STOCHASTIC COALESCENCE IN LOGARITHMIC TIME

PO-SHEN LOH AND EYAL LUBETZKY

Abstract. The following distributed coalescence protocol was introduced by Dahlia Malkhi in 2006 motivated by applications in social networking. Initially there are n agents wishing to coalesce into one cluster via a decentralized stochastic process, where each round is as follows: Every cluster flips a fair coin to dictate whether it is to issue or accept requests in this round. Issuing a request amounts to contacting a cluster randomly chosen *proportionally to its size*. A cluster accepting requests is to select an incoming one *uniformly* (if there are such) and merge with that cluster. Empirical results by Fernandess and Malkhi suggested the protocol concludes in $O(\log n)$ rounds with high probability, whereas numerical estimates by Oded Schramm, based on an ingenious analytic approximation, suggested that the coalescence time should be super-logarithmic.

Our contribution is a rigorous study of the stochastic coalescence process with two consequences. First, we confirm that the above process indeed requires super-logarithmic time w.h.p., where the inefficient rounds are due to oversized clusters that occasionally develop. Second, we remedy this by showing that a simple modification produces an essentially optimal distributed protocol: If clusters favor their *smallest* incoming merge request then the process *does* terminate in $O(\log n)$ rounds w.h.p., and simulations show that the new protocol readily outperforms the original one. Our upper bound hinges on a potential function involving the logarithm of the number of clusters and the cluster-susceptibility, carefully chosen to form a supermartingale. The analysis of the lower bound builds upon the novel approach of Schramm which may find additional applications: Rather than seeking a single parameter that controls the system behavior, instead one approximates the system by the Laplace transform of the entire cluster-size distribution.

1. Introduction

The following stochastic distributed coalescence protocol was proposed by Dahlia Malkhi in 2006, motivated by applications in social networking and the reliable formation of peer-to-peer networks (see [10] for more on these applications). The objective is to coalesce n participating agents into a single hierarchal cluster reliably and efficiently. To do so without relying on a centralized authority, the protocol first identifies each agent as a cluster (a singleton), and then proceeds in rounds as follows:

(1) Each cluster flips a fair coin to determine whether it will be *issuing* a merge-request or *accepting* requests in the upcoming round.
(2) Issuing a request amounts to selecting another cluster randomly *proportionally to its size*.
(3) Accepting requests amounts to choosing an incoming request (if there are any) *uniformly* at random and proceeding to merge with that cluster.

In practice, each cluster is in fact a layered tree whose root is entrusted with running the protocol, e.g. each root decides whether to issue/accept requests in a given round etc. When attempting to merge with another cluster, the root of cluster $\mathcal{C}_i$ simply chooses a vertex v uniformly out of $[n]$, which then propagates the request to its root. This therefore corresponds to choosing the cluster $\mathcal{C}_j$ proportionally to $|\mathcal{C}_j|$. This part of the protocol is well-justified by the fact that agents within a cluster typically have no information on the structure of other clusters in the system.

A second feature of the protocol is the symmetry between the roles of issuing/accepting requests played by the clusters. Clearly, every protocol enjoying this feature would have (roughly) at most half of its clusters become acceptors in any given round, and as such could terminate within $O(\log n)$ rounds. Furthermore, on an intuitive level, as long as all clusters are of roughly the same size (as is the case initially), there are few "collisions" (multiple clusters issuing a request to the same cluster) each round and hence the effect of a round is similar to that of merging clusters according to a random perfect matching. As such, one might expect that the protocol should conclude with a roughly balanced binary tree in logarithmic time.

Indeed, empirical evidence by Fernandess and Malkhi [11] showed that this protocol seems highly efficient, typically taking a logarithmic number of rounds to coalesce. However, rigorous performance guarantees for the protocol were not available.

While there are numerous examples of stochastic processes that have been successfully analyzed by means of identifying a single tractable parameter that controls their behavior, here it appears that the entire distribution of the cluster-sizes plays an essential role in the behavior of the system. Demonstrating this is the following example: Suppose that the cluster $\mathcal{C}_1$ has size $n - o(\sqrt{n})$ while all others are singletons. In this case it is easy to see that with high probability all of the merge-requests will be issued to $\mathcal{C}_1$, who will accept at most one of them (we say an event holds *with high probability*, or w.h.p. for brevity, if its probability tends to 1 as $n \to \infty$). Therefore, starting from this configuration, coalescence will take at least $n^{1/2-o(1)}$ rounds w.h.p., a polynomial slowdown. Of course, this scenario is extremely unlikely to arise when starting from n individual agents, yet possibly other mildly unbalanced configurations *are* likely to occur and slow the process down.

In 2007, Oded Schramm proposed a novel approach to the problem, approximately reducing it to an analytic problem of determining the asymptotics of a recursively defined family of real functions. Via this approximation framework Schramm then gave numerical estimates suggesting that the running time of the stochastic coalescence protocol is w.h.p. super-logarithmic. Unfortunately, the analytical problem itself seemed highly nontrivial and overall no bounds for the process were known.

1.1. **New results.** In this work we study the stochastic coalescence process with two main consequences. First, we provide a rigorous lower bound confirming that this process w.h.p. requires a super-logarithmic number of rounds to terminate. Second, we identify the vulnerability in the protocol, namely the choice of which merge-request a cluster should approve: While the original choice seems promising in order to maintain the balance between clusters, it turns out that typical deviations in cluster-sizes are likely to be amplified by this rule and lead to irreparably unbalanced configurations. On the other hand, we show that a simple modification of this rule to favor the smallest incoming request is already enough to guarantee coalescence in $O(\log n)$ rounds w.h.p. (Here and in what follows we let $f \lesssim g$ denote that $f = O(g)$ while $f \asymp g$ is short for $f \lesssim g \lesssim f$.)

Theorem 1. *The uniform coalescence process $\mathcal{U}$ coalesces in $\tau_c(\mathcal{U}) \gtrsim \log n \cdot \frac{\log\log n}{\log\log\log n}$ rounds w.h.p. Consider a modified* size-biased *process $\mathcal{S}$ where every accepting cluster $\mathcal{C}_i$ has the following rule:*

- *Ignore requests from clusters of size larger than $|\mathcal{C}_i|$.*
- *Among other requests (if any) select one issued by a cluster $\mathcal{C}_j$ of smallest size.*

Then the coalescence time of the size-biased process satisfies $\tau_c(\mathcal{S}) \asymp \log n$ w.h.p.

Observe that the new protocol is easy to implement efficiently in practice as each root can keep track of the size of its cluster and can thus include it as part of the merge-request.

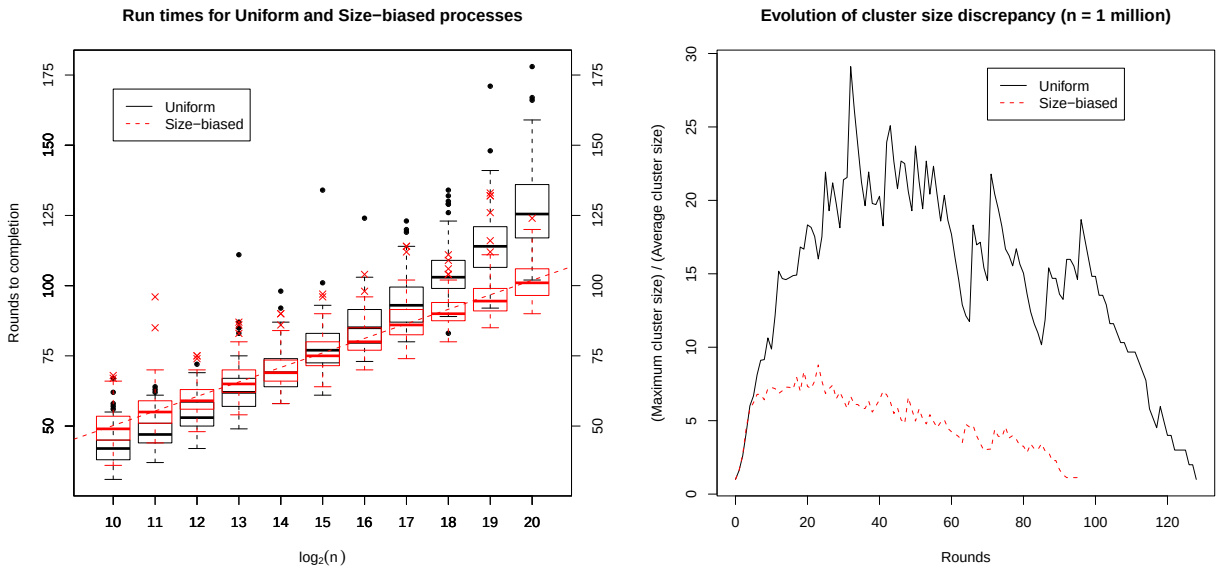

FIGURE 1. The left plot compares the running times for the two processes. Statistics are derived from 100 independent runs of each process, for each $n \in \{1024, 2048, \ldots, 2^{20}\}$. The right plot tracks the ratio between the maximum and average cluster-sizes, through a single run of each process, for $n = 10^6$. There, the uniform process took 128 rounds, while the size-biased process finished in 96.

1.2. **Empirical results.** Our simulations show that the running time of the size-biased process is approximately $5 \log_2 n$. Moreover, they further demonstrate that the new size-biased process empirically performs substantially better than the uniform process even for fairly small values of n, i.e. the improvement appears not only asymptotically in the limit but already for ordinary input sizes. These results are summarized in Figure 1, where the plot on the left clearly shows how the uniform process diverges from the linear (in logarithmic scale) trend corresponding to the runtime of the size-biased process. The rightmost plot identifies the crux of the matter: the uniform process rapidly produces a highly skewed cluster-size distribution, which slows it down considerably.

1.3. **Related work.** There is extensive literature on stochastic coalescence processes whose various flavors fit the following scheme: The clusters act via a continuous-time process where the coalescence rate of two clusters with given masses x, y (which can be either discrete or continuous) is dictated up to re-scaling by a rate kernel K. A notable example of this is Kingman's coalescent [18], which corresponds to the kernel $K(x, y) = 1$ and has been intensively studied in mathematical population genetics; see e.g. [7] for more on Kingman's coalescent and its applications in genetics. Other rate kernels that have been thoroughly studied include the additive coalescent $K(x, y) = x + y$ which corresponds to Aldous's continuum random tree [1], and the multiplicative coalescent $K(x, y) = xy$ that corresponds to Erdős-Rényi random graphs [9] (see the books [4, 17]). For further information on these as well as other coalescence processes, whose applications range from physics to chemistry to biology, we refer the reader to the excellent survey of Aldous [2].

A major difference between the classical stochastic coalescence processes mentioned above and those studied in this work is the synchronous nature of the latter ones: Instead of individual merges whose occurrences are governed by independent exponentials, here the process is comprised of rounds where all clusters act simultaneously and the outcome of a round (multiple disjoint merges)

is a function of these combined actions. This framework introduces delicate dependencies between the clusters, and rather than having the coalescence rate of two clusters be given by the rate kernel K as a function of their masses, here it is a function of the entire cluster distribution. For instance, suppose nearly all of the mass is in one cluster $\mathcal{C}_i$ (which thus attracts almost all merge requests); its coalescence rate with a given cluster $\mathcal{C}_j$ in the uniform coalescence process $\mathcal{U}$ clearly depends on the total number of clusters at that given moment, and similarly in the size-biased coalescence process $\mathcal{S}$ it depends on the sizes of all other clusters, viewed as competing with $\mathcal{C}_j$ over this merge. In face of these mentioned dependencies, the task of analyzing the evolution of the clusters along the high-dimensional stochastic processes $\mathcal{U}$ and $\mathcal{S}$ becomes highly nontrivial.

In terms of applications and related work in Computer Science, the processes studied here have similar flavor to those which arose in the 1980's, most notably the Random Mate algorithm introduced by Reif, and used by Gazit [15] for parallel graph components and by Miller and Reif [20] for parallel tree contraction. However, as opposed to the setting of those algorithms, a key difference here is the fact that as the process evolves through time each cluster is oblivious to the distribution of its peers at any given round (including the total number of clusters for that matter). Therefore for instance it is impossible for a cluster to sample from the uniform distribution over the other clusters when issuing its merge request.

For another related line of works in Computer Science, recall that the coalescence processes studied in this work organize n agents in a hierarchic tree, where each merged cluster reports to its acceptor cluster. This is closely related to the rich and intensively studied topic of Randomized Leader Elections (see e.g. [6, 12, 22, 23, 27]), where a computer network comprised of n processors attempts to single out a leader (in charge of communication, etc.) by means of a distributed randomized process generating the hierarchic tree. Finally, studying the dynamics of randomly merging sets is also fundamental to understanding the average-case performance of disjoint-set data structures (see e.g. the works of Bollobás and Simon [5], Knuth and Schönhage [19] and Yao [26]). These structures, which are of fundamental importance in computer science, store collections of disjoint sets, and support two operations: (i) taking the union of a pair of sets, and (ii) determining which set a particular element is in. See e.g. [14] for a survey of these data structures. The processes studied here precisely consider the evolution of a collection of disjoint sets under random merge operations, and it is plausible that the tools used here could contribute to advances in that area.

1.4. **Main techniques.** As we mentioned above, the main obstacle in the coalescence processes studied here is that since requests go to other clusters with probability proportional to their size, the largest clusters can create a bottleneck, absorbing all requests yet each granting only one per round. An intuitive approach for analyzing the size-biased process $\mathcal{S}$ would be to track a statistic that would warn against this scenario, with the most obvious candidate being the size of the largest cluster. However, simulations indicate that this alone will be insufficient as the largest cluster does in fact grow out of proportion in typical runs of the process. Nevertheless, the distribution of large clusters turns out to be sparse. The key idea is then to track a smoother parameter involving the *susceptibility*, which is essentially the second moment of the cluster-size distribution.

To simplify notation normalize the cluster-sizes w_i to sum to 1 so that the initial distribution consists of n clusters of size $\frac{1}{n}$ each. With this normalization, the susceptibility χ_t is defined as $\sum_i w_i^2$, the sum of squares of cluster-sizes after the t-th round. (We note in passing that this parameter has played a central role in the study of the phase-transition in Percolation and Random Graphs, see e.g. [16, 25].) The proof that the size-biased protocol is optimal hinges on a carefully

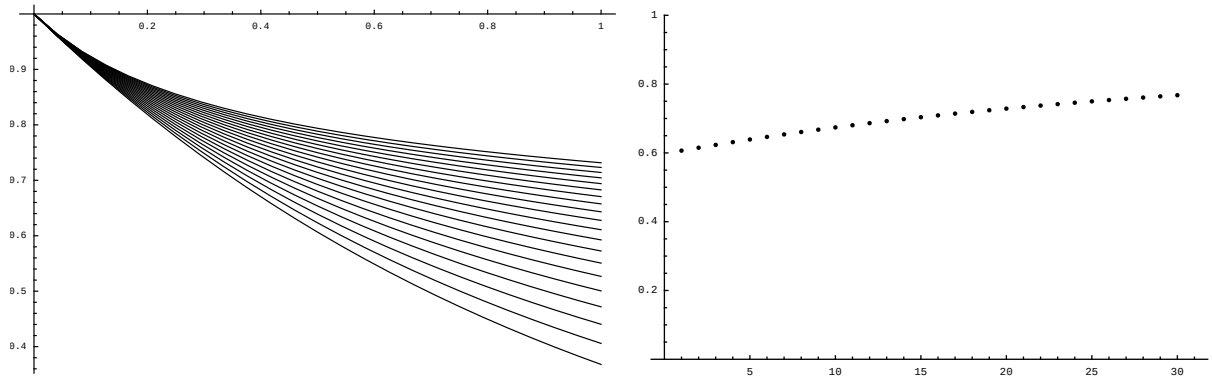

FIGURE 2. Numerical estimations by Oded Schramm for the functions $G_t(s)$ from his analytic approximation of the uniform coalescence process. The left plot features $G_t(s)$ for $t = \{0, 2, \ldots, 40\}$ and $s \in [0, 1]$ and demonstrates how these increase with t. The right plot focuses on $G_t(\frac{1}{2})$ and suggests that $G_t(\frac{1}{2}) \to 1$ and that in turn the coalescence rate should be super-logarithmic.

chosen potential function $\Phi_t = \chi_t \kappa_t + C \log \kappa_t$, where κ_t denotes the number of clusters after the t-th round and C is an absolute constant chosen to turn Φ_t into a supermartingale. In Sections 3 and 4 we will control the evolution of Φ_t and prove our upper bound on the running time of the size-biased process.

