

BACKTRACK: AN O(1) EXPECTED TIME ALGORITHM FOR THE GRAPH COLORING PROBLEM *

Herbert S. WILF

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

Communicated by Kenneth C. Sevcik

Received 28 July 1983

Revised 17 November 1983

Fix a number K , of colors. We consider the usual backtrack algorithm for the decision problem of K -colorability of a graph G . We show that the algorithm operates in average time that is $O(1)$, as the number of vertices of G approaches infinity. For instance, a backtrack search tree for 3-coloring a graph has an average of about 197 nodes, averaged over all graphs of all sizes.

Keywords: Algorithm, graph coloring, backtrack, backtracking, average complexity

Consider the following NP-complete problem:

“Given a graph G and a positive integer K . Can the vertices of G be properly colored in K (or fewer) colors?”

We will show that the backtrack search tree for this problem has an *average* of $O(1)$ nodes, as $V = |V(G)|$ grows without bound. Therefore, the NP-complete problem can be solved by an algorithm that certainly delivers the correct answer, and that does so in an average of $O(1)$ time in V .

This means, of course, that typically the algorithm will never even look at most of the input data and will halt with “no”. While it is clear that this must happen frequently, it is noteworthy that it happens often enough that the few cases of exponentially long search time do not disturb the conclusion that the average search tree is of bounded size.

We will write $H_L(G)$ for the subgraph that is induced by vertices $1, 2, \dots, L$ of the vertex-labelled graph G . \mathcal{G}_n will denote the set of all such graphs G of n vertices. By the *search tree* of G we mean the tree whose nodes are on levels $L = 0, 1, \dots, V$, and in which there is a node at level L correspond-

ing to every proper coloring of $H_L(G)$ in K colors. A node v' at level L is connected by an edge of the search tree to a node v'' at level $L + 1$ iff the colors of vertices $1, 2, \dots, L$ are the same at v' and v'' . Level 0 contains a single root node, corresponding to the empty coloring.

If, for example, G is the graph given in Fig. 1, then the complete backtrack search tree for G , in colors 1, 2, 3, is as shown in Fig. 2. This search tree has 46 nodes.

Among all graphs $G \in \mathcal{G}_n$ the complete graph K_n has the smallest search tree, and it has

$$f_n(K) = 1 + K + K(K - 1) + K(K - 1)(K - 2) + \dots + K!$$

nodes, if $n \geq K$.

Again, among all graphs $G \in \mathcal{G}_n$, the completely disconnected graph \bar{K}_n has the largest

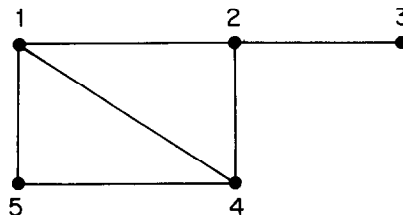


Fig. 1.

* Research supported by the National Science Foundation.

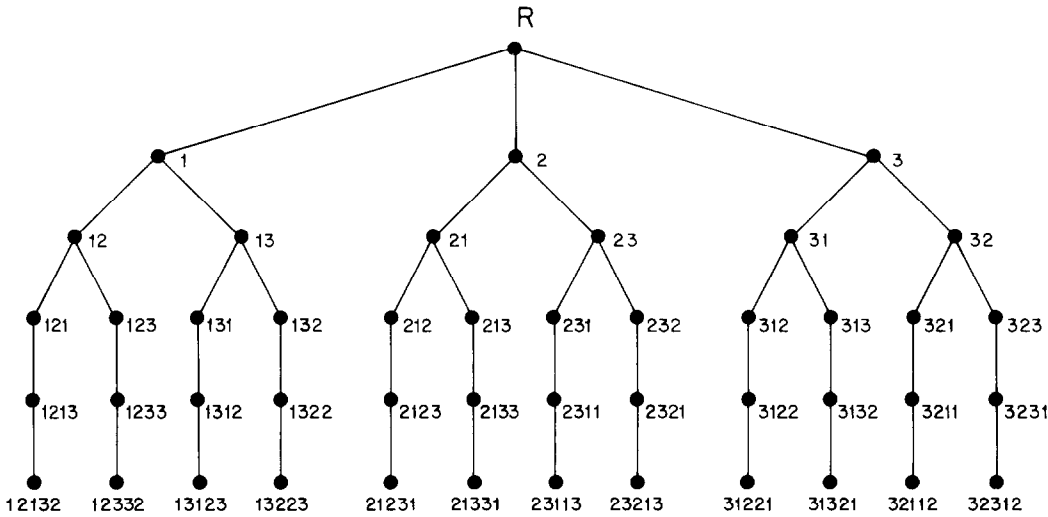


Fig. 2.

search tree, and it has

$$F_n(K) = 1 + K + K^2 + \dots + K^n$$

nodes.

Hence, for all graphs $G \in \mathcal{G}_n$ the number

$$\beta(G) = \beta(G; K)$$

of nodes in the backtrack search tree of G satisfies

$$f_n(K) \leq \beta(G) \leq F_n(K).$$

Note that $F_n(K)$ grows exponentially fast with n , but that $f_n(K) = O(1)$.

So much for the extremes. Now the question is, on the *average*, how big is $\beta(G)$ for $G \in \mathcal{G}_n$?

