

# Asymptotics for Push on the Complete Graph

Rami Daknama<sup>1</sup>      Konstantinos Panagiotou<sup>1</sup>      Simon Reisser<sup>1</sup>

<sup>1</sup>Ludwig-Maximilians-Universität München

24th March, 2020

## Abstract

We study the popular randomized rumour spreading protocol *push*. Initially, a node in a graph possesses some information, which is then spread in a round based manner. In each round, each informed node chooses uniformly at random one of its neighbours and passes the information to it. The central quantity to investigate is the *runtime*, that is, the number of rounds needed until every node has received the information.

The *push* protocol and variations of it have been studied extensively. Here we study the case where the underlying graph is complete with  $n$  nodes. Even in this most basic setting, specifying the limiting distribution of the runtime as well as determining related quantities, like its expectation, have remained open problems since the protocol was introduced.

In our main result we describe the limiting distribution of the runtime. We show that it does not converge, and that it becomes, after the appropriate normalization, asymptotically periodic both on the  $\log_2 n$  as well as on the  $\ln n$  scale. In particular, the limiting distribution converges only if we restrict ourselves to suitable subsequences of  $\mathbb{N}$ , where simultaneously  $\log_2 n - \lfloor \log_2 n \rfloor \rightarrow x$  and  $\ln n - \lfloor \ln n \rfloor \rightarrow y$  for some fixed  $x, y \in [0, 1)$ . On such subsequences we show that the expected runtime is  $\log_2 n + \ln n + h(x, y) + o(1)$ , where  $h$  is explicitly given and numerically  $|\sup h - \inf h| \approx 2 \cdot 10^{-4}$ .

This “double oscillatory” behaviour has its origin in two key ingredients that were also implicit in previous works: first, an intricate discrete recursive relation that describes how the set of informed nodes grows, and second, a coupon collector problem with batches of size  $n$  that takes the lead when the protocol is almost finished. Rounding in the recursion introduces the periodicity on the  $\log_2 n$  scale – as it is the case in many discrete systems – and rounding in the batched problem is the source of the second periodicity.

## 1 Introduction

We consider the well-known and well-studied rumour spreading protocol *Push*. It has applications in replicated databases [6], multicast [1] and blockchain technology [19]. *Push* operates on graphs and proceeds in rounds as follows. In the beginning, one node has a piece of information. In subsequent rounds each informed node chooses a neighbour independently and uniformly at random and informs it. For a graph  $G = (V, E)$  with  $|V| = n$  and a node  $v \in V$  we denote by  $X(G, v)$  the (random) number of rounds needed to inform all nodes, where at the beginning of the first round only  $v$  knows the information. We call  $X(G, v)$  the *runtime* (on  $G$  with start node  $v$ ). The most basic case, and the one that we study here, is when  $G$  is the complete graph  $K_n$ . Since in that case the initially informed node makes no difference, we will abbreviate  $X(K_n, v) = X_n$  for any starting node  $v$ .

**Related Work** There are several works studying the runtime of *push* on the complete graph. The first paper considering this protocol is by Frieze and Grimmett [12], who showed that with high probability (whp), that is, with probability  $1 - o(1)$  as  $n \rightarrow \infty$ , that

$$X_n = \log_2 n + \ln n + o(\ln n).$$

Moreover, they obtained bounds for (very) large deviations of  $X_n$  from its expectation. In [20], Pittel improved upon the results in [12], in particular, he showed that for any  $f : \mathbb{N} \rightarrow \mathbb{R}^+$  with  $f = \omega(1)$ , whp,

$$|X_n - \log_2 n - \ln n| \leq f(n).$$

The currently most precise result in this context was obtained by Doerr and Künnemann [7], who considered in great detail the distribution of  $X_n$ . They showed that  $X_n$  can be stochastically bounded (from both sides) by coupon collector type problems. This gives a lot of control regarding the distribution of  $X_n$ , and it allowed them to derive, for example, very sharp bounds for tail probabilities. Apart from that, it enabled them to consider related quantities, as for example the expectation of  $X_n$ . Among other results, their bounds on the distribution of  $X_n$  imply that

$$\lfloor \log_2 n \rfloor + \ln n - 1.116 \leq \mathbb{E}[X_n] \leq \lceil \log_2 n \rceil + \ln n + 2.765, \quad (1.1)$$

which pins down the expectation up to a constant additive term. Besides on complete graphs, *push* has been extensively studied on several other graph classes. For example, Erdős-Rényi random graphs [9, 10], random regular graphs and expander graphs [11, 18, 5]. More general bounds that only depend on some graph parameter have also been derived, e.g. the diameter [9], graph conductance [17, 3, 4, 13] and node expansion [4, 22, 15, 14].

**Results** In order to state our main result we need some definitions first. Set

$$g = g^{(1)} : [0, 1] \rightarrow [0, 1], \quad x \mapsto xe^{x-1} \quad \text{and} \quad g^{(i)} : [0, 1] \rightarrow [0, 1], \quad g^{(i)} = g \circ g^{(i-1)}, i \geq 2.$$

As we will see later, the function  $g$  describes, for a wide range of the parameters, the evolution of the number of uninformed nodes; in particular, if at the beginning of some round there are  $xn$  uninformed nodes, then at the end of the same round there will be (roughly)  $g(x)n$  uninformed nodes, and after  $i$  rounds there will be (roughly)  $g^{(i)}(x)n$  uninformed nodes. This fact is not new – at least for bounded  $i$  – and has been observed long ago, see for example [20, Lem. 2]. For  $x \in \mathbb{R}$  define the function

$$c(x) = -x + \lim_{a \rightarrow \infty, a \in \mathbb{N}} \lim_{b \rightarrow \infty, b \in \mathbb{N}} -a + b + \ln(g^{(b)}(1 - 2^{-a-x})), \quad (1.2)$$

whose actual meaning will become clear later. We will show that the double limit exists, so that this indeed defines a function  $c : \mathbb{R} \rightarrow \mathbb{R}$ . Moreover, we will show that  $c$  is continuous and periodic with period 1, that is, if we write  $\{x\} = x - \lfloor x \rfloor$  then  $c(x) = c(\{x\})$ , and that (numerically)  $|\sup c - \inf c| \approx 10^{-9}$ , cf. Figure 1. The Gumbel distribution will play a prominent role in our considerations. We say that a real valued random variable  $G$  follows a  $\text{Gum}(\alpha)$  distribution with parameter  $\alpha \in \mathbb{R}$ ,  $G \sim \text{Gum}(\alpha)$ , if for all  $x \in \mathbb{R}$

$$P[G \leq x] = e^{-e^{-x-\alpha}}, \quad x \in \mathbb{R}.$$

Finally, let  $\gamma$  denote the Euler-Mascheroni constant. With all these ingredients we can now state our main result, which specifies – see also below – the distribution of the runtime of *push* on the complete graph.

**Lemma 1.1.** *Let  $G \sim \text{Gum}(\gamma)$ . Then, as  $n \rightarrow \infty$*

$$\sup_{k \in \mathbb{N}} \left| P[X_n \geq k] - P\left[\left[G + \log_2 n + \ln n + \gamma + c(\{\log_2 n\})\right] \geq k\right] \right| = o(1).$$

This lemma does not look completely innocent, and it actually has striking consequences. It readily implies the following result, which establishes that the limiting distribution  $X_n$  is periodic both on the  $\log_2 n$  and on the  $\ln n$  scale. In order to formulate it, we need a version of the Gumbel distribution where we restrict ourselves to integers only. More specifically, we say that a random variable  $G$  follows a *discrete Gumbel* distribution,  $G \sim \text{dGum}(\alpha)$ , if the domain of  $G$  is  $\mathbb{Z}$  and

$$P[G \leq k] = e^{-e^{-k-\alpha}}, \quad k \in \mathbb{Z}.$$

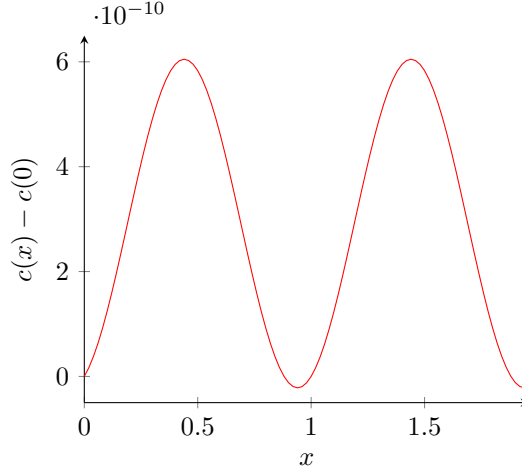


Figure 1: The function  $c(x) - c(0)$ ,  $c(0) \approx 0.105$ , plotted for values of  $x$  between 0 and 2. The periodic nature of the function and its small amplitude are evident.

**Theorem 1.2.** *Let  $x, y \in [0, 1)$  and  $(n_i)_{i \in \mathbb{N}}$  be a strictly increasing sequence of natural numbers, such that  $\log_2 n_i - \lfloor \log_2 n_i \rfloor \rightarrow x$  and  $\ln n_i - \lfloor \ln n_i \rfloor \rightarrow y$  as  $i \rightarrow \infty$ . Then in distribution, as  $i \rightarrow \infty$*

$$X_{n_i} - (\lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor) \rightarrow \text{dGum}(-x - y - c(x)).$$

Some remarks are in place. First, it is a priori not obvious (at least it was not to us) that subsequences as required in the theorem indeed exist. They do, and the fundamental reason for this is that real numbers can be approximated arbitrarily well by rational numbers; we include a short proof of the existence in the Appendix. Second, it is a priori not clear that  $x + c(x)$  is not constant for  $x \in [0, 1)$ . If it was constant, Theorem 1.2 would imply that the limiting distribution of  $X_n$  is periodic on the  $\ln n$  scale only. Although we didn't manage to *prove* that  $x + c(x)$  is not constant, we have strong numerical evidence that it indeed is not so. In particular, as we shall also see later, the double limit in the definition of  $c$  converges exponentially fast and thus it is not difficult to obtain accurate estimates for it and explicit error bounds. We leave it as an open problem to study the behavior of  $c$  more accurately.

Our next result addresses moments of  $X_n$ . Bounds given in previous works, for example in [7], guarantee that  $X_n - \log_2 n - \ln n$  and all integer powers of it are uniformly integrable. This allows us to conclude that the expectation and all of its moments also converge.

**Theorem 1.3.** *Let  $x, y \in [0, 1)$  and  $(n_i)_{i \in \mathbb{N}}$  be a strictly increasing sequence of natural numbers, such that  $\log_2 n_i - \lfloor \log_2 n_i \rfloor \rightarrow x$  and  $\ln n_i - \lfloor \ln n_i \rfloor \rightarrow y$  as  $i \rightarrow \infty$ . Then for all  $k \in \mathbb{N}$ , as  $i \rightarrow \infty$*

$$\mathbb{E}[(X_{n_i} - (\lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor))^k] \rightarrow \mathbb{E}[(\text{dGum}(-x - y - c(x)))^k].$$

For  $x, y \in [0, 1)$  and a strictly increasing sequence of natural numbers  $(n_i)_{i \in \mathbb{N}}$  such that  $\{\log_2 n_i\} \rightarrow x$  and  $\{\ln n_i\} \rightarrow y$  Theorem 1.3 immediately implies that, as  $i \rightarrow \infty$ ,

$$\mathbb{E}[X_{n_i}] = \log_2 n_i + \ln n_i + h(x, y) + o(1),$$

where we abbreviated  $h(x, y) = \mathbb{E}[\text{dGum}(-x - y - c(x))] - x - y$ , cf. Figure 2 for a visualization of  $h$ . Similarly, to obtain an expression for the variance of the runtime, see that

$$\begin{aligned} \text{Var}[X_{n_i}] &= \text{Var}[X_{n_i} - (\lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor)] \\ &= \mathbb{E}[(X_{n_i} - (\lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor))^2] - \mathbb{E}[X_{n_i} - (\lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor)]^2 \end{aligned}$$

