



University of Dundee

The harmonious chromatic number of almost all trees

Edwards, Keith

Published in: Combinatorics, Probability and Computing

DOI: 10.1017/S0963548300001462

Publication date: 1995

Document Version Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA): Edwards, K. (1995). The harmonious chromatic number of almost all trees. Combinatorics, Probability and Computing, 4(1), 31-46. 10.1017/S0963548300001462

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
You may not further distribute the material or use it for any profit-making activity or commercial gain.
You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Harmonious Chromatic Number of Almost All Trees

KEITH EDWARDS

Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, U.K. Received 3 September 1993; revised 6 July 1994

A harmonious colouring of a simple graph G is a proper vertex colouring such that each pair of colours appears together on at most one edge. The harmonious chromatic number h(G) is the least number of colours in such a colouring.

For any positive integer m, let Q(m) be the least positive integer k such that $\binom{k}{2} \ge m$. We show that for almost all unlabelled, unrooted trees T, h(T) = Q(m), where m is the number of edges of T.

1. Introduction

A harmonious colouring of a simple graph G is a proper vertex colouring such that each pair of colours appears together on at most one edge. Formally, a harmonious colouring is a function c from a colour set C to the set V(G) of vertices of G such that for any edge e of G, with endpoints x, y say, $c(x) \neq c(y)$, and for any pair of distinct edges e, e', with endpoints x, y and x', y', respectively, then $\{c(x), c(y)\} \neq \{c(x'), c(y')\}$. The harmonious chromatic number h(G) is the least number of colours in such a colouring. A survey on harmonious colourings is given by B. Wilson [14]. In this paper we consider the harmonious chromatic number of trees. First we give a definition.

Definition. Let m be a positive integer. Then we define Q(m) to be the least positive integer k such that $\binom{k}{2} \ge m$. It is easily calculated that

$$Q(m) = \left\lceil \frac{1}{2}(1 + \sqrt{8m+1}) \right\rceil.$$

If T is any tree with m edges, it is immediate that the harmonious chromatic number of T is at least Q(m). (Indeed the same is true for any graph.) It is shown in [2] that the general problem of determining whether h(T) = Q(m) for a tree T is NP-complete, so we cannot expect to find a nice characterization of those trees T for which h(T) = Q(m). Nevertheless, we will show that almost all trees have this property, in the normal sense that the proportion of trees on n vertices with the property tends to 1 as n tends to infinity. We refer here to trees in the usual graph theoretic sense, that is unlabelled and unrooted trees.

In outline, we proceed as follows. We take d to be a fixed positive integer, and let T be a tree with m edges and maximum degree at most d. It was shown in [1] that $h(T) = Q(m) + o(m^{1/2})$. Here we show that for any positive ε , there is an integer $N(d, \varepsilon)$ such that if T is any tree with n vertices (and so m = n - 1 edges), at least εn of which are leaves, and $n \ge N(d, \varepsilon)$, then h(T) = Q(m). We then use a slight modification of this result to show that h(T) = Q(m) for almost all trees, a result that was essentially conjectured by Frank, Harary and Plantholt [3].

We will need the following definition.

Definition. A partial harmonious colouring of a graph G is a colouring of a subset of the vertices of G that is harmonious on the graph induced by the coloured vertices and such that no uncoloured vertex has two or more coloured neighbours with the same colour.

2. Trees with many leaves

We consider, for some fixed d and ε , trees that have maximum degree at most d, and whose leaves form a proportion at least ε of the vertex set. We will show that such a tree satisfies h(T) = Q(m) provided it is sufficiently large.

The idea of the proof is as follows: start with Q(m) colours, giving $\binom{Q(m)}{2}$ colour pairs each of which can occur at most once on an edge of the tree. We note that if instead of a tree T we had a forest F consisting of r disjoint copies of a forest with 2r + 1 edges, we could easily colour F in a highly symmetrical way. Now by removing a small proportion (*i.e.* much smaller than ε) of the vertices of T, we can split it up to form a forest consisting of a number of copies of a smaller forest. We then remove the leaves. We are then left again with a forest consisting of a number of copies of a smaller forest. We colour this in a highly symmetrical way using less than Q(m) colours. We then have to colour the vertices originally removed; this process to some extent destroys the symmetry. However, sufficient structure remains, both in the colouring and in the set of unused colour pairs, for it to be possible to colour the leaves using only the remaining colour pairs.

Theorem 2.1. Let d be a positive integer and let $\varepsilon > 0$. Then there exists an integer $N = N(d, \varepsilon)$ such that if T is any tree with $n \ge N$ vertices (and m = n - 1 edges), maximum degree of T at most d and at least ε n leaves, then T satisfies

$$h(T) = \left\lceil \frac{1}{2}(1 + \sqrt{8n - 7}) \right\rceil = Q(m).$$

Proof. Suppose that T is a tree with n vertices and at least εn leaves, and that the maximum degree if T is at most d. Let C be the least integer such that $\binom{C}{2} \ge n-1$, the number of edges of T, so that $C = \left\lceil \frac{1}{2}(1 + \sqrt{8n-7}) \right\rceil$. Then

$$\frac{C-3}{2} \le \frac{n}{C} \le \frac{C+1}{2}.$$

Now let η be a constant depending on ε , to be chosen later. By removing a set S of at most ηn vertices of T we can split T up into components each of size at most $c(\eta)$, where $c(\eta)$ is a constant depending only on η . (See for example Corollary 3.7 of [1].) Now consider T - S. This consists of a forest each component of which is a tree of size at most $c(\eta)$. Let $C' = C(1 - \alpha)$, where α is chosen so that C' is odd and

