Randomized Strategies for the Plurality Problem

Daniel Kráľ[*] Jiří Sgall[†] Tomáš Tichý[†]

Abstract

We consider a game played by two players, Paul and Carol. At the beginning of the game, Carol fixes a coloring of n balls. At each turn, Paul chooses a pair of the balls and asks Carol whether the balls have the same color. Carol truthfully answers his question. Paul's goal is to determine the most frequent (plurality) color in the coloring by asking as few questions as possible. The game is studied in the probabilistic setting when Paul is allowed to choose his next question randomly.

We give asymptotically tight bounds both for the case of two colors and many colors. For the balls colored by two colors, we provide a strategy for Paul to determine the plurality color with the expected number of $2n/3 + O(\sqrt{n \log n})$ questions and a lower bound $2n/3 - O(\sqrt{n})$ on the expected number of Paul's questions. For the balls colored by k colors, we prove a lower bound $\Omega(kn)$ on the expected number of questions.

[*]Institute for Mathematics, Technical University Berlin, Strasse des 17. Juni 136, D-10623 Berlin, Germany. E-mail: kral@math.tu-berlin.de. The author is a postdoctoral fellow at TU Berlin within the framework of the European training network COMBSTRU. On leave from Department of Applied Mathematics and Institute for Theoretical Computer Science (project 1M0021620808 of MŠMT ČR), Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00 Prague, Czech Republic. E-mail: kral@kam.mff.cuni.cz.

[†]Mathematical Institute of the Academy of Sciences of the Czech Republic, Žitná 25, CZ-11567 Prague, Czech Republic. E-mail: sgall,tichy@math.cas.cz. Partially supported by Institute for Theoretical Computer Science, Prague (project 1M0021620808 of MŠMT ČR) and grant 201/05/0124 of GA ČR.

1 Introduction

We study a two-player game played by Paul and Carol. Paul wants to determine a certain property of the input based on Carol's answers. At the beginning of the game, Carol fixes a coloring of n balls by k colors. Paul does not know the coloring. At each step, he chooses two balls and asks Carol whether the balls have the same color. Carol truthfully answers the posed question. Paul wants to ask the least number of questions (in the worst case) to determine the desired property of the coloring.

In the probabilistic setting, Paul can choose his next question randomly, depending on the previous course of the game. As in the previous literature, we consider Las Vegas strategies, which means that at the end of the game Paul is not allowed to make any error. His goal is to minimize the expected number of questions asked for the worst case (most difficult) coloring.

1.1 Previous work

The basic problem studied before is the *Majority problem*: Paul wants to find a ball b such that the number of balls with the same color is greater than $n/2$, or to declare that there is no such ball. If the balls are colored just with two colors and no randomness is allowed, Paul needs to ask at least $n - \nu(n)$ questions for some coloring, and Paul has a strategy such that he asks at most $n - \nu(n)$ questions [10, 4] for any coloring, where $\nu(n)$ is the number of 1's in the binary representation of n. If the number of colors is unrestricted, then the necessary and sufficient number of questions to resolve the Majority problem is $\lceil 3n/2 \rceil - 2$, see [6]. Other variants of the Majority problem were considered in [1, 8].

Another problem is the *Partition problem* in which Paul wants to partition the balls according to their colors. Dvořák et al. [5] showed that in the deterministic case, the necessary and sufficient number of questions to resolve the Partition problem with n balls of k colors is $(k-1)n - \binom{k}{2}$. They also showed that in the probabilistic setting the necessary and sufficient expected number of questions to resolve the problem for n balls of three colors is $5n/3 - 8/3 + o(1)$.

Another variant, the *Plurality problem* was introduced by Aigner et al. [2]. Here Paul seeks a ball with the *plurality color*, i.e., the color such that the number of balls with this color exceeds the number of balls of any other color; if no such ball exists, Paul declares that there is a tie. Paul's goal is only to point to a single ball with this color, he does not want to