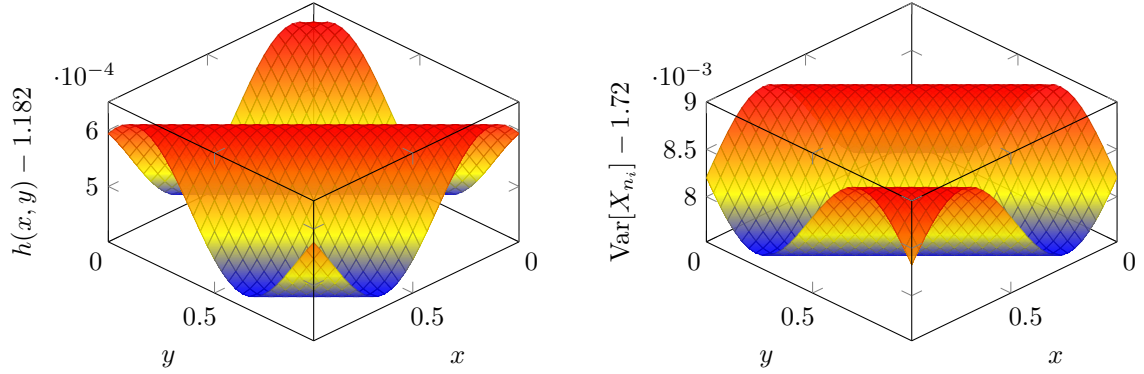


Figure 2: Let  $(n_i)_{i \in \mathbb{N}}$  be a sequence of natural numbers such that  $\{\log_2 n_i\} \rightarrow x$  and  $\{\ln n_i\} \rightarrow y$  for  $x, y \in [0, 1)$ . The left figure shows the function  $h(x, y)$  (appearing in the expectation of  $X_{n_i}$ ) for values of  $x$  and  $y$  between 0 and 1. The right figure shows  $\text{Var}[X_{n_i}]$  as a function of  $x, y$ .

and using Theorem 1.3, consequently

$$\text{Var}[X_{n_i}] = \mathbb{E}[\text{dGum}(-x - y - c(x))^2] - \mathbb{E}[\text{dGum}(-x - y - c(x))]^2 + o(1).$$

To determine the expectation and variance of the runtime we need to consider various moments of the discrete Gumbel distribution. To this end, let  $X$  be an integer valued random variable with finite  $k$ th moment, then

$$\mathbb{E}[X^k] = \sum_{\ell \in \mathbb{Z}} \ell^k P[X = \ell] = \sum_{\ell \in \mathbb{Z}} \ell^k (P[X \leq \ell] - P[X \leq \ell - 1]),$$

and therefore, for all  $\alpha \in \mathbb{R}$  and  $k \in \mathbb{N}$ ,

$$\mathbb{E}[\text{dGum}(\alpha)^k] = \sum_{\ell \in \mathbb{Z}} \ell^k \left( e^{-e^{-\ell-\alpha}} - e^{-e^{-\ell-\alpha+1}} \right).$$

This sum converges exponentially fast, both for  $\ell \rightarrow \infty$  and  $\ell \rightarrow -\infty$ , and thus allows for effective numerical treatment. In summary, improving (1.1), we get for all  $n \in \mathbb{N}$  the numerical bounds

$$\log_2 n + \ln n + 1.18242 \leq \mathbb{E}[X_n] \leq \log_2 n + \ln n + 1.18263,$$

as  $\inf_{0 \leq x, y \leq 1} h(x, y) = 1.18242 \dots$ ,  $\sup_{0 \leq x, y \leq 1} h(x, y) = 1.18262 \dots$  and

$$1.7277 \leq \text{Var}[X_n] \leq 1.7289.$$

These numerical bounds are (essentially) best possible, see also Figure 2. Higher moments can be estimated similarly.

Let us close this section with a final remark on the function  $c$  defined in (1.2), as this might be helpful in future works. This function is defined as the limit of a sequence in two parameters  $a, b$ ; the main reason for this is its combinatorial origin, which will become apparent in the proofs. However, all that is actually important is that  $b$  is large enough, in the sense that the difference  $b - a \rightarrow \infty$  as  $a \rightarrow \infty$ . So, if we write  $h$  for an integer function that diverges to infinity, then we could define

$$d(x) = -x + \lim_{a \rightarrow \infty, a \in \mathbb{N}} h(a) + \ln(g^{(a+h(a))}(1 - 2^{-a-x})).$$

Then  $c(x) = d(x)$  (which we state without proof, as we do not need it here), and  $c$  can be represented as a limit of an (one-dimensional) sequence.

**Outline** In the next section we give an outline of the proof of our main results. At the beginning of the rumour spreading process *push* is dominated by an exponential growth of the informed nodes (Lemma 2.2). For the main part, where most nodes get informed, it closely follows a deterministic recursion (Lemma 2.1) and at the end it is described by a coupon collector type problem (Lemma 2.3). Based on these lemmas we give the rigorous proof of our claims in Section 3. The proof to these three important lemmas can also be found there, in Subsections 3.3-3.5. Subsections 3.6 and 3.7 contain all other proofs.

**Further Notation** Unless stated otherwise, all asymptotic behaviour in this paper is for  $n \rightarrow \infty$ . Consider a graph  $G = (V, E)$ . For  $t \in \mathbb{N}_0 (= \mathbb{N} \cup \{0\})$  we denote by  $I_t \subseteq V$  the set of informed nodes at the end of round  $t$ ; in particular  $|I_0| = 1$ . Analogously we write  $U_t = V \setminus I_t$  for the set of uninformed nodes. For an event  $A$ , we sometimes write  $P_A[\cdot]$  instead of  $P[\cdot \mid A]$  to denote the conditional probability and we write  $\mathbb{E}_A[\cdot] = \mathbb{E}[\cdot \mid A]$ . If we condition on  $I_t$ , then we also abbreviate  $P[\cdot \mid I_t] = P_t[\cdot]$  and  $\mathbb{E}[\cdot \mid I_t] = \mathbb{E}_t[\cdot]$ .

## 2 Proof Overview

Let us start the proof of Lemma 1.1 about the distribution of the runtime of *push* on  $K_n$  with a simple observation, that is more or less explicit also in previous works. Note that as long as the *total number* of pushes performed is  $o(\sqrt{n})$ , then whp no node will be informed twice – this is a simple consequence of the famous birthday paradox. That is, whp as long as  $|I_t| = o(\sqrt{n})$ , every node in  $I_t$  will inform a currently uninformed node and thus  $|I_{t+1}| = 2|I_t|$ . In particular, whp

$$|I_{t_0}| = 2^{t_0}, \quad \text{where } t_0 := \lfloor 0.49 \cdot \log_2 n \rfloor. \quad (2.1)$$

Soon after round  $t_0$  things get more complicated. We continue with a definition. Apart from the functions  $g^{(i)}$  defined in the previous section, we will also need the following functions. Set

$$f = f^{(1)} : [0, 1] \rightarrow [0, 1], \quad x \mapsto 1 - e^{-x}(1 - x) \quad \text{and} \quad f^{(i)} : [0, 1] \rightarrow [0, 1], \quad f^{(i)} = f \circ f^{(i-1)}, i \geq 2.$$

Some elementary properties of  $f$  are:  $f$  is strictly increasing and concave, and  $f^{(b)}(x) \rightarrow 1$  as  $b \rightarrow \infty$  for all  $x \in (0, 1]$ . Moreover,  $f^{(i)}(x) = 1 - g^{(i)}(1 - x)$  for all  $x \in [0, 1]$  and  $i \in \mathbb{N}$ . It is also not difficult to establish, see also [20] and Lemma 3.5 below, that  $f$  captures the behavior of the expected number of informed nodes after one round of the protocol. Moreover,  $|I_{t+1}|$  is typically close to  $f(|I_t|/n)n$ . Here we will need a more explicit qualitative control of how  $|I_t|$  behaves, since our aim is to specify the limiting distribution. We show the following statement, which implies that if we start in round  $t_0$  (set  $T = t_0$  in that lemma) then whp for *all* succeeding rounds  $t_0 + t$  the number of informed nodes is close to  $f^{(t)}(|I_{t_0}|/n)n$ .

**Lemma 2.1.** *Let  $0 < c < 0.49$  and  $T \geq c \log_2 n$ . Then*

$$P_T \left[ \bigcap_{t \in \mathbb{N}_0} \left\{ \left| |I_{T+t}| - f^{(t)}(|I_T|/n)n \right| \leq n^{1-c/4} \right\} \right] = 1 - O(n^{-c^2/10}).$$

Thus, the key to understanding  $|I_t|$  is to understand how  $f$  behaves when iterated very many times. Note that when the number of informed nodes is  $xn$  for some very small  $x$ , then the  $e^{-x}$  term in the definition of  $f$  can be approximated by  $1 - x$  and therefore  $f(x) \approx 1 - (1 - x)^2 \approx 2x$ . This crude estimate suggests that the number of informed nodes doubles every round as long as there are only few informed nodes, and we know already that the doubling is perfect if  $xn = o(\sqrt{n})$ . Our next lemma actually shows that the doubling continues to be *almost* perfect, as long as the total number of nodes is not close to  $n$ .

**Lemma 2.2.** *Let  $a, T \in \mathbb{N}$  be such that  $2^{-a} < 0.1$  and  $T \leq \lfloor 0.49 \cdot \log_2 n \rfloor$ . Set  $t_1 := \lfloor \log_2 n \rfloor - a$ . Then*

$$\left| 2^{t_1} - f^{(t_1-T)}(2^T/n)n \right| \leq 2^{-2a+1}n.$$

Combining the previous lemmas we have thus established that for any  $a \in \mathbb{N}$  with  $2^{-a} < 0.1$  whp

$$(1 - 2^{-a+2}) \cdot 2^{t_1} \leq |I_{t_1}| \leq 2^{t_1}, \quad t_1 := \lfloor \log_2 n \rfloor - a. \quad (2.2)$$

Here we can think of  $a$  being very large (but fixed) and then the two bounds are very close to each other; in particular,  $|I_{t_1}| \approx 2^{\lfloor \log_2 n \rfloor - a}$  and thus  $I_t$  contains a linear number of nodes. Up to that point we have studied the behaviour of the process up to time  $t_1$ . Next we perform another  $b$  steps, where again  $b$  is fixed. Applying Lemma 2.1 once more and using that  $f^{(b)}(x)$  is increasing and is less than 1 for  $x < 1$  yields with room to spare that whp

$$\left(1 - n^{-1/6}\right) f^{(b)}((1 - 2^{-a+2})2^{t_1}/n) \leq n^{-1} |I_{t_2}| \leq \left(1 + n^{-1/6}\right) f^{(b)}(2^{t_1}/n), \quad t_2 := t_1 + b. \quad (2.3)$$

In essence, this says that if we write  $x = \log_2 n - \lfloor \log_2 n \rfloor = \{\log_2 n\}$ , then (we begin getting informal and obtain that)

$$|I_{t_2}| \approx f^{(b)}(2^{t_1}/n)n = f^{(b)}(2^{-a-x})n, \quad \text{where } t_2 = \lfloor \log_2 n \rfloor - a + b.$$

In particular, choosing a priori  $b$  large enough makes the fraction  $|I_{t_2}|/n$  arbitrarily close to 1, that is, almost all nodes except for a tiny fraction are informed. All in all, up to time  $t_2$  we have very fine control of the number of informed nodes, and we also see how the quantity  $\{\log_2 n\}$  slowly sneaks in.

After time  $t_2$  the behavior changes once more. In this regime there is an interesting connection to the well-known Coupon Collector Problem (CCP), which was also exploited in [7]. In order to formulate the connection, note that the number of pushes that are needed to inform one uninformed node, having  $N$  informed nodes, is (in distribution) equal to the number of coupons needed to draw the  $(N+1)$ st distinct coupon. The CCP is very well understood, and it is a classic result that, appropriately normalized, the number of coupons tends to a Gumbel distribution. However, translating the number of required pushes to the number of rounds – the quantity we are interested in – is not straightforward. In particular, the number of pushes in one round depends on the current number of informed nodes. On the other hand, after round  $t_2$  there are  $n - o(n)$  informed nodes, so that we may hope to approximate the remaining number of rounds with  $n^{-1}$  times the number of coupons in the CCP. The next lemma establishes the precise bridge between the two problems. There, for two sequences of random variables  $(X_n)_{n \in \mathbb{N}}$  and  $(Y_n)_{n \in \mathbb{N}}$  we write  $X_n \lesssim Y_n$  if there is a function  $h : \mathbb{N} \rightarrow \mathbb{R}^+$  with  $h = o(1)$  such that  $P[X_n \geq x] \leq P[Y_n \geq x] + h(n)$  for all  $n \in \mathbb{N}, x \in \mathbb{R}$ ;  $X_n \gtrsim Y_n$  is defined with “ $\geq$ ” instead of “ $\leq$ ”.

