

On Reliability of the Extrema Propagation Technique in Random Environment

Jacek Cichoń  

Department of Fundamentals of Computer Science, Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland

Dawid Dworzański  

Department of Fundamentals of Computer Science, Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland

Karol Gotfryd  

Department of Fundamentals of Computer Science, Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland

Abstract

We study the reliability of the following simple mechanism for spreading information in a communication network in the presence of random message loss. Initially, some nodes have information that they want to distribute throughout the network. Each node that has received the information tries to broadcast it to all its neighbors. However, due to interference or communication failures, each transmission between two nodes is broken independently with some fixed probability.

This transmission mechanism is the basis for the extrema propagation technique, proposed and analyzed in [2, 3, 10]. Extrema propagation is a simple but powerful method of spreading the extreme values stored by the nodes. In a fully reliable environment, only the number of rounds equal to the network diameter is required for all nodes to be informed. It was shown empirically in [2] that it also performs well in the presence of link failures and message loss. Extrema propagation is an algorithmic framework that was adopted for designing efficient method for distributed data aggregation, such as estimation of cardinalities, sums, and averages, estimation of data distribution via histograms and random sampling (cf. [3, 19]).

In this paper, we propose a formal network model in which random transmission failures occur and analyze the operation time of the extrema propagation technique for any connected network graph. We provide new general probabilistic upper bounds on the number of rounds needed to spread the extreme values that depend only on the number of nodes, the diameter of the network, and the probability of successful transmission. For some special families of graphs, we also derive specific, more accurate estimates. Our theoretical and experimental results confirm the reliability and efficiency of the extrema propagation technique in faulty or noisy environments.

2012 ACM Subject Classification Networks → Mobile ad hoc networks; Computing methodologies → Distributed algorithms; Theory of computation → Random network models; Networks → Network reliability

Keywords and phrases wireless ad-hoc networks, extrema propagation, distributed data aggregation, fault tolerant aggregation, randomly evolving networks

Digital Object Identifier 10.4230/LIPIcs.OPODIS.2024.29

1 Introduction

Extrema Propagation Technique (EPT), introduced in [3], is a simple but powerful method to find the maximal or minimal value of observations stored in nodes of the network. Here is a brief overview of this procedure. Assume that each node $v \in V$ of a connected graph $\mathcal{G} = (V, E)$ has some real number X_v and suppose that the edges E of the graph \mathcal{G} describe communication links between nodes. In the initial round of the considered algorithm, each node copies the value X_v into the local variable Min_v and sends it to all its neighbors.



© Jacek Cichoń, Dawid Dworzański, and Karol Gotfryd;
licensed under Creative Commons License CC-BY 4.0

28th International Conference on Principles of Distributed Systems (OPODIS 2024).

Editors: Silvia Bonomi, Letterio Galletta, Etienne Rivière, and Valerio Schiavoni; Article No. 29; pp. 29:1–29:16



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Then, in each subsequent round, each node stores all the values received from its neighbors, calculates its minimum M_v and if $M_v < Min_v$, then sends M_v to all its neighbors and makes a substitution $Min_v = M_v$. It is clear that after D rounds, where D denotes the diameter of the graph \mathcal{G} , the algorithm stabilizes (no messages are sent over the network) and that $Min_v = M$ for all $v \in V$, where $M = \min\{X_v : v \in V\}$.

The great advantage of this method is its ease of implementation. A trivial use of this method is to determine extreme values in sensor networks. Less trivial applications involve the use of systems of independent variables. For example, assume that the values $(X_v)_{v \in V}$ are independent random variables that have a uniform distribution in the interval $[0, 1]$. Then the expected value of the random variable $\min\{X_v : v \in V\}$ is $1/(N + 1)$, where $N = |V|$. This fact can be used to derive an estimator of the number of vertices in the network. Here is another observation: If $(X_v)_{v \in V}$ are independent random variables with exponential distribution with parameters $(\lambda_v)_{v \in V}$, then the random variable $M = \min\{X_v : v \in V\}$ is also exponentially distributed with parameter $\lambda = \sum_{v \in V} \lambda_v$. This property can be used to derive an estimator of the sum of observed parameters $\{\lambda_v : v \in V\}$. These interesting applications were discussed and analyzed in [2, 6].

The upper limit on the EPT operation time given by the diameter of the communication graph is true assuming full connection reliability. However, communication interference may occur in real systems. The aim of this paper is to provide simple formulas for the upper estimate of the EPT operation time, assuming that communication interruptions of individual channels are independent and occur with the same probability. Our formulas depend only on the number of nodes, the diameter of the graph, and the probability of successful transmission.

EPT also has reasonable communication complexity. For example, in [10], it was shown that in specific random settings the average number of messages sent by any node is $O(\log |V|)$. However, it should be noted that these results are true assuming complete reliability of the connections. We will briefly expand on this topic in the last section of this paper.

The results shown in this paper are applicable to each of the examples discussed in [2, 6] in situations where random interferences may occur during message exchange.

1.1 Notation and basic facts

We say that a random variable X follows a geometric distribution ($X \sim \text{Geo}(p)$) if X takes values in the set of positive integers and $\Pr(X = k) = (1 - p)^{k-1}p$ for each $k \geq 1$. If $X \sim \text{Geo}(p)$, then $\mathbf{E}(X) = \frac{1}{p}$ and $\mathbf{Var}(X) = \frac{1-p}{p^2}$. A random variable Y follows an exponential distribution with parameter $\lambda > 0$ ($X \sim \text{Exp}(\lambda)$) if Y takes values in the set of positive reals and for each $x \in \mathbb{R}$, $x \geq 0$, we have $\Pr(Y > x) = \exp(-\lambda x)$. If $\exp(-\lambda) = 1 - p$, then $\text{Geo}(p)$ can be regarded as a discrete version of the $\text{Exp}(\lambda)$ distribution.

We say that a random variable X follows a distribution $\text{MGeo}(n, p)$ if there exists a family Y_1, \dots, Y_n of independent random variables such that $Y_i \sim \text{Geo}(p)$ for each $i \in \{1, \dots, n\}$ and $X = \max\{Y_1, \dots, Y_n\}$. Using elementary arguments it can be showed that if $X \sim \text{MGeo}(n, p)$ then

$$\mathbf{E}(X) = \sum_{k=1}^n \binom{n}{k} \frac{(-1)^{k+1}}{1 - q^k}, \quad (1)$$

where $q = 1 - p$. The special case of this equation applied to $n = 2$ tells us that if $X \sim \text{MGeo}(2, p)$, then $\mathbf{E}(X) = \frac{1}{p} \left(2 - \frac{1}{2-p}\right)$. From Eq. (1) it is possible to determine the asymptotics of $\mathbf{E}(X)$ for small values of n .

Similarly, we define the distribution $\text{MExp}(n, \lambda)$: it is a maximum of n independent copies of random variables with distribution $\text{Exp}(\lambda)$. It is quite easy to derive a closed formula for the expected value of a random variable with $\text{MExp}(n, \lambda)$ distribution: it equals H_n/λ , where H_n denotes the n -th harmonic number. Formulas for expected values of random variables with $\text{MGeo}(n, p)$ for arbitrary n are much more complicated (see, for example, [24]). But the above-mentioned connection between geometric distribution and exponential distribution gives the following upper and lower bounds for a random variable $X \sim \text{MGeo}(n, p)$ (see [15] for details):

$$\frac{H_n}{\ln \frac{1}{1-p}} \leq \mathbf{E}(X) \leq \frac{H_n}{\ln \frac{1}{1-p}} + 1, \quad (2)$$

which will be sufficient for our purposes. Recall that $H_n = \log n + \gamma + O(\frac{1}{n})$, where $\gamma = 0.5772\dots$ is the Euler-Mascheroni constant. We also notice that $(\ln \frac{1}{1-p})^{-1} = \frac{1}{p} - \frac{1}{2} + O(p)$ is as $p \rightarrow 0$.

