Communication Complexity of Quasirandom Rumor Spreading*

Petra Berenbrink[1], Robert Elsässer[2], and Thomas Sauerwald[1]

[1] Simon Fraser University, Burnaby, Canada
{petra,tsauerwa}@cs.sfu.ca
[2] University of Freiburg, Germany
elsa@informatik.uni-freiburg.de

Abstract. We consider rumor spreading on random graphs and hypercubes in the quasirandom phone call model. In this model, every node has a list of neighbors whose order is specified by an adversary. In step i every node opens a channel to its ith neighbor (modulo degree) on that list, beginning from a randomly chosen starting position. Then, the channels can be used for bi-directional communication in that step. The goal is to spread a message efficiently to all nodes of the graph.

We show three results. For random graphs (with sufficiently many edges) we present an address-oblivious algorithm with runtime $O(\log n)$ that uses at most $O(n \log\log n)$ message transmissions. For hypercubes of dimension $\log n$ we present an address-oblivious algorithm with runtime $O(\log n)$ that uses at most $O(n(\log\log n)^2)$ message transmissions. For hypercubes we also show a lower bound of $\Omega(n \log n / \log\log n)$ on the total number of message transmissions required by any $O(\log n)$ time address-oblivious algorithm in the standard random phone call model. Together with a result of [8], our results imply that for random graphs and hypercubes the communication complexity of the quasirandom phone call model is significantly smaller than that of the standard phone call model. This seems to be surprising given the small amount of randomness used in our model.

1 Introduction

In this paper we consider rumor spreading (a.k.a. randomized broadcast) in random graphs and hypercubes. This problem is motivated by overlay topologies in peer to peer (P2P) systems, in which each node possesses a list of neighboring peers. Our goal is to develop time-efficient rumor spreading algorithms which produce a minimal number of message transmissions and use a small amount of randomness. Since P2P networks are decentralized platforms for sharing data and computing resources, it is very important to provide efficient, simple, and robust rumor spreading algorithms for P2P overlays. Minimization of the number of transmission (communication complexity) is important for applications

* The second author was partially supported by the German Research Foundation (DFG) under contract EL 399/2-1.

M. de Berg and U. Meyer (Eds.): ESA 2010, Part I, LNCS 6346, pp. 134–145, 2010.

such as the maintenance of replicated databases in which often huge amounts of broadcasts are necessary to deal with frequent updates in the system.

We assume the *quasirandom phone call model*, a variant of the *standard phone call model*. Let us first introduce the standard phone call model (also known as random phone call model, see [5]). In this model, each node v may perform perform the following actions in every step: 1) create a new rumor to be spread, 2) establish a communication channel between itself and one randomly chosen neighbor, 3) transmit a message over incident channels (opened by v or by some neighbor of v) and 4) close the channel opened in the current step. Note that open channels can be used for bi-directional communications. *Calling nodes* (i.e., the nodes that opened the channels) can send their messages to their neighbors. These are called `push` transmissions. *Called nodes* can also perform so called pull transmissions, i.e., they send the message to the calling nodes. These transmissions are simply called `pull` transmissions. In the phone call model (both standard and quasirandom) it is assumed that nodes can combine several rumors to one larger message. Nodes can send messages over all their open channels in one time step.

The major challenge for rumor spreading algorithms in the phone call model is to decide whether or not a node should forward the rumor over an open communication channel. An algorithm is called *address-oblivious* (see [15]) if the decision of node v to send a rumor over an open channel (v, w) or (w, v) does not depend on w. However, this decision can depend on the communication partners chosen in earlier rounds or on decisions made so far. Hence, according to such an algorithm a node has to decide whether to use a channel without knowing if the rumor is already known by the neighbor in question. If there are only very few rumors in the network, then many communication channels may be established without ever being used for transmissions. Thus, the phone call model is especially of interest in situations where rumors are frequently generated. Then, the cost of establishing communication channels amortizes over all message transmissions.

In the case of the *quasirandom* phone call model it is assumed that every node has a cyclic list of all its neighbors, whose order is specified by an adversary. At the beginning, each node v chooses a random position in the list, independently of the other nodes. Assume that $1 \leqslant \ell \leqslant d$ is the random choice of node v, where d is the degree of v. Then v communicates in step i with the neighbor $(i + \ell) \bmod d + 1$ from the list. To create the list we assume that the adversary has total knowledge about the topology of the network, but cannot foresee any node's random choice w.r.t. the position selected at the beginning (cf. [6]).

