Quasirandom Rumor Spreading: Expanders, Push vs. Pull, and Robustness*

Benjamin Doerr[1] Tobias Friedrich[1,2] Thomas Sauerwald[2]

[1] Max-Planck-Institut für Informatik, Saarbrücken, Germany
[2] International Computer Science Institute, Berkeley, CA, USA

Abstract Randomized rumor spreading is an efficient protocol to distribute information in networks. Recently, a quasirandom version has been proposed and proven to work equally well on many graphs and better for sparse random graphs. In this work we show three main results for the quasirandom rumor spreading model.

We exhibit a natural expansion property for networks which suffices to make quasirandom rumor spreading inform all nodes of the network in logarithmic time with high probability. This expansion property is satisfied, among others, by many expander graphs, random regular graphs, and Erdős-Rényi random graphs.

For all network topologies, we show that if one of the push or pull model works well, so does the other. We also show that quasirandom rumor spreading is robust against transmission failures. If each message sent out gets lost with probability f, then the runtime increases only by a factor of $\mathcal{O}(1/(1-f))$.

1 Introduction

Randomized rumor spreading or *random phone call protocols* are simple randomized epidemic algorithms designed to distribute a piece of information in a network. They build on the simple approach that informed nodes call random neighbors and make them informed (*push model*), or that uninformed nodes call random neighbors and become informed if the neighbor was (*pull model*). In spite of the simple concept, these algorithms succeed in distributing information extremely fast. In contrast to many natural deterministic approaches, they are also highly robust against transmission failures.

Such algorithms have been applied successfully both in the context where a single news has to be distributed from one processor to all others (cf. [11]), and in the one where news may be injected at various nodes at different times. The latter problem occurs when maintaining data integrity in a distributed databases, e.g., name servers in large corporate networks [3, 15]. For a more extensive, but still concise discussion of various central aspects of this area, we refer the reader to the paper by Karp, Schindelhauer, Shenker, and Vöcking [14].

* Tobias Friedrich and Thomas Sauerwald were partially supported by postdoctoral fellowships from the German Academic Exchange Service (DAAD).

1.1 Quasirandom Rumor Spreading

Rumor spreading protocols often assume that all nodes have access to a central clock. The protocols then proceed in rounds, in each of which each node independent from the others can perform certain actions. In the classical randomized rumor spreading protocols, in each round each node contacts a neighbor chosen independently and uniformly at random. In the push model, the latter node becomes informed if the first was, and vice versa in the pull model.

In [4], we proposed a quasirandom version of randomized rumor spreading. It assumes that each node has a (cyclic) list of its neighbors. Except that each node starts at a random position in the list, this list describes the order in which the node contacts its neighbors. We make no particular assumption on the structure of the list. This allows to use any list that is already present to technically organize the communication. In such, this protocol is simpler than the classical, fully-random model.

Surprisingly, even though the amount of independent randomness is greatly reduced, similar or even better results could be shown. For the push model, with high probability $\log n$ rounds suffice to inform all nodes of an n–vertex hypercube or random graph $G \in \mathcal{G}(n,p)$, if $p \geqslant (\ln(n)+\omega(1))/n$. The same results are known for the classical model [8], except that for random graphs this only holds for $p \geqslant (1+\varepsilon)\ln(n)/n$, $\varepsilon > 0$ constant. For smaller p, $\Theta((\log n)^2)$ rounds are necessary.

These theoretical results are complemented by an experimental investigation [5], which observes that the quasirandom model typically needs less time than the fully-random one, e.g., by more than 10% for the 12-dimensional hypercube.

1.2 Our Results

In this paper, we greatly expand the first results of [4]. We *(a)* exhibit a natural expansion property that guarantees that quasirandom rumor spreading succeeds in $\mathcal{O}(\log n)$ iterations, *(b)* prove the surprising result that for each graph, the quasirandom push and pull model need the same time to inform all nodes with probability $1 - n^{-\Theta(1)}$, and *(c)* demonstrate that robustness is no problem for the quasirandom model in spite its greatly reduced use of independent randomness.

Our expansion properties (see Definition 3.1) are fulfilled by expander graphs (defined via the second largest eigenvalue, see Definition 3.4). In consequence, random regular graphs fulfill these properties with probability $1 - o(1)$. However, regularity is not necessary. These expansion properties are also satisfied by random graphs $\mathcal{G}(n,p)$ with probability $1 - o(1)$, where p can be as small as $(\ln(n)+\omega(1))/n$. Such graphs typically have vertices of constant degree and of logarithmic degree. Hence our result also subsumes (and improves in terms of the failure probability) the result on random graphs in [4], which gave a runtime of $\mathcal{O}(\log n)$ with probability $1 - o(1)$.

