Choice numbers of graphs; a probabilistic approach

Noga Alon *

Department of Mathematics

Raymond and Beverly Sackler Faculty of Exact Sciences

Tel Aviv University, Tel Aviv, Israel

and Bellcore, Morristown, NJ 07960, USA

Abstract

The *choice number* of a graph G is the minimum integer k such that for every assignment of a set $S(v)$ of k colors to every vertex v of G, there is a proper coloring of G that assigns to each vertex v a color from $S(v)$. Applying probabilistic methods it is shown that there are two positive constants c_1 and c_2 such that for all $m \geq 2$ and $r \geq 2$ the choice number of the complete r-partite graph with m vertices in each vertex class is between $c_1 r \log m$ and $c_2 r \log m$. This supplies the solutions of two problems of Erdős, Rubin and Taylor, as it implies that the choice number of almost all the graphs on n vertices is $o(n)$ and that there is an n vertex graph G such that the sum of the choice number of G with that of its complement is at most $O(n^{1/2}(\log n)^{1/2})$.

*Research supported in part by a United States Israel BSF Grant

1 Introduction

All graphs considered here are finite, undirected and simple (i.e., have no loops and no parallel edges). A graph $G = (V, E)$ is *k-choosable* if for every family of sets $\{S(v) : v \in V\}$, where $|S(v)| = k$ for all $v \in V$, there is a proper vertex-coloring of G assigning to each vertex $v \in V$ a color in $S(v)$. The *choice number* of G, denoted by $ch(G)$, is the minimum integer k so that G is k-choosable. Obviously, this number is at least the chromatic number $\chi(G)$ of G.

The study of choice numbers of graphs was initiated by Vizing in [8] and by Erdős, Rubin and Taylor in [7], and has been studied in various papers (see, e.g., [6], [3] and their references). In the present short paper we apply probabilistic arguments and solve two of the problems raised in 1979 in the original paper [7]. Our main result supplies a sharp estimate to the choice numbers of complete multipartite graphs with equal color classes. For two positive integers m and r let K_{m*r} denote the complete r-partite graph with m vertices in each vertex class. For $r = 1$, K_{m*r} has no edges and hence, obviously, $ch(K_{m*1}) = 1$ for all m. Another trivial observation is the fact that $ch(K_{1*r}) = r$ for all r. In [7] it is shown that $ch(K_{2*r}) = r$ for all r. The following theorem determines, up to a constant factor, the choice number of K_{m*r} for all the remaining cases.

Theorem 1.1 *There exist two positive constants c_1 and c_2 such that for every $m \geq 2$ and for every $r \geq 2$*

$$c_1 r \log m \leq ch(K_{m*r}) \leq c_2 r \log m.$$

Two applications of this theorem are the following.

Corollary 1.2 *There exists a positive constant b so that for every n there is an n-vertex graph G so that*

$$ch(G) + ch(G^c) \leq b n^{1/2} (\log n)^{1/2},$$

where G^c is the complement of G.

The second corollary deals with the choice numbers of random graphs. It is convenient to consider the common model $G_{n,1/2}$ (see, e.g., [5]), in which the graph is obtained by taking each pair of the n labelled vertices $1, 2, \ldots, n$ to be an edge, randomly and independently, with probability $1/2$. (It is not too difficult to obtain similar results for other models of random graphs as well).

Corollary 1.3 *There exists a positive constant c so that for the random graph $G_{n,1/2}$ on n vertices, the probability that $ch(G_{n,1/2}) \leq cn\frac{\log\log n}{\log n}$ tends to 1 as n tends to infinity.*

The last corollary shows that for almost all the graphs G on n vertices, $ch(G) = o(n)$ as n tends to infinity. This solves a problem raised in [7]. Corollary 1.2 settles another problem raised in [7], where the authors ask if there exists a constant $\epsilon > 0$ so that for every n-vertex graph G, $ch(G) + ch(G^c) > n^{1/2+\epsilon}$.

The rest of the paper is organized as follows. In sections 2 and 3 we prove the main result-Theorem 1.1. Section 2 contains the proof of the required upper bound for $ch(K_{m*r})$, and section 3 includes the lower bound. The proof relies heavily on probabilistic arguments and employs a splitting technique similar to the one used in [1], together with some additional ideas. In section 4 we derive the two corollaries mentioned above.

