On Message Complexity of Extrema Propagation Techniques*

Jacek Cichoń, Jakub Lemiesz, and Marcin Zawada

Institute of Mathematics and Computer Science
Wrocław University of Technology
Poland

Abstract. In this paper we discuss the message complexity of some variants of the Extrema Propagation techniques in wireless networks. We show that the average message complexity, counted as the number of messages sent by each given node, is $\mathrm{O}(\log n)$, where n denotes the size of the network.
We indicate the connection between our problem and the well known and deeply studied problem of the number of records in a random permutation. We generalize this problem onto an arbitrary simple and locally finite graphs, prove some basic theorems and find message complexity for some classical graphs such us lines, circles, grids and trees.

1 Introduction

We analyze a synchronous model of communication. At each round each node (1) receives messages from its neighbors; (2) makes some calculations and finally (3) sends, if necessary, some messages to its neighbors.

Our main goal is to investigate the message complexity of algorithms based on Extrema Propagation Techniques discussed and analyzed in [1], [2]. This technique can be treated as a framework for the construction of efficient algorithms in a distributed environment. For example, in [3] this technique was adopted to an algorithm for approximate estimation of a size of the network. The last algorithm was later improved in [4], where the balls and urns model used in [3] was replaced with independent Bernoulli trials in order to obtain a provable precision of proposed algorithm. In this paper we will show that the message complexity of one node for algorithms based on the Extrema Propagation Technique is logarithmic in the network size.

In Section 2 we consider a distributed algorithm which computes minimum from random numbers generated by nodes. In Section 3 we extend our discussion to a distributed algorithm which determines kth order statistics from numbers generated randomly by nodes, which was used in [3] for the estimation of the cardinality of a wireless network. Theorem 1 and Proposition 1 are known. The remaining results are presumably original.

* Supported by grant N N206 369739 of the Polish National Science Center

We assume that each node in considered networks can calculate a random real number uniformly in the interval $[0,1]$ and that this generators are independent.

1.1 Notation and Basic Definitions

We model a network as a simple directed graph $\mathcal{G}=(V,E)$, i.e. V is a nonempty set and $E \subseteq V \times V \setminus \{(v,v) : v \in V\}$. By $d(x,y)$ we denote the length of a shortest directed path from x to y. If there is no such path, then we put $d(x,y) = \infty$. Let $x \in V$ and $r \geq 0$. We put $\mathrm{B}(x,r) = \{y \in V : d(x,y) < r\}$, $\mathrm{D}(x,r) = \{y \in V : d(x,y) \leq r\}$ and $\mathrm{S}(x,r) = \{y : d(x,y) = r\}$. Observe that $\mathrm{D}(x,0) = \mathrm{S}(x,0) = \{x\}$. The diameter of a graph $\mathcal{G}$ is the number $\Delta = \sup\{d(x,y) : x,y \in V\}$.

In this paper we shall consider only locally finite graphs, i.e. we shall assume that for all $x \in V$ and $r \geq 0$ we have $|\mathrm{D}(x,r)| < \infty$.

Let us recall that the nth harmonic number is defined by $\mathrm{H}_n = \sum_{k=1}^{n} \frac{1}{k}$ and that $\mathrm{H}_n = \ln(n) + \mathrm{O}(1)$. We will also use the standard extension of the function H_n to the complex plane defined, for example, by the formula $\mathrm{H}_z = \sum_{j\geq 1} z/(j(z+j))$. The Euler Beta function is defined by the formula $\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ for $\Re(a) > 0$ and $\Re(b) > 0$. We will use the following identity $\mathrm{B}(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. By $(x)^{\underline{k}}$ we denote the factorial power of x, i.e. $(x)^{\underline{k}} = \prod_{j=0}^{k-1}(x-j)$. By $|A|$ we denote the cardinality of the set A.

2 Propagation of Minimal Number

We start our investigations from the following algorithm (see [1], [2]) of propagation of minimal value of randomly generated real numbers (the pseudo-code of this algorithm is shown at Listing 1):

1. Initially each node $x \in V$ selects independently at random a real ξ_x from the interval [0,1] according to uniform distribution and sends it to all $y \in V$ such that $\{x,y\} \in E$.
2. At each round each node listens to information sent by nodes $S \subseteq \{y : (x,y) \in E\}$ and if $S \neq \emptyset$ and $\xi_x > \min\{\xi_y : y \in S\}$ then
   (a) it puts $\xi_x = \min\{\xi_y : y \in S\}$
   (b) it sends ξ_x to all $y \in V$ such that $\{x,y\} \in E$.

Let $\mathcal{G} = (V,E)$ be the communication graph of considered network, i.e. $\{x,y\} \in E$ if the node x can directly communicate with the node y. Let us assume for the moment that the graph $\mathcal{G}$ is strongly connected. Let Δ denotes the diameter of the graph $\mathcal{G}$. It is easy to see that after Δ rounds for all nodes $x \in V$ we have $\xi_x = \min\{\xi_y : y \in V\}$. Therefore, this algorithm may be used, for example, for leader election in connected networks.