For all these graphs, we show that with probability $1 - n^{-\gamma}$, where γ can be an arbitrary constant, the quasirandom rumor spreading model succeeds in informing all vertices from a single initially informed one in $\mathcal{O}(\log n)$ rounds. This result holds independent of how the cyclic lists look like. To the best of our knowledge, this is first attempt to analyze rumor spreading on several diverse graph classes (some of which are even far from being regular) altogether.

We then show two results that hold for all graphs. The first concerns the pull model, where vertices call others to retrieve information. This model is traditionally regarded less frequently in the literature, though some beautiful results exist. In particular, it is known that combining both push and pull model can lead to a drastic reduction of the number of messages needed (some restrictions apply to the underlying model, though). The first result of this type is Karp et al. [14].

The pull model has quite different characteristics from the push model. For example, in the push model, the number of informed vertices can at most double each round. In the pull model, an increase by a factor of $\Delta(G)$ is possible.

In spite of these differences, we can show the following surprising result. If for some graph, one of the two quasirandom broadcasting models has the property that for all lists and all starting vertices all other vertices become informed in time T with probability $1-n^{-\Theta(1)}$, then the other variant has this property as well. In consequence, our result that the expansion property implies efficient broadcasting holds as well for the pull model.

We finally analyze the robustness of quasirandom rumor spreading. By robustness we mean that we want the protocol still to work well, even if some transmissions get lost. Since quasirandom rumor spreading uses much fewer independent random bits, some colleagues after publication of [4] expressed the concern that robustness might be a problem here. However, we are now able to show that such problems do not occur. We prove that if each transmission independently fails with probability $f < 1$, the time needed to inform all vertices with high probability increases only by a factor of $\mathcal{O}(1/(1-f))$. This again holds for all graphs.

Due to lack of space, several of our proofs are abbreviated or deferred to the full version of the paper.

2 Precise Model and Preliminaries

In the quasirandom model, each vertex $v \in V$ is equipped with a cyclic permutation $\pi_v \colon \Gamma(v) \to \Gamma(v)$ of its neighbors $\Gamma(v)$. This can also be seen as a list of its neighbors.

At the start of the protocol each vertex chooses a first neighbor i_v uniformly at random from $\Gamma(v)$. This is the neighbor it contacts at time $t = 1$. In each following time step $t = 2, 3, \ldots$, the vertex v contacts a vertex $\pi_v^{t-1}(i_v)$. For the quasirandom push[1] model the result of one vertex contacting another one is as follows. If v was informed at time $t-1$, then $\pi_v^t(i_v)$ becomes informed at time t.

This model slightly deviates from the description in the introduction, where each vertex chooses the starting point on its list only when it gets informed. However, the two variants are clearly equivalent and in the following it will be advantageous to assume that all vertices start contacting their neighbors already when they are uninformed.

We shall analyze how long it takes until a rumor known to a single vertex is broadcasted to all other vertices. We adapt a worst-case view in that we aim at bounds that are independent of the starting vertex and of all the lists present in the model. For the

[1] Here we focus on the push model. In the *pull model* the result of a node u contacting a vertex v is opposite, that is, if v is informed, u gets informed. The differences are discussed in Section 5.

quasirandom model the probability space consists of the initial positions of the fixed neighborhood lists of all vertices.

In the analysis it will occasionally be convenient to assume that a vertex after receiving the rumor does not transfer it on for a certain number of time steps. We call this a *delayed model*. Clearly, delaying only results in other vertices receiving the rumor later. Consequently, the random variable describing the broadcast time of this model strictly dominates the original one. This, of course, also holds if several vertices delay the propagation of the rumor.

We will also need chains of contacting vertices. So, a vertex $u_1 \in V$ contacts another vertex $u_m \in V$ within the time-interval $[a, b]$, if there is a path $(u_1, u_2, \dots, u_m)$ in G and $t_1 < t_2 < \cdots < t_{m-1} \in [a, b]$ such that for all $j \in [1, m-1]$, $\pi_{u_j}^{t_j}(i_{u_j}) = u_{j+1}$.

