



Broadcast in the rendezvous model[☆]

Philippe Duchon, Nicolas Hanusse, Nasser Saheb, Akka Zemmari*

LaBRI – CNRS – Université Bordeaux I, 351 Cours de la Liberation, 33405 Talence, France

Received 9 July 2004; revised 21 April 2005
Available online 10 March 2006

Abstract

In many large, distributed or mobile networks, broadcast algorithms are used to update information stored at the nodes. In this paper, we propose a new model of communication based on *rendezvous* and analyze a *multi-hop distributed algorithm* to *broadcast* a message in a *synchronous* setting. In the *rendezvous model*, two neighbors u and v can communicate if and only if u calls v and v calls u simultaneously. Thus, nodes u and v obtain a rendezvous at a meeting point. If m is the number of meeting points, the network can be modeled by a graph of n vertices and m edges. At each round, every vertex chooses a random neighbor and there is a rendezvous if an edge has been chosen by its two extremities. Rendezvous enable an exchange of information between the two entities. We get sharp lower and upper bounds on the time complexity in terms of number of rounds to broadcast: we show that, for any graph, the expected number of rounds is between $\ln n$ and $O(n^2)$. For these two bounds, we prove that there exist some graphs for which the expected number of rounds is either $O(\ln(n))$ or $\Omega(n^2)$. For specific topologies, additional bounds are given.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Algorithms and data structures; Distributed algorithms; Graph; Broadcast; Rendezvous model

[☆] A preliminary version of this paper has appeared in the 21st Symposium on Theoretical Aspects of Computer Science (STACS), Montpellier, France, March 2004.

* Corresponding author .

E-mail addresses: duchon@labri.fr (P. Duchon), hanusse@labri.fr (N. Hanusse), saheb@labri.fr (N. Saheb), zemmari@labri.fr (A. Zemmari) .

1. Introduction

Among the numerous algorithms to broadcast in a synchronized setting, we are witnessing a new tendency of distributed and randomized algorithms, also called *gossip-based algorithms*: at each instant, any number of broadcasts can take place simultaneously and we do not give any priority to any particular one. In each round, a node chooses a random neighbor and tries to exchange some information. Due to the simplicity of gossip-based algorithm, such an approach provides reliability and scalability. Contrary to deterministic schemes for which messages tend to route in a particular subgraph (for instance a tree), a gossip-based algorithm can be fault-tolerant (or efficient for a dynamic network) since in a strongly connected network, many paths can be used to transmit a message to almost every node.

The majority of results deal with the *uniform random phone call* for which a node chooses a neighbor uniformly at random. However, such a model does not take into account that a given node could be “called” by many nodes simultaneously implying a potential congestion. A more embarrassing situation is the one of the *radio networks* in which a node should be called simultaneously by a unique neighbor otherwise the received messages are in collision. In the *rendezvous model*, every node chooses a neighbor and if two neighbors choose themselves mutually, they can exchange some information. The *rendezvous model* is useful if a physical meeting is needed to communicate as in the case of robots network.

Although the *rendezvous model* can be used in different settings, we describe the problem of broadcasting a message in a network of robots. A robot is an autonomous entity with a bounded amount of memory having the capacity to perform some tasks and to communicate with other entities by radio when they are geographically close. Examples of use of such robots are numerous: exploration [1,7], navigation (see Survey of [17]), capture of an intruder [3], search for information, help to handicapped people or rescue, cleaning of buildings, ... The literature contains many efficient algorithms for one robot and multiple robots are seen as a way to speed up the algorithms. However, in a network of robots [4], the coordination of multiple robots implies complex algorithms. *Rendezvous* between robots can be used in the following setting: consider a set of robots distributed on a geometric environment. Even if two robots sharing a region of navigation (called neighbors) might communicate, they should also be close enough. It may happen that their own tasks do not give them the opportunity to meet (because their routes are deterministic and never cross) or it may take a long time if they navigate at random. A solution consists in deciding on a meeting point for each pair of neighbor robots. If two neighbors are close to a given meeting point at the same time, they have a *rendezvous* and can communicate.

Although there exist many algorithms to broadcast messages, we only deal with algorithms working under a very weak assumption: each node or robot only knows its neighbors or its own meeting points. This implies that the underlying topology is *unknown*. Depending on the context, we might also be interested in *anonymous* networks in which the labeling of the nodes (or history of the visited nodes) is not used. By anonymous, we mean that unique identities are not available to distinguish nodes (processors) or edges (links). In a robot network, the network can have two (or more) meeting points with the same label if the environment contains two pairs of regions that do not overlap. The anonymous setting can be encountered in dynamic, mobile or heterogeneous networks.

1.1. Related works

How to broadcast efficiently a message with a very poor knowledge on the topology of an *anonymous* network? Depending on the context, this problem is related to the way a “rumor” or an “epidemic” spreads in a graph. In the literature, a node is *contaminated* if it knows the rumor. The broadcast algorithm highly depends on the communication model. For instance, in the *k-ports model*, a node can send a message to at most k neighbors. Thus, our rendezvous model is a 1-port model.

The performance of a broadcast algorithm is measured by the time required to contaminate all the nodes, the amount of memory stored at each node or the total number of messages. In this article, we analyze the time complexity in a synchronous setting of a rendezvous algorithm (although several broadcast algorithms including ours can work in an asynchronous setting, the theoretical time complexity is usually analyzed in a synchronous model).

Many broadcast algorithms exist (see the survey by Hedetniemi et al. [9]) but few of them are related to our model. The closest model is the one of Feige et al. [8]. The authors prove general lower and upper bounds ($\log_2 n$ and $O(n \ln n)$) on the time to broadcast a message *with high probability*¹ in any unknown graph. A contaminated node chooses a neighbor uniformly at random but no rendezvous are needed. In our model, the time complexity increases since a rendezvous has to be obtained to communicate. For a family of small-world graphs and other models (2-ports model but a node can only transmit a given message a bounded number of times), Comellas et al. [6] showed that a broadcast can always be done. A recent work of Karp et al. [11] deals with the *random phone call model*. In each round, each node u chooses another node v uniformly at random (more or less as in [8]) but the transmission of a rumor is done either from the caller to the called node (*push transmission algorithm*) or from the called node to the caller (*pull transmission algorithm*). The underlying topology is the complete graph and they prove that any rumor broadcasted in $O(\ln n)$ rounds needs to send $\omega(n)$ messages on expectation.

