Proceedings of Symposia in Pure Mathematics
Volume 34, 1979

THE PROBABILITY OF AN INTERVAL GRAPH, AND WHY IT MATTERS

Joel E. Cohen, János Komlós, and Thomas Mueller[1]

ABSTRACT. An interval graph is the intersection graph of a family of intervals of the real line. Interval graphs have been used for inference in several sciences, including archeology, ecology, genetics, and psychology. In these applications, the strength of inference depends on the probability that a random graph is an interval graph. Using a definition of a random graph due to Erdős and Rényi [8], we obtain exact probabilities and asymptotic and Monte Carlo estimates of the probabilities, for varying numbers of vertices and edges. We also obtain the asymptotic probability that a random graph is a circular arc graph, which is the intersection graph of a family of arcs on the circle. Some mathematical questions arising in scientific inference remain unanswered.

1. INTRODUCTION. Since interval graphs were introduced 21 years ago, they have been widely used in science and mathematics. In Section 2, we review the definition and characterizations of interval graphs. In Section 3, we describe applications of interval graphs. Some of these applications lead naturally to the question, what is the probability that a random graph is an interval graph? In Section 4, we define a random graph. Using exact analysis, asymptotic theory, and Monte Carlo simulation, we estimate the probability that a random graph is an interval graph. We also find the asymptotic probability that a random graph is a circular arc graph. Finally, in Section 5, we present some problems which remain unsolved.

2. DEFINITION AND CHARACTERIZATIONS. Let G be a graph with v labelled vertices $a_1, \ldots, a_v$ and E unlabelled, undirected edges $e_1, \ldots, e_E$. Each edge may be written as an unordered pair of vertices $(a_i, a_j) = (a_j, a_i)$, $i \neq j$, and the edge is said to connect the vertices a_i and a_j.

AMS(MOS) subject classifications (1970). Primary 05C30, 60C05; Secondary 55A15, 62E25, 92A05, 92A10.

[1] J.E.C. and T.M. were supported in part by grant DEB(BMS) 74-13276 of the U.S. National Science Foundation.

Loops and multiple edges are excluded. G is an interval graph when there is a collection $S_1, \dots, S_v$ of open, closed, or mixed intervals of the real line such that there is an edge between a_i and a_j, $i \neq j$, if and only if S_i and S_j overlap, that is $S_i \cap S_j \neq \emptyset$. Thus G is an interval graph if and only if G is the intersection graph of some family of intervals of the line.

H is a <u>subgraph</u> of G if the vertices of H are a subset of the vertices of G and the edges of H are a subset of the edges of G.

A subgraph H of G is an <u>induced</u> subgraph of G if there is an edge between two vertices of H whenever there is an edge between those two vertices in G.

Lekkerkerker and Boland [14] characterize interval graphs when v is finite in two ways.

First, G is an interval graph if and only if it contains no induced subgraph of the form pictured in Figure 1. (The graph I in Figure 1

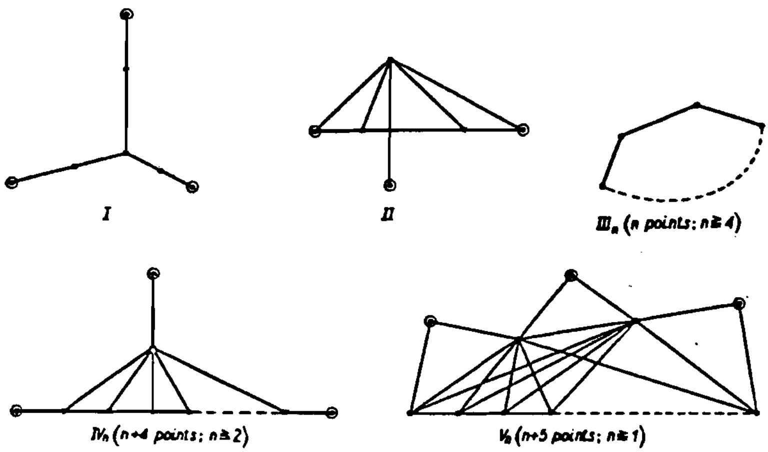

Fig. 1. The 5 graphs (I, II) or classes of graphs (III_n, IV_n, V_n) forbidden as induced subgraphs according to the characterization of Lekkerkerker and Boland [14].

plays a special role in the asymptotic probability that a random graph is an interval graph.)

A reader who is interested only in new results could now proceed directly to Section 4.

The second characterization requires three additional definitions. A path is an induced subgraph of G defined by a sequence of not necessarily distinct vertices $a_1, \dots, a_k$ such that (a_i, a_{i+1}) is an edge in

G, i = 1, . . . , k-1. An irreducible path is a path in which $a_i \neq a_j$ for $i \neq j$ and no vertex in the path is joined to any vertex in the path other than those immediately preceding or following it, when such exist. A cycle is any path of the form $a_1, \ldots, a_k, a_1$.

Then a graph G is an interval graph if and only if (a) G contains no irreducible cycle with more than three vertices; and (b) for any three distinct vertices of G no two of which are joined by an edge, at least one of the vertices is connected to every path between the other two vertices.