The analysis of the uniform process $\mathcal{U}$ is delicate and relies on rigorizing and analyzing the novel framework of Schramm [24] for approximating the problem by an analytic one. We believe this technique is of independent interest and may find additional applications in the analysis of high-dimensional stochastic processes. Instead of seeking a single parameter to summarize the system behavior, one instead measures the system using the Laplace transform of the entire cluster-size distribution:

Definition 1.1. *For any integer $t \geq 0$ let $\mathcal{F}_t$ be the σ-algebra generated by the first t rounds of the process. Conditioned on $\mathcal{F}_t$, define the functions $F_t(s)$ and $G_t(s)$ on the domain $\mathbb{R}$ as follows. Let κ be the number of clusters and let $w_1, \ldots, w_\kappa$ be the normalized cluster-sizes after t rounds. Set*

$$F_t(s) = \sum_{i=1}^{\kappa} \exp(-w_i\, s)\,, \qquad G_t(s) = \frac{1}{\kappa} F_t(\kappa\, s)\,. \tag{1.1}$$

As we will further explain in Section 2, the Laplace transform F_t simultaneously captures all the moments of the cluster-size distribution, in a manner analogous to the moment generating function of a random variable. This form is particularly useful in our application as we will see in Section 5 that the specific evaluation $G_t\big(\frac{1}{2}\big)$ governs the expected coalescence rate. Furthermore, it turns out that it is possible to estimate values of F_t (and G_t) recursively. Although the resulting recursion is nonstandard and highly complex, a somewhat intricate analysis eventually produces a lower bound for the uniform process.

1.5. **Organization.** The rest of this paper is organized as follows. In Section 2 we describe Schramm's analytic approach for approximating the uniform process $\mathcal{U}$. Sections 3 and 4 are devoted to the size-biased process $\mathcal{S}$: In the former we prove that $\mathbb{E}[\tau_c(\mathcal{S})] = O(\log n)$ and in the latter we build on this proof together with additional ideas to show that $\tau_c(\mathcal{S}) = O(\log n)$ w.h.p. The final section, Section 5, builds upon Schramm's aforementioned framework to produce a super-logarithmic lower bound for $\tau_c(\mathcal{U})$.

2. Schramm's analytic approximation framework for the uniform process

In this section we describe Schramm's analytic approach as it was presented in [24] for analyzing the uniform coalescence process $\mathcal{U}$, as well as the numerical evidence that Schramm obtained based on this approach suggesting that $\tau_c(\mathcal{U})$ is super-logarithmic. Throughout this section we write approximations loosely as they were sketched by Schramm and postpone any arguments on their validity (including concentration of random variables, etc.) to Section 5, where we will turn elements from this approach into a rigorous lower bound on $\tau_c(\mathcal{U})$.

Let $\mathcal{F}_t$ denote the σ-algebra generated by the first t rounds of the coalescence process $\mathcal{U}$. The starting point of Schramm's approach was to examine the following function conditioned on $\mathcal{F}_t$:

$$F_t(s) = \sum_{i=1}^{\kappa_t} \exp(-w_i\, s)\,,$$

where κ_t is the number of clusters after t rounds and $w_1, \ldots, w_{\kappa_t}$ denote the normalized cluster-sizes at that time (see Definition 1.1). The benefit that one could gain from understanding the behavior of $F_t(s)$ is obvious as $F_t(0)$ recovers the number of clusters at time t.

More interesting is the following observation of Schramm regarding the role that $F_t(\kappa_t/2)$ plays in the evolution of the clusters. Conditioned on $\mathcal{F}_t$, the probability that the cluster $\mathcal{C}_i$ receives a merge request from another cluster $\mathcal{C}_j$ is $\frac{1}{2}w_i$ (the factor $\frac{1}{2}$ accounts for the choice of $\mathcal{C}_j$ to issue rather than accept requests). Thus, the probability that $\mathcal{C}_i$ will receive *any* incoming request in round $t+1$ and independently decide to be an acceptor is

$$\frac{1}{2}\left[1-(1-w_i/2)^{\kappa_t-1}\right] \approx \frac{1}{2}\left[1-\exp(-w_i\kappa_t/2)\right]\,.$$

On this event, $\mathcal{C}_i$ will account for one merge at time $t+1$, and summing this over all clusters yields

$$\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] \approx \kappa_t - \frac{1}{2}\sum_{i=1}^{\kappa_t}\left[1-\exp(-w_i\kappa_t/2)\right] = \frac{1}{2}\left[\kappa_t + F_t(\kappa_t/2)\right]\,,$$

or equivalently, re-scaling $F_t(s)$ into $G_t(s) = (1/\kappa_t)F_t(\kappa_t s)$ as in Eq. (1.1),

$$\mathbb{E}\left[\kappa_{t+1}/\kappa_t \mid \mathcal{F}_t\right] \approx \frac{1+G_t(\frac{1}{2})}{2}\,. \tag{2.1}$$

In order to have $\tau_c(\mathcal{U}) \asymp \log n$ the number of clusters would need to typically drop by at least a constant factor at each round. This would require the ratio in (2.1) to be bounded away from 1, or equivalently, $G_t(\frac{1}{2})$ should be bounded away from 1.

Unfortunately, the evolution of the sequence $G_t(\frac{1}{2}) = (1/\kappa_t)F_t(\kappa_t/2)$ appears to be quite complex and there does not seem to be a simple way to determine its limiting behavior. Nevertheless, Schramm was able to write down an approximate recursion for the expected value of F_{t+1} in terms of multiple evaluations of F_t by observing the following: On the above event that $\mathcal{C}_i$ chooses to accept the merge request of some other cluster $\mathcal{C}_j$, by definition of the process $\mathcal{U}$ the identity of the cluster $\mathcal{C}_j$ is uniformly distributed over all $\kappa_t - 1$ clusters other than $\mathcal{C}_i$. Hence,

$$\mathbb{E}\left[F_{t+1}(s) - F_t(s) \mid \mathcal{F}_t\right] \approx \sum_i \frac{1}{2}\left(1-e^{-w_i\kappa_t/2}\right)\frac{1}{\kappa_t}\sum_{j\neq i}\left(e^{-(w_i+w_j)s} - e^{-w_i s} - e^{-w_j s}\right)\,.$$

Ignoring the fact that the last sum in the approximation skips the diagonal terms $j=i$, one arrives at a summation over all $1 \le i,j \le \kappa_t$ of exponents similar to those in the definition of F_t with an

argument of either s, $\kappa_t/2$, or $s+\kappa_t/2$, which after rearranging gives

$$\mathbb{E}[F_{t+1}(s) \mid \mathcal{F}_t] \approx \frac{1}{2}F_t(s+\kappa_t/2) + \frac{1}{2\kappa_t}F_t(s)\Big[F_t(s) + F_t(\kappa_t/2) - F_t(s+\kappa_t/2)\Big]\,.$$

To turn the above into an expression for $G_{t+1}(s)$ one needs to evaluate $F_{t+1}(\kappa_{t+1}s)$ rather than $F_{t+1}(\kappa_t s)$, to which end the approximation $\kappa_{t+1} \approx \frac{1}{2}[1+G_t(\frac{1}{2})]\kappa_t$ can be used based on (2.1). Additionally, for the starting point of the recursion, note that the initial configuration of $w_i = 1/\kappa_0$ for all $1 \le i \le \kappa_0$ has $G_0(s) = \exp(-s)$. Altogether, Schramm obtained the following deterministic analytic recurrence, whose behavior should (approximately) dictate the coalescence rate:

$$\begin{cases} g_0(s) = & \exp(-s)\,, \\ g_{t+1}(s) = & \dfrac{1}{2\alpha}\left[g_t(\alpha s)^2 - g_t(\alpha s + \frac{1}{2})g_t(\alpha s) + g_t(\alpha s + \frac{1}{2}) + g_t(\frac{1}{2})g_t(\alpha s)\right] \text{ where } \alpha = \frac{1}{2}[1+g_t(\frac{1}{2})]\,. \end{cases}$$

In light of this, aside from the task of assessing how good of an approximation the above defined functions g_t provide for the random variables G_t along the uniform coalescence process $\mathcal{U}$, the other key question is whether the sequence $g_t(\frac{1}{2})$ converges to 1 as $t \to \infty$, and if so, at what rate.

For the latter, as the complicated definition of g_{t+1} attests, analyzing the recursion of g_t seems highly nontrivial. Moreover, a naive evaluation of $g_t(\frac{1}{2})$ involves exponentially many terms, making numerical simulations already challenging. The computer-assisted numerical estimates performed by Schramm for the above recursion, shown in Figure 2, seemed to suggest that indeed $g_t(\frac{1}{2}) \to 1$ (albeit very slowly), which should lead to a super-logarithmic coalescence time for $\mathcal{U}$. However, no rigorous results were known for the limit of $g_t(\frac{1}{2})$ or its stochastic counterpart $G_t(\frac{1}{2})$.

As we show in Section 5, in order to turn Schramm's argument into a rigorous lower bound on $\tau_c(\mathcal{U})$, we move our attention away from the sought value of $G_t(\frac{1}{2})$ and focus instead on $G_t(1)$. By manipulating Schramm's recursion for G_t and combining it with additional analytic arguments and appropriate concentration inequalities, we show that as long as κ_t is large enough and $G_t(\frac{1}{2}) < 1-\delta$ for some fixed $\delta > 0$, then typically $G_{t+1}(1) > G_t(1) + \varepsilon$ for some $\varepsilon(\delta) > 0$. Since by definition $0 \le G_t(1) \le 1$ this can be used to show that ultimately $G_t(\frac{1}{2}) \to 1$ w.h.p., and a careful quantitative version of this argument produces the rigorous lower bound on $\tau_c(\mathcal{U})$ stated in Theorem 1.

3. Expected running time of the size-biased process

The goal of this section is to prove that the expected time for the size-biased process to complete has logarithmic order, as stated in Proposition 3.1. Following a few simple observations on the process we will prove this proposition using two key lemmas, Lemmas 3.4 and 3.5, whose proofs will appear in §3.2 and §3.3 respectively. In Section 4 we extend the proof of this proposition using some additional ideas to establish that the coalescence time is bounded by $O(\log n)$ w.h.p.

Proposition 3.1. *Let $\tau_c = \tau_c(\mathcal{S})$ denote the coalescence time of the size-biased process $\mathcal{S}$. Then there exists an absolute constant $C > 0$ such that $\mathbb{E}_1[\tau_c] \le C \log n$, where $\mathbb{E}_1[\cdot]$ denotes expectation w.r.t. an initial cluster distribution comprised of n singletons.*

Throughout Sections 3 and 4 we refer only to the size-biased process and use the following notation. Define the filtration $\mathcal{F}_t$ to be the σ-algebra generated by the process up to and including the t-th round. Let κ_t denote the number of clusters after the conclusion of round t, noting that with these definitions we are interested in bounding the expected value of the stopping time

$$\tau_c = \min\{t : \kappa_t = 1\}\,. \tag{3.1}$$

As mentioned in the introduction, we normalize the cluster-sizes so that they sum to 1. Finally, the susceptibility χ_t denotes the sum of squares of the cluster-sizes at the end of round t.

Observe that by Cauchy-Schwarz, if $w_1, \ldots, w_{\kappa_t}$ are the cluster-sizes at the end of round t (and as such $\chi_t = \sum_i w_i^2$) then we always have

$$\chi_t \, \kappa_t \geq \left(\sum_{i=1}^{\kappa_t} w_i \right)^2 = 1 \,, \tag{3.2}$$

with equality iff all clusters have the same size. Indeed, the susceptibility χ_t measures the variance of the cluster-size distribution. When χ_t is smaller (closer to κ_t^{-1}), the distribution is more uniform. We further claim that

$$\chi_{t+1} \leq 2\chi_t \quad \text{for all } t \,. \tag{3.3}$$

To see this, note that if a cluster of size a merges with a cluster of size b the susceptibility increases by exactly $(a+b)^2 - (a^2+b^2) = 2ab \leq a^2 + b^2$. Since each round only involves merges between disjoint pairs of clusters, this immediately implies that the total additive increase in susceptibility is bounded by the current sum of squares of the cluster sizes, i.e., the current susceptibility χ_t.

Before commencing with the proof of Proposition 3.1, we present a trivial linear bound for the expected running time of the coalescence process, which will later serve as the final step in our proof. Here and in what follows, $\mathbb{P}_w$ and $\mathbb{E}_w$ denote probability and expectation given the initial cluster distribution w. While the estimate featured here appears to be quite crude when w is uniform, recall that in general τ_c can in fact be linear in the initial number of clusters w.h.p., e.g. when w is comprised of one cluster of mass $1 - 1/\sqrt{n}$ and $\sqrt{n}$ other clusters of mass $1/n$ each.

Lemma 3.2. *Starting from κ clusters with an arbitrary cluster distribution $w = (w_1, \ldots, w_\kappa)$ we have $\mathbb{E}_w[\tau_c] \leq 8\kappa$. Furthermore, $\mathbb{P}_w(\tau_c > 16\kappa) \leq e^{-\kappa/4}$.*

Proof. Consider an arbitrary round in which at least 2 clusters still remain. We claim that the probability that there is at least one merge in this round is at least $\frac{1}{8}$. Indeed, let $\mathcal{C}_1$ be a cluster of minimal size: The probability that it decides to send a request is $\frac{1}{2}$, and since there are at least two clusters and $\mathcal{C}_1$ is the smallest one, the probability that this request goes to some $\mathcal{C}_j$ with $j \neq 1$ is at least $\frac{1}{2}$. Finally, the probability that $\mathcal{C}_j$ is accepting requests is again $\frac{1}{2}$. Conditioned on these events, $\mathcal{C}_j$ will definitely accept some request (possibly not the one from $\mathcal{C}_1$ as another cluster of the same size as $\mathcal{C}_1$ may have sent it a request) leading to at least one merge, as claimed.

The process terminates when the total cumulative number of merges reaches $\kappa - 1$. Therefore, the time of completion is stochastically dominated by the sum of $\kappa - 1$ geometric random variables with success probability $\frac{1}{8}$, and in particular $\mathbb{E}_w[\tau_c] \leq 8(\kappa - 1)$.

