Randomized Rounding and Rumor Spreading with Stochastic Dependencies

Dissertation zur Erlangung des Grades
„Doktor der Naturwissenschaften“
der Naturwissenschaftlich-Technischen Fakultäten
der Universität des Saarlandes

vorgelegt von
Anna Huber

Saarbrücken, 2010

Tag des Kolloquiums: 22. September 2010
Dekan der Naturwissenschaftlich-Technischen Fakultät I:
Prof. Dr. Holger Hermanns
Mitglieder des Prüfungsausschusses:
Vorsitzender: Prof. Dr. Markus Bläser
Betreuer: Prof. Dr. Benjamin Doerr
Zweitgutachter: Prof. Dr. Kurt Mehlhorn
Promovierter akademischer Mitarbeiter: Dr. Tobias Friedrich

Contents

Abstract

Randomness is an important ingredient of modern computer science. The present thesis is concerned with two uses of randomness, viz. randomized roundings and randomized rumor spreading algorithms.

The theorem of Beck and Fiala (1981) asserts that for every hypergraph and every set of vertex weights there is a rounding of the vertex weights such that the additive rounding error for all hyperedges is bounded by the maximum degree. In Chapter 2 this theorem will be extended to randomized roundings, that is, to roundings that are efficiently generated at random in such a way that each value is rounded up with probability equal to its fractional part.

The larger part of this thesis deals with randomized rumor spreading algorithms. These are protocols for disseminating information on graphs. The classical randomized rumor spreading was introduced and first investigated by Frieze and Grimmett on the complete graph (1985). In Chapter 3 a generalization of their results both in terms of the model used and in terms of the underlying graph will be shown.

In Chapter 4 a quasirandom rumor spreading protocol introduced by Doerr, Friedrich, and Sauerwald (2008) will be considered. We present a detailed analysis of its evolution and show that its performance and robustness match performance and robustness of the randomized rumor spreading protocol.

The unifying idea is to use dependencies so as to obtain results that are superior or equal to those obtained via independent randomness.

Zusammenfassung

Die Verwendung von Zufallselementen ist ein wichtiger Bestandteil der modernen Informatik. Die vorliegende Arbeit untersucht zwei Bereiche, in denen randomisierte Methoden Verwendung finden, nämlich randomisierte Rundungen und randomisierte Algorithmen zur Gerüchteverbreitung.

Der Satz von Beck und Fiala (1981) sagt aus, dass es für jeden Hypergraphen und für jeden Satz von Knotengewichten eine Rundung gibt derart, dass der Rundungsfehler pro Kante vom Maximalgrad beschränkt wird. Im ersten Teil der Arbeit wird dieser Satz auf den Fall randomisierter Rundungen verallgemeinert, das heißt auf zufällige Rundungen, bei denen jede Zahl mit der Wahrscheinlichkeit entsprechend ihren Nachkommastellen aufgerundet wird.

Der zweite, größere Teil der Arbeit handelt von randomisierten Algorithmen zur Gerüchteverbreitung. Das klassische „Randomized Rumor Spreading“ wurde von Frieze und Grimmett (1985) eingeführt. Ihre Ergebnisse werden in Kapitel 3 sowohl hinsichtlich des Modells als auch hinsichtlich des zugrundegelegten Graphen verallgemeinert.

In Kapitel 4 wird ein quasizufälliges Modell zur Gerüchteverbreitung betrachtet und gezeigt, dass es bezüglich Laufzeit und Robustheit dem klassischen Modell gleichwertig ist.

Gemeinsam liegt beiden Teilen der Arbeit die Idee zugrunde, stochastische Abhängigkeiten zu nutzen um Ergebnisse zu erzielen, die den unter Verwendung stochastischer Unabhängigkeit erzielten gleichwertig oder überlegen sind.

Introduction

Probabilistic techniques and randomized algorithms play a fundamental role in modern computer science. Randomized algorithms are popular because of their simplicity, elegance and efficiency. They are frequently faster on the average than deterministic algorithms and they are also often much easier to implement.

The larger part of the present thesis deals with one important example of randomized algorithms, viz. with randomized rumor spreading protocols. These are protocols for disseminating information in graphs.

Information spreading in large networks is an important topic of study with widespread applications, several of them in distributed systems. Consider, for instance, the task of maintaining replicated databases on name servers in large networks [DGH$^+$87, FPRU90]. Here, the goal is to propagate updates that originate at some specific vertex to all other vertices in the network. This is typically done by means of information exchange between pairs of nodes in the following manner. A pair of neighboring vertices checks whether their copies of the database are in agreement, and subsequently performs the necessary updates. In order to guarantee fast dissemination of the information, it is important that these pairs of vertices are chosen suitably. Moreover it is desirable and often a requirement that the broadcasting algorithms be simple, resilient against failures, and that they operate locally, i. e., the vertices should not require knowledge of the global network topology. Similar broadcasting scenarios have been investigated in the mathematics of infectious diseases (see e. g. [Het00]). Here the desired event is, of course, that not too many nodes become infected. These requirements have also motivated the study of *gossip-based* multicast protocols in distributed networks (see, for instance, [ADH05] and references therein).

A simple, yet powerful approach to disseminate information in graphs is *randomized rumor spreading*, which was introduced by Frieze and Grim-

mett [FG85]. To start with, one vertex of a finite, undirected, connected graph has some piece of information ("rumor"). In each round, every vertex that knows the rumor tells it to a neighbor chosen uniformly at random. As a result, the neighbor vertex now also knows the rumor and begins to gossip in the next round. Besides being self-organized, this approach has two crucial advantages. (i) It is fast. For many important network topologies, $O(\log n)$ rounds suffice to inform all n nodes with high probability. Frieze and Grimmett showed that on the complete graph, the time needed to inform all n vertices is within $(1 \pm o(1))(\log_2 n + \ln n)$ with probability $1 - o(1)$. In Section 3.3 it will be shown that this also holds for random graphs $G_{n,p}$ where $p \geq \frac{\alpha(n) \ln n}{n}$ and $\alpha : \mathbb{N} \to \mathbb{R}$ is any function that tends to infinity. (ii) It is robust on sufficiently dense graphs. In Chapter 3 it will be shown that, for a faulty version of the classical randomized rumor spreading model where the transmission of messages only succeeds with some probability $q \in]0, 1]$, the time needed to inform all vertices of a random graph as above is within $(1 \pm o(1)) \left(\log_{1+q} n + \frac{1}{q} \ln n \right)$ with probability $1 - o(1)$.

There is also a recently growing interest in quasirandomness. This concept describes processes that imitate a particular property of a random process by using less or no randomness. It occurs in several areas of mathematics and computer science. A prominent example are low-discrepancy point sets and quasi-Monte Carlo methods (see e. g. Niederreiter [Nie92]), which proved to be superior to standard random sample point methods in numerical integration.

For randomized rumor spreading, there also exists a quasirandom counterpart. It was introduced by Doerr, Friedrich, and Sauerwald [DFS08]. The basic setup is the same as in the randomized rumor spreading model, where in each time-step every informed vertex contacts one of its neighbors. However, the choices of these neighbors are not stochastically independent. Instead, each vertex has a fixed, cyclic list of its neighbors which dictates the order in which the vertex contacts them. The first neighbor to be contacted is determined by choosing a starting position in this cyclic list at random, independently of the choices of the other vertices. From that point onwards, in each round the vertex contacts one new vertex per round in the order demanded by the list. In that way, all random decisions of one vertex are completely dependent.

In Chapter 4, a detailed analysis of the evolution of the quasirandom protocol on the complete graph will be presented and it will be shown that it

evolves essentially in the same way as the randomized protocol. Section 4.2 presents results on the runtime including lower order terms. More precisely, the number of stages that are needed until all n vertices are informed lies with probability $1 - o(1)$ between $\log_2 n + \ln n - 4 \ln \ln n$ and $\log_2 n + \ln n + \omega(n)$ for any function $\omega : \mathbb{N} \to \mathbb{R}$ that tends to infinity. In Section 4.3 the robustness of the protocol on the complete graph will be investigated and it will be shown that, also in this aspect, quasirandom rumor spreading performs as well as randomized rumor spreading. In particular it will be shown that, for a faulty version of the quasirandom rumor spreading in which the transmission of messages only succeeds with some probability $p \in]0, 1]$, the time needed to inform all n vertices is at most $(1 + o(1))(\log_{1+p} n + \frac{1}{p} \ln n)$ with probability $1 - o(1)$.

Another field where probabilistic techniques play an important role is the method of randomized rounding. This will be investigated in Chapter 2 of the thesis. There a classical result of Beck and Fiala [BF81] about rounding on hypergraphs will be extended to randomized rounding. The theorem of Beck and Fiala states that for every hypergraph and every set of associated vertex weights there is a rounding of the vertex weights such that the sum of the rounding errors inside every hyperedge is bounded from above by the maximum degree of the hypergraph. It will be shown that this theorem can be extended to randomized rounding. A randomized rounding is a random vector which takes only roundings of the original vector as values and which has the original vector as expectation. Similar to the concept of quasirandom rumor spreading, here the main idea is to use clever dependencies to get results that are superior to independent randomized rounding.

Chapter 1 contains some probabilistic tools that are required for the remainder of the thesis.

Chapter 1

Probabilistic Tools

1.1 Chernoff Bounds

Chernoff bounds [Che52] are a basic tool that we will use several times in the course of the thesis. They provide exponentially small bounds for the probability that a binomially distributed random variable deviates significantly from its expected value. This classic result can be found for example in [MU05] in the following form.

1.1 Theorem (Chernoff Bounds). *Let $X_1, \ldots, X_n$ be independent random variables, taking values in $\{0, 1\}$. Let $X := \sum_{i=1}^{n} X_i$ and let $\delta \in]0, 1[$. Then*

$$\Pr\left(X \leq (1-\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{1}{2}\delta^2\,\mathrm{E}(X)},$$

and

$$\Pr\left(X \geq (1+\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{1}{3}\delta^2\,\mathrm{E}(X)}.$$

For random variables which are no longer restricted to two values, but are still independent and bounded, Hoeffding showed that similar bounds hold. Theorem 2 from [Hoe63] yields directly the following upper and lower tail bounds.

1.2 Theorem (Hoeffding Bounds). *Let $X_1, \ldots, X_n$ be independent random variables, and for every $i \in [n] := \mathbb{N}_{\leq n}$ let $a_i, b_i \in \mathbb{R}$ such that $0 \leq a_i < b_i$ and X_i takes values in $[a_i, b_i]$. Let $X := \sum_{i=1}^{n} X_i$, $c := \sum_{i=1}^{n} (b_i - a_i)^2$ and $\delta > 0$.*

Then

$$\Pr\left(X \leq (1-\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{2\delta^2\,\mathrm{E}(X)^2}{c}},$$

and

$$\Pr\left(X \geq (1+\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{2\delta^2\,\mathrm{E}(X)^2}{c}}.$$

There are places where we would like to use Chernoff bounds, but we do not have independence of the random variables. Here results of Panconesi and Srinivasan [PS97, Sri01] show that we may use the classic Chernoff bounds even under the more general assumption that the random variables are negatively correlated. This is defined as follows.

1.3 Definition. *The random variables $X_1, \ldots, X_n$, taking values in Ω, are called* negatively correlated, *if for every subset $I \subseteq [n]$ and every $\omega \in \Omega$ we have*

$$\Pr\left(\bigwedge_{i\in I} X_i = \omega\right) \leq \prod_{i\in I} \Pr\left(X_i = \omega\right).$$

The results of Panconesi and Srinivasan lead to the following generalization of Theorem 1.1.

1.4 Theorem (Chernoff Bounds for negative correlation). *Let $X_1, \ldots, X_n$ be negatively correlated random variables, taking values in $\{0,1\}$. Let $X := \sum_{i=1}^{n} X_i$ and let $\delta \in]0,1[$. Then*

$$\Pr\left(X \leq (1-\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{1}{2}\delta^2\,\mathrm{E}(X)},$$

and

$$\Pr\left(X \geq (1+\delta)\,\mathrm{E}(X)\right) \leq e^{-\frac{1}{3}\delta^2\,\mathrm{E}(X)}.$$

1.2 Sharper Concentration Bounds

A more general tool that we shall apply several times is an inequality by Hoeffding and Azuma [Hoe63, Azu67] which provides strong bounds for the probability that a function defined on a set of independent random variables deviates significantly from its expected value when the value of the function is affected only slightly by changes to only one of its arguments. We will use it in the following version, stated for example in [McD89, Lemma 1.2] or [JŁR00, Corollary 2.27].

1.5 Theorem (Hoeffding–Azuma inequality). *Let $Z_1, \dots, Z_N$ be independent random variables taking values in the sets $\Lambda_1, \dots, \Lambda_N$, respectively. Let $\Lambda = \Lambda_1 \times \cdots \times \Lambda_N$. Let $f : \Lambda \to \mathbb{R}$ be a measurable function, and set $X = f(Z_1, \dots, Z_N)$. Assume that there are quantities c_k, $k = 1, \dots, N$ satisfying the following:*

a. *If $z, z' \in \Lambda$ differ only in the kth coordinate, then $|f(z) - f(z')| \le c_k$.*

Then, denoting $c := \sum_{i=1}^{N} c_i^2$, for every $x \ge 0$ we have that

$$\Pr(|X - \mathrm{E}(X)| \ge x) \le 2e^{-\frac{x^2}{2c}}. \tag{1}$$

Note that the above inequality gives meaningful bounds only if the expectation of X is much larger than $\sqrt{c}$. This condition is unfortunately not always given in our intended applications. In such cases, we will use an estimate given by Talagrand [Tal95], which gives a much stronger tail bound, provided that an additional assumption is satisfied. Intuitively, the statement claims that if the value of X is "witnessed" by only a "small" number of its arguments, then X is sharply concentrated. However, there is a little caveat: the concentration is not guaranteed to be around the expectation, but instead around a *median* of X. A median is a number m such that $\Pr(X < m) \le \frac{1}{2}$ and $\Pr(X > m) \le \frac{1}{2}$. As we shall see below, this is not a significant problem as typically a median is very close the expected value. The following version of Talagrand's inequality can be found in [JŁR00, Theorem 2.29].

1.6 Theorem (Talagrand's Inequality). *Suppose that the preconditions of Theorem 1.5 are satisfied. Additionally, assume that there is an increasing function $\psi : \mathbb{R} \to \mathbb{R}$ satisfying the following condition.*

b. *Let $z \in \Lambda$ and $r \in \mathbb{R}$ such that $f(z) \ge r$. Then there exists a set $J \subseteq [N]$ with $\sum_{i \in J} c_i^2 \le \psi(r)$ such that for all $y \in \Lambda$ with $y_i = z_i$ when $i \in J$, we have $f(y) \ge r$.*

Then, if m is a median of X, for every $x \ge 0$ we have

$$\Pr(|X - m| \ge x) \le 4e^{-\frac{x^2}{4\psi(m+x)}}. \tag{2}$$

The next statement gives a sufficient condition that ensures that a median is very close to the expected value. It can be found in [JŁR00, Example 2.33].

1.7 Proposition. *Let X be a random variable that satisfies the preconditions of Theorem 1.6 with $\psi(r) \leq \lceil r \rceil$. Then*

$$|m - \mathrm{E}(X)| = O\big(\sqrt{\mathrm{E}(X)}\big). \tag{3}$$

The presentation of the above inequalities is as in [JŁR00], where also many applications are presented.

1.3 The Lovász Local Lemma

The following lemma, due to Erdős and Lovász [EL75], can be found for example in [AS08, Corollary 5.1.2] in the following form.

1.8 Lemma (Lovász Local Lemma). *Let $m \in \mathbb{N}$ and let $E_1, \ldots, E_m$ be events with $\Pr(E_i) < p$ for all $i \in [m]$. If each E_i is mutually independent of all but at most d of the other events E_j and if $ep(d+1) \leq 1$, then with a positive probability none of the events $E_1, \ldots, E_m$ occurs.*

Chapter 2

Randomized Rounding on Hypergraphs

2.1 Introduction

The concept of randomized rounding, introduced by Raghavan and Thompson [RT87, Rag88], plays an important role in several areas of computer science. In combinatorial optimization, the problem of approximating an integer program is approached by first solving its relaxation linear program and then rounding the resulting solution. It is also applied in routing problems, see [KLR$^+$87], and in scheduling, see [KMPS05]. We will furthermore give some applications of our main result to controlled roundings in Section 2.3.

The theorem of Beck and Fiala [BF81] states that for every $d \in \mathbb{N}_{\geq 2}$, every finite hypergraph $H = (V, \mathcal{E})$ with maximum degree d and every set of weights $x \in \mathbb{R}^V$ there is a rounding $y \in \mathbb{Z}^V$ of x such that

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} y_i \right| \leq d - 1 \quad \text{for all } E \in \mathcal{E}.$$

It will be shown that this theorem can be extended to randomized rounding. In the case $d \geq 9$ we get the same result (Theorem 2.8), that is, there is a randomized rounding Y of x so that

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| \leq d - 1 \quad \text{for all } E \in \mathcal{E}.$$

We will prove this in Section 2.4.

In the general case $d \geq 2$ we get the only slightly weaker result

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| < d \quad \text{for all } E \in \mathcal{E},$$

see Theorem 2.3.

The idea of the proofs is to use dependent randomized rounding, as opposed to the independent randomized rounding introduced by Raghavan and Thompson. With this idea we can guarantee bounds on the rounding errors.

Related work. A one-sided version of the theorem of Beck and Fiala in the special case of matrices was shown in [KLR+87]. This version was generalized to randomized rounding by Kumar, Marathe, Parthasarathy, and Srinivasan [KMPS05].

Indication of source. The results presented in this chapter are the outcome of joint work with Benjamin Doerr and Christian Klein.

2.2 The General Case

2.1 Definition (Randomized Rounding). *For every finite index set V and every vector $x \in \mathbb{R}^V$, a discrete random vector $Y = (Y_i)_{i \in V}$ is called an* intermediate rounding vector *of x, if Y_i takes only values in $\left[\lfloor x_i \rfloor , \lceil x_i \rceil \right]$ for every $i \in V$ and if $\mathrm{E}(Y) = x$. It is called a* randomized rounding *of x, if furthermore Y_i takes only the values $\lfloor x_i \rfloor$ and $\lceil x_i \rceil$ for every $i \in V$.*

For every intermediate rounding vector Y let

$$U(Y) := \left\{ i \in V \;\middle|\; Y_i \in \left] \lfloor x_i \rfloor , \lceil x_i \rceil \right[\right\}$$

be the subset of indices whose x-values are still unrounded. Note that this subset is itself a random variable.

2.2 Lemma. *Let $H = (V, \mathcal{E})$ be a finite hypergraph and $x \in \mathbb{R}^V$ be a set of weights. If Y is an intermediate rounding vector of x and $\mathcal{G} \subseteq \mathcal{E}$ a random variable such that $|\mathcal{G}| < |U(Y)|$, then there is an intermediate rounding vector Z of x with $U(Y) \supset U(Z)$ and*

$$\sum_{i \in E} Y_i = \sum_{i \in E} Z_i \quad \text{for every } E \in \mathcal{G}. \tag{1}$$

Proof. We may assume $x \in [0,1]^V$. Let Y be an intermediate rounding vector of x, so Y takes only values in $[0,1]^V$. Let $\mathcal{G} \subseteq \mathcal{E}$ be a random variable such that $|\mathcal{G}| < |U(Y)|$. By linear algebra, we know that we can choose $\varepsilon \in \mathbb{R}^{U(Y)} \setminus \{0\}$ with the property

$$\sum_{i \in E \cap U(Y)} \varepsilon_i = 0 \quad \text{for all } E \in \mathcal{G}.$$

We define the random variables t_1 and t_2 as

$$t_1 := \max\left\{ t \in \mathbb{R} \;\middle|\; Y|_{U(Y)} - t\varepsilon \in [0,1]^{U(Y)} \right\},$$

$$t_2 := \max\left\{ t \in \mathbb{R} \;\middle|\; Y|_{U(Y)} + t\varepsilon \in [0,1]^{U(Y)} \right\}.$$

We then define the random vector Z as follows. For $i \in V \setminus U(Y)$ let $Z_i := Y_i$, furthermore let

$$Z|_{U(Y)} := \begin{cases} Y|_{U(Y)} - t_1\varepsilon & \text{with probability } \frac{t_2}{t_1+t_2}, \\ Y|_{U(Y)} + t_2\varepsilon & \text{with probability } \frac{t_1}{t_1+t_2}. \end{cases}$$

For $i \in V \setminus U(Y)$ we have $Z_i = Y_i$ and so $\mathrm{E}(Z_i|Y_i) = Y_i$. For $i \in U(Y)$ we compute $\mathrm{E}(Z_i|Y_i) = (Y_i - t_1\varepsilon_i)\frac{t_2}{t_1+t_2} + (Y_i + t_2\varepsilon_i)\frac{t_1}{t_1+t_2} = Y_i$, so $\mathrm{E}(Z|Y) = Y$. As $\mathrm{E}(Y) = x$, also $\mathrm{E}(Z) = x$ holds. So Z is an intermediate rounding vector of x.

From the maximality of t_1 and t_2 we know
$U(Y) \supset \{i \in V \mid Y_i - t_1\varepsilon_i \in]0,1[\}$ and likewise
$U(Y) \supset \{i \in V \mid Y_i + t_2\varepsilon_i \in]0,1[\}$, and so $U(Y) \supset U(Z)$ holds.

Let $E \in \mathcal{G}$. If $Z|_{U(Y)} = Y|_{U(Y)} - t_1\varepsilon$ one has

$$\sum_{i \in E}(Y_i - Z_i) = \sum_{i \in E \cap U(Y)}(Y_i - Z_i) = \sum_{i \in E \cap U(Y)} t_1\varepsilon_i = t_1 \sum_{i \in E \cap U(Y)} \varepsilon_i = 0.$$

If $Z|_{U(Y)} = Y|_{U(Y)} + t_2\varepsilon$ alike, so (1) holds. □

2.3 Theorem (Randomized Beck-Fiala). *Let $d \in \mathbb{N}_{\geq 2}$. For every finite hypergraph $H = (V, \mathcal{E})$ with maximum degree d, and for every set of weights $x \in \mathbb{R}^V$, there exists a randomized rounding Y of x such that*

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| < d \quad \textit{for all } E \in \mathcal{E}.$$

Proof. We inductively define a sequence $\left(Y^{(k)}\right)_{k\in\mathbb{N}_0}$ of intermediate rounding vectors using Lemma 2.2. Let

$$Y^{(0)} := x.$$

Let $k \in \mathbb{N}_0$ and $Y^{(k)}$ be already defined. Let

$$U_k := U(Y^{(k)}),$$

$$\mathcal{G}_k := \{E \in \mathcal{E} \mid |E \cap U_k| \geq d+1\}.$$

In the case where $\mathcal{G}_k = \emptyset$ let $Y^{(k+1)}$ be a randomized rounding of $Y^{(k)}$ with independent components.

In the case where $\mathcal{G}_k \neq \emptyset$, as d is the maximum degree and every element of $\mathcal{G}_k$ contains at least $d+1$ vertices, we have $|\mathcal{G}_k| < |U_k|$ and thus we can apply Lemma 2.2 with $Y := Y^{(k)}$ and $\mathcal{G} := \mathcal{G}_k$. Let $Y^{(k+1)}$ be the intermediate rounding vector Z assured by Lemma 2.2 to exist.

As $U_0 \supset U_1 \supset \ldots$, we can find a $k \leq |V|$ with $U_k = \emptyset$. This means that $Y_i^{(k)}$ takes only the values $\lfloor x_i \rfloor$ and $\lceil x_i \rceil$ for every $i \in V$. As $Y^{(k)}$ is an intermediate rounding vector, it thus is a randomized rounding of x.

We will show that furthermore $Y^{(k)}$ fulfills the desired property

$$\left|\sum_{i\in E} x_i - \sum_{i\in E} Y_i^{(k)}\right| < d \quad \text{for all } E \in \mathcal{E}.$$

Let $E \in \mathcal{E}$ and $\ell := \min\{m \in \mathbb{N}_0 \mid E \cap U_m \leq d\}$. Then from (1) one has $\sum_{i\in E} Y_i^{(0)} = \sum_{i\in E} Y_i^{(\ell)}$ and so

$$\left|\sum_{i\in E} x_i - \sum_{i\in E} Y_i^{(k)}\right| = \left|\sum_{i\in E}\left(Y_i^{(\ell)} - Y_i^{(k)}\right)\right| = \left|\sum_{i\in E\cap U_\ell}\left(Y_i^{(\ell)} - Y_i^{(k)}\right)\right| < d.$$

So $Y := Y^{(k)}$ satisfies the assertion of the theorem. □

2.3 Application: Controlled Roundings

We will now give some applications of our main result to controlled roundings. The method of controlled roundings describes the fundamental problem of

rounding a table of confidential data in such a way that the readability of the table is improved and the anonymity of the respondents is guaranteed. At the same time, the integrity of the data shall be preserved. This means that row, column, face and possibly other summations of the entries should stay as near as possible to the original values. Nargundkar and Saveland [NS72] first used what we call randomized rounding for this purpose. For a detailed discussion of the problem see [KAG90, KGAB90, KGA90]. We present three applications of Theorem 2.3 for the case of two- and three-dimensional data tables. We get the following results.

2.4 Application. *Let $m, n \in \mathbb{N}$ and $A = (a_{ij})_{\substack{i\in[m]\\ j\in[n]}} \in \mathbb{R}^{m\times n}$. Then there exists a randomized rounding Y of A with*

$$\left|\sum_{i=1}^{m} a_{ij} - \sum_{i=1}^{m} Y_{ij}\right| < 2 \quad \text{for all} \quad j \in [n] \text{ and}$$

$$\left|\sum_{j=1}^{n} a_{ij} - \sum_{j=1}^{n} Y_{ij}\right| < 2 \quad \text{for all} \quad i \in [m].$$

Proof. We apply Theorem 2.3 to the hypergraph

$$H := ([m] \times [n], \{\{i\} \times [n] \mid i \in [m]\} \cup \{[m] \times \{j\} \mid j \in [n]\}).$$

□

2.5 Application. *Let $m, n \in \mathbb{N}$ and $A = (a_{ij})_{\substack{i\in[m]\\ j\in[n]}} \in \mathbb{R}^{m\times n}$. Then there exists a randomized rounding Y of A with*

$$\left|\sum_{j=1}^{n} a_{ij} - \sum_{j=1}^{n} Y_{ij}\right| < 3 \quad \text{for all} \quad i \in [m],$$

$$\left|\sum_{i=1}^{m} a_{ij} - \sum_{i=1}^{m} Y_{ij}\right| < 3 \quad \text{for all} \quad j \in [n] \text{ and}$$

$$\left|\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\right| < 3.$$

Proof. We apply Theorem 2.3 to the hypergraph

$$H := ([m] \times [n], \{\{i\} \times [n] \mid i \in [m]\} \cup \{[m] \times \{j\} \mid j \in [n]\} \cup \{[m] \times [n]\}).$$

□

2.6 Application. *Let $m, n, p \in \mathbb{N}$ and $A = (a_{ijk})_{\substack{i\in[m]\\ j\in[n]\\ k\in[p]}} \in \mathbb{R}^{m\times n}$. Then there exists a randomized rounding Y of A with*

$$\left|\sum_{i=1}^{m} a_{ijk} - \sum_{i=1}^{m} Y_{ijk}\right| < 6 \quad \textit{for all} \quad j \in [n], k \in [p],$$

$$\left|\sum_{j=1}^{n} a_{ijk} - \sum_{j=1}^{n} Y_{ijk}\right| < 6 \quad \textit{for all} \quad i \in [m], k \in [p],$$

$$\left|\sum_{k=1}^{p} a_{ijk} - \sum_{k=1}^{p} Y_{ijk}\right| < 6 \quad \textit{for all} \quad i \in [m], j \in [n],$$

$$\left|\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijk} - \sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ijk}\right| < 6 \quad \textit{for all} \quad k \in [p],$$

$$\left|\sum_{i=1}^{m}\sum_{k=1}^{p} a_{ijk} - \sum_{i=1}^{m}\sum_{k=1}^{p} Y_{ijk}\right| < 6 \quad \textit{for all} \quad j \in [n] \textit{ and}$$

$$\left|\sum_{j=1}^{n}\sum_{k=1}^{p} a_{ijk} - \sum_{j=1}^{n}\sum_{k=1}^{p} Y_{ijk}\right| < 6 \quad \textit{for all} \quad i \in [m].$$

Proof. We apply Theorem 2.3 to the hypergraph

$$\begin{aligned} H := ([m] \times [n] \times [p] \quad , \quad & \{[m] \times \{j\} \times \{k\} \mid j \in [n], k \in [p]\} \\ \cup \quad & \{\{i\} \times [n] \times \{k\} \mid i \in [m], k \in [p]\} \\ \cup \quad & \{\{i\} \times \{j\} \times [p] \mid i \in [m], j \in [n]\} \\ \cup \quad & \{[m] \times [n] \times \{k\} \mid k \in [p]\} \\ \cup \quad & \{[m] \times \{j\} \times [p] \mid j \in [n]\} \\ \cup \quad & \{\{i\} \times [n] \times [p] \mid i \in [m]\}). \end{aligned}$$

□

2.4 Hypergraphs of Maximum Degree at Least 9

In this section, we will show the exact analog of the theorem of Beck and Fiala for hypergraphs with maximum degree at least 9. We start with the special case of uniform and regular hypergraphs in the following lemma. Later, in the proof of Theorem 2.8, we will reduce the general case to this case with the help of Lemma 2.2.

2.7 Lemma. *Let $n \in \mathbb{N}_{\geq 9}$. For every finite n-uniform n-regular hypergraph $H = (V, \mathcal{E})$ and every set of weights $x \in \mathbb{R}^V$ there is a randomized rounding Y of x so that*

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| \leq n - 1 \quad \text{for all } E \in \mathcal{E}.$$

Proof. With a result by Doerr [Doe00, Doe06, Doe07] we may assume $x = (\frac{1}{2}, \ldots, \frac{1}{2})$ and then it is sufficient to show the existence of a randomized rounding Y of x with the property

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| \leq \frac{n-2}{2} < \frac{n-1}{2} \quad \text{for all } E \in \mathcal{E}.$$

To that purpose it suffices to find a subset $S \subseteq V$ with $\emptyset \neq S \cap E \neq E$ for all $E \in \mathcal{E}$, because then we can define Y as follows. With probability $\frac{1}{2}$ we set $Y|_S = 0, Y|_{V \setminus S} = 1$ and with probability $\frac{1}{2}$ we set $Y|_S = 1, Y|_{V \setminus S} = 0$. With this definition we get for every $E \in \mathcal{E}$

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| \leq \left| \frac{n}{2} - 1 \right| = \frac{n-2}{2}.$$

To show the existence of such an $S \subseteq V$ we apply the Lovász Local Lemma (Lemma 1.8).