**Lemma 2.3.** *Let  $G \sim \text{Gum}(\gamma)$ ,  $b > 2a \in \mathbb{N}$  and assume that  $\ell \cdot n \leq |I_{\lfloor \log_2 n \rfloor - a + b}| \leq u \cdot n$  for some  $\ell, u \in [0, 1)$ . Then*

$$X_n - \lfloor \log_2 n \rfloor + a - b \gtrsim \left\lceil \ln n + \ln \left( \frac{1}{u} - 1 \right) + \gamma + \right\rceil$$

and

$$X_n - \lfloor \log_2 n \rfloor + a - b \lesssim \left\lceil \ln n + \ln \left( \frac{1}{\ell} - 1 \right) + \ln \left( \frac{\ell}{e\ell - e + 1} \right) + \gamma + G \right\rceil.$$

Note that the previous discussion guarantees that  $\ell, u$  in Lemma 2.3 are very close to 1 and very close to each other. So, the term  $\ln(\ell/(e\ell - e + 1))$  is very close to 0. We obtain that in distribution

$$X_n - \lfloor \log_2 n \rfloor + a - b \approx \left\lceil \ln n + \ln \left( \frac{1}{u} - 1 \right) + \gamma + G \right\rceil, \quad \text{where } u = f^{(b)}(2^{-a-x}).$$

and equivalently with  $x = \log_2 n - \lfloor \log_2 n \rfloor$

$$X_n \approx \left\lceil \log_2 n + \ln n - a + b + \ln \left( g^{(b)}(2^{-a-x}) \right) - x + \gamma + G \right\rceil. \quad (2.4)$$

Here we now encounter the mysterious function  $c$  from (1.2). The next lemma collects some important properties of it that will turn out to be very helpful.

**Lemma 2.4.** *The function*

$$c(x) = \lim_{a \rightarrow \infty} \lim_{a \in \mathbb{N}} \lim_{b \rightarrow \infty} \lim_{b \in \mathbb{N}} -a + b + \ln(g^{(b)}(1 - 2^{-a-x})) - x$$

*is well-defined, continuous and periodic with period 1.*

With all these facts at hand, the proof of Lemma 1.1 is completed by considering the random variable on the right-hand side of (2.4); in particular, the dependence on  $y = \ln n - \lfloor \ln n \rfloor$  arises naturally. The complete details of the proof, which is based on Lemmas 2.1–2.3 and follows the strategy outlined here can be found in Section 3 (together with the proofs of the lemmas).

As described in the introduction, apart from the limiting distribution we are interested in the asymptotic expectation of the runtime. A key ingredient towards the proof of Theorem 1.3 is uniform integrability, which can be shown by using the distributional bounds from [7]. Uniform integrability is a sufficient condition that convergence in distribution also implies convergence of the means.

**Lemma 2.5** (uniform integrability). *Let  $k \in \mathbb{N}$  and set  $Y_n := X_n - \lfloor \log_2 n \rfloor - \lfloor \ln n \rfloor$ . Then  $Y_n^k$  is uniformly integrable, that is*

$$\lim_{N \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}[|Y_n|^k \mid \mathbb{1}[|Y_n|^k > N]] = 0.$$

### 3 Proof of the Main Result

In this section we complete the proof of Lemma 1.1 outlined in Section 2. Afterwards we give the (short) proofs for Theorems 1.2 and 1.3.

#### 3.1 Proof of Lemma 1.1

As the outline was indeed rigorous until (2.3) we take the proof up from there. Choose the quantities  $a, b \in \mathbb{N}$  such that  $2a < b$  and recall that  $t_1 = \lfloor \log_2 n \rfloor - a$ . Set furthermore for brevity

$$\ell = \left(1 - n^{-1/6}\right) f^{(b)}((1 - 2^{-a+2})2^{t_1}/n) \quad \text{and} \quad u = \left(1 + n^{-1/6}\right) f^{(b)}(2^{t_1}/n).$$

Then (2.3) states that, for  $t_2 = \lfloor \log_2 n \rfloor - a + b$ ,

$$\ell \leq n^{-1} |I_{t_2}| \leq u,$$

and Lemma 2.3 yields, for  $Y_n = X_n - \lfloor \log_2 n \rfloor + a - b$ , that

$$Y_n \lesssim \left\lceil \ln n + \ln \left( \frac{1}{\ell} - 1 \right) + \ln \left( \frac{\ell}{e\ell - e + 1} \right) + \gamma + G \right\rceil$$

and

$$Y_n \gtrsim \left\lfloor \ln n + \ln \left( \frac{1}{u} - 1 \right) + \gamma + G \right\rfloor.$$

The next lemma establishes that both  $\ell, u$  tend to 1 as  $a$  gets large, and moreover that the difference  $\ln(1/\ell - 1) - \ln(1/u - 1)$  can be made arbitrarily small. Its proof can be found in Subsection 3.7.

**Lemma 3.1.** *For  $\ell, u$  defined as above, where  $b > 2a$*

$$\lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} |\ln \ell| = \lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} |\ln u| = \lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} \left| \ln \left( \frac{\ell}{e\ell - e + 1} \right) \right| = 0.$$

*Furthermore,*

$$\lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} |\ln(1 - \ell) - \ln(1 - u)| = 0.$$

Thus, as  $n \rightarrow \infty$ ,

$$\ln(1-u) = \ln\left(1 - f^{(b)}\left(2^{t_1}/n\right)\right) + o(1) = \ln\left(g^{(b)}\left(1 - 2^{-a-\{\log_2 n\}}\right)\right) + o(1).$$

Let  $\varepsilon > 0$ . Lemma 3.1 readily implies that there are  $a_0, n_0 \in \mathbb{N}$  such that for all  $a > a_0$  and  $n > n_0$ ,

$$Y_n \lesssim \left\lceil \ln n + \ln\left(g^{(b)}\left(1 - 2^{-a-\{\log_2 n\}}\right)\right) + \gamma + G - \varepsilon \right\rceil$$

and similarly also

$$Y_n \lesssim \left\lceil \ln n + \ln\left(g^{(b)}\left(1 - 2^{-a-\{\log_2 n\}}\right)\right) + \gamma + G + \varepsilon \right\rceil.$$

Lemma 2.4 guarantees that there is an  $a_1 \geq a_0$  such that for all  $a \geq a_1$

$$\left| \ln\left(g^{(b)}\left(1 - 2^{-a-\{\log_2 n\}}\right)\right) - a + b - (c(\{\log_2 n\}) + \{\log_2 n\}) \right| \leq \varepsilon.$$

Thus for all  $a > a_1$  and  $n > n_0$

$$X_n \lesssim \lceil \log_2 n + \ln n + c(\{\log_2 n\}) + \gamma + G - 2\varepsilon \rceil,$$

as well as

$$X_n \lesssim \lceil \log_2 n + \ln n + c(\{\log_2 n\}) + \gamma + G + 2\varepsilon \rceil.$$

Thus we are left with getting rid of the  $\varepsilon$  terms in the previous equations. The following lemma accomplishes exactly that and therefore implies the claim of Lemma 1.1. Its proof can be found in Subsection 3.7.

**Lemma 3.2.** *Let  $h : \mathbb{N} \rightarrow \mathbb{R}^+$  and  $G \sim \text{Gum}(\gamma)$ . Then*

$$\forall \varepsilon > 0 : X_n \lesssim [h(n) + G + \varepsilon] \implies X_n \lesssim [h(n) + G].$$

*The respective statement also holds for “ $\gtrsim$ ”.*

### 3.2 Proof of Theorems 1.2 and 1.3

*Proof of Theorem 1.2.* Recall that  $\{z\} = z - \lfloor z \rfloor, z \in \mathbb{R}$ . Let  $(n_i)_{i \in \mathbb{N}}$  be a strictly increasing subsequence of  $\mathbb{N}$  such that  $\{\log_2 n_i\} \rightarrow x$  and  $\{\ln n_i\} \rightarrow y$ . Substituting  $k = \lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor + 1 + t$  for any  $t \in \mathbb{Z}$  we get that

$$\begin{aligned} & P\left[\lceil G + \log_2 n_i + \ln n_i + \gamma + c(\{\log_2 n_i\}) \rceil \geq k\right] \\ &= P\left[\lceil G + \log_2 n_i + \ln n_i + \gamma + c(\{\log_2 n_i\}) \rceil \geq \lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor + 1 + t\right] \\ &= P\left[\lceil G + \{\log_2 n_i\} + \{\ln n_i\} + \gamma + c(\{\log_2 n_i\}) \rceil > t\right] \\ &= P\left[G + \{\log_2 n_i\} + \{\ln n_i\} + \gamma + c(\{\log_2 n_i\}) > t\right]. \end{aligned}$$

Thus using Lemma 1.1, Lemma 2.4 and Lemma 3.2 we get that, as  $i \rightarrow \infty$ ,

$$\sup_{t \in \mathbb{Z}} \left| P[X_{n_i} \geq \lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor + 1 + t] - P[G + x + y + \gamma + c(x) > t] \right| = o(1).$$

Using the distribution function of  $G \sim \text{Gum}(\gamma)$  we get

$$P[X_{n_i} \geq \lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor + 1 + t] \xrightarrow{i \rightarrow \infty} 1 - \exp(-\exp(-t + x + y + c(x))),$$

that is,

$$P[X_{n_i} \leq \lfloor \log_2 n_i \rfloor + \lfloor \ln n_i \rfloor + t] \xrightarrow{i \rightarrow \infty} P(\text{dGum}(-x - y - c(x)) \leq t).$$

□



Next we prove Theorem 1.3.

*Proof of Theorem 1.3.* Lemma 2.5 states that  $(X_n - \lfloor \log_2 n \rfloor - \lfloor \ln n \rfloor)^k$  is uniformly integrable and Theorem 1.2 established its convergence in distribution to  $(\text{dGum}(-x - y - c(x)))^k$ . Together this implies

$$\mathbb{E}[(X_n - \lfloor \log_2 n \rfloor - \lfloor \ln n \rfloor)^k] \rightarrow \mathbb{E}[(\text{dGum}(-x - y - c(x)))^k].$$

□

### 3.3 Proof of Lemma 2.1

The number of informed nodes,  $|I_t|$ , fulfils a so-called self-bounding property, for reference see [2]. One striking consequence thereof is the following bound.

**Lemma 3.3** ([5]). *For any  $t \in \mathbb{N}$ ,*

$$\text{Var}[|I_{t+1}| \mid I_t] \leq \mathbb{E}[|I_{t+1}| \mid I_t].$$

This bound on the variance and Chebychev's inequality ensure that the number of informed nodes is highly concentrated around its expectation as soon as enough nodes are informed. Moreover, even stronger concentration results are possible, as self-bounding functions admit exponential concentration inequalities, see e.g. [2]. Here, Chebychev is sufficient for our application.

