Random Interval Graphs

Iliopoulos Vasileios

Contents

1 Abstract

In this thesis, which is supervised by Dr. David Penman, we examine random interval graphs. Recall that such a graph is defined by letting $X_1, \ldots X_n, Y_1, \ldots Y_n$ be $2n$ independent random variables, with uniform distribution on $[0,1]$. We then say that the ith of the n vertices is the interval $[X_i, Y_i]$ if $X_i < Y_i$ and the interval $[Y_i, X_i]$ if $Y_i < X_i$. We then say that two vertices are adjacent if and only if the corresponding intervals intersect.

We recall from our MA902 essay that fact that in such a graph, each edge arises with probability 2/3, and use this fact to obtain estimates of the number of edges. Next, we turn to how these edges are spread out, seeing that (for example) the range of degrees for the vertices is much larger than classically, by use of an interesting geometrical lemma. We further investigate the maximum degree, showing it is always very close to the maximum possible value $(n-1)$, and the striking result that it is equal to $(n-1)$ with probability exactly 2/3. We also recall a result on the minimum degree, and contrast all these results with the much narrower range of values obtained in the alternative 'comparable' model $G(n, 2/3)$ (defined later).

We then study clique numbers, chromatic numbers and independence numbers in the Random Interval Graphs, presenting (for example) a result on independence numbers which is proved by considering the largest chain in the associated interval order.

Last, we make some brief remarks about other ways to define random interval graphs, and extensions of random interval graphs, including random dot product graphs and other ways to define random interval graphs. We

also discuss some areas these ideas should be usable in. We close with a summary and some comments.

2 Acknowledgements

I would like to thank my supervisor, Dr. David Penman for his motivation and support. Also, I would like to express my sincere thanks to all my teachers in this academic year. All mistakes and errors in this work are my own.

3 Basic results

3.1 Definitions

We recall here the definitions of interval graph, and random interval graph, from our essay [2].

Definition 3.1 *Let a graph G have n vertices $\{1, 2, \ldots n\}$. To create an Interval Graph, to each vertex i we assign a finite interval I_i of the real line. This transformation yields n intervals of the real line. We then say that two different vertices of G are adjacent, in the initial graph, if the corresponding intervals have non-empty intersection. That is, $i \sim j \leftrightarrow I_i \cap I_j \neq \emptyset$.*

Definition 3.2 *A Random Interval Graph is formed as follows. Suppose the vertex set is $\{1, 2 \ldots n\}$. Let $2n$ independent and identically distributed continuous random variables, $X_1, X_2 \ldots X_n$ and $Y_1, Y_2 \ldots Y_n$, from the uniform distribution in $[0, 1]$, be given. The interval I_i will be $[X_i, Y_i]$, if $X_i < Y_i$ or $[Y_i, X_i]$, when $X_i > Y_i$. (The case where any two random variables are equal has probability 0, by properties of continuous distributions). We then say that the random interval graph is the interval graph formed for these vertices from the intervals I_i.*

We use Δ_n to denote the set of all possible random interval graphs on n vertices.

The main aim of this essay will be to prove various basic properties of these graphs, based on the two papers [9] by E. R. Scheinerman and [3]. We will first show some equivalent formulations of the model, which will be useful in proving theorems, and will then prove some results about various aspects of the graphs.

3.2 An Equivalent model

In [9] it is observed that we do not have to use the uniform random variables X_i and Y_i above. He observed that it is enough to suppose that the intervals have as their endpoints the numbers 1,2...2n in some random order, with all the $(2n)!$ possible orderings equally likely. The reason why this is equivalent is that, as we observed when defining random interval graphs, the probability that two of the X_i and Y_j are equal is zero, so they can be any set of unequal numbers. Also, because the X_i and Y_i are all independent, all $(2n)!$ possible orderings of them have the same probability. (Independence implies exchangeability - that is, the property that the probability that the X_is and Y_is take certain values is the same as the probability that the image of them all under some permutation in S_{2n}, the symmetric group on $2n$ letters, take these values). This concludes a proof of the following result, which will be used later in the proof that there is vertex of degree $(n-1)$ with probability $2/3$.

Lemma 3.1 *An equivalent definition of random interval graphs is to say they have vertex set $\{1, 2 \ldots n\}$ and that the intervals attached to the vertices have as their $2n$ endpoints the numbers $1, 2 \ldots 2n$ in some order, the order being chosen uniformly at random so that all $(2n)!$ possible orders are equally likely.*

Proof. See above.

3.3 Edges in interval graphs

The simplest question about a graph is how many edges it has. By Theorem 10 in [2], we know that the probability of any particular edge arising in a random interval graph is 2/3. We first extend this result to obtain an estimate of the total number of edges in a random interval graph, namely that it is near to $n^2(1 + o(1))/3$. This is the first main result in E. R. Scheinerman's paper [9].

It will be helpful to give an overview of the proof. It is easy to show that the expected number of edges is exactly $n(n-1)/3$ using the fact about the probability that an individual edge arises is 2/3. What we need to do is to show that it is very likely to be very close to this value - the precise sense of this will be made clear in the statement of the result. (The result is a limit result: many interesting results in probability, e.g. the law of large numbers and the Central Limit Theorem, are of this form). In order to show that it is very likely to be very close to $n(n-1)/3$, we shall look at the variance of the number of edges and show that it is small compared with $n(n-1)/3$. The way this will be done is by writing the number of edges as the sum of indicator variables, one for each edge. Evaluating the variance then involves considering the various possible values of the expectation of $X_{ij}X_{k\ell}$ for various possible i, j, k, ℓ. We shall see that in most cases they are independent, so the total variance is really rather small.

Theorem 3.2 *Almost all graphs in Δ_n have $n^2/3 + o(n^2)$ edges. More for-*

mally, letting X denote the number of edges in the graph,

$$\lim_{n\to\infty} P\left(X = \frac{n^2}{3}(1 + o(1))\right) = 1.$$

Proof. For every pair (i,j) with $1 \leq i, j \leq n$ and $i \neq j$, let the indicator random variable X_{ij} be equal to 1 if $i \sim j$ and be equal to 0 otherwise. (In other words, X_{ij} takes the value of 1, if the intervals corresponding to the vertices i and j intersect, so that the edge is present).

Now let $X = \Sigma_{1 \leq i < j \leq n} X_{ij}$. Then each edge which is present contributes 1 to this sum, and those absent contribute 0, thus X counts the total number of edges in the graph.

As we previously saw, $P(X_{ij} = 1) = 2/3$. Because X_{ij} is an indicator variable (i.e. takes only the values 1 and 0) we have

$$\mathbf{E}(X_{ij}) = \sum_{r=0}^{\infty} rP(X_{ij} = r) = 0 + 1 \times P(X_{ij} = 1) = P(X_{ij} = 1)$$

Thus

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}(\sum_{1 \leq i < j \leq n} X_{ij}) \\
&= \sum_{1 \leq i < j \leq n} \mathbf{E}(X_{ij}) \\
&= \sum_{1 \leq i < j \leq n} P(X_{ij} = 1) \\
&= \frac{n(n-1)}{2} \frac{2}{3} \\
&= \frac{n(n-1)}{3}
\end{aligned}$$

To estimate the number more precisely, we need also to have some grip on

the variance of X, namely

$$\text{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$$

$$= \mathbf{E}(\sum_{1 \leq i < j \leq n} X_{ij} \sum_{1 \leq k < \ell \leq n} X_{kl}) - (\frac{n(n-1)}{3})^2$$

(the last term is because we just worked out $\mathbf{E}(X)$)! Now we consider various cases for i, j, k and ℓ, counting - or at least estimating - carefully how many such cases there are.

Case 1. i, j, k, ℓ are all distinct. As $i < j$ and $k < \ell$, there are $\binom{n}{2}\binom{(n-2)}{2}/2$ possibilities (as the first two are chosen from all n, the second two from the remaining $(n-2)$. Thus we get $n(n-1)(n-2)(n-3)/4$ cases. In each of them, $\mathbf{E}(X_{ij}X_{kl}) = \mathbf{E}(X_{ij})\mathbf{E}(X_{kl}) = 2/3 \times 2/3 = 4/9$. This is because the two edges are independent, as they have no vertex in common.

Case 2. There is some overlap between set $\{i, j\}$ and $\{k, \ell\}$. There are however only a small number of these. Indeed since at least two of i, j, k and ℓ have to be equal, we are only choosing (at most) three numbers. Thus there are at most $\binom{n}{3}$ such cases. In each of these cases, the contribution to the sum made by $\mathbf{E}(X_{ij}X_{kl})$ is at most 1, simply because X_{ij} and X_{kl} are ≤ 1.

Then, calculating, we get

$$\text{Var}(X) = \mathbf{E}(\sum_{1 \leq i < j \leq n} X_{ij} \sum_{1 \leq k < \ell \leq n} X_{kl}) - (\frac{n(n-1)}{3})^2$$

$$\leq \frac{n(n-1)(n-2)(n-3)}{4} \frac{4}{9} \text{ from Case 1}$$

$$+\binom{n}{3} \times 1 \text{ from Case 2}$$

$$-(\frac{n(n-1)}{3})^2$$

$$= \frac{n(n-1)(n-2)(n-3)}{9} + \binom{n}{3} - \frac{n^2(n-1)^2}{9}$$

$$= \frac{n(n-1)}{9}[(n-2)(n-3) + \frac{9(n-2)}{6} - n(n-1)]$$

The above is, since the two terms in the square bracket involving n^2 cancel with each other, of the form

$$\frac{n(n-1)}{9}[Cn + D] \text{ for suitable constants } C, D$$

$$\leq En^3$$

for a suitable constant E.