Throughout the paper, we use the following graph-theoretical notation. Let $n = |V|$ denote the number of vertices. For a vertex v of a graph $G = (V, E)$, let $\Gamma(v) := \{u \in V \mid \{u, v\} \in E\}$ the set of its *neighbors* and by $\deg(v) := |\Gamma(v)|$ its *degree*. For any $S \subseteq V$, let $\deg_S(v) := |\Gamma(v) \cap S|$. Let $\delta := \min_{v \in V} \deg(v)$ be the *minimum degree*, $d := 2|E|/n$ be the *average degree*, and $\Delta := \max_{v \in V} \deg(v)$ be the *maximum degree*. The *distance* $\operatorname{dist}(x, y)$ between vertices x and y is the length of the shortest path from x to y. The *diameter* $\operatorname{diam}(G)$ of a connected graph G is the largest distance between two vertices. We will also use $\Gamma^k(u) := \{v \in V \mid \operatorname{dist}(u, v) = k\}$ and $\Gamma^{\leqslant k}(u) := \{v \in V \mid \operatorname{dist}(u, v) \leqslant k\}$. For sets S we define $\Gamma(S) := \{v \in V \mid \exists u \in S, (u, v) \in E\}$ as the set of neighbors of S.

All logarithms $\log n$ are natural logarithms to the base e. As we are only interested in the asymptotic behavior, we will sometimes assume that n is sufficiently large.

3 Expanding Graphs

Instead of analyzing specific graphs, we distill three simple properties. For these properties we can prove that the quasirandom rumor spreading model succeeds in a logarithmic runtime to inform all vertices. This is independent of which vertex is initially informed and independent of the order of the lists. The properties are as follows.

Definition 3.1 (expanding graphs) *We call a connected graph expanding if the following properties hold:*

- **(P1)** *For all constants C_α with $0 < C_\alpha \leqslant d/2$ there is a constant $C_\beta \in (0, 1)$ such that for any connected $S \subseteq V$ with $3 \leqslant |S| \leqslant C_\alpha\, n/d$, we have $|\Gamma(S) \setminus S| \geqslant C_\beta\, d\, |S|$.*
- **(P2)** *There are constants $C_\delta \in (0, 1)$ and $C_\omega > 0$ such that for any $S \subseteq V$, the number of vertices in S^c which have at least $C_\delta d(|S|/n)$ neighbors in S is at least $|S|^c - \frac{C_\omega n^2}{d|S|}$.*
- **(P3)** *$d = \Omega(\Delta)$. If $d = \omega(\log n)$ then $d = O(\delta)$.*

We will now describe the properties in detail and argue why each of them is intrinsic for the analysis. **(P1)** describes a vertex expansion which means that connected sets have a neighborhood which is roughly in the order of the average degree larger than the set itself. Without this property, the broadcasting process could end up in a set with a tiny

neighborhood and slow down thereby too much. Note that in **(P1)**, C_β depends on C_α. As C_α has to be a constant, the upper limit on C_α only applies for constant d.

(P2) is a certain edge-expansion property implying that a large portion of uninformed vertices have a sufficiently number of informed neighbors. This avoids that the broadcasting process stumbles upon a point when it has informed many vertices but most of the remaining uninformed vertices have very few informed neighbors and therefore only a small chance to get informed. Note that **(P2)** is only useful for $|S| = \Omega(n/d)$ and $|S| \leqslant n/2$.

The last property **(P3)** demands a certain regularity of the graph. It is trivially fulfilled for regular graphs, which most definitions of expanders require. The condition $d = \Omega(\Delta)$ for the case $d = \mathcal{O}(\log n)$ does not limit any of our graph classes below. If the average degree is at most logarithmic, **(P3)** applies no further restrictions. Otherwise, we require δ, d and Δ to be of the same order of magnitude. Without this condition, there could be an uninformed vertex with δ informed neighbors of degree $\omega(\delta)$ which does not get informed in logarithmic time with a good probability. With an additional factor of Δ/δ this could be resolved, but as we aim at a logarithmic bound, we require $\delta = \Theta(\Delta)$ for $d = \omega(\log n)$. Note that we do not require $d = \omega(1)$, but the proof techniques for constant and non-constant average degrees will differ in Section 4.

We describe three important graph classes which are expanding, i.e., satisfy all three properties of Definition 3.1, with high probability.

