SELECTED COMBINATORIAL RESEARCH PROBLEMS

BY

V. CHVATAL

D. A. KLARNER

D. E. KNUTH

STAN-CS-72-292

JUNE 1972

COMPUTER SCIENCE DEPARTMENT

School of Humanities and Sciences

STANFORD UNIVERSITY

SELECTED COMBINATORIAL RESEARCH PROBLEMS

by

V. Chvatal, D. A. Klarner, and D. E. Knuth

Abstract

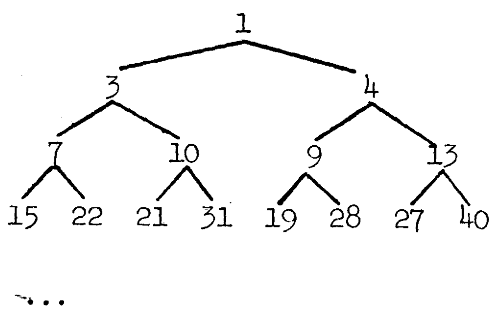
Thirty-seven research problems are described, covering a wide range of combinatorial topics. Unlike Hilbert's problems, most of these are not especially famous and they might be "do-able" in the next few years.

(Problems 1-16 were contributed by Klarner, 17-26 by Chvátal, 27-37 by Knuth. All cash awards are Chvátal's responsibility.)

This research was sup-ported in part by the National Science Foundation under grant number GJ-992, and the Office of Naval Research under contract number N-00014-67-A-0112-0057 NR 044-402. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Problem 1.

Consider the set $\langle 2x+1, 3x+1: 1\rangle$ defined to be the smallest set of natural numbers which contains 1 and is closed under the operations $x \to 2x+1$ or $3x+1$. The set can be constructed by iterating these operations as indicated in the following tree.



Michael Fredman showed in his thesis that this set has density 0 in the set of all natural numbers; hence, $S = \langle 2x+1, 3x+1: 1\rangle$ does not contain an infinite arithmetic progression. Let N denote the set of all natural numbers. Is it true that $N \backslash S$ may be expressed as a disjoint union of infinite arithmetic progressions?

Problem 2. Milner's Problem (Robin Milner at Stanford A. I. project)

Let B_n denote the set of all binary sequences of length n. Suppose $m \leq n$, $\bar{a} \in B_n$, $\bar{b} \in B_m$, and let $\nu(\bar{a}, \bar{b})$ denote the number of subsequences of $\bar{a}$ equal to $\bar{b}$. The m-list of $\bar{a} \in B_n$ consists of a knowledge of the numbers $\nu(\bar{a}, \bar{b})$ for all $\bar{b} \in B_m$. How large must m be such that the m-lists for all elements $\bar{a} \in B_n$ are distinct? This is Milner's problem. Chvátal, Rivest, and Klarner have obtained some results on this problem. A related problem is the following. There are many identities connecting the ν's. For example, let $B_{m,n}$ denote a $2^m \times 2^n$ matrix with $b_{i,j}$ defined to be $\nu(\bar{i}, \bar{j})(n-m)!$ where

$\bar{i}$ and $\bar{j}$ denote the binary sequences of length n and m used to represent i and j respectively. Then it is easy to check that $B_{r,s}B_{s,t} = B_{r,t}$. Many other identities exist which seem to be algebraically independent of these. For example, $\nu(\bar{a},(11)) = \binom{\nu(\bar{a},(1))}{2}$. The problem is to find a basis for all algebraic identities relating the numbers $\nu(\bar{a},\bar{b})$ for fixed $\bar{a}$, as $\bar{b}$ ranges over all binary sequences.

Problem 3.

Recently, Ron Rivest and David Klarner succeeded in showing that $\alpha < 4.65$, where $\alpha = \lim_{n\to\infty}(a(n))^{1/n}$ and a(n) denotes the number of connected square-celled animals with n cells. In fact, we designed a procedure for calculating numbers $\alpha_1, \alpha_2, \ldots$ such that $\alpha < \alpha_{i+1} < \alpha_i$ for all i . We were unable to prove, but conjecture that

$$\lim_{i\to\infty} \alpha_i = \alpha \ .$$

Prove or disprove our conjecture. Try to beat our upper bound $\alpha < 4.65$.

Reference: D. Klarner and R. Rivest, "A procedure for improving the upper bound for the number of n-ominoes," CS 263, Computer Science Department, Stanford University, February 1972.

Problem 4.

Give a "sieve formula" for enumerating planted plane trees having certain subtrees excluded. The n-omino enumeration problem is a special case of this problem.

Problem 5.

More on plane trees. A famous problem in probability theory (solved, by the way) asks for the probability that a candidate always has at least j/k of the votes cast. Here is a related enumeration problem. How many binary sequences $(a_1, a_2, \dots, a_{kn})$ of length kn containing exactly jn ones satisfy the conditions

$$a_1 + \dots + a_{kn} \geq jm \quad \text{for } m = 1, \dots, n \ ?$$

When j = 1 , the solution is

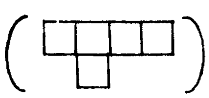
$$\frac{\binom{kn+k}{n+1}}{(kn+k-n)} .$$

Problem 6.

Give a simple proof that if a rectangle is cut into three congruent n-ominoes, then the n-omino is a rectangle.

Problem 7.

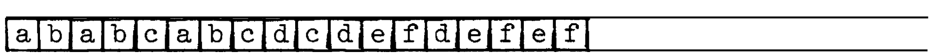
Find the smallest number x > 0 , such that copies of the Y-pentomino



pack a 12 x 5x rectangle. Klarner holds the record with x = 16 .

Problem 8.

Every 3-celled animal on the line packs some interval. An example of a 3-celled animal and an interval it packs:



the interval

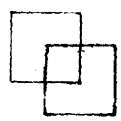
the animal

the animal's reflection

Translations of the animal and its reflection are used in the packing. Here gaps between the cells are 1 and 2, and the length of the smallest interval the animal can pack has length $\ell(1,2) = 18$. If the gaps between the cells are m and n in a three-celled animal, it can be shown that the length of the smallest interval the animal can pack $\ell(m,n)$ is bounded above by $1+3^{m+n}$. This proof depends on the following algorithm: Suppose $m < n$, let A denote the animal with gap m on the left, and let B denote the animal with gap m on the right. We pack a one-way infinite strip of cells as follows. Fill the first cell with the left-most cell of A. Fill the left-most unfilled cell in the strip with the left-most cell of A, if there is overlap remove A and try B. It is an interesting exercise to show that this procedure results in a packing of an interval whose length is not greater than $1+3^{m+n}$. Let $\ell(m,n)$ denote the length of the interval packed by this algorithm. Give a nice upper bound on $\ell(m,n)$, and find out if it satisfies some kind of recurrence relation.