The difference between the exact value of $\mathbf{E}(X)$ and the number $H_n/\ln(1/(1-p))$ was investigated by several authors (see, e.g., [15, 28]). It appears that there are no limit distributions for this difference and that it has an oscillatory behavior. Among others, it was shown that asymptotically the difference is $\frac{1}{2} + O(p^2)$ as $p \rightarrow 0$.

We will use the following version of the geometric-arithmetic mean inequality:

$$x_1 \cdots x_n \leq ((x_1 + \dots + x_n)/n)^n, \quad n \in \mathbb{N}, \quad x_1, \dots, x_n \geq 0 \quad (3)$$

as well as the Bernoulli inequality:

$$(1+x)^n \geq 1+nx \quad \text{if } n \in \mathbb{N} \text{ and } x \in \mathbb{R}, x > -1. \quad (4)$$

1.2 Related work

The extrema propagation technique (EPT), briefly described at the beginning of this section, was proposed and analyzed in [2, 3]. It is a fast and robust probabilistic method for estimating sums and cardinalities in distributed environments. It can be used in scenarios where neither the network topology nor its size is known to the nodes. The extreme propagation technique is efficient in terms of its message complexity. As shown in [10], the average number of messages sent by any station in any network with n nodes is of order $O(\log n)$. EPT is one of the main building blocks of algorithms for distributed calculation of approximate histograms of sensory values, average estimation (including averaging of time-series data), and distributed random sampling, as discussed in [9, 19, 25].

The authors in [3] show that appropriately adjusted variants of the basic EPT procedure will also perform well in unreliable or noisy environments, where message loss, slow connections, and link failures occur. In recent years, the problem of fault-tolerant distributed data aggregation has received much attention. Many novel and efficient algorithms for both synchronous and asynchronous settings that are resistant to transmission loss or changes in network topology caused by node or link failures have been proposed and analyzed. The most important techniques are surveyed, among others, in [21, 22]. Examples are flow-updating [1, 23], the push-flow algorithm [16], some variants of the push-sum protocol for distributed consensus and average estimation [17, 20], or methods for estimating the distribution of sensory values [5, 27].

The probabilistic model of message loss or link failures proposed by us in Sect. 2 is closely related to the concept of dynamic networks. That is, wireless networks or other distributed systems whose topology varies over time, as nodes enter or leave the network, and some

communication links are active only in some time intervals. There is a vast body of literature on highly dynamic distributed systems, in which some formal models based on randomly evolving graphs are introduced and protocols for fast information dissemination (including variations of randomized rumor spreading) and efficient data aggregation are proposed and analyzed; see, e.g., [4, 8, 12, 18, 26] and references therein. In [7] the authors surveyed existing formal models for dynamic networks and introduced the concept of time-varying graphs, which unifies many of the classical models. They also analyzed the most important properties of this framework.

In [13, 14] the idea of edge-Markovian dynamic graphs was introduced and discussed. This framework allows for modeling stochastic time dependence in communication networks. In an edge-Markovian dynamic graph each edge follows an independent two-state Markov chain switching between being active and non-active. The authors also analyzed the rate at which the information is spread across such networks. The study in [14] considered only complete graphs, resulting in tight asymptotic bounds for a wide range of parameters in such graphs.

1.3 Organization of the paper

The rest of this paper is organized as follows. The formal model of networks and communication along with some illustrative examples is presented in Sect. 2. In Sect. 3 we formulate and prove our main results that provide probabilistic upper bounds on the time of information propagation and discuss their accuracy. In Sect. 4 we derive more precise estimates for certain families of graphs. The results of the experimental evaluation of the accuracy of proposed estimates are presented in Sect. 5. In Sect. 6 we conclude our work and suggest some directions for future research.

2 Network Model

Let $\mathcal{G} = (V, E)$ be a graph, and $p \in (0, 1)$. By $\text{Bin}(\mathcal{G}, p)$ we denote the discrete-time stochastic process defined as a double-indexed sequence $(X_{e,t})$, where $e \in E$ and $t \in \{1, 2, \dots\}$, of pairwise independent Bernoulli random variables with probability of success p . Unless otherwise stated, we assume that all network graphs considered in this paper are simple, finite, and connected.

The process $\text{Bin}(\mathcal{G}, p)$ can be described as a sequence (\mathcal{G}_t) of graphs where

$$\mathcal{G}_t = (V, \{e \in E : X_{e,t} = 1\}) .$$

Fix a process $\mathcal{H} = \text{Bin}(\mathcal{G}, p)$ and a non-empty subset A of V . We define the process of flooding the information from the set A as follows: we put $A_0 = A$ and

$$A_{t+1} = A_t \cup \{y \in V : (\exists x \in A_t)(\{x, y\} \in E \wedge X_{\{x,y\},t} = 1)\} .$$

The set A_t consists of nodes that have been informed (activated) up to time t (inclusive), assuming that at each time instant each transmission between two nodes is successful independently with probability p . We are interested in the probabilistic properties of *flooding time with source A* , defined as the random variable

$$T(\mathcal{G}, A, p) = \min\{t : A_t = V\} ,$$

i.e. $T(\mathcal{G}, A, p)$ is the first moment when all nodes of the graph \mathcal{G} are informed, given that at time 0 the initially informed nodes form the set A (source).

We will now discuss a certain property of the proposed model, which we will use repeatedly in future considerations. That is, suppose that $\mathcal{G} = (V, E)$ is a graph. Let $\mathcal{X} = (X_{e,t})$ be a sequence of random binary variables from $\text{Bin}(\mathcal{G}, p)$. Let $E' \subseteq E$ and $\mathcal{G}' = (V, E')$. Let $\mathcal{X}' = (X'_{e,t})$ be the restriction of \mathcal{X} to the edges of E' . Let (A_t) be the flooding process defined by \mathcal{X} and let (A'_t) be the flooding process defined by \mathcal{X}' . Then, by an easy induction argument we have $A'_t \subseteq A_t$ for all $t \geq 1$. Therefore, $T(\mathcal{G}, A, p) \leq T(\mathcal{G}', A, p)$. So, for example, if (V, E') is a spanning tree of a connected graph (V, E) , then the value of $\mathbf{E}(T(\mathcal{G}', A, p))$ is an upper bound on $\mathbf{E}(T(\mathcal{G}, A, p))$.

► **Example 2.1.** Fix a positive integer n . Consider the star graph $\mathcal{S} = (V, E)$ with $V = \{0, \dots, n\}$ and $E = \{\{0, k\} : k \in \{1, \dots, n\}\}$. Consider the process $\text{Bin}(\mathcal{S}, p)$ and the set $A = \{0\}$. Notice that in this case we can look independently at the process of disseminating information from vertex 0 to vertex k for each $k \in \{1, \dots, n\}$ and that each such process has a geometric distribution with parameter p . Therefore, the random variable $T(\mathcal{S}, A, p)$ has the distribution $\text{MGeo}(n, p)$, so

$$\mathbf{E}(T(\mathcal{S}, \{0\}, p)) = \frac{H_n}{\ln \frac{1}{1-p}} + \epsilon_n$$

for some $\epsilon_n \in [0, 1]$. From this formula, we deduce that when n is fixed and $p \rightarrow 0$ then we get $\mathbf{E}(T(\mathcal{S}, \{0\}, p)) = H_n(\frac{1}{p} - \frac{1}{2}) + O(p) + \epsilon_n$.