$$15d^2(4d\eta)^{\frac{1}{2}} \ge \alpha \ge 14d^2(4d\eta)^{\frac{1}{2}}$$

We will classify the components of T-S as follows. Call a leaf of one of these components special if (i) it is a leaf of the original tree T, and (ii) it is at distance at least 4 in T from any element of S. The components are classified first according to isomorphism type; then each isomorphism type is further classified according to which of its leaves are special. It is clear that in this way each of the components of T-S may be assigned to one of a finite number, say P, of classes. Now we ensure that the number of components of T-S in each class is a multiple of C', by discarding at most C'-1 of each class if necessary and adding the vertices of these to S. Note that any special leaf in the remaining components remains special when this is done. The number of vertices added to S is at most $c(\eta)PC'$. The size of S is now at most $\eta n + c(\eta)PC'$, which is less than $2\eta n$ provided n is large enough. Let $\phi = |S|/n$. Then $\phi \leq 2\eta$. Now T-S consists of C' copies of a forest. Furthermore, if a vertex v in one copy is a special leaf, the corresponding vertex in each of the other copies is also a special leaf. Let $\psi = 2\phi d$, then the total degree of the vertices in S is at most ψn . (Note that we could take $\psi = \phi d$ here, but for a technical reason it is convenient to take $\psi = 2\phi d$.) Note that $\psi \leq 4d\eta$ and hence $\alpha \geq 14d^2\psi^{1/2}$.

The number of vertices of T that are at distance at most 3 in T from an element of S is at most $n(\phi + \psi(1 + d + d^2))$, so T - S contains at least $(\varepsilon - (\phi + \psi(1 + d + d^2)))n$ special leaves. We will choose η small enough so that this is at least $(1/2)\varepsilon n$. Let the number of special leaves be βn , so $\beta \ge (1/2)\varepsilon$. We now delete these special leaves, to form a new forest F that consists of C' copies of a forest F'. We refer to these leaves below as the deleted special leaves. Now let the number of edges of F be kC'. Then k satisfies

$$(1-\beta)n \ge kC' \ge (1-\psi-\beta)n-1,$$

so

$$(1-\beta)\frac{n}{C'} \ge k \ge (1-\psi-\beta)\frac{n}{C'} - 1.$$

Note that the number of edges in F' is k. Now since

$$\frac{C-3}{2(1-\alpha)} \le \frac{n}{C'} \le \frac{C+1}{2(1-\alpha)},$$

we have

$$\frac{C'}{2(1-\alpha)^2} - \frac{3}{2(1-\alpha)} \le \frac{n}{C'} \le \frac{C'}{2(1-\alpha)^2} + \frac{1}{2(1-\alpha)}.$$

Hence

$$(1-\beta)\left(\frac{C'}{2(1-\alpha)^2}+\frac{1}{2(1-\alpha)}\right) \ge k \ge (1-\psi-\beta)\left(\frac{C'}{2(1-\alpha)^2}-\frac{3}{2(1-\alpha)}\right)-1.$$

Define t by setting

$$k = \frac{1}{2}(C'-1) - tC' = \frac{1}{2}(1-2t)C' - \frac{1}{2}$$

Substituting k = (1/2)(1-2t)C' - 1/2 above and rearranging, we obtain

$$1 - \frac{(1 - \psi - \beta)}{(1 - \alpha)^2} + O(\frac{1}{C'}) \ge 2t \ge 1 - \frac{(1 - \beta)}{(1 - \alpha)^2} + O(\frac{1}{C'})$$

from which

$$\beta + \psi - 2\alpha(1-\beta-\psi) + O(\frac{1}{C'}) + O(\alpha^2) \ge 2t \ge \beta + 2\alpha(1-\beta) + O(\frac{1}{C'}) + O(\alpha^2).$$

Now since $\psi = O(\alpha^2)$ and $1 - \beta > 0$, it follows that for any $\xi > 0$, we can ensure that $\beta > 2t > \beta - \xi$ provided that η (and hence α , ϕ and ψ) is small enough, and n (and hence C') is large enough.

Now colour F' as follows: label the edges of F' with the integers $tC'+1, \ldots, (1/2)(C'-1)$, in some order. Then colour the vertices of F' with colours from $0, \ldots, C'-1$ so that for any edge, the difference $|c_1 - c_2|$ between the colours c_1, c_2 of the endpoints equals the label. (This is easily done.) Then, to colour F, colour the first copy of F' as described, and for $i \ge 2$, colour copy i by adding (modulo C') i-1 to the colour of the corresponding vertex in copy 1. Note that the C' copies of a given vertex of F' receive the colours $0, \ldots, C'-1$ in F, and that the colour pairs i, j that are used in this colouring are precisely those for which the cyclic difference between i and j is in the set $\{tC'+1,\ldots,(1/2)(C'-1)\}$. Hence the pairs still unused are those with difference in the set $\{1,\ldots,tC'\}$. Furthermore, the number of deleted special leaves adjacent to a vertex of colour i is the same for all i, and is $\beta n/C' = \gamma C'$ say. Then, provided that α is sufficiently small and n is large enough, we can ensure that $2\gamma > \beta > 2\gamma - \xi$, and hence $\gamma > t > \gamma - \xi$.

Define the graph U^* on vertices $\{0, ..., C' - 1\}$ by joining a pair of vertices (colours) if this pair of colours has not been used in colouring F. We will refer to U^* again later.

Now we colour the set S. (This procedure is similar to that described in Theorem 3.5 of [1].) This may involve changing the colours of some vertices from T - S. Set

$$N(S) = \{v | (v, x) \in E \text{ for some } x \in S\} - S$$

$$N^{2}(S) = \{v | (v, x) \in E \text{ for some } x \in N(S)\} - (S \cup N(S)).$$

Note that $|N(S)| \le \psi n$, and $|N^2(S)| \le (d-1)|N(S)| \le d\psi n$.

There are three stages in colouring S.

Stage 1: We ensure that no colour occurs more than $\left[\psi^{1/2}n^{1/2}\right]$ times on $N^2(S)$, as follows: if colour c occurs more than $\left[\psi^{1/2}n^{1/2}\right]$ times, pick $\left[\psi^{1/2}n^{1/2}\right]$ of the vertices with colour c and recolour them with a new colour not previously used for any vertex. Repeat this until no colour occurs more than $\left[\psi^{1/2}n^{1/2}\right]$ times. Now since each new colour occurs exactly $\left[\psi^{1/2}n^{1/2}\right]$ times, and $\left|N^2(S)\right| \leq d\psi n$, we have used at most $d\psi^{1/2}n^{1/2}$ new colours. Stage 2: We now discard the colours of N(S), and recolour the vertices of $S \cup N(S)$. To do this we consider just the forest induced by the vertices of $S \cup N(S)$ and ignore the rest of the tree.

