Imbalances in k-Colorations

P. Erdös
Hungarian Academy of Sciences
Budapest, Hungary

J. Spencer
Department of Mathematics
University of California
Los Angeles, California

1. INTRODUCTION

The following problem is due to Paul Erdös [1]. Color the edges of a complete graph K_n on n vertices red and blue. What is the largest t such that we may always find a complete subgraph in which $|\#\text{ red edges} - \#\text{ blue edges}| \geq t$?

We need a more precise and more general formulation. For any set V, define

$$V^k = \{W : W \subseteq V, \ |W| = k\}. \tag{1}$$

Note that V^2 is the complete graph generated by V. V^k is called the complete k-graph generated by V. The elements of V^k are called k-edges. We color the k-edges. A coloring of a set A, $|A| = n$, is given by a map

$$g_k : A^k \to \{+1, -1\}. \tag{2}$$

The values +1, -1 may be thought of as Red and Blue. The subscript k indicates a function on k-edges and will be dropped when there is no confusion. The function g_k induces another function, also denoted by g_k, on the subsets of A given by

$$g_k(B) = \sum_{\substack{W \subseteq B \\ |W|=k}} g_k(W) \tag{3}$$

Set

$$H_k(n) = \min_{g_k} \max_{B \subseteq A} |g_k(B)| \tag{4}$$

Networks, 1: 379-385

where $A = \{1, \ldots, n\}$ and g_k ranges over all functions satisfying (2). Clearly $H_2(n)$ is the t required in the opening paragraph. Erdös [1] showed $\frac{n}{4} \leq H_2(n) \leq cn^{3/2}$. We prove

Theorem: For $k \geq 1$, and n sufficiently large

$$C_k\, n^{(k+1)/2} \leq H_k(n) \leq C_k'\, n^{(k+1)/2} \tag{5}$$

where the C_k, C_k' are positive absolute constants.

2. THE PROOF

The case $k = 1$ is trivial, $H_1(n) = \{\frac{n}{2}\}$. We sketch the proof of the upper bound.

Fix $B \subseteq A$, $|B| = b$. Letting g_k be random, $g_k(B)$ is the sum of $\binom{b}{k}$ values $g_k(W)$. The values $g_k(W)$ are +1 with probability $\frac{1}{2}$, -1 with probability $\frac{1}{2}$ and independent. Thus the distribution of $g_k(B)$ may be approximated by a normal curve of mean 0 and $\sigma = \binom{b}{k}^{1/2}/2 \leq n^{k/2}$. Thus

$$\text{Prob}\,[|g_k(B)| \geq cn^{(k+1)/2}] \leq e^{-c^2 n/2}. \tag{6}$$

As there are 2^n choices of B

$$\text{Prob}\,[\max_{B\subseteq A} |g_k(B)| \geq cn^{(k+1)/2}] \leq 2^n\, e^{-c^2 n/2}. \tag{7}$$

For $c = \sqrt{2 \log 2}$ the right hand side of (7) is less than unity so there does exist g_k such that $\max |g_k(B)| \leq cn^{(k+1)/2}$. A more careful proof, using that fact that most $|B| \sim \frac{n}{2}$ will show the upper bound of (5) for

$$C_k' \sim \frac{\sqrt{\log 2}}{2^{(k+1)/2} k!^{1/2}}\,. \tag{8}$$

Let us define

$$g_k(B_1^{a_1} B_2^{a_2} \ldots B_t^{a_t}) = \sum g_k(W) \tag{9}$$

where the sum is over all $W \subseteq A$, $|W| = n$, $|W \cap B_i| = a_i$ for $1 \le i \le t$. This shall only be defined when the B_i are disjoint and $\sum_{j=1}^{t} a_j = k$.

Now we give a quick proof of the lower bound (5) for $k = 2$. Applying the methods of [3] we find sets $B_1, B_2 \subseteq A$ with

$$g_2(B_1B_2) \ge cn^{3/2} \tag{10}$$

for some absolute constant c. But

$$g_2(B_1) + g_2(B_2) + g_2(B_1B_2) = g_2(B_1 \cup B_2) \tag{11}$$

so

$$|g_2(V)| \ge \frac{c}{3} n^{3/2} \text{ for } V = B_1, B_2, \text{ or } B_1 \cup B_2. \tag{12}$$

Now we prove our theorem for all $k \ge 2$.

Lemma 1: Fix $k \ge 2$. Then there exists $d_1, \ldots, d_k > 0$, t_o, such that for $t > t_o$ and A_j pairwise disjoint, $|A_j| = t$, $1 \le j \le k$ we have

$$|\{(B_1, \ldots, B_i) : B_j \subseteq A_j, |g_i(B_1 \ldots B_i)| \ge t^{i/2}\}| \ge d_i 2^{ti} \tag{13}$$

for all g_i, $1 \le i \le k$.

