HAL open science

Stochastic analysis of rumor spreading with multiple pull operations in presence of non-cooperative nodes

Sébastien Kilian, Emmanuelle Anceaume, Bruno Sericola

To cite this version:

Sébastien Kilian, Emmanuelle Anceaume, Bruno Sericola. Stochastic analysis of rumor spreading with multiple pull operations in presence of non-cooperative nodes. ASMTA 2023 - 27th International Conference on Analytical & Stochastic Modelling Techniques & Applications, Jun 2023, Florence, Italy. pp.1-15. hal-04166945

HAL Id: hal-04166945

https://cnrs.hal.science/hal-04166945v1

Submitted on 20 Jul 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.

Distributed under a Creative Commons Attribution 4.0 International License

Stochastic analysis of rumor spreading with multiple pull operations in presence of non-cooperative nodes

Sébastien Kilian[1], Emmanuelle Anceaume[2], and Bruno Sericola[1]

[1] Centre Inria de l'Univ. de Rennes, Irisa
[2] CNRS, Univ. Rennes, Irisa

Abstract. The recent rise of interest in distributed applications has highlighted the importance of effective information dissemination. The challenge lies in the fact that nodes in a distributed system are not necessarily synchronized, and may fail at any time. This has led to the emergence of randomized rumor spreading protocols, such as push and pull protocols, which have been studied extensively. The k-pull operation, which allows an uninformed node to ask for the rumor from a fixed number of other nodes in parallel, has been proposed to improve the pull algorithm's effectiveness. This paper presents and studies the performance of the k-pull operation in the presence of a certain fraction f of non-cooperative nodes. Our goal is to understand the impact of k on the propagation of the rumor despite the presence of a fraction f of non-collaborative nodes.

1 Introduction

The dissemination of information in distributed systems has been an active area of research in recent years. With the rise of distributed applications, efficient and robust methods for information dissemination have become increasingly important. In a distributed system, the nodes are not necessarily synchronized, and can fail at any time. This makes the dissemination of information a challenging problem. This problem, often called *rumor spreading* or *gossip spreading*, is the process of sending a message to all the nodes in a network [5,7]. The nodes in the network are anonymous, meaning that they can not be designated in advance, and any two nodes cannot tell whether they already interacted together or not. Different variants of randomized rumor spreading protocols have been studied. The push protocol provides a single operation, called the push operation, that allows an informed node to contact some randomly chosen node and sends it the rumor. The pull protocol, on the other hand allows, through the pull operation, an uninformed node to contact some random node to ask for the rumor. The same node can perform both operations according to whether it knows or not the rumor, which corresponds to the push-pull protocol.

One of the important questions raised by these randomized rumor spreading protocols is the spreading time, that is the time needed for all the nodes of the network to become informed.

Several models have been considered to answer this question. The synchronous 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 neighbors 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 this model, the spreading time of a rumor is defined as the number of synchronous rounds necessary for all the nodes to become informed. Analyses have been conducted when the underlying communication graph is complete (e.g., [10,11], and in different topologies (e.g., [2, 4, 9, 15]), in the presence of link or nodes failures as in [8], in dynamic graphs as in [3]. Another alternative consists for the nodes to make more than one call during the push-pull operations [16]. In large scale and open networks, assuming that all nodes act in synchrony is a strong assumption since it requires that all the nodes have access to some global synchronization mechanism and that message delays are upper bounded. Several authors, including [1, 6, 12, 14, 17], suppose that nodes asynchronously trigger operations with randomly chosen nodes. In this model, the spreading time of a rumor is defined as the number of operations necessary for all the nodes to know the rumor. In [17], the authors model a multiple call by tuning the clock rate of each node with a given probabilitity distribution.

Regarding the type of interaction, the pull algorithm has attracted very little attention because this algorithm was long considered inefficient to spread a rumor within a large scale network [19]. However, it is 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. In the line of Panagiotou et al.'s work [16], Robin et al [18] have extended the pull operation with the k-pull operation, which allows an uniformed node to ask for the rumor to a fixed number $k-1$ of other nodes in parallel.

The objective of this paper is to push further this line of inquiry by presenting and studying the performance of the k-pull operation in presence of a certain fraction f of non collaborative nodes. A non-cooperative node is a node that refuses to learn and thus to propagate the rumor. Our aim is to understand the impact of k on the propagation of the rumor despite the presence of a proportion f of non-collaborative nodes.

The remaining of the paper is organized as follows. Section 2 presents the k-pull protocol in presence of non-cooperative nodes. Section 3 analyses the rumor spreading time when $k = 2$. Section 4 presents a numerical study of the influence of larger values of k on the expected spreading time and on the distribution of the k-pull operation. Finally Section 5 presents future works.

