtain

$$\sum_{i=1}^{n-k} \frac{1}{i} f_k^2\left(\frac{i}{n}\right) + \sum_{i=k}^{n-1} \frac{1}{i} f_k^2\left(1-\frac{i}{n}\right) \underset{n\longrightarrow\infty}{\sim} \left(f_k^2(0) + f_k^2(1)\right)\ln(n) = \left(1 + \frac{1}{(k-1)^2}\right)\ln(n)$$

which implies that

$$\lim_{n\longrightarrow\infty} \zeta_{n,3} = \lim_{n\longrightarrow\infty} 2\left(1 + \frac{1}{(k-1)^2}\right)\frac{\ln(n)}{n} = 0.$$

Concerning $\zeta_{n,2}$, since we have

$$\zeta_{n,2} = \sum_{i=1}^{n-k} \frac{1}{i^2} f_k^2\left(\frac{i}{n}\right),$$

applying Lemma 5 with function $g(x) = f_k^2(x)$, we obtain

$$\lim_{n\longrightarrow\infty} \zeta_{n,2} = f_k^2(0)\frac{\pi^2}{6} = \frac{1}{(k-1)^2}\frac{\pi^2}{6}.$$

Finally, for term $\zeta_{n,1}$, we have

$$\zeta_{n,1} = \sum_{i=1}^{n-k} \frac{1}{(n-i)^2} f_k^2\left(\frac{i}{n}\right) = \sum_{i=k}^{n-1} \frac{1}{i^2} f_k^2\left(1-\frac{i}{n}\right)$$

Applying again Lemma 5 with function $g(x) = f_k^2(1-x)$, we obtain

$$\lim_{n\longrightarrow\infty} \zeta_{n,1} = f_k^2(1) \sum_{i=k}^{\infty} \frac{1}{i^2} = \sum_{i=k}^{\infty} \frac{1}{i^2}.$$

Putting these three limits together leads to

$$\lim_{n\longrightarrow\infty} \zeta_n = \frac{1}{(k-1)^2}\frac{\pi^2}{6} + \sum_{i=k}^{\infty} \frac{1}{i^2},$$

which completes the proof. □

We analyze now the limiting value of η_n when n tends to infinity.

Lemma 7. *For all $k \geq 2$, we have*

$$\lim_{n\longrightarrow\infty} \eta_n = \frac{1}{(k-1)^2}\frac{\pi^2}{6} + \sum_{i=k}^{\infty} \frac{1}{i^2}.$$

Proof. Using again the function f_k, we obtain

$$\eta_n = \sum_{i=1}^{n-k} \frac{f_k^2\left(\frac{i}{n-k}\right)}{(n-i)^2\left[1-\left(1-\frac{i}{n-k}\right)\right]^2} = \frac{(n-k)^2}{n^2} \sum_{i=1}^{n-k} \frac{n^2}{i^2(n-i)^2} f_k^2\left(\frac{i}{n-k}\right). \quad (14)$$

The function f_k being Lipschitz on interval $[0,1]$, defining $C_k = \max_{z\in[0,1]} |(f_k^2)'(z)|$ where $(f_k^2)'$ is the derivative of f_k^2, we get, using Relations (12) and (14),

$$\begin{aligned}
\left|\eta_n - \frac{(n-k)^2}{n^2}\zeta_n\right| &\leq \frac{(n-k)^2}{n^2} \sum_{i=1}^{n-k} \frac{n^2}{i^2(n-i)^2} \left|f_k^2\left(\frac{i}{n-k}\right) - f_k\left(\frac{i}{n}\right)\right| \\
&\leq C_k \frac{(n-k)^2}{n^2} \sum_{i=1}^{n-k} \frac{n^2}{i^2(n-i)^2} \frac{ik}{n(n-k)} \\
&\leq C_k \frac{k(n-k)}{n^2} \sum_{i=1}^{n-k} \frac{n^2}{i^2(n-i)^2}.
\end{aligned}$$

Using Relation (13), we have

$$\lim_{n\longrightarrow\infty}\sum_{i=1}^{n-k}\frac{n^2}{i^2(n-i)^2}=\sum_{i=1}^{\infty}\frac{1}{i^2}+\sum_{i=k}^{\infty}\frac{1}{i^2}=\frac{\pi^2}{3}-\sum_{i=1}^{k-1}\frac{1}{i^2}.$$