However, the results of random call phone [8,11] do not imply the presented results in the rendezvous model:

- The classes of graphs for which the broadcast runs fast or slow are different in the rendezvous model and in the random phone call model. For instance, the lower bound is $\Omega(\ln(n))$ in both models. Now, consider the complete graph, its broadcast time $O(n \ln(n))$ is close to the lower bound in the random phone call model whereas it becomes $\theta(n \ln(n))$ in the rendezvous model.
- We deal with the expected broadcast time. Depending on the topology, this time can be either equal or different to the broadcast time with high probability.

In the radio network setting (n -ports model), some algorithms and bounds exist whether the topology is known or unknown (see the survey of Chlebus [5]). However, the model of communication is different from ours: simultaneously, a node can send a message to all of its neighbors and a node can receive a message if and only if a unique neighbor send a message. Two kinds of algorithms are

¹ High probability means with probability $1 - O(n^{-c})$ for some positive constant c .

proposed in the radio model: with or without collision detection. In our model, there is no problem of collision.

Rendezvous in a broadcast protocol are used in applications like Dynamic Host Configuration Protocol but to the best of our knowledge, the analysis of a *randomized* rendezvous algorithm to broadcast in a network is new. The random rendezvous model was introduced in [15] in which the authors compute the expected number of rendezvous per round in a randomized algorithm. Their algorithm is a solution to implement synchronous message passing in an anonymous network that passes messages asynchronously [18]. Many concurrent programming languages including CSP and Ada use this method to define a communication between pairs of asynchronous processes. Angluin [2] proved that there is no deterministic algorithm for this problem (see the paper of Lynch [12] containing many problems having no deterministic solutions in distributed computing). In [16], the rendezvous are used to elect randomly a leader in an anonymous graph.

1.2. The model

Let $G = (V, E)$ be a connected and undirected graph of n vertices and m edges. For convenience and with respect to the problem of spreading an epidemic, a vertex is *contaminated* if it has received the message sent by an *initial vertex* v_0 .

The model can be implemented in a fully distributed way. The complexity analysis, however based on the concept of *rounds*, is commonly used in similar studies [8,15,16]. In our article, a *round* is the following sequence:

- for each $v \in V$, choose uniformly at random an incident edge;
- if an edge (v_i, v_j) has been chosen by v_i and v_j , there is a rendezvous;
- if there is a rendezvous and if only v_i is contaminated, then v_j becomes contaminated.

T_G is the *broadcast time* or *contamination time*, that is the number of rounds until all vertices of graph G are contaminated. T_G is an integer-valued random variable; in this paper, we concentrate the study on its expectation $\mathbf{E}(T_G)$.

Some remarks can be made on our model. As explained in Section 1, the rendezvous process (the first two steps of the round) keeps repeating forever and could be seen as a way of maintaining connectivity. Several broadcasts can take place simultaneously and we do not give any priority to any one of them, even if we study a broadcast starting from a given vertex v_0 .

We concentrate our effort on $\mathbf{E}(T_G)$ and we do not require that the algorithm finds out when the rumor sent by v_0 has reached all the nodes. However, some hints can be given: we can stop the broadcast algorithm (do not run the third step of the round) using a local control mechanism in each node of the network: if identities of the nodes are available (non anonymous networks), each node keeps into its memory a list of contaminated neighbors for each rumor and when this list contains all the neighbors, the process may stop trying to contaminate them (with the same rumor). If the network is anonymous and the number of nodes n is known, then it is possible to prove that in $O(n^2 \ln(n))$ rounds with high probability, all the neighbors of a contaminated node know the rumor.

In our algorithm, nodes of large degree and a large diameter increase the contamination time. Taking two adjacent nodes v_i and v_j of degrees d_i and d_j respectively, the expected number of rounds

to contaminate v_j from v_i is $d_i d_j$. For instance, take two stars of $n/2$ leaves. Join each center by an edge. In the rendezvous model, the expected broadcast time is $\Theta(n^2)$ whereas in [8]’s model, it will be $\Theta(n \ln(n))$ on expectation and with high probability. Starting from this example, $\mathbf{E}(T_G)$ can easily be upper bounded by $O(n^3)$ but we find a tighter upper bound.

1.3. Our results

The main result of the paper is to prove in Section 2 that for any graph G , $\log_2 n \leq \mathbf{E}(T_G) \leq O(n^2)$. More precisely, for any graph G of maximal degree Δ , $\mathbf{E}(T_G) = O(\Delta n)$. This main result is far from obvious.

In Section 3, we show that there are some graphs for which the expected broadcast time asymptotically matches either the lower bound or the upper bound up to a constant factor. For instance, for the complete balanced binary tree, $\mathbf{E}(T_G) = O(\log_2 n)$ whereas $\mathbf{E}(T_G) = \Omega(n^2)$ for the double star graph (two identical stars joined by one edge). For graphs of bounded degree Δ and diameter D , we also prove in Section 3 that $\mathbf{E}(T_G) = O(D\Delta^2 \ln \Delta)$. This upper bound is tight since for Δ -ary complete trees of diameter D , $\mathbf{E}(T_G) = \Omega(D\Delta^2 \ln \Delta)$. The complete graph was proved [15] to have the least expected number of rendezvous per round; nevertheless, its expected broadcast time is $\Theta(n \ln n)$.

2. Arbitrary graphs

The first section presents some terminology and basic lemmas that are useful for the main results.