► **Example 2.2.** Fix a positive integer n . Consider the following graph $\mathcal{G} = (V, E)$ where $V = \{0, \dots, n, n+1\}$ and $E = \{\{0, k\} : k \in \{1, \dots, n\}\} \cup \{\{k, n+1\} : k \in \{1, \dots, n\}\}$ (we will call this graph a *double star graph*; see Fig. 1). Consider the process $\text{Bin}(\mathcal{G}, p)$ and sets $A = \{0\}$ and $B = \{0, n+1\}$. We will estimate the expected flooding time $\mathbf{E}(T)$, where $T = T(\mathcal{G}, A, p)$.

Let $T_1 = \min\{t : X_{\{1, n+1\}, t} = 1 \wedge (\exists s < t)(X_{\{0, 1\}, s} = 1)\}$. It is clear that $\mathbf{E}(T_1) = 2/p$. The random variable T_1 can be interpreted as the first moment when node $n+1$ was informed by node 1 (may be because this node was informed by another node before this time). From time T_1 our process behaves similarly to the process of flooding the information in \mathcal{G} from the set B , hence it is similar to the process of Example 2.1 with probability p replaced by $1 - (1-p)^2$ (each of at most n uninformed nodes can be activated independently either by 0 or $n+1$). Therefore

$$\begin{aligned} \mathbf{E}(T) &= \sum_{t \geq 2} \mathbf{E}(T|T_1 = t) \Pr(T_1 = t) \leq \sum_{t \geq 2} (t + \mathbf{E}(T_B)) \Pr(T_1 = t) \\ &= \mathbf{E}(T_1) + \mathbf{E}(T_B) = \frac{2}{p} + \mathbf{E}(T_B) \end{aligned}$$

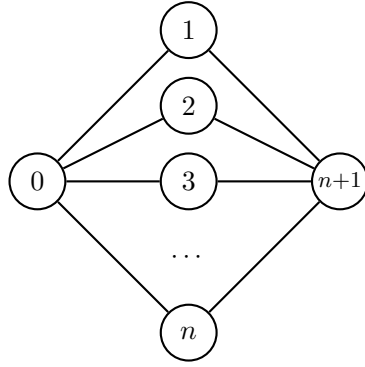
from which we conclude that

$$\mathbf{E}(T) \leq \frac{H_n}{2 \ln \frac{1}{1-p}} + \frac{2}{p} + \epsilon_n$$

for some $\epsilon_n \in [0, 1]$. If n is fixed and $p \rightarrow 0$ then $\mathbf{E}(T) = H_n(\frac{1}{2p} - \frac{1}{4}) + \frac{2}{p} + 1 + O(p)$. Therefore, for large values of n , the considered process runs approximately twice as fast as the process for the star graph.

3 Upper bounds

In this section, we prove the main results of our paper – new upper bounds on the time of information propagation in any connected communication graph (Theorems 3.4 and 3.5).



■ **Figure 1** A double star graph with $n + 2$ vertices from Example 2.2.

These bounds depend only on some global characteristics of the network, i.e., the number of nodes, its diameter, and probability of successful transmission. We also discuss their precision and applications. Before formulation of these results, we formulate two auxiliary lemmas.

► **Lemma 3.1.** *Let \mathcal{S} be a star graph with vertices $\{0, 1, \dots, n\}$ and root 0 (as in Example 2.1). Let $\eta, p \in (0, 1)$ and denote $q = 1 - p$. Then*

$$\Pr(T(\mathcal{S}, \{0\}, p) \leq t) \geq 1 - \eta$$

if and only if

$$t \geq \frac{1}{\ln(1/q)} \ln \frac{1}{1 - (1 - \eta)^{1/n}} .$$

Proof. Let E denote the event $\{T(\mathcal{S}, \{0\}, p) \leq t\}$, that is, “all the vertices are active after t steps”. Then $\Pr(E) = (1 - q^t)^n$. Therefore,

$$\begin{aligned} ((1 - q^t)^n \geq 1 - \eta) &\iff (1 - q^t \geq (1 - \eta)^{1/n}) \iff (q^t \leq 1 - (1 - \eta)^{1/n}) \iff \\ &\iff \left(t \ln q \leq \ln(1 - (1 - \eta)^{1/n})\right) \iff \left(t \geq \frac{1}{\ln q} \ln(1 - (1 - \eta)^{1/n})\right) . \quad \blacktriangleleft \end{aligned}$$

► **Lemma 3.2.** *Let $\eta \in (0, 1)$ and $n \geq 1$ be an integer. Then*

$$\frac{1}{1 - (1 - \eta)^{1/n}} \leq \frac{n}{\eta} .$$

Proof. Let $x = 1 - (1 - \eta)^{1/n}$. Observe that $x \in (0, 1)$. Then $\eta = 1 - (1 - x)^n$. From the Bernoulli inequality (4) we get $(1 - x)^n \geq 1 - nx$, so $1 - (1 - x)^n \leq nx$. Therefore $(1 - (1 - x)^n)^{-1} \geq 1/(nx)$, so

$$\frac{n}{\eta} = \frac{n}{1 - (1 - x)^n} \geq \frac{n}{nx} = \frac{1}{x} = \frac{1}{1 - (1 - \eta)^{1/n}} . \quad \blacktriangleleft$$

Directly from Lemmas 3.1 and 3.2 we get the following result.

► **Corollary 3.3.** *Let \mathcal{S} be the star graph with root 0 and vertices $\{0, 1, \dots, n\}$ and let $\eta, p \in (0, 1)$. Then*

$$t \geq \frac{1}{\ln(1/q)} \ln \left(\frac{n}{\eta}\right) \implies \Pr(T(\mathcal{S}, \{0\}, p) \leq t) \geq 1 - \eta ,$$

where $q = 1 - p$.

We now formulate and prove our main results characterizing the probabilistic upper bounds on the flooding time in any connected network graph with a single source node.

► **Theorem 3.4.** *Suppose that $\mathcal{G} = (V, E)$ is a connected graph with $n + 1$ vertices, where $n > 0$ and diameter D . Let $a \in V$. Then for every $\varepsilon \in (0, 1)$ we have*

$$\Pr \left(T(\mathcal{G}, \{a\}, p) \leq \frac{D \ln \left(\frac{n}{\varepsilon} \right)}{\ln \left(\frac{1}{1-p} \right)} \right) \geq 1 - \varepsilon .$$

Proof. Fix a connected graph $\mathcal{G} = (V, E)$ with $n + 1$ vertices. Let d denote the distance function in this graph (that is, the length of the shortest path between the vertices). Fix a vertex $a \in V$ and let $L = \max\{d(a, v) : v \in V\}$ be the eccentricity of a . Clearly $L \leq D$.