Random Graphs $\mathcal{G}(n,p)$, $p \geqslant (\log n + \omega(1))/n$. Here, we show that sparse and dense random graphs are expanding with probability $1 - o(1)$. We use the popular random graph model $\mathcal{G}(n,p)$ introduced by Erdős and Rényi [7] where each edge of an n-vertex graph is picked independently with probability p. We distinguish two kinds of random graphs with slightly different properties:

Definition 3.2 (sparse and dense random graph) *We call a random graph $\mathcal{G}(n,p)$* sparse *if $p = (\log n + f_n)/n$ with $f_n = \omega(1)$ and $f_n = \mathcal{O}(\log n)$, and* dense *if $p = \omega(\log(n)/n)$.*

Note that our definition of sparse random graph coincides with the one of Cooper and Frieze [2] who set $p = c_n \log(n)/n$ with $(c_n - 1)\log n \to \infty$ and $c_n = \mathcal{O}(1)$.

Theorem 3.3 *Sparse and dense random graphs are expanding with probability $1 - o(1)$.*

By setting $p = 1$ this also shows that complete graphs are expanding.

Expander Graphs. In order to define a (regular) expander graph formally (see Hoory, Linial, and Wigderson [12] for a survey on expander graphs), we have to introduce a bit of notation. For a d-regular graph, its adjacency matrix A of G is symmetric and has real eigenvalues $d = \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Define $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$.

Definition 3.4 (expander) *We call a d-regular graph $G = (V,E)$ expander if $\lambda(G) \leqslant \min\{d/C, C'\sqrt{d}\}$, where $C > 1, C' > 0$ are arbitrary constants.*

Hence for the case when $d = \mathcal{O}(1)$, Definition 3.4 requires that $\lambda(G) \leqslant d/C$, $C > 1$, which is the classic (algebraic) definition of expander graphs. For larger d, we require the bound $\lambda(G) \leqslant C'\sqrt{d}$. We point out that graphs that satisfy the even stronger condition $\lambda \leqslant 2\sqrt{d-1}$ are called *Ramanujan graphs* and the construction of them has received a lot of attention in mathematics and computer science (cf. Hoory et al. [12]).

Theorem 3.5 *Let G be a d-regular expander. Then G is expanding.*

Random Regular Graphs. Random regular graphs are a natural extension of the classic Erdős-Rényi-Graph model, satisfying the additional property of being regular.

Definition 3.6 (random regular graph) *For any even d, a random d-regular graph $G \in \mathcal{G}(n,d)$ is chosen uniformly among all labeled d-regular graphs with n vertices.*

Using a result of [9], we can prove that a random d-regular graph is also an expander in the sense of Definition 3.4. Hence, applying Theorem 3.5 gives the following.

Theorem 3.7 *A random d-regular graph $G \in \mathcal{G}(n,d)$ is expanding w. p. $1 - o(1)$.*

3.1 Previous Results on Expanding Graphs

We summarize what is known for the runtime of the fully-random and the quasirandom model for expanding graphs as defined in Definition 3.1.

The complete graph is the simplest expanding graph. Frieze and Grimmett [10] and later Pittel [16] analyzed the fully random model on this graph. It was shown that with probability $1 - o(1)$, $\log_2 n + \ln n + \omega(1)$ rounds suffice. For the quasirandom model, [4] proved a bound of $\mathcal{O}(\log n)$ rounds with probability $1 - o(1)$.

Feige et al. [8] showed that on random graphs $G(n,p)$, $p \geq (1+\varepsilon)\log n/n$, the fully random model satisfies a runtime bound of $\mathcal{O}(\log n)$ with probability $1-n^{-1}$. They also showed that this failure probability can be achieved for $p = (\log n + \mathcal{O}(\log\log n))/n$ only in $\Omega(\log^2 n)$ rounds. For the quasirandom model, [4] showed that a runtime of $O(\log n)$ holds with probability $1 - o(1)$, already if $p \geq (\log n + \omega(1))/n$.

For expanders with $\Delta/\delta = \mathcal{O}(1)$, it was shown in [17] that the fully random model completes its broadcast campaign in $\mathcal{O}(\log n)$ rounds with probability $1 - 1/n$. For the quasirandom model, no such results have been known so far.

The situation is the same for random regular graphs. Berenbrink, Elsässer, and Friedetzky [1] investigated the fully-random model on random regular graphs and proved, amongst other results, an upper bound of $\mathcal{O}(\log n)$ with probability $1 - n^{-1}$. However, the runtime of the quasirandom model was not considered therein.

As a unified answer to these open questions, this work shows for all aforementioned graphs a runtime of the quasirandom model of $\mathcal{O}(\log n)$ with probability $1-n^{-\gamma}$, where $\gamma \geqslant 1$ is an arbitrary constant.