**Lemma 3.4.** *Let  $0 < c \leq 1$ , let  $t_0 \in \mathbb{N}$  and assume that  $|I_{t_0}| \geq n^c$ . For  $t \in \mathbb{N}$  and  $\varepsilon > 0$  let  $C_t$  denote the event that*

$$||I_{t+1}| - \mathbb{E}_t[|I_{t+1}|]| \leq (\mathbb{E}_t[|I_{t+1}|])^{1/2+\varepsilon}.$$

*Then*

$$P_{t_0} \left[ \bigcap_{t \geq t_0} C_t \right] = 1 - O(n^{-c\varepsilon}).$$

*Proof.* From [7, Corollary 3.2] it is known that for any  $r > 0$

$$P[X_n \geq \lceil \log_2 n \rceil + \ln n + 2.188 + r] \leq 2e^{-r}.$$

Thus it suffices (with lots of room to spare) to show

$$P_{t_0} \left[ \bigcup_{t_0 \leq t \leq \log^2 n} \overline{C_t} \right] = O(n^{-3c\varepsilon/2}). \quad (3.1)$$

By using Chebychev's inequality and Lemma 3.3,

$$P_t[\overline{C_t}] = P_t \left[ ||I_{t+1}| - \mathbb{E}_t[|I_{t+1}|]| > \mathbb{E}_t[|I_{t+1}|]^{1/2+\varepsilon} \right] \leq \frac{\text{Var}[|I_{t+1}|]}{\mathbb{E}_t[|I_{t+1}|]^{1+2\varepsilon}} \leq \mathbb{E}_t[|I_{t+1}|]^{-2\varepsilon}.$$

Since  $\mathbb{E}_t[|I_{t+1}|] \geq |I_{t+1}| \geq |I_{t_0}|$  the claim follows from (3.1) and the union bound. □

Lemma 3.5 establishes a connection between the expected value of  $|I_{t+1}|$  and our previously defined function  $f$ , see below Equation 2.1. This has also been observed (though not so precise) in [20] and we include a quick proof for completeness.

**Lemma 3.5.** *Let  $t \in \mathbb{N}$  and  $n \geq 3$ . Then*

$$f(|I_t|/n)n \leq \mathbb{E}_t[|I_{t+1}|] \leq f(|I_t|/n)n + 5.$$

*Proof.* Each uninformed node  $u \in U_t$  remains uninformed if all  $|I_t|$  informed nodes do not push to  $u$ . Since all these events are independent, we obtain that the probability that  $u$  remains uninformed is  $(1 - 1/(n-1))^{|I_t|}$ . Thus by linearity of expectation

$$\mathbb{E}_t[|I_{t+1}|] = |I_t| + (n - |I_t|) \left( 1 - \left( 1 - \frac{1}{n-1} \right)^{|I_t|} \right) = n - (n - |I_t|) \left( 1 - \frac{1}{n-1} \right)^{|I_t|}.$$

For a lower bound we use  $1 - x \leq e^{-x}$  and get

$$\mathbb{E}_t[|I_{t+1}|] \geq n - (n - |I_t|)e^{-|I_t|/(n-1)} \geq n - (n - |I_t|)e^{-|I_t|/n} = f(|I_t|/n)n.$$

For an upper bound we use  $1 - x \geq e^{-x-x^2}$  for all  $x \leq 1/2$

$$\mathbb{E}_t[|I_{t+1}|] \leq n - (n - |I_t|)e^{-|I_t|/(n-1)-|I_t|/(n-1)^2} \leq n - (n - |I_t|)e^{-|I_t|/n} \exp\left(-\frac{2|I_t|}{(n-1)^2}\right)$$

and again using  $1 - x \leq e^{-x}$

$$\mathbb{E}_t[|I_{t+1}|] \leq n - (n - |I_t|)e^{-|I_t|/n} \left( 1 - \frac{2|I_t|}{(n-1)^2} \right) \leq f(|I_t|/n)n + 5.$$

□

Lemma 3.6 is an auxiliary result that we use in the proof of Lemma 2.1. It shows that  $f$  is concave and has decreasing derivative on the interval  $[0, 1]$ , the stated property is a direct consequence.

**Lemma 3.6.** *Let  $0 < x_1 \leq x_2 < 1$ . Then  $|f(x_2) - f(x_1)| \leq (2 - x_1)e^{-x_1}(x_2 - x_1)$ .*

*Proof.* It is  $f'(x) = (2-x)e^{-x}$  and  $f''(x) = (x-3)e^{-x}$ ; in particular,  $f'$  is monotonically decreasing and takes only positive values on  $[x_1, x_2]$ . Furthermore

$$\max_{x \in [x_1, x_2]} f'(x) = (2 - x_1)e^{-x_1}$$

and therefore, as a direct consequence of the mean value theorem, we have

$$|f(x_2) - f(x_1)| \leq (x_2 - x_1) \max_{x \in [x_1, x_2]} f'(x) = (2 - x_1)e^{-x_1}(x_2 - x_1).$$

□

We state a simple corollary for later reference.

**Corollary 3.7.** *Let  $i \in \mathbb{N}$  and  $r, s \in [0, 1/2]$ . Then  $f^{(i)}(r+s) \leq f^{(i)}(r) + 2^i s$ .*

Having these lemmas as ingredients we can prove the main result of this subsection. Lemma 3.5 shows that the expectation of  $|I_{t+1}|$  is given by  $f(|I_t|/n)n$  and Lemma 3.4 shows that  $|I_{t+1}|$  is closely concentrated around its expectation in (nearly) all rounds. To then prove that  $|I_{t+\tau}|$  is close to  $f^{(\tau)}(|I_t|/n)n$  for any  $\tau \in \mathbb{N}$  we need to make sure that the errors in the concentration and the approximation of the expectation are not blown up by repeated applications of  $f$ . We will show that  $f$  can indeed increase the error in each step by a factor that can be as large as  $\sqrt{2}$ , but luckily this only happens when  $|I_{t+\tau}| = o(n)$  and thus the accumulated error will remain small (as  $|I_t|$  nearly doubles in this regime).

*Proof of Lemma 2.1.* Let  $0 < \varepsilon < c/10$ , and assume, with foresight, that  $n \geq n_0$ , where  $n_0$  satisfies the inequalities

$$\sqrt{2} + 10n_0^{-\varepsilon} < \sqrt{2 + \varepsilon} \quad \text{and} \quad n_0^c \geq 25.$$

As  $T \geq c \log_2 n$  and because of (2.1), that is,  $|I_t| = 2^t$  for all  $t \leq \lfloor 0.49 \log_2 n \rfloor$ , we have  $|I_T| \geq n^c$ . Consequently we can apply Lemma 3.4 and thus get with probability  $1 - O(n^{-c\varepsilon})$

$$||I_{t+1}| - \mathbb{E}_t[|I_{t+1}|]| \leq \mathbb{E}_t[|I_{t+1}|]^{1/2+\varepsilon}, \quad \text{for all } t \geq T. \quad (3.2)$$

For the rest of this proof we assume that (3.2) holds. Set

$$\alpha_{T+t} = f^{(t)}(|I_T|/n), \quad t \in \mathbb{N}_0.$$

We will first argue that

$$||I_t| - \alpha_t n| \leq \alpha_t^{1/2+\varepsilon} n^{1/2+2\varepsilon} \sqrt{2+\varepsilon}^{t-T} =: d_t. \quad (3.3)$$

for all  $t \geq T$  such that  $d_t \leq n^{1-\varepsilon}$ . Note that this is obviously true for  $t = T$ . For the induction step we argue that

$$||I_{t+1}| - \alpha_{t+1} n| \leq \alpha_{t+1}^{1/2+\varepsilon} n^{1/2+2\varepsilon} \sqrt{2+\varepsilon}^{t+1-T} = d_{t+1}. \quad (3.4)$$

To see this, we use Lemma 3.5, (3.2) and the fact that  $|I_{t+1}| \leq 2|I_t|$  (in this order) to obtain the bound

$$||I_{t+1}| - f(|I_t|/n)n| \leq ||I_{t+1}| - \mathbb{E}_t[|I_{t+1}|]| + 5 \leq (2|I_t|)^{1/2+\varepsilon} + 5.$$

Then we apply Lemma 3.6 to estimate the difference of  $f(|I_t|/n)$  and  $\alpha_{t+1} = f(\alpha_t)$ , and infer from (3.3), using  $e^x \leq 1 + 2x$  for all  $0 \leq x \leq 1$ , that

$$\begin{aligned} |f(|I_t|/n)n - \alpha_{t+1}n| &\leq ||I_t| - \alpha_t n| (2 - \min\{\alpha_t, |I_t|/n\}) e^{-\min\{\alpha_t, |I_t|/n\}} \\ &\leq d_t (2 - \alpha_t + d_t/n) e^{-\min\{\alpha_t, |I_t|/n\}} \\ &\leq d_t (2 - \alpha_t) e^{-\alpha_t + d_t/n} + d_t^2/n \\ &\leq d_t (2 - \alpha_t) e^{-\alpha_t} + 5d_t^2/n. \end{aligned}$$

All in all we have argued that for all  $t$  such that  $d_t \leq n^{1-\varepsilon}$

$$\begin{aligned} ||I_{t+1}| - \alpha_{t+1}n| &\leq ||I_{t+1}| - f(|I_t|/n)n| + |f(|I_t|/n)n - \alpha_{t+1}n| \\ &\leq (2|I_t|)^{1/2+\varepsilon} + 5 + d_t (2 - \alpha_t) e^{-\alpha_t} + 5d_t^2/n \\ &\leq 2(\alpha_t n + d_t)^{1/2+\varepsilon} + 5 + d_t (2 - \alpha_t) e^{-\alpha_t} + 5d_t n^{-\varepsilon}. \end{aligned}$$

Our assumptions on  $\varepsilon$  and  $n$  imply that  $d_t^{1/2+\varepsilon} \leq d_t n^{-\varepsilon}$ . Moreover,  $\alpha_T n \geq n^c \geq 25$ , and thus

$$\begin{aligned} ||I_{t+1}| - \alpha_{t+1}n| &\leq 3(\alpha_t n)^{1/2+\varepsilon} + d_t (2 - \alpha_t) e^{-\alpha_t} + 7d_t n^{-\varepsilon} \\ &\leq d_t (2 - \alpha_t) e^{-\alpha_t} + 10d_t n^{-\varepsilon}. \end{aligned} \quad (3.5)$$

To understand (3.5) consider the auxiliary function

$$H(x) = \sqrt{\frac{f(x)}{x}} - \frac{f'(x)}{\sqrt{2}} = \sqrt{\frac{1 - (1-x)e^{-x}}{x}} - \frac{(2-x)e^{-x}}{\sqrt{2}}.$$

As  $(1-x)e^{-x} = 1 - 2x + O(x^2)$  as  $x \rightarrow 0$  we have that  $\lim_{x \rightarrow 0} (1 - (1-x)e^{-x})/x = 2$  and thus  $\lim_{x \rightarrow 0} H(x) = 0$ . Furthermore is  $H$  an increasing function on the interval  $[0, 1]$  as,

$$H' = \frac{1}{2} \left( \frac{2(1-x)e^{-x}}{x^2} \right)^{-1/2} + \frac{(3-x)e^{-x}}{\sqrt{2}} \geq 0 \quad \text{for } x \leq 1.$$

Therefore  $H(x) \geq 0$  for all  $0 \leq x \leq 1$  and consequently, using  $\alpha_{t+1} > \alpha_t$ ,

$$\left( \frac{\alpha_{t+1}}{\alpha_t} \right)^{1/2+\varepsilon} \geq \left( \frac{\alpha_{t+1}}{\alpha_t} \right)^{1/2} \geq \frac{(2 - \alpha_t) e^{-\alpha_t}}{\sqrt{2}}.$$

Since  $d_t = \alpha_t^{1/2+\varepsilon} n^{1/2+2\varepsilon} \sqrt{2+\varepsilon}^{t-T}$ , applying the previous bound to (3.5) implies (3.4) for all  $n \geq n_0$ , that is, all  $n$  such that  $\sqrt{2} + 10n^{-\varepsilon} < \sqrt{2+\varepsilon}$ . This completes the induction step and the proof of (3.3) is completed.