Gilmore and Hoffman [10] characterize interval graphs whether v is finite or infinite. Some further definitions are required. By a cycle, Gilmore and Hoffman mean any finite sequence of vertices $a_1, \ldots, a_k$ of G such that all of the edges (a_i, a_{i+1}), $1 \leq i \leq k-1$ and (a_k, a_1) are in G and such that if one traces out the given sequence of vertices, one does not travel along the same edge twice in the same direction. This definition does not exclude visiting a vertex twice or traveling along an edge once in each direction. If $a_1, \ldots, a_k$ is the given sequence of vertices in a cycle, the cycle is odd if k is odd. A triangular chord of this cycle is any one of the edges (a_i, a_{i+2}), $1 \leq i \leq k-2$, or (a_{k-1}, a_1) or (a_k, a_2). If G is any graph, the complementary graph G^c has the same vertices as G but has an edge connecting two vertices if and only if that edge does not occur in G.

With these definitions, a graph G is an interval graph if and only if every quadrilateral in G has a diagonal and every odd cycle in G^c has a triangular chord.

Fulkerson and Gross [9] give a matrix-theoretic characterization of finite interval graphs. This characterization has been the basis of most machine computation involving interval graphs, including the Monte Carlo results described below.

Again, some definitions are needed. A clique of a graph G is a subgraph of G which is complete, that is, in which every pair of vertices is joined by an edge. If the family of subgraphs of G which are cliques is partially ordered by set inclusion, the maximal elements of the partial ordering are called dominant cliques. The dominant clique versus vertex matrix of G is a matrix with one row for each dominant clique and one column for each vertex of G. The element in the ith row and jth column is 1 if the jth vertex is a vertex of the ith dominant clique, and is 0 otherwise. Since two vertices of G are joined by an edge if and only if they are both vertices in some dominant clique, the dominant clique versus vertex matrix specifies G uniquely, and vice versa. A (0,1) matrix has the consecutive 1's property if and only if there is some permutation of the rows after which the 1's in each column occur consecutively. Thus a matrix has the consecutive 1's property if some (possibly null) reordering of the rows results in no two 1's in a given column being separated by a 0. Fulkerson and Gross give an explicit algorithm for testing whether a (0,1) matrix has the consecutive 1's property.

Then G is an interval graph if and only if the dominant clique versus vertix matrix of G has the consecutive 1's property.

3. APPLICATIONS. According to Berge [2], Hajós [11] originated interval graphs with the following, apparently hypothetical, problem: Suppose each student at a university visits the library for exactly one interval during the day, and reports at the end of the day the other students who were there while he was. If each vertex of the graph G corresponds to one student, and two vertices are joined by an edge if and only if the two corresponding students were in the library simultaneously, then G is an interval graph. The problem is to characterize which graphs could or could not arise from such a process.

Independently of Hajós, Benzer [1] posed the formally identical problem of deciding whether, at the level of genetic fine structure, mutations in the rII region of the virus, bacteriophage T4, are linked together in a linear structure. Recombination experiments can indicate whether any two mutant regions overlap. If each vertex of a graph G corresponds to one mutant and two vertices are joined by an edge if and only if the two corresponding mutant regions overlap, then the genetic fine structure is compatible with a linear order if and only if G is an interval graph. For complete overlap data on 19 mutants and incomplete data on a total of 145 mutants of T4, a linear model is adequate. Electron microscopic and autoradiographic pictures of the genetic material of T4 confirm its physical linearity.

Benzer [1] investigates the probability that randomly generated data would be compatible with a linear one-dimensional structure and obtains an upper bound. He raises the important question of whether his data could discriminate a linear one-dimensional structure from a plausible alternative, e.g. branched, structure. This brilliant paper remains worth reading as a source of mathematical problems and biological ideas.

Roberts [16] reviews applications of interval graphs to the psychological theory of preference, archeology, developmental psychology, and the timing of traffic lights.

In a preference experiment, a person is asked whether he prefers one wine, for example, to another or is indifferent between them. If each wine corresponds to a vertex of a graph G, and two vertices are joined if and only if the individual is indifferent between the corresponding wines, then G is an interval graph if and only if there is some linear ordering along which each wine can be assigned to an interval. See also the related work of Hubert [13].

In archeology, the problem is to establish a seriation, or chronological order, for graves or strata on the basis of the artifacts found in them. Each artifact is classified into one of a finite number of styles. It is supposed that each style of artifact was put in graves during one time interval. It is also supposed that if artifacts of two different styles are found together in a grave, then the time intervals during which they were made overlapped. The graph G with vertices corresponding to styles, and edges between every pair of styles found in a common grave, is an interval graph if and only if some assignment of a time interval to each style is possible.

In the application of interval graphs to developmental psychology, it is assumed that traits arise in all children in a single sequence with temporal overlapping among traits. Children are assessed for the simultaneous presence of various traits. If the vertices of a graph G correspond to the traits and are joined by an edge if the corresponding traits appear simultaneously in some child, then G is an interval graph if and only if it is possible to assign some time interval to each trait.

In timing traffic lights, the problem is: given a graph G with vertices corresponding to the streams of traffic at an intersection, and edges between two vertices if the corresponding streams could safely be permitted to flow at the same time, G is an interval graph if and only if each traffic stream is permitted to flow during a single time interval in one cycle of the traffic lights.

Though many of these substantive applications require special extensions and refinements of theory, discussed by Roberts [16], the concept of interval graphs captures the lion's share of the formal structure.

Booth and Lueker [4] and Booth [3] review applications of interval graphs to the assignment of records to tracks on a computer's disk memory for efficient information retrieval, Gaussian elimination schemes for sparse symmetric positive definite matrices (see also Tarjan [17]), and table-driven parsing. Interval graphs have been used to determine the feasibility of proposed schedules of tasks in the building of large ships (Alan J. Hoffman, personal communication, 20 March 1978).