2.1. Generalities on the broadcast process

The rendezvous process induces a *broadcast process*, that is, for each nonnegative integer t , we get a (random) set of vertices, V_t , which is the set of vertices that have been reached by the broadcast after t rounds. The sequence $(V_t)_{t \in \mathbb{N}}$ is a *homogeneous, increasing Markov process* with state space $\{U : \emptyset \subsetneq U \subset V\}$. Any state U contains the initial vertex v_0 and the subgraph induced by U is connected. State V is its sole absorbing state; thus, for each graph G , this process reaches state V (that is, the broadcast is complete) in finite expected time.

The transition probabilities for this Markov chain (V_k) depend on the rendezvous model. Specifically, if U and U' are two nonempty subsets of V , the transition probability $p_{U,U'}$ is 0 if $U \not\subseteq U'$, and, if $U \subseteq U'$, $p_{U,U'}$ is the probability that, in a given round, $U' - U$ is the set of vertices not in U that have a rendezvous with a vertex in U . Thus, the *loop probability* $p_{U,U}$ is the probability that each vertex in U either has no rendezvous, or has one with another vertex in U .

In the sequel, what we call the *broadcast sequence* is the sequence of *distinct* states visited by the broadcast process between the initial state $\{v_0\}$ and the final absorbing state V . A *possible broadcast sequence* is any sequence of states that has a positive probability of being the broadcast sequence; this is any sequence $\mathcal{X} = (X_1, \dots, X_m)$ such that $X_1 = V_0 = \{v_0\}$, $X_m = V$, and $p_{X_k, X_{k+1}} > 0$ for all k .

By d_u we denote the degree of vertex u . For a bounded degree graph, Δ is the maximal degree of the graph. By D we denote the diameter of the graph.

If $X_k = V_t$ is the set of the k contaminated vertices at time t then Y_k is the set of remaining vertices. We define the cut C_k as the set of edges that have one endpoint in X_k and the other in Y_k .

For any edge $a = (u, v) \in E$, $\mathbf{P}(a) = (d_u d_v)^{-1}$ (respectively, $\mathbf{P}(\bar{a})$) is the probability that edge a will obtain (respectively, not obtain) a rendezvous at a given round. The product $(d_u d_v)^{-1}$ is also called the *weight* of the edge a .

We also define two values for any set of edges $C \subset E$: $\mathbf{P}(\mathcal{E}_C)$ (respectively, $\mathbf{P}(\overline{\mathcal{E}_C})$) where \mathcal{E}_C is the event of obtaining a rendezvous in a round for at least one edge (respectively, no edge) in C ; and $\pi(C) = \sum_{a \in C} \mathbf{P}(a)$. Since $\pi(C)$ is the expected number of rendezvous in C , it is much easier to deal with in computations. Obviously, $\mathbf{P}(\mathcal{E}_C) \leq \pi(C)$ holds for any C . Lemma 2 provides a *lower bound* for $\mathbf{P}(\mathcal{E}_C)$ of the form $\Omega(\pi(C))$ provided $\pi(C)$ is not too large.

With these notations, for any set of vertices U , $p_{U,U} = 1 - \mathbf{P}(\mathcal{E}_{C_U})$, where C_U is the set of edges that have exactly one endpoint in U (the *cut* defined by the partition $(U, V - U)$).

Lemma 1. *Let C be any given subset of E . For any $a \in E$, we have $\mathbf{P}(a \mid \overline{\mathcal{E}_C}) \geq \mathbf{P}(a)$.*

Proof. Partition C into $C_1 \cup C_2$, where $C_1 = \{e' \mid e' \in C, e' \text{ incident to } a\}$ and $C_2 = C \setminus C_1$. Then we have:

$$\begin{aligned} \mathbf{P}(a \mid \overline{\mathcal{E}_C}) &= \mathbf{P}(a \mid \overline{\mathcal{E}_{C_1}} \wedge \overline{\mathcal{E}_{C_2}}) \\ &= \frac{\mathbf{P}(a \wedge \overline{\mathcal{E}_{C_1}} \mid \overline{\mathcal{E}_{C_2}})}{\mathbf{P}(\overline{\mathcal{E}_{C_1}} \mid \overline{\mathcal{E}_{C_2}})}. \end{aligned}$$

Since $\mathbf{P}(\overline{\mathcal{E}_{C_1}} \mid \overline{\mathcal{E}_{C_2}}) \leq 1$, we have:

$$\mathbf{P}(a \mid \overline{\mathcal{E}_{C_1}} \wedge \overline{\mathcal{E}_{C_2}}) \geq \mathbf{P}(a \wedge \overline{\mathcal{E}_{C_1}} \mid \overline{\mathcal{E}_{C_2}}).$$

Once there is a rendezvous on the edge a , there will be no rendezvous on the edges of C_1 . So we have:

$$\mathbf{P}(a \wedge \overline{\mathcal{E}_{C_1}} \mid \overline{\mathcal{E}_{C_2}}) = \mathbf{P}(a \mid \overline{\mathcal{E}_{C_2}}),$$

yielding

$$\mathbf{P}(a \mid \overline{\mathcal{E}_{C_1}} \wedge \overline{\mathcal{E}_{C_2}}) \geq \mathbf{P}(a \mid \overline{\mathcal{E}_{C_2}}).$$

The edge a being adjacent to none of the edges in C_2 , the fact that there is no rendezvous on this edges does not affect the probability of a rendezvous on the edge a . Therefore:

$$\mathbf{P}(a \mid \overline{\mathcal{E}_{C_1}} \wedge \overline{\mathcal{E}_{C_2}}) \geq \mathbf{P}(a).$$

Thus, for any $a \in E \setminus C$

$$\mathbf{P}(a \mid \overline{\mathcal{E}_C}) \geq \mathbf{P}(a). \quad \square$$

Lemma 2. *For any $C \subset E$, $\mathbf{P}(\mathcal{E}_C) \geq \lambda \min(1, \pi(C))$ with $\lambda = 1 - e^{-1}$ where $e = \exp(1)$.*

Proof. Assume that at time t , we have $V_t = X_k$, that is k nodes are contaminated.