2 The k-pull protocol in presence of non-cooperative nodes

We consider a system with interacting subpopulations of informed, uninformed, and non-cooperative nodes. The total number of nodes is equal to n. The communication graph among the n nodes is complete. Nodes are anonymous meaning that that they do not use identifiers and thus cannot determine whether any two interactions have occurred with the same nodes or not. However, for ease of presentation the nodes are numbered $1, 2, \ldots, n$. We assume a thoroughly mixed population, so that nodes encounter each other at random, with uniform probability. Initially, a single node knows the rumor and wishes to propagate it (this is not a non-cooperative node) to uninformed nodes. However, among uninformed nodes, a constant fraction f of them are non-cooperative, i.e., they do not want to learn the rumor and thus to propagate it further.

The k-pull protocol is defined as follows. At each discrete time t, a single uninformed node s contacts $k-1$ distinct nodes, chosen at random uniformly among the $n-1$ other nodes, and applies the following rule:

- If at least one of the $k-1$ contacted nodes knows the rumor and s is not non-cooperative then node s becomes informed.

Note that despite the fact that non-cooperative nodes do not want to learn the rumor, they trigger the k-pull operation. Their motivation is to increase the spreading time of the rumor. At any time we suppose that $nf \geq k$, otherwise we come back to Robin et al.'s analysis [18]. Note that in practice, nf is an integer, but this is not necessary for the analysis. The protocol halts by itself once all the $n(1-f)$ nodes are informed.

2.1 The k-pull protocol in absence of non-cooperative nodes

To analyze the k-pull protocol, the authors in [18] have introduced the discrete-time stochastic process $Y = \{Y_t,\ t \geq 0\}$ where Y_t represents the number of informed nodes at time t. Stochastic process Y is a discrete-time homogeneous Markov chain with n states where states $1, \ldots, n-1$ are transient and state n is absorbing. When the Markov chain Y is in state i at time t, then at time $t+1$, either it remains in state i if none of the $k-1$ chosen nodes know the rumor or it transits to state $i+1$ if at least one of the $k-1$ chosen nodes know the rumor. Let P be the transition probability matrix of Markov chain Y. The non zero entries of matrix P are thus $P_{i,i}$ and $P_{i,i+1}$, for any $i = 1, \ldots, n-1$. We denote by $T_{k,n}$ the random variable defined by

$$T_{k,n} := \inf\{t \geq 0 \mid Y_t = n\},$$

which represents the spreading time, that is the total number of k-pull operations needed for all the nodes in the network to know the rumor. The spreading time distribution can thus be expressed as a sum of independent random variables $S_{k,n}(i)$, where $S_{k,n}(i)$ is the sojourn time of Markov chain Y in state i. For all

$i = 1, \ldots, n-k$, $S_{k,n}(i)$ follows a geometric distribution with parameter $P_{i,i+1}$ which we denote more simply by $p_{k,n}(i)$. It is shown in [18] that

$$p_{k,n}(i) = 1 - P_{i,i} = 1 - \prod_{h=1}^{k-1} \left(1 - \frac{i}{n-h}\right). \tag{1}$$

2.2 The k-pull protocol in presence of nf non-cooperative nodes

We keep the same notation used in the previous subsection. We just suppose that when $f \neq 0$, we have $nf \geq k$. In presence of a proportion f of non-cooperative node, when $Y_t = i$, i.e. when i nodes are informed of the rumor at time t, we have $Y_{t+1} = i$ if and only if, at time $t+1$, either the selected node is a non-cooperative node (with probability $nf/(n-i)$) or the selected node is not a non-cooperative node (with probability $1 - nf/(n-i)$) and the set of $k-1$ chosen nodes (i.e. $k-1$ among $n-1$) must be chosen among the $n-1-i$ non informed nodes which corresponds to the situation where $f = 0$. More formally, if M_t denotes the status of the selected node at time t (1 if it is non-cooperative and 0 otherwise), we have, for every $t \geq 0$, using (1)

$$\begin{aligned}
P_{i,i} &= \mathbb{P}\{Y_{t+1} = i \mid Y_t = i\} \\
&= \mathbb{P}\{Y_{t+1} = i \mid M_t = 1, Y_t = i\}\mathbb{P}\{M_t = 1 \mid Y_t = i\} \\
&\quad + \mathbb{P}\{Y_{t+1} = i \mid M_t = 0, Y_t = i\}\mathbb{P}\{M_t = 0 \mid Y_t = i\} \\
&= \frac{nf}{n-i} + \left(1 - \frac{nf}{n-i}\right) \prod_{h=1}^{k-1} \left(1 - \frac{i}{n-h}\right)
\end{aligned}$$

Observe that Y has now $n(1-f)$ states where states $1, \ldots, n(1-f)-1$ are transient and state $n(1-f)$ is absorbing. In the same way, we introduce the notation

$$p_{k,n}(i) = 1 - P_{i,i} = \left(1 - \frac{nf}{n-i}\right) \left(1 - \prod_{h=1}^{k-1} \left(1 - \frac{i}{n-h}\right)\right) \tag{2}$$

and we have

$$T_{k,n} = \sum_{i=1}^{n(1-f)-1} S_{k,n}(i), \tag{3}$$

where $S_{k,n}(i)$ has a geometric distribution with parameter $p_{k,n}(i)$.