In this paper we show three results. For random graphs (with sufficiently many edges) we present an algorithm with runtime $O(\log n)$ that uses at most $O(n \log \log n)$ message transmissions. For $\log n$-dimensional hypercubes we devise an algorithm with runtime $O(\log n)$ that uses at most $O(n \cdot (\log \log n)^2)$ message transmissions. Both algorithms are oblivious and complete rumor spreading in the quasirandom phone call model with probability $1 - n^{-\Omega(1)}$. For hypercubes we show a lower bound of $\Omega(n \log n / \log \log n)$ on the number of message transmissions required by any $O(\log n)$ time oblivious algorithm in the standard phone call model. In [8] Elsässer shows a similar lower bound for oblivious rumor spreading algorithms on random graphs in the standard phone call model. Hence our results imply that for random graphs and hypercubes the communication complexity of the quasirandom phone call model is substantially smaller than of the standard phone call model. This seems to be surprising given the small amount of randomness used in our model.

1.1 Related Work

Due to space constraints, we consider only results which focus on the analytical study of `push` and `push & pull` algorithms.

Runtime. Most rumor spreading studies analyze the runtime of the `push` algorithm in the standard phone call model for different graph classes. For complete graphs of size n, Pittel [17] shows that (with probability $1 - o(1)$) it is possible to spread a rumor in time $\log_2(n) + \ln(n) + f(n)$, where $f(n)$ is a slow growing function, improving a result of Frieze and Grimmett [12]. In [11], Feige et al. determine asymptotically optimal upper bounds for the runtime on $G(n, p)$ graphs (i.e., traditional Erdös-Rényi random graphs [10]), bounded degree graphs, and hypercubes, which all hold w.h.p.[1]. Recently, Fountoulakis et al. [13] prove a tighter bound for the runtime on sufficiently dense $G(n, p)$ graphs, similar to the result of [17] for complete graphs. Very recently, Chierichetti et al. [4] show that the runtime of the combined `push & pull` model is $O(\Phi^{-1} \cdot \log n \cdot \text{polylog}(\Phi^{-1}))$ w.h.p. for any graph G, where Φ denotes the conductance of G.

In [6], Doerr et al. analyze the so called quasirandom rumor spreading. They show that for hypercubes and $G(n, p)$ graphs $O(\log n)$ steps suffice to inform every node, w.h.p. These bounds are similar to the ones in the standard phone call model (`push` model). The results of [6] are extended to further graph classes with good expansion properties in [7]. Note that in [6, 7] the authors mainly concentrate on the runtime efficiency, and the best known algorithms there require $\Theta(n \log n)$ message transmissions in hypercubes and $G(n, p)$ graphs.

Number of Message Transmissions. Karp et al. [15] note that in complete networks the `pull` approach is inferior to the `push` approach until roughly $n/2$ nodes receive the rumor. Then the `pull` approach becomes superior. They present a `push & pull` algorithm, together with a termination mechanism, which bounds the number of total transmissions to $O(n \log \log n)$ (w.h.p.), and show that this result is asymptotically optimal.

For sparser graphs and the standard phone call model it is not possible to get an oblivious algorithm that uses $O(n \log \log n)$ message transmissions, together with a runtime of $O(\log n)$. In [8] Elsässer considers random $G(n, p)$ graphs and shows a lower bound of $\Omega(n \log n / \log(pn))$ message transmissions for oblivious rumor spreading algorithms with a runtime of $O(\log n)$. For $p > \log^2 n/n$ he

[1] W.h.p. or "with high probability" means with probability at least $1 - n^{-c}$ for some constant $c > 0$.

develops an oblivious algorithm that spreads a rumor in time $O(\log n)$ using $O(n \cdot (\log\log n + \log n/\log(pn)))$ transmissions, w.h.p.

In [9] the authors consider a simple modification of the standard phone call model, called RANDOM[4], where every node is allowed to open a channel to *four different* randomly chosen neighbors in every time step. For $G(n, p)$ graphs with $p > \log^2 n/n$, they show that this modification results in a reduction of the number of message transmissions down to $O(n \log\log n)$. Similar results are shown for random d-regular graphs in [1].

The authors of [2] present an extension of RANDOM[4] which they call RR model. In their model each node has a randomly ordered cyclic list with all its neighbors. In step i, the node opens a communication channel to the ith neighbor in its list. The RR model is the same as the quasirandom model except that the adversarial order is replaced by the random order. The authors present an oblivious algorithm for graphs with very good edge and node expansion properties which has a runtime $O(\log n)$ and which uses $O(n\sqrt{\log n})$ message transmissions, w.h.p. The authors establish a lower bound of $\Omega(n\sqrt{\log n/\log d})$ on the number of message transmissions for oblivious rumor spreading algorithms (assuming a runntime of $O(\log n)$), showing that their upper bound is tight up to a $\sqrt{\log\log n}$ factor if d is polylogarithmic in n. Since on these graphs all time efficient algorithms known so far may lead to a communication overhead of $\Theta(n \log n/\log\log n)$, this result shows that avoiding the re-opening of previously used channels makes it possible to reduce the number of message transmissions per node by almost a quadratic factor.