It follows that

$$\lim_{n\longrightarrow\infty}\left|\eta_n-\frac{(n-k)^2}{n^2}\zeta_n\right|=0.$$

This result together with the result of Lemma 6 leads to

$$\lim_{n\longrightarrow\infty}\eta_n=\lim_{n\longrightarrow\infty}\zeta_n=\frac{1}{(k-1)^2}\frac{\pi^2}{6}+\sum_{i=k}^{\infty}\frac{1}{i^2},$$

which completes the proof. □

Theorem 8. *For all $k\geq 2$, we have*

$$\lim_{n\longrightarrow\infty}\mathbb{V}(\Theta_{k,n})=\left(1+\frac{1}{(k-1)^2}\right)\frac{\pi^2}{6\lambda^2}.$$

Proof. Relations (10) and (11) give

$$\frac{\eta_n}{\lambda^2}+\frac{1}{\lambda^2}\sum_{i=1}^{k-1}\frac{1}{i^2}\leq\mathbb{V}(\Theta_{k,n})\leq\frac{\zeta_n}{\lambda^2}+\frac{1}{\lambda^2}\sum_{i=1}^{k-1}\frac{1}{i^2}.$$

The use of Lemma 6 and Lemma 7 leads to the desired result. □

We illustrate this result through Table 2, which gives the variance of the rumor spreading time $\mathbb{V}(\Theta_{k,n})$ when $\lambda=1$, for different values of k and n and its limiting value given by Theorem 8. For that purpose, we denote this limiting value by V_k, that is

$$V_k=\left(1+\frac{1}{(k-1)^2}\right)\frac{\pi^2}{6\lambda^2}.$$

We observe that $\mathbb{V}\left(\Theta_{k,n}\right)$ is very close to its limiting value V_k. Observe also that $\mathbb{V}\left(\Theta_{k,n}\right)$ decreases with k.

4 Distribution of the rumor spreading time

This section provides explicit limiting distributions such as $\Theta_{k,n}-\mathbb{E}(\Theta_{k,n})$ when n tends to infinity. We introduce the notation $\mu_{k,n}(i)=\lambda(n-i)p_{k,n}(i)$. Recall that $U_{k,n}(i)$ is exponentially distributed with rate $\mu_{k,n}(i)$ and that

$$\Theta_{k,n}=\sum_{i=1}^{n-1}U_{k,n}(i).$$

k	$\mathbb{V}\left(\Theta_{k,10^3}\right)$	$\mathbb{V}\left(\Theta_{k,10^4}\right)$	$\mathbb{V}\left(\Theta_{k,10^5}\right)$	V_k
2	3.3111727246	3.2929244424	3.2902659327	3.2898681337
3	2.0598946906	2.0567171079	2.0562398628	2.0561675836
4	1.8290154857	1.8279392316	1.8277382394	1.8277045187
5	1.7481896847	1.7478599785	1.7477614057	1.7477424460
6	1.7107521018	1.7107894334	1.7107427649	1.7107314295
7	1.6903987592	1.6906492342	1.6906334201	1.6906266798
8	1.6781158082	1.6785034176	1.6785078427	1.6785041498
9	1.6701368648	1.6706190746	1.6706376972	1.6706361616
10	1.6646618989	1.6652127502	1.6652418289	1.6652418948

Table 2 Values of $\mathbb{V}\left(\Theta_{k,n}\right)$ and V_k for $k = 2, \ldots, 10$, $\lambda = 1$ and $n = 10^3, 10^4, 10^5$.

The main result of this section is Theorem 10 whose proof needs the following lemma.

Lemma 9. *For all $k \geq 2$, we have*

$$\lim_{m \longrightarrow \infty} \limsup_{n \longrightarrow \infty} \sum_{i=m}^{n} \frac{1}{(2n+1-i)^2 p_{k,2n+1}^2(i)} = 0.$$

and

$$\lim_{m \longrightarrow \infty} \limsup_{n \longrightarrow \infty} \sum_{i=m}^{n} \frac{1}{i^2 p_{k,2n+1}^2(2n+1-i)} = 0.$$