It was shown in [1] that a tree with *m* edges and maximum degree at most *d* can be given a harmonious colouring with at most $2(2m)^{1/2} + d$ colours, such that for each colour

34

used, the total degree of the vertices receiving that colour is at most $(2m)^{1/2} + d$. Hence we can colour $S \cup N(S)$ with at most $2(2(\phi + \psi)n)^{1/2} + d \le 4\psi^{1/2}n^{1/2} + d$ new colours, and for each colour the degree sum of the vertices with that colour does not exceed $(2(\phi + \psi)n)^{1/2} + d \le 2\psi^{1/2}n^{1/2} + d$.

We now again discard the colours of N(S). (The reason for colouring $S \cup N(S)$ rather than just S above was to ensure that no vertex of N(S) has two neighbours that receive the same colour. Thus we still have a partial harmonious colouring.)

Stage 3: Now recolour N(S) with new colours not previously used.

We colour N(S) sequentially. Order the vertices arbitrarily. We ensure that no colour set on N(S) ever has size > $\left[\psi^{1/2}n^{1/2}\right]$. Thus when we colour an element $v \in N(S)$ some colours are unavailable.

- (a) The colours on the coloured neighbours of v are unavailable. This excludes at most d colours.
- (b) If w ∈ N²(S) ∪ N(S) is adjacent to v and is already coloured, and w' ∈ N²(S) ∪ N(S) is the same colour as w, we cannot use the colour of any neighbour of w'. There are at most d choices for w, then at most [ψ^{1/2}n^{1/2}] for w', and then at most d for a neighbour of w'. Hence this excludes at most d² [ψ^{1/2}n^{1/2}] colours.
- (c) If $w \in S$ is adjacent to v and is already coloured, and $w' \in S$ is the same colour as w, we cannot use the colour of any neighbour of w'. There are at most d choices for w, and for each, the total degree of the vertices of S with the same colour as w is at most $2\psi^{1/2}n^{1/2} + d$, so this excludes at most $2d\psi^{1/2}n^{1/2} + d^2$ colours.
- (d) We must ensure that no uncoloured vertex in N(S) and no vertex in S gets two neighbours in N(S) of the same colour. This excludes at most d(d-1) colours.
- (e) We must exclude any colour that has already occurred $\left[\psi^{1/2}n^{1/2}\right]$ times on N(S); there are at most $|N(S)| / \left[\psi^{1/2}n^{1/2}\right]$ such colours. So we exclude at most $\psi^{1/2}n^{1/2}$.

Hence in total at most $d + d^2 \left[\psi^{1/2} n^{1/2} \right] + 2d\psi^{1/2} n^{1/2} + d^2 + d(d-1) + \psi^{1/2} n^{1/2}$ colours are excluded, so we can colour N(S) with at most $d^2 \left[\psi^{1/2} n^{1/2} \right] + 2d\psi^{1/2} n^{1/2} + 2d^2 + \psi^{1/2} n^{1/2} + 1$ colours.

The three stages together introduce at most

$$d^{2} \left[\psi^{1/2} n^{1/2} \right] + 3d\psi^{1/2} n^{1/2} + 2d^{2} + 5\psi^{1/2} n^{1/2} + d + 1 \le 14d^{2}\psi^{1/2} n^{1/2} \le \alpha n^{1/2} \le \alpha C$$

new colours in total, so we have still used at most C in total. We may also have changed the colour of elements of N(S) and $N^2(S)$, of which there are at most $d\psi n$. Note, however, that the colours of all vertices adjacent to a special leaf remain unchanged, since by the definition of a special leaf such vertices are at distance at least 3 from any element of S, so are not in $S \cup N(S) \cup N^2(S)$. For each new colour, the total degree of the vertices with that colour is at most $\max(2\psi^{1/2}n^{1/2} + d, d\left[\psi^{1/2}n^{1/2}\right]) \leq 3d\psi^{1/2}n^{1/2} \leq 3d\psi^{1/2}(C+1)/\sqrt{2}$. For each colour x in $\{C', \ldots, C-1\}$, we pick a vertex v_x of F' that is adjacent to at least one deleted special leaf. It is easy to ensure that these vertices are pairwise non-adjacent, provided α is small enough. Now for each x, there is a copy of v_x in each of the C' copies of F' that form F. For each x, choose a subset S_x of these copies of v_x so that the number of deleted special leaves adjacent to some element of S_x is at most $\gamma C'$ and at least $\gamma C'-d$. Recolour the elements of S_x with colour x.

Now for each colour $c \in \{0, \dots, C'-1\}$, we know that before this recolouring there were exactly $\gamma C'$ deleted special leaves adjacent to some vertex of colour c; denote by R_c the set of vertices of colour c that were adjacent to a deleted special leaf. For any x, c, at most one element of S_x can be an element of R_c , because prior to recolouring the elements of S_x had distinct colours, while those of R_c all had colour c. Thus altogether at most αC elements of R_c are recoloured, and these have total degree at most $d\alpha C$. Thus the number of deleted special leaves adjacent to some vertex of colour c is still at least $\gamma C' - d\alpha C = (\gamma - d\alpha/(1-\alpha))C'$. Let the number of deleted special leaves adjacent to some vertex of colour *i* be r_i . From above, each r_i satisfies $\gamma C' - d\alpha C \le r_i \le \gamma C'$. Now consider the set of pairs of colours that still do not occur on any edge. Define a graph U on vertex set $V = \{0, ..., C-1\}$ as follows: join a pair of vertices (colours) c_1, c_2 if and only if there is so far no edge of T with endpoints coloured c_1 and c_2 . Note that each of the 'extra' colours $C', \ldots, C-1$ has been used on vertices with total degree at most $3d\psi^{1/2}n^{1/2} + \gamma C'$, hence provided ψ , α and γ are small enough, each such colour will be joined in U to at least C/2 colours in $\{0, \ldots, C'-1\}$. Then by Lemma 2.2 we can complete the colouring of T with C colours if and only if

for every set
$$X \subseteq \{0, \ldots, C-1\}, \sum_{i \in X} r_i \leq |P_X|$$

where P_X is the number of edges of U with at least one endpoint in X. We will show that this is indeed so.