The algorithms of [2, 8, 9, 15] spread the rumor using `push` transmissions until a constant fraction of the nodes receives the rumor (we call these nodes *informed* in the following). Then the algorithms spread the rumor via `pull` transmissions until every node is informed. To save on communications, the algorithms of [1, 2, 8, 9] only allow each node v a certain number of transmissions which depends on the age the rumor had at the time v received it for the first time.

1.2 Model

In this paper we consider random graphs $G(n, p) = (V, E)$ and hypercubes H_d of dimension d. A random graph $G(n, p)$ consists of n nodes. The probability that any pair of nodes is connected is p. For simplicity, we assume $(\log^2 n)/n \leqslant p \leqslant 2^{o(\sqrt{\log n})}/n$ in this extended abstract, although our results can be generalized to a larger regime of p. The expected number of edges for $G(n, p)$ is $pn \cdot (n-1)$. Let $d(v)$ be the degree of node v and $N(v)$ be the set of neighbors in V. For $S \subset V$, let $N(S)$ be the set of neighbors of nodes in S. Let α be the node expansion value of $G(n, p)$. Then $\alpha = \min_{S \in V, |S| \leqslant n/4} N(S)/|S|$. It is known that for our choice of p, α is a constant close to 1 w.h.p. ([3]).

The d-dimensional hypercube H_d consists of $n = 2^d$ many nodes. A binary string of length 2 is assigned to every node and two nodes are connected if their binary strings differ in *exactly* one bit. Hence, the degree of any node of H_d is d. Note that hypercubes have much smaller expansion than random graphs.

We assume that every node has an estimation of n which is accurate to within a constant factor. We also assume that all nodes have access to a global clock, and that they work synchronously. As communication model we assume a variant of the phone call model. In the standard phone call model (see [5]) in each step t every node can create an arbitrary amount of rumors to be spread. To measure the communication cost we only count the number of message transmissions, i.e., opening a channel is not counted. Following [1, 2, 8, 15], we assume here that new pieces of information are generated frequently in the network, and then the cost of establishing communication channels amortizes over all message transmissions. However, we only concentrate on the distribution and lifetime of a single rumor.

The quasirandom variant of the phone call model considered in this paper was introduced in [6]. In the quasirandom phone call model every node v has a list $\tilde{L}_v = \tilde{L}_v[0], \tilde{L}_v[1], \ldots \tilde{L}_v[d(v)-1]$ of length $d(v)$ with all its neighbors. The order of that list is arbitrary, i.e., it may be determined by an adversary. For spreading the rumor, every node v chooses a random position i_v in the list, independently of the other nodes. For its j-th communication v will open a channel to node $L_v[(i_v + j - 1) \bmod d(v)]$. We define $L_v = L_v[0], L_v[1], \ldots, L_v[d(v)]$ as the list beginning at neighbor i_v.

Nodes that received the rumor will be called *informed*. By I_t (H_t) we denote the set of informed (uninformed) nodes in step t. Furthermore, let I_t^+ be the set of nodes that reveive the rumor *for the first time* in step t. These nodes will also be called *newly informed* nodes.

1.3 Our Contribution

In this paper we show the following results. For random graphs we present an oblivious algorithm (in the quasirandom model) that spreads a rumor in time $O(\log n)$ using $O(n \log\log n)$ message transmissions, w.h.p. Compared to [6], we reduce the number of message transmissions by a factor of $\log n/\log\log n$. Moreover, our upper bound in the quasirandom model is significantly smaller than the lower bound for the standard phone call model (cf. [8]).

For the hypercube we show a result that is slightly weaker than our result for random graphs. We present an oblivious algorithm (which is similar to the algorithm for random graphs) that spreads a rumor in time $O(\log n)$ using $O(n \cdot (\log\log n)^2)$ message transmissions, w.h.p. We also show that in the standard phone call model, any oblivious algorithm with runtime $O(\log n)$ requires $\Omega(n \log n/\log\log n)$ message transmissions. The communication complexity of this problem has not been analyzed before, neither in the standard nor in the quasirandom phone call model. Therefore the best known algorithms require $O(\log n)$ time, but produce $\Omega(n \log n)$ message transmissions. In comparison to that, we reduce the number of message transmissions by a factor of $\log n/(\log\log n)^2$. Again, our algorithm outperforms the lower bound on the communication complexity in the standard phone call model.

Our two results demonstrate that on two important networks, rumor spreading can be done much more efficiently in the quasirandom phone model than in the standard phone call model. From a higher level, the results provide evidence that avoiding previously chosen communication partners is more important than choosing all communication partners independently and uniformly at random.

2 Random Graphs

In this section we present an algorithm with runtime $\mathcal{O}(\log n)$ and communication complexity $\mathcal{O}(n \log\log n)$ for random graphs. Note that all the results presented in this section can be generalized to expanders in which the girth is large ($\Omega(\log\log n)$). The details are omitted in this extended abstract.