By the same reasoning, the total number of merges that occurred in the first t rounds clearly stochastically dominates a binomial variable $\mathrm{Bin}(t, \frac{1}{8})$ as long as $t \leq \tau_c$. Therefore,

$$\mathbb{P}_w(\tau_c > 16\kappa) \leq \mathbb{P}\left(\mathrm{Bin}(16\kappa, \tfrac{1}{8}) \leq \kappa - 1\right) \leq e^{-\kappa/4} \,,$$

where the last inequality used the well-known Chernoff bounds (see e.g. [17, Theorem 2.1]). ■

3.1. **Proof of Proposition 3.1 via two key lemmas.** We next present the two main lemmas on which the proof of the proposition hinges. The key idea is to design a potential function comprised of two parts Φ_1, Φ_2 while identifying a certain event A_t such that the following holds: $\mathbb{E}[\Phi_1(t+1) - \Phi_1(t) \mid \mathcal{F}_t \,, A_t] < c_1 < 0$ and $\mathbb{E}[\Phi_2(t+1) - \Phi_2(t) \mid \mathcal{F}_t] < c_2$ where c_1, c_2 are absolute constants, and a similar statement holds conditioned on A_t^c when reversing the roles of Φ_1 and Φ_2.

At this point we will establish that an appropriate linear combination of Φ_1, Φ_2 is a supermartingale, and the required bound on τ_c will follow from Optional Stopping. Note that throughout the proof we make no attempt to optimize the absolute constants involved. The event A_t of interest is defined as follows:

Definition 3.3. *Let A_t be the event that the following two properties hold after the t-th round:*
(i) At least $\kappa_t/2$ clusters have size at most $1/(600\kappa_t)$.
(ii) The cluster-size distribution satisfies $\sum_i w_i \mathbb{1}_{\{w_i < 41/\kappa_t\}} < 4 \cdot 10^{-5}$.

The intuition behind this definition is that Property (i) boosts the number of tiny clusters, thereby severely retarding the growth of the largest clusters, which will tend to see incoming requests from these tiny clusters. Property (ii) ensures that most of the mass of the cluster-size distribution is on relatively large clusters, of size at least 41 times the average.

Examining the event A_t will aid in tracking the variable $\chi_t \kappa_t$, the normalized susceptibility (recall from (3.2) that this quantity is always at least 1 and it equals 1 whenever all clusters are of the same size). The next lemma, whose proof appears in §3.2, estimates the expected change in this quantity and most notably shows that it is at most $-\frac{1}{200}$ if we condition on A_t.

Lemma 3.4. *Let $\Phi_1(t) = \chi_t \kappa_t$ and suppose that at the end of the t-th round one has $\kappa_t \geq 2$. Then*

$$\mathbb{E}[\Phi_1(t+1) - \Phi_1(t) \mid \mathcal{F}_t] \leq 5 \tag{3.4}$$

and furthermore

$$\mathbb{E}\left[\Phi_1(t+1) - \Phi_1(t) \mid \mathcal{F}_t\,, A_t\,, \chi_t < 3 \cdot 10^{-7}\right] \leq -\tfrac{1}{200}\,. \tag{3.5}$$

Fortunately, when A_t does not hold the behavior in the next round can still be advantageous in the sense that in this case the number of clusters tends to fall by at least a constant fraction. This is established by the following lemma, whose proof is postponed to §3.3.

Lemma 3.5. *Let $\Phi_2(t) = \log \kappa_t$ and suppose that after the t-th round one has $\kappa_t \geq 2$. Then*

$$\mathbb{E}\left[\Phi_2(t+1) - \Phi_2(t) \mid \mathcal{F}_t\,, A_t^c\right] < -2 \cdot 10^{-7}\,. \tag{3.6}$$

We are now in a position to derive Proposition 3.1 from the above two lemmas.

Proof of Proposition 3.1. Define the stopping time τ to be

$$\tau = \min\left\{i : \chi_t \geq 3 \cdot 10^{-7}\right\}\,.$$

Observe that the susceptibility is initially $1/n$, its value is 1 once the process arrives at a single cluster (i.e. at time τ_c) and until that point it is nondecreasing, hence $\mathbb{E}\tau \leq \mathbb{E}\tau_c < \infty$ by Lemma 3.2. Further define the random variable

$$Z_t = \chi_t \kappa_t + 3 \cdot 10^7 \log \kappa_t + \frac{t}{200}\,.$$

We claim that $(Z_{t\wedge\tau})$ is a supermartingale. Indeed, consider $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t\,, \tau > t]$ and note that the fact that $\tau > t$ implies in particular that $\kappa_t \geq 2$ since in that case $\chi_t < 3 \cdot 10^{-7} < 1$.

- If A_t holds then by (3.5) the conditional expected change in $\chi_t \kappa_t$ is below $-\frac{1}{200}$, while $\log \kappa_t$ can only decrease (as κ_t is non-increasing), hence $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t\,, A_t\,, \tau > t] \leq Z_t$.
- If A_t does not hold then by (3.4) the conditional expected change in $\chi_t \kappa_t$ is at most $+5$ whereas the conditional expected change in $\log \kappa_t$ is below $-2 \cdot 10^{-7}$ due to (3.6). By the scaling in the definition of Z_t these add up to give $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t\,, A_t^c\,, \tau > t] \leq Z_t - \frac{199}{200}$.

Altogether, $(Z_{t\wedge\tau})$ is indeed a supermartingale. As its increments are bounded and the stopping time τ is integrable we can apply the Optional Stopping Theorem (see e.g. [8, Chapter 5]) and get

$$\mathbb{E}Z_\tau \le Z_0 = \chi_0\kappa_0 + 3\cdot 10^7 \log\kappa_0 = O(\log n)\,. \tag{3.7}$$

At the same time, by definition of τ we have $\chi_\tau \ge 3\cdot 10^{-7}$ and so

$$Z_\tau = \chi_\tau\kappa_\tau + 3\cdot 10^7\log\kappa_\tau + \frac{\tau}{200} \ge 3\cdot 10^{-7}\left(\kappa_\tau + \tau/8\right)\,. \tag{3.8}$$

Taking expectation in (3.8) and combining it with (3.7) we find that

$$\mathbb{E}\left[\tau + 8\kappa_\tau\right] \le O(\log n)\,.$$

Finally, conditioned on the cluster distribution at time τ we know by Lemma 3.2 that the expected number of additional rounds it takes the process to conclude is at most $8\kappa_\tau$, thus $\mathbb{E}[\tau_c] \le \mathbb{E}[\tau + 8\kappa_\tau]$. We can now conclude that $\mathbb{E}[\tau_c] = O(\log n)$, as required. ■

3.2. **Proof of Lemma 3.4: Estimating the normalized susceptibility when A_t holds.** The first step in controlling the product $\chi_t\,\kappa_t$ is to quantify the coalescence rate in terms of the susceptibility, as achieved by the following claim.

Claim 3.6. *Suppose that at the end of the t-th round one has $\kappa_t \ge 2$. Then*

$$\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] \le \kappa_t - (46\chi_t)^{-1} \tag{3.9}$$

and furthermore

$$\mathbb{P}\left(\kappa_{t+1} < \kappa_t - (100\chi_t)^{-1} \mid \mathcal{F}_t\,, \chi_t < 3\cdot 10^{-7}\right) \ge 1 - e^{-100}\,.$$

Proof. To simplify the notation let $\kappa = \kappa_t$, $\chi = \chi_t$ and $\kappa' = \kappa_{t+1}$ throughout the proof of the claim. Further let the clusters $\mathcal{C}_i$ be indexed in increasing order of their sizes and let $w_i = |\mathcal{C}_i|$.

Recall that the number of merges in round $t+1$ is precisely the number of clusters which decide to accept requests and then receive at least one incoming request from a cluster of size no larger than itself. Consider the probability of the latter event for a cluster $\mathcal{C}_i$ with $i > \lfloor\kappa/2\rfloor$. Since the clusters are ordered by size there are at least $\lfloor\kappa/2\rfloor$ clusters of size at most w_i and each will send a request to $\mathcal{C}_i$ independently with probability $w_i/2$ (the factor of 2 is due to the probability of issuing rather than receiving requests this round). The probability that none of these clusters do so is thus at most $(1-w_i/2)^{\lfloor\kappa/2\rfloor} \le e^{-w_i\kappa/6}$ (where we used the fact that $\lfloor\kappa/2\rfloor \ge \kappa/3$ for any $\kappa \ge 2$), and altogether the probability that $\mathcal{C}_i$ accepts a merge request from one of these clusters is at least $\frac{1}{2}\big(1 - e^{-w_i\kappa/6}\big)$. Summing over these clusters we conclude that

$$\mathbb{E}[\kappa - \kappa' \mid \mathcal{F}_t] \ge \sum_{i>\lfloor\kappa/2\rfloor} \frac{1}{2}\left(1 - e^{-w_i\kappa/6}\right) \ge \sum_{i=1}^{\kappa} \frac{1}{4}\left(1 - e^{-w_i\kappa/6}\right)\,,$$

where the last inequality follows from the fact that the summand is increasing in w_i and hence the sum over the $\lceil\kappa/2\rceil$ largest clusters should be at least as large as the sum over the $\lfloor\kappa/2\rfloor$ smallest ones. Next, observe that by concavity, for all $0 \le w_i \le 6\chi$ the final summand is at least $w_i \cdot \frac{1}{4}\big(1 - e^{-\chi\kappa}\big)/(6\chi)$ which in turn is at least $w_i \cdot \frac{1}{4}\big(1 - e^{-1}\big)/(6\chi)$ by Eq. (3.2). As this last expression always exceeds $w_i/(38\chi)$ we get

$$\mathbb{E}[\kappa - \kappa' \mid \mathcal{F}_t] \ge \frac{1}{38\chi}\sum_{w_i \le 6\chi} w_i\,. \tag{3.10}$$

We now aim to show that much of the overall mass is spread on clusters of size at most 6χ. To this end recall that by definition $\chi = \sum w_i^2$ while $\sum_i w_i = 1$, hence we can write $\chi = \mathbb{E}Y$ where Y is the random variable that accepts the value w_i with probability w_i for $i = 1, \dots, \kappa$. This gives that

$$\sum_{w_i \le 6\chi} w_i = \mathbb{P}(Y \le 6\mathbb{E}Y) > \frac{5}{6}\,,$$

(with the final bound is due to Markov's inequality) and revisiting (3.10) we obtain that

$$\mathbb{E}[\kappa - \kappa' \mid \mathcal{F}_t] > \frac{1}{38\chi} \cdot \frac{5}{6} > \frac{1}{46\chi}\,,$$

establishing inequality (3.9).

To complete the proof of the claim it suffices to show that the random variable $X = \kappa - \kappa'$ is suitably concentrated, to which end we use Talagrand's inequality (see, e.g., [21, Chapter 10]). In its following version we say that a function $f : \prod_i \Omega_i \to \mathbb{R}$ is C-Lipschitz if changing its argument ω in any single coordinate changes $f(\omega)$ by at most C, and that f is r-certifiable if for every s and ω with $f(\omega) \ge s$ there exists a subset I of at most rs coordinates such that every ω' that agrees with ω on the coordinates indexed by I also has $f(\omega') \ge s$. In the context of a product space $\Omega = \prod_i \Omega_i$ these definitions carry to the random variable that f corresponds to via the product measure.

Theorem 3.7 (Talagrand's inequality)**.** *If X is a C-Lipschitz and r-certifiable random variable on $\Omega = \prod_{i=1}^n \Omega_i$ then $\mathbb{P}\big(|X - \mathbb{E}X| > t + 60C\sqrt{r\mathbb{E}X}\big) \le 4\exp\big(-t^2/(8C^2 r\mathbb{E}X)\big)$ for any $0 \le t \le \mathbb{E}X$.*

Observe that round $t+1$, conditioned on $\mathcal{F}_t$, is clearly a product space as the actions of the individual clusters are independent: Formally, each cluster chooses either to accept requests or to send a request to a random cluster. Changing the action of a single cluster can only affect X, the number of merges in round $t+1$, by at most one merge and so X is 1-Lipschitz. Also, if $X \ge s$ then one can identify s clusters which accepted merge requests from smaller clusters. By fixing the decisions of the $2s$ clusters comprising these merges (the acceptors together with their corresponding requesters) we must have $X \ge s$ regardless of the other clusters' actions, as the s acceptors will accept (possibly different) merge-requests no matter what. Thus, X is also 2-certifiable.

Let $\mu = \mathbb{E}X$ and assume now that $\chi < 3 \cdot 10^{-7}$. By the first part of the proof (Eq. (3.9)) it then follows that $\mu \ge (46\chi)^{-1} > 70000$, in which case Talagrand's inequality gives

$$\mathbb{P}\left(|X - \mu| > \frac{\mu}{6} + 60\sqrt{2\mu}\right) \le 4\exp\big(-(\mu/6)^2/(16\mu)\big) = 4e^{-\mu/576} < e^{-100}\,.$$

Also, note that our above bound $\mu > 70000 > 2 \cdot 180^2$ implies that

$$60\sqrt{2\mu} < \mu/3\,,$$

so in fact the probability of X falling below $\mu - \left(\frac{\mu}{6} + \frac{\mu}{3}\right)$ is at most e^{-100}. As $\mu \ge (46\chi)^{-1}$ we conclude that $\kappa - \kappa' = X > (100\chi)^{-1}$ with probability at least $1 - e^{-100}$, as required. ■

As the above claim demonstrated the effect of the susceptibility on the coalescence rate, we move to study the evolution of the susceptibility. The critical advantage of the size-biased process is that large clusters grow more slowly than small clusters. The intuition behind this is that larger clusters tend to receive more requests, and since clusters choose to accept their smallest incoming request, these clusters typically have more choices to minimize over. It turns out that this effect is enough to produce a useful quantitative bound on the growth of the susceptibility.

Claim 3.8. *Suppose that after the t-th round $\kappa_t \geq 2$. Then*

$$\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t] \leq \chi_t + \frac{5}{\kappa_t}\,. \tag{3.11}$$

Proof. Set $\kappa = \kappa_t$ and $\chi = \chi_t$. Let the clusters $\mathcal{C}_i$ be indexed in increasing order of their sizes and let $w_i = |\mathcal{C}_i|$. For each cluster $\mathcal{C}_i$ let the random variable X_i be the size of the smallest cluster that it receives a merge request from, as long as that cluster is no larger than itself, and not itself; otherwise (the case where $\mathcal{C}_i$ receives no merge requests from another cluster of size less than or equal to its own) set $X_i = 0$. Under these definitions we have

$$\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t] = \chi + \sum_{i=1}^{\kappa} w_i \mathbb{E}[X_i]\,, \tag{3.12}$$

since each $\mathcal{C}_i$ is an acceptor with probability $\frac{1}{2}$ and if it indeed accepts a request from a cluster of size X_i then the susceptibility will increase by exactly $(w_i + X_i)^2 - (w_i^2 + X_i^2) = 2w_iX_i$.