An application of interval graphs to ecology leads naturally to the problem of determining the probability that a random graph is an interval graph (Cohen [5], [6], [7]).

In ecology, a food web W is a directed graph that tells which kinds of organisms nourish which other kinds of organisms in a community of species. Each labelled vertex in W corresponds to a kind of organism. Each arrow or directed edge (a_i, a_j) from vertex i to vertex j specifies a flow of energy or biomass (food, in short) from the ith kind of organism to the jth kind of organism.

Another description of communities of species represents each kind of organism by a multidimensional hypervolume in a hypothetical ecological niche space. In niche space, each dimension corresponds to some environmental variable or some variable characterizing the food consumed by the organisms. The multidimensional hypervolume associated with each kind of organism is called its niche. The projection of ecological niche space onto the dimensions characterizing the food consumed is called trophic niche space, and the same projection of a niche is called the trophic niche.

An elementary question about niche space is: what is the minimum dimensionality of a niche space necessary to represent, or to describe completely, the overlaps among observed niches? This question remains unanswered for niche space in general. The use of interval graphs gives a partial answer for trophic niche space.

We define the predators in a food web graph W as the set of all kinds of organisms which consume some kind of organism in W, or more formally as the set of vertices a_j such that (a_i, a_j) is an arrow in W, for some a_i. Cannibalism is not excluded. Prey are defined as the set of all kinds of organisms that are consumed by some kind of organism, i.e., all a_i such that (a_i, a_j) is an arrow in W, for some vertex a_j. The trophic niche overlap graph G(W) is defined as an undirected graph in which the vertices are the predators in a food web graph W. Two predators are joined by an undirected edge when there is some kind of prey that both predators eat. That is, $(a_j, a_k) = (a_k, a_j)$ is an edge in G(W) if and only if there exists some a_i in W such that both (a_i, a_j) and (a_i, a_k) are arrows in W.

If the trophic niche of a kind of organism is a connected region in trophic niche space, then it is possible for trophic niche overlaps to be described in a one-dimensional space if and only if the trophic niche overlap graph G(W) is an interval graph.

An analysis of the niche overlap graphs of 30 real food webs suggests that a niche space of dimension one suffices to describe the trophic niche overlaps implied by real food webs in single habitats.

To determine whether chance alone might explain why observed trophic niche overlap graphs are interval graphs, it is necessary to give precise meaning to "chance alone." Seven probabilistic models are described and compared with data in Cohen [7]. Six of these are models of food web graphs W; associated with each such model is an induced probability distribution on G(W). One model assigns a probability distribution directly to niche overlap graphs G(W). In the next section we shall study the probability that a random graph G is an interval graph under this model.

We brought the problem of enumerating interval graphs to the attention of Frank Harary (Harary and Palmer [12]), but to our knowledge there have been no results other than those we present here.

The probabilities we calculate are relevant not only to ecology, but to any scientific application where interval graphs are used for inference about the dimensionality of some hypothetical underlying structure.

4. PROBABILITY THAT A RANDOM GRAPH IS AN INTERVAL GRAPH. The model of a random graph which we shall study is that defined by Erdős and Rényi [8]. Their systematic investigation of the properties and structures of random graphs is a basic source of results and methods in this area. They define a random graph with v labelled vertices and E edges as one in which the E edges are chosen randomly (without replacement) among the $\binom{v}{2}$ possible edges so that all $C_{v,E}$ possible graphs are equiprobable, where

$$C_{v,E} = \binom{\binom{v}{2}}{E} .$$

We also use an equivalent construction of a random graph with E edges on v labelled vertices. If k edges are already chosen, choose one of the remaining $\binom{v}{2} - k$ edges, each with equal probability $1/[\binom{v}{2}-k]$, $k = 0, 1, \ldots, E-1$.

If A is a property which a graph either possesses or does not possess, and $A_{v,E}$ is the number of graphs on v labelled vertices and E edges which possess the property, then the probability $P_{v,E}(A)$ that a random graph has property A is defined as $A_{v,E}/C_{v,E}$.

Let T be the event that a graph is an interval graph and -T be the event that a graph is not an interval graph. We first obtain the exact probability $P_{v,E}(T)$ that a random graph with v labelled vertices and E edges is an interval graph for a limited range of values of v and E. Then we develop some asymptotic results. Finally, we attempt to connect the two kinds of results with Monte Carlo simulations.

THEOREM 1. All graphs with $E < 4$ or $E = \binom{v}{2} - k$, $k = 0$ or 1, are interval. For $v > 3$,

(i) $P_{v,4}(T) = 1 - 48/[(v^2-v-4)(v+2)(v+1)]$

(ii) $P_{v,5}(T) = 1 - 240(v-4)(v+(23/5))/[(v^2-v-4)(v+2)(v+1)(v^2-v-8)]$

(iii) $P_{v,6}(T) = 1 - (v-4)(720v^3+11424v^2-122064v+261600)/[(v+1)(v+2) \cdot(v^2-v-4)(v^2-v-8)(v^2-v-10)]$

(iv) $P_{v,\binom{v}{2}-3}(T) = (32v-88)/[(v+1)(v^2-v-4)]$

(v) $P_{v,\binom{v}{2}-2}(T) = 4/(v+1)$

PROOF. (i) A graph with 4 edges can fail to be interval if and only if the 4 edges form a quadrilateral. We can think of placing a quadrilateral on v vertices as first placing a complete subgraph (with 6 edges) on 4 vertices (in $\binom{v}{4}$ possible ways) and then deleting from each such complete subgraph 2 edges so that the remaining 4 form a quadrilateral (in 3 possible ways). The total number of possible ways of placing the 4 edges is $C_{v,4}$. Thus $P_{v,4}(-T) = 3\binom{v}{4}/C_{v,4}$, from which the given $P_{v,4}(T)$ follows.