	Deterministic strategies		Probabilistic strategies	
Problem	Lower bound	Upper bound	Lower bound	Upper bound
Majority 2 colors	$n-\nu(n)$	$n-\nu(n)$	$2n/3-o(n)$ [*]	$2n/3+o(n)$ [*]
Plurality 2 colors 3 colors k colors	 $n-\nu(n)$ $3n/2-O(1)$ $kn/40$	 $n-\nu(n)$ $5n/3+O(1)$ $O(kn)$	 $2n/3-o(n)$ [*] $3n/2-o(n)$ $\Theta(kn)$ [*]	 $2n/3+o(n)$ [*] $3n/2+O(1)$ $kn/2+o(n)$
Partition 2 colors 3 colors k colors	 $n-1$ $2n-3$ $(k-1)n-\binom{k}{2}$	 $n-1$ $2n-3$ $(k-1)n-\binom{k}{2}$	 $n-1$ $5n/3-8/3$ $(k-1)(n-k)/4$ [*]	 $n-1$ $5n/3-8/3+o(1)$ $kn/2+o(n)$

Table 1: The summary of the previous and new results. The results obtained in this paper are marked [*].

find all such balls. Aigner et al. [2] give a deterministic strategy to solve the Plurality problem with n balls of three colors asking at most $\lfloor 5n/3\rfloor-2$ questions and showed that Carol can force Paul to ask at least $3\lfloor n/2\rfloor-2$ questions. In addition, if the number k of colors of the balls is not fixed to be three, Aigner et al. [3] provide lower and upper bounds of order $\Theta(kn)$ for deterministic strategies (the lower bound is $kn/40$, but they did not try to tune the multiplicative constant). In the probabilistic setting, Dvořák et al. [5] showed that the necessary and sufficient expected number of questions to resolve the problem with n balls of three colors is $3n/2+o(n)$.

1.2 Our results

We establish new lower and upper bounds for the Plurality problem in the probabilistic setting (see Table 1). For n balls of two colors, we show that the necessary and sufficient expected number of questions to resolve the Plurality problem is $\frac{2n}{3}+o(n)$. Note that for two colors, the Majority and Plurality problems coincide, and our bounds apply to the Majority problem, too.

For the general number k of colors, we show a lower bound $\Omega(kn)$ on the expected number of questions of Paul to resolve the Plurality problem with n balls of k colors; the constant at the linear term in Ω is approximately 2/27. This improves the lower bound $\Omega(kn)$ of Aigner et al. [3] from deterministic to probabilistic strategies. Since it seems that the argument of [3] does not easily translate to the probabilistic setting, we use a different method in our proof. Since the number of questions necessary to resolve the Partition problem must be larger than the number of questions for the Plurality problem, our bound also yields a lower bound $\Omega(kn)$ on the expected number of questions to resolve the Partition problem. However, for this problem, we provide a simpler and better lower bound of $(k-1)n/4$ on the number of questions.

2 Notation

We now introduce notation used in the paper. The base of all the logarithms is 2, i.e., $\log x$ always means $\log_2 x$. We use a way of representing Paul's information about the colors of the balls from [5]: The game of Paul and Carol can be viewed as a game on a graph whose vertices correspond to the balls. At the beginning, the graph has no edges. At each turn, Paul chooses a pair of nonadjacent vertices and adds that edge to the graph. Carol colors the edge by red if the balls corresponding to its end-vertices have the same color, or by blue if their colors are different. This edge-colored graph represents the state of Paul's knowledge and is referred to as *Paul's graph*. A coloring of the balls is *consistent* with Paul's graph if the colors of every pair of balls corresponding to the vertices joined by a red (blue) edge are the same (different). In the Partition problem with k colors, the game ends when there is a unique coloring (up to a permutation of the colors) of the balls consistent with Paul's graph. In the Plurality problem with k colors, the game ends when there is a vertex v such that in any coloring consistent with Paul's graph the ball corresponding to v has the plurality color.

3 Plurality Problem with Two Colors

3.1 Upper Bound

The natural idea for an upper bound is to compare random pairs of balls. If the colors are different, the pair can be discarded, as it does not influence the result. We then recurse on the pairs of balls of the same color. This strategy performs well if the numbers of balls of both the colors are about the same. However, on general instances the expected number of comparisons may be too large. To circumvent this problem, Paul first applies the previous strategy only for a half of the balls, obtaining not only the plurality color, but also a good estimate of the number of balls of each color. He then applies the same strategy to classify a fraction of the remaining balls so that with high probability the plurality color is determined with a small number of additional comparisons.

The algorithm below determines the number of balls of each of the two colors and finds a representative of the larger color class (in case that there is no tie).