Now we use Chebyshev's inequality, which says that, for any random variable X,

$$P(|X - \mathbf{E}(X)| \geq \epsilon \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

See [8]. Here we thus deduce that, for any $\epsilon = cn^2$, we have

$$P(|X - \mathbf{E}(X)| \geq cn^2 \leq \frac{\text{Var}(X)}{(cn^2)^2}$$

$$\leq \frac{En^3}{c^2n^4}$$

which, for large n, tends to 0. Thus with probability tending to 1, we do indeed get $(1 + o(1))n^2/3$ edges. $\bullet$

It is perhaps worth noting that it is natural to compare a random interval graph with another well-known model of random graphs. This is the so-called

Erdős-Rényi model $G(n, p)$ where there are n vertices and each edge arises with probability p *independently* of all other edges. This model is discussed in great detail in [1]. Clearly the most reasonable such model to compare random interval graphs with is $G(n, 2/3)$ as we know that in random interval graphs $2/3$ is the probability of each edge arising. The above result does have a simple analogue for Erdős-Rényi random graphs, namely that an Erdős-Rényi random graph has about $n^2/3$ edges. The proof in this case is much simpler: Indeed the law of large numbers, [11] says that the number of edges, divided by the total number of possible edges $n(n-1)/2$, is close to the expectation of any one of the indicators, namely $2/3$. Thus the number of edges is close to $n(n-1)/3$. The reason why this case is so much easier is that the edges in the Erdős-Rényi graph are independent, so that standard results like the law of large numbers, [11] apply to them. (It will emerge later that the edges are not independent, when we show that various things in a random interval graph usually take very different values from their values in $G(n, 2/3)$.)

4 Degrees in Random Interval Graphs

4.1 Introduction

We now turn our attention to the more detailed distribution of where the edges are. We shall consider the degrees of vertices and show that degrees are much more spread out that in the Erdős-Rényi model $G(n, 2/3)$. Indeed in the Erdős-Rényi model $G(n, 2/3)$, we shall see that, for all $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(\text{all vertices have degree between } \frac{2n(1-\epsilon)}{3} \text{ and } \frac{2n(1+\epsilon)}{3}\right) = 1.$$

That is, almost all degrees are about $2n/3$. However in random interval graphs, we shall see in various ways that degrees are much more 'spread out'. For example, the probability that there is a vertex of degree $n - 1$ (the maximum possible degree) is $2/3$.

4.2 Degrees of Graphs and some results

We recall first a definition from [2].

Definition 4.1 *The degree of a node of a graph is the number of vertices which are adjacent to this vertex. If v is a vertex, then the degree of v is denoted by* $\deg(v)$.

For a graph G, we define $\Delta(G)$ to be the maximum degree: that is,

$$\Delta(G) = \max_{1 \le i \le n} d(v_i).$$

Similarly we define the minimum degree

$$\delta(G) = \min_{1 \leq i \leq n} d(v_i).$$

We saw in the last section that a random interval graph has about $n^2/3$ edges. Our next task is to show how these edges are spread out. We will see a theorem from [9] about the degrees of random interval graphs.

Theorem 4.1 *Let $G \in \Delta_n$ and $v \in V(G)$. For a known $x \in [0,1]$, we have for $x \geq 1/2$*

$$\lim_{n \to \infty} P(d(v) \leq xn) = 1 - (1-x)\frac{\pi}{2}$$

and for $x < 1/2$ we have

$$\lim_{n \to \infty} P(d(v) \leq xn) = 1 - (1-x)(\pi/2 - 2\cos^{-1}[1/\sqrt{2-2x}]) - \sqrt{1-2x}$$

Note what the theorem means. It says that, for example, taking x to be (say) 0.01, the probability that there is a vertex of degree $\leq 0.01n$ is, in the limit as $n \to \infty$, strictly positive. This is very different from what happens in $G(n, 2/3)$ where, as mentioned earlier, all the degrees are close to $2n/3$. Similarly it says that the probability that a vertex does not have degree $\leq 0.99n$ - that is, that its degree is at least $0.99n$ - is, in the limit, a non-zero number. So the degrees are indeed much more spread out than in $G(n, 2/3)$.

The proof will rely on the following geometrical lemma. We give a rather detailed proof of the Lemma as no details are provided in Scheinerman's article [9]. Some of the details here were suggested to me by Dr. Penman [6].

Lemma 4.2 *For an interval $I = [x, z] \subset [0, 1]$ let the radius of I, $\rho(I)$ be*

14

$\sqrt{a^2 + (1-b)^2}$, where $a = \min\{x, z\}$ and $b = \max\{x, z\}$. Assume that x and z are independent, and uniformly distributed on $[0, 1]$. Then, for $y \leq 1/2$

$$P(\rho^2(I) \leq y) = \frac{y\pi}{2}.$$

When $y > 1/2$, we have that

$$P(\rho^2(I) \leq y) = y(\frac{\pi}{2} - 2cos^{-1}[\frac{1}{\sqrt{2y}}]) + \sqrt{2y - 1}.$$

Proof. The required probability is the probability that $\rho(I) \leq \sqrt{y}$, which is the size of the set of points in the square $[0, 1]^2$ which are within a distance $\sqrt{y}$ from (at least) one of the points $(0, 1)$ or $(1, 0)$. Let us consider all the possible cases for $0 \leq y \leq 1$. The equation of the circle centered at $(1, 0)$ is $(x - 1^2) + z^2 = y$. In the second circle, centered at point $(0, 1)$ we have the equation $x^2 + (z - 1^2) = y$. So if they intersect at (x, z), then

$$(x - 1)^2 + z^2 = x^2 + (z - 1)^2 = y$$
$$\Rightarrow x^2 - 2x + 1 + z^2 = x^2 + z^2 - 2z + 1$$
$$\Rightarrow -2z = -2x \Rightarrow z = x$$

Thus, these points satisfy:

$$(x - 1)^2 + x^2 = y$$
$$\Rightarrow 2x^2 - 2x + 1 = y$$
$$\Rightarrow x^2 - x + \frac{1 - y}{2} = 0$$

15

For such points to exist, we need that the discriminant is greater than 0. Thus,

$$\Rightarrow (-1)^2 - 4\frac{(1-y)}{2} \geq 0$$
$$\Rightarrow 1 - 2(1-y) \geq 0$$
$$\Rightarrow (2y - 1) \geq 0$$
$$\Rightarrow y \geq \frac{1}{2}$$

Now let consider the two cases for the value of the radius.

Case 1. Let $y < \frac{1}{2}$, then $\sqrt{y} < 1/\sqrt{2}$ is the radius of the two circles centered at either the point $(0, 1)$ or $(1, 0)$. Because $y < \frac{1}{2}$, these two circles do not intersect by the above analysis. Then the area we want is two quarters of disjoint circles which have radius equal to $\sqrt{y}$. Having in mind that the area of a circle with radius χ is equal to $\pi \times \chi^2$, then the area of a quarter of circle is $\frac{1}{4}\pi \times \sqrt{y}^2 = \frac{1}{4}\pi \times y$. So the required area is equal to $\frac{2\pi \times y}{4} = \frac{\pi \times y}{2}$, as stated.

Case 2. $y \geq 1/2$. In this case, the circles have intersection and by the above we have for $x = z$, that:

$$x^2 - x + \frac{(1-y)}{2} = 0 \tag{1}$$
$$\Rightarrow (x - \frac{1}{2})^2 + \frac{(1-y)}{2} - \frac{1}{4} = 0 \tag{2}$$

16

$$\Rightarrow (x - \frac{1}{2})^2 + \frac{(1 - 2y)}{4} = 0 \tag{3}$$

$$\Rightarrow (x - \frac{1}{2})^2 + (\frac{1}{4} - \frac{y}{2}) = 0 \tag{4}$$

$$\Rightarrow x = \frac{1}{2} \pm \sqrt{\frac{y}{2} - \frac{1}{4}} \tag{5}$$

In this case, as previously noted, the two circles intersect in two points, A and B say. We consider the triangle with vertices $(0, 1)$, A and B and the triangle with vertices $(1, 0)$, A and B, also the arcs of the two circles between A and B. Let θ be the angle formed at $[0, 1]$ and $[1, 0]$ by the two triangles (by symmetry, the two angles are the same). The area of each sector of each circle is equal to

$$\pi \times y \frac{\theta}{2\pi} = \frac{\theta y}{2}.$$

Also the area of the triangle is $\frac{1}{2} \times |CA| \times |CB| \times \sin(\theta)$, where C Is the point $(0, 1)$ and A,B are the intersection points, between the two circles.

Then the total area of intersection of the two sectors is:

$$2[\frac{\theta \times y}{2} - \frac{1}{2}|CA||CB|\sin\theta]$$
$$= \theta y - |CA||CB|\sin\theta$$

To obtain the angle θ, recall that the vertices are: $C = (0, 1)$, $B = (\frac{1}{2} + \sqrt{\frac{y}{2} - \frac{1}{4}}, \frac{1}{2} + \sqrt{\frac{y}{2} - \frac{1}{4}})$ and $A = (\frac{1}{2} - \sqrt{\frac{y}{2} - \frac{1}{4}}, \frac{1}{2} - \sqrt{\frac{y}{2} - \frac{1}{4}})$.

