Combinatorics, Probability and Computing (1999) 8, 131–159. Printed in the United Kingdom

On Random Intersection Graphs: The Subgraph Problem

MICHAŁ KAROŃSKI[1]†, EDWARD R. SCHEINERMAN[2]‡
and KAREN B. SINGER-COHEN[3]‡

[1] Faculty of Mathematics and Computer Science,
Adam Mickiewicz University, Poznań, Poland
and
Department of Mathematics and Computer Science,
Emory University, Atlanta, GA 30322, USA
(e-mail: karonski@math.amu.edu.pl and michal@mathcs.emory.edu)

[2] Department of Mathematical Sciences, The Johns Hopkins University,
Baltimore, MD 21218–2689, USA
(e-mail: ers@cs.jhu.edu)

[3] Department of Mathematical Sciences, The Johns Hopkins University,
Baltimore, MD 21218–2689, USA
and
School of Mathematics, University of Minnesota,
Minneapolis, MN 55455, USA
(e-mail: singer@math.umn.edu)

A new model of random graphs – *random intersection graphs* – is introduced. In this model, vertices are assigned random subsets of a given set. Two vertices are adjacent provided their assigned sets intersect. We explore the evolution of random intersection graphs by studying thresholds for the appearance and disappearance of small induced subgraphs. An application to gate matrix circuit design is presented.

1. Introduction

1.1. The model

In most models of random graphs, the edges enjoy all the attention and the vertices are passive bystanders. In Erdős–Rényi random graph theory, we are given n vertices, and flip coins to see where the edges go – the appearance of one edge is independent of any other. Such a model is useful when the 'relations' between 'objects' are independent of one another.

† Research supported in part by the Komitet Badań Naukowych, grant 2 P03A 023 09.
‡ Research supported in part by the Office of Naval Research.

In this paper, we explore a model of random graphs in which the vertices are the focus. We independently assign to each vertex a random structure and then assess the adjacency of two vertices by comparing their assigned structures. To do this, we use the concept of an *intersection graph*.

Let G be a (finite, simple) graph. We say that G is an *intersection graph* if we can assign to each vertex $v \in V(G)$ a set S_v so that $vw \in E(G)$ (we write $v \sim w$) exactly when $S_v \cap S_w \neq \emptyset$. In this case, we say G is the intersection graph of the family of sets $\mathscr{S} = \{S_v : v \in V(G)\}$. It is easy to check that every graph is an intersection graph (see [15]).

If one restricts the choices for the sets S_v, various classes of graphs can be defined; the best known example is the class of *interval graphs*, in which the S_v must be real intervals. (See [17, 18] for a discussion of *random interval graphs*.)

We are now ready to define *random intersection graphs*. Let n, m be positive integers and let $p \in [0, 1]$. For every positive integer k with $1 \leqslant k \leqslant n$, let S_k be a random subset of $M = \{1, 2, \ldots, m\}$ formed by selecting each element of M independently with probability p. Thus the probability that we choose a particular set S for k is $p^s(1-p)^{m-s}$, where $s = |S|$. Finally, let $G(n, m, p)$ be the intersection graph of the S_ks.

Thus $G(n, m, p)$ has n vertices $\{1, 2, \ldots, n\}$. We assign to each vertex k a random subset (as described above) $S_k \subseteq \{1, 2, \ldots, m\}$ and we have $i \sim j$ if and only if $S_i \cap S_j \neq \emptyset$.

Now, given two vertices u and v of $G(n, m, p)$ the probability that there is an edge connecting them is $\Pr\{u \sim v\} = 1 - (1-p^2)^m$, since the probability that S_u and S_v are disjoint is simply $(1-p^2)^m$. It follows that the expected number of edges in $G(n, m, p)$ is $\binom{n}{2}\left[1-(1-p^2)^m\right] \asymp n^2mp^2$ (provided $mp^2 \to 0$ as $n \to \infty$). If we take $p = 1/(\omega_n n\sqrt{m})$, where ω_n denotes hereafter a function that goes to infinity with n, then the expected number of edges goes to 0 in the limit and with high probability[1] $G(n, m, p)$ is edgeless. Further, it follows from our results below that, when $p = \omega_n/(n\sqrt{m})$, then with high probability $G(n, m, p)$ has edges.

On the other hand, the expected number of *non-edges* in $G(n, m, p)$ is $\binom{n}{2}(1-p^2)^m \asymp n^2\exp\{-mp^2\}$. Thus, if we take $p = \sqrt{\frac{2\log n + \omega_n}{m}}$, then with high probability $G(n, m, p)$ is a complete graph. Further, when $p = \sqrt{(2\log n - \omega_n)/m}$, we show below that with high probability $G(n, m, p)$ is not complete.