It is well-known, see for instance [20], that the distribution of $T_{k,n}$ is given, for every integer $t \geq 0$, by

$$\mathbb{P}\{T_{k,n} > t\} = \alpha Q^t \mathbb{1}, \tag{4}$$

where α is the row vector containing the initial probabilities of states $1, \ldots, n(1-f)-1$, that is $\alpha_i = \mathbb{P}\{Y_0 = i\} = 1_{\{i=1\}}$, Q is the matrix obtained from the transition matrix P by only keeping the transition probabilities between transient states, i.e. by removing the last line and the last column of P and $\mathbb{1}$ is the column vector of dimension $n(1-f)-1$ with all its entries equal to 1.

3 Analysis of the rumor spreading time when $k = 2$

In this section, we analyze the two first moments and the distribution of the rumor spreading time $T_{2,n}$ and their asymptotic behavior when n goes to infinity. We denote by H_n the harmonic series defined, for every $n \geq 1$, by $H_n = \sum_{i=1}^{n} 1/i$ and we recall that the Euler-Mascheroni constant γ is given by $\gamma = \lim_{n \longrightarrow \infty} (H_n - \ln(n))$, which is approximately equal to 0.5772156649.

3.1 Asymptotic mean and variance of the rumor spreading time

In the case where $k = 2$, we have from relation (2),

$$p_{2,n}(i) = \left(1 - \frac{nf}{n-i}\right) \frac{i}{n-1} = \frac{(n(1-f) - i)\, i}{(n-1)(n-i)}. \tag{5}$$

The asymptotic expected rumor spreading time is obtained in the following theorem.

Theorem 1.

$$\mathbb{E}(T_{2,n}) \underset{n \longrightarrow \infty}{\sim} \frac{(1+f)n\ln(n)}{1-f}$$

and

$$\lim_{n \longrightarrow \infty} \left(\frac{\mathbb{E}(T_{2,n})}{n} - \frac{(1+f)\ln(n)}{1-f} \right) = \frac{(1+f)\,(\gamma + \ln(1-f))}{1-f}.$$

Proof. The expected value of the spreading time $T_{2,n}$ is given, using Relation (3), by

$$\begin{aligned}
\mathbb{E}(T_{2,n}) &= \sum_{i=1}^{n(1-f)-1} \frac{1}{p_{2,n}(i)} = (n-1) \sum_{i=1}^{n(1-f)-1} \frac{n-i}{(n(1-f)-i)\, i} \\
&= (n-1) \left[\sum_{i=1}^{n(1-f)-1} \frac{1}{i} + nf \sum_{i=1}^{n(1-f)-1} \frac{1}{(n(1-f)-i)\, i} \right].
\end{aligned}$$

Observing that

$$\frac{1}{(n(1-f)-i)\, i} = \frac{1}{n(1-f)} \left(\frac{1}{i} + \frac{1}{n(1-f)-i} \right), \tag{6}$$

we obtain

$$\begin{aligned}
\mathbb{E}(T_{2,n}) &= (n-1) \left[\sum_{i=1}^{n(1-f)-1} \frac{1}{i} + \frac{2f}{1-f} \sum_{i=1}^{n(1-f)-1} \frac{1}{i} \right] \\
&= \frac{(1+f)(n-1)H_{n(1-f)-1}}{1-f} \\
&\underset{n \longrightarrow \infty}{\sim} \frac{(1+f)n\ln(n)}{1-f}.
\end{aligned} \tag{7}$$

Moreover, we have

$$\begin{aligned}
\frac{\mathbb{E}(T_{2,n})}{n} - \frac{(1+f)\ln(n)}{1-f} &= \frac{1+f}{1-f}\left[\frac{(n-1)H_{n(1-f)-1}}{n} - \ln(n)\right] \\
&= \frac{1+f}{1-f}\left[H_{n(1-f)-1} - \ln(n) - \frac{H_{n(1-f)-1}}{n}\right] \\
&= \frac{1+f}{1-f}\left[H_{n(1-f)-1} - \ln(n(1-f)-1)\right] \\
&\quad + \frac{1+f}{1-f}\left[\ln(n(1-f)-1) - \ln(n) - \frac{H_{n(1-f)-1}}{n}\right] \\
&= \frac{1+f}{1-f}\left[H_{n(1-f)-1} - \ln(n(1-f)-1)\right] \\
&\quad + \frac{1+f}{1-f}\left[\ln(1-f-1/n) - \frac{H_{n(1-f)-1}}{n}\right].
\end{aligned}$$

The second term in square brackets tends to $\ln(1-f)$ when n tends to infinity, so by definition of γ we have

$$\lim_{n\longrightarrow\infty}\left(\frac{\mathbb{E}(T_{2,n})}{n} - \frac{(1+f)\ln(n)}{1-f}\right) = \frac{(1+f)\left(\gamma + \ln(1-f)\right)}{1-f},$$

which completes the proof.

We consider now the variance of $T_{2,n}$ and its limiting value when n goes to infinity.