2.1 Our Algorithm

We assume that the rumor we want to spread is generated at time 0, i.e., at time t the age of the rumor equals t. The algorithm describes the behavior of the nodes w.r.t. one specific rumor. Each node is, depending on the age of the rumor, in one of the following phases:

In the following algorithm, ρ is a sufficiently large constant.

Phase 0: [age $\leqslant \lceil \rho \log n \rceil$] The node which generates the rumor performs `push` in each step of this phase. No other node transmits the rumor in this phase.

Phase 1: [$\lceil \rho \log n \rceil + 1 \leqslant$ age $\leqslant 2 \cdot \lceil \rho(\log n + 320) \rceil$] Nodes that received the rumor in Phase 0 use the first 320 steps of this phase to perform `push` in each of these steps. If a node receives a rumor for the *first* time in step $t \in \{\lceil \rho \log n \rceil + 1, \ldots, 2 \cdot \lceil \rho \log n \rceil\}$, then the node perform `push` in the *exactly* 320 next steps.

Phase 2: [$2\lceil \rho(\log n + 320) \rceil + 1 \leqslant$ age $\leqslant 2 \cdot \lceil \rho \log n + \rho \log\log n \rceil$] Every informed node performs `push` in every step of this phase.

Phase 3: [$2\lceil \rho \log n + \rho \log\log n \rceil + 1 \leqslant$ age $\leqslant 3 \cdot \lceil \rho \log n \rceil$] Every node which becomes informed in this phase performs `pull`, i.e., it sends the message over all incoming channels. All other informed nodes perform `pull` over all incoming channels with a probability of $1/\log n$.

Phase 4: [$3\lceil \rho \log n \rceil + 1 \leqslant$ age $\leqslant 3 \cdot \lceil \rho \log n + \rho \log\log n \rceil$] All informed nodes perform `pull` transmissions.

It is easy to see that at the end of Phase 0, exactly $\rho \log n + 1$ nodes are informed (Observation 1). In Phase 1 we inform half of the nodes (see Lemma 1). At the end of Phase 2 we have $n \cdot (1 - 2 \log\log n / \log n)$ informed nodes, w.h.p. (Lemma 2). Phase 3 and Phase 4 are analyzed in Lemma 3. There we show that w.h.p. at the end of Phase 4 all nodes are informed.

2.2 Analysis of the Algorithm

For a graph $G(n,p)$ and our choice of p the degree of each node is $np \cdot (1 \pm o(1))$, with probability $1 - n^{-3}$. For simplicity we ignore the $1 \pm o(1)$ factor in our analysis and assume $d = pn$.

Theorem 1. *Assume that* $G = G(n,p)$ *with* $(\log^2 n)/n \leqslant p \leqslant 2^{o(\sqrt{\log n})}/n$. *The algorithm above spreads a rumor in* G *in time* $O(\log n)$ *using* $O(n \log\log n)$ *message transmissions, w.h.p.*

In the rest of this section we will prove the above theorem. The proof is split into several lemmata. It is easy to see that in Phase 0 the node that generated the rumor informs $\rho \log n$ different neighbors, which results in the following observation.

Observation 1. *At the end of Phase 0 there are* $\rho \log n$ *informed nodes.*

Now we concentrate on Phase 1 and show the following lemma.

Lemma 1. *With probability* $1 - n^{-2}$, *at least* $n/2$ *nodes are informed at the end of Phase 1.*

Proof. Assuming that the nodes all have a degree d we show that

1. After the first $\rho \cdot (\log n)/2$ steps at least $6n/d$ nodes are informed, where $\rho > 8$.
2. After $\rho \cdot ((\log n)/2 - 320)$ additional steps we have at least $n/40$ informed nodes.
3. After the last $320 \cdot \rho$ steps we have $n/2$ informed nodes for ρ large enough.

Part 1). This follows from Claim A.1 of [2].

Part 2). In this case the number of informed nodes lies in the range $[6n/d, n/40]$. We show inductively that with a very high probability the number of informed nodes grows by a factor of 2.1 every 160 steps. To do so we divide the time into $\ell = (\rho \cdot ((\log n)/2 - 320)/160$ subphases. For $0 \leqslant i \leqslant \ell$, subphase τ_i starts in step $\rho \cdot (\log n)/2 + 160i + 1$ and ends in step $\rho \cdot (\log n)/2 + 160(i+1)$. Let $I^+_{\tau_i}$ be the newly informed nodes in Subphase τ_i, and I_{τ_i} are the informed nodes at the *beginning* of Subphase τ_i. Note that all nodes in $I^+_{\tau_i}$ perform a `push` transmissions in Subphase τ_{i+1}.