Algorithm 1

Initialization:

$\xi := Random(0,1)$
broadcast $\langle\xi\rangle$ to neighbors

At each round:

gather $\{\eta_i\}_{i\in S}$ from all neighbors
$x = \min\{\eta_i : i \in S\}$
if $x < \xi$ **then**
    $\xi := x$
    broadcast $\langle\xi\rangle$ to neighbors
end if

The first goal of our paper is to investigate the message complexity of this algorithm. Let us fix a graph $\mathcal{G} = (V, E)$ and $x \in V$. We say that the node x transmits at the round r if the part (1) or the part $(2b)$ of the considered algorithm is executed during the rth round. Let $M_{x,r}$ denote the event "node x transmits at the rth round". Notice that $\Pr[M_{x,0}] = 1$ (each node transmits at initialization step) and that for $r > 0$ we have

$$\Pr[M_{x,r}] = \Pr[\min\{\xi_b : b \in \mathrm{S}(x,r)\} < \min\{\xi_b : b \in \mathrm{B}(x,r)\})] .$$

Theorem 1. *Let $\mathcal{G} = (V, E)$ be a simple directed graph and let $x \in V$. Suppose that $\mathrm{S}(x,r) \neq \emptyset$. Then the events $M_{x,1}, \ldots M_{x,r}$ are independent and*

$$\Pr[M_{x,r}] = \frac{|\mathrm{S}(x,r)|}{|\mathrm{D}(x,r)|} .$$

This theorem can be deduced from [5]. We give here a short and self contained proof of it.

Proof. Let $(\xi_v)_{v\in V}$ be a family of independent uniformly distributed random variables in the interval $(0,1)$. Suppose that the theorem is true for a number r and that $\mathrm{S}(x,r+1) \neq \emptyset$. Notice that the event $M_{x,r+1}$ holds if and only if $\min_{v\in\mathrm{S}(x,r+1)} \xi_v < \min_{v\in\mathrm{D}(x,r)} \xi_v$.

Let $a = |\mathrm{S}(x,r+1)|$, $b = |\mathrm{B}(x,r+1)|$, let $X = \min_{v\in\mathrm{S}(x,r+1)} \xi_v$ and $Y = \min_{v\in D(x,r)} \xi_v$. Then $\Pr[X > t] = (1-t)^a$ for $t \in (0,1)$, therefore the function $\phi_X(t) = a(1-t)^{a-1}$ is the density function of the random variable X. Hence

$$\Pr[M_{x,r+1}] = \int_0^1 \Pr[\mathrm{X} < \mathrm{Y}|\mathrm{X} = t]\phi_X(t)dt = \int_0^1 (1-t)^b a(1-t)^{a-1}dt = \frac{a}{a+b} ,$$

therefore $\Pr[M_{x,r+1}] = \frac{a}{a+b} = |\mathrm{S}(x,r+1)/|\mathrm{D}(x,r+1)|$.

Let C denote a conjunction of events of a form $(\pm M_{x,1}) \wedge \ldots \wedge (\pm M_{x,r})$, where $+M_{x,i}$ denotes $M_{x,i}$ and $-M_{x,i}$ denotes $\neg M_{x,i}$. Observe that

$$(C \wedge M_{x,r+1} \wedge (\mathrm{X} = t)) \leftrightarrow (C \wedge (\mathrm{Y} > t) \wedge (\mathrm{X} = t))$$

and that $\Pr[C \wedge (Y > t)] = \Pr[C] \cdot (1-t)^b$. Therefore

$$\Pr[C \wedge M_{x,r+1}] = \int_0^1 \Pr[C \wedge M_{x,r+1} | X = t] a(1-t)^{a-1} dt =$$
$$\int_0^1 \Pr[C](1-t)^b a(1-t)^{a-1} dt = \Pr[C] \frac{a}{a+b} ,$$

hence the event $M_{x,r+1}$ is independent from events $\{M_{x,1}, \ldots, M_{x,r}\}$. $\square$

Let $\mathrm{MC}_{x;\mathcal{G}}$ denote the number of times when the part (1) or $(2b)$ of the considered algorithm is executed. Notice that the energy consumption of sending message is much higher than the cost of listening. Hence this number may be treated as the message complexity of the considered algorithm for the node x. Observe that

$$\mathrm{MC}_{x;\mathcal{G}} = \sum_{r \geq 0} \mathbf{1}_{M_{x,r}} .$$

Therefore the random variable $\mathrm{MC}_{x;\mathcal{G}}$ can be expressed as a sum $\sum_{r \geq 0} \xi_r$ of independent Bernoulli random variables with mean $\frac{|S(x,r)|}{|D(x,r)|}$. Hence

$$\mathrm{E}\left(\mathrm{MC}_{x;\mathcal{G}}\right) = \sum_{r=0}^{\infty} \frac{|S(x,r)|}{|D(x,r)|} \qquad (1)$$

and

$$\mathrm{var}\left(\mathrm{MC}_{x;\mathcal{G}}\right) = \sum_{r=1}^{\infty} \frac{|S(x,r)|}{|D(x,r)|} \left(1 - \frac{|S(x,r)|}{|D(x,r)|}\right) . \qquad (2)$$

Example. Let us consider the line graph $\mathcal{L}_n$, i.e. let $n > 0$, $V = \{1, 2, \ldots, n\}$ and $E = \{(a,b) \in V \times V : 1 \leq a, b \leq n, |a-b| = 1\}$. Note that $|\mathrm{D}(1,r)| = r+1$, $|\mathrm{S}(1,r)| = 1$ for $r < n$. Therefore the random variable $L_n = \mathrm{MC}_{1;\mathcal{L}_n}$ has the same distribution as a sum $\sum_{k=1}^{n} X_k$ of independent random variables, where X_k is a Bernoulli random variable such that $\mathrm{E}\left(X_k\right) = \frac{1}{k}$. Hence the random variable L_n has the same distribution as a well studied number of records in random permutation (see [6]). Thus $\mathrm{E}\left(L_n\right) = \mathrm{H}_n$ and (see e.g. [7]) the normalized random variable $(L_n - \mathrm{H}_n)/\sqrt{\mathrm{H}_n}$ converges in distribution to the standard normal distribution.