(ii) Figure 2 gives the 3 ways in which 5 edges may be placed on 5 or 6 vertices to form a graph which is not an interval graph. For any

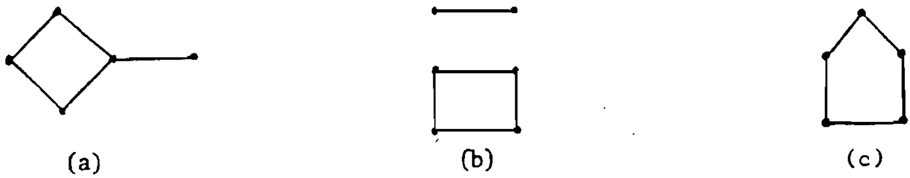

Fig. 2. The 3 ways 5 edges may be placed on 5 or 6 vertices to form a graph which is not an interval graph.

set of 5 labelled vertices, pattern (a) can be chosen in $5 \times 4 \times 3 = 60$ ways. For any set of 6 labelled vertices, pattern (b) can be chosen in $15 \times 3 = 45$ ways. For any set of 5 labelled vertices, pattern (c) can be chosen in 12 possible ways. Thus $P_{v,5}(-T) = [60\binom{v}{5} + 12\binom{v}{5} + 45\binom{v}{6}]/C_{v,5}$.

(iii) Figure 3 gives 11 of the 12 ways 6 edges may be placed on 5 or more vertices to form a graph which is not an interval graph. For each pattern with n labelled vertices, $n = 5, 6, 7, 8$, there are $\binom{v}{n}$ choices of the vertices. Elementary counting shows that the number of ways of assigning 6 edges to form each pattern is: (a) $60 = 5 \times 4 \times 3$; (b) $10 = \binom{5}{2}$; (c) $60 = 6!/(2\times6)$; (d) $360 = 12 \times 5 \times 6$; (e) $360 = \binom{6}{2} \times 2 \times 3 \times 4$; (f) $180 = 6!/(2\times2)$; (g, which is identical to graph I in Figure 1) 840 (see

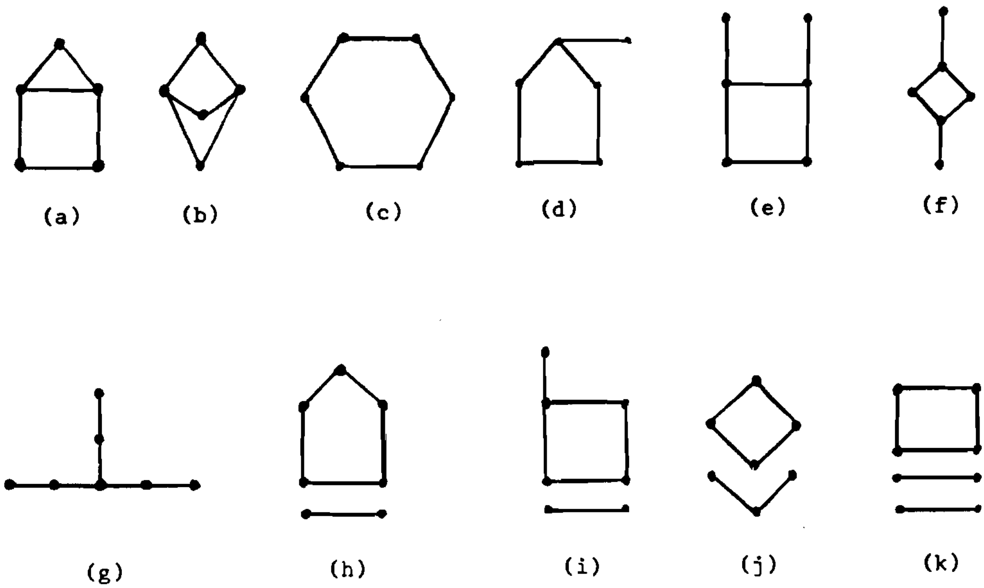

Fig. 3. Eleven of the 12 ways 6 edges may be placed on 5 or more vertices to form a graph which is not an interval graph.

proof of Theorem 3 below); (h) $252 = 12\times\binom{7}{2}$; (i) $1260 = 60\times\binom{7}{2}$; (j) $315 = \binom{7}{4}\times3\times3$; (k) $630 = \binom{8}{4}\times3\times3$. The twelfth pattern, identical to graph IV_2 in Figure 1, may be chosen in $120 = \binom{6}{3}\times3!$ ways. Multiplying each of these numbers by the appropriate $\binom{v}{n}$, summing, and dividing by $C_{v,6}$ gives $P_{v,6}(-T)$, and collecting powers of v yields (iii). The final result shows correctly that $P_{4,6}(T) = 1$.

