COMBINATORICA
Akadémiai Kiadó — Springer-Verlag

COMBINATORICA 8 (4) 1988 357—371

RANDOM INTERVAL GRAPHS

E. R. SCHEINERMAN

Received June 23, 1986
Revised August 28, 1987

In this paper we introduce a notion of *random interval graphs:* the intersection graphs of real, compact intervals whose end points are chosen at random. We establish results about the number of edges, degrees, Hamiltonicity, chromatic number and independence number of almost all interval graphs.

1. Introduction

We call a graph an *interval graph* if it is possible to assign real intervals to its vertices so that adjacency corresponds to non-empty intersection. More precisely, denote by $\mathcal{I}$ the set of all real compact intervals. An *interval representation* for a graph $G=(V, E)$ is a function $f\colon V\to\mathcal{I}$ which has the property that for every pair of distinct vertices v, w, we have v adjacent to w (notation: $v\sim w$) if and only if $f(v)\cap f(w)\neq\emptyset$. The class of interval graphs consists of those graphs with interval representations. In this paper we discuss random interval graphs.

In the common models of random graphs, interval graphs play a minor role. Interval graphs die out early in the graphical evolution. Hence, interval graphs are "unimportant" from a random graph point of view; they appear with probability tending to 0 (see [4]). We therefore are led to create a separate model for random *interval* graphs which we discuss in Section 2. In Section 3 we discuss properties of sets of random intervals which we need for Section 4 in which we deduce many of the properties enjoyed by almost all interval graphs. We conclude, in Section 5, with a comparison of our results with those of the Erdős—Rényi model for random graphs.

2. The model

In this section we introduce two equivalent models for random interval graphs. Let $\mathcal{G}$ denote the set of interval graphs and $\mathcal{G}_n$ denote the set of all interval graphs on n labelled vertices $\{1, \ldots, n\}$. As a notational convenience, when x and y are real numbers, we write $[x, y]$ to stand for the set $\{tx+(1-t)y\colon 0\leqq t\leqq 1\}$. Thus both $[0, 1]$ and $[1, 0]$ stand for the unit interval. Define $\Psi\colon \mathcal{I}^n\to\mathcal{G}_n$ by assigning to the list $(I_1, I_2, \ldots, I_n)$ the unique labeled interval graph on vertex set $\{1, \ldots, n\}$ such that $i\to I_i$ is an interval representation (i.e., $i\sim j$ iff $I_i\cap I_j\neq\emptyset$).

AMS subject classification (1980): 05 C 30.

Let n be a positive integer. In both models, we choose n intervals at random and form the corresponding interval graph. Thus we consider probability models on subsets of $\mathscr{I}^n$ and then the probability assigned to an interval graph $G\in\mathscr{G}_n$ is defined as the probability of the set $\Psi^{-1}(G)$.

In the first model, we restrict our attention to the subset of $\mathscr{I}^n$ consisting of those lists of n intervals, all of whose end points are distinct elements of $\{1, ..., 2n\}$. There are $(2n)!/2^n$ such lists of intervals, which we take as equiprobable. Thus random interval graphs are formed by selecting n intervals at random, without repetition and with distinct end points in $\{1, ..., 2n\}$. While this model has the advantages of being very natural and defined in a strictly combinatorial manner, it suffers from the fact that the choices of the intervals assigned to vertices are highly dependent.

In our second model, the subset of $\mathscr{I}^n$ we consider are those lists of intervals contained in $[0, 1]$. Indeed, we identify this set with $[0, 1]^{2n}$ endowed with Lebesgue measure. Let $\Phi\colon [0, 1]^{2n}\to\mathscr{I}^n$ by

$$\Phi(x_1, ..., x_{2n}) = ([x_1, x_2], ..., [x_{2n-1}, x_{2n}]).$$

Note that Φ is a 2^n-to-1 map almost everywhere. Denote the image of Φ by

$$\mathscr{I}^n_{01} = \{(I_1, ..., I_n)\in\mathscr{I}^n\colon\ I_i \subset [0, 1] \text{ for all } i\}.$$

Thus we have

$$[0, 1]^{2n} \xrightarrow{\Phi} \mathscr{I}^n_{01} \subset \mathscr{I}^n \xrightarrow{\Psi} \mathscr{G}_n$$

and the probability assigned to a graph $G\in\mathscr{G}_n$ is this second model is the volume of $\Phi^{-1}(\Psi^{-1}(\{G\}))$. Equivalently, in this second model we generate random interval graphs by considering $2\,n$ random variables $X_1, Y_1, ..., X_n, Y_n$ which are independent and uniform on $[0, 1]$, and then forming the interval graph whose representation consists of the intervals $[X_1, Y_1], ..., [X_n, Y_n]$. This second model has the advantage that the random interval selected for one vertex is independent of the interval selected for another.

Fortunately, it is easy to see that the probabilities assigned to a graph in the two models are equal. This is because, in the second model, the $2n$ random variables $X_1, ..., Y_n$ are all distinct (with probability 1) and each of the $(2n)!$ possible orders for their values are equally likely. Indeed, any continuous distribution for the $2n$ random variables would result in the same model.

We therefore consider $\mathscr{G}_n$ as a sample space whose probability measure $P=P_n$ is defined in either of the first or the second sense above. We now pose and answef fundamental questions about properties of random interval graphs. A property or an interval graph can be identified with the subset $Q\subset\mathscr{G}$ of those interval graphs which have that property. We say that Q holds for almost all interval graphs if

$$\lim_{n\to\infty} P_n(\mathscr{Q}\cap\mathscr{G}_n) = 1.$$

Given a random interval graph, is it connected? How many edges does it have? What is its maximum (minimum) degree? Is it Hamiltonian? What is its chromatic number? What is its independence number?