Problem 9.

Does every 4-celled animal in the plane pack the plane? Does every 5-celled animal in the plane pack the plane? There is at least one 6-celled animal that does not pack the plane, namely,



A 3n-celled animal like this one can be constructed which does not pack E_n. Thus, if every n-celled animal packs E_k, then $n < 3k$. Improve this upper bound if possible.

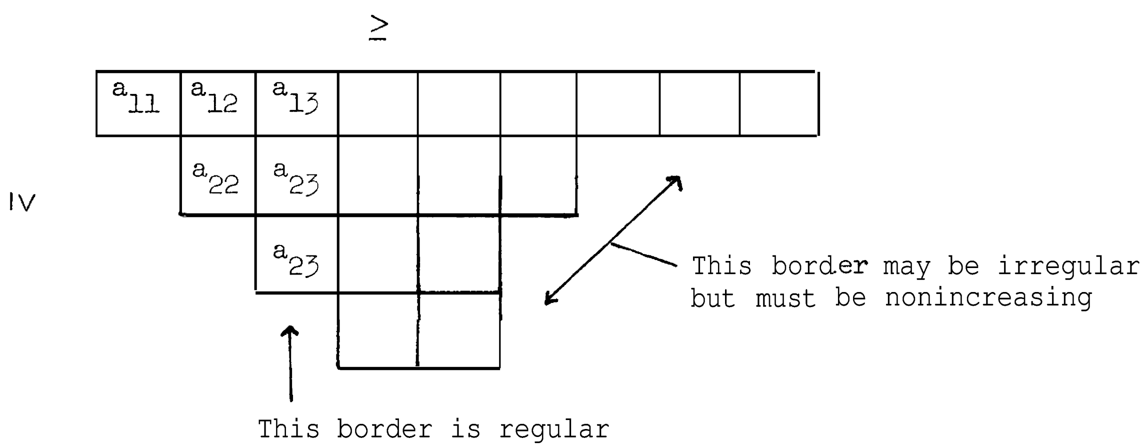
Problem 10. (R. Rado)

Consider sets of squares in the plane having sides parallel to the x- or y-axis. Let a(S) denote the area covered by the union of suchaset S . Is it true that"

$$\max_{\substack{T \subseteq S \\ \square\text{'s in } T \text{ disjoint}}} \frac{a(T)}{a(S)} \geq \frac{1}{4} \quad ?$$

Problem 11. (R. Stanley)

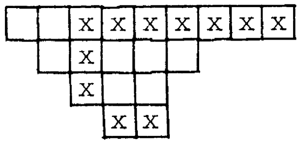
Consider partitions of n which satisfy a tableau condition:



The entries a_{ij} do not decrease in the rows or columns, and their total is n . Let T denote the shape of the array, and let $\nu_T(n)$ denote the number of ways of filling in the array subject to these conditions. Prove that the generating function is as follows:

$$\prod_{(i,j)\in T} \frac{1}{1 - x^{d_{i,j}}} = \sum_{n=0}^{\infty} \nu_T(n) x^n$$

where the numbers $d_{i,j}$ are defined as in the following example:



To find d_{13} begin in cell (1,3) and count all cells in its row to the right of (1,3), count also the cells in the column below (1,3), and if one can "turn the corner" at the bottom of this column, count the cells in this row as well. Thus, $d_{1,5} = 8$, $d_{2,4} = 5$, $d_{3,5} = 2$, etc.

Problem 12. An extremal problem (see problem 5728 of Amer. Math. Monthly, 1970).

The "octahedron" in E_n has 2^{n-1} different pairs of parallel hyperplanes spanned by two n-sets whose union comprises the vertex set of the octahedron. Prove that the octahedron is an optimal configuration of 2n points in E_n having the property that the points span many pairs of parallel hyperplanes.

Problem 13.

R. C. Read (J. London Math. Soc., 1963, 99-104) enumerated classes of isomorphic self-complementary linear graphs with 4n vertices and classes of isomorphic self-complementary directed graphs with 2n vertices. It turns out that these numbers are equal. Give a "natural" one-one correspondence between the two sets.

Problem 14.

Recently, Klarner showed that the set $S = (m_1x_1 + \ldots + m_rx_r : 1)$ (that is, the smallest set of natural numbers which contains 1 and is closed under the operation $m_1x_1 + \ldots + m_rx_r$ where $m_1, \ldots, m_r$ are given natural numbers) is a finite union of infinite arithmetic progressions provided (i) $r \geq 2$, (ii) $(m_1, \ldots, m_r) = 1$, and (iii) $(m_1 \ldots m_r, m_1 + \ldots + m_r) = 1$. Does the conclusion still follow if we drop hypothesis (iii)?

Problem 15.

Hautus and Klarner gave a simple characterization of all uniform $\{m \times n\}$-colorings of the square plane lattice provided $(m,n) = 1$. We were unable to describe the uniform colorings when $(m,n) > 1$. Any nice theorems about these designs?

Problem 16. (Due to Leo Moser.)

Can the whole plane be tiled by using exactly one square each of sides 1,2,3,4,... ?

Problem 17.

The ordinary game of tic-tat-toe is an instance of a positional game played on a hypergraph $H = (V,E)$. Here V (the set of vertices of H) is a finite set and E (the set of edges of H) is a set of subsets of v. Two players take turns to claim a previously unclaimed vertex of H. If a player claims all the vertices of an edge of H, he wins. If all the vertices of H have been claimed but no one has yet won then the game is a draw. An easy argument (Hales and Jewett, "Regularity and positional games," Trans. Amer. Math. Soc. 106 (1963), 222-229) shows that the second player cannot have a winning strategy. Besides, if the game results in a draw then there is a partition $V = V_1 \cup V_2$ such that no V_i contains an edge (in that case, H is called 2-colorable).