Next, note that since we ordered the clusters by increasing order of size, each of the first $\lfloor \kappa/2 \rfloor$ clusters has size at most $2/\kappa$ (otherwise the last $\lceil \kappa/2 \rceil$ clusters would combine to a total mass larger than 1). We will use this fact to bound $\mathbb{E}[X_i \mid \mathcal{F}_t]$ by considering two situations:

1. If $\mathcal{C}_i$ receives an incoming request from at least one of the first $\lfloor \kappa/2 \rfloor$ clusters (including itself) then $X_i \leq 2/\kappa$ by the above argument. The probability of this is precisely $1 - \left(1 - \frac{w_i}{2}\right)^{\lfloor \kappa/2 \rfloor}$ as each of the first $\lfloor \kappa/2 \rfloor$ clusters $\mathcal{C}_j$ independently sends a request to $\mathcal{C}_i$ with probability $w_i/2$ (with the factor of 2 due to the decision of $\mathcal{C}_j$ whether or not to issue requests).
2. If $\mathcal{C}_i$ gets no requests from the first $\lfloor \kappa/2 \rfloor$ clusters then use the trivial bound $X_i \leq w_i$.

Combining the two cases we deduce that

$$\mathbb{E}X_i \leq \left(1 - \left(1 - \frac{w_i}{2}\right)^{\lfloor \kappa/2 \rfloor}\right)\frac{2}{\kappa} + \left(1 - \frac{w_i}{2}\right)^{\lfloor \kappa/2 \rfloor} w_i\,. \tag{3.13}$$

We claim that $\mathbb{E}X_i$ is in fact always at most $5/\kappa$. To see this, first note that if $w_i \leq 2/\kappa$ then this immediately holds, e.g. since $X_i \leq w_i$. Consider therefore the case where $w_i > 2/\kappa$. Since (3.13) is a weighted average of $2/\kappa$ and $w_i > 2/\kappa$, it increases whenever the weight on w_i is increased. As

$$\left(1 - \frac{w_i}{2}\right)^{\lfloor \kappa/2 \rfloor} \leq e^{-(w_i/2)\lfloor \kappa/2 \rfloor} \leq e^{-w_i\kappa/6}\,,$$

we have that in this case

$$\mathbb{E}X_i \leq \left(1 - e^{-w_i\kappa/6}\right)\frac{2}{\kappa} + e^{-w_i\kappa/6}\, w_i \leq \frac{1}{\kappa}\left(2 + w_i\kappa\, e^{-w_i\kappa/6}\right)\,.$$

One can easily verify that the function $f(x) = xe^{-x/6}$ satisfies $f(x) \leq 3$ for all x, hence we conclude that $\mathbb{E}X_i \leq 5/\kappa$ in all cases, as claimed. Plugging this into equation (3.12) we obtain that

$$\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t] \leq \chi + \frac{5}{\kappa}\sum_{i=1}^{\kappa} w_i = \chi + \frac{5}{\kappa}\,,$$

as required. ∎

While the last claim allows us to limit the growth of the susceptibility, this bound is unfortunately too weak in general. For instance, when used in tandem with Claim 3.6 it results in the the susceptibility growing out of control, while the number of clusters decreases slower and slower. Crucially however, conditioned on the event A_t (as given in Definition 3.3) we can refine these bounds to show that the growth of χ_{t+1} slows down dramatically, as the following claim establishes.

Claim 3.9. *Suppose that at the end of the t-th round $\kappa_t \geq 2$. Then*

$$\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t, A_t] \leq \chi_t + (201\kappa_t)^{-1}\,. \tag{3.14}$$

Proof. Let $\kappa = \kappa_t$ and $\chi = \chi_t$, and define the random variables X_i as in the proof of Claim 3.8. By the same reasoning used to deduce inequality (3.13), only now using Property (i) of A_t according to which each of the smallest $\lceil \kappa/2 \rceil$ clusters has size at most $1/(600\kappa_t)$, we have

$$\mathbb{E}X_i \leq \left(1 - \left(1 - \frac{w_i}{2}\right)^{\lceil \kappa/2 \rceil}\right)\frac{1}{600\kappa} + \left(1 - \frac{w_i}{2}\right)^{\lceil \kappa/2 \rceil} w_i\,. \tag{3.15}$$

Recall that equation (3.12) established that $\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t] = \chi + \sum_{i=1}^{\kappa} w_i \mathbb{E}X_i$. This time we will need to bound this sum more delicately by splitting it into two parts based on whether or not $w_i < 41/\kappa$. In the case $w_i < 41/\kappa$ we can use the trivial bound $X_i \leq w_i$ to arrive at

$$\sum_i w_i \mathbb{1}_{\{w_i < 41/\kappa\}} \mathbb{E}X_i < \sum_i w_i \mathbb{1}_{\{w_i < 41/\kappa\}} \frac{41}{\kappa} < 4 \cdot 10^{-5} \cdot \frac{41}{\kappa}\,,$$

where the last inequality is by Property (ii) of A_t. For the second part of the summation we use the same weighted mean argument from the proof of Claim 3.8 to deduce that when $w_i > (600\kappa)^{-1}$ the r.h.s. of (3.15) increases with the weight on w_i, which in turn is at most $\left(1 - \frac{w_i}{2}\right)^{\lceil \kappa/2 \rceil} \leq \exp(-w_i\kappa/4)$. In particular, in case $w_i \geq 41/\kappa$ we have

$$\mathbb{E}X_i \leq \left(1 - e^{-w_i\kappa/4}\right)\frac{1}{600\kappa} + e^{-w_i\kappa/4} w_i \leq \frac{1}{\kappa}\left(\frac{1}{600} + w_i\kappa\, e^{-w_i\kappa/4}\right) \leq \frac{1}{\kappa}\left(\frac{1}{600} + 41 e^{-41/4}\right)$$

(here we used the fact that the function $xe^{-x/4}$ is decreasing for $x \geq 41$). Combining our bounds,

$$\sum_{i=1}^{\kappa} w_i \mathbb{E}X_i \leq \frac{1}{\kappa}\left(4 \cdot 10^{-5} \cdot 41 + \sum_i w_i \mathbb{1}_{\{w_i \geq 41/\kappa\}}\left(\frac{1}{600} + 41 e^{-41/4}\right)\right) < \frac{1}{201\kappa}\,,$$

since $\sum_i w_i = 1$. Together with equation (3.12) this completes the proof. ∎

Combining the bound on κ_{t+1} in Claim 3.6 with the bounds on χ_{t+1} from Claims 3.8 and 3.9 will now result in the statement of Lemma 3.4.

Proof of Lemma 3.4. For convenience let $\kappa = \kappa_t$ and $\chi = \chi_t$, as well as $\kappa' = \kappa_{t+1}$ and $\chi' = \chi_{t+1}$. The first statement of the lemma is an immediate consequence of Claim 3.8, since $\kappa' \leq \kappa$ and so

$$\mathbb{E}[\chi'\kappa' \mid \mathcal{F}_t] \leq \kappa\, \mathbb{E}[\chi' \mid \mathcal{F}_t] \leq \kappa\left(\chi + \frac{5}{\kappa}\right) = \chi\kappa + 5\,.$$

For the second statement, since we can break down $\chi'\kappa'$ into

$$\chi'\kappa' = \chi'\left(\kappa - \frac{1}{100\chi}\right) + \chi'\left(\kappa' - \kappa + \frac{1}{100\chi}\right)\mathbb{1}_{\{\kappa' \geq \kappa - \frac{1}{100\chi}\}} + \chi'\left(\kappa' - \kappa + \frac{1}{100\chi}\right)\mathbb{1}_{\{\kappa' < \kappa - \frac{1}{100\chi}\}}\,,$$

noticing that the last expression in the r.h.s. is at most 0, and recalling that $0 < \chi \leq \chi' \leq 2\chi$ (due to Eq. (3.3)) and $1 \leq \kappa' \leq \kappa$, we now obtain that $\mathbb{E}\left[\chi'\kappa' \mid \mathcal{F}_t\,, A_t\,, \chi < 3 \cdot 10^{-7}\right]$ is at most

$$\begin{aligned}
&\mathbb{E}\left[\chi'\left(\kappa - \frac{1}{100\chi}\right) \,\middle|\, \mathcal{F}_t\,, A_t\,, \chi < 3 \cdot 10^{-7}\right] + \mathbb{E}\left[2\chi \cdot \frac{1}{100\chi}\mathbb{1}_{\{\kappa' \geq \kappa - \frac{1}{100\chi}\}} \,\middle|\, \mathcal{F}_t\,, A_t\,, \chi < 3 \cdot 10^{-7}\right] \\
&= \left(\kappa - \frac{1}{100\chi}\right)\mathbb{E}\left[\chi' \,\middle|\, \mathcal{F}_t\,, A_t\,, \chi < 3 \cdot 10^{-7}\right] + \frac{1}{50}\mathbb{P}\left(\kappa' \geq \kappa - \frac{1}{100\chi} \,\middle|\, \mathcal{F}_t\,, A_t\,, \chi < 3 \cdot 10^{-7}\right).
\end{aligned}$$

Applying Claims 3.6 and 3.9 now gives

$$\mathbb{E}\left[\chi'\kappa' \mid \mathcal{F}_t\,, A_t\,, \chi < 3\cdot 10^{-7}\right] \le \left(\kappa - \frac{1}{100\chi}\right)\left(\chi + \frac{1}{201\kappa}\right) + \frac{1}{50}\,e^{-100}$$
$$< \chi\kappa - \frac{1}{100} + \frac{1}{201} + \frac{1}{50}e^{-100} < \chi\kappa - \frac{1}{200}\,,$$

completing the proof. ■

3.3. **Proof of Lemma 3.5: Estimating the number of components when A_t fails.** We wish to show that whenever either one of the two properties specified in A_t does not hold, the expected number of clusters drops by a constant factor.

Suppose that Property (i) of A_t fails: In this case a constant fraction of the clusters have size which is at least a constant fraction of the average size $1/\kappa_t$. We will show that each such cluster receives an incoming request (from another cluster of no larger size) in the next round with a probability that is uniformly bounded from below. Consequently, we will be able to conclude that the number of clusters shrinks by at least a constant factor in expectation.

Claim 3.10. *Suppose that at the end of the t-th round $\kappa_t \ge 2$ and Property (i) of A_t does not hold, i.e. more than $\kappa_t/2$ clusters have size greater than $(600\kappa_t)^{-1}$. Then*

$$\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] \le (1 - 5\cdot 10^{-5})\,\kappa_t\,. \tag{3.16}$$

Proof. Let $\kappa = \kappa_t$ and $\kappa' = \kappa_{t+1}$ and as usual, order the clusters by increasing order of size. Consider an arbitrary cluster $\mathcal{C}_i$ which is one of the last $\lceil \kappa/2 \rceil$ clusters, and let w_i denote its size. If $\mathcal{C}_i$ opts to accept requests in this round (with probability $\frac{1}{2}$) and any of the first $\lfloor \kappa/2 \rfloor$ clusters sends it a request, it will contribute a merge in this round. This occurs with probability

$$\frac{1}{2}\left(1 - \left(1 - \frac{w_i}{2}\right)^{\lfloor \kappa/2 \rfloor}\right) \ge \frac{1}{2}\left(1 - e^{-w_i\kappa/6}\right) > \frac{1}{2}(1 - e^{-1/3600}) > 10^{-4}\,,$$

where we used our assumption that $w_i \ge (600\kappa)^{-1}$. Thus, the probability that $\mathcal{C}_i$ contributes to a merge is at least 10^{-4}. We conclude that the expected number of merges in this round is at least $10^{-4}\lceil \kappa/2 \rceil$, from which the desired result follows. ■

Now suppose that Property (ii) of A_t fails: Here at least a constant proportion of the mass of the cluster-size distribution falls on clusters with size at most a constant multiple of the average size. Such clusters behave nicely as in this window the relation between the cluster-size and the typical number of incoming requests can be bounded by a linear function. Again, this will result in a constant proportion of clusters merging in the next round in expectation.

Claim 3.11. *Suppose that at the end of the t-th round $\kappa_t \ge 2$ and Property (ii) of A_t does not hold, i.e. $\sum_i w_i \mathbb{1}_{\{w_i < 41/\kappa_t\}} \ge 4\cdot 10^{-5}$ where w_i denotes the size of $\mathcal{C}_i$. Then*

$$\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] \le \left(1 - 2\cdot 10^{-7}\right)\kappa_t\,. \tag{3.17}$$

Proof. Let $\kappa = \kappa_t$ and $\kappa' = \kappa_{t+1}$. Order the clusters by size and let r be the number of clusters which are smaller than $41/\kappa$. Since clearly at most $\kappa/41$ clusters can have size at least $41/\kappa$ we have $r \ge \left\lceil \frac{40}{41}\kappa \right\rceil$. Notice that since $\kappa \ge 2$ this implies that in particular $\lfloor r/2 \rfloor \ge \kappa/3$. By the same arguments as before, each cluster $\mathcal{C}_i$ with $\lfloor r/2 \rfloor < i \le r$ will accept a merge request from a smaller cluster with probability at least

$$\frac{1}{2}\left(1 - \left(1 - \frac{w_i}{2}\right)^{\lfloor r/2 \rfloor}\right) \ge \frac{1}{2}\left(1 - e^{-(w_i/2)\lfloor r/2 \rfloor}\right) \ge \frac{1}{2}\left(1 - e^{-w_i\kappa/6}\right)\,.$$

Since we are concentrating our attention on the clusters of size $w_i < 41/\kappa$, concavity implies that the last expression is actually at least

$$\frac{1}{2}\left(1-e^{-41/6}\right)\frac{w_i}{41/\kappa} > \frac{w_i\kappa}{100}\,.$$

We conclude that the expected number of merges in this round is at least

$$\sum_{i=\lfloor r/2\rfloor+1}^{r}\frac{w_i\kappa}{100} \geq \frac{\kappa}{100}\cdot\frac{1}{2}\sum_{i=1}^{r} w_i \geq \frac{\kappa}{100}\cdot\frac{1}{2}\cdot 4\cdot 10^{-5} = 2\cdot 10^{-7}\,\kappa\,,$$

where we used the fact that the w_i's are sorted in increasing order to relate the sum over the cluster indices $\lfloor r/2\rfloor+1,\ldots,r$ to the one over the first r clusters. This gives the desired result. ■

Proof of Lemma 3.5. The proof readily follows from the combination of Claims 3.10 and 3.11. Indeed, these claims establish that whenever the event A_t fails we have

$$\mathbb{E}\left[\kappa_{t+1} \mid \mathcal{F}_t\,, A_t^c\right] \leq \left(1-2\cdot 10^{-7}\right)\kappa_t\,.$$

Therefore, by the concavity of the logarithm, Jensen's inequality implies that

$$\mathbb{E}[\log\kappa_{t+1} \mid \mathcal{F}_t\,, A_t^c] \leq \log\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t\,, A_t^c] \leq \log\kappa_t + \log(1-2\cdot 10^{-7}) < \log\kappa_t - 2\cdot 10^{-7}\,,$$

as required. ■

4. Optimal upper bound for size-biased process

We now prove the upper bound in Theorem 1 by building upon the ideas of the previous section. Recall that in the proof of Proposition 3.1 we defined the sequence

$$Z_t = \chi_t\kappa_t + M\log\kappa_t + \frac{t}{200} \qquad \text{where } M = 3\cdot 10^7\,,$$

established that it was a supermartingale and derived the required result from Optional Stopping. That approach was only enough to produce a bound on $\mathbb{E}[\tau_c]$, the expected completion time. For the stronger result on the typical value of τ_c we will analyze (Z_t) more delicately. Namely, we estimate its increments in $\mathcal{L}^2$ to qualify an application of an appropriate Bernstein-Kolmogorov large-deviation inequality for supermartingales due to Freedman [13].