3. Properties of random representations

In this section we discuss some of the probabilistic tools and properties of $\mathscr{I}_{01}^n$ which enable us to deduce the structure of random interval graphs.

We begin with three "deviations" results (see [2], chapter 1). The simplest of these is Chebyshev's inequality. For a random variable X,

Chebyshev's Inequality. $P(|X-E(X)|\geqq t)\leqq \mathrm{Var}(x)/t^2$. ∎

Now suppose, for the next two results, that $X=X_1+\ldots+X_n$ where the X_i are independent 0—1 random variables with $p=P(X_i=1)$.

"Medium" Deviations Lemma. *If*

$$1 \leqq h < \min\left\{\frac{np(1-p}{10}, \frac{(pn)^{2/3}}{2}\right\} \quad \text{and} \quad t = \frac{h}{\sqrt{np(1-p)}}$$

then

$$P(|X-np| \geqq h) \leqq \frac{1}{t}\exp\left\{\frac{-t^2}{2}\right\}. \quad ∎$$

"Large" Deviations Lemma. *If* $p<1/2$ *and* $\varepsilon>0$ *then*

$$P(|X-np| \geqq \varepsilon np) \leqq \frac{a_\varepsilon \exp\{-b_\varepsilon pn\}}{\sqrt{np}}$$

where a_ε *and* b_ε *are positive constants which depend only on* ε *but not on* p *or* n. ∎

We infer properties of almost all interval graphs by analysis of the set of all representations, $\mathscr{I}_{01}^n$. Here we consider some important properties enjoyed by almost all representations, that is, those properties which fail only on a subset of $\mathscr{I}_{01}^n$ whose probability tends to 0 as $n\to\infty$.

Proposition 3.1. *In almost all representations, all intervals have length at least* $1/n^2$.

Proof. The probability that an interval in [0, 1] has length less than $1/n^2$ is at most $2/n^2$. Given $\mathbf{I}=(I_1, \ldots, I_n)\in\mathscr{I}_{01}^n$, the probability that some interval has length less than $1/n^2$ is at most $2/n\to 0$. ∎

Given n, we define a *grid value* to be a number which is an integer multiple of $1/n^2$. Proposition 3.1 implies that in almost all representations, all intervals contain a grid value.

For real x, define a non-negative integer valued random variable, $\mathscr{D}_x$, on $\mathscr{I}_{01}^n$ as follows: Given $\mathbf{I}=(I_1, I_n)\in\mathscr{I}_{01}^n$, put $\mathscr{D}_x(\mathbf{I})$ equal to the number of intervals I_i with $x\in I_i$. We call $\mathscr{D}_x$ the *depth* of the representation $\mathbf{I}$ at x.

Denote by η any function of n for which $\eta n^k\to 0$ as $n\to\infty$ for any fixed k.

Proposition 3.2. *Let* $\omega_n\to\infty$ *arbitrarily slowly. In almost all representations, every grid value* $x\in[\omega_n \log n/n, 1-(\omega_n \log n/n)]$ *satisfies:*

$$(1) \qquad |\mathscr{D}_x-2n(x-x^2)| \leqq (nx)^{0.6}.$$

Proof. We show that the probability (1) fails for a fixed grid value x (in the interval specified) is η. [Without loss of generality, we may assume $x\leqq 1/2$.] Let $X_i=1$ if

$x \in I_i$ and 0 otherwise. Thus $\mathscr{D}_x = \sum X_i$. Now $p = P(X_i = 1) = 1 - x^2 - (1-x)^2 = 2x - 2x^2$. Hence $E(\mathscr{D}_x) = 2n(x - x^2)$. Let $h = (nx)^{0.6}$ and $t = h/\sqrt{np(1-p)}$. By the "medium" deviations result, (1) fails with probability at most

$$\frac{1}{t}\exp\left\{\frac{-t^2}{2}\right\} \leq \frac{\sqrt{2}}{(nx)^{0.1}}\exp\left\{\frac{-(nx)^{0.2}}{4}\right\} = o(1)n^{-\omega_n/4} = \eta.$$

Since there are fewer than n^2 possible grid values, (1) holds for all grid values [in the interval specified] with probability at least $1 - n^2\eta \to 1$. ▮

A useful concept is the radius of an interval. Let $I = [x, y] \subset [0, 1]$. The *radius* of I, denoted ϱI, is $\sqrt{a^2 + (1-b)^2}$ where $a = \min\{x, y\}$ and $b = \max\{x, y\}$. This is the distance from (x, y) to $(0, 1)$ or $(1, 0)$ whichever is smaller. We write $\varrho^2 I$ for $(\varrho I)^2$.

Proposition 3.3. *For* $I \in \mathscr{I}_{01}$ *and* $0 \leq y \leq 1$ *we have*

$$P(\varrho^2 I \leq y) = \frac{\pi}{2} y \quad \textit{when} \quad y \leq \frac{1}{2}$$

and

$$P(\varrho^2 I \leq y) = y\left\{\frac{\pi}{2} - 2\cos^{-1}\left[\frac{1}{\sqrt{2y}}\right]\right\} + \sqrt{2y-1} \quad \textit{when} \quad y > \frac{1}{2}.$$

Proof. The required probability equals the area of the square $[0, 1]^2$ within a distance $\sqrt{y}$ of either $(0, 1)$ or $(1, 0)$. If $y \leq 1/2$ this consists of two disjoint quarter circles and if $y > 1/2$ the quarter circles overlap. See Figure 1. ▮

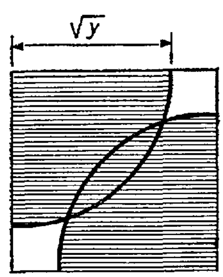

Fig. 1

4. Properties of Random Interval Graphs

In this section we dicuss the fundamental properties of random interval graphs. We begin with a simple edge count.

