Majority and Plurality Problems

Balázs Patkós

Rényi Institute

Workshop at Zhejiang Normal University, Jinhua, April 20, 2012

Joint work with D. Gerbner, G.O.H. Katona and D. Pálvölgyi

THE CONTEXT

THE CONTEXT

We are given n balls in an urn, each colored with one color.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

The aim is either to decide whether there exists a majority/plurality ball or even to show one (if there exists one).

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

The aim is either to decide whether there exists a majority/plurality ball or even to show one (if there exists one).

Note that if the number of colors is two, then the majority and the plurality problems coincide.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

The aim is either to decide whether there exists a majority/plurality ball or even to show one (if there exists one).

Note that if the number of colors is two, then the majority and the plurality problems coincide.

A ball is said to be k-majority if its color class contains at least k balls.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

The aim is either to decide whether there exists a majority/plurality ball or even to show one (if there exists one).

Note that if the number of colors is two, then the majority and the plurality problems coincide.

A ball is said to be k-majority if its color class contains at least k balls. By definition a ball is majority if and only if it is k-majority for $k = \lfloor n/2 \rfloor + 1$.

THE CONTEXT

We are given n balls in an urn, each colored with one color.

A majority ball is one such that its color class has size strictly larger than $n/2$.

A plurality ball is one such that its color class is strictly larger than any other color class.

The aim is either to decide whether there exists a majority/plurality ball or even to show one (if there exists one).

Note that if the number of colors is two, then the majority and the plurality problems coincide.

A ball is said to be k-majority if its color class contains at least k balls. By definition a ball is majority if and only if it is k-majority for $k = \lfloor n/2 \rfloor + 1$. It is natural to consider only the case when $k > n/2$ since then k-majority is a strengthening of majority.

A query is a pair of balls and the answer to the query tells us whether the two balls have the same color or not.

A query is a pair of balls and the answer to the query tells us whether the two balls have the same color or not.

Two main sorts of algorithms:

1) adaptive: the ith query might depend on the answers to the first $i-1$ queries,

A query is a pair of balls and the answer to the query tells us whether the two balls have the same color or not.

Two main sorts of algorithms:

1) adaptive: the ith query might depend on the answers to the first $i-1$ queries,

2) non-adaptive: all queries should be asked in advance.

A query is a pair of balls and the answer to the query tells us whether the two balls have the same color or not.

Two main sorts of algorithms:

1) adaptive: the ith query might depend on the answers to the first $i-1$ queries,

2) non-adaptive: all queries should be asked in advance.

Any non-adaptive strategy is adaptive, therefore the best non-adaptive strategy always uses at least as many queries as the best non-adaptive strategy.

History:

Theorem (Fisher and Salzberg)

If the number of possible colors is unknown, then the minimum number of queries in an adaptive search for a majority ball is $\lfloor 3n/2 \rfloor - 2$.

History:

Theorem (Fisher and Salzberg)

If the number of possible colors is unknown, then the minimum number of queries in an adaptive search for a majority ball is $\lfloor 3n/2 \rfloor - 2$.

Theorem (Saks and Werman)

If the number of colors is two, then the minimum number of queries needed in an adaptive search to find a majority ball is $n - b(n)$, *where* $b(n)$ *is the number of 1's in the binary representation of* n.

For a set Q of queries let the query graph G_Q be the graph where the vertices correspond to balls and two vertices are joined by an edge if and only if there exists a query in Q that asks for the comparison of the two corresponding balls.

For a set Q of queries let the query graph G_Q be the graph where the vertices correspond to balls and two vertices are joined by an edge if and only if there exists a query in Q that asks for the comparison of the two corresponding balls.

Theorem (Aigner, De Marco, and Montangero)

For any fixed positive integer c, if the number of possible colors is at most c, then the minimum number of queries needed in an adaptive search for a plurality ball is $O_c(n)$.

Let $M_c^*(n,k)$ ($M_c(n,k)$) denote the minimum number of queries needed to find a k-majority ball in a non-adaptive (adaptive) search if the number of colors is at most c and the number of balls is n.

Let $M^*_c(n,k)$ ($M_c(n,k)$) denote the minimum number of queries needed to find a k-majority ball in a non-adaptive (adaptive) search if the number of colors is at most c and the number of balls is n.

Theorem (Aigner)

For $n \geq 3$

$$M^*_2(n,k) = \begin{cases} n-1 & \textit{if } n < 2k-1 \\ n-2 & \textit{if } n = 2k-1. \end{cases}$$

Let $M_c^*(n,k)$ ($M_c(n,k)$) denote the minimum number of queries needed to find a k-majority ball in a non-adaptive (adaptive) search if the number of colors is at most c and the number of balls is n.

