A Polylogarithmic Gossip Algorithm for Plurality Consensus

Mohsen Ghaffari
MIT
ghaffari@mit.edu

Merav Parter
MIT
parter@mit.edu

Abstract

Consider n anonymous nodes each initially supporting an opinion in $\{1, 2, \ldots, k\}$ and suppose that they should all learn the opinion with the largest support. Per round, each node contacts a random other node and exchanges B bits with it, where typically B is at most $O(\log n)$.

This basic distributed computing problem is called the *plurality consensus* problem (in the *gossip* model) and it has received extensive attention. An efficient plurality protocol is one that converges to the plurality consensus as fast as possible, and the standard assumption is that each node has memory at most polylogarithmic in n. The best known time bound is due to Becchetti et al. [SODA'15], reaching plurality consensus in $O(k \log n)$ rounds using $\log(k+1)$ bits of local memory, under some mild assumptions. As stated by Becchetti et al., achieving a poly-logarithmic time complexity remained an open question.

Resolving this question, we present an algorithm that with high probability[1] reaches plurality consensus in $O(\log k \, \log n)$ rounds, while having message and memory size of $\log k + O(1)$ bits. This even holds under considerably more relaxed assumptions regarding the initial bias (towards plurality) compared to those of prior work. The algorithm is based on a very simple and arguably natural mechanism.

[1]As standard, *with high probability* (w.h.p.) indicates that an event has probability $1 - 1/n^c$ for a constant $c \geq 2$.

1 Introduction and Related Work

Coordination problems in distributed systems require nodes to agree on a common action or value. These problems are referred to as consensus, agreement, or voting and are among the most fundamental problems in distributed computing [Lyn96, Tsi84, TBA86]. In this paper, we consider the following basic agreement problem, often called plurality consensus problem[2] in the gossip model:

> `There are` n `nodes each supporting an opinion in` $\{1, \dots, k\}$`. Per round each node contacts one random other node and exchanges` B `bits with it, where typically` B `is at most` $O(\log n)$`. Eventually, all nodes should hold the opinion with the largest initial support.`

This problem, or close variants of it, come up in computer networks such as sensor networks [AAD+06] and peer-to-peer networks [AFJ06], as well as in a wider range of settings that goes much beyond the classical distributed computing settings: e.g., social networks [MS], biological systems [CKFL05, SKJ+08, CCN12, BSDDS14], and chemical reaction networks [CDS+13, Dot14].

State of the Art: Gossip algorithms for plurality have received extensive attention, see e.g. [BCN+14, BCN+15a, BCN+15b, AAE08, PVV09]. The key performance measures of the algorithms are: (1) *time complexity*, and (2) space measured by the *message size* and *memory size* per node. In the standard regime where message and memory sizes are at most $\text{poly}(\log n)$, the best known algorithm is due to Becchetti et al. [BCN+15a], achieving a time complexity of $O(k \log n)$, under some mild assumptions, while having $\log(k+1)$ bit message/memory. A faster solution was known before, however at the expense of a significantly larger message/memory size. Particularly, the general approach of Kempe et al. [KDG03] can be used to obtain a plurality algorithm with $O(\log n)$ time-complexity but with $O(k \log n)$ message/memory size. Also, a number of other fast algorithms are known but all limited to very small k, e.g., $k = 2$ [AAE08, PVV09]. Given that each of the known solutions lacks in one of the dimensions, Bechetti et al. [BCN+15a] state:

> "*A major open question for the plurality consensus problem is whether a plurality protocol exists that converges in polylogarithmic time and uses only polylogarithmic local memory.*"

Our Result: In this paper, we answer this question in the affirmative. We particularly present an algorithm that has nearly optimal bounds in both dimensions:

Theorem 1.1. *There is a gossip algorithm that solves plurality w.h.p. in* $O(\log k \cdot \log n)$ *rounds using* $\log k + O(1)$ *bits of local memory, assuming that the initial bias* $p_1 - p_2 = \Omega(\sqrt{\log n / n})$*. Here* p_1 *and* p_2 *are, respectively, the fractions of nodes holding the plurality and second largest opinions.*

This initial bias assumption $p_1 - p_2 = \Omega(\sqrt{\log n / n})$ is due to the concentration considerations[3], and it is at least as relaxed as the assumptions in prior work. In fact under the stronger assumptions of [BCN+15a] (to be stated later), our algorithm converges in $O(\log k \log \log n + \log n)$ rounds.

As we will see, the algorithm is based on a quite simple and intuitive mechanism and one can even argue that dynamics with a similar flavor occur in nature. We also note that the result might be viewed as somewhat surprising, given the indications in the prior work and especially when put in contrast to what Becchetti et al. [BCN+15a] write in their conclusion:

> "*We believe that this distance* [referring to the monochromatic distance, which can be as large as $\Omega(k)$] *might represent a general lower bound on the convergence time of* <u>*any*</u> *plurality dynamics which uses only* $\log k + O(1)$ *bits of local memory.*"

[2] This problem has also been called *proportionate agreement* and in the special case of $k = 2$ *majority agreement*.

[3] Even when flipping n fair coins, there is a $1/n^{0.1}$ probability that the number of heads is outside $n/2 \pm \Omega(\sqrt{n \log n})$, which means a relative deviation of $\pm\Omega(\sqrt{\log n / n})$. This is essentially why we need to assume $p_1 - p_2 = \Omega(\sqrt{\log n / n})$.

A Remark—Measuring Memory Size: We note that because of the practical scenarios in which gossip algorithms are applied, it is more standard to measure the memory size in terms of the *number of states* rater than the *number of bits*. Here we briefly discuss the reasons. One justification for considering the number of states is that many of these dynamics are often viewed as Finite-state Automota, for which minimizing the number of states is a classical optimization objective (see, e.g., [Hop71, Yu05]). In addition, states of nodes have a clear physical interpretation in practical settings such as in molecular programming: in particular, in chemical reaction networks, the encoding of the state is not binary but rather (closer to) unary [CSWB09], and hence minimizing the number of the states is a more sensible objective. Although these two ways of measuring are equivalent when talking about exact bounds—as the number of states is simply exponential in the number of bits—they differ when discussing relaxed notions of optimality. Particularly, a constant factor improvement in the bit complexity, which might be regarded negligible, translates to a polynomial improvement in the number of states, which would be considered significant.

In fact, when presenting our result, we start with a simpler algorithm with $\log k + O(\log\log k)$ bit complexity, and then present a more complex one with bit complexity of $\log k + O(1)$. Our first algorithm converges within $O(\log k \cdot \log n)$ rounds but uses $\log k + O(\log\log k)$ bits. Note that this already gives an answer to the open problem raised by Bechetti et al. [BCN+15a], stated above. Yet, even though the message size of this simple algorithm in terms of the number bits is optimal up to an *additive* $O(\log\log k)$ term, with respect to the *number of states*, its $O(k\log k)$ state complexity bound is an $O(\log k)$ factor off from the trivial lower bound k. In our second algorithm, we improve this to a *state complexity* of $O(k)$, which is optimal up to a constant.

1.1 Our approach, in the context of the previous ones

Plurality protocols can be roughly divided into two main classes: what we call *"reading"* protocols and what we call *plurality amplification* protocols (or amplification protocols, for short).

A natural approach for solving plurality, in fact perhaps the first approach that comes to the mind from a standard algorithm design viewpoint, is to estimate the frequencies of the nodes of each opinion and pick the most frequent one. In a more relaxed definition, in a reading style protocol, for each opinion i, we would have at least one node v that learns some estimation for the frequency of this opinion in the population. Kempe et al. [KDG03] provided a general approach for designing reading type protocols for computing aggregate functions (e.g. sum or average). As noted by Becchetti et al. [BCN+15a], adopting the approach of Kemp et al. to the plurality problem yields an $O(\log n)$ time complexity but with $O(k\log n)$ message/memory size. Generally, it seems implausible that reading style algorithms would lead to a plurality algorithm for the random gossip model with polylogarithmic size messages and polylogarithmic time complexity[4].

The second class of plurality amplification protocols do not try to estimate the initial distribution, but rather attempt to *modify* this initial distribution into another distribution of the same plurality opinion, but with a *larger* bias towards plurality. In other words, throughout the execution of an amplification protocol, there is a shift from the initial distribution (with possibly very weak bias) towards the target configuration (where we have consensus on plurality) by repetitively amplifying the bias towards the plurality opinion via some *positive feedback mechanism*. The standard plurality amplification protocols are based on simple majority rules. For example, Becchetti et al. [BCN+14] consider an update rule where each node polls the opinion of three random nodes and adopts the majority opinion among these. This algorithm uses only $\Theta(\log k)$ bits of local memory, but unfortunately has running time $O(\min\{k\log n, n^{1/3}\cdot\log^{2/3} n\})$. Other known time efficient algorithms exist for small values of k, e.g., $k = 2$ [PVV09, DGM+11]. Recently, Becchetti et al. [BCN+15a] analyzed

[4]We note however that if the gossip model is relaxed to include non-random meetings, a rather simple "reading" style algorithm would achieve this objective, see Appendix D.

the well-known *Undecided-State Dynamics*, in which each decided node meeting a node of different opinion becomes undecided (i.e., forgets its opinion) and every undecided node meeting a decided node adopts its opinion. This protocol provides the current best bounds: they showed that the undecided-state dynamics converges within $O(k \log n)$ rounds, using $\log(k+1)$ bits of local memory, assuming[5] $k = O((n/\log n)^{1/6})$ and $p_1 - p_2 = \Omega(1/n^{1/6})$.

The high level idea of our algorithm: We want to create a dynamic with a '*strong and fast positive feedback*' so that, roughly speaking, the opinions with larger support get amplified while those with smaller support get attenuated. In more colloquial terms, we seek a (fast) '*rich get richer*' effect. At the high level, our protocol consists of two main alternating steps: A *selection* step, which is inspired by the classical Darwinian evolution, and a sequence of *recovery* steps. The selection process imposes a constraint on the system that potentially may lead many nodes to lose their opinion and become undecided. Specifically, a node survives the selection process only if it meets a node of its *own* opinion. By that, we intentionally cut down the number of decided nodes significantly while opinions with larger support survive through this process slightly better. Particularly, the plurality opinion gets slightly less attenuation in comparison to the others, hence the proportion of the plurality opinion to that of any other opinion grows. Since repetitive applications of selections would eventually diminish the decided population, our system evolution incorporates some healing process between two consecutive selections. In the healing phase, the fraction of undecided nodes shrinks back to (almost) 0 but the ratios between the number of nodes holding different opinions is roughly preserved. We repeat through this cycle of selection and healing, thus continually *amplifying the plurality*, till reaching plurality consensus. For further discussion on other rich get richer mechanisms, see Appendix A.

Other related work: The plurality problem bears some resemblance to problems studied in some other settings: The restricted case of $k = 2$ has been extensively studied in the setting of population protocols [AAD⁺06] and many "plurality amplification" type algorithms using small number of states (e.g., 3 or 4) have been suggested [AAE08, DV12, MNRS14, AGV15]. This case of binary consensus ($k = 2$) has been also studied in the voter model [DW83, HP01, WH04, Lig12]. The plurality problem also resembles the *heavy hitters* problem in the streaming model, see e.g. [BICS10].

2 The Algorithm: Take 1

We next present a simpler variant of our algorithm, which reaches plurality in $O(\log k \log n)$ rounds, using messages of size $\log k + O(1)$ bits, but with per node memory size of $\log k + \log\log k + O(1)$ bits. Note that this result already answers the open question stated in the introduction.

2.1 The Algorithm Description

The algorithm is an extremely simple and clean dynamics, as follows: Let us call nodes holding one of opinions $\{1, 2, \ldots, k\}$ *decided* while nodes holding no opinion are called *undecided*.

The algorithm works in phases, each having $R = O(\log k)$ rounds, as follows:

- **Round 1 of each phase:** `- Relative Gap Amplification -`
  - a decided node keeps its opinion *only if* it contacts a node with the same opinion,
  - undecided nodes remains undecided.
- **Rounds 2 to R of each phase:** `- Healing -`
  - decided nodes keep their opinion,
  - an undecided node v that contacts a decided node u adopts u's opinion.

[5] They assume $p_1 \geq (1+\alpha)p_2$ for some constant α and $k \leq O(n^{1/6})$. This implies $p_1 - p_2 = \Omega(p_1) = \Omega(1/n^{1/6})$.

When Implementing this algorithm, each message has size $\log(k+1)$ bits as it contains only one opinion in $\{0, 1, \ldots, k\}$, where 0 represents being undecided. As for the memory, the only additional thing that a node needs to remember besides its opinion is the round number in the phase, i.e. modulo $R = O(\log k)$, and that takes $\log \log k + O(1)$ bits.

We next provide some intuition for why this dynamic converges to plurality within $O(\log n)$ phases and then the next subsection presents the related formal analysis.

Basic notations: To keep track of the state of the system, we use a vector $\mathbf{p} = (p_1, p_2, \ldots, p_k) \in [0,1]^k$ where each entry p_i denotes the fraction of the nodes holding opinion $i \in \{1, 2, \ldots, k\}$. Thus, the fraction of undecided nodes is $1 - \sum_{i=1}^{k} p_i$. Without loss of generality, in our analysis, we assume that the opinions are renumbered such that initially $p_1 > p_2 \geq \cdots \geq p_k$.