Now $|AB|^2 = |AC|^2 + |CB|^2 - 2|AC||CB|\cos\theta$ by the cosine rule, [4]. The vectors are

$$\overrightarrow{AB} = (2\sqrt{\frac{y}{2} - \frac{1}{4}}, 2\sqrt{\frac{y}{2} - \frac{1}{4}})$$

17

$$\overrightarrow{CB} = (\sqrt{\frac{y}{2} - \frac{1}{4}} + \frac{1}{2}, \sqrt{\frac{y}{2} - \frac{1}{4}} - \frac{1}{2})$$

$$\overrightarrow{AC} = (-\sqrt{\frac{y}{2} - \frac{1}{4}} + \frac{1}{2}, -\sqrt{\frac{y}{2} - \frac{1}{4}} - \frac{1}{2})$$

$$\Rightarrow AB^2 = [2\sqrt{\frac{y}{2} - \frac{1}{4}}]^2 + [2\sqrt{\frac{y}{2} - \frac{1}{4}}]^2$$

$$= 8(\frac{y}{2} - \frac{1}{4}) = 4y - 2$$

Similarly,

$$|CB|^2$$

$$= (\sqrt{\frac{y}{2} - \frac{1}{4}} + \frac{1}{2})^2 + (\sqrt{\frac{y}{2} - \frac{1}{4}} - \frac{1}{2})^2$$

$$= 2[\frac{y}{2} - \frac{1}{4}] + \frac{1}{2} = y$$

and then $|CA|^2 = y$ just by symmetry. Then the cosine rule, [4] becomes

$$(4y - 2) = y + y - 2\sqrt{y}\sqrt{y}\cos\theta$$

$$\Rightarrow 4y - 2 = 2y - 2y\cos\theta$$

$$\Rightarrow \cos\theta = \frac{2y - 4y + 2}{2y}$$

$$\Rightarrow \theta = \cos^{-1}[-1 + \frac{1}{y}]$$

(Note that, as $y \geq 1/2$, $-1 + 1/y \leq 1$ as required). Then, the formula for the area of the intersection of the two quarter-circles is (note that here we use $\cos^{-1}$ to mean the inverse function to cos, what many people call arccos:

in particular, $\cos^{-1}(x)$ does *not* mean $1/\cos(x)$)

$$y\cos^{-1}[-1+\frac{1}{y}] - \sqrt{y}\sqrt{y}\sin[\cos^{-1}(-1+\frac{1}{y})]$$

$$= y\cos^{-1}[-1+\frac{1}{y}] - y\sqrt{1-(-1+\frac{1}{y})^2}$$

$$= y\cos^{-1}[-1+\frac{1}{y}] - \sqrt{y^2-(-y+1)^2}$$

$$= y\cos^{-1}[-1+\frac{1}{y}] - \sqrt{2y-1}$$

To explain the working in the last paragraph; We used $\cos^2\theta + \sin^2\theta = 1$, [4] and also the fact that in our situation the angle θ is clearly between 0 and $\pi/2$ so that both cos and sin are positive, with the result that $\sin(x) = \sqrt{1-\cos^2(x)}$. Thus $\sin[\cos^{-1}(-1+\frac{1}{y})]$ is equal to $\sqrt{1-\cos^2(\cos^{-1}(-1+\frac{1}{y}))}$ which is of course equal to $\sqrt{1-(-1+\frac{1}{y})^2}$.

What we have just worked out is the formula for the area of the intersection of the two quarter-circles. Thus the shaded area in Scheinerman's picture is the areas of the two individual quarter circles minus the area of their intersection, which of course is

$$\frac{\pi y}{2} - \left(y\cos^{-1}[-1+\frac{1}{y}] - \sqrt{2y-1}\right) = \frac{\pi y}{2} - y\cos^{-1}[-1+\frac{1}{y}] + \sqrt{2y-1}$$

It only remains to confirm that this formula we have just derived is the same as the one given in Scheinerman's article, namely

$$y(\frac{\pi}{2} - 2cos^{-1}[\frac{1}{\sqrt{2y}}]) + \sqrt{2y-1}.$$

For this in turn it suffices to show that

$$2\cos^{-1}[\frac{1}{\sqrt{2y}}] = \cos^{-1}[-1 + \frac{1}{y}].$$

To see this, take cosines of both sides, and recalling the identity $\cos(2x) = 2\cos^2(x) - 1$, [4] we see the left-hand side is:

$$2\cos^2(\cos^{-1}[\frac{1}{\sqrt{2y}}]) - 1$$
$$= 2[\frac{1}{\sqrt{2y}}]^2 - 1$$
$$= -1 + \frac{1}{y}$$

as required. ●

Proof of Scheinerman's theorem. (see Theorem 4.2 in [9]: we provide some more details). Let v corresponds to vertex 1: no generality is lost by this, as no vertex is favored by the set-up. Let J_i be the interval assigned to vertex i, for each $i = 2\dots n$. Then let $I_i = 1$, if 1 is adjacent to i and be equal to 0 otherwise. Thus $X = \sum_{i=2}^{n} I_i$ is the degree of the vertex 1.

Suppose now that $\rho(I_1) = r$ is fixed. We claim that then:

$$p = P(I_i = 1 | \rho(I_1) = r) = 1 - r^2.$$

When we have proved this, it is then obvious that, conditional on $\rho(I_1) = r$, the expectation of X is $(n-1)p$ and its variance (again conditioned on the value of $\rho(I_1)$) is $(n-1)p(1-p)$ since the I_i are independent of each other given I_1 and the radius. Indeed, if $i \neq j$, then the two random variables X_i

and Y_i giving the two endpoints of I_i are independent of X_j and Y_j giving the endpoints of I_j.

Thus we can use Chebyshev's inequality again on the random variable X, [8] and deduce that:

$$P(|X - (n-1)p| \geq n^{2/3}) \leq \frac{\text{Var}(X)}{n^{4/3}}$$

$$\leq \frac{np(1-p)}{n^{4/3}} \to 0$$

Then, $d(1) = np + o(n)$, under the hypothesis that $\rho(I_1) = r$. Thus for any $\epsilon > 0$ we have:

$$P(d(1) \leq xn) = \begin{cases} 1 - o(1) \text{ for } r < \sqrt{1-x} - \epsilon \\ o(1) \text{ for } r > \sqrt{1-x} + \epsilon \end{cases}$$

using our formula for the value of p for a given value of r. So now we need to remove the conditioning on the value of r, which we do in the usual manner:

$$P(d(1) \leq xn) = \int_0^1 P(d(1) \leq xn|p(I_1) = r)dP(\rho(I_1) \leq r) \text{ by the law of total probability}$$

But we have just worked out the distribution function for the probability that $\rho(I_1) \leq \sqrt{y}$ in the Lemma. So this is:

$$[1 - o(1)]P(\rho^2(I_1) \leq 1 - x) + \epsilon O(1) + o(1)$$

$$\Rightarrow P(\rho^2(I_1) \leq 1 - x).$$

Now the result follows using the Lemma, with y replaced by $1 - x$.

Thus the only thing that remains to be proved is that, if $\rho(I_1) = r$, then the probability that any other interval intersects is $1 - r^2$. To obtain this, recall that in [9], the radius of an interval $[x, y] \subset [0, 1]$ is equal to $\sqrt{a^2 + (1 - b)^2} = r$, where $a = \min\{x, y\}$ and $b = \max\{x, y\}$. Also, we should remind ourselves that the $(2n)$ random points follow the uniform distribution in $(0, 1)$. Let $I_1 = [a, b]$. We calculate the possibility of the existence of i interval, denoted by $I_i = [X_i, Y_i]$, which does not intersect with I_1. To happen this, both X_i, Y_i must be smaller than a or both must be greater than b. The first possibility is: $P(X_i, Y_i \leq a) = \frac{a-0}{1-0}\frac{a-0}{1-0} = a^2$, since they are independent, uniformly distributed random variables in $(0, 1)$. The second possibility is equal to:

$$P(X_i, Y_i \geq b)$$

$$= P(X_i \geq b)P(Y_i \geq b)$$

$$= (1 - P(X_i \leq b))(1 - P(Y_i \leq b))$$

$$= (1 - \frac{b - 0}{1 - 0})(1 - \frac{b - 0}{1 - 0})$$

$$= (1 - b)(1 - b)$$

$$= (1 - b)^2.$$

Thus, the probability of $I_1 \cap I_i = \emptyset$, for $i = 2, 3 \ldots n$ is equal to $a^2 + (1-b)^2 = r^2$. Hence the complement probability (I_1 and I_i have a non-null intersection) is $1 - r^2$. $\bullet$

4.3 Maximum degree

In the last section, we saw there is a non-zero probability that a vertex has very high degree (e.g. at least $0.99n$). In this section, we first sharpen this. We then give a result from [3] which [9] attempted to prove, but did not succeed. That result is, that with probability $2/3$, there is a vertex in a random interval graph whose degree is $n - 1$. This is of course the largest possible degree any vertex in the graph can have.

The first result (which is Theorem 4.4. in Scheinerman) is easy.

