Algorithmica (2007) 48: 147–157
DOI: 10.1007/s00453-007-0060-0

Algorithmica

Oblivious and Adaptive Strategies for the Majority and Plurality Problems[1]

Fan Chung,[2] Ron Graham,[2] Jia Mao,[2] and Andrew Yao[3]

Abstract. In the well-studied *Majority problem*, we are given a set of n balls colored with two or more colors, and the goal is to use the minimum number of color comparisons to find a ball of the majority color (i.e., a color that occurs for more than $n/2$ times). The *Plurality problem* has exactly the same setting while the goal is to find a ball of the dominant color (i.e., a color that occurs most often). Previous literature regarding this topic dealt mainly with adaptive strategies, whereas in this paper we focus more on the oblivious (i.e., non-adaptive) strategies. Given that our strategies are oblivious, we establish a linear upper bound for the *Majority problem* with arbitrarily many different colors assuming a majority label exists. We then show that the *Plurality problem* is significantly more difficult by establishing quadratic lower and upper bounds. In the end we also discuss some generalized upper bounds for adaptive strategies in the *k-color Plurality problem*.

Key Words. Majority problem, Plurality problem, Oblivious strategy, Adaptive strategy, Ramanujan graph.

1. Introduction. The earliest variant of the *Majority problem* was first raised by Moore in 1982 in connection with problems in the design of fault-tolerant computer systems. (It appeared in an equivalent setting of finding the majority vote among n processors with minimum number of paired comparisons [18].) In the colored-ball setting, we are given a set of n balls, each of which is colored in one of $k \in Z^+$ possible colors $\varphi = \{c_1, c_2, \ldots, c_k\}$. We can choose any two balls a and b and ask questions of the form "Do a and b have the same color?" Our goal is to identify a ball of the *majority color* (i.e., this color occurs more than half of the time) or determine that there is no majority color, using the minimum possible number of questions.

We can view this problem as a game played between two players: **Q**, the Questioner, and **A**, the adversary. **Q**'s role is to ask a sequence of queries $Q(a, b) :=$ "Is $\varphi(a) = \varphi(b)$?" **A** can answer each such query with the hope of extending the game as long as possible before **Q** can finally identify a ball of the majority color or determine that there is no majority. For the case when $k = 2$, a number of proofs were given (see [5], [19], and [22]) showing that $n - w_2(n)$ color comparisons are necessary and sufficient in the worst case, where $w_2(n)$ is the number of 1's in the binary representation of n.

[1] The research by Fan Chung was supported in part by NSF Grant DMS 0100472 and ITR 0205061. Ron Graham's research was supported in part by NSF Grant CCR-0310991, and that by Jia Mao was supported in part by an NSF Graduate Fellowship. The research by Andrew Yao was supported in part by CCR-0310466 and CCF-0426582.
[2] Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093, USA. {fan,rgraham,jiamao}@cs.ucsd.edu.
[3] Department of Computer Science, Tsinghua University, Beijing 100084, People's Republic of China. andrewcyao@tsinghua.edu.cn.

Received February 18, 2005; revised August 30, 2006. Communicated by L. Wang.
Online publication May 8, 2007.

Recently, several variants of this problem were also analyzed by Aigner [1]. One natural generalization is the so-called *Plurality problem* where the goal is to identify a ball of the *dominant color* (i.e., this color occurs more often than any other color). Linear upper and lower bounds were given for $k = 3$ colors for this variant in [2].

In general, two types of strategies can be considered for **Q**. These are *adaptive* strategies in which each query can depend on the answers given to all previous queries, and *oblivious* (or non-adaptive) strategies in which all the queries must be specified before **A** is required to answer any of them. Clearly, in the oblivious case, **A** has more opportunity to be evasive. To the best of our knowledge, except for the case when $k = 2$ [1], very little is known about the bounds for oblivious strategies for other variants of the *Majority problem*.