2.1 Arbitrary Finite Graphs

It is clear that $1 \leq \mathrm{MC}_{x;\mathcal{G}} \leq 1 + \max\{r : \mathrm{S}(x,r) \neq \emptyset\}$. Here we give other bounds:

Theorem 2. *For any finite graph $\mathcal{G} = (V, E)$ and any vertex $x \in V$ we have*

$$2 - \frac{1}{N} \leq \mathrm{E}\left(\mathrm{MC}_{x;\mathcal{G}}\right) \leq \mathrm{H}_N ,$$

where $N = |\mathrm{B}(x, \infty)|$.

Notice that $\mathrm{B}(x,\infty)$ is the set of all nodes from which the node x can obtain any message hence it is the connected component of the graph $\mathcal{G}$ to which the node x belongs.

Proof. Let $Z = \mathrm{B}(x,\infty)\setminus\{x\}$. Then $\Pr[\min_{z\in Z}\xi_z < \xi_x] = \frac{N-1}{N}$ and any message ξ_a such that $\xi_a = \min_{z\in Z}\xi_z$ will be eventually transmitted to the node x. This proves the first inequality.

We prove the second inequality using a series of simple transformations of the original graph (V,E). Observe that if $|\mathrm{S}(x,r)|\leq 1$ for all $r\geq 1$ then (V,E) is a line graph with the vertex x at its end. Suppose hence that $|\mathrm{S}(x,r)|\geq 2$ for some $r\geq 1$. Let $a\in\mathrm{S}(x,r)$. We perform the following transformations:

1. remove all edges adjacent to a,
2. remove all edges joining $\mathrm{S}(x,r)$ with $\mathrm{S}(x,r+1)$,
3. add new edges $\{\{a,x\} : x\in\mathrm{S}(x,r))\cup\mathrm{S}(x,r+1)\}$

Let $\mathcal{H} = (V,E')$ be the resulting graph. Due to inequality ($b\geq 1, c\geq 2$)

$$\frac{c}{b+c} < \frac{c-1}{b+c-1} + \frac{1}{b+c}$$

we have $\mathrm{E}\left(\mathrm{MC}_{x;\mathcal{G}}\right) < \mathrm{E}\left(\mathrm{MC}_{x;\mathcal{H}}\right)$. After a finite number of such transformations we obtain the line graph with n vertices with the vertex x at its end which was discussed at the end of the previous section. □

Let us note that the lower limit from Theorem 2 is reached in the complete graph, and that the upper limit is achieved by the boundary vertex in the line graph.

2.2 Infinite Graphs

Let us consider an arbitrary locally finite graph $\mathcal{G} = (V,E)$ and $x\in V$. Let $L_{x,n} = \sum_{r=0}^{n} X_r$, where X_n are independent Bernoulli random variables such that $\mathrm{E}\left(X_r\right) = p_r$ where $p_r = |\mathrm{S}(x,r)|/|\mathrm{D}(x,r)|$. Then $\mathrm{E}\left(L_{x,n}\right) = 1 + p_1 + \ldots + p_n$. Since $\mathrm{var}\left(X_r\right) < 1$ for each r we may apply the Strong Law of Large Numbers (see e.g. [8], Thm. 22.4) to the sequence X_r and deduce that

$$\Pr[\lim_{n\to\infty}\frac{1}{n}(L_{x,n} - \mathrm{E}\left(L_{x,n}\right)) = 0] = 1\ .$$

Example. Suppose that $|\mathrm{D}(x,r)| = 2^{r^2}$. Then $|\mathrm{S}(x,r)| = 2^{r^2} - 2^{(r-1)^2}$ for $r>0$, so $\mathrm{E}\left(L_{x,n}\right) = n + \frac{1}{3} + \frac{2}{3}\frac{1}{4^n}$. Hence

$$\Pr[\lim_{n\to\infty}\frac{L_{x,n}}{n} = 1] = 1\ .$$

The following result is a reformulation of a result formulated in Exercise 20.12 from [8]:

Proposition 1. *If* $\mathrm{S}(x,r) \neq \emptyset$ *for each* r*, then* $\Pr[\lim_{n\to\infty} L_{x,n} = \infty] = 1$.

Proof. Let (V,E) be a fixed infinite, locally finite graph. Let $(\xi_v)_{v\in V}$ be a family of independent random variables uniformly distributed in $[0,1]$. Notice that

$$\lim_{n\to\infty} L_{x,n} < \infty \equiv (\exists r)\left(\min_{v\in \mathrm{D}(x,r)} \xi_v < \min_{v\in V\setminus \mathrm{B}(x,r)} \xi_v\right) .$$

For each fixed r the event $\min_{v\in \mathrm{D}(x,r)} \xi_v < \min_{v\in V\setminus \mathrm{B}(x,r)} \xi_v$ has probability null, since the set $\mathrm{D}(x,r)$ is finite and the set $V \setminus \mathrm{B}(x,r)$ is infinite. □

The trivial inequality $L_{x,n} \leq 1+n$ can be improved:

Proposition 2. $\mathrm{E}\left(L_{x,n}\right) \leq 1 + n\left(1 - \sqrt[n]{\frac{1}{|\mathrm{D}(x,n)|}}\right)$