Let $S \subseteq V$ be chosen at random, that means for all $v \in V$ let v be in S with probability $\frac{1}{2}$ independently. For each $E \in \mathcal{E}$ the event that E has empty or full cut with S has probability 2^{-n+1} and depends only on events concerning hyperedges cutting E, so at most $d := n(n-1)$. As $n \geq 9$ and the function $f(x) := 2^{-x+1}x(x-1)$ is monotonely decreasing on $\mathbb{R}_{\geq 9}$, we have

$$e2^{-n+1}d = e2^{-n+1}n(n-1) \leq e2^{-9+1}9(9-1) < 1$$

and so we can apply the Lovász Local Lemma which yields the existence of an $S \subseteq V$ with the desired properties. □

2.8 Theorem (Randomized Beck-Fiala). *Let $d \in \mathbb{N}_{\geq 9}$. For every finite hypergraph $H = (V, \mathcal{E})$ with maximum degree d and every set of weights $x \in \mathbb{R}^V$ there is a randomized rounding Y of x such that*

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i \right| \leq d - 1 \quad \text{for all } E \in \mathcal{E}.$$

Proof. We will recursively define intermediate rounding vectors $Y^{(k)}$, $k \in \mathbb{N}_0$ of x and show that there is a $k \in \mathbb{N}_0$ such that $Y := Y^{(k)}$ is a randomized rounding of x satisfying the assertion of the theorem. Let

$$Y^{(0)} := x.$$

Let $k \in \mathbb{N}_0$ and $Y^{(k)}$ be already defined with the property

$$\sum_{i \in E} x_i = \sum_{i \in E} Y_i^{(k)} \quad \text{for all } E \in \mathcal{E}. \tag{2}$$

Let

$$U_k := U(Y^{(k)}) \qquad \text{and}$$
$$\mathcal{G}_k := \{E \in \mathcal{E} \mid |E \cap U_k| \geq d\}.$$

As d is the maximum degree, we have $|\mathcal{G}_k| \leq |U_k|$. We do a case distinction whether this inequality is strict or not.

Case (a): $|\mathcal{G}_k| < |U_k|$. We apply Lemma 2.2 with $Y := Y^{(k)}$ and $\mathcal{G} := \mathcal{G}_k$. Let $Y^{(k+1)}$ be the intermediate rounding vector Z assured by Lemma 2.2 to exist.

As long as case (a) applies, we have $V \supseteq U_0 \supset U_1 \supset \ldots$ by Lemma 2.2. So we can find a $k \leq |V|$ with $U_k = \emptyset$, which means that we are in case (b).

Case (b): $|\mathcal{G}_k| = |U_k|$. We can apply Lemma 2.7 to the hypergraph $H = (U_k, \{E \cap U_k \,|\, E \in \mathcal{G}_k\})$, as this is a d-uniform d-regular hypergraph. As weights we take $Y^{(k)}$. Note that here the hypergraph as well as the weights are random variables, so we apply Lemma 2.7 pointwise. Let $Y^{(k+1)}$ be the randomized rounding Y assured by Lemma 2.7 to exist. We have

$$\left| \sum_{i \in E} Y_i^{(k)} - \sum_{i \in E} Y_i^{(k+1)} \right| \leq d - 1 \quad \text{for all } E \in \mathcal{G}_k$$

by Lemma 2.7, and thus, by definition of $\mathcal{G}_k$, also for all $E \in \mathcal{E}$. Together with equation (2), it follows that

$$\left| \sum_{i \in E} x_i - \sum_{i \in E} Y_i^{(k+1)} \right| \leq d - 1 \quad \text{for all } E \in \mathcal{E}.$$

So $Y := Y^{(k+1)}$ satisfies the assertion of the theorem. □

2.5 Open Question

Comparing Theorems 2.3 and 2.8 to the original theorem of Beck and Fiala in its strict version, the question is left whether the original theorem of Beck and Fiala in its strict version can also be extended to randomized rounding for hypergraphs with maximum degree less than 9, that is, if the assertion of Theorem 2.8 also holds for $d \in \{2, \ldots, 8\}$.

Chapter 3

Randomized Rumor Spreading

3.1 Introduction

The study of the dissemination of information within networks has become a topic of intense research in recent years, mainly due to the development and the widespread applications of networks such as the internet or various types of distributed networks. In the latter, for example, there might be a given server/node that wants to pass a certain message to every other node so that the whole network is informed about a certain situation. The main issue one is trying to address is, given a network and a piece of information that is currently held by one of the nodes of the network, what is an efficient method for spreading this all over the network as quickly and as reliably as possible?

A simple but nonetheless nontrivial method for the spread of information within a connected network is *randomized rumor spreading*. This method proceeds in steps or rounds, and as a consequence, we must assume some form of synchronization. We assume that initially only one node of the network possesses a piece of information. During the first step, this vertex informs one uniformly chosen neighbor. Now, if $\mathcal{I}(t)$ denotes the set of informed vertices after the first t steps, then during the $(t+1)$st step every node in $\mathcal{I}(t)$ chooses one of its neighbors uniformly at random, independently of every other vertex in $\mathcal{I}(t)$, and informs it.

This model of dissemination was initially studied by Frieze and Grimmett [FG85] on the complete graph with n vertices, where they proved that all nodes are informed in $(1+o_p(1))(\log_2 n + \ln n)$ steps, where $o_p(1)$ denotes a random variable which converges to 0 in probability, as $n \to \infty$ (that is, for

every $\varepsilon > 0$ we have $\Pr(|o_p(1)| > \varepsilon) = o(1)$). Later, Pittel [Pit87] improved on this, showing that in fact the randomized broadcasting protocol informs all vertices within $\log_2 n + \ln n + O_p(1)$ steps, where $O_p(1)$ denotes a random variable with $|O_p(1)| \leq \omega(n)$ with probability $1 - o(1)$ for every $\omega(n)$ such that $\omega(n) \to \infty$ as $n \to \infty$.

Upper bounds for arbitrary connected graphs were given by Feige, Peleg, Raghavan, and Upfal [FPRU90]. They provide the general bounds of $12n \log n$ and $O(\Delta(G)(\mathrm{diam}(G) + \log n))$ for arbitrary n-vertex graphs. Also, they determined the correct order of magnitude in the case of hypercubes as well as $G_{n,p}$ random graphs, where the edge probability p exceeds the connectivity threshold (see [Bol01]). They determine a broadcast time of $\Theta(\log n)$ with probability $1 - 1/n$. Subsequent work by Sauerwald [Sau07], Elsässer and Sauerwald [ES07] and Berenbrink, Elsässer and Friedetzky [BEF08] shows that the $\Theta(\log n)$ bound also holds for expander graphs, Cayley graphs and random regular graphs.

Note that regardless of the underlying network topology, every dissemination protocol needs at least $\log_2 n$ time-steps, as the number of informed vertices can at most double in each round. Consequently, all the results mentioned above state that randomized rumor spreading is, up to multiplicative constants, an optimal protocol for disseminating information in several important types of networks.

However, it was not at all well-understood, how much the structure of the underlying network affects the performance of randomized rumor spreading. Although, for example, we know from the above-mentioned results in [FPRU90] that on a random graph $G_{n,p}$ the protocol requires with high probability at most $C \ln n$ rounds, for some $C > 0$, we have a priori no bounds that quantify how much slower (or faster?) the protocol is compared to the case where the network is the complete graph. In particular, it is not clear in which way the average degree of the underlying graph influences the speed of the protocol. In Section 3.3 we determine the leading constant for $G_{n,p}$ random graphs, where $p \geq \frac{\alpha(n) \ln n}{n}$ for a function α that tends to infinity as n grows. This result states that the number of steps is essentially unaffected by the density of the underlying graph, thus confirming the robustness and the efficiency of randomized rumor spreading.

For the practical side of the randomized rumor spreading protocol, we refer the interested reader to the aforementioned paper [FPRU90] as well as the paper by Karp, Shenker, Schindelhauer, and Vöcking [KSSV00] for a gen-

eral discussion, or the works of Demers et al. [DGH+88] and Kempe, Dobra, and Gehrke [KDG03] for particular applications. Randomized broadcasting is among the most fundamental and well-studied communication primitives in distributed computing, and has also applications in several other disciplines, like e.g. in mathematical theories of epidemics. A particularly popular example [DGH+88] is the maintenance of consistency in a distributed database, which is replicated at many hundreds or thousands of sites in a large, heterogeneous network. Obviously, efficient broadcasting algorithms are crucial in order to ensure that all copies of the database converge quickly and effectively to the same content. There is an enormous amount of literature devoted to the theoretical and experimental evaluation of broadcasting algorithms on several different underlying networks. Our interest in considering random graphs in Section 3.3 is motivated, among other reasons, by P2P (peer-to-peer) systems. The idea of using random graphs appears in some "real-life" networks, like the popular *Gnutella* network [Inc00], or the *Juxtapose* protocol [BW01], which was originally developed by Sun Microsystems. Meanwhile, a considerable amount of work by several research groups aimed at designing many diverse networks for P2P systems that resemble properties of random graphs, see e.g. [JPS06, PRU03, LS03], and at developing protocols that perform efficiently on random (nearly) regular networks [BEF08, Els06, EGS08]. The most relevant properties of P2P networks, and more generally, of communication networks, are high expansion, connectivity, small average degree, and, approximate regularity of the degrees of the nodes. The random graph model has these properties.

The main drawback of the randomized broadcasting method is the amount of randomness used, as in each round each informed vertex must make a random choice. Doerr, Friedrich, and Sauerwald [DFS08] introduced a quasi-random analogue of this method, in order to reduce significantly the amount of random bits that every node uses. This protocol will be investigated in Chapter 4.

The generalized model. Here we describe a general version of randomized rumor spreading, where errors may occur during the transmissions. Initially some information is placed on one of the nodes. In each succeeding step, every informed node passes the information to another node, that it chooses uniformly at random and independently among its neighbors, with probability $q \in]0, 1]$.

As far as we are aware, the only previous results on the robustness of randomized rumor spreading are due to Elsässer and Sauerwald [ES09]. They assert that the broadcast time for all graphs in this lossy model is at most a factor of $O(1/q)$ larger than in the model without transmission failures.

Our results. In Section 3.2 it will be shown that for any $\varepsilon > 0$ this lossy model needs at least $(1-\varepsilon)\left(\log_{1+q} n + \frac{1}{q}\ln n\right)$ rounds to inform with probability $1-e^{-\Omega(n^{\varepsilon/6})}$ all the nodes of the complete graph on n vertices. This work is under submission [DHL], together with the work described in Section 4.3. A short version has been previously published in [DHL09].

In Section 3.3 we show that if $p \geq \frac{\alpha(n)\ln n}{n}$, where $\alpha(n)$ is any function that tends to infinity as n grows, then the protocol with faulty transmissions on the random graph $G_{n,p}$ broadcasts the message within $(1\pm\varepsilon)\left(\log_{1+q} n + \frac{1}{q}\ln n\right)$ time-steps with probability $1-o(1)$. In other words, in almost every network of density d such that $d \geq \alpha(n)\ln n$, the lossy model broadcasts a message as fast as in a fully connected network and the speed is only affected by the success probability q. This is quite surprising in the sense that the time needed remains essentially unaffected by the fact that most of the links are missing. This work has been previously published in [FHP10], parts of the special case $q = 1$ of it in [FHP09].

Precise description of the model and notations. Let $G = (V, E)$ be a graph on n vertices and let $q \in]0, 1]$. Our model of *randomized rumor spreading with transmission success probability* q works as follows. We assume that initially only one vertex of G is informed. During the first round, this vertex attempts to inform one uniformly chosen neighbor and is successful with probability q. Now, if $\mathcal{I}(t)$ denotes the set of informed vertices after the first t time-steps, then during the $(t+1)$st round every node in $\mathcal{I}(t)$ chooses one of its neighbors uniformly at random, independently of every other vertex in $\mathcal{I}(t)$, and informs it. We set $I(t) := |\mathcal{I}(t)|$. Furthermore, by $\mathcal{U}(t)$ we will denote the set of uninformed vertices after t steps, by $U(t)$ its size, by $\mathcal{N}(t)$ we will denote the set of vertices that are newly informed at a given time-step t and by $N(t)$ its size. Finally, for two real numbers a, b we will write $a \pm b$ for the interval of reals $[a-b, a+b]$, and with slight abuse of notation we will write $X = a \pm b$ to denote $X \in a \pm b$.

3.2 Inclusion of Transmission Failures into the Model: Lower Bound for the Complete Graph

In this section, we analyze the randomized rumor spreading model with transmission success probability q on the complete graph, as this is shorter and easier and may serve as an introduction to the techniques we use. Furthermore, we get a higher probability of validity for our lower bound than in the general case of random graphs in Section 3.3. We provide the following lower bound for the broadcast time.

3.1 Theorem. *Let $\varepsilon > 0$ and $q \in]0,1]$. With probability $1 - e^{-\Omega(n^{\varepsilon/6})}$, the number of rounds we need to inform all the nodes of the complete graph on n vertices using the randomized rumor spreading protocol with transmission success probability q is at least*

$$(1-\varepsilon)\left(\log_{1+q} n + \tfrac{1}{q}\ln n\right).$$

The key to this proof is to split up the rumor spreading process into three phases. The first phase is composed of the rounds that occur between the start of the process and the end of the first round at which point $n^{\varepsilon/2}$ nodes are informed. The second phase begins directly after Phase 1 terminates, and continues until the end of the first round after which $n/4$ nodes are informed. Within each round of Phase 2, the number of informed nodes will grow by a multiplicative factor. The last phase begins directly after Phase 2 terminates, and continues until all the nodes are informed. In this phase we observe a type of coupon collector process.

In order to establish the lower bound posited above, we give lower bounds for the durations of Phases 2 and 3.

3.2 Lemma. *Let $\varepsilon > 0$. With probability $1 - e^{-\Omega(n^{\varepsilon/6})}$, we need more than $(1-\varepsilon)\log_{1+q} n$ rounds to complete Phase 2.*

Proof. Let t_1 denote the number of rounds needed to inform the first $n^{\varepsilon/2}$ nodes. Note that this means that $n^{\varepsilon/2} \leq I(t_1) < 2n^{\varepsilon/2}$. Let $t \geq t_1$. We have $\mathrm{E}(N(t+1)) \leq qI(t)$. Enumerate the nodes in $\mathcal{I}(t)$ from 1 to $I(t)$, and define the indicator random variables $X_1, \ldots X_{I(t)}$ such that for $i \in \{1, \ldots, I(t)\}$ we have

$$X_i = \begin{cases} 1 & \text{if vertex } i \text{ has successfully contacted another vertex} \\ 0 & \text{otherwise.} \end{cases}$$

In this context a successful contact refers only to the transmission of the rumor, regardless of whether or not the contacted vertex was already informed or is also contacted by another vertex. Therefore the random variables $X_1, \ldots, X_{I(t)}$ are independent.

If $X := \sum_{i=1}^{I(t)} X_i$, then $\mathrm{E}(X) = qI(t)$. It is intuitive that $N(t+1) \leq X$, because X not only counts all the nodes in $\mathcal{N}(t+1)$, but also counts nodes multiple times if they are contacted by multiple nodes, and counts nodes that are contacted in round $t+1$ that have already been informed in previous rounds. Therefore, any upper bound we can find on the size of X also holds as an upper bound for $N(t+1)$. But because of the independence of $X_1, \ldots, X_{I(t)}$, the random variable X is a lot easier to handle.

Using Chernoff bounds, we see that

$$\begin{aligned} \Pr\left(X > (1+n^{-\varepsilon/6})\,\mathrm{E}(X)\right) &\leq \exp\left(-\tfrac{1}{3}n^{-\varepsilon/3}qI(t)\right) \\ &\leq \exp\left(-\tfrac{1}{3}n^{-\varepsilon/3}qn^{\varepsilon/2}\right) \\ &= \exp\left(-\tfrac{1}{3}qn^{\varepsilon/6}\right) \\ &= e^{-\Omega(n^{\varepsilon/6})}. \end{aligned}$$

So with probability $1 - e^{-\Omega(n^{\varepsilon/6})}$, the number of nodes informed after $t+1$ rounds satisfies

$$I(t+1) \leq I(t) + (1+n^{-\varepsilon/6})pI(t) \leq (1+n^{-\varepsilon/6})(1+q)I(t).$$

We can therefore infer by using recursion that for every $k \in \mathbb{N}$ we have

$$I(t+k) \leq (1+n^{-\varepsilon/6})^k(1+q)^k I(t)$$

with probability $1 - ke^{-\Omega(n^{\varepsilon/6})}$. Pick $k = (1-\varepsilon)\log_{1+q} n$. Under the assumption that n is sufficiently large, we compute

$$\begin{aligned} I(t_1 + (1-\varepsilon)\log_{1+q} n) &\leq (1+n^{-\varepsilon/6})^{(1-\varepsilon)\log_{1+q} n}(1+q)^{(1-\varepsilon)\log_{1+q} n} 2n^{\varepsilon/2} \\ &= (1+n^{-\varepsilon/6})^{(1-\varepsilon)\log_{1+q} n} n^{1-\varepsilon} 2n^{\varepsilon/2} \\ &\leq \exp\left(n^{-\varepsilon/6}(1-\varepsilon)\log_{1+q} n\right) 2n^{1-\varepsilon/2} \\ &\leq 4n^{1-\varepsilon/2} \\ &< n/4 \end{aligned}$$

with probability $1 - e^{-\Omega(n^{\varepsilon/6})}$. So $(1-\varepsilon)\log_{1+q} n$ rounds are, with probability $1 - e^{-\Omega(n^{\varepsilon/6})}$, not enough to complete Phase 2. □

3.3 Lemma. *Let $\varepsilon > 0$. With probability $1 - e^{-\Omega(n^{\varepsilon})}$, we need more than $(1-\varepsilon)\frac{1}{q}\ln n$ rounds to complete Phase 3.*

Proof. Let t_2 denote the number of rounds needed to inform the first $n/4$ nodes. This means that we have $n/4 \leq I(t_2) < n/2$. Let v be a node that is still uninformed by time t_2. Then

$$\Pr\left(\text{node } v \text{ is uninformed at time } t_2 + k\right) \geq \left(1 - \frac{p}{n-1}\right)^{k(n-1)}.$$

Enumerate the uninformed nodes at time t_2 from 1 to $U(t_2)$. Define the indicator random variables $X_1, \ldots, X_{U(t_2)}$ such that for $i \in \{1, \ldots, U(t_2)\}$ we have

$$X_i = \begin{cases} 1 & \text{if node } i \text{ is uninformed at time } t_2 + (1-\varepsilon)\frac{1}{q}\ln n \\ 0 & \text{otherwise.} \end{cases}$$

Let $X := \sum_{i=1}^{U(t_2)} X_i$. This is the number of uninformed vertices at time $t_2 + (1-\varepsilon)\frac{1}{q}\ln n$. Since $U(t_2) = n - I(t_2) \geq n/2$ and because for small enough

$x > 0$ we have $1 - x \geq e^{-x-x^2}$, we obtain for the expected value

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{i=1}^{U(t_2)} \Pr\left(X_i = 1\right) \\
&\geq \frac{n}{2}\left(1 - \frac{q}{n-1}\right)^{(n-1)(1-\varepsilon)\frac{1}{q}\ln n} \\
&\geq \frac{n}{2}\exp\left(-(1-\varepsilon)\ln n - \frac{q}{n-1}(1-\varepsilon)\ln n\right) \\
&\geq \tfrac{1}{4} n e^{-(1-\varepsilon)\ln n} \\
&= \tfrac{1}{4} n^{\varepsilon},
\end{aligned}$$

again assuming that n is sufficiently large.

The random variables $X_1, \ldots, X_{U(t_2)}$ are negatively correlated. Hence by Lemma 1.4 we get

$$\Pr\left(X = 0\right) \leq \Pr\left(X \leq (1 - \tfrac{1}{2})\,\mathrm{E}(X)\right) \leq e^{-\mathrm{E}(X)/8} \leq e^{-n^{\varepsilon}/32}.$$

So with probability at least $1 - e^{-n^{\varepsilon}/32}$, we see that $(1-\varepsilon)\frac{1}{q}\ln n$ rounds are insufficient to complete Phase 3. □

Proof of Theorem 3.1. Follows immediately from Lemmas 2 and 3. □

Indication of source. The content of this section is under submission [DHL], together with the work described in Section 4.3. A short version has been previously published in [DHL09].

3.3 Performance and Robustness on Random Graphs

In this section we investigate the classical Erdős-Rényi graph $G_{n,p}$, which is obtained by including each of the possible $\binom{n}{2}$ edges that connect any two out of n labeled vertices with probability p, independently of all other edges [ER60]. Let $G_{n,p} = (V, E)$. For the ease of the notation we will assume that $V = [n]$ and that the information is initially placed on node 1. We define $T(G_{n,p})$ as the number of rounds needed by the randomized rumor spreading protocol with transmission success probability q until all vertices have been

informed. Note that our definitions then imply that $\mathcal{I}(0) = \{1\}$. The main result of this section is the following.

3.4 Theorem. *Let $\alpha : \mathbb{N} \to \mathbb{R}$ be any function with the properties $\lim_{n\to\infty} \alpha(n) = \infty$ and $0 < \alpha(n) \leq \ln^{1/9} n$ for every $n \in \mathbb{N}$. Let $p \geq \frac{\alpha(n) \ln n}{n}$. Then w.h.p.*

$$\left| T(G_{n,p}) - \left(\log_{1+q} n + \tfrac{1}{q} \ln n \right) \right| < \alpha(n)^{-1/7} \ln n.$$

In other words, if the average degree of $G_{n,p}$ is slightly larger than $\ln n$, then the broadcast time of randomized rumor spreading essentially remains unaffected by the density of the random graph. In particular, if $q = 1$, the broadcast time coincides with the broadcast time on the complete graph, which was shown in [FG85] to be very close to $\log_2 n + \ln n$. Consequently, the number of rounds needed is not influenced by the fact that most of the links are missing.

The term "w.h.p" (with high probability) denotes a probability of $1-o(1)$. To avoid any confusion, we want to note that in Theorem 3.4 it refers to three independent probability spaces: first, the space from which we sample the underlying network, second, the space of the random choices performed by the nodes, and third, the random faults that occur during the transmission.

Proof ideas & techniques. Before we proceed to a detailed exposition of our proof, let us outline the general strategy. Theorem 3.4 is proved by bounding, for each stage performed by the rumor spreading model, simultaneously from above and from below the number of informed nodes. In particular we show that in the first $(1 - o(1)) \log_{1+q} n$ rounds the number of informed nodes increases roughly by a factor of $1 + q$ in each round. As a result we are able to show that after $(1+o(1)) \log_{1+q} n$ rounds there will be εn informed nodes in total, where $\varepsilon : \mathbb{N} \to \mathbb{R}$ with the property $\lim_{n\to\infty} \varepsilon(n) = 0$ will be defined in the course of our proof. Then things evolve very fast: After only a small number of further steps, the number of nodes having the information will be already at least $(1 - \varepsilon)n$. After that, we show that approximately $\frac{1}{q} \ln n$ additional rounds are necessary and sufficient to spread the information to everybody.

The analysis of the last stages is particularly challenging from a technical point of view, as the number of informed nodes increases only slowly towards the end of the process. In such cases, it is usually difficult to control the deviations of several involved random variables from their expectations. We

achieve this control with the aid of a modern and powerful tool from probability theory called *Talagrand's inequality* [Tal95], which – to our knowledge – has not been applied before in the context of distributed computing problems. We believe that it could be widely applicable to the analysis of existing or future randomized protocols with several different degrees of dependency.

Outline. In Section 3.3.1 we collect and prove the basic properties of $G_{n,p}$ that will be important in the proof of Theorem 3.4, and introduce some necessary notation that will be used throughout. Finally, we present the "core" of the proofs, where the general strategy described above is converted to a rigorous argument.

Notations. For any graph G with vertex set V and any $v \in V$ let $\Gamma_G(v)$ be the set of neighbors of v in G. If the underlying graph is clear, we will write $\Gamma(v)$. Moreover, for $S, S' \subseteq V$ we will denote by $e_G(S, S')$ the number of edges with one endpoint in each of S, S'.

3.3.1 Properties of $G_{n,p}$

Let $\alpha : \mathbb{N} \to \mathbb{R}$ be any function with the properties $\lim_{n\to\infty} \alpha(n) = \infty$ and $\alpha(n) \leq \frac{n}{\ln n}$ for every $n \in \mathbb{N}$ and let $p \geq \frac{\alpha(n) \ln n}{n}$. In this section we collect a few properties of $G_{n,p}$ that we will use in the proof of Theorem 3.4.

Note that for any $S \subset V$, the expected number of neighbors of any $v \in V \setminus S$ in S is $p|S|$. The next lemma says that in $G_{n,p}$ for all large enough S almost all vertices have roughly the right degree in S.

3.5 Lemma. *The random graph $G_{n,p}$ has w.h.p. the following property. For any subset S of its vertices satisfying $|S| \geq \frac{n}{\alpha(n)}$, there is a set $X_S \subset V \setminus S$ that contains at most $\frac{4(\ln \alpha(n)+2)n}{\ln n}$ vertices such that*

$$\forall v \in (V \setminus S) \setminus X_S : |\Gamma_{G_{n,p}}(v) \cap S| = (1 \pm \alpha(n)^{-1/2}) p|S|.$$

Proof. Let S be any fixed subset of the vertices such that $|S| \geq \frac{n}{\alpha(n)}$. Let $\varepsilon := \alpha(n)^{-1/2}$. We call a vertex $v \in V \setminus S$ *violating* with respect to S, if the number of its neighbors in S is $> (1+\varepsilon)p|S|$ or $< (1-\varepsilon)p|S|$. Assume there exist more than $t := \frac{4(\ln \alpha(n)+2)n}{\ln n}$ vertices that are violating, and denote by X_S the set consisting of those vertices.

Note that the expected number of neighbors in S of a vertex is $p|S|$. By applying the Chernoff bounds, we obtain that the probability that a vertex is violating is for large n at most $e^{-\varepsilon^2 p|S|/4}$. Moreover, the events that two distinct vertices are violating are independent, which implies that the probability that there are t violating vertices is bounded from above by $e^{-\frac{\varepsilon^2 p|S|}{4}t}$. Hence, as there are $\binom{n}{k} \leq (\frac{en}{k})^k \leq (e\alpha(n))^k$ ways to choose a set S of size k, the probability that there is a set of size k such that there are t violating vertices with respect to it is at most

$$\exp\left\{k(\ln\alpha(n)+1) - \frac{\varepsilon^2 pk}{4}\cdot t\right\}$$
$$= \exp\left\{k\left(\ln\alpha(n)+1-\frac{\varepsilon^2 p}{4}\cdot\frac{4(\ln\alpha(n)+2)n}{\ln n}\right)\right\}.$$

This, combined with the bound $p \geq \frac{\alpha(n)\ln n}{n}$, can be estimated from above by e^{-k}. The proof is completed by summing this expression up for all $k \geq \frac{n}{\alpha(n)}$. □

The next statement considers a similar setting as before, with the difference that now S might be very small. Here we show that the number of vertices that have many neighbors in S is only $o(|S|)$.

3.6 Lemma. *For any $\varepsilon \geq \alpha(n)^{-1/2}$, the random graph $G_{n,p}$ has w.h.p. the following property. For any subset S of its vertices such that $|S| \leq \frac{n}{\alpha(n)}$ there is a set X_S containing at most $|S|\varepsilon^{-1}\alpha(n)^{-1}$ vertices, such that*

$$\forall v \in (V \setminus S) \setminus X_S : |\Gamma_{G_{n,p}}(v) \cap S| \leq \varepsilon pn.$$

Proof. The proof is similar to the proof of Lemma 3.5, except that here we have to deal with small sets S. We give the whole proof for the sake of completeness. We assume that $|S| \geq \varepsilon pn$, for otherwise the statement holds trivially.

Let S be any fixed subset of the vertices such that $|S| \leq \frac{n}{\alpha(n)}$. We call a vertex $v \in V \setminus S$ *violating* with respect to S, if the number of its neighbors in S is $> \varepsilon pn$. Assume there exist more then $t := \frac{|S|}{\varepsilon\alpha(n)}$ vertices that are violating, and denote by X_S the set consisting of those vertices.

The expected number of neighbors in S of a vertex $v \in V \setminus S$ is $p|S| = o(\varepsilon pn)$. A straightforward application of the Chernoff bounds then implies that the probability that a particular vertex is violating is for large n at most

$e^{-\varepsilon pn}$. Hence, as the events that distinct vertices are violating with respect to S are independent, the probability that there are t such vertices is at most $e^{-\varepsilon pn \cdot t}$.

Note that the number of ways to choose S of size k is $\binom{n}{k} \leq (\frac{en}{k})^k$. In conclusion, the probability that there is an S with t violating vertices is at most

$$\left(\frac{e}{k}\right)^k \exp\left\{k \ln n - \varepsilon pn \cdot t\right\} \leq \left(\frac{e}{k}\right)^k \exp\left\{k\left(\ln n - \ln n\right)\right\} = \left(\frac{e}{k}\right)^k.$$

The proof is completed by summing this expression up for all $\varepsilon pn \leq k \leq \frac{n}{\alpha(n)}$. □

Finally, we need the following statement about the distribution of the edges in $G_{n,p}$. The lemma is a straightforward application of Chernoff bounds, and quite standard in the classical random graph theory. We include a short proof for completeness.

3.7 Lemma. *The following holds w.h.p.*

$$\forall S \subseteq V : e_{G_{n,p}}(S, V \setminus S) = |S|(n - |S|)p\left(1 \pm \sqrt{\tfrac{8}{\alpha(n)}}\right).$$

Proof. For any $S \subseteq V$, the quantity $e_{G_{n,p}}(S, V \setminus S)$ is binomially distributed with expectation $|S|(n - |S|)p$. Call S *bad*, if $e_{G_{n,p}}(S, V \setminus S)$ deviates from its expected value by more than $\sqrt{4|S|^2(n - |S|)p \ln n}$. As the statement of the lemma is symmetric in $S, V \setminus S$, it is sufficient to show it for S such that $|S| \leq n/2$. For any fixed such S we have

$$\frac{\sqrt{4|S|^2(n - |S|)p \ln n}}{|S|(n - |S|)p} = \sqrt{\frac{4 \ln n}{np(1 - |S|/n)}} \leq \sqrt{\frac{8 \ln n}{np}} \leq \sqrt{\frac{8}{\alpha(n)}}.$$

By applying the Chernoff bounds we obtain that the probability that S is bad is for large n at most

$$\exp\left\{-\frac{4|S|^2(n - |S|)p \cdot \ln n}{3|S|(n - |S|)p}\right\} = \exp\left\{-\frac{4}{3}|S| \ln n\right\}.$$

As the number of ways to choose S of size k is at most n^k, we infer by summing over all $1 \leq k \leq n/2$ that w.h.p. there is no bad set S in $G_{n,p}$. □

Note that in the special case that $|S| = 1$ in the above lemma, i.e., S contains just a single vertex v, we obtain that

$$|\Gamma_{G_{n,p}}(v)| = e_{G_{n,p}}(S, V \setminus S) = (1 \pm 3\alpha(n)^{-1/2})pn.$$

We will use this fact without further reference.