There are two cases:

Case (i): $|X| \le (1-t)C'$. Let $Y = X \cap \{0, ..., C'-1\}$ and $Z = X \cap \{C', ..., C-1\}$. Let Q_Y be the set of edges of U that have both endpoints in $\{0, ..., C'-1\}$ and at least one endpoint in Y.

Now recall the graph U^* defined earlier, with vertex set $\{0, \ldots, C'-1\}$ and vertices c_1, c_2 joined in U^* if and only if the pair of colours c_1, c_2 did not appear on any edge in the colouring of F. The changes to the colouring made after colouring F do not cause any vertex to be coloured (or recoloured) with a colour from $\{0, \ldots, C'-1\}$, so every edge of U^* is also an edge of U. Thus if P_Y^* is the set of edges of U^* with at least one endpoint in Y, then $|Q_Y| \ge |P_Y^*|$. Now if we can show that

$$|Y|\gamma C' \leq |P_Y^{\bullet}|,$$

we have

$$\sum_{i \in X} r_i \leq |X| \gamma C'$$

$$= |Y| \gamma C' + |Z| \gamma C'$$

$$\leq |P_Y^{\bullet}| + |Z| \frac{C}{2}$$

$$\leq |Q_Y| + |Z| \frac{C}{2}$$

$$\leq |P_X|,$$

since there are at least |Z| C/2 edges in P_X that have an endpoint in Z and so are not in Q_Y .

Hence it suffices to show that for all $Y \subseteq \{0, ..., C'-1\}$ with $|Y| \leq (1-t)C'$, we have

$$|Y|\gamma C' \leq |P_Y^*|$$

There are two subcases: Subcase (a): |Y| < tC'. Then

$$|P_Y^{\bullet}| \geq 2tC'|Y| - {Y \choose 2}$$

$$\geq |Y|(2tC' - \frac{1}{2}tC') \geq |Y|\gamma C'$$

provided that $3t/2 \ge \gamma$, which is true if $(1/2)t \ge \xi$. We can choose ξ such that this is so. Subcase (b): $tC' \le |Y| \le (1-t)C'$. By Lemma 2.3, we have

$$|P_Y^{\bullet}| \ge |Y| tC' + \frac{(tC')^2}{4}.$$

To ensure that this is at least $|Y|\gamma C'$, we need

$$|Y| \gamma C' \leq |Y| tC' + \frac{(tC')^2}{4}.$$

This will be true if

$$Y | (\gamma - t) \le \frac{(t^2 C')}{4}$$

Now since $|Y| \le C'(1-t)$ and $(\gamma - t) < \xi$, it suffices that

$$C'(1-t)\xi \leq (\frac{t^2}{4})C'$$

or

$$\xi \leq \left(\frac{t^2}{4(1-t)}\right).$$

Since ξ and t are chosen independently, this can clearly be achieved. **Case (ii):** |X| > C'(1-t). Let $W = \{0, \dots, C-1\} - X$, so that

$$|W| = C - |X| < C'\left(\frac{1}{1-\alpha} - (1-t)\right) = C'\left(t + \frac{\alpha}{1-\alpha}\right).$$

Then P_X consists of all the edges of U except those with both ends in W. Certainly

$$\sum_{i\in V} |r_i| \le |P_V| = |E(U)|,$$

since the left-hand side is the number of edges with one end still uncoloured, and the right-hand side is the number of unused pairs of colours. Hence

$$\begin{split} \sum_{i \in X} r_i &\leq |E(U)| - \sum_{i \in W} r_i \\ &\leq |P_X| + \binom{|W|}{2} - \sum_{i \in W} r_i \\ &\leq |P_X| + \binom{|W|}{2} - |W| \left(\gamma - \frac{d\alpha}{1 - \alpha}\right) C'. \end{split}$$

Hence we have

$$\sum_{i\in X} r_i \le |P_X|$$

provided that

$$\frac{(|W|-1)}{2} \leq \left(\gamma - \frac{d\alpha}{1-\alpha}\right)C',$$

which is true if

$$C'\left(t+\frac{\alpha}{1-\alpha}\right) \leq 2\left(\gamma-\frac{d\alpha}{1-\alpha}\right)C'+1,$$

or if

$$t \leq 2\gamma - (2d+1)\left(\frac{\alpha}{1-\alpha}\right).$$

Since $t \leq \gamma$, this is true provided $\gamma \geq (2d+1)(\alpha/(1-\alpha))$. But $\gamma > \beta/2 \geq \varepsilon/4$ while $\alpha \leq 15d^2(4d\eta)^{1/2}$, so this is true provided η is small enough.

Remark. The proof of the theorem contains a number of inequalities that must be satisfied, and which relate the quantities ε , η , α , β , γ , ϕ , ψ , ξ and t. We must choose η sufficiently small, relative to ε , to allow all the other quantities to be chosen to satisfy the constraints. Note, however, that each of the inequalities is implied by one or more inequalities of the form $x \leq y^k/P(d)$, where k is an integer, and P a polynomial with positive coefficients, and x and y are two of the small quantities listed above. It follows that η need be chosen no smaller than $\varepsilon^k/P(d)$ for some fixed k and P.

Theorem 2.1 applies in particular to complete r-ary trees on a certain number of levels – a case that has been considered by several authors [7, 9, 10]. For any $r \ge 2$, we have that any complete r-ary tree with sufficiently many levels satisfies h(T) = Q(m). We have not investigated exactly how many levels are needed for this to be true.