Algorithm COUNT

(1) if there is no ball, return 0 and no representative;
(2) if there is a single ball, return 1 and set the representative r to this ball;
(3) randomly permute the n balls;
(4) for each $i \leq n/2$, compare the $(2i-1)$-th and $2i$-th balls (in order determined by the random permutation);
(5) let R be the set of balls containing one ball from each compared pair where Carol answered that the balls have the same color;
(6) apply Algorithm COUNT recursively to the set R;
let r be the representative and m' be the number of balls of the plurality color (in case of tie, no r is found);
(7) let $m = \lfloor n/2 \rfloor - |R| + 2m'$;
(8) if n is odd then
    if there is no representative r then
        set r to be the n-th ball and let $m := m + 1$;
    else
        compare the n-th ball and r,
        if they have the same color, then let $m := m + 1$;
(9) output m and r (if r was defined);

We summarize the properties of Algorithm COUNT in the next lemma:

Lemma 3.1. *Let n balls be colored so that there are a and b balls of each color and $a \geq b$. Then Algorithm* COUNT *correctly determines the number of balls with the plurality color and provides its representative (if there is no tie). The expected number of Paul's questions is at most $a + b/3 + \log n = n - 2b/3 + \log n$.*

Proof. The pairs where Carol answered that the balls have different colors contain equal number of balls of each color. Therefore, the plurality color for the original coloring is the plurality color among the balls of R, if n is even. If n is odd, we also have to adjust the number of balls with the plurality color in Step (8). Overall, Algorithm Count always correctly determines the number of balls with the plurality color and finds its representative.

Let $m_{a,b}$ be the expected number of comparisons on an instance with a and b balls of each color. We show by induction on n that

$$m_{a,b} \leq a + \frac{b}{3} + \log(a+b).$$

The bound trivially holds if n is one or two. Assume that $n = a + b \geq 3$. The probability that a ball of the second color is paired with a ball of the first color, over the random permutation of n balls, is at least $1/2$: for n even it is $a/(a+b-1) \geq 1/2$, and for n odd, counting the probability that the ball is not matched, the probability is $a/(a+b) \geq 1/2$. Thus the expected number of pairs of balls with different colors is at least $b/2$. Let a' and b' be the numbers of balls of each of the two colors in R. Since the expected number of pairs of balls with different colors is at least $b/2$, we have $\mathbf{Exp}[a'] \leq (a - b/2)/2 = a/2 - b/4$ and $\mathbf{Exp}[b'] \leq (b - b/2)/2 = b/4$. In particular, the expected number of balls with the plurality color in R is at most $a/2 - b/4$. Since the plurality color in R is always the same as in the original problem, we conclude by induction:

$$\begin{aligned}
m_{a,b} &\leq \frac{a+b}{2} + 1 + \mathbf{Exp}[m_{a',b'}] \\
&\leq \frac{a+b}{2} + 1 + \mathbf{Exp}[a'] + \frac{\mathbf{Exp}[b']}{3} + \log\frac{a+b}{2} \\
&\leq \frac{a+b}{2} + 1 + \frac{a}{2} - \frac{b}{4} + \frac{b}{12} + \log(a+b) - 1 \\
&= a + \frac{b}{3} + \log(a+b)
\end{aligned}$$

The bound from the statement of the lemma now readily follows. □

We now present our algorithm for the Plurality problem:

Algorithm Plurality
(1) randomly permute the n balls;
(2) apply Algorithm Count to the first $\lfloor n/2 \rfloor + 1$ balls;
let a' and b' be the numbers of balls of each color, $a' \geq b'$, and let r' be the representative of the plurality color (if it exists)

(3) apply Algorithm COUNT to the next $2b'$ balls (or all the remaining balls if $2b' + \lfloor n/2 \rfloor + 1 > n$);
let a'' and b'' be the numbers of balls of each color, $a'' \geq b''$, and let r'' be the representative of the plurality color (if it exists)
(4) if no representative (r' or r'') exists, output "TIE" and halt;
(5) if exactly one of the representatives r' and r'' exists, output this representative as the ball with the plurality color and halt;
(6) compare r' and r'' and if they have the same color, output r' as the ball with the plurality color and halt;
(7) otherwise continue by comparing all the remaining balls to r'—this determines the color of all the balls relatively to r' and thus eventually determines the correct answer—stop and output the answer as soon as it is known.

Theorem 3.2. *Algorithm* PLURALITY *correctly determines the plurality color. For every coloring of the balls, the expected number of Paul's questions does not exceed* $2n/3 + O(\sqrt{n \log n})$.