Let $E_i = \{y \in V : d(a, y) = i\}$ and $n_i = |E_i|$ for $i = 0, \dots, L$. Notice that $\sum_{i=1}^L n_i = n$. Fix $\varepsilon \in (0, 1)$ and let $\eta \in (0, 1)$ be such that

$$(1 - \eta)^L = 1 - \varepsilon . \tag{5}$$

From the Bernoulli inequality (4) we get

$$\varepsilon \leq L\eta . \tag{6}$$

Consider the process $\text{Bin}(\mathcal{G}, p)$ that starts with $A_0 = \{a\}$ and assume that at time t_i we have $E_i \subseteq A_{t_i}$ (that is, all nodes at distance i from a have been informed up to time t_i). Letting $q = 1 - p$, from Corollary 3.3 we deduce that at time

$$t_{i+1} = t_i + \frac{1}{\ln(1/q)} \ln \left(\frac{n_{i+1}}{\eta} \right)$$

we have $\Pr(E_{i+1} \subseteq A_{t_{i+1}}) \geq 1 - \eta$ for any $i = 1, \dots, L$. In fact, each node in E_{i+1} is connected to some informed node in E_i , and the bound of the star graph carries over to this case. Let

$$T = \sum_{i=0}^{L-1} \frac{1}{\ln(1/q)} \ln \left(\frac{n_{i+1}}{\eta} \right)$$

and notice that

$$\Pr(T(\mathcal{G}, \{a\}, p) \leq T) \geq (1 - \eta)^L = 1 - \varepsilon .$$

We have

$$\begin{aligned} T &= \frac{1}{\ln(1/q)} \sum_{i=1}^L \ln \left(\frac{n_i}{\eta} \right) = \frac{1}{\ln(1/q)} \ln \left(\prod_{i=1}^L \frac{n_i}{\eta} \right) \\ &\leq \frac{1}{\ln(1/q)} \ln \left(\frac{\sum_{i=1}^L n_i}{L\eta} \right)^L = \frac{L}{\ln(1/q)} \ln \left(\frac{n}{L\eta} \right) \\ &\leq \frac{L}{\ln(1/q)} \ln \left(\frac{n}{\varepsilon} \right) \leq \frac{D}{\ln(1/q)} \ln \left(\frac{n}{\varepsilon} \right) \end{aligned}$$

(we used the geometric-arithmetic mean inequality (3) in the second line and inequality (6) in the third line). Therefore,

$$\Pr \left(T(\mathcal{G}, \{a\}, p) \leq \frac{D}{\ln(1/q)} \ln \left(\frac{n}{\varepsilon} \right) \right) \geq \Pr(T(\mathcal{G}, \{a\}, p) \leq T) \geq 1 - \varepsilon . \quad \blacktriangleleft$$

29:8 On Reliability of the Extrema Propagation Technique

► **Theorem 3.5.** *Suppose that $\mathcal{G} = (V, E)$ is a connected graph with $n + 1$ vertices, where $n > 0$. Let $a \in V$ and $L = \max\{d(a, v) : v \in V\}$. Then*

$$\mathbf{E}(T(\mathcal{G}, \{a\}, p)) \leq \frac{L \ln \frac{ne}{L}}{\ln \frac{1}{1-p}} + L.$$

Proof. The proof is based on a similar reasoning as in the case of Theorem 3.4. Let $T_i = \min\{t : E_i \subseteq A_t\}$, $R_i = T_i - T_{i-1}$ and $T = \sum_{i=1}^L R_i$. From Example 2.1 we can deduce that $\mathbf{E}(R_i) \leq \frac{H_{n_i}}{\ln(1/(1-p))} + 1$. Using the inequality $H_n \leq \ln n + 1$ and letting $q = 1 - p$, we have

$$\begin{aligned} E(T) &\leq \sum_{i=1}^L \left(\frac{H_{n_i}}{\ln(1/q)} + 1 \right) \leq \frac{1}{\ln(1/q)} \sum_{i=1}^L (\ln n_i + 1) + L \\ &= \frac{1}{\ln(1/q)} \left(\ln \left(\prod_{i=1}^L n_i \right) + L \right) + L \leq \frac{1}{\ln(1/q)} \left(L \ln \frac{\sum_{i=1}^L n_i}{L} + L \right) + L \\ &= \frac{1}{\ln(1/q)} \left(L \ln \frac{n}{L} + L \right) + L = \frac{L \ln \frac{ne}{L}}{\ln(1/q)} + L. \quad \blacktriangleleft \end{aligned}$$

The upper bound on the expected flooding time given by Theorem 3.5 depends on the eccentricity L of the initially informed vertex a . However, one can easily derive a generalized version that holds for all vertices in a given graph.

► **Corollary 3.6.** *Suppose that $\mathcal{G} = (V, E)$ is a connected graph with $n + 1$ vertices, where $n > 0$, and diameter D . Then for any $a \in V$*

$$\mathbf{E}(T(\mathcal{G}, \{a\}, p)) \leq \frac{D \ln \frac{ne}{D}}{\ln \frac{1}{1-p}} + D. \quad (7)$$

Proof. The eccentricity L of any given vertex $a \in V$ is bounded above by the diameter D . Since the function $f(x) = x \ln(\frac{ne}{x})$ is monotonically increasing on the interval $(0, n]$, the bound (7) holds for every $a \in V$. \blacktriangleleft

► **Example 3.7.** Let us consider once again the star graph \mathcal{S} from Example 2.1 with $n + 1$ vertices and the source $A = \{0\}$. We then have $L = 1$, so from Theorem 3.5 we get

$$\mathbf{E}(T(\mathcal{S}, \{0\}, p)) \leq \frac{\ln(n) + 1}{\ln \frac{1}{1-p}} + 1.$$

As shown in Example 2.1, $\mathbf{E}(T(\mathcal{S}, \{0\}, p)) = \frac{H_n}{\ln(1/(1-p))} + \epsilon_n$ for some $\epsilon_n \in [0, 1]$, thus the upper bound is almost tight in this case.

► **Example 3.8.** Let \mathcal{P}_n be the path graph with $n + 1$ vertices $\{0, 1, \dots, n\}$ and edges $\{\{i, i + 1\} : 0 \leq i \leq n - 1\}$. The upper bound from Theorem 3.5 gives

$$\mathbf{E}(T(\mathcal{P}_n, \{0\}, p)) \leq \frac{n}{\ln \frac{1}{1-p}} + n.$$

Using the inequality $\ln(1/(1-p)) \geq p$ (when $p \in [0, 1)$) we obtain the formula $\frac{n}{p} + n$ which is consistent with the exact value $\frac{n}{p}$ of the expectation of a random variable with a negative binomial distribution with parameters n and p up to a constant factor n .

The above examples confirm that the upper estimate of the expected flooding time is relatively accurate. However, the next example shows that in some cases this bound is quite far from the actual value.

► **Example 3.9.** Let \mathcal{K}_n denote the complete graph with n vertices. Let $\mathcal{G}_n = \mathcal{K}_{n+1}$ and let a be some fixed vertex. From Theorem 3.5 we get

$$\mathbf{E}(T(\mathcal{G}_n, \{a\}, p)) \leq \frac{\ln n + 1}{\ln \frac{1}{1-p}} + 1 ,$$

but in Theorem 4.1 we will show that $\lim_{n \rightarrow \infty} \mathbf{E}(T(\mathcal{G}_n, \{a\}, p)) = 2$.