Lemmas

We now give the three lemmas used in the proof of Theorem 2.1. For any subsets X, Y of the vertices of a graph G, denote by E(X) the set of edges in G that join two vertices in X, and by E(X, Y) the set of edges that join a vertex in X and a vertex in Y.

Lemma 2.1. Let G = (V, E) be an undirected graph, and for each $v \in V$, let a(v) be a non-negative integer. Then it is possible to orient the edges of G so that for each $v \in V$, the outdegree $d^+(v)$ is at least a(v), if and only if

$$\sum_{x \in X} a(x) \le |E(X)| + |E(X, V - X)| \text{ for each } X \subseteq V.$$

Proof. A proof is given in [2].

Lemma 2.2. Suppose that we have a graph consisting of C subgraphs, S_1, \ldots, S_C where, for $i = 1, \ldots, C$, S_i is a collection of disjoint stars, with all the centres coloured i, and r_i other vertices (leaves). Suppose also that we have a set P of unordered pairs of the colours

38

 $\{1, ..., C\}$. Then the graph can be harmoniously coloured with the colours $\{1, ..., C\}$, without changing the colours of the centres, and such that the pair of colours on the endpoints of each edge is an element of P, if and only if

for every set
$$S \subseteq \{1, \ldots, C\}, \sum_{i \in S} r_i \leq |P_S|,$$

where P_S is the subset of pairs in P that contain at least one element of S.

Proof. (Only if) This is trivial.

(If) We use Lemma 2.1. Let G be the graph on C vertices 1, ..., C obtained by joining vertex i to vertex j if and only if $(i, j) \in P$. Let $a(i) = r_i$. By Lemma 2.1, if

$$\sum_{i \in S} a(i) \le |E(S)| + |E(S, \{1, \dots, C\} - S)|$$

for each $S \subseteq \{1, ..., C\}$, we can orient the edges of G so that for each *i*, the out-degree of vertex *i* is at least r_i . But we know that for each $S \subseteq \{1, ..., C\}$,

$$\sum_{i\in S} |r_i| \leq |P_S|,$$

and clearly $|P_S| = |E(S)| + |E(S, \{1, \dots, C\} - S)|$.

Hence G can be oriented so that for each *i*, the out-degree of vertex *i* is at least r_i . So we can choose r_i out-neighbours of *i* and colour the leaves of S_i with these colours. It is clear that a harmonious colouring of the required type results.

Lemma 2.3. Let G be a graph on vertex set $V = \{0, ..., n-1\}$ with vertices i, j joined if and only if $(j-i) \mod n \in \{-k, ..., -1, 1, ..., k\}$, where $n \ge 2k + 1$. Let $S \subseteq V$ with $k \le |S| \le n-k$, and let P_S be the set of edges with at least one endpoint in the set S. Then

$$|P_S| \ge |S| \, k + \frac{k^2}{4}.$$

Proof. First orient the edges of G as follows: edge (i, j) is oriented from i to j if $(j-i) \mod n \in \{1, ..., k\}$. Define the k-section of V starting at i, Q_i , to be the set $\{i, ..., i+k-1\}$. Let T = V - S. There are two cases.

Case (i): First suppose that for some *i*, Q_i contains at least 3k/4 elements of *T*. Then consider the first *k* elements of *S* starting at *i* and working in the forward direction. Let these vertices form a set *R*. There are k^2 edges oriented into an element of *R*. Of these, at most $\binom{k}{2}$ originate from some element of *R*. Of the remainder, any which originate in *S* must end at an element of Q_i , hence, since the number of elements of *R* in Q_i is at most k/4, there are at most $k^2/4$ of these. Hence there are at least $k^2/4$ that start at a vertex outside *S*. To these can be added the |S|k edges oriented away from *S* to give the desired lower bound for $|P_S|$.

Case (ii): Now we must suppose that each Q_i contains at most 3k/4 elements of T, and hence at least k/4 elements of S. But then we can find, as before, |S|k edges oriented away from S. In addition, there are at least k elements of T, and the k-section starting

at each contains at least k/4 elements of S. This gives at least a further $k^2/4$ members of P_S , and completes the proof.

Remark. The conclusion of Lemma 2.3 can be strengthened [8] to

$$|P_{\mathcal{S}}| \ge |\mathcal{S}| \, k + \binom{k+1}{2},$$

which is the best possible.

3. Random trees

In this section we use a slightly modified form of Theorem 2.1 to show that almost all trees T satisfy h(T) = Q(m). More precisely, we show that the proportion of all trees on n vertices for which the equality holds tends to 1 as n tends to infinity. First we need to establish some properties of random trees.

Lemma 3.1. Almost all trees on n vertices have at least n/4 leaves.

Proof. This follows immediately from the work of Robinson and Schwenk [12], who prove that the number of leaves in a random (unlabelled, unrooted) tree has mean about (0.438)n and variance about (0.192)n.

Lemma 3.2. Almost all trees on n vertices have maximum degree at most $K_0 \log n$, for some constant $K_0 > 0$.

Proof. First consider, for each $n \ge 0$, the set of rooted forests with *n* edges and no isolated vertices, so that each component has at least 1 edge and a vertex that is distinguished as its root. (For n = 0 we allow the empty forest with no vertices.) Let the number of such forests be \bar{F}_n , and let

$$\bar{F}(x) = \sum_{n=0}^{\infty} \bar{F}_n x^n.$$

Now let \overline{T}_n be the number of rooted trees with $n \ge 1$ edges, and let

$$\bar{T}(x) = \sum_{n=1}^{\infty} \bar{T}_n x^n.$$

Then clearly $\tilde{T}_n = T_{n+1}$, where T_n is the number of rooted trees on *n* vertices.