Proof. Let $p_r = |\mathrm{S}(x,r)|/|\mathrm{D}(x,r)|$ and $q_r = 1 - p_r$. Then

$$q_r = |\mathrm{D}(x,r-1)|/|\mathrm{D}(x,r)|$$

for $r > 0$. Therefore $q_1 \cdot \ldots \cdot q_n = 1/|\mathrm{D}(x,r)|$. From the inequality of arithmetic and geometric means we get $q_1 + \ldots + q_n \geq n/\sqrt[n]{|\mathrm{D}(x,n)|}$. Hence, $\mathrm{E}\left(L_{x,n}\right) = 1 + \sum_{r=1}^{n}(1-q_r) \leq 1 + n - n/\sqrt[n]{|\mathrm{D}(x,n)|}$. □

Example. Let us consider an infinite complete binary tree $\mathcal{T} = (V,E)$ and let $x \in V$ be its root. Note that $|\mathrm{D}(x,r)| = 2^{r+1} - 1$. Then, from Proposition 2 by the simple calculation we get $\mathrm{E}\left(L_{x,n}\right) \leq f(n)$, where

$$f(n) = 1 + n\left(1 - \frac{1}{2\sqrt[n]{2}}\right) = \frac{n}{2} + 1 + \frac{\ln 2}{2} + \mathrm{O}\left(\frac{1}{n}\right) .$$

Note that $1 + \frac{\ln 2}{2} \approx 1.3466$. One can verify that this upper bound is sharp. Namely, by some technical manipulation in this case we are able to show that $\mathrm{E}\left(L_{x,n}\right) = \frac{n}{2} + \alpha + \mathrm{O}\left(2^{-n}\right)$, where $\alpha \approx 1.3033$.

There are a lot of examples of infinite graphs where $|\mathrm{S}(x,r)|/|\mathrm{D}(x,r)| = \Theta(\frac{1}{r})$ when r runs to infinity - natural examples of such graphs are the grid-like graphs of arbitrary dimension.

Theorem 3. *Suppose that* $|\mathrm{S}(x,r)|/|\mathrm{D}(x,r)| = \Theta(\frac{1}{r})$*. Then* $\mathrm{E}\left(L_{x,n}\right) = \Theta(\ln n)$ *and the random variable* $(L_{x,n} - \mathrm{E}\left(L_{x,n}\right))/\sqrt{\mathrm{var}\left(L_{x,n}\right)}$ *converges in distribution to the standard normal variable.*

Proof. Let $x_r = |\mathrm{S}(x,r)|/|\mathrm{D}(x,r)|$. From the assumption $x_r = \Theta(\frac{1}{r})$ we deduce that $\sum_{r=0}^{n} x_r = \Theta(\ln n)$ and $\sum_{r=0}^{n} x_r^2 = \mathrm{O}\left(1\right)$. This implies that $\mathrm{E}\left(L_n\right) = \Theta(\ln n)$ and $\mathrm{var}\left(L_n\right) = \sum_r x_r(1-x_r) = \Theta(\ln n)$. Thus the Lindeberg condition (see e.g. [8]) is satisfied, so we may apply Central Limit Theorem. □

2.3 Examples

In this section we shall study message complexity of Algorithm 1 on some classical graphs. Equation 1 gives a possibility to estimate the expected value and variance of the random variable $\mathrm{MC}_{x;\mathcal{G}}$ for many graphs with any required precision. For example, if $\mathcal{C}_n$ denotes the complete graph with n vertices then for any $x \in \mathcal{C}_n$ we have $\mathrm{E}\left(\mathrm{MC}_{x;\mathcal{C}_n}\right) = 2 - \frac{1}{n}$ and $\mathrm{var}\left(\mathrm{MC}_{x;\mathcal{C}_n}\right) = \frac{n-1}{n^2}$.

Line graph Let us consider once again the line graph $\mathcal{L}_n$, i.e. let V=$\{1,2,\ldots,n\}$ and $E = \{(a,b) \in V \times V : 1 \leq a,b \leq n, |a-b| = 1\}$. For an arbitrary number $1 \leq a \leq n/2$ we have

$$\mathrm{E}\left(\mathrm{MC}_{a;\mathcal{L}_n}\right) = 1 + \sum_{k=1}^{a-1} \frac{2}{2k+1} + \sum_{k=2a}^{n} \frac{1}{k} .$$

Hence

$$\mathrm{E}\left(\mathrm{MC}_{\lfloor \frac{n}{2} \rfloor;\mathcal{L}_n}\right) = \mathrm{H}_n - \ln\frac{e}{2} + \mathrm{O}\left(\frac{1}{n}\right) \approx \mathrm{H}_n - 0.306853 + \mathrm{O}\left(\frac{1}{n}\right) .$$

At Fig. 1 we plot the diagram of the function $f(a) = \mathrm{E}\left(\mathrm{MC}_{a;\mathcal{L}_{100}}\right)$ for $a = 1,\ldots,100$. We observe that the maximum value of f is achieved at ends of the graph $\mathcal{L}_n$.