Let $MO_*(n)$ denote the minimum number of queries needed by **Q** in the *Majority problem* for arbitrarily many colors (i.e., k is not fixed) over all oblivious strategies, and let $PO_k(n)$ (resp. $PA_k(n)$) denote the corresponding minimum in the *Plurality problem* for fixed k colors over all oblivious (resp. adaptive) strategies. In this paper we establish a linear upper bound for $MO_*(n)$ assuming a majority color exists, quadratic bounds for $PO_k(n)$, and also a generalized linear upper bound for $PA_k(n)$.

2. Oblivious Strategy for the Majority Problem. Consider the case where the number of possible colors k is unrestricted. In principle, this is a more challenging situation for **Q**. At least the upper bounds we have in this case are weaker than those for $k = 2$ colors [11]. A linear upper bound can be shown assuming the existence of a majority color. We also remark that without such an assumption, a quadratic lower bound can be proven to be very close to the worst-case $\binom{n}{2}$ upper bound using similar arguments as in the proof of Theorem 2 in Section 3.[4] In fact, this quadratic lower bound can also be extended to the general case where the goal is to identify a ball whose color has appeared more than $t \geq \lceil n/2 \rceil$ times.

THEOREM 1. *For all n,*

$$MO_*(n) \leq (1 + o(1))21n,$$

assuming a majority color exists.

PROOF. We use an auxiliary graph to represent the game configuration at each step. To avoid the confusion between coloring of the vertices and coloring of the edges, in this proof we consider another equivalent setting of the game. Given a set of n elements, each of which has a label from the label set $\varphi = \{l_1, l_2, \ldots\}$, **Q** wants to identify one element of the majority label (or in the case of a tie, claim that there is none) using only pairwise equal/unequal label comparisons of elements. In the auxiliary graph, the n elements are the nodes of this undirected graph. Any equal/unequal label comparison query corresponds to the selection of two nodes, and the answer given by **A** corresponds to a colored edge drawn between them. We let the color *blue* represent the *equal* answer,

[4] The bound of $\binom{n-1}{2}$ in [20] serves as a natural lower bound here.

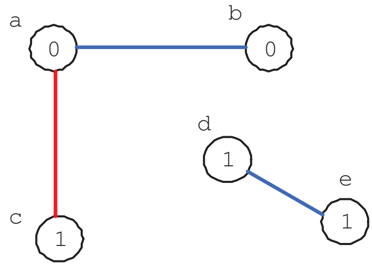

Fig. 1. An auxiliary graph on a set of $n = 5$ elements $\{a, b, c, d, e\}$ each with a binary label. The *blue* edges denote the "equal" answer for label comparisons and the *red* edges denote the "unequal" answers. Note that the actual labeling is not revealed to **Q** until the end of game.

and *red* the *unequal* answer. So the vertex set of G is fixed, but the edge set is growing because every round of query and answer adds a colored edge to this graph. See Figure 1 for an example.

If the existence of a majority label is known a priori, we denote the majority label by $l \in \varphi$. The set of the remaining possible labels is still unknown. In the game configuration graph H, the vertices correspond to the elements, the queries correspond to edges, and the answers are represented by edge colors. The vertex set of H is $V(H) = \{v_1, \dots, v_n\}$. An edge $\{v_i, v_j\}$ in H corresponds to the query $Q(v_i, v_j) :=$ "Is $\varphi(v_i) = \varphi(v_j)$?" The edge is colored *blue* if they are equal, and *red* if they are not equal.

By a *valid assignment* φ on $V(H)$ we mean that:

(i) $\{v_i, v_j\}$ is blue $\Rightarrow \varphi(v_i) = \varphi(v_j)$,
(ii) $\{v_i, v_j\}$ is red $\Rightarrow \varphi(v_i) \neq \varphi(v_j)$,
(iii) $|\varphi^{-1}(l)| > n/2$.

An oblivious strategy for **Q** is simply a layout of edges in H such that **Q** can correctly identify an element of a majority label regardless of the edge coloring **A** specifies. Intuitively, H should be reasonably well connected so that the elements of a majority label always form a large embedded blue component. This notion of good connectivity leads us naturally to the so-called *expander graphs* [3].