2 The upper bound

In this section we prove the following proposition, which establishes the upper bound for $ch(K_{m*r})$, asserted in Theorem 1.1. Here and in the rest of the paper we omit all the floor and ceiling signs whenever these are not crucial, to simplify the notation. All the logarithms are in the natural base e, unless otherwise specified.

Proposition 2.1 *There exists a positive constant c so that for all positive integers $m \geq 2$ and r, $ch(K_{m*r}) \leq cr \log m$.*

Proof Since rm is a trivial upper bound for $ch(K_{m*r})$ and since for, say, $c \geq 4$, $rm \leq cr \log m$ for all m satisfying $m \leq c$ we may asume that $m > c$ (where $c \geq 4$ will be chosen later). Let $V_1, V_2, \ldots V_r$ be the vertex classes of $K = K_{m*r}$, where $|V_i| = m$ for all i, and let $V = V_1 \cup \ldots \cup V_r$ be the set of all vertices of K. For each $v \in V$, let $S(v)$ be a set of at least $cr \log m$ distinct colors. We must show that there is a proper coloring of K assigning to each vertex v a color from $S(v)$. Since $ch(K_{m*r})$ is a non-decreasing function of r we may and will assume that r is a power of 2.

We consider two possible cases.

Case 1: $r \leq m$.

Let $S = \cup_{v \in V} S(v)$ be the set of all colors. Put $R = \{1, 2, \ldots, r\}$ and let $f : S \mapsto R$ be a random function, obtained by choosing, for each color $c \in S$, randomly and independently, the value of $f(c)$

according to a uniform distribution on R. The colors c for which $f(c) = i$ will be the ones to be used for coloring the vertices in V_i. To complete the proof for this case it thus suffices to show that with positive probability for every i, $1 \leq i \leq r$, and for every vertex $v \in V_i$ there is at least one color $c \in S(v)$ so that $f(c) = i$.

Fix an i and a vertex $v \in V_i$. The probability that there is no color $c \in S(v)$ so that $f(c) = i$ is clearly

$$\begin{aligned} (1-\frac{1}{r})^{|S(v)|} &\leq (1-\frac{1}{r})^{cr\log m} \\ &\leq e^{-c\log m} \leq \frac{1}{m^c} < \frac{1}{rm}, \end{aligned}$$

where the last inequality follows from the fact that $r \leq m$ and $c \geq 4 > 2$. There are rm possible choices of i, $1 \leq i \leq r$ and $v \in V_i$, and hence, the probability that for some i and some $v \in V_i$ there is no $c \in S(v)$ so that $f(c) = i$ is smaller than 1, completing the proof in this case.

Case 2: $r > m$.

This case is more difficult, and requires a splitting trick similar to the one used in [1]. As before, define $R = \{1, 2, \ldots, r\}$ and let $S = \cup_{v \in V} S(v)$ be the set of all colors. Put $R_1 = \{1, 2, \ldots, r/2\}$ and $R_2 = \{r/2+1, \ldots, r\}$. Let $f : S \mapsto \{1, 2\}$ be a random function obtained by choosing, for each $c \in S$ randomly and independently, $f(c) \in \{1, 2\}$ according to a uniform distribution. The colors c for which $f(c) = 1$ will be used for coloring the vertices in $\cup_{i \in R_1} V_i$, whereas the colors c for which $f(c) = 2$ will be used for coloring the vertices in $\cup_{i \in R_2} V_i$.

For every vertex $v \in V$, put $S^0(v) = S(v)$, and define $S^1(v) = S^0(v) \cap f^{-1}(1)$ if v belongs to $\cup_{i \in R_1} V_i$, and $S^1(v) = S^0(v) \cap f^{-1}(2)$ if v belongs to $\cup_{i \in R_2} V_i$. Observe that in this manner the problem of finding a proper coloring of K in which the color of each vertex v is in $S(v) = S^0(v)$ has been decomposed into two independent problems. These are the problems of finding proper colorings of the two complete $r/2$-partite graphs on the vertex classes $\cup_{i \in R_1} V_i$ and $\cup_{i \in R_2} V_i$, by assigning to each vertex v a color from $S^1(v)$. Let $s_0 = cr\log m$ be the number of colors in each original list of colors assigned to a vertex. We claim that for all sufficiently large c, with high probability,

$$|S^1(v)| \geq \frac{1}{2}s_0 - \frac{1}{2}s_0^{2/3} \tag{1}$$

for all $v \in V$. This is because for every fixed vertex v, $|S^1(v)|$ is a Binomial random variable, and by the standard known tail estimates for such variables (cf., e.g., [2]), for every fixed v

$$Pr(|S^1(v)| < \frac{1}{2}s_0 - \frac{1}{2}s_0^{2/3}) \leq e^{-\frac{1}{2}c^{1/3}r^{1/3}(\log m)^{1/3}}.$$