There is no new contaminated vertex (and hence $|V_{t+1}| = k$), if and only if there is no rendezvous in C during one round. Let $\{e_1, e_2, \dots, e_l\}$ denote the set of edges of C . Then

$$\begin{aligned} \mathbf{P}(\overline{\mathcal{E}}_C) &= \mathbf{P}(\overline{e}_1 \wedge \overline{e}_2 \wedge \dots \wedge \overline{e}_l) \\ &= \mathbf{P}(\overline{e}_1) \mathbf{P}(\overline{e}_2 \mid \overline{e}_1) \dots \mathbf{P}(\overline{e}_l \mid \overline{e}_1 \wedge \overline{e}_2 \wedge \dots \wedge \overline{e}_{l-1}) \\ &= (1 - \mathbf{P}(e_1))(1 - \mathbf{P}(e_2 \mid \overline{e}_1)) \dots (1 - \mathbf{P}(e_l \mid \overline{e}_1 \wedge \overline{e}_2 \wedge \dots \wedge \overline{e}_{l-1})). \end{aligned}$$

From Lemma 1, we have $\mathbf{P}(e_i \mid \overline{e}_1 \wedge \overline{e}_2 \wedge \dots \wedge \overline{e}_{i-1}) \geq \mathbf{P}(e_i)$. Hence

$$\mathbf{P}(\overline{\mathcal{E}}_C) \leq \prod_{i=1}^l (1 - \mathbf{P}(e_i))$$

and

$$\mathbf{P}(\mathcal{E}_C) \geq 1 - \prod_{i=1}^l (1 - \mathbf{P}(e_i)).$$

Now, since $1 - x \leq e^{-x}$, this becomes

$$\mathbf{P}(\mathcal{E}_C) \geq 1 - \prod_{i=1}^l e^{-\mathbf{P}(e_i)} = 1 - e^{-\pi(C)}.$$

The function $x \mapsto 1 - e^{-x}$ is increasing and concave, so that $1 - e^{-x} \geq \min(\lambda, \lambda x) = \lambda \min(1, x)$ (with $\lambda = 1 - e^{-1}$) holds for all $x \geq 0$. This proves the lemma. \square

Corollary 3.

$$\frac{1}{\mathbf{P}(\mathcal{E}_C)} \leq \frac{e}{e-1} \max\left(1, \frac{1}{\pi(C)}\right) \leq \frac{e}{e-1} \left(1 + \frac{1}{\pi(C)}\right). \tag{1}$$

Lemma 4. For any graph G , any integer k and any $p \in (0, 1)$, if $\mathbf{P}(T_G > k) \leq p$ then $\mathbf{E}(T_G) \leq k/(1 - p)$.

Proof. Cut the broadcast process into “segments” of k rounds, and consider the “broadcast-or-reset” process such that, at the beginning of each segment when the broadcast has not yet occurred, the set of contaminated vertices is reset to the initial vertex. Let X be the index of the segment in which the broadcast-or-reset process terminates. The hypothesis implies that X is geometrically distributed with parameter at least $1 - p$, so that $\mathbf{E}(X) \leq 1/(1 - p)$.

The broadcast-or-reset process cannot terminate before the broadcast process, so that $T_G \leq kX$. Taking expectations yields

$$\mathbf{E}(T_G) \leq \frac{k}{1 - p}. \tag{2}$$

Since the number of contaminated vertices can be at most doubled at each round, we have the following trivial lower bound. \square

Theorem 5. *For any graph G , $T_G \geq \log_2 n$ with probability 1.*

2.2. *The general upper bound*

We will prove the following:

Theorem 6. *For any connected graph G with n vertices and maximum degree Δ , the broadcast time T_G satisfies*

$$\mathbf{E}(T_G) \leq \frac{e}{e-1}(n-1)(6\Delta+1). \tag{3}$$

The proof of this theorem is a bit involved; we will sketch it before stating and proving a few lemmas.

The probability distribution for the full broadcast time T_G is not known, but, when *conditioned* by the sequence of states visited by the broadcast process, it becomes a sum of independent geometric random variables, for which the parameters are known exactly (Lemma 7). Thus, the conditional expectation of the broadcast time becomes the weight of some trajectory, which is defined as a sum of weights for the visited states. Each individual weight is upper bounded by an expression that only depends on individual rendezvous probabilities (Lemma 2 and Corollary 3), and then a uniform upper bound is obtained for the conditional expectations (Lemma 9); this uniform upper bound then straightforwardly translates into an upper bound for the (unconditional) expected broadcast time.

The next lemma is stated in a more general setting than our broadcasting process.

Lemma 7. *Let $(M_t)_{t \in \mathbb{N}}$ be a homogeneous Markov chain with finite state space S and transition probabilities $(p_{x,y})_{x,y \in S}$.*

Let $(T_k)_{k \in \mathbb{N}}$ denote the increasing sequence of stopping times defined by

$$\begin{aligned} T_0 &= 0 \\ T_{k+1} &= \inf\{t > T_k : M_t \neq M_{T_k}\}, \end{aligned}$$

and let $(M'_k)_{k \in \mathbb{N}}$ be the “trajectory” chain defined by

$$M'_k = \begin{cases} M_{T_k} & \text{if } T_k < \infty, \\ M'_{k-1} & \text{if } T_k = \infty. \end{cases}$$

Then, for any sequence x_0, \dots, x_N such that $x_{k+1} \neq x_k$ and $p_{x_k, x_{k+1}} > 0$ for $0 \leq k \leq N-1$, conditioned on $M'_k = x_k$ for $0 \leq k \leq N$, $\mathcal{T} = (T_{k+1} - T_k)_{0 \leq k \leq N-1}$ is distributed as a vector of independent geometric random variables with respective parameters $1 - p_{x_k, x_k}$.