Proof. We first show that Algorithm PLURALITY correctly determines the plurality color. If there is no representative r', then all the balls are examined, and the plurality color is correctly determined. If the representative r' exists and either there is no r'' or the color of r'' is the same as the color of r', the number of balls with the same color as r' is at least $a' + a'' \geq \lfloor n/2 \rfloor + 1$ and thus the color of r' is the plurality color. Otherwise, all the balls are examined in the last step and the plurality color is correctly determined.

Next, we analyze the expected number of Paul's questions. Suppose that the number of balls with the plurality color is a and let b be the number of balls with the other color. If $b \leq n/12$, then $b' \leq b$ and the entire number of examined balls does not exceed $n/2 + 2b' + 1 \leq 2n/3 + 1$ even in the worst case. In the rest of the proof, we assume that $b \geq n/12$.

Since the number of Paul's question never exceeds n, the contribution of the cases which happen with probability $O(1/n)$ is only a constant. Therefore, we restrict our analysis to the case when such events do not happen. This excludes the cases when the sample of the first $n/2$ or the next $2b'$ balls is bad. More precisely, let B' be the number of balls of the smaller color among $\lfloor n/2 \rfloor + 1$ balls examined in Step (2). The expectation of B' is $b/2 + O(1)$ and by Chernoff's bound, the probability that $|B' - b/2| \geq \sqrt{n \log n}$ is at most $O(1/n)$ (see [7, 9], for example). Thus we assume for the rest of the proof that $|B' - b/2| < \sqrt{n \log n}$.

We now distinguish two cases. The first case is $b \geq n/2 - 25\sqrt{n \log n}$. Then $|B' - n/4| \leq 14\sqrt{n \log n}$ and, no matter if r' has the plurality color, we have $b' \geq n/4 - 14\sqrt{n \log n}$. Therefore, there are at most $O(\sqrt{n \log n})$ balls not examined in Steps (2) and (3), and thus $b'' \geq n/4 - O(\sqrt{n \log n})$. By Lemma 3.1, the expected number of questions in the two runs of Algorithm COUNT is at most:

$$\begin{aligned} \frac{n}{2} - \frac{2b'}{3} + 2b' - \frac{2b''}{3} + O(\log n) &< \frac{n}{2} - \frac{n}{6} + \frac{n}{2} - \frac{n}{6} + O(\sqrt{n \log n}) \\ &= \frac{2n}{3} + O\left(\sqrt{n \log n}\right) . \end{aligned}$$

Since there are at most $O(\sqrt{n \log n})$ additional questions, the theorem follows.

The second case is $b < n/2 - 25\sqrt{n \log n}$. Then $B' \leq b/2 + \sqrt{n \log n} < n/4$ (using our previous assumption about B') and thus r' has the plurality color, $b' = B'$, and there are at most $2b' < n/2$ balls examined in Step (3). In addition, $b' > b/2 - \sqrt{n \log n} > n/25$ for large n.

Let B'' be the number of balls of the smaller color among the $2b'$ balls in Step (3). After removing the first $\lfloor n/2 \rfloor + 1$ balls, there are $b - B' = b/2 \pm \sqrt{n \log n}$ balls of the smaller color. Hence, $\mathbf{Exp}[B'']$ is $2b'b/n \pm \sqrt{n \log n}$ and the probability that $|B'' - 2b'b/n| \geq 2\sqrt{n \log n}$ is at most the probability that $|B'' - \mathbf{Exp}[B'']| \geq \sqrt{n \log n}$ which in turn is at most $O(1/n)$, by Chernoff's bound (using also $b' > n/25$). Similarly as before, the contribution of this case to the expected number of Paul's questions is a constant. Thus we assume for the rest of the proof that $|B'' - 2b'b/n| < 2\sqrt{n \log n}$.