Theorem 4.1. *Almost all graphs in* $\mathscr{G}_n$ *have* $n^2/3 + o(n^2)$ *edges.*

Proof. For $G \in \mathscr{G}_n$ and $1 \leq i < j \leq n$ put X_{ij} equal 1 if $i \sim j$ and 0 otherwise. By considering the end points of the intervals assigned to i and j we see that $P(X_{ij} = 1) = 2/3$.

Now $X=\sum X_{ij}$ is the number of edges in G. One checks that

$$E(X) = \frac{2}{3}\binom{n}{2} = \frac{n^2}{2}+o(n^2)$$

and $\mathrm{Var}(X)=O(n^3)$.

Chebyshev's inequality gives,

$$P\left(\left|X-\frac{n^2}{3}\right| \geqq n^{7/4}\right) \leqq \frac{\mathrm{Var}(X)}{n^{7/2}} \to 0. \quad \blacksquare$$

Given two vertices of a random interval graph, the probability they are adjacent is 2/3. How does $\mathcal{G}_n$ compare with the usual random graph model in which $p_{\mathrm{edge}}=2/3$? In the Erdős—Rényi model almost all graphs have $n^3/3+o(n^2)$ edges as well. However, in $\mathcal{G}_n$ the variance in the number of edges is $O(n^3)$ which is greater than the variance of $O(n^2)$ in $p_{\mathrm{edge}}=2/3$. We compare the two models in more detail in Section 5.

Observe that the adjacency $1\sim 2$ is dependent on the adjacency of 2 with 3. Indeed, $P(1\sim 2|2\sim 3)=11/15>2/3=P(1\sim 2)$. However, given two pairs of vertices, they will almost always be disjoint and the dependency does not cause drastic effects in the number of edges. On the other hand, it does have far reaching effects on the degrees of vertices.

Theorem 4.2. *Let $G\in\mathcal{G}_n$ and $v\in V(G)$. For fixed $x\in[0,1]$ we have*

$$\lim_{n\to\infty} P(d(v) \leqq xn) = 1-(1-x)\frac{\pi}{2} \quad \text{when} \quad x \geqq \frac{1}{2}$$

and

$$\lim_{n\to\infty} P(d(v) \leqq xn) = 1-(1-x)\left\{\frac{\pi}{2}-2\cos^{-1}\left[\frac{1}{\sqrt{2-2x}}\right]\right\}-\sqrt{1-2x} \quad \text{when} \quad x < \frac{1}{2}.$$

Proof. Without loss of generality, we assume v is vertex 1. Let I_i be the interval assigned to vertex i in a representation of G. For $i=2,\dots,n$ put $X_i=1$ if $I_i\cap I_1\neq\emptyset$ (i.e., $i\sim 1$) and 0 otherwise. Observe that X_i is independent of X_j for $i\neq j$ and that $X=\sum_{i\geqq 2} X_i$ is the degree of vertex 1.

Now suppose $\varrho(I_1)=r$ is fixed. Put $p=P(X_i=1)=1-r^2$. Now $E(X)=(n-1)p$ and $\mathrm{Var}(X)=(n-1)p(1-p)$. Hence $P(|X-np|\geqq n^{2/3})\leqq \mathrm{Var}(X)n^{-4/3}\to 0$. Thus $d(1)=np+o(n)$ in almost all random interval graphs under the assumption $\varrho(I_1)=r$. Thus for any $\varepsilon>0$

$$P\big(d(1) \leqq xn | \varrho(I_1) = r\big) = \begin{cases} 1-o(1) & \text{for } r < \sqrt{1-x}-\varepsilon \\ o(1) & \text{for } r > \sqrt{1-x}+\varepsilon \end{cases}$$

Finally,

$$P(d(1) \leqq xn) = \int_0^1 P\big(d(1) \leqq xn | \varrho(I_1) = r\big)\, dP\big(\varrho(I_1) \leqq r\big) =$$

$$= [1-o(1)]P\big(\varrho^2(I_1) \leqq 1-x\big)+\varepsilon O(1)+o(1) \to P\big(\varrho^2(I_1) \leqq 1-x\big).$$

The result follows by Proposition 3.3. $\blacksquare$

Recall that δ and Δ denote the minimum and maximum degrees of a graph. The above result implies:

Corollary. *For every* $\varepsilon > 0$, *almost all interval graphs satisfy* $\delta < \varepsilon n$ *and* $\Delta > (1-\varepsilon)n$. ∎

We now provide more detailed information about the minimum and maximum degrees of a random interval graph. First, we show that the minimum degrees of random interval graphs converge in distribution to a Rayleigh [13] distribution:

Theorem 4.3. *Let k be a fixed, non-negative real number. We have,*

$$\lim_{n\to\infty} P(\delta < k\sqrt{n}) = 1 - e^{-k^2/2}.$$

Proof. Denote by K_z the number of intervals in a representation which intersect $[0, z)$. If $z = A/\sqrt{n}$, where A is a constant, one readily shows

$$P(|K_z - 2nz| \geqq n^{0.3}) \leqq \eta$$

using the "medium" deviations result. Put $M(z) = [z, 1-z]$ and $T(z) = [0, 1] - M(z)$.

By Proposition 3.1, we may assume every interval contains a grid value. Choose $\varepsilon > 0$. Let v be a fixed vertex and I_v its interval.