Circle Let $\mathcal{C}_n$ denote the circle graph with n vertices. If $n = 2k+1$ then for each $x \in \mathcal{C}_n$ we have

$$\mathrm{E}\left(\mathrm{MC}_{x;\mathcal{C}_n}\right) = 1 + \sum_{a=1}^{k} \frac{2}{2a+1} = \mathrm{H}_{k+\frac{1}{2}} + \log\frac{4}{e} = \mathrm{H}_n - \ln\frac{e}{2} + \mathrm{O}\left(\frac{1}{n}\right) .$$

For $n = 2k$ we obtain a similar formula

$$\mathrm{E}\left(\mathrm{MC}_{x;\mathcal{C}_n}\right) = \mathrm{H}_{k-\frac{1}{2}} + \log\frac{4}{e} + \frac{1}{n} = \mathrm{H}_n - \ln\frac{e}{2} + \mathrm{O}\left(\frac{1}{n}\right) .$$

Grid Let $\mathcal{G}_n$ denote the grid graph with vertices $V = \{1,\ldots,n\} \times \{1,\ldots,n\}$ and edges $E = \{\{(x,y),(x',y')\} : |x-x'| + |y-y'| = 1\}$. Theorem 2 implies that $\mathrm{E}\left(\mathrm{MC}_{(a,b);\mathcal{G}_n}\right) \leq \mathrm{H}_{n^2}$ for each vertex $(a,b) \in V$.

Proposition 3. *Let $n = 2k-1$ and $N = n^2$. Then*

(a) $\mathrm{E}\left(\mathrm{MC}_{(1,1);\mathcal{G}_n}\right) = \mathrm{H}_N - \delta_1 + \mathrm{O}\left(\frac{1}{\sqrt{N}}\right)$ *where* $\delta_1 \approx 0.729637$,

(b) $\mathrm{E}\left(\mathrm{MC}_{(k,k);\mathcal{G}_n}\right) = \mathrm{H}_N - \delta_2 + \mathrm{O}\left(\frac{1}{\sqrt{N}}\right)$ *where* $\delta_2 \approx 1.415467$.

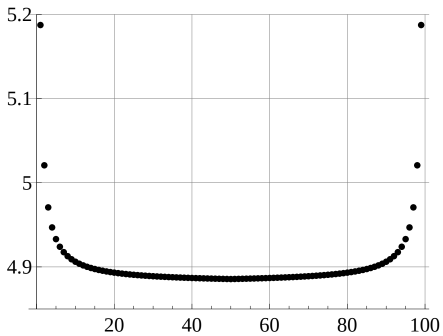

Fig. 1. Plot of $\mathrm{E}\left(\mathrm{MC}_{a;\mathcal{L}_{100}}\right)$ for $a=1,\ldots,100$

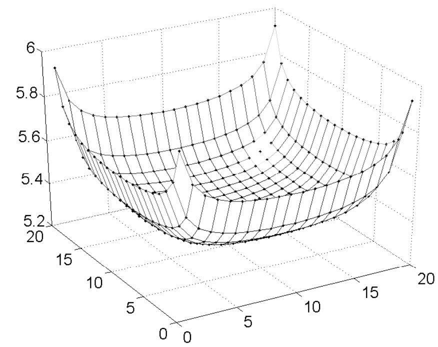

Fig. 2. Plot of $\mathrm{E}\left(\mathrm{MC}_{(a,b);\mathcal{G}_{20}}\right)$ for $a,b\in\{1,\ldots,20\}$

Proof. (a) Let us consider the vertex $v=(1,1)$ and let us define $\mathcal{S}_{r_1}^{r_2}(x)=\sum_{r=r_1}^{r_2}\frac{|S(x,r)|}{|D(x,r)|}$. Then $\mathrm{E}\left(\mathrm{MC}_{v;\mathcal{G}_n}\right)=\mathcal{S}_0^{n-1}(v)+\mathcal{S}_n^{2n-2}(v)$ where (see Fig. 3)

$$\mathcal{S}_0^{n-1}(v)=\sum_{r=0}^{n-1}\frac{r+1}{\frac{1}{2}(r+1)(r+2)}=2(\mathrm{H}_{n+1}-1)$$

and

$$\mathcal{S}_n^{2n-2}(v)=\sum_{r=1}^{n-1}\frac{n-r}{\frac{1}{2}n(n+1)+r\left(n-\frac{r+1}{2}\right)}=\ln 2+\mathrm{O}\left(\frac{1}{n}\right).$$

Hence,

$$\mathrm{E}\left(\mathrm{MC}_{v;\mathcal{G}_n}\right)=2\mathrm{H}_{n+1}-2+\ln 2+\mathrm{O}\left(\frac{1}{n}\right)=\mathrm{H}_{n+1}+(\gamma-2+\ln 2)+\mathrm{O}\left(\frac{1}{\sqrt{N}}\right).$$

(b) Let us now consider the vertex $v=(k,k)$. In a similar way we split the required sum into two parts $\mathrm{E}\left(\mathrm{MC}_{v;\mathcal{G}_n}\right)=1+\mathcal{S}_1^{k-1}(v)+\mathcal{S}_k^{n-1}(v)$ and check that

$$\mathcal{S}_1^{k-1}(v)=\sum_{r=1}^{k-1}\frac{4r}{1+2r(r+1)}=\mathrm{H}_N+c+\mathrm{O}\left(\frac{1}{\sqrt{N}}\right),$$

where $c=-3.108614341\ldots$ and

$$\mathcal{S}_k^{n-1}(v)=\sum_{r=1}^{k-1}\frac{4(k-r)}{(1-2k+2k^2)+(-2+4k)r-2r^2}=\ln 2+\mathrm{O}\left(\frac{1}{\sqrt{N}}\right).$$

Finally we have $1-3.108614341+\ln 2=-1.415467160\ldots$.