Now let $Z(S_k)$ be the cycle index of the symmetric group S_k . $Z(S_k)$ is a polynomial in k variables s_1, \ldots, s_k . $Z(S_0)$ is taken to be 1, and in general the cycle index of a symmetric group may be determined by the recurrence (see for example [5], page 36)

$$Z(S_n) = n^{-1} \sum_{k=1}^n s_k Z(S_{n-k}).$$

It now follows easily from Polya's theorem that

$$\bar{F}(x) = \sum_{k=0}^{\infty} Z(S_k; \bar{T}(x)),$$

where $Z(S_k; \bar{T}(x))$ is obtained from $Z(S_k)$ by substituting $s_i = \bar{T}(x^i)$, for each *i*. Now from the identity (see for example [5], page 52)

$$\sum_{k=0}^{\infty} Z(S_k; p(x)) = \exp\left(\sum_{k=1}^{\infty} p(x^k)/k\right),$$

where p(x) is any power series, it follows that

$$\overline{F}(x) = \exp\left(\sum_{k=1}^{\infty} \overline{T}(x^k)/k\right).$$

We will use the following notation: if p(x) is any power series, we will denote by $[x^n]p(x)$ the coefficient of x^n in p(x).

Now consider $\sum_{k=1}^{\infty} \bar{T}(x^k)/k$. From Otter's theorem [11], we know that $T_n \sim K_1 n^{-3/2} \rho^{-n}$ for some constant $K_1 > 0$, where $\rho \approx 0.338$, so it follows that for some constant $K_2 > 0$, $\bar{T}_n \leq K_2 n^{-3/2} \rho^{-n}$ for all $n \geq 1$. Now $[x^n] \sum_{k=1}^{\infty} \bar{T}(x^k)/k$ is $\sum_{k=1}^n (1/k) \bar{T}_{n/k}$ (take $\bar{T}_r = 0$ if r is not an integer), and

$$\sum_{k=1}^{n} \frac{1}{k} \bar{T}_{n/k} \leq K_2 \sum_{k=1}^{n} \frac{1}{k} \left(\frac{k}{n}\right)^{3/2} \left(\frac{1}{\rho}\right)^{n/k}.$$

Now if $1 \le k \le n$, then $n/k \le n - k + 1$, hence

$$\sum_{k=1}^{n} \frac{1}{k} \bar{T}_{n/k} \leq \frac{K_2 \rho^{-n}}{n^{3/2}} \sum_{k=1}^{n} k^{\frac{1}{2}} \left(\frac{1}{\rho}\right)^{-k+1}$$
$$\leq \frac{K_2 \rho^{-n}}{n^{3/2}} \sum_{k=1}^{\infty} k^{\frac{1}{2}} \left(\frac{1}{\rho}\right)^{-k+1}$$
$$= \frac{K_3 \rho^{-n}}{n^{3/2}}$$
$$\leq \frac{K_3 \rho^{-n}}{n}$$

for some integer constant $K_3 > 0$. Hence we have, for all $n \ge 0$,

$$[x^n]\sum_{k=1}^{\infty} \bar{T}(x^k)/k \le [x^n]\sum_{n=1}^{\infty} \frac{K_3 \rho^{-n} x^n}{n} = [x^n](-K_3 \log(1-\frac{x}{\rho})).$$

Since both sides of this inequality are non-negative, it follows that, for all $n \ge 0$,

$$[x^{n}] \exp\left(\sum_{k=1}^{\infty} \bar{T}(x^{k})/k\right) \leq [x^{n}] \exp(-K_{3} \log(1-\frac{x}{\rho})) = [x^{n}](1-\frac{x}{\rho})^{-K_{3}}.$$

But the left-hand side above is just $[x^n]\overline{F}(x)$, which is \overline{F}_n , so

K. J. Edwards

$$\bar{F}_n \leq [x^n](1-\frac{x}{\rho})^{-K_3} \\ = [x^n] \sum_{n=0}^{\infty} {\binom{n+K_3-1}{K_3-1}} \rho^{-n} x^n \\ = {\binom{n+K_3-1}{K_3-1}} \rho^{-n}$$

Hence $\bar{F}_n \leq An^{K_3-1}\rho^{-n}$ for some constant A > 0.

Now let F_n^* be the number of rooted forests with no isolated vertices, and at most n edges, so that

$$F_n^* = \sum_{i=0}^n \bar{F}_i$$

$$\leq \sum_{i=0}^n A i^{K_3 - 1} \rho^{-i}$$

$$\leq B n^{K_3} \rho^{-n}$$

for some constant B > 0.

Consider now the set of all unrooted trees on n vertices with at least one vertex of degree greater than $K_0 \log n$. In each tree, choose one such vertex v, and delete v and its incident edges, and any isolated vertices that result. Also designate each of the remaining neighbours of v as the root of its component. In this way we obtain a rooted forest with no isolated edges and at most $n - 1 - K_0 \log n$ edges. Further, this construction is clearly reversible, since to recover the original tree we simply add a new vertex v, join it to each root, and add new vertices joined to v to ensure that the tree has n vertices. Hence the number of trees on n vertices with a vertex of degree greater than $K_0 \log n$ is at most

$$F_{n-1-K_0\log n}^* \leq B(n-1-K_0\log n)^{K_3}\rho^{-(n-1-K_0\log n)}.$$

But Otter [11] proved that the number of trees on *n* vertices is at least $K_4 n^{-5/2} \rho^{-n}$ for some constant $K_4 > 0$. Hence the proportion of these that have a vertex of degree greater than $K_0 \log n$ is at most

$$\frac{B(n-1-K_0\log n)^{K_3}\rho^{-(n-1-K_0\log n)}}{K_4n^{-5/2}\rho^{-n}} < \frac{B\rho}{K_4}n^{K_3}n^{5/2}n^{-K_0\log(1/\rho)},$$

which tends to 0 as *n* tends to infinity, provided we set $K_0 = K_3 + 3$ for example.

Remark. Lemma 3.2 also follows from a recent theorem of Goh and Schmutz [4], who show that the maximum degree of a random tree is typically about $c_1 \log n$, where c_1 is about 0.9227.

Lemma 3.3. Let k be a positive integer. Then there is a constant b_k such that the expected total degree of the vertices of degree at least k in a random tree on n vertices is less than $2b_kn$ provided n is large enough. Furthermore, provided k is large enough, $b_k \leq 7k\rho^k/(1-\rho)^2$.