We show by induction that for $0 \leqslant i \leqslant \ell$ we have $|I^+_{\tau_i}| \geqslant 2.1 \cdot |I_{\tau_i}|$, which then implies that $|I^+_{\tau_i}| \geqslant |I_{\tau_{i+1}}|/2$.

Fix a subphase τ_{i+1}. One can show that there are $n/6$ uninformed nodes at the beginning of the subphase such that, with probability $1 - \varepsilon^n$, all of these nodes have at least $|I^+_{\tau_i}| \cdot d/(2n)$ neighbors in the set of nodes $I^+_{\tau_i}$. (Due to space limitations, we do not prove this claim.) Hence, such an uninformed node remains uninformed in the time interval τ_{i+1} with probability at most $(1-160/d)^{|I^+_{\tau_i}| \cdot d/(2n)}$. This holds since the first positions are chosen independently and uniformly at random, and a neighbor misses a specific node in 160 steps with probability $1 - 160/d$. Thus,

$$\begin{aligned}\mathbf{E}\left[|I^+_{\tau_{i+1}}|\right] &\geqslant \left(1-\left(1-\frac{160}{d}\right)^{|I^+_{\tau_i}|\cdot d/(2n)}\right)\cdot\frac{n}{6}\\ &\geqslant \left(1-\left(\frac{1}{e}\right)^{80|I^+_{\tau_i}|/n}\right)\cdot\frac{n}{6} \geqslant \left(1-\left(\frac{1}{e}\right)^{40\cdot|I_{\tau_{i+1}}|/n}\right)\cdot\frac{n}{6}\\ &\geqslant \left(1-\left(1-\frac{1}{n/(40\cdot|I_{\tau_{i+1}}|)+1}\right)\right)\cdot\frac{n}{6} > 2.2\cdot|I_{\tau_{i+1}}|\end{aligned}$$

Here, the third equation uses the induction hyphothesis. Using Azuma-Hoeffding ([16]) we obtain with probability $1 - o(n^{-3})$ that $|I^+_{\tau_{i+1}}| \geqslant 2.1 \cdot |I_{\tau_{i+1}}|$.

Part 3). Now the number of informed nodes lies in the range $[n/40, n/2]$. We divide the time into $\ell = 2\rho$ subphases. For $0 \leqslant i \leqslant \ell$, subphase τ_i starts in step $\rho \cdot (\log n/2 - 320) + 160i + 1$ and ends in step $\rho \cdot (\log n/2 - 320) + 160(i + 1)$. Our goal is to show inductively that for all but the last phase $|I^+_{\tau_i}| \geqslant 2.1 \cdot |I_{\tau_i}|$. In the last phase we inform enough nodes so that half of the nodes are informed at the end of this phase.

Similar to Part 2) we fix a subphase τ_{i+1} and define $H_{\tau_{i+1}}$ as the number of uninformed nodes at the *beginning* of Subphase τ_{i+1} One can show that $|H_{\tau_{i+1}}|/2$ of the uninformed nodes have at least $|I^+_{\tau_i}|d/(2n)$ neighbors in the set of nodes $I^+_{\tau_i}$, with probability $1 - \varepsilon^n$ (again, we omit the prrof of this claim due to space limitations). Such an uninformed node remains uninformed in τ_{i+1} with probability at most $(1 - 160/d)^{|I^+_{\tau_i}|d/(2n)}$. Thus,

$$\begin{aligned}\mathbf{E}\left[|I^+_{\tau_{i+1}}|\right] &\geqslant \left(1-\left(1-\frac{160}{d}\right)^{|I^+_{\tau_i}|d/(2n)}\right)\cdot\frac{|H_{\tau_{i+1}}|}{2}\\ &\geqslant \left(1-\left(\frac{1}{e}\right)^{80|I^+_{\tau_i}|/n}\right)\cdot\frac{|H_{\tau_{i+1}}|}{2} \geqslant \left(1-\left(\frac{1}{e}\right)^{40|I_{\tau_{i+1}}|/n}\right)\cdot\frac{|H_{\tau_{i+1}}|}{2}.\end{aligned}$$

The remainder of the proof is a case analysis depending on $|I_{\tau_{i+1}}|$. If $n/40 \leqslant |I_{\tau_{i+1}}| \leqslant n/10$, then

$$\left(1-\left(\frac{1}{e}\right)^{40|I_{\tau_{i+1}}|/n}\right)\cdot\frac{|H_{\tau_{i+1}}|}{2} \geqslant \left(1-\left(\frac{1}{e}\right)\right)\cdot\frac{9n}{20} \geqslant \frac{2.2\cdot n}{10}.$$

Using the method of bounded independent differences [16] one can show that with probability $1 - o(n^{-3})$ it holds that $|I^+_{\tau_{i+1}}| \geqslant 2.1 \cdot |I(\tau_{i+1})|$. For $n/10 < |I_{\tau_{i+1}}| \leqslant n/6$

$$\left(1-\left(\frac{1}{e}\right)^{40|I_{\tau_{i+1}}|/n}\right)\cdot\frac{|H_{\tau_{i+1}}|}{2} \geqslant \left(1-\left(\frac{1}{e}\right)^{4}\right)\cdot\frac{5n}{12} \geqslant \frac{2.2\cdot n}{6}.$$

Then, with probability $1 - o(n^{-3})$ we have $|I^{+}_{\tau_{i+1}}| \geqslant 2.1 \cdot |I_{\tau_{i+1}}|$ [16].