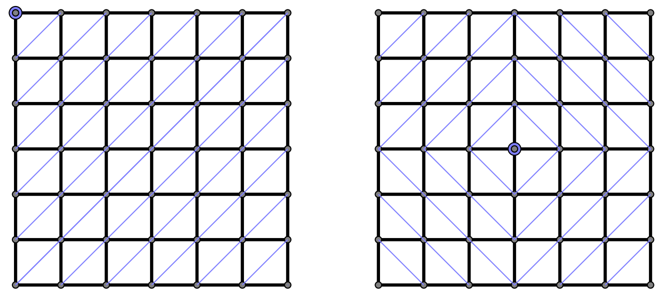

Fig. 3. Division of the graph $\mathcal{G}_7$ into layers depending on the distance from the vertex $(1,1)$ and from the vertex $(4,4)$.

At Fig. 2 we plot the diagram of the function $f(a,b) = \mathrm{E}\left(\mathrm{MC}_{(a,b);\mathcal{G}_{20}}\right)$ for all $a,b = 1,\ldots,20$. We may observe that the maximum value of f is achieved at "corners" of the graph $\mathcal{G}_{20}$.

3 Propagation of Order Statistics

In [3] and [4] a protocol for wireless networks which propagate kth order statistics of real numbers randomly generated by nodes was used for estimation of the size of a network. Here is the description of the transmission part of this algorithm:

1. Initially each node $v \in V$ sets $X_v[1..k] = (1,1,\ldots,1)$, selects a random number $\xi_v \in (0,1)$, puts $X_v[1] = \xi_v$ and sends X_v to all its neighbors.
2. At the beginning of each round the node v makes a copy $Y = X_v$; next with each obtained array Z from a neighbor the node v makes the following operation: $X_v = \mathrm{sort}(X_v \oplus Z)[1,\ldots,k]$ (where $\oplus$ denotes the concatenation of arrays); finally, at the end of the round, if $X_v \neq Y$ then node v sends the array X_v to all its neighbors.

The pseudo-code of this algorithm is shown at Listing 2. Let us note that the case $k = 1$ was considered in previous section.

Lemma 1. *Let $A, B \subseteq V$, $|A| = a$, $|B| = b$, $A \cap B = \emptyset$. Suppose that $1 \leq k \leq a$. Let $(\xi_v)_{v \in A \cup B}$ be a family of independent random variables uniformly distributed in $(0,1)$. Let $\xi_{1:a} \leq \ldots \leq \xi_{a:a}$ be the order statistics generated by $(\xi_v)_{v \in A}$. Then*

$$\Pr[\min_{v \in B} \xi_v < \xi_{k:a}] = 1 - \frac{\binom{a}{k}}{\binom{a+b}{k}} .$$

Algorithm 2

Initialization:

```
X := (1, ..., 1)   [k entries]
X[1] := Random(0,1);
broadcast X to neighbors
```

At each round:

```
Y := X;
for every obtained array Z from neighbors do
    append Z to X;
    sort X;
    X := X[1...k];
end for
if X ≠ Y then
    broadcast X to neighbors;
end if
```

Proof. Let us recall (see e.g. [9]) that the density of the kth order statistic derived from a sequence $(\xi_1, \ldots, \xi_a)$ of independent random variables uniformly distributed in $(0,1)$ is given by the formula $f_{k:a}(t) = \mathrm{B}(k, a-k+1)^{-1} t^{k-1}(1-t)^{a-k}$. Let $\eta = \min_{v \in B} \xi_v$. Notice that $\Pr[\eta < t] = 1-(1-t)^b$ and that η is independent from $\xi_{k:a}$. Therefore

$$\Pr[\eta < \xi_{k:a}] = \int_0^1 (1-(1-t)^b) \frac{1}{\mathrm{B}(k, a-k+1)} t^{k-1}(1-t)^{a-k} dt =$$

$$1 - \frac{1}{\mathrm{B}(k, a-k+1)} \int_0^1 t^{k-1}(1-t)^{a+b-k} dt = 1 - \frac{\mathrm{B}(k, a+b-k+1)}{\mathrm{B}(k, a-k+1)} .$$

□

Let $\mathrm{MC}_x^{(k)}$ denote the number of rounds in which the node $x \in V$ sends a message. Observe that if $|\mathrm{D}(x,\infty)| \leq k$ then

$$\mathrm{MC}_x^{(k)} = 1 + \max\{r : \mathrm{S}(x,r) \neq \emptyset\} .$$

Theorem 4. *Suppose that $|\mathrm{D}(x,\infty)| > k$. Let $s = \min\{r : |\mathrm{D}(x,r)| \geq k\}$ and $m = \max\{r : \mathrm{S}(x,r) \neq \emptyset\}$. Then*

$$\mathrm{MC}_x^{(k)} = \sum_{r=0}^{m} \xi_r ,$$

where $(\xi_r)_{r=0,\ldots,m}$ is a sequence of independent Bernoulli trials such that $\mathrm{E}(\xi_r) = 1$ for $r \leq s$ and

$$\mathrm{E}(\xi_r) = 1 - \frac{\binom{|\mathrm{D}(x,r-1)|}{k}}{\binom{|\mathrm{D}(x,r)|}{k}}$$

for $r > s$.