Proof. The proof is straightforward. Let $\mathbf{t} = (t_0, \dots, t_{N-1})$ be any vector of positive integers. The event $\{\mathcal{T} = \mathbf{t} \wedge M'_k = x_k, 0 \leq k \leq N\}$ has probability

$$\mathbf{P}(M_0 = x_0) \prod_{i=0}^{N-1} p_{x_i, x_i}^{t_i-1} p_{x_i, x_{i+1}}. \tag{4}$$

Summing over all possible vectors \mathbf{t} , we get for the probability that the trajectory matches x_0, \dots, x_N :

$$\mathbf{P}(M'_k = x_k, 0 \leq k \leq N) = \mathbf{P}(M_0 = x_0) \prod_{i=0}^{N-1} \frac{p_{x_i, x_{i+1}}}{1 - p_{x_i, x_i}}. \tag{5}$$

Dividing (4) by (5) yields

$$\mathbf{P}(\mathcal{T} = \mathbf{t} \mid M'_k = x_k, 0 \leq k \leq N) = \prod_{i=0}^{N-1} p_{x_i, x_i}^{t_i - 1} (1 - p_{x_i, x_i}), \tag{6}$$

which is indeed the distribution of a vector of independent geometric variables with the claimed parameters. \square

Corollary 8.

Let \mathcal{V} denote the trajectory of the loopless broadcast process (denoted M' in the statement of Lemma 7). Let $\mathcal{X} = (X_1, \dots, X_m)$ be any possible broadcast sequence, and $\mathcal{C} = (C_1, \dots, C_{m-1})$ the corresponding sequence of cuts. Then

$$\mathbf{E}(T_G \mid \mathcal{V} = \mathcal{X}) = \sum_{k=1}^{m-1} \frac{1}{\mathbf{P}(\mathcal{E}_{C_k})}.$$

Proof. Lemma 7 ensures that, conditioned on $\mathcal{V} = \mathcal{X}$, T_G is distributed as the sum of independent geometric random variables G_1, \dots, G_{m-1} , where G_k has parameter $1 - p_{X_k, X_k} = \mathbf{P}(\mathcal{E}_{C_k})$, which implies expectation $1/\mathbf{P}(\mathcal{E}_{C_k})$. Linearity of expectation yields the claim. \square

Lemma 9. Define the weight of any possible broadcast sequence \mathcal{X} as

$$w(\mathcal{X}) = \sum_{k=1}^{m-1} \frac{1}{\pi(C_k)}. \tag{7}$$

Then

$$w(\mathcal{X}) \leq 6(n - 1)\Delta. \tag{8}$$

Proof. We begin by noting that, since we are looking for a uniform upper bound on the weight, we can assume that $m = n$, which is equivalent to $|X_k| = k$ for all k (recall that in the slowest process, we have at most one new vertex contaminated per round). If such is not the case in a sequence \mathcal{X} , then we can obtain another possible sequence \mathcal{X}' with a higher weight by inserting an additional set X' between any two consecutive sets X_k and X_{k+1} such that $|X_{k+1} - X_k| \geq 2$, with $p_{X_k, X'}$ and $p_{X', X_{k+1}}$ meeting the condition that they are both positive; such an X' always exists, because each edge of every graph has positive probability of being the only rendezvous edge in a given round. This will just add a positive term to the weight of the sequence; thus, the sequence with the maximum weight satisfies $m = n$.

To prove that $\sum_{k=1}^{n-1} 1/\pi(C_k) \leq 6(n - 1)\Delta$, we prove that the integer interval $[1, n - 1]$ can be partitioned into a sequence of smaller intervals, such that, in each interval, the average value of $1/\pi(C_k)$ is at most 6Δ .

Assume that integers 1 to $k - 1$ have been thus partitioned, and let us consider C_k . If $\pi(C_k) \geq 1/(4\Delta)$ (that is, $1/\pi(C_k) \leq 4\Delta < 6\Delta$), we put k into an interval by itself and move on to $k + 1$. We now assume $\pi(C_k) < 1/(4\Delta)$, and set $1/\pi(C_k) = \alpha\Delta$ with $\alpha > 4$.

Let v be the next vertex to be reached by the broadcast after X_k , that is, $\{v\} = X_{k+1} - X_k$. This vertex must have at least one neighbor u in X_k .

Let $d \geq 1$ denote the number of neighbors of v that are in X_k . Each edge incident to v has weight at least $1/(d_v\Delta)$, and d of them are in C_k , so that we have $d/(d_v\Delta) \leq \pi(C_k) = 1/(\alpha\Delta)$, or equivalently,

$$d \leq d_v/\alpha. \tag{9}$$

Thus, $v \in X_{k+1}$ has $d_v - d$ neighbors in $Y_{k+1} = V - X_{k+1}$. Since at most one of them is added to X at each step of the sequence, this means that, for $0 \leq j \leq d_v - d$, Y_{k+1+j} contains at least $d_v - d - j$ neighbors of v . In other words, C_{k+1+j} contains at least $d_v - d - j$ edges that are incident to v , each of which has weight at least $1/(d_v\Delta)$. Consequently,

$$\frac{1}{\pi(C_{k+1+j})} \leq \frac{d_v\Delta}{d_v - d - j} \tag{10}$$

holds for $0 \leq j \leq d_v - d$.