We use a special class of expander graphs $X^{p,q}$, called Ramanujan graphs, which can be constructed for any primes p and q congruent to 1 modulo 4. They are the best expanders in the spectral sense [17].

$X^{p,q}$ has the following properties:

(i) $X^{p,q}$ has $n = \frac{1}{2}q(q^2 - 1)$ vertices.
(ii) $X^{p,q}$ is regular of degree $p + 1$.
(iii) The adjacency matrix of $X^{p,q}$ has the large eigenvalue $\lambda_0 = p + 1$ and all other eigenvalues λ_i satisfying $|\lambda_i| \leq 2\sqrt{p}$.

We use the following discrepancy inequality (see [3] and [9]) for a d-regular graph H of n vertices with eigenvalues satisfying

$$\max_{i \neq 0} |\lambda_i| \leq \delta.$$

For any subset $X, Y \subseteq V(H)$, we have

$$\left|e(X,Y) - \frac{d}{n}|X|\,|Y|\right| \le \frac{\delta}{n}\sqrt{|X|(n-|X|)|Y|(n-|Y|)}, \tag{1}$$

where $e(X, Y)$ denotes the number of edges between X and Y.

Applying (1) to $X^{p,q}$, we obtain, for all $X, Y \subseteq V(X^{p,q})$,

$$\left|e(X,Y) - \frac{p+1}{n}|X|\,|Y|\right| \le \frac{2\sqrt{p}}{n}\sqrt{|X|(n-|X|)|Y|(n-|Y|)}, \tag{2}$$

where $n = \frac{1}{2}q(q^2-1) = |V(X^{p,q})|$.

The oblivious strategy for $\mathbf{Q}$ is first to construct a Ramanujan graph $X^{p,q}$ on the vertex set $V(H) = \{v_1, \ldots, v_n\}$. Let φ be a valid assignment of $V(H)$ and consider the subgraph M of $X^{p,q}$ induced by $\varphi^{-1}(\ell)$ (the majority-labeled vertices of $X^{p,q}$ under the mapping φ).

CLAIM. *Suppose $p \ge 38$. Then M has a connected component C with size at least $c'n$, where $c' > \frac{1}{3}$.*

PROOF. We use (2) with $X = C$, the largest connected component of M, and $Y = \varphi^{-1}(\ell)\backslash X$. Write $|\varphi^{-1}(\ell)| = \alpha n$ and $|C| = \beta n$. Since $e(X, Y) = 0$ for this choice, then by (2) we have

$$\begin{aligned}
(p+1)^2|X|\,|Y| &\le 4p(n-|X|)(n-|Y|),\\
(p+1)^2\beta(\alpha-\beta) &\le 4p(1-\beta)(1-\alpha+\beta),\\
\beta(\alpha-\beta) &\le \frac{4(1-\alpha)p}{(p-1)^2}.
\end{aligned}$$

There are two possibilities:

$$\beta \ge \frac{1}{2}\left(\alpha + \sqrt{\alpha^2 - \frac{16(1-\alpha)p}{(p-1)^2}}\right) \quad \text{or} \quad \beta \le \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \frac{16(1-\alpha)p}{(p-1)^2}}\right).$$

Subcase (a).

$$\begin{aligned}
\beta &\ge \frac{1}{2}\left(\alpha + \sqrt{\alpha^2 - \frac{16(1-\alpha)p}{(p-1)^2}}\right)\\
&> \frac{1}{4}\left(1 + \sqrt{1 - \frac{32p}{(p-1)^2}}\right) \qquad \left(\text{since } \alpha \ge \tfrac{1}{2}\right)\\
&> \tfrac{1}{3} \qquad (\text{since } p \ge 38)
\end{aligned}$$

as desired.