For $|I_{\tau_{i+1}}| \geqslant n/6$ we get

$$|I_{\tau_{i+1}}| + \left(1 - \left(\frac{1}{e}\right)^{40|I_{\tau_{i+1}}|/n}\right) \cdot \frac{|H_{\tau_{i+1}}|}{2}$$
$$\geqslant |I_{\tau_{i+1}}| + \left(1 - \left(\frac{1}{e}\right)^{40/6}\right) \cdot \left(\frac{n - |I_{\tau_{i+1}}|}{2}\right) \geqslant \frac{41n}{80}.$$

Again, we obtain with probability $1 - o(n^{-3})$ that $|I_{\tau_{i+2}}| > n/2$. □

Lemma 2. *Assume* $\rho \geqslant 30$*. With probability* $1 - n^{-2}$*, there are at most* $(n \cdot 2 \log\log n/ \log n)$ *uninformed nodes at the end of Phase 2.*

Proof. Note that in this phase every informed node performs a `push` transmission in every step. Let T be a random variable defined as the time step in which $|I_T| > n/2$ for the first time (this happens w.h.p. in Phase 1). Let τ be the time interval $[T - 160, T]$. For the sake of this proof we assume that only the nodes of I^{+}_{τ} perform `push` transmissions in this phase. Due to Lemma 1, $|I^{+}_{\tau}| > n/5$.

One can show that, with probability $1 - \varepsilon^n$ ($\varepsilon > 0$ is a constant) there are at most $n \cdot \log\log n/ \log n$ nodes in H_T which have fewer than $d/10$ neighbors in I^{+}_{τ}. After $\rho \log\log n$ additional steps each of the other (uninformed) nodes remains uninformed with probability at least

$$\left(1 - \frac{\rho \log\log n}{d}\right)^{d/10} \leqslant e^{-\rho \log\log n/10} < \log^{-3} n,$$

for $\rho \geqslant 30$. Thus, if there are at most $n \cdot \log\log n/ \log n$ nodes in H_T which have fewer than $d/10$ neighbors in I^{+}_{τ}, the expected number of new informed nodes in Phase 2 is at least

$$\left(|H_T| - \frac{n \log\log n}{\log n}\right) \cdot (1 - \log^{-3} n).$$

Then, using [16] one can show that with probability at least $1 - n^{-2}$, the number of newly informed nodes in this phase is at least

$$\left(|H_T| - \frac{2n \log\log n}{\log n}\right).$$

Hence with probability at least $1 - n^{-2}$, the number of uninformed nodes after this phase is at most $n \cdot 2 \log\log n/ \log n$. □

Finally, we concentrate on Phases 3 and 4.

Lemma 3. *Assume* $\rho \geqslant 30$*. With probability* $1 - n^{-2}$ *all nodes are informed at the end of Phase 4.*

Proof. For a node u and time interval $\tau = [t, t']$, let $L_u(\tau)$ be the set of nodes chosen by u in steps $\tau = t, t+1, \ldots, t'$. Define $t_2 = 3\rho \cdot (\log n + \log\log n)$ as the end of Phase 4, $t_1 = 3 \cdot \rho \log n$ as the beginning of Phase 4, and $t_0 = 2\rho(\log n + \log\log n)$ as the beginning of Phase 3.

First we consider Phase 4 and divide the time interval $[t_1 + 1, t_2]$ into $k' = (t_2 - t_1)/320$ subintervals of length 320. For any $0 \leqslant i \leqslant k' - 1$ we define

$$\tilde{\tau}_i = [t_2 - 320i, t_2 - 320 \cdot (i+1) + 1].$$

For a node v, let

$$U_0(v) = L_v[\tilde{\tau}_0] \text{ and } U_i(v) = \cup_{w \in U_{i-1}} L_w[\tilde{\tau}_i].$$

We can visualize $\cup_{i \leqslant k'-1} U_i(v)$ as tree of depth $k'-1$ rooted in v. The level i nodes are the nodes in $U_i(v)$. Then, one can show that $|U_{k'-1}(v) \cap H_{t_0}| = \Omega(\log^3 n)$ with probability $1 - o(n^{-3})$.