4 Fast transmission kernel

The upper bound from Eq. (7) for the flooding time for some dense graphs is poor. In fact, it was derived by splitting the graph into layers E_i consisting of vertices at distance i from the source, $1 \leq i \leq L$, and summing up the times required to inform all nodes from E_{i+1} after all nodes from E_i have been informed, as if none of them had received the information up to that time and was connected to only one node from E_i (as mentioned in Sect. 2, the analysis can be restricted to some spanning tree rooted at the source node). This is an overestimate, especially in dense graphs, in which the information is disseminated much faster. In the extreme case, for the complete graph \mathcal{K}_{n+1} , it gives $\mathbf{E}(T(\mathcal{K}_{n+1}, \{a\}, p)) \leq \frac{\ln n + 1}{\ln \frac{1}{1-p}} + 1$. However, it can be shown that for fixed p and large n we have $\mathbf{E}(\mathcal{K}_n, \{a\}, p) \approx 2$.

► **Theorem 4.1.** *Let \mathcal{K}_{n+1} denote the complete graph with vertices $V = \{0, \dots, n\}$ and let $p \in (0, 1)$. Then*

$$\lim_{n \rightarrow \infty} \mathbf{E}(T(\mathcal{K}_n, \{0\}, p)) = 2 .$$

Here is an imprecise explanation of the above result: notice that $\mathbf{E}(|A_1|) = 1 + np$ and that $\mathbf{E}(|A_2| \mid |A_1| = k) = (n+1)(1 - (1-p)^k) + k(1-p)^k$; moreover if $k = np$ then $(1-p)^k \leq e^{-np^2}$, so if $np^2 \rightarrow \infty$ then $\mathbf{E}(|A_2|) \approx n + 1$. Notice that this result can be summarized as follows: for fixed p and large n , the graph \mathcal{K}_n behaves like the cycle graph \mathcal{C}_3 . Now we will present a refinement of the above heuristic.

Proof. Let $V_n = \{1, \dots, n\}$, $T_n = \mathbf{E}(T(\mathcal{K}_{n+1}, \{0\}, p))$ and denote $q = 1 - p$. Let $X_0 = \{0\}$, $X_1 = \{y : X_{\{0,y\},1} = 1\}$ and $X_2 = \{y : (\exists x \in X_0 \cup X_1)(X_{\{x,y\},2} = 1)\} \setminus (X_0 \cup X_1)$. We will use the inequality $\mathbf{E}(T_n) \leq \frac{H_n}{\ln(1/q)} + 1$. It holds because a complete graph contains a star graph and we can use the bound from Eq. (2). We will bound expected values from two sides.

Notice that $\Pr(X_1 = V_n) = p^n$, so $\Pr(T_n = 1) = p^n$, hence

$$\begin{aligned} \mathbf{E}(T_n) &= \mathbf{E}(T_n \mid X_1 = V_n) \Pr(X_1 = V_n) + \mathbf{E}(T_n \mid X_1 \neq V_n) \Pr(X_1 \neq V_n) \\ &= p^n + \mathbf{E}(T_n \mid X_1 \neq V_n) (1 - p^n) \geq p^n + 2(1 - p^n) = 2 - p^n , \end{aligned}$$

hence $\liminf_{n \rightarrow \infty} \mathbf{E}(T_n) \geq 2$.

Next, we note that

$$\mathbf{E}(T_n) = \mathbf{E}\left(T_n \mid |X_1| \leq \frac{np}{2}\right) \Pr\left(|X_1| \leq \frac{np}{2}\right) + \mathbf{E}\left(T_n \mid |X_1| > \frac{np}{2}\right) \Pr\left(|X_1| > \frac{np}{2}\right) .$$

For the first term of this formula, we see $\mathbf{E}(|X_1|) = np$. From the Chebyshev inequality applied to the binomial distribution, we get $\Pr(|X_1| \leq \frac{1}{2}np) \leq \frac{4q}{np}$. Thus, we get

$$\lim_{n \rightarrow \infty} \mathbf{E}\left(T_n \mid |X_1| \leq \frac{np}{2}\right) \Pr\left(|X_1| \leq \frac{np}{2}\right) \leq \lim_{n \rightarrow \infty} \left(\left(\frac{H_n}{\ln(1/q)} + 1 \right) \frac{4q}{np} \right) = 0 .$$

29:10 On Reliability of the Extrema Propagation Technique

For the second term we notice that

$$\mathbf{E} \left(T_n \mid |X_1| > \frac{np}{2} \right) \Pr \left(|X_1| > \frac{np}{2} \right) = \sum_{k > (np)/2} \mathbf{E} (T_n \mid |X_1| = k) \Pr(|X_1| = k) .$$

Fix $k > \frac{np}{2}$ and suppose that $|X_1| = k$. Then $|X_2|$ follows the binomial distribution $\text{Bin}(n-k, 1-(1-p)^{k+1})$. Let $Z_k = n-k-|X_2|$. Then $\mathbf{E}(Z_k) = (n-k)(1-p)^{k+1}$. The function $f(x) = (n-x)(1-p)^{x+1}$ decreases on the interval $[0, n]$, so for $k > \frac{np}{2}$ we have

$$\mathbf{E}(Z_k) \leq n(1-p/2)(1-p)^{(np/2)+1} ,$$

so from the Markov inequality we get

$$\Pr(Z_k > 0) = \Pr(Z_k \geq 1) \leq n(1-p/2)(1-p)^{(np/2)+1} .$$

Notice that if $Z_k = 0$ then $T_n = 2$. Hence,

$$\begin{aligned} \mathbf{E}(T_n \mid |X_1| = k) &= \\ \mathbf{E}(T_n \mid |X_1| = k \wedge Z_k = 0) \Pr(Z_k = 0) &+ \mathbf{E}(T_n \mid |X_1| = k \wedge Z_k > 0) \Pr(Z_k > 0) \leq \\ 2 \Pr(Z_k = 0) + \left(\frac{H_n}{\ln(1/q)} + 1 \right) \Pr(Z_k > 0) &\leq 2 + \left(\frac{H_n}{\ln(1/q)} + 1 \right) n(1-p/2)(1-p)^{np/2} , \end{aligned}$$

which implies that

$$\begin{aligned} \sum_{k > (np)/2} \mathbf{E}(T_n \mid |X_1| = k) \Pr(|X_1| = k) &\leq \\ \sum_{k > (np)/2} \left(2 + \left(\frac{H_n}{\ln(1/q)} + 1 \right) n(1-p/2)(1-p)^{np/2} \right) \Pr(|X_1| = k) &= \\ \left(2 + \left(\frac{H_n}{\ln(1/q)} + 1 \right) n(1-p/2)(1-p)^{np/2} \right) \Pr \left(|X_1| > \frac{np}{2} \right) , \end{aligned}$$

so

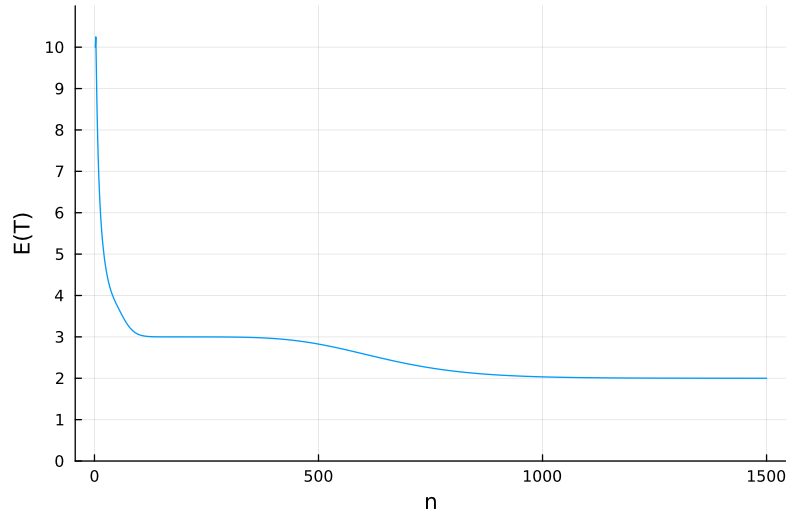
$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{E}(T_n) &= \limsup_{n \rightarrow \infty} \mathbf{E} \left(T_n \mid |X_1| > \frac{np}{2} \right) \Pr \left(|X_1| > \frac{np}{2} \right) \leq \\ \lim_{n \rightarrow \infty} \left(2 + \left(\frac{H_n}{\ln(1/q)} + 1 \right) n(1-p/2)(1-p)^{np/2} \right) \Pr \left(|X_1| > \frac{np}{2} \right) &= 2 . \quad \blacktriangleleft \end{aligned}$$