Actually our arguments yield also the following statement, which is stronger than (3.3) when there are “many” informed nodes. In particular, for all  $t' \in \mathbb{N}$  such that  $(2 - \alpha_{t'})e^{-\alpha_{t'}} < 1 - \varepsilon$  Equation (3.5) also yields for all  $n \geq n_0$

$$|I_{T+t'} - \alpha_{T+t'} n| \leq d_{t'} \quad \Rightarrow \quad |I_{T+t'+1} - \alpha_{T+t'+1} n| \leq d_{t'},$$

meaning that the absolute error does not increase any more after round  $t'$ . (Actually the error decreases by a factor of at least  $\varepsilon$  after that round, but we do not need this.) To complete the proof we show that we can choose  $t'$  such that  $d_{t'} \leq n^{1-c/4}$  and  $(2 - \alpha_{t'})e^{-\alpha_{t'}} < 1 - \varepsilon$ . To this end, consider

$$T' = \lfloor \log_2 n \rfloor - 4 - T$$

and applying Lemma 2.2 to  $\alpha_{T'}$  yields

$$\alpha_{T+T'} = f^{(T')}(|I_T|/n) \geq f^{(\lfloor \log_2 n \rfloor - 4 - T)}(2^T/n) \geq 2^{-4}(1 - 2^{-8+1})$$

and furthermore, a simple computation yields that  $\alpha_{T+T'+5} \geq 3/4$ . Thus

$$(2 - \alpha_{T+T'+5})e^{-\alpha_{T+T'+5}} \leq (2 - 3/4)e^{-3/4} < 1 - \varepsilon$$

and we set  $t' := T' + 5$ . Moreover,

$$d_{t'} \leq n^{1/2+2\varepsilon} \sqrt{2+\varepsilon}^{t'} \leq (2+\varepsilon)n^{1/2+2\varepsilon+(1-c)\log_2(2+\varepsilon)/2}.$$

Note that  $\log_2(2+\varepsilon) \leq 1+\varepsilon$ . Plugging this into the exponent yields that if  $\varepsilon < c/10$  and  $n$  is large enough then  $d_{t'} \leq n^{1-c/4} (\leq n^{1-\varepsilon})$ , as claimed.  $\square$

### 3.4 Proof of Lemma 2.2

We begin with showing the basic inequality

$$2x(1-x) \leq f(x) \leq 2x. \quad (3.6)$$

To see this, note that  $e^{-x} \leq 1 - x + x^2/2$  for  $x \in [0, 1]$  and so

$$f(x) = 1 - e^{-x}(1-x) \geq 1 - \left(1 - x + \frac{x^2}{2}\right)(1-x) \geq x \left(2 - \frac{3}{2}x\right) \geq 2x - 2x^2,$$

which establishes the first inequality in (3.6). The other inequality follows directly from the simple bound  $e^{-x} \geq 1 - x$ .

Let us write  $z_0 = 2^{t_0}/n$  and  $z_i = f(z_{i-1}) = f^{(i)}(z_0)$ ; we want to bound  $z_{t_1-t_0}$ , where  $t_1 = \lfloor \log_2 n \rfloor - a$  and  $t_0 \leq \lfloor 0.49 \log_2 n \rfloor$ . Clearly  $z_i \leq 2^i z_0$ , which shows the upper bound in Lemma 2.2. Using (3.6) we obtain by induction

$$z_i \geq 2^i z_0 \cdot \prod_{j=0}^{i-1} (1 - 2^j z_0), \quad i \in \mathbb{N}.$$

Further, using the bound  $1 - x \geq e^{-x-x^2/2(1-x)}$ , valid for any  $x \in [0, 1]$  we obtain

$$z_i \geq 2^i z_0 \cdot \exp \left\{ -z_0 \sum_{0 \leq j < i} 2^j - z_0^2 \sum_{0 \leq j < i} \frac{4^j}{2(1-2^j z_0)} \right\}$$

Note that our assumptions guarantee that  $2^{t_1-t_0} z_0 = 2^{-a} < 0.1$ , and so for any  $1 \leq i \leq t_1 - t_0$

$$z_i \geq 2^i z_0 \cdot \exp \{ -2^i z_0 - (2^i z_0)^2 \} \geq 2^i z_0 \cdot (1 - 2^{-a} - 2^{-2a}).$$

Finally note that  $1 - y - y^2 \geq 1 - 2y$  for any  $y \in [0, 1]$ , and so the last term is bounded by  $2^i z_0 \cdot (1 - 2^{-2a+1})$ , which coincides with the lower bound claimed in Lemma 2.2.

**Corollary 3.8.** *For all  $x \in [0, 1]$  and  $i \in \mathbb{N}$*

$$2^i x (1 - 2^i x - 2^{2i} x^2) \leq f^{(i)}(x) \leq 2^i x.$$

### 3.5 Proof of Lemma 2.3

A main tool in the forthcoming proof is the following result, which states that a sum of normalized independent geometric random variables converges to a Gumbel distributed random variable.

**Theorem 3.9** ([8]). *Let  $T_1, \dots, T_{n-1}$  be independent random variables such that  $T_i \sim \text{Geo}((n-i)/(n-1))$  for  $1 \leq i < n$ . Then, in distribution*

$$n^{-1} \sum_{1 \leq i < n} (T_i - \mathbb{E}[T_i]) \rightarrow \text{Gum}(\gamma).$$

Unfortunately we can not apply directly Theorem 3.9 to our setting, as we will have to deal with a sum of independent geometric random variables that are not normalized with the ‘correct’ factor  $n^{-1}$ . However, the next well-known statement assures that if the error is small enough we will still converge to the same limiting distribution

**Theorem 3.10** (Slutsky’s Theorem, see, e.g., [23, p. 19]). *Let  $(X_n)_{n \in \mathbb{N}}$ ,  $(Y_n)_{n \in \mathbb{N}}$  and  $(Z_n)_{n \in \mathbb{N}}$  be sequences of real-valued random variables. Suppose that  $X_n \rightarrow X$  in distribution and that there are constants  $a, b \in \mathbb{R}$  such that  $Y_n \rightarrow a$  and  $Z_n \rightarrow b$  in probability. Then  $Y_n X_n + Z_n \rightarrow aX + b$  in distribution.*

We now show a more general version of Theorem 3.9 that is applicable to our setting.

**Lemma 3.11.** *Let  $T_1, \dots, T_{n-1}$  be independent random variables such that  $T_i \sim \text{Geo}((n-i)/(n-1))$  for  $1 \leq i < n$ . Let furthermore  $\varepsilon > 0$  and  $s : \mathbb{N} \rightarrow [1, n]$  be a function such that  $s(n-i) \geq (1 - o(1))(n - c \cdot i)$  for any positive integer  $i < \varepsilon n$ . Then, in distribution*

$$\sum_{(1-\varepsilon)n \leq i < n} \frac{T_i - \mathbb{E}[T_i]}{s(i)} \rightarrow \text{Gum}(\gamma).$$

*Proof.* Let  $D_i = T_i - \mathbb{E}[T_i]$  be the centralised version of  $T_i$ . Then

$$\sum_{(1-\varepsilon)n \leq i < n} \frac{D_i}{s(i)} = \sum_{1 \leq i < n} \frac{D_i}{n} - \sum_{1 \leq i < (1-\varepsilon)n} \frac{D_i}{n} + \left( \sum_{(1-\varepsilon)n \leq i < n} \frac{D_i}{s(i)} - \sum_{(1-\varepsilon)n \leq i < n} \frac{D_i}{n} \right).$$

A direct application of Theorem 3.9 guarantees that the first sum converges to  $\text{Gum}(\gamma)$  in distribution. To complete the proof it is sufficient to argue that in probability

$$\sum_{1 \leq i < (1-\varepsilon)n} \frac{D_i}{n} \rightarrow 0 \quad \text{and} \quad \sum_{(1-\varepsilon)n \leq i < n} \left( \frac{D_i}{s(i)} - \frac{D_i}{n} \right) \rightarrow 0, \quad (3.7)$$

from which the claim in the lemma follows immediately from Theorem 3.10. Since the  $D_i$ ’s are centralised

$$\mathbb{E} \left[ \sum_{1 \leq i < (1-\varepsilon)n} \frac{D_i}{n} \right] = 0,$$

and using that  $\text{Var}[T_i] = ((n-1)(i-1))/(n-i)^2$  for all  $i < n$

$$\text{Var} \left[ \sum_{1 \leq i < (1-\varepsilon)n} \frac{D_i}{n} \right] = \sum_{1 \leq i < (1-\varepsilon)n} \frac{\text{Var}[T_i]}{n^2} = \sum_{1 \leq i < (1-\varepsilon)n} \frac{1}{n^2} \frac{(n-1)(i-1)}{(n-i)^2} \leq \sum_{1 \leq i < (1-\varepsilon)n} \frac{1}{(\varepsilon n)^2} = o(1).$$

Thus Chebychev's inequality directly implies that

$$\sum_{1 \leq i < (1-\varepsilon)n} \frac{D_i}{n} \rightarrow 0 \quad \text{in probability.}$$

It remains to treat the second term in (3.7). We compute the variance as before

$$\text{Var} \left[ \sum_{(1-\varepsilon)n \leq i < n} \frac{D_i}{s(i)} - \frac{D_i}{n} \right] = \sum_{(1-\varepsilon)n \leq i < n} \left( \frac{1}{s(i)} - \frac{1}{n} \right)^2 \frac{(n-1)(i-1)}{(n-i)^2} \leq \sum_{1 \leq i \leq \varepsilon n} \left( \frac{1}{s(n-i)} - \frac{1}{n} \right)^2 \frac{n^2}{i^2}.$$

However, this is also  $o(1)$ , as  $s(n-i) \geq (1-o(1))(n-c \cdot i)$  for all integers  $i \leq \varepsilon n$  by assumption, and therefore

$$0 \leq \frac{1}{s(i)} - \frac{1}{n} \leq \frac{1}{(1+o(1))(n-c \cdot i)} - \frac{1}{n} = (1+o(1)) \frac{c \cdot i + o(n)}{n^2}, \quad i \leq \varepsilon n.$$

In summary we have shown that

$$\text{Var} \left[ \sum_{(1-\varepsilon)n \leq i < n} \left( \frac{D_i}{s(i)} - \frac{D_i}{n} \right) \right] = o(1) \quad \text{and clearly} \quad \mathbb{E} \left[ \sum_{(1-\varepsilon)n \leq i < n} \left( \frac{D_i}{s(i)} - \frac{D_i}{n} \right) \right] = 0.$$

Thus Chebychev's inequality implies also the second statement in (3.7) and the proof is complete.  $\square$

A further ingredient that we shall exploit is the following fact. If a sequence of random variables  $X_n \rightarrow X$  in distribution with distribution functions  $F_n \rightarrow F$  and if  $F$  is continuous everywhere, then the convergence of  $F_n$  to  $F$  is even uniform.

**Theorem 3.12** (Polya's Theorem, [21, Theorem 1]). *For each  $n \in \mathbb{N}$  let  $X_n$  be a real-valued random variable with distribution function  $F_n$ . Assume that  $X_n \rightarrow X$  in distribution. If  $X$  has continuous distribution function  $F$ , then*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0.$$

We need one more auxiliary lemma that gives an upper bound on the informed nodes when going one round backwards in order to later convert the number of Coupons into the number of rounds that are needed to finish the protocol. Appropriately, Lemma 3.4 assures that in all rounds the number of informed nodes is tightly concentrated around its expectation, which in turn is described by  $f$ , thus applying  $f^{-1}$  will give a good bound.

**Lemma 3.13.** *Let  $t_0 \in \mathbb{N}$  and  $0 < \varepsilon < 1/6$ . Let  $C_t$  be the event that  $||I_{t+1}| - E_t[|I_{t+1}|]| \leq (E_t[|I_{t+1}|])^{1/2+\varepsilon}$ , as given in Lemma 3.4. Then for  $n$  large enough the event  $\bigcap_{t \geq t_0} C_t$  implies for all  $t \geq t_0$*

$$|I_t| \geq (1 - n^{-1/3}) \cdot e \cdot (|I_{t+1}| - (1 - 1/e)n).$$

*Proof.* Lemma 3.5 and  $C_t$  together give that

$$|I_{t+1}| \leq E_t[|I_{t+1}|] + (E_t[|I_{t+1}|])^{1/2+\varepsilon} = f(|I_t|/n)n + o(n^{2/3}).$$

Using the definition of  $f(x) = 1 - (1-x)e^{-x}$  and that  $|I_t| \leq n$  for all  $t$  we get that

$$|I_{t+1}| \leq n - e^{-|I_t|/n}(n - |I_t|) + o(n^{2/3}) \leq (1 - 1/e)n + |I_t|/e + o(n^{2/3}).$$

Rearranging yields the claimed statement.  $\square$

Let us briefly outline the proof of Lemma 2.3. We have already shown bounds for the number of informed nodes after  $\lfloor \log_2 n \rfloor - a + b$  rounds in (2.3). Starting from these bounds we will use the Coupon Collector Problem to compute the number of *pushes* that are needed to inform all remaining uninformed nodes. This will yield sums of independent geometric random variables (one summand for each uninformed node). Using Lemma 3.13 we will translate these numbers of pushes into numbers of rounds, which results in an almost correctly normalised sum of geometric random variables that Lemma 3.11 assures to converge to a Gumbel distribution. We will end up with upper and lower bounds to the distribution function of *push*.