Let x be a grid value nearest $(k+\varepsilon)/(2\sqrt{n})$ Now if I_v intersects $M(x)$ we have

$$d(v) \geqq \mathcal{D}_x + O(n^{0.3}) > k\sqrt{n}$$

fails with probability η. Thus,

$$I_v \cap M(x) \neq \emptyset \Rightarrow d(v) \geqq k\sqrt{n} \tag{1}$$

for all $v \in V(G)$ fails with probability $n\eta \to 0$.

Now let y be a grid value nearest $(k-\varepsilon)/(2\sqrt{n})$. If $I_v \subset T(y)$, then either $I_v \subset [0, y)$ or else $I_v \subset (1-y, 1]$. The number of intervals with an end point in $[0, y)$ is $2ny + O(n^{0.3})$ with probability $1-\eta$. (Likewise for $(1-y, 1]$.) Thus $d(v) \leqq 2ny + O(n^{0.3}) < k\sqrt{n}$. Hence, $I_v \cap M(y) \neq \emptyset \Rightarrow d(v) < k\sqrt{n}$, or

$$I_v \cap M(y) \neq \emptyset \Leftarrow d(v) \geqq k\sqrt{n} \tag{2}$$

for all $v \in V(G)$ fails with probability $n\eta \to 0$.

Now we compute $P(\delta \geqq k\sqrt{n})$. Using (1) we have,

$$P(\delta \geqq k\sqrt{n}) = P(d(v) \geqq k\sqrt{n} \text{ for all } v) \leqq P(I_v \cap M(x) \neq \emptyset \text{ for all } v) + o(1) =$$
$$= (1-2x^2)^n + o(1) \leqq e^{-k^2/2} + o(1) + \varepsilon a_k$$

where a_k depends only on k.

Now we use (2) to estimate

$$P(\delta \geqq k\sqrt{n}) = P(d(v) \geqq k\sqrt{n} \text{ for all } v) \geqq$$
$$\geqq P(I_v \cap M(y) \neq \emptyset \text{ for all } v) + o(1) = (1-2y^2)^n + o(1) \geqq e^{-k^2/2} + o(1) + \varepsilon b_k$$

with b_k depending only on k. Hence,

$$\left|P(\delta \geqq k\sqrt{n}) - e^{-k^2/2}\right| \leqq \varepsilon(|a_k|+|b_k|)+o(1).$$

Since ε was arbitrary, we have

$$P(\delta < k\sqrt{n}) \to 1-e^{-k^2/2}. \quad \blacksquare$$

Because the minimum degree (when normalized by dividing by $\sqrt{n}$) has a Rayleigh distribution (with parameter=1), we know that its mean is asymptotically $\sqrt{\pi n/2}$.

Next we analyze the maximum degree. It is convenient to use the notation $\Delta^*(G)=n-1-\Delta(G)$ where G is a graph with n vertices; $\Delta^*(G)$ is the minimum degree of G's complement.

Theorem 4.4. *Let $\omega_n \to \infty$ arbitrarily slowly. In almost all interval graphs, $\Delta \geqq n-\omega_n$.*

Proof. Put $x=\sqrt{\omega_n/n}/2$. First one checks that in almost all representations, at least $n-\omega_n/2+o(\omega_n)$ intervals intersect $[x, 1-x]$ and then one verifies, by the second moment method, that there must be an interval containing $[x, 1-x]$ which corresponds to a vertex of degree at least $n-\omega_n$. $\blacksquare$

For fixed k, what is the probability that $\Delta^*=k$ (i.e., $\Delta=n-1-k$)? This appears to be a difficult problem. Using a computer we generated 1,000 random interval graphs on $n=10{,}000$ vertices and obtained the following results:

Δ^*	Per Cent
0	63.6
1	24.6
2	7.8
3	2.7
4	0.9
5	0.2
6	0.2

Table 1. Δ^* values of 1,000 random interval graphs.

We now focus on the problem of evaluating $P(\Delta^*=0)$.

Proposition 4.5. *Let N_n denote the number of vertices of denree $n-1$ in a random interval graph on n vertices. The r^{th} factorial moments of N_n are:*

$$E_r(N_n) = E[N_n(N_n-1)\ldots(N_n-r+1)] \to 2^r\left[\left(\frac{r}{2}\right)!\right]^2$$

Note that for r large, the above limit is asymptotic to

$$\pi r\left(\frac{r}{e}\right)^r.$$

Proof. For a vertex v of a graph on n vertices, write $d^*(v)=n-1-d(v)$. Thus $\Delta^*(G)=\min_{v\in V(G)} d^*(v)$.

For fixed r let A be the event $d^*(1)=d^*(2)=\ldots=d^*(r)=0$ and let $I_i=[x_i, y_i]$ denote vertex i's interval for $i=1, \ldots, r$. We compute $P(A)$ by conditioning on the order of the x's and y's. If for some $1\leqq i<j\leqq r$ we had i not adjacent to j then A could not hold, hence we only consider orders for the x's and y's for which

$$(*) \qquad \max_i \{\min(x_i, y_i)\} < \min_i \{\max(x_i, y_i)\}.$$

Of the $(2r)!$ orders for the x's and y's there are $2^r(r!)^2$ orders which satisfy $(*)$. We now assume, without loss of generality, that $x_i<y_i$ for $i=1, \ldots, r$. Given a fixed order for the x's and y's in which $x_i<y_j$ for all i, j and which satisfies $(*)$, we see that the conditional probability of A is $[1-x^2-(1-y)^2]^{n-r}$ where $x=\max_i \{x_i\}$ and $y=\min_i \{y_i\}$. Let B be the event

$$0 \leqq x_1 < x_2 < \ldots < x_r < y_r < \ldots < y_1 \leqq 1$$

and so $P(A)=2^r(r!)^2P(A \text{ and } B)$. Further, let C be the event $B\cap\{x_r<1/2<y_r\}$. We now calculate,

$$P(A \text{ and } B) = \int_B [1-x_r^2-(1-y_r)^2]^{n-r} d\mathcal{L}^{2r} =$$

$$= \int_C [1-x_r^2-(1-y_r)^2]^{n-r} d\mathcal{L}^{2r} + \eta =$$

$$= \int_0^{1/2}\int_0^{1/2} \left[\frac{1}{(r-1)!}\right]^2 [1-x^2-z^2]^{n-r} x^{r-1} z^{r-1}\, dx\, dz + \eta.$$