(iv) A graph with $\binom{v}{2} - 3$ edges may be viewed as a complete graph from which 3 edges have been selected for omission. If these 3 edges connect a total of 3 vertices, there is exactly one way the 3 edges can be chosen for each set of 3 vertices; each graph obtained from a complete graph by omitting such a triangle is an interval graph. If the 3 omitted edges involve exactly 4 points, the 3 omitted edges must constitute a pattern of the form (a_1, a_2), (a_1, a_3), (a_1, a_4) (which can be chosen in $16 = \binom{6}{3} - 4$ ways) in order for the result to be an interval graph. Thus $P_{v,\binom{v}{2}-3}(T) = [16\binom{v}{4} + \binom{v}{3}]/C_{v,\binom{v}{2}-3}$.

(v) To obtain an interval graph by omitting 2 edges from a complete graph, the 2 edges must form a pattern like (a_1, a_2), (a_1, a_3). For each set of 3 labelled vertices there are 3 ways of choosing such a pattern. Thus $P_{v,\binom{v}{2}-2}(T) = 3\binom{v}{3}/C_{v,\binom{v}{2}-2} = 4/(v+1)$. This proves Theorem 1.

Theorem 1 covers all possible numbers E of edges only for graphs with up to $v = 5$ vertices.

To carry out the asymptotic analysis, we recall some more concepts and results from Erdős and Rényi [8].

$f(v)$ is called a threshold function for the property A if for any $\varepsilon > 0$ there are positive numbers δ, Δ, and v_0 such that for $v > v_0$, $E \le \delta f(v)$ we have $P_{v,E}(A) < \varepsilon$, and for $E \ge \Delta f(v)$ we have $P_{v,E}(A) > 1 - \varepsilon$.

If a graph G has v vertices and E edges, the degree d of the graph is $2E/v$, which is the average degree of the vertices of G. G is said to be balanced if no subgraph H of G has a larger degree than G itself.

THEOREM A (Erdős and Rényi [8, p. 23]). Let $v \ge 2$ and E be positive integers. Let $B_{v,E}$ denote an arbitrary non-empty class of connected balanced graphs with v vertices and E edges. The threshold function for the property that a random graph on n labelled vertices contains at least one subgraph isomorphic with some element of $B_{v,E}$ is $n^{2-v/E} = n^{2-2/d}$, where d is the degree of each graph in $B_{v,E}$.

Here we give without proof a slight generalization of Theorem A which can be obtained by using the inclusion-exclusion formula and calculations along the lines of [8].

We say that a graph G is strongly balanced if any subgraph of G has a smaller degree than G.

Let B be an arbitrary finite class of strongly balanced graphs $G_1, \ldots, G_m$ all having the same degree d. Let the number of (labelled) vertices of G_i be v_i and the number of edges be E_i. Let B_i denote the number of graphs with v_i labelled vertices which are isomorphic to G_i.

THEOREM 2. Let A_k denote the event that a random graph contains exactly k subgraphs each isomorphic to some element of B. Assume that, as the number v of labelled vertices of a random graph is increased, the number $E(v)$ of edges is also increased so that $\lim_{v\to\infty} E(v)/v^{2-2/d} = c$. Here d is the degree of each graph in B, while $E(v)$ and v refer to the edges and vertices of the random graph. Then $P_{v,E(v)}(A_k)$, $k = 0, 1, 2, \ldots$, is, asymptotically, a Poisson distribution

$$P_{v,E(v)}(A_k) \sim \lambda^k e^{-\lambda}/k!$$

where we define $p = E(v)/\binom{v}{2}$ and

$$\lambda = \Sigma_{i=1}^{m} \binom{v}{v_i} B_i p^{E_i} (1-p)^{\binom{v_i}{2}-E_i} . \qquad (1)$$

In particular,

$$P_{v,E(v)}(A_0) \sim e^{-\lambda} .$$

λ is the asymptotic expected number of subgraphs isomorphic to some graph G_i in B. In the alternative model (called $\Gamma^{**}_{v,E}$ in [8, p. 20]) of a random graph in which each edge is chosen independently according to a Bernoulli trial with probability p of success, λ is the precise expected number of subgraphs isomorphic to some graph G_i in B. As v increases, $\lambda \to \lambda^*$ where

$$\lambda^* = \Sigma_{i=1}^{m} B_i (2c)^{E_i} / v_i! , \qquad (2)$$

and we have the precise statement

$$\lim_{v\to\infty} P_{v,E(v)}(A_k) = (\lambda^*)^k e^{-\lambda^*}/k! .$$

THEOREM 3. Assume that $\lim_{v\to\infty} E(v)/v^{5/6} = c$. Then $\lim P_{v,E(v)}(T) = e^{-\lambda^*}$, where $\lambda^* = 32c^6/3$. Furthermore, $P_{v,E}(T) \sim e^{-\lambda}$ for large v and E as long as E^6/v^5 is not too large, where $\lambda = \binom{v}{7}(7!/6)p^6(1-p)^{15} \sim 32E^6/(3v^5)$, $p = E/\binom{v}{2}$.

A circular arc graph is the intersection graph of a family of arcs on a circle. Thus the definition of a circular arc graph is the same as that of an interval graph except that "intervals of the real line" is replaced by "arcs of a circle." Tucker [18] has given a matrix characterization of circular arc graphs.

COROLLARY. Let S be the event that a random graph is a circular arc graph. Then Theorem 3 is true when T is replaced by S.

To prove Theorem 3 from Theorem 2, we require a lemma which is a corollary of the first characterization of interval graphs by Lekkerkerker and Boland [14] given in Section 2, but which is also easy to check directly.