The right-hand side of (10) increases with j , and for $j = \lfloor d_v/4 \rfloor$ (Eq. (9) and $\alpha > 4$),

$$\begin{aligned} \frac{d_v\Delta}{d_v - d - \lfloor d_v/4 \rfloor} &\leq \frac{d_v\Delta}{d_v - 2\lfloor d_v/4 \rfloor} \\ &\leq \frac{d_v\Delta}{\lceil d_v/2 \rceil} \\ &\leq 2\Delta. \end{aligned}$$

Summing (10) over $0 \leq j \leq \lfloor d_v/4 \rfloor$, we obtain

$$\sum_{j=0}^{\lfloor d_v/4 \rfloor} \frac{1}{\pi(C_{k+1+j})} \leq 2\Delta \left(1 + \frac{d_v}{4}\right). \tag{11}$$

Since $d_v \geq \alpha$, we also have $1/\pi(C_k) \leq d_v\Delta$. Adding this to inequality (11), we now get

$$\begin{aligned} \frac{1}{\pi(C_k)} + \sum_{0 \leq j \leq \lfloor d_v/4 \rfloor} \frac{1}{\pi(C_{k+1+j})} &\leq \Delta \left(\alpha + 2 + \frac{d_v}{2}\right) \\ &\leq \Delta \left(2 + \frac{3d_v}{2}\right). \end{aligned}$$

There are $2 + \lfloor d_v/4 \rfloor \geq 1 + \frac{d_v}{4}$ terms in the left-hand side of this inequality, so that the average value of $1/\pi(C_i)$, when i ranges over $[k, k + 1 + \lfloor d_v/4 \rfloor]$, is at most

$$\Delta \frac{2 + \frac{3d_v}{2}}{1 + \frac{d_v}{4}} \leq 6\Delta. \tag{12}$$

This concludes the recursion, and the proof. \square

Proof (Theorem 6).

Let \mathcal{X} be any possible broadcast sequence as in Lemma 9. Applying Corollary 3 to $C = C_k$ and summing over k , we get

$$\sum_k \frac{1}{\mathbf{P}(\mathcal{E}_{C_k})} \leq \frac{e}{e-1} \left(n-1 + \sum_k \frac{1}{\pi(C_k)} \right). \tag{13}$$

By Lemma 9, the right-hand side of (13) is at most

$$\frac{e}{e-1} (n-1 + 6\Delta(n-1)) = \frac{e(n-1)(6\Delta+1)}{e-1}. \tag{14}$$

By Lemma 7, the left-hand side of (13) is the conditional expectation of T_G . The upper bound remains valid upon taking a convex linear combination, so that we get, as claimed,

$$\mathbf{E}(T_G) \leq \frac{e(n-1)(6\Delta+1)}{e-1}. \tag{15}$$

Note. It should be clear that the constants are not best possible, even with our method of proof. They are, however, quite sufficient for our purpose, which is to obtain a uniform bound on the expected broadcast time.

The complete characterization of the distribution of T_G seems difficult and is left open. \square

3. Specific graphs

Theorems 5 and 6 provide lower and upper bounds on the expected contamination time for any graph. In this section, we prove that there exists some graphs for which the bounds can be attained.

The well-known coupon-collector problem (that is the number of trials required to obtain n different coupons if each round one is chosen randomly and independently. See [14] for instance) implies the next lemma:

Lemma 10. *For a star S of n leaves, $\mathbf{E}(T_S) = n \ln n + O(n)$.*

3.1. The l -star graphs

An l -star graph S_l is a graph built with a chain of $l+2$ vertices. Then, to each vertex different from the extremities, $\Delta-2$ leaves are added. Let S_l be a l -star graphs with $n = l(\Delta-1) + 2$ vertices. According to Theorem 6, $\mathbf{E}(T_{S_l}) = O(\Delta n) = O(\frac{n^2}{l})$. On the other hand, the expected number of rounds to get a rendezvous between the centers of two adjacent stars is Δ^2 and, therefore, the expected number of rounds for contaminating all the centers is $\Omega(l\Delta^2) = \Omega(n\Delta)$. As a corollary to this result we have

Proposition 11. *There exists an infinite family \mathcal{F} of graphs with n vertices and maximal degree Δ such that, for any $G \in \mathcal{F}$, $\mathbf{E}(T_G) = \Omega(\Delta n)$.*

It follows that the general upper bound $O(n^2)$ given by Theorem 6 is tight for the any l -star graph with $l \geq 2$ constant.

3.2. Matching the lower bound

To prove that the $\Omega(\ln(n))$ bound is tight, we prove an upper bound that only involves the maximum degree Δ and the diameter D .

Theorem 12. *Let G be any graph with maximum degree $\Delta \geq 3$ and diameter D . Then the expected broadcast time in G , starting from any vertex, is at most $4\Delta^2(\ln 2 + D + D \ln \Delta)$.*

Our proof of this theorem will make use of the following lemma.

Lemma 13. *Fix a constant $p > 0$, and let Z_k denote the sum of k independent geometric random variables with parameter p .*

Then, for any $t \geq k/p$, we have

$$\mathbf{P}(Z_k > t) \leq \exp\left(-\frac{tp}{2}\left(1 - \frac{k}{tp}\right)^2\right).$$

Proof. We will use Hoeffding’s inequality, as recalled in [13], Theorem 2.3: if $(X_i)_{1 \leq i \leq t}$ are independent random variables such that $0 \leq X_i \leq 1$ holds with probability 1 for each i , and $X = \sum_{i=1}^t X_i$ has expected value μ , then, for every positive ϵ ,

$$\mathbf{P}(X \leq (1 - \epsilon)\mu) \leq \exp\left(-\frac{\epsilon^2\mu}{2}\right). \tag{16}$$

Let $(X_i)_{i \geq 1}$ be a sequence of independent Bernoulli trials, each one with probability of success p . Since the index of the first success in such a sequence is geometric with parameter p , the index of the k -th success in this sequence is distributed as Z_k . Thus,

$$\mathbf{P}(Z_k > t) = \mathbf{P}\left(\sum_{i=1}^t X_i < k\right). \tag{17}$$

Here we have $\mu = tp$, so that the right-hand side of (17) is of the form $\mathbf{P}(X \leq (1 - \epsilon)\mu)$ with $\epsilon = 1 - \frac{k}{tp}$, provided $t \geq k/p$. Applying (16) yields the claimed upper bound. \square

Proof (Theorem 12). We prove that the probability for the broadcast time to exceed half of the claimed bound is at most $1/2$ and then use Lemma 4.