An important element in our proof is the modification of the above given variable Z_t into an overestimate Y_t which allows far better control over the increments in $\mathcal{L}^2$. This is defined as follows:

$$Y_0 = Z_0 = \chi_0\kappa_0 + M\log\kappa_0 = 1 + M\log n\,, \tag{4.1}$$

$$Y_{t+1} = \begin{cases} Y_t + \left(\Xi_{t+1}\wedge\log^{2/3} n\right) + M\log\frac{\kappa_{t+1}}{\kappa_t} + \frac{1}{200} & \text{if } \tau_c > t\,,\\ Y_t & \text{if } \tau_c \leq t\,,\end{cases}$$

where

$$\Xi_{t+1} = \chi_{t+1}\Big(\kappa_{t+1}\vee\Big(\kappa_t - \frac{1}{\chi_t}\Big)\Big) - \chi_t\kappa_t\,.$$

The purpose of the $\left(\kappa_t - \frac{1}{\chi_t}\right)$ term is to limit the potential decrease from negative Ξ. In this section, we will need two-sided estimates (in addition to one-sided bounds such as those used in the previous section) due to the fact that we must control the $\mathcal{L}^2$ increments.

It is clear that $Y_{t+1} - Y_t \geq Z_{t+1} - Z_t$ as long as $t < \tau_c$ and $\Xi_{t+1} \leq \log^{2/3} n$. Therefore, setting

$$\bar\tau = \min\{t : \Xi_{t+1} > \log^{2/3} n\}\,,$$

it follows that

$$Y_t \geq Z_t \quad \text{for all } t \leq \tau_c \wedge \bar{\tau}. \tag{4.2}$$

In what follows we will establish a large deviation estimate for (Y_t), then use this overestimate for Z_t to show that w.h.p. $\tau_c = O(\log n)$. We thus focus our attention on the sequence (Y_t).

Lemma 4.1. *The sequence (Y_t) is a supermartingale.*

Proof. Since by definition $Y_t = Y_{t\wedge\tau_c}$ it suffices to consider the times $t < \tau_c$. As we clearly have $\big(\kappa_{t+1} \vee \big(\kappa_t - \frac{1}{\chi_t}\big)\big) \leq \kappa_t$ and Claim 3.8 established that $\mathbb{E}[\chi_{t+1} \mid \mathcal{F}_t] \leq \chi_t + \frac{5}{\kappa_t}$ we can deduce that

$$\mathbb{E}[\Xi_{t+1} \mid \mathcal{F}_t] \leq 5\,. \tag{4.3}$$

Combined with Lemma 3.5 as in the proof of Proposition 3.1, it then follows that

$$\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t\,, A_t^c] \leq 0\,.$$

We turn to consider $\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t\,, A_t]$. Since $\kappa_{t+1} \leq \kappa_t$ holds for all t, it suffices to show that

$$\mathbb{E}[\Xi_{t+1} \mid \mathcal{F}_t\,, A_t] \leq -\frac{1}{200}\,.$$

Indeed, as in the proof of Lemma 3.4 we write

$$\begin{aligned}\Xi_{t+1} &\leq \chi_{t+1}\Big(\kappa_t - \frac{1}{100\chi_t}\Big) + \chi_{t+1}\Big[\Big(\kappa_{t+1} \vee \Big(\kappa_t - \frac{1}{\chi_t}\Big)\Big) - \kappa_t + \frac{1}{100\chi_t}\Big]\mathbb{1}_{\{\kappa_{t+1} \geq \kappa_t - \frac{1}{100\chi_t}\}} - \chi_t\kappa_t \\ &\leq \chi_{t+1}\Big(\kappa_t - \frac{1}{100\chi_t}\Big) + 2\chi_t \cdot \frac{1}{100\chi_t}\mathbb{1}_{\{\kappa_{t+1} \geq \kappa_t - \frac{1}{100\chi_t}\}} - \chi_t\kappa_t\,,\end{aligned}$$

which as before gives rise to the following:

$$\mathbb{E}[\Xi_{t+1} \mid \mathcal{F}_t,\, A_t] < -\frac{1}{100} + \frac{1}{201} + \frac{1}{50}\,e^{-100} < -\frac{1}{200}\,,$$

and we conclude that (Y_t) is indeed a supermartingale, as required. ■

Lemma 4.2. *The increments of the supermartingale (Y_t) are uniformly bounded in $\mathcal{L}^2$. Namely, for every t we have $\mathbb{E}[(Y_{t+1} - Y_t)^2 \mid \mathcal{F}_t] < 2M^2$ where $M = 3 \cdot 10^7$.*

Proof. First observe that

$$(Y_{t+1} - Y_t)^2 \leq 3(\Xi_{t+1})^2 + 3\Big(M \log \frac{\kappa_{t+1}}{\kappa_t}\Big)^2 + 3\Big(\frac{1}{200}\Big)^2\,. \tag{4.4}$$

Since $\frac{1}{2}\kappa_t \leq \kappa_{t+1} \leq \kappa_t$, we have $-M\log 2 \leq M \log \frac{\kappa_{t+1}}{\kappa_t} \leq 0$, hence the last two expressions above sum to at most $\frac{3}{2}M^2$ (with room to spare) and it remains to bound $\mathbb{E}[(\Xi_{t+1})^2 \mid \mathcal{F}_t] = O(1)$ for a suitably small implicit constant.

Observe that when $\Xi_{t+1} \geq 0$ we must have $|\Xi_{t+1}| \leq \chi_{t+1}\kappa_t - \chi_t\kappa_t$ since $\big(\kappa_{t+1} \vee \big(\kappa_t - \frac{1}{\chi_t}\big)\big) \leq \kappa_t$. Conversely, if $\Xi_{t+1} \leq 0$ then necessarily $|\Xi_{t+1}| \leq \chi_t\kappa_t - \chi_{t+1}\big(\kappa_t - \frac{1}{\chi_t}\big) \leq 1$, with the last inequality due to the fact that $\kappa_t \geq 1/\chi_t$ and $\chi_{t+1} \geq \chi_t$. Combining the cases we deduce that in particular

$$|\Xi_{t+1}| \leq \kappa_t(\chi_{t+1} - \chi_t) + 1\,.$$

By Claim 3.8 we have $\mathbb{E}[\chi_{t+1} - \chi_t \mid \mathcal{F}_t] \leq 5/\kappa_t$, hence we get

$$\mathbb{E}[(\Xi_{t+1})^2 \mid \mathcal{F}_t] \leq \kappa_t^2\,\mathbb{E}[(\chi_{t+1} - \chi_t)^2 \mid \mathcal{F}_t] + 1 + 2\kappa_t(5/\kappa_t) \leq \kappa_t^2\,\mathbb{E}[(\chi_{t+1} - \chi_t)^2 \mid \mathcal{F}_t] + 11\,. \tag{4.5}$$

It remains to show that $\mathbb{E}[(\chi_{t+1}-\chi_t)^2 \mid \mathcal{F}_t] = O(1/\kappa_t^2)$. To do so, let $w_1, \ldots, w_{\kappa_t}$ be the cluster-sizes after the t-th round and recall that by Eq. (3.12) and the arguments following it we have

$$\mathbb{E}[(\chi_{t+1}-\chi_t)^2 \mid \mathcal{F}_t] = \mathbb{E}\left[\left(\sum_{i=1}^{\kappa_t} 2w_i X_i I_i\right)^2\right],$$

where each X_i is a non-negative random variable satisfying $\mathbb{E}X_i \le 5/\kappa_t$ (marking the size of another cluster of no larger size that issued a request to $\mathcal{C}_i$ or 0 if there was no such cluster) and each I_i is a Bernoulli($\frac{1}{2}$) variable independent of X_i (indicating whether or not $\mathcal{C}_i$ chose to accept requests). Since $\sum w_i = 1$, it follows from convexity that

$$\left(\sum_{i=1}^{\kappa_t} w_i X_i I_i\right)^2 \le \sum_{i=1}^{\kappa_t} w_i X_i^2 I_i\,,$$

hence, taking expectation while recalling that I_i and X_i are independent,

$$\mathbb{E}[(\chi_{t+1}-\chi_t)^2 \mid \mathcal{F}_t] \le 4\sum_{i=1}^{\kappa_t} w_i(\mathbb{E}X_i^2)\mathbb{P}(I_i) = 2\sum_{i=1}^{\kappa_t} w_i \mathbb{E}X_i^2\,,$$

and it remains to bound $\mathbb{E}X_i^2$. Following the same argument that led to (3.13) now gives

$$\mathbb{E}X_i^2 \le \left(1-\left(1-\frac{w_i}{2}\right)^{\lfloor \kappa_t/2\rfloor}\right)\left(\frac{2}{\kappa_t}\right)^2 + \left(1-\frac{w_i}{2}\right)^{\lfloor \kappa_t/2\rfloor} w_i^2\,.$$

As before, we now deduce that either $w_i \le 2/\kappa_t$, in which case clearly $\mathbb{E}X_i^2 \le 4/\kappa_t^2$, or we have

$$\mathbb{E}X_i^2 \le \left(1-e^{-w_i\kappa_t/6}\right)\frac{4}{\kappa_t^2} + e^{-w_i\kappa_t/6}w_i^2 \le \frac{1}{\kappa_t^2}\left(4+e^{-w_i\kappa_t/6}(w_i\kappa_t)^2\right)\,.$$

Since $x^2\exp(-x/6) < 20$ for all $x \ge 0$ it then follows that $\mathbb{E}X_i^2 < 24/\kappa_t^2$ (with room to spare). Either way we deduce that

$$\mathbb{E}[(\chi_{t+1}-\chi_t)^2 \mid \mathcal{F}_t] < 2\sum_i \left(w_i \cdot 24/\kappa_t^2\right) = 48/\kappa_t^2\,,$$

and so, going back to (4.5),

$$\mathbb{E}[(\Xi_{t+1})^2 \mid \mathcal{F}_t] < 48 + 11 < 60\,. \tag{4.6}$$

Using this bound in (4.4) we can conclude the proof as we have

$$\mathbb{E}[(Y_{t+1}-Y_t)^2 \mid \mathcal{F}_t] < 3\,\mathbb{E}[(\Xi_{t+1})^2 \mid \mathcal{F}_t] + \tfrac{3}{2}M^2 < 2M^2\,. \qquad \blacksquare$$

By now we have established that (Y_t) is a supermartingale which satisfies $Y_{t+1}-Y_t \le L$ for a value of $L = \log^{2/3} n + \frac{1}{200}$ and that in addition $\mathbb{E}[(Y_{t+1}-Y_t)^2 \mid \mathcal{F}_t] \le 2M^2$. We are now in a position to apply the following inequality due to Freedman [13]; we note the this result was originally stated for martingales yet its proof, essentially unmodified, extends also to supermartingales.

Theorem 4.3 ([13, Theorem 1.6])**.** *Let (S_i) be a supermartingale with respect to a filter $(\mathcal{F}_i)$. Suppose $S_i - S_{i-1} \le L$ for all i, and write $V_t = \sum_{i=1}^t \mathbb{E}[(S_i-S_{i-1})^2 \mid \mathcal{F}_{i-1}]$. Then for any $s, v > 0$,*

$$\mathbb{P}\left(\{S_t \ge S_0 + s\ ,\ V_t \le v\} \text{ for some } t\right) \le \exp\left(-\tfrac{1}{2}s^2/(v+Ls)\right)\,.$$

By the above theorem and a standard application of Optional Stopping, for any $s > 0$, integer t and stopping time τ we have $\mathbb{P}(Y_{t\wedge\tau} \geq Y_0 + s) \leq \exp\left(-\frac{1}{2}s^2/(2M^2t + Ls)\right)$. In particular, letting

$$t_0 = 500M \log n$$

and plugging $s = \log^{3/4} n$ and $\tau = \bar{\tau}$ in the last inequality we deduce that

$$\mathbb{P}(Y_{t_0\wedge\bar{\tau}} \geq Y_0 + \log^{3/4} n) \leq \exp\left(-(\tfrac{1}{2} - o(1)) \log^{1/12} n\right) = o(1)\,.$$

Hence, recalling the value of Y_0 from (4.1) we have w.h.p.

$$Y_{t_0\wedge\bar{\tau}} \leq 1 + M \log n + \log^{3/4} n \leq 2M \log n\,, \tag{4.7}$$

where the last inequality holds for sufficiently large n.

In order to compare t_0 and $\bar{\tau}$, recall from Eq. (4.3) that $\mathbb{E}[\Xi_{t+1} \mid \mathcal{F}_t] \leq 5$ whereas we established in Eq. (4.6) that $\mathbb{E}[(\Xi_{t+1})^2 \mid \mathcal{F}_t] < 60$. By Chebyshev's inequality,

$$\mathbb{P}(\Xi_{t+1} \geq \log^{2/3} n \mid \mathcal{F}_t) = O\big(\mathbb{E}\left[(\Xi_{t+1})^2 \mid \mathcal{F}_t\right] \log^{-4/3} n\big) = O\big(\log^{-4/3} n\big)\,.$$

In particular, a union bound implies that

$$\mathbb{P}(\bar{\tau} \leq t_0) = O(\log^{-1/3} n)\,.$$

Revisiting (4.7) this immediately implies that w.h.p.

$$Y_{t_0} \leq 2M \log n\,,$$

and since $Y_{t_0\wedge\bar{\tau}\wedge\tau_c} \geq Z_{t_0\wedge\bar{\tau}\wedge\tau_c}$ (due to (4.2)) we further have that w.h.p.

$$Y_{t_0\wedge\tau_c} \geq Z_{t_0\wedge\tau_c} \geq \frac{t_0 \wedge \tau_c}{200}\,.$$

Therefore, we must have $\tau_c < t_0$ w.h.p., otherwise the last two inequalities would contradict our choice of $t_0 = 500M \log n$. This concludes the proof. ■

5. Super-logarithmic lower bound for the uniform process

In this section we use the analytic approximation framework introduced by Schramm to prove the super-logarithmic lower bound stated in Theorem 1 for the coalescence time of the uniform process. Recall that a key element in this framework is the normalized Laplace transform of the cluster-size distribution, namely $G_t(s) = (1/\kappa_t)F_t(\kappa_t s)$ where $F_t(s) = \sum_{i=1}^{\kappa_t} e^{-w_i s}$ (see Definition 1.1). The following proposition, whose proof entails most of the technical difficulties in our analysis of the uniform process, demonstrates the effect of $G_t(\frac{1}{2})$ and $G_t(1)$ on the coalescence rate.