Theorem 2.

$$\mathbb{V}(T_{2,n}) \underset{n\longrightarrow\infty}{\sim} \frac{(1+f^2)\pi^2 n^2}{6(1-f)^2},$$

Proof. Using Relation (3), the variance of $T_{2,n}$ is given by

$$\begin{aligned}
\mathbb{V}(T_{2,n}) &= \sum_{i=1}^{n(1-f)-1} \frac{1-p_{2,n}(i)}{\left(p_{2,n}(i)\right)^2} \\
&= (n-1)^2 \sum_{i=1}^{n(1-f)-1} \left(\frac{n-i}{i\left(n(1-f)-i\right)}\right)^2 - \mathbb{E}(T_{2,n}). \qquad (8)
\end{aligned}$$

Using relation (6), we write

$$\begin{aligned}
\frac{n-i}{i\left(n(1-f)-i\right)} &= \frac{1}{i} + \frac{nf}{i\left(n(1-f)-i\right)} = \frac{1}{i} + \frac{f}{1-f}\left(\frac{1}{i} + \frac{1}{n(1-f)-i}\right) \\
&= \left(\frac{1}{1-f}\right)\frac{1}{i} + \left(\frac{f}{1-f}\right)\frac{1}{n(1-f)-i}
\end{aligned}$$

and thus

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

Using again relation (6) for the third term of this last expression, we get

$$\left(\frac{n-i}{i\,(n(1-f)-i)}\right)^2 = \frac{1}{(1-f)^2}\left(\frac{1}{i^2}+\frac{f^2}{(n(1-f)-i)^2}\right) + \frac{2f}{(1-f)^3 n}\left(\frac{1}{i}+\frac{1}{n(1-f)-i}\right).$$

It follows that

$$\sum_{i=1}^{n(1-f)-1}\left(\frac{n-i}{i\,(n(1-f)-i)}\right)^2 = \frac{1+f^2}{(1-f)^2}\sum_{i=1}^{n(1-f)-1}\frac{1}{i^2}+\frac{4f}{(1-f)^3 n}\sum_{i=1}^{n(1-f)-1}\frac{1}{i}.$$

Inserting this result in (8), we obtain using (7)

$$\mathbb{V}(T_{2,n}) = \frac{(1+f^2)(n-1)^2}{(1-f)^2}\sum_{i=1}^{n(1-f)-1}\frac{1}{i^2} + \left(\frac{4f(n-1)}{(1-f)^3 n}-\frac{(1+f)}{1-f}\right)(n-1)H_{n(1-f)-1}.$$

The second term of this sum is in $O(n\ln(n))$, thus

$$\mathbb{V}(T_{2,n}) \underset{n\longrightarrow\infty}{\sim} \frac{(1+f^2)\pi^2 n^2}{6(1-f)^2},$$

which completes the proof. ■

3.2 Asymptotic distribution of the rumor spreading time

This section provides the explicit limiting distribution of $(T_{2,n}-\mathbb{E}(T_{2,n}))/n$ when n tends to infinity. To prove the main result of this section, we need the following lemma.

Lemma 1.

$$\lim_{m\longrightarrow\infty}\limsup_{\ell\longrightarrow\infty}\frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\frac{1}{p^2_{2,(2\ell+1)/(1-f)}(i)} = 0.$$

and

$$\lim_{m\longrightarrow\infty}\limsup_{\ell\longrightarrow\infty}\frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\frac{1}{p^2_{k,(2\ell+1)/(1-f)}(2\ell+1-i)} = 0.$$

Proof. From (5), we have, by taking $n=(2\ell+1)/(1-f)$,

$$p_{2,(2\ell+1)/(1-f)}(i) = \frac{(2\ell+1-i)\,i}{\left(\frac{2\ell+1}{1-f}-1\right)\left(\frac{2\ell+1}{1-f}-i\right)}.$$

Introducing the notation

$$\Delta_{\ell,m} = \frac{(1-f)^2}{(2\ell+1)^2} \sum_{i=m}^{\ell} \frac{1}{p^2_{2,(2\ell+1)/(1-f)}(i)}$$

we obtain, after some algebra, and since $f \leq 1$,

$$\begin{aligned}
\Delta_{\ell,m} &= \frac{(2\ell+f)^2}{(1-f)^2(2\ell+1)^2} \sum_{i=m}^{\ell} \left(\frac{1}{i} + \frac{f}{2\ell+1-i} \right)^2 \\
&\leq \frac{1}{(1-f)^2} \sum_{i=m}^{\ell} \left(\frac{1}{i} + \frac{1}{2\ell+1-i} \right)^2 \qquad (9) \\
&= \frac{1}{(1-f)^2} \left(\sum_{i=m}^{2\ell+1-m} \frac{1}{i^2} + \frac{2}{2\ell+1} \sum_{i=m}^{2\ell+1-m} \frac{1}{i} \right) \\
&\leq \frac{1}{(1-f)^2} \left(\sum_{i=m}^{2\ell+1-m} \frac{1}{i^2} + \frac{2(1+\ln(2\ell+1-m)}{2\ell+1} \right).
\end{aligned}$$