The total number of vertices is $rm < r^2$. Since $r > m > c$ and c can be chosen to be a sufficiently large constant (independent of r and m), one can easily check that for all $r > m > c$:

$$r^2 \cdot e^{-\frac{1}{2}c^{1/3}r^{1/3}(\log m)^{1/3}} << 1.$$

Therefore, with high probability (1) holds for all $v \in V$, as claimed. Let s_1 denote the minimum cardinality of a set $S^1(v)$, for $v \in V$. As shown above we can make sure that

$$s_1 \geq \frac{1}{2}s_0 - \frac{1}{2}s_0^{2/3}.$$

We have thus reduced the problem of showing that the choice number of K_{m*r} is at most s_0 to that of showing that the choice number of $K_{m*(r/2)}$ is at most s_1.

Repeating the above decomposition technique (which we can repeat as long as $r/2^i > m$) we obtain, after j iterations, a sequence s_i, where $s_0 = cr\log m$ and

$$s_{i+1} \geq s_i/2 - s_i^{2/3}/2 \quad for \quad 1 \leq i < j. \tag{2}$$

In order to show that the choice number of $K = K_{m*r}$ is at most s_0, it suffices to show that for some i the choice number of $K_{m*(r/2^i)}$ is at most s_i.

Let the number of iterations j be chosen so that j is the minimum integer satisfying $r/2^j \leq m$. Clearly, in this case, $r/2^j > m/2 \geq c/2$. We claim that

$$s_j \geq \frac{s_0}{2^{j+1}}$$

provided c is sufficiently large. To see this, define $z_i = s_i^{1/3}$ and observe that by equation (2)

$$z_{i+1}^3 \geq \frac{z_i^3 - z_i^2}{2} \geq \frac{(z_i - 1)^3}{2},$$

implying that

$$z_{i+1} \geq \frac{z_i - 1}{2^{1/3}}, \qquad (1 \leq i < j).$$

Define $t_i = z_i + x$, where x is chosen so that $x = \frac{1+x}{2^{1/3}}$. Then $0 < x = \frac{1}{2^{1/3}-1} < 4$ and

$$t_{i+1} - x \geq \frac{t_i - x - 1}{2^{1/3}},$$

i.e.,

$$t_{i+1} \geq \frac{t_i}{2^{1/3}}.$$

Therefore,

$$t_j \geq \frac{t_0}{2^{j/3}} = \frac{z_0 + x}{2^{j/3}} \geq \frac{z_0}{2^{j/3}} = (\frac{s_0}{2^j})^{1/3}.$$

Hence

$$s_j = z_j^3 = (t_j - x)^3 \geq (t_j - 4)^3 \geq ((\frac{s_0}{2^j})^{1/3} - 4)^3 = \frac{s_0}{2^j} - O((\frac{s_0}{2^j})^{2/3}). \tag{3}$$

However, since $r/2^j \geq m/2 \geq c/2$ and $c \geq 4$ it follows that

$$\frac{s_0}{2^j} = \frac{cr \log m}{2^j} \geq \frac{c^2 \log m}{2} > c$$

and thus if c is sufficiently large then the right hand side of (3) is at least $s_0/2^{j+1}$, as claimed.

To complete the proof of the proposition observe, now, that it suffices to show that the choice number of $K_{m*(r/2^j)}$ is at most s_j. However, $r/2^j \leq m$ and

$$s_j \geq s_0/2^{j+1} \geq \frac{c}{2}\frac{r}{2^j} \log m.$$

For a sufficiently large c the result thus follows from Case 1. This completes the proof. $\Box$

3 The lower bound

In this section we complete the proof of Theorem 1.1 by proving the following result.

Proposition 3.1 *There exists a positive constant d so that for all integers m and $r \geq 2$, $ch(K_{m*r}) > dr \log m$.*

We note that the above result for the special case $r = 2$ is derived in [7] from the known estimates on the minimum possible number of edges in m-uniform hypergraphs which do not have property B. Our basic approach here is similar, but certain additional ideas are needed.