Convergence intuition: A rough intuition for why the algorithm is designed this way and why it converges to plurality consensus fast, based merely on expectations, is as follows: in the first round of each phase, each node v keeps its opinion only if it contacted a node with the same opinion. Note that this means, even if v contacted an undecided node, v forgets its opinion and becomes undecided. At the end of this round, the fraction of nodes holding an opinion i changes from p_i to p_i^2, in expectation. Thus, the ratio between the plurality opinion and an opinion $i \in \{2, \ldots, k\}$ goes up from $\frac{p_1}{p_i}$ to $(\frac{p_1}{p_i})^2$. This already is the desired *'rich get richer'* effect, in terms of the ratios. However, as a side-effect of this relative *gap amplification*, the fraction of decided nodes has dropped down significantly to about $\sum_{i=1}^{k} p_i^2$, which can be as small as $\Theta(1/k)$. During rounds 2 to R of the phase, we fix this issue by growing the fraction of decided nodes back to about 1. As we will show, during this *healing* period, the ratio between different opinions remains essentially the same. Thus, looking over phases, we can say that the ratio $\frac{p_1}{p_i}$ starts with at least $1 + 1/n$ and gets essentially squared in each phase. That means, this ratio reaches 2 within $O(\log n)$ phases and from there to n in $O(\log \log n)$ additional phases. Due to the integralities, once the ratio passes n, it actually means that $p_i = 0$. When this happens for all the opinions $i \in \{2, \ldots, k\}$, all of them have been filtered out and only the plurality opinion remains alive[6]. From there, within additional $O(\log n / \log k)$ phases, the system reaches plurality consensus where all nodes hold the plurality opinion.

A Remark—Interpretations of the Term Dynamics: When calling our algorithms "dynamics", we use this word as a general and relative rule of thumb for simplicity and succinctness, rather than a precise and well-defined notion. In fact, there is no such precise definition and in the prior work, the word *dynamics* has been used to refer to algorithms with a range of complexity: On one extreme, there are simple update rules such as 3-majority or the undecided-state dynamics, and on the other extreme, there are algorithms such as that of [BCN+15a, Section 4] which involves time-dependent actions, assumes knowledge of global parameters of the network, and even relies on non-random interactions. Our algorithms fall somewhere within this range. Particularly, they include basic time-dependent behavior and use a minimal knowledge of parameters, but all interactions are random. We believe that these algorithm are simple enough for implementation in most practical settings of gossiping algorithms.

2.2 Analysis

We next provide the formal analysis, proving that:

Theorem 2.1. *The above algorithm reaches plurality consensus w.h.p. within $O(\log k \cdot \log n)$ rounds, assuming that the initial bias $p_1 - p_2 \geq \sqrt{C \log n / n}$, for a sufficiently large constant C. Furthermore, if initially $p_1/p_2 \geq 1 + \delta$ for some constant $\delta > 0$, then the convergence will happen in fact within $O(\log k \log \log n + \log n)$ rounds, w.h.p.*

[6] One case say this is virtually the same as '*survival of the fittest*' in Herbert Spencer's terminology for describing Darwin's *natural selection* concept [Spe64]. In our setting, the largest opinion is effectively the fittest.

Let $\mathsf{bias} = p_1 - p_2$ be the difference between the fraction of the plurality opinion and the second largest opinion. Finally, to keep track of the the ratio between the plurality opinion and the second largest opinion, let us define

$$\mathsf{gap} = \min\{\frac{p_1}{\sqrt{10\log n/n}}, \frac{p_1}{p_2}\} \tag{1}$$

Note that when $p_2 \leq \sqrt{10\log n/n}$, there is a large constant gap between p_1 and each of the other opinions as $p_1 \geq \sqrt{C\log n/n}$. Since once all other opinions are below $\sqrt{10\log n/n}$, we will not be able to bound their deviations from expected by an $(1 - o(1))$ factor, in such cases, we will try to base the analysis mainly on the fast growth of p_1, rather than the growth of p_1/p_2. The above definition of gap allows us to unify the two statements.

The three main transitions of the dynamics: Our dynamics can be characterized by three main transitions: The first transition occurs during the first $O(\log n)$ phases, by the end of which, the relative gap between the plurality opinion and the second largest opinion exceeds a 2 factor. The second transition occurs within $O(\log\log n)$ additional phases, where all the non-plurality opinions become extinct. At that point, we will have $p_1 \geq 2/3$ and $\sum_{i\geq 2} p_i = 0$. Finally, the last transition occurs within $O(\log n/\log k)$ additional phases where all nodes hold the plurality opinion. The second and third transitions are proven in Lemma 2.5 to Lemma 2.8, and their proofs are deferred to Appendix B.

We begin by showing that in every phase, three conditions hold: two of these are safety conditions guaranteeing that the fraction of decided nodes is sufficiently large and that the initial bias is maintained; the third condition is a progress condition that guarantees the growth of the gap.

Lemma 2.2. *Consider a phase j and let us use $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ to denote the vector of opinions at the beginning of this phase and $\mathbf{p}^{new}$ to denote the same vector at the end of the phase. Assume that $\sum_{i=1}^{k} p_i \geq 2/3$, $\mathsf{bias} \geq \sqrt{C\log n/n}$ and $p_1 \leq 2/3$. Then, with high probability, we have:*

(S1) $\sum_{i=1}^{k} p_i^{new} \geq 2/3$,

(S2) $\mathsf{bias}^{new} = p_1^{new} - p_2^{new} \geq \sqrt{C\log n/n}$, *and most importantly*

(P) either $p_1^{new} \geq 2/3$ *or* $\mathsf{gap}^{new} \geq \mathsf{gap}^{1.4}$ *where* gap^{new} *is defined as in Eq. (1) with* p_1^{new}, p_2^{new}.

Proof. We begin with (S1). We first claim that due to the gap-amplification step of phase j, the fraction of decided nodes has been dropped down to $\Omega(1/k + \log n/n)$, and then show that this drop is recovered during the $R = O(\log k)$ rounds of the healing step. To see this, note that the expected fraction of decided nodes at the end of the gap amplification round is $\sum_{i=1}^{k} p_i^2$. Since $\sum_{i=1}^{k} p_i \geq 2/3$, we have $\sum_{i=1}^{k} p_i^2 \geq k \cdot ((\sum_{i=1}^{k} p_i)/k)^2 \geq 4/(9k)$. In addition, since $p_1 \geq \sqrt{\log n/n}$, $p_1^2 = \Omega(\log n/n)$. We thus get that $\sum_{i=1}^{k} p_i^2 = \Omega(1/k + \log n/n)$. We now claim that the $\Omega(1/k + \log n/n)$ fraction of decided nodes after the gap-amplification is increased to at least $2/3$ over the R rounds of the healing process. The argument is simple and similar to the usual process of gossiping a rumor. Particularly, in every round $r \in \{2, \ldots, R\}$ of phase j, so long as the fraction of decided nodes is less than $2/3$, this fraction grows in expectation by a factor of at least $4/3$ and hence, it will grow by factor of $6/5$ with high probability. We therefore get that within $O(\log k)$ rounds, the fraction of decided nodes $\sum_{i=1}^{k} p_i$ reaches at least $2/3$ as required. Thus, property (S1) holds.

Next, we prove Claim (P) and then, we reuse some of the arguments to establish (S2). For (P), we distinguish between two cases depending on the magnitude of the second largest opinion p_2.

Case (P.a), when $p_2 \geq \sqrt{10\log n/n}$: After the first gap-amplification round, in expectation, there are np_1^2 nodes with opinion 1 and np_2^2 nodes with opinion 2. We want to show that the ratio between the actual numbers is very close to what we would have according to these expectations, i.e., $(\frac{p_1}{p_1})^2$, and remains so throughout rounds 2 to R. Since we might start with a bias as small as $\mathsf{bias} = \Theta(\sqrt{\log n/n})$, the deviations from expectations should be analyzed carefully. Throughout, we focus on the two leading opinions of this phase and denote by the random variables x_r and y_r, the number of nodes in these two opinions at round $r \in \{1, \ldots, R\}$ of this phase. By Chernoff, it holds that w.h.p.

$$x_1 \in np_1^2\left(1 \pm \frac{\sqrt{5\log n/n}}{p_1}\right) \quad \text{and} \quad y_1 \in np_2^2\left(1 \pm \frac{\sqrt{5\log n/n}}{p_2}\right). \tag{2}$$

We now use this to bound the ratio x_1/y_1:

$$\frac{x_1}{y_1} \geq \left(\frac{p_1}{p_2}\right)^2 \cdot \left(1 + \frac{3\sqrt{5\log n/n}}{p_2}\right)^{-2} \geq \left(\frac{p_1}{p_2}\right)^2 (1 + \frac{p_1 - p_2}{p_2})^{-0.2} = \left(\frac{p_1}{p_2}\right)^{1.8}, \tag{3}$$

where the last inequality follows because at the start of the phase we have $\mathsf{bias} = p_1 - p_2 \geq \sqrt{C\log n/n}$. We next argue that throughout rounds $r \in \{2, \ldots, R\}$, this ratio between the plurality opinion and the second opinion remains almost the same, except for minimal deviations similar to above which would be tolerable and still leave the $\mathtt{gap}$ amplified to at least $(p_1/p_2)^{1.4}$.

First, we provide an initial intuition about the change in the number of decided nodes of each of these opinions throughout rounds $r \in \{2, \ldots, R\}$ and the potential deviations. Let us use (the random variable) q_r to denote the fraction of undecided nodes at the end of round r. For each r, we expect $q_r x_r$ and $q_r y_r$ new nodes to adopt, respectively, the plurality opinion and the second largest opinion. Thus, $\mathbb{E}[x_{r+1}] = x_r(1+q_r)$ and $\mathbb{E}[y_{r+1}] = y_r(1+q_r)$. Here, the expectation is based on the randomness in round $r+1$. If the random variables happen to be sharply around their expectations, the ratio remains essentially preserved. We next argue that this is true even despite the possible deviations.

We break the argument into two parts: Let r^* be the first round in which the fraction of undecided nodes is below 1/2—i.e., r^* is the minimum r satisfying that $q_r < 1/2$. We first show in Claim 2.3 that the ratio remains essentially preserved by round r^*, and then we use a reasoning with a similar flavor to argue in Claim 2.4 that a similar thing happens even for up to round R.

Claim 2.3. *W.h.p., we have* $\frac{x_{r^*}}{y_{r^*}} \geq (\frac{p_1}{p_2})^{1.7}$.

Proof. For each round $r \in \{1, \ldots, r^*\}$, Chernoff bound shows that the random variables of how many nodes get added to each of the opinions can deviate from their expectations by at most $\pm(\sqrt{5x_r q_r \log n})$ and $(\pm\sqrt{5y_r q_r \log n})$. Here, we have used that $x_r q_r \geq x_1/2 \geq 5\log n$ and $y_r q_r \geq y_1/2 \geq 5\log n$. Hence, we can say w.h.p.,

$$\frac{x_{r+1}}{y_{r+1}} \geq \frac{x_r(1+q_r) - \sqrt{5x_r q_r \log n}}{y_r(1+q_r) + \sqrt{5y_r q_r \log n}} \geq \frac{x_r}{y_r} \cdot \left(\frac{1 - \sqrt{5q_r \log n/x_r}}{1 + \sqrt{5q_r \log n/y_r}}\right).$$

Let us define $\mathtt{DEV}(x_r) = \sqrt{5q_r \log n/x_r}$ and $\mathtt{DEV}(y_r) = \sqrt{5q_r \log n/y_r}$. Then, we can write

$$\frac{x_{r^*}}{y_{r^*}} \geq \frac{x_1}{y_1} \prod_{r=1}^{r^*} \left(\frac{1 - \mathtt{DEV}(x_r)}{1 + \mathtt{DEV}(y_r)}\right).$$

Recall that we want to argue that despite the deviations of each round $r \in \{1, \ldots, r^*\}$, the ratio remains essentially preserved. For that, we need to analyze the deviation product in the above inequality. Noting that since p_1 and p_2 are both at least $\Omega(\sqrt{\log n/n})$, we know that for each r, $\mathtt{DEV}(x_r)$ and $\mathtt{DEV}(y_r)$ are both smaller than 0.1. This allows us to write

$$\prod_{r=1}^{r^*} \left(\frac{1 - \mathtt{DEV}(x_r)}{1 + \mathtt{DEV}(y_r)}\right) \geq 4^{-2(\sum_{r=1}^{r^*} \mathtt{DEV}(x_r) + \sum_{r=1}^{r^*} \mathtt{DEV}(y_r))}$$

Thus, we need to upper bound the summations $\sum \mathtt{DEV}(x_r)$ and $\sum \mathtt{DEV}(y_r)$. For convenience in doing this without repeating the argument, let us use $z \in \{x, y\}$ and consider the z_r for $r \in \{1, \ldots, r^*\}$. We next show that w.h.p., $\sum_r^{r^*} \mathtt{DEV}(z_r) = \sum_{r=1}^{r^*} \sqrt{5q_r \log n/z_r} = O(\frac{\sqrt{\log n}}{\sqrt{z_1}})$. During rounds $r \in \{2, r^*\}$, in each round we expect z_{r+1} to be $z_r(1+q_r) \geq 3z_r/2$. Since $z_r q_r = \Omega(\log n)$, with high probability, we will indeed have $z_{r+1} \geq 4z_r/3$. Hence, we have $\sum_{r=1}^{r^*} \sqrt{5q_r \log n/z_r} \leq \sum_{r=1}^{r^*} \sqrt{5 \log n/z_r} < O(\sqrt{5 \log n/z_1})$ where the last inequality holds because, during these rounds, the sum is a geometric series with a constant decay factor. Now we can use the inequality established in the above paragraph to write:

$$
\begin{aligned}
\frac{x_{r^*}}{y_{r^*}} &\geq \frac{x_1}{y_1} \prod_{r=1}^{r^*} \left(\frac{1 - \mathtt{DEV}(x_r)}{1 + \mathtt{DEV}(y_r)} \right) \geq \frac{x_1}{y_1} 4^{-2(\sum_{r=1}^{r^*} \mathtt{DEV}(x_r) + \sum_{r=1}^{r^*} \mathtt{DEV}(y_r))} \geq \frac{x_1}{y_1} 4^{-O(\sqrt{\log n/x_1} + \sqrt{\log n/y_1})} \\
&\geq \frac{x_1}{y_1}(1 - O(\sqrt{\log n/x_1} + \sqrt{\log n/y_1})) \geq (\frac{p_1}{p_2})^{1.8}(1 - O(\sqrt{\log n/np_1^2} + \sqrt{\log n/np_2^2})) \\
&\geq (\frac{p_1}{p_2})^{1.8}(1 + (p_1 - p_2)/p_2)^{-0.1} = (\frac{p_1}{p_2})^{1.7}
\end{aligned}
$$

□

Claim 2.4. *W.h.p., we have* $\frac{x_R}{y_R} \geq (\frac{p_1}{p_2})^{1.4}$.