3.3.2 Proof of Theorem 3.4

Let G be any graph with vertex set V and let $p \geq \frac{\alpha(n)\ln n}{n}$, where $\alpha(n) \leq \ln^{1/9} n$ is any positive function such that $\lim_{n\to\infty} \alpha(n) = \infty$. Fix

$$\varepsilon := \alpha(n)^{-1/2}.$$

We say that G is *p-typical* if it satisfies the following three conditions:

(I) For any $S \subseteq V$ such that $|S| \geq \frac{n}{\alpha(n)}$ there is a $X_S \subset V \setminus S$ such that $|X_S| \leq \frac{4(\ln\alpha(n)+2)n}{\ln n}$ and

$$\forall v \in (V \setminus S) \setminus X_S : |\Gamma_G(v) \cap S| = (1 \pm \varepsilon)p|S|.$$

(II) For any $S \subseteq V$ such that $|S| \leq \frac{n}{\alpha(n)}$ there is a $X_S \subset V \setminus S$ such that $|X_S| \leq \frac{|S|}{\varepsilon\alpha(n)}$ and

$$\forall v \in (V \setminus S) \setminus X_S : |\Gamma_G(v) \cap S| \leq \varepsilon pn.$$

(III) For all $S \subseteq V$

$$e_G(S, V \setminus S) = |S|(n - |S|)p\left(1 \pm \sqrt{8}\varepsilon\right).$$

We will denote by $\mathcal{T}_n(p)$ the set of p-typical graphs on V. Note that Lemmas 3.5-3.7 guarantee that $G_{n,p}$ is w.h.p. p-typical. Hence, we shall restrict our attention only to graphs in $\mathcal{T}_n(p)$.

Let us denote by $T_1(G)$ the first point in time where at least εn vertices are informed and $T_2(G)$ the first point in time where at least $(1-\varepsilon)n$ vertices are informed. Our aim is to give bounds on $T(G)$ by bounding $T_1(G)$, $T_2(G) - T_1(G)$ and $T(G) - T_2(G)$ uniformly for every $G \in \mathcal{T}_n(p)$. The following three lemmas do so. In the proofs we will several times assume that n is sufficiently large so that the claimed inequalities hold, without explicitly mentioning that.

3.8 Lemma. *Uniformly for $G \in \mathcal{T}_n(p)$, with probability $1 - o(1)$ it holds that*

$$|T_1(G) - \log_{1+q} n| \leq 18\sqrt{\varepsilon}\log_{1+q} n.$$

Proof. Let $G \in \mathcal{T}_n(p)$ and let $T_1' = T_1'(G) \leq T_1(G) = T_1$ be the first time where the number of informed vertices exceeds $\ln^{1/4} n$. By Property (III) we have that if n is large enough, then for any $0 \leq t < T_1'$, each vertex in $\mathcal{I}(t)$ has at least $\frac{1}{2}\ln n$ neighbors outside $\mathcal{I}(t)$. For n large enough, the number of steps needed until a new vertex is informed is stochastically bounded from above by a geometric random variable with parameter $q/2$. Therefore, the probability that at least k steps are needed until a new vertex is added in the set of informed vertices is at most $(1-q/2)^k$. With $b := 1/(1-q/2)$ and $k := \lceil \log_b \ln n \rceil$, this probability is at most $1/\ln n$. We call the set of steps between two consecutive steps at which the set of informed vertices increases an *idle period*. Therefore, the probability that each one of the idle periods before T_1' lasts for less than k steps is at least $\left(1 - \frac{1}{\ln n}\right)^{\ln^{1/4} n} \geq 1 - \frac{1}{\ln^{3/4} n}$, as there are no more than $\ln^{1/4} n$ idle periods before T_1'. Consequently, $T_1' \leq k \ln^{1/4} n$ with probability at least $1 - \ln^{-3/4} n$. Observe that for n that is sufficiently large the bound on T_1' is at most $9\sqrt{\varepsilon}\log_{1+q} n$.

Next, we will show that with probability $1 - o(1)$, we have $T_1 - T_1' \in (1 \pm 9\sqrt{\varepsilon})\log_{1+q} n$, thus yielding the bound of the lemma. In particular, we will show that

$$\begin{aligned} &\Pr\big(I(t+1) \in (1+q \pm 7\sqrt{\varepsilon})I(t) \mid \ln^{1/4} n \leq I(t) < \varepsilon n\big) \\ &\quad \geq 1 - o\left((\ln n)^{-1}\right). \end{aligned} \tag{1}$$

The proof of the lemma is then completed by a repeated application of the above inequality. In particular, either there is a $t < (1+8\sqrt{\varepsilon})\log_{1+q} n + T_1'$ such that $I(t) \geq \varepsilon n$, in which case there is nothing to show, or, with probability $1 - o(1)$,

$$I\left(\lceil (1+8\sqrt{\varepsilon})\log_{1+q} n \rceil + T_1'\right) \geq (1+q-7\sqrt{\varepsilon})^{(1+8\sqrt{\varepsilon})\log_{1+q} n} \geq \varepsilon n.$$

For all $t < (1-8\sqrt{\varepsilon})\log_{1+q} n + T_1'$ we have

$$I(t+1) \in (1+q \pm 7\sqrt{\varepsilon})I(t)$$

with probability $(1 - o((\ln n)^{-1}))^{(1-8\sqrt{\varepsilon})\log_{1+q} n} = 1 - o(1)$, and therefore

$$I\left(\lfloor (1-8\sqrt{\varepsilon})\log_{1+q} n \rfloor + T_1'\right) \leq 2\ln^{1/4} n (1+q+7\sqrt{\varepsilon})^{(1-8\sqrt{\varepsilon})\log_{1+q} n} < \varepsilon n.$$

So we showed that

$$\begin{aligned} &(1-9\sqrt{\varepsilon})\log_{1+q} n \leq \lfloor (1-8\sqrt{\varepsilon})\log_{1+q} n \rfloor \leq T_1 - T_1' \\ &\leq \lceil (1+8\sqrt{\varepsilon})\log_{1+q} n \rceil \leq (1+9\sqrt{\varepsilon})\log_{1+q} n. \end{aligned}$$

In the remainder we prove equation (1). For every vertex $v \in I(t)$ we define an indicator random variable N_v that equals 1 if v informs a vertex in $\mathcal{U}(t)$. Moreover, for every pair of distinct vertices $v, v' \in \mathcal{I}(t)$ let $C_{v,v'}$ be the indicator variable that is equal to 1 if v and v' inform the same vertex in $\mathcal{U}(t)$. By the inclusion-exclusion principle we obtain that

$$\sum_{v \in \mathcal{I}(t)} N_v - \sum_{\substack{v,v' \in \mathcal{I}(t) \\ v \neq v'}} C_{v,v'} \;\leq\; N(t+1) \;\leq\; \sum_{v \in \mathcal{I}(t)} N_v.$$

Note that

$$\begin{aligned} \mathrm{E}\left(N_v\right) &= q \frac{|\Gamma(v) \cap \mathcal{U}(t)|}{|\Gamma(v)|} \quad \text{and} \\ \mathrm{E}\left(C_{v,v'}\right) &= q^2 \frac{|\Gamma(v) \cap \Gamma(v') \cap \mathcal{U}(t)|}{|\Gamma(v)||\Gamma(v')|}. \end{aligned} \tag{2}$$

We shall now show that $N(t+1) \in (1 \pm 7\sqrt{\varepsilon})I(t)q$ holds with the desired probability, which will complete the proof of equation (1). Before we proceed, let us make two auxiliary preparatory remarks. Note that by Property (III) of G we obtain for sufficiently large n that

$$\forall v \in V : |\Gamma(v)| = (1 \pm 3\varepsilon)pn.$$

This, together with equation (2) implies with a double counting argument that

$$\begin{aligned} \sum_{\substack{v,v' \in \mathcal{I}(t) \\ v \neq v'}} \mathrm{E}\left(C_{v,v'}\right) &= \sum_{\substack{v,v' \in \mathcal{I}(t) \\ v \neq v'}} q^2 \frac{|\Gamma(v) \cap \Gamma(v') \cap \mathcal{U}(t)|}{(1 \pm 7\varepsilon)(pn)^2} \\ &= \frac{(1 \pm 8\varepsilon)}{(pn)^2} q^2 \sum_{u \in \mathcal{U}(t)} \binom{|\Gamma(u) \cap \mathcal{I}(t)|}{2}. \end{aligned}$$

We will use these facts in the remainder without further reference.

Recall that $I(t) \geq \ln^{1/4} n$. We will first give tight bounds on the expectation of $N(t+1)$, and then apply the Hoeffding–Azuma inequality to show that $N(t+1)$ is sufficiently sharply concentrated around $\mathrm{E}\left(N(t+1)\right)$. By

using equation (2) we obtain for large n that

$$\begin{aligned} \mathrm{E}\left(\sum_{v\in\mathcal{I}(t)} N_v\right) &= q\sum_{v\in\mathcal{I}(t)} \frac{|\Gamma(v)\cap\mathcal{U}(t)|}{(1\pm 3\varepsilon)pn} \\ &= q\frac{(1\pm 4\varepsilon)e_G(\mathcal{I}(t), V\setminus\mathcal{I}(t))}{pn} \overset{(III)}{=} (1\pm 8\varepsilon)I(t)q. \end{aligned} \tag{3}$$

Moreover, recall that

$$\sum_{\substack{v,v'\in\mathcal{I}(t)\\ v\neq v'}} \mathrm{E}\left(C_{v,v'}\right) = \frac{q^2(1\pm 8\varepsilon)}{(pn)^2}\cdot\sum_{u\in\mathcal{U}(t)}\binom{|\Gamma(u)\cap\mathcal{I}(t)|}{2}. \tag{4}$$

We are going to estimate the last sum from above as follows. As $G\in\mathcal{T}_n(p)$ we may infer the following.

- If $I(t)\leq\frac{n}{\alpha(n)}$, then, by (II), there is a subset $\mathcal{X}\subset V\setminus\mathcal{I}(t)$ such that $|\mathcal{X}|\leq\sqrt{\varepsilon}I(t)$ and

$$\forall v\in(V\setminus\mathcal{I}(t))\setminus\mathcal{X} : |\Gamma(v)\cap\mathcal{I}(t)|\leq\varepsilon pn.$$

- If $\frac{n}{\alpha(n)}\leq I(t)\leq\varepsilon n=\frac{n}{\alpha(n)^{1/2}}$, then, by (I), there is $\mathcal{X}\subset V\setminus\mathcal{I}(t)$ such that $|\mathcal{X}|\leq\frac{4(\ln\alpha(n)+2)n}{\ln n}$ and

$$\forall v\in(V\setminus\mathcal{I}(t))\setminus\mathcal{X} : |\Gamma(v)\cap\mathcal{I}(t)|\leq(1+\varepsilon)pI(t)\leq 2\varepsilon pn.$$

So, in both cases we have for all $v\in\mathcal{U}(t)\setminus\mathcal{X}$ that $|\Gamma(v)\cap\mathcal{I}(t)|\leq 2\varepsilon pn$, and $|\mathcal{X}|\leq\sqrt{\varepsilon}I(t)$. Moreover, by exploiting property (III) of G we obtain that for all $v\in\mathcal{X}$ it holds $|\Gamma(v)\cap\mathcal{I}(t)|\leq 2pn$. Using this, we can bound from above the sum on the right-hand-side of equation (4) by splitting it into two parts as follows.

$$\begin{aligned} \sum_{u\in\mathcal{U}(t)}|\Gamma(u)\cap\mathcal{I}(t)|^2 &= \sum_{u\in\mathcal{U}(t)\setminus\mathcal{X}}|\Gamma(u)\cap\mathcal{I}(t)|^2+\sum_{u\in\mathcal{X}}|\Gamma(u)\cap\mathcal{I}(t)|^2 \\ &\leq \sum_{u\in\mathcal{U}(t)\setminus\mathcal{X}}|\Gamma(u)\cap\mathcal{I}(t)|^2+|\mathcal{X}|\ (2pn)^2 \\ &\leq \sum_{u\in\mathcal{U}(t)\setminus\mathcal{X}}|\Gamma(u)\cap\mathcal{I}(t)|^2+\sqrt{\varepsilon}I(t)\ (2pn)^2. \end{aligned}$$

Note that $0 \leq |\Gamma(u) \cap \mathcal{I}(t)| \leq 2\varepsilon pn$ for every $u \in \mathcal{U}(t) \setminus \mathcal{X}$. Moreover $\sum_{u \in \mathcal{U}(t)} |\Gamma(u) \cap \mathcal{I}(t)| = e_G(\mathcal{I}(t), V \setminus \mathcal{I}(t))$. By the convexity of x^2, an upper bound on the above sum can be obtained if we choose $|\Gamma(u) \cap \mathcal{I}(t)| = 2\varepsilon pn$ for $e_G(\mathcal{I}(t), V \setminus \mathcal{I}(t))/(2\varepsilon pn)$ different u's, and $|\Gamma(u) \cap \mathcal{I}(t)| = 0$ otherwise. We obtain

$$\sum_{u \in \mathcal{U}(t)} |\Gamma(u) \cap \mathcal{I}(t)|^2 \leq \frac{I(t)(n - I(t))p(2\varepsilon pn)^2}{\varepsilon pn} + \sqrt{\varepsilon} I(t)\ (2pn)^2 \leq \frac{9}{2}\sqrt{\varepsilon} p^2 n^2 I(t).$$

By plugging this into equation (4) we obtain that

$$\sum_{\substack{v,v' \in \mathcal{I}(t) \\ v \neq v'}} \mathrm{E}\left(C_{v,v'}\right) \leq 5q^2\sqrt{\varepsilon} I(t).$$

Finally, combined with equation (3) this gives

$$(1 - 6\sqrt{\varepsilon})I(t)q \leq \mathrm{E}\left(N(t+1)\right) \leq (1 + 6\sqrt{\varepsilon})I(t)q.$$

To complete the proof we will bound the probability that $N(t+1) \notin I(t)q(1 \pm 7\sqrt{\varepsilon})$ by using the Hoeffding–Azuma inequality. Note that $N(t+1)$ can change by at most 1 if we modify one of the choices made by some vertex in $\mathcal{I}(t)$. So by applying Theorem 1.5 with $c_i = 1$ and $N = I(t)$ we obtain

$$\Pr\left(N(t+1) \notin I(t)q(1 \pm 7\sqrt{\varepsilon})\right) \leq \Pr\left(N(t+1) \notin \mathrm{E}\left(N(t+1)\right) \pm \sqrt{\varepsilon} I(t)q\right) \leq 2e^{-\frac{q^2\varepsilon}{2}\ln^{1/4} n}.$$

□

In the next lemma we will consider the "intermediate" phase of the model between $T_1(G)$ and $T_2(G)$ for $G \in \mathcal{T}_n(p)$. Our general strategy is to bound the number $N(t)$ of vertices that get informed in the current round t from below. For this, we first estimate $\mathrm{E}(N(t))$ and then we use concentration inequalities (Theorem 1.5) to show that with sufficiently high probability $N(t)$ is very close to $\mathrm{E}(N(t))$.

3.9 Lemma. *Uniformly for all $G \in \mathcal{T}_n(p)$, with probability $1 - o(1)$ it holds that*

$$T_2(G) - T_1(G) \leq 9\varepsilon^{-1} \ln \varepsilon^{-1},$$

and there are at least $\frac{\varepsilon n}{2e}$ uninformed vertices at $T_2(G)$.

Proof. Let $G \in \mathcal{T}_n(p)$. We will show that for $T_1(G) \le t < T_2(G)$

$$I(t+1) \ge I(t)\left(1 + \frac{q\varepsilon}{4}\right), \tag{5}$$

with probability at least $1 - e^{-q\varepsilon^6 n/8}$. Let us abbreviate $b = 8\varepsilon^{-1}\ln\varepsilon^{-1}$. To see that this is sufficient for the first claim, note that if "$T_2(G) - T_1(G) \le b$", then there is nothing to prove. On the other hand, if "$T_2(G) - T_1(G) > b$", then with (conditional) probability at least $(1 - e^{-q\varepsilon^6 n/8})^{b+1} = 1 - o(1)$, for $\lceil b \rceil$ consecutive steps after $T_1(G)$ the recursion (5) holds. In turn, this implies with $1 + x > e^{x/2}$, which is valid for small enough $x > 0$, that

$$I(T_1(G) + \lceil b \rceil) \ge I(T_1(G)) \cdot \left(1 + \frac{\varepsilon}{4}\right)^{\lceil b \rceil} > \varepsilon n e^{b\varepsilon/8} > (1-\varepsilon)n.$$

Thus we obtain $T_2(G) - T_1(G) \le b + 1 \le 9\varepsilon^{-1}\ln\varepsilon^{-1}$.

Now we turn to the proof of (5). Let t be such that $T_1(G) \le t < T_2(G)$. We will show that $N(t+1)$ is not much smaller than its expected value. But first let us calculate $\mathrm{E}\,(N(t+1))$. The model implies that the probability that any $v \in V \setminus \mathcal{I}(t)$ *does not* belong to $\mathcal{N}(t+1)$ is precisely

$$\prod_{u \in \Gamma(v) \cap \mathcal{I}(t)} \left(1 - \frac{q}{|\Gamma(u)|}\right).$$

Next we make use of property (I) in the definition of $\mathcal{T}_n(p)$: All vertices in $V \setminus \mathcal{I}(t)$, apart from an exceptional set $\mathcal{X} = \mathcal{X}_t \subset V \setminus \mathcal{I}(t)$ that contains at most $4(\ln\alpha(n)+2)n/\ln n$ vertices, have $(1 \pm \varepsilon)pI(t)$ neighbors in $\mathcal{I}(t)$. Using this and the above fact we infer that $\mathrm{E}\,(N(t+1))$ is in

$$\sum_{v \in \mathcal{U}(t) \setminus \mathcal{X}} \left(1 - \prod_{u \in \Gamma(v) \cap \mathcal{I}(t)} \left(1 - \frac{q}{|\Gamma(u)|}\right)\right) \pm |\mathcal{X}|.$$

Next we derive tight bounds for the product above. Firstly, observe that property (III) implies that for all $u \in \mathcal{I}(t)$

$$|\Gamma(u)| = np\,(1 \pm 3\varepsilon)\,. \tag{6}$$

Also, the definition of $\mathcal{X}$ implies for $v \in (V \setminus \mathcal{I}(t)) \setminus \mathcal{X}$

$$|\Gamma(v) \cap \mathcal{I}(t)| = (1 \pm \varepsilon)\,pI(t). \tag{7}$$

Recall that for $x > 0$ small enough we have $e^{-x-x^2} \leq 1 - x \leq e^{-x}$. So the bounds in (6) and (7) imply that for $v \in \mathcal{U}(t) \setminus \mathcal{X}$ we have

$$\begin{aligned}&\prod_{u \in \Gamma(v) \cap \mathcal{I}(t)} \left(1 - \frac{q}{|\Gamma(u)|}\right) \\ &= \exp\left(-\frac{qI(t)}{n}\left(1 \pm 5\varepsilon\right)\right)\left(1 - O\left(\frac{1}{np}\right)\right).\end{aligned}$$

As $|\mathcal{U}(t) \setminus \mathcal{X}| = (n - I(t))(1 \pm \varepsilon)$, we obtain

$$\mathrm{E}\left(N(t+1)\right) = n\left(1 - \frac{I(t)}{n}\right)\left(1 - e^{-\frac{qI(t)}{n}}\right)\left(1 \pm O(\varepsilon)\right). \tag{8}$$

We will bound the probability that $|N(t+1) - \mathrm{E}\left(N(t+1)\right)| > \varepsilon\, \mathrm{E}\left(N(t+1)\right)$ using the Hoeffding–Azuma inequality. Firstly, note that as $\varepsilon < \frac{I(t)}{n} \leq 1 - \varepsilon$, we have $\mathrm{E}\left(N(t+1)\right) \geq q\varepsilon^2 n/2$, for n sufficiently large. Moreover, if we change only one of the random choices of the vertices in $\mathcal{I}(t)$, then $N(t+1)$ changes by at most 1. Thus, a simple application of Theorem 1.5 with $c_k = 1$ and $N = I(t)$ yields

$$\begin{aligned}&\Pr\left(|N(t+1) - \mathrm{E}\left(N(t+1)\right)| > \varepsilon\, \mathrm{E}\left(N(t+1)\right)\right) \\ &\quad \leq 2\exp\left(-\frac{\varepsilon^2\, \mathrm{E}^2\left(N(t+1)\right)}{2I(t)}\right) \leq 2\exp\left(-\frac{q^2\varepsilon^6 n}{8}\right).\end{aligned}$$

So, for n sufficiently large, we obtain that

$$N(t+1) = n\left(1 - \frac{I(t)}{n}\right)\left(1 - e^{-\frac{qI(t)}{n}}\right)\left(1 \pm \sqrt{\varepsilon}\right) \tag{9}$$

with probability at most $2\exp\left(-\frac{q^2\varepsilon^6 n}{8}\right)$. This identity enables us to write a recursive formula concerning the evolution of the number of informed vertices. Recall that for all $0 < x < 1$, we have $1 - e^{-x} \geq x/2$. (9) implies that

$$\begin{aligned}I(t+1) &\geq I(t) + n\left(1 - \frac{I(t)}{n}\right)\frac{qI(t)}{2n}\left(1 - \sqrt{\varepsilon}\right) \\ &= I(t)\left(1 + \frac{q}{2}\,\left(1 - \frac{I(t)}{n}\right)\left(1 - \sqrt{\varepsilon}\right)\right).\end{aligned} \tag{10}$$

Since $I(t) \leq (1-\varepsilon)n$, it follows that for n large enough

$$\frac{q}{2}\left(1-\frac{I(t)}{n}\right)\left(1-\sqrt{\varepsilon}\right) \geq \frac{\varepsilon q}{2}\left(1-\sqrt{\varepsilon}\right) \geq \frac{\varepsilon q}{4}.$$

By substituting this bound into (10) we obtain (5).

What remains is to show the second statement of the lemma. This follows readily from (9). Indeed, observe first that $U(t) = n\left(1-\frac{I(t)}{n}\right)$. So, for n large enough

$$\begin{aligned} U(T_2(G)) &= U(T_2(G)-1) - N(T_2(G)) \\ &\overset{(9)}{\geq} U(T_2(G)-1)e^{-qI((T_2(G)-1))/n}(1-e\sqrt{\varepsilon}) \geq \frac{\varepsilon n}{2e^q}. \end{aligned}$$

□

Finally, we proceed by bounding $T(G) - T_2(G)$, for $G \in \mathcal{T}_n(p)$. Recall that the main strategy in the previous argument was to show that the number $N(t+1)$ of vertices that become informed by $\mathcal{I}(t)$ in round $t+1$ is close to its expected value. To achieve this, we exploited the fact that in G, apart from a small exceptional set set, all vertices have the "right" degree in $\mathcal{I}(t)$. This argument is unfortunately not applicable in the proof of the next lemma: for $t > T_2(G)$, the set $V \setminus \mathcal{I}(t)$ of not yet informed vertices can become much smaller than $\mathcal{X}$, which makes our bounds useless. So we need to argue somehow differently. An additional difficulty is that we are not able to apply the Hoeffding–Azuma inequality in a meaningful way. Note that for $t > T_2(G)$ the quantity $I(t)$ is already of linear order, but the number $N(t+1)$ of newly informed vertices at step $t+1$ may become very small. In this case, the Hoeffding–Azuma inequality gives a trivial upper bound and thus the need for a stronger concentration inequality.

3.10 Lemma. *Uniformly for all $G \in \mathcal{T}_n(p)$, with probability $1-o(1)$*

$$\left|(T(G)-T_2(G)) - \tfrac{1}{q}\ln n\right| \leq \varepsilon^{1/3}\ln n.$$

Proof. We will split the interval between $T_2(G)$ and $T(G)$ into two subintervals. In particular, let $T'(G)$ be the first time after $T_2(G)$ where at most $\ln^{1/2} n$ uninformed vertices remain. We will give separate bounds for $T'(G) - T_2(G)$ and $T(G) - T'(G)$.

Let us start with the latter case, as it is the easier of the two. Let $U(t)$ denote the number of vertices that are still uninformed after the tth round. Let $t \geq T'(G)$. Since $np > \alpha(n) \ln n$, it follows from property (III) that for every $v \in \mathcal{U}(t)$ we have for n large enough $|\Gamma(v) \cap \mathcal{I}(t)| \geq np(1-3\varepsilon) - \ln^{1/2} n \geq np(1-4\varepsilon)$. So, the probability that a given uninformed vertex remains uninformed in the next round is for large n at most

$$\left(1 - \frac{q}{np(1+3\varepsilon)}\right)^{np(1-4\varepsilon)} \leq e^{-q\frac{1-4\varepsilon}{1+3\varepsilon}} \leq \frac{2}{e^q}.$$

Therefore, the probability that such a vertex remains uninformed for at least $\ln^{1/2} n$ steps after $T'(G)$ is at most $(2/e^q)^{\ln^{1/2} n}$. This implies that the expected number of vertices that remain uninformed for at least $\ln^{1/2} n$ steps after $T'(G)$ is at most $\ln^{1/2} n (2/e^q)^{\ln^{1/2} n} \leq (2/e^q)^{\ln^{1/3} n} = o(1)$. That is, with probability at least $1 - (2/e^q)^{\ln^{1/3} n}$, we have $T(G) - T'(G) < \ln^{1/2} n$.

The bound on $T'(G) - T_2(G)$ is significantly more complex. We will show that if t is such that $U(t) > \ln^{1/2} n$, then

$$U(t+1) = U(t)e^{-q}\left(1 \pm 50\sqrt{\varepsilon}\right), \tag{11}$$

with probability at least $1 - e^{-q\varepsilon \ln^{1/2} n/40} \geq 1 - e^{-\ln^{1/3} n}$. So if $T'(G) - T_2(G) > \frac{b_1}{q}$, where $b_1 := \ln n + 55\sqrt{\varepsilon} \ln n$, then with conditional probability at least $(1 - e^{-\ln^{1/3} n})^{\frac{b_1}{q}} = 1 - o(1)$ we have

$$U\left(T_2(G) + \left\lceil \tfrac{b_1}{q} \right\rceil\right) \leq U(T_2(G))e^{-b_1}\left(1 + 50\sqrt{\varepsilon}\right)^{\frac{1}{q}(b_1+1)}.$$

For large n

$$\left(1 + 50\sqrt{\varepsilon}\right)^{\frac{1}{q}(b_1+1)} \leq e^{\frac{55}{q}\sqrt{\varepsilon}\ln n}.$$

Also, $U(T_2(G)) \leq \varepsilon n$, which together with the above facts implies that

$$U\left(T_2(G) + \tfrac{1}{q}(b_1+1)\right) \leq \varepsilon n.$$

So, we may conclude that

$$T'(G) < T_2(G) + \tfrac{1}{q}(b_1+1).$$

Similarly, if we assume that $T'(G)-T_2(G) < \frac{b_2}{q}$, where $b_2 := \ln n - 55\sqrt{\varepsilon}\ln n$, then with conditional probability at least $(1-e^{-q\varepsilon\ln^{1/2} n/40})^{\frac{b_2}{q}+1} = 1-o(1)$ we have

$$U\left(T_2(G)+\left\lceil\tfrac{b_2}{q}\right\rceil\right) \geq U(T_2(G))e^{-b_2-1}\left(1-50\sqrt{\varepsilon}\right)^{\frac{b_2}{q}+1}.$$

As $U(T_2(G)) \geq \frac{\varepsilon n}{2e}$ holds by Lemma 3.9 with probability $1-o(1)$, a similar calculation as above shows that

$$U\left(T_2(G)+\left\lceil\tfrac{b_2}{q}\right\rceil\right) \geq \varepsilon e^{\frac{\sqrt{\varepsilon}}{q}\ln n} \gg \ln^{1/2} n.$$

Thus, $\left|T'(G)-T_2(G)-\frac{1}{q}\ln n\right| \leq \frac{1}{q}55\sqrt{\varepsilon}\ln n + 2$, which concludes the proof of the lemma.

It remains to show equation (11). As an auxiliary preparation we will show that "most" vertices in $\mathcal{U}(t)$ have the "right" degree in $\mathcal{I}(t)$, by arguing that if this was not the case, then there would be a significant deviation in the number of edges between $\mathcal{I}(t)$ and $\mathcal{U}(t)$. More precisely, let

$$\mathcal{X} = \left\{v \in \mathcal{U}(t) : |\Gamma(v)\cap\mathcal{I}(t)| < (1-3\sqrt{\varepsilon})pn\right\}.$$

In what follows, we argue that

$$|\mathcal{X}| \leq 3\sqrt{\varepsilon}(n-I(t)). \tag{12}$$

Indeed, as we assumed that $G \in \mathcal{T}_n(p)$, property (III) guarantees that $e_G(\mathcal{I}(t), V\setminus\mathcal{I}(t)) \geq I(t)(n-I(t))p(1-3\varepsilon)$. Moreover, it also implies that every vertex v has degree at most $(1+3\varepsilon)pn$. Therefore

$$e_G(\mathcal{I}(t), V\setminus\mathcal{I}(t)) < |\mathcal{X}|(1-3\sqrt{\varepsilon})pn + (1+3\varepsilon)(n-I(t)-|\mathcal{X}|)pn.$$

By putting the upper and the lower bounds together we obtain

$$I(t)(n-I(t))p(1-3\varepsilon) \leq -3|\mathcal{X}|pn(\sqrt{\varepsilon}+\varepsilon) + (1+3\varepsilon)(n-I(t))pn,$$

which implies with $I(t) \geq (1-\varepsilon)n$ that

$$1-4\varepsilon \leq -3\frac{|\mathcal{X}|}{n-I(t)}(\sqrt{\varepsilon}+\varepsilon) + (1+3\varepsilon).$$

An elementary calculation shows that the claimed equation (12) holds.