LEMMA. If a graph G contains as an induced subgraph the graph I pictured in Figure 1, then G is not an interval graph. If G is a forest (disjoint union of trees) and does not contain I as a subgraph, then G is an interval graph.

PROOF. The first part is obvious since I is not an interval graph. For the second part, consider a tree not containing I. Then it must be of the form of a caterpillar exemplified by Figure 4. To see this, simply lay

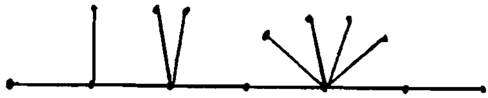

Fig. 4. An example of a caterpillar, which is a tree not containing a forbidden subgraph of form I.

down the tree along its longest path. Now it is easy to construct a set of intervals with intersection graph corresponding to any caterpillar. The construction for the example in Figure 4 is given in Figure 5. This proves the lemma.

Fig. 5. A set of intervals of the real line for which the intersection graph is the caterpillar in Fig. 4.

PROOF OF THEOREM 3. Let T_1 be the event that a graph does not contain a subgraph of the form of I in Figure 1. Let T_2 be the event that a graph is a forest which does not contain a subgraph of the form I. As before let T denote the event that a graph is an interval graph. By the lemma, T_2 implies T which in turn implies T_1. We shall calculate the asymptotic probability of T_1 and show that asymptotically the probability of T_2 is the same. It follows that the asymptotic probability of T equals the asymptotic probability of T_1.

Apply Theorem 2 to the one-element family of graphs $B = \{G_1\}$ where G_1 is the graph I with 7 vertices, 6 edges and $d = 12/7$ pictured in Figure 1. T_1 is identical to the event that the graph contains no element of the family B. Then B_1, the number of possible ways of placing $G_1 = I$ on 7 labelled vertices, is just $7\cdot\binom{6}{3}\cdot 3! = 7!/6$. Thus $P_{v,E(v)}(T_1) \sim e^{-\lambda}$, where all the required constants in Eq. (1) are now known.

To show that $\lim_{v\to\infty} P_{v,E(v)}(T_1) - P_{v,E(v)}(T_2) = 0$, we observe that the event $T_1 - T_2$ is contained in the event C that the graph is not a forest, i.e., in the event that the graph contains a cycle. The threshold function for the event that a random graph contains a cycle with $k \geq 3$ edges is v [8, p. 23]. Since we let $\lim_{v\to\infty} E/v^{5/6} = c$ and since $\lim v^{5/6}/v = 0$, it is immediate that $\lim P_{v,E(v)}(C) = 0$. This proves Theorem 3.

PROOF OF COROLLARY. The lemma used in proving Theorem 3 applies to circular arc graphs as well as to interval graphs.

In Theorem 3, while λ is the exact expected number of subgraphs G_1 in a random graph under the model $\Gamma^{**}_{v,E}$ in which edges are drawn independently, $e^{-\lambda}$ is only an approximation to $P_{v,E}(T)$ even when edges are drawn independently.

There is an apparent discrepancy between the results of Theorems 1 and 3. According to Theorem 3, $P_{v,E}(T)$ decreases monotonically as E increases, for any fixed v. But Figure 6 illustrates how $P_{v,E}(T)$ varies over the whole

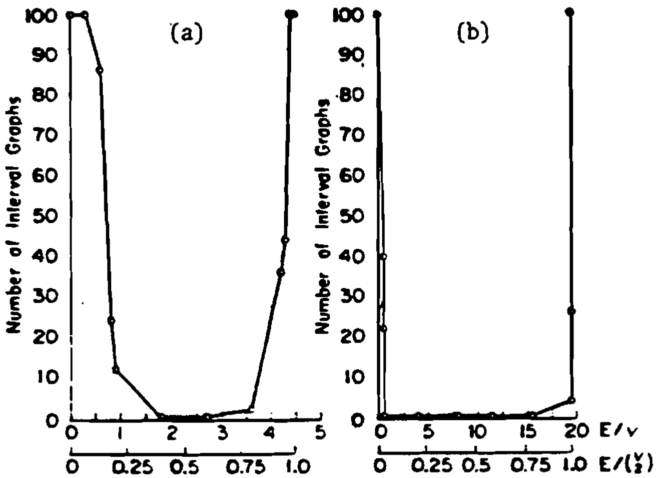

Fig. 6. The number of pseudo-random graphs which are interval graphs, in Monte Carlo samples of 100 for each value of E and v, as a function of E/v (upper abscissa) or as a function of $E/\binom{v}{2}$ (lower abscissa), where E is the number of edges and v is the number of vertices; (a) v = 10, (b) v = 40. These simulations used a different pseudo-random number generator from those in Table 1 and the numerical results differ slightly. (From Cohen [7].)

range of E, for v = 10 (on the left) and v = 40 (on the right), and the exact results in Theorem 1 show that as E approaches $\binom{v}{2}$, $P_{v,E}(T)$ must rise to approach 1. The discrepancy arises because the asymptotic analysis is valid only for values not much larger than $v^{5/6}$, which vanishes relative to $\binom{v}{2}$ for large v. The methods developed here could usefully be applied to finding an asymptotic estimate of $P_{v,E}(T)$ for edges in the range $v^{5/6} << E < \binom{v}{2} - g(v)$ where $g(v) \to \infty$ as $v \to \infty$. It is clear that $P_{v,E}(T)$ vanishes very rapidly in this range.