Proof. Notice that while $|\mathrm{D}(x,r)| \leq k$ then from each sphere $\mathrm{S}(x,j)$ where $j \leq r$ some new information about the kth statistic will be obtained with probability 1. If $r > s$ then the node x has gathered at least k different values from nodes from ball $\mathrm{B}(x,r)$. Hence, its register changes its contents if $\min_{v\in \mathrm{S}(x,r)} \xi_v < X_v[k]$. So we may apply Lemma 1 and deduce that this happens with probability

$$1 - \frac{\binom{|\mathrm{B}(x,r)|}{k}}{\binom{|\mathrm{B}(x,r)|+|\mathrm{S}(x,r)|}{k}} = 1 - \frac{\binom{|\mathrm{D}(x,r-1)|}{k}}{\binom{|\mathrm{D}(x,r)|}{k}} .$$

The proof of independence of constructed random variables follows the same lines as in the proof of Theorem 1 of the corresponding fact. □

Example. Let us again consider the vertex $x = 1$ of the line graph $\mathcal{L}_n$=$\{1,\ldots,n\}$. Recall that $|\mathrm{D}(1,r)| = r+1$. Let us suppose that $k < n$. Therefore we have $\min\{r : |\mathrm{D}(1,r)| \geq k\} = k-1$ and from Theorem 4 we get

$$\mathrm{E}\left(\mathrm{MC}_1^{(k)}\right) = k + \sum_{r=k}^{n-1}\left(1 - \frac{\binom{r}{k}}{\binom{r+1}{k}}\right) = k + \sum_{r=k}^{n-1}\frac{k}{r+1} = k\left(\mathrm{H}_n - \mathrm{H}_k + 1\right) .$$

The next result is a generalization of Theorem 2 onto the case of kth order statistics.

Theorem 5. *Let $k \geq 2$. Suppose that $N = |\mathrm{D}(x,\infty)| > k$. Then*

$$2 \leq \mathrm{E}\left(\mathrm{MC}_x^{(k)}\right) \leq k\left(\mathrm{H}_N - \mathrm{H}_k + 1\right) .$$

Proof. The proof follows the same lines as the proof of Theorem 2: we transform original graph as long as we get a line graph and use the inequality

$$1 - \frac{\binom{a}{k}}{\binom{a+b}{k}} < 1 - \frac{\binom{a}{k}}{\binom{a+b-1}{k}} + 1 - \frac{\binom{a+b-1}{k}}{\binom{a+b}{k}}$$

which holds for $k \geq 2$. □

Let us now consider an infinite graph. Suppose that s and the sequence $(\xi_r)_{r=0,1,\ldots}$ are defined similarly as in Theorem 4 and let us denote $L_{x,n}^k = \sum_{r=0}^n \xi_r$. Theorems 4 and 6 presented below can be proved in the analogous way as the corresponding theorems in Section 2.2.

Proposition 4. $\mathrm{E}\left(L_{x,n}^k\right) \leq 1 + s + (n-s)\left(1 - \sqrt[n-s]{\frac{|\mathrm{D}(x,s)|^{\underline{k}}}{|\mathrm{D}(x,n)|^{\underline{k}}}}\right)$

Proof. Note that

$$\mathrm{E}\left(L_{x,n}^k\right) = 1 + s + (n-s) - \sum_{r=s+1}^{n}(1-\xi_r) .$$

From the inequality of arithmetic and geometric means we get

$$\sum_{r=s+1}^{n}(1-\xi_r) \geq (n-s)\sqrt[n-s]{\frac{|\mathrm{D}(x,s)|^{\underline{k}}}{|\mathrm{D}(x,n)|^{\underline{k}}}}\,.$$

Theorem 6. *Suppose that* $|\mathrm{S}(x,r)|/|\mathrm{D}(x,r)| = \Theta(\frac{1}{r})$. *Then* $\mathrm{E}\left(L^k_{x,n}\right) = \Theta(\ln n)$ *and the random variable* $(L^k_{x,n} - \mathrm{E}\left(L^k_{x,n}\right))/\sqrt{\mathrm{var}\left(L^k_{x,n}\right)}$ *converges in distribution to the standard normal variable.*

Proof. Note that from the fact that $|\mathrm{S}(x,r)|/|\mathrm{D}(x,r)| = \Theta(1/r)$ we can easily deduce that $1 - \frac{|\mathrm{D}(x,r-1)|^k}{|\mathrm{D}(x,r)|^k} = \Theta(1/r)$. One can also check that above relation holds if we replace power k by the falling factorial $\underline{k}$.

3.1 Examples

Circle Let us consider the circle graph $\mathcal{C}_N$ where $N = 2n+1$. Let x be any vertex from this graph. Then $|\mathrm{D}(x,r)| = 2r+1$ for $r \leq n$, so

$$\mathrm{E}\left(\mathrm{MC}^{(k)}_x\right) = m + \sum_{r=m+1}^{n}\left(1 - \frac{\binom{2r-1}{k}}{\binom{2r+1}{k}}\right)\,,$$

where $m = \left\lceil \frac{k-1}{2} \right\rceil$. After some simplification we get

$$\mathrm{E}\left(\mathrm{MC}^{(k)}_x\right) = m + \sum_{r=m+1}^{n}\left(\frac{k}{2r} + \frac{k}{2r+1} - \frac{k^2}{2r(2r+1)}\right) =$$

$$m + k(\mathrm{H}_N - \mathrm{H}_{2m+1}) - k^2\sum_{r=m+1}^{n}\frac{1}{2r(2r+1)} =$$

$$m + k(\mathrm{H}_N - \mathrm{H}_{2m+1}) + k^2(\mathrm{H}_m - \mathrm{H}_{m+\frac{1}{2}}) + \mathrm{O}\left(\frac{1}{N}\right)\,.$$

Therefore, $\mathrm{E}\left(\mathrm{MC}^{(k)}_x\right) \approx \frac{k}{2} + k(\mathrm{H}_N - \mathrm{H}_k) - k$, so $\mathrm{E}\left(\mathrm{MC}^{(k)}_x\right) \approx k(\mathrm{H}_N - \mathrm{H}_k - \frac{1}{2})$.