Now let $v \in \mathcal{U}(t) \setminus \mathcal{X}$. The probability that v becomes informed in the next round is

$$1 - \prod_{u \in \Gamma(v) \cap \mathcal{I}(t)} \left(1 - \frac{q}{|\Gamma(u)|}\right) = 1 - \left(1 - \frac{q}{pn(1 \pm 3\varepsilon)}\right)^{pn(1 \pm 3\sqrt{\varepsilon})} = 1 - \frac{1}{e^q}\left(1 \pm 7\sqrt{\varepsilon}\right).$$

By linearity of expectation, for n large enough we obtain

$$\begin{aligned} \mathrm{E}\left(N(t+1)\right) &= (n - I(t))\left(1 - \frac{1}{e^q}\right)(1 - 7\sqrt{\varepsilon}) \pm 3\sqrt{\varepsilon}(n - I(t)) \\ &= (n - I(t))\left(1 - \frac{1}{e^q}\right)(1 \pm 14\sqrt{\varepsilon}). \end{aligned} \tag{13}$$

Next we will show that $N(t+1)$ is with sufficiently high probability close to its expected value. Note that the Hoeffding–Azuma inequality does not give any meaningful bounds, as the number of the independent random variables is $I(t) \geq (1-\varepsilon)n$, while the expected value of $N(t+1)$ is proportional only to $n - I(t)$. The latter will eventually become so small that the exponent in the Hoeffding–Azuma inequality is $o(1)$, thus yielding a trivial bound. To bypass this problem, we will use Talagrand's inequality (Theorem 1.6). Note first that the bounded differences condition a. is satisfied, that is, changing one random choice can change $N(t+1)$ by at most 1. Regarding the second condition, b., note that if $N(t+1) = r$, then there must be at least r vertices in $\mathcal{I}(t)$ that have informed the vertices in $\mathcal{N}(t+1)$. Therefore, we may take $\psi(r) = \lceil r \rceil$ and with $m(N(t+1))$ denoting the median of $N(t+1)$, we deduce for any $x > 0$ that

$$\begin{aligned} \Pr\left(|N(t+1) - m(N(t+1))| > x\right) &\leq 4\exp\left(-\frac{x^2}{4(m(N(t+1)) + x)}\right) \\ &\leq 4\exp\left(-\frac{x^2}{4(2\,\mathrm{E}(N(t+1)) + x)}\right), \end{aligned} \tag{14}$$

where in the last inequality we have used the fact that $\mathrm{E}(N(t+1)) \geq m(N(t+1))\Pr(N(t+1) > m(N(t+1))) \geq m(N(t+1))/2$, which implies that $m(N(t+1)) \leq 2\,\mathrm{E}(N(t+1))$. However, we need to argue about the distance of $m(N(t+1))$ from $\mathrm{E}(N(t+1))$. We will use Proposition 1.7. The triangle

inequality yields:

$$\begin{aligned}
&|N(t+1) - \mathrm{E}(N(t+1))| \\
&\quad = \quad |N(t+1) - m(N(t+1)) + m(N(t+1)) - \mathrm{E}(N(t+1))| \\
&\quad \leq \quad |N(t+1) - m(N(t+1))| + |\,\mathrm{E}(N(t+1)) - m(N(t+1))| \\
&\quad \overset{\text{Proposition 1.7}}{=} \quad |N(t+1) - m(N(t+1))| + O\left(\sqrt{\mathrm{E}(N(t+1))}\right).
\end{aligned}$$

Since $\alpha(n) \leq \ln^{1/9} n$, we have $\sqrt{\mathrm{E}(N(t+1))} = o(\sqrt{\varepsilon}\,\mathrm{E}(N(t+1)))$. Therefore, for sufficiently large n

$$|N(t+1) - \mathrm{E}(N(t+1))| > x \Rightarrow |N(t+1) - m(N(t+1))| > x - \sqrt{\varepsilon}\,\mathrm{E}(N(t+1)).$$

Using this in (14) with $x = \sqrt{\varepsilon}\,\mathrm{E}\,(N(t+1))$ we obtain

$$\Pr(|N(t+1) - \mathrm{E}(N(t+1))| > 2\sqrt{\varepsilon}\,\mathrm{E}(N(t+1))) \leq 4\exp\left(-\frac{\varepsilon\,\mathrm{E}(N(t+1))}{4(2+\sqrt{\varepsilon})}\right).$$

Since $n - I(t) \geq \ln^{1/2} n$, by (13) we obtain that, say, $\mathrm{E}(N(t+1)) \geq \frac{q \ln^{1/2} n}{3}$. So, for large n, the above bound becomes

$$\Pr(|N(t+1) - \mathrm{E}(N(t+1))| > 2\sqrt{\varepsilon}\,\mathrm{E}(N(t+1))) \leq \exp\left(-\frac{q\varepsilon \ln^{1/2} n}{40}\right).$$

By putting everything together we obtain that with probability at least $1 - e^{-q\varepsilon \ln^{1/2} n/40}$

$$N(t+1) = (n - I(t))\left(1 - \frac{1}{e^q}\right)\left(1 \pm 16\sqrt{\varepsilon}\right).$$

So there remain $(U(t))e^{-q}\,(1 \pm 50\sqrt{\varepsilon})$ vertices uninformed in $\mathcal{U}(t)$ after one additional step of the protocol. This proves equation (11). □

The bounds obtained in Lemmas 3.8–3.10 imply Theorem 3.4, thus concluding our proof.

Indication of source. The content of this section has been previously published in [FHP10], parts of the special case $q = 1$ of it in [FHP09].

Chapter 4

Quasirandom Rumor Spreading on the Complete Graph

4.1 Introduction

The results presented in Chapter 3 show that randomized rumor spreading is a very powerful approach to dissemination problems. However, taking all decisions independently at random also has some unwanted effects. For example, a vertex may contact one of its neighbors twice before contacting all of its other neighbors. This may only be a minor problem for dense graphs such as the complete graph, but for sparse graphs it may increase the broadcast time significantly. For example, let G be a star on n vertices, i.e., a graph with one central vertex such that all other vertices have this vertex as their only neighbor. Clearly, any dissemination process where each vertex can send out at most one transmission per round needs at least $n-1$ rounds, simply because the central vertex has $n-1$ neighbors that cannot be informed by other vertices. However, due to the coupon collector effect, randomized rumor spreading on the star needs $\Theta(n \log n)$ rounds.

This type of imbalance in informing one's neighbors may be avoided by choosing the destination of the current transmission uniformly at random from those neighbors which have not yet been contacted by the originating vertex. However, this requires tracking all previously sent messages and is therefore less desirable. Motivated by the paradigm of quasirandomness, Doerr, Friedrich, and Sauerwald [DFS08] suggested the following *quasirandom rumor spreading* protocol. In this model, each vertex is equipped with a cyclic permutation (list) of its neighbors. As before, the protocol proceeds

in rounds, and all informed vertices participate in the dissemination process. However, each vertex only directs its first transmission to a random neighbor. Subsequently, it informs the successors of the first addressee on its list. We shall not make any assumptions about the structures of these cyclic lists. Now each vertex makes only one random choice, namely during the round at which it is informed.

The work of Doerr, Friedrich, and Sauerwald initiated a study of the quasirandom rumor spreading model, and showed that this model is efficient in the following sense. For the complete graph, the hypercube and the random graph they showed that all nodes are informed within $O(\log n)$ rounds, with probability $1 - o(1)$, independently of the particular choice of the lists.

We will sketch briefly the connection of this work to the general concept of quasirandomness. The main underlying idea is to imitate particular properties of a random process deterministically. This concept occurs in several areas of mathematics and computer science. Prominent examples are low-discrepancy point sets and Quasi-Monte Carlo Methods (see, e. g., [Nie92]). A particular example that inspired quasirandom rumor spreading is a quasirandom analogue of random walks, the rotor-router model introduced by Priezzhev, Dhar, Dhar, and Krishnamurthy [PDDK96]. This was later popularized by Propp (see, e. g., [CS06, CDFS08, CDST07, HLM$^+$08]) and became known as the *Propp machine*. To imitate the property of a random walk that many visits to a vertex result in a balanced number of moves going from it to each of its neighbors, each vertex is equipped with a rotor always pointing to a neighbor together with a cyclic permutation of the neighbors. A walk arises from leaving the current vertex in the rotor direction and then updating the rotor to the next neighbor according to the order given by the permutation. Some beautiful results exist on this model. Particularly, Cooper and Spencer [CS06] showed that if an arbitrary large population of particles does such a quasirandom walk on an infinite grid $\mathbb{Z}^d$, then under some mild conditions the number of particles on any vertex at each time deviates from the corresponding expected value by only a constant c_d. This constant is independent of the number of particles, their initial position, and the cyclic permutations used by the rotors. For example in the case $d = 1$, that is the graph being the infinite path, the best possible constant is $c_1 \approx 2.29$ [CDST07]. For the two-dimensional grid, the best possible constant c_2 satisfies $c_2 < 8.03$ [DF09].

Before analyzing the quasirandom protocol, let us discuss it from an implementation point of view. From a theory perspective, we immediately note

that the quasirandom model requires each vertex to store the permutation of its neighbors, which may utilize up to $\Theta(n \log n)$ bits. This is not necessary for the fully random model. However, we may assume that in most networks each vertex already has some list or array of its neighbors, since the information regarding how to contact a neighbor must be stored somewhere. In this scenario, the use of the lists does not increase the complexity. Rather, it appears that the quasirandom protocol needs less resources. In particular, it requires significantly fewer random bits. This is beneficial if we consider randomness costly, and useful if we want to trace an actual run of the protocol.

The core question to be answered is whether the quasirandom protocol works well even if we are not permitted to design the lists. Surprisingly, the answer is yes. For any selection of lists that can be present at each vertex, $O(\log n)$ rounds suffice with high probability to inform all the vertices of a complete graph K_n, a hypercube Q_n, an expander graph on n vertices (some extra conditions are needed here), or a random graph $G(n,p)$ with $p \geq (1+\varepsilon)(\ln n)/n$ [DFS08, DFS09]. Naturally, the lower bound of $\log_2 n$ rounds valid for the fully random model also holds for the quasirandom model. Once again these bounds fall within the right order of magnitude.

In some settings, we observe better broadcast times than in the classical model. One example is the random graph with edge probability $p = (\ln n + \omega(1))/n$ only minimally above the connectivity threshold. Nevertheless, with probability $1 - o(1)$, the random graph is such that with high probability the quasirandom protocol needs only $O(\log n)$ rounds independent of the starting point [DFS08]. This is a notable advantage over the fully random model. Feige, Peleg, Raghavan, and Upfal [FPRU90] demonstrate that for $p = (\ln n + O(\log \log n))/n$, the random graph with probability $1 - o(1)$ is such that $\Theta(\log^2 n)$ rounds are necessary to spread the rumor with high probability. The bounds obtained for arbitrary graphs are also superior for the quasirandom model. For the fully random model, we saw in Section 3.1 that $12n \ln n$ and $O(\Delta(G)(\mathrm{diam}(G) + \log n))$ rounds suffice to inform all vertices of an n-vertex graph G with high probability [FPRU90]. For the quasirandom model, it is easily proven that after $2n - 3$ or $\Delta(G)\,\mathrm{diam}(G)$ rounds, all vertices are informed with probability one.

The above results show that the broadcast time of the quasirandom rumor spreading protocol on the complete graph is quite well understood. Together with the experimental investigation [DFKS09], all results indicate that the quasirandom protocol achieves comparable or better broadcast times than the

random model. In Section 4.3 we investigate the equally important aspect of robustness. Since it would typically seem that robustness of randomized algorithms is caused by the large number of independent random decisions taken by the algorithm, one might expect that the quasirandom protocol is less robust. The experimental evaluation in [DFKS09] debunks this assertion. For both the hypercube and the complete graph on 2^{12} vertices, it was observed that if messages sent across the network using either protocol get lost with probability $\frac{1}{2}$, the broadcast time increases by a factor of between 1.8 and 1.9.

The only theoretical result pertaining to robustness is the one in [DFS09]. Let G be a graph, $T \in \mathcal{N}$ and $\gamma \geq 1$ such that the quasirandom protocol independent of the starting vertex with probability $1-n^{-\gamma}$ succeeds in informing all other nodes within T rounds. Then in the presence of transmission failures (independently chosen with probability $1-p$), independent of the starting vertex, $4\gamma(1/p)T$ rounds of a modified quasirandom protocol suffice to inform all vertices with probability $1-2n^{-\gamma}$. The modification of the protocol needed to prove this result is that the recipient of a message returns a feedback message to the sender (which also gets lost with probability $1-p$). Whenever the sender does not receive a feedback message, he tries to reach the same addressee in the next round. With this modification, however, the result is slightly weaker, in particular because the feedback modification makes the protocol significantly less simple.

In addition, the robustness result in [DFS09] leaves room for constant factor differences between the random and the quasirandom models in the presence of transmission faults.

Our results. The focus of this chapter is to investigate how long it takes until some rumor initially known only to a single vertex is broadcast to all other vertices. We adopt a worst-case view in that we aim at bounds that are independent of all the lists. Let $S(n)$ denote the number of rounds that are needed until all vertices of the complete graph with n vertices are informed. Note that always $\log_2 n \leq S(n) \leq n$. As mentioned above, Doerr, Friedrich, and Sauerwald [DFS08] proved that $S(n) \leq C \ln n$ with probability $1-o(1)$, for some constant $C > 0$. In [ADHP09], we proved sharper bounds analogous to those by Frieze and Grimmett [FG85], that is, we showed that with probability $1-o(1)$ we have $S(n) = (1 \pm o(1))(\log_2 n + \ln n)$. This result will be strengthened in Section 4.2. In particular, we show that for

any function $\omega : \mathbb{N} \to \mathbb{R}$ such that $\omega(n) \to \infty$ as $n \to \infty$ we have

$$\log_2 n + \ln n - 4 \ln \ln n \leq S(n) \leq \log_2 n + \ln n + \omega(n)$$

with probability $1 - o(1)$. Together with the result by Pittel [Pit87] described in Section 3.1, this result demonstrates that irrespectively of the choice of lists quasirandom rumor spreading is as fast as randomized rumor spreading. At the same time it reduces the number of random bits from $O(\log^2 n)$ to only $\lceil \log_2 n \rceil$ per vertex. This work has been published in [FH09a, FH09b].

In Section 4.3, we offer a detailed investigation of the robustness of quasirandom rumor spreading on the complete graph. We use the following model of lossy communication, which was analyzed in [HKP+05]. We assume that each message reaches its target with a certain probability $p \in]0, 1]$ independently for all transmissions. Note that we do not assume that the sender is notified of a transmission failure. We show that for all $\varepsilon > 0$ and $p \in]0, 1]$ the quasirandom rumor spreading protocol with arbitrary lists, despite independent message losses occurring with probability $1 - p$, succeeds with probability at least $1 - n^{-p\varepsilon/40}$ in informing all other vertices from a given vertex in time at most $(1 + \varepsilon)(\log_{1+p} n + \frac{1}{p} \ln n)$. A short version of this work appeared in [DHL09], see also [DHL]. For $p = 1$, this result coincides with the bound of $(1 + o(1))(\log_2 n + \ln n)$ proved in [ADHP09].

This result is interesting for two reasons. Firstly, it shows that the quasirandom protocol is even more robust than previous results indicate. Note that the above bound is strictly better than $\frac{1}{p}(1 + o(1))(\log_2 n + \ln n)$, that is, $\frac{1}{p}$ times the bound for the case without faulty transmissions. For example, for $p = 1/2$, the runtime increases by a factor of about 1.828 only. Secondly, our results imply that the quasirandom protocol is at least as robust as the classical one, as we showed that the corresponding slow-down for the fully random protocol is at least this factor in Section 3.2.

Precise description of the model and notations. Let $G = (V, E)$ be a finite and connected graph and let $n := |V|$ be the number of vertices. We associate with each vertex $v \in V$ a cyclic ordering (list) of the neighbors of v. As randomized rumor spreading, quasirandom rumor spreading also proceeds in *rounds* or *steps*. We assume that each vertex $v \in V$ has a pointer which always points to the vertex which is to be informed at the next step if v is informed.

We assume that in the beginning only one vertex $s \in V$ is aware of the information. Initially, it selects a position in its list uniformly at random and puts its pointer there. During the first round, s informs the vertex which is at this position and, subsequently, moves its pointer to the next position. That is, if the pointer was in position i, it moves to position $i+1 \mod d(s)$, where $d(s)$ denotes the degree of s. Now the newly informed vertex chooses at random a position in its own cyclic ordering. More generally, assume that after $t \geq 1$ rounds there are $I(t)$ informed vertices, and let $\mathcal{I}(t)$ denote their set (also $\mathcal{I}(0) = \{s\}$). At the $(t+1)$st step, each $v \in \mathcal{I}(t)$ informs the vertex that is indicated by its pointer (if the latter is uniformed) and, then, it moves the pointer to the next position. Let $\mathcal{N}(t+1)$ denote the set of newly informed vertices, and let $N(t+1) = |\mathcal{N}(t+1)|$. So we have $\mathcal{I}(t+1) = \mathcal{I}(t) \cup \mathcal{N}(t+1)$. At the end of the $(t+1)$st round, each $v \in \mathcal{N}(t+1)$ selects uniformly at random a position in its list and places its pointer there.

4.2 Quasirandom Rumor Spreading as Fast as Randomized Rumor Spreading

In this section, we present a tight analysis of the rumor spreading under the quasirandom model on the complete graph and show that its evolution is very close to the evolution of the randomized model. We will prove an (almost) analogue of the bound that Pittel [Pit87] gave for the randomized model (see Section 3.1). Let $G = (V, E)$ be the underlying complete graph and let $n := |V|$ be the number of vertices. For the ease of the notation let us assume $V = [n]$. For the simplicity of our calculations, we assume that any given vertex can also contact itself. We associate with each vertex $v \in V$ a cyclic ordering of V, which we denote by $\ell(v)$. This determines the order in which a vertex informs the other vertices according to the process described above for general graphs. We let $\mathcal{L} = \{\ell(v)\}_{v \in V}$ and refer to the pair $\mathcal{Q}_n := (V, \mathcal{L})$ as a *quasirandom rumor spreading configuration*. We assume without loss of generality that the initially informed vertex is 1. Let

$$S(n) := \min\{t \geq 0 \ : \ I(t) = n\},$$

which is the number of steps needed until all vertices have been informed in a quasirandom rumor spreading configuration $\mathcal{Q}_n$. The main theorem of this section is as follows.

4.1 Theorem. *Given a quasirandom rumor spreading configuration $\mathcal{Q}_n$ and any function $\omega : \mathbb{Z}^+ \to \mathbb{R}$ with $\lim_{n\to\infty} \omega(n) = \infty$, with probability $1 - o(1)$,*

$$\log_2 n + \ln n - 4 \ln \ln n \leq S(n) \leq \log_2 n + \ln n + \omega(n).$$

Sketch of the proof. The proof of the theorem is based on splitting the set of rounds into consecutive phases. We show that during the first three phases the number of informed vertices at each step nearly doubles. These phases last until the number of informed vertices becomes very close to n (but still $o(n)$) and yield the $\log_2 n$ term in the above theorem. Thereafter, there is an intermediate phase where the vast majority of the vertices are informed, leaving no more than $ne^{-\omega(n)/2}$ uninformed vertices with probability $1-o(1)$. During the subsequent phase, which lasts for approximately $\frac{1}{2} \ln n$ rounds, we show that in each round the number of uninformed vertices decreases approximately by a factor of e^{-1}. We deduce the upper bound in Theorem 4.1 by looking at the contents of the lists of the informed vertices within length $\frac{1}{2} \ln n$ after the current position of their pointer and proving that they cover all the uninformed vertices with high probability. In other words, if we let the system run for another $\frac{1}{2} \ln n$ steps, then all vertices will have been informed with probability $1-o(1)$. As far as the lower bound is concerned, we condition on the number of uninformed vertices just after the third phase (recall that this number is still almost equal to n) and then we couple the process with a process in which all the uninformed vertices make their random choices simultaneously, and we look at the segments of length (approximately) $\ln n - 4 \ln \ln n$ after the positions of the pointers. The number of vertices which do not belong to any of these segments is stochastically smaller than the number of uniformed vertices after $\log_2 n + \ln n - 4 \ln \ln n$ steps in the original process. We use Chebyshev's inequality to show that the former is positive with probability $1 - o(1)$ and the lower bound in Theorem 4.1 follows. In summary, we show that the quasirandom broadcasting method has essentially the same evolution as the randomized method, as presented in detail by Pittel in [Pit87].

Organization of the proof. In Section 4.2.1, we analyze the first three phases. Here we show that during each round the number of informed vertices almost doubles. In Section 4.2.2, we present the basic tools with which we analyze the fourth and fifth phases. In Section 4.2.3, we present the proof

of the upper bound of Theorem 4.1, and we conclude with the proof of the lower bound in Section 4.2.4.

4.2.1 Nearly Doubling of $I(t)$ During the Early Stages

Phase 1: $t < \frac{1}{2}\log_2 \frac{n}{\ln^3 n}$

With Phase 1 we denote the rumor spreading process up to round $t_1 := \left\lfloor \frac{1}{2}\log_2 \frac{n}{\ln^3 n} \right\rfloor$. We show that during each of the steps $t < t_1$ the number of informed vertices actually doubles with probability $1 - o(1)$. Therefore, at the end of this phase the number of informed vertices is 2^{t_1}.

First, note that $t_1 < \frac{1}{2}\log_2 n < \ln n$. With $L := \lfloor 3 \ln n \rfloor$, let $\ell_L(v)$ denote the segment of length L in $\ell(v)$ that starts at the position which v selects randomly. That is, if v selects position i, then $\ell_L(v)$ consists of the vertices in $\ell(v)$ which are located at positions $\{i, i+1 \mod n, \ldots, i+L-1 \mod n\}$.

Observe that the number of informed vertices during each of the first t_1 steps doubles if

1. for all distinct $v, v' \in \mathcal{I}(t_1)$ we have $\ell_L(v) \cap \ell_L(v') = \emptyset$, and
2. for all $v \in \mathcal{I}(t_1)$ we have $1 \notin \ell_L(v)$.

Note that this implies that every $v \in \mathcal{I}(t_1) \setminus \{1\}$ is contacted by exactly one vertex from $\mathcal{I}(t_1)$, and therefore there exists exactly one $v' \in \mathcal{I}(t_1)$ such that $v \in \ell_L(v')$.

We will show inductively that this event occurs with probability $1 - o(1)$. For this purpose, we let $\mathcal{E}_t$ be the event that is defined by these two conditions taken up to step t instead of t_1.

Note that, as far as the first round is concerned, it suffices to show that the second part of the event occurs with high probability, as there is only one vertex initially, namely vertex 1, which is aware of the rumor. So

$$\Pr(\mathcal{E}_0) = \Pr(1 \notin \ell_L(1)) \geq 1 - \frac{3 \ln n}{n}. \tag{1}$$

Let us assume now that for $1 \leq t \leq t_1$ the event $\mathcal{E}_{t-1}$ is realized. Then $N(t) = I(t-1) = 2^{t-1}$. Let us fix an ordering on $\mathcal{N}(t)$ according to which we expose the random choices of the vertices in $\mathcal{N}(t)$. If $v_i \in \mathcal{N}(t)$ is the ith vertex according to this ordering, then

$$\Pr\left(1 \in \ell_L(v_i)\right) \leq \frac{3 \ln n}{n}. \tag{2}$$

Also

$$\Pr\left(\exists v' \in \mathcal{I}(t-1) \cup \{v_1, \ldots, v_{i-1}\} : \ \ell_L(v') \cap \ell_L(v_i) \neq \emptyset\right) \leq \frac{2 \cdot 2^{t-1} 9 \ln^2 n}{n}. \tag{3}$$

Since the random choices of the vertices in $\mathcal{N}(t)$ are independent, we have for n large enough,

$$\begin{aligned} \Pr(\mathcal{E}_t \mid \mathcal{E}_{t-1}) &\overset{(2),(3)}{\geq} \left(1 - \frac{3\ln n}{n} - \frac{2^t 9 \ln^2 n}{n}\right)^{N(t)} \geq \left(1 - \frac{2^t 10 \ln^2 n}{n}\right)^{N(t)} \\ &\overset{N(t)=2^{t-1}}{=} \left(1 - \frac{2^t 10 \ln^2 n}{n}\right)^{2^{t-1}} \geq 1 - \frac{4^t 10 \ln^2 n}{n}. \end{aligned} \tag{4}$$

Thus for n large enough,

$$\begin{aligned} \Pr(\mathcal{E}_{t_1}) = \Pr(\mathcal{E}_0) \prod_{t=1}^{t_1} \Pr(\mathcal{E}_t \mid \mathcal{E}_{t-1}) &\overset{(1),(4)}{\geq} \left(1 - \frac{3\ln n}{n}\right) \prod_{t=1}^{t_1} \left(1 - \frac{4^t 10 \ln^2 n}{n}\right) \\ &\overset{\frac{4^{t_1} \ln^2 n}{n} = o(1)}{\geq} \left(1 - \frac{3\ln n}{n}\right) \exp\left(-\sum_{t=1}^{t_1} \frac{4^t 20 \ln^2 n}{n}\right) \\ &\geq \left(1 - \frac{3\ln n}{n}\right) \exp\left(-\frac{20 \ln^2 n}{n} \sum_{t=0}^{t_1} 4^t\right) \\ &\geq \left(1 - \frac{3\ln n}{n}\right) \exp\left(-\frac{4 \cdot 4^{t_1} 20}{3} \frac{\ln^2 n}{n}\right) \\ &\overset{4^{t_1} \leq \frac{n}{\ln^3 n}}{\geq} \left(1 - \frac{3\ln n}{n}\right) \exp\left(-\frac{27}{\ln n}\right) \geq 1 - \frac{28}{\ln n}. \end{aligned} \tag{5}$$

Note that on $\mathcal{E}_{t_1}$, we have $\frac{1}{2}\sqrt{\frac{n}{\ln^3 n}} \leq I(t_1) \leq \sqrt{\frac{n}{\ln^3 n}}$.

Phase 2: $\frac{1}{2}\log_2 \frac{n}{\ln^3 n} \leq t \leq \log_2\left(\frac{n}{\ln^6 n}\right)$

We set $t_2 := \left\lceil \log_2\left(\frac{n}{\ln^6 n}\right)\right\rceil$. We will approximate $\mathcal{I}(t)$ from below by a suitable subset of informed vertices. First, let us set $\mathcal{I}'(t_1) := \mathcal{I}(t_1)$ and $I'(t_1) := |\mathcal{I}'(t_1)|$. Note that conditional on $\mathcal{E}_{t_1}$, $N(t_1+1) = I'(t_1)$. Assume that we have defined $\mathcal{I}'(t-1) \subseteq \mathcal{I}(t-1)$, which is such that for every distinct $v', v \in \mathcal{I}'(t-1)$ we have $\ell_L(v) \cap \ell_L(v') = \emptyset$ and $1 \notin \ell_L(v)$. Let $\mathcal{N}'(t)$ be the

set of vertices which are contacted by the vertices in $\mathcal{I}'(t-1)$ during the tth round, and let $N'(t) := |\mathcal{N}'(t)|$. Note that $N'(t) = I'(t-1) := |\mathcal{I}'(t-1)|$. Assume that the vertices of $\mathcal{N}'(t)$ make their random choices according to a particular ordering which we fix. Let $\mathcal{B}_t$ be the set of vertices in $\mathcal{N}'(t)$ defined as follows. If v_i is the ith vertex in the ordering, then $v_i \in \mathcal{B}_t$ if either there exists $v' \in \mathcal{I}'(t-1) \cup \{v_1, \dots, v_{i-1}\}$ such that the segment $\ell_L(v_i)$ overlaps with $\ell_L(v')$ or $\ell_L(v_i)$ contains 1. We set $\mathcal{I}'(t) := \mathcal{I}'(t-1) \cup (\mathcal{N}'(t) \setminus \mathcal{B}_t)$.

Now, for $t_1 < t \le t_2$, we let $\mathcal{E}_t$ be the event that for all $t_1 \le s \le t$ we have $I'(s) \ge 2I'(s-1) - \frac{I'(s-1)}{\ln^2 n}$, and also $\mathcal{E}_{t_1}$ is realized. We will show the following lemma.

4.2 Lemma. *For all $t_1 < t \le t_2$, $\Pr(\mathcal{E}_t \mid \mathcal{E}_{t-1}) \ge 1 - \frac{18}{\ln^2 n}$.*

Proof. We will show that $I'(t) \ge 2I'(t-1) - \frac{I'(t-1)}{\ln^2 n}$ with conditional probability at least $1 - \frac{18}{\ln^2 n}$. In other words, it suffices to show that, conditional on $\mathcal{E}_{t-1}$, we have $B_t := |\mathcal{B}_t| \le \frac{I'(t-1)}{\ln^2 n}$ with this probability. If $v_i \in \mathcal{N}'(t)$ denotes the
ith vertex in the ordering of $\mathcal{N}'(t)$, then the probability that $\ell_L(v_i)$ contains 1 is at most $\frac{3\ln n}{n}$. Also, the probability that $\ell_L(v_i)$ is not disjoint from $\ell_L(v')$ for some $v' \in \mathcal{I}'(t-1) \cup \{v_1, \dots, v_{i-1}\}$ is at most $\frac{(I'(t-1)+i-1)9\ln^2 n}{n}$. Thus the union of these events occurs with probability at most $\frac{(I'(t-1)+i)9\ln^2 n}{n}$ for $n \ge 3$. So if we condition on specific realizations of $\mathcal{I}'(t-1)$ and $\mathcal{N}'(t)$, then

$$\begin{aligned}
&\mathrm{E}(B_t \mid \mathcal{I}'(t-1), \mathcal{N}'(t), \mathcal{E}_{t-1}) \le \frac{9\ln^2 n}{n} \sum_{i=1}^{N'(t)} (I'(t-1) + i) \\
&\le \frac{9\ln^2 n}{n} \left(I'(t-1)N'(t) + N'(t)^2 \right) = \frac{9\ln^2 n}{n} 2I'(t-1)^2 = \frac{18\ln^2 n}{n} I'(t-1)^2,
\end{aligned} \tag{6}$$

where in the penultimate equality we used the fact that $N'(t) = I'(t-1)$.

Therefore by Markov's inequality and using the trivial upper bound $I'(t-1) \le 2^{t-1}$ we get

$$\begin{aligned}
&\Pr\left(B_t \ge \frac{I'(t-1)}{\ln^2 n} \;\middle|\; \mathcal{I}'(t-1), \mathcal{N}'(t), \mathcal{E}_{t-1} \right) \le \frac{18\ln^2 n}{n} \frac{I'(t-1)^2}{I'(t-1)/\ln^2 n} \\
&\quad \le \frac{18\ln^4 n}{n} I'(t-1) \le \frac{18\ln^4 n}{n} \frac{n}{\ln^6 n} = \frac{18}{\ln^2 n}.
\end{aligned} \tag{7}$$

Averaging over all realizations of $\mathcal{I}'(t-1)$ and $\mathcal{N}'(t)$ such that $\mathcal{E}_{t-1}$ is realized, we deduce the lemma. □

Therefore

$$\begin{aligned}
\Pr(\mathcal{E}_{t_2}) &= \Pr(\mathcal{E}_{t_1}) \prod_{t=t_1+1}^{t_2} \Pr(\mathcal{E}_t \mid \mathcal{E}_{t-1}) \\
&\overset{\text{Lemma 4.2,(5)}}{\geq} \left(1-\frac{28}{\ln n}\right)\left(1-\frac{18}{\ln^2 n}\right)^{t_2-t_1} \\
&\overset{t_2\leq 2\ln n}{\geq} \left(1-\frac{28}{\ln n}\right)\left(1-\frac{18}{\ln^2 n}\right)^{2\ln n} \\
&\geq \left(1-\frac{28}{\ln n}\right)\left(1-\frac{36}{\ln n}\right) \;\geq\; 1-\frac{64}{\ln n}. \qquad (8)
\end{aligned}$$

Also, on $\mathcal{E}_{t_2}$ we have

$$\begin{aligned}
I'(t_2) &\geq I(t_1)2^{t_2-t_1}\left(1-\frac{1}{2\ln^2 n}\right)^{t_2-t_1} \overset{I(t_1)=2^{t_1}}{=} 2^{t_2}\left(1-\frac{1}{2\ln^2 n}\right)^{t_2-t_1} \\
&\geq 2^{t_2}\left(1-\frac{1}{2\ln^2 n}\right)^{2\ln n} \geq 2^{t_2}\left(1-\frac{1}{\ln n}\right) \geq \frac{n}{\ln^6 n}\left(1-\frac{1}{\ln n}\right).
\end{aligned}$$

Since $I(t) \geq I'(t)$ we deduce that on $\mathcal{E}_{t_2}$,

$$I(t_2) \geq \frac{n}{\ln^6 n}\left(1-\frac{1}{\ln n}\right). \qquad (9)$$

Phase 3: $t_2 < t \leq t_2 + \left\lceil \log_2\left(\frac{\ln^6 n}{\omega(n)}\right)\right\rceil + 1$

We set

$$t_3 := t_2 + \left\lceil \log_2\left(\frac{\ln^6 n}{\omega(n)}\right)\right\rceil + 1.$$

Here ω is a real-valued function on the set of positive integers that tends to infinity, slowly enough for our calculations to work. For convenience we will drop the argument and also write ω for the function value $\omega(n)$ in the following.