In the following we consider two cases. In the first case, we assume that $\cup_{i \leqslant k'-1} U_i(v) \cap I_{t_0} \neq \emptyset$ for some node v. Then v is informed in Phase 4 since all informed nodes perform `pull` transmissions in that phase. In the second case, let $U_{k'-1}(v) \cap I_{t_0} = \emptyset$. For this case we show that in Phase 4 v will be the root of a communication tree consisting of nodes which are still all uninformed in step t_0. Then we will show that w.h.p. at least one of the leaves of the tree will be informed in Phase 3. The rumor will be propagated to v via the path between v and the informed leaf.

Now we need some additional definitions. We divide the time interval $[t_0+1, t_1]$ into $k'' = (t_1 - t_0)/160$ rounds of length 160. For any $0 \leqslant i \leqslant k'' - 1$

$$\tilde{\tau}'_i = [t_1 - 160i, t_1 - 160 \cdot (i+1) + 1].$$

For $0 \leqslant i \leqslant \rho \log n$, let

$$\begin{aligned} \tilde{U}^H_{-1}(v) &= U_{k'-1}(v) \\ \tilde{U}^H_i(v) &= \cup_{w \in \tilde{U}^H_{i-1}(v)} L_w[\tilde{\tau}'_i] \cap H_{t_0} \\ \tilde{U}^I_i(v) &= \cup_{w \in \tilde{U}^H_{i-1}(v)} L_w[\tilde{\tau}'_i] \cap I_{t_0}. \end{aligned}$$

A node $\tilde{w}_i \in \tilde{U}^I_i(v)$ is connected to a node $\tilde{w}_{-1} \in \tilde{U}^H_{-1}(v)$ by a path $P = (\tilde{w}_i, \ldots, \tilde{w}_0, \tilde{w}_{-1})$, where $\tilde{w}_{i-1}, \ldots, \tilde{w}_0, \tilde{w}_{-1} \in H_{t_0}$, and $\tilde{w}_{j+1} \in L_{\tilde{w}_j}(\tilde{\tau}'_{j+1})$. Now define

$$\tilde{U}^H_{0\to i}(v) = \cup_{j=0}^{i} \tilde{U}^H_j(v) \qquad \text{and} \qquad \tilde{U}^I_{0\to i}(v) = \cup_{j=0}^{i} \tilde{U}^I_j(v).$$

Since $|\tilde{U}^H_{-1}(v)| = \Omega(\log^3 n)$, we can apply the same techniques as in Lemma 1 and obtain that

$$|\tilde{U}^H_i(v) \cup \tilde{U}^I_i(v)| \leqslant 2.1 \cdot |\tilde{U}^H_{i-1}(v)|$$

for any $i \geqslant 1$ as long as $|\tilde{U}^I_{i-1}(v)| = O(\log^2 n)$ and $|\tilde{U}^H_i(v)| < n/40$. However, since $|H_{t_0}| \leqslant 2n \log\log n / \log n$, there exists some $i < k''$ such that $|\tilde{U}^I_{0\to i}(v)| >$

$\rho \log^2 n$. Then, we can argue that every node $u \in \tilde{U}^I_{0\to i}(v)$ performs pull transmissions with probability $1/\log n$. Since for every u a path $(u, \tilde{w}_s, \ldots, \tilde{w}_0, \ldots, v)$ with $s < k''$ exists, which consists of nodes of H_{t_0}, all the nodes on this path will perform pull transmisions. Hence, there is a node $u \in L_{\tilde{w}_s}(\tilde{\tau'}_{s+1})$ which transmits the rumor at the right time, with probability

$$1 - \left(1 - \frac{1}{\log n}\right)^{\rho \log^2 n} = 1 - o(n^{-3}),$$

if ρ is large enough. □

Let us now prove Theorem 1. The correctness (every node gets informed w.h.p.) follows from the lemmata above. It remains to analyze the total number of message transmissions. In Phase 0, the algorithm uses $\mathcal{O}(\log n)$ message transmissions. In Phases 1, 2 and 4, the algorithm uses $\mathcal{O}(n \log\log n)$ message transmissions. By Lemma 2, we know that after Phase 2 at most $\mathcal{O}(n \log\log n/(\log n))$ uninformed nodes remain. These nodes generate at most $\mathcal{O}(n \log\log n)$ message transmissions in Phase 3. Using a Chernoff bound, we can show that the nodes that are informed at the end of Phase 2 use at most $\mathcal{O}(n)$ message transmissions. Hence the total number of message transmissions is $\mathcal{O}(n \log\log n)$. □

3 Hypercubes

In this section we first present a lower bound on the communication complexity for rumor spreading on hypercubes in the standard phone call model and then an upper bound (which is much smaller) in the quasirandom phone call model.