Grid Let us consider the set $V_n = \{(x,y) \in \mathbf{N}\times\mathbf{N} : |x|+|y| \leq n\}$, $E_n = \{((x,y),(x'y')) : |x-x'|+|y-y'| = 1\}$ and the vertex $v = (0,0)$. Let $\mathcal{G}_n = (V_n, E_n)$. Then, for all $r \leq n$ we have $|\mathrm{D}(v,r)| = 1 + 2r + 2r^2$. Applying Theorem 4 to the graph $\mathcal{G}_n$ we get

$$\mathrm{E}\left(\mathrm{MC}^{(2)}_v\right) = 2 + \sum_{r=2}^{n}\left(1 - \frac{(1-2r+2r^2)^{\underline{2}}}{(1+2r+2r^2)^{\underline{2}}}\right) = 2 + \sum_{r=2}^{n}\frac{2+8r^2}{1+3r+4r^2+2r^3}\,.$$

After some transformations we obtain

$$\mathrm{E}\left(\mathrm{MC}^{(2)}_v\right) = 4\cdot\mathrm{H}_n - 5.62667\ldots + \mathrm{O}\left(\frac{1}{n}\right)\,.$$

Notice that the average message complexity in this case very close to the upper bound

$$2(\mathrm{H}_{2+2n+2n^2} - \mathrm{H}_2 + 1) = 4 \cdot \mathrm{H}_n - 0.768137\ldots + \mathrm{O}\left(\frac{1}{n}\right) .$$

given by Theorem 5. In a similar way we can show that for an arbitrary k

$$\mathrm{E}\left(\mathrm{MC}_v^{(k)}\right) = (2k) \cdot \mathrm{H}_n + \mathrm{O}(1) ,$$

in the graph $\mathcal{G}_n$ when $n \to \infty$.

Tree

Let $\mathcal{T}_n$ be a complete binary tree of depth n rooted at node v. Let us recall that $|\mathrm{D}(v,r)| = 2^{r+1} - 1$ and let us set $s = \min\{r : 2^{r+1} - 1 \geq k\}$. Then, from Theorem 4 we have

$$\mathrm{E}\left(\mathrm{MC}_v^{(k)}\right) = 1 + s + \sum_{r=s+1}^{n} \left(1 - \frac{(2^r - 1)^{\underline{k}}}{(2^{r+1} - 1)^{\underline{k}}}\right) = 1 + n - \sum_{r=s+1}^{n} \left(2^{-k} + \mathrm{O}\left(2^{-r}\right)\right) .$$

Hence, we obtain

$$\mathrm{E}\left(\mathrm{MC}_v^{(k)}\right) = (1 - 2^{-k})n + \mathrm{O}(1) = \alpha_k \mathrm{H}_{(2^{n+1}-1)} + \mathrm{O}(1) ,$$

where $\alpha_k \leq 1.4427$. Observe that in the upper bound given by Theorem 5 the corresponding constant is equal to k.

4 Summary

We analyzed a message complexity of two algorithms based on the Extrema Propagation Techniques - a simple algorithm and an algorithm gathering kth order statistics. We showed that the average message complexity for each node in both algorithms is of order $\mathrm{O}(\log n)$, where n denotes the size of the network.

Note that while considering the records of i.i.d. continuous random variables only the relative order of their outcomes matters (see e.g. [10]). Hence, it is straightforward to observe that the presented results hold for any random variables with a common continuous distribution function. Thus, they can be widely applied. For instance, in [2] Shah et al. consider a general framework for a distributed computing of separable functions, which is based on finding the minimum of exponential random variables.

References

1. Baquero, C., Almeida, P.S., Menezes, R.: Fast estimation of aggregates in unstructured networks. In: Proceedings of the 2009 Fifth International Conference on Autonomic and Autonomous Systems. (2009) 88–93. Available from: `http://gsd.di.uminho.pt/members/cbm/ps/IEEEfastFinalICAS2009.pdf`

2. Mosk-Aoyama, D., Shah, D.: Computing separable functions via gossip. In: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing. PODC '06 (2006) 113–122
3. Cichon, J., Lemiesz, J., Zawada, M.: On cardinality estimation protocols for wireless sensor networks. In Frey, H., Li, X., Rührup, S., eds.: ADHOC-NOW. Volume 6811 of Lecture Notes in Computer Science., Springer (2011) 322–331
4. Cichoń, J., Lemiesz, J., Szpankowski, W., Zawada, M.: Two-phase cardinality estimation protocols for sensor networks with provable precision. In: IEEE WCNC 2012 Conference Proceeding, IEEE Xplore (2012)
5. Yang, M.C.K.: On the distribution of the inter-record times in an increasing population. J. Appl. Probab. (1975) 148–154
6. Rényi, A.: Théorie des 'el'ements saillants d'une suite d'observations. Ann. Fac. Sci. Univ.Clermont-Ferrand (8) (1962) 7–13
7. Steele, J.M.: The bohnenblust—spitzer algorithm and its applications. J. Comput. Appl. Math. **142** (2002) 235–249. Available from: `http://portal.acm.org/citation.cfm?id=586795.586814`
8. Billingsley, P.: Probability and Measure. 3 edn. Wiley-Interscience (1995)
9. Arnold, B., Balakrishnan, N., Nagaraja, H.: A First Course in Order Statistics. John Wiley & Sons, New York (1992)
10. Devroye, L.: Applications of the theory of records in the study of random trees. Acta Informatica **26** (1988) 123–130