Theorem 4.3 *In a random interval graph, let ω_n be any function which tends to infinity with n. (One should think of it as doing so very slowly). Then*

$$\lim_{n\to\infty} P\left(\text{a random interval graph has } \Delta \geq n - \omega(n) \right) = 1.$$

Proof. Some details of that proof were suggested to me by Dr. Penman [6], as Scheinerman's proof for that Theorem is rather short. Let $x = \frac{1}{2}\sqrt{\frac{\omega_n}{n}}$. Then

$$P(\geq n - \frac{\omega_n}{2} + o(\omega_n) \text{ intervals intersect } [x, 1-x])$$
$$= 1 - P(\geq n - \frac{\omega_n}{2} + o(\omega_n) \text{ intervals don't intersect} [x, 1-x])$$
$$= 1 - P(\text{interval } I_1 \text{ doesn't intersect } [x, 1-x])^n$$
$$= 1 - [(\frac{1}{2}\sqrt{\frac{\omega_n}{n}} + \frac{1}{2}\sqrt{\frac{\omega_n}{n}})^2]^n$$

as the probability a random interval does not intersect $[x, 1-x]$ is the probability that both ends are less than x (probability $x \times x = x^2$, using the

independence of the ends) or that both ends are greater than $1 - x$ (probability $[1 - (1 - x)]^2 = x^2$). The above is

$$= 1 - (\frac{\omega_n}{n})^n$$

$$\to 1 \text{ as } n \geq \omega_n \text{ when } n \to \infty$$

Thus at least $n - \omega_n/2 + o(\omega_n)$ intervals intersect $[x, 1-x]$. Our proof will now be complete if we can show that, with probability tending to 1 as $n \to \infty$, there is at least one of the random intervals which contains $[x, 1-x]$, as then such an interval will be a vertex of degree $\geq n - \omega(n)$.

To this end, recall the fact that random intervals are independent, as all the possible orderings of their endpoints are equally likely. Let X_i be an indicator variable, denoting whether the interval $I_i = [A_i, B_i]$ contains $[x, 1-x]$. Then $X = \sum_{i=1}^{n} X_i$ is the number of our n random intervals which contain $[x, 1 - x]$ and our objective is to show that $X > 0$ with probability tending to 1. To this end we use Chebyshev's inequality, [8]

$$P(|X - \mathbf{E}(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

$$\Rightarrow P(X = 0) \leq P(|X - \mathbf{E}(X)| \geq \mathbf{E}(X))$$

$$\leq \frac{Var(X)}{(\mathbf{E}(X))^2}$$

Now X_i is a Bernoulli variable and the various X_i are independent of each other, simply because the distinct intervals are independent of each other. Thus X is binomial, with parameters n and the success probability a. We

thus have, by standard results about Binomial random variables,

$$\mathbf{E}(X) = na \text{ and } \mathrm{Var}(X) = \sum_{i=1}^{n} Var(X_i) = na(1-a)$$

and thus by Chebyshev, [8]

$$P(X = 0) \leq \frac{Var(X)}{(\mathbf{E}(X))^2}$$
$$= \frac{na(1-a)}{n^2 a^2} = \frac{(1-a)}{na}$$

Thus if we can show that $P(X = 0)$ tends to zero, this will show that $P(X > 0)$ tend to 1.

We need to calculate a. Recall

$$a = P(X_i = 1) = P(\text{a given random interval contains } [x, 1-x])$$

Letting the two (random) ends of the interval be A and B, we have that this probability is

$$a = P(\{A \leq x \text{ and } B \geq 1-x\} \cup \{A \geq 1-x \text{ and } B \leq x\})$$
$$= P(A \leq x \text{ and } B \geq 1-x)) + P(A \geq 1-x \text{ and } B \leq x)$$

as the two events involved are mutually exclusive. This in turn gives

$$a = P(A \leq x)P(B \geq 1-x) + P(A \geq 1-x)P(B \leq x\}) \text{ by independence}$$

$$= x(1 - x) + x(1 - x) = 2x(1 - x)$$
$$= 2\frac{1}{2}\sqrt{\frac{\omega_n}{n}}(1 - \frac{1}{2}\sqrt{\frac{\omega_n}{n}})$$
$$= \sqrt{\frac{\omega_n}{n}}(1 - \frac{1}{2}\sqrt{\frac{\omega_n}{n}})$$

Hence, combining the above results,

$$P(X = 0) \le \frac{1 - a}{na} \le \frac{1}{na}$$
$$= \frac{1}{n\sqrt{\omega_n/n}\left(1 - \sqrt{\omega_n/(2n)}\right)}$$
$$= \frac{1}{n^{1/2}\sqrt{\omega_n}(1 - \sqrt{\omega_n/(2n)})}$$
$$\le \frac{2}{n^{1/2}\sqrt{\omega(n)}} \text{ for large enough n}$$

using the fact that

$$1 - \frac{1}{2}\sqrt{\frac{\omega_n}{n}} \ge \frac{1}{2}$$

for large enough n. Thus

$$P(X = 0) \le \frac{2}{n^{1/2}\omega(n)} \to 0 \text{ as } n \to \infty$$

and thus with probability tending to 1, $X > 0$, that is there is some interval containing $[x, 1 - x]$ as required. ●

The final result is the following striking fact. We emphasize that, unlike most of the results in this chapter, this has nothing to do with a limit: it is an *exact* result, not depending on the number of vertices. The proof comes from [3]. Scheinerman made a lot of effort in his paper [9] to prove a result

along these lines, but did not quite succeed.

Theorem 4.4 *Let G be a random interval graph. Then*

$$P(\Delta(G) = n - 1) = \frac{2}{3}.$$

Proof. Instead of calculating the moderately difficult expression

$$1 - 4n(n-1) \int_0^1 \int_0^{1-y} xy(1 - x^2 - y^2 - 2xy)^{n-2} - 1\, dxdy$$

writers [3] used an efficient combinatorial proof, as we shall see. They take random pairs of integers $1, 2, \ldots, 2n$. Once the intervals are selected by some random pairing of the $2n$ numbers, they label the endpoints $A(1), B(1), \ldots, A(n-2), B(n-2)$ in the following way. Let the endpoints $\{1, \ldots, n\}$ be at the left side and respectively the endpoints $\{n+1, \ldots, 2n\}$, be at the right side. Let also $A(1) = n$ and $B(1)$ is its mate. Suppose that we have assigned through $A(j)$, $B(j)$. We label the next endpoints, by the following rules:

Case 1. If $B(j)$ is on the left side, then let $A(j+1)$ be the leftmost point on the right side that has not yet been labeled. Let $B(j+1)$ be its mate.

Case 2. If $B(j)$ is on the right side, then let $A(j+1)$ be the rightmost point on the left side that has not yet been labeled. Let $B(j+1)$ be its mate.

Endpoints are being labeled from the center outwards. Then, if $A(j) < B(j)$, it is $A(j+1) < B(j)$

If $A(j) > B(j)$, it is $A(j+1) > B(j)$

Then, we will either have:

$$A(j) < A(j+1)$$

$$\text{or } A(j) > A(j+1)$$

Furthermore, if $A(j) < B(j)$, then $A(j+1) < B(j+1)$

If $A(j) > B(j)$, then $A(j+1) > B(j+1)$

With this way, starting from the center labeling the endpoints, we deduce that either an equal number of points have been assigned in both sides, or two more points have been assigned on the left than on the right side. Since the last endpoints assigned are $A(n-2)$ and $B(n-2)$, from the total number of points, which is equal to $2n$, there are four remaining points unlabeled, namely $a < b < c < d$. Having a specific ordering, it is considered all the possible ways of pairing them to consist two random intervals. a can be matched with b, c or d, with equal probabilities. Thus, we have 3 possible cases. We easily observe that in two cases of pairing the remaining points, the corresponding intervals intersect and in one case they are disjoint. Let now a and b be on left and c and d on the right. If a is paired with c, then the random interval $[a, c]$ meets all the others. Also the same happens when a is paired with d. This is because, we assumed the points $\{1, 2 \ldots n\}$ lie on left and the remaining points $\{n+1 \ldots 2n\}$ lie on the right and by construction of labeling them. On the other hand, if a is paired with b, then $[a, b] \cap [c, d] = \emptyset$. Suppose that an interval $[e, f]$ intersects all the others. Also let $A(j) = e$ and

$B(j) = f$. In this case, where a and b are on the left, $A(j)$ lies between b and c. Thus, $[e, f]$ cannot intersect both $[a, b]$ and $[c, d]$. Furthermore, consider the case where only a is on left. Since $[e, f]$ meets $[c, d]$ we have $f > c$, hence $f = B(j)$. Again, by construction if a is paired with b, then $[a, b] \cap [c, d] = \emptyset$. On the other hand, in cases where a is paired with c or d the corresponding intervals intersect. The probability of pairing a with c or d, in the specific ordering of the four endpoints, which is $a < b < c < d$ is 2/3. Finally, we see that the probability in a family of n random intervals, the maximum degree has value $n - 1$ (i.e the probability that an interval meets all the others) is 2/3. ●

4.4 Minimum degree

The previous subsection makes it clear that the maximum degree in a random interval graph is much bigger than $2n/3$, which is roughly the expected number of neighbors of each vertex. As we will see at the end of this chapter, this is different from $G(n, 2/3)$ where all degrees are about $2n/3$. We now say a little about minimum degrees, presenting a Theorem from [9]. We omit the proof of this result, as it is moderately difficult.

Theorem 4.5 *Let k be a fixed, non-negative real number and δ denotes the minimum degree of the graph. We have,*

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

Note that Theorem 4.1 implies the result below, which is noted in [9].

Corollary 4.6 *For every $\epsilon > 0$ sufficiently small, almost all interval graphs*

satisfy $\delta < \epsilon \cdot n$ and $\Delta > (1 - \epsilon) \cdot n$.