The analysis in this phase refines the idea that was used in the analysis in the previous phase. Quite informally, we will work with a subset $\tilde{\mathcal{I}}(t)$ of the set of informed vertices $\mathcal{I}(t)$, which has the property that any two vertices v and v' belonging to this set are such that

1. if v and v' are distinct, the segments of length $t_3 - t$ in $\ell(v)$ and $\ell(v')$ which start at the positions of the pointers of v and v' after step t are disjoint; and

2. this segment in $\ell(v)$ does not contain v'.

We denote the segment of length $t_3 - t$ in $\ell(v)$ starting at the position of the pointer of v after step t by $\ell(v; t+1, t_3)$. We will give an inductive definition of the set $\tilde{\mathcal{I}}(t)$. Note first that the vertices which belong to the set $\mathcal{I}'(t_2)$ satisfy the above conditions. Thus, we may set $\tilde{\mathcal{I}}(t_2) := \mathcal{I}'(t_2)$. Assume now that we have defined the set $\tilde{\mathcal{I}}(t-1)$ for $t > t_2$. Note that the above properties imply that if $\tilde{\mathcal{N}}(t)$ denotes the set of vertices contacted at step t by the vertices of $\tilde{\mathcal{I}}(t-1)$, then $\tilde{N}(t) := |\tilde{\mathcal{N}}(t)| = |\tilde{\mathcal{I}}(t-1)| =: \tilde{I}(t-1)$. The set $\tilde{\mathcal{I}}(t)$ is defined as the union of the sets $\tilde{\mathcal{I}}_1(t) \subseteq \tilde{\mathcal{I}}(t-1)$ and $\tilde{\mathcal{I}}_2(t) \subseteq \tilde{\mathcal{N}}(t)$ which in turn are defined as follows:

1. The set $\tilde{\mathcal{I}}_1(t)$ consists of those vertices $v \in \tilde{\mathcal{I}}(t-1)$ such that for all $v' \in \tilde{\mathcal{N}}(t)$ we have $\ell(v; t+1, t_3) \cap \ell(v'; t+1, t_3) = \emptyset$ and also $v \notin \ell(v'; t+1, t_3)$.

2. The set $\tilde{\mathcal{I}}_2(t)$ consists of those vertices $v \in \tilde{\mathcal{N}}(t)$ such that for all $v' \in \tilde{\mathcal{N}}(t)$ we have $\ell(v; t+1, t_3) \cap \ell(v'; t+1, t_3) = \emptyset$, if $v \neq v'$, and also $v \notin \ell(v'; t+1, t_3)$.

The main lemma for Phase 3 concerns the rate of growth of $\tilde{\mathcal{I}}(t)$ within this phase.

4.3 Lemma. *Conditional on $\mathcal{E}_{t_2}$, with probability $1 - o(1)$, for all t with $t_2 \leq t < t_3 - 1$ we have*

$$\tilde{I}(t+1) \geq 2\tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t)^2}{n}\right).$$

Proof. We will show the lemma by induction on t. For $t_2 < t \leq t_3$ we define $\mathcal{E}_t$ to be the event that for all s such that $t_2 \leq s \leq t$,

$$\tilde{I}(s) \geq 2\tilde{I}(s-1)\left(1 - \frac{\tilde{I}(s-1)(t_3 - s + 1)^2}{n}\right),$$

and that the event $\mathcal{E}_{t_2}$ is also realized. Note that the event of the lemma is $\mathcal{E}_{t_3-1}$. We will show that

$$\Pr(\mathcal{E}_{t+1} \mid \mathcal{E}_t) \geq 1 - 2e^{-n^{1/4}}. \tag{10}$$

Since the events $\{\mathcal{E}_t\}_{t_2 \le t \le t_3 - 1}$ form a decreasing family and since $t_3 - t_2 \le \ln n$, we then have

$$\begin{aligned} \Pr(\mathcal{E}_{t_3-1} \mid \mathcal{E}_{t_2}) = \prod_{t=t_2}^{t_3-2} \Pr(\mathcal{E}_{t+1} \mid \mathcal{E}_t) &\ge \left(1 - 2e^{-n^{1/4}}\right)^{t_3 - t_2} \\ &\ge \left(1 - 2e^{-n^{1/4}}\right)^{\ln n} \ge 1 - e^{-n^{1/5}} \end{aligned} \tag{11}$$

if n is large enough. Now let us fix some t which satisfies $t_2 \le t < t_3 - 1$, and let us condition on $\mathcal{E}_t$. To estimate $\Pr(\mathcal{E}_{t+1} \mid \mathcal{E}_t)$, we need only estimate the (conditional) probability that

$$\tilde{I}(t+1) \ge 2\tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t)^2}{n}\right).$$

Let $\tilde{I}_i(t+1) := |\tilde{\mathcal{I}}_i(t+1)|$ for $i = 1, 2$. As $\tilde{\mathcal{I}}(t+1)$ is the disjoint union of the sets $\tilde{\mathcal{I}}_1(t+1)$ and $\tilde{\mathcal{I}}_2(t+1)$, it suffices to show that each of the following events occurs with sufficiently high (conditional) probability:

$$\tilde{I}_1(t+1) \ge \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t)^2}{n}\right) \tag{12}$$

and

$$\tilde{I}_2(t+1) \ge \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t)^2}{n}\right). \tag{13}$$

Proof of (12). Note that the size of $\tilde{\mathcal{I}}_1(t+1)$ is a function of the independent random choices of the vertices in $\tilde{\mathcal{N}}(t+1)$. We will bound the probability of (12) from below using the Hoeffding–Azuma inequality (Theorem 1.5). Here all probabilities and expected values are conditional on $\mathcal{E}_t$. But first we shall give a lower bound on the (conditional) expected value of $\tilde{I}_1(t+1)$. Recall that a vertex $v \in \tilde{\mathcal{I}}(t)$ belongs to $\tilde{\mathcal{I}}_1(t+1)$ if for all $v' \in \tilde{\mathcal{N}}(t+1)$ we have $\ell(v; t+2, t_3) \cap \ell(v'; t+2, t_3) = \emptyset$ and also $v \notin \ell(v'; t+2, t_3)$. The former fails with probability at most $\frac{\tilde{N}(t+1)(t_3-(t+2)+1)^2}{n} = \frac{\tilde{I}(t)(t_3-t-1)^2}{n}$. The latter fails with probability at most $\frac{\tilde{N}(t+1)(t_3-t-1)}{n} = \frac{\tilde{I}(t)(t_3-t-1)}{n}$. Thus, for a vertex $v \in \tilde{\mathcal{I}}(t)$,

$$\begin{aligned} \Pr\left(v \in \tilde{\mathcal{I}}_1(t+1)\right) &\ge 1 - \frac{\tilde{I}(t)(t_3 - t - 1)}{n} - \frac{\tilde{I}(t)(t_3 - t - 1)^2}{n} \\ &\ge 1 - \frac{\tilde{I}(t)(t_3 - t - \frac{1}{2})^2}{n}. \end{aligned}$$

In turn, the expected value of $\tilde{I}_1(t+1)$ conditional on $\mathcal{E}_t$ is

$$\mathrm{E}\left(\tilde{I}_1(t+1)\right) \geq \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t - \frac{1}{2})^2}{n}\right). \tag{14}$$

In order to prove (12), it suffices to bound the probability of the event $\tilde{I}_1(t+1) < \mathrm{E}\big(\tilde{I}_1(t+1)\big) - \tilde{I}^{2/3}(t)$. Indeed, by (14),

$$\mathrm{E}\left(\tilde{I}_1(t+1)\right) - \tilde{I}^{2/3}(t) \geq \tilde{I}(t) - \frac{\tilde{I}^2(t)(t_3 - t - \frac{1}{2})^2}{n} - \tilde{I}^{2/3}(t).$$

We have the following.

4.4 Proposition.

$$\frac{\tilde{I}^2(t)(t_3 - t - \frac{1}{2})^2}{n} + \tilde{I}^{2/3}(t) \leq \frac{\tilde{I}^2(t)(t_3 - t)^2}{n}.$$

Proof. The inequality is equivalent to

$$\left(t_3 - t - \frac{1}{2}\right)^2 + \frac{n\tilde{I}^{2/3}(t)}{\tilde{I}^2(t)} \leq (t_3 - t)^2.$$

Note that $(t_3-t)^2-(t_3-t-\frac{1}{2})^2 = t_3-t-\frac{1}{4} \geq 1$. But the ratio $\frac{n\tilde{I}^{2/3}(t)}{\tilde{I}^2(t)} = \frac{n}{\tilde{I}^{4/3}(t)}$ is $o(1)$. Indeed, on $\mathcal{E}_t$, $\tilde{I}(s) \geq \tilde{I}(s-1)$ for all s with $t_2 + 1 \leq s \leq t$, as $1 - \tilde{I}(s-1)(t_3 - s + 1)^2/n \geq 1 - 2^{s-1}(t_3 - s + 1)^2/n \geq 1/2$, which can be shown by applying elementary methods.

Thus, by (9),

$$\tilde{I}(t) \geq \tilde{I}(t_2) \geq \frac{n}{(2\ln^6 n)} \tag{15}$$

for n sufficiently large. Therefore $\frac{n}{\tilde{I}^{4/3}(t)} \leq \frac{2^{4/3}\ln^8 n}{n^{1/3}} = o(1)$, and this concludes the proof of the proposition. □

Therefore

$$\mathrm{E}\left(\tilde{I}_1(t+1)\right) - \tilde{I}^{2/3}(t) \geq \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3 - t)^2}{n}\right).$$

We will bound $\Pr\Big(\tilde{I}_1(t+1) < \mathrm{E}\left(\tilde{I}_1(t+1)\right) - \tilde{I}^{2/3}(t)\Big)$ using the Hoeffding–Azuma inequality. Note that if we change the choice of one vertex in $\tilde{\mathcal{N}}(t+1)$,

then $\tilde{I}_1(t+1)$ can change by at most $t_3 - t$. Also recall that $\tilde{N}(t+1) = \tilde{I}(t)$ and $t_3 - t \le \ln n$. Therefore the Hoeffding–Azuma inequality yields for n sufficiently large:

$$\begin{aligned}\Pr\left(\tilde{I}_1(t+1) < \mathrm{E}\left(\tilde{I}_1(t+1)\right) - \tilde{I}^{2/3}(t)\right) &\le 2\exp\left(-\frac{\tilde{I}^{4/3}(t)}{2\tilde{I}(t)(t_3-t)^2}\right)\\ &\le 2\exp\left(-\frac{\tilde{I}^{1/3}(t)}{2\ln^2 n}\right) \overset{(15)}{\le} 2\exp\left(-\frac{n^{1/3}}{2^{4/3}\ln^4 n}\right) \le \exp\left(-n^{1/4}\right).\end{aligned} \tag{16}$$

So (12) holds with probability at least $1 - \exp\left(-n^{1/4}\right)$.

Proof of (13). The proof of (13) is also based on the application of the Hoeffding–Azuma inequality. We begin with a lower bound on the expected value of $\tilde{I}_2(t+1)$ conditional on $\mathcal{E}_t$. Again all probabilities and expected values are conditional on $\mathcal{E}_t$. We will give a lower bound on the probability that a given vertex $v \in \tilde{\mathcal{N}}(t+1)$ belongs to $\tilde{\mathcal{I}}_2(t+1)$. First, we will expose the random choice of v, and we will condition on the event that there is no $v' \in \tilde{\mathcal{N}}(t+1)$ which belongs to $\ell(v; t+2, t_3)$. The probability that such a vertex exists is at most $\frac{\tilde{N}(t+1)(t_3-t-1)}{n} = \frac{\tilde{I}(t)(t_3-t-1)}{n}$. Having fixed the choice of v, the probability that, for a given vertex $v' \in \tilde{\mathcal{N}}(t+1)$ which is different from v, the segments $\ell(v; t+2, t_3)$ and $\ell(v'; t+2, t_3)$ are disjoint is at least $1 - \frac{(t_3-t-1)^2}{n}$. Since the random choices of the vertices in $\tilde{\mathcal{N}}(t+1)$ are independent, we obtain

$$\begin{aligned}\Pr\left(v \in \tilde{\mathcal{I}}_2(t+1)\right) &\ge \left(1 - \frac{\tilde{I}(t)(t_3-t-1)}{n}\right)\left(1 - \frac{(t_3-t-1)^2}{n}\right)^{\tilde{N}(t+1)-1}\\ &\overset{\tilde{N}(t+1)=\tilde{I}(t)}{\ge} \left(1 - \frac{\tilde{I}(t)(t_3-t-1)}{n}\right)\left(1 - \frac{(t_3-t-1)^2}{n}\right)^{\tilde{I}(t)}\\ &\ge \left(1 - \frac{\tilde{I}(t)(t_3-t-1)}{n}\right)\left(1 - \frac{\tilde{I}(t)(t_3-t-1)^2}{n}\right)\\ &\ge 1 - \frac{\tilde{I}(t)}{n}\left((t_3-t-1) + (t_3-t-1)^2\right) \ge 1 - \frac{\tilde{I}(t)(t_3-t-\frac{1}{2})^2}{n}.\end{aligned}$$

Therefore,

$$\mathrm{E}\left(\tilde{I}_2(t+1)\right) \ge \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t_3-t-\frac{1}{2})^2}{n}\right). \tag{17}$$

As in the proof of (12), we will use the Hoeffding–Azuma inequality to bound the probability that $\tilde{I}_2(t+1) < \mathrm{E}(\tilde{I}_2(t+1)) - \tilde{I}^{2/3}(t)$. This is indeed sufficient to show that inequality (13) holds with high probability. Proposition 4.4 yields

$$\mathrm{E}\left(\tilde{I}_2(t+1)\right) - \tilde{I}^{2/3}(t) \geq \tilde{I}(t)\left(1 - \frac{\tilde{I}(t)(t-t_3)^2}{n}\right).$$

Let us now see what happens to $\tilde{I}_2(t+1)$ when we change the choice of one vertex in $\tilde{\mathcal{N}}(t+1)$. Observe first that for any $v, v' \in \tilde{\mathcal{I}}_2(t+1)$ we have $\ell(v; t+2, t_3) \cap \ell(v'; t+2, t_3) = \emptyset$, if $v \neq v'$, and $v' \notin \ell(v; t+2, t_3)$. Therefore, if we change the choice of one vertex in $\tilde{\mathcal{N}}(t+1)$ we may "destroy" and also "create" at most $t_3 - (t+2) + 1 + 1$ of them (the last term comes from the fact that the vertex whose choice we change might also be "destroyed" or "created"). So $\tilde{I}_2(t+1)$ changes by at most $t_3 - t$. Therefore, applying the Hoeffding–Azuma inequality as in (16), we deduce that for n large enough,

$$\Pr\left(\tilde{I}_2(t+1) < \mathrm{E}\left(\tilde{I}_2(t+1)\right) - \tilde{I}^{2/3}(t)\right) \leq \exp\left(-n^{1/4}\right).$$

In turn, this implies that (13) holds with probability at least $1 - \exp\left(-n^{1/4}\right)$. □

The recursive relation in the above lemma implies the following bound on $\tilde{I}(t)$:

4.5 Lemma. *If $\mathcal{E}_{t_3-1}$ is realized, then for all $t \in \{t_2, \ldots, t_3 - 1\}$ we have*

$$\tilde{I}(t) \geq 2^{t-t_2}\tilde{I}(t_2) - 2 \cdot \frac{2^{2(t-t_2)}\tilde{I}^2(t_2)(t_3 - t + 2)^2}{n}.$$

Proof. We will show this by induction on t. Clearly, for $t = t_2$ this holds. Assume now that it also holds for some t with $t_2 \leq t \leq t_3 - 2$. As $\mathcal{E}_{t_3-1}$ holds,

$$\tilde{I}(t+1) \geq 2\tilde{I}(t) - 2\frac{\tilde{I}^2(t)(t_3 - t)^2}{n}. \tag{18}$$

We will use the induction hypothesis to bound $\tilde{I}(t)$ from below, as well as

the trivial upper bound $\tilde{I}(t) \leq 2^{t-t_2}\tilde{I}(t_2)$. Thus (18) becomes

$$\begin{aligned}
&\tilde{I}(t+1)\\
&\geq 2^{t+1-t_2}\tilde{I}(t_2) - 2\cdot\frac{2^{2(t-t_2)+1}\tilde{I}^2(t_2)(t_3-t+2)^2}{n} - 2\cdot\frac{2^{2(t-t_2)}\tilde{I}^2(t_2)(t_3-t)^2}{n}\\
&= 2^{t+1-t_2}\tilde{I}(t_2) - 2\cdot\frac{2^{2(t-t_2)+2}\tilde{I}^2(t_2)}{n}\left(\frac{1}{2}(t_3-t+2)^2 + \frac{1}{4}(t_3-t)^2\right)\\
&\geq 2^{t+1-t_2}\tilde{I}(t_2) - 2\cdot\frac{2^{2(t-t_2)+2}\tilde{I}^2(t_2)(t_3-t+1)^2}{n},
\end{aligned}$$

which concludes the proof of the lemma. □

This lemma implies that on $\mathcal{E}_{t_3-1}$,

$$\tilde{I}(t_3-1) \geq 2^{t_3-t_2-1}\tilde{I}(t_2) - 18\cdot\frac{2^{2(t_3-t_2-1)}\tilde{I}^2(t_2)}{n}.$$

So as $\frac{\ln^6 n}{\omega} \leq 2^{t_3-t_2-1} \leq 2\cdot\frac{\ln^6 n}{\omega}$ and $\frac{n}{\ln^6 n}\left(1-\frac{1}{\ln n}\right) \leq \tilde{I}(t_2) \leq 2^{t_2} \leq 2\cdot\frac{n}{\ln^6 n}$ on $\mathcal{E}_{t_3-1}$, we have

$$\begin{aligned}
\tilde{I}(t_3-1) &\geq \frac{\ln^6 n}{\omega}\frac{n}{\ln^6 n}\left(1-\frac{1}{\ln n}\right) - 18\cdot\frac{4\left(\frac{\ln^6 n}{\omega}\right)^2 4\left(\frac{n}{\ln^6 n}\right)^2}{n}\\
&= \frac{n}{\omega}\left(1-\frac{1}{\ln n}\right) - 288\cdot\frac{n}{\omega^2}\\
&= \frac{n}{\omega}\left(1-\frac{1}{\ln n}-\frac{288}{\omega}\right) \quad = \quad \frac{n}{\omega}\left(1-o(1)\right).
\end{aligned}$$

The definition of $\tilde{\mathcal{I}}(t_3-1)$ implies that for n large enough,

$$I(t_3) \geq \tilde{I}(t_3-1) + \tilde{N}(t_3) = 2\tilde{I}(t_3-1) \geq \frac{n}{\omega}. \tag{19}$$

4.2.2 The Final Stages

Let $\mathcal{U}(t)$ denote the set of uninformed vertices after the first t rounds, that is, $\mathcal{U}(t) := V_n \setminus \mathcal{I}(t)$, and let $U(t)$ denote its size. For $t \geq 0$ and $s > t$ we set

$$\mathcal{U}_t(s) := \mathcal{U}(t) \setminus \bigcup_{v\in\mathcal{I}(t)} \ell(v; t+1, s-1)$$

(where we assume that $\ell(v;t+1,t)=\emptyset$), and we let $U_t(s) := |\mathcal{U}_t(s)|$.

These quantities are fairly important, as they help us to describe the evolution of the process after time t_3. In particular, observe that $\mathcal{U}_t(t+1) \setminus \mathcal{U}_t(t+2) = \mathcal{N}(t+1)$. Thus if we have estimates for $U_t(t+1)$ and $U_t(t+2)$ we will be able to estimate $N(t+1)$ as well. Let $T := t_3 + \omega^2 + \lceil \ln n \rceil + \omega$ and without loss of generality assume that ω takes integral values. Throughout this subsection we will be using the fact that $T < 3\ln n$, for any n sufficiently large. Also we set $T' := t_3 + \omega^2 + \left\lfloor \frac{1}{2}\ln n \right\rfloor$. Our aim is to keep track of $U_t(s)$ for all $t = 0,\dots,T'$ and all $t < s \le T$. Eventually, we will show that the expected value of $U_{T'}(T)$ is $o(1)$, which implies that $U_{T'}(T)$ is zero with probability $1-o(1)$. In turn, the definition of the set $\mathcal{U}_{T'}(T)$ implies that if the process continues after step T' until step T, then all vertices will be informed and this will conclude the proof of the upper bound on $S(n)$.

For the lower bound, we will set $T_- := t_3 + \omega^2 + \ln n - 4\ln\ln n$ and will use a second moment argument to show that $\mathcal{U}_{T'}(T_-)$ is nonempty with probability $1-o(1)$. Of course this does not imply that if we let the process continue until time T_-, there will still be uninformed vertices. However, we will also show that with probability $1-o(1)$ none of the vertices in $\mathcal{U}_{T'}(T_-)$ is informed between rounds T' and T_-, which will conclude the proof of the lower bound on $S(n)$.

We begin with an estimate of a random variable, which, as we will see later, approximates the conditional expectation of $U_t(s)$ given the history of the process until step t. For $0 \le t < s$,

$$\overline{U}_t(s) := (n-1)\prod_{r=0}^{t}\left(1-\frac{s-r-1}{n}\right)^{N(r)}, \tag{20}$$

where $N(0)=1$. We will show the following.

4.6 Lemma. *If n is large enough, then for all $0 < t < s \le T$,*

$$\overline{U}_t(s) \le (n-1)\left(1-\frac{1}{n}\right)^{(s-t-1)I(t)+\sum_{r=0}^{t-1} I(r)} \tag{21}$$

and

$$\overline{U}_t(s) \ge (n-1)\left(1-\frac{1}{n}\right)^{(s-t-1)I(t)+\sum_{r=0}^{t-1} I(r)}\left(1-\frac{10\ln^2 n}{n^2}\right)^{I(t)}. \tag{22}$$

Proof. For the upper bound observe first that

$$\overline{U}_t(s) = (n-1)\prod_{r=0}^{t}\left(1-\frac{s-r-1}{n}\right)^{N(r)} \leq (n-1)\left(1-\frac{1}{n}\right)^{\sum_{r=0}^{t}(s-r-1)N(r)}.$$

Now, we write

$$\sum_{r=0}^{t}(s-r-1)N(r) = (s-1)\sum_{r=0}^{t}N(r) - \sum_{r=1}^{t}rN(r) = (s-1)I(t) - \sum_{r=1}^{t}rN(r).$$

But note that

$$\sum_{r=1}^{t}rN(r) = \sum_{r=0}^{t-1}(I(t)-I(r)) = tI(t) - \sum_{r=0}^{t-1}I(r).$$

Therefore

$$\sum_{r=0}^{t}(s-r-1)N(r) = (s-t-1)I(t) + \sum_{r=0}^{t-1}I(r) \tag{23}$$

and (21) follows.

To show (22) observe first that

$$1-\frac{s-r-1}{n} \geq \left(1-\frac{1}{n}\right)^{s-r-1} - \frac{(s-r-1)^2}{n^2},$$

which follows, for example, from Bonferroni inequalities (see [Bol01, Theorem 1.10, p. 17]). Thus for n sufficiently large,

$$\begin{aligned}1-\frac{s-r-1}{n} &\geq \left(1-\frac{1}{n}\right)^{s-r-1}\left(1-\frac{s^2}{n^2(1-1/n)^s}\right)\\ &\overset{s\leq T<3\ln n}{\geq} \left(1-\frac{1}{n}\right)^{s-r-1}\left(1-\frac{10\ln^2 n}{n^2}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\overline{U}_t(s) &\geq (n-1)\prod_{r=0}^{t}\left(1-\frac{1}{n}\right)^{N(r)(s-r-1)}\left(1-\frac{10\ln^2 n}{n^2}\right)^{N(r)}\\ &\overset{(23)}{=} (n-1)\left(1-\frac{1}{n}\right)^{(s-t-1)I(t)+\sum_{r=0}^{t-1}I(r)}\left(1-\frac{10\ln^2 n}{n^2}\right)^{I(t)},\end{aligned}$$

and the lower bound follows. □

Let $\mathcal{D}_t$ be the event that for all $0 \le r \le t$ and for all $r < s \le T' + 1$ we have

$$U_r(s) \in \overline{U}_r(s) \left(1 \pm \frac{1}{\ln^3 n}\right)^{r+1}. \tag{24}$$

We now define

$$\mathcal{A}_t := \begin{cases} \mathcal{D}_t, & t < t_3, \\ \mathcal{D}_t \cap \mathcal{E}_{t_3}, & t \ge t_3. \end{cases} \tag{25}$$

Note that $\mathcal{A}_0$ occurs with probability 1 as $U_0(s) \in \{n - s, n - s + 1\} \in \overline{U}_0(s)\left(1 \pm \frac{1}{\ln^3 n}\right)$. We will show by induction on t that $\mathcal{A}_{T'}$ occurs with probability $1 - o(1)$. To this aim we need the following lemma.

4.7 Lemma. *For all $0 \le t < T'$ we have*

$$\Pr(\overline{\mathcal{D}_{t+1}} \mid \mathcal{A}_t) \le e^{-n^{\frac{1}{5\omega}}}.$$

Proof. We will show that $\mathcal{D}_{t+1}$ occurs with probability at least $1 - e^{-n^{\frac{1}{5\omega}}}$; that is, we will show that with probability at least $1 - e^{-n^{\frac{1}{5\omega}}}$ we have for every s with $t + 1 < s \le T' + 1$,

$$U_{t+1}(s) \in \overline{U}_{t+1}(s) \left(1 \pm \frac{1}{\ln^3 n}\right)^{t+2}. \tag{26}$$

In fact, we will show that $U_{t+1}(s)$ is concentrated around its expected value conditional on its history up to step t.

Let a_t be a realization of the process up to step t fulfilling $\mathcal{A}_t$. Note that

$$\begin{aligned} \mathrm{E}(U_{t+1}(s) \mid a_t) &= U_t(s) \left(1 - \frac{s - t - 2}{n}\right)^{N(t+1)} \\ &\in \overline{U}_t(s) \left(1 \pm \frac{1}{\ln^3 n}\right)^{t+1} \left(1 - \frac{s - t - 2}{n}\right)^{N(t+1)} = \overline{U}_{t+1}(s) \left(1 \pm \frac{1}{\ln^3 n}\right)^{t+1}, \end{aligned} \tag{27}$$

where $\mathcal{N}(t + 1)$ is determined by a_t. Thus, it suffices to show that with conditional probability at least $1 - e^{-n^{\frac{1}{5\omega}}}$ for all s with $t + 1 < s \le T' + 1$,

$$|U_{t+1}(s) - \mathrm{E}(U_{t+1}(s) \mid a_t)| \le \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n}. \tag{28}$$

Note that $U_{t+1}(s) = U_t(s) - L_{t+1}(s)$, where $L_{t+1}(s)$ is a nonnegative random variable that is equal to the number of vertices which are removed from $\mathcal{U}_t(s)$ during the $(t+1)$st round. Thus

$$|U_{t+1}(s) - \mathrm{E}(U_{t+1}(s) \mid a_t)| = |L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)|\,.$$

So it suffices to show that with probability at least $1 - e^{-n^{\frac{1}{5\omega}}}$ for all s with $t+1 < s \le T'+1$,

$$|L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)| \le \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n}. \tag{29}$$

In the present situation, we will use Talagrand's inequality. In particular, this is useful when the number of informed vertices has become linear. There, the number of vertices which make a random choice is quite large compared to the conditional expectation of $L_{t+1}(s)$ and, for example, the Hoeffding–Azuma inequality would give trivial bounds. Thus, we need to apply a stronger tool such as Talagrand's inequality (see Theorem 1.6).

We will apply it to $L_{t+1}(s)$. Note that $L_{t+1}(s)$ is a function of the independent random choices of the vertices in $\mathcal{N}(t+1)$. Note also that if we change only one of these random choices, then $L_{t+1}(s)$ can change by at most $s - 1 - (t+2) + 1 < s \le T < 3\ln n$. Therefore, we may take $c_k := 3\ln n$ for all k. Regarding condition b in the above theorem and the definition of ψ, observe that if $L_{t+1}(s) \ge r$, then there must be some vertices in $\mathcal{N}(t+1)$ which force $L_{t+1}(s)$ to be at least r and these must be no more than r. In other words, if a vertex v is removed from $\mathcal{U}_t(s)$, then there must be some vertex $v' \in \mathcal{N}(t+1)$ for which $v \in \ell(v'; t+2, s-1)$. So if r vertices are removed from $\mathcal{U}_t(s)$ after the exposure of the random choices of the vertices in $\mathcal{N}(t+1)$, then there must be at most r of them that certify this. So we may take $\psi(r) = 9r\ln^2 n$.

However, we would like to show the concentration of $L_{t+1}(s)$ around its expected value. If m now denotes the (conditional) median of $L_{t+1}(s)$, then the triangle inequality implies that

$$|L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)| \le |L_{t+1}(s) - m| + |m - \mathrm{E}(L_{t+1}(s) \mid a_t)|.$$

But $|m - \mathrm{E}(L_{t+1}(s) \mid a_t)|$ is not very large. In fact, we show the following.

4.8 Proposition.

$$|m - \mathrm{E}(L_{t+1}(s) \mid a_t)| = O\left(\ln n\sqrt{\mathrm{E}(L_{t+1}(s) \mid a_t)} + \ln^2 n\right).$$

Proof. The proof of this proposition is very similar to the argument that is used in Example 2.33 on page 41 in [JŁR00], but we include it for completeness.

We set $X := L_{t+1}(s)$ conditional on a_t. Talagrand's inequality (see Theorem 1.6) yields

$$\Pr\left(|X - m| \geq x\right) \leq 4\exp\left(-\frac{x^2}{36(m+x)\ln^2 n}\right).$$

But $x^2/(m+x) \geq x^2/(2m)$ for all $0 \leq x \leq m$, and $x^2/(m+x) \geq x/2$ for $x > m$. So we write

$$\begin{aligned}|m - \mathrm{E}(X)| \leq \mathrm{E}\left(|X-m|\right) &= \int_0^\infty \Pr\left(|X-m| \geq x\right) dx \\ &\leq 4\int_0^m e^{-\frac{x^2}{72m\ln^2 n}} dx + 4\int_m^\infty e^{-\frac{x}{72\ln^2 n}} dx = O\left(\ln n\sqrt{m} + \ln^2 n\right).\end{aligned}$$

Since $m \leq 2\,\mathrm{E}(X)$, the proposition follows. □

We will show that

$$\ln n\sqrt{\mathrm{E}(L_{t+1}(s) \mid a_t)} \text{ and } \ln^2 n \text{ are } o\left(\frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n}\right). \tag{30}$$

In turn, this will imply that for n sufficiently large,

$$\begin{aligned}|L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)| &> \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n} \\ &\Rightarrow |L_{t+1}(s) - m| > \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{2\ln^3 n}.\end{aligned} \tag{31}$$

To show the first part of (30) it suffices to show that

$$\frac{\sqrt{\mathrm{E}(L_{t+1}(s) \mid a_t)}}{\mathrm{E}(U_{t+1}(s) \mid a_t)} = o\left(\frac{1}{\ln^4 n}\right) \text{ , or equivalently}$$

$$\frac{\mathrm{E}(L_{t+1}(s) \mid a_t)}{\mathrm{E}^2(U_{t+1}(s) \mid a_t)} = o\left(\frac{1}{\ln^8 n}\right).$$

As $\mathrm{E}(L_{t+1}(s) \mid a_t) \leq U_t(s)$, equation (30) will follow if we show the following proposition.