The stochastic process from Theorem 4.1 can be modeled as a Markov chain with the absorbing state as the state in which all nodes are informed. Using a formula for the expected number of steps before absorption, we calculated exact numerical values for $p = 0.1$. The results are shown in Fig. 2.

► **Example 4.2.** Fix $L > 1$ and consider a graph \mathcal{G} with vertices divided into L blocks B_1, \dots, B_L , each of cardinality 2, and one additional vertex s . We add two edges between s and the vertices in B_1 and connect each vertex in B_i with each vertex in B_{i+1} , $i = 1, \dots, L-1$. This graph, which we call a *double path graph*, has $n = 2L + 1$ vertices and diameter L (see Fig. 3). From Eq. (7) we get the following bound:

$$\mathbf{E}(T(\mathcal{G}, \{s\}, p)) \leq \frac{L(1 + \ln 2)}{\ln \frac{1}{1-p}} + L .$$

Notice that for fixed L and small p this is close to $1.69L/p$. However, in this case, we can easily get a much better upper bound. That is, the time T_0 needed to spread the information

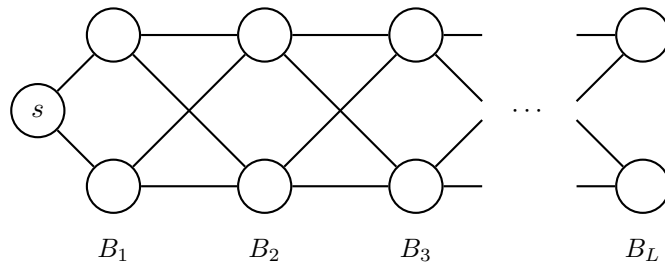


■ **Figure 2** Values of $\mathbf{E}(T(\mathcal{K}_n, \{0\}, 0.1))$ for n from 2 to 1500. The results converge to 2, as stated in Theorem 4.1. The first three values are 10, 10.2493 and 9.62651.

from node s to both nodes of the first block B_1 is bounded above by a random variable with distribution $\text{MGeo}(2, p)$, so from Eq. (1) we get $\mathbf{E}(T_0) = \frac{1}{p}(2 - \frac{1}{2-p})$. Next, assuming that both nodes in B_i are informed, the time T_i needed to spread the information to both nodes in B_{i+1} is bounded above by a random variable with distribution $\text{MGeo}(2, 1 - (1 - p)^2)$, so $\mathbf{E}(T_i) \leq \frac{1}{p(2-p)}(2 - \frac{1}{1+(1-p)^2})$. Putting all these inequalities together we get

$$\mathbf{E}(T(\mathcal{G}, \{s\}, p)) \leq \frac{1}{p} \left(2 - \frac{1}{2-p}\right) + \frac{L-1}{p(2-p)} \left(2 - \frac{1}{1+(1-p)^2}\right).$$

This upper bound for large L and small p is close to $\frac{3}{4}(L+1)/p$.



■ **Figure 3** A double path graph with $n = 2L + 1$ vertices from Example 4.2.

► **Example 4.3.** For the process from Example 2.2, Eq. (7) gives us the following bound:

$$T(\mathcal{G}, A, p) \leq \frac{2 \ln((n+1)e/2)}{\ln(1/(1-p))} + 2.$$

For small p this is close to $2 \ln((n+1)e/2)/p - \ln((n+1)e/2) + 2$, whereas the bound from Example 2.2 is close to $(H_n/2 + 2)/p + 1 - H_n/4$.

29:12 On Reliability of the Extrema Propagation Technique

The examples given above show that in many cases we are able to derive upper bounds on the time of information propagation in a graph better than those that are formulated in the universal Theorems 3.4 and 3.5.

In many natural examples of communication graphs, dense fragments can be distinguished, in which the transfer of information, even in the presence of interference or random faults, is relatively fast. Such fragments can be complete graphs or, for example, fairly dense subsets similar to the graph in Example 4.2. They can be called *fast communication kernels*. In the following, we formulate a theorem that makes it easier to estimate the flooding time of information in such graphs.

► **Theorem 4.4.** *Let $\mathcal{G} = (V, E)$ be a connected graph with diameter D , $V = V_1 \cup V_2$, $n_1 = |V_1|$ and $V_1 \cap V_2 = \emptyset$. Let $\mathcal{G}_2 = (V_2, E \cap [V_2]^2)$. Suppose that*

1. $(\forall v \in V_1)(d(v, V_2) \leq l)$
2. $(\forall v \in V_2)(\mathbf{E}(T(\mathcal{G}_2, \{v\}, p)) \leq T^*)$

Then for all $v \in V$ we have

$$\mathbf{E}(T(\mathcal{G}, \{v\}, p)) \leq \frac{D}{p} \left(2 + \ln \left(\frac{n_1 e}{D} \right) \right) + T^* .$$

Proof. Let us consider $v \in V_1$ and let T_1 be the random variable denoting the time needed to reach the set V_2 by a process of spreading information from the source v . Let P be a path from v to some node from V_2 . The expected time to send the information only along the path P is equal to l/p , where l is the length of P . So $\mathbf{E}(T_1) \leq l/p \leq D/p$. The time T_2 needed to spread the information throughout the whole set V_2 from the reached node is bounded by T^* . At time $T_1 + T_2$ we may treat all nodes of V_2 as a single node, so for the bound on the remaining time T_3 we may use Theorem 3.5, getting $\mathbf{E}(T_3) \leq D \left(\frac{\ln(n_1 e / D)}{\ln(1/(1-p))} + 1 \right)$. Finally, using the inequality $\ln(1/(1-p)) \geq p$ we obtain the final inequality. ◀

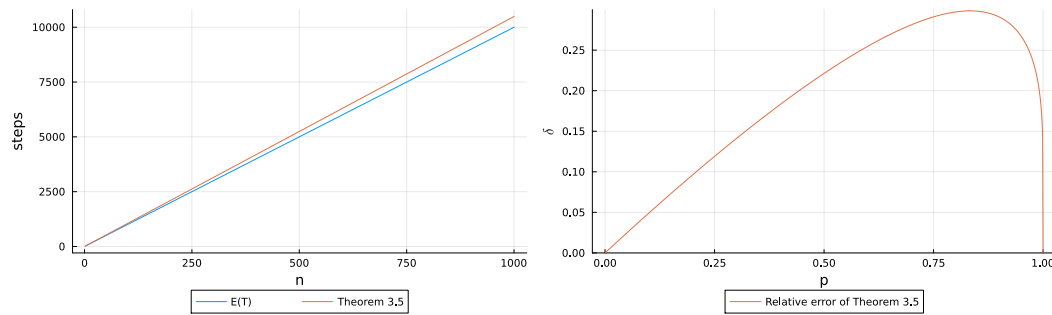
► **Example 4.5.** Consider the graph which consists of a complete graph \mathcal{K}_n with k disjoint paths of length bounded by D connected by a single vertex with \mathcal{K}_n . If $k = 1$ then this graph is called a lollipop graph. From Theorem 4.1 and Theorem 4.4 we deduce that for large n the upper bound is close to $\frac{D}{p}(3 + \ln k) + 2$.