Given positive integers n, k with $k \leq n$ we define a hypergraph W(n,k) by setting $V = \{1,2,\ldots,n\}$ and letting a set $A \subset V$ to be an element of E if, and only if, $|A| = k$ and the elements of A form an arithmetic progression. Van der Waerden (Beweis einer Baudetschen Vermutung, Nieuw Archief v. Wiskunde 15 (1928), 212-216) proved that given any k there is always an n such that W(n,k) is not 2-colorable. Let N(k) be the smallest such n. It is easy to show that N(2) = 3 and N(3) = 9; one has N(4) = 35 (see Chvátal, "Some unknown van der Waerden numbers," Combinatorial Structures and Their Applications (R. K. Guy et al., Eds.), Gordon and Breach, New York, 1970). As far as I know, the value of N(5) is still unknown. The existing upper bounds on N(k) are beyond the range of algebraic expressions. The existence of N(k) implies the existence of the smallest n = n(k) such that the first player has a winning strategy on W(n,k). Obviously, we have $n(k) \leq N(k)$. One has n(3) = 5 and n(4) = 13 (see Chvátal, "Hypergraphs and Ramseyian theorems," Thesis, University of Waterloo, 1970). Apparently, N(k) is a rather poor upper bound for n(k).

What is the value of n(5)? Can you find a decent upper bound for n(k)? Is n(k) always odd? If so, is $\frac{1}{2}(n(k)+1)$ a winning first move? Is there a winning strategy for the first player on W(n,k) for all $n \geq n(k)$?

Problem 18.

A k-graph is a hypergraph (V,E) with $|A| = k$ for all $A \in E$. Let $m(k)$ be the smallest $|E|$ in a k-graph which is not 2-colorable. Obviously, $m(2) = 3$. It is not difficult to show that $m(3) = 7$; the edges of the corresponding 3-graph are the lines of a projective plane of order two. One has

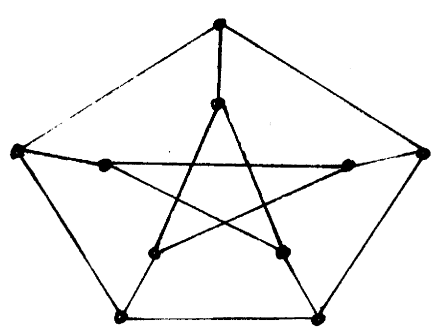
$$2^k(1+2k^{-1})^{-1} < m(k) < \lceil k^2 2^k e^{1/2} \log 2/(1+(1+2p)) \rceil .$$

(Herzog and Schönheim, "The B_r property and chromatic numbers of generalized graphs,'* J. Combinatorial Theory 12 (1972), 41-49), improving $2^{k-1} < m(k) < k^2 2^{k+1}$ due to Erdös.)

Erdös repeatedly asks for the value of $m(4)$. Perhaps a computer would help.

Problem 19.

A graph G is called hypohamiltonian if it contains no hamiltonian circuit (that is, a circuit passing through all the vertices of G), but given any vertex u of G , the vertex-deleted subgraph G−u has a hamiltonian circuit. The smallest hypohamiltonian graph is the Petersen graph.



Herz, Duby and Vigué ("Recherche Systématique des Graphes Hypohamiltonians," Theory of Graphs (P. Rosenstiehl, Ed.), 1966) used a computer to search for hypohamiltonian graphs with 11 or 12 vertices and found that there are none. However, they discovered one with 13 and another one with 15 vertices. Since then, the existence of hypohamiltonian graphs with n vertices has been demonstrated for all $n \geq 13$ except for

$$n = 14, 17, 19, 20, 25$$

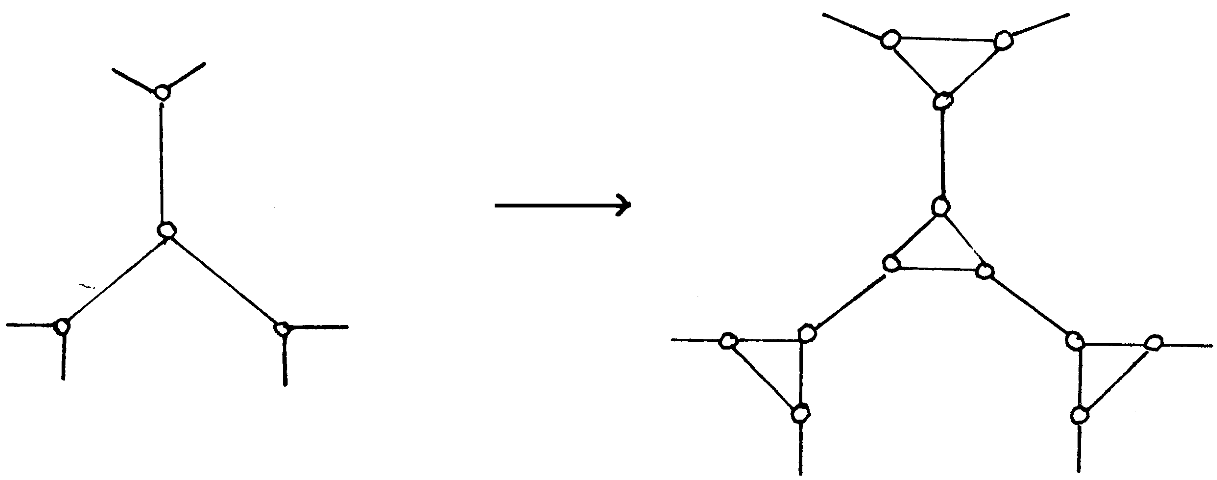
(see Chvátal, '*Flip-flops in hypohamiltonian graphs,'* to appear in Canad. Math. Dull.). Perhaps it is time to settle at least the case $n = 14$ (computers could help).

The hypohamiltonian graphs offer a number of amusing questions. It seems that these graphs never contain a circuit of length three or four. However, so far no one has found any graph F such that no hypohamiltonian graph contains F . I offer $5.00 for an example of a planar hypohamiltonian graph or a proof that there is none.

Problem 20.

A graph G is called t-tough if deletion of any m points from G results in a graph that is either connected or else has at most m/t components. It is not difficult to see that every hamiltonian graph is 1-tough but the converse is not true (the Petersen graph is $\frac{4}{3}$ -tough). I offer $10 ● & for the proof that every t-tough graph is hamiltonian and $10 ●$^{t+1}$m for an example of a t-tough graph ($t > \frac{3}{2}$) which is not hamiltonian.

Fleischner ("Square of a block is hamiltonian," to appear in J. Combinatorial Theory) proved that the squares of a 2-connected graph is always hamiltonian. (The square G^2 of a graph G is defined to be the graph having the same vertices as G ; vertices u , v are adjacent in G^2 if and only if they have distance at most two in G .) Since the square of a k-connected graph is always k-tough, it is desirable to prove that every 2-tough graph is hamiltonian. An example of a $\frac{3}{2}$ -tough nonhamiltonian graph is obtained when in the Petersen graph, each vertex is replaced by a triangle as indicated below.