The proofs of Claim 2.4, case (P.b), and property (S2) are deferred to Appendix B. □

Lemma 2.5. *After $O(\log n)$ phases from the start, we have* $\mathtt{gap} \geq 2$. *If initially we had $p_1 > (1+\delta)p_2$ for some constant δ, then if fact within $O(1)$ phases we would have* $\mathtt{gap} \geq 2$.

Lemma 2.6. *If at the start of a phase $p_1 \geq \frac{2}{3}$, then w.h.p. we have $p_1^{new} \geq \frac{2}{3}$ also at the end of it.*

Lemma 2.7. *After* $\mathtt{gap} \geq 2$, *within $O(\log \log n)$ phases, w.h.p., we have $p_1 \geq 2/3$ and $\sum_{i \geq 2} p_i = 0$.*

Lemma 2.8. *After $p_1 \geq 2/3$ and $\sum_{i \geq 2} p_i = 0$, within $O(\log n/\log k)$ phases, w.h.p., $p_1 = 1$.*

3 The Algorithm: Take 2

In the algorithm presented in the previous section, each node has a memory size of $\log k + \log \log k + O(1)$ bits, where $\log \log k + O(1)$ bits of it are used solely to remember the round number, modulo $R = O(\log k)$. Here, we explain how to remove this $\log \log k$ overhead. Thus, we get a dynamics in which each node has at most $O(k)$ states—i.e., within constant factor of the trivially optimal—instead of the $O(k \log k)$ states of the previous algorithm.

Intuitive Discussions: If we are to reduce the memory from $\log k + \log \log k + O(1)$ to $\log k + O(1)$, we cannot require each node to know both an opinion and the round number modulo R. Nodes not knowing the time create the risk that nodes might not be "*in sync*" in regards to whether at a given point of time, they should be performing *gap amplification* or *healing*. We try to minimize this effect and keep the vast majority of the nodes essentially "*in sync*".

Notice that when we allow memory size of $\log k + O(1)$, we have (in fact far more than) enough space for keeping the time modulo R; we just cannot ask each node to keep both time and an opinion. To overcome this, we simply *split the responsibilities among the nodes*: particularly, at the start, each node tosses a fair random coin and decides to be a *clock-node* or a *game-player*, each with probability 1/2. Clock-nodes forget their initial opinion and instead keep the time (say modulo $4R$, thus needing at most $\log k + O(1)$ bits). On the other hand, the game-players will try to follow the approach of the take 1 algorithm, using the help of the clock-nodes.

There are three subtleties this intuitive idea: First, note that now the *game-players*, who actually play the main role in getting to plurality consensus, do not know the time and thus cannot perform

Algorithm 1 The algorithm run at each game-player node v

```
In each round, node v does as follows (assuming u is the node that v contacted):
if u.isClock = true then
    if phase_v ≠ end-game or (phase_v = end-game and phase_u = 0) then
        phase_v ← phase_u;
else
    switch phase_v do
        case 0:                                                    - Time Buffer 1-
            sampled ← false; forget ← false;
        case 1:                                  - Gap Amplification (Sampling) -
            if sampled = false and v.opinion ≠ u.opinion then forget ← true;
            sampled ← true;
        case 2:                                                    - Time Buffer 2-
            if forget = true then v.opinion ← 0; forget ← false;
        case 3:                                                        - Healing -
            if v.opinion = 0 then v.opinion ← u.opinion;
                sampled ← false; forget ← false;
        case end-game:                                                - End Game -
            if v.opinion ≠ 0 and v.opinion ≠ u.opinion then v.opinion ← 0;
            if v.opinion = 0 then v.opinion ← u.opinion;
```

the take 1 algorithm *per se*. Furthermore, they cannot really expect to receive the precise time from the clock-nodes and remember it, as that would again require $\log k + \log\log k + O(1)$ memory bits. We will make the clock nodes report the time only in (integer) factors of $R = O(\log k)$, and modulo say $4R$. That is, the reported phase number is always in $\{0, 1, 2, 3\}$. Reporting these numbers and remembering them will thus require only $O(1)$ bits. Second, now the *game-players* do not really know the precise time, and thus they cannot update the phase numbers on their own. They will have to instead rely completely on receiving phase-updates from *clock-nodes*. Third, we want that eventually all the nodes to hold the plurality opinion, including the *clock-nodes*. Hence, each clock-node needs to at some point forget it's *time-keeping responsibilities* and try to learn the plurality opinion. This has to be done in a judicious manner so that the system as a whole has enough time-keepers, so long as they are needed. We next present the algorithm that deals with these challenges.

The Algorithm Description: The algorithm performed by the game-player nodes is described in Algorithm 1 and the algorithm performed by the clock-nodes is presented in Algorithm 2.

We next provide some intuitive explanation for what these algorithms are doing. The algorithm run the by game-player nodes can be viewed as *interpolating* between two algorithms: the dynamic presented in the previous section, which we will call hereafter the Gap Amplification (`GA`) dynamic, and the Undecided State (`Undecided`) dynamic[7] as used in [BCN+15a]. We have the `Undecided` dynamic for the time that the number of clock-nodes that are following their time-keeping responsibilities goes down, and thus we cannot follow the `GA` dynamic. In reality, as long as p_1 is far from 1, there will be a large population of clock-nodes that are still performing their time-keeping responsibilities, and thus the `GA` dynamic will be the dominant part of the system (formalizing this will take some care). Hence, during that time, the system will enjoy the fast gap amplification behavior of `GA`. Once p_1 gets close to 1, the number of active clock-nodes drops and the game-players gradually move to the `Undecided` dynamic. At this time, we no longer need the power of the `GA` dynamic; regardless of how game-player nodes are split between the two dynamics, as we will show, the system evolves almost as good as the `Undecided` dynamic and gets to plurality consensus fast.

Let us now discuss the `GA` dynamic, which is the core of the algorithm of the game-players, and

[7] We note that in fact any other protocol with a minimal plurality amplification behavior would suffice for this part; we choose the `Undecided` dynamics just for convenience.

Algorithm 2 The algorithm run at each clock-node v

```
In each round, node v does as follows (assuming u is the node that v contacted):
switch v.status do
    case counting:                                                          - Time Keeping -
        v.opinion ← null;
        v.time ← v.time + 1 mod 4R;
        v.phase ← ⌈v.time/R⌉ mod 4;
        if u.isClock = false and u.opinion = 0 then v.consensus ← false;
        if u.isClock = true and u.consensus = false then v.consensus ← false;
        if v.time mod 4R = 0 then
            if v.consensus = true then v.status ← end-game;
            v.consensus ← true;
    case end-game:                                                          - End Game -
        v.time ← null;
        v.phase ← end-game;
        if u.isClock = false then v.opinion ← u.opinion
        else if u.status = 0 and u.consensus = false then
            v.status ← counting;
            v.opinion ← null;
            v.time ← u.time; v.phase ← u.phase; v.consensus ← u.consensus;
```

it is essentially a simulation of the take 1 algorithm. We have a few small changes, in comparison to the take 1 algorithm, which are added to deal with the subtleties described above: Particularly, since the game-player nodes will not be able to know the precise time, to mitigate the effects of asynchrony, we extend the *gap amplification* to a full phase (instead of one round) and we also add some time buffer between the *gap amplification* period and the *healing* period. Thus, now the execution of the algorithm can be broken to intervals made of 4 consecutive phases, which we call a "long-phase". Particularly, a long-phase is made of 4 phases, the two main ones being gap amplification and healing phases, interspersed with two (essentially) time buffer phases. Although the gap amplification period is now one full phase, we still make each node act based on a single sample: that is, in the first round in this phase that the game-player node v contacts another game-player u, node v decides whether to keep its opinion or not, and it remains with this decision. If v decided to forget its opinion, then it remembers this decision (via a Boolean variable called $forget$) and once v sees the next phase (by meeting a clock-node of phase 2), it becomes undecided. With a slight informality, from now on, we will refer to the dynamic described by phases $\{0, 1, 2, 3\}$ as the `GA` (gap amplification) dynamic. A game-player node goes out of the `GA` dynamic only if it meets a clock-node that has moved to the end-game (to be explained while discussing the algorithm of the clock-nodes). Game-players that move out of the `GA` dynamic go to the `Undecided` dynamic, as described the pseudocode in the `end-game` switch case. However, each such game-player node might come back to the `GA` dynamic later on, if it meets a clock-node that reports phase 0.

Let us now discuss the algorithm for the clock-nodes. As mentioned above, an important challenge is for the clock-nodes to recognize that the game-players have reached consensus, so that they can switch to adopt an opinion then. Since the `GA` dynamics might take $O(\log n \log k)$ rounds to reach plurality consensus, we cannot require the clock-nodes to keep the precise time until this period has passed, as that would require $\Omega(\log \log n)$ memory bits. Instead, we make the clock-nodes use the existence of undecided nodes as a signal to detect that the game-players still have not reached consensus. Although this is not a perfect check, it will be good enough for our purposes, i.e., so long as p_1 is far from 1. Particularly, note that in the `GA` dynamic, so long as we are far from reaching consensus (while p_1 is far from 1), there will be a considerable fraction of undecided nodes (at some point during at least one of the phases). We make the clock-nodes remain active—i.e., keeping time—as long as they meet an undecided node directly or indirectly. Here, indirectly means by meeting a clock-node that is aware of the existence of undecided nodes. This information spreading essentially in the usual gossip spreading style and via the boolean variable *consensus*, for which $consensus = false$ indicates that an undecided node still exists, i.e., an indication for that the system has not reached plurality consensus yet. If during a long-phase, a clock-node receives no information about the existence of undecided nodes, it neglects its time keeping task, by forgetting time, and attempts to move to the end-game reporting phase (possibly temporarily). Here, the phase number it reports is a special symbol "end-game". Also, at this time, this clock-node that now has forgotten time adopts an opinion, always remembering the opinion of the last game-player that it met. However, this clock-node that has become "deactivated" from time-keeping by switching to the end-game mode might get reactivated again: particularly, if this "deactivated" clock-node v meets an active clock-node u who is aware of the existence of undecided nodes (has $u.consensus = false$), then v gets reactivated again, keeping track of time, using the clock that it gets from u, and reporting the phase number in $\{0, 1, 2, 3\}$.

The analysis of the algorithm are deferred to Appendix C. We next provide some intuition for the analysis, for the cases in which $k \geq n^{0.1}$. The analysis for the complementary case of $k < n^{0.1}$ also follows a similar outline, roughly speaking, but typically requires more care.

Analysis Intuition, for when $k \geq n^{0.1}$: Here the length of each phase is $R = \Theta(\log k) = \Theta(\log n)$, which allows us to say that some good events happen w.h.p. without needing to keep track of their precise probability, e.g., this is sufficient time for information to spread to the whole system.

Specifically, the first analysis step is to establish that as long as $p_1 \leq 1 - \Theta(\log n/n)$, all clock-nodes will continue to perform their time-keeping responsibilities. Thus, during this period, all game-player nodes will be performing the `GA` dynamics without going to the end-game of the `Undecided` dynamics. The (formally inductive) argument for why all clock-nodes will remain is roughly as follows: Consider one long-phase (4 consecutive phases). Because game-player nodes are following the `GA` dynamic and there are $\Omega(\log n)$ non-plurality nodes, during phase 2 of this long-phase, there will be $\Theta(\log n)$ undecided nodes w.h.p. Thus, w.h.p., there will be clock-nodes that notice this undecided population. This information about the existence of undecided nodes will spread among the clock-nodes following essentially the pattern of a usual rumor spreading. Particularly, w.h.p. within $O(\log n)$ rounds all clock-nodes will be aware of the existence of undecideds and thus, at the end of the long-phase, none of the clock-nodes switches to the end-game mode. Given this, during this time of $p_1 \leq 1 - \Theta(\log n/n)$, all game-players are acting according to the `GA` dynamic. This allows us to use the analysis of take 1, modulo minor details, and argue that in $O(\log n)$ phases, w.h.p., we reach $p_1 \geq 1 - \Theta(\log n/n)$.

Once p_1 reaches $1 - \Theta(\log n/n)$, some clock-nodes might start to switch to the end-game mode and thus some game-player nodes might act based on the `Undecided` dynamic. However, now that p_1 is very close to 1, we do not need the power of `GA`. Here, even the `Undecided` dynamic would be pushing us towards consensus on plurality quickly enough, and we can actually prove that the system (which is an uncontrolled mix of `GA` and `Undecided`) does not behave worse than the worst of them. Particularly, we can first see that w.h.p. the system remains in this regime of p_1 very close to 1, and then use this to establish that non-plurality opinions get filtered out within $O(1)$ phases and that plurality reaches totality among game-players within $O(1)$ phases. By the end of that time, all clock-nodes switch to the end-game, as they will not see undecided game-players anymore. Since clock-nodes in the end-game adopt the opinion of the last game-player they have seen, within $O(1)$ phases, all nodes will hold the plurality opinion.

References

[AAD+06] Dana Angluin, James Aspnes, Zoë Diamadi, Michael J Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed computing*, 18(4):235–253, 2006.

[AAE08] Dana Angluin, James Aspnes, and David Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, 2008.

[AFJ06] Dana Angluin, Michael J Fischer, and Hong Jiang. Stabilizing consensus in mobile networks. In *Distributed Computing in Sensor Systems*, pages 37–50. Springer, 2006.

[AGV15] Dan Alistarh, Rati Gelashvili, and Milan Vojnovic. Fast and exact majority in population protocols. In *the Proc. of the Int'l Symp. on Princ. of Dist. Comp. (PODC)*, pages 47–56, 2015.

[BBV08] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.