5 Experiments

We conducted a series of experiments that exemplify our results. We compare the estimate of expected value of the flooding time acquired from experiments with general bounds from Theorem 3.5 and specific bounds derived for a given problem. To compare bounds, we used the relative error of the bound and the estimate of the expected value of the flooding time, defined as the ratio of the difference between them to the estimate of the expected value of the flooding time. We choose a small value of p as 0.1 to simulate a noisy environment in which a significant fraction of messages are lost.

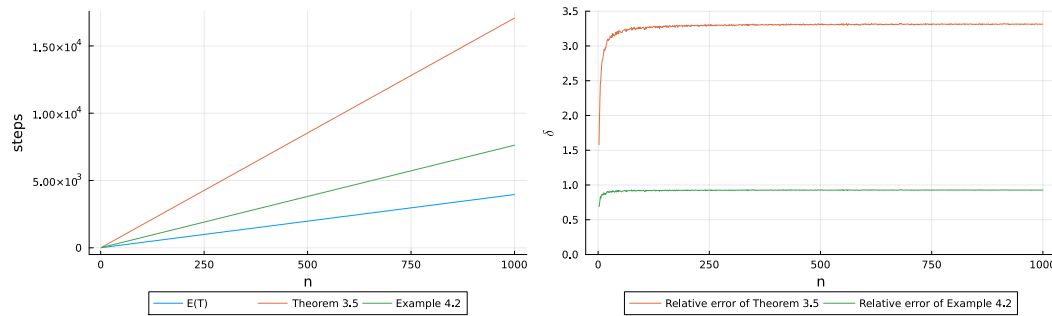
For the path graph from Example 3.8 we can compare the bound from Theorem 3.5 with the exact value of $E(T)$. In Figure 4 we see that the bound as well as $E(T)$ grow linearly as the number of nodes n . The relative error is independent of n and reaches a maximum at $p = 0.833587$ of $\delta = 0.298426$.

For the double path graph from Example 4.2 we set the value of p to 0.1 and average over 1000 simulation runs for values of n from 2 to 1000, the results are shown in Figure 5. We compare two bounds, the first from Theorem 3.5 and the second from the theoretical analysis presented in Example 4.2. We can observe in Fig. 4a that both estimates grow linearly as the number of nodes n increases, but the specific upper bound acquired in Example 4.2 has a lower slope. Relative errors stabilize for larger n .



(a) Expected value of flooding time($p = 0.1$). (b) Relative error of bound.

■ **Figure 4** Comparison of the bounds from Theorem 3.5, the analysis from Example 3.8 and the expected value of flooding time for the path graph for n from 2 to 1000. The Relative error of bound is independent of n .



(a) Expected value of flooding time. (b) Relative error of bounds.

■ **Figure 5** Results of experiments for the double path graph; simulation was run 1000 times for each n from 2 to 1000. Value of p was set to 0.1.

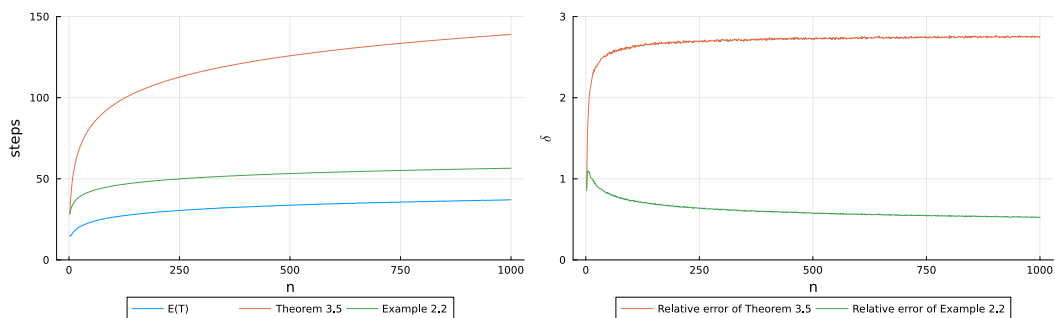
We performed similar experiments for the double star graph from Example 2.2 (the average is taken over 10 000 runs). The results are shown in Fig. 6. In this case, the estimate from Theorem 3.5 is worse than the upper bound derived in Example 2.2. However, note that both bounds capture the logarithmic asymptotic of the experimental data.

6 Conclusions and future work

In this work, we showed simple upper bounds for the necessary number of rounds needed to broadcast information in the connected graph in the case when the transmission disruption probabilities are the same and independent. These formulas are easy to use because they depend only on the diameter of the graph, the number of its vertices, the probability of error-free transmission, and the desired accuracy. We also showed in Sect. 4 how in some situations it is possible to obtain better upper bounds in non-uniform networks.

The upper bounds given in Sect. 3 use very little information about the graph structure. The examples considered by us suggest that the quality of these constraints depends on the density of the graphs. One of the goals of further research is to find simple formulas, similar to those in Section 3, that take into account another characteristic of the graph such as the minimum and maximum degrees of vertices.

29:14 On Reliability of the Extrema Propagation Technique



(a) Expected value of flooding time.

(b) Relative error of bounds.

■ **Figure 6** Results of experiments for the model described in Example 2.2. Simulation was conducted 10 000 times for each n from 2 to 1000. Value of p was set to 0.1.

The basic implementation of EPT-based algorithms is very simple for interference-free communication. It is based on the following simple idea: If in one round you hear a number smaller than the minimum observed so far, then in the next round send it to all your neighbors. This approach obviously fails if the communication is not completely error-free. A trivial but correct solution to this problem is to broadcast the observed minimum to all neighbors until the last round of the algorithm. The sufficient number of rounds can be determined using Theorem 3.4. We plan to use methods from [11] to improve this elementary idea.

Another possible direction of future research is to consider more complex models of transmission failures that, e.g., do not assume the full independence of communication errors or allow different probabilities of successful transmission for different edges. Such models can be better-suited for describing real-world systems, in which interference may occur in some fragment of the network at the same time (so transmission failures are spatially or temporally correlated), or different parts of the system may have different physical properties (e.g., certain areas may be affected by stronger disturbances). We expect that the methods from this paper can also be applied to these situations.