4.9 Proposition.

$$\frac{U_t(s)}{\mathrm{E}^2(U_{t+1}(s) \mid a_t)} \leq \frac{1}{n^{\frac{1}{3\omega}} \ln^8 n} = o\left(\frac{1}{\ln^8 n}\right) \text{ and } \ln^5 n = o\left(\mathrm{E}(U_{t+1}(s) \mid a_t)\right).$$

Proof. We begin with the proof of the second part of the proposition, as the necessary bounds will also be useful for the proof of the first part.

As $\mathcal{A}_t$ holds, using (21) and the fact that $t \leq T' < 3 \ln n$, we obtain for n sufficiently large,

$$U_t(s) \leq \overline{U}_t(s) \left(1 + \frac{1}{\ln^3 n}\right)^{t+1} \leq 2n \left(1 - \frac{1}{n}\right)^{(s-t-1)I(t) + \sum_{r=0}^{t-1} I(r)}. \tag{32}$$

By equation (27) and for n sufficiently large,

$$\begin{aligned}
&\mathrm{E}(U_{t+1}(s) \mid a_t) \geq \overline{U}_{t+1}(s) \left(1 - \frac{1}{\ln^3 n}\right)^{t+1} \\
&\overset{(22)}{\geq} (n-1) \left(1 - \frac{1}{n}\right)^{(s-t-2)I(t+1) + \sum_{r=0}^{t} I(r)} \left(1 - \frac{10 \ln^2 n}{n^2}\right)^{I(t+1)} \left(1 - \frac{1}{\ln^3 n}\right)^{t+1} \\
&\geq (n-1) \left(1 - \frac{1}{n}\right)^{(s-t-2)I(t+1) + \sum_{r=0}^{t} I(r)} \left(1 - \frac{10 \ln^2 n}{n^2}\right)^{n} \left(1 - \frac{1}{\ln^3 n}\right)^{3 \ln n} \\
&\geq \frac{n}{2} \left(1 - \frac{1}{n}\right)^{(s-t-2)I(t+1) + \sum_{r=0}^{t} I(r)}.
\end{aligned} \tag{33}$$

To prove the second part of the proposition, we take a further lower bound on $\mathrm{E}(U_{t+1}(s) \mid a_t)$ by giving an upper bound on $(s-t-2)I(t+1) + \sum_{r=0}^{t} I(r)$. Note that, by using the bound $I(r) \leq 2^r$, we get $\sum_{r=0}^{t_3-1} I(r) \leq 2^{t_3} \leq 8\frac{n}{\omega}$. We will consider two different cases.

First, if $t < t_3$, we get

$$\begin{aligned}
(s-t-2)I(t+1) + \sum_{r=0}^{t} I(r) &\leq (s-t-2)I(t_3) + 8\frac{n}{\omega} \leq (s-t-1)8\frac{n}{\omega} \\
&\leq 8T'\frac{n}{\omega} \leq n\frac{\ln n}{2}
\end{aligned}$$

if n is large enough. By (33) this yields

$$\mathrm{E}(U_{t+1}(s) \mid a_t) \geq \frac{n}{2}\left(1-\frac{1}{n}\right)^{(s-t-2)I(t+1)+\sum_{r=0}^{t} I(r)} \geq \frac{n}{2}\left(1-\frac{1}{n}\right)^{n\frac{\ln n}{2}} \geq \frac{\sqrt{n}}{4}.$$

This implies the second part of the proposition for the case $t < t_3$.

In the case where $t \geq t_3$, we first split the second sum into two parts and write $\sum_{r=0}^{t} I(r) = \sum_{r=0}^{t_3-1} I(r) + \sum_{r=t_3}^{t} I(r)$. For the first part, we use the bound $\sum_{r=0}^{t_3-1} I(r) \leq 8\frac{n}{\omega}$ as above. For the second sum we use the bound $I(r) \leq n$. This implies that

$$(s-t-2)I(t+1) + \sum_{r=t_3}^{t} I(r) \leq (s-t-2+t-t_3+1)\,n \leq (s-t_3)\,n.$$

Thus by (33) and for n large enough,

$$\mathrm{E}(U_{t+1}(s) \mid a_t) \geq \frac{n}{2}\left(1-\frac{1}{n}\right)^{(s-t_3)n+\frac{8n}{\omega}} \geq \frac{n}{4}\left(1-\frac{1}{n}\right)^{(s-t_3)n}. \tag{34}$$

Using $s - t_3 \leq \frac{1}{2}\ln n + \omega^2$, this yields

$$\mathrm{E}(U_{t+1}(s) \mid a_t) \geq \frac{n}{4}\left(1-\frac{1}{n}\right)^{\left(\frac{1}{2}\ln n+\omega^2\right)n} \geq \frac{\sqrt{n}\; e^{-\omega^2}}{8}. \tag{35}$$

This implies the second part of the proposition. We will use equations (34) and (35) also later in the proof of the first part of the proposition.

For the first part of the proposition we will consider three different cases as follows.

$t < t_3$: In this case, using $I(r) \leq 2^r,\ t < t_3$, and $s - t - 1 \leq T' < 3\ln n$,

we write

$$\begin{aligned}
&\frac{U_t(s)}{\mathrm{E}^2(U_{t+1}(s) \mid a_t)} \\
&\overset{(32),(33)}{\leq} \frac{8}{n}\left(1-\frac{1}{n}\right)^{(s-t-1)I(t)+\sum_{r=0}^{t-1} I(r)-2(s-t-2)I(t+1)-2\sum_{r=0}^{t} I(r)} \\
&\leq \frac{8}{n}\left(1-\frac{1}{n}\right)^{-2(s-t-2)I(t+1)-2\sum_{r=0}^{t} I(r)} \leq \frac{8}{n}\left(1-\frac{1}{n}\right)^{-2(s-t-2)2^{t_3}-2\cdot 2^{t_3}} \\
&\leq \frac{8}{n}\left(1-\frac{1}{n}\right)^{-16(s-t-1)n/\omega} \leq \frac{8}{n}\, e^{\frac{16(s-t-1)}{\omega}} \leq \frac{8}{n}\, e^{\frac{48\ln n}{\omega}} \leq \frac{1}{n^{\frac{1}{3\omega}}\ln^8 n}.
\end{aligned} \tag{36}$$

$t_3 \leq t \leq t_4 := t_3 + \omega^2$: By (32) and since $\mathcal{E}_{t_3}$ is realized,

$$\begin{aligned}
U_t(s) &\leq 2n\left(1-\frac{1}{n}\right)^{(s-t-1)I(t)} \leq 2n\left(1-\frac{1}{n}\right)^{(s-t-1)I(t_3)} \\
&\overset{(19)}{\leq} 2n\left(1-\frac{1}{n}\right)^{(s-t-1)n/\omega}.
\end{aligned}$$

Combining this with (34) we obtain for n large enough,

$$\begin{aligned}
&\frac{U_t(s)}{\mathrm{E}^2(U_{t+1}(s) \mid a_t)} \leq \frac{32}{n}\left(1-\frac{1}{n}\right)^{(s-t-1)n/\omega-2(s-t_3)n} \leq \frac{32}{n}\, e^{-(s-t-1)/\omega+2(s-t_3)} \\
&\leq \frac{32}{n}\, e^{-(T'-t-1)/\omega+2(T'-t_3)} \leq \frac{32}{n}\, e^{-(T'-(t_3+\omega^2)-1)/\omega+2(T'-t_3)} \\
&\leq \frac{32}{n}\, e^{-(\frac{1}{2}\ln n-1)/\omega+2\left(\frac{1}{2}\ln n+\omega^2\right)} \leq \frac{33e^{2\omega^2}}{n^{\frac{1}{2\omega}}} \leq \frac{1}{n^{\frac{1}{3\omega}}\ln^8 n}.
\end{aligned} \tag{37}$$

$t_4 < t \leq T' - 1$: Here we bound the following ratio separately:

$$\begin{aligned}
\frac{U_t(s)}{\mathrm{E}(U_{t+1}(s) \mid a_t)} &\overset{(32),(33)}{\leq} 4\left(1-\frac{1}{n}\right)^{-(s-t-2)(I(t+1)-I(t))} \\
&\leq 4\left(1-\frac{1}{n}\right)^{-(s-t-2)U(t)} \leq 4\left(1-\frac{1}{n}\right)^{-(s-t-2)U(t_4)}.
\end{aligned} \tag{38}$$

We will first give an upper bound on $U(t_4)$. Recall that $U(t) = U_t(t+1)$. Thus we are able to give an upper bound on $U(t_4)$ by giving an upper bound

on $U_{t_4}(t_4+1)$. On $\mathcal{A}_t$, we have for n sufficiently large,

$$\begin{aligned} U(t_4) = U_{t_4}(t_4+1) &\overset{\mathcal{A}_t,(21)}{\leq} n\left(1-\frac{1}{n}\right)^{\sum_{r=0}^{t_4-1} I(r)} \left(1+\frac{1}{\ln^3 n}\right)^{t_4+1} \\ &\leq 2n\left(1-\frac{1}{n}\right)^{\sum_{r=t_3}^{t_4-1} I(r)} \leq 2n\left(1-\frac{1}{n}\right)^{\omega^2 I(t_3)} \\ &\overset{\mathcal{E}_{t_3},(19)}{\leq} 2n\left(1-\frac{1}{n}\right)^{\omega^2\frac{n}{\omega}} \leq 2ne^{-\omega} \leq ne^{-\omega/2}. \end{aligned} \tag{39}$$

Substituting this into (38), we obtain

$$\begin{aligned} \frac{U_t(s)}{\mathrm{E}(U_{t+1}(s) \mid a_t)} &\leq 4\left(1-\frac{1}{n}\right)^{-(s-t-2)ne^{-\omega/2}} \leq 4e^{(s-t)e^{-\omega/2}} \\ &\overset{s-t\leq\frac{1}{2}\ln n}{\leq} 4e^{\frac{e^{-\omega/2}}{2}\ln n} = 4n^{\frac{1}{2e^{\omega/2}}}. \end{aligned} \tag{40}$$

Now equations (40) and (35) yield

$$\frac{U_t(s)}{\mathrm{E}^2(U_{t+1}(s) \mid a_t)} \leq \frac{32\ e^{\omega^2}}{n^{\frac{1}{2}-\frac{1}{2e^{\omega/2}}}} \leq \frac{1}{n^{\frac{1}{3\omega}}\ln^8 n}, \tag{41}$$

and this concludes the proof of the proposition. □

We will apply Talagrand's inequality with $x = \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{2\ln^3 n}$. We use the observation (see, for example, page 42 in [JŁR00]) that $m \leq 2\,\mathrm{E}(L_{t+1}(s) \mid a_t) \leq 2U_t(s)$ and also $x \leq \mathrm{E}(U_{t+1}(s) \mid a_t) \leq U_t(s)$. Thus $\psi(m+x) \leq 27\ U_t(s)\ln^2 n$. Therefore by (2) and (31) and for n sufficiently large,

$$\begin{aligned} &\Pr\left(|L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)| > \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n} \;\middle|\; a_t\right) \\ &\leq 4\exp\left(-\frac{\mathrm{E}^2(U_{t+1}(s) \mid a_t)}{16\cdot 27\ U_t(s)\ln^8 n}\right) \overset{\text{Proposition 4.9}}{\leq} e^{-n^{\frac{1}{4\omega}}}. \end{aligned} \tag{42}$$

Thus for n sufficiently large,

$$\begin{aligned} &\Pr(\overline{\mathcal{D}_{t+1}} \mid a_t) \ \leq \\ &\Pr\left(\exists s \in \{t+2\ldots T'\} : |L_{t+1}(s) - \mathrm{E}(L_{t+1}(s) \mid a_t)| > \frac{\mathrm{E}(U_{t+1}(s) \mid a_t)}{\ln^3 n} \;\middle|\; a_t\right) \\ &\overset{(42)}{\leq} T'e^{-n^{\frac{1}{4\omega}}} < 3\ln n\ e^{-n^{\frac{1}{4\omega}}} \leq e^{-n^{\frac{1}{5\omega}}}. \end{aligned} \tag{43}$$

Averaging over all a_t such that $\mathcal{A}_t$ holds, we deduce $\Pr(\overline{\mathcal{D}_{t+1}} \mid \mathcal{A}_t) \le e^{-n^{\frac{1}{5\omega}}}$. □

By Lemma 4.7, for $t < t_3 - 1$,

$$\Pr(\mathcal{A}_{t+1} \mid \mathcal{A}_t) \ge 1 - e^{-n^{\frac{1}{5\omega}}}, \tag{44}$$

and therefore

$$\begin{aligned} \Pr(\mathcal{A}_{t_3-1}) = \Pr(\mathcal{A}_0) \prod_{t=0}^{t_3-2} \Pr(\mathcal{A}_{t+1} \mid \mathcal{A}_t) &\ge \left(1 - e^{-n^{\frac{1}{5\omega}}}\right)^{t_3} \\ &\ge \left(1 - e^{-n^{\frac{1}{5\omega}}}\right)^{\log_2 n} = 1 - o(1). \end{aligned} \tag{45}$$

So

$$\begin{aligned} \Pr(\overline{\mathcal{A}_{t_3}}) = \Pr(\overline{\mathcal{D}_{t_3} \cap \mathcal{E}_{t_3}}) &\le \Pr(\overline{\mathcal{D}_{t_3}}) + \Pr(\overline{\mathcal{E}_{t_3}}) \overset{(8),(11)}{\le} \Pr(\overline{\mathcal{D}_{t_3}}) + o(1) \\ &\le \Pr(\overline{\mathcal{D}_{t_3}} \mid \mathcal{A}_{t_3-1}) + \Pr(\overline{\mathcal{A}_{t_3-1}}) + o(1) \\ &\overset{(45)}{=} \Pr(\overline{\mathcal{D}_{t_3}} \mid \mathcal{A}_{t_3-1}) + o(1) \overset{\text{Lemma 4.7}}{=} o(1). \end{aligned} \tag{46}$$

Now note that for any $t \ge t_3$,

$$\Pr(\overline{\mathcal{A}_{t+1}} \mid \mathcal{A}_t) = \Pr(\overline{\mathcal{D}_{t+1}} \mid \mathcal{A}_t) \le e^{-n^{\frac{1}{5\omega}}}.$$

So using (46) and the above inequality,

$$\begin{aligned} \Pr(\mathcal{A}_{T'}) = \Pr(\mathcal{A}_{t_3}) \prod_{t=t_3}^{T'-1} \Pr(\mathcal{A}_{t+1} \mid \mathcal{A}_t) &\ge (1 - o(1))\left(1 - e^{-n^{\frac{1}{5\omega}}}\right)^{T'} \\ &\ge (1 - o(1))\left(1 - e^{-n^{\frac{1}{5\omega}}}\right)^{3 \ln n} = 1 - o(1). \end{aligned} \tag{47}$$

Now we will study the evolution of $I(t)$ on the event $\mathcal{A}_{T'}$. Recall that $t_4 = t_3 + \omega^2$. We first prove the following lemma.

4.10 Lemma. *On $\mathcal{A}_{T'}$, if n is sufficiently large, then for any t such that $t_4 \le t \le T' - 1$, we have*

$$\frac{U(t+1)}{U(t)} \le \frac{2}{e}.$$

Proof. As we mentioned above, $U(t) = U_t(t+1)$. But also observe that $U(t+1) = U_t(t+2)$. Therefore

$$\frac{U(t+1)}{U(t)} = \frac{U_t(t+2)}{U_t(t+1)}.$$

Thus we may use the estimates on $U_t(t+2)$ and $U_t(t+1)$ on the event $\mathcal{A}_{T'}$. So we have

$$\frac{U_t(t+2)}{U_t(t+1)} \overset{\mathcal{A}_{T'},(24)}{\leq} \frac{\overline{U}_t(t+2)}{\overline{U}_t(t+1)} \left(1 + \frac{1}{\ln^3 n}\right)^{t+1} \left(1 - \frac{1}{\ln^3 n}\right)^{-t-1}. \tag{48}$$

By Lemma 4.6 (equations (21) and (22)), if n is large enough, then

$$\begin{aligned} \frac{\overline{U}_t(t+2)}{\overline{U}_t(t+1)} &\leq \frac{(1-1/n)^{\sum_{r=0}^{t} I(r)}}{(1-1/n)^{\sum_{r=0}^{t-1} I(r)}} \left(1 - \frac{10\ln^2 n}{n^2}\right)^{-n} \\ &\leq \left(1 - \frac{1}{n}\right)^{I(t)} \left(1 - \frac{10\ln^2 n}{n^2}\right)^{-n} \\ &\leq \left(1 - \frac{1}{n}\right)^{I(t_4)} \left(1 - \frac{10\ln^2 n}{n^2}\right)^{-n}, \end{aligned} \tag{49}$$

where we used $I(t) \geq I(t_4)$ in the last inequality.

But by (39), $I(t_4) \geq n - ne^{-\omega/2}$. Thus

$$\left(1 - \frac{1}{n}\right)^{I(t_4)} \leq \left(1 - \frac{1}{n}\right)^{n - ne^{-\omega/2}} \leq e^{-1}e^{e^{-\omega/2}}. \tag{50}$$

Now we combine (48), (49), and (50) and deduce that for n sufficiently large,

$$\begin{aligned} \frac{U_t(t+2)}{U_t(t+1)} &\leq e^{-1}e^{e^{-\omega/2}} \left(1 + \frac{1}{\ln^3 n}\right)^{t+1} \left(1 - \frac{1}{\ln^3 n}\right)^{-t-1} \left(1 - \frac{10\ln^2 n}{n^2}\right)^{-n} \\ &\overset{t<T'<3\ln n}{\leq} \frac{2}{e}. \end{aligned}$$

□

4.2.3 Proof of Theorem 4.1: The Upper Bound

We are now ready to conclude the proof of the upper bound in Theorem 4.1. We will condition on $\mathcal{A}_{T'}$ and will calculate $\mathrm{E}(U_{T'}(T) \mid \mathcal{A}_{T'})$, proving that it is $o(1)$. So the upper bound in Theorem 4.1 will follow from Markov's inequality.

Let $v \in V_n \setminus \{1\}$, and let $(a_0, \ldots, a_{T'})$ be a realization of the process up to time T' fulfilling the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$. More precisely, a_i is an ordered set of $i+1$ sets with the jth set containing ordered pairs, where the first element is a vertex informed at step j and the second is its random choice. Moreover, these $i+1$ sets describe a feasible realization of the process until the ith round with the constraints that the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$ are not violated. Then

$$\Pr(a_0) = \Pr(a_0,\ v \in \mathcal{U}_0(T)) = \Pr(v \in \mathcal{U}_0(T)) \Pr(a_0 \mid v \in \mathcal{U}_0(T)), \tag{51}$$

and, since a_{t+1} satisfies the event "$v \in \mathcal{U}_{t+1}(T)$", one has

$$\begin{aligned} \Pr(a_{t+1} \mid a_t) &= \Pr(a_{t+1},\ v \in \mathcal{U}_{t+1}(T) \mid a_t) \\ &= \Pr(v \in \mathcal{U}_{t+1}(T) \mid a_t) \Pr(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t). \end{aligned} \tag{52}$$

So summing over all realizations $(a_0, \ldots, a_{T'})$ of the process up to time T' fulfilling the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$, we write

$$\begin{aligned} \Pr(v \in \mathcal{U}_{T'}(T),\ \mathcal{A}_{T'}) &= \sum_{(a_0,\ldots,a_{T'})} \Pr(a_0, \ldots, a_{T'}) \\ &= \sum_{(a_0,\ldots,a_{T'})} \Pr(a_0) \prod_{t=0}^{T'-1} \Pr(a_{t+1} \mid a_t) \\ &\overset{(51),(52)}{=} \sum_{(a_0,\ldots,a_{T'})} \Pr(v \in \mathcal{U}_0(T)) \Pr(a_0 \mid v \in \mathcal{U}_0(T)) \\ &\qquad \times \prod_{t=0}^{T'-1} \Pr(v \in \mathcal{U}_{t+1}(T) \mid a_t) \Pr(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t). \end{aligned} \tag{53}$$

First, we will obtain a uniform bound on

$$\Pr(v \in \mathcal{U}_0(T)) \prod_{t=0}^{T'-1} \Pr(v \in \mathcal{U}_{t+1}(T) \mid a_t).$$

4.11 Proposition. *If n is large enough, then for all realizations $(a_0, \dots, a_{T'})$ of the process up to step T' fulfilling the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$ we have*

$$\Pr\left(v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(v \in \mathcal{U}_{t+1}(T) \mid a_t\right) \leq \frac{2e^{-\omega}}{n}.$$

Proof. Let $N(r, a_{r-1})$ denote the number of newly informed vertices at step r given the history of the process up to step $r-1$; note that the latter determines this set. But as $(a_0, \dots, a_{T'})$ fulfills $\mathcal{A}_{T'}$,

$$\begin{aligned}
&\Pr\left(v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(v \in \mathcal{U}_{t+1}(T) \mid a_t\right) \\
&\qquad = \left(1 - \frac{T-1}{n}\right)^{N(0)} \prod_{r=1}^{T'} \left(1 - \frac{T-r-1}{n}\right)^{N(r,a_{r-1})} \qquad (54) \\
&\qquad \leq \left(1 - \frac{1}{n}\right)^{(T-1)N(0) + \sum_{r=1}^{T'} (T-r-1)N(r,a_{r-1})}.
\end{aligned}$$

To bound this from above, we will give a lower bound on the exponent:

$$\begin{aligned}
(T-1)N(0) + \sum_{r=1}^{T'} (T-r-1)N(r,a_{r-1}) &\overset{(23)}{\geq} (T-T'-1)I(T') + \sum_{r=0}^{T'-1} I(r) \\
&\geq (T-T'-1)I(T') + \sum_{r=t_4}^{T'-1} I(r) \\
&= (T-T'-1)(n-U(T')) + \sum_{r=t_4}^{T'-1} (n-U(r)) \\
&= (T-t_4-1)n - (T-T'-1)U(T') - \sum_{r=t_4}^{T'-1} U(r).
\end{aligned} \qquad (55)$$

Therefore (54) yields

$$\begin{aligned}
&\Pr\left(v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(v \in \mathcal{U}_{t+1}(T) \mid a_t\right) \\
&\qquad \leq \left(1-\frac{1}{n}\right)^{(T-t_4-1)n-(T-T'-1)U(T')-\sum_{r=t_4}^{T'-1}U(r)} \\
&\qquad \leq e^{-\ln n-\omega} e^{\frac{(T-T'-1)U(T')+\sum_{r=t_4}^{T'-1}U(r)}{n}} \\
&\qquad \leq \frac{e^{-\omega}}{n} e^{\frac{(T-T')U(T')+\sum_{r=t_4}^{T'-1}U(r)}{n}}.
\end{aligned} \tag{56}$$

By Lemma 4.10 for n sufficiently large,

$$U(T') \leq \left(\frac{2}{e}\right)^{T'-t_4} U(t_4) \overset{(39)}{\leq} 2\left(\frac{2}{e}\right)^{\frac{\ln n}{2}} n e^{-\frac{\omega}{2}}, \tag{57}$$

and also

$$\sum_{r=t_4}^{T'-1} U(r) \leq \sum_{r=t_4}^{T'-1} \left(\frac{2}{e}\right)^{r-t_4} U(t_4) \leq \frac{e}{e-2} U(t_4) \overset{(39)}{\leq} n\ \frac{e}{e-2}\ e^{-\frac{\omega}{2}}. \tag{58}$$

So (57) and (58) imply that

$$\frac{(T-T')U(T') + \sum_{r=t_4}^{T'-1} U(r)}{n} \leq 6\ln n \left(\frac{2}{e}\right)^{\frac{\ln n}{2}} e^{-\frac{\omega}{2}} + \frac{e}{e-2}\ e^{-\frac{\omega}{2}} = o(1).$$

Substituting this bound into (56), we obtain

$$\Pr\left(v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(v \in \mathcal{U}_{t+1}(T) \mid a_t\right) \leq \frac{2e^{-\omega}}{n}.$$

□

So (53) becomes

$$\begin{aligned}
&\Pr\left(v \in \mathcal{U}_{T'}(T),\ \mathcal{A}_{T'}\right) \\
&\quad \leq \frac{2e^{-\omega}}{n} \sum_{(a_0,\ldots,a_{T'})} \Pr\left(a_0 \mid v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t\right).
\end{aligned} \tag{59}$$

But for the sum on the right-hand side, we have the following.

4.12 Proposition. *For the sum over all realizations $(a_0, \dots, a_{T'})$ of the process up to time T' fulfilling the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$, we have*

$$\sum_{(a_0,\dots,a_{T'})} \Pr\left(a_0 \mid v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t\right) \leq 1.$$

Proof. Consider the following probability space: We run the process from 0 to T' as before, except that in each step t each vertex $u \in \mathcal{N}_t$ selects its starting position uniformly at random from those positions such that $v \notin \ell(u; t+1, T-1)$. In other words, u selects a position in its list such that v is avoided in the segment of length $T-t-1$ starting at this position. We will denote the probabilities in this probability space by $\widetilde{\Pr}$. All realizations $(a_0, \dots, a_{T'})$ of the process up to time T' fulfilling the events "$v \in \mathcal{U}_{T'}(T)$" and $\mathcal{A}_{T'}$ are in this modified probability space and we have

$$\Pr\left(a_0 \mid v \in \mathcal{U}_0(T)\right) = \widetilde{\Pr}\left(a_0\right)$$

and in general, for all $0 \leq t \leq T'-1$,

$$\Pr\left(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t\right) = \widetilde{\Pr}\left(a_{t+1} \mid a_t\right).$$

So

$$\begin{aligned}\Pr\left(a_0 \mid v \in \mathcal{U}_0(T)\right) &\prod_{t=0}^{T'-1} \Pr\left(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t\right) \\ &= \widetilde{\Pr}\left(a_0\right) \prod_{t=0}^{T'-1} \widetilde{\Pr}\left(a_{t+1} \mid a_t\right) = \widetilde{\Pr}\left(a_0, \dots, a_{T'}\right)\end{aligned}$$

and

$$\begin{aligned}\sum_{(a_0,\dots,a_{T'})} &\Pr\left(a_0 \mid v \in \mathcal{U}_0(T)\right) \prod_{t=0}^{T'-1} \Pr\left(a_{t+1} \mid v \in \mathcal{U}_{t+1}(T),\ a_t\right) \\ &= \sum_{(a_0,\dots,a_{T'})} \widetilde{\Pr}\left(a_0, \dots, a_{T'}\right) \leq 1.\end{aligned}$$

□

So (59) becomes

$$\Pr\left(v \in \mathcal{U}_{T'}(T),\ \mathcal{A}_{T'}\right) \leq \frac{2e^{-\omega}}{n}. \tag{60}$$

Bayes' rule now yields for n sufficiently large,

$$\Pr\left(v \in \mathcal{U}_{T'}(T) \mid \mathcal{A}_{T'}\right) = \frac{\Pr\left(v \in \mathcal{U}_{T'}(T),\ \mathcal{A}_{T'}\right)}{\Pr(\mathcal{A}_{T'})} \overset{(47),(60)}{\leq} \frac{3e^{-\omega}}{n},$$

and therefore

$$\mathrm{E}(U_{T'}(T) \mid \mathcal{A}_{T'}) = \sum_{v \in V_n \setminus \{1\}} \Pr\left(v \in \mathcal{U}_{T'}(T) \mid \mathcal{A}_{T'}\right) \leq 3e^{-\omega} = o(1). \tag{61}$$

We now have

$$\begin{aligned} \Pr(S(n) > T) &\leq \Pr(S(n) > T \mid \mathcal{A}_{T'}) + \Pr(\overline{\mathcal{A}_{T'}}) \\ &\overset{(47)}{\leq} \Pr(U_{T'}(T) > 0 \mid \mathcal{A}_{T'}) + o(1) \\ &\leq \mathrm{E}(U_{T'}(T) \mid \mathcal{A}_{T'}) + o(1) \overset{(61)}{=} o(1), \end{aligned} \tag{62}$$

and this concludes the proof of the upper bound in Theorem 4.1.

4.2.4 Proof of Theorem 4.1: The Lower Bound

Here we set

$$T_- := t_3 + \omega^2 + \ln n - 4 \ln \ln n.$$

First, we will show that with probability $1 - o(1)$ none of the vertices in $\mathcal{U}_{T'}(T_-)$ will be informed until step T_-. We will then conclude the proof of the lower bound on $S(n)$, proving that $U_{T'}(T_-) > 0$ with probability $1 - o(1)$. We will show the latter by means of a second moment argument. In both steps, we will work conditioning on the event $\mathcal{A}_{T'}$.

Let us begin with the first step. Recall that by (57) on $\mathcal{A}_{T'}$ we have for large n

$$U(T') \leq 2 \left(\frac{2}{e}\right)^{\frac{\ln n}{2}} n\ e^{-\frac{\omega}{2}} \leq \left(\frac{2}{e}\right)^{\frac{\ln n}{2}} n.$$

The probability that a given vertex in $\mathcal{U}_{T'}(T_-)$ is informed during one of the subsequent steps, that is, from step $T' + 1$ up to T_-, is no more than

$1-\left(1-\frac{T_--T'}{n}\right)^{U(T')}$. But as $T_- - T' \le \frac{1}{2}\ln n$, the latter probability is no more than

$$\begin{aligned}1-\left(1-\frac{T_- - T'}{n}\right)^{U(T')} &\le 1-\left(1-\frac{\ln n}{2n}\right)^{\left(\frac{2}{e}\right)^{\frac{\ln n}{2}} n}\\ &\le \frac{\ln n}{2n}\left(\frac{2}{e}\right)^{\frac{\ln n}{2}} n \le \ln n\ \left(\frac{2}{e}\right)^{\frac{\ln n}{2}}.\end{aligned}$$

Also, we repeat the calculations which led to (61), replacing T by T_- (note that $T_- - t_4 = \ln n - 4\ln\ln n$). We obtain

$$\mathrm{E}(U_{T'}(T_-) \mid \mathcal{A}_{T'}) \le 3e^{4\ln\ln n}.$$

So, conditional on $\mathcal{A}_{T'}$, the expected number of vertices in $\mathcal{U}_{T'}(T_-)$ which are informed between steps T' and (including) T_- is at most $3e^{4\ln\ln n}\ln n\ \left(\frac{2}{e}\right)^{\frac{\ln n}{2}} = 3\ln^5 n\left(\frac{2}{e}\right)^{\frac{\ln n}{2}} = o(1)$. In other words, with (conditional) probability $1-o(1)$, no vertex in $\mathcal{U}_{T'}(T_-)$ is informed during these steps.

Now we conclude the proof of the lower bound, showing that conditional on $\mathcal{A}_{T'}$ with probability $1-o(1)$ we have $U_{T'}(T_-) > 0$. Since $\mathcal{A}_{T'}$ itself occurs with probability $1-o(1)$, this implies that with probability $1-o(1)$, running the process for T_- steps is not enough to inform all vertices.