Proposition 5.1. *Let $\varepsilon_t = 1 - G_t(\frac{1}{2})$ and $\zeta_t = G_t(1)$. There exists an absolute constant $C > 0$ such that, conditioned on $\mathcal{F}_t$, with probability at least $1 - C\kappa_t^{-100}$ we have*

$$\left|\kappa_{t+1} - (1 - \varepsilon_t/2)\kappa_t\right| \leq \kappa_t^{2/3}\,, \tag{5.1}$$

$$\zeta_{t+1} \geq \zeta_t + \varepsilon_t^{13/\varepsilon_t} - 8\kappa_t^{-1/3}\,. \tag{5.2}$$

We postpone the proof of this proposition to §5.4 in favor of showing how the relations that it establishes between $\kappa_t, G_t(1), G_t(\frac{1}{2})$ can be used to derive the desired lower bound on τ_c. We claim that as long as $\kappa_t, G_t(\frac{1}{2}), G_t(1)$ satisfy Eq. (5.1),(5.2) and $t = O\big(\log n \cdot \frac{\log\log\log n}{\log\log n}\big)$ then $\kappa_t \geq n^{3/4}$; this deterministic statement is given by the following lemma:

Lemma 5.2. *Set $T = \frac{1}{75}\log n \cdot \frac{\log\log n}{\log\log\log n}$ for a sufficiently large n and let $\kappa_0, \dots, \kappa_T$ be a sequence of integers in $\{1, \dots, n\}$ with $\kappa_0 = n$. Further let ε_t and ζ_t for $t = 0, \dots, T$ be two sequences of reals in $[0, 1]$ and suppose that for all $t < T$ the three sequences satisfy inequalities (5.1) and (5.2). Then $\kappa_t > n^{3/4}$ for all $t \le T$.*

Observe that the desired lower bound on the coalescence time of the uniform process $\mathcal{U}$ is an immediate corollary of Proposition 5.1 and Lemma 5.2. Indeed, condition on the first t rounds where $0 \le t < T = \frac{1}{75}\log n \cdot \frac{\log\log n}{\log\log\log n}$ and assume $\kappa_t > n^{3/4}$. Proposition 5.1 implies that Eqs. (5.1), (5.2) hold except with probability $O(\kappa_t^{-100}) = o(n^{-1})$. On this event Lemma 5.2 yields $\kappa_{t+1} > n^{3/4}$, extending our assumption to the next round. Accumulating these probabilities for all $t < T$ now shows that $\mathbb{P}(\kappa_T > n^{3/4}) = 1 - o(T/n)$ and in particular $\tau_c > T$ w.h.p., as required.

Proof of lemma. The proof proceeds by induction: Assuming that $\kappa_i > n^{3/4}$ for all $i \le t < T$ we wish to deduce that $\kappa_{t+1} > n^{3/4}$.

Repeatedly applying Eq. (5.2) and using the induction hypothesis we find that

$$\zeta_{t+1} \ge \zeta_0 + \sum_{i=0}^{t}\left(\varepsilon_i^{13/\varepsilon_i} - 8\kappa_i^{-1/3}\right) > \sum_{i=0}^{t}\left(\varepsilon_i^{13/\varepsilon_i}\right) - 8(t+1)\left(n^{3/4}\right)^{-1/3} = \sum_{i=0}^{t}\left(\varepsilon_i^{13/\varepsilon_i}\right) - n^{-\frac{1}{4}+o(1)} \quad (5.3)$$

since $t \le T = n^{o(1)}$. Following this, we claim that the set $I = \left\{0 \le i \le t : \varepsilon_i \ge 15\frac{\log\log\log n}{\log\log n}\right\}$ has size at most $(\log n)^{\frac{9}{10}}$. Indeed, as $x^{1/x}$ is monotone increasing for all $x \le e$, every such $i \in I$ has

$$\varepsilon_i^{13/\varepsilon_i} \ge \left(15\frac{\log\log\log n}{\log\log n}\right)^{\frac{13}{15}\frac{\log\log n}{\log\log\log n}} = (\log n)^{-\frac{13}{15}+o(1)} > (\log n)^{-\frac{9}{10}},$$

where the last inequality holds for large n. Hence, if we had $|I| > 2(\log n)^{9/10}$ then it would follow from (5.3) that $\zeta_{t+1} > 2 - o(1)$, contradicting the assumption of the lemma for large enough n.

Moreover, by the assumption that $\varepsilon_i \in [0, 1]$ we have $\frac{1}{2} \le (1 - \varepsilon_i/2) \le 1$ for all i. Together with the facts that $\kappa_{i+1} \ge (1 - \varepsilon_i/2)\kappa_i - \kappa_i^{2/3}$ for all $i \le t$ due to (5.1) while $\kappa_i \le n$ for all i we now get

$$\begin{aligned}\kappa_{t+1} &\ge \kappa_0 \prod_{i=0}^{t}(1 - \varepsilon_i/2) - \sum_{i=0}^{t}\kappa_i^{2/3} \ge \left(1 - \frac{1}{2}\cdot\frac{15\log\log\log n}{\log\log n}\right)^t 2^{-|I|}\, n - (t+1)n^{2/3} \\ &\ge e^{-15\frac{\log\log\log n}{\log\log n}T} 2^{-|I|}\, n - Tn^{2/3},\end{aligned}$$

where the last inequality used the fact that $t < T$ as well as the inequality $1 - x/2 > e^{-x}$, valid for all $0 < x < 1$. Now, $2^{-|I|} = n^{-o(1)}$ since $|I| \le 2(\log n)^{9/10}$ and by the definition of T we obtain that

$$\kappa_{t+1} \ge e^{-\frac{1}{5}\log n} n^{1-o(1)} - n^{\frac{2}{3}+o(1)} = n^{\frac{4}{5}-o(1)} > n^{3/4}$$

for sufficiently large n, as claimed. This concludes the proof. ■

The remaining sections are devoted to the proof of Proposition 5.1 and are organized as follows. In §5.1 we will relate $G_t(\frac{1}{2})$ to the expected change in κ_t. While unfortunately there is no direct recursive relation for the sequence $\{G_t(\frac{1}{2}) : t \ge 0\}$, in §5.2 we will approximate $\mathbb{E}[F_{t+1}(\kappa_t s) \mid \mathcal{F}_t]$ (closely related to $G_{t+1}(s) = (1/\kappa_{t+1})F_{t+1}(\kappa_{t+1}s)$) in terms of several evaluations of G_t. We will then refine our approximation of $G_{t+1}(\frac{1}{2})$ in §5.3 by examining $F_{t+1}(s)$ at a point $s \approx \mathbb{E}[\frac{1}{2}\kappa_{t+1} \mid \mathcal{F}_t]$. Finally, these ingredients will be combined into the proof of Proposition 5.1 in §5.4.

5.1. **Relating $G_t(\frac{1}{2})$ to the coalescence rate.** The next lemma shows that the value of $G_t(\frac{1}{2})$ governs the expected number of merges in round $t+1$.

Lemma 5.3. *Suppose that after t rounds we have $\kappa_t \geq 2$ clusters and set $\varepsilon_t = 1 - G_t(\frac{1}{2})$. Then*

$$\left|\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] - (1-\varepsilon_t/2)\kappa_t\right| \leq \tfrac{1}{4}\,. \tag{5.4}$$

This emphasizes the importance of tracking the value of $G_t(\frac{1}{2})$, as one could derive a lower bound on the coalescence time by showing that $G_t(\frac{1}{2})$ is sufficiently close to 1 (that is, ε_t is suitably small). In order to prove this lemma we first require two straightforward facts on the functions involved.

Claim 5.4. *The following holds for all t with probability 1. The function $G_t(\cdot)$ is convex, decreasing and 1-Lipschitz on the domain $\mathbb{R}^+$. Furthermore, $G_t(s) \geq e^{-s}$ for any s.*

Proof. Denote the cluster-sizes at the end of round t by $w_1, \ldots, w_{\kappa_t}$. Recall that by definition $G_t(s) = (1/\kappa_t)F_t(\kappa_t s) = (1/\kappa_t)\sum_i e^{-w_i\kappa_t s}$ is an arithmetic mean of negative exponentials of s, hence convex and decreasing. Moreover, its first derivative is $G_t'(s) = F_t'(\kappa_t s)$ and in particular

$$G_t'(0) = F_t'(0) = -\sum_i w_i = -1\,.$$

Since $G_t'(s)$ is increasing and negative we deduce that G_t is indeed 1-Lipschitz. Finally, since the negative exponential function is convex, Jensen's inequality concludes the proof by yielding

$$G_t(s) = \frac{1}{\kappa_t}\sum_i e^{-w_i\kappa_t s} \geq e^{-(1/\kappa_t)\sum_i w_i\kappa_t s} = e^{-s}\,. \qquad \blacksquare$$

Claim 5.5. *For any real numbers $0 \leq x \leq 1$ and $\kappa > 0$ we have $(1-x)^\kappa \geq e^{-\kappa x} - (e\kappa)^{-1}$.*

Proof. Fix $\kappa > 0$ and consider the function $f(x) = \kappa\big(e^{-\kappa x} - (1-x)^\kappa\big)$. The desired inequality is equivalent to having $f(x) \leq 1/e$ for all $0 \leq x \leq 1$, hence it suffices to bound $f(x)$ at all local maxima, then compare that bound to its values at the endpoints $f(0) = 0$ and $f(1) = \kappa e^{-\kappa}$.

It is easy to verify that any local extrema x^* must satisfy $(1-x^*)^{\kappa-1} = e^{-\kappa x^*}$, and so

$$f(x^*) = \kappa e^{-\kappa x^*}(1-(1-x^*)) = (\kappa x^*)e^{-\kappa x^*}\,.$$

Since $ye^{-y} \leq 1/e$ for any $y \in \mathbb{R}$, both $f(x^*)$ and $f(1)$ are at most $1/e$, as required. $\blacksquare$

Proof of Lemma 5.3. Let $\kappa = \kappa_t$ and $\kappa' = \kappa_{t+1}$, and as usual let $w_1, \ldots, w_\kappa$ denote the cluster-sizes at the end of t rounds. Recalling the definition of the uniform coalescence process, the number of pairs of clusters that merge in round $t+1$ is equal to the number of clusters which

(a) select to be acceptors in this round, and
(b) receive at least one incoming request in this round.

(Compare this simple characterization with the number of merges in the *size-biased* process, where one must also consider the cluster-sizes of the incoming requests relative to the size of the acceptor.) A given cluster $\mathcal{C}_i$ becomes an acceptor with probability $\frac{1}{2}$, and conditioning on this event we are left with $\kappa - 1$ other clusters, each of which may send a request to the cluster $\mathcal{C}_i$ with probability $w_i/2$ (the factor of 2 accounts for the choice to issue rather than accept requests this round) independently of its peers. Altogether we conclude that the probability that $\mathcal{C}_i$ accepts an incoming request is exactly $\frac{1}{2}\big(1-(1-w_i/2)^{\kappa-1}\big)$ and so the expected total number of merges is

$$\mathbb{E}[\kappa - \kappa' \mid \mathcal{F}_t] = \frac{1}{2}\sum_i \Big(1-(1-w_i/2)^{\kappa-1}\Big)\,.$$

Therefore,

$$\mathbb{E}[\kappa-\kappa' \mid \mathcal{F}_t] \geq \frac{1}{2}\sum_i \left(1-e^{-\frac{1}{2}w_i(\kappa-1)}\right) = \frac{1-G_t\left(\frac{1}{2}\frac{\kappa-1}{\kappa}\right)}{2}\kappa \geq \frac{1-G_t\left(\frac{1}{2}\right)}{2}\kappa - \frac{1}{4}\,,$$

where the last inequality is due to G_t being 1-Lipschitz as was established in Claim 5.4. For an upper bound on the expected number of merges we apply Claim 5.5, from which it follows that

$$\mathbb{E}[\kappa-\kappa' \mid \mathcal{F}_t] \leq \frac{1}{2}\sum_i \left(1-e^{-\frac{1}{2}w_i(\kappa-1)}+\frac{1}{e\kappa}\right) \leq \frac{1}{2}\sum_i \left(1-e^{-\frac{1}{2}w_i\kappa}\right)+\frac{1}{2e} = \frac{1-G_t\left(\frac{1}{2}\right)}{2}\kappa+\frac{1}{2e}\,.$$

Combining these bounds gives the required result. ■

5.2. **Recursive approximation for F_t.** Despite the fact that there is no direct recursion for the values of $G_t(\frac{1}{2})$, it turns out that on the level of expectation one can recover values of its counterpart F_{t+1} from several different evaluations of G_t. Note that this still does not provide an estimate for the expected value of G_{t+1}, as the transformation between the F_{t+1} and G_{t+1} unfortunately involves the number of clusters at time $t+1$, thereby introducing nonlinearity to the approximation.

Lemma 5.6. *Suppose that after t rounds $\kappa_t \geq 2$ and let $\varepsilon_t = 1-G_t(\frac{1}{2})$. Then*

$$\mathbb{E}[F_{t+1}(\kappa_t s) \mid \mathcal{F}_t] > (1-\varepsilon_t/2)\kappa_t\left[\frac{\alpha}{\alpha+\beta}G_t(s)+\frac{\beta}{\alpha+\beta}G_t\left(s+\tfrac{1}{2}\right)\right]-2\,, \tag{5.5}$$

where

$$\alpha=\alpha(s,t)=G_t(s)+G_t\left(\tfrac{1}{2}\right)\,, \qquad \beta=\beta(s,t)=1-G_t(s)\,.$$

Remark. Although the approximation in Eq. (5.5) may look intractable, its structure is in fact quite useful: The leading factor $(1-\varepsilon_t/2)\kappa_t$ is essentially $\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t]$ from Lemma 5.3, which is particularly convenient as we will need to divide by κ_{t+1} to pass from F_{t+1} to G_{t+1}.

Proof of lemma. As before, let $\kappa=\kappa_t$ and denote the cluster-sizes by $w_1,\dots,w_\kappa$. We account for the change $F_{t+1}(s)-F_t(s)$ as follows: Should the clusters $\mathcal{C}_i$ and $\mathcal{C}_j$ merge in round $t+1$, this would contribute exactly $e^{-(w_i+w_j)s}-e^{-w_is}-e^{-w_js}$ to $F_{t+1}(s)-F_t(s)$. Thus, $\mathbb{E}[F_{t+1}(s)-F_t(s) \mid \mathcal{F}_t]$ is simply the sum of these expressions, weighted by the probabilities that the individual pairs merge.