[BCN+14] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, Riccardo Silvestri, and Luca Trevisan. Simple dynamics for plurality consensus. In *Proceedings of the 26th ACM symposium on Parallelism in algorithms and architectures*, pages 247–256. ACM, 2014.

[BCN+15a] L Becchetti, A Clementi, E Natale, F Pasquale, and R Silvestri. Plurality consensus in the gossip model. In *Pro. of ACM-SIAM Symp. on Disc. Alg. (SODA)*, pages 371–390, 2015.

[BCN+15b] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Stabilizing consensus with many opinions. *to appear at SODA'16, arXiv preprint arXiv:1508.06782*, 2015.

[BICS10] Radu Berinde, Piotr Indyk, Graham Cormode, and Martin J Strauss. Space-optimal heavy hitters with strong error bounds. *ACM Transactions on Database Systems (TODS)*, 35(4):26, 2010.

[BSDDS14] Ohad Ben-Shahar, Shlomi Dolev, Andrey Dolgin, and Michael Segal. Direction election in flocking swarms. *Ad Hoc Networks*, 12:250–258, 2014.

[CCN12] Luca Cardelli and Attila Csikász-Nagy. The cell cycle switch computes approximate majority. *Scientific reports*, 2, 2012.

[CDS+13] Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature nanotechnology*, 8(10):755–762, 2013.

[CKFL05] Iain D Couzin, Jens Krause, Nigel R Franks, and Simon A Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–516, 2005.

[CSWB09] Matthew Cook, David Soloveichik, Erik Winfree, and Jehoshua Bruck. Programmability of chemical reaction networks. In *Algorithmic Bioprocesses*, pages 543–584. Springer, 2009.

[DGM+11] Benjamin Doerr, Leslie Ann Goldberg, Lorenz Minder, Thomas Sauerwald, and Christian Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures*, pages 149–158. ACM, 2011.

[Dot14] David Doty. Timing in chemical reaction networks. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 772–784. SIAM, 2014.

[DV12] Moez Draief and Milan Vojnovic. Convergence speed of binary interval consensus. *SIAM Journal on Control and Optimization*, 50(3):1087–1109, 2012.

[DW83] Peter Donnelly and Dominic Welsh. Finite particle systems and infection models. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 94, pages 167–182. Cambridge Univ Press, 1983.

[HHMW10] Nils Lid Hjort, Chris Holmes, Peter Müller, and Stephen G Walker. *Bayesian nonparametrics*, volume 28. Cambridge University Press, 2010.

[Hop71] John Hopcroft. An $n \log n$ algorithm for minimizing states in a finite automaton. *Theory of Machines and Computations*, pages 189–196, 1971.

[HP01] Yehuda Hassin and David Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Information and Computation*, 171(2):248–268, 2001.

[KDG03] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proc. of the Symp. on Found. of Comp. Sci. (FOCS)*, pages 482–491, 2003.

[Lig12] Thomas Liggett. *Interacting particle systems*, volume 276. Springer Science & Business Media, 2012.

[Lyn96] Nancy A Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.

[MNRS14] George B Mertzios, Sotiris E Nikoletseas, Christoforos L Raptopoulos, and Paul G Spirakis. Determining majority in networks with local interactions and very small local memory. In *Automata, Languages, and Programming*, pages 871–882. Springer, 2014.

[MS] Elchanan Mossel and Grant Schoenebeck. Reaching consensus on social networks.

[PVV09] Etienne Perron, Dinkar Vasudevan, and Milan Vojnovic. Using three states for binary consensus on complete graphs. In *the Proc. of IEEE INFOCOM*, pages 2527–2535, 2009.

[SKJ+08] David JT Sumpter, Jens Krause, Richard James, Iain D Couzin, and Ashley JW Ward. Consensus decision making by fish. *Current Biology*, 18(22):1773–1777, 2008.

[Spe64] H. Spencer. *The Principles of Biology*. Number v. 1 in Spencer, Herbert: A system of synthetic philosophy. Williams and Norgate, 1864.

[Spr78] DA Sprott. Urn models and their applicationan approach to modern discrete probability theory. *Technometrics*, 20(4):501–501, 1978.

[TBA86] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE transactions on automatic control*, 31(9):803–812, 1986.

[Tsi84] John Nikolas Tsitsiklis. Problems in decentralized decision making and computation. Technical report, DTIC Document, 1984.

[WH04] Fang Wu and Bernardo A Huberman. Social structure and opinion formation. *arXiv preprint cond-mat/0407252*, 2004.

[Yu05] Sheng Yu. State complexity: Recent results and open problems. *Fundamenta Informaticae*, 64(1-4):471–480, 2005.

A A remark regarding other rich get richer mechanisms

The above plurality amplification might resemble the well known *rich get richer* effect, a recurrent theme in the dynamics of many real life systems. Such phenomena are usually explained by the existence of some self-reinforcing mechanism which provides the system some sort of a positive feedback. Examples include Polya's urn [Spr78], the Dirichlet process, the Chinese restaurant [HHMW10] and the preferential attachment processes [BBV08]. These processes usually share the following cascade: starting from an empty population, in each step a new individual arrives and its destiny (e.g., which opinion to hold) is determined by a random choice which has some *bias* towards the stronger trends in the current system. Our plurality algorithm follows the general theme of rich-get-richer processes but deviates from the aforementioned mechanisms in the sense that our population size is fixed and the amplification is done within the "closed" system. We note that the interplay between selection and healing presented in this paper, and the fast amplification provided by them, may be of broader interest for the study of rich-get-richer mechanisms in "closed" systems.

B Missing Proofs of Section 2

Theorem B.1. ***(Simple Corollary of Chernoff Bound)*** *Suppose $X_1, X_2, \ldots, X_\ell \in [0,1]$ are independent random variables, and let $X = \sum_{i=1}^{\ell} X_i$ and $\mu = \mathbb{E}[X]$. If $\mu \geq 5\log n$, then w.h.p. $X \in \mu \pm \sqrt{5\mu \log n}$, and if $\mu < 5\log n$, then w.h.p. $X \leq \mu + 5\log n$.*

Proof of Claim 2.4. Notice that even in these rounds, it is true that per round r, $\mathbb{E}[x_{r+1}] = x_r(1+q_r)$ and $\mathbb{E}[y_{r+1}] = y_r(1+q_r)$. However, here $x_r q_r$ and $y_r q_r$ might drop below $O(\log n)$, which means every time when we write the additive deviations, we should still add a possible $O(\log n)$ term. Particularly, for each $r \in \{r^*+1, \ldots, R\}$ we can say w.h.p.,

$$\frac{x_{r+1}}{y_{r+1}} \geq \frac{x_r(1+q_r) - (\sqrt{3x_r q_r \log n} + 5\log n)}{y_r(1+q_r) + (\sqrt{5y_r q_r \log n} + 5\log n)} \geq \frac{x_r}{y_r} \cdot \left(\frac{1 - (\sqrt{5q_r \log n / x_r} + 5\log n / x_r)}{1 + \sqrt{5q_r \log n / y_r} + 5\log n / y_r}\right).$$

Let us define $\texttt{DEV}(x_r) = \sqrt{5q_r \log n / x_r} + 5\log n / x_r$ and $\texttt{DEV}(y_r) = \sqrt{5q_r \log n / y_r} + 5\log n / y_r$, for $r \in \{r^*+1, \ldots, R\}$. Then, we can write

$$\frac{x_R}{y_R} \geq \frac{x_{r^*}}{y_{r^*}} \prod_{r=r^*+1}^{R} \left(\frac{1 - \texttt{DEV}(x_r)}{1 + \texttt{DEV}(y_r)}\right).$$

Noting that since p_1 and p_2 are both at least $\Omega(\sqrt{\log n / n})$, we know that for each r, $\texttt{DEV}(x_r)$ and $\texttt{DEV}(y_r)$ are both smaller than 0.1. This allows us to write

$$\prod_{r=r^*+1}^{R} \left(\frac{1 - \texttt{DEV}(x_r)}{1 + \texttt{DEV}(y_r)}\right) \geq 4^{-(\sum_{r=r^*+1}^{R} \texttt{DEV}(x_r) + \sum_{r=r^*+1}^{R} \texttt{DEV}(x_r))}$$

Thus, we need to upper bound the summations $\sum \texttt{DEV}(x_r)$ and $\sum \texttt{DEV}(y_r)$ in the above inequality. Again, for convenience in doing this without repeating the argument, let us use $z \in \{x, y\}$ and consider the z_r for $r \in \{r^*+1, \ldots, R\}$. We next show that w.h.p.,

$$\sum_{r=r^*+1}^{R} \texttt{DEV}(z_r) = \sum_{r=r^*+1}^{R} \sqrt{5q_r \log n/z_r} + 5\log n/z_r = O(\frac{\sqrt{\log n}}{\sqrt{z_{r^*}}} + \frac{\log k \log n}{z_{r^*}}).$$

First, note that the z_r values cannot decrease in the healing phase, and hence for each for $r' \in \{r^*, \ldots, R\}$, we have $z_{r'} \geq z_{r^*}$. Thus, $\sum_{r=r^*}^{R}(\sqrt{5\log n q_r/z_r} + 5\log n/z_r) \leq \sqrt{5\log n/z_{r^*}} \cdot \sum_{r=r^*+1}^{R}(\sqrt{q_r}) + 5R\log n/z_{r^*}$. We next bound the term $\sum_{r=r^*+1}^{R}(\sqrt{q_r})$. Note that in round r^*+1, we expect the number of undecided nodes to change from $n/2$ to $n/4$ and with high probability, it will be at most $n/3$. Similarly, so long as $nq_r \geq \Omega(\log n)$, with high probability, we have $q_{r+1} \leq 3q_r/4$. Thus, the summation $\sum_{r=r^*+1}^{R}(\sqrt{q_r})$ is also a summation of a geometric series, so long as $q_r \geq \Omega(\log n/n)$. For rounds in which $q_r \leq O(\log n/n)$, the summation of $\sqrt{q_r}$ is clearly at most $O(\log^2 n/n) = o(1)$. Therefore, we get that $\sum_{r=r^*}^{R} \sqrt{q_r} \leq 4q_{r^*} + o(1) \leq 5$. This means $\sum_{r=r^*+1}^{R}(\sqrt{5\log n q_r/z_r} + 5\log n/z_r) \leq 10\sqrt{\log n}/\sqrt{z_{r^*}} + 5R\log n/z_{r^*}$.

From the above, it thus follows that the $\sum_{r=r^*+1}^{R} \texttt{DEV}(x_r) + \sum_{r=r^*+1}^{R} \texttt{DEV}(x_r) \leq O(\frac{\sqrt{\log n}}{\sqrt{x_{r^*}}} + \frac{\log k \log n}{x_{r^*}}) + O(\frac{\sqrt{\log n}}{\sqrt{y_{r^*}}} + \frac{\log k \log n}{y_{r^*}})$. We know that w.h.p. $x_{r^*} = \Omega(np_1)$ and $y_{r^*} = \Omega(np_2^2/p_1)$. The reason for this is as follows: Note that at round r^*, the total fraction of decided nodes has dropped below $1/2$, which means, ignoring a 2-factor, x_{r^*} is the same as the fraction among decided nodes that have opinion 1. Now, at the end of the gap amplification round, this fraction is w.h.p. $\frac{\Theta(np_1^2)}{\Theta(\sum i=1^k np_i^2)} \geq \frac{\Theta(np_1^2)}{\Theta(np_1)} \geq \Omega(p_1)$. Now, during the rounds $r \in \{1, \ldots, r^*\}$, even in a coarse and pessimistic estimate, this fraction deviates by at most a constant factor, because of the deviation factor analysis that we provided above for rounds $r \in \{1, \ldots, r^*\}$. Hence, in round r^*, about $\Omega(p_1)$ fraction of decided nodes have opinion 1, which since at r^* more than $n/2$ nodes are decided, means $x_{r^*} = \Omega(np_1)$. A similar argument can be used to show that $y_{r^*} = \Omega(n\frac{p_2^2}{\sum i=1^k p_i^2}) = \Omega(np_2^2/(p_1^2+p_2)) = \Omega(\min\{np_2^2/p_1^2, np_2\})$.