Proof. Let $a_k(n)$ be the expected proportion of vertices in a tree on *n* vertices that have degree *k*. It was shown by Robinson and Schwenk [12] that $a_k(n)$ converges to a constant a_k as *n* tends to infinity. They also established (equation 32) that

$$a_{k} = \frac{2}{b^{2}\rho} \left(\sum_{i=2}^{\infty} D^{(k)}(\rho^{i}) + \rho Z(S_{k-1}; T(\rho)) \right).$$

Here $T(x) = \sum_{n\geq 1} T_n x^n$ and $D^{(k)}(x) = \sum_{n\geq 1} D_n^{(k)} x^n$, where T_n is the number of rooted trees on *n* vertices, and $D_n^{(k)}$ is the total number of vertices of degree *k* (excluding the root in the case k = 1) in all planted trees with *n* vertices in addition to the root. Also, as in Lemma 3.2, $\rho \approx 0.338$ is the radius of convergence of *T* (and *D*), and *b* is a constant about 2.68. Now let $B_k(n)$ be the expected total degree of all vertices of degree at least *k* in a tree on *n* vertices, and let $b_k(n) = B_k(n)/n$. Then clearly

$$b_k(n) = \sum_{i\geq k} ia_i(n).$$

Also, it is clear that $b_1(n) = 2 - 2/n$. Now let

$$b_k = \sum_{i\geq k} ia_i$$

It is shown by Schwenk [13] that provided k is large enough, $a_k \leq 7\rho^k$. Hence, if k is large enough, we have

$$b_k = \sum_{i \ge k} ia_i \le \sum_{i \ge k} 7i\rho^i = \frac{7\rho}{(1-\rho)^2} (k\rho^{k-1} - (k-1)\rho^k) \le \frac{7k\rho^k}{(1-\rho)^2}$$

The result will now follow if we can show that $b_k(n) \rightarrow b_k$ as $n \rightarrow \infty$. Now

$$b_k - b_k(n) = b_1 - \sum_{i=1}^{k-1} ia_i - b_1(n) + \sum_{i=1}^{k-1} ia_i(n)$$
$$= b_1 - 2 + 2/n + \sum_{i=1}^{k-1} i(a_i(n) - a_i),$$

which converges to $b_1 - 2$. Thus it suffices to show that $b_1 = 2$. To see this, we note first that Robinson and Schwenk [12] showed (equation 36) that

$$D^{(k)}(x) = \frac{T(x) \sum_{i \ge 2} D^{(k)}(x^i) + xZ(S_{k-1}; T(x))}{1 - T(x)}.$$

Hence we have

$$D^{(k)}(x)(1-T(x)) = T(x) \sum_{i\geq 2} D^{(k)}(x^i) + xZ(S_{k-1};T(x)).$$

Multiplying both sides of this equation by k and summing, we obtain

$$(1-T(x))\sum_{k\geq 1} kD^{(k)}(x) = T(x)\sum_{k\geq 1} k\sum_{i\geq 2} D^{(k)}(x^i) + \sum_{k\geq 1} kxZ(S_{k-1};T(x)).$$

But

$$\sum_{k\geq 1} kD^{(k)}(x) = \sum_{n\geq 1} (2n-1)T_n x^n = 2xT'(x) - T(x).$$

Thus we have

$$T(x)\sum_{k\geq 1} k\sum_{i\geq 2} D^{(k)}(x^i) + \sum_{k\geq 1} kxZ(S_{k-1};T(x)) = 2xT'(x)(1-T(x)) - T(x)(1-T(x)).$$

Now let $x \to \rho^-$, and use the fact that $T(\rho) = 1$ and $\lim_{x\to\rho^-} T'(x)(1-T(x)) = (1/2)b^2$ (see, for example, [5], equation (9.5.13), page 211, and equation (9.5.24), page 212), to obtain

$$\sum_{k\geq 1} k\left(\sum_{i\geq 2} D^{(k)}(\rho^{i}) + \rho Z(S_{k-1}; T(\rho))\right) = \rho b^{2},$$

from which

$$b_1 = \sum_{k \ge 1} ka_k = \frac{2}{b^2 \rho} \sum_{k \ge 1} k \left(\sum_{i \ge 2} D^{(k)}(\rho^i) + \rho Z(S_{k-1}; T(\rho)) \right) = 2,$$

as required.

We now prove a slightly modified form of Theorem 2.1. We fix the value of ε as 1/4.

Lemma 3.4. Let d be a positive integer. Then provided d is large enough, there exists an integer N = N(d) such that the following holds: let T be any tree with $n \ge N$ vertices satisfying (i) the maximum degree is at most $K_0 \log n$, where $K_0 > 0$ is a constant, (ii) the total degree of the vertices of degree greater than d is at most $n/2^d$, (iii) T has at least n/4 leaves. Then

$$h(T) = Q(m).$$

Proof. The proof is a slight modification of that of Theorem 2.1 with $\varepsilon = 1/4$. First note that from the remark following Theorem 2.1, we can take $\eta = 1/4^k P(d)$, where k is a constant and P a fixed polynomial with positive coefficients. Clearly we can choose D such that for any $d \ge D$, $2^{-d} \le 1/2(4^k)P(d) = \eta/2$. We show that the result holds for any such d.

We will place all vertices of degree greater than d into the set S. This will increase the size of S from at most $\eta n + c(\eta)PC'$ to at most $3\eta n/2 + c(\eta)PC'$, but this is still less than $2\eta n$, provided n is large enough. As before, we take $\phi = |S|/n$ and $\psi = 2\phi d$. Now the total degree of the vertices in S is at most $\phi dn + n/2^d \le n(\phi d + \eta/2) \le \psi n$, as before.