Let us calculate the probability that $\mathcal{C}_i$ accepts an incoming request from the cluster $\mathcal{C}_j$. First let R_i denote the event that $\mathcal{C}_i$ accepts an incoming request from *some* cluster, which was shown in the proof of Lemma 5.3 to satisfy $\mathbb{P}(R_i \mid \mathcal{F}_t)=\frac{1}{2}\big(1-(1-w_i/2)^{\kappa-1}\big)$. Crucially, the fact that acceptors select an incoming request to merge with via a uniform law now implies that, given R_i, the identity of the cluster that $\mathcal{C}_i$ merges with is uniform over the remaining $\kappa-1$ clusters by symmetry. In particular, the probability that $\mathcal{C}_i$ accepts a merge request from $\mathcal{C}_j$ equals $\mathbb{P}(R_i \mid \mathcal{F}_t)/(\kappa-1)$ and so

$$\mathbb{E}[F_{t+1}(s)-F_t(s) \mid \mathcal{F}_t] = \sum_{i\neq j}\left(e^{-(w_i+w_j)s}-e^{-w_is}-e^{-w_js}\right)\frac{1}{2}\left(1-(1-w_i/2)^{\kappa-1}\right)\frac{1}{\kappa-1}\,. \tag{5.6}$$

The term $(1-w_i/2)^{\kappa-1}$ is greater or equal to $e^{-(\kappa-1)w_i/2}-[e(\kappa-1)]^{-1} \geq e^{-w_i\kappa/2}-[e(\kappa-1)]^{-1}$ by Claim 5.5. Since $e^{-(w_i+w_j)s}-e^{-w_is}-e^{-w_js}$ is always negative by convexity, this gives

$$\mathbb{E}[F_{t+1}(s)-F_t(s) \mid \mathcal{F}_t] \geq \sum_{i\neq j}\left(e^{-(w_i+w_j)s}-e^{-w_is}-e^{-w_js}\right)\frac{1}{2}\left(1-e^{-w_i\kappa/2}+\frac{1}{e(\kappa-1)}\right)\frac{1}{\kappa-1}\,.$$

Next, observe that

$$e^{-(w_i+w_j)s}-e^{-w_is}-e^{-w_js} = -1+\left(1-e^{-w_is}\right)\left(1-e^{-w_js}\right) \geq -1\,, \tag{5.7}$$

hence we can sum the effect of the term $1/(e(\kappa-1))$ over all $\kappa(\kappa-1)$ indices $i \neq j$ and get

$$\mathbb{E}[F_{t+1}(s) - F_t(s) \mid \mathcal{F}_t] \geq \frac{1}{2(\kappa-1)} \sum_{i \neq j} \left[\left(e^{-(w_i+w_j)s} - e^{-w_i s} - e^{-w_j s} \right) \left(1 - e^{-w_i \kappa/2} \right) \right] - \frac{\kappa}{2e(\kappa-1)} \,.$$

Note that the last expression has magnitude at most $1/e$ due to the assumption $\kappa \geq 2$. Furthermore, each of the $\kappa(\kappa-1)$ summands in the summation over $i \neq j$ has magnitude at most 1, hence we may replace the factor $1/(\kappa-1)$ with $1/\kappa$ in front of the summation at a maximal cost of $\frac{1}{2}(\frac{1}{\kappa-1} - \frac{1}{\kappa})\kappa(\kappa-1) = \frac{1}{2}$, giving

$$\begin{aligned} \mathbb{E}[F_{t+1}(s) - F_t(s) \mid \mathcal{F}_t] &\geq \frac{1}{2\kappa} \sum_{i \neq j} \left[\left(e^{-(w_i+w_j)s} - e^{-w_i s} - e^{-w_j s} \right) \left(1 - e^{-w_i \kappa/2} \right) \right] - \frac{1}{2} - \frac{1}{e} \\ &> \frac{1}{2\kappa} \sum_{i,j} \left[\left(e^{-(w_i+w_j)s} - e^{-w_i s} - e^{-w_j s} \right) \left(1 - e^{-w_i \kappa/2} \right) \right] - 2 \,, \end{aligned} \tag{5.8}$$

where the last inequality is due to each of the κ diagonal terms $i = j$ having magnitude at most 1.

Since (5.5) addresses $F_{t+1}(\kappa s)$ rather than $F_{t+1}(s)$ we now focus on the following summation:

$$\begin{aligned} &\sum_{i,j} \left(e^{-(w_i+w_j)\kappa s} - e^{-w_i \kappa s} - e^{-w_j \kappa s} \right) \left(1 - e^{-w_i \kappa/2} \right) \\ &\quad = \sum_{i,j} \left(e^{-w_i \kappa s - w_j \kappa s} - e^{-w_i(\kappa s + \frac{\kappa}{2}) - w_j \kappa s} - e^{-w_i \kappa s} + e^{-w_i(\kappa s + \frac{\kappa}{2})} - e^{-w_j \kappa s} + e^{-w_j \kappa s - w_i \kappa/2} \right) \\ &\quad = F_t(\kappa s)^2 - F_t\left(\kappa s + \frac{\kappa}{2}\right) F_t(\kappa s) - \kappa\, F_t(\kappa s) + \kappa\, F_t\left(\kappa s + \frac{\kappa}{2}\right) - \kappa\, F_t(\kappa s) + F_t\left(\frac{\kappa}{2}\right) F_t(\kappa s) \,. \end{aligned}$$

Using this in (5.8), noting that the term $-F_t(\kappa s)$ cancels out, we find that

$$\begin{aligned} \mathbb{E}[F_{t+1}(\kappa s) \mid \mathcal{F}_t] &> \frac{1}{2\kappa} \left[F_t(\kappa s)^2 - F_t\left(\kappa s + \frac{\kappa}{2}\right) F_t(\kappa s) + \kappa\, F_t\left(\kappa s + \frac{\kappa}{2}\right) + F_t\left(\frac{\kappa}{2}\right) F_t(\kappa s) \right] - 2 \\ &= \frac{\kappa + F_t\left(\frac{\kappa}{2}\right)}{2} \left[\frac{F_t(\kappa s) + F_t\left(\frac{\kappa}{2}\right)}{\kappa + F_t\left(\frac{\kappa}{2}\right)} \cdot \frac{F_t(\kappa s)}{\kappa} + \frac{\kappa - F_t(\kappa s)}{\kappa + F_t\left(\frac{\kappa}{2}\right)} \cdot \frac{F_t\left(\kappa s + \frac{\kappa}{2}\right)}{\kappa} \right] - 2 \\ &= \frac{1 + G_t\left(\frac{1}{2}\right)}{2} \kappa \left[\frac{\alpha}{\alpha+\beta} G_t(s) + \frac{\beta}{\alpha+\beta} G_t\left(s + \tfrac{1}{2}\right) \right] - 2 \,, \end{aligned}$$

where $\alpha = G_t(s) + G_t\left(\frac{1}{2}\right)$ and $\beta = 1 - G_t(s)$, thus establishing (5.5). ■

5.3. **Quantifying the convexity correction in the recursion for F_t.** Examine the recursion established in Lemma 5.6. In order to derive lower bounds on the F_t's, we recognize the second factor in the r.h.s. of (5.5) as a weighted arithmetic mean of two evaluations of G_t. Recalling that G_t is a convex combination of negative exponentials, we will now estimate the "convexity correction" between G_t and its weighted mean. It is precisely this increment which will allow us to show that G_t rises toward 1 at a nontrivial rate, as the following lemma demonstrates:

Lemma 5.7. *Suppose after t rounds $\kappa_t \geq 2$ and let $\varepsilon_t = 1 - G_t(\frac{1}{2})$ and $\kappa^* = (1 - \varepsilon_t/2)\kappa_t$. Then*

$$\mathbb{E}[F_{t+1}(\kappa^*) \mid \mathcal{F}_t] \geq \left[G_t(1) + \varepsilon_t^{13/\varepsilon_t} \right] \kappa^* - 2 \,. \tag{5.9}$$

Indeed, by Lemma 5.3 we recognize that κ^* is approximately $\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t]$, hence, postponing for the moment concentration arguments, one sees that Eq. (5.9) resembles the form of Eq. (5.2). Our first step in proving this lemma will be to establish a lower bound similar to (5.9) which replaces the $\varepsilon_t^{1/\varepsilon_t}$ term by the convexity correction between G_t and its weighted mean from Eq. (5.5).

Claim 5.8. *Suppose after t rounds $\kappa_t \geq 2$ and let $\varepsilon_t = 1 - G_t(\frac{1}{2})$ and $\kappa^* = (1-\varepsilon_t/2)\kappa_t$. Let $h(s)$ be the secant line intersecting $G_t(s)$ at $s_1 = \kappa^*/\kappa_t$ and $s_2 = s_1 + \frac{1}{2}$. Let $\theta = \frac{\alpha}{\alpha+\beta}s_1 + \frac{\beta}{\alpha+\beta}s_2$ where $\alpha = G_t(s_1) + G_t(\frac{1}{2})$ and $\beta = 1 - G_t(s_1)$, and let $\Delta = h(\theta) - G_t(\theta)$. Then*

$$\mathbb{E}[F_{t+1}(\kappa^*) \mid \mathcal{F}_t] \geq [G_t(1) + \Delta]\kappa^* - 2 \tag{5.10}$$

and in addition

$$\frac{\varepsilon_t}{4} \leq \theta - s_1 \leq \frac{1}{4}\,. \tag{5.11}$$

Proof. Applying Lemma 5.6 with $s = s_1$ and rewriting its statement in terms of h, θ, Δ give

$$\mathbb{E}[F_{t+1}(\kappa^*) \mid \mathcal{F}_t] > (1-\varepsilon_t/2)\kappa_t h(\theta) - 2 = h(\theta)\kappa^* - 2 = [G_t(\theta) + \Delta]\kappa^* - 2\,.$$

Since we established in Claim 5.4 that G_t is decreasing, (5.10) will follow from showing that $\theta \leq 1$. Note that θ is a weighted mean between $s_1 = 1 - \varepsilon_t/2$ and $s_2 = s_1 + \frac{1}{2}$, and so it is not immediate that $\theta \leq 1$. To show that this is the case, we argue as follows.

Recalling the definition of θ, we wish to show that $\alpha s_1 + \beta(s_1 + \frac{1}{2}) \leq \alpha + \beta$ where $\alpha = G_t(s_1) + G_t(\frac{1}{2})$ and $\beta = 1 - G_t(s_1)$. Observe that $\alpha + \beta = 1 + G_t(\frac{1}{2}) = 2 - \varepsilon_t = 2s_1$ by definition. Therefore, $\theta \leq 1$ if and only if $(\alpha+\beta)s_1 + \beta/2 \leq 2s_1$, or equivalently

$$2\left(G_t(\tfrac{1}{2}) - 1\right)s_1 + 1 \leq G_t(s_1)\,.$$

We claim that indeed

$$2\left(G_t(\tfrac{1}{2}) - 1\right)s + 1 \leq G_t(s) \quad \text{for any } s \geq \tfrac{1}{2}\,, \tag{5.12}$$

which would in particular imply that it holds for $s = s_1 > \frac{1}{2}$ since $s_1 = 1 - \varepsilon_t/2$ with $\varepsilon_t < 1$. In order to verify (5.12) observe that its l.h.s. is an affine function of s whereas the r.h.s. is convex, and that equality holds for $s = 0$ (recall that $G_t(0) = 1$) and $s = \frac{1}{2}$. Thus, the affine l.h.s. does not exceed the convex r.h.s. for any $s \geq \frac{1}{2}$, as required. We now conclude that $\theta \leq 1$, establishing (5.10).

It remains to prove (5.11). Since θ is a weighted arithmetic mean of s_1 and $s_2 = s_1 + \frac{1}{2}$, the upper bound will follow once we show that the weight on s_1 exceeds the weight on $s_2 + \frac{1}{2}$, that is, when $\alpha > \beta$ or equivalently

$$2G_t(s_1) + G_t(\tfrac{1}{2}) > 1\,.$$

This indeed holds, as Claim 5.4 established that $G_t(s) \geq e^{-s}$ and therefore the l.h.s. above is at least $2e^{-s_1} + e^{-1/2} \geq 2/e + e^{-1/2} > \frac{5}{4}$, where we used the fact that $s_1 = 1 - \varepsilon_t/2 \leq 1$.

For the lower bound in (5.11), recall from Claim 5.4 that G_t is decreasing and $G_t(0) = 1$, which together with the aforementioned fact that $s_1 \geq \frac{1}{2}$ gives

$$\frac{\beta}{\alpha+\beta} = \frac{1 - G_t(s_1)}{1 + G_t(\frac{1}{2})} \geq \frac{1 - G_t(\frac{1}{2})}{2} = \frac{\varepsilon_t}{2}\,.$$

The proof is now concluded by noting that $\theta - s_1 = \frac{1}{2} \cdot \frac{\beta}{\alpha+\beta}$ by definition. ∎

Next, we will provide a lower bound on the convexity correction in terms of the difference between two evaluations of G_t.

Claim 5.9. *Let $s \leq 1$ and let h be the secant line intersecting G_t at s and $s + \frac{1}{2}$. For any $0 \leq \delta \leq \frac{1}{4}$,*

$$h(s+\delta) - G_t(s+\delta) \geq \frac{\delta^2}{2}\left[G_t(\tfrac{1}{2}) - G_t(1)\right]^2\,.$$

Proof. Let g denote the secant line intersecting G_t at s and $s+2\delta$. Since $\delta \le \frac{1}{4}$ and G_t is a decreasing convex function,

$$G_t(s+\delta) < g(s+\delta) \le h(s+\delta)\,.$$

It thus suffices to show the following to deduce the statement of the claim:

$$g(s+\delta) - G_t(s+\delta) \ge \frac{\delta^2}{2}\left[G_t(\tfrac{1}{2}) - G_t(1)\right]^2\,, \tag{5.13}$$

which has a particularly convenient l.h.s. due to the fact that $g(s+\delta) = \frac{1}{2}[G_t(s) + G_t(s+2\delta)]$ by definition. Now let $\kappa = \kappa_t$ and let $w_1, \ldots, w_\kappa$ be the cluster-sizes at the end of round t. We have

$$\begin{aligned}\frac{1}{2}\big[G_t(s) + G_t(s+2\delta)\big] - G_t(s+\delta) &= \frac{1}{2\kappa}\sum_i \left[e^{-w_i\kappa s} - 2e^{-w_i\kappa(s+\delta)} + e^{-w_i\kappa(s+2\delta)}\right] \\ &= \frac{1}{2\kappa}\sum_i e^{-w_i\kappa s}\left(1 - e^{-w_i\kappa\delta}\right)^2\,. \end{aligned}\tag{5.14}$$

By Cauchy-Schwarz, the r.h.s. of (5.14) satisfies

$$\frac{1}{2\kappa}\sum_i e^{-w_i\kappa s}\left(1-e^{-w_i\kappa\delta}\right)^2 \ge \frac{1}{2}\left[\frac{1}{\kappa}\sum_i e^{-w_i\kappa s/2}\big(1-e^{-w_i\kappa\delta}\big)\right]^2 = \frac{1}{2}\left[G_t\Big(\frac{s}{2}\Big) - G_t\Big(\frac{s}{2}+\delta\Big)\right]^2\,. \tag{5.15}$$

Set $K = \lceil 1/2\delta \rceil$, noting that $K \le 1/\delta$ as $\delta \le \frac{1}{4}$. Since G_t is a decreasing convex function we have

$$G_t\left(\frac{s}{2}\right) - G_t\left(\frac{s}{2}+\delta\right) \ge G_t\left(\frac{s}{2}+(j-1)\delta\right) - G_t\left(\frac{s}{2}+j\delta\right) \qquad \text{for any } j \ge 1\,,$$

and summing these equations for $j = 1, \ldots, K$ yields

$$G_t\left(\frac{s}{2}\right) - G_t\left(\frac{s}{2}+\delta\right) \ge \frac{1}{K}\left[G_t\left(\frac{s}{2}\right) - G_t\left(\frac{s}{2}+K\delta\right)\right] \ge \frac{1}{K}\left[G_t\Big(\frac{s}{2}\Big) - G_t\Big(\frac{s}{2}+\frac{1}{2}\Big)\right]\,,$$

which is at least $(1/K)\big[G_t(\frac{1}{2}) - G_t(1)\big]$ once again since $s \le 1$ and G_t is convex and decreasing. Therefore, since $K \le 1/\delta$ we can conclude that

$$G_t\left(\frac{s}{2}\right) - G_t\left(\frac{s}{2}+\delta\right) \ge \delta\left[G_t(\tfrac{1}{2}) - G_t(1)\right]\,,$$

which together with (5.14),(5.15) now establishes (5.13) and thus completes the proof. ■

The above claim quantified the convexity correction in terms of $G_t(\frac{1}{2}) - G_t(1)$, and next we wish to estimate this quantity in terms of the key parameter $\varepsilon_t = 1 - G_t(\frac{1}{2})$, which governs the coalescence rate as was established by Lemma 5.3.