Hence,

$$P(A) = 2^r r^2 \int_0^{1/2}\int_0^{1/2} (1-x^2-z^2)^{n-r} x^{r-1} z^{r-1}\, dx\, dz + \eta$$

$$= 2^r r^2 \iint (1-s^2)^{n-r} s^{2r-2} \sin^{r-1}\theta \cos^{r-1}\theta\, s\, ds\, d + \eta$$

$$= 2^r r^2 \left\{\frac{1}{2} B(n-r+1, r)\right\}\left\{\frac{1}{2} B\left(\frac{r}{2}, \frac{r}{2}\right)\right\} + \eta$$

[where B is the Beta function]

$$= \frac{2^r((r/2)!)^2}{(n)_r} + \eta.$$

This gives, $E_r(N)=(n)_r P(A)\to 2^r[(r/2)!]^2$. ∎

What is $\lim P(\Delta^*=0)$? Here we argue that it exists and posit a conjecture for its value.

Observe that $P(\Delta^*=0)=1-P(N_n=0)$, so it is enough to compute $\lim_{n\to\infty} P(N_n=0)$. One can use the following formula for Stirling numbers of the second kind:

$$S(r,k)=\frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^r$$

to bound the r^{th} moments of N_n:

$$E(N_n^r)=\sum_{k=0}^{r}S(r,k)E_k(N_n)\leqq(r+1)2^r r^r\sqrt{\pi r/2.}$$

Hence,

$$\limsup_{r\to\infty}\frac{(E(N_n^r))^{1/r}}{r}\leqq 1<\infty.$$

Now we see that the sequence of random variables $\{N_n\}$ has finite moments which converge to finite limits. Hence there exists a limiting distribution which, by Carleman's theorem, is unique. (See [1], Section 30.) Hence there is a unique non-negative integer valued random variable N with $N_n \xrightarrow{d} N$. Therefore, $\lim_{n\to\infty} P(N_n=0)$ exists and equals $P(N=0)$.

Now put $p_i=P(N=i)$ and consider the probability generating function $f(x)=p_0+p_1x+p_2x^2+\dots$. Now we expand f about $x-1$ and we obtain:

$$(**)\qquad f(x)=\sum_{k=0}^{\infty}\frac{f_k(x-1)^k}{k!}$$

where $f_k=E_k(N)=2^k[(k/2)!]^2$. We see that the coefficient of $(x-1)^k$ in $(**)$ is $f_k/k!\sim\sqrt{\pi k/2}$.

We want to compute $f(0)=P(N=0)$. Using $(**)$ we found $|f(0)-1/3|\leqq$ $\leqq 10^{-4}$. We therefore conjecture that $P(\Delta^*=0)\to 2/3$. We posit:

Conjecture. For fixed non-negative integers k, $\lim_{n\to\infty} P(\Delta=n-1-k)=2/3^{k+1}$.

Note added in revision: The referee, using a better method for evaluating f, computed $|f(0)-1/3|\leqq 10^{-32}$. Recently, Andrew Barbour (personal communication) succeeded in verifying $f(0)=1/3$ analytically. The conjecture above, for $k>0$, remains open.

It is easy to use the degree results to show that almost all interval graphs are connected. Here we analyze the degrees to show the stronger result that almost all interval graphs have a *Hamiltonian cycle,* a cycle which contains every vertex of the graph.

Theorem 4.6. *Almost all interval graphs are Hamiltonian.*

Proof. We use the following sufficient condition for Hamiltonicity which is a weakened version of a theoremdue to Chvatal[3]. Let D_n denote the number of vertices of degree at most k. In a graph G with n vertices, if

$$(1)\qquad D_k<k \quad\text{for all}\quad k \quad\text{with}\quad 1\leqq k\leqq\frac{n}{2}$$

then G is Hamiltonian. We show that (1) holds for almost all interval graphs. We begin by asserting claims about almost all representations. The first is a variant of Proposition 3.2:

Claim 1. *In almost all representations, every pair of grid values* $x, y \in [0, 1]$, *with* $1-\varrho^2[x, y] \geqq \log n/\sqrt{n}$, *the number of intervals which intersect* $[x, y]$, *denoted* $\mathscr{D}_{[x,y]}$, *satisfies,*

(2) $$\mathscr{D}_{[x,y]} > 0.95n(1-\varrho^2[x, y])$$

and if x, y *are grid values for which* $1-\varrho^2[x, y] < 2\log n/\sqrt{n}$ *then*

(3) $$\mathscr{D}_{[x,y]} < \frac{n}{11}$$

Moreover, if $x=y$, *and* $\log n/(2\sqrt{n}) \leqq x \leqq 1-\log n/(2\sqrt{n})$

(4) $$\mathscr{D}_x > 1.9nx(1-x).$$

Fix x, y satisfying the conditions of the first part of the claim. Let p denote the probability that an interval intersects $[x, y]$, hence $p=1-\varrho^2[x, y]$. Taking $\varepsilon=0.05$, we apply the "large" deviation result to calculate,

$$P(\mathscr{D}_{[x,y]} \leqq 0.95np) \leqq o(1)n^{-b\sqrt{n}} = \eta.$$

The number of possible x, y pairs is at most n^4 and so the probability that some pair satisfies $\mathscr{D}_{[x,y]} \leqq 0.95np$ is at most $n^4\eta \to 0$. This gives (2).