Theorem (Aigner)

For $n \geq 3$

$$M_2^*(n,k) = \begin{cases} n-1 & \textit{if } n < 2k-1 \\ n-2 & \textit{if } n = 2k-1. \end{cases}$$

Thus the theorem of Saks and Werman compared to this theorem of Aigner tells us that if n is not a power of 2, then there exists a strictly better adaptive algorithm than the best non-adaptive one, but the gain is at most logarithmic (in case $n = 2^l - 1$ for some integer l).

Theorem (GKPP)

Suppose $c > 2$, $n > k > n/2$ and $n > 1$. Then a query graph G_Q solves the k-majority problem in the non-adaptive case if and only if G_Q is $(n - k + 1)$-connected.

Theorem (GKPP)

Suppose $c > 2$, $n > k > n/2$ *and* $n > 1$. *Then a query graph* G_Q *solves the* k*-majority problem in the non-adaptive case if and only if* G_Q *is* $(n-k+1)$*-connected.*

The proof uses weighted versions of the problem.

Theorem (GKPP)

Suppose $c > 2$, $n > k > n/2$ and $n > 1$. Then a query graph G_Q solves the k-majority problem in the non-adaptive case if and only if G_Q is $(n-k+1)$-connected.

The proof uses weighted versions of the problem.

Corollary (GKPP)

Let $M^*_c(n)$ denote the minimum number of queries needed to find a majority ball in a non-adaptive search if the number of colors is at most c and the number of balls is n, and suppose $n \geq c > 2$. Then $M^*_c(n) = \lceil \lceil n/2 \rceil n/2 \rceil$.

Let $\mu(n)$ denote the largest integer l such that 2^l divides n, i.e. the exponent of two in the prime factorization of n.

Theorem (Aigner)

For any pair of integers $n \geq k > n/2$, the inequality $M_2(n,k) \geq n - 1 - \mu(\binom{n-1}{k-1})$ holds.

Let $\mu(n)$ denote the largest integer l such that 2^l divides n, i.e. the exponent of two in the prime factorization of n.

Theorem (Aigner)

For any pair of integers $n \geq k > n/2$, the inequality $M_2(n,k) \geq n - 1 - \mu(\binom{n-1}{k-1})$ holds.

Note that this result is a generalization of the theorem of Saks and Werman as if n is even, then for $k = \lfloor n/2 + 1 \rfloor$ we have $\mu(\binom{n-1}{n/2}) = b(n) - 1$, where $b(n)$ is the number of 1's in the binary representation of n.

Proposition (GKPP)

Let $k > n/2$. Then

$$M_2(n,k) \geq n - 1 - \mu\left(\sum_{i=k}^{n} \binom{n}{i}\right).$$

Proposition (GKPP)

Let $k > n/2$. Then

$$M_2(n,k) \geq n-1-\mu\left(\sum_{i=k}^{n}\binom{n}{i}\right).$$

Proposition (GKPP)

Let $k > n/2$. Then $M_2(n,k) \geq n-2-\mu(\sum_{i=k}^{n}\binom{n-1}{i-1})$.

Proposition (GKPP)

Let $k > n/2$. Then

$$M_2(n,k) \geq n-1-\mu\left(\sum_{i=k}^{n}\binom{n}{i}\right).$$

Proposition (GKPP)

Let $k > n/2$. Then $M_2(n,k) \geq n-2-\mu(\sum_{i=k}^{n}\binom{n-1}{i-1})$.

A little calculation involving the μ-function shows that the above propositions yield the theorem of Aigner.

Plurality: earlier non-adaptive results by Chung, Graham, Mao and Yao.

Plurality: earlier non-adaptive results by Chung, Graham, Mao and Yao.

Let $P_c^*(n)$ denote the minimum number of queries that is needed in a non-adaptive search to determine if there exists a plurality color and if so, then to show one ball of that color.

Plurality: earlier non-adaptive results by Chung, Graham, Mao and Yao.

Let $P_c^*(n)$ denote the minimum number of queries that is needed in a non-adaptive search to determine if there exists a plurality color and if so, then to show one ball of that color.