The rest of the proof goes through unchanged, except that in Stage 2 of the colouring process, the maximum degree of the forest induced by the vertices of $S \cup N(S)$ may be greater than d. However, it is at most $K_0 \log n$, hence the number of colours added here is at most $2\psi^{1/2}n^{1/2} + K_0 \log n$. Correspondingly, in Stage 3, part (c), we have to exclude $2d\psi^{1/2}n^{1/2} + dK_0 \log n$ colours. However, for large n these increases are small compared to the total number of colours added, and for n large enough we still have that the total

44

number of colours added by the three stages is at most $14d^2\psi^{1/2}n^{1/2} \le \alpha n^{1/2} \le \alpha C$. The rest of the proof is unchanged.

We can now state and prove the main theorem.

Theorem 3.1. For almost all trees T, the harmonious chromatic number h(T) of T is Q(m) where m is the number of edges of T.

Proof. Let ε be any positive number. We will show that provided *n* is sufficiently large, the proportion of trees *T* on *n* vertices for which h(T) = Q(m) is at least $1 - \varepsilon$. Choose $d \ge D$, where *D* is the integer defined in Lemma 3.4, such that $14d2^d \rho^d/(1-\rho)^2 \le \varepsilon/2$. Now, by Lemma 3.3, the expected total degree of the vertices of degree greater than *d* in a tree on *n* vertices is at most $(14d\rho^d/(1-\rho)^2)n$, hence by the Markov inequality, the proportion of trees on *n* vertices for which this total degree is more than $n/2^d$ is at most $\varepsilon/2$. Also, by Lemma 3.1 and Lemma 3.2, the proportion of trees on *n* vertices that have fewer than n/4 leaves or maximum degree greater than $K_0 \log n$ is at most $\varepsilon/2$ if *n* is large enough. But by Lemma 3.4, all the remaining trees, that is a proportion at least $1-\varepsilon$ of the trees on *n* vertices, satisfy h(T) = Q(m) provided *n* is large enough.

4. Line-distinguishing colourings

A line-distinguishing colouring of a graph is like a harmonious colouring except that it need not be proper. Thus, for each colour there is at most one edge with both ends with that colour. The line-distinguishing chromatic number ld(G) is the least number of colours in such a colouring.

It is easy to see that if a graph is coloured with k colours, it cannot have more than $\binom{k}{2} + k = \binom{k+1}{2}$ edges. It follows that for any graph G, $ld(G) \ge Q(m) - 1$.

Each of the results proved above has an analogue for line-distinguishing colourings, with an almost identical proof. Thus we have the following analogues of Theorem 2.1 and Theorem 3.1.

Theorem 4.1. Let d be a positive integer and let $\varepsilon > 0$. Then there exists an integer $N = N(d, \varepsilon)$ such that if T is any tree with $n \ge N$ vertices (and m = n-1 edges), maximum degree at most d and at least ε leaves, then T satisfies

$$ld(T) = Q(m) - 1.$$

Theorem 4.2. For almost all trees T, the line-distinguishing chromatic number ld(T) of T is Q(m) - 1, where m is the number of edges of T.

Frank, Harary and Plantholt [3] conjectured the weaker result that for almost all trees, ld(T) = Q(m) - 1 or Q(m).

5. Concluding remarks

We have seen that for bounded degree trees with at least some fixed proportion of vertices being leaves, we have that h(T) = Q(m) if T is large enough, and that almost all trees satisfy h(T) = Q(m).

It is natural to consider also bounded degree trees with at most some fixed proportion of vertices being leaves. Note that this case is more complicated, because there are arbitrarily long paths that require Q(m) + 1 colours for a harmonious colouring. We will consider this case in a later paper.

We might also consider whether, for other classes of graphs, we can show that for almost all members of the class, h(G) = Q(m). Note that this is not true for graphs in general, as almost all graphs have diameter 2 and so h(G) = |V(G)|. In view of the result in [1] that planar graphs satisfy h(G) = Q(m)(1 + o(1)), a more appropriate class for study might be planar graphs with minimum degree at least 3.

Acknowledgments

I would like to thank Colin McDiarmid for useful comments on this work, John Moon for some information about random trees, and the referee for a number of helpful comments.

References

- Edwards, K. J. and McDiarmid, C. J. H. (1994) New Upper Bounds on Harmonious Colorings. J. Graph Theory. 18 257-267.
- [2] Edwards, K. J. and McDiarmid, C. J. H. (to appear) The Complexity of Harmonious Colouring for Trees. *Disc. Appl. Maths.*.
- [3] Frank, O., Harary, F. and Plantholt, M. (1982) The line-distinguishing chromatic number of a graph. Ars Combinatoria 14 241-252.
- [4] Goh, W. M. Y. and Schmutz, E. (1994) Unlabelled Trees: Distribution of the Maximum Degree. Random Struct. Alg. 5 411-440.
- [5] Harary, F. and Palmer, E. M. (1973) Graphical Enumeration. Academic Press.
- [6] Hopcroft, J. and Krishnamoorthy, M.S. (1983) On the harmonious coloring of graphs. SIAM J. Alg. Disc. Methods 4 306-311.
- [7] Lu, Z. (1993) The harmonious chromatic number of a complete binary and trinary tree. Disc. Maths. 118 165-172.
- [8] McDiarmid, C. J. H. Personal communication.
- [9] Miller, Z. and Pritikin, D. (1991) The harmonious colouring number of a graph. Disc. Maths. 93 211-228.
- [10] Mitchem, J. (1989) On the harmonious chromatic number of a graph. Disc. Maths. 74 151-157.
- [11] Otter, R. (1948) The number of trees. Ann. Math. 49 583-599.
- [12] Robinson, R. W. and Schwenk, A. J. (1975) The Distribution of Degrees in a Large Random Tree. Disc. Maths. 12 359-372.
- [13] Schwenk, A. J. (1977) An Asymptotic Evaluation of the Cycle Index of a Symmetric Group. Disc. Maths. 18 71-78.
- [14] Wilson, B. (1990) Line distinguishing and harmonious colourings. In: Nelson, R. and Wilson, R. J. (eds.) Graph Colourings, Pitman Research Notes in Mathematics 218 115–133, Longman Scientific & Technical.