Subcase (b).

$$\begin{aligned}\beta &\leq \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \frac{16(1-\alpha)p}{(p-1)^2}}\right)\\ &\leq \frac{8(1-\alpha)p}{\alpha(p-1)^2}.\end{aligned}$$

Thus, we can choose a subset F of some of the connected components whose union $\bigcup F$ has size $xn = |\bigcup F|$ satisfying

$$\frac{\alpha}{2} - \frac{4(1-\alpha)p}{\alpha(p-1)^2} \leq x < \frac{\alpha}{2} + \frac{4(1-\alpha)p}{\alpha(p-1)^2}. \tag{3}$$

Now we apply the discrepancy inequality (2) again by choosing $X = \bigcup F$ and $Y = \varphi^{-1}(l)\backslash X$. We have

$$\begin{aligned}(p+1)^2 x(\alpha - x) &\leq 4p(1-x)(1-\alpha+x),\\ x(\alpha - x) &\leq \frac{4(1-\alpha)p}{(p-1)^2}.\end{aligned}$$

However, it is easily checked that because of (3) this is not possible for $\alpha \geq \frac{1}{2}$ and $p \geq 38$. Hence, subcase (b) cannot occur. This proves the claim. □

Now we are ready to prove Theorem 1. We will show that when $p \geq 38$, an element of the majority label can always be identified after all the queries are answered. Suppose we have an arbitrary blue/red coloring of the edges of $H = X^{p,q}$ with $p \geq 38$, and φ a valid assignment on $V(H) = V(X^{p,q})$. Consider the connected components formed by the blue edges of $X^{p,q}$. By the Claim there is at least one blue component of size at least $\frac{1}{3}n$ since $p \geq 38$. Call any such blue component *large*.

If there is only one large component then we are done, i.e., every element in it must be of the majority label. Since $p \geq 38$, there cannot be three large blue components. So the only remaining case is that we have exactly two large blue components, say S_1 and S_2. Again, if either $S_1 \subseteq \varphi^{-1}(l)$ or $S_2 \subseteq \varphi^{-1}(l)$ is forced, then we are done. So we can assume there is a valid assignment φ_1 with $S_1 \subseteq \varphi_1^{-1}(l)$, $S_2 \subseteq \varphi_1^{-1}(\neg l)$, and a valid assignment φ_2 with $S_2 \subseteq \varphi_2^{-1}(l)$, $S_1 \subseteq \varphi_2^{-1}(\neg l)$ (where $\neg l$ denotes any label in the label set except l).

We write $S_i' = \varphi_i^{-1}(l)\backslash S_i$, $i = 1, 2$. Clearly we must have $A := S_1' \cap S_2' \neq \emptyset$. Also note that $|A| \leq n - |S_1| - |S_2| < (16p/(p-1)^2)n$.

Define $B_1 = S_1'\backslash A$, $B_2 = S_2'\backslash A$. Observe that there can be no edge between A and $S_1 \cup S_2 \cup B_1 \cup B_2$. Now we are going to use (2) again, this time choosing $X = A$, $Y = S_1 \cup S_2 \cup B_1 \cup B_2$. Note that

$$n > |Y| = |\varphi_1^{-1}(l)| - |A| + |\varphi_2^{-1}(l)| - |A| > n - 2|A|.$$

Since $e(X, Y) = 0$, we have, by (2),

$$\begin{aligned}(p+1)^2|X|\,|Y| &\leq 4p(n-|X|)(n-|Y|),\\ (p+1)^2|A|(n-2|A|) &\leq 4p(n-|A|)2|A|.\end{aligned}$$

However, this implies

$$\begin{aligned}(p+1)^2(n-2|A|) &\le 8p(n-|A|),\\ \text{i.e.,}\qquad n((p+1)^2-8p) &\le 2|A|((p+1)^2-4p)\\ &\le 2|A|(p-1)^2\\ &< 32pn,\\ (p+1)^2-8p &< 32p,\end{aligned}$$

which is impossible for $p \ge 38$.

Setting $p = 41$ (so that $X^{p,q} = X^{41,q}$ is regular of degree $p+1 = 42$), we see that $X^{41,q}$ has $21n$ edges. This shows that Theorem 1 holds when $n = \frac{1}{2}q(q^2-1)$ for a prime $q \equiv 1 \pmod 4$.