Let u be the initial vertex for the broadcast. For each other vertex v , pick a path γ_{uv} from u to v with length at most D . Since all degrees are at most Δ , each edge in γ_{uv} has a rendezvous probability at least $1/\Delta^2$. Hence, the broadcast time from u to v along the path γ_{uv} (that is, the time until the first edge has a rendezvous, then the second edge, and so on) is distributed as the sum of independent geometric random variables with parameters equal to the rendezvous probabilities, and is thus stochastically dominated by the sum of D independent geometric random variables with parameter $1/\Delta^2$.

Let T_{uv} denote the time until broadcast reaches v when the initial vertex is u ; Lemma 13 and the above discussion imply that, for any t ,

$$\mathbf{P}(T_{uv} > t) \leq e^{-\frac{t}{2\Delta^2}\left(1 - \frac{D\Delta^2}{t}\right)^2}. \tag{18}$$

Let n denote the number of vertices in G . Moore’s bound ensures that $n - 1 \leq \Delta^D$.

It is routine to check that, if $t > 2\Delta^2(\ln 2 + D + D \ln \Delta)$, then $t(1 - D\Delta^2/t)^2 > 2\Delta^2(\ln 2 + D \ln \Delta) \geq 2\Delta^2 \ln(2n - 2)$. Thus, for each of the $n - 1$ vertices $v \neq u$, we get

$$\mathbf{P}(T_{uv} > t) \leq e^{-\ln(2n-2)} = \frac{1}{2n - 2}, \tag{19}$$

so that, summing over v , we get

$$\mathbf{P}(T_u > t) \leq \frac{1}{2}. \quad \square \tag{20}$$

Corollary 14. *There exists an infinite family of graphs \mathcal{F} such that, for any $G \in \mathcal{F}$, $\mathbf{E}(T_G) = O(\ln(|V|))$.*

Proof. For any integers $\Delta \geq 3$ and any h , the complete Δ -ary tree with diameter $2h$ has $\Delta \cdot (\Delta - 1)^{h-1} > (\Delta - 1)^h$ leaves (which implies that $\ln |V| \geq h \ln(\Delta - 1)$), and Theorem 12 states that its expected broadcast time is no larger than

$$4\Delta^2 (\ln 2 + 2h + 2h \ln \Delta) < 8\Delta^2 \left(1 + \frac{\ln \Delta}{\ln(\Delta - 1)}\right) \ln |V| + 4\Delta^2 \ln 2.$$

For any fixed $\Delta \geq 3$, this is $O(\ln |V|)$. \square

3.3. The complete graph

It seems also interesting to point out that the complete graph K_n has the minimal (see [15]) expected rendezvous number in a round:

$$\mathbf{E}(N_{K_n}) = \frac{\binom{n}{2}}{(n - 1)^2},$$

which is asymptotically $\frac{1}{2}$. We prove in this section that its expected broadcast time is however $O(n \ln n)$, which is *significantly shorter* than that of the l -star graph with l constant which is $\Omega(n^2)$.

Lemma 15. $\mathbf{E}(T_{K_n}) \leq 2\lambda^{-1}n \ln n + O(n)$.

Proof. We bound the expected contamination time from above, by allowing at most one new contaminated vertex per round. In this new and pessimistic contamination process, we sum up the expected time to increase the number of contaminated vertices by one:

$$E(T_{K_n}) \leq \sum_{k=1}^{n-1} \frac{1}{\mathbf{P}(\mathcal{E}_{C_k})}.$$

Since $\pi(C_k) = \frac{k(n-k)}{(n-1)^2} < 1$, we can apply Lemma 2 with $\mathbf{P}(\mathcal{E}_{C_k}) \geq \lambda\pi(C_k)$. It turns out that

$$E(T_{K_n}) \leq \lambda^{-1}(n - 1)^2 \sum_{k=1}^{n-1} \frac{1}{k(n - k)} = \lambda^{-1}(n - 1)^2 \frac{2}{(n - 1)} H_{n-1}.$$

Since $H_n \sim \ln n + 0.57721\dots + o(1)$ we obtain $E(T_{K_n}) \leq 2\lambda^{-1}(n-1) \ln(n-1) + O(n)$. \square

Moreover, we have:

Lemma 16. *With probability $1 - n^{-1/2}$, $T_{K_n} \geq \frac{1}{2}n \ln n$.*

Proof. Let v_t denote the number of contaminated vertices at round t . Then, we have

$$\mathbf{E}(v_{t+1} \mid v_t = k) = k \left(1 + \frac{n-k}{(n-1)^2} \right)$$

and then

$$\mathbf{E}(v_{t+1} \mid v_t) = v_t \left(1 + \frac{n-v_t}{(n-1)^2} \right) = v_t \left(1 + \frac{n}{(n-1)^2} \right) - \frac{v_t^2}{(n-1)^2} \leq v_t \left(1 + \frac{n}{(n-1)^2} \right),$$

yielding

$$\mathbf{E}(v_t) \leq \left(1 + \frac{n}{(n-1)^2} \right)^t.$$

For any $\alpha < 1$ positive real value, we have $(1 + \frac{n}{(n-1)^2})^t \leq \alpha n$ whenever $t \ln \left(1 + \frac{n}{(n-1)^2} \right) \leq \ln(\alpha n)$. It follows that if $t < t_0(n, \alpha) = (\ln n + \ln \alpha) \frac{(n-1)^2}{n}$ then $\mathbf{E}(v_t) \leq \alpha n$. By the Markov inequality, we have

$$\mathbf{P}(v_t \geq n) = \mathbf{P}(v_t = n) \leq \frac{\mathbf{E}(v_t)}{n} \leq \alpha.$$

By definition, $v_t = n$ if and only if $T_{K_n} \leq t$. Hence

$$\mathbf{P}(T_{K_n} > t) \geq 1 - \alpha$$

and then using again the Markov inequality, we have

$$\mathbf{E}(T_{K_n}) \geq t \mathbf{P}(T_{K_n} > t) \geq (1 - \alpha)t.$$

With $\alpha = \frac{1}{\sqrt{n}}$ and $t = t_0(n, \alpha)$, we obtain:

$$\mathbf{E}(T_{K_n}) \geq \frac{1}{2} \ln n \frac{(n-1)^2}{n} \left(1 - \frac{1}{\sqrt{n}} \right)$$

yielding

$$\mathbf{P} \left(T_{K_n} \geq \frac{1}{2} \ln n \frac{(n-1)^2}{n} \right) \geq 1 - \frac{1}{\sqrt{n}},$$

i.e., with probability $1 - n^{-1/2}$, we have

$$T_{K_n} \geq \frac{1}{2} n \ln n. \quad \square$$

Lemmas 16 and 15 imply:

Proposition 17. $\mathbf{E}(T_{K_n}) = \Theta(n \ln n)$.