Let $\mathcal{D}'_{t_3}$ be the event that for all $0 \le r \le t_3$ and for all $r < s \le T_-$ we have

$$U_r(s) \in \overline{U}_r(s)\left(1 \pm \frac{1}{\ln^3 n}\right)^{r+1}. \tag{63}$$

We now define

$$\mathcal{A}'_{t_3} := \mathcal{D}'_{t_3} \cap \mathcal{E}_{t_3}. \tag{64}$$

We can prove the following proposition.

4.13 Proposition.

$$\Pr(\mathcal{A}'_{t_3}) = 1 - o(1).$$

Indeed, the proof goes exactly as the proof of the fact that $\Pr(\mathcal{A}_{t_3}) = 1 - o(1)$. We can apply Talagrand's inequality (equation (2)), since the assertion of Proposition 4.9 holds in this case. In particular, (36) still holds for $t \le t_3$ if we consider $s \le T_-$ (instead of $s \le T'$), and this is sufficient for Proposition 4.9. We omit the details.

Now, on $\mathcal{A}'_{t_3}$ we have with (22)

$$\begin{aligned}
U_{t_3}(T_-) &\geq \overline{U}_{t_3}(T_-)\left(1-\frac{1}{\ln^3 n}\right)^{t_3+1}\\
&\geq (n-1)\left(1-\frac{1}{n}\right)^{(T_- -t_3-1)I(t_3)+\sum_{r=0}^{t_3-1} I(r)}\left(1-\frac{10\ln^2 n}{n^2}\right)^{I(t_3)}\left(1-\frac{1}{\ln^3 n}\right)^{t_3+1}\\
&= n\left(1-\frac{1}{n}\right)^{(T_- -t_3-1)I(t_3)+\sum_{r=0}^{t_3-1} I(r)}(1-o(1))\,.
\end{aligned}$$

But $\sum_{r=0}^{t_3-1} I(r) \leq \sum_{r=0}^{t_3-1} 2^r \leq \frac{8n}{\omega}$ and $(1-\frac{1}{n})^{\frac{8n}{\omega}} = 1-o(1)$. Therefore

$$U_{t_3}(T_-) \geq n\left(1-\frac{1}{n}\right)^{(T_- -t_3)I(t_3)}(1-o(1)) = ne^{-(T_- -t_3)\frac{I(t_3)}{n}}(1-o(1))\,. \tag{65}$$

Now, let us fix a certain realization of $\mathcal{U}_{t_3}(T_-)$ such that $\mathcal{A}'_{t_3}$ is fulfilled and consider the following "imaginary" setting: after step t_3, every vertex from $\mathcal{U}(t_3)$ chooses a random segment of length $T_- - t_3$ of its list. Let $\tilde{\mathcal{U}}_{T'}(T_-)$ be the set of vertices from $\mathcal{U}_{t_3}(T_-)$ which are in none of these segments, and let $\tilde{U}_{T'}(T_-) := |\tilde{\mathcal{U}}_{T'}(T_-)|$.

The random variable $\tilde{U}_{T'}(T_-)$ is stochastically smaller than $U_{T'}(T_-)$ because, in the original setting, $U_{T'}(T_-)$ is determined by the vertices informed during steps t_3+1 up to T' (there are at most $U(t_3)$ of them) and for each of them only a random segment of length less than $T_- - t_3$ is taken into account. So it suffices to show that $\tilde{U}_{T'}(T_-) > 0$ with probability $1-o(1)$.

Now we calculate $\mathrm{E}(\tilde{U}_{T'}(T_-))$:

$$\begin{aligned}
\mathrm{E}(\tilde{U}_{T'}(T_-)) &= U_{t_3}(T_-)\left(1-\frac{T_- - t_3}{n}\right)^{U(t_3)}\\
&= U_{t_3}(T_-)e^{-(T_- -t_3)\frac{U(t_3)}{n}}(1-o(1))\,.
\end{aligned} \tag{66}$$

It is easy to see that $\mathrm{E}(\tilde{U}_{T'}(T_-)) \to \infty$ as $n \to \infty$. In particular, we have

$$\begin{aligned}
\mathrm{E}(\tilde{U}_{T'}(T_-)) &\overset{(66)}{=} U_{t_3}(T_-)e^{-(T_- - t_3)\frac{U(t_3)}{n}}\,(1-o(1)) \\
&\overset{(65)}{\geq} ne^{-(T_- - t_3)\frac{I(t_3)}{n}}e^{-(T_- - t_3)\frac{U(t_3)}{n}}\,(1-o(1)) \\
&= ne^{-(T_- - t_3)}\,(1-o(1)) \\
&= ne^{-\omega^2 - \ln n + 4\ln\ln n}\,(1-o(1)) \\
&= e^{-\omega^2}\ln^4 n\,(1-o(1))\,. \qquad (67)
\end{aligned}$$

We will show the following lemma.

4.14 Lemma. *With probability* $1-o(1)$,

$$\tilde{U}_{T'}(T_-) > 0.$$

Therefore we have the following corollary.

4.15 Corollary.

$$\Pr(U_{T'}(T_-) > 0 \;\mid\; \mathcal{A}'_{t_3}) = 1-o(1).$$

Since $\mathcal{A}'_{t_3}$ and $\mathcal{A}_{T'}$ both occur with probability $1-o(1)$, repeated application of Bayes' rule may show that also

$$\Pr(U_{T'}(T_-) > 0 \;\mid\; \mathcal{A}_{T'}) = 1-o(1).$$

This concludes the proof of the lower bound. So the only work left now is the proof of Lemma 4.14. □

Proof of Lemma 4.14. Chebyshev's inequality yields

$$\begin{aligned}
\Pr(\tilde{U}_{T'}(T_-) = 0) &\leq \Pr\left(\left|\tilde{U}_{T'}(T_-) - \mathrm{E}(\tilde{U}_{T'}(T_-))\right| \geq \mathrm{E}(\tilde{U}_{T'}(T_-))\right) \\
&\leq \frac{\mathrm{Var}(\tilde{U}_{T'}(T_-))}{\mathrm{E}^2(\tilde{U}_{T'}(T_-))}.
\end{aligned}$$

To show that this ratio is $o(1)$, it suffices to show that

$$\mathrm{E}\left(\tilde{U}^2_{T'}(T_-)\right) \leq (1+o(1))\,\mathrm{E}^2\left(\tilde{U}_{T'}(T_-)\right). \qquad (68)$$

As

$$\mathrm{E}\left(\tilde{U}^2_{T'}(T_-)\right) = \sum_{u,v\in\mathcal{U}_{t_3}(T_-)} \Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-)), \qquad (69)$$

we will estimate $\mathrm{E}\big(\tilde{U}^2_{T'}(T_-)\big)$ by estimating first $\Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-))$ for any given $u,v \in \mathcal{U}_{t_3}(T_-)$.

If the distance of u,v in $\ell(x)$ was larger than $T_- - t_3$ for all x, then $\Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-))$ would be bounded from above by the square of an expression similar to that in (54), yielding (68). However, this is not the case for all pairs u,v. Nonetheless, we can show that it is true for most of the pairs as follows.

4.16 Proposition. *Let $\mathcal{B}$ be the set of distinct unordered pairs $\{u,v\} \in \binom{\mathcal{U}_{t_3}(T_-)}{2}$ for which there are no more than $U(t_3) - U(t_3)/\ln^2 n$ vertices $x \in \mathcal{U}(t_3)$ such that v and u are at distance at least T_- in $\ell(x)$. Then*

$$|\mathcal{B}| \le \frac{2T_- \ln^2 n}{U_{t_3}(T_-)} \binom{U_{t_3}(T_-)}{2}.$$

Proof. We write $|\mathcal{B}| = \lambda\binom{U_{t_3}(T_-)}{2}$ and count the pairs $(x,\{u,v\}) \in \mathcal{U}(t_3) \times \binom{\mathcal{U}_{t_3}(T_-)}{2}$, which are such that v and u are at distance at least T_- in $\ell(x)$. Then the number of such pairs is at least $U(t_3)U_{t_3}(T_-)(U_{t_3}(T_-) - 2T_- + 2)/2$. Let us also express this number in terms of $|\mathcal{B}|$. Then this is at most $\lambda\binom{U_{t_3}(T_-)}{2}(U(t_3) - U(t_3)/\ln^2 n) + (1-\lambda)\binom{U_{t_3}(T_-)}{2}U(t_3)$. In other words,

$$\begin{aligned} U(t_3)U_{t_3}(T_-)&\frac{U_{t_3}(T_-) - 2T_- + 2}{2} \\ &\le \lambda\binom{U_{t_3}(T_-)}{2}U(t_3)\left(1 - \frac{1}{\ln^2 n}\right) + (1-\lambda)\binom{U_{t_3}(T_-)}{2}U(t_3) \end{aligned}$$

or

$$\begin{aligned} U(t_3)&\frac{U_{t_3}(T_-)^2}{2}\left(1 - \frac{2T_-}{U_{t_3}(T_-)}\right) \\ &\le \lambda\binom{U_{t_3}(T_-)}{2}U(t_3)\left(1 - \frac{1}{\ln^2 n}\right) + (1-\lambda)\binom{U_{t_3}(T_-)}{2}U(t_3). \end{aligned}$$

Dividing both sides by $\binom{U_{t_3}(T_-)}{2}$, we obtain

$$U(t_3)\frac{U_{t_3}(T_-)}{U_{t_3}(T_-) - 1}\left(1 - \frac{2T_-}{U_{t_3}(T_-)}\right) \le \lambda U(t_3)\left(1 - \frac{1}{\ln^2 n}\right) + (1-\lambda)U(t_3),$$

and so

$$U(t_3)\left(1 - \frac{2T_-}{U_{t_3}(T_-)}\right) \le \lambda U(t_3)\left(1 - \frac{1}{\ln^2 n}\right) + (1-\lambda)U(t_3).$$

Now we divide both sides by $U(t_3)$ and obtain

$$1 - \frac{2T_-}{U_{t_3}(T_-)} \leq \lambda \left(1 - \frac{1}{\ln^2 n}\right) + (1-\lambda) = 1 - \frac{\lambda}{\ln^2 n},$$

which yields

$$\lambda \leq \frac{2T_- \ln^2 n}{U_{t_3}(T_-)}.$$

□

Now we write the sum in (69) as follows:

$$\mathrm{E}\left(\tilde{U}^2_{T'}(T_-)\right) = 2 \sum_{\{u,v\}\in\mathcal{B}} \Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-)) + 2 \sum_{\{u,v\}\in\overline{\mathcal{B}}} \Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-)) + \sum_{v\in\mathcal{U}_{t_3}(T_-)} \Pr(v \in \tilde{\mathcal{U}}_{T'}(T_-)). \tag{70}$$

We will treat the three sums separately. The third sum is

$$\sum_{v\in\mathcal{U}_{t_3}(T_-)} \Pr(v \in \tilde{\mathcal{U}}_{T'}(T_-)) = \mathrm{E}(\tilde{U}_{T'}(T_-)) = o(\mathrm{E}^2(\tilde{U}_{T'}(T_-))), \tag{71}$$

since $\mathrm{E}(\tilde{U}_{T'}(T_-)) \to \infty$.

To bound the first sum note that if $u, v \in \tilde{\mathcal{U}}_{T'}(T_-)$, then for all $x \in U(t_3)$ the random position that x chooses must be more than $T_- - t_3$ positions before the position of v. In other words,

$$\Pr(u,v \in \tilde{\mathcal{U}}_{T'}(T_-)) \leq \Pr(v \in \tilde{\mathcal{U}}_{T'}(T_-)) \leq \left(1 - \frac{T_- - t_3}{n}\right)^{U(t_3)}.$$

So we have

$$\begin{aligned}
&\sum_{\{u,v\}\in\mathcal{B}} \Pr(u,v\in\tilde{\mathcal{U}}_{T'}(T_-)) \\
&\qquad\leq \quad |\mathcal{B}|\left(1-\frac{T_- - t_3}{n}\right)^{U(t_3)} \\
&\qquad\overset{\text{Proposition 4.16}}{\leq} \quad \frac{2T_-\ln^2 n}{U_{t_3}(T_-)}\binom{U_{t_3}(T_-)}{2}\left(1-\frac{T_- - t_3}{n}\right)^{U(t_3)} \\
&\qquad\leq \quad \frac{T_-\ln^2 n}{U_{t_3}(T_-)}U_{t_3}(T_-)^2\left(1-\frac{T_- - t_3}{n}\right)^{U(t_3)} \\
&\qquad\overset{(66)}{=} \quad \frac{T_-\ln^2 n}{U_{t_3}(T_-)}\left(1-\frac{T_- - t_3}{n}\right)^{-U(t_3)}\mathrm{E}^2(\tilde{U}_{T'}(T_-)).
\end{aligned}$$

Now, we use (65) to bound $U_{t_3}(T_-)$ from below, thus obtaining for n large enough

$$\begin{aligned}
&\sum_{\{u,v\}\in\mathcal{B}} \Pr(u,v\in\tilde{\mathcal{U}}_{T'}(T_-)) \\
&\qquad\overset{(65)}{\leq} \quad \frac{T_-\ln^2 n}{n}e^{(T_- - t_3)\frac{I(t_3)}{n}}\left(1-\frac{T_- - t_3}{n}\right)^{-U(t_3)}(1+o(1))\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)) \\
&\qquad\leq \quad \frac{T_-\ln^2 n}{n}e^{(T_- - t_3)\frac{I(t_3)}{n}+(T_- - t_3)\frac{U(t_3)}{n}}(1+o(1))\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)) \\
&\qquad\overset{U(t_3)=n-I(t_3)}{=} \quad \frac{T_-\ln^2 n}{n}e^{T_- - t_3}(1+o(1))\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)) \\
&\qquad\overset{T_-<3\ln n}{\leq} \quad \frac{3\ln^3 n}{n}e^{\omega^2+\ln n-4\ln\ln n}(1+o(1))\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)) \\
&\qquad\leq \quad \frac{4e^{\omega^2}\ln^3 n}{\ln^4 n}\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)) = o(\mathrm{E}^2(\tilde{U}_{T'}(T_-))). \qquad (72)
\end{aligned}$$

Finally, recall that for each pair $\{u,v\}\in\overline{\mathcal{B}}$ there are at least $U(t_3) - U(t_3)/\ln^2 n$ vertices $x\in\mathcal{U}(t_3)$ such that v and u are at distance at least T_- in $\ell(x)$. For each such pair $\{u,v\}$ and for each such $x\in\mathcal{U}(t_3)$, x will not inform v and u within the first $T_- - t_3$ steps if its random choice is more than $T_- - t_3$ places before the position of u and v in $\ell(x)$. This excludes $2(T_- - t_3)$

positions on $\ell(x)$. Thus each summand in the second sum is

$$\begin{aligned}&\Pr(u,v\in\tilde{\mathcal{U}}_{T'}(T_-))\\&\quad\le\left(1-\frac{2(T_--t_3)}{n}\right)^{U(t_3)\left(1-\frac{1}{\ln^2 n}\right)}=(1+o(1))\left(1-\frac{2(T_--t_3)}{n}\right)^{U(t_3)}.\end{aligned}$$

So we obtain

$$\begin{aligned}2(1+o(1)) \qquad & \sum_{\{u,v\}\in\overline{\mathcal{B}}}\Pr(u,v\in\tilde{\mathcal{U}}_{T'}(T_-))\\&= 2(1+o(1))\binom{U_{t_3}(T_-)}{2}\left(1-\frac{2(T_--t_3)}{n}\right)^{U(t_3)}\\&\le (1+o(1))U_{t_3}(T_-)^2\left(1-\frac{T_--t_3}{n}\right)^{2U(t_3)} \qquad (73)\\&= (1+o(1))\,\mathrm{E}^2(\tilde{U}_{T'}(T_-)).\end{aligned}$$

(74)

We combine equations (72), (71), and (74) to bound $\mathrm{E}(\tilde{U}_{T'}(T_-))$ as in (70), thus deducing (68). □

□

Indication of source. The content of this section has been previously published in [FH09b]. An extended abstract has been published in [FH09a].

4.3 Robustness

In this section we investigate the *quasirandom rumor spreading model with transmission success probability p*. Additionally to the procedure of quasirandom rumor spreading described above, we assume that each message reaches its target only with a certain probability $p \in]0,1]$ independently for all transmissions. Note that we do not assume that the sender is notified of a transmission failure. Our goal is to show that the rumor is disseminated in the quasirandom model at least as quickly as it is in the random model, and so we will focus on determining the following upper bound on the broadcast time.

4.17 Theorem. *For every $\varepsilon > 0$ and $p \in]0,1]$, the number of rounds we need to inform all vertices of the complete graph on n vertices using the quasirandom rumor spreading model with transmission success probability p is at most*

$$(1+\varepsilon)\left(\log_{1+p} n + \tfrac{1}{p}\ln n\right)$$

with probability at least $1 - n^{-p\varepsilon/40}$.

Unfortunately, since the rumor spreading process is saturated with many dependencies, determining the runtime for the the quasirandom model is not straightforward. As in [DFS08], we try to overcome this difficulty by suitably simplifying the random experiment, in particular, by assuming that certain vertices stop informing (*ignoring*), and that other vertices do not immediately start their own informing process after becoming informed (*delaying*). Delaying turns out to be useful as it gives us some influence on when a vertex uses its one random choice. Nodes that have been informed but have not yet begun informing new nodes play an important role in our analysis. We call them *newly informed* vertices.

To obtain bounds that are precise up to the leading constant, however, we have to be careful that our delaying and ignoring techniques do not slow down the rumor spreading process too much. For this reason, we partition the set of rounds that are necessary to inform all the nodes in the graph into two different types of phases. For both types of phases, the set of nodes that are initially active is the set of newly informed nodes.

Lazy phases were also used in the time analysis of [DFS08]. Only nodes that are considered active at the beginning of the phase are considered active for the remainder of the phase. Nodes that are contacted during the phase, although they are still considered to be informed, remain inactive, and are therefore unable to spread the rumor themselves for the continuation of the phase.

Since lazy phases neglect the rumor spreading potential of a significant portion of the nodes, we also need *busy phases*. Here, all nodes informed during the busy phase are active for the remainder of the phase. In other words, nodes newly informed during the busy phase have the ability to spread the rumor in each subsequent round until the termination of the phase. By choosing the lengths of the busy phases suitably, we balance the difficulties with the inherent dependencies and the losses due to ignoring informed vertices at the end of each phase.

As a result of implementing phases in which vertices that can spread the rumor in the original model are now inactive, we are only delaying the point in time at which all the vertices are informed. Therefore, the upper bound for the quasirandom model with lazy and busy phases holds as an upper bound for the original quasirandom model.

We will split the rumor spreading process into lazy and busy phases in the following way. We start with two lazy phases of $\frac{1}{2}\varepsilon \ln n$ rounds each. The main purpose of these two phases, which are easy to analyze, is to inform a set of vertices that is sufficiently large enough to maximize the effectiveness of the subsequent busy phases. We then perform a logarithmic number of busy phases, each composed of a constant number of rounds. This process results in a constant fraction of informed nodes, and we only need two more lazy phases to render the entire network informed.

4.3.1 The First Lazy Phase

The first lazy phase lasts for $\frac{1}{2}\varepsilon \ln n$ rounds. Our goal is to prove the following.

4.18 Lemma. *Let $\varepsilon > 0$. After one lazy phase of length $\frac{1}{2}\varepsilon \ln n$, at least $\frac{1}{3}p\varepsilon \ln n$ nodes are newly informed with probability at least $1 - n^{-p\varepsilon/36}$.*

Proof. Let $t_1 := \frac{1}{2}\varepsilon \ln n$. At time $t = 0$ one node, v_0, is informed. We perform a lazy phase of length t_1. This means that v_0 contacts each of the first t_1 nodes from its list with probability p. Therefore,

$$\mathrm{E}(N(t_1)) = \tfrac{1}{2}p\varepsilon \ln n.$$

Using Chernoff bounds we see that

$$\begin{aligned}
\Pr\left(N(t_1) < \tfrac{1}{3}p\varepsilon \ln n\right) &= \Pr\left(N(t_1) < \left(1 - \tfrac{1}{3}\right)\mathrm{E}\left(N(t_1)\right)\right) \\
&\leq \exp\left(-\mathrm{E}(N(t_1))/18\right) \\
&= \exp\left(-p\varepsilon \ln n/36\right) \\
&= n^{-p\varepsilon/36}.
\end{aligned}$$

□

4.3.2 The Second Lazy Phase

The second lazy phase begins at time $t_1 + 1$ and terminates after $\frac{1}{2}\varepsilon \ln n$ rounds. Our goal is to prove the following.

4.19 Lemma. *Let $\varepsilon > 0$. If, at some point t_1 in our model, we have $\frac{1}{3}p\varepsilon \ln n \leq N(t_1) \leq \frac{1}{2}\varepsilon \ln n$ and $I(t_1) \leq \frac{1}{2}\varepsilon \ln n + 1$, then after one lazy phase of length $\frac{1}{2}\varepsilon \ln n$, at least $\left(\frac{1}{3}p\varepsilon \ln n\right)^2$ nodes are newly informed with probability at least $1 - n^{-\gamma}$ for any $\gamma \in [0, 1[$.*

Proof. Let $t_1 \in \mathbb{N}$ be such that $\frac{1}{3}p\varepsilon \ln n \leq N(t_1) \leq \frac{1}{2}\varepsilon \ln n$ and $I(t_1) \leq \frac{1}{2}\varepsilon \ln n + 1$ and let $t_2 := t_1 + \frac{1}{2}\varepsilon \ln n$. Enumerate the nodes of $\mathcal{N}(t_1)$ from 1 to $N(t_1)$, and impose an artificial ordering on the set so that each node i calls $\frac{1}{2}\varepsilon \ln n$ of its neighbors, determined from its cyclic list and its initial random decision, before node $i+1$ attempts any contact. For each $i \in \{1, \ldots, N(t_1)\}$, let X_i be the indicator random variable of the following event: Vertex i informs none of the vertices which were already informed by a vertex from 1 to $i - 1$, nor a vertex from $\mathcal{I}(t_1)$. In other words, if $X_i = 1$ for all $i \in \{1, \ldots, N(t_1)\}$, then $N(t_2)$ is equal to the number of contacts made during this phase.

When vertex i first attempts contact, at most $|\mathcal{I}(t_1) \setminus \{i\}| + \frac{1}{2}(i-1)\varepsilon \ln n \leq \frac{1}{2}i\varepsilon \ln n$ other vertices are already informed. Therefore,

$$\Pr(X_i = 0) \leq \frac{\left(\frac{1}{2}\varepsilon \ln n\right)\left(\frac{1}{2}i\varepsilon \ln n\right)}{n-1} \leq \frac{\left(\frac{1}{2}\varepsilon \ln n\right)^3}{n-1}. \tag{75}$$

Using a simple union bound, we conclude that

$$\begin{aligned}\Pr(\forall i \in \{1, \ldots, N(t_1)\} : X_i = 1) &= 1 - \Pr(\exists i \in \{1, \ldots, N(t_1)\} : X_i = 0) \\ &\geq 1 - \sum_{i=1}^{N(t_1)} \Pr(X_i = 0) \\ &\geq 1 - \frac{\left(\frac{1}{2}\varepsilon \ln n\right)^4}{n-1}.\end{aligned}$$

Now that we have shown that the chances of contacting an already informed vertex in this phase are sufficiently small, all that is left to do is to determine how many contacts are made during the phase.

Every node in $\mathcal{N}(t_1)$ attempts to contact $\frac{1}{2}\varepsilon \ln n$ nodes, so there are $\frac{1}{2}\varepsilon \ln n N(t_1)$ possible contacts made during the phase. Each of these is independently successful with probability p. Let Y be the random variable denoting the number of contacts that are actually made during the phase. Then we have

$$\mathrm{E}(Y) = \tfrac{1}{2}p\varepsilon \ln n N(t_1) \geq \left(\tfrac{1}{2}p\varepsilon \ln n\right)\left(\tfrac{1}{3}p\varepsilon \ln n\right). \tag{76}$$

Using Chernoff bounds, we see that

$$\begin{aligned}\Pr\left(Y < \left(\tfrac{1}{3}p\varepsilon \ln n\right)^2\right) &\leq \Pr\left(Y < \left(1 - \tfrac{1}{3}\right)\mathrm{E}(Y)\right)\\ &\leq e^{-\mathrm{E}(Y)/18}\\ &\leq e^{-(p\varepsilon \ln n)^2/108}\\ &= n^{-p^2\varepsilon^2(\ln n)/108}.\end{aligned}$$

Therefore, at least $\left(\frac{1}{3}p\varepsilon \ln n\right)^2$ vertices are informed during this phase with probability at least $1 - \frac{\left(\frac{1}{2}\varepsilon \ln n\right)^4}{n-1} - n^{-p^2\varepsilon^2(\ln n)/108} \geq 1 - n^{-\gamma}$ for any fixed $\gamma \in [0, 1[$. □

4.3.3 The Busy Phases

A sufficient number of nodes are informed of the rumor in the first lazy phase, and so we are ready to commence the set of busy phases. As we have mentioned earlier, the idea of these phases is that nodes informed during each busy phase are able to spread the rumor during subsequent rounds of this phase. Because of the dependencies, these phases require a more refined analysis. Our goal is to inform a constant fraction of the nodes in the network by the time we complete this sequence of phases.

The Analysis of a Single Busy Phase

In order to determine the cumulative effect of the busy phases, we must first analyze the impact of a single busy phase composed of k rounds starting after time-step t. The theorem we present below is the heart of the precise analysis of the quasirandom model. The idea of the proof is to investigate the part of the process originating from each single node in $\mathcal{N}(t)$. A single such process can be analyzed with moderate difficulty. Unfortunately, there may

be "conflicts" among these partial processes, that is, several of these partial processes may inform the same node, possibly at different times. However, we show that only few of these conflicts occur. By completely ignoring all parts that are contained in a conflict, we manage to analyze the busy phase.

Let $\varepsilon' > 0$, $k \in \mathbb{N}$, $p \in]0, 1]$ and $\zeta' \leq \frac{2^{-k}}{k}(2e)^{-\frac{2^{k-1}}{p^3(1+p)^{k-3}}-k-1}$. We will prove the following statement.

4.20 Theorem. *Let $t \in \mathbb{N}$ such that in our model at point t we have $N(t) \geq (p\varepsilon' \ln n)^2$ and $I(t) \leq \zeta' n$. For any constant $c > 0$, if we perform a busy phase of length k, then at the conclusion of this busy phase, the number of newly informed vertices satisfies with probability at least $1 - n^{-c}$ the inequality*

$$N(t + k) \geq p(1 + p)^{k-2} N(t).$$

Proof. Let $\zeta := 2^k \zeta'$. Let t be such that we have $N(t) \geq (p\varepsilon' \ln n)^2$ and $I(t) \leq \zeta' n$. Note that

$$I(t + k) \leq \zeta n.$$

Enumerate the nodes of $\mathcal{N}(t)$ from 1 to $N(t)$. For each $i \in \{1, \ldots, N(t)\}$, we define the set of *descendants of* i, denoted $\mathcal{D}_i$, such that for any node $v \in V$,

(i) If v is contacted within rounds $t + 1, \ldots, t + k$ by i, then $v \in \mathcal{D}_i$, and

(ii) If v is contacted within rounds $t + 1, \ldots, t + k$ by u and $u \in \mathcal{D}_i$, then $v \in \mathcal{D}_i$.

We say that set $\mathcal{D}_i$ is *conflict-free* if the following two conditions hold for each node $v \in \mathcal{D}_i$.

(i) There exists exactly one node $u \in \mathcal{D}_i \cup \{i\}$ that attempts to contact v within rounds $t + 1, \ldots, t + k$.

(ii) No node $w \in \mathcal{I}(t + k) \setminus (\mathcal{D}_i \cup \{i\})$ contacts v before round $t + k + 1$. Note that this implies that the node u specified in condition 1 informs v of the rumor if it makes contact.

Otherwise, we call $\mathcal{D}_i$ *conflicting*.

We will now determine the likelihood that for all $i \in \{1, \ldots, N(t)\}$, the set $\mathcal{D}_i$ is conflict-free.

The first condition of conflict-freeness demands that for each node $v \in \mathcal{D}_i$ there exists exactly one node $u \in \mathcal{D}_i \cup \{i\}$ that attempts to contact v within rounds $t+1, \ldots, t+k$. To bound the probability that this condition fails, we impose an ordering on the random decisions of the vertices in $\mathcal{D}_i \cup \{i\}$. For every such decision d, the probability that d creates a conflict with any previous decision, i.e., that the node in question attempts to contact a node that is targeted by a different node in $\mathcal{D}_i \cup \{i\}$, is bounded from above by $\frac{k}{n-1}2^k$. So the probability that among all decisions a conflict is created is bounded from above by $\frac{k}{n-1}(2^k)^2$.

The probability that the second condition of conflict-freeness fails is

$$\Pr\left(\mathcal{D}_i \cap \left(\mathcal{D}_1 \cup \cdots \cup \mathcal{D}_{i-1} \cup \mathcal{D}_{i+1} \cup \cdots \cup \mathcal{D}_{N(t)} \cup \mathcal{I}(t)\right) \neq \emptyset\right).$$

Let d be a random decision of a vertex in $\mathcal{D}_i \cup \{i\}$.

For all outcomes of random decisions up to round $t+k$ other than those of vertices in $\mathcal{D}_i \cup \{i\}$, the probability that d creates a conflict with any such decision can be *uniformly* bounded from above by $\left(\zeta n - 2^k\right)\frac{k}{n-1}$. So

$$\begin{aligned}
&\Pr\left(\mathcal{D}_i \cap \left(\mathcal{D}_1 \cup \cdots \cup \mathcal{D}_{i-1} \cup \mathcal{D}_{i+1} \cup \cdots \cup \mathcal{D}_{N(t)} \cup \mathcal{I}(t)\right) \neq \emptyset\right) \\
&\quad \leq |\mathcal{D}_i \cup \{i\}| \cdot \left(\zeta n - 2^k\right)\frac{k}{n-1} \\
&\quad \leq 2^k\left(\zeta n - 2^k\right)\frac{k}{n-1}.
\end{aligned}$$

Using a union bound, the probability that $\mathcal{D}_i$ is conflicting is bounded from above by $\frac{k}{n-1}(2^k)^2 + 2^k\left(\zeta n - 2^k\right)\frac{k}{n-1} = 2^k\zeta n\frac{k}{n-1} \leq 2^{k+1}\zeta k$.

Now we want to bound the number of conflicting sets of descendants from above. Let $q = \frac{p^3(1+p)^{k-3}}{2^{k-1}}$. Then the probability that there are at least $qN(t)$ conflicting sets of descendants is

$$\begin{aligned}
\sum_{\substack{M \subseteq \mathcal{N}(t) \\ |M| \geq qN(t)}} \Pr\left(\forall i \in M: \quad \mathcal{D}_i \text{ is conflicting}\right) &\leq \sum_{\substack{M \subseteq \mathcal{N}(t) \\ |M| \geq qN(t)}} \prod_{i \in M} \Pr\left(\mathcal{D}_i \text{ is conflicting}\right) \\
&\leq \sum_{\substack{M \subseteq \mathcal{N}(t) \\ |M| \geq qN(t)}} \left(2^{k+1}k\zeta\right)^{|M|} \;\leq\; 2^{N(t)}\left(2^{k+1}k\zeta\right)^{qN(t)} \\
&= \left(2\left(2^{k+1}k\zeta\right)^{\frac{p^3(1+p)^{k-3}}{2^{k-1}}}\right)^{N(t)} \;\leq\; e^{-N(t)} \;\leq\; n^{-c}
\end{aligned}$$

for any $c > 0$.