Proof. Introducing the notation

$$\Delta_{m,n}(k) = \sum_{i=m}^{n} \frac{1}{(2n+1-i)^2 \left[1 - \left(1 - \frac{i}{2n+1}\right)^{k-1}\right]^2}$$

we obtain, using inequality (5),

$$\sum_{i=m}^{n} \frac{1}{(2n+1-i)^2 p_{k,2n+1}^2(i)} \leq \Delta_{m,n}(k).$$

By definition of function f_k we write

$$\Delta_{m,n}(k) = \sum_{i=m}^{n} \frac{(2n+1)^2}{(2n+1-i)^2 i^2} f_k^2\left(\frac{i}{2n+1}\right).$$

Observing that function f_k is increasing on interval $[0,1]$ and that $f_k(1) = 1$, we obtain using Relation (13), with $2n+1$ instead n,

$$\Delta_{m,n}(k) \leq \sum_{i=m}^{n} \frac{(2n+1)^2}{(2n+1-i)^2 i^2}$$

$$
\begin{aligned}
&= \sum_{i=m}^{n} \frac{1}{(2n+1-i)^2} + \sum_{i=m}^{n} \frac{1}{i^2} + \frac{2}{2n+1} \sum_{i=m}^{n} \left(\frac{1}{i} + \frac{1}{2n+1-i} \right) \\
&= \sum_{i=n+1}^{2n+1-m} \frac{1}{i^2} + \sum_{i=m}^{n} \frac{1}{i^2} + \frac{2}{2n+1} \left(\sum_{i=m}^{n} \frac{1}{i} + \sum_{i=n+1}^{2n+1-m} \frac{1}{i} \right) \\
&= \sum_{i=m}^{2n+1-m} \frac{1}{i^2} + \frac{2}{2n+1} \left(\sum_{i=m}^{2n+1-m} \frac{1}{i} \right) \leq \sum_{i=m}^{2n+1-m} \frac{1}{i^2} + \frac{2(1+\ln(2n+1-m)}{2n+1}.
\end{aligned}
$$

The $\lim_{m \longrightarrow \infty} \limsup_{n \longrightarrow \infty}$ of both terms is 0 because $\sum_i 1/i^2$ is a converging series. This proves the first relation.

Concerning the second relation, introducing the notation

$$
\Lambda_{m,n}(k) = \sum_{i=m}^{n} \frac{1}{i^2 \left[1 - \left(1 - \frac{2n+1-i}{2n+1} \right)^{k-1} \right]^2}
$$

we obtain in the same way, using inequality (5),

$$
\sum_{i=m}^{n} \frac{1}{i^2 p_{k,2n+1}^2 (2n+1-i)} \leq \Lambda_{m,n}(k)
$$

and, using function f_k,

$$
\Lambda_{m,n}(k) = \sum_{i=m}^{n} \frac{(2n+1)^2}{(2n+1-i)^2 i^2} f_k^2 \left(1 - \frac{i}{2n+1} \right).
$$

As we did for term $\Delta_{m,n}(k)$, we have

$$
\Lambda_{m,n}(k) \leq \sum_{i=m}^{n} \frac{(2n+1)^2}{(2n+1-i)^2 i^2},
$$

which in turn leads to the same result. □

We are now able to prove the following theorem.

Theorem 10. *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*

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

We then have

$$\Theta_{k,n} - \mathbb{E}(\Theta_{k,n}) \xrightarrow{\mathcal{L}} \frac{1}{(k-1)\lambda} W^{(1)} + \frac{1}{\lambda} W^{(2)} \text{ as } n \longrightarrow \infty$$

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

Proof. Recalling that $\mu_{k,n}(i) = \lambda(n-i)p_{k,n}(i)$, we have, for all $n \geq 2$, $i \in \{1, \ldots, n-1\}$ and $x \geq 0$,

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

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

Observing now that for each fixed i and k, we have, from Relation (1),

$$p_{k,n}(i) = i \sum_{h=1}^{k-1} \frac{1}{n-h} + o(1/n),$$

we obtain

$$\lim_{n\longrightarrow\infty} \mu_{k,n}(i) = \lambda \lim_{n\longrightarrow\infty} (n-i)p_{k,n}(i) = (k-1)\lambda i.$$

From Relation (1), we also trivially get

$$\lim_{n\longrightarrow\infty} \mu_{k,n}(n-i) = \lambda i \lim_{n\longrightarrow\infty} p_{k,n}(n-i) = \lambda i.$$