Some results on toughness are contained in a forthcoming paper of mine, to appear in Discrete Mathematics.

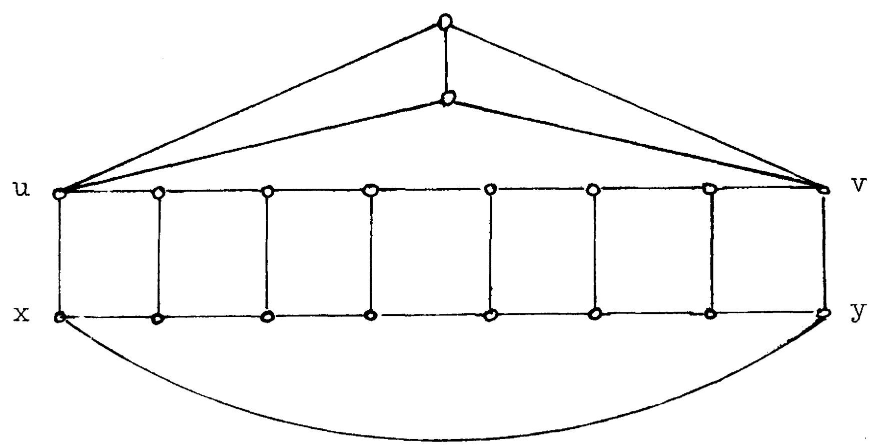
Problem 21.

A unit distance graph is one whose vertices can be represented by points in the Euclidean plane in such a way that adjacent vertices are represented by points having distance one. Obviously, the unit distance graphs can be characterized by forbidden subgraphs. It is easy to show that the unit distance graphs contain neither K_4 nor $K_{2,3}$.

K_4 $K_{2,3}$

Let us denote by $f(n)$ the largest possible number of edges in a unit distance graph with n vertices. One has $f(3) = 3$, $f(4) = 5$, $f(6) = 9$. Obviously, if $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are unit distance graphs then $G_1 \times G_2$ (defined as $(V_1 \times V_2, E)$ with $\{(x_1, x_2), (y_1, y_2)\} \in E$ iff either $x_1 = y_1$, $\{x_2, y_2\} \in E_2$ or else $x_2 = y_2$, $\{x_1, y_1\} \in E_1$) is a unit distance graph. Therefore $f(mn) \geq mf(n) + nf(m)$ and so $f(n) > cn \log n$. On the other hand, the absence of $K_{2,3}$ in unit distance graphs implies quite easily that $f(n) < cn^{3/2}$. Erdős asked whether $f(n) = o(n^{3/2})$.

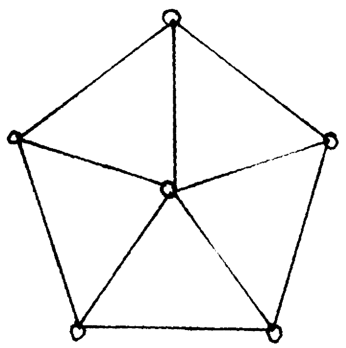
Klarner has observed that, for any representation of the circuit C_4 of length four, the opposite edges must be parallel, and that this fact can be used to construct more forbidden subgraphs. For instance, the graph below is not a unit distance graph: xu is parallel to yv and so dist(u,v) = dist(x,y) = 1.



Nevertheless, there are graphs with n vertices and $cn^{3/2}$ edges which do not contain even a C_4. These can be obtained by assigning vertices to all the points and lines of a projective plane and joining a point-vertex to a line-vertex if and only if the line passes through the point. Thus a geometry of order m gives rise to a (bipartite) graph having total of $2(m^2+m+1)$ vertices and $(m+1)(m^2+m+1)$ edges, containing no C_4 . I suspect that these graphs are not unit distance graphs but have not proved it even for $m = 2$.

Problem 22.

For any finite graph F (without isolated vertices) we define $r(F)$ to be the smallest N such that, for every graph G with N vertices, either G or its complement $\bar{G}$ contains F . Obviously, $r(F) \leq r(K_n)$ where F has n vertices and K_n is the complete graph with n vertices. Hence the existence of $r(F)$ for every F follows from Ramsey's theorem. Erdős conjectures that K_n minimizes $r(F)$ among all n-chromatic graphs F and suggests to test this conjecture on the wheel W_6 .



W_6

We have $r(K_4) = 18$; the unique graph G with 17 vertices such that $K_4 \not\subset G$, $K_4 \not\subset \bar{G}$ is the graph with vertices $\{0,1,\ldots,16\}$, two of them (i,j) being adjacent iff $|i-j|$ is a quadratic residue mod 17 .

Certainly $r(W_6) \geq 17$. Indeed, there is a graph G_0 with eight vertices such that G_0 contains no K_3 and $\bar{G}_0$ contains no K_4 . Replacing each vertex x_i in G_0 by a pair of adjacent vertices x_i^1, x_i^2 and joining x_i^s to x_j^t $(i \neq j)$ if and only if x_i is adjacent to x_j

in G_0, we obtain a graph G with 16 vertices such that $W_6 \not\subset G$, $W_6 \not\subset \bar{G}$.

Can you prove $r(W_6) \geq 18$?

Problem 23.

Among many equivalent formulations of the four-color conjecture, there is a recent one which deserves special interest. Unlike in most other cases, the proof of the equivalence is nontrivial and so it may constitute the first step towards the solution of 4CC. Given any graph $G = (V,E)$ and a set $S \subset V$, we denote by ∂S the number of edges having exactly one endpoint in S. A function $w: V \to \{-2, +2\}$ is called a balanced coloring if

$$\partial S \geq \sum_{x \in S} w(x)$$

for all $S \subset V$. Bondy ("Balanced colourings and the four colour conjecture," to appear in Proc. Amer. Math. Soc.) proved that the four color conjecture is equivalent to the following "balanced coloring conjecture":

Every bridgeless cubic planar graph admits a balanced coloring. (A cubic graph is one where each vertex meets exactly three edges; a bridgeless graph is one which remains connected after the deletion of an arbitrary edge.) One can think of the vertices x with $w(x) = -2$ as being colored blue and those with $w(x) = 2$ as colored red. A balanced coloring of a cubic graph has two simple but interesting properties:

(i) the number of blue vertices equals the number of red ones,

(ii) there is no nonchromatic path with three vertices.