To determine how large v must be for the asymptotic analysis to provide a good approximation to $P_{v,E}(T)$, we estimate $P_{v,E}(T)$ by Theorem 3 and by Monte Carlo simulation, for v = 10, 40, 100, and 200, and for each v, for values of $E = [kv^{5/6}]$, where [ ] is the integer part and k = 1/3, 1/2, 2/3, 5/6, 1, and 4/3 (Table 1).

Where $E \leq 5$, it makes little sense to use the asymptotic estimate of $P_{v,E}(T)$ since this estimate is based on the event T_1 that a graph does not contain the graph G_1 with 6 edges. Fortunately, exact results are available from Theorem 1 for $E \leq 6$. When $E \leq 6$, Table 1 includes both the exact $P_{v,E}(T)$ and asymptotic estimates for comparison with the Monte Carlo results.

To obtain the Monte Carlo estimates, the algorithm of Fulkerson and Gross [9] was programmed in APL. (In future calculations where it is necessary to determine whether each of many graphs is an interval graph, this algorithm should be replaced by the much faster algorithm described by Booth and Lueker [4]. In their algorithm the number of steps is linear in v + E.) For each combination of v and E, 100 pseudo-random graphs were generated. Unfortunately, the algorithm used to generate pseudo-random numbers in the version of APL which we used is unknown. We corrected the shortcomings of the pseudo-random number generator described in Cohen ([7], Chapter 5) by changing the software. The Monte Carlo estimates of $P_{v,E}(T)$ given in Table 1 are the proportions of these generated graphs which were interval graphs for each v and E.

If p is any one of these estimated proportions, then an estimate of the standard deviation of p is $0.1[p(1-p)]^{1/2}$, which never exceeds 0.05. It is reassuring that the exact probabilities $P_{v,E}(T)$, where known, are never more than 2.5 standard deviations from the corresponding Monte Carlo estimates (M). With v = 40 vertices and E = 18 edges the standard deviation of the estimate p = 0.25 of $P_{40,18}(T)$ is approximately 0.0433. The asymptotic estimate of 0.189 differs from the Monte Carlo estimate by 1.4 standard deviations, which suggests that the asymptotic theory is useful for a graph with as few as 40 vertices and up to 18 edges; for 21 edges, the asymptotic theory appears to underestimate $P_{v,E}(T)$ relative to the Monte

Table 1. Estimates of the probability $P_{v,E}(T)$ that a random graph with v labelled vertices and E edges is an interval graph, according to asymptotic theory (A) and 100 Monte Carlo simulations (M)

$v =$	10				40			100			200		
$v^{5/6} =$	6.813				21.630			46.416			82.704		
Edges	E	$P_{v,E}(T)$	A	M	E	A	M	E	A	M	E	A	M
$[\frac{1}{3}v^{5/6}]$	2	1	.9996#	1.00	7	.993	.99	15	.990	.98	27	.988	.94
$[\frac{1}{2}v^{5/6}]$	3	1	.997#	1.00	10	.944	.88	23	.881	.83	41	.867	.71
$[\frac{2}{3}v^{5/6}]$	4	.996	.988#	.99	14	.671	.73	30	.544	.52	55	.440	.31
$[\frac{5}{6}v^{5/6}]$	5	.977	.968#	.91	18	.189	.25	38	.086	.17	68	.055	.08
$[v^{5/6}]$	6	.927	.936	.86	21	.019	.10	46	.001	.02	82	.000	0
$[\frac{4}{3}v^{5/6}]$	9	*	.797	.11	28	.000	0	61	.000	0	110	.000	*

v = number of vertices

E = number of edges

$P_{v,E}(T)$ = exact probability of an interval graph

A = asymptotic probability of an interval graph

$= \exp(-(7!/6)\binom{v}{7}p^6(1-p)^{15})$, where $p = E/\binom{v}{2}$

M = Monte Carlo probability of an interval graph based on 100 trials for each case

* = not computed

$[kv^{5/6}]$ = integer part of $kv^{5/6}$

= in these cases, $P(T_1) = 1$ when edges are sampled without replacement

Carlo estimates. For $v = 100$ vertices, all the asymptotic estimates are within 2.5 standard deviations of the corresponding Monte Carlo estimates; for $v = 200$, there are some larger deviations. Overall, we conclude that the asymptotic theory is useful for graphs with 100 or more vertices, as long as the number of edges does not greatly exceed $v^{5/6}$. Calculations not shown here of the asymptotic probability using $\exp(-(32/3)(E^6/v^5))$ approximate the Monte Carlo estimates within 3.2 standard deviations in all cases when $v = 200$, though not so well with smaller values of v. Thus λ^* is useful for $v \geq 200$.

The asymptotic estimates in Theorem 3 could probably be improved by taking into account cycles as forbidden induced subgraphs. However, the results would be more complicated and would no longer apply to circular arc graphs, which may have cycles as induced subgraphs.

Benzer [1] observed $E = 61$ (which is one half the number of off-diagonal 0's in his Figure 5) with $v = 19$. Using λ in Theorem 3, the $e^{-221942}$ asymptotic estimate of $P_{19,61}(T)$ is ~~10^{-51}~~. From the rate of decrease of the Monte Carlo estimates of $P_{10,E}(T)$ in Table 1 as E increases beyond $10^{5/6}$, and because $E = 61$ is much less than $\binom{19}{2} = 171$, it appears likely that the chance that Benzer observed an interval graph by chance alone is nearly 0.