Defining $R_{k,n}(i) = U_{k,n}(i) - \mathbb{E}(U_{k,n}(i))$ we obtain, since $\mathbb{E}(U_{k,n}(i)) = 1/\mu_{k,n}(i)$,

$$R_{k,n}(i) = U_{k,n}(i) - \mathbb{E}(U_{k,n}(i)) = \frac{\mu_{k,n}(i)U_{k,n}(i) - 1}{\mu_{k,n}(i)} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{(k-1)\lambda i} \text{ as } n \longrightarrow \infty. \quad (15)$$

In the same way, we get

$$R_{k,n}(n-i) = \frac{\mu_{k,n}(n-i)U_{k,n}(n-i) - 1}{\mu_{k,n}(n-i)} \xrightarrow{\mathcal{L}} \frac{Z_i - 1}{\lambda i} \text{ as } n \longrightarrow \infty. \quad (16)$$

Suppose that n is odd, *i.e.*, that $n = 2\ell + 1$. Defining

$$V_{k,\ell} = \sum_{i=1}^{\ell} R_{k,2\ell+1}(i) \text{ and } \overline{V}_{k,\ell} = \sum_{i=1}^{\ell} R_{k,2\ell+1}(2\ell + 1 - i),$$

we have

$$\Theta_{k,2\ell+1} - \mathbb{E}(\Theta_{k,2\ell+1}) = \sum_{i=1}^{2\ell} [U_{k,2\ell+1}(i) - \mathbb{E}(U_{k,2\ell+1}(i))]$$

$$= \sum_{i=1}^{2\ell} R_{k,2\ell+1}(i) = V_{k,\ell} + \overline{V}_{k,\ell}. \tag{17}$$

Observe that the random variables $V_{k,\ell}$ and $\overline{V}_{k,\ell}$ are independent.

The rest of the proof consists in checking the hypothesis of the principle of accompanying laws of Theorem 9.1.13 of [25]. We introduce the notation

$$V_{k,\ell,m} = \sum_{i=1}^{m-1} R_{k,2\ell+1}(i) \text{ and } \overline{V}_{k,\ell,m} = \sum_{i=1}^{m-1} R_{k,2\ell+1}(2\ell+1-i).$$

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

$$\begin{aligned}
\mathbb{E}\left((V_{k,\ell} - V_{k,\ell,m})^2\right) &= \mathbb{E}\left(\left[\sum_{i=m}^{\ell} R_{k,2\ell+1}(i)\right]^2\right) = \mathbb{V}\left(\sum_{i=m}^{\ell} R_{k,2\ell+1}(i)\right) \\
&= \sum_{i=m}^{\ell} \mathbb{V}(R_{k,2\ell+1}(i)) = \sum_{i=m}^{\ell} \mathbb{V}(U_{k,2\ell+1}(i)) = \sum_{i=m}^{\ell} \frac{1}{\mu_{k,2\ell+1}^2(i)} \\
&= \frac{1}{\lambda^2} \sum_{i=m}^{\ell} \frac{1}{(2\ell+1-i)^2 p_{k,2\ell+1}^2(i)}
\end{aligned}$$

and, in the same way,

$$\mathbb{E}\left((\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m})^2\right) = \sum_{i=m}^{\ell} \frac{1}{\mu_{k,2\ell+1}^2(2\ell+1-i)} = \frac{1}{\lambda^2} \sum_{i=m}^{\ell} \frac{1}{i^2 p_{k,2\ell+1}^2(2\ell+1-i)}$$

Using Lemma 9, we have

$$\lim_{m\longrightarrow\infty} \limsup_{\ell\longrightarrow\infty} \mathbb{E}((V_{k,\ell} - V_{k,\ell,m})^2) = \lim_{m\longrightarrow\infty} \limsup_{\ell\longrightarrow\infty} \mathbb{E}((\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m})^2) = 0.$$

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

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

and

$$\mathbb{P}\{\left|\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m}\right| \geq \varepsilon\} = \mathbb{P}\{(\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m})^2 \geq \varepsilon^2\} \leq \frac{\mathbb{E}((\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m})^2)}{\varepsilon^2}.$$