*Proof of Lemma 2.3.* In this proof we will establish a connection between the Coupon Collector Problem and the behavior of *push*. Let  $v \in V$  be the node that was initially informed. Instead of every informed node choosing one of its neighbours uniformly at random, we now assume that it samples one node in  $V \setminus \{v\}$  uniformly at random. This defines an equivalent model, as for all  $u \in V$  the probability to choose any specific node in  $V \setminus \{u, v\}$  does not change (it equals  $1/(n-1)$  in both models) and choosing  $u$  or  $v$  makes no difference for the distribution of the set of informed nodes. Thus *push* is the same as drawing coupons out of a pool of  $n-1$  different coupons, but doing so in batches with size being the number of distinct coupons already collected plus one, the ‘plus one’ representing the initially informed node  $v$ . It is widely known and easy to see that assuming  $1 \leq i \leq n-1$  coupons (including  $v$ ) have already been collected, then

$$T_i \sim \text{Geo}\left(\frac{n-i}{n-1}\right), \quad 1 \leq i \leq n-1. \quad (3.8)$$

describes the number of coupons one needs to draw in order to draw the next,  $(i+1)$ st new, distinct coupon. Thus in order to collect all  $n$  coupons one needs to draw  $\sum_{i=1}^{n-1} T_i$  coupons, where the summands are independent random variables. However, we are not particularly interested in the total number of coupons drawn, but in the number of batches needed. If a batch has size  $s \leq n-1$ , then this batch is worth  $s$  coupons, or vice versa, each coupon drawn in this batch is worth  $1/s$  batches. Thus we need to estimate the size of the batch that contained all coupons that were needed to draw the  $(i+1)$ st distinct coupon, or if these coupons were contained in multiple batches, then we bound all those involved – we call these batches the batches that are *linked to*  $i+1$ . Let  $L_i$  be the smallest and  $U_i$  the largest size of a batch linked to the  $(i+1)$ st coupon. Then certainly  $U_i \leq i$ , as at the time that the  $(i+1)$ st distinct coupon gets collected there are obviously at most  $i$  distinct collected coupons. Using our assumption  $\ell \cdot n \leq |I_{\lfloor \log_2 n \rfloor - a + b}| \leq u \cdot n$  we thus obtain

$$\left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i}{U_i} \right\rceil \leq X_n - (\lfloor \log_2 n \rfloor - a + b) \leq \left\lceil \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i}{L_i} \right\rceil. \quad (3.9)$$

Abbreviating  $Y_n = X_n - (\lfloor \log_2 n \rfloor - a + b)$  and recalling that  $U_i \leq i$  yields

$$Y_n \geq \left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i}{U_i} \right\rceil = \left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i}{i} \right\rceil.$$

As the  $T_i$  are independent and geometrically distributed, we can apply Lemma 3.11 and for  $G \sim \text{Gum}(\gamma)$  we obtain with Theorem 3.12

$$\sup_{k \in \mathbb{Z}} \left| P \left[ \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i - \mathbb{E}[T_i]}{i} \geq k \right] - P[G \geq k] \right| = o(1)$$

and therefore

$$\left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i}{i} \right\rceil = \left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{\mathbb{E}[T_i]}{i} + \sum_{i=\lfloor un \rfloor}^{n-1} \frac{T_i - \mathbb{E}[T_i]}{i} \right\rceil \asymp \left\lceil \sum_{i=\lfloor un \rfloor}^{n-1} \frac{1}{i(1-i/n)} + G \right\rceil.$$

The partial fraction decomposition  $(i(1 - i/n))^{-1} = (n - i)^{-1} + i^{-1}$  allows us to simplify

$$\left[ \sum_{i=\lfloor un \rfloor}^{n-1} \frac{1}{i(1 - i/n)} + G \right] = \left[ \sum_{i=\lfloor un \rfloor}^{n-1} \frac{1}{n - i} + \sum_{i=\lfloor un \rfloor}^{n-1} \frac{1}{i} + G \right] = \left[ \sum_{i=1}^{n-\lfloor un \rfloor} \frac{1}{i} + \sum_{i=\lfloor un \rfloor}^{n-1} \frac{1}{i} + G \right].$$

Expressing these partial harmonic sums using the asymptotic expansion for the  $n$ th harmonic number

$$\sum_{1 \leq k \leq n} k^{-1} = H_n = \ln n + \gamma + O(1/n) \quad (3.10)$$

we get, using Lemma 3.2,

$$\begin{aligned} Y_n &\lesssim \lceil \ln(n - un) + \gamma + \ln n + \gamma - \ln(un) - \gamma + G + O(1/n) \rceil \\ &= \left\lceil \ln n + \ln \left( \frac{n(1 - u)}{un} \right) + \gamma + G + O(1/n) \right\rceil \\ &\lesssim \lceil \ln n + \ln(1/u - 1) + \gamma + G \rceil. \end{aligned}$$

We now look at the upper bound in (3.9). For all  $\lfloor \ell n \rfloor \leq i \leq n - 1$  we specify an appropriate bound for  $L_i$ . To obtain it, assume that  $t$  is the round in which the  $i$ th vertex was informed. Then all batches that are linked to the  $(i + 1)$ st coupon have size at least  $|I_t|$ , i.e.  $L_i \geq |I_t|$ , as the  $(i + 1)$ st distinct coupon is drawn after the  $i$ th distinct coupon, i.e., it cannot be drawn in any round  $t' < t$ . However, we do not know  $|I_t|$ , but we certainly can say that  $|I_{t+1}| \geq i$ . So, Lemma 3.13, guarantees that whp

$$|I_t| \geq (1 - n^{-1/3}) \cdot e \cdot (i - (1 - 1/e)n) \quad \text{for all } i \in \{\lfloor \ell n \rfloor, \dots, n - 1\}.$$

(Note that  $t = t(i)$  in that statement.) In particular, whp

$$L_i \geq |I_t| \geq (1 - n^{-1/3}) \cdot (n - e \cdot (n - i)) \quad \text{for all } i \in \{\lfloor \ell n \rfloor, \dots, n - 1\}.$$

Let  $C$  be the event that Lemma 3.13 conditions on, that is that  $|I_t|$  (for all  $t \in \mathbb{N}$ ) is closely concentrated around its expectation. Let  $k \in \mathbb{N}$  and

$$B = \left\{ \left[ \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i}{L_i} \right] \geq k \right\}.$$

Then  $P(B) = P(C \cap B) + o(1)$  and as

$$\{C \cap B \geq k\} \Rightarrow \left\{ \left[ \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i}{(1 - n^{-1/3})(n - e \cdot (n - i))} \right] \geq k \right\}$$

we get, recalling  $Y_n = X_n - (\lfloor \log_2 n \rfloor - a + b)$ , that

$$Y_n \leq \left[ \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i}{L_i} \right] \lesssim \left[ \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i}{(1 - n^{-1/3})(n - e \cdot (n - i))} \right].$$

Again applying Lemma 3.11 and Theorem 3.12, for  $G \sim \text{Gum}(\gamma)$  and  $c = e$ , we obtain

$$\begin{aligned} Y_n &\lesssim \left[ \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{\mathbb{E}[T_i]}{(1 - n^{-1/3})(n - e \cdot (n - i))} + \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{T_i - \mathbb{E}[T_i]}{(1 - n^{-1/3})(n - e \cdot (n - i))} \right] \\ &\lesssim \left[ (1 + O(n^{-1/3})) \sum_{i=\lfloor \ell n \rfloor}^{n-1} \frac{1}{(n - e \cdot (n - i))(1 - i/n)} + G \right]. \end{aligned}$$



Let  $c = 1 - 1/e$ . Using that  $((n - e \cdot (n - i))(1 - i/n))^{-1} = (n - i)^{-1} + (i - cn)^{-1}$  gives

$$Y_n \lesssim \left[ (1 + O(n^{-1/3})) \left( \sum_{i=\lceil \ell n \rceil}^{n-1} \frac{1}{n-i} + \sum_{i=\lceil \ell n \rceil}^{n-1} \frac{1}{i-cn} \right) + G \right].$$

Using index shifts, the asymptotic expansion for the harmonic number (3.10) and Lemma 3.2 yields

$$\begin{aligned} Y_n &\lesssim \left[ (1 + O(n^{-1/3})) \left( \sum_{i=1}^{n-\lceil \ell n \rceil} \frac{1}{i} + \sum_{i=\lceil \ell n \rceil - \lfloor cn \rfloor}^{n-1-\lfloor cn \rfloor} \frac{1}{i} \right) + G + o(1) \right] \\ &\lesssim \left[ \ln n + \ln(1/\ell - 1) - \ln(1/\ell) + \gamma + \ln \left( \frac{1-c}{\ell-c} \right) + G \right]. \end{aligned}$$

□

### 3.6 Proof of Lemma 2.4

In this subsection we investigate the double limit

$$\lim_{a \rightarrow \infty, a \in \mathbb{N}} \lim_{b \rightarrow \infty, b \in \mathbb{N}} -a + b + \ln(g^{(b)}(1 - 2^{-a-x})) - x$$

where  $g(x) = xe^{x-1}$ . We will show that this limit exists and defines a continuous function  $c(x)$ . It being periodic with period 1 is an immediate consequence of substituting  $a \rightarrow a + 1$  in the limit. A similar proof would also yield that  $c$  is continuously differentiable, but we only need continuity in the proof of our main theorem.

Before we actually prove Lemma 2.4 let us state two auxiliary statements first. In Definition 3.14, we quantify “exponentially fast convergence” and in Lemma 3.15 we state some simple properties.

**Definition 3.14** (Exponentially fast convergence). *Let  $(a_n)_{n \in \mathbb{N}}$  be a real-valued sequence and let  $c \in (0, 1)$ . If there is an  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$  we have  $|a_{n+1}| < c|a_n|$ , then we say that  $a_n$  converges exponentially fast to zero at rate  $c$  with start number  $n_0$ .*

**Lemma 3.15.** *a) Let  $c \in (0, 1)$  and let  $(a_n)_{n \in \mathbb{N}}$  be a real-valued sequence that converges exponentially fast to zero at rate  $c$ . Then  $\sum_{n \geq 1} a_n$  converges absolutely.*

*b) Let  $c \in (0, 1)$ ,  $n_0 \in \mathbb{N}$  and let  $(h_n)_{n \in \mathbb{N}}$  denote a sequence of functions with  $h_n : [0, 1] \rightarrow \mathbb{R}$  such that for any  $x \in [0, 1]$  the sequence  $(h_n(x))_{n \in \mathbb{N}}$  converges exponentially fast to zero at rate  $c$  with start number at most  $n_0$ . Define  $h : [0, 1] \rightarrow \mathbb{R}$  by  $h(x) := \sum_{n \geq 1} h_n(x)$ . Then the sequence of functions  $(\sum_{j=1}^n h_j)_{n \in \mathbb{N}}$  converges uniformly to  $h$ , i.e.*

$$\lim_{n \rightarrow \infty} \sup_{x \in [0, 1]} \left| h(x) - \sum_{j=1}^n h_j(x) \right| = 0.$$