Using the case condition and $b' > n/25$ for large n, we have:

$$\begin{aligned} B'' &< 2b'\frac{b}{n} + 2\sqrt{n \log n} \\ &< b' \cdot \frac{n - 50\sqrt{n \log n}}{n} + 2\sqrt{n \log n} \\ &< b' - \frac{b' \cdot 50\sqrt{n \log n}}{n} + 2\sqrt{n \log n} \\ &< b' - \frac{n \cdot 50\sqrt{n \log n}}{25 \cdot n} + 2\sqrt{n \log n} \;=\; b' . \end{aligned}$$

Thus r'' is a representative of the plurality color, $b'' = B''$, and the algorithm terminates in Step (6). By Lemma 3.1, the expected number of Paul's questions does not exceed:

$$\begin{aligned}
\frac{n}{2} - \frac{2b'}{3} + 2b' - \frac{2b''}{3} + O(\log n) &\leq \frac{n}{2} + b'\left(\frac{4}{3} - \frac{4b}{3n}\right) + O\left(\sqrt{n \log n}\right) \\
&\leq \frac{n}{2} + \frac{b}{2}\left(\frac{4}{3} - \frac{4b}{3n}\right) + O\left(\sqrt{n \log n}\right) \\
&\leq \frac{2n}{3} + O\left(\sqrt{n \log n}\right) .
\end{aligned}$$

The last inequality holds since the quadratic function of b is maximized for $b = n/2$. □

3.2 Lower Bound for Two Colors

The proofs of the lower bounds presented in this subsection and in the next section are based on Yao's principle [9, 11] (its easy direction). Instead of showing that for every probabilistic strategy of Paul, there is a coloring with large expected number of questions, we construct a probability distribution on the colorings such that the expected number of Paul's questions (with respect to the distribution) for any deterministic strategy is large. It follows that for any probabilistic strategy, the expected number of questions is large, averaging both over the distribution and the randomness of the strategy. Consequently, for any probabilistic strategy, there exists a fixed input that needs at least the same expected number of questions.

In this subsection, we choose the input distribution to be the uniform distribution on all the 2^n colorings of n balls. In Paul's graph G, every component G_i of G uniquely determines the distribution of the colors among the balls corresponding to its vertices, up to their swap. The *advantage* of a component G_i, $\mathrm{Adv}(G_i)$, is the (absolute value of) the difference between the numbers of balls in its two parts. The *advantage* of the graph G is defined as $\mathrm{Adv}(G) = \sqrt{\sum_i \mathrm{Adv}(G_i)^2}$ (note that this is consistent in case of a single component). A component G_i is said to be *balanced* if $\mathrm{Adv}(G_i) = 0$.

A question of Paul is called *essential* if it corresponds to an edge between two different components of Paul's graph. In any strategy for two colors, the answers of questions that are not essential are determined by the previous answers. Therefore non-essential questions can be omitted. From now on we assume, without loss of generality, that all Paul's questions are essential.

Lemma 3.3. *Fix a deterministic strategy of Paul. The expected advantage of the final graph of Paul with respect to the uniform distribution on the* 2^n

colorings of n balls is at most $\sqrt{n}$.

Proof. We show that $\mathbf{Exp}[\mathrm{Adv}(G)^2] = n$ where G is the final graph of Paul. Since $\mathbf{Exp}[X^2] \geq (\mathbf{Exp}[X])^2$ for every random variable X, the statement of the lemma then follows.

The deterministic strategy of Paul can be viewed as a decision binary tree whose nodes correspond to Paul's questions. We prove by induction on the number of leaves of a decision binary tree that the average advantage of Paul's graph in the leaves of any decision tree is $\sqrt{n}$. The claim holds for the trivial decision tree: since Paul's graph G is comprised of n isolated vertices, $\mathrm{Adv}(G) = \sqrt{n}$.

Consider a leaf v of the decision tree. We add a new query corresponding to v (the node v has now two children). Suppose that p is the probability that the node v is reached in the original tree. Since the query is essential and the considered distribution is the uniform distribution on the 2^n colorings, each of the two children of v is reached with probability $p/2$ in the new tree.

Suppose that the edge corresponding to the question joins two components G_i and G_j with $\mathrm{Adv}(G_i) = x$ and $\mathrm{Adv}(G_j) = y$. Their original contribution to the average advantage was $p(x^2 + y^2)$. After the additional query, we have a single component, with contribution either $(x+y)^2$ or $(x-y)^2$, each with probability $p/2$. The new contribution is thus $((x+y)^2 + (x-y)^2)p/2 = (x^2+y^2)p$, i.e., it is equal to the original contribution. Since the contributions of the graphs corresponding to the other leaves of the decision tree remain the same, we conclude that the average advantage of Paul's graph corresponding to the leaves of the tree remains $\sqrt{n}$. □

Note that Lemma 3.3 in particular implies that it is very unlikely that the final graph of Paul contains a component with advantage significantly larger than $\sqrt{n}$.