References

- 1 Paulo Sérgio Almeida, Carlos Baquero, Martín Farach-Colton, Paulo Jesus, and Miguel A. Mosteiro. Fault-tolerant aggregation: Flow-Updating meets Mass-Distribution. *Distributed Computing*, 30(4):281–291, August 2017. doi:10.1007/s00446-016-0288-5.
- 2 Carlos Baquero, Paulo Sérgio Almeida, Raquel Menezes, and Paulo Jesus. Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes. *IEEE Trans. Parallel Distrib. Syst.*, 23:668–675, 2012. doi:10.1109/TPDS.2011.209.
- 3 Carlos Baquero, Paulo Sérgio Almeida, and Raquel Menezes. Fast Estimation of Aggregates in Unstructured Networks. In *Fifth International Conference on Autonomic and Autonomous Systems*, ICAS 2009, pages 88–93. IEEE Computer Society, 2009. doi:10.1109/ICAS.2009.31.
- 4 Hervé Baumann, Pierluigi Crescenzi, and Pierre Fraigniaud. Parsimonious flooding in dynamic graphs. In *Proceedings of the 28th ACM Symposium on Principles of Distributed Computing*, PODC '09, pages 260–269, New York, NY, USA, 2009. Association for Computing Machinery. doi:10.1145/1582716.1582757.
- 5 Miguel Borges, Paulo Jesus, Carlos Baquero, and Paulo Sérgio Almeida. Spectra: Robust Estimation of Distribution Functions in Networks. In Karl Michael Göschka and Seif Haridi, editors, *Distributed Applications and Interoperable Systems, Proceedings*, DAIS 2012, pages 96–103. Springer, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-30823-9_8.

- 6 Jorge C. S. Cardoso, Carlos Baquero, and Paulo Sérgio Almeida. Probabilistic Estimation of Network Size and Diameter. In *2009 Fourth Latin-American Symposium on Dependable Computing*, LADC '09, pages 33–40. IEEE Computer Society, September 2009. doi:10.1109/LADC.2009.19.
- 7 Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5):387–408, 2012. doi:10.1080/17445760.2012.668546.
- 8 Arnaud Casteigts, Michael Raskin, Malte Renken, and Viktor Zamaraev. Sharp Thresholds in Random Simple Temporal Graphs. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 319–326, 2022. doi:10.1109/FOCS52979.2021.00040.
- 9 Jacek Cichoń and Karol Gotfryd. Average Counting via Approximate Histograms. *ACM Trans. Sen. Netw.*, 14(2):8:1–8:32, March 2018. doi:10.1145/3177922.
- 10 Jacek Cichoń, Jakub Lemiesz, and Marcin Zawada. On Message Complexity of Extrema Propagation Techniques. In Xiang-Yang Li, Symeon Papavassiliou, and Stefan Rührup, editors, *Proceedings of the 11th International Conference on Ad-Hoc, Mobile, and Wireless Networks*, volume 7363 of *ADHOC-NOW 2012*, pages 1–13, Berlin, Heidelberg, 2012. Springer. doi:10.1007/978-3-642-31638-8_1.
- 11 Jacek Cichoń, Maciej Gębala, and Marcin Zawada. Fault tolerant protocol for data collecting in wireless sensor networks. In *2017 IEEE Symposium on Computers and Communications (ISCC)*, pages 483–486, 2017. doi:10.1109/ISCC.2017.8024575.
- 12 Andrea Clementi, Pierluigi Crescenzi, Carola Doerr, Pierre Fraigniaud, Francesco Pasquale, and Riccardo Silvestri. Rumor spreading in random evolving graphs. *Random Struct. Algorithms*, 48(2):290–312, March 2016. doi:10.1002/rsa.20586.
- 13 Andrea Clementi, Angelo Monti, Francesco Pasquale, and Riccardo Silvestri. Information Spreading in Stationary Markovian Evolving Graphs. *IEEE Trans. Parallel Distributed Syst.*, 22(9):1425–1432, 2011. doi:10.1109/TPDS.2011.33.
- 14 Andrea E.F. Clementi, Claudio Macci, Angelo Monti, Francesco Pasquale, and Riccardo Silvestri. Flooding Time in edge-Markovian Dynamic Graphs. In *Proceedings of the Twenty-Seventh ACM Symposium on Principles of Distributed Computing*, PODC '08, pages 213–222, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1400751.1400781.
- 15 Bennett Eisenberg. On the expectation of the maximum of IID geometric random variables. *Statistics & Probability Letters*, 78(2):135–143, 2008. doi:10.1016/j.spl.2007.05.011.
- 16 Wilfried Gansterer, Gerhard Niederbrucker, Hana Straková, and Stefan Schulze Grotthoff. Scalable and fault tolerant orthogonalization based on randomized distributed data aggregation. *Journal of Computational Science*, 4:480–488, November 2013. doi:10.1016/j.jocs.2013.01.006.
- 17 Balázs Gerencsér and Julien M. Hendrickx. Push-Sum With Transmission Failures. *IEEE Transactions on Automatic Control*, 64(3):1019–1033, 2019. doi:10.1109/TAC.2018.2836861.
- 18 George Giakkoupis, Thomas Sauerwald, and Alexandre Stauffer. Randomized Rumor Spreading in Dynamic Graphs. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *41st International Colloquium on Automata, Languages and Programming (ICALP)*, pages 495–507. Springer Berlin Heidelberg, 2014. doi:10.1007/978-3-662-43951-7_42.
- 19 Karol Gotfryd and Jacek Cichoń. On distributed data aggregation and the precision of approximate histograms. *Journal of Parallel and Distributed Computing*, 180:104722, 2023. doi:10.1016/j.jpdc.2023.104722.
- 20 Christoforos N. Hadjicostis and Themistoklis Charalambous. Average Consensus in the Presence of Delays in Directed Graph Topologies. *IEEE Transactions on Automatic Control*, 59(3):763–768, 2014. doi:10.1109/TAC.2013.2275669.
- 21 Paulo Jesus. *Robust Distributed Data Aggregation*. LAP LAMBERT Academic Publishing, 2016.

29:16 On Reliability of the Extrema Propagation Technique

- 22 Paulo Jesus, Carlos Baquero, and Paulo Sergio Almeida. A Survey of Distributed Data Aggregation Algorithms. *Commun. Surveys Tuts.*, 17(1):381–404, January 2015. doi:10.1109/COMST.2014.2354398.
- 23 Paulo Jesus, Carlos Baquero, and Paulo Sérgio Almeida. Flow updating: Fault-tolerant aggregation for dynamic networks. *Journal of Parallel and Distributed Computing*, 78:53–64, 2015. doi:10.1016/j.jpdc.2015.02.003.
- 24 Barry H. Margolin and Herbert S. Winokur, Jr. Exact Moments of the Order Statistics of the Geometric Distribution and their Relation to Inverse Sampling and Reliability of Redundant Systems. *Journal of the American Statistical Association*, 62(319):915–925, September 1997. doi:10.2307/2283679.
- 25 Saptadi Nugroho, Alexander Weinmann, Christian Schindelbauer, and Andreas Christ. Averaging Emulated Time-Series Data Using Approximate Histograms in Peer to Peer Networks. In *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Trust-worthiness. The PAAMS Collection*, pages 339–346, Cham, 2020. Springer International Publishing. doi:10.1007/978-3-030-51999-5_28.
- 26 Ali Pourmiri and Bernard Mans. Tight Analysis of Asynchronous Rumor Spreading in Dynamic Networks. In *Proceedings of the 39th Symposium on Principles of Distributed Computing, PODC '20*, pages 263–272, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3382734.3405720.
- 27 Jan Sacha, Jeff Napper, Corina Stratan, and Guillaume Pierre. Adam2: Reliable Distribution Estimation in Decentralised Environments. In *Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems, ICDCS '10*, pages 697–707, Washington, DC, USA, 2010. IEEE Computer Society. doi:10.1109/ICDCS.2010.16.
- 28 Wojciech Szpankowski and Vernon Rego. Yet another application of a binomial recurrence order statistics. *Computing*, 43(4):401–410, 1990. doi:10.1007/BF02241658.