This result is an immediate corollary of Theorem 4.1.

4.5   Degrees of vertices in $G(n, 2/3)$

In the previous subsections, we studied the minimum and maximum degrees of Random Interval Graphs. We now give, for contrast, the result for the Erdős-Rényi model, where the probability of an edge arising is constant, equal to 2/3 and the edges are independent. Here it will turn out that all the degrees are 'about' $2n/3$.

Theorem 4.7 *In $G(n, \frac{2}{3})$ and for $\epsilon > 0$ sufficiently 'small', all vertices have degree between about $(\frac{2}{3} - \epsilon)n$ and $(\frac{2}{3} + \epsilon)n$. More precisely,*

$$\lim_{n \to \infty} P(\text{all vertex degrees} \in [(\frac{2}{3} - \epsilon)n, (\frac{2}{3} + \epsilon)n]) = 1$$

Some details of this proof were suggested by [6]. Also, for the proof of this Theorem, we use the 'Large Deviations' Lemma from Scheinerman, [9].

Lemma 4.8 *If p is constant and $\epsilon > 0$, then*

$$P(|X - np| \geq \epsilon np) \leq \frac{a_\epsilon e^{-b_\epsilon pn}}{\sqrt{np}}$$

Where a_ϵ, b_ϵ are positive constants, which depend only on ϵ: that is, not on n or p.

Proof of Theorem 4.7. Let the random variable X denotes the number of vertices having degrees not in the interval

$$[(\frac{2}{3} - \epsilon)n, (\frac{2}{3} + \epsilon)n].$$

X is the sum of n independent random variables $X_i, i = 1, 2 \ldots n$ where X_i is 1 if vertex i has degree not in the stated range and is zero otherwise. We again aim to show that $X = \sum_{i=1}^{n} X_i$ is 0 with probability tending to one as $n \to \infty$, and to prove this we will use the fact that

$$P(X > 0) = \sum_{i=1}^{\infty} P(X = i) \leq \sum_{i=1}^{\infty} iP(X = i) \leq \mathbf{E}(X).$$

So we aim to show $\mathbf{E}(X) \to 0$, for which in turn it suffices, as

$$\mathbf{E}(X) = \mathbf{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbf{E}(X_i)$$
$$= n\mathbf{E}(X_1) \text{ as the } X_i \text{ are identically distributed}$$
$$= nP(X_1 = 1)$$

to show that $P(X_1 = 1)$ is $o(n)$.

But the degree of vertex i in $G(n, \frac{2}{3})$ has binomial distribution with parameters $n - 1$ and $\frac{2}{3}$. Thus, using the Large Deviations Result, [9]

$$P[\text{degree of vertex i} \notin [(\frac{2}{3} - \epsilon)n, (\frac{2}{3} + \epsilon)n]]$$
$$\leq \frac{a_\epsilon e^{-b_\epsilon \frac{2}{3}n}}{\sqrt{\frac{2n}{3}}}$$
$$\Rightarrow E(X_i) = P(X_i = 1) \times 1 + P(X_i = 0) \times 0 = P(X_i = 1)$$

31

$$E(X_i) = P(X_i = 1) \leq \frac{a_\epsilon e^{-b_\epsilon \frac{2}{3}n}}{\sqrt{\frac{2n}{3}}}$$

$$\Rightarrow E(X) = \sum_{i=1}^{n} E(X_i) \leq n \frac{a_\epsilon e^{-b_\epsilon \frac{2}{3}n}}{\sqrt{\frac{2n}{3}}}$$

And this is indeed $o(n)$ for large n, just because the exponential terms converge more rapidly than polynomial. $\bullet$

Remark. In fact the above argument can be sharpened quite substantially: a random graph $G(n, 2/3)$ has the property that there is an explicit constant C such that

$$\lim_{n \to \infty} P\left(\frac{2n}{3} - C\sqrt{n \log(n)} \leq \delta(G) \leq \Delta(G) \leq \frac{2n}{3} + C\sqrt{n \log(n)}\right) = 1.$$

The proof is in [1] (but is much harder than the above proof).

5 Cliques, independent sets and chromatic numbers in Random Interval Graphs

5.1 Cliques

In this section we study the clique number $\omega(G)$ of a random interval graph G. Recalling from [2], the clique number of a graph G, denoted as $\omega(G)$ is the order of the largest complete subgraph of G. The basic result is from [9], Theorem 4.7.

Theorem 5.1 *The clique number of a random interval graph is usually about $n/2$. More precisely,*

$$\lim_{n \to \infty} P\left(\omega(G) = \frac{n}{2} + o(n)\right) = 1.$$

Proof. First note that a maximum clique consists of a family of intervals, each pair of which intersect. In our essay, [2] we showed that such a family of intervals has some point x say, which is in all the intervals. This result is known as Helly's Theorem.

It is intuitively obvious that the point with the best chance of being in several intervals is $x = 1/2$: let us be more formal about this now, using the argument from [3]. The probability that a random interval $[X, Y]$ does not contain x is the probability that both X and Y are less than x, which has probability x^2, or the (exclusive) possibility that both are greater than x, which is $(1-x)^2$: thus the probability that it does contain x is $1 - x^2 - (1-x)^2 = 2x - 2x^2$. This is maximized for $x = 1/2$, as the derivative of $2x - 2x^2$ is $2 - 4x$ which is zero exactly when $x = 1/2$, when the probability that an

interval contains x is $2 \cdot 1/2 - 2 \cdot (1/2)^2 = 1/2$. In a different sense, since the endpoints of the n random intervals follow the the uniform distribution in $[0,1]$, then with probability equal to 1 will be distinct. Moreover, the mean number of a variable, which follows the uniform distribution is equal to $\frac{0+1}{2} = \frac{1}{2}$. As a result, the expected number of intervals is indeed $n/2$. The number of n random intervals containing $1/2$ is binomial: $\text{Bin}(n, 1/2)$. That is because each random interval in $(0,1)$, either it will contain $1/2$ with probability p say, or it will not contain it, with probability $1 - p$. We have n independent Bernoulli trials so the number of intervals containing $1/2$ is indeed a Binomial random variable. This takes a value very close to $n/2$, as we saw above: in particular, it is $n/2 + o(n)$. This gives a lower bound on the clique number.

We now need to show that it is not more than $n/2$. Consider intervals of the form $I_i = [i/n^2, (i+1)/n^2]$. We need the following technical result (which Scheinerman calls the 'medium deviations lemma' for binomial random variables):

Lemma 5.2 *If $X = \sum_{i=1}^n X_i$ where $X_i = 1$ with probability p and is zero otherwise, and the X_i are independent, then for*

$$1 \leq h < \min\left\{ \frac{np(1-p)}{10}, \frac{(pn)^{2/3}}{2} \right\}$$

then

$$P\left(|X - np| \geq h \right) \leq \frac{\sqrt{np(1-p)}}{h} \exp\left(-\frac{h^2}{2np(1-p)}\right).$$

The above lemma comes from [9].

How does this help us? Let us work out the probability that more than $n/2 + o(n)$ of the n random intervals intersect I_i. (We are aiming to show that this does not happen). The probability that a random interval $[X, Y]$ does not intersect I_i is the probability that both X and Y are less than i/n^2 (which is i^2/n^4), or that both are bigger than $(i+1)/n^2$ which has probability $(1 - \frac{i+1}{n^2})^2$. Thus the success probability (i.e the probability that it does intersect I_i) is

$$
1 - \frac{i^2}{n^4} - \left(1 - \frac{i+1}{n^2}\right)^2
$$
$$
= 1 - \frac{i^2}{n^4} - (1 - \frac{2(i+1)}{n^2} + \frac{(i+1)^2}{n^4})
$$
$$
= \frac{2(i+1)}{n^2} - \frac{i^2}{n^4} - \frac{(i+1)^2}{n^4}
$$

Hence the number X of the n random intervals which intersect I_i is binomial with n trials and success probability

$$
\frac{2(i+1)}{n^2} - \frac{i^2}{n^4} - \frac{(i+1)^2}{n^4}.
$$

We need to work out which value of i maximizes this expression. Let us treat the more general problem of which (continuous) value of x maximizes

$$
f(x) = \frac{2(x+1)}{n^2} - \frac{x^2}{n^4} - \frac{(x+1)^2}{n^4}.
$$

35

Differentiating, we get

$$f'(x) = \frac{2}{n^2} - \frac{2x}{n^4} - \frac{2(x+1)}{n^4}$$
$$= \left(\frac{2}{n^2} - \frac{2}{n^4}\right) - \frac{4x}{n^4}$$

so the turning point is at $f'(x) = 0$,

$$x = \frac{n^4}{4}\left(\frac{2}{n^2} - \frac{2}{n^4}\right) = \frac{n^2 - 1}{2}.$$

Since $f''(x) < 0$, this turning point is a maximum. Note that at it we have

$$f(x) = \frac{n^2 + 1}{n^2} - \frac{(n^2 - 1)^2}{4n^4} - \frac{(n^2 + 1)^2}{4n^4}$$
$$= 1 + \frac{1}{n^2} - \frac{1}{4} + \frac{1}{2n^2} - \frac{1}{4n^4} - \frac{1}{4} - \frac{1}{2n^2} - \frac{1}{4n^4}$$
$$= \frac{1}{2} + \frac{1}{n^2} - \frac{1}{2n^4}$$
$$= \frac{1}{2} + \frac{2n^2 - 1}{2n^4}$$

The point about this number is that it is very close to $1/2$.