Thus we may restrict our attention to values of p in the range between $1/(n\sqrt{m})$ and $\sqrt{2\log n/m}$. As p increases from the former to the latter, we witness the evolution of the structure of $G(n, m, p)$.

An alternative view of random intersection graph generation is given by its *representation matrix* $R(n, m, p)$. This matrix is an $n \times m$ matrix whose rows represent the vertices of $G(n, m, p)$ and whose columns represent the elements of the universal set $M = \{1, \ldots, m\}$. The entries in $R(n, m, p)$ are 0s and 1s; each entry is independently 1 with probability p (and 0 with probability $1-p$). From the random representation matrix $R(n, m, p)$ we derive the graph $G(n, m, p)$ by deeming two vertices to be adjacent if and only if their

[1] As is customary in random graph theory, by *with high probability* we mean that the probability that $G = G(n, m, p)$ has the stated property tends to 1 as $n \to \infty$.

corresponding rows have a 1 in a common column (*i.e.*, their dot product is nonzero). Note that a given graph G may arise from many different representation matrices.

More formally, we let $\mathscr{R}(n,m,p)$ denote the sample space of all $n \times m$ 0,1-matrices with the probability of a particular matrix set to $p^a(1-p)^b$, where a is the number of 1s and b is the number of 0s in the particular matrix. Now we let $\mathscr{G}(n,m,p)$ be the sample space of all graphs on n labelled vertices $\{1,2,\ldots,n\}$. The probability of a particular graph G in $\mathscr{G}(n,m,p)$ is the sum of the probabilities of all matrices in $\mathscr{R}(n,m,p)$ that represent G.

Thus the random intersection graph $G(n,m,p)$ is an element of the sample space $\mathscr{G}(n,m,p)$.

We have seen how $G = G(n,m,p)$ arises from $R = R(n,m,p)$ by concentrating on the *rows* of R. A dual view of G is afforded by examining the *columns* of R. Consider a given column of R. The 1s in this column correspond to a collection of pairwise adjacent vertices in G, that is, a *clique*[2] of G. This said, we can think of $G(n,m,p)$ as being generated by the following random process. For $j = 1,\ldots,m$ do the following. Let C_j be a random subset of $V(G) = \{1,\ldots,n\}$ with each $i \in C_j$ independently with probability p. Having generated the sets $\{C_1, C_2, \ldots, C_m\}$, we declare vertices u and v to be adjacent exactly when they are together in a common C_j. In other words, the family $\mathscr{C} = \{C_1, C_2, \ldots, C_m\}$ is a *clique cover* of G, that is, a family of cliques of G with the property that every edge of G is induced in at least one of the C_js. The concept of clique covers is central to our discussion of subgraphs of $G(n,m,p)$.

The representation matrix $R(n,m,p)$ has yet another interpretation. In fact it can be viewed as the adjacency matrix of a random binomial bipartite graph $B(n,m,p)$ in which edges occur independently between vertices in the two parts $N = \{1,2,\ldots,n\}$ and $M = \{1,2,\ldots,m\}$ with probability p. A random intersection graph with the vertex set N is recovered from $B(n,m,p)$ as follows. We put an edge between two vertices u and v of $G(n,m,p)$ if and only if there is a vertex z in the M-part of $B(n,m,p)$ such that both $\{u,z\}$ and $\{v,z\}$ are edges of $B(n,m,p)$. Such an approach provides a useful relationship between the classical binomial model of a random graph with independent edges and the random intersection graph $G(n,m,p)$ where the edges are no longer independent.

We are interested in studying the properties of $G(n,m,p)$ for n large. We therefore have two 'parameters' that we can adjust: m and p. As we discuss below, when m is very small compared to n, the model is not particularly interesting, and when m is exceedingly large (compared to n) the behavior of $G(n,m,p)$ is essentially the same as for an Erdős–Rényi random graph. The 'right' balance is achieved when we take $m = \lfloor n^\alpha \rfloor$ where α is a positive constant, and this is the m we shall use. (From now on we drop the $\lfloor \cdot \rfloor$.)

1.2. Overview of results and applications

In Erdős–Rényi random graph theory, a basic question concerns the appearance of subgraphs during the evolution of a random graph. In particular, considering a fixed graph H, we ask the question: for which values of p is H with high probability an (induced) subgraph of the random graph? The answer depends on the *maximum average*

[2] A *clique* is a set of pairwise adjacent vertices. Cliques, for us, are not necessarily maximal cliques.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

degree of H (the maximum of the average degree of all of H's subgraphs). For an overview, see [1, 5, 9].

One way to state the classical result is that a random graph $G(n, p)$ contains H with high probability exactly when the expected number of copies of H and all its subgraphs all go to infinity.

This paper studies the analogous problem for random intersection graphs. We show that for a fixed graph H there are two thresholds, τ_1 and τ_2. We show that $H \leqslant G(n, m, p)$ (we write $H \leqslant G$ to mean H is an induced subgraph of G) exactly when p is asymptotically between these thresholds. One of the curious special features of our model is that – in contradistinction to the Erdős–Rényi model – it is possible that the expected number of copies of H and all its subgraphs is very large, but the probability that $H \leqslant G(n, m, p)$ is very small.

We apply our results on subgraphs to answer the question: when is $G(n, m, p)$ with high probability an interval graph? In Section 3 we explain the importance of this question.

We believe that this 'vertex-biased' approach to random graphs can have important applications. Often the relationship between two 'objects' is not independent of the pair of objects. Objects that are 'closer' might be more likely to be related. For example, physical proximity is important in the spread of disease; the probability that a disease spreads from person A to person B is *not* independent of the two people chosen.

An application scenario well suited to our model of random graphs involves processors in a distributed setting. These processors 'compete' for shared resources (such as disks, printers, pages of memory, *etc*.) If each processor is oblivious to the actions of the others, then a reasonable protocol for the processor to follow is to try to secure its resources by making random selections. The graph $G(n, m, p)$ nicely models this situation. The n vertices are the processors, the m elements of the universal set are the resources, and processor i selects resource j with probability p. The edges of the resulting graph $G(n, m, p)$ represent the resulting pairwise conflicts that need to be resolved.

Another application of our model is to the *gate matrix layout* problem, which is discussed more fully in Section 3.

1.3. Probabilistic lemmas

Our proofs rely on the following probability results.

Lemma 1. *Let t be a fixed positive integer and let E denote an experiment with $t + 1$ mutually exclusive possible outcomes $\{0, \dots, t\}$. Let p_j denote the probability that we observe outcome j. We perform n independent (with $n \to \infty$) trials of this experiment and let N_j denote the number of times we observe outcome j. Furthermore, suppose that for $1 \leqslant j \leqslant t$ we have $p_j \to 0$ as $n \to \infty$ (hence $p_0 \to 1$). Finally, let $a_1, a_2, \dots, a_t$ be fixed, nonnegative integers. Then*

$$\frac{\Pr\{N_1 = a_1 \wedge N_2 = a_2 \wedge \cdots \wedge N_t = a_t\}}{\Pr\{N_1 = a_1\} \Pr\{N_2 = a_2\} \cdots \Pr\{N_t = a_t\}} \to 1 \qquad \text{as } n \to \infty.$$

Note: only the p_i are assumed to vary with n; the quantities t and a_j (with $1 \leqslant j \leqslant t$) are fixed.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Proof. Let $a_0 = n - a_1 - a_2 - \cdots - a_t$. The ratio in the statement of the lemma equals

$$\begin{aligned} \frac{\binom{n}{a_0,a_1,\ldots,a_t} p_0^{a_0} p_1^{a_1} \cdots p_t^{a_t}}{\prod_{j=1}^t \left[\binom{n}{a_j} p_j^{a_j} (1-p_j)^{n-a_j}\right]} &\sim \frac{(1-p_1-p_2-\cdots-p_t)^{n-a_1-\cdots-a_t}}{\prod_{j=1}^t (1-p_j)^{n-a_j}} \\ &\sim \frac{\exp\{-(p_1+\cdots+p_t)(n-a_1-\cdots-a_t)\}}{\prod_{j=1}^t \exp\{-p_j(n-a_j)\}} \\ &\sim \frac{\exp\{-(p_1+\cdots+p_t)n\}}{\prod_{j=1}^t \exp\{-p_j n\}} \\ &= 1. \end{aligned}$$

□

The conclusion of the lemma can be restated: the events $N_j = a_j$ (for $1 \leqslant j \leqslant t$) are asymptotically independent. Furthermore, it follows from the lemma that

$$\frac{\Pr\{N_1 \geqslant a_1 \wedge N_2 \geqslant a_2 \wedge \cdots \wedge N_t \geqslant a_t\}}{\Pr\{N_1 \geqslant a_1\} \Pr\{N_2 \geqslant a_2\} \cdots \Pr\{N_t \geqslant a_t\}} \to 1 \qquad \text{as } n \to \infty$$

as well.

Next we consider asymptotic expressions for $\Pr\{N_j \geqslant a_j\}$ that are readily derived from basic properties of the binomial distribution.

Lemma 2. *Suppose we perform n (with $n \to \infty$) Bernoulli trials of an experiment, and the probability of success is p with $p \to 0$ as $n \to \infty$. Let N be the number of successes and let a be a fixed nonnegative integer.*

(1) *If $np \to 0$ then $\Pr\{N \geqslant a\} \sim \Pr\{N = a\} \sim (np)^a/a!$.*
(2) *If $a > 0$ and if there is a constant $\varepsilon > 0$ so that $\varepsilon \leqslant np \leqslant 1/\varepsilon$, then there is a constant $\delta > 0$ so that $\delta \leqslant \Pr\{N \geqslant a\} \leqslant 1 - \delta$.*
(3) *If $np \to \infty$ then $\Pr\{N \geqslant a\} \to 1$.*

Proof. In every case we have $\Pr\{N \geqslant a\} \geqslant \Pr\{N = a\}$. For (1) we compute:

$$\Pr\{N \geqslant a\} = \sum_{j=a}^{n} \Pr\{N = j\} = \sum_{j=a}^{n} \binom{n}{j} p^j (1-p)^{n-j}.$$

Note that in the latter sum the ratio of the successive terms is

$$\frac{\binom{n}{j} p^j (1-p)^{n-j}}{\binom{n}{j+1} p^{j+1} (1-p)^{n-j-1}} = \frac{(j+1)(1-p)}{(n-j)p} > \frac{1}{2np} \to \infty.$$

Bounding the sum by a geometric series we have

$$\Pr\{N = a\} \leqslant \Pr\{N \geqslant a\} \leqslant \Pr\{N = a\} \left(\frac{1}{1-2np}\right) \sim \Pr\{N = a\}.$$

For (2) we have $\Pr\{N \geqslant a\} \geqslant \Pr\{N = a\} = \binom{n}{a} p^a (1-p)^{n-a} \geqslant \binom{n}{a} (\varepsilon/n)^a [1 - 1/(\varepsilon n)]^{n-a} \sim \varepsilon^a e^{-1/\varepsilon}/a!$, which is a positive constant. On the other hand,

$$\Pr\{N \geqslant a\} = 1 - \sum_{j=0}^{a-1} \Pr\{N = j\} \leqslant 1 - \sum_{j=0}^{a-1} \binom{n}{j} \left(\frac{\varepsilon}{n}\right)^j \left[1 - \frac{1}{\varepsilon n}\right]^{n-j} \sim 1 - \sum_{j=0}^{a-1} \frac{\varepsilon^j}{e^{1/\varepsilon} j!},$$

which is strictly less than 1.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Finally, for (3) we note that for fixed j we have $\Pr\{N = j\} = \binom{n}{j} p^j (1-p)^{n-j} \leqslant (np)^j e^{-np} \to 0$ since $np \to \infty$. Hence,

$$\Pr\{N \geqslant a\} = 1 - \sum_{j=0}^{a-1} \Pr\{N = j\} = 1 - o(1). \qquad \square$$

2. Subgraph thresholds

2.1. Thresholds

In this paper we show that for all fixed graphs H there is the 'birth' threshold $\tau_1(H)$ such that, if $p \ll \tau_1(H)$, then with high probability $G(n, m, p)$ does not contain H as a subgraph, while for $p \gg \tau_1(H)$, H is with high probability a subgraph of our random graph. With induced subgraphs there is more to the story. If H is any fixed graph, then the 'birth' threshold for H being an induced subgraph of $G(n, m, p)$ coincides with the 'birth' threshold for H as a subgraph. However the property 'H is an induced subgraph of G' is not monotone; hence, when our random graph becomes dense enough, H will disappear from it. Therefore in this case there are two thresholds, $\tau_1(H)$ and $\tau_2(H)$, associated with H. If $p \ll \tau_1(H)$ or $p \gg \tau_2(H)$, then with high probability $H \not\leqslant G$. However, if $\tau_1(H) \ll p \ll \tau_2(H)$, then with high probability $H \leqslant G$.

Let us first introduce basic notions and a notation used in the paper. Let H be any fixed graph. A *clique cover* of a graph H is a collection of vertex sets such that each induces a complete subgraph (*clique*) of H and, for every edge vw of H, v and w are together in at least one common member of the collection. In other words, the cliques induced by the vertex sets exactly cover the edges of H. We say that $\mathscr{C}$ is *reducible* if, for some $C \in \mathscr{C}$, the edges induced by C are contained in the union of the edges induced by $\mathscr{C}$; otherwise $\mathscr{C}$ is *irreducible*.

If $\mathscr{C} = \{C_1, C_2, \ldots, C_k\}$ is a particular clique cover of H (with $|C_i| \geqslant 1$ for all $i = 1, 2, \ldots, k$) then $|\mathscr{C}|$ denotes the number of cliques in $\mathscr{C}$, $\sum \mathscr{C}$ the sum of clique sizes in $\mathscr{C}$, and $\mathscr{C}'$ stands for $\{C \in \mathscr{C} : |C| > 1\}$.

Furthermore, for $S \subset V(H)$, define *restricted clique covers* as follows:

$$\mathscr{C}[S] := \{C_i \cup S : |C_i \cup S| \geqslant 1, i = 1, 2, \ldots, k\},$$

that is, the clique cover of S that results from restricting the cliques of $\mathscr{C}$ to the vertices that are in S, and

$$\mathscr{C}'[S] := \{C_i \cup S : |C_i \cup S| \geqslant 2, i = 1, 2, \ldots, k\},$$

that is, the clique cover of S that results from restricting the cliques of $\mathscr{C}$ to the vertices that are in S, ignoring all resulting cliques of size 1.

Using these restricted clique covers, let us define:

$$\begin{aligned}
\tau(H, \mathscr{C}, S) &= 1 \big/ \left(n^{|S|/\Sigma\mathscr{C}[S]} m^{|\mathscr{C}[S]|/\Sigma\mathscr{C}[S]} \right), \\
\tau'(H, \mathscr{C}, S) &= 1 \big/ \left(n^{|S|/\Sigma\mathscr{C}'[S]} m^{|\mathscr{C}'[S]|/\Sigma\mathscr{C}'[S]} \right), \\
\tau(H, \mathscr{C}) &= \max_S \{\tau(H, \mathscr{C}, S), \tau'(H, \mathscr{C}, S)\}, \\
\tau_1(H) &= \min_{\mathscr{C}} \tau(H, \mathscr{C}),
\end{aligned}$$

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

where $\mathscr{C}$ is a clique cover of H, and S is a non-empty subset of $V(H)$. When $\mathscr{C}'[S]$ is empty, we put the corresponding τ' term equal to 0.

Furthermore, let $d(H) = |E(H)|/|V(H)|$, while $d^*(H) = \max_{L \leqslant H} d(L)$

We are now ready to state our main result, which considers three segments of the random graph evolution: the 'appearance' period, the period in which all small subgraphs have probability 1 of being present as induced subgraphs, and the 'disappearance' period.

Theorem 3. *Let H be a fixed graph.*

(a) *Suppose $mp^2 \to 0$. Then*

$$\lim_{n\to\infty} \Pr(H \leqslant G(n,m,p)) = \begin{cases} 0 & \text{if} \quad p/\tau_1(H) \to 0, \\ 1 & \text{if} \quad p/\tau_1(H) \to \infty. \end{cases}$$

(b) *Suppose $\epsilon \leqslant mp^2 \leqslant 1/\epsilon$. Then*

$$\lim_{n\to\infty} \Pr(H \leqslant G(n,m,p)) = 1.$$

(c) *Suppose $p = \sqrt{\frac{\log n + \omega_n}{d^*(\overline{H})m}}$ and $mp^2 \to \infty$. Then*

$$\lim_{n\to\infty} \Pr(H \leqslant G(n,m,p)) = \begin{cases} 1 & \text{if} \quad \omega_n \to +\infty, \\ 0 & \text{if} \quad \omega_n \to -\infty. \end{cases}$$

Proof. Let $X(H)$ denote the number of copies of H in $G = G(n,m,p)$. Now, if $E(X(H)) \to 0$ as $n \to \infty$, it follows from Markov's inequality that $\Pr\{H \leqslant G\} \to 0$. Furthermore, if L is an induced subgraph of H, and $E(X(L)) \to 0$ as $n \to \infty$, it also follows that $\Pr\{H \leqslant G\} \to 0$. This is the exact same situation as in Erdős–Rényi random graphs.

On the other hand, suppose $E(X(L)) \to \infty$ for all induced subgraphs $L \leqslant H$. In the Erdős–Rényi model this is sufficient to conclude that $H \leqslant G$ with high probability. However, in our model this is not sufficient. Thus the expected number of copies of H in G is not the full story. Nonetheless, it is the beginning of the story, so we concentrate on how to compute it.

Let $\pi(H)$ denote the probability that H is induced on vertices 1 through h *in that order*, that is, the identity map is an isomorphism of H onto the first h vertices of G. Thus, the expected number of copies of H in G is

$$E(X(H)) = \binom{n}{h} \frac{h!}{|\mathrm{aut}(H)|} \pi(H)$$

and it only remains to compute $\pi(H)$.

Let us refine our $X(H)$ notation. Given a clique cover $\mathscr{C}$ of H, let $X(H,\mathscr{C})$ denote the number of copies of H induced in G that are represented by clique cover $\mathscr{C}$. Likewise, let $\pi(H,\mathscr{C})$ denote the probability that H is induced on the first h vertices (in order) of G with clique cover $\mathscr{C}$.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

We have now reduced our problem to computing $\pi(H,\mathscr{C})$ We shall show first that, if $mp^2 \to 0$ and $\mathscr{C}$ is a clique cover of a graph H on h vertices, then

$$\pi(H,\mathscr{C}) \begin{cases} \sim m^{|\mathscr{C}|}p^{\Sigma\mathscr{C}}, & mp \to 0, \quad \text{or} \\ \asymp m^{|\mathscr{C}'|}p^{\Sigma\mathscr{C}'}, & mp \geqslant \varepsilon > 0. \end{cases}$$

Hence it follows that

$$EX(H,\mathscr{C}) \asymp \begin{cases} n^h m^{|\mathscr{C}|}p^{\Sigma\mathscr{C}}, & mp \to 0, \quad \text{or} \\ n^h m^{|\mathscr{C}'|}p^{\Sigma\mathscr{C}'}, & mp \geqslant \varepsilon > 0. \end{cases}$$

Let $\mathscr{C} = \{C_1, \dots, C_t\}$ be a clique cover of a fixed graph H on h vertices. Consider the h rows of $R(n,m,p)$ corresponding to H. The columns in these rows must correspond to cliques in $\mathscr{C}$ (or else contain at most one 1). Thus there are t kinds of columns that are *mandatory* and, say, s kinds of columns that are *forbidden*. The probability that a particular column corresponds to a mandatory clique C_i is $p^{|C_i|}(1-p)^{h-|C_i|} \sim p^{|C_i|}$. Let $N_1, \dots, N_t$ denote the number of columns corresponding to the cliques in $\mathscr{C}$ and let $N_{t+1}, \dots, N_{t+s}$ denote the number of columns of the forbidden types. Thus

$$\pi(H,\mathscr{C}) = \Pr\{N_1 > 0 \wedge \cdots N_t > 0 \wedge N_{t+1} = 0 \wedge \cdots \wedge N_{s+t} = 0\},$$

which, by Lemma 1, is asymptotic to

$$\Pr\{N_1 > 0\} \cdots \Pr\{N_t > 0\} \Pr\{N_{t+1} = 0\} \cdots \Pr\{N_{s+t} = 0\}.$$

Now we apply Lemma 2 to the first t terms. For $1 \leqslant i \leqslant t$ we have

$$\Pr\{N_i > 0\} \begin{cases} \sim mp^{|C_i|}, & |C_i| \geqslant 2, \text{ or } |C_i| = 1 \text{ and } mp \to 0, \\ \geqslant \delta, & |C_i| = 1 \text{ and } mp \geqslant \varepsilon, \end{cases}$$

where δ and ε are positive constants.

Next, for $t+1 \leqslant i \leqslant t+s$, the kind of column we are forbidding has $a \geqslant 2$ ones. Thus,

$$\Pr\{N_i = 0\} = \left(1 - p^a q^{h-a}\right)^m \sim \exp\left\{-mp^a q^{h-a}\right\} \sim 1$$

since $mp^2 \to 0$ (and $a \geqslant 2$).

Combining these results we get either

$$\pi(H,\mathscr{C}) \sim m^{|\mathscr{C}|}p^{\Sigma\mathscr{C}} \qquad \text{provided } mp \to 0,$$

or else,

$$\pi(H,\mathscr{C}) \asymp m^{|\mathscr{C}'|}p^{\Sigma\mathscr{C}'} \qquad \text{provided } mp \geqslant \varepsilon > 0.$$

We now claim that reducible clique covers are less likely to occur than irreducible ones. Suppose $mp^2 \to 0$, and $\mathscr{C}$ is a reducible clique cover of H. Thus there is a $C \in \mathscr{C}$ so that $\mathscr{C}^* = \mathscr{C} - \{C\}$ is also a clique cover of H. Now if C is a 1-clique we clearly have $\pi(H,\mathscr{C}) \leqslant \pi(H,\mathscr{C}^*)$ since C is *permitted* in the $\mathscr{C}^*$ representation, but *required* in the $\mathscr{C}$ representation. Otherwise, $|C| \geqslant 2$ and we have $\pi(H,\mathscr{C}) \leqslant mp^2\pi(H,\mathscr{C}^*) \ll \pi(H,\mathscr{C}^*)$ (since $mp^2 \to 0$).

To prove part (a) of our theorem, assume as before that $\mathscr{C}$ is an irreducible clique cover of H and let S be a non-empty subset of $V(H)$. Then the *restriction* of $\mathscr{C}$ to S is the multiset $\mathscr{C}[S] = \{C \cap S \neq \emptyset : C \in \mathscr{C}\}$.

We then let $\pi(H, \mathscr{C}, S)$ be the probability that a fixed set of $|S|$ rows generates $\mathscr{C}[S]$, *i.e.*, that for each $C \in \mathscr{C}[S]$ there is a separate column in the rows corresponding to S with 1s exactly for the elements of C.

Let $X(H, \mathscr{C}, S)$ be the number of subsets of rows of $R(n, m, p)$ that generate $\mathscr{C}[S]$. Thus $EX(H, \mathscr{C}, S) \asymp n^{|S|}\pi(H, \mathscr{C}, S)$. Therefore

$$EX(H, \mathscr{C}, S) \asymp \begin{cases} x = n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]}, & mp \to 0, \quad \text{and} \\ x' = n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]}, & mp \not\to 0. \end{cases}$$

where $\mathscr{C}'[S] = \{C \cap S : C \in \mathscr{C}, |C \cap S| > 1\}$.

Our next step in deriving a formula for $\tau_1(H)$ – the appearance threshold for H – is to show that the following statements hold. First, if for some $S \subseteq V(H)$ we have

$$n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]} \to 0 \text{ or } n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]} \to 0$$

as $n \to \infty$, then we also have $\Pr\{X(H, \mathscr{C}) > 0\} \to 0$. Second, if for all $S \subseteq V(H)$ we have

$$n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]} \to \infty \text{ and } n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]} \to \infty$$

as $n \to \infty$, then we also have $\Pr\{X(H, \mathscr{C}) > 0\} \to 1$.

Since $X(H, \mathscr{C}, S) = 0 \Longrightarrow X(H, \mathscr{C}) = 0$, it is enough to show that $EX(H, \mathscr{C}, S) \to 0$ for some $S \subseteq V(H)$. We have four possible cases to consider, depending on whether an x or x' tends to 0, and depending on whether or not mp tends to 0.

Observe that x and x' differ by a power of mp, namely $x = (mp)^{\ell}x'$ for some integer $\ell \geqslant 0$ (where ℓ is the number of 1-cliques in $\mathscr{C}[S]$).

Suppose first that $mp \to 0$. If $x \to 0$ (for some S) then, since $x \asymp EX(H, \mathscr{C}, S)$, we are done. Otherwise, if some $x' \to 0$, then, since $x = (mp)^{\ell}x'$ we also have $x \to 0$, and, again, we are done. On the other hand, suppose $mp \geqslant \varepsilon > 0$. If some $x' \to 0$ then, since $x' \asymp EX(H, \mathscr{C}, S)$, we are done. Otherwise, if some $x \to 0$, then since $x' = x/(mp)^{\ell} \leqslant x/\varepsilon^{\ell}$ we must have $x' \to 0$, and again we are done. Thus, if any expression of the form $n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]}$ or $n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]}$ tends to 0, $X(H, \mathscr{C}) = 0$, almost surely.

Now, suppose that for all S we have $n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]}$ and $n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]}$ tending to infinity. Let $\mu = EX(H, \mathscr{C}) = EX(H, \mathscr{C}, V(H))$, so $\mu \to \infty$. We show that $\Pr\{X(H, \mathscr{C}) > 0\} \to 1$ by the second moment method. Write

$$E[X(H, \mathscr{C})^2] = \sum_A \sum_B E[Z_A Z_B]$$

where the sums are over all h element subsets of $[n]$ and Z_A is 1 when the rows of $R(n, m, p)$ corresponding to A generate a copy of H with clique cover $\mathscr{C}$ and 0 otherwise. When $A \cap B = \emptyset$, note that Z_A and Z_B are independent; there are $\binom{n}{h}\binom{n-h}{h}$ such pairs (A, B). Thus,

$$E[X(H, \mathscr{C})^2] \sim \mu^2 + \sum_{A \cap B \neq \emptyset} E[Z_A Z_B].$$

We wish to show that $E[X(H, \mathscr{C})^2] \sim \mu^2$, so we need to show that the above summation is $o(\mu^2)$. There are $O(n^{2h-s})$ pairs (A, B) for which $|A \cap B| = s$, so it is enough to prove that when $|A \cap B| = s$ we have

$$(n^{2h-s}E[Z_A Z_B])/\mu^2 \to 0. \qquad (*)$$

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Let $\mathscr{C}_A$ be the clique cover $\mathscr{C}$ in terms of the specific assignment of cliques to labelled vertex sets as indicated by Z_A (since Z_A is associated with a particular order for laying down the clique cover on a set of h vertices). Define $\mathscr{C}_B$ similarly.

Let $\mathscr{C}^*$ be the cover of $G[A \cup B]$ generated by taking the union of the collections $\mathscr{C}_A$ and $\mathscr{C}_B$. By this definition, some edge sets of $A \cap B$ may get covered more than once, for example, if they are covered by different cliques in $\mathscr{C}_A$ and $\mathscr{C}_B$. The other sets will get covered exactly once.

For the purpose of comparing the numerator and denominator of the ratio

$$(n^{2h-s}\mathbf{E}[Z_A Z_B])/\mu^2,$$

think of writing the asymptotic expression for μ^2 as $\mu_A\mu_B$ (with $\mu_A = \mu_B$), where the clique terms in the two products μ_A and μ_B are ordered in the same way as the cliques in $\mathscr{C}_A$ and $\mathscr{C}_B$.

By definition of a union, there is a clique in $\mathscr{C}^*$ for every clique of $\mathscr{C}_A$, yielding terms of the form $mp^{|C|}$ in both the numerator and denominator for each such clique $|C|$. Thus the final product of clique probabilities for A in the numerator cancels with the clique probability part of μ_A in the denominator.

And what about those terms stemming from cliques in $\mathscr{C}_B$? For a clique C in $\mathscr{C}^*$ that consists only of vertices from B (*i.e.*, $C \cap A = \emptyset$), the numerator and μ_B will both have terms of the form $mp^{|C|}$, which will cancel each other. The only remaining type of clique probability term in the numerator will be for those cliques of $\mathscr{C}^*$ that contain vertices of $A \cap B$ but are not cliques of $\mathscr{C}_A$ (if they were in $\mathscr{C}_A$ they would already have been cancelled out with the cliques of A). Since these are additional cliques of $\mathscr{C}_B$, they match terms in the denominator's μ_B, and cancel with them.

The left-over clique probability terms are then all in the μ_B part of the denominator, and correspond to cliques on vertices of $A \cap B$ that are in both covers $\mathscr{C}_A$ and $\mathscr{C}_B$. For this reason, there was only one copy of the term in the numerator originally, and it was cancelled out when $\mathscr{C}_A$ was examined. There was, however, a copy in each of μ_A and μ_B, and the one in μ_B remains.

Letting $C_1, C_2, \ldots, C_b$ refer to the cliques of $\mathscr{C}_B$ generating these remaining terms, the ratio for $(n^{2h-s}\mathbf{E}(Z_A \cap A_B])/\mu^2$ can be simplified to

$$\frac{n^{2h-s}}{n^{2h}}(1)\left(\prod_{i=1}^{b}\frac{1}{mp^{|C_i|}}\right) = \frac{1}{n^s\prod_{i=1}^{b} mp^{|C_i|}}.$$

Each C_i here is a clique of $\mathscr{C}_B$ on $S = (A \cap B) \subseteq B$. So this is a partial set of the cliques in $\mathscr{C}[S]$. By assumption, $n^s m^{|\mathscr{C}[S]|} p^{\Sigma\mathscr{C}[S]} \to \infty$. But this expression can be written as a product of the current term $n^s \prod_{i=1}^{b} mp^{|C_i|}$ with additional terms of the form mp^a, each for some $a \geqslant 2$, and all of which tend to 0. Thus $n^s \prod_{i=1}^{b} mp^{|C_i|} \gg n^s m^{|\mathscr{C}[S]|} p^{\Sigma\mathscr{C}[S]} \to \infty$, and so $1/(n^s \prod_{i=1}^{b} mp^{|C_i|})$ tends to 0 as $n \to \infty$.

We can now derive a formula for the appearance threshold for H. Check that we have selected $\tau_1(H)$ so that if $p \ll \tau_1(H)$ then for every $\mathscr{C}$ there is an $S \subseteq V(H)$ for which $n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]}$ or $n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]}$ tends to 0. Conversely, if $p \gg \tau_1(H)$ then there is a $\mathscr{C}$ so that for all $S \subseteq V(H)$ we have $n^{|S|}m^{|\mathscr{C}[S]|}p^{\Sigma\mathscr{C}[S]}$ and $n^{|S|}m^{|\mathscr{C}'[S]|}p^{\Sigma\mathscr{C}'[S]}$ tending to infinity.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Thus we have shown that if H is a fixed graph and $mp^2 \to 0$, then if $p/\tau_1(H) \to 0$ then with high probability H is not an induced subgraph of $G(n,m,p)$ while if $\tau_1(H)/p \to 0$ then with high probability H is an induced subgraph of $G(n,m,p)$.

To prove part (b) of our theorem, that is, to show that when mp^2 is bounded away from 0 then the probability that H is an induced subgraph of $G(n,m,p)$ tends to 1 as $n \to \infty$, let $\mathscr{C} = E(H)$, that is, consider the clique cover of H consisting of all pairs of adjacent vertices in H. Let N_