The main part of the proof is the following lemma.

Lemma 3.2 *There exists a positive constant c so that for all integers m and $r \geq 2$ there is a set S of cardinality $|S| = cr \log m$ and a family $\mathcal{F}$ of m subsets of S, each of size at least $\frac{1}{20}|S|$, so that there is no $X \subset S$ of size $|X| \leq c \log m$ that intersects each member of $\mathcal{F}$. Moreover, in case $m \geq r$ there exists such a family $\mathcal{F}$ in which the size of each member $f \in \mathcal{F}$ is at least $\frac{1}{10}|S|$.*

Proof Suppose, first, that $m \geq r$. Let S be a set of size $cr \log m$, where $c \leq 1$ will be chosen later, and let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a random family of m subsets of S obtained by choosing each F_i randomly and independently among the subsets of cardinality $\frac{1}{10}|S|$ of S, according to a uniform distribution.

Let X be a fixed subset of cardinality at most $c \log m$. For each i, $1 \leq i \leq m$, the probability that X does not intersect F_i is

$$\frac{\binom{|S \setminus X|}{\frac{1}{10}cr\log m}}{\binom{|S|}{\frac{1}{10}cr\log m}} \geq \frac{\binom{cr\log m(1-\frac{1}{r})}{\frac{1}{10}cr\log m}}{\binom{cr\log m}{\frac{1}{10}cr\log m}}$$

$$\geq (\frac{0.9cr\log m - c\log m}{0.9cr\log m})^{0.1cr\log m} \geq (1 - \frac{10}{9r})^{0.1cr\log m}$$

$$\geq ((1 - \frac{10}{18})^2)^{0.1c\log m} \geq \frac{1}{m^{1/2}},$$

where here we used the fact that $c \leq 1$.

Since the m members of $\mathcal{F}$ have been chosen independently, the probability that X intersects all of them is at most

$$(1 - \frac{1}{m^{1/2}})^m \leq e^{-m^{1/2}}.$$

The number of possible subsets X of cardinality at most $c \log m$ of S is at most

$$(cr\log m + 1)^{c\log m} \leq (m^2+1)^{c\log m} \leq e^{3c(\log m)^2}.$$

Therefore, the probability that there exists a subset X of size at most $c \log m$ of S that intersects each member of $\mathcal{F}$ does not exceed

$$e^{3c(\log m)^2} e^{-m^{1/2}}$$

which is smaller than 1 provided c is sufficiently small. This completes the proof for the case $m \geq r$.

In case $r > m$ we just modify the construction given above for $r = m$ by replacing each element of the ground set by a set of size roughly r/m. Here are the details. By the validity of the result for $r = m$, there is a set T of size $t = cm \log m$ and a family $\mathcal{G} = \{G_1, \ldots, G_m\}$ of m subsets of T of size

$\frac{1}{10}|T|$ each, so that there is no subset of cardinality at most $c \log m$ of T that intersects each member of $\mathcal{G}$. Let $S_1, \ldots, S_t$ be pairwise disjoint sets, where $\sum_{i=1}^{t} |S_i| = cr \log m$ and $\lfloor r/m \rfloor \leq |S_i| \leq \lceil r/m \rceil$ for all $1 \leq i \leq t$. Define $S = \cup_{i=1}^{t} S_i$, and let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be the family of m subsets of S defined by $F_i = \cup_{j \in G_i} S_j$ for $1 \leq i \leq m$. Observe that $|F_i| \geq \frac{1}{10} t \lfloor r/m \rfloor \geq \frac{1}{20} cr \log m$ for all i. Moreover, there is no subset X of size at most $c \log m$ of S that intersects each member of $\mathcal{F}$, since if there is such a set the subset Y of T defined by $Y = \{j : X \cap S_j \neq \emptyset\}$ is of size at most $c \log m$ and it intersects each member of $\mathcal{G}$, contradicting the choice of $\mathcal{G}$. This completes the proof of the lemma. □

Proof of Proposition 3.1 We prove the assertion with $d = \frac{1}{20}c$, where c is the constant in Lemma 3.2. Let S be a set of size $cr \log m$; this will be our set of colors. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ be a family of subsets of S satisfying the assertion of Lemma 3.2. Denote the set of vertices of $K = K_{m*r}$ by $V = V_1 \cup V_2 \cup \ldots \cup V_r$, where $V_i = \{v_i^1, v_i^2, \ldots, v_i^m\}$ is the i^{th} vertex class. Define, now, $S(v_i^j) = F_j$. We claim that there is no proper vertex-coloring of K which assigns to each vertex v of K a color from $S(v)$. Since $|F_j| \geq \frac{1}{20}|S| = dr \log m$, this will show that $ch(K_{m*r}) > dr \log m$, and hence complete the proof.