For simplicity, let us now distinguish between two cases, regarding the value of p_1/p_2. First, we consider the possibly sensitive regime when $p_1/p_2 \in [1, 500]$. In the complementary case, the concentrations we need will be far more relaxed. For the case $p_1/p_2 \in [1, 500]$, from above we get that

$$\begin{aligned}
\frac{x_R}{y_R} &\geq \frac{x_{r^*}}{y_{r^*}} \cdot \prod_{r=r^*+1}^{R} \left(\frac{1-\texttt{DEV}(x_r)}{1+\texttt{DEV}(y_r)}\right) \geq \frac{x_{r^*}}{y_{r^*}} \cdot 4^{-(\sum_{r=r^*+1}^{R} \texttt{DEV}(x_r) + \sum_{r=r^*+1}^{R} \texttt{DEV}(x_r))} \\
&\geq \frac{x_{r^*}}{y_{r^*}} \cdot 4^{-O(\frac{\sqrt{\log n}}{\sqrt{x_{r^*}}} + \frac{\log k \log n}{x_{r^*}} + \frac{\sqrt{\log n}}{\sqrt{y_{r^*}}} + \frac{\log k \log n}{y_{r^*}})} \\
&\geq \frac{x_{r^*}}{y_{r^*}} \cdot 4^{-O(\frac{\sqrt{\log n}}{\sqrt{np_1}} + \frac{\log k \log n}{np_1} + \frac{\sqrt{\log n}}{\sqrt{np_2}} + \frac{\log k \log n}{np_2})} \\
&\geq \frac{x_{r^*}}{y_{r^*}} \cdot 4^{-O(\frac{\sqrt{\log n}}{\sqrt{np_2}} + \frac{\log k \log n}{np_2})} \geq \frac{x_{r^*}}{y_{r^*}} \cdot 4^{-O(\frac{\sqrt{\log n}}{\sqrt{np_2^2}})} \geq \frac{x_{r^*}}{y_{r^*}} \cdot (1 - O(\frac{\sqrt{\log n}}{\sqrt{np_2^2}})) \\
&\geq (\frac{p_1}{p_2})^{1.7}(1 + \frac{p_1 - p_2}{p_2})^{-0.3} = (\frac{p_1}{p_2})^{1.4}
\end{aligned}$$

Now suppose that $p_1/p_2 > 500$. Here, it suffices to bound the deviations by a small constant. It is easy to see that w.h.p. $y_R \leq 5y_{r^*} + O(\log^2 n)$, because during these rounds, as discussed above, q_r decays exponentially with r. Thus, $\frac{x_R}{y_R} \geq \frac{x_{r^*}}{y_R} \geq \frac{x_{r^*}}{5y_{r^*} + O(\log^2 n)} \geq \frac{x_{r^*}}{y_{r^*}}/(5 + o(1)) > (\frac{p_1}{p_2})^{1.7}/5.1 \geq (\frac{p_1}{p_2})^{1.4}$. This completes the proof of (P3.a). Next, we consider the complementary case where p_2 is small. $\square$

Case (P.b), when $p_2 \le \sqrt{10 \log n/n}$: Now we come back to the case that $p_2 = \max_{i=2}^k p_i \le \sqrt{10\log n/n}$. Note that in this case, the multiplicative factor deviations of the opinion 2 from its expected values might be large. Recall the definition $\mathsf{gap} = \min\{\frac{p_1}{\sqrt{10\log n/n}}, \min_{i=2}^k \frac{p_1}{p_i}\}$. Also, note that by the above arguments, the ratio between $\frac{p_1}{p_2}$ will grow to at least $\mathsf{gap}^{1.4}$ because that would happen even if we manually raise the initial p_2 to $\sqrt{10\log n/n}$, thus allowing us to use the above concentration arguments, and the whole process is clearly monotonic meaning that raising the initial p_2 cannot increase the ratio $\frac{p_1^{new}}{p_2^{new}}$. However, to show that gap_{new} increases to $\mathsf{gap}^{1.4}$, in this regime of $p_2 \le \sqrt{10\log n/n}$, the minimizer in the definition of gap is the $\frac{p_1}{\sqrt{10\log n/n}}$ term which means we should show that p_1 will grow accordingly.

Here, since $p_1 \ge \sqrt{C\log n/n}$, we have $\mathsf{gap} \ge \frac{\sqrt{C\log n/n}}{\sqrt{10\log n/n}} \gg 20$. On the other hand, by repeating the arguments given above, we know that p_1^{new} will be within at most a $1 \pm o(1)$ factor of its expectation, which is $\frac{p_1^2}{\sum_{i=1}^k p_i^2} = \frac{p_1^2}{p_1^2 + \sum_{i=2}^k p_i^2} \ge \frac{p_1^2}{p_1^2 + p_{i^*}}$. Thus, the growth factor from p_1 to p_1^{new} is at least $(1 - o(1))\frac{p_1}{p_1^2+p_2} \ge \frac{1-o(1)}{p_1 + 1/\mathsf{gap}}$. If $1/\mathsf{gap} \le 0.45 p_1$, then $p_1^{new} > 2/3$. Otherwise, the growth is at least $\frac{1-o(1)}{3.3/\mathsf{gap}} \ge \mathsf{gap}/4 \ge \mathsf{gap}^{1/2}$, which finishes the proof of case (P.b) and thus also property (P).

Property (S2): Finally, we use the above concentration arguments to prove property (S2), by showing that w.h.p. $\mathsf{bias}^{new} \ge p_1 - p_2$. First suppose that $\frac{p_1}{p_2} < 10$. Then, we know that p_1^{new} and p_2^{new} will be within $1 + o(1)$ factors of their expectations (as analyzed above), which are, respectively, $\frac{p_1^2}{\sum_{j=1}^k p_j^2}$ and $\frac{p_1^2}{\sum_{j=1}^k p_j^2}$. Thus,

$$p_1^{new} - p_2^{new} \ge (1 - o(1))\frac{p_1^2}{\sum_{j=1}^k p_j^2} - \frac{p_1^2}{\sum_{j=1}^k p_j^2} = (1 - o(1))(p_1 - p_2) \cdot \frac{(p_1 + p_2)}{\sum_{j=1}^k p_j^2}$$
$$\ge (1 - o(1))(p_1 - p_2) \cdot \frac{(p_1 + p_i)}{p_1 \sum_{j=1}^k p_j} \ge (1 - o(1))(p_1 - p_2) \cdot \frac{(p_1 + p_2)}{p_1} \ge (1.1 - o(1))(p_1 - p_2).$$

On the other hand, when $\frac{p_1}{p_2} > 10$, we know the $\frac{p_1^{new}}{p_2^{new}} \ge \mathsf{gap}^{new} \ge \mathsf{gap}^{1.4} > 2.5\mathsf{gap}$. We also know that w.h.p. p_1 does not decrease in this case. This implies that $\mathsf{bias} = p_1^{new} - p_2^{new}$ also does not decrease, thus proving property (S2).

Proof of Lemma 2.5. Initially, $\mathsf{gap} \ge \min\{\frac{p_1}{\sqrt{10\log n/n}}, \min_{i=2}^k \frac{p_1}{p_i}\} \ge \min\{3, \frac{p_1}{p_2}\} > 1 + \frac{1}{\sqrt{n}}$. Note that once $p_1 \ge 2/3$, we also have $\mathsf{gap} \ge 2$. Define $\gamma = \mathsf{gap} - 1$. In every phase that $p_1^{new} \ge 2/3$ does not happen, Lemma 2.2 gives $\mathsf{gap}^{new} \ge \mathsf{gap}^{1.4}$. This implies $\gamma^{new} \ge (1+\gamma)^{1.4} - 1 \ge 6\gamma/5$. Hence, after $O(\log n)$ phases, γ becomes larger than 1 which means $\mathsf{gap} \ge 2$. If initially we had $p_1 > (1+\delta)p_2$, then initially $\mathsf{gap} \ge 1 + \delta$ and thus it takes it only $O(1)$ phases to reach $\mathsf{gap} \ge 2$. □

Proof of Lemma 2.6. At the end of the gap amplification round, w.h.p., the fraction of decided nodes of opinion 1 is $(1 - o(1))4/9$ and the fraction of decided nodes of all other opinions is at most $(1+o(1))1/9$. That is, w.h.p., the fraction among decided nodes that have opinion 1 is at least $(1-o(1))\frac{4/9}{4/9+1/9} > 3/4$. During each round r of the healing rounds, both the fraction of plurality nodes and the total fraction of decided nodes are expected to grow by $(1+q_r)$, where q_r is the fraction of undecided nodes at the start of that round. Since the number of plurality opinion nodes and also the total number of decided nodes are $\Omega(n)$, the actual growth in the two fractions will be within $1 \pm o(\frac{1}{\log n})$ factor of their expectations. Thus, the fraction among the decided node that have opinion 1 remains preserved up to at most a $1 \pm o(\frac{1}{\log n})$ factor. This means, even after all the $O(\log k)$ rounds of healing, the fraction among the decided nodes that have opinion 1 is at least $(1-o(1/\log n))^{O(\log k)}3/4 = (1-o(1))3/4$. On the other hand, at that point the probability for a node to remain undecided is at most $(1/3)^{O(\log k)} \leq 0.01/k$, which means w.h.p., at least 0.99 fraction of nodes will be decided. Hence, w.h.p., at least $0.98 \times (1-o(1))3/4 > 2/3$ fraction of nodes will have opinion 1. □

Proof of Lemma 2.7. By Lemma 2.2, in each phase j, either we get $p_1 \geq 2/3$ or $\texttt{gap}^{new} \geq \texttt{gap}^{1.4}$. Since we start with a $\texttt{gap} \geq 2$, within $O(\log\log n)$ phases, the latter condition cannot continue to hold which means we would have $p_1 \geq 2/3$. At this point, by Lemma 2.6, it will always remain true that $p_1 \geq 2/3$.

Then, for any other opinion $i \in \{2, \ldots, k\}$, in each phase, the change satisfies the following inequality: $\mathbb{E}[p_i^{new}] = \frac{p_i^2}{\sum_{i=1}^k p_i^2} \geq \frac{p_i^2}{4/9} = 2.25p_i^2$. Here, the expectation is based only on the randomness of this phase. Hence, within $O(\log\log n)$ phases, we expect p_i to fall below $1/n^c$, for any desirable constant c. By Markov's inequality, this means at that point, we actually have $p_i = 0$ w.h.p. □

Proof of Lemma 2.8. Consider the first phase after we have $p_1 \geq 2/3$ and $\sum_{i\geq 2} p_i = 0$. First note that by Lemma 2.6, w.h.p. $p_1 \geq 2/3$ from now on. Now, let q be the fraction of undecided nodes. Notice that after the first round, q increases slightly to q' where $\mathbb{E}[q'] = 1-(1-q)^2 \leq 2q$. During the next $R-1 = O(\log k)$ rounds of this phase, a node remains undecided only if all nodes that it contacted where undecided and that has probability at most $(5/9)^{R-1} \leq 1/(4k)$. Thus, $\mathbb{E}[q^{new}] \leq \frac{q}{2k}$. This expectation relies only on the randomness of this phase. Thus, the expected value of q after $O(\log n/\log k)$ phases is $1/n^c$—for any arbitrary c, by adjusting the constant in the number of phases. Hence, by Markov's inequality, the probability that $q \neq 0$, which would imply $q \geq 1/n$, is at most $1/n^{c-1}$. That is, w.h.p., $q = 0$ and thus $p_1 = 1$. Clearly, from that point onwards, it remains true that $p_1 = 1$. □

Proof of Theorem 2.1. By Lemma 2.5, within $O(\log n)$ phases from the beginning of the execution, $\texttt{gap} \geq 2$. In fact if we initially have $p_1 \geq (1+\delta)p_2$ for a constant δ, this part takes $O(1)$ phases. Then, after additional $O(\log\log n)$ phases, by Lemma 2.5, we have $p_1 \geq 2/3$ and all other nodes are undecided. Finally, by Lemma 2.8, w.h.p., within $O(\log n/\log k)$ phases, plurality achieves totality. □

C Analysis of the Take 2 Algorithm, from Section 3

For simplicity, we divide the analysis into two cases, (C1) in which $k \geq n^{0.1}$, and (C2) where $k < n^{0.1}$. Case (C1) is simpler because the length of every phase $R = \Theta(\log k) = \Theta(\log n)$, which allows us to say that some "good" events happen with high probability, without needing to keep track of their exact probability and how it affects the whole system's evolution. In each of the cases (C1) and (C2), we further distinguish between the case where p_1 is large and close to 1 or small and away from 1 (to be formalized). For the case where p_1 is small, we will argue that in each long-phase there will be a time with a large enough number of undecided nodes to be noticed by the clock population, which thus keeps all or the vast majority of the clocks active. Because of this, we will be able to say that almost all the game-player nodes remain in the `GA` dynamic, which thus allows us to show that the system evolves with fast gap amplifications similar to what we saw in the analysis of the take 1 algorithm. For the remaining case where p_1 gets close to 1, we will not have the aformentioned abundance of active clocks. However, since at that point p_1 is already close to 1, we do not need the power/speed of the `GA` dynamic and we will argue that regardless of how game-player nodes are split between the two dynamics, even in the worst case, the system evolves almost as good as the `Undecided` dynamic and gets to plurality consensus fast.

Throughout the analysis, we will refer to each span of $4R$ rounds as a long-phase, which is made of 4 phases, each having R rounds. We also use $p_1, \ldots, p_k, q$ to denote the fraction of the game-player nodes that hold each of the k opinions and the fraction of undecided game-player nodes, respectively. Here, the fractions are computed with respect to the *game-player* nodes only. Throughout, we refer to the nodes that hold opinion 1 by *plurality nodes* and the remaining decided nodes as the *non-plurality decided nodes.*

C.1 Analysis for $k \geq n^{0.1}$

As mentioned above, some of the arguments are easier in this case because the length of each phase R is $\Theta(\log k) = \Theta(\log n)$, which allows us to say that some good events happen with high probability without needing to track their exact probability. For instance, $\Theta(\log n)$ rounds suffice for the information regarding the existence of undecided nodes to spread from one clock-node to all, w.h.p.

The analysis is divided into two parts, with respect to the value of the fraction p_1 of the plurality opinion. Recall that this fraction is with respect to the game-player nodes. When p_1 is sufficiently small (to be made precise later), we will show that all clock-nodes remain active w.h.p., and hence we can focus on analyzing the GA dynamics. In the complementary case, when p_1 gets close to 1, we will not be able to say that all the clock-nodes remain active, as they start switching to the end-game. Fortunately, during this period, we also do not need the full power of the GA dynamics. We will see that, despite the fact that some uncontrolled number of clock-nodes are in the end-game, within $O(1)$ phases all opinions other than the plurality disappear, and then within an additional O(1) additional phase, all nodes have the plurality opinion, w.h.p.

C.1.1 Case $k \geq n^{0.1}$ and $p_1 \leq O(1 - \log n/n)$

We begin by showing that in this regime, w.h.p., the entire clock population remains active:

Lemma C.1. *Suppose that $k \geq n^{0.1}$. Then, so long as $p_1 \leq 1 - \frac{20 \log n}{n}$, with high probability, all clock-nodes remain in the time-counting state.*

Proof. The proof in done in an inductive manner, over long-phases. The base case of the first long-phase is trivial and thus we only focus on the inductive step. That is, assuming that at the beginning of a long-phase all clocks are active, we show that w.h.p they all remain active even at the end of the phase. Notice that clock-nodes become inactive—switching to the end-game mode—only at the transitions from one long-phase to the next, i.e., when the round number is a multiple of $4R$.