It would be nice to prove that every bridgeless cubic planar graph admits a coloring with properties (i), (ii). Besides, it may be useful to study balanced colorings in the class of all graphs, not necessarily bridgeless cubic planar ones.

Problem 24.

Let $d_1 \geq d_2 \geq \cdots \geq d_n$ be nonnegative integers. What are the necessary and sufficient conditions for the existence of a planar graph with n vertices having degrees $d_1, d_2, \ldots, d_n$? This question appears to be quite deep. When the planarity assumption is dropped, the answer becomes quite simple: the sum of all d_i's must be even and the inequality

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min(k, d_i)$$

satisfied for each $k = 1,2,\ldots,n$ (Erdös and Gallai, "Gráfok elöírt fokú pontokkal," Mat. Lapok 11 (1961), 264-274; also in Harary, Graph Theory, Addison-Wesley, Reading, Mass. 1969). The only additional condition which is known to be necessary in the planar case is

$$\sum_{i=1}^{k} d_i \leq \begin{cases} k(n-1), & 1 \leq k \leq 2, \\ 2n+6k-16, & 3 \leq k \leq \frac{1}{3}(n+4), \\ 3n+3k-12, & \frac{1}{3}(n+4) \leq k \leq n. \end{cases} \tag{1}$$

(Bowen, "On sums of valencies in planar graphs," Canad. Math. Bull. 9 (1966), 111-114; and Chvátal, "Planarity of graphs with given degrees of vertices," Nieuw Archief voor Wiskunde 17 (1969), 47-60).

Unlike the general (purely combinatorial) case, the planar problem exhibits peculiar irregularities. When dealing with the simplest case, $d_1 = d_2 = \cdots = d_n = d$, Euler's formula (resp. (1) with $k = n$)

forces $d \leq 5$ and $n \geq \frac{12}{6-d}$. These conditions, together with the trivial dn = even , turn out to be sufficient apart from two exceptionalcases: $d = 4$, $n = 7$ and $d = 5$, $n = 14$. It would be interesting to go deeper into the structure of the problem and find more subtle additional conditions that would exclude the two exceptional cases.

Problem 25.

A finite family F of finite sets is called an independence system if

$$X \in F ,\, Y \subset X \Rightarrow Y \in F \ .$$

A family of sets is called intersecting if it contains no two disjoint sets. It is called a star if all of its sets have at least one element in common. For example, the family

$$\emptyset , \{1\} , \{2\} , \{3\} , \{1,2\} , \{1,3\} , \{2,3\}$$

is an independence system; its subfamily

$$\{1\} , \{1,2\} , \{1,3\}$$

is a star (and therefore intersecting), the subfamily

$$\{1,2\} , \{1,3\} , \{2,3\}$$

is intersecting but not a star. I offer $10.00 for a proof or a disproof of the following conjecture.

Among the largest intersecting subfamilies of an independence system, there is always a star.

Naturally, I am not going to pay anything for the degenerate "counterexample" $F = \{\{\emptyset\}\}$. Decent as we are, we are interested in independence systems F with $|F| \geq 2$. Without loss of generality, we can assume that the sets in F are sets of positive integers. Then we write $X < Y$ if, and only if, there is a one-to-one mapping

$f: X \to Y$ with $f(t) > t$ for each $t \in X$. I can prove the above conjecture for rather special independence systems, namely, those which satisfy

$$X \in F, Y < X \Rightarrow Y \in F .$$

In its full generality, the problem appears to be quite difficult. I am rather skeptical about the use of counting arguments. It would be interesting to prove the conjecture for independence systems whose maximal sets are lines of a projective plane.

<u>Problem 26</u>.

In the 1930's, Miss Esther Klein asked whether there is a function F(n) such that from any F(n) points in the plane (no three collinear) one can always choose $n+1$ of them which are the vertices of a convex polygon. Erdös and Szekeres ("A combinatorial problem in geometry," Compositio Math. 2 (1935), 463-470) proved $F(n) \leq \binom{2n-2}{n-1}+1$; subsequently Szekeres married Miss Klein. Komlós and I proved a more general result ("Some combinatorial theorems on monotonicity," Canad. Math. Bull. 14 (1971), 151-157) which goes as follows. Let f be an arbitrary real-valued function defined on the edges of a directed graph D which contains no directed cycles. If the vertices of D cannot be colored in $\binom{2n-2}{n\ 1}$ colors then there is a directed path with n edges $e_1, e_2, \ldots, e_n$ such that the sequence $f(e_1), f(e_2), f(e_2), \ldots, f(e_n)$ is monotone. (The specialization is clear: the points in plane can be ordered by their first coordinate, D becomes a transitive tournament and f is the slope function.) The bound $F(n) \leq \binom{2n-2}{n-1}+1$ is not sharp; one has $F(3) = 5$, $F(4) = 9$. I wonder if the last equality carries through to the abstract setting. The abstract version of the problem (more messy but more faithful than the first one) can be set up as follows.

Let f be an arbitrary real-valued function defined on the set

$$\{(i,j): 1 \leq i < j \leq N\} ; \qquad (2)$$

perhaps we should assume

$$\min(f(i,j),f(j,k)) \leq f(i,k) \leq \max(f(i,j),f(j,k))$$

for all $1 \leq i < j < k \leq N$. By an n-gon, we shall mean a pair of sequences $(i_1,i_2,\ldots,i_n)$, $(j_1,j_2,\ldots,j_s)$ with $r,s \geq 2$ such that

$$i_1 < i_2 < \ldots < i_r$$

$$j_1 < j_2 < \ldots < j_s$$

$$f(i_1, i_2) \leq f(i_2, i_3) \leq \ldots \leq f(i_{r-1}, i_r)$$

$$f(j_1,j_2) \geq f(j_2,j_3) \geq . \geq f(j_{s-1},j_s)$$

and $i_1 = j_1$, $i_r = j_s$, $r+s-2 = n$. Let G(n) denote the smallest N such that every function f defined on (2) gives rise to an (n+1)-gon. Obviously, $F(n) \leq G(n)$; the theorem of Komlós and myself shows that $G(n) \leq \binom{2n-2}{n-1}+1$. It is not difficult to show that G(3) = 5 . Indeed, let f be an arbitrary real-valued function defined on $\{(i,j): 1 \leq i < j < 5\}$. Without loss of generality, we can assume $f(1,2) \leq f(2,3)$. Now, let us assume that f gives rise to no 4-gon. Then we necessarily have - step by step -

$f(2,3) > f(3,4)$	because of	(1,2, 3,4) , (1,4)
$f(3,4) < f(4,5)$	because of	(2,5),(2,3,4,5)
$f(2,3) > f(3,5)$	because of	(1,2,3,5), (1,5)
$f(2,4) > f(4,5)$	because of	(2,4,5), (2,3,5)
$f(1,2) < f(2,4)$	because of	(1,5),(1,2,4,5)
$f(1,3) < f(3,4)$	because of	(1,2,4), (1,3,4)

and so finally (1,3,4,5),(1,5) is a 4-gon; contradiction.