Now if $p < 2\log n/\sqrt{n}$ then $E(\mathscr{D}_{[x,y]}) < 2\sqrt{n}\log n$ and $\operatorname{Var}(\mathscr{D}_{[x,y]}) = np(1-p) < 2\sqrt{n}\log n$. The "large" deviations result gives

$$P\left(\mathscr{D}_{[x,y]} > \frac{n}{11}\right) \leqq P(|\mathscr{D}_{[x,y]} - E(\mathscr{D}_{[x,y]})| > E(\mathscr{D}_{[x,y]})) \leqq \eta.$$

Since (much) fewer than n^4 grid value end point intervals satisfy $1-\varrho^2 < 2\log n/\sqrt{n}$, (3) holds with probability exceeding $1-n^4\eta \to 1$.

Equation (4) follows from equation (2) by noting $\varrho[x, x] = 2x(1-x)$.

Let $T(x) = [0, 1]-[x, 1-x]$ and let K_α denote the number of intervals contained in T_α.

Claim 2. *In almost every representation, every grid value* x *for which* $\log n/\sqrt{n} \leqq x \leqq 1/2$ *satisfies*

(5) $$K_x < 2.1nx^2.$$

Let p denote the probability that an interval is contained in $T(x)$, hence $p=2x^2$. Applying the "large" deviation result for a fixed x in the range specified gives,

$$P(K_x \geqq 2.1nx^2) \leqq O(1)^{-b\log n} = \eta.$$

There are at most n^2 possible values for x, hence the probability that $K_\alpha > 2.1nx^2$ for some x is at most $n^2\eta \to 0$.

Claim 3. *In almost all representations, for all grid values* x *with* $0.47 \leqq x \leqq 0.9$, *the number of intervals with* $\varrho^2 I \geqq x$ *is at most*

(6) $$\frac{n(1-x)}{2}.$$

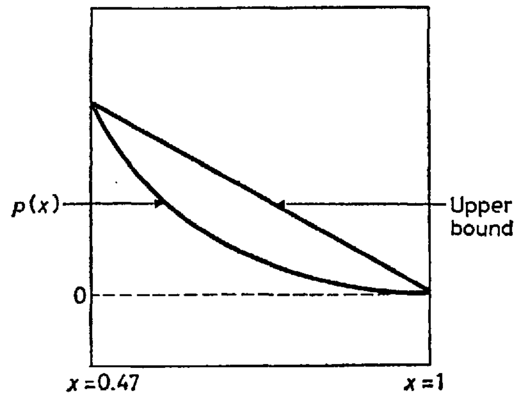

Fig. 2

For fixed x, lep p denote the probability that an interval I satisfies $\varrho^2 I \geqq x$. This probability is given exactly in Proposition 3.3. Elementary, but tedious, calculations show that $p < (1-x)/2.01$ in the range specified; a graph plotting both p and the lower bound versus x is shown in Figure 2. Let X_x denote the number of intervals with $\varrho^2 \geqq x$. The "large" deviation result gives,

$$P\left(X_x \geqq \frac{n(1-x)}{2}\right) \leqq P(X_x \geqq 1.005np) \leqq e^{-bn} = \eta.$$

Since there are fewer that n^2 possible values for x, the probability that $X \geqq n(1-x)/2$ is at most $n^2\eta \to 0$.

We now restrict our attention to those representations for which the conclusions to the three claims and Proposition 3.1 hold; this includes almost all representations.

Case A: $1 \leqq k \leqq 2\sqrt{n}/\log n$.

By Theorem 4.3, in almost all interval graphs, $\delta > 2\sqrt{n}/\log n$. Thus $D_k = 0 < k$.

Case B: $2\sqrt{n}/\log n \leqq k \leqq n/10$.

Put $x = [0.6kn]/n^2$, the grid value nearest and at least $0.6k/n$. In particular, $x \geqq 1/(\sqrt{n}\log n)$. By Proposition 3.1, we know that every interval in the representation contains a grid value. We infer that if $d(v) \leqq k$ then its interval, I_v, is contained in $T(x)$ since the depth at each grid value between x and $1-x$ is at least (using (4)) $1.9nx(1-x) \geqq 1.9n(0.6k/n)(0.93) > 1.06k$. (Note that the estimate $1-x > 0.93$ depends on $k < n/10$.) Hence, $D_k \leqq K_x < 2.1nx^2 < 0.8k^2/n < 0.1k < k$.

Case C: $n/10 \leqq k \leqq n/2$.

Let $d(v) \leqq k$ and $r = \varrho I_v$, where I_v is v's interval. By case (B) fewer than $0.1(n/10) < 0.1k$ vertices have degree less than k when $k < n/10$. Suppose $d(v) \geqq n/10$. Let J be the smallest interval with grid value end points such that $J \supset I_v$. Let $s = \varrho J$. Note that $s^2 - r^2 \leqq O(n^{-2})$. Since $d(v) \geqq n/10$, the number of intervals meeting J is at least $n/10$, hence by claim 1, equation (3), $1 - s^2 > 2\log n/\sqrt{n}$. Now let I^* be the largest interval with grid value end points such that $I^* \subset I_v$ and let $r^* = \varrho I^*$. Observe $s^2 - r^{*2} \leqq O(n^{-2})$ hence $1 - r^{*1} > \log n/\sqrt{n}$. By (2), $\mathscr{D}_{I^*} > 0.95n(1-r^{*2})$ and $k \geqq d(v) \geqq \mathscr{D}_{I^*} - 1 > 0.95n(1-r^{*2})$. From this one computes $r^2 > 1 - 1.055k/n$. Thus the number of vertices with degree at most k but no less than $n/10$ is bounded above by the number of vertices with radius exceeding $\sqrt{1-1.055k/n}$. Put x equal to the largest grid value below $1-1.055k/n$, hence $0.47 \leqq y \leqq 0.9$. Using claim 3,

$$D_k \leqq 0.1k + \frac{1.06k}{2} < k.$$

Thus, combining cases (A), (B) and (C), we see that in almost all representations, $D_k < k$ for all $1 \leqq k \leqq n/2$. Therefore, almost all interval graphs are Hamiltonian. ∎