Lemma 3.4. *Suppose that G is a final graph of Paul for the Plurality problem with two colors. If G contains m unbalanced components, then the advantage of G is at least m, i.e., $\mathrm{Adv}(G) \geq m$.*

Proof. For $m = 0$, the statement is trivial. If $m > 0$, there is a consistent input with plurality color, thus Paul has to choose a ball which always has the plurality color. Suppose Paul chooses a ball b from a component G_0 of G. Consider a coloring of the balls such that the balls of the larger parts of the remaining (at least) $m-1$ unbalanced components are colored with the color different from the color of b. Since the color of b is the plurality color in this coloring, the advantage of G_0 must be at least m. Therefore, the advantage of the entire graph G is at least m. $\square$

Theorem 3.5. *For every probabilistic strategy of Paul for the Plurality problem with n balls of two colors, there exists a coloring of the balls such that Paul asks at least $2n/3 - O(\sqrt{n})$ questions on average.*

Proof. As explained in the beginning of the section, by Yao's principle, it is enough to establish that the expected number of Paul's questions for any deterministic strategy is at least $2n/3 - O(\sqrt{n})$ with respect to the uniform distribution on the 2^n colorings of n balls.

Fix a deterministic strategy of Paul. We can assume that each question of Paul is essential. Upon each query, the number of unbalanced components in Paul's graph decreases on average by at most $3/2$: the number can decrease by more than 1 only if two unbalanced components merge into a single balanced component. Since both answers of Carol have probability $1/2$, the probability that a balanced component is created is at most $1/2$, and the expected decrease of the number of unbalanced components is at most $3/2$.

Let q be the expected number of Paul's questions. Since each question decreases the number of unbalanced components by at most $3/2$ on average, the expected number of unbalanced components in the final graph of Paul is at least $n-3q/2$. By Lemma 3.4, the average advantage of the final graph is at least $n-3q/2$. On the other hand, the average advantage of the final graph is at most $\sqrt{n}$. We conclude that $q \geq 2n/3 - O(\sqrt{n})$. $\square$

4 Lower Bounds for Large Numbers of Colors

In this section, we prove the lower bound $\Omega(kn)$ for the Partition and the Plurality problems with n balls of k colors in the probabilistic setting. Since any strategy for the Partition problem also yields a solution of the Plurality Problem, it would be sufficient to give a lower bound only for the Plurality problem. However, the proof of the lower bound for the Partition problem is simpler, we achieve a better constant, and the bound holds for all values of n and k (in the case of Plurality problem, the bound holds only if n is sufficiently large compared to the number k of colors).

4.1 Partition Problem

Let G be a graph of Paul. We define a potential of Paul's graph that roughly measures the progress achieved by Paul during the game. The *partition potential* of a vertex v of G is equal to 0, if v is adjacent to a red edge (recall that red edges correspond to equal comparisons), and to $\max\{0, k-1-\deg_G(v)\}$, otherwise.

Lemma 4.1. *If G is a final graph of Paul for the Partition problem such that it is consistent with some coloring that uses all k colors, then the partition potential of every vertex of G is zero.*

Proof. Suppose for a contradiction that the potential of a vertex v is not 0, i.e., there is no red edge incident with v in G and $\deg_G(v) < k-1$. Consider a coloring c of all the vertices of G that uses all the k colors (such a coloring exists by the assumption of the lemma) and uncolor the vertex v. Note that the coloring uses still at least $k-1$ colors. There are now at least two different colors that the vertex v can be assigned, as at most $\deg_G(v) < k-1$ colors for v are forbidden. Since at least one of such colors is used by another vertex, the partitions of the vertices corresponding to the possible extensions of c to v are different. □

Theorem 4.2. *Any probabilistic strategy for the Partition problem with n balls of k colors requires asking at least $(k-1)(n-k)/4$ questions on average for some coloring of the balls.*

Proof. By Yao's principle, it is enough to show that every deterministic strategy of Paul for the Partition problem with n balls of k colors requires asking at least $(k-1)(n-k)/4$ questions on average for a fixed probability distribution on the colorings. The distribution which we consider is the following: color the first k balls by mutually different colors and use the uniform distribution on all the k^{n-k} possible colorings of the remaining $n-k$ balls. Paul's graph initially contains blue edges among any two of the first k vertices and no other edges; thus the initial partition potential is $(k-1)(n-k)$. By Lemma 4.1 and the fact that any used coloring uses all k colors, the final potential is 0. We show that after each question, on average, the sum of the partition potentials of the vertices of Paul's graph decreases by at most four. The theorem then follows.