Theorem (GKPP)

For any pair of integers n and c, the following holds:

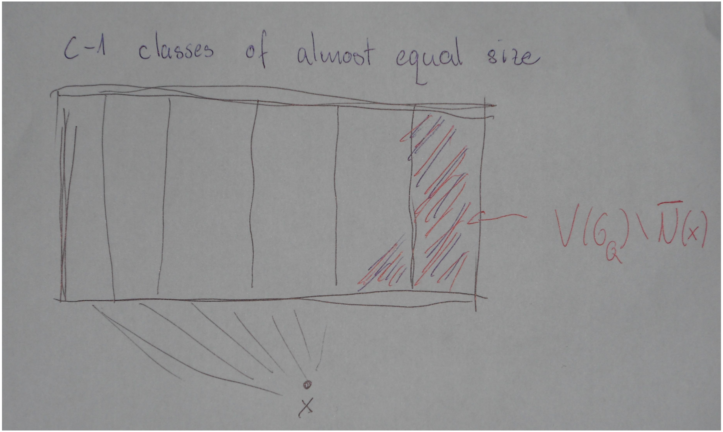
$$\left\lceil \frac{1}{2}\left(n-1-\frac{n-1}{c-1}\right)n \right\rceil \le P_c^*(n) \le \frac{c-2}{2(c-1)}n^2+n.$$

The lower bound follows easily from the following lemma

The lower bound follows easily from the following lemma

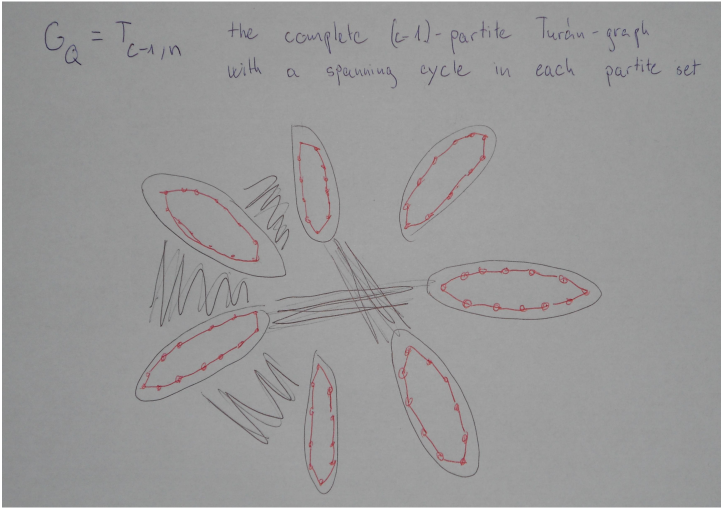
Lemma

If Q is a set of queries that solve the problem, then the minimum degree in G_Q is larger than $n-1-\lceil\frac{n-1}{c-1}\rceil$. Furthermore, if $n-1\equiv 1 \mod c-1$, then the minimum degree in G_Q is larger than $n-1-\lfloor\frac{n-1}{c-1}\rfloor$.


c-1 classes of almost equal size
x

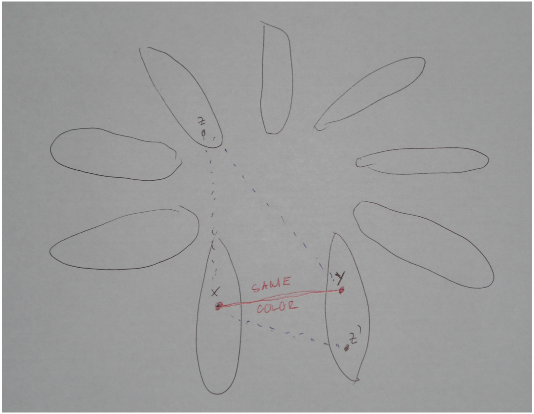
A construction for the upper bound:

A construction for the upper bound:



Proposition

If for some $x \in V_i, y \in V_j$ with $i \neq j$ the answer to the query (x, y) is SAME COLOR, then after receiving answers to all queries we are able to determine the whole color class of x and y.



Let k denote the number of color classes $C_1, ..., C_k$ that intersect at least two of the V_i's and let l denote the number of partite sets that are not contained in $K := \cup_{i=1}^{k} C_i$.

Let k denote the number of color classes $C_1, ..., C_k$ that intersect at least two of the V_i's and let l denote the number of partite sets that are not contained in $K := \cup_{i=1}^k C_i$.

Any of the remaining $c - k$ color classes is contained in one of the partite sets not covered by K, thus $l \leq c - k$.