Is G(4) = 9 and, more generally, G(n) = F(n) for all n ?

Problem 27. Number Systems

The problem is to determine all sets D of ten real numbers such that every positive or, (if you prefer, nonnegative, or arbitrary) real x can be represented as .

$$x = \sum_{-\infty < k \leq n} a_k 10^k \ , \qquad a_k \epsilon D \ . \qquad (*)$$

It is clear that if $D = \{d_1, \ldots, d_{10}\}$ has this property, then so does $\alpha D = \{\alpha d_1, \ldots, \alpha d_{10}\}$ for any real $\alpha > 0$. This will be implicitly understood below.

If $0 \epsilon D$ and all d_i are ≥ 0 then we can deduce that $D = \{0, 1, \ldots, 9\}$. It is also known that D can be chosen to be $\{x, x+1, x+2, \ldots, x+9\}$ for any x , $-9 < x \leq 1$.

As an instance of the latter, take the symmetric case where $x = -4\frac{1}{2}$. Then

$$D = \{-4\tfrac{1}{2}, -3\tfrac{1}{2}, -2\tfrac{1}{2}, \ldots, 3\tfrac{1}{2}, 4\tfrac{1}{2}\} \ ;$$

the number 0 , for example, now admits the representation

$$\tfrac{1}{2} \cdot (-4\tfrac{1}{2})(-4\tfrac{1}{2}) \ \ldots \ =$$

$$\tfrac{1}{2} + \sum_{n \geq 1} (-4\tfrac{1}{2}) 10^{-n} = \tfrac{1}{2} + \frac{-4\frac{1}{2}}{9} \ .$$

An interesting feature of this system, noted by Claude Shannon, is that rounding-off is equivalent to truncation.

Another interesting case is when x = +1 . The following example shows how a positive number written in ordinary decimal notation can be transformed into the new system:

LO 0 2 3 0 4 5	decimal #
.0 1 1 1 1 1 1 -	
.9 9 1 1 9 3 4	
.0 1 1 1 1 1 1 +	add back without carries
.9 10 2 2 10 4 5	representation in new system

It can easily be shown that this method works in general. For other x replace the 1's in the above example by x's .

Finally, it should be remarked that we require D to have ten elements, since Ron Graham proved that if it had fewer, the set of numbers representable as (*) is a set of measure zero. It is not difficult, however, to construct sets D with as few as three elements which give rise via (*) to a set of numbers dense on the positive real axis, even if we stipulate that $a_k = 0$ for all $k < 0$. Such an example is $D = \{0,1,\alpha\}$ where $\alpha = -\sum_{n>1} 10^{-n^2}$.

Problem 28. Sorting by Deques

The problem is to investigate the permutations that can be obtained from a general deque starting with the permutation 1,2,...,n as input. For example, one would like to know a simple test for deciding whether or not a given permutation can be so obtained. Another interesting question is to count the number of permutations thus obtainable, by recurrence, generating function, and/or asymptotic formula.

For a definition of deque, related problems, and a description of techniques that have been found useful in attacking them, see Section 2.2.1 and the following problems in D. Knuth "Fundamental Algorithms" (p. 234).

Note: It was remarked that $\sqrt[n]{a_n}$ exists. Possible "canonical" sequences of the four-operations

insert next input at left
insert next input at right
output the leftmost element
output the rightmost element

(with exactly one canonical sequence per obtainable permutation) are being investigated. Vaughan Pratt found that there exist four special permutations of every odd length ≥ 5 such that a given permutation is obtainable if and only if it contains none of these special permutations.

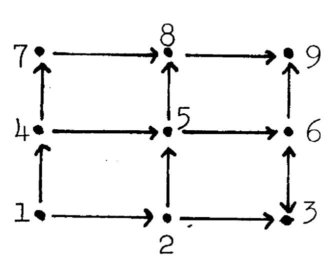
Problem 29. Dragon Curve

The "dragon design" is obtained by repeatedly folding a sheet of paper in one direction and then unfolding it so that each of the creases forms a fixed angle θ. For a more precise definition and many interesting properties of this curve, see the article: 'Number representations and dragon curves' by Chandler Davis and D. E. Knuth, J. Recreational Math. 3 (1970), 66-81, 133-149.

It is experimentally observed that there is a greatest angle θ_0, between $90°$ and $100°$, such that for $\theta_0 < \theta \leq 180°$ the dragon curve does not have any self-intersections. The problem is to determine θ_0. Some experimentation indicates that the "crucial points", where self-intersection is likely, occur at the $\delta(5 \cdot 2^n)$ and $\delta(11 \cdot 2^n)$.

Problem 30. Posets and Permutations

We consider a partially ordered set P, such as the one drawn below.



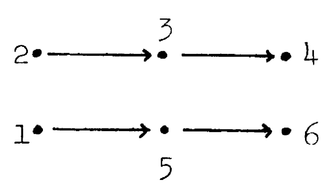
where we use arrows to indicate the ordering relations.

We also label the points of P in a particular way, consistent with their relative order with the integers $1,\ldots,n$; that is, we require that if $i \to j$, then $i \leq j$.

The problem now is to find an efficient algorithm for determining the permutation of the labels $1,2,\ldots,n$ which gives a labelling still consistent with the partial order of P , but which has a <u>maximal number of inversions</u>. (An inversion of a permutation $p_1 \ldots p_n$ is two numbers $\ldots p_i \ldots p_j \ldots$ with $i < j$ but $p_i > p_j$.)

For the above example we easily find that the desired permutation is uniquely $p_1 \ldots p_9 = 147258369$.

Another way to veiw this problem (obtaining the inverse permutation, which has the same number of inversions) is to consider any method of removing the points of the graph one at a time, never removing a point until its predecessors have been removed. The idea is to maximize the number of inversions in the output. The example



shows that it is not sufficient simply to remove the largest possible element first (1, 5, 6, 2, 3, 4 has more inversions than 2, 3, 4, 1, 5, 6).

[This problem was inspired by computer sorting.]