Generally, since in each round each game-player meets a clock-node with probability $1/2$, we get that with high probability, each game-player node meets a clock-node within the first $5 \log n$ rounds of each phase. Because of this, for instance, by the end of the phase 0, all game-player nodes have activated their gap amplification, by setting $sampled = false$ and $forget = false$. Similarly we know that w.h.p., each game player node v will have one round in phase 1 in which v samples another game-player node u. Note that (once transitioned to phase 2), node v keeps its opinion only if the sampled node u has the same opinion as v. Hence, the number of the undecided nodes after $5 \log n$ rounds of phase 2 is, w.h.p., lower bounded by the number of nodes that hold the plurality opinion and did not sample a plurality opinion node in their sampling round of the gap amplification phase. This is in expectation $\frac{n}{2} p_1(1-p_1) \geq 9 \log n$ undecided nodes, which means with high probability we have at least $6 \log n$ undecided nodes. Thus, in the round $5 \log n + 1$ of phase 2, each clock-node meets an undecided node with probability at least $\frac{6 \log n}{n}$, which means we expect $3 \log n$ clocks to have notice the existence of undecideds, and thus with high probability, there are at least $\log n$ clock-nodes that do so and set their $v.consensus = false$. Now, in each of the next rounds, this $v.consensus = false$ spreads among the clock-nodes in a pattern similar to usual spreading of a gossip, and thus reaches all clocks with high probability in $O(\log n)$ rounds. We just give a brief sketch for this spreading: in each round, the fraction of the clock-nodes that have $v.consensus = false$ increases by a constant factor, w.h.p, until it reaches about $1/2$. From there on, the number of clock-nodes that have $v.consensus = true$ drops exponentially, and thus falls below 1 within $O(\log n)$ rounds. This is because, the probability that a clock-node has met one of these false-consensus clocks after $O(\log n)$ rounds is at most $1/n^{10}$. Hence, at the end of phase 3, w.h.p., all clock-nodes have $v.consensus = false$, which means they will all remain active for (the beginning of) the next long-phase. □

Equipped with the maintenance of the clock-nodes, the analysis now carries a similar line to that of of the Take 1 algorithm. Here we just mention a few words about the possible differences:

Lemma C.2. *Suppose that $k \geq n^{0.1}$. Then, so long as $p_1 \leq 1 - \frac{20 \log n}{n}$, with high probability, the fraction of nodes of different opinions evolves as analyzed in the previous section, and thus particularly, within $O(\log k \log n)$ rounds, we have $p_1 \geq 1 - \frac{20 \log n}{n}$.*

Proof. (Sketch) First note that since w.h.p. all clock are active, w.h.p. every node meets at least one clock-node within the first $5 \log n$ rounds of each phase. Thus, each game-player node receives the correct phase number in each phase, in fact within the first $1/10^{th}$ rounds of the phase, w.h.p. This implies that the system essentially simulates the setting where each node was also keeping the time itself (as in the take 1 algorithm). The only change required to be taken into consideration is that in gap-amplification and healing phases, each game player v has a probability of $1/2$ to meet a clock-node and if that happens, the opinion of v would not change in this round. However, it is easy to see that this 2-factor change in the probabilities, which is independent of the node's opinion, only slows down the process by a constant factor and otherwise has no effect on the proof. □

C.1.2 Case $k \geq n^{0.1}$ and $p_1 \geq 1 - O(\log n/n)$

We now turn to consider the case where p_1 has become very close to 1. Here, we can no longer assume that there are many active clock-nodes. For the analysis in this case, we will always consider both of the dynamics and each time present bounds on the behavior of the overall system based on the worst case behavior possible. In Lemma C.3, we show that this regime of $p_1 \approx 1$ is in fact an absorbing state (in a certain sense, to be formalized) meaning that, w.h.p., the system will not exit this regime. Then Lemma C.4 uses this fact to show that within $O(1)$ long-phases, all opinions other than the plurality get filtered out, and Lemma C.5 shows that then within $O(1)$ additional long-phase, all nodes hold the plurality opinion.

Lemma C.3. *Suppose that at the start of a long-phase, we have $\sum_{i \geq 2} p_i \leq \alpha \log n/n$ and $q \leq \beta \log^2 n/n$, for sufficiently large constants α and β (in comparison to the constant in R), and where $\beta \gg \alpha$. Then, w.h.p., in each of the rounds of this long-phase, we have $\sum_{i \geq 2} p_i \leq 2\alpha \log n/n$ and $q \leq 2\beta \log^2 n/n$. Moreover, w.h.p., we have $\sum_{i \geq 2} p_i \leq \alpha \log n/n$ and $q \leq \beta \log^2 n/n$ at the end of this long-phase as well.*

Proof. The proof is provided in two parts. We first use a somewhat coarse argument to show that w.h.p. in all rounds, we will have $\sum_{i \geq 2} p_i \leq 5\alpha \log n/n$ and $q \leq 10\beta \log^2 n/n$. We then provide a slightly more fine-grained argument, which also uses the guarantee established by the coarse argument, to show that at the end of the long-phase $\sum_{i \geq 2} p_i \leq \alpha \log n/n$ and $q \leq \beta \log^2 n/n$.

Consider one long-phase, and each of the rounds during the first two phases (according to the global time, which might be unknown to some nodes). The increase in $\sum_{i \geq 2} p_i$ in this one round comes from undecided nodes that met a node with opinion in $\{p_2, \dots, p_k\}$. The expected number of such nodes in one round is $nq \sum_{i \geq 2} p_i$ where q and p_i are the values at the start of that particular round. Thus, so long as we are still in the regime that $\sum_{i \geq 2} p_i \leq 5\alpha \log n/n$ and $q \leq 10\beta \log^2 n/n$, this expectation would be at most $O(\log^3 /n)$. That means, with high probability, the number of nodes that join $\{p_2, \dots, p_k\}$ in this one round is at most $O(1)$. On the other hand, the increase in the number of undecided nodes during these two phases comes from nodes that were acting based on the end-game and met a node of a different opinion. Per round, the expected number of such nodes is at most $2n\sum_{i\geq 2} p_i$, where p_i of that round is used. This is at most $2\alpha \log n/n$, so long as we are in the aforementioned regime of $\sum_{i\geq 2} p_i \leq 5\alpha \log n/n$. Now by repeated application of this argument for the $2R$ rounds of the first two phases, we get that throughout these two phases, we always remain in the regime of $\sum_{i\geq 2} p_i \leq \alpha \log n/n + O(R) \leq 2\alpha \log n/n$ and $q \leq \beta \log^2 n/n + O(R)\alpha \log n/n \leq 2\beta \log^2 n/n$.

While the number of new nodes that join the non-plurality opinions during phase 2 and 3 behaves (at most) similar to above, the increase in the undecided nodes in these rounds can come also from nodes that were acting based on the gap amplification dynamic and met a node not holding the same opinion in their sampling round in phase 1. The number of the nodes in the latter group is at most $n(1 - p_1)$ where, in the worst case, p_1 is the minimum value it can take during phase 2. Hence, the total increase in q because of these undecided nodes is expected to be at most $2\beta \log^2 n/n + 2\alpha \log n/n$ and is thus at most $3\beta \log^2 n/n$ with high probability. Modulo this one-time addition of $3\beta \log^2 n/n$ to q, again per each round of phases 2 and 3 the growth in $\sum_{i\geq 2} p_i$ is at most $O(1)$ per round and the increase in q is at most $2n\sum_{i\geq 2} p_i$ (so long as we are still in the regime that $\sum_{i\geq 2} p_i \leq 5\alpha \log n/n$ and $q \leq 10\beta \log^2 n/n$). We can thus apply this argument for $2R$ rounds repeatedly and get that we always remain in this regime of $\sum_{i\geq 2} p_i \leq 5\alpha \log n/n$ and $q \leq 10\beta \log^2 n/n$.

Now that we have this coarse analysis that shows the dynamic to always remain in the aforementioned regime, we re-examine the dynamic in a more fine-grained way, with the goal of establishing that by the end of the long-phase, with high probability, we actually have $\sum_{i\geq 2} p_i \leq \alpha \log n/n$ and $q \leq \beta \log^2 n/n$, i.e., the conditions at the start of the long-phase.

For a node v to hold a non-plurality opinion at the end of this long-phase, one of two things has to happen: either v had an opinion in $\{p_2, \ldots, p_k\}$, acted according to the gap amplification dynamic, and sampled some node of the same opinion, or v became undecided at some point and then switched to this non-plurality opinion. Note that potentially there could be a third case where the node followed the `Undecided` dynamic and kept it's non-plurality opinion but with high probability there is no such node as with high probability, each node v has two consecutive rounds in which it meets a plurality opinion node. Now per round, the expected number in the former case (following `GA` and keeping the opinion) is at most $n\sum_i p_i^2 \leq n(\sum_i p_i)^2$ and the expected number of the latter case (switching from undecided to a non-plurality opinion) is at most $nq\sum_{i\geq 2} p_i$. Taking the largest possible values from the regime that we established above, this is an expectation of at most $O(\frac{\log^3 n}{n})$ per round. Hence, the actual number is at most $O(1)$, with high probability. Even summing up over all the $4R$ rounds, the total would be at most $O(R) \leq \alpha \log n/n$, thus establishing that at the end of the long-phase, we have that $\sum_{i\geq 2} p_i \leq \alpha \log n/n$.

On the other hand, for a node v to be undecided at the end of the long-phase v, one of two things has to happen: either v was undecided at the end of phase 2 and it never met a decided node during phase 3, or v at some point became decided but later on switched to undecided because of meeting a node of an opposite opinion. W.h.p. there is no node in the first group because per round each node meets a decided node with probability at least p_1 and there are $O(\log n)$ rounds. On the other hand, the expected number of nodes in the second group is at most $n(\sum_{i\geq 2} p_i)$ per round, and thus $2n(\sum_{i\geq 2} p_i) + 5\log n$, w.h.p. Considering the established regime that $\sum_{i\geq 2} p_i \leq \alpha \log n/n$, we get that over the R rounds, this is at most $8R\alpha \log n/n + 20R \log n/n \leq \beta \log^2 n/n$. That is, at the end of the long-phase, we also have $q \leq \beta \log^2 n/n$. □

Lemma C.4. *Starting from $p_1 \geq 1 - 20\log n/n$, w.h.p., within $O(1)$ phases, we have $\sum_{i\geq 2} p_i = 0$.*

Proof. Consider the first time at the start of a long-phase such that $p_1 \geq 1 - 20\log n/n$. First note that by Lemma C.3, w.h.p., during all rounds of the next $O(1)$ phases, we have $p_1 \geq 1 - O(\log^2 n)/n$. Now a game-player node v can have a non-plurality opinion at the end of one long-phase only if, either it acted according to the gap amplification rule and it sampled a node of its own opinion, or it was undecided at some point and met a node of non-plurality opinion. Thus, at the end of the long-phase $\mathbb{E}[\sum_{i\geq 2} p_i^{new}] \leq \sum_{i\geq 2} p_i^{new} \cdot \frac{O(\log^2 n)}{n}$. Hence, the expected number of such game-player nodes after $O(1)$ phases is less than $1/n^c$, for any desirably large constant c, which by Markov's inequality means with high probability $\sum_{i\geq 2} p_i = 0$. □

Lemma C.5. *Once $\sum_{i\geq 2} p_i = 0$, within $O(1)$ long-phase, the plurality opinion obtains totality.*

Proof. First, consider just one long-phase, after reaching $\sum_{i\geq 2} p_i = 0$. Note that since $p_1 \geq 1 - O(\log^2 n/n)$ at the end of phase 2, throughout phase 3, with high probability, each game-player node will meet a node of the plurality opinion. Since there is no opposite opinion left, all game-player nodes remain with the plurality opinion at the end.

Now consider the long-phases after the one long-phase discussed above. Note that the clock-nodes will switch to end-game w.h.p at the end of the next long-phase, at the latest, as they would not detect an undecided game-player node then. Once they switch, they take the opinion of the last game-player node that they have met and that is the plurality opinion as we know. Hence, at that time, all nodes hold the plurality opinion. □

C.2 Analysis for $k < n^{0.1}$

We now turn to consider the case where k is small. Since the length of each phase is $\Theta(\log k)$, there are two difficulties that may arise: (a) even if all clock-nodes are active, a node might not meet any clock-node during the course of a phase, and (b) the information on the existence of undecided nodes may not spread to all the clock-nodes and hence some clock-nodes would move to the end-game prematurely. Again, we distinguish between two regimes of p_1. In the small regime, we would claim that a significant fraction $1 - O(1/k^5)$ of the clock-nodes are active and then we show this to be sufficient for the dynamics to have gap amplifications at nearly the same rate as in the Take 1 algorithm. We then consider the high regime of p_1 and use arguments similar in nature to those of the $k \geq n^{0.1}$ to show that p_1 reaches 1 soon, even considering the worst that can happen per round, in each of the gap amplification and undecided state dynamics.