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

Concerning the second one, from (5), we have, by taking $n = (2\ell+1)/(1-f)$,

$$p_{2,(2\ell+1)/(1-f)}(2\ell+1-i) = \frac{i\,(2\ell+1-i)}{\left(\frac{2\ell+1}{1-f} - 1\right)\left(\frac{2\ell+1}{1-f} - (2\ell+1-i)\right)}.$$

Introducing the notation

$$\Lambda_{\ell,m} = \frac{(1-f)^2}{(2\ell+1)^2} \sum_{i=m}^{\ell} \frac{1}{p^2_{k,(2\ell+1)/(1-f)}(2\ell+1-i)}$$

we obtain as we did for $\Delta_{\ell,m}$,

$$\begin{aligned}
\Lambda_{\ell,m} &= \frac{(2\ell+f)^2}{(1-f)^2(2\ell+1)^2} \sum_{i=m}^{\ell} \left(\frac{f}{i} + \frac{1}{2\ell+1-i} \right)^2 \\
&\leq \frac{1}{(1-f)^2} \sum_{i=m}^{\ell} \left(\frac{1}{i} + \frac{1}{2\ell+1-i} \right)^2,
\end{aligned}$$

which is exactly (9). This completes the proof. ■

We are now able to prove the following theorem, where $\xrightarrow{\mathcal{L}}$ means the convergence in law.

Theorem 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*

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

We then have

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

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

Proof. For a fixed value of i, we have $\lim_{n\longrightarrow\infty} p_{2,n}(i) = 0$. It follows that for every $x \geq 0$, we have

$$\mathbb{P}\{p_{2,n}(i)S_{k,n} > x\} = \mathbb{P}\{S_{2,n} > x/p_{2,n}(i) > x\} = (1 - p_{2,n}(i))^{\lfloor x/p_{k,n}(i) \rfloor}$$

which tends to e^{-x} when n tends to infinity, since $p_{2,n}(i)$ tends to 0. If Z_i is a random variable exponentially distributed with rate 1, we have shown that $p_{2,n}(i)S_{2,n} \xrightarrow{\mathcal{L}} Z_i$ when $n \longrightarrow \infty$. Moreover since the $(S_{2,n}(i))_{i=1,\ldots,n(1-f)-1}$ are independent, the $(Z_i)_{i=1,\ldots,n(1-f)-1}$ are also independent. In the same way, we have $\lim_{n\longrightarrow\infty} np_{2,n}(i) = (1-f)i$. Defining $R_{2,n}(i) = S_{2,n}(i) - \mathbb{E}\left(S_{2,n}(i)\right)$ we obtain, since $\mathbb{E}\left(S_{2,n}(i)\right) = 1/p_{2,n}(i)$,

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

In the same way, replacing i by $n(1-f) - i$ in (2) leads to

$$p_{2,n}\left(n(1-f) - i\right) = \frac{i\left(n(1-f) - i\right)}{(n-1)(nf+i)}.$$

It follows that

$$\lim_{n\longrightarrow\infty} p_{2,n}(n(1-f) - i) = 0 \text{ and } \lim_{n\longrightarrow\infty} np_{2,n}(n(1-f) - i) = \frac{i(1-f)}{f}$$

and thus

$$\frac{R_{2,n}(n(1-f) - i)}{n} = \frac{p_{2,n}(n(1-f) - i)S_{2,n}(n(1-f) - i) - 1}{np_{2,n}(n(1-f) - i)} \xrightarrow{\mathcal{L}} \frac{(Z_i - 1)f}{(1-f)i}. \tag{11}$$

Suppose that $n(1-f)$ is odd, i.e. $n(1-f) = 2\ell + 1$. We then have from (3)

$$\begin{aligned}
\frac{T_{2,n} - \mathbb{E}(T_{2,n})}{n} &= \frac{1-f}{2\ell+1}\left(\sum_{i=1}^{\ell} R_{2,n}(i) + \sum_{i=\ell+1}^{2\ell} R_{2,n}(i)\right) \\
&= \frac{1-f}{2\ell+1}\left(\sum_{i=1}^{\ell} R_{2,n}(i) + \sum_{i=1}^{\ell} R_{2,n}(2\ell+1-i)\right) \\
&= V_\ell + \overline{V}_\ell,
\end{aligned} \tag{12}$$

where

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

Observe that the random variables $V_{2,\ell}$ and $\overline{V}_{2,\ell}$ are independent. The rest of the proof consists in checking the hypothesis of the principle of accompanying laws of Theorem 3.1.14 of [24]. We introduce the notation