We shall first require

Lemma 2: Fix $c_1 > 0$. There exists $c_2 > 0$, t_o, such that $t \ge t_o$ implies that for any choice of real x_j, $1 \le j \le t$, satisfying $|x_j| \ge 1$ for $1 \le j \le c_1 t$, we have

$$\left|\sum_{j \in V} x_j\right| > \sqrt{t} \tag{14}$$

for at least $c_2 2^t$ choices of $V \subseteq \{1, \ldots, t\}$.

Proof: For $V \subseteq \{1, \ldots, t\}$ set $\varphi(V) = \sum_{j \in V} x_j$, $V_1 = V \cap \{j : 1 < j < c_1 t\}$, $V_2 = V - V_1$. Then $\varphi(V) = \varphi(V_1) + \varphi(V_2)$ so (14) does not hold if $\varphi(V_1) \in [\varphi(V_2) - \sqrt{t}, \varphi(V_2) + \sqrt{t}]$. By a theorem of Erdös [2] for V_2 fixed this holds for at most

$$\sum_{\left|r - \frac{c_1 t}{2}\right| \leq \sqrt{t}} \binom{[c_1 t]}{r} < (1 - c_2) 2^{[c_1 t]} \tag{15}$$

values of V_1, where c_2 is a positive constant dependent only on c_1. Summing over all V_2 yields Lemma 2. Q.E.D.

Proof of Lemma 1: We use induction on i. For i = 1 set $x_j = g_1(\{j\})$ and apply Lemma 2. Now assume Lemma 1 holds for i - 1. Any point $a \in A$ generates, with any g_i, a coloring $g_{i-1}^{(a)}$ on $A - \{a\}$. The coloring is given by

$$g_{i-1}^{(a)}(W) = g_i(W \cup \{a\}).$$

Set

$$V = \{((B_1, \ldots, B_{i-1}), a) : B_j \subseteq A_j, a \in B_i,$$

$$g_{i-1}^{(a)}(B_1 \ldots B_{i-1}) \geq t^{(i-1)/2}\}.$$

We double count

$$|V| = \sum_a |\{(B_1, \ldots, B_{i-1}) : ((B_1, \ldots, B_{i-1}), a) \in V\}| \tag{16}$$

$$= \sum_{B_1, \ldots, B_{i-1}} |\{a : ((B_1, \ldots, B_{i-1}), a) \in V\}|. \tag{17}$$

By induction the inner summation (16) is at least $d_{i-1} 2^{t(i-1)}$. Thus $|V| \geq t d_{i-1} 2^{t(i-1)}$. The sum (17) has $2^{t(i-1)}$ addends, each bounded by t. Thus for at least $(d_{i-1}/2) 2^{t(i-1)}$ choices of $(B_1, \ldots, B_{i-1})$ we have $|\{a : ((B_1, \ldots, B_{i-1}) a) \in V\}| \geq d_{i-1} t/2$. Fix such a $(B_1, \ldots, B_{i-1})$. Set

$$x_a = \frac{g_i(B_1 \ldots B_{i-1}\{a\})}{t^{(i-1)/2}}, \quad a \in B_i.$$

By assumption $|x_a| \geq 1$ for at least $d_{i-1} t/2$ of the a. By Lemma 2 there exists c_2 such that

$$|g(B_1 \ldots B_{i-1}B_i)| = |\sum_{a\in B_i} g(B_1 \ldots B_{i-1}\{a\})|$$
$$= t^{(i-1)/2} |\sum_{a\in B_i} x_a|$$
$$\geq t^{i/2}$$

for $c_2 2^t$ choices of B_i. As this is true for at least $(d_{i-1}/2)2^{t(i-1)}$ choices of $(B_1, \ldots, B_{i-1})$ we may show (15) for $d_i = d_{i-1}c_2/2$, completing the induction. Q.E.D.

Now let $g = g_k$ be any coloring on A, $|A| = n$. For $t = [\frac{n}{k}]$ find disjoint $A_1, \ldots, A_k \subseteq A$, $|A_i| = t$. From the proof of Lemma 1 we find, and fix $B_1, \ldots, B_{k-1}$ and $\delta > 0$ such that

$$|\{a : |g(B_1 \ldots B_{k-1} \cdot \{a\})| \geq t^{(k-1)/2}\}| \geq 2\delta t. \tag{18}$$

Either δt a's have $g(B_1 \ldots B_{k-1} \cdot \{a\}) \geq t^{(k-1)/2}$ or δt a's have $g(B_1 \ldots B_{k-1} \cdot \{a\}) \leq -t^{(k-1)/2}$. By symmetry (between g and -g) assume the former. Set $B_k = \{a : g(B_1 \ldots B_{k-1} \cdot \{a\}) > n^{(k-1)/2}\}$. Then

$$g(B_1 \ldots B_k) = \sum_{a\in B_k} g(B_1 \ldots B_{k-1} \cdot \{a\})$$
$$\geq \delta\, t^{(k+1)/2}$$
$$\geq \varepsilon\, n^{(k+1)/2} \tag{19}$$

where $\varepsilon \sim \delta/k^{(k+1)/2} > 0$, independent of n.