C.2.1 Case $k < n^{0.1}$ and $p_1 \leq 1 - 1/k^5$

Lemma C.6. *W.h.p, in every long-phase, at least $1 - 1/(100k^8)$ fraction of the clocks are active.*

Proof. We begin by showing that in phase 2 there are $\Omega(n/k^5)$ undecided nodes. This is because, for each node v not having the plurality opinion at the start of the long-phase, v will see at least one active clock-node in each of phases 0 and 1, with probability at least $1 - (1/2)^R \geq 1 - 1/k^{10}$, and then it will sample one node, which has the same opinion as v with probability at most $O(1/k^5)$, due to the assumption $p_1 \leq 1 - 1/k^5$. Therefore, node v will forget its opinion during the first $O(1)$ rounds of the phase 2 with probability at least $\Omega(1)$. This means, overall the at least n/k^5 nodes that do not hold the plurality opinion at the start of the long-phase, we expect at least $\Omega(n/k^5)$ of them to be undecided within $O(1)$ rounds after the start of phase 2. Since $\frac{n}{k^5} \gg \log n$, this means w.h.p., there are at least $\Omega(n/k^5)$ such undecided nodes. Now, at that time, each clock-node meets one of these undecided nodes with probability at least $\Omega(1/k^5)$. Thus, with high probability, within one round, at least $\Omega(n/k^5)$ clock-nodes are aware of the existence of the undecided nodes. Now, this information spreads following the usual doubling-spreading of a gossip, among the clock-nodes: particularly, per round, the fraction of clock-nodes having this information grows by a constant factor, until it reaches $1/2$ which happens within $O(\log k)$ rounds. From there on, the probability that a clock-node does not get informed about the existence of undecideds after $40(\log k + 2)$ additional rounds is at most

$(3/4)^{40(\log k+2)} \leq 1/(100k^{10})$. Note that this is even true about clock-nodes that have already switched to the end-game in the past. Now, since $k < n^{0.1}$ Chernoff bound again shows that the actual number of the clock-nodes that do not receive this information is no more than $n/(100k^8)$. That is, at the end of the long-phase, at most $1/(100k^8)$ fraction of clock-nodes might switch to the end game. □

Lifting the gap amplification analysis: We use the large mass of the active clock-nodes established in Lemma C.6 to argue that the dynamic evolves per long-phase almost the same was as that of the take 1 algorithm. That is, roughly speaking in each long-phase the ratio of the plurality opinion to the largest other opinion gets squared. Thus, within $O(\log n \log k)$ phases, we reach $p_1 > 1 - 1/k^5$. At that point, the analysis would follow a different track (as we describe later in Appendix C.2.2) and will not rely on the existence/abundance of active clock-nodes, since the number of them might go down rapidly as p_1 approaches 1.

For simplicity, let us first talk about the gap amplification in the less sensitive cases where the gap we already have $\frac{p_1}{p_2} \geq 1 + \Omega(1/k^2)$. We examine the complementary case later. Consider one long-phase and call a node bad if it meets an end-game clock-node in any round of this long-phase or if it does not see any clock-node during one of the 4 phases of this long-phase. Note that for each node, the probability to be bad is at most $\frac{4R}{100k^5} + 4(1/2)^R \leq O(1/k^8)$. Hence, except for $O(\frac{1}{k^8})$ fraction of the p_1 nodes, the rest follow the algorithm by taking steps in each of the phases of the algorithm. This means, at the end of phase 2, we would have at least $(1 - O(1/k^8))np_1^2$ nodes in opinion 1, and this even holds w.h.p. as $np_1^2 \geq n^{0.8}$ (because $p_1 \geq 1/k$ and $k \leq n^{0.1}$). On the other hand, a $1 - O(1/k^8)$ fraction of the nodes of opinion 2 take their steps in each of the phases of the algorithm, and thus the number of nodes of opinion 2 that keep their opinion by the end of phase 2 is at most $(1-O(1/k^8))np_2^2+O(1/k)np_2$. There might be some further bad nodes that join opinion 2 because they are performing the end-game, but note that a node can join opinion 2 only if either it was previously in that opinion and it stayed there or it was undecided and it met a node of opinion 2. This means the number of bad node that join opinion 2 by the end of phase 2 is at most $np_2 O(1/k^8) + O(\log n)$. Thus, even with these very coarse bounds, the total number of nodes in opinion 2 at the end of phase 2 is at most $np_2^2 + O(1/k^8)np_2 + O(\log n)$. Hence, the ratio between the two opinions at this time is at least $\frac{(1-O(1/k^8))np_1^2}{np_2^2+O(1/k^8)np_2+O(\log n)} \geq (p_1/p_2) \cdot \min\{(p_1/p_2)^{0.9}, k^2\}$. Here, the inequality holds because $p_1 \geq 1/k$, $k \leq n^{0.1}$, and $\frac{p_1}{p_2} \geq 1 + \Omega(1/k^2)$. Note that this is essentially as good of a gap amplification as we need. Now that the gap ratio has been amplified by the end of phase 2, we get to the healing phase. The healing again works (essentially) as in take 1, except for a small difference because of bad nodes: per round, the number of nodes of opinion 1 might grow by a $1 - O(1/k^8)$ factor less than the growth in that of the nodes of opinion 2. However, this is a loss factor of $1 - O(\frac{\log k}{k^8})$ over all the $O(\log k)$ rounds of the healing phase and thus, the ratio at the end of the long-phase remains at least $(p_1/p_2) \min\{(p_1/p_2)^{0.8}, k\}$. That is, overall the long-phase provides a gap amplification as we want, except that now the growth of the gap is capped to k. Note that limiting the growth of the gap by a k factor can increase the number of the required phases by at most $O(\log n/\log k)$ phases, i.e., $O(\log n)$ rounds, which gets absorbed in the asymptotic notations of our complexity.

Now we discuss the case when $p_1/p_2 \in (1, 1+O(1/k^2))$, which requires a bit more care in analyzing how the bad nodes can affect the gap amplification. We first analyze the ratio of the nodes of the plurality opinion to that of the second largest opinion, at the end of phase 1, i.e., right before the time that the gap amplification results are activated (by forgetting the opinions).

Consider the first round and let γ be the probability of a game-player node acting based on the end-game (because of meeting an end-game clock-node). Notice that γ is the same for all nodes (particularly it's the same regardless of whether the node's opinion), and that $\gamma \leq O(\log k/k^8)$, if fact for any of the rounds of the long-phase. Now note that during this round, no node changes its opinion, except for those bad nodes that are acting based on the end-game. We can easily see that, even in the end-game, at least in expectation, the ratio of the number of nodes of the plurality opinion to that of the second largest opinion does not decrease (in one round). This is because, in the end-game, the undecided nodes join each of the opinions with probabilities proportional to the number of the nodes of those opinions, and on the other hand, the probability of a decided node becoming undecided is smaller for plurality opinion nodes, compared to the non-plurality opinion nodes. However, we still might have deviations from the expectations, which would end up shrinking the ratio between the two largest opinions. This additive deviation is at most $\pm O(\sqrt{n\gamma \log n})$ for each of the opinions, as even the whole number of nodes that act based on the end game is at most $n\gamma$. Hence, at the end of the first two phases, the ratio between the two opinions is at least

$$\frac{np_1 - O(\sqrt{n\gamma \log n})}{np_2 + O(\sqrt{n\gamma \log n})} \geq \frac{np_1(1 - O(\frac{\sqrt{\log n/n}}{k^2}))}{np_2(1 + O(\frac{\sqrt{\log n/n}}{k^2}))} \geq (\frac{p_1}{p_2})(1 - O(\frac{\sqrt{\log n/n}}{k^2})),$$

where the inequality holds because $p_1 \geq 1/k$, $p_1/p_2 \in (1, 1 + O(1/k^2))$, and $\gamma \leq O(\log k/k^5)$. Noting that during the first two phases, the only nodes that change their opinion are those that act based on the end-game, we can repeat the above argument for the $O(\log k)$ rounds of the first two phases, and conclude that at the end of the first two phases, the ratio between the two largest opinions is at least $(\frac{p_1}{p_2})(1 - O(\frac{\sqrt{\log n/n}}{k^2}))^{O(\log k)} \geq (\frac{p_1}{p_2})(1 - O(\frac{\log k\sqrt{\log n/n}}{k^2}))$.

Since $p_1 - p_2 \geq \Omega(\sqrt{\log n/n})$, the above means during the first two phases, the ratio between the two largest opinions does not get attenuated significantly. We now show that during phase 2 and by the end of it, the ratio grows considerably, to at least $(p_1/p_2)^{1.6}$.

Now during phase 2, during the main algorithm, decided nodes that their sample in phase 1 had a different opinion (or was undecided) forget their opinion. However, there are two other effects, because of bad nodes: (1) we again have a small-fraction of nodes that are acting based on the end-game, (2) there is a group of bad nodes that although they are not following the end-game, they did not act based on the main algorithm properly, i.e., either they did not sample a game-player or they do not get to forget their opinion, even though they had sampled a node of a different opinion. We will be able to bound the effect of the first group (end-game players), round by round, as we did in the above paragraph. For the latter group, note that for a node v fall in this group, it should be true that, in at least one of the first three phases, node v does not meet (either) any clock-node (or any game-player node). The probability of this is at most $O(1) \cdot (1/2)^R \leq 1/k^{10}$.

Note that if we did not have the bad nodes, the (expected) ratio between the two largest opinions would be $(p_1/p_2)^2$. However, the actual ratio has some (small) influence from the bad nodes. Among the bad nodes, the ratio is at least as large as the ratio that we had at the start of phase 2. This is because, similar to the discussions above, the end-game nodes do not decrease the ratio (in expectation), and for the bad nodes of the second group, which do not get to forget their opinion, the ratio is at most as if they did change their opinion. Furthermore, note that the probability of a node to become a bad node of group 2 is an $\alpha = O(1/k^8)$ but this event is independent of the node's opinion. Thus, we can conclude that, even considering the possible deviations, the ratios are at least

$$\frac{p_1^2(1-\alpha) + \alpha p_1 \pm 10\sqrt{np_1^2 \log n} \pm O(\sqrt{n \log n}/k^2)}{p_2^2(1-\alpha) + \alpha p_2 \pm 10\sqrt{np_2^2 \log n} \pm O(\sqrt{n \log n}/k^2)} \geq \min\{(p_1/p_2)^{1.6}, k(p_1/p_2)\}.$$

The inequality holds because $p_1 - p_2 \geq \Omega(\sqrt{\log n/n})$, $p_1 \geq 1/k$ and $\alpha \leq O(1/k^8)$. Hence, at the end of phase 2, we have the desired gap amplification in the ratios[8].

[8]Note that again the amplification is capped to k factor, but as we saw before, this can increase the time complexity by only an additive $O(\log n/\log k)$ phases, which is just $O(\log n)$ rounds

Now during phase 3, the healing gets performed, which effectively shrinks the fraction of undecided nodes to at most $O(1/k)$. While doing this, the ratio of the multiplicative gap between the two largest opinions remains essentially preserved, as analyzed in the take 1 algorithm. The only difference here with regards to the analysis of the take 1 algorithm is that, here the healing process is initially about a 1/2 factor slower as at the start of the phase a node gets informed of this phase only with probability 1/2. However, this only affects when an undecided node becomes ready to adopt an opinion, and does not affect the ratios between the two opinions, as still, each undecided node adopts the opinion of the first decided node that it meets. Again, there is a small effect because of nodes that play according to the end-game, but as analyzed above, this can shrink the ratios by at most a $(1 - O(\frac{\log k\sqrt{\log n/n}}{k^2}))$ factor, which given that $p_1 - p_2 \geq \Omega(\sqrt{\log n/n}$, $p_1 \geq 1/k$ and $p_2 \in (1, 1 + O(1/k^2))$, means the ratio at the end of the long phase is at least $\min\{(p_1/p_2)^{1.5}, (k/2)(p_1/p_2)\}$. This shows the desired gap amplification that happens during one phase, and gives that within $O(\log n \log k)$ phases, p_1 reaches $p_1 \geq 1 - 1/k^5$.

C.2.2 Case $k < n^{0.1}$ and $p_1 > 1 - 1/k^5$

Here we show that once p_1 gets very close to 1, particularly passing $1-1/k^5$, within $O(\log n/\log k)$ more phases, we reach plurality consensus. In this part of the analysis, we will not rely on the existence (or abundance) of active clock-nodes, as the number of the active clock-nodes might be (or become) very small. The style of the arguments here is to a large extent similar to Appendix C.1.2. In Lemma C.7, we show that w.h.p. the system will not leave this regime of $p_1 \approx 1$. Then Lemma C.8 uses this fact to show that within $O(\log n/\log k)$ long-phases, all opinions other than the plurality get filtered out, and Lemma C.9 shows that then within $O(\log n/\log k)$ additional long-phase, all nodes hold the plurality opinion.

Lemma C.7. *Suppose that at the start of a long-phase, we have $\sum_{i\geq 2} p_i \leq O(1/k^5)$ and $q \leq O(1/k^3)$. Then, w.h.p., in each of the rounds of this long-phase, we have $\sum_{i\geq 2} p_i \leq O(1/k^4)$ and $q \leq O(1/k^2)$ and at the end of this long-phase we have $\sum_{i\geq 2} p_i \leq O(1/k^5)$ and $q \leq O(1/k^3)$.*

Proof. The proof is provided in two parts. We first use a somewhat coarse argument to show that w.h.p. in all rounds, we will have $\sum_{i\geq 2} p_i \leq O(1/k^4)$ and $q \leq O(1/k^2)$. We then provide a slightly more fine-grained argument, which also uses the guarantee established by the coarse argument, to show that at the end of the long-phase $\sum_{i\geq 2} p_i \leq o(1/k^5)$ and $q \leq o(1/k^3)$.

Consider the global time, and each of the rounds during the first two phases. The increase in $\sum_{i\geq 2} p_i$ in this round comes from undecided nodes that met a node with opinion in $\{p_2, \ldots, p_k\}$. The expected number of such nodes in one round is $nq\sum_{i\geq 2} p_i$ where q and p_i are the values at the start of that particular round. Thus, so long as we are still in the regime that $\sum_{i\geq 2} p_i \leq O(1/k^4)$ and $q \leq O(1/k^2)$, this expectation would be at most $O(n/k^6)$. Since $n \geq k^{0.1}$, that means with high probability the number of nodes that join $\{p_2, \ldots, p_k\}$ in this one round is at most $O(n/k^6)$. On the other hand, the increase in the number of undecided nodes during these two phases comes from nodes that were acting based on the end-game and met a decided node of a different opinion. Per round, the expected number of such nodes is at most $2n\sum_{i\geq 2} p_i$, where p_i of that round is used. This expectation is at most $O(n/k^4)$, so long as we are in the aforementioned regime, and thus with high probability the actual number is also at most $O(n/k^4)$. By repeated application of this argument for the $2R = O(\log k)$ rounds of the first two phases, we get that throughout these two phases, always $\sum_{i\geq 2} p_i \leq O(1/k^5) + O(\log k)/k^6 \leq O(1/k^4)$ and $q \leq O(1/k^3) + O(\log k)/k^4 \leq O(1/k^2)$.