Next we establish the chromatic number of almost all interval graphs:

Theorem 4.7. *Almost all interval graphs* $G \in \mathscr{G}_n$ *have* $\chi(G) = n/2 + o(n)$.

Proof. Because G is perfect (see [9]), its chromatic number equals the number of vertices in its largest clique. By the Helly property [10], the intervals representing this clique have nonempty intersection. For each interval of the form $I_i = [i/n^2, (i+1)/n^2]$ (where i is an integer), one can use the "medium" deviation result to check that the probability more than $n/2 + o(n)$ intervals meet I_i is η. Hence, in almost all representations, no point in [0, 1] is contained in more than $n/2 + o(n)$ intervals. However, $\mathscr{D}_{1/2} = n/2 + o(n)$. Hence the largest clique in G has size $n/2 + o(n)$ and the result follows. ∎

Finally, we have the following information concerning $\alpha(G)$, the size of the largest independent set of vertices in G.

Theorem 4.8. *There exist absolute constants* c_1, c_2 *such that almost all interval graphs* $G \in \mathscr{G}_n$ *satisfy*

$$c_1 \sqrt{n} \leqq \alpha(G) \leqq c_2 \sqrt{n}$$

Proof. Here we establish this result with $c_1 = (4\sqrt{2})/9 - \varepsilon$ (for any $\varepsilon > 0$) and $c_2 = e/\sqrt{2}$. (This gives $c_1 \approx 0.628$ and $c_2 \approx 1.922$.)

First we establish the uper bound. For $S \subset V(G)$, let X_S denote the random variable which is 1 if S is an independent set of vertices and which is 0 otherwise. Let k be an integer with $k \geqq e\sqrt{n/2}$. Let $X = \sum_{|S|=k} X_S$. Thus $X > 0$ if and only if $\alpha(G) \geqq k$. Observe that

$$P(X_S = 1) = \frac{k!\,2k}{(2k!)} \sim \frac{1}{\sqrt{2}} \left(\frac{e}{2k}\right)^k \quad \text{where} \quad k = |S|.$$

Hence,

$$P(\alpha(G) \geqq k) = P(X > 0) \leqq E(X) = \binom{n}{k} E(X_S) \leqq \frac{n^k}{k!\sqrt{2}} \left(\frac{e}{2k}\right)^k \to 0.$$

Hence, in almost every interval graph, $\alpha(G) < (e/\sqrt{2})\sqrt{n}$.

Next we establish the lower bound. Let $|I|$ denote the length of the interval I. Given $\lambda \in [0, 1]$ and selecting $I \in \mathscr{I}_{01}$ we have $P(|I| \leqq \lambda) = 2\lambda - \lambda^2$. This probability equals the area of the shaded part of Figure 3. In the sequel we assume that λ depends on n such that $\lambda \to 0$ and $n\lambda \to \infty$. In particular, we choose $\lambda = \sqrt{2/(9n)}$. Fix $\varepsilon > 0$.

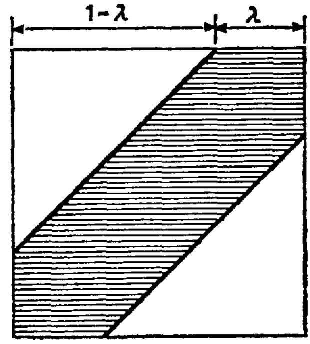

Fig. 3

Let X denote the number of intervals in G's representation which have length at most λ. It is easy to check that $X=2n\lambda+o(n\lambda)$. We now consider only interval graphs with a representation containing at least $m=\lfloor(2-\varepsilon)n\lambda\rfloor$ intervals of length at most λ. (The probability that a graph has no such representation tends to 0.) Let $I_1, \ldots, I_m$ denote intervals in G's representation which have length at most λ. Next we compute

$$p = P\big(I_1 \cap I_2 \neq \emptyset \big| |I_1| \leqq \lambda \text{ and } |I_2| \leqq \lambda\big) = \frac{P(I_1 \cap I_2 \neq \emptyset \text{ and } |I_2| \leqq \lambda \big| |I_1| \leqq \lambda)}{P(|I_2| \leqq \lambda)}$$

$$\leqq \frac{P(I_1 \cap I_2 \neq \emptyset \text{ and } |I_2| \leqq \lambda \big| |I_1| = \lambda)}{P(|I_2| \leqq \lambda)} \leqq \frac{3}{2}\lambda$$

because the denominator equals $2\lambda-\lambda^2$ and the numerator is bounded above by $3\lambda^2$ as shown in Figure 4.