5. UNSOLVED PROBLEMS. There remain unsolved mathematical problems related to random interval graphs. Solutions would be useful in scientific inference.

First, the problem of calculating $P_{v,E}(T)$ is really a special case of calculating the probability distribution of the minimum dimension of Euclidean space necessary to represent a random graph as the intersection graph of a family of sets of some given form. For example, the minimum number of dimensions necessary to represent a graph G by the intersections of boxes, or rectangular parallelipipeds with edges parallel to the axes, is called the boxicity of G, and never exceeds $\lfloor v/2 \rfloor$ (Roberts [15]). If $[B = k]$ is the event that the boxicity of a given graph is k, $[B = 0]$ is the event that the graph is a complete graph, since all vertices can be represented by coincident points, and $T = [B = 0 \text{ or } B = 1]$. It would be useful to know the probability $P_{v,E}[B = k]$ that a random graph is of boxicity k, $k = 0, 1, \ldots, \lfloor v/2 \rfloor$. For other families of sets, such as convex sets, the problem may be easier; every graph is the intersection graph of convex sets in 3 or fewer dimensions (Wegner [19]).

Second, there is still no theory for the probability of an interval graph when the probability distribution is defined initially on the directed graphs corresponding to food webs in ecology, rather than on the undirected niche overlap graphs as in the theory of Erdős and Rényi [8]. For example, in ecology it appears (Cohen [7]) that a useful model of a directed graph W is as follows: given a set of v labelled vertices (corresponding to v predators) which may have positive in-degree, and a not necessarily disjoint set of u labelled vertices (corresponding to u prey) which may have positive out-degree, an arrow from a prey vertex to a predator vertex in W actually occurs with probability p, and fails to occur with probability 1-p, independently and with identical probability for each pair of prey and predator vertices. From each such random directed graph W, the random undirected graph G(W) on the v (predator) vertices joins two vertices with an undirected edge if and only if there is a prey vertex in W from which arrows go to both predator vertices. The problem is to find the probability that G(W) is an interval graph, or better, the probability distribution of the boxicity of G.

ACKNOWLEDGMENTS. J.E.C. thanks members of the Mathematical Institute of the Hungarian Academy of Sciences for kind hospitality during a visit in 1973. The visit, under the auspices of an exchange program between the Hungarian Academy of Sciences and the National Academy of Sciences of the U.S.A., led to this collaborative work. William T. Trotter, Jr. corrected slips in an earlier version. Anne Whittaker prepared this typescript.

BIBLIOGRAPHY

1. S. Benzer, On the topology of the genetic fine structure, Proc. Natl. Acad. Sci. U.S.A. 45 (1959), 1607-1620.

2. C. Berge, Perfect graphs, in D. R. Fulkerson (ed.), Studies in Graph Theory, pp. 1-22, Mathematical Association of America, 1975.

3. K. S. Booth, PQ-tree algorithms, Ph.D. thesis, Lawrence Livermore Laboratory, University of California, Livermore, 1975.

4. K. S. Booth and G. S. Lueker, Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms, J. Comput. Syst. Sci. 13 (1976), 335-379.

5. J. E. Cohen, Interval graphs and food webs: a finding and a problem, RAND Document 17696-PR, 30 August 1968.

6. ________, Food webs and the dimensionality of trophic niche space, Proc. Natl. Acad. Sci. 74 (1977), 4533-4536.

7. ________, Food Webs and Niche Space, Princeton University Press, Princeton, 1978.

8. P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960), 17-61.

9. D. R. Fulkerson and O. A. Gross, Incidence matrices and interval graphs, Pac. J. Math. 15 (1965), 835-855.

10. P. C. Gilmore and A. J. Hoffman, A characterization of comparability graphs and of interval graphs, Can. J. Math. 16 (1964), 539-548.

11. G. Hajós, Uber eine Art von Graphen, Int. Math. Nachr. 11 (1957), 65.

12. F. Harary and E. M. Palmer, A survey of graphical enumeration problems, in J. N. Srivastava et al (eds.), A Survey of Combinatorial Theory, North Holland, New York, 1973.

13. L. Hubert, Problems of seriation using a subject by item response matrix, Psychol. Bull. 81 (1974), 976-983.

14. C. G. Lekkerkerker and J. C. Boland, Representation of a finite graph by a set of intervals on the real line, Fund. Math. Polska Akad.

Nauk. 51 (1962), 45-64.

15. F. S. Roberts, On the boxicity and cubicity of a graph, in W. T. Tutte (ed.), Recent Progress in Combinatorics, pp. 301-310, Academic Press, New York, 1969.

16. __________, Discrete Mathematical Models, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.

17. R. E. Tarjan, Graph theory and Gaussian elimination, in J. R. Bunch and D. J. Rose (eds.), Sparse Matrix Computations, pp. 3-22, Academic Press, New York, 1976.

18. A. C. Tucker, Matrix characterizations of circular-arc graphs, Pac. J. Math. 39 (1971), 535-545.

19. G. Wegner, Eigenschaften der Nerven Homologisch-einfacher Familien in R^n, Doctoral Dissertation, Göttingen, 1967.

THE ROCKEFELLER UNIVERSITY, 1230 YORK AVENUE, NEW YORK, NY 10021, U.S.A.

MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI KUTATÓ INTÉZETE, BUDAPEST V., REÁLTANODA U. 13-15, HUNGARY