We now (rather arbitrarily) consider 0.6. Note that $n^{0.6} < (np)^{2/3}/2$ and $n^{0.6} < np(1-p)/10$ for large enough values of n. The same holds with 0.6 replaced by 0.61. (This is checked so that we can apply the medium deviations lemma in a minute). We have

$$P(X \geq n/2 + n^{0.6})$$
$$\leq P(\text{Bin}(n, 1/2 + (2n^2 - 1)/n^4) \geq n/2 + n^{0.6})$$

since the success probability for the latter binomial is the largest possible value of the success probability under the constraints.

This we now apply the medium deviations result to, taking $h = n^{0.6}$ and $t = h/\sqrt{np(1-p)}$ where $p = 1/2 + (2n^2 - 1)/n^4$. Note that given $\epsilon > 0$, $p < 1/2 + \epsilon$ for all large enough n. Thus $p(1-p)$ is $\geq (1/2 + \epsilon)(1/2 - \epsilon) = 1/4 - \epsilon^2$, and on the other hand is $\leq 1/4$. Thus t is between $\sqrt{1/4 - \epsilon^2}n^{0.1}$ and $1/2n^{0.1}$. Thus

$$\leq P(\text{Bin}(n, 1/2) \geq n/2 + n^{0.6}) \leq \frac{1}{t}\exp(\frac{-t^2}{2})$$

and this is, by the above estimates for t,

$$\leq \frac{1}{\sqrt{1/4 - \epsilon^2}n^{0.1}}e^{-(1/4-\epsilon^2)n^{0.2}/2} \to 0 \text{ as } n \to \infty$$

and this completes the proof. •

The following simple consequence is also contained in Theorem 4.7 of [9].

Corollary 5.3 *The chromatic number of a random interval graph satisfies*

$$\lim_{n\to\infty} P\left(\chi(G) = \frac{n}{2} + o(n)\right) = 1.$$

Informally: it is usually approximately n/2.

Proof. Interval graphs are perfect, and all perfect graphs have $\omega(G) = \chi(G)$, as we saw in our essay [2]. The result now follows from the previous theorem. •

Here is a comparison of the result with what happens for $G(n, 2/3)$, though we do not give the proof. It turns out that the clique number of $G(n, 2/3)$ is (with probability tending to 1 as $n \to \infty$) about $2 \log_{3/2}(n)$, which of course is far smaller than for the random interval graphs. Also, in Erdős-Rényi, the chromatic number is (again, with probability tending to 1 as $n \to \infty$) about $n/(2 \log_3(n))$, which is of course much larger than the clique number. In particular, the Erdős-Rényi graph is very far from being a perfect graph, since the chromatic number is (for large n) so much larger than the clique number. Note that in both the case of the clique number and the case of hte chromatic number, we get noticeably larger answers for random interval graphs than we do for Erdős-Rényi graphs. We refer to [1] for detailed statements of these two results.

Here we present a reformulation from [3] of the result we have just proved about the clique number.

Theorem 5.4 *Let the random variable A_n denotes the size of the largest set of pairwise intersecting intervals in a family of n random intervals. There exists a function $f(n)$ such that:*

$$\lim_{n \to \infty} \frac{f(n)}{n} = 0$$
$$and \ \lim_{n \to \infty} P(\frac{n}{2} - f(n) \le A_n \le \frac{n}{2} + f(n)) = 1$$

5.2   Independent sets in random interval graphs

Recall from [2] that an independent set in a random interval graph is the same as a chain in the random interval order associated with it. We thus investigate

the length of the longest chain in the partial order, so as to obtain a result on the size of the largest independent set. Let a graph with n vertices where the set of edges is equal to the emptyset, then the corresponding random interval graph consists of n disjoint intervals. Thus, the size of the maximum chain is equal to n. The ordered intervals form a chain, in the sense that they have a null intersection. On the other hand, in a complete graph where all the edges are existent, the chain is equal to the null set, as all the intervals intersect. Again, this material is based on [3] sharpening results in [9]. Recalling from [2] a partially ordered set is a non-empty set, which has the mathematical property of order. Taking randomly two elements, say x and y, then if we have $x < y$ or $y < x$ we define these elements as being *comparable* forming a *chain.* In other case, we define them as *incomparable*, forming an *antichain.* In the case of a random interval order, where for two intervals $[a, b]$ and $[c, d]$, we say $[a, b] \prec [c, d]$ if $b < c$. Thus an independent set in the random interval graph corresponds to a set of non-intersecting intervals.

Theorem 5.5 *Let Y_n denote the maximum number of pairwise disjoint intervals in a family of n random intervals. Then,*

$$\lim_{n \to \infty} \frac{Y_n}{\sqrt{n}} = \frac{2}{\sqrt{\pi}}$$

in probability.

The proof of this Theorem comes from [3].

Proof. We generate a Poisson process with intensity 1 in the upper right quadrant. Thus, since the probability function for a discrete Random Vari-

able X, which follows the Poisson distribution is $\frac{e^{-\lambda}\lambda^x}{x!}$, where λ is the expected value of X, the probability that, for any positive number s, the region $\{(x,y) : 0 \leq x, y \leq s\}$ does not contain any point of the process is equal to e^{-s^2}.

We are going to choose an infinite chain of points of the process. Let $C = \{(\ell_1, u_1), (\ell_2, u_2), \ldots\}$, the points being chosen as follows. Let (ℓ_1, u_1) be the point that minimizes $\max\{\ell_1, u_1\}$ and thereafter (ℓ_k, u_k) is the point above (ℓ_{k-1}, u_{k-1}) that minimizes $\max\{\ell_k, u_k\}$. Thus the points of the chain are chosen subject to the restrictions that they are monotonically increasing and the difference between every two 'neighboring' points of the chain minimum. Thinking of each point in the chain as defining an interval, it is easy to see that thus this chain is built up from the bottom by always choosing the next interval to be the one with least possible upper endpoint. It is not hard to check by induction the intuitively reasonable claim that, in any finite collection of intervals, this chain will have the maximum possible length.

Then, if S is a variable, whose value is $\max(x_1, y_1)$ the mass density function of S is given by:

$$f(s) = \frac{d}{ds}(1 - e^{-s^2}) = 2se^{-s^2}.$$

This is because we have

$$P(\max(x_1, y_1) \leq s) = 1 - P(x_1 \geq s \text{ and } y_1 \geq s)$$
$$= 1 - P(\text{the square } (0,0), (0,s), (s,0) \text{ and } (s,s) \text{ contains no point of the process})$$
$$= 1 - e^{-s^2} \text{ as observed above, using that the intensity is } 1$$

Thus $F(s) = P(S \leq s) = 1 - e^{-s^2}$. Thus its density $f(s)$ is the derivative of this with respect to s, which is indeed as stated.

Therefore, we have

$$
\begin{aligned}
E(S) &= \int_0^\infty 2se^{-s^2}s.ds \\
&= \int_0^\infty t^{\frac{1}{2}}e^{-t}dt \\
&= \Gamma(3/2) = \frac{\sqrt{\pi}}{2}
\end{aligned}
$$

We now claim the differences

$$X_1 = \max(\ell_1, u_1) - 0, \, X_2 = \max(\ell_2, u_2) - \max(\ell_1, u_1), \, X_3 = \max(\ell_3, u_3) - \max(\ell_2, u_2), \ldots$$

are independent and identically distributed with mean $\frac{\sqrt{\pi}}{2}$. To see this, we have just proved this for the first difference. Now we, so to speak, move the origin to (ℓ_1, u_1) and use the homogeneity of the Poisson process to get that the variables are identically distributed. Independence follows from the independence properties of the Poisson process.

Therefore, by the Law of Large Numbers, [11] for any $\epsilon > 0$ sufficiently small,

$$
\lim_{n\to\infty} P\left((1-\epsilon)\frac{\sqrt{\pi}}{2} < \frac{X_1 + \ldots X_n}{n} < (1+\epsilon)\frac{\sqrt{\pi}}{2}\right) = 1
$$
$$
\Rightarrow \lim_{n\to\infty} P\left((1-\epsilon)\frac{\sqrt{\pi}}{2} < \frac{\max(x_m, y_m)}{m} < (1+\epsilon)\frac{\sqrt{\pi}}{2}\right) = 1
$$

just by simplifying the telescoping sum in the definition of the X_i.

41

Let $r(n)$ denotes the minimum r, such the area $[0, r]^2$ contains exactly n points of the Poisson Process. Then these points determine n random intervals. (Conditional on the number of points of a Poisson process in a certain area being given, the points themselves are uniformly distributed over that area).

Recall that we are studying Y_n, the size of the largest independent set in a random interval graph with n intervals, that is the longest chain of intervals in the corresponding interval order. Thus the above remarks show that we can identify Y_n with the largest m such (ℓ_m, u_m) lies in the area $[0, r(n)]^2$.