<u>Problem 31</u>. Partitions into subintervals

It is a known result that

(1) If Θ is irrational then the numbers $\{n\Theta\} = n\Theta \bmod 1$, $n = 0,1,2,\ldots$ are dense in $[0,1]$ and in fact evenly distributed (H. Weyl) and that

(2) having introduced the points $\{k\Theta\}$, $k = 0,1,\ldots,n-1$, the next point $\{n\Theta\}$ will then fall in the middle of one of the largest remaining intervals. It is furthermore true that at each stage we have at most <u>three</u> distinct sizes of subintervals (V. T. Sós).

We are concerned here with proving the following generalization of the above. Let Θ and $\alpha_1,\ldots,\alpha_{n-1}$ be any reals, and let $k_0,\ldots,k_{n-1}$ be any positive integers. Then R. Graham has conjectured that the numbers

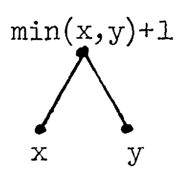
$$\begin{array}{llll} \{\Theta\} & , \{2\Theta\} & \ldots & \{k_0\Theta\} \\ \{\Theta+\alpha_1\} & , \{2\Theta+\alpha_1\} & \ldots & \{k_1\Theta+\alpha_1\} \\ & \ldots & & \\ \{\Theta+\alpha_{n-1}\} & , \{2\Theta+\alpha_{n-1}\} & \ldots & \{k_{n-1}\Theta+\alpha_{n-1}\} \end{array}$$

will subdivide $[0,1]$ into subintervals of at most $3n$ distinct sizes.

<u>Problem 32</u>. Counting leftist trees

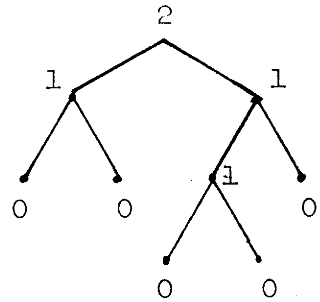
In what follows we consider only binary trees (with left distinguished from right); a leaf node has no sons, and a nonleaf node has 2 sons. We assign to each vertex of a binary tree a weight by proceeding as follows. We first assign 0 to all leaves. Then we climb up the tree by the rule that if the two sons of a node have been given weights x and y , the node itself receives the weight $\min(x,y)+1$.

Now such a weighted binary tree is called <u>leftist</u> if for each node, the son situated to the left has a weight at least as great as the son situated to the right. The constraints can be symbolized by



$x \geq y$.

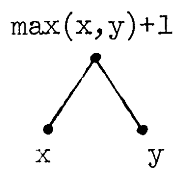
The picture below shows a leftist tree.



Intuitively leftist trees are characterized by the fact that to go from any node to the nearest leaf it suffices to proceed always to the right. The problem then is to compute the asymptotic growth of a_n , the number of leftist trees with n leaves.

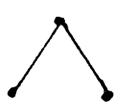
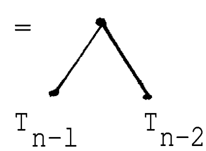
Problem 33. Counting balanced trees

A similar problem can be asked about the situation where we modify the above as follows. If the two sons of a node have weights x and y , then the node itself has weight $\underline{\max}(x,y)+1$. A binary tree so weighted is called balanced if the weights of the two sons of any node differ by at most one. Thus, the constraints now are



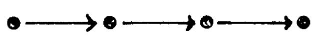
$$|x-y| \leq 1 \ .$$

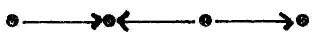
It is known that the "Fibonacci trees" are the balanced trees with the minimal number of nodes for a given height h . (The height of a tree is the weight at the root.) The Fibonacci trees T_n are given by

$T_0 = \bullet$, $T_1 =$  and $T_n =$ 

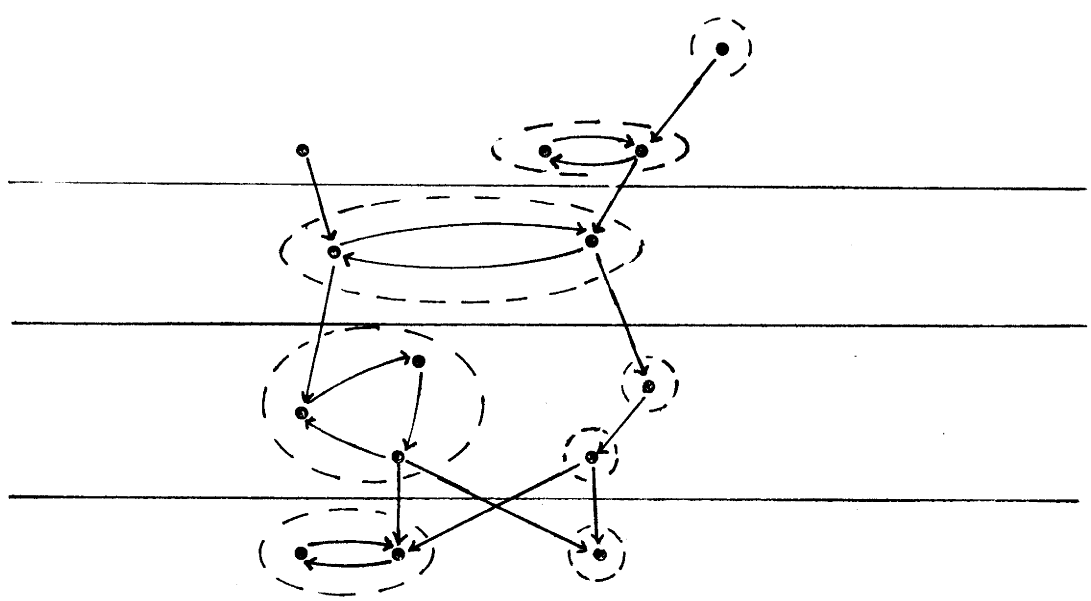
Also the number of distinct balanced trees with height h is known to grow as C^{2^h} where C is a positive (unknown) constant. (Balanced trees are sometimes known by the less desirable term "AVL trees".) What is the asymptotic number of balanced trees with n leaves?

Problem 34. A basic equivalence relation on graphs.

Find an efficient algorithm for computing the weak components of a directed graph. (Efficient in the sense that it takes O(max(vertices, edges)) steps.) The weak components are the finest partition such that, if all nodes in each component were collapsed together, the graph would be linear and the ordering would be a linear ordering, i.e.,  is acceptable but

 is not.