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

Using the fact that $\mathbb{E}(R_{2,n}(i)) = 0$ and the $R_{2,n}(i)$ are independent, we have

$$\begin{aligned}
&\mathbb{E}\left((V_\ell - V_{\ell,m})^2\right) = \mathbb{V}\left(\frac{1-f}{2\ell+1}\sum_{i=m}^{\ell} R_{2,n}(i)\right)\\
&= \frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\mathbb{V}(R_{2,n}(i)) = \frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\mathbb{V}(S_{2,n}(i))\\
&= \frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\frac{1-p_{2,(2\ell+1)/(1-f)}(i)}{p^2_{2,(2\ell+1)/(1-f)}(i)} \le \frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\frac{1}{p^2_{2,(2\ell+1)/(1-f)}(i)}
\end{aligned}$$

and, in the same way,

$$\mathbb{E}\left((\overline{V}_\ell - \overline{V}_{\ell,m})^2\right) \le \frac{(1-f)^2}{(2\ell+1)^2}\sum_{i=m}^{\ell}\frac{1}{p^2_{2,(2\ell+1)/(1-f)}(2\ell+1-i)}.$$

Using Lemma 1, we have

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

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

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

and

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

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

$$\lim_{m\longrightarrow\infty}\limsup_{\ell\longrightarrow\infty}\mathbb{P}\{|V_\ell - V_{\ell,m}| \ge \varepsilon\} = 0 \tag{13}$$

$$\lim_{m\longrightarrow\infty}\limsup_{\ell\longrightarrow\infty}\mathbb{P}\{\left|\overline{V}_\ell - \overline{V}_{\ell,m}\right| \ge \varepsilon\} = 0. \tag{14}$$

Let us introduce the notation

$$W_m = \frac{1}{1-f}\sum_{i=1}^{m-1}\frac{Z_i-1}{i} \text{ and } \overline{W}_m = \frac{f}{1-f}\sum_{i=1}^{m-1}\frac{Z_i-1}{i}.$$

Using (10) and (11) and the fact that the $R_{k,n}(i)$ are independent, we have

$$V_{\ell,m} \xrightarrow{\mathcal{L}} W_m \text{ and } \overline{V}_{\ell,m} \xrightarrow{\mathcal{L}} \overline{W}_m \text{ as } \ell \longrightarrow \infty. \tag{15}$$

The hypothesis of the principle of accompanying laws of Theorem 3.1.14 of [21] are properties (13), (14) and (15). We can thus conclude that

$$V_\ell \xrightarrow{\mathcal{L}} \frac{1}{1-f}W \text{ and } \overline{V}_\ell \xrightarrow{\mathcal{L}} \frac{f}{1-f}W \text{ as } \ell \longrightarrow \infty.$$

This means, from relation (12), that

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

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

Corollary 1. *For all $x \in \mathbb{R}$ we have*

$$\lim_{n\longrightarrow\infty} \mathbb{P}\left\{\frac{T_{2,n}-\mathbb{E}(T_{2,n})}{n} \le x\right\} = \int_0^\infty \exp\left(-t - t^{-f}e^{-(1-f)x-\gamma(1+f)}\right)dt.$$

Proof. L. Gordon has proved in [13] that

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

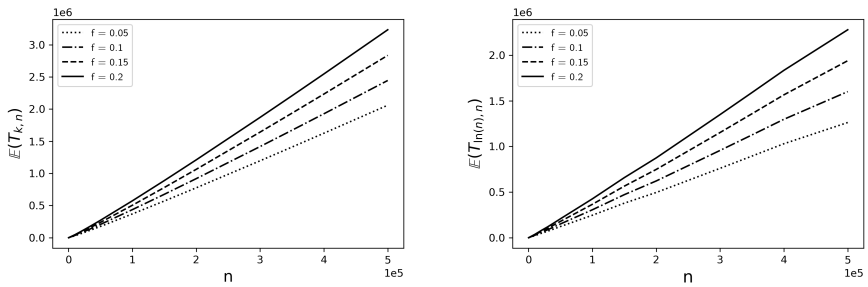
where (Z_i) are i.i.d. exponential with rate 1. Thus, by definition of W in Theorem 3, we have

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

Introducing $W^{(1)} \overset{\mathcal{L}}{=} -\gamma - \ln(Z_1)$ and $W^{(2)} \overset{\mathcal{L}}{=} -\gamma - \ln(Z_2)$, we obtain from Theorem 3, for all $x \in \mathbb{R}$,

$$\begin{aligned}
\lim_{n\longrightarrow\infty} \mathbb{P}\left\{\frac{T_{2,n}-\mathbb{E}(T_{2,n})}{n} \le x\right\} &= \mathbb{P}\left\{\frac{1}{1-f}W^{(1)} + \frac{f}{1-f}W^{(2)} \le x\right\} \\
&= \mathbb{P}\left\{-\ln(Z_1) - f\ln(Z_2) \le (1-f)x + \gamma(1+f)\right\} \\
&= \mathbb{P}\left\{Z_1 Z_2^f \ge e^{-(1-f)x-\gamma(1+f)}\right\} \\
&= \int_0^\infty \mathbb{P}\left\{Z_1 \ge t^{-f}e^{-(1-f)x-\gamma(1+f)}\right\}e^{-t}dt \\
&= \int_0^\infty \exp\left(-t - t^{-f}e^{-(1-f)x-\gamma(1+f)}\right)dt,
\end{aligned}$$

which completes the proof. ■



(a) $k = 5$. (b) $k = \ln(n)$.