*Proof.* a) is elementary. We prove b). Let  $\varepsilon > 0$ . We show that there is an  $n_1 \in \mathbb{N}$  such that for all  $n \geq n_1$  and for all  $x \in [0, 1]$  it holds  $\left| \sum_{j=1}^n h_j(x) - h(x) \right| < \varepsilon$ . For  $n \geq n_0$  it is

$$\left| \sum_{j=1}^n h_j(x) - h(x) \right| = \left| \sum_{j=n+1}^{\infty} h_j(x) \right| \leq \sum_{j=n+1}^{\infty} |h_j(x)| \leq |a_{n_0}| \sum_{j=n+1}^{\infty} c^j = |a_{n_0}| \frac{c^{n+1}}{1-c}$$

which implies that an  $n_1$  as required exists. □

*Proof of Lemma 2.4.* We show first, that for  $a$  fixed and any  $x \in [0, 1]$  the limit

$$\lim_{b \rightarrow \infty, b \in \mathbb{N}} b + \ln \left( g^{(b)} (1 - 2^{-a-x}) \right)$$

exists and the convergence is uniform. Inductively we get

$$\begin{aligned} b + \ln \left( g^{(b)} (1 - 2^{-a-x}) \right) &= b + \ln \left( g^{(b-1)} (1 - 2^{-a-x}) \right) + g^{(b-1)} (1 - 2^{-a-x}) - 1 \\ &= 1 + \ln (1 - 2^{-a-x}) - 2^{-a-x} + \sum_{j=1}^{b-1} g^{(j)} (1 - 2^{-a-x}) \end{aligned} \quad (3.11)$$

which, according to Lemma 3.15 a), converges for  $b \rightarrow \infty$  because  $g^{(j)} (1 - 2^{-a-x})$  converges exponentially fast to zero at rate at most  $\exp(-2^{-a-1}) < 1$  and start number 1 for  $j \rightarrow \infty$  in the sense of Definition 3.14. For  $x \in [0, 1]$ , according to Lemma 3.15 b), the convergence is even uniform with respect to  $x$ . By the Uniform Limit Theorem we thus showed that

$$\gamma_a(x) = -a + \sum_{j \geq 1} g^{(j)} (1 - 2^{-a-x}) \quad \text{is continuous for } a \in \mathbb{N}. \quad (3.12)$$

To complete the proof we show that the sequence of continuous functions  $(\gamma_a)_{a \in \mathbb{N}}$  converges uniformly. But first we make an observation. Let  $a' > a \in \mathbb{N}$  and  $x \in [0, 1]$ , then, using  $g(x) = 1 - f(1 - x)$ ,

$$\begin{aligned} \gamma_{a'}(x) &= -a' + \sum_{j \geq 1} g^{(j)} (1 - 2^{-a'-x}) = -a' + \sum_{j=1}^{a'-a} g^{(j)} (1 - 2^{-a'-x}) + \sum_{j \geq 1} g^{(j)} \left( g^{(a'-a)} (1 - 2^{-a'-x}) \right) \\ &= -a - \sum_{j=1}^{a'-a} f^{(j)} (2^{-a'-x}) + \sum_{j \geq 1} g^{(j)} \left( 1 - f^{(a'-a)} (2^{-a'-x}) \right). \end{aligned}$$

Furthermore, we can bound the repeated application of  $f$  using Corollary 3.8 and therefore

$$0 \leq \sum_{j=1}^{a'-a} f^{(j)} (2^{-a'-x}) \leq 2^{-a-x+1} \quad \text{and} \quad 2^{-a-x} (1 - 2^{-a-x} - 2^{-2a-2x}) \leq f^{(a'-a)} (2^{-a'-x}) \leq 2^{-a-x}.$$

Thus there is  $x' \in [0, 1]$  such that  $|x - x'| \leq 2^{-a}$  and  $\gamma_{a'}(x) = \gamma_a(x') + O(2^{-a})$ .

With this at hand we show uniform convergence of  $(\gamma_a)_{a \in \mathbb{N}}$ . In particular, for any  $0 < \varepsilon < 1/8$  we will show that there is some  $N \in \mathbb{N}$  such that  $\sup_{x \in [0, 1]} |\gamma_a(x) - \gamma_{a'}(x)| \leq \varepsilon$  for all  $a' > a > N$ . To achieve this we use our previous observation and obtain that

$$\begin{aligned} \sup_{x \in [0, 1]} |\gamma_a(x) - \gamma_{a'}(x)| &\leq \sup_{x \in [0, 1], |x - x'| \leq 2^{-a}} |\gamma_a(x) - \gamma_a(x')| + O(2^{-a}) \\ &= \sup_{x \in [0, 1], |x - x'| \leq 2^{-a}} \left| \sum_{j \geq 1} (g^{(j)} (1 - 2^{-a-x'}) - g^{(j)} (1 - 2^{-a-x})) \right| + O(2^{-a}). \end{aligned}$$

We bound this sum by splitting it into three parts. There is  $M_1 \in \mathbb{N}$  such that for any  $a > M_1$  there is  $N_1 \in \mathbb{N}$  ( $N_1$  depending on  $a$  and  $\varepsilon$ ) such that

$$\varepsilon \leq f^{(N_1)} (2^{-a-1}) \leq f^{(N_1+1)} (2^{-a}) \leq 8\varepsilon. \quad (3.13)$$

That is,  $N_1$  is the number of iterations such that  $f^{(N_1)} (2^{-a}) \approx \varepsilon$ , in particular  $N_1 \leq a$ , as  $f^{(a)} (2^{-a-1}) \geq 1/8$  by Corollary 3.8 and the fact that  $f$  is increasing. Furthermore, using again

that  $g^{(j)}(1 - 2^{-a-x})$  converges exponentially fast to zero with rate at most  $\exp(-2^{-a-1}) < 1$  for  $j \rightarrow \infty$ , there is  $c \in \mathbb{N}$  depending only on  $\varepsilon$  such that for  $N_2 := N_1 + c$

$$0 \leq \sup_{x \in [0,1]} \sum_{j \geq N_2} g^{(j)}(1 - 2^{-a-x}) \leq \varepsilon \quad \text{for all } a > M_1. \quad (3.14)$$

Then, abbreviating  $h^{(j)} = g^{(j)}(1 - 2^{-a-x'}) - g^{(j)}(1 - 2^{-a-x})$ , we can write

$$\sum_{j=1}^{\infty} (g^{(j)}(1 - 2^{-a-x'}) - g^{(j)}(1 - 2^{-a-x})) = \sum_{j=1}^{N_1} h^{(j)} + \sum_{j=N_1+1}^{N_2} h^{(j)} + \sum_{j>N_2} h^{(j)}. \quad (3.15)$$

In the rest of the proof estimate these sums individually, starting with the first one. Again using (3.11) and  $f(x) = 1 - g(1 - x)$  we have as  $a \rightarrow \infty$

$$\begin{aligned} \sum_{j=1}^{N_1} g^{(j)}(1 - 2^{-a-x'}) - g^{(j)}(1 - 2^{-a-x}) &= \ln g^{(N_1)}(1 - 2^{-a-x'}) - \ln g^{(N_1)}(1 - 2^{-a-x}) + O(2^{-a}) \\ &= \ln \left( 1 - f^{(N_1)}(2^{-a-x'}) \right) - \ln \left( 1 - f^{(N_1)}(2^{-a-x}) \right) + O(2^{-a}). \end{aligned}$$

By our choice of  $N_1$ , see (3.13), and the elementary inequalities  $z/(1+z) \leq \ln(1+z) \leq z$  for all  $z > -1$  this yields the upper bound

$$\sup_{x \in [0,1]} \left| \sum_{j=1}^{N_1} h^{(j)} \right| \leq \varepsilon + \frac{8\varepsilon}{1+8\varepsilon} + O(2^{-a}) \quad \text{for all } a > M_1. \quad (3.16)$$

We continue with the second sum in (3.15). Corollary 3.7 yields

$$\left| \sum_{j=N_1+1}^{N_2} h^{(j)} \right| = \left| \sum_{j=N_1+1}^{N_2} (f^{(j)}(2^{-a-x}) - f^{(j)}(2^{-a-x'})) \right| \leq (2^{-a-x} - 2^{-a-x'}) \sum_{j=N_1+1}^{N_2} 2^j.$$

Thus, as  $N_1 \leq a$  and  $N_2 = N_1 + c$ , where  $c$  depends on  $\varepsilon$  only, and our assumption  $|x - x'| \leq 2^{-a}$  there is  $M_2 \geq M_1$  such that

$$\left| \sum_{j=N_1+1}^{N_2} h^{(j)} \right| \leq (2^{2^{-a}} - 1) \cdot 2^{-a} \cdot \sum_{j=N_1+1}^{N_2} 2^j \leq (2^{2^{-a}} - 1) \cdot 2^{c+1} \leq \varepsilon \quad \text{for all } a > M_2. \quad (3.17)$$

In summary, (3.15) gives

$$\sup_{x \in [0,1]} |\gamma_a(x) - \gamma_{a'}(x)| \leq \sup_{x \in [0,1], |x-x'| \leq 2^{-a}} \left| \sum_{j=1}^{N_1} h^{(j)} + \sum_{j=N_1+1}^{N_2} h^{(j)} + \sum_{j>N_2} h^{(j)} \right| + O(2^{-a}).$$

and for  $a > M_2 > M_1$ , applying (3.16), (3.17) and (3.14) yields the uniform convergence of  $(\gamma_a)_{a \in \mathbb{N}}$ .  $\square$

### 3.7 Other Proofs

In this subsection we complete the rigorous treatment of our main theorems and give the last two remaining proofs. First we prove Lemma 2.5, which states that  $X_n - \lfloor \log_2 n \rfloor - \lfloor \ln n \rfloor$  is uniformly integrable.

*Proof of Lemma 2.5.* Doerr and Künnemann show in [7, Cor. 3.2 and Thm. 4.1] that for all  $r \in \mathbb{N}$

$$P[X_n \geq \lfloor \log_2 n \rfloor + \ln n + 2.188 + r] \leq 2e^{-r} \quad \text{and} \\ P[X_n \leq r] \leq P\left[\lfloor \log_2 n \rfloor - 1 + \frac{C_n(\lceil n/2 \rceil)}{n} \leq r\right],$$

where  $C_n(\lceil n/2 \rceil)$  is the number of rounds a coupon collector needs to draw the last  $n/2$  out of  $n$  coupons. These two bounds together with common deviation bounds for the coupon collector problem imply, see e.g. [8], that

$$P[Y_n \notin [\lfloor \log_2 n \rfloor + \lfloor \ln n \rfloor \pm (r + 5)] \leq 4e^{-r}.$$

Using this inequality we get that for any  $N \in \mathbb{N}$

$$\mathbb{E}\left[|Y_n|^k \mid \mathbb{1}[|Y_n|^k > N]\right] \leq \sum_{t \geq \sqrt[k]{N}} (t + 5)^k 4e^{-t},$$

which implies the claim.  $\square$

We close the section with the proof of Lemmas 3.1 and 3.2.

*Proof of Lemma 3.1.* First we observe that the  $(1 - n^{-1/6})$  error term in the definition of  $\ell, u$  is negligible as it factors out as a small additional term. Thus it suffices to consider  $\ell = f^{(b)}(L)$  and  $u = f^{(b)}(U)$  where  $L = (1 - 2^{-a+2})2^{-a-x}$  and  $U = 2^{-a-x}$  for some  $x \in (0, 1]$ . We assume that  $a \geq 3$ .