Suppose the claim is false and let $f : V \mapsto S$ be a proper coloring of K in which $f(v) \in S(v)$ for all $v \in V$. For each i, $1 \leq i \leq r$, define $X_i = \{c \in S : f(v) = c \quad for \quad some \quad v \in V_i\}$. Since f is a proper vertex coloring and K is a complete r-partite graph, the sets X_i are pairswise disjoint. It follows that there exists an index i so that $|X_i| \leq |S|/r = c \log m$. By the choice of the family $\mathcal{F}$ this implies that there is a member F_j of $\mathcal{F}$ that does not intersect X_i. But $S(v_i^j) = F_j$ and hence the color $f(v_i^j)$ of v_i^j must be a member of F_j that belongs to X_i. This contradiction shows that the assumption that there is a proper coloring as above is false, and completes the proof of the proposition and hence that of Theorem 1.1 as well. □

4 Applications

Corollaries 1.2 and 1.3 are simple consequences of Theorem 1.1.

Proof of Corollary 1.2 Define $m = \sqrt{n}\sqrt{\log n}$ and $r = n/m = \frac{\sqrt{n}}{\sqrt{\log n}}$ and let G be the graph K_{m*r}. The complement G^c of G is a disjoint union of r cliques of size m each, and thus $ch(G^c) =$

$m = O(\sqrt{n}\sqrt{\log n})$. By Theorem 1.1, $ch(G) = O(r \log m) = O(\sqrt{n}\sqrt{\log n})$. Therefore

$$ch(G) + ch(G^c) = O(\sqrt{n}\sqrt{\log n}),$$

as needed. $\Box$

Proof of Corollary 1.3 As proved by Bollobás in [4], almost surely (i.e., with probability that tends to 1 as n tends to infinity), the random graph $G = G_{n,1/2}$ has chromatic number

$$(1 + o(1))n/2 \log_2 n.$$

It is also known (and easy, cf., e.g., [5], [2]) that almost surely G contains no independent set of size greater than $2 \log_2 n$. Therefore, for , say $r = n/\log_2 n$ and $m = 2\log_2 n$, almost surely G has a proper coloring with r colors in which no color appears more than m times. It follows that G is almost surely a subgraph of K_{m*r} and hence, by Theorem 1.1, almost surely

$$ch(G) \leq ch(K_{m*r}) = O(r \log m) = O(n\frac{\log\log n}{\log n}),$$

completing then proof. $\Box$

Note that the tight result of [4] for the chromatic number of the random graph is not needed here and the well known $O(n/\log n)$ estimate (cf., e.g., [5], [2]) suffices for the last proof. It will be interesting to decide if the right order of magnitude of $ch(G_{n,1/2})$ is closer, almost surely, to its chromatic number $\Theta(n/\log n)$ (which is an obvious lower bound for it) or to the upper bound given in the last corollary.

Acknowledgement I would like to thank H. Taylor for helpful discussions.

References

[1] N. Alon, *The strong chromatic number of a graph*, Random Structures and Algorithms 3 (1992), 1-7.

[2] N. Alon and J. H. Spencer, **The Probabilistic Method**, Wiley, 1991.

[3] N. Alon and M. Tarsi, *Colorings and orientations of graphs*, Combinatorica, in press.

[4] B. Bollobás, *The chromatic number of random graphs*, Combinatorica 8 (1988), 49-55.

[5] B. Bollobás, **Random Graphs**, Academic Press, 1985.

[6] A. Chetwynd and R. Häggkvist, *A note on list colorings*, J. Graph Theory 13 (1989), 87-95.

[7] P. Erdős, A. L. Rubin and H. Taylor, *Choosability in graphs*, Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium XXVI, 1979, 125-157.

[8] V.G. Vizing, *Coloring the vertices of a graph in prescribed colors* (in Russian), Diskret. Analiz. No. 29, Metody Diskret. Anal. v. Teorii Kodov i Shem 101 (1976), 3-10.