Assume that Paul asks for the comparison of the balls corresponding to the vertices v and w of G. Let p be the partition potential of v. If $p = 0$ (this includes the case that v corresponds to one of the first k balls), the potential of v cannot further decrease. If $p > 0$, the vertex v is not incident with a red edge. Any precoloring of all the vertices of G except for v, that is consistent with G, can be extended in at least $k - \deg_G(v)$ different ways to v. Therefore, the probability that the edge vw will be red is at most $1/(k - \deg_G(v)) = 1/(p+1) \leq 1/p$. In such case, the partition potential of v drops down from p to zero. Otherwise, the potential of v is decreased by one. Therefore, the potential of v decreases by at most 2 on average. Similarly, the potential of w decreases by at most 2 on average. Since the potentials of the remaining vertices do not change, the average decrease (after each question) of the sum of the partition potentials of all the vertices of Paul's graph is at most four. □

It seems interesting (and maybe not too hard) to close the gap between lower and upper bounds for the Partition problem with n balls with $k \geq 4$ colors in the probabilistic setting. Dvořák et al. [5] provided a strategy such that Paul asks $\frac{k^2+k-2}{2k}n + O(1) \approx kn/2$ questions on average. On the other hand, we established a lower bound $(k-1)(n-k)/4 \approx kn/4$ for the Partition problem in the probabilistic setting. Let us recall that there are matching lower and upper bounds of order $(k-1)n - O(k^2) \approx kn$ in the deterministic setting.

4.2 Plurality Problem

We first introduce notation used throughout this subsection. A coloring of n balls with k colors is *nice* if there exists the plurality color and the number of balls of any of the colors does not exceed $n/k + n^{2/3}$. Note that if k is fixed, then Chernoff's bound implies that a random coloring of n balls with k colors is nice with probability $1 - o(1)$ (as n tends to infinity). A vertex of Paul's graph is *free* if it is incident with no red edge and with less than $\lfloor k/3 \rfloor$ blue edges. The other vertices are said to be *non-free*. The *plurality potential* of a vertex v of Paul's graph is $\lfloor k/3 \rfloor - \deg_G(v)$ if v is free, and 0 otherwise.

Lemma 4.3. *If G is a final graph of Paul for the Plurality problem of n balls with $k \geq 3$ colors such that G is consistent with some nice coloring, then G contains at most $n/3 + 2kn^{2/3}$ free vertices. In particular, the sum of the plurality potentials of the vertices of G is at most $\lfloor k/3 \rfloor \cdot n/3 + o(n)$ for a fixed k.*

Proof. Assume for a contradiction that G has more than $n/3 + 2kn^{2/3}$ free vertices. Depending on Paul's answer, we construct a coloring that is consistent with G but for which Paul's answer is incorrect. We distinguish two cases.

In the first case, suppose that Paul claims that a ball b has the plurality color. Color the ball b and all the balls corresponding to non-free vertices in G as in the nice coloring that exists by the assumption of the theorem. Since the coloring is nice, at most $n/k + n^{2/3}$ balls have the same color as b. Fix any $\lfloor k/3 \rfloor$ colors distinct from the color of the ball b. Recolor all the balls corresponding to the free vertices greedily by the fixed $\lfloor k/3 \rfloor$ colors in such a way that the resulting coloring is consistent with G. In case that the vertex corresponding to b is free, the ball b keeps its original color. Note that such a recoloring is always possible because each free vertex is incident with at most $\lfloor k/3 \rfloor - 1$ blue edges (that forbids at most $\lfloor k/3 \rfloor - 1$ colors at each such vertex). Since the (at least $n/3 + 2kn^{2/3}$) balls corresponding to free vertices are colored with $\lfloor k/3 \rfloor$ colors, there exists a color used for more than $n/k + n^{2/3}$ such balls. Then, the color of b is not the plurality color in the constructed coloring, contradicting Paul's claim.