Now, because the Poisson process has density 1, we have

$$\lim_{n \to \infty} P\left((1 - \epsilon)\sqrt{n} < \ell(n) < (1 + \epsilon)\sqrt{n}\right) = 1$$

using the Law of Large Numbers [11]. Thus, if now we let

$$m_1 = [(1 - \epsilon)(\frac{2}{\sqrt{\pi}})\sqrt{n}] \text{ and}$$
$$m_2 = [(1 + \epsilon)(\frac{2}{\sqrt{\pi}})\sqrt{n}]$$

we see by the previous results that, for n sufficiently large, the point (ℓ_{m1}, u_{m1}) will lie inside the square $[0, r(n)]^2$ and the point (ℓ_{m2}, u_{m2}) will lie outside of this area, with probability tending to 1. Thus we indeed get

$$\lim_{n \to \infty} P\left((1 - \epsilon)\frac{2}{\sqrt{\pi}} < \frac{Y_n}{\sqrt{n}} < (1 + \epsilon)\frac{2}{\sqrt{\pi}}\right) = 1$$

which completes the proof. ●

We again compare this with the result for $G(n, 2/3)$. Here it turns out that the independence number is, with probability tending to 1 as $n \to \infty$, about $2 \log_3(n)$. Again we refer to [1] for a proof of this fact. Note again that this number is much smaller in the $G(n, 2/3)$ than in the random interval graph.

5.3 Comparison of different models

Here we discuss the differences, which arise in two different models of random graphs, namely the Erdős-Rényi model and the random interval graphs. In the first model, the possibility of an edge arising is a constant equal to 2/3, independent from the number of edges. Moreover from [9] we present other features of random interval graphs, as the value of minimum, maximum degree for the two models, chromatic and independence numbers. As we previously saw, for the Erdős-Rényi model the minimum and maximum degrees are $\frac{2}{3}n - o(n)$ and $\frac{2}{3}n + o(n)$ respectively and all the degrees are roughly close to $2n/3$. On the other hand, from [9] we have that the minimum degree is equal to $O(\sqrt{n})$ and the maximum degree is about $n - 1$ for the ordinary model. Moreover chromatic and independence numbers are $O(\frac{n}{\log n})$ and logarithmic, $O(\log n)$ for Erdős-Rényi model, see [9]. Also, the clique number is equal to independence number. Furthermore, in this model random graphs are not perfect, as the chromatic number is not equal with the clique number. In random interval graphs chromatic number and clique number is about to $\frac{n}{2}$, as we saw in above sections. Finally, from [9] independence number is equal to $O(\sqrt{n})$. This comparison between these models clearly shows the differences arising in their characteristic values, which determine their properties.

6 Variants

6.1 Introduction

In this rather miscellaneous section, we discuss an alternative way to define random interval graphs, a recent generalization by Scheinerman of these graphs and some applications of them.

6.2 Scheinerman's generalization

Scheinerman has recently introduced a common generalization of both random interval graphs and the Erdős-Rényi graphs, namely random dot product graphs. Also from [10] he gives various definitions of interval graphs.

To understand Scheinerman's idea, we first introduce *Intersection Graphs*. Suppose we have a finite set of n vertices, V_n. At each vertex $v \in V_n$, we have a subset $S_v \subseteq \mathbf{R}$ (here, as usual, $\mathbf{R}$ is the set of the real numbers). We now say that two vertices are adjacent if and only if the corresponding sets have a non-null intersection. In mathematical notation

$$v \sim w \iff S_v \cap S_w \neq \emptyset.$$

So an interval graph is a special kind of intersection graph, with the set S_v for each vertex v being an interval of the real line.

Moreover, in [10] random intersection graphs are introduced, by assigning randomly sets S_v to the vertices, and then we again say that two vertices are adjacent if their corresponding sets intersect. The usual way to assign these sets is to say that each S_v is a subset of $\{1, 2, \ldots k\}$ with, for each v,

$P(i \in S_v) = p$ say, independently for $1 \leq i \leq k$ and each vertex choosing its subset independently. However there are other possibilities.

A further model studied in [10] are *Threshold Graphs*. Here, For every vertex v, we assign a number x_v. Then, two vertices intersect if and only if the sum of the corresponding numbers is ≥ 1. Again, in mathematical notation, $v \sim w \Leftrightarrow x_v + x_w \geq 1$. Again we can have random threshold graphs by generating the x_v in some random way.

The main business of [10] is to give a new definition of a model of random graphs which combines all these definitions, by using dot products. Here, vertex v is assigned a d-dimensional vector of real numbers X_v. Then, two vertices are adjacent, if the corresponding inner product of the vectors is ≥ 1. Mathematically,

$$v \sim w \iff X_v \bullet X_w \geq 1$$

where $\bullet$ denotes inner (dot) product:

$$(x_1, x_2, \ldots x_d) \bullet (y_1, y_2, \ldots y_d) = \sum_{i=1}^{d} x_i y_i.$$

The idea behind the definition of random dot products is that various ways of defining random interval graphs can be replaced by the random dot product. Indeed, we have that each vertex i is randomly assigned a d-dimensional vector X_i. Here $d \in \mathbf{N}$ is fixed. The vectors themselves can be generated from some d-dimensional distribution: this could be each component chosen independently, but there are other possibilities as well. We now say that $i \sim j$ with probability $f(X_i \bullet X_j)$ for some fixed, and carefully chosen, function f.

This is a general definition, which generalizes several of the definitions above:

Erdős-Rényi graphs: generalized because if we take, for every vertex v, $X_v = \mathbf{x} = (x, x, \ldots x)$ where $\mathbf{x} \bullet \mathbf{x} = p$, and $f(x) = x$, a moment's thought will show that we recover the $G(n, p)$ model.

Random intersection graphs with each $S_v \subseteq \{1, 2, \ldots k\}$: because, if we take X_i to be the vector whose jth component is 1 if $j \in S_v$, and whose jth component is 0 otherwise, then the property that two vertices v and w are adjacent if and only if $S_v \cap S_w \neq \emptyset$ can be written as the property that $v \sim w$ with probability $f(X_v \bullet X_w)$, where $f(t)$ is 0 if $t = 0$ and is 1 otherwise.

(Note: Observant readers will have observed that this is only a generalization, in the strict sense, of the random intersection graphs in the case when each $S_v \subseteq \{1, 2 \ldots k\}$, whereas of course to get intersection graphs to generalize interval graphs we have to have the S_v being infinite sets, namely certain intervals of the real line. However it is certainly a generalization in spirit of the idea).

An attractive feature of this very general definition is that we can combine random and non-random ideas in giving the definitions of the vectors, according to the situation we are working in.

Scheinerman [10] starts by giving some results for the case when $d = 1$ and the 'vectors' (really, in this case, scalars, so we will denote them by the lower case letter) x_i are uniformly distributed on $[0, 1]$. He takes $f(t) = t^r$

for some fixed r: these assumptions will remain in force throughout this paragraph. Now we have

$$P[i \sim j] = f(x_i \bullet x_j) = \int_0^1 \int_0^1 (x_i x_j)^r dx_i dx_j = \frac{1}{(1+r)^2}$$

since this is the average, over all possible values of x_i and x_j, of $f(x_i x_j)$. This is, just by simple integrations,

$$\int_{x_i=0}^1 x_i^r dx_i \int_{x_j=0}^1 x_j^r dx_j$$
$$= [\frac{x_i^{r+1}}{r+1}]_{x_i=0}^1 [\frac{x_i^{r+1}}{r+1}]_{x_j=0}^1$$
$$= \left(\frac{1}{r+1} - 0\right)\left(\frac{1}{r+1} - 0\right)$$
$$= \frac{1}{(1+r)^2}$$

Thus, the expected number of edges is $\frac{n(n-1)}{2}(1+r)^{-2}$, since it is the expectation of a sum of $n(n-1)/2$ indicator variables of whether each edge is present, each indicator having expectation $1/(1+r)^2$. He also presents a short calculation, the details of which we omit, showing that if $a \sim b$ and $b \sim c$, then conditional on this information $P(a \sim c)$ is larger than it would be unconditionally: that is, there is a 'clustering'effect. He believes, but cannot at present prove, that the degrees in the graph follow a 'power law': that is, letting $N(d)$ denote the number of vertices of degree d, a plot of $\log(N_d)$ against $\log(d)$ should be a roughly straight line with negative gradient. Scheinerman observes that various large networks arising in real life have been observed to have this property (at least roughly). He further ob-

tains that the expected number of isolated vertices is $C_r n^{(r-1)/r}$ for a suitable constant $C_r > 0$. In particular, these graphs are not connected: however, they do have a very large component and a few isolated vertices. Further they have diameter at most 6. (The diameter of a graph is the worst case of the distance between two points in it).

Moreover, Scheinerman [10] introduces the 'inverse problem'. Given a graph on a specific set of vertices, which vectors are more suitable to model his graph? An obvious approach is to say that the best choice of Xs are those which maximize the likelihood function. This is doable in dimension 1, though in higher dimensions it becomes very unpleasant fairly rapidly. He thus suggests an alternative approach based on matrix theory, the Gram Matrix Approach. In detail: given $G_1, G_2 \ldots G_m$ let $A = \frac{1}{m} \sum_{j=1}^{m} A(G_j)$

$$a_{i,j} \approx P[i \sim j] = x_i x_j (i \neq j)$$

$$X = [x_1, x_2 \ldots, x_n]$$

$$A = X^t X$$

6.3 Prisner's definition

In [7], E. Prisner proposes the following question:

'What other reasonable models, apart from Scheinerman's, are there for random interval graphs? For example, suppose we choose n unit intervals (i.e. intervals of length 1) which are chosen from the interval $[0, m]$ and are chosen

uniformly and at random?'

We are not aware of any substantial work on this model.