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

$$\lim_{m\longrightarrow\infty} \limsup_{\ell\longrightarrow\infty} \mathbb{P}\{|V_{k,\ell} - V_{k,\ell,m}| \geq \varepsilon\} = \lim_{m\longrightarrow\infty} \limsup_{\ell\longrightarrow\infty} \mathbb{P}\{\left|\overline{V}_{k,\ell} - \overline{V}_{k,\ell,m}\right| \geq \varepsilon\} = 0. \tag{18}$$

Let us introduce the notation

$$W_{k,m} = \frac{1}{(k-1)\lambda} \sum_{i=1}^{m-1} \frac{Z_i - 1}{i} \text{ and } \overline{W}_{k,m} = \frac{1}{\lambda} \sum_{i=1}^{m-1} \frac{Z_i - 1}{i}.$$

Using (15) and (16) and the fact that the $R_{k,n}(i)$ are independent, we have

$$V_{k,\ell,m} \xrightarrow{\mathcal{L}} W_{k,m} \text{ and } \overline{V}_{k,\ell,m} \xrightarrow{\mathcal{L}} \overline{W}_{k,m} \text{ as } \ell \longrightarrow \infty. \tag{19}$$

The hypothesis of the principle of accompanying laws of Theorem 9.1.13 of [25] are properties (18) and (19). We can thus conclude that

$$V_{k,\ell} \xrightarrow{\mathcal{L}} \frac{1}{(k-1)\lambda} W \text{ and } \overline{V}_{k,\ell} \xrightarrow{\mathcal{L}} \frac{1}{\lambda} W \text{ as } \ell \longrightarrow \infty.$$

This means, from Relation (17), that

$$\Theta_{k,2\ell+1} - \mathbb{E}(\Theta_{k,2\ell+1}) \xrightarrow{\mathcal{L}} \frac{1}{(k-1)\lambda} W^{(1)} + \frac{1}{\lambda} W^{(2)} \text{ as } \ell \longrightarrow \infty,$$

where $W^{(1)}$ and $W^{(2)}$ are independent and identically distributed as W. The same reasoning applies in the case where $n = 2\ell$. □

Corollary 11. *For all $x \in \mathbb{R}$ and $k \geq 2$, we have*

$$\lim_{n \longrightarrow \infty} \mathbb{P}\left\{\Theta_{k,n} - \mathbb{E}(\Theta_{k,n}) \leq x\right\} = \int_0^\infty \exp\left(-t - t^{-(k-1)} e^{-(k-1)\lambda x - k\gamma}\right) dt.$$

Proof. L. Gordon has proved in [26] 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. Thus, by definition of W in Theorem 10, we have

$$W \overset{\mathcal{L}}{=} -\gamma - \ln(Z_1).$$

Introducing $W^{(1)} \overset{\mathcal{L}}{=} -\gamma - \ln(Z_1)$ and $W^{(2)} \overset{\mathcal{L}}{=} -\gamma - \ln(Z_2)$, we obtain from Theorem 10, for all $x \in \mathbb{R}$,

$$\begin{aligned}\lim_{n \longrightarrow \infty} \mathbb{P}\left\{\Theta_{k,n} - \mathbb{E}(\Theta_{k,n}) > x\right\} &= \mathbb{P}\left\{\frac{1}{(k-1)\lambda} W^{(1)} + \frac{1}{\lambda} W^{(2)} > x\right\} \\ &= \mathbb{P}\left\{-\frac{1}{\lambda}\left(\frac{k\gamma}{k-1} + \frac{\ln(Z_1)}{k-1} + \ln(Z_2)\right) \leq x\right\}\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left\{\frac{\ln(Z_1)}{k-1} + \ln(Z_2) \geq -\lambda x - \frac{k\gamma}{k-1}\right\} \\
&= \mathbb{P}\left\{Z_1^{1/(k-1)} Z_2 \geq e^{-\lambda x - k\gamma/(k-1)}\right\} \\
&= \int_0^\infty \mathbb{P}\left\{Z_1 \geq t^{-(k-1)} e^{-(k-1)\lambda x - k\gamma}\right\} e^{-t} dt \\
&= \int_0^\infty \exp\left(-t - t^{-(k-1)} e^{-(k-1)\lambda x - k\gamma}\right) dt,
\end{aligned}
$$

which completes the proof. □

Corollary 12. *For all $x \in \mathbb{R}$ and $k \geq 2$, we have*

$$
\lim_{n \longrightarrow \infty} \mathbb{P}\left\{\Theta_{k,n} - \frac{k\ln(n)}{(k-1)\lambda} \leq x\right\} = \int_0^\infty \exp\left(-t - \frac{t^{-(k-1)} e^{-(k-1)\lambda x}}{k-1}\right) dt
$$