4 Analysis of the Quasirandom Push Model

In this section, we prove the following theorem, which is one of the three main results.

Theorem 4.1 *Let $\gamma \geqslant 1$ be a constant. The probability that the quasirandom push model started at an arbitrary vertex of an expanding graph informs all other vertices within $\mathcal{O}(\log n)$ rounds is $1 - \mathcal{O}(n^{-\gamma})$.*

To analyze the propagation process, we decompose it in a forward and a backward part. In the forward part we show that one informed vertex informs $n - \mathcal{O}(n/d)$ vertices in $\mathcal{O}(\log n)$ steps (cf. Theorem 4.2). In the backward part we show that if a vertex is uninformed, $\mathcal{O}(\log n)$ steps earlier, at least $\omega(n/d)$ vertices must be uninformed as well (cf. Theorem 4.7). Combining both yields Theorem 4.1.

We will show that all this holds with probability $1 - n^{-\gamma}$ for any $\gamma \geqslant 1$. As Theorem 4.1 is easy to show for $d = \mathcal{O}(1)$, we handle this case separately in Section 4.3 and now consider the case $d = \omega(1)$. This makes the proofs of the lemmas of this section shorter. Therefore in Sections 4.1 and 4.2 we may use the following adjusted property:

(P3') $d = \omega(1)$ and $d = \Omega(\Delta)$. If $d = \omega(\log n)$ then $d = \mathcal{O}(\delta)$.

As the precise constants will be crucial in parts of the following proofs, we use the following notation. Constants with a lower case Greek letter index (e.g., C_α and C_β) stem from Definition 3.1. Constants without an index or with a numbered index (e.g., C and C_1) are local constants in lemmas. K is used to denote a number of time steps.

4.1 Forward Analysis

Theorem 4.2 *Let $\gamma \geqslant 1$ be a constant. The probability that the quasirandom push model started in a fixed vertex u does not inform $n - \mathcal{O}(n/d)$ vertices within $\mathcal{O}(\log n)$ rounds, is at most $n^{-\gamma}$.*

In our analysis we will use the following two notations for sets of informed vertices. Let I_t be the set of vertices that know the rumor after the t-th step. Let $N_t \subseteq I_t$ be the set of "newly informed" vertices that know the rumor after the t-th step, but have not spread this information yet. The latter set will be especially important as these are the vertices which have preserved their independent random choice.

Each of the following Lemmas 4.3–4.6 examines one phase consisting of several steps. Within each phase, we will only consider information spread from vertices N_{t-1} that became informed in the previous phase.

Let u be (newly) informed at time step 0. To get a sufficiently large set of (newly) informed vertices to start with, we first show how to obtain a set N_t of size $\Theta(\log n)$ within $t = \mathcal{O}(\log n)$ steps. If $d = \omega(\log n)$, it suffices to inform enough vertices in $\Gamma(u)$.Otherwise, we use that **(P1)** implies that the neighborhoods $\Gamma^k(u)$ grow exponentially with k. Since within Δ steps, $\Gamma^k(u)$ can be informed if $\Gamma^{k-1}(u)$ was informed beforehand, the claim follows in this case. More precisely:

Lemma 4.3 *Let $C > 0$ be an arbitrary constant. Then with probability 1 there is a time step $t = \mathcal{O}(\log n)$ such that*

- $|N_t| \geqslant C \log n$,
- $|I_t \setminus N_t| = o(|N_t|)$.

We can now assume that we have a set N_t of size $\Omega(\log n)$. The next step aims at informing $\Omega(n/d)$ vertices. For the very dense case of $d = \Omega(n/\log n)$ it can obviously be skipped. Note that in the following we can always assume that we have not informed *too many* vertices as the number of informed vertices will always at most double in each time step. The following lemma shows that given a set of informed vertices matching the conditions of **(P1)** within a constant number of steps the set of informed vertices increases by a factor strictly larger than one.

Lemma 4.4 *For all constants $\gamma \geqslant 1$ and $C_\alpha > 0$ there are constants $K > 1$, $C_1 > 1$, $C_2 > 1$, and $C_3 \in (3/4, 1)$ such that for all time steps t, if*

- $C_1 \log n \leqslant |I_t| \leqslant C_\alpha\,(n/d)$,
- $|N_t| \geqslant C_3\,|I_t|$,

then with probability $1 - n^{-\gamma}$,

- $|I_{t+K}| \geqslant C_2\,|I_t|$,
- $|N_{t+K}| \geqslant C_3\,|I_{t+K}|$.