6.4 Applications of Random Interval Graphs

The obvious applicability of random interval graphs is to scheduling and assignment problems. For example, suppose each of n people in an office has an interval each day when he is free for a meeting. The exact size of this interval, and its position in the period of the working day (late, or early, or whatever) will vary from day to day, so can be modeled as random. Then, if we have a random interval graph whose vertices are the n individuals and where two vertices (individuals) are adjacent if and only if their random intervals intersect, then we are saying that these two individuals will have a chance to meet on that day: and, for example, the largest clique in the random interval graph will be the largest number of people who can all meet. Similarly, the largest independent set will be the largest set of people, no two of whom can meet. Of course, the assumption that the intervals are uniformly distributed over the working day is probably not very realistic: for example, most people will be unavailable for some time over lunch, and in practice they will have several time intervals at which they are available, rather than just one. ('I am available between 9.00 and 10.00, and between 2.00 and 3.00'). However it is a reasonable first model.

Similarly, if we have a series of jobs to carry out in a factory. Suppose for example we are making a car or similar. Various tasks - say, n of them - have to be carried out during the production process (for example: painting the

outside, installing the radio, checking the braking system, etc.), and usually we cannot be doing more than one of these things at once. There will be time intervals during a working day in which the people qualified to carry out the various tasks (brake testing, painting etc.) are available: again, these will be hard to predict in advance, so can be modeled as random. Again, we will then want to have a large independent set in the graph, as that means we have the corresponding time intervals are disjoint, so there are no clashes. That is, by doing one of the jobs in its time interval, we are not reducing our chance of getting one of the other jobs done that day. (If the painter and the radio installer are available in disjoint time intervals, then we know that we can just get on with doing the painting and this will not reduce our chances of getting the radio installed that day as well).

Moreover, from my essay, [2] interval graphs are widely used in resource allocation problems. That is, we want to allocate a fixed amount of assets in production activity, in order to maximize profit. This is a problem in the field of combinatorial optimization. Another application of interval graphs is their usefulness in many problems of this field of discrete Mathematics. An example is the traveling salesman's problem. A salesman leaves his home and he is willing to visit n towns. He then has to consider $n!$ alternative feasible tours. We want to find the optimal tour, so as to visit each town only once and to minimize the relevant cost of traveling. Obviously, this is a challenging problem, as the set of feasible tours is too vast. (For example, if he has to visit 4 towns, the number of feasible tours is 24. If he has to visit say 6 towns, then we have to consider 6!=720 different tours!). Consider, to each town we assign a vertex. Then, two vertices are adjacent if the salesman

leaves the town and goes to the other. Then, to each vertex let's assign an interval. Clearly, if two intervals intersect, there is an edge arising. But, it is a cost associated to the salesman tour. So, we aim to find the tour having minimum cost, so all the n towns will be visited. Also interval graphs have many other applications, as we shall see.

The textbook by McKee and McMorris [5] in Chapter 3 contains references to various applications of interval graphs in Biology, Psychology and Computing. In Biology, for example, from [5] a prominent application of Interval Graphs is the physical mapping of DNA. From a DNA sequence, some fragments, which are called *clones* are obtained and the goal is to reconstruct the placement of the clones: that is, where they are on the DNA string. Thus a clone is an interval of a line of DNA.

To turn this into a problem about random interval graphs, we assign to each clone a vertex. Two different clones are adjacent if and only if their corresponding intervals intersect. This clearly gives an interval graph.

Also, interval graphs are used in social sciences. For example, in Psychology, [5] they are widely used as tools, measuring notions, which determine different psychological theories. (Most theory of measurement is based on physical science: however, in the social sciences, different theories may be more appropriate).

An example from [5] is that a person has a set A of alternatives solutions to choose. For simplicity, suppose that elements of A are different makes of cars. Our person has preferences among the the different makes of cars: for example, he might prefer stylish cars or cars having low cost of service, etc). Then, a real-valued function f on the set A, such that for $a, b \in A$,

he prefers alternative a than b when $f(a) > f(b) + \delta$, where δ is a positive constant representing a threshold - a 'just noticeable' difference between the two kinds of cars. Then, we define a binary relation R on A to be an interval order on the set A of alternatives, if it satisfies the two following axioms. (We are thinking here of aRb as meaning that a is preferable to b).

Axiom 1: For all $a \in A$, not aRa

Axiom 2: For all $a, b, c, d \in A$, if aRb and cRd, then either aRd or cRb

The motivation is that a car is not preferable to itself (clearly), which gives Axiom 1. Similarly, if a is preferable to b and c preferable to d, it seems reasonable that at least one of a is preferable to d and c is preferable to c should hold.

Furthermore, interval graphs are used in Computing. They are used in scheduling problems. From [5] we have an interesting application of this class of problems. Suppose that we want to find an arrangement, in order to construct a timetable for different courses in a University. We have a fixed number of rooms available for teaching purposes and we know the number of teachers. We aim to construct an efficient timetable, so to be no overlap between teaching hours for every lesson. We assign various courses to vertices. Then, two vertices are adjacent, when the corresponding intervals intersect. When this is the case, we have two different courses at the same hour. Thus, we want to find the minimum number of rooms needed, in order all the courses to be taught. In "graph language" we want to find the chromatic number. That is the minimum number of colors needed, so two

connected vertices have different colors, [5]. Moreover, in problems related to information retrieval, we use interval graphs. Suppose that Φ denotes set of files, which contain information and Q is the set of queries for retrieving information. Then, Φ and Q satisfy the *consecutive retrieval property* if the files relevant to each query can be stored consecutively in a linear form, so not to be overlap, [5]. We easily deduce from the above the extensive use of interval graphs in various, different sciences, from areas of applied Mathematics to social sciences, as Psychology.

7 Conclusions: areas for further work

In this project, we examined Random Interval Graphs, emphasizing Scheinerman's definition. We presented some results on the number of edges and then considered the degrees of vertices in the graphs, observing that these are much more spread out that in the alternative Erdős-Rényi model of random graphs: for example, the maximum degree of a random interval graph is very likely to be close to $n - 1$, and indeed is equal to $n - 1$ with probability $2/3$. We then considered cliques, independent sets and chromatic numbers of random interval graphs, obtaining asymptotic estimates of each of the quantities involved and comparing their values with the values in the Erdős-Rényi model. Finally, we wrote about other ways of defining Random Interval Graphs, such as Scheinerman's random dot product graphs, emphasizing the 1-dimensional case of Scheinerman's theory. Finally we outlined some areas of application.

There are some questions left unanswered by our work. For example, it would be desirable to investigate measures of connectivity (such as vertex-connectivity or edge-connectivity) in random interval graphs. Also: what is the diameter of a random interval graph? The obvious guess would be that it is 2, since if we have two vertices v and w, one would hope that one of the many vertices of high degree (close to $n - 1$) will be adjacent to both of them. Certainly the probability that the diameter is 2 is at least $2/3$, since if there is a vertex of degree $n - 1$ is till be adjacent to both v and w (if either v or w is a vertex of degree $n - 1$, then it is clearly adjacent to the other). The result that the probability is at least $2/3$ is now just a consequence of

the fact that with probability $2/3$ there is a vertex of degree $n-1$. However the guess that the diameter of a random interval graph is 2 with probability tending to 1 does not seem to follow immediately from what we have proven, because it is not quite clear that we can avoid the situation where there are two vertices v and w of low degree and all the vertices of high degree fail to be adjacent to at least one of v and w.

Another topic is the existence of Hamilton cycles in a random interval graph. Scheinerman ([9]) shows, by a rather long and difficult argument, that with probability tending to 1, a random interval graph is Hamiltonian: that is, it has a cycle which passes through every vertex of the graph. A $G(n, 2/3)$ is also Hamiltonian: indeed $G(n, 2/3)$ has the stronger property that it has (with probability tending to 1) $\lfloor \delta(G)/2 \rfloor$ edge-disjoint Hamilton cycles: we refer to [1] for a proof of this fact. (Two cycles are edge-disjoint if and only if there is no edge which is in both cycles). Note that $\lfloor \delta(G)/2 \rfloor$ is the largest number of edge disjoint Hamilton cycles we could have in a graph G, because each Hamilton cycle will use up two edges in passing through a vertex v of degree $\delta(G)$. Is it true that a random interval graph has $\lfloor \delta(G)/2 \rfloor$ edge-disjoint Hamilton cycles? Again this does not seem to be obvious.

We hope that we have given a reasonable material of random interval graphs, some interesting results on them, and some ideas of how they might be useful.

References

[1] Bollobas, B. *Random Graphs.* Academic Press (1985).

[2] Iliopoulos V.*Introduction to Interval Graphs.* MA902 essay. May 2005 (unpublished).

[3] Justicz J., Scheinerman E. R, Winkler P. M. Random Intervals. The American Mathematical Monthly, Vol. 97, No. 10 (Dec. 1990), 881-889.

[4] Manura, David. Webpage
`www.math2.org/math/trig/identities.htm`

[5] McKee, T. A; McMorris, F. R. Topics in intersection graph theory. SIAM monographs on discrete maths and applications 2, Philadelphia, 1999.

[6] Dr. D. B. Penman, personal communications.

[7] Prisner, E. Webpage
`http://www.math.uni-hamburg.de/spag/gd/mitarbeiter/prisner/Pris/Random.html`

[8] Weisstein, E. W. `www.mathworld.wolfram.com/ChebyshevInequality.html`

[9] E. R Scheinerman. Random Interval Graphs. Combinatorica Vol 8 357-371 (1988).

[10] E. R Scheinerman. `www.ipam.ucla.edu/publications/gss2005/gss2005_498.ppt`

[11] `www.wikipedia.org/wiki/Law_of_large_numbers`.