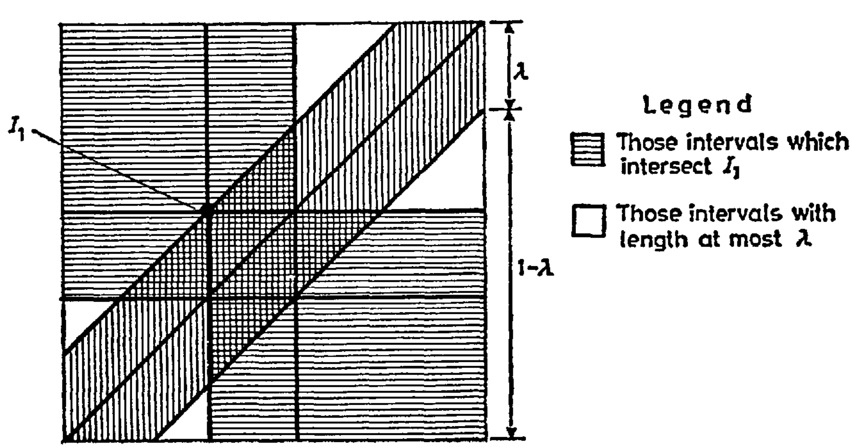

Fig. 4

For $1 \leqq i < j \leqq m$ let Y_{ij} be the random variable which is 1 if $I_i \cap I_j \neq \emptyset$ and 0 otherwise, and put $Y = \sum Y_{ij}$. Now

$$E(Y) = \binom{m}{p} p \leqq 3n^2 \lambda^3$$

and

$$E(Y^2) = \binom{m}{2} p + 6 \binom{m}{3} E(Y_{12} Y_{23}) + 6 \binom{m}{4} p^2 = F(Y)^2 + o(n^4 \lambda^6)$$

hence $\operatorname{Var}(Y) = o(n^4 \lambda^6)$. Now Chebyshev's inequality

$$P(|Y - E(Y)| \leqq t) \leqq \frac{\operatorname{Var}(Y)}{t^2} = \frac{o(n^4 \lambda^6)}{t^2} \to 0$$

when $t = \varepsilon n^2 \lambda^3$. Thus for almost all interval graphs,

$$Y \leqq (3+\varepsilon) n^2 \lambda^3.$$

For any graph H let $c(H)$ denote the number of connected components in H. Let G^* denote the induced subgraph of G on those vertices whose intervals have length at most λ. We conclude,

$$\alpha(G) \geqq \alpha(G^*) \geqq c(G^*) \geqq m - Y \geqq (2-\varepsilon) n\lambda - (3+\varepsilon) n^2 \lambda^3 \geqq \left(\frac{4}{9}\sqrt{2} - \varepsilon\right) \sqrt{n}. \quad \blacksquare$$

After generating several large random interval graphs on a computer we posit the following conjecture:

Conjecture. *For almost all interval graphs $G \in \mathcal{G}_n$ we have*

$$\alpha(G) = \sqrt{n} + o(\sqrt{n}).$$

Table 2. Erdős—Rényi random graphs vs. random interval graphs

Property	$p_{edge} = 2/3$	Random Interval Graphs
Number of edges	$\frac{1}{3} n^2 + o(n^2)$	$\frac{1}{3} n^2 + o(n^2)$
Hamiltonian?	Yes	Yes
Minimum Degree	$\frac{2}{3} n - o(n)$	$O(\sqrt{n})$
Maximum Degree	$\frac{2}{3} n + o(n)$	$n - O(1)$
Chromatic Number	$O\left(\frac{n}{\log n}\right)$	$\frac{n}{2} + o(n)$
Independence Number	$O(\log n)$	$O(\sqrt{n})$
Largest Clique	$O(\log n)$	$\frac{n}{2} + o(n)$

5. Comparison with Standard Models

Earlier we noted that the probability that two vertices of a random interval graph are adjacent equals 2/3. The table below summarizes our result and compares them with those of the Erdős—Rényi model in which $p_{\text{edge}}=2/3$.

Acknowledgements. The author would like to thank John Wierman for many useful comments and interesting discussions. Thanks also to Alan Karr, Robert Serfling and Michal Karonski. Research supported in part by Office of Naval Research contract number N00014-85-K0622.

References

[1] P. Billingsley, *Probability and Measure,* Wiley, New York (1979).
[2] B. Bollobás, *Random Graphs,* Academic Press, New York (1985).
[3] V. Chvatal, On Hamiltonian ideals, *J. Comb. Theory B* **12** (1972), 163—168.
[4] J. E. Cohen, The asymptotic probability that a random graph is a unit interval graph, indifference graph or proper interval graph, *Discrete Math.* **40** (1982), 21—24.
[5] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* **6** (1959), 290—297.
[6] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17—61.
[7] P. Erdős and D. West, A note on the interval number of a graph, *Discrete Math.* **55** (1985), 129—133.
[8] P. Gilmore and A. Hoffman, A characterization of comparability graphs and of interval graphs, *Canad. J. Math.* **16** (1964), 539—548.
[9] M. Golumbic, *Algorithmic Graphs Theory and Perfect Graphs,* Academic Press, New York (1980).
[10] E. Helly, Uber Mengen Kurper mit gemeinschaftlichen Punkten, *J. Deutsch Math. Verein* **32** (1923), 175—176.
[11] C. Lekkerkerker and J. Boland, Representation of a finite graph by a set of intervals on the real line, *Fund. Math.* **51** (1974), 45—64.
[12] E. Palmer, *Graphical Evolution: An Introduction to the Theory of Random Graphs,* Wiley, New York (1985).
[13] Lord, Rayleigh, On the resultant of a large number of vibrations of the same pitch and of arbitrary phase, *Philosophical Magazine and Journal of Science* **10** (1880), 73—78.

Edward R. Scheinerman

Department of Mathematical Sciences
The John Hopkins University
Baltimore, Maryland 21218 U.S.A.