To avoid the process dying out, it is important that a large fraction of the vertices is newly informed in each phase. With every application of Lemma 4.4, the number of informed vertices increases by a factor of $C_2 > 1$ which depends on C_β which in turn depends on C_α. As the precondition of the next Lemma 4.5 is $|I_t| = 16\,C_\omega(n/d)$, we choose $C_\alpha = 16\,C_\omega$ in every application of Lemma 4.4. This implies a constant $C_2 > 0$ in every phase and therefore at most $\log_{C_2}\left(16\,C_\omega\,(n/d)\right) = \mathcal{O}(\log n)$ applications of Lemma 4.4 suffice to reach $16\,C_\omega(n/d)$ informed vertices with a constant fraction of them newly informed.

The next aim is informing a linear number of vertices. Note that as long as that is not achieved, **(P2)** says that there is always a large set of uninformed vertices which have many neighbors in N_t. Lemma 4.5 below shows that under these conditions, a phase of a constant number of steps suffices to triple the number of informed vertices.

Lemma 4.5 *For all constants $\gamma \geqslant 1$ there are constants $K > 1$, $C > 1$, and $C_\omega > 0$ such that for all time steps t, if*

- $C \log n \leqslant |I_t| \leqslant n/16$,
- $|I_t| \geqslant 16\,C_\omega(n/d)$,
- $|N_t| \geqslant 3/4\,|I_t|$,

then with probability $1 - n^{-\gamma}$,

- $|I_{t+K}| \geqslant 3\,|I_t|$,
- $|N_{t+K}| \geqslant 3/4\,|I_{t+K}|$.

Applying Lemma 4.5 at most $\mathcal{O}(\log n)$ times, a linear fraction of the vertices gets informed. In a final phase of $\mathcal{O}(\log n)$ steps, one can then inform all but $\mathcal{O}(n/d)$ vertices as shown in the following Lemma 4.6.

Lemma 4.6 *Let $\gamma \geqslant 1$ be a constant and t be a time step such that $|N_t| = \Theta(n)$. Then with probability $1 - n^{-\gamma}$, $|I_{t+\mathcal{O}(\log n)}| = n - \mathcal{O}(n/d)$.*

Combining all above phases, a union bound gives that $|I_{\mathcal{O}(\log n)}| = n - \mathcal{O}(n/d)$ with probability $1 - \mathcal{O}(\log(n)\,n^{-\gamma})$, and Theorem 4.2 follows.

4.2 Backward Analysis

The forward analysis has shown that within $\mathcal{O}(\log n)$ steps, at most $\mathcal{O}(n/d)$ vertices stay uninformed. We now analyze the reverse. The question here is, how many vertices have to be uninformed at time $t - \mathcal{O}(\log n)$ if there is an uninformed vertex at time t? We will show that this is at least $\omega(n/d)$ which finishes the overall proof. For this, we introduce a further piece of notation needed in the backward analysis. We denote by $U_{[t_1,t_2]}(w)$ the set of nodes that contact the vertex w within the time-interval $[t_1, t_2]$.

Theorem 4.7 *Let $\gamma \geqslant 1$ be a constant. If the quasirandom rumor spreading process does not inform a fixed vertex w at some time t, then there are $\omega(n/d)$ uninformed vertices at time $t - \mathcal{O}(\log n)$ with probability at least $1 - n^{-\gamma}$.*

To prove Theorem 4.7, we fix an arbitrary vertex w and a time t. Ignoring some technicalities, our aim is now to lower bound the number of vertices which have to be uniformed at times $< t$ to keep w uniformed at time t. As before in Lemma 4.3, we first show the set of uninformed vertices is at least of logarithmic size.

For $d = \mathcal{O}(\log n)$ this follows from **(P1)** as all vertices of $\Gamma^{\mathcal{O}(\log\log n/\log d)}(w) = \Omega(\log n)$ contact w within $\mathcal{O}(\log n)$ steps. For $d = \omega(\log n)$, simple Chernoff bounds show that enough vertices of $\Gamma(w)$ contact w within $\mathcal{O}(\log n)$ steps.