Theorem. *If $\beta_n = \beta_n(K)$ is the average of $\beta(G)$ over all $G \in \mathcal{G}_n$ for a fixed K , then*

$$\lim_n \beta_n = h(K) < \infty \tag{1}$$

exists for every $K = 1, 2, \dots$.

To prove this theorem, we first have the following.

Lemma 1. *Let C be one of the K^L possible assignments of K colors to L abstract ‘vertices’ $1, 2, \dots, L$. The number of graphs G of L vertices such that G admits C as a proper vertex coloring cannot exceed*

$$Q(L) = 2^{L^2(1-1/K)/2}.$$

Proof. Suppose that the color classes of C have cardinalities s_1, \dots, s_K . A graph G of L vertices admits C as a proper coloring iff the edges of G run between different color classes. The number of edges that G *might* have is therefore

$$\sum_{1 \leq i < j \leq K} s_i s_j = \frac{1}{2} \left\{ L^2 - \sum_{i=1}^K s_i^2 \right\}$$

and the number of G that admits C is exactly

$$2^{\{L^2 - \sum_{i=1}^K s_i^2\}/2}.$$

This is maximum when all s_i are equal to L/K , and therefore the number of such G cannot exceed $Q(L)$. \square

Lemma 2. *If $P(K, G)$ is the number of proper colorings of the vertices of G in K colors, then for the sum $M_L(K)$ of $P(K, G)$ over all graphs G of L vertices we have the estimate*

$$M_L(K) \leq K^L 2^{L^2(1-1/K)/2}.$$

Proof. Define a function f on each pair (C, H) consisting of one of the K^L K -colorings, C , of L abstract ‘vertices’, and a graph H of L vertices, as follows: $f(C, H) = 1$ if C is a proper coloring of H and $= 0$ otherwise.

Then we have

$$\begin{aligned}
 M_L(K) &= \sum_{H \in \mathcal{G}_L} P(K, H) \\
 &= \sum_{H \in \mathcal{G}_L} \left\{ \sum_C f(C, H) \right\} \\
 &= \sum_C \left\{ \sum_{H \in \mathcal{G}_L} f(C, H) \right\} \\
 &\leq K^L \max_C \left\{ \sum_{H \in \mathcal{G}_L} f(C, H) \right\}.
 \end{aligned}$$

Hence $M_L(K)$ is at most K^L times the maximum number of graphs that can belong to any given coloring of $1, \dots, L$ in K colors, and the result follows by Lemma 1. \square

Proof of the theorem. Now we prove the theorem. For the average search tree size we have

$$\begin{aligned}
 \beta_n &= 2^{-\binom{n}{2}} \sum_{\mathcal{G}_n} \beta(G) \\
 &= 2^{-\binom{n}{2}} \sum_{\mathcal{G}_n} \sum_{L=0}^n P(K, H_L(G)) \\
 &= 2^{-\binom{n}{2}} \sum_{L=0}^n \sum_{H \in \mathcal{G}_L} \left\{ \sum_{H_L(G)=H} P(K, H) \right\}.
 \end{aligned}$$

Consider the innermost sum. How many graphs G of n vertices have a given graph H , of L vertices, as their $H_L(G)$? Of the $\frac{1}{2}n(n-1)$ possible edges of G , $\frac{1}{2}L(L-1)$ of them are determined by H . Hence there are exactly

$$2^{\{n(n-1)/2 - L(L-1)/2\}}$$

such graphs G . Hence we have

$$\begin{aligned}
 \beta_n &= 2^{-\binom{n}{2}} \sum_{L=0}^n \sum_{H \in \mathcal{G}_L} P(K, H) 2^{\binom{n}{2} - \binom{L}{2}} \\
 &= \sum_{L=0}^n 2^{-\binom{L}{2}} \left\{ \sum_{H \in \mathcal{G}_L} P(K, H) \right\}.
 \end{aligned}$$

Now if we use the result of Lemma 2, we get for the average number of nodes in a backtrack search tree, an 'exact formula' and an estimate, namely

$$\beta_n = \sum_{L \leq n} 2^{-\binom{L}{2}} M_L(K), \tag{2a}$$

$$\beta_n \leq \sum_{L=0}^{\infty} 2^{L/2} K^L 2^{-L^2/2K}. \tag{2b}$$

The proof of the theorem is complete, since the infinite series is clearly convergent, and from (2a) the monotonicity of $\{\beta_n\}$ is clear. \square

The first few values of $h(K)$, computed from (2a), are approximately 2.64, 13.9, 197, 7733, 755000 for $K = 1, \dots, 5$, respectively.

Remark. The estimation of $M_L(K)$ in Lemma 2 was inspired by the work of Read [1] and Wright [2], who studied the function $M_L(K)$ ("the number of K -colored graphs on L labeled nodes") above and estimated its growth with much more precision than we needed for our present purposes of algorithmic analysis.

In a forthcoming paper, to be published elsewhere, E.A. Bender and the present author investigate in more detail the properties of the distribution of search tree sizes.

References

[1] R.C. Read, The number of k -coloured graphs on labelled nodes, *Canad. J. Math.* 12 (1960) 410-414.
 [2] E.M. Wright, Counting coloured graphs, *Canad. J. Math.* 13 (1961) 683-693.
 [3] R.C. Read and E.M. Wright, Coloured graphs: A correction and extension, *Canad. J. Math.* 22 (1970) 594-596.