Theorem 2. *Let $G = (V, E)$ be a d-dimensional hypercube with $n = 2^d$ nodes. Any algorithm in the standard phone call model with runtime $O(\log n)$ requires $\Omega(n \log n/\log\log n)$ message transmissions, with probability at least $1 - n^{-\omega(1)}$.*

The proof of Theorem 2 is omitted due to space limitations.

3.1 Our Algorithm

In this section we present our upper bound on the communication complexity for rumor spreading in the quasirandom phone call model. In the algorithm below, the total number of message transmissions is $\mathcal{O}(n(\log\log n)^2)$, which can be shown as in the proof of Theorem 1 above.

Phase 1: [$1 \leqslant age \leqslant \lceil \rho \log n \rceil$] If a node receives a rumor for the *first* time in step $t \in \{\lceil \rho \log n \rceil + 1, \ldots, 2 \cdot \lceil \rho \log n \rceil\}$, then the node performs `push` in the *exactly* $C \log\log n$ next steps.

Phase 2: [$\lceil \rho \log n \rceil + 1 \leqslant age \leqslant 2 \cdot \lceil \rho \log n \rceil$] Every node which becomes informed in this phase performs `pull` over each incoming channel. All other informed nodes perform `pull` with a probability of $1/\log n$ over each incoming channel.

Phase 3: [$2\lceil \rho \log n \rceil + 1 \leqslant age \leqslant 2 \cdot \lceil \rho \log n + \rho(\log\log n)^2 \rceil$] All informed nodes perform `pull` transmissions in every step of this phase.

3.2 Analysis of the Algorithm

Theorem 3. *Assume that H_d is a hypercube of dimension $\log n$. The algorithm above spreads a rumor in H_d in time $O(\log n)$ using $O(n(\log\log n)^2)$ message transmissions, w.h.p.*

The proof of Theorem 3 is omitted due to space limitations. The analysis is similar to the one for random graphs. However, the lack of strong expansion properties makes it more difficult and one has to resort to the special structure and the symmetries of hypercubes.

References

[1] Berenbrink, P., Elsässer, R., Friedetzky, T.: Efficient Randomised Broadcasting in Random Regular Networks with Applications in Peer-to-Peer Systems. In: Proc. of PODC 2008, pp. 155–164 (2008)
[2] Berenbrink, P., Elsässer, R., Sauerwald, T.: Randomised Broadcasting: Memory vs. Randomness. In: Proc. of LATIN 2010 (to appear, 2010)
[3] Bollobás, B.: Random Graphs. Academic Press, London (1985)
[4] Chierichetti, F., Lattanzi, S., Panconesi, A.: Almost Tight Bounds for Rumour Spreading with Conductance. In: Proc. of STOC 2010 (to appear, 2010)
[5] Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., Terry, D.: Epidemic algorithms for replicated database maintenance. In: Proc. of PODC 1987, pp. 1–12 (1987)
[6] Doerr, B., Friedrich, T., Sauerwald, T.: Quasirandom Rumor Spreading. In: Proc. of SODA 2008, pp. 773–781 (2008)
[7] Doerr, B., Friedrich, T., Sauerwald, T.: Quasirandom rumor spreading: expanders, push vs. pull, and robustness. In: Proc. of ICALP 2009, pp. 366–377 (2009)
[8] Elsässer, R.: On the communication complexity of randomized broadcasting in random-like graphs. In: Proc. of SPAA 2006, pp. 148–157 (2006)
[9] Elsässer, R., Sauerwald, T.: The power of memory in randomized broadcasting. In: Proc. of SODA 2008, pp. 218–227 (2008)
[10] Erdős, P., Rényi, A.: On random graphs I. Publ. Math. Debrecen 6, 290–297 (1959)
[11] Feige, U., Peleg, D., Raghavan, P., Upfal, E.: Randomized broadcast in networks. Random Structures and Algorithms 1(4), 447–460 (1990)
[12] Frieze, A.M., Grimmett, G.R.: The shortest-path problem for graphs with random arc-lengths. Discrete Applied Mathematics 10, 57–77 (1985)
[13] Fountoulakis, N., Huber, A., Panagiotou, K.: Reliable broadcasting in random networks and the effect of density. In: Proc. of INFOCOM 2010 (to appear, 2010)
[14] Hagerup, T., Rüb, C.: A guided tour of Chernoff bounds. Information Processing Letters 36(6), 305–308 (1990)
[15] Karp, R., Schindelhauer, C., Shenker, S., Vöcking, B.: Randomized rumor spreading. In: Proc. of FOCS 2000, pp. 565–574 (2000)
[16] McDiarmid, C.: On the method of bounded differences. In: Surveys in Combinatorics. London Math. Soc. Lectures Notes, vol. 141, pp. 148–188. Cambridge Univ. Press, Cambridge (1989)
[17] Pittel, B.: On spreading rumor. SIAM Journal on Applied Mathematics 47(1), 213–223 (1987)