Example:



The dotted lines indicate strong components and the horizontal lines indicate weak components.

Formally, x and y are in the same weak component if they are in the same strong component or if one can get from x to y and back by a sequence of "non-path" steps. Nodes a and b are said to be connected by a non-path if there is no path from a to b .

Reference: Graham, Knuth, and Motzkin, Discrete Math. 2 (1972), 17-30.

Problem 35. Greatest common substrings.

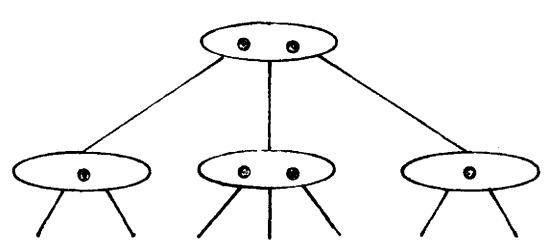
It is possible to find the longest common subsequence of two sequences of a's and b's in a time proportional to the product of their lengths. Can one do better? Note: aba is a subsequence of aabbbba .

Problem 36. Permutations as substrings. (Due to R. M. Karp.)

What is the shortest string of $\{1,2,\ldots,n\}$ containing all permutations on n elements as subsequences? (For n = 3 , 1 2 1 3 1 2 1 ; for n = 4 , 1 2 3 4 1 2 3 1 4 3 2 1 ; for n = 5 , M. Newey claims the shortest has length 19 .)

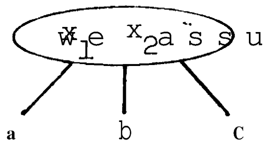
Problem 37. Random growth of 3-2 trees.

Analyze the probability that various numbers of splits will occur during random insertion into (3,2) trees as the trees become large. A (3,2) tree is a tree in which every node may be a leaf or else it has 2 or 3 sons. One may write (3,2) trees as follows:

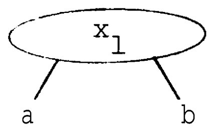


The dots indicate data items, one or two in each cell. Downward arcs leaving a node indicate possible results of comparing another data item with those in a node. For example, a node containing only one item x can yield only two results for new incoming y , either $y < x$ or $y > x$.

In a node of form



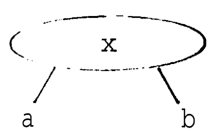
we assume $x_1 < x_2$. Also, all items y in subtree a are smaller than x_1 ; in b they are between x_1 and x_2 ; and in c they are larger than x_2 . Similarly in



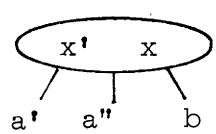
all items in subtree a are smaller than x_1 and all items in subtree b are larger than x_1 .

Now, we consider only (3,2) trees with this ordering on their items and whose leaves are all at the same level. We insert by introducing items randomly at the positions of the leaves. (The leaves represent equiprobable gaps between existing items.)

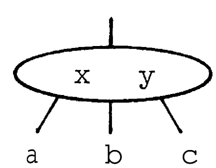
We <u>insert</u> by a sequence of operations local to various nodes along the path to the root. For example, to insert x' in place of leaf a in



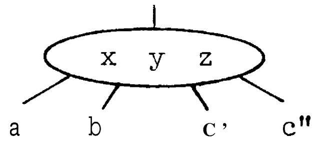
we get



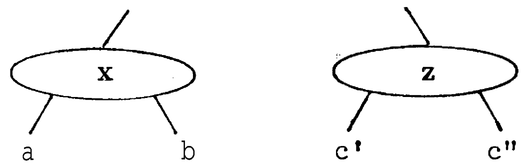
to insert z in place of leaf c in



gives

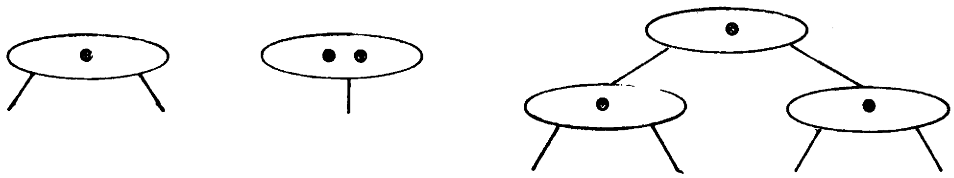


which splits into

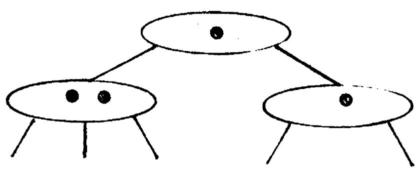


and y is inserted in the node above in the same way. (If there was no node above, we place y in a new root node.)

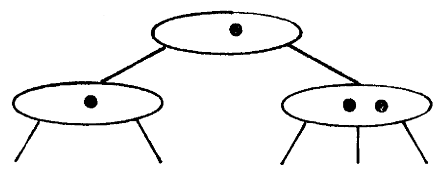
Thus the first three steps in the growth of a 3-2 tree are always



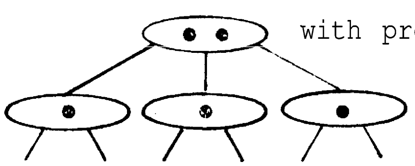
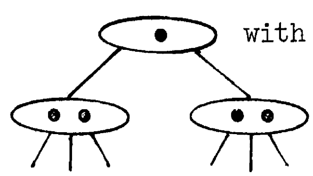
and the fourth step is either



or

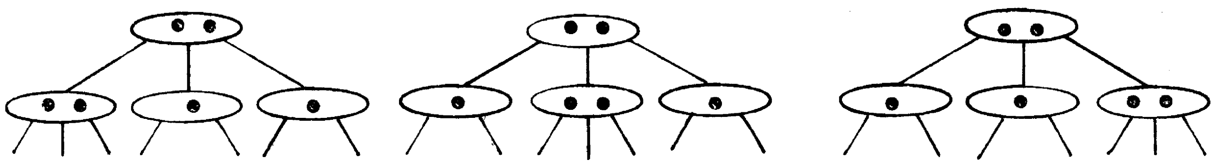
 .

By symmetry, we may choose the former. Now the fifth step yields

 with probability $\frac{3}{5}$,  with probability $\frac{2}{5}$.

The question is, how many splittings will occur on the n-th random step, on the average, and how many tree nodes will there be? This is one of the few important "basic" algorithms that hasn't been analyzed yet.

Note: After six steps the tree is either



and it appears that all are equivalent with respect to further operations. Or are they?