The above neglects a minor detail, i.e., there might be nodes that their last phase update remained stuck at 2 or 3, and they might join $\{p_2, \ldots, p_k\}$ or they might become undecided per round. However, that would mean this node did not meet any clock-node during the phase 0 or 1 and since the total number of clock-nodes (active or in end-game) is about 1/2, the total number of such 'bad' nodes is expected to be at most $O(1/k^8)$ and adding this to the above numbers, even per round, does not change the asymptotic notations.

While the number of new nodes that join the non-plurality opinions during phase 2 and 3 behaves (at most) similar to above, the increase in the undecided nodes in these rounds can come also from nodes that were acting based on the gap amplification dynamic and met a node not holding the same opinion in their sampling round in phase 1. The number of the nodes in the latter group is at most $n(1-p_1)$ where, in the worst case, p_1 is the minimum value it can take during phase 2. Hence, the total increase in q because of these undecided nodes is expected to be at most $O(n/k^2)$ and is thus at most $O(n/k^2)$ with high probability. Modulo this one-time addition of $O(1/k^2)$ to q, again per each round of phases 2 and 3 the growth in $\sum_{i\geq 2} p_i$ is at most $O(1/k^6)$ per round and the increase in q is at most $O(1/k^6)$ (so long as we are still in the regime that $\sum_{i\geq 2} p_i \leq O(1/k^4)$ and $q \leq O(1/k^2)$). We can thus apply this argument for $2R$ rounds repeatedly and get that we always remain in this regime of $\sum_{i\geq 2} p_i \leq O(1/k^4)$ and $q \leq O(1/k^2)$. This finishes the coarse grained argument, for remaining within this regime in all rounds of one long-phase.

Now that we have this coarse analysis that shows the dynamic to always remain in this regime, we re-examine the dynamic in a more fine-grained way, with the goal of establishing that by the end of the long-phase, with high probability, we actually have $\sum_{i\geq 2} p_i \leq o(1/k^5)$ and $q \leq o(1/k^3)$.

For a node v to hold a non-plurality opinion at the end of this long-phase, one of three things has to happen: (1) either v had an opinion in $\{p_2, \ldots, p_k\}$, acted according to the gap amplification dynamic, and sampled some node of the same opinion, or (2) v became undecided at some point and then switched to this non-plurality opinion, (3) there was a phase in which v did not see any clock-node. Per round, the expected number in the first case is at most $n\sum_i p_i^2 \leq n(\sum_i p_i)^2$, the expected number of the second case is at most $nq\sum_{i\geq 2} p_i$, and the expected number of the third case is at most n/k^{10}. Taking the largest possible values from the regime that we established above, this is an expectation of at most $O(n/k^6)$ per round in total. Hence, the actual number is at most $O(n/k^6)$, w.h.p., as $k \leq n^{0.1}$. Summing up over all the $4R$ rounds, we get that at the the end of the long-phase, we have $\sum_{i\geq 2} p_i \leq O(\log k/k^6) \leq o(1/k^5)$.

On the other hand, for a node v to be undecided at the end of the long-phase v, one of three things has to happen: (1) either v was undecided at the end of phase 2 and it never met a decided node during phase 3, or (2) v at some point became decided but later on switched to undecided because of meeting a node of an opposite opinion, (3) there was a phase in which v did not meet any clock node (active or in the end-game). The number of the nodes in the first and third groups is expected to be at most $O(n/k^8)$, and thus also with high probability at most $O(n/k^8)$. The reason for the first group is that per round each node meets a decided node with probability at least p_1 and there are $R = \Theta(\log k)$ rounds in a phase. The reason for the third group is similar, per round each node meets a clock with probability 1/2 and there are $R = \Theta(\log k)$ rounds in a phase. On the other hand, the expected number of nodes in the second group is at most $n(\sum_{i\geq 2} p_i)$ per round. Considering the established fact that we always have $\sum_{i\geq 2} p_i = O(1/k^4)$, this expectation is at most $O(n/k^4)$ and thus the actual random number is also w.h.p. $O(n/k^4)$. We thus get that over the $O(R)$ rounds, this is at most $O(n\log k/k^4) = o(n/k^3)$. That is, w.h.p., at the end of the long-phase, $q = o(1/k^3)$. □

Lemma C.8. *After $p_1 \geq 1 - 1/k^5$, in $O(\log n/\log k)$ phases, w.h.p., we have $\sum_{i\geq 2} p_i = 0$.*

Proof. Consider the first time at the start of a long-phase such that $p_1 \geq 1 - 1/k^5$. Note that by Lemma C.3, with high probability, during all rounds of the next $O(\log n/\log k)$ phases, we have $p_1 \geq 1 - O(1)/k^2$. Now a node v can have a non-plurality opinion at the end of one long-phase only if, either it acted according to the gap amplification rule and it sampled a node of its own opinion, or it was undecided at some point and met a node of non-plurality opinion, or it had this opinion at the start of the long-phase but never changed its opinion. Thus, at the end of the long-phase $\mathbb{E}[\sum_{i\geq 2} p_i^{new}] \leq \sum_{i\geq 2} p_i^{new} \cdot \frac{O(1)}{k^2}$. Hence, the expected number of such nodes after $O(\log n/\log k)$ phases is less than $1/n^c$, for any desirable constant c, which by Markov's inequality means w.h.p. there is no such node. That is, after $O(\log n/\log k)$ phases, w.h.p., $\sum_{i\geq 2} p_i = 0$. □

Lemma C.9. *After $O(\log n/\log k)$ phases from the time that $p_1 \geq 1 - 1/k^5$ and $\sum_{i\geq 2} p_i = 0$, with high probability, we have $p_1 = 1$.*

Proof. First, consider each long-phase, after reaching $\sum_{i\geq 2} p_i = 0$. Note that since $p_1 \gg 2/3$ at the end of phase 2, the probability that a node remains undecided throughout phase 3 is at most $1/k^{10}$. Now, per long-phase, the number of undecided might grow at most by a 3 factor (during the phase 2 of the gap-amplification), but it will decrease then by a $O(1/k^{10})$ factor in the consequent phase 3, regardless of what fraction of nodes act based on the undecided state dynamic and what fraction follow the gap amplification. From this, we get that after $O(\log n/\log k)$, we expect to see at most $1/n^c$ undecided nodes for a desirably large c, which by Markov's inequality means with high probability there is no undecided node.

Furthermore, note that the clock-nodes will switch to end-game w.h.p at the end of the next long-phase at the latest, as they would not detect an undecided node then. Regardless of when each of them switches, once a clock-node switches, it takes the opinion of the (last) game-player node that it has met. Clearly, at this time, that is the plurality opinion. Per long-phase each clock-node sees a game-player node with probability at least $1 - 1/k^{10}$ which means with high probability, within additional $O(\log n/\log k)$ long-phases, all nodes hold the plurality opinion. □

D A Fast Plurality Algorithm for the Non-Random Gossip Model

We describe here an $O(\log n)$-time algorithm with memory/message size $O(\log n)$, assuming $k \leq O(\sqrt{n})$ and $p_1 - p_2 \geq \Omega(\sqrt{\log n/n})$. However, this algorithm assumes a non-random gossip model: that is, the contacts are not necessarily random in each round and a node v can contact one other node u that either v contacted u before or u contacted v before, conditioned on that v remembers u's identifier (considering its memory limitation). We remark that this is not the standard gossip model. Though, a similar model was assumed in [BCN+15a, Section 4] as well.

Plurality Estimation via Sampler Trees:

1. Elect a leader, in $O(\log n)$ rounds as follows: let every node pick a $4\log n$-bit random ID. For $O(\log n)$ rounds, each time each node v updates the maximum observed ID to be the maximum of its current value and that of the node u that v contacted.

2. Pick k sampling roots as follows: starting from the leader, every time each sampling root holds a sub-interval $I \subset [1, k]$ and passes along about half of it to the first node that it meets that is not a sampling root node, until we have $|I| = 1$. Particularly, when the sampling root meets the first other node, it passes along $[\lfloor k/2 \rfloor + 1, k]$ to that node and keeps $[1, \lfloor k/2 \rfloor]$ for itself. Next time that each of these sampling roots meets some new node, it passes along the higher half of the responsibility interval and keeps the lower half for itself. If the sampling roots meets someone which already has some responsibilities, no passing along happens. Within $O(\log n)$ rounds, w.h.p. the responsibilities are spread so that there are exactly k counters. one for each number $i \in \{1, 2, \ldots, k\}$.

3. Each sampling root node i, for every $i \in \{1, \ldots, k\}$, is responsible for opinion i whose statistics is to be collected using a *sampler tree* rooted at this node: grow from each counter i a sampler tree T_i of size $S = \frac{n}{2k}$ essentially by repeating the same idea as above but where a subinterval $[1, S]$ is

being split, for each sampling root. Note that different intervals are spread for different sampling roots simultaneously. Again, if a subinterval-holding node contacts one that is also holding a subinterval, nothing happens. Within $O(\log n)$ rounds, w.h.p., we this process terminates and each node holds at most one sub-interval. Let each node v remember the first node u that contacted v passing a responsibility to it, as its *parent* in the corresponding *sampler tree.*

4. In the next communication round, make each node v in tree T_i meet a random other node u and remember whether u holds opinion i or not.

5. Rewind the tape of the meetings in step 3 (i.e., the process of growing sampler trees), running it in reverse-time order where each node contacts its parent in the appropriate round. In each tree T_i, perform a converge-cast of the number of sampled nodes n_i. By the end of $O(\log n)$ rounds, for each i, the sampling root of the tree T_i knows n_i.

6. Rebuild the sampler trees[9] by growing from each sampling root i the sampler trees $\widetilde{T}_i$ of size $S_i = n_i \cdot k/4$, by repeating the same idea as in step 3.

7. Repeat step 4 and 5. Let $\widetilde{n}_i$ be the number of nodes of opinion i sampled by the nodes of $\widetilde{T}_i$.

8. Each sampler root i computes the estimated fractions $\widetilde{p}_i = \widetilde{n}_i/S_i$.

9. Compute the i for which $\widetilde{p}_i$ is maximum, using a standard spreading of maximum, similar to step 1, in $O(\log n)$ rounds. This is the w.h.p. the plurality opinion.

Analysis:

Claim D.1. *Within $O(\log n)$ rounds, w.h.p., the following holds:*
(a) k sampling roots are computed in Step (2).
(b) k sampler trees are constructed in Step (4), and
(c) k sampler trees are constructed in Step (6).

Proof. Since the number of sampler roots is k, there are at most $k \ll n/2$ nodes that are holding a sub-interval. Therefore, each node that holds an interval has probability of at least $1/2$ to meet a free node in a given round and the interval is bisected. Therefore, within $O(\log n)$ rounds, all splits have been occurred and thus the k sampling roots have been elected.

Since in Claim (b) the total number of nodes that are supposed to participate in different sampler trees is $n/2$, the same reasoning holds.

Finally, consider Claim (c). We show that w.h.p. the total number of nodes $\sum_j S_j$ in different sampler trees $\widetilde{T}_j$ is at most $n/2$. In expectation, $\mathbb{E}(\sum_j S_j) = (k/4) \cdot \sum_j n_j = n/4$. Hence, w.h.p., $\sum_j S_j \le n/2$ as required and the remaining proof continues as for Claim (a). □

Lemma D.2. *Assuming that the initial bias is $p_1 - p_2 = \Omega(\sqrt{\log n/n})$, w.h.p., the root holds the plurality opinion.*

Proof. We show that w.h.p. $\widetilde{p}_1 > \widetilde{p}_j$ for every $j \in \{2, \ldots, k\}$, which proves the lemma.

Let us first consider the sensitive case where p_j is not too far from p_1, particularly where $Sp_j \ge 10 \log n$. In the first sampling phase, the expected number of nodes in tree T_i that meet a node of opinion i is $Sp_i = \frac{n}{2k} p_i$. For the plurality opinion, this is $np_1/2k \ge \Omega(\log n)$. Hence, $n_1 = \Theta(np_1/k)$, w.h.p. As $Sp_j \ge 10 \log n$, we can similarly say $n_j = \Theta(np_j/k)$.

[9] This rebuilding is done in order to refine the estimations for opinions with large support (by increasing the number of samples, which thus decreases the relative deviation).

In the second sampling phase, the number of sampled nodes of opinion i as observed by the sampling tree $\widetilde{T}_i$ is $\widetilde{n}_i \in [S_i \cdot p_i \pm \sqrt{5 S_i p_i \cdot \log n}]$. Hence, we can write

$$\widetilde{p}_1 = \frac{\widetilde{n}_1}{S_1} \le p_1 - \sqrt{\frac{5 p_1 \log n}{S_1}} \le p_1 - \Theta(\sqrt{\frac{\log n}{n}}) \quad \text{and} \quad \widetilde{p}_j = \frac{\widetilde{n}_j}{S_j} \ge p_j + \sqrt{\frac{5 p_j \log n}{S_j}} \ge p_j + \Theta(\sqrt{\frac{\log n}{n}}),$$

which given that $p_1 - p_2 \ge p_1 - p_j \ge \Omega(\sqrt{\log n / n})$, means $\widetilde{p}_1 > \widetilde{p}_j$.

Now we consider the complementary case where $S p_j \le 10 \log n$. Since $S p_j \le 10 \log n$, w.h.p., we have $n_j \le 15 \log n$. If $n_j = 0$, we already done. Suppose $n_j \ge 1$. Then, $\mathbb{E}[\widetilde{n}_j] = n_j p_j k/4 = n_j \frac{5k^2 \log n}{n} \ge 100 \log n$. Hence, w.h.p., $\widetilde{n}_j \le (1.05)\mathbb{E}[\widetilde{n}_j] \le 1.05 n_j p_j k/4$. That means, $\widetilde{p}_j \le 1.05 p_j$. Since $p_1 = \Omega(\sqrt{\log n / n})$, and $S p_j \le 10 \log n$, we get that $\widetilde{p}_j < \widetilde{p}_1$, thus completing the proof. $\square$