Lemma 4.8 *Let $\gamma \geqslant 1$ and $C \geqslant 1$ be constants, w a vertex, and $t_2 = \Omega(\log n)$ a time step. Then with probability $1 - 2\,n^{-\gamma}$ there is a time step $t_1 = t_2 - \mathcal{O}(\log n)$ such that*

$$|U_{[t_1,t_2]}(w)| \geqslant C\ \log n.$$

We now know that within a logarithmic number of time steps, there are at least $\log n$ vertices which have contacted w. Very similar to Lemmas 4.4 and 4.5 in the forward analysis, we can increase the set of vertices that contact w by a multiplicative factor within a constant number of time steps. The following lemma again mainly draws on **(P1)**. For the very dense case of $d = \Omega(n/\log n)$ it can again be ignored.

Lemma 4.9 *For all constants $\gamma \geqslant 1$ there is a time step K such that for all vertices w and time steps t_1, t_2, if*

$$\log n \leqslant |U_{[t_1,t_2]}(w)| = \mathcal{O}(n/d),$$

then with probability $1 - n^{-\gamma}$,

$$|U_{[t_1-K,t_2]}(w)| \geqslant 4\,|U_{[t_1,t_2]}(w)|.$$

Using Lemma 4.9 at most $\mathcal{O}(\log n)$ times reaches a set of vertices that contact w of size $\Omega(n/d)$. If we have already reached $\omega(n/d)$, we are done. Otherwise, the following Lemma 4.10 shows that a phase consisting of $\mathcal{O}(\log n)$ suffices to reach it. This is the only lemma which substantially draws on **(P3')**.

Lemma 4.10 *Let $\gamma \geqslant 1$ be a constant, w a vertex, and t_1, t_2 time steps such that*

$$|U_{[t_1,t_2]}(w)| = \Theta(n/d).$$

Then with probability $1 - n^{-\gamma}$,

$$|U_{[t_1-\mathcal{O}(\log n),t_2]}(w)| = \omega(n/d).$$

This finishes the backward analysis and shows that $\omega(n/d)$ vertices have to uninformed to keep a single vertex uninformed within $\mathcal{O}(\log n)$ steps. Together with the forward analysis which proved that only $\mathcal{O}(n/d)$ vertices remain uninformed after $\mathcal{O}(\log n)$ steps, this finishes the overall proof of Theorem 4.1 for $d = \omega(1)$.

4.3 Analysis for Graphs with Constant Degree

It remains to show that the quasirandom push model also succeeds on expanding graphs with constant degree $d = \mathcal{O}(1)$. For this, it is easy to see (cf. [4, Theorem 2]) that for any graph the quasirandom push model succeeds in time $\leqslant \Delta \cdot \mathrm{diam}(G)$ with probability 1.[2] Naturally, the diameter of expanding graphs can be bounded easily.

Lemma 4.11 *For any expanding graph G with $d = \mathcal{O}(1)$, $\mathrm{diam}(G) = \mathcal{O}(\log n)$.*

Plugging Lemma 4.11 into the bound $\Delta \cdot \mathrm{diam}(G)$ yields Theorem 4.1 for $d = \mathcal{O}(1)$.

5 Quasirandom Pull Model

In this section, we analyze a second broadcasting model called *pull model*. Its only difference to the push model defined in Section 2 is that here non-informed vertices call other vertices to gain the information. More precisely, let G be a graph equipped with a list of neighbors for each vertex as defined in Section 2. Again, each vertex chooses a random initial neighbor and contacts its neighbors in the order of the list.

The only difference is the result of such a contact. Assume that u and v are vertices in G and that u contacts v. Then if v is informed, u becomes informed (and not vice versa). To avoid misunderstanding, we say that v *asks* u instead of *contacts*, to stress the fact that v is the vertex possibly becoming informed.

Besides being equally natural as the push model, there is a second reason to analyze both models. For the fully-random broadcasting scenario, it is well known that typically the push model works more efficiently when still many vertices are uninformed, whereas the pull model is more efficient in informing the remaining few vertices after the majority is already informed. Combining the two models in a non-trivial manner allows to develop protocols that still work in time $\mathcal{O}(\log n)$, but need fewer messages sent than $\mathcal{O}(n \log n)$. See Karp et al. [14] for more details.

While we shall not go that far and discuss optimizing the number of messages sent, we shall prove bounds for the pull model along. In fact, we shall prove the surprising result that both models are equally efficient. This comes unexpected in the light, e.g., of the following simple example.

Assume that G is a star, that is, a tree with one central vertex c which is neighbor to all other vertices. If c is initially informed, then the quasirandom push model needs exactly $n - 1$ rounds with probability one. For the pull model, things are very different as in this case the quasirandom pull model[3] needs only a single round. At first sight

[2] The corresponding bound for the fully-random model is $\mathcal{O}(\Delta \cdot (\mathrm{diam}(G) + \log n))$ with probability $1 - n^{-1}$ [8, Theorem 2.2].