If $\frac{1}{2}q_i(q_i^2-1) < n < \frac{1}{2}q_{i+1}(q_{i+1}^2-1) = n'$ where q_i and q_{i+1} are consecutive primes of the form 1 (mod 4), we can simply augment our initial set $V(H)$ to a slightly larger set $V'(H)$ of size n' by adding $n'-n = \delta(n)$ additional elements of the majority label. Standard results from number theory show that $\delta(n) = o(n^{3/5})$, for example. Since the Ramanujan graph query strategy of **Q** actually identifies $\Omega(n')$ elements of the majority label l from $V'(H)$ (for fixed p) then it certainly identifies an element of the majority label of our original set $V(H)$.

This proves Theorem 1 for all n. □

Instead of using Ramanujan graphs, we can consider random graphs $G(n, p)$ on n vertices and the probability of each pair to be chosen as an edge is p. Because random graphs with the best achievable spectral gap are known to exist abundantly, the same proof using discrepancy inequalities still works. For the claims to hold, the degree for such a random graph only has to be greater than or equal to 39. So, we can at least reduce the constant 21 to 19.5 if we do not require explicit constructions.

3. Oblivious Strategies for the Plurality Problem. The *Plurality problem* generalizes the *Majority problem* where the goal is to identify a ball whose color occurs most often or show that there is no dominant color. When there are only $k = 2$ possible colors, the *Plurality problem* degenerates to the *Majority problem* with two colors, and hence there are tight bounds for both adaptive ($n - w_2(n)$) and oblivious ($n-2$ for n odd, $n-3$ for n even) strategies [10].

In general, it seems clear that the *k-color Plurality problem* should take more queries than the corresponding *Majority problem*. However, exactly how much more difficult it is compared with the *Majority problem* was not so clear to us at the beginning. Similar arguments using concentration inequalities in random graphs seemed possible for achieving a linear upper bound. In the following section we prove the contrary by establishing a quadratic lower bound, even for the case when $k = 3$. Also note that the lower bound would remain quadratic even if we assume the existence of a plurality color through slight modification on the proof of Theorem 2. Intuitively, this is because the existence of a majority color gives us much more information than the existence of a plurality color.

3.1. *Lower Bound*

THEOREM 2. *For the Plurality problem with $k = 3$ colors, the number of queries needed for any oblivious strategy satisfies*

$$PO_3(n) > \frac{n^2}{6} - \frac{3n}{2}.$$

PROOF. Consider any query graph G with n vertices and at most $n^2/6 - 3n/2$ edges. Therefore there must exist a vertex v with $\deg(v) \leq n/3 - 3$. Denote the neighborhood of v by $N(v)$ (which consists of all vertices adjacent to v in G), and the remaining graph by $H = G \backslash (N(v) \cup \{v\})$. Hence, H has at least $2n/3 + 2$ vertices.

Now split H into three parts H_1, H_2, and X where $|H_1| = |H_2| + 1$ and $|X| \leq 1$. Assign color 1 to all vertices in H_1, color 2 to all vertices in H_2, color 3 to all vertices in $N(v)$ and X, and color 1 or 2 to v. Note that based on either one of the two possible color assignments, all query answers are forced.

Since color 3 cannot possibly be the dominant color, we see that whether color 1 is the dominant color or there is no dominant color (because of a tie) solely depends on the color of v, which the Questioner cannot deduce from the query answers.

This proves that the lower bound to the number of queries needed for the oblivious strategy is $n^2/6 - 3n/2$. □

This quadratic lower bound also applies to all $k \geq 3$ colors for the *Plurality problem* using oblivious strategies, since we do not need to use any additional colors beyond 3 for this argument.

3.2. *Upper Bound*. A trivial upper bound is the maximum number of possible queries we can ask, which is $\binom{n}{2}$. In this section we show that for fixed k colors and n sufficiently large, essentially $(1 - 1/k)\binom{n}{2}$ queries suffice for oblivious strategies, using probabilistic arguments. Again we use the equivalent element/label setting as specified in Theorem 1 to avoid confusion between edge coloring and vertex coloring.