Proof. Observe first that we have

$$
\mathbb{P}\left\{\Theta_{k,n} - \frac{k\ln(n)}{(k-1)\lambda} \leq x\right\} = \mathbb{P}\left\{\Theta_{k,n} - \mathbb{E}(\Theta_{k,n}) \leq x + \frac{k\ln(n)}{(k-1)\lambda} - \mathbb{E}(\Theta_{k,n})\right\}.
$$

From Theorem 4, we have

$$
\lim_{n \longrightarrow \infty} \left(\frac{k\ln(n)}{(k-1)\lambda} - \mathbb{E}(\Theta_{k,n})\right) = -\frac{\gamma k - \ln(k-1)}{(k-1)\lambda}.
$$

It follows from Corollary 11 that

$$
\lim_{n \longrightarrow \infty} \mathbb{P}\left\{\Theta_{k,n} - \frac{k\ln(n)}{(k-1)\lambda} \leq x\right\} = \int_0^\infty \exp\left(-t - \frac{t^{-(k-1)} e^{-(k-1)\lambda x}}{k-1}\right) dt.
$$

which completes the proof. □

We illustrate the result of Corollary 11 using simulations experiments. We introduce the notation

$$
F_k(x) = \int_0^\infty \exp\left(-t - t^{-(k-1)} e^{-(k-1)\lambda x - k\gamma}\right) dt
$$

We plot in Figure 1 the cumulative distribution function $\mathbb{P}\left\{\Theta_{k,n} - \mathbb{E}(\Theta_{k,n}) \leq x\right\}$ for the values $k = 2$ and $k = 10$ when $n = 10^4$ nodes using a sample of 10^4 values resulting from the distribution of $\Theta_{k,n}$. We also plot in Figure 1 the corresponding limits $F_2(x)$ and $F_{10}(x)$. We observe that the limiting distribution $F_k(x)$ is very close to the simulation results which tells us that the convergence speed seems to be quite good. Indeed, the maximal absolute value of the difference observed between $\mathbb{P}\left\{\Theta_{k,10^4} - \mathbb{E}(\Theta_{k,10^4}) \leq x\right\}$ and $F_k(x)$ is less than 0.006.

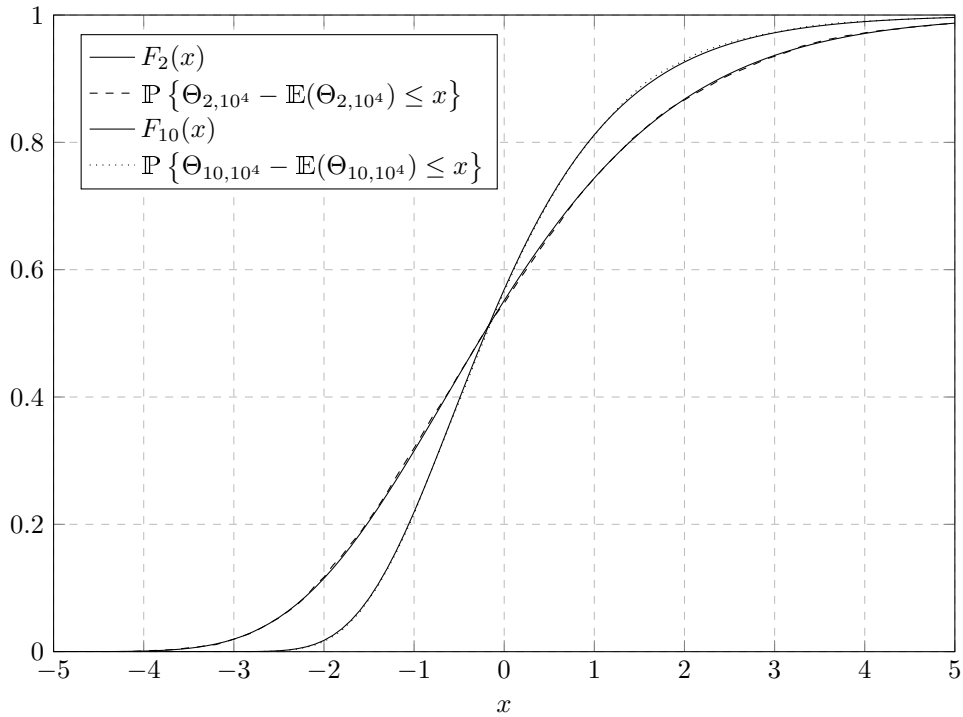

Fig. 1 Cumulative distribution function of $\Theta_{k,n} - \mathbb{E}(\Theta_{k,n})$ for $k = 2$, $k = 10$ and $n = 10^4$ and their limiting distribution when $n \longrightarrow \infty$