Claim 5.10. *For any t we have $G_t(\frac{1}{2}) - G_t(1) \ge \varepsilon_t^{5/\varepsilon_t}$, where $\varepsilon_t = 1 - G_t(\frac{1}{2})$.*

Proof. We first claim that

$$G_t(s) - G_t(2s) \le \sqrt{G_t(2s) - G_t(4s)} \quad \text{for any } s > 0\,. \tag{5.16}$$

Indeed, let $\kappa = \kappa_t$, let $w_1, \ldots, w_\kappa$ be the cluster-sizes after time t and define

$$X = G_t(0) - G_t(s) = \frac{1}{\kappa} \sum_i \left(1 - e^{-w_i \kappa s}\right) ,$$

$$Y = G_t(s) - G_t(2s) = \frac{1}{\kappa} \sum_i e^{-\kappa s} \left(1 - e^{-w_i \kappa s}\right) ,$$

$$Z = G_t(2s) - G_t(3s) = \frac{1}{\kappa} \sum_i e^{-2\kappa s} \left(1 - e^{-w_i \kappa s}\right) .$$

By Cauchy-Schwarz, $Y \le \sqrt{XZ}$. Moreover, $XZ \le Z \le G_t(2s) - G_t(4s)$ since G_t is decreasing and $G_t(0) = 1$, and combining these inequalities now establishes (5.16).

Let $\gamma = G_t(\frac{1}{2}) - G_t(1)$ and let $r \ge 2$. A repeated application of (5.16) reveals that

$$G_t\left(2^{-k}\right) - G_t\left(2^{-(k-1)}\right) \le \gamma^{1/2^{k-1}} \quad \text{for } k = 1, 2, \ldots, r ,$$

and summing these equations we find that

$$G_t\left(2^{-r}\right) - G_t\left(\tfrac{1}{2}\right) \le \sum_{k=1}^{r} \gamma^{1/2^{k-1}} \le r\gamma^{1/2^{r-1}} .$$

On the other hand, since G_t is 1-Lipschitz we also have $G_t(2^{-r}) \ge G_t(0) - 2^{-r} = 1 - 2^{-r}$.

At this point, recalling that $\varepsilon_t = 1 - G_t(\frac{1}{2})$ and combining it with the above bounds gives

$$\varepsilon_t - 2^{-r} \le G_t\left(2^{-r}\right) - G_t\left(\tfrac{1}{2}\right) \le r\gamma^{1/2^{r-1}} . \tag{5.17}$$

The above inequality is valid for any integer $r \ge 2$ and we now choose $r = \lceil \log_2(4/3\varepsilon_t) \rceil$, or equivalently r is the least integer such that $2^{-r} \le \frac{3}{4}\varepsilon_t$. One should notice that indeed $r \ge 2$ since we have $\varepsilon_t < \frac{1}{2}$, which in turn follows from the fact $G_t(s) \ge e^{-s}$ (see Claim 5.4) yielding

$$\varepsilon_t \le 1 - e^{-1/2} < \tfrac{2}{5} . \tag{5.18}$$

Revisiting Eq. (5.17) and using the fact that $2^{-r} \le \frac{3}{4}\varepsilon_t$ we find that $\varepsilon_t/4 \le r\gamma^{1/2^{r-1}}$, and after rearranging $\gamma \ge (\varepsilon_t/4r)^{2^{r-1}}$. Moreover, by definition $r \le \log_2(8/3\varepsilon_t)$ and as one can easily verify that $4\log_2(8/3x) < x^{-11/4}$ for all $0 < x \le \frac{2}{5}$ (which by (5.18) covers the range of ε_t), we have $r < \frac{1}{4}\varepsilon_t^{-11/4}$. The choice of r further implies that $2^{r-1} < 4/3\varepsilon_t$ and combining these bounds gives

$$\gamma > \left(\frac{\varepsilon_t}{4r}\right)^{4/3\varepsilon_t} > \left(\varepsilon_t^{15/4}\right)^{4/3\varepsilon_t} = \varepsilon_t^{5/\varepsilon_t} ,$$

as claimed. ■

We are now ready to establish Eq. (5.9), the quantitative bound on the convexity correction in the weighted mean of Eq. (5.5).

Proof of Lemma 5.7. By Claim 5.8, in order to prove Eq. (5.9) it suffices to show that $\Delta \ge \varepsilon_t^{13/\varepsilon_t}$ with Δ as defined in the statement of that claim. Using Eq. (5.11) of Claim 5.8 we can write $\Delta = h(s_1 + \delta) - G_t(s_1 + \delta)$ where h is the secant line defined in that claim, $s_1 = 1 - \varepsilon_t/2$ and δ satisfies $\varepsilon_t \le 4\delta \le 1$. Therefore, Claim 5.9 implies that $\Delta \ge \frac{1}{2}(\varepsilon_t/4)^2[G_t(\frac{1}{2}) - G_t(1)]^2$. Applying Claim 5.10 we find that

$$\Delta \ge \frac{1}{2}\left(\frac{\varepsilon_t}{4}\right)^2 \left(\varepsilon_t^{5/\varepsilon_t}\right)^2 \ge \varepsilon_t^{13/\varepsilon_t} ,$$

where we consolidated the constant factors into the exponent using the fact that $x^2/32 > x^{3/x}$ for all $0 < x \le \frac{2}{5}$ while bearing in mind that by (5.18) indeed $\varepsilon_t < \frac{2}{5}$. ■

5.4. **Proof of Proposition 5.1.** Let $\kappa = \kappa_t$ and note that w.l.o.g. we may assume that κ is sufficiently large by choosing the constant C from the statement of the proposition appropriately.

Let $w_1, \dots, w_\kappa$ denote the cluster-sizes. As argued before, given $\mathcal{F}_t$ one can realize round $t+1$ of the process by a κ-dimensional product space, where clusters behave independently as follows:

(1) For each i, the cluster $\mathcal{C}_i$ decides whether to send or accept requests via a fair coin toss.
(2) When sending a request $\mathcal{C}_i$ selects its recipient cluster randomly (proportionally to the w_j's).
(3) When accepting requests $\mathcal{C}_i$ generates a random real number between 0 and 1 to be used to select the incoming merge-request it will grant (uniformly over all the incoming requests).

As such, conditioned on $\mathcal{F}_t$ the variable κ_{t+1} is clearly 1-Lipschitz w.r.t. the above product space since changing the value corresponding to the action of one cluster can affect at most one merge. Thus, by a standard well-known coupling argument (see e.g. [3]) the increments of the corresponding Doob martingale are bounded by 1 (i.e. $|M_{i+1} - M_i| \le 1$ where $M_i = \mathbb{E}[\kappa_{t+1} \mid \mathcal{F}'_i]$ with $\mathcal{F}'_i$ being the σ-algebra generated by the actions of clusters $1, \dots, i$ and $\mathcal{F}_t$). Hoeffding's inequality now gives

$$\mathbb{P}\Big(\big|\kappa_{t+1} - \mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t]\big| > a \;\Big|\; \mathcal{F}_t\Big) \le 2\exp(-a^2/2\kappa) \quad \text{for any } a > 0\,.$$

Letting $\kappa^* = (1+\varepsilon_t/2)\kappa$ we recall from Lemma 5.3 that $|\mathbb{E}[\kappa_{t+1} \mid \mathcal{F}_t] - \kappa^*| \le \frac{1}{4}$ and obtain that

$$\mathbb{P}\big(|\kappa_{t+1} - \kappa^*| > \kappa^{2/3} \mid \mathcal{F}_t\big) \le 2\exp\big(-\tfrac{1}{2}(\kappa^{2/3} - \tfrac{1}{4})^2/\kappa\big) = 2\exp\big(-\tfrac{1}{2}\kappa^{1/3} + O(\kappa^{-1/3})\big) < \kappa^{-100}\,, \quad (5.19)$$

where the last inequality holds for any sufficiently large κ, thus establishing (5.1).

To obtain Eq. (5.2), recall from (5.7) that $-1 \le e^{-(w_i+w_j)s} - e^{-w_i s} - e^{-w_j s} \le 0$, implying that the random variable $F_{t+1}(\kappa^*)$ is 1-Lipschitz w.r.t. the aforementioned κ-dimensional product space. Furthermore, $\mathbb{E}[F_{t+1}(\kappa^*) \mid \mathcal{F}_t] \ge \big[G_t(1) + \varepsilon_t^{13/\varepsilon_t}\big]\kappa^* - 2$ due to Lemma 5.7, and by the same argument as before we conclude from Hoeffding's inequality that

$$\mathbb{P}\left(F_{t+1}(\kappa^*) < \big[G_t(1) + \varepsilon_t^{13/\varepsilon_t}\big]\kappa^* - \kappa^{2/3} \mid \mathcal{F}_t\right) \le \exp\big(-\tfrac{1}{2}\kappa^{1/3} + O(\kappa^{-1/3})\big) < \kappa^{-100}\,.$$

Rewriting this inequality in terms of G_{t+1}, with probability at least $1 - \kappa^{-100}$ we have

$$G_{t+1}\left(\frac{\kappa^*}{\kappa_{t+1}}\right) \ge \big[G_t(1) + \varepsilon_t^{13/\varepsilon_t}\big]\frac{\kappa^*}{\kappa_{t+1}} - \frac{\kappa^{2/3}}{\kappa_{t+1}} \ge G_t(1) + \varepsilon_t^{13/\varepsilon_t} - \frac{2|\kappa_{t+1} - \kappa^*| + \kappa^{2/3}}{\kappa_{t+1}}\,,$$

where we used that $[G_t(1) + \varepsilon_t^{13/\varepsilon_t}](\kappa^* - \kappa_{t+1}) \ge -(G_t(0)+1)|\kappa_{t+1} - \kappa^*| = -2|\kappa_{t+1} - \kappa^*|$ due to $G_t(s)$ being decreasing in s. Moreover, since G_{t+1} is 1-Lipschitz as was shown in Claim 5.4, in this event we have

$$G_{t+1}(1) \ge G_{t+1}\left(\frac{\kappa^*}{\kappa_{t+1}}\right) - \left|1 - \frac{\kappa^*}{\kappa_{t+1}}\right| \ge G_t(1) + \varepsilon_t^{13/\varepsilon_t} - \frac{3|\kappa_{t+1} - \kappa^*| + \kappa^{2/3}}{\kappa_{t+1}}\,.$$

Finally, recalling from Eq. (5.19) that $|\kappa_{t+1} - \kappa^*| \le \kappa^{2/3}$ except with a probability of at most κ^{-100}, we can conclude that with probability at least $1 - 2\kappa^{-100}$

$$G_{t+1}(1) \ge G_{t+1}\left(\frac{\kappa^*}{\kappa_{t+1}}\right) - \left|1 - \frac{\kappa^*}{\kappa_{t+1}}\right| \ge G_t(1) + \varepsilon_t^{13/\varepsilon_t} - 4\frac{\kappa^{2/3}}{\kappa_{t+1}} \ge G_t(1) + \varepsilon_t^{13/\varepsilon_t} - 8\kappa^{-1/3}\,,$$

where the last inequality used the fact that $\kappa_{t+1} \ge \kappa/2$ by definition of the coalescence process (since the merging pairs of clusters are always pairwise-disjoint). This yields (5.2) and therefore completes the proof of the proposition. ■

Acknowledgments

We thank Yuval Peres and Dahlia Malkhi for suggesting the problem and for useful discussions. The starting point of our work is attributed to the analytic approximation framework of Oded Schramm, and E. L. is indebted to Oded for enlightening and fruitful discussions on his approach.

This work was initiated while P.-S. L. was an intern at the Theory Group of Microsoft Research, and he thanks the Theory Group for its hospitality.

References

[1] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab. **1** (1991), no. 2, 228–266.

[2] D. Aldous, *Deterministic and stochastic models for coalescence (aggregation, coagulation): a review of the mean-field theory for probabilists*, Bernoulli **5** (1997), 3–48.

[3] N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., John Wiley & Sons Inc., Hoboken, NJ, 2008.

[4] B. Bollobás, *Random graphs*, 2nd ed., Cambridge University Press, 2001.

[5] B. Bollobás and I. Simon, *Probabilistic analysis of disjoint set union algorithms*, SIAM J. Comput. **22** (1993), no. 5, 1053–1074.

[6] J. Cooper and N. Linial, *Fast perfect-information leader-election protocols with linear immunity*, Combinatorica **15** (1995), 319–332. Early version appeared in Proc. of the the 25th Annual ACM Symposium on Theory of Computing (STOC 1993), 662–671.

[7] R. Durrett, *Probability models for DNA sequence evolution*, 2nd ed., Probability and its Applications (New York), Springer, New York, 2008.

[8] R. Durrett, *Probability: theory and examples*, 3rd ed., Duxbury Press, 2004.

[9] P. Erdős and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. **5** (1960), 17–61.

[10] Y. Fernandess and D. Malkhi, *On spreading recommendations via social gossip*, Proc. of the 20th annual symposium on Parallelism in algorithms and architectures (SPAA 2008), 91–97.

[11] Y. Fernandess and D. Malkhi, *private communication* (2007).

[12] J. A. Fill, H. M. Mahmoud, and W. Szpankowski, *On the distribution for the duration of a randomized leader election algorithm*, Ann. Appl. Probab. **6** (1996), no. 4, 1260–1283.

[13] D. A. Freedman, *On tail probabilities for martingales*, Ann. Probability **3** (1975), 100–118.

[14] Z. Galil and G. F. Italiano, *Data structures and algorithms for disjoint set union problems*, ACM Comput. Surv. **23** (1991), 319–344.

[15] H. Gazit, *An optimal randomized parallel algorithm for finding connected components in a graph*, Proc. of the 27th Annual Symposium on Foundations of Computer Science (FOCS 1986), 492–501.

[16] G. Grimmett, *Percolation*, 2nd ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 321, Springer-Verlag, Berlin, 1999.

[17] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[18] J. F. C. Kingman, *The coalescent*, Stochastic Process. Appl. **13** (1982), no. 3, 235–248.

[19] D. E. Knuth and A. Schönhage, *The expected linearity of a simple equivalence algorithm*, Theoret. Comput. Sci. **6** (1978), no. 3, 281–315.

[20] G. Miller and J. H. Reif, *Parallel tree contraction and its applications*, Proc. of the 26th Annual Symposium on Foundations of Computer Science (FOCS 1985), 478–489.

[21] M. Molloy and B. Reed, *Graph colouring and the probabilistic method*, Springer-Verlag, Berlin, 2002.

[22] R. Ostrovsky, S. Rajagopalan, and U. Vazirani, *Simple and Efficient Leader Election in the Full Information Model*, Proc. of the 26th Annual ACM Symposium on Theory of Computing (STOC 1994), 234–242.

[23] A. Russell and D. Zuckerman, *Perfect information leader election in* $\log^* n + O(1)$ *rounds*, J. Comput. System Sci. **63** (2001), 612–626. Special issue on FOCS 1998 (Palo Alto, CA).

[24] O. Schramm, *An analytic approach to Malkhi's stochastic coalescence process*, Microsoft Research Theory Seminar (Redmond, WA, August 2008). Also: *private communication* (2007).

[25] J. H. Spencer and N. Wormald, *Birth control for giants*, Combinatorica **27** (2007), 587–628.

[26] A. C.-C. Yao, *On the average behavior of set merging algorithms (extended abstract)*, Proc. of the 8th annual ACM symposium on Theory of Computing (STOC 1976), 192–195.

[27] D. Zuckerman, *Randomness-optimal sampling, extractors, and constructive leader election*, Proc. of the 28th Annual ACM Symposium on Theory of Computing (STOC 1996), 286–295.

Po-Shen Loh
Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
E-mail address: ploh@cmu.edu

Eyal Lubetzky
Theory Group of Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA.
E-mail address: eyal@microsoft.com