Let $\mathcal{X}_i \subseteq \mathcal{D}_i$ denote the set of vertices that are descendants of i and are contacted in round $t+k$ and let $X_i := |\mathcal{X}_i|$. We have

$$\begin{aligned} N(t+k) &\geq \sum_{\substack{i=1 \\ \mathcal{D}_i \text{ is conflict-free}}}^{N(t)} X_i \quad = \sum_{\substack{i=1 \\ \mathcal{D}_i \text{ is conflict-free}}}^{N(t)} (X_i | \mathcal{D}_i \text{ is conflict-free}) \\ &= \sum_{i=1}^{N(t)} (X_i | \mathcal{D}_i \text{ is conflict-free}) - \sum_{\substack{i=1 \\ \mathcal{D}_i \text{ is conflicting}}}^{N(t)} (X_i | \mathcal{D}_i \text{ is conflict-free}) . \end{aligned}$$

Since the largest number of descendants of i that can be informed in round $t+k$ is 2^{k-1}, this yields that

$$N(t+k) \geq \sum_{i=1}^{N(t)} (X_i | \mathcal{D}_i \text{ is conflict-free}) - \sum_{\substack{i=1 \\ \mathcal{D}_i \text{ is conflicting}}}^{N(t)} 2^{k-1}.$$

We can say the following about the expectation of X_i for conflict-free sets of descendants.

4.21 Proposition. *We have*

$$\mathrm{E}\,(X_i | \mathcal{D}_i \textit{ is conflict-free}) = p(1+p)^{k-1}.$$

Proof. We prove this proposition by induction on the number of rounds that have occurred. Assume that $\mathcal{D}_i$ is conflict-free. At time $t+1$, the probability that i has successfully contacted a new node is p, and therefore the expected number of nodes informed by i is p. Let $j \in \{1, \ldots, k-1\}$ be such that for all $r \in \{1, \ldots, j\}$ the expected number of descendants of i informed in round $t+r$ is $p(1+p)^{r-1}$.

Then the expected number of descendants of i informed in round $t+j+1$ is

$$p\left(1 + \sum_{r=1}^{j} p(1+p)^{r-1}\right) = p(1+p)^j.$$

□

Let $X := \sum_{i=1}^{N(t)} (X_i | \mathcal{D}_i \text{ is conflict-free})$. Note that the *conditional* random variables $(X_i | \mathcal{D}_i \text{ is conflict-free})$ are independent and bounded by 2^{k-1}.

So we can use Hoeffding bounds (Theorem 1.2) and get

$$\begin{aligned}
&\Pr\left(X < p(1+p+p^2)(1+p)^{k-3}N(t)\right)\\
&\quad= \Pr\left(X < p(1+p)^{k-1}N(t)\left(1-\frac{p}{(1+p)^2}\right)\right)\\
&\quad= \Pr\left(X < \mathrm{E}(X)\left(1-\frac{p}{(1+p)^2}\right)\right)\\
&\quad\le \exp\left(-\frac{2p^2\,\mathrm{E}(X)^2}{2^{2k-2}(1+p)^4N(t)}\right) \;=\; \exp\left(-\frac{p^4(1+p)^{2k-6}N(t)}{2^{2k-3}}\right)\\
&\quad\le n^{-c}
\end{aligned}$$

for any $c > 0$.

This means that we have

$$X \ge p(1+p+p^2)(1+p)^{k-3}N(t)$$

with probability at least $1 - n^{-c}$.

So with probability at least $1 - n^{-c}$,

$$\begin{aligned}
N(t+k) &\ge X - \sum_{\substack{i=1\\ \mathcal{D}_i \text{ is conflicting}}}^{N(t)} 2^{k-1}\\
&\ge p(1+p+p^2)(1+p)^{k-3}N(t) - p^3(1+p)^{k-3}N(t)\\
&= p(1+p)^{k-2}N(t).
\end{aligned}$$

This proves Theorem 4.20. □

We also have to ensure that after the performance of one busy phase, a big enough fraction of the informed vertices is newly informed, since only the newly informed vertices are active in the next phase. We show that this holds in the proof of the following corollary.

4.22 Corollary. *Define t such that at point t we have $N(t) \ge (p\varepsilon' \ln n)^2$ and $I(t) \le \min\left\{\zeta' n, \frac{2^k-1}{p(1+p)^{k-2}-1}N(t)\right\}$. If we perform a busy phase of length k, then for any $c > 0$ at the conclusion of this busy phase we have*

$$I(t+k) \le \frac{2^k-1}{p(1+p)^{k-2}-1}N(t+k)$$

with probability at least $1 - n^{-c}$.

Proof. If we perform a busy phase of length k, by Theorem 4.20 we get

$$\begin{aligned} I(t+k) &= I(t) + \sum_{i=1}^{k} N(t+i) \\ &\leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t) + \sum_{i=1}^{k} 2^{i-1} N(t) \\ &= \frac{(2^k - 1)p(1+p)^{k-2}}{p(1+p)^{k-2} - 1} N(t) \\ &\leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t+k) \end{aligned}$$

with probability at least $1 - n^{-c}$ for any $c > 0$. □

Assembling of the Busy Phases

Now that we have analyzed a single busy phase, we can put these phases together to obtain a constant fraction of informed nodes. Let $\varepsilon > 0$, $p \in]0, 1]$ and

$$k := \frac{1+\varepsilon}{\varepsilon} \left(\log_{1+p} \frac{1}{p} + 2 \right).$$

As in the previous subsection, let $\zeta \leq \frac{1}{k}(2e)^{-\frac{2^{k-1}}{p^3(1+p)^{k-3}} - k - 1}$, and $\zeta' := 2^{-k}\zeta$. We show the following.

4.23 Theorem. *Let $\varepsilon' > 0$. Let t_2 be such that in our model at point t_2 we have $N(t_2) \geq (p\varepsilon' \ln n)^2$ and $I(t_2) \leq \min\left\{\zeta' n, \frac{2^k-1}{p(1+p)^{k-2}-1} N(t_2)\right\}$.*

Let ℓ denote the smallest integer such that if we perform ℓ busy phases with k rounds we have $I(t_2 + \ell k) \geq \zeta' n$.

Then

$$\ell \leq \frac{(1+\varepsilon)\log_{1+p} n}{k}$$

and

$$I(t_2 + \ell k) \leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t_2 + \ell k)$$

hold with probability at least $1 - n^{-c}$ for any $c > 0$.

Proof. Since $p(1+p)^{k-2} \geq 1$ and $N(t_2) \geq (p\varepsilon' \ln n)^2$, we can use Theorem 4.20 inductively and obtain that for all $s \in \{1, \dots, \ell\}$ we have

$$N(t_2 + sk) \geq p(1+p)^{k-2} N(t_2 + (s-1)k) \geq \cdots \geq \left(p(1+p)^{k-2}\right)^s N(t_2)$$

with probability at least $1 - sn^{-c}$ for any $c > 0$, and therefore

$$I(t_2 + \ell k) \geq \left(p(1+p)^{k-2}\right)^\ell,$$

with probability at least $1 - \ell n^{-c}$ for any $c > 0$.

Since $n \geq \zeta n \geq I(t_2 + \ell k)$, we have

$$\begin{aligned} \log_{1+p} n &\geq \log_{1+p} \zeta n \\ &> \log_{1+p} \left(p(1+p)^{k-2}\right)^\ell \\ &= \ell(k - 2 + \log_{1+p} p), \end{aligned}$$

which implies that

$$\ell \leq \frac{\log_{1+p} n}{k - 2 + \log_{1+p} p} = \frac{(1+\varepsilon)\log_{1+p} n}{k}$$

with probability at least $1 - n^{-c}$ for any $c > 0$.

Since $I(t_2) \leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t_2)$, by an inductive application of Corollary 4.22, we get that

$$I(t_2 + \ell k) \leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t_2 + \ell k)$$

holds with probability at least $1 - n^{-c}$ for any $c > 0$. □

4.3.4 Second To Last Phase

Now that we have a small constant fraction of newly informed nodes, a lazy phase of a constant number of rounds suffices to yield a large fraction of newly informed nodes.

4.24 Lemma. *Let $\varepsilon \in]0,1[$ and $k := \frac{1+\varepsilon}{\varepsilon}\left(\log_{1+p}\frac{1}{p}+2\right)$. Let t_3 be such that in our model at round t_3 we have $I(t_3) \le \frac{2^k-1}{p(1+p)^{k-2}-1}N(t_3)$, and that there exist $\zeta,\zeta' \in]0,1[$ such that $\zeta' n \le I(t_3) \le \zeta n$ holds. Let $S := \frac{2^k \ln(1/\zeta)}{p\zeta'}$.*

After one lazy phase of S rounds starting at time t_3, at least $(1-3\zeta)\,n$ nodes will be newly informed with probability $1-e^{-\Omega(n)}$.

Proof. We perform one lazy phase of S rounds starting at time t_3. Let $v_0 \in V \setminus \mathcal{I}(t_3)$. Then

$$\begin{aligned}\Pr(\text{no } v \in \mathcal{N}(t_3) \text{ contacts } v_0 \text{ in this phase}) &= \left(1-\frac{pS}{n-1}\right)^{N(t_3)}\\ &\le \exp\left(-\frac{pSN(t_3)}{n-1}\right) \le \exp\left(-\frac{Sp(p(1+p)^{k-2}-1)I(t_3)}{2^k(n-1)}\right)\\ &\le \exp\left(-\frac{SpI(t_3)}{2^k(n-1)}\right) \le \exp\left(-\frac{Sp\zeta'}{2^k}\right) = \zeta.\end{aligned}$$

We now calculate the expected number of newly informed nodes after S rounds. Let $t_4 := t_3 + S$.

$$\begin{aligned}\mathrm{E}\left(N(t_4)\right) &= |V \setminus \mathcal{I}(t_3)| \cdot \Pr\left(v_0 \in V \setminus \mathcal{I}(t_3) \text{ is informed in this phase}\right)\\ &\ge (n - I(t_3))\,(1-\zeta)\\ &\ge (n-\zeta n)\,(1-\zeta)\\ &\ge (1-2\zeta)n.\end{aligned}$$

We will now use the Hoeffding–Azuma inequality (see Theorem 1.5). Number the nodes of $\mathcal{N}(t_3)$ from 1 to $N(t_3)$. Then for all $i \in \{1,\dots,N(t_3)\}$, define the random variable $\mathcal{X}_i$ as the set of vertices that i contacts in the S lazy rounds. Now we can define the function f such that

$$f(\mathcal{X}_1,\dots,\mathcal{X}_{N(t_3)}) := \left|\bigcup_{i=1}^{N(t_3)} \mathcal{X}_i \setminus \mathcal{I}(t_3)\right| = N(t_4).$$

By this definition, we see that

$$f(x_1,\dots,x_i,\dots,x_{N(t_3)}) - f(x_1,\dots,x_i',\dots,x_{N(t_3)}) \le S.$$

Therefore, we can bound the probability that we inform less than $(1-3\zeta)n$ vertices in this phase as follows.

$$\begin{aligned}\Pr\left(N(t_4) < (1-3\zeta)\,n\right) &= \Pr\left(N(t_4) < (1-2\zeta)n - \zeta n\right)\\ &\le \Pr\left(|N(t_4) - \mathrm{E}(N(t_4))| \ge \zeta n\right) \le 2\exp\left(-\frac{2\zeta^2 n^2}{\sum_{i=1}^{N(t_3)} S^2}\right)\\ &\le 2\exp\left(-\frac{2\zeta^2 n^2}{\zeta n S^2}\right) = e^{-\Omega(n)}.\end{aligned}$$

□

4.3.5 The Final Phase

The last phase of the protocol is again a lazy phase. We now use the large fraction of newly informed nodes from the previous phase to inform the few remaining nodes.

4.25 Lemma. *Let $\varepsilon \in]0,1[$ and $\eta \le \frac{\varepsilon}{4}$. Let t_4 be such that in our model at round t_4 we have $N(t_4) \ge (1-\eta)\,n$. After one lazy phase of $\frac{(3+\varepsilon)}{3p}\ln n$ rounds starting at time t_4, all the nodes will be informed with probability $1 - O(n^{-\varepsilon(1-\varepsilon)/12})$.*

Proof. We will perform one lazy phase of $\frac{(3+\varepsilon)}{3p}\ln n$ rounds starting at time t_4. Let $v_0 \in V \setminus \mathcal{I}(t_4)$. Then

$$\begin{aligned}\Pr(\text{no } v \in \mathcal{N}(t_4) \text{ contacts } v_0 \text{ in this phase}) &= \left(1 - \frac{p(3+\varepsilon)\ln n}{3p(n-1)}\right)^{N(t_4)}\\ &\le \exp\left(\frac{-(3+\varepsilon)\ln n N(t_4)}{3(n-1)}\right) \le \exp\left(\frac{-(3+\varepsilon)(1-\eta)n\ln n}{3(n-1)}\right)\\ &\le \exp\left(-\frac{(3+\varepsilon)(1-\eta)\ln n}{3}\right) \le n^{-(1+\varepsilon(1-\varepsilon)/12)}.\end{aligned}$$

So the probability that all the nodes become informed is

$$\begin{aligned}&\Pr(\forall v \in V \setminus \mathcal{I}(t_4) :\ v \text{ becomes informed})\\ &\quad = 1 - \Pr\left(\exists v \in V \setminus \mathcal{I}(t_4) :\ v \text{ does not get informed}\right)\\ &\quad \ge 1 - \sum_{v \in V \setminus \mathcal{I}(t_4)} \Pr\left(v \text{ does not get informed}\right)\\ &\quad \ge 1 - \eta n n^{-(1+\varepsilon(1-\varepsilon)/12)} = 1 - O(n^{-\varepsilon(1-\varepsilon)/12}).\end{aligned}$$

□

4.3.6 Proof of Theorem 4.17

Let $\varepsilon \in]0,1[$, $p \in]0,1]$ and $k := \frac{1+\varepsilon}{\varepsilon}\left(\log_{1+p}\frac{1}{p}+2\right)$. Furthermore, let $\zeta := \min\left\{\frac{1}{k}(2e)^{-\frac{2^{k-1}}{p^3(1+p)^{k-3}}-k-1}, \frac{\varepsilon}{12}\right\}$ and $\zeta' := 2^{-k}\zeta$.

We start a delayed quasirandom rumor spreading protocol with message success probability p and with one initially informed vertex. We first perform one lazy phase of length $t_1 := \frac{1}{2}\varepsilon \ln n$. By Lemma 4.18 this yields that $N(t_1) \geq \frac{1}{3}p\varepsilon \ln n$ holds with probability at least $1 - n^{-p\varepsilon/36}$. Of course, after one lazy phase of length t_1 we have with probability one $N(t_1) \leq t_1$ and $I(t_1) \leq t_1 + 1$. So we can apply Lemma 4.19 and get $N(t_2) \geq \left(\frac{1}{3}p\varepsilon \ln n\right)^2$, this phase succeeds with probability at least $1 - n^{-\gamma}$ for any $\gamma \in [0,1[$. Furthermore we have with probability one $I(t_2) \leq \left(\frac{1}{2}\varepsilon \ln n\right)^2 + \frac{1}{2}\varepsilon \ln n + 1 \leq \zeta' n$ as well as $I(t_2) = I(t_1) + N(t_2) \leq \frac{2^k-1}{p(1+p)^{k-2}-1}N(t_2)$ for any sufficiently large n. So we can apply Theorem 4.23 with $\varepsilon' := \frac{\varepsilon}{3}$. This gives us an $\ell \leq \frac{(1+\varepsilon)\log_{1+p} n}{k}$ such that if we set $t_3 := t_2 + \ell k$, then for any $c > 0$ we have with probability at least $1 - n^{-c}$

$$\zeta' n \leq I(t_3) \leq \zeta n \qquad \text{and} \qquad I(t_3) \leq \frac{2^k - 1}{p(1+p)^{k-2} - 1} N(t_3).$$

So with probability at least $1 - n^{-c}$ the preconditions of Lemma 4.24 are fulfilled. Therefore, if we set $S := \frac{2^k \ln(1/\zeta)}{p\zeta'}$ and $t_4 := t_3 + S$, we get $N(t_4) \geq (1 - 3\zeta)\, n$ with probability at least $1 - n^{-c}$. We can consequently apply Lemma 4.25 with $\eta := 3\zeta$. We conclude that after $\frac{(3+\varepsilon)}{3p}\ln n$ more rounds all the nodes will be informed with probability $1 - O(n^{-\varepsilon(1-\varepsilon)/12})$.

Overall, we perform at most

$$\tfrac{1}{2}\varepsilon \ln n + (1+\varepsilon)\log_{1+p} n + S + \tfrac{3+\varepsilon}{3p}\ln n \leq (1+\varepsilon)\left(\tfrac{1}{p}\ln n + \log_{1+p} n\right)$$

rounds in our delayed quasirandom rumor spreading protocol with message success probability p.

The overall failure probability is at most

$$n^{-p\varepsilon/36} + n^{-\gamma} + n^{-c} + e^{-\Omega(n)} + O(n^{-\varepsilon(1-\varepsilon)/12}) \leq n^{-p\varepsilon/40}.$$

Indication of source. The content of this section is under submission [DHL], together with the work described in Section 3.2. A short version has been previously published in [DHL09].

Conclusion and Outlook

In Chapter 2 it was shown that the theorem of Beck and Fiala [BF81] generalizes to randomized rounding except for a difference of one in the upper bound for hypergraphs with maximum degree in $\{2, \ldots, 8\}$. Even if this difference is not very important for most applications, it would still be interesting from the theoretic point of view to know if the theorem of Beck and Fiala generalizes exactly.

In Chapters 3 and 4 we provided an analysis of the evolution of rumor spreading under the randomized and the quasirandom models including the aspects of performance, robustness, and density of the underlying graph and a detailed comparison of both models. There are, however, still various interesting open questions for further research in the rumor spreading context. Our investigation of random graphs presented in Section 3.3 was the first precise analysis including leading constants of the runtime of randomized rumor spreading on a non-trivial graph class. Afterwards, Fountoulakis and Panagiotou [FP10] determined the leading constant for random regular graphs. Tight bounds for other graph classes, for example hypercubes or graphs with good expansion properties, are not known.

In Section 4.2, we presented a detailed analysis of the quasirandom rumor spreading model on the complete graph and showed that its evolution is very close to the evolution of the randomized rumor spreading model. Combined with the previous work in [Pit87] about randomized rumor spreading on complete graphs, our work demonstrates that quasirandom rumor spreading is as fast as randomized rumor spreading not only to leading order in n but also including the leading constant and lower order terms. At the same time passing to the quasirandom version of the model greatly reduces the number of random bits needed. Together with [ADHP09], these are the first results presenting bounds on of the runtime of quasirandom rumor spreading including leading constants. Although from the application point of view the upper bound appears more important, it would be interesting to know

whether a corresponding lower bound also holds, that is, whether one can replace the $4 \ln \ln n$ term in Theorem 4.1 by $\omega(n)$. This would then establish the full equivalence between the randomized method and the quasirandom method on the complete graph.

Also for the robustness question investigated in Chapter 3 and Section 4.3, this is the first work precise including leading constants. The Sections 3.2 and 4.3 together ensure that quasirandom rumor spreading on the complete graph is not only as fast as randomized rumor spreading but also as robust. In Section 3.3 the leading constant for faulty randomized rumor spreading on random graphs has been determined. To complete the correspondence between the randomized method and the quasirandom method regarding robustness on random graphs, it would be necessary to determine also the leading constant for faulty quasirandom rumor spreading on random graphs. Furthermore, it would be interesting to conceive of a precise analysis of both faulty randomized rumor spreading and faulty quasirandom rumor spreading also on other graph classes. We believe that for most natural network topologies, the quasirandom variant is as robust as the fully random one.

For the robustness of randomized rumor spreading on general graphs Elsässer and Sauerwald showed in [ES09] that the broadcast time for all graphs in this lossy model is at most a factor of $O(1/q)$ larger than it is in the model without transmission failures as described in Section 3.1. It would be interesting to obtain a similar result for quasirandom rumor spreading. However, we cannot expect the same generality: On a star graph, for example, quasirandom rumor spreading runs in linear time while the runtime of faulty quasirandom rumor spreading is of order $n \log n$.

Another problem would be to consider the number of messages sent instead of the number of rounds performed. Counting the number of rounds answers the question of how long it takes to inform all vertices. Counting the number of messages sent answers the question of how much it costs to inform all vertices, if with every transmission there is associated a fixed cost.

A more general setting for further research would be to apply the concept of quasirandomness to other randomized algorithms or models. As in the two areas of research presented above, the central question to be answered is how much randomness is needed, and how much of the randomness can be discarded, for the efficiency and other desired properties of the random process to be maintained.

Bibliography

[ADH05] A. Alavena, A. Demers, and J. E. Hopcroft. Correctness of a gossip based membership protocol. In *Proceedings of the 24th Symposium on Principles of Distributed Computing (PODC)*, pages 292–301, 2005.

[ADHP09] S. Angelopoulos, B. Doerr, A. Huber, and K. Panagiotou. Tight bounds for quasirandom rumor spreading. *The Electronic Journal of Combinatorics*, 16(#R102), 2009.

[AS08] N. Alon and J. H. Spencer. *The Probabilistic Method*. Interscience Series in Discrete Mathematics and Optimization. Wiley, third edition, 2008.

[Azu67] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Mathematical Journal*, 19(3):357–367, 1967.

[BEF08] P. Berenbrink, R. Elsässer, and T. Friedetzky. Efficient randomised broadcasting in random regular networks with applications in peer-to-peer systems. In *Proceedings of the 27th Symposium on Principles of Distributed Computing (PODC)*, pages 155–164, 2008.

[BF81] J. Beck and T. Fiala. "Integer-making" theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[Bol01] B. Bollobás. *Random Graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, second edition, 2001.

[BW01] S. Botros and S. Waterhouse. Search in JXTA and other distributed networks. In *Proceedings of the 1st IEEE Interna-*

tional Conference on Peer-to-Peer Computing (P2P), pages 30–35, 2001.

[CDFS08] J. Cooper, B. Doerr, T. Friedrich, and J. Spencer. Deterministic random walks on regular trees. In *Proceedings of the 19th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 766–772, 2008.

[CDST07] J. Cooper, B. Doerr, J. Spencer, and G. Tardos. Deterministic random walks on the integers. *European Journal of Combinatorics*, 28(8):2072–2090, 2007.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[CS06] J. Cooper and J. Spencer. Simulating a random walk with constant error. *Combinatorics, Probability & Computing*, 15(6):815–822, 2006.

[DF09] B. Doerr and T. Friedrich. Deterministic random walks on the two-dimensional grid. *Combinatorics, Probability & Computing*, 18(1-2):123–144, 2009.

[DFKS09] 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)*, pages 145–153, 2009.

[DFS08] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 773–781, 2008.

[DFS09] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading: Expanders, push vs. pull, and robustness. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 5555 of *LNCS*, pages 366–377, 2009.

[DGH+87] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for

replicated database maintenance. In *Proceedings of the 6th Symposium on Principles of Distributed Computing (PODC)*, pages 1–12, 1987.

[DGH+88] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. *Operating Systems Review*, 22(1):8–32, 1988.

[DHL] B. Doerr, A. Huber, and A. Levavi. Strong robustness of randomized rumor spreading protocols. Submitted. Available from arXiv:1001.3056.

[DHL09] B. Doerr, A. Huber, and A. Levavi. Strong robustness of randomized rumor spreading protocols. In *Proceedings of the 20th International Symposium on Algorithms and Computation (ISAAC)*, volume 5878 of *LNCS*, pages 812–821, 2009. Short version of [DHL].

[Doe00] B. Doerr. Linear and hereditary discrepancy. *Combinatorics, Probability & Computing*, 9(4):349–354, 2000.

[Doe06] B. Doerr. Generating randomized roundings with cardinality constraints and derandomizations. In *Proceedings of the 23rd Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 3884 of *LNCS*, pages 571–583, 2006.

[Doe07] B. Doerr. Randomly rounding rationals with cardinality constraints and derandomizations. In *Proceedings of the 24th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 4393 of *LNCS*, pages 441–452, 2007.

[EGS08] R. Elsässer, L. Gasieniec, and T. Sauerwald. On radio broadcasting in random geometric graphs. In *Proceedings of the 22nd International Symposium on Distributed Computing (DISC)*, volume 5218 of *LNCS*, pages 212–226, 2008.

[EL75] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets*, pages 609–627, 1975.

[Els06] R. Elsässer. On randomized broadcasting in power law networks. In *Proceedings of the 20th International Symposium on Distributed Computing (DISC)*, volume 4167 of *LNCS*, pages 370–384, 2006.

[ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 5(1-2):17–61, 1960.

[ES07] R. Elsässer and T. Sauerwald. Broadcasting vs. mixing and information dissemination on cayley graphs. In *Proceedings of the 24th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 4393 of *LNCS*, pages 163–174, 2007.

[ES09] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[FG85] A. M. Frieze and G. R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10(1):57–77, 1985.

[FH09a] N. Fountoulakis and A. Huber. Quasirandom broadcasting on the complete graph is as fast as randomized broadcasting. In *Proceedings of the European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB)*, volume 34 of *ENDM*, pages 553–559, 2009. Short version of [FH09b].

[FH09b] N. Fountoulakis and A. Huber. Quasirandom rumor spreading on the complete graph is as fast as randomized rumor spreading. *SIAM Journal on Discrete Mathematics*, 23(4):1964–1991, 2009.

[FHP09] N. Fountoulakis, A. Huber, and K. Panagiotou. The speed of broadcasting in random networks: Density does not matter. In *Proceedings of the 23th International Symposium on Distributed Computing (DISC)*, volume 5805 of *LNCS*, pages 529–530, 2009. Brief announcement. Full version available from arXiv:0904.4851.

[FHP10] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable broadcasting in random networks and the effect of density. In *Proceedings of the 29th Conference on Computer Communications (IEEE INFOCOM)*, pages 1–9, 2010.

[FP10] N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and expanders. In *Proceedings of the 14th International Workshop on Randomization and Computation (RANDOM)*, 2010.

[FPRU90] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.

[Het00] H.W. Hethcore. Mathematics of infectious diseases. *SIAM Review 42*, pages 599–653, 2000.

[HKP+05] J. Hromkovič, R. Klasing, A. Pelc, P. Ružička, and W. Unger. *Dissemination of information in communication networks*. Texts in Theoretical Computer Science. Springer-Verlag, 2005.

[HLM+08] A. E. Holroyd, L. Levine, K. Mészáros, Y. Peres, J. Propp, and D. B. Wilson. Chip-firing and rotor-routing on directed graphs. In V. Sidoravicius and M. E. Vares, editors, *In and Out of Equilibrium II*, volume 60 of *Progress in Probability*, pages 331–364. Birkhäuser, 2008.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[Inc00] Clip2.com Inc. Gnutella: To the bandwidth barrier and beyond. Published online at http://cs-www.cs.yale.edu/homes/arvind/cs425/doc/gnutella.html, 2000.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Interscience Series in Discrete Mathematics and Optimization. Wiley, 2000.

[JPS06] S. Jagannathan, G. Pandurangan, and S. Srinivasan. Query protocols for highly resilient peer-to-peer networks. In *Proceedings*

of the 19th International Conference on Parallel and Distributed Computing Systems (ISCA PDCS), pages 247–252, 2006.

[KAG90] J. P. Kelly, A. A. Assad, and B. L. Golden. The controlled rounding problem: Relaxations and complexity issues. *OR Spektrum*, 12(3):129–138, 1990.

[KDG03] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 482–491, 2003.

[KGA90] J. P. Kelly, B. L. Golden, and A. A. Assad. Using simulated annealing to solve controlled rounding problems. *INFORMS Journal on Computing*, 2(2):174–185, 1990.

[KGAB90] J. P. Kelly, B. L. Golden, A. A. Assad, and E. K. Baker. Controlled rounding of tabular data. *Operations Research*, 38(5):760–772, 1990.

[KLR+87] R. M. Karp, F. T. Leighton, R. L. Rivest, C. D. Thompson, U. V. Vazirani, and V. V. Vazirani. Global wire routing in two-dimensional arrays. *Algorithmica*, 2(1):113–129, 1987.

[KMPS05] V. S. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan. Approximation algorithms for scheduling on multiple machines. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 254–263, 2005.

[KSSV00] R. M. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized Rumor Spreading. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 565–574, 2000.

[LS03] C. Law and K.-Y. Siu. Distributed construction of random expander networks. In *Proceedings of the 22th Conference on Computer Communications (IEEE INFOCOM)*, pages 2133–2143, 2003.

[McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics*, volume 141 of *London Mathematical Society*

Lecture Note Series, pages 148–188. Cambridge University Press, 1989.

[MU05] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[Nie92] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1992.

[NS72] M. S. Nargundkar and W. Saveland. Random-rounding to prevent statistical disclosures. In *Proceedings of the American Statistical Association*, pages 382–385, 1972.

[PDDK96] V. B. Priezzhev, D. Dhar, A. Dhar, and S. Krishnamurthy. Eulerian walkers as a model of self-organized criticality. *Physical Review Letters*, 77(25):5079–5082, 1996.

[Pit87] B. Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[PRU03] G. Pandurangan, P. Raghavan, and E. Upfal. Building low-diameter peer-to-peer networks. *IEEE Journal on Selected Areas in Communications*, 21(6):995–1002, 2003.

[PS97] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff–Hoeffding bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.

[Rag88] P. Raghavan. Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *Journal of Computer and System Sciences*, 37(2):130 – 143, 1988.

[RT87] P. Raghavan and C. D. Tompson. Randomized rounding: A technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7(4):365 – 374, 1987.

[Sau07] T. Sauerwald. On mixing and edge expansion properties in randomized broadcasting. In *Proceedings of the 18th International Symposium on Algorithms and Computation (ISAAC)*, volume 4835 of *LNCS*, pages 196–207, 2007.

[Sri01] A. Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 588–597, 2001.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 81(1):73–205, 1995.

Acknowledgements

First of all, I would like to thank my thesis advisor Benjamin Doerr for introducing me to the world of randomized methods in computer science, for suggesting an interesting research project, and for his helpful advice and continuous encouragement.

I am very grateful to the Max-Planck-Institute for Informatics for the opportunity to prepare my thesis here. Furthermore I would like to thank my working group, the Algorithms and Complexity department, for the inspiring research environment and in particular its head Kurt Mehlhorn for creating a pleasant atmosphere in the group, for the freedom he has always granted us, and for his encouragement.

A great deal of the work reported in this thesis is the outcome of joint research efforts. For their collaboration, and for agreeing that our common results be included in my dissertation, I thank my coworkers Nikolaos Fountoulakis, Christian Klein, Ariel Levavi, and Konstantinos Panagiotou.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, den 28. Juli 2010

Anna Huber