[3] The corresponding bound for the fully-random push model is $\Theta(n \log n)$ with probability $1 - n^{-1}$ while the fully-random pull model also succeeds within a single round.

this might suggest that the push and pull models are not related at all, but as the initial position is chosen worst-case we can in fact show the following result.

Theorem 5.1 *Let G be a graph such that for all lists and all initially informed vertices, the quasirandom push model needs T rounds to inform all other vertices with probability $1 - n^{-\gamma}$, where $\gamma \geqslant 1$. Then the quasirandom pull model, again for all lists and starting vertices, within T rounds informs all other nodes with probability $1 - n^{-\gamma+1}$. The same holds with the roles of push and pull reversed.*

6 Robustness

To demonstrate that quasirandom rumor spreading is robust against failures, we consider a natural model similar to the one considered in [6, 13]. Here, each sent message reaches its destination only with a certain probability. These failures are assumed to be stochastically independent.

We assume that each vertex sends an acknowledgment to each neighbor from which it has received a correctly sent message. However, also the acknowledged message may not be sent correctly. Only after a vertex has received the acknowledged message, it continues to broadcast the message to the next neighbor on its list, otherwise it tries to send the message to the same neighbor in the next round again. We assume that a node sends a message and receives an acknowledgment with probability f, $0 < f < 1$ (again independently of all other messages). Note that we do not require f to be a constant.

We believe that the assumption of acknowledged messages is inline with practical considerations since the required communication only increases by a constant factor.

Theorem 6.1 *Let G be any graph and let T such that the quasirandom model started at an arbitrary vertex needs T rounds to inform all vertices with probability $1-n^{-\gamma}$, where $\gamma \geqslant 1$. Then the quasirandom model started at an arbitrary vertex needs $4\gamma/(1-f)\,T$ rounds to inform all vertices with probability $1 - 2n^{-\gamma}$.*

7 Conclusion and Outlook

In this work, we made significant progress to understand quasirandom rumor spreading. In particular, we answered the question if it is robust affirmatively, and showed that both push and pull model achieve logarithmic broadcast times on the large class of expanding graphs. From the broader view-point of randomized algorithmics, this work again shows that a reduced amount of randomness can yield superior algorithms, but still can be analyzed with theoretical means.

One open problem is to analyze to what extent push and pull model can be combined to reduce the number of messages needed. We should note, though, that the quasirandom push model naturally never needs more than $2m = nd$ messages. Hence for really sparse networks, e.g., constant-degree expander graphs, the standard quasirandom model is both time and message efficient.

Bibliography

[1] P. Berenbrink, R. Elsässer, and T. Friedetzky. Efficient randomized broadcasting in random regular networks with applications in peer-to-peer systems. In *27th ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 155–164, 2008.

[2] C. Cooper and A. M. Frieze. The cover time of sparse random graphs. *Random Struct. Algorithms*, 30:1–16, 2007.

[3] A. J. Demers, D. H. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. E. Sturgis, D. C. Swinehart, and D. B. Terry. Epidemic algorithms for replicated database maintenance. *Operating Systems Review*, 22:8–32, 1988.

[4] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 773–781, 2008.

[5] B. Doerr, T. Friedrich, M. Künnemann, and T. Sauerwald. Quasirandom rumor spreading: An experimental analysis. In *Proceedings of the 10th Workshop on Algorithm Engineering and Experiments (ALENEX)*, pp. 145–153, 2009.

[6] R. Elsässer and T. Sauerwald. On the Runtime and Robustness of Randomized Broadcasting. In *17th International Symposium on Algorithms and Computation (ISAAC)*, pp. 349–358, 2006.

[7] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[8] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1:447–460, 1990.

[9] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica*, 11:331–362, 1991.

[10] A. Frieze and G. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.

[11] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18:319–349, 1988.

[12] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43:439–561, 2006.

[13] J. Hromkovič, R. Klasing, A. Pelc, P. Ružička, and W. Unger. *Dissemination of Information in Communication Networks*. 2005.

[14] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized Rumor Spreading. In 41*st IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 565–574, 2000.

[15] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *44th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 482–491, 2003.

[16] B. Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47: 213–223, 1987.

[17] T. Sauerwald. On Mixing and Edge Expansion Properties in Randomized Broadcasting. In *18th International Symposium on Algorithms and Computation (ISAAC)*, pp. 196–207, 2007.