5 Conclusion

We considered the asynchronous k-pull protocol in continuous time and analyzed the total amount of time needed for all the nodes to know a rumor that a single node possesses initially. We obtained asymptotic values of both the expectation and the variance of this total amount of time. We also provided explicit limiting results concerning its distribution. All this work has been done by considering that the network of nodes is a complete network which means to each node may be asked for the rumor. Clearly, this analysis will not hold for more general networks and this is one of our objectives for future works.

6 Declarations

Funding : There has been no support for this work.

Conflicts of interest/Competing interests : there are neither Conflicts of interest nor Competing interests for this work.

Availability of data and material : Not applicable.

Authors' contributions : The four authors, who are François Castella, Bruno Sericola, Emmanuelle Anceaume and Yves Mocquard, contributed equally to this work.

References

[1] Demers, A.J., Greene, D.H., Hauser, C.H., Irish, W., Larson, J., Shenker, S.J., Sturgis, H.E., Swinehart, D.C., Terry, D.B.: Epidemic algorithms for replicated database maintenance. PODC 87, 1–12 (1987)

[2] Doerr, B., Kostrygin, A.: Randomized rumor spreading revisited. In: Chatzigiannakis, I., Indyk, P., Kuhn, F., Muscholl, A. (eds.) ICALP 2017, vol. 80, pp. 1–14 (2017)

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

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

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

[6] Giakkoupis, G.: 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)

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

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

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

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

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

[12] Clementi, A., Crescenzi, P., Doerr, C., Fraigniaud, P., Pasquale, F., Silvestri, R.: Rumor spreading in random evolving graphs. Random structures and Algorithms

48(2), 290–312 (2015)

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

[14] Panagiotou, K., Pourmiri, A., Sauerwald, T.: Faster rumor spreading with multiple calls. The electronic journal of combinatorics **22** (2015)

[15] Giakkoupis, G., Nazari, Y., Woelfel, P.: How asynchrony affects rumor spreading time. In: Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pp. 185–194 (2016)

[16] Mocquard, Y., Robert, S., Sericola, B., Anceaume, E.: Analysis of the propagation time of a rumour in large-scale distributed systems. In: Proceedings of the IEEE International Symposium on Network Computing and Applications (NCA) (2016)

[17] Acan, H., Collevecchio, A., Mehrabian, A., Wormald, N.: On the push & pull protocol for rumour spreading. Trends in Mathematics **6**, 3–10 (2017)

[18] Doerr, B., Fouz, M., Friedrich, T.: Experimental analysis of rumor spreading in social networks. MedAlg 2012, 159–173 (2012)

[19] Pourmiri, A., Ramezani, F.: Brief announcement: Ultra-fast asynchronous randomized rumor spreading. SPAA 2019 (2019)

[20] Daley, D.J., Gani, J.: Epidemic Modelling: An Introduction. Cambridge Studies in Mathematical Biology. Cambridge University Press, Cambridge (1999)

[21] Sanghavi, S., Hajek, B., Massoulié, L.: Gossiping with multiple messages. IEEE Transactions on Information Theory **53**(123) (2007)

[22] Yao, G., Bi, J., Wang, S., Zhang, Y., Li, Y.: A pull model IPv6 duplicate address detection. LCN 2010 (2010)

[23] Defants: The Defants vSIRT platform. https://www.defants.com/en/

[24] Robin, F., Sericola, B., Anceaume, E., Mocquard, Y.: Stochastic analysis of rumor spreading with multiple pull operations. Methodology and Computing in Applied Probability Journal (2021)

[25] Stroock, D.W.: Probability Theory: An Analytic View, 2nd edn. Cambridge University Press, Cambridge (2010)

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