In the second case, suppose that Paul claims that there is a tie. Color all the balls corresponding to non-free vertices in G as in the nice coloring. Fix any $\lfloor k/3 \rfloor$ colors and use them to color all the remaining balls greedily and consistently with G, as in the previous case. If there is no tie, we have a coloring that contradicts Paul's answer. Otherwise, similarly as in the previous case, there exists a color used for at least $n/k + 2n^{2/3}$ such balls and thus any color class that has the maximum number of balls contains a ball corresponding to a free vertex. Choose from all but one of the maximal color classes such a vertex and recolor them one by one by a color not used for any ball corresponding to a free vertex; there is always a choice of at least $\lfloor k/3 \rfloor$ colors, as at most $\lfloor k/3 \rfloor$ colors are used for the original coloring the free vertices and at most $\lfloor k/3 \rfloor$ free vertices are recolored. This coloring has no tie, as no color used for recoloring can have more than $n/k + n^{2/3}$ balls, contradicting Paul's claim. □

Theorem 4.4. *Let $k \geq 3$ be a fixed integer. For any probabilistic strategy for the Plurality problem with n balls of k colors, there exists a coloring such that Paul asks at least $\lfloor k/3 \rfloor \cdot 2n/9 - o(n)$ questions on average.*

Proof. We again apply Yao's principle and show that every deterministic strategy for the Plurality problem requires asking at least $\lfloor k/3 \rfloor \cdot 2n/9 - o(n)$ questions on average for the uniform distribution on k^n colorings of n

balls with k colors. Fix a deterministic strategy for the Plurality problem. Since a random coloring is nice with probability $1 - o(1)$, we may analyze the strategy only for nice colorings. The initial potential is $\lfloor k/3 \rfloor n$. By Lemma 4.3, the sum of the plurality potentials of the vertices of Paul's final graph is at least $\lfloor k/3 \rfloor \cdot n/3 + o(n)$. We show that the expected decrease of this sum is at most 3 after every single question of Paul. This implies the required bound.

Consider a moment when Paul asked for the comparison of the balls corresponding to the vertices v and w of G. If v is non-free, then its potential does not change after Carol's answer. Assume that v is free. For any precoloring of all the vertices of G except for v, there are at least $k - \lfloor k/3 \rfloor \geq 2k/3$ ways how to extend the precoloring to v. Therefore, the probability that Carol answers that the balls have the same color (with respect to the uniform distribution on all the coloring) is at most $3/2k$. In such case, the plurality potential of v decreases by at most $k/3$. Otherwise, it decreases by one. In particular, the plurality potential of v decreases by at most $1 + 3/2k \cdot k/3 = 1.5$ on average. Similarly, the plurality potential of w decreases by at most 1.5 on average. Since the plurality potentials of the remaining vertices of G do not change, the average decrease of the sum of the plurality potentials is at most three. □

References

[1] Aigner M.: Variants of the majority problem, Discrete Applied Mathematics **137** (2004) 3–25.

[2] Aigner M., De Marco G., Montangero M.: The plurality problem with three colors, in: Diekert V., Habib M. (eds.): STACS 2004, LNCS 2296, Springer-Verlag, Berlin, Heidelberg, 2004, 513–521.

[3] Aigner M., De Marco G., Montangero M.: The plurality problem with three colors and more, submitted.

[4] Alonso L., Reingold E. M., Schott R.: Determining the majority, Information Processing Letters **47** (1993) 253–255.

[5] Dvořák Z., Jelínek V., Král' D., Kynčl J., Saks M.: Three optimal algorithms for balls of three colors, in Diekert V., Durand B. (eds.): STACS 2005, LNCS 3404, Springer-Verlag, Berlin, Heidelberg, 2005, 206–217. accepted to STACS 2005.

[6] Fisher M., Salzberg S.: Finding a majority among n votes, Journal of Algorithms **3** (1982) 375–379.

[7] Hagerup T., Rüb C.: A guided tour Chernoff bounds, Inform. Process. Letters **33** (1989) 305–308.

[8] De Marco G., Pelc A.: Randomized algorithms for determining the majority on graphs, in: Rovan B., Vojtás P. (eds.): MFCS 2003, LNCS 2747, Springer-Verlag, Berlin, Heidelberg, 2003, 368–377.

[9] Motwani R., Raghavan P.: Randomized Algorithms, Cambridge University Press, 1995.

[10] Saks M., Werman M.: On computing majority by comparisons, Combinatorica **11** (1991) 383–387.

[11] Yao A. C.-C.: Probabilistic computations: Towards a unified measure of complexity, in: Proc. of the 17th Annual Symposium on Foundations of Computer Science (FOCS), 1977, 222–227.