3.4. Δ -regular balanced rooted trees with bounded diameter

Lemma 18. *Let G be a Δ -regular balanced complete rooted tree of depth 2. The expected time for the root to contaminate its children is $\Theta(\Delta^2 \ln \Delta)$.*

Proof. This lemma is a variation of the coupon collector problem (See Lemma 10). Let v be the root of tree. The probability of rendezvous in one round for vertex v is $\frac{1}{\Delta}$. Let T_k be the number of rounds required for v to obtain a rendezvous with a new vertex knowing that k is the number of its children already contaminated. This event occurs with probability $\frac{\Delta-k}{\Delta^2}$ and implies $\mathbf{E}(T_k) = \frac{\Delta^2}{\Delta-k}$. Hence, we have

$$\mathbf{E}(T_d) = \sum_{k=0}^{\Delta-1} \mathbf{E}(T_k) = \sum_{k=0}^{\Delta-1} \frac{\Delta^2}{\Delta-k} = \Delta^2 \sum_{k=1}^{\Delta} \frac{1}{k} = \Delta^2 H_{\Delta} = \Theta(\Delta^2 \ln \Delta). \quad \square$$

Theorem 19. *Let G be a Δ -regular balanced complete rooted tree of depth $D/2$ with D even. $\mathbf{E}(T_G) = \Omega(D\Delta^2 \ln \Delta)$.*

Proof. Suppose the broadcast starts from the root v_0 . Let us construct a path $v_0, v_1, v_2, \dots, v_{D/2}$ such that v_i is the last contaminated child of v_{i-1} . T_{v_i} denotes the number of rounds to contaminate v_j by its parent v_{i-1} once v_{i-1} is contaminated. Since $T_G \geq \sum_{i=1}^{D/2} T_{v_i}$ and from Lemma 18, for every $1 \leq i \leq D/2$, $\mathbf{E}(T_{v_i}) = \Theta(\Delta^2 \ln \Delta)$, we have $\mathbf{E}(T_G) \geq \sum_{i=1}^{D/2} \mathbf{E}(T_{v_i}) = \Omega(D\Delta^2 \ln \Delta)$. \square

Theorem 19 proves that there exists a graph for which the upper bound of Theorem 12 is tight.

References

- [1] S. Albers, M. Henzinger, Exploring unknown environments, *SIAM Journal on Computing* 29 (4) (2000) 1164–1188.
- [2] D. Angluin, Local and global properties in networks of processors, in: *Proceedings of the 12th Symposium on theory of computing*, 1980, pp. 82–93.
- [3] L. Barriere, P. Flocchini, P. Fraigniaud, N. Santoro, Capture of an intruder by mobile agents, in: *14th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, 2002, pp. 200–209.
- [4] M.A. Bender, D.K. Slonim, The power of team exploration: two robots can learn unlabeled directed graphs, in: S. Sadeddine (Ed.), *Proceedings of the 35rd Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, CA, 1994, pp. 75–85.
- [5] B. Chlebus, *Handbook on Randomized Computing*, in: P.M. Pardalos, S. Rajasekaran, J.H. Reif, J.D.P. Rolim, (eds.), *Randomized communication in radio networks*, Kluwer Academic (2001). <http://citeseer.nj.nec.com/489613.html>.
- [6] F. Comellas, J. Ozón, J.G. Peters, Deterministic small-world communication networks, *Information Processing Letters* 76 (1–2) (2000) 83–90.
- [7] X. Deng, T. Kameda, C.H. Papadimitriou, How to learn an unknown environment i: the rectilinear case, *Journal of the ACM* 45 (2) (1998) 215–245.
- [8] U. Feige, D. Peleg, P. Raghavan, E. Upfal, Randomized broadcast in networks, *Random Structures and Algorithms* 1 (1990).
- [9] S.M. Hedetniemi, S.T. Hedetniemi, A.L. Liestman, A survey of gossiping and broadcasting in communication networks, *Networks* 18 (1988) 319–349.
- [10] M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed, *Probabilistic Methods for Algorithmic Discrete Mathematics*, Springer, Berlin, 1998.

- [11] R.M. Karp, C. Schindelhauer, S. Shenker, B. Vocking, Randomized rumor spreading, *IEEE Symposium on Foundations of Computer Science* (2000) 565–574.
- [12] N. Lynch, A hundred impossibility proofs for distributed computing, in: *Proceedings of the 8th ACM Symposium on Principles of Distributed Computing PODC*, ACM Press, New York, NY, 1989, pp. 1–28.
- [13] C. McDiarmid, Concentration, pp. 195–248. In Habib et al. [10], 1998.
- [14] R. Motwani, P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, 1995.
- [15] Y. Metivier, N. Saheb, A. Zemmari, Randomized rendezvous, *Trends in Mathematics* (2000) 183–194.
- [16] Y. Metivier, N. Saheb, A. Zemmari, Randomized local elections, *Information Processing Letters* 82 (2002) 313–320.
- [17] N. Rao, S. Karet, W. Shi, S. Iyenagar, *Robot Navigation in Unknown Terrains: Introductory Survey of Non-Heuristic Algorithms*. 1993. <http://citeseer.nj.nec.com/rao93robot.html>.
- [18] G. Tel, *Introduction to Distributed Algorithms*, Cambridge University Press, Cambridge, 2000.