Fig. 1: Expected spreading time $T_{k,n}$ as a function of the number n of nodes in the system for different values of f, when $k = 5$ on the left, and $k = \ln(n)$ on the right.

4 General case

Generalizing to higher values of k has shown to be more difficult. Indeed, as a node can interact with a parameterized number of other nodes, its behavior during an interaction becomes harder to predict. This can be observed in (1) where k has an influence on the number of terms in the product of the relation. Consequently, we numerically study the influence of larger values of k on the expected spreading time and on its distribution.

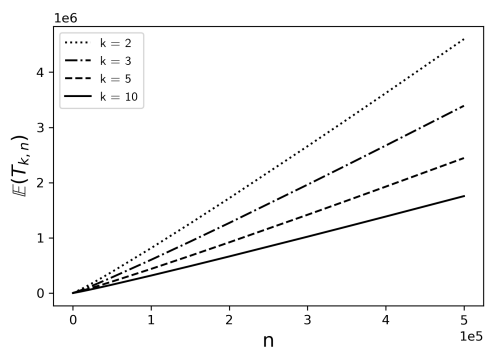
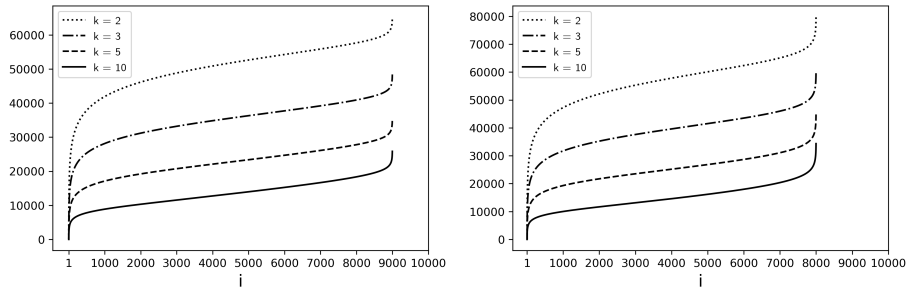


Fig. 2: Expected spreading time $T_{k,n}$ as a function of the number n of nodes in the system for different values of k when $f = 0.1$.

As expected, Figure 1a shows that the mean spreading time increases with the proportion f. A realistic assumption would be to have k as a function of n. Figure 1b shows that having k as a function of n does not provide significant



(a) $f = 0.1$. (b) $f = 0.2$.

Fig. 3: $\sum_{j=1}^{i} \mathbb{E}(T_{k,j})$: Expected spreading time to reach i informed nodes, as a function of i, with $n = 10,000$ nodes.

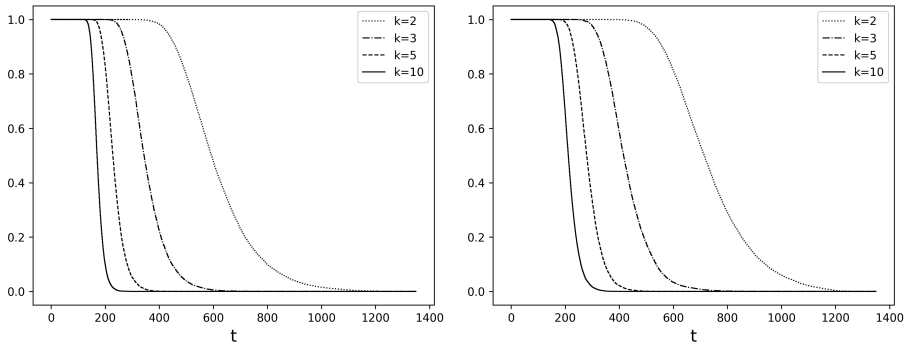
improvement for small values of n. More interestingly, k has a significant influence on the spreading time, independently from the size of the network, which is illustrated by Figure 2. Using a k-pull operation helps to mitigate the influence of non-cooperative node compared to a regular asynchronous pull protocol ($k = 2$).

One of the major caveats of the pull operation is that the diffusion of the rumor during the first interactions is very slow. This phenomenon is made worse with the presence of non-cooperative nodes. This is due to the fact that the few informed and cooperative nodes need to be selected to be able to propagate the rumor, as opposed to the push operation. This is highlighted in Figure 3a for $k = 2$, where we can see that the propagation is quite slow during the first interactions. On the other hand, when $k > 2$, the k-pull operation mitigates this problem as shown in Figures 3a and 3b for different proportions f of non-cooperative nodes.

Figure 4 shows the cumulative distribution of $T_{k,n}$ (see (4)). This distribution shows not only that higher values of k lead to faster spreading times, but also that this spreading time is more predictable. Indeed, the decrease of the function $\mathbb{P}\{T_{k,n} > t\}$ is significantly faster for higher values of k. Moreover, the influence of the non-cooperative nodes is weaker for high values of k. Figure 4a and Figure 4b provide this intuition, which is confirmed in Table 1. This table gives the smallest value of t such that $\mathbb{P}\{T_{k,n} > t\} < \varepsilon$.