We start by showing an analogue to Corollary 3.7 but concerning  $g$ . For all  $r \geq s \in [0, 1]$ , using  $1 - x \leq e^{-x}$ ,

$$g(r - s) = (r - s)e^{r-s-1} \geq re^{r-s-1} - se^{r-1} \geq g(r) - s(1 + r)e^{r-1}$$

and consequently

$$g^{(i)}(r - s) \geq g^{(i)}(r) - s((1 + r)e^{r-1})^i \quad \text{for all } r \geq s \in [0, 1] \text{ and } i \in \mathbb{N}. \quad (3.18)$$

This completes our preparations. In order to show that  $(1 - \ell)/(1 - u) \rightarrow 1$  as  $a \rightarrow \infty$  we argue that  $\ell$  and  $u$  are very close together and approach 1 as  $a$  (and  $b > 2a$ ) gets big. We start by bounding the distance between  $\ell$  and  $u$ . Applying Corollary 3.7 to  $U = L + 2^{-2a-x+2}$  we get that

$$f^{(a)}(U) = f^{(a)}(L + 2^{-2a-x+2}) \leq f^{(a)}(L) + 2^{-a-x+2} \quad (3.19)$$

and Corollary 3.8 bounds  $f^{(a+1)}(U)$  from below with  $2^{-x-1}(1 - 2^{-x-1} - 2^{-2x-2}) \geq 1/8$ , thus  $f^{(a+2)}(U) \geq 1/2$ , and therefore we get using the monotonicity of  $f$

$$\frac{1}{2} \leq f^{(a+2)}(U) \leq f^{(a+3)}(L) \leq f^{(a+3)}(U) \stackrel{(3.19)}{\leq} f^{(a+3)}(L) + 2^{-a-x+5}. \quad (3.20)$$

We switch our focus to  $g$ . Observe that  $z := 3e^{-1/2}/2 < 1$  and, using (3.18),

$$g^{(b-a-3)}\left(1 - f^{(a+3)}(L) - 2^{-a-x+5}\right) \geq g^{(b-a-3)}\left(1 - f^{(a+3)}(L)\right) - 2^{-a-x+5} \cdot z^{b-a-3}.$$

This implies, using (3.20) and the previous equation, that

$$g^{(b-a-3)}\left(1 - f^{(a+3)}(U)\right) \geq g^{(b-a-3)}\left(1 - f^{(a+3)}(L)\right) - 2^{-a-x+5} \cdot z^{b-a-3},$$

and therefore, as  $1 - f^{(b)}(L) \geq 1 - f^{(b)}(U) = g^{(b-a-3)}(1 - f^{(a+3)}(U))$ ,

$$|u - \ell| = |f^{(b)}(U) - f^{(b)}(L)| \leq 2^{-a-x+5} z^{b-a-3} \rightarrow 0 \quad \text{as } a \rightarrow \infty, b - a \rightarrow \infty. \quad (3.21)$$

Next we show that  $u, \ell$  approach 1. Using  $g(x) = 1 - f(1 - x)$ , (3.20),  $g$  being increasing and  $g(x) \leq xe^{-1/2}$  for all  $x \leq 1/2$  (in that order), we get

$$g^{(b)}(1 - L) = g^{(b-a-3)} \left( 1 - f^{(a+3)}(L) \right) \leq g^{(b-a-3)} \left( \frac{1}{2} \right) \leq \frac{1}{2} e^{-(b-a-3)/2}, \quad b > a + 3.$$

Moreover, using that  $f(x) \leq 2x$  and  $g(x) \geq x/e$ ,

$$g^{(b)}(1 - U) = g^{(b-a)} \left( 1 - f^{(a)}(U) \right) \geq (1 - 2^{-x}) e^{-(b-a)}.$$

Thus, these two bounds together give

$$1 - \frac{1}{2} e^{-(b-a-3)/2} \leq f^{(b)}(L) \leq f^{(b)}(U) \leq 1 - (1 - 2^{-x}) e^{-(b-a)} \quad \text{for all } b > a + 3. \quad (3.22)$$

We just showed that  $u, \ell \rightarrow 1$  as  $a$  (and  $b$ ) tends to infinity. This yields that  $\ln u, \ln \ell$  and  $\ln(\ell/(e\ell - e + 1))$  tend to 0, leaving us with the term  $\ln((1 - \ell)/(1 - u))$ . The fact  $U \leq f(L)$  (and so  $f^{(b-2)}(U) \leq f^{(b-1)}(L)$ ) implies that

$$\begin{aligned} \frac{1 - \ell}{1 - u} &= \frac{g^{(b)}(1 - L)}{g^{(b)}(1 - U)} = \frac{\exp(g^{(b-1)}(1 - L) - 1) \cdot \exp(g^{(b-2)}(1 - L) - 1)}{\exp(g^{(b-1)}(1 - U) - 1) \cdot \exp(g^{(b-2)}(1 - U) - 1)} \cdot \frac{g^{(b-2)}(1 - L)}{g^{(b-2)}(1 - U)} \\ &\leq \frac{\exp(g^{(b-2)}(1 - L) - 1)}{\exp(g^{(b-1)}(1 - U) - 1)} \cdot \frac{g^{(b-2)}(1 - L)}{g^{(b-2)}(1 - U)}. \end{aligned}$$

Applying the same estimate to the latter fraction inductively we get for any  $c \in \mathbb{N}$

$$\frac{1 - \ell}{1 - u} \leq \frac{\exp(g^{(c)}(1 - L) - 1)}{\exp(g^{(b-1)}(1 - U) - 1)} \cdot \frac{g^{(c)}(1 - L)}{g^{(c)}(1 - U)} \leq \exp(g^{(c)}(1 - L)) \cdot \frac{g^{(c)}(1 - L)}{g^{(c)}(1 - U)}.$$

Set  $c = \lceil a(1 + \ln 2) \rceil$ . Using (3.21) and (3.22), where we set  $b = c$ , we obtain (for large enough  $a$ ) that  $|g^{(c)}(1 - L) - g^{(c)}(1 - U)| \leq 2^{-a-x+5} z^{c-a-3}$  as well as  $f^{(c)}(U) \leq 1 - (1 - 2^{-x}) e^{-(c-a)}$  and  $f^{(c)}(L) \geq 1 - e^{-(c-a-3)/2}/2$ . Thus

$$\begin{aligned} \frac{1 - \ell}{1 - u} &\leq \exp(1 - f^{(c)}(L)) \left( 1 + \frac{2^{-a-x+5} z^{c-a-3}}{1 - f^{(c)}(U)} \right) \\ &\leq \exp\left(\frac{1}{2} e^{-(c-a-3)/2}\right) \left( 1 + \frac{2^{-a-x+5} z^{c-a-3}}{(1 - 2^{-x}) e^{-c+a}} \right). \end{aligned}$$

Using  $e^x \leq 1 + 2x, x \in [0, 1]$  this yields the bounds

$$1 \leq \frac{1 - \ell}{1 - u} \leq \left( 1 + \sqrt{2}^{-a+4} \right) \left( 1 + \frac{2^{-x+5} e}{1 - 2^{-x}} \cdot z^{a \ln 2 - 2} \right) \quad \text{for all } a \in \mathbb{N}.$$

Therefore, as  $0 < z < 1$  we obtain  $\frac{1 - \ell}{1 - u} \rightarrow 1$ , and consequently  $\ln((1 - \ell)/(1 - u)) \rightarrow 0$ , as  $a \rightarrow \infty$ .  $\square$

Lemma 3.2 states that disturbing a Gumbel distributed random variable by a small amount does not significantly alter its distribution.

*Proof of Lemma 3.2.* Observe that  $\lceil h(n) + G \pm \varepsilon \rceil \neq \lceil h(n) + G \rceil$  is equivalent to

$$G \in [j - h(n) - \varepsilon, j - h(n) + \varepsilon] \quad \text{for some } j \geq 1.$$

But as  $G$  is absolutely continuous, for any  $\delta > 0$  we can choose  $\varepsilon$  small enough such that

$$P \left[ G \in \bigcup_{j \geq 1} [j - h(n) - \varepsilon, j - h(n) + \varepsilon] \right] \leq \delta.$$

$\square$

## References

- [1] K. P. Birman, M. Hayden, O. Ozkasap, Z. Xiao, M. Budiu, and Y. Minsky. Bimodal multicast. *ACM Transactions on Computer Systems (TOCS)*, 17(2):41–88, 1999.
- [2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.
- [3] F. Chierichetti, S. Lattanzi, and A. Panconesi. Almost tight bounds for rumour spreading with conductance. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 399–408. ACM, 2010.
- [4] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumour spreading and graph conductance. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1657–1663. SIAM, 2010.
- [5] R. Daknama, K. Panagiotou, and S. Reisser. Robustness of Randomized Rumour Spreading. In *27th Annual European Symposium on Algorithms (ESA 2019)*, volume 144 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:15, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [6] 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 sixth annual ACM Symposium on Principles of distributed computing*, pages 1–12. ACM, 1987.
- [7] B. Doerr and M. Künnemann. Tight analysis of randomized rumor spreading in complete graphs. In *2014 Proceedings of the Eleventh Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 82–91. SIAM, 2014.
- [8] P. Erdős and A. Rényi. On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 6, pages 215–220, 1961.
- [9] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures & Algorithms*, 1(4):447–460, 1990.
- [10] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable Broadcasting in Random Networks and the Effect of Density. In *INFOCOM 2010. 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 15-19 March 2010, San Diego, CA, USA*, pages 2552–2560, 2010.
- [11] N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and expanders. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 560–573, 2010.
- [12] 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.
- [13] G. Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *28th International Symposium on Theoretical Aspects of Computer Science, STACS 2011, March 10-12, 2011, Dortmund, Germany*, pages 57–68, 2011.
- [14] G. Giakkoupis. Tight Bounds for Rumor Spreading with Vertex Expansion. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 801–815, 2014.
- [15] G. Giakkoupis and T. Sauerwald. Rumor spreading and vertex expansion. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 1623–1641, 2012.

- [16] G. H. Hardy, E. M. Wright, et al. *An introduction to the theory of numbers*. Oxford University Press, 1979.
- [17] D. Mosk-Aoyama and D. Shah. Fast Distributed Algorithms for Computing Separable Functions. *IEEE Trans. Information Theory*, 54(7):2997–3007, 2008.
- [18] K. Panagiotou, X. Pérez-Giménez, T. Sauerwald, and H. Sun. Randomized Rumour Spreading: The Effect of the Network Topology. *Combinatorics, Probability & Computing*, 24(2):457–479, 2015.
- [19] C. Patsonakis and M. Roussopoulos. Revisiting Asynchronous Rumor Spreading in the Blockchain Era. In *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 284–293, Dec 2019.
- [20] B. Pittel. On Spreading a Rumor. *SIAM J. Appl. Math.*, 47(1):213–223, 1987.
- [21] G. Pólya. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift*, 8(3):171–181, 1920.
- [22] T. Sauerwald and A. Stauffer. Rumor Spreading and Vertex Expansion on Regular Graphs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 462–475, 2011.
- [23] R. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley, 2009.

## A Existence of Subsequence

Let  $x, y \in [0, 1]$ . In this section we show that there is an unbounded sequence of natural numbers  $(n_i)_{i \in \mathbb{N}}$  such that  $\log_2 n_i - \lfloor \log_2 n_i \rfloor \rightarrow x$  and  $\ln n_i - \lfloor \ln n_i \rfloor \rightarrow y$  as  $i \rightarrow \infty$ . To this end, set  $z = y - x \ln 2$ . According to a Theorem of Kronecker, see e.g. [16, Thm. 440], for all  $i \in \mathbb{N}$ , there are  $p_i, q_i \in \mathbb{N}$  such that

$$|q_i \ln 2 - p_i - z| \leq i^{-1}. \quad (\text{A.1})$$

Actually even more is true: there are infinitely many  $p_i, q_i \in \mathbb{N}$  that solve (A.1). To see this, assume that there are only finitely many, then there is  $k, \ell \in \mathbb{N}$  such that  $k \ln 2 = \ell + z$ , otherwise there would be some  $i \in \mathbb{N}$  where (A.1) has no solution. However, according to a Theorem of Hurwitz, see e.g. [16, Thm. 193], there are infinitely many  $r_j, s_j \in \mathbb{N}$  such that

$$|r_j \ln 2 - s_j| \leq r_j^{-2}.$$

But then

$$|r_j \ln 2 - s_j| = |(r_j + k) \ln 2 - (s_j + \ell) - z| \leq r_j^{-2},$$

a contradiction, thus there are infinitely many solutions to (A.1). We continue with that equation, which we can restate, as  $i \rightarrow \infty$ ,

$$q_i \ln 2 + x \ln 2 = p_i + y + O(i^{-1}).$$

Taking the exponential on both sides thus yields, as  $i \rightarrow \infty$ ,

$$2^{q_i + x} = e^{p_i + y + O(i^{-1})}.$$

Set  $n_i = \lfloor 2^{q_i + x} \rfloor$  for all  $i \in \mathbb{N}$ , where we choose  $q_i$  such that  $q_i \geq i$  from the infinitely many solutions to (A.1). Then  $n_i \in \mathbb{N}$  for all  $i \in \mathbb{N}$  and

$$\log_2 n_i - \lfloor \log_2 n_i \rfloor = x + O(2^{-i}) \quad \text{as well as} \quad \ln n_i - \lfloor \ln n_i \rfloor = y + O(i^{-1}).$$

Thus the subsequence of natural numbers that is induced by  $\log_2 n_i - \lfloor \log_2 n_i \rfloor \rightarrow x$  and  $\ln n_i - \lfloor \ln n_i \rfloor \rightarrow y$  is non-empty and unbounded.