We consider the usual random graphs $G(n, p)$ where n is the number of vertices and p is the probability for any particular edge to be included in the graph. We also use the standard notion "almost surely" to denote "probability goes to 1 as $n \to \infty$". Our upper bound is based on the following lemma.

LEMMA 1. *If $p \geq 1 - 1/k + \varepsilon$ for some $\varepsilon > 0$, a random graph $G \in G(n, p)$ almost surely has the property that for any subset S of vertices of size at least n/k, the graph $G(n, p)[S]$ induced by S is connected.*

PROOF. Consider $G \in G(n, 1 - 1/k + \varepsilon)$ where $\varepsilon > 0$, the expected vertex degree is $np = (1 - 1/k + \varepsilon)n$. Using concentration inequalities (see Theorem 4 of [12]), almost surely the degree for any vertex is lower bounded by $(1 - 1/k + \varepsilon/2)n$. Given any subset of vertices S with size $|S| \geq n/k$ and any vertex v in S, by degree concentration v must be adjacent to at least $(\varepsilon/2)n$ other vertices in S. Therefore there is no connected component of $G(n, p)[S]$ with order smaller than $(\varepsilon/2)n$.

Now consider $G \in G(n, p)$ with $p > 0$ fixed. Then almost surely for any two disjoint vertex sets T and U in G such that $|T|, |U| = m = \Omega(\log n)$, there is an edge joining a vertex in T to a vertex in U. This is because the probability that this fails to hold is upper bounded by

$$n^{2m}(1-p)^{m^2} = e^{m(2\log n + m\log(1-p))},$$

which goes to 0 when $n \to \infty$ because $m = \Omega(\log n)$.

Therefore G almost surely has the property that for any S with size at least n/k, the components in $G(n, p)[S]$ are so large (i.e., of size $(\varepsilon/2)n$) that they all have to be connected. This proves the lemma. □

An oblivious strategy for **Q** can then be specified by a random graph G with this property. Regardless of how **A** colors the query edges, for each potential dominant label (i.e., it has occurred for at least n/k times), **A** cannot avoid forming an induced blue component of every vertex with that label and the calculation for **Q** to figure out the correct answer is the following:

Q's calculation:

- Remove all the red edges from H to obtain H' and then remove all isolated vertices from H'.
- Identify all connected components in H' with size at least n/k.
- If one such component has size strictly bigger than any other component, all elements in it must have the dominant label; otherwise, there is no dominant label.

Hence the following upper bound holds for **Q**'s oblivious strategies. Note that the previous probabilistic arguments need n to be sufficiently large. In particular, the threshold value is a function of ε, say $n_0(\varepsilon)$.

THEOREM 3. *For every* $\varepsilon > 0$,

$$PO_k(n) < \left(1 - \frac{1}{k} + \varepsilon\right)\binom{n}{2}$$

provided $n > n_0(\varepsilon)$.

4. Adaptive Strategies for the Plurality Problem. Aigner et al. [2] showed linear bounds for adaptive strategies for the *Plurality problem* with $k = 3$ colors. In this section we first note a linear upper bound for general k in this case, and then strengthen it using a generalized argument.

THEOREM 4. *For the Plurality problem with k colors where $k \in Z^+$, the minimum number of queries needed for any adaptive strategy satisfies*

$$PA_k(n) \leq (k-1)n - \frac{k(k-1)}{2}.$$

PROOF. There are k possible colors for the given n balls. We use k buckets, each for a different possible color. All buckets are empty initially. The first ball s_1 is put in the first bucket b_1. The second ball is compared against a ball from b_1; if they have the same color, it is put in b_1, otherwise, it is put in a new bucket b_2. Similarly, the ith ball has to be compared against a ball from every non-empty bucket (at most $(i-1) \leq k-1$ many of them). Therefore the number of comparisons is no more than

$$1 + 2 + \cdots + (k-1) + (k-1)(n-k) = (k-1)n - \frac{k(k-1)}{2}. \qquad \square$$

In [2] it was proved that $PA_3(n) \leq \frac{5}{3}n - 2$. We now give a generalized proof for all $k \geq 3$ in this setting.

THEOREM 5. *For the Plurality problem with k colors where $k \in Z^+$, the minimum number of queries needed for any adaptive strategy is*

$$PA_k(n) \leq \left(k - \frac{1}{k} - 1\right) n - 2.$$

PROOF. We denote the comparison of ball a against b by $(a : b)$, and define a *color class* to be a set of balls having the same color. There are two phases in this game. Given n balls $\{s_1, s_2, \ldots, s_n\}$, in Phase I the Questioner handles one ball at a time (except for the first query) and keeps a state vector v_i after ball s_i is handled. Each v_i is simply the list of color class cardinalities, in non-increasing order, $(a_{i1}, a_{i2}, \ldots, a_{ik})$ where $a_{i1} \geq a_{i2} \geq \cdots \geq a_{ik}$. The Questioner also keeps a representative ball for each of the largest $(k-1)$ color classes (if they are non-empty) for comparisons and updates this list accordingly whenever there is a change in the state vector.

CLAIM. *At every state, the Questioner has a strategy such that the total number t_i of comparisons up to v_i (inclusive) satisfies*

$$t_i \leq (k-1)a_{i1} + (k-2)\sum_{j=2}^{k-1} a_{ij} + (k-1)a_{ik} - 2.$$

PROOF. We proceed by induction. After the first comparison, $v_2 = (1, 1, 0, \ldots)$ or $(2, 0, \ldots)$, so $t_2 = 1 \leq (k-1) + (k-2) - 2 \leq 2(k-1) - 2$ because $k \geq 3$.

For $2 \leq i \leq n$, let $v_i = (a_{i1}, a_{i2}, \ldots, a_{ik})$ be the state vector and let $\{A_{i1}, A_{i2}, \ldots, A_{i(k-1)}\}$ be the set of corresponding representative balls (some may be null if the color class has cardinality 0). Now, ball s_{i+1} is to be handled as follows:

1. If $a_{i(k-1)} \neq a_{ik}$, we compare s_{i+1} with the representative balls in the following order:

$$(s_{i+1} : A_{i2}), (s_{i+1} : A_{i3}), \dots, (s_{i+1} : A_{i(k-1)}), (s_{i+1} : A_{i1})$$

   with a total number of no more than $(k-1)$ comparisons. Note whenever the Adversary answers *Yes*, we know to which color class s_{i+1} belongs, and hence, we can skip the remaining comparisons.
2. Otherwise, compare s_{i+1} with the representative balls in the following order:

$$(s_{i+1} : A_{i1}), (s_{i+1} : A_{i2}), \dots, (s_{i+1} : A_{i(k-2)})$$

   with a total number of no more than $(k-2)$ comparisons. If all these $(k-2)$ answers are *No*, we can increment $a_{i(k-1)}$ and set representative ball $A_{i(k-1)} := s_{i+1}$.

After identifying to which color class s_{i+1} belongs, only one of the numbers in v_i gets incremented by 1 and possibly moved forward, to maintain the non-increasing order in v_{i+1}. Using the above strategy, we can ensure that no more than $(k-2)$ comparisons have been used in this round unless a_{i1} or a_{ik} gets incremented, in which case their positions in the list do not change. Therefore, by the inductive hypothesis, we have

$$t_{i+1} \leq (k-1)a_{(i+1)1} + (k-2)\sum_{j=2}^{k-1} a_{(i+1)j} + (k-1)a_{(i+1)k} - 2.$$

This proves the claim. □

At state v_i, let r_i be the number of the remaining balls that have not been involved in any queries. Phase I ends when one of the following happens:

(A) $a_{i1} = a_{i2} = \cdots = a_{ik}$
(B) $r_i = a_{i1} - a_{i2} - 1$
(C) $r_i = a_{i1} - a_{i2}$

(Note that one of (A), (B), or (C) will eventually occur. Detailed arguments are similar as in [2].) To prove the theorem, we use induction where the cases for $n \leq 3$ are easy to verify. More comparisons may be needed in Phase II depending on in which case Phase I ends. If Phase I ends in case (A), we use the induction hypothesis; in case (B), no more comparisons are needed because A_{i1} is a Plurality ball; in case (C), we need no more than r_i more comparisons to identify A_{i1} or A_{i2} as a Plurality ball. In all cases, we can show (using the claim) with arguments similar to those in [2] that

$$PA_k(n) \leq (k-1)n - \frac{n}{k} - 2 = \left(k - \frac{1}{k} - 1\right)n - 2.$$

This proves the theorem. □

5. Conclusion. In this paper we established upper and lower bounds for oblivious and adaptive strategies used to solve the *Majority* and *Plurality problems*. These problems originally arose from applications in fault tolerant system design. However, the interactive nature of these problems also places them in the general area of interactive computing. It is therefore desirable to develop more techniques to solve this type of problem efficiently as well as to understand the limits of our ability in doing so.

Acknowledgements. The authors are grateful to Steve Butler and Kevin Costello for their contribution to one of the proofs. We also thank the anonymous reviewers for their careful review and helpful comments.

References

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

[2] M. Aigner, G. De Marco, and M. Montangero, The Plurality problem with three colors and more, *Theoret. Comput. Sci.* **337** (2005), 319–330.

[3] N. Alon, Eigenvalues and expanders, *Combinatorica* **6** (1986), 86–96.

[4] L. Alonso and E. Reingold, Determining plurality, Manuscript, 2006.

[5] L. Alonso, E. Reingold, and R. Schott, Determining the majority, *Inform. Process. Lett.* **47** (1993), 253–255.

[6] L. Alonso, E. Reingold, and R. Schott, Average-case complexity of determining the majority, *SIAM J. Comput.* **26** (1997), 1–14.

[7] P. M. Blecher, On a logical problem, *Discrete Math.* **43** (1983), 107–110.

[8] B. Bollobás, *Random Graphs*, 2nd edn., Cambridge Studies in Advanced Mathematics, 73, Cambridge University Press, Cambridge, 2001.

[9] F. R. K. Chung, *Spectral Graph Theory*, CBMS Lecture Notes, American Mathematical Society, Providence, RI.

[10] F. R. K. Chung, R. L. Graham, J. Mao, and A. C. Yao, Finding favorites, *Electronic Colloquium on Computational Complexity* (*ECCC*) (078), 2003.

[11] F. R. K. Chung, R. L. Graham, J. Mao, and A. C. Yao, Oblivious and adaptive strategies for the Majority and Plurality problems, *Proc. COCOON* 2005, pp. 329–338.

[12] F. R. K. Chung and L. Lu, Concentration inequalities and martingale inequalities—a survey, *Internet Math.* **3**(1) (2006), 79–127.

[13] Z. Dvorak, V. Jelinek, D. Kral, J. Kyncl, and M. Saks, Three optimal algorithms for balls of three colors, *Proc. STACS* 2005.

[14] P. Erdős and A. Rényi, On random graphs. I, *Publ. Math. Debrecen* **6** (1959), 290–291.

[15] M. J. Fischer and S. L. Salzberg, Finding a majority among n votes, *J. Algorithms* **3** (1982), 375–379.

[16] D. Kral, J. Sgall, and T. Tichỳ, Randomized strategies for the plurality problem, Manuscript, 2005.

[17] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, *Combinatorica* **8** (1988), 261–277.

[18] J. Moore, Proposed problem 81-5, *J. Algorithms* **2** (1981), 208–210.

[19] M. Saks and M. Werman, On computing majority by comparisons, *Combinatorica* **11**(4) (1991), 383–387.

[20] N. Srivastava, and A. D. Taylor, Tight bounds on plurality, *Inform. Proc. Lett.* **96** (2005), 93–95.

[21] A. Taylor and W. Zwicker, Personal communication.

[22] G. Wiener, Search for a majority element, *J. Statist. Plann. Inference* **100** (2002), 313–318.