5 Discussion

In this paper, we have considered the presence of non-cooperative nodes. Such nodes have an impact on the spreading time of a rumor, but do not endanger the content of the rumor. We are currently studying the impact of a proportion f of malicious nodes whose objective is to modify the rumor and propagate this



(a) $f = 0.1$. (b) $f = 0.2$.

Fig. 4: $\mathbb{P}\{T_{k,n} > t\}$ with $n = 100$ nodes.

	$\varepsilon = 0.1$		$\varepsilon = 0.01$		$\varepsilon = 0.001$	
k	$f = 0.1$	$f = 0.2$	$f = 0.1$	$f = 0.2$	$f = 0.1$	$f = 0.2$
$k = 2$	790	928	1020	1099	1131	1146
$k = 3$	448	536	582	689	699	829
$k = 5$	283	347	345	426	410	496
$k = 10$	200	258	232	313	264	354

Table 1: Values of $t = \inf\{t \geq 0 \mid \mathbb{P}\{T_{k,n} > t\} < \varepsilon\}$ for different values of k, f and ε with $n = 100$ nodes.

modified rumor further. Concretely, when a malicious node is chosen during a k-pull operation, if this node knows the rumor, it will modify it and send it to the initiator of the k-pull operation. The objective is to analyze the necessary (and sufficient) conditions for the propagation of the initial rumor in presence of a fraction f of malicious nodes. The value of k should be predominant to enable the initiator of a k-pull operation to choose which rumor to learn during a k-pull operation.

References

1. H. Acan, A. Collevecchio, A. Mehrabian, and N. Wormald. On the push & pull protocol for rumour spreading. *Trends in Mathematics*, 6:3–10, 2017.
2. F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumor spreading in social networks. *Theoretical Computer Science*, 412(24):2602–2610, 2011.
3. A. Clementi, P. Crescenzi, C. Doerr, P. Fraigniaud, F. Pasquale, and R. Silvestri. Rumor spreading in random evolving graphs. *Random structures and Algorithms*, 48(2):290–312, 2015.
4. S. Daum, F. Kuhn, and Y. Maus. Rumor spreading with bounded indegree. In *Proceedings of the International Colloquium on Structural Information and Communication Complexity (SIROCCO)*, 2016.

5. A. J. Demers, D. H. Greene, C. H. Hauser, W. Irish, J. Larson, S. J. Shenker, H. E. Sturgis, D. C. Swinehart, and D. B. Terry. Epidemic algorithms for replicated database maintenance. *PODC 87*, pages 1–12, 1987.
6. B. Doerr, M. Fouz, and T. Friedrich. Experimental analysis of rumor spreading in social networks. *MedAlg 2012*, pages 159–173, 2012.
7. B. Doerr and A. Kostrygin. Randomized Rumor Spreading Revisited. In Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, editors, *ICALP 2017*, volume 80, pages 138:1–138:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.
8. F. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.
9. N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and expanders. *Random Structures and Algorithms*, 43(2):201–220, 2013.
10. A. Frieze and G. Grimmet. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10(1):57–77, 1985.
11. G. 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.
12. G. Giakkoupis, Y. Nazari, and P. Woelfel. How asynchrony affects rumor spreading time. *PODC '16*, pages 185–194, 2016.
13. L. Gordon. Bounds for the distribution of the generalized variance. *The Annals of Statistics*, 17(4):1684–1692, 1989.
14. Y. Mocquard, S. Robert, B. Sericola, and E. Anceaume. Analysis of the propagation time of a rumour in large-scale distributed systems. *NCA 2016*, 2016.
15. K. Panagiotou, X. Perez-Gimenez, T. Sauerwald, and H. Sun. Randomized rumor spreading: the effect of the network topology. *Combinatorics, Probability and Computing*, 24(2):457–479, 2015.
16. K. Panagiotou, A. Pourmiri, and T. Sauerwald. Faster rumor spreading with multiple calls. *The electronic journal of combinatorics*, 22, 2015.
17. A. Pourmiri and F. Ramezani. Brief announcement: Ultra-fast asynchronous randomized rumor spreading. *SPAA 2019*, 2019.
18. F. Robin, B. Sericola, E. Anceaume, and Y. Mocquard. Stochastic analysis of rumor spreading with multiple pull operations. *Methodology and Computing in Applied Probability*, 24:2195–2211, 2022.
19. S. Sanghavi, B. Hajek, and L. Massoulié. Gossiping with multiple messages. *IEEE Transactions on Information Theory*, 53(123), 2007.
20. B. Sericola. *Markov Chains. Theory, Algorithms and Applications.* John Wiley & Sons, 2013.
21. D. W. Stroock. *Probability Theory: An Analytic View.* Cambridge University Press, second edition, 2010.
22. G. Yao, J. Bi, S. Wang, Y. Zhang, and Y. Li. A pull model IPv6 duplicate address detection. *LCN 2010*, 2010.