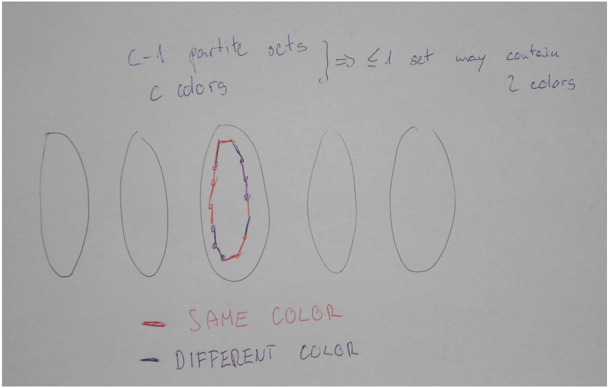
Let k denote the number of color classes $C_1, ..., C_k$ that intersect at least two of the V_i's and let l denote the number of partite sets that are not contained in $K := \cup_{i=1}^{k} C_i$.

Any of the remaining $c - k$ color classes is contained in one of the partite sets not covered by K, thus $l \leq c - k$.

Moreover, if $l = c - k$, then to cover all partite sets, all $c - k$ colors need to be used in different partite sets and thus they should be of the form $V_j \setminus K$. As we are able to determine all color classes, the proof is finished in this case.

From now on we assume $l \leq c - k - 1$. Let us suppose first that $k = 0$.

From now on we assume $l \leq c - k - 1$. Let us suppose first that $k = 0$.



Then as all color classes are included in one of the partite sets we have that all but at most one partite sets form one color class each and the last partite set is the union of at most two color classes which we can identify due to the additional spanning cycle.

Thus we may suppose that $l \leq c - k - 1$ and $k \geq 1$ hold. This means that the number of already covered partite sets is $c - 1 - l \geq k$, thus $|K| \geq k\lfloor \frac{n}{c-1} \rfloor$ and therefore at least one of $C_1, ..., C_k$ has size at least $\lfloor \frac{n}{c-1} \rfloor$.

Thus we may suppose that $l \leq c-k-1$ and $k \geq 1$ hold. This means that the number of already covered partite sets is $c-1-l \geq k$, thus $|K| \geq k\lfloor \frac{n}{c-1} \rfloor$ and therefore at least one of $C_1, ..., C_k$ has size at least $\lfloor \frac{n}{c-1} \rfloor$.

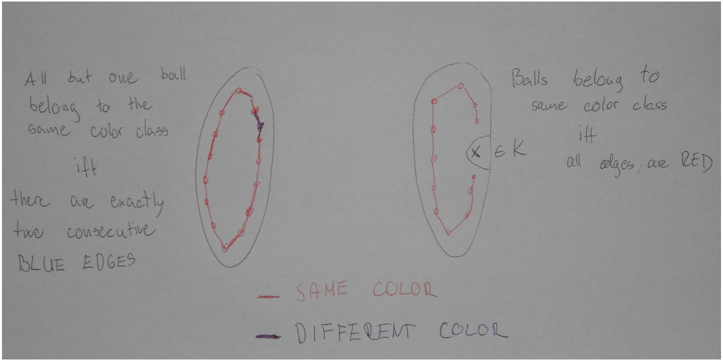
Therefore it is enough to prove that we are able to identify all other color classes that have size at least $\lfloor \frac{n}{c-1} \rfloor$.

Thus we may suppose that $l \leq c - k - 1$ and $k \geq 1$ hold. This means that the number of already covered partite sets is $c - 1 - l \geq k$, thus $|K| \geq k\lfloor \frac{n}{c-1} \rfloor$ and therefore at least one of $C_1, \ldots, C_k$ has size at least $\lfloor \frac{n}{c-1} \rfloor$.

Therefore it is enough to prove that we are able to identify all other color classes that have size at least $\lfloor \frac{n}{c-1} \rfloor$.

As all remaining color classes are contained in one of the V_i's, their size is at most $\lceil \frac{n}{c-1} \rceil$. Thus if a V_j contains at least two points from K, then it cannot contain a color class of size at least $\lfloor \frac{n}{c-1} \rfloor$.

On the other hand if a V_j contains at most one point of $\cup_{i=1}^k C_i$, then, due to the spanning cycle, we are able to tell whether it contains a color class of size $\lceil \frac{n}{c-1} \rceil$ or $\lfloor \frac{n}{c-1} \rfloor$.



Theorem (GKPP)

(i) $P_3^*(2k) = k(k+1)$,
(ii) $\frac{1}{2}(k+1)(2k+1) \leq P_3^*(2k+1) \leq \frac{1}{2}(k+1)(2k+1) + k - 1$.

Thank you for your attention!