To prove our result we first need a result in polynomial approximations. If G is a polynomial in, say, s variables we set $|G|$ = the maximum absolute value of a coefficient of G and $||G|| = \max\{G(x_1, \ldots, x_s) : 0 \leq x_i \leq 1 \text{ for } 1 \leq i \leq s\}$.

Lemma 3: There exists $\varepsilon = \varepsilon(s) > 0$ such that if G is a polynomial in s variables with degree at most s then

$$||G|| \geq \varepsilon\, |G| \tag{20}$$

Proof: Set $T = \{G : |G| = 1\}$. With the $|\cdot|$ metric, T is compact, $||\cdot||$ is continuous, non zero, so there exists ε, $|G| = 1 \Rightarrow ||G|| \geq \varepsilon$. But any $G = |G|\, G_1$, $G_1 \in T$, so $||G|| = |G|\, ||G_1|| \geq \varepsilon\, |G|$.

It should be noted that by other methods explicit bounds on $\varepsilon(s)$ may be found.

Proof of Theorem: We need transfer the imbalance (19) of a product into the imbalance of a set. For $1 \leq i \leq k$ let W_i range over all subsets of B_i, $|W_i| = [x_i |B_i|]$, where $0 \leq x_i \leq 1$ will be determined later.

$$g(W_1 \cup \ldots \cup W_k) = \sum g(W_1^{a_1} \ldots W_k^{a_k}) \tag{21}$$

where the summation ranges over all nonnegative integers a_i, $\sum_{i=1}^{k} a_i = k$. Fix the a_i. Set

$$\nu(V_1, \ldots, V_k, W_1, \ldots, W_k) = \begin{cases} 1 & \text{if } V_i \subseteq W_i \text{ for } 1 \leq i \leq k \\ 0 & \text{otherwise.} \end{cases}$$

Then from (9) the expected value

$$E[g(W_1^{a_1} \ldots W_k^{a_k})] = E[\sum g(V_1 \cup \ldots \cup V_k)\nu(V_1, \ldots, V_k, W_1, \ldots, W_k)]$$

(the summation over all $V_i \subseteq B_i$, $|V_i| = a_i$)

$$= \sum g(V_1 \cup \ldots \cup V_k) E[\nu(V_1, \ldots, V_k, W_1, \ldots, W_k)]$$

$$= \sum g(V_1 \cup \ldots \cup V_k) \operatorname{Prob}[V_i \subseteq W_i,\ 1 \leq i \leq k \mid V_i \subseteq B_i,\ |V_i| = a_i]$$

$$= \prod_{i=1}^{k} x_i^{a_i} \sum g(V_1 \cup \ldots \cup V_k)$$

(an approximation valid as k is fixed and n sufficiently large)

$$= g(B_1^{a_1} \ldots B_k^{a_k}) \prod_{i=1}^{k} x_i^{a_i} \tag{22}$$

Setting

$$c_{a_1 \ldots a_k} = g(B_1^{a_1} \ldots B_k^{a_k})/n^{(k+1)/2}$$

we have, using (22) in (21)

$$E(g(W_1 \cup \ldots \cup W_n)) = n^{(k+1)/2} \sum c_{a_1 \ldots a_k} x_1^{a_1} \ldots x_k^{a_k}$$

where by (19), $|c_{1 \ldots 1}| \geq \varepsilon$. By Lemma 3 we find, and fix, $x_1, \ldots, x_k$ so that

$$E(g(W_1 \cup \ldots \cup W_n)) = \varepsilon_1 n^{(k+1)/2}$$

where $|\varepsilon_1| \geq |c_{1 \ldots 1}| \varepsilon(k) \geq \varepsilon\ \varepsilon(k)$ which depends only on k. By the definition of expected value we find, and fix, $W_1' \ldots, W_n', |W_i'| = [x_i |B_i|]$, such that

$$|g(W_1' \cup \ldots \cup W_n')| \geq |E(g(W_1 \cup \ldots \cup W_n))| \geq \varepsilon_1 n^{(k+1)/2}$$

proving our theorem.

REFERENCES

1. Erdös, P., "Ramsey és Van der Waerden tételével Kapcsolatos Kombinatorikai Kérdesekröl," *Mat Lapok* 14, 1963, pp. 29-37, (in Hungarian).

2. Erdös, P., "On a Lemma of Littlewood and Offord," *Bulletin Amer. Math. Soc.* 51, 1945, pp. 898-902.

3. Spencer, J. H., "Optimal Ranking of Tournaments," *Networks*, Vol. 1, No. 2.

Any views expressed in this paper are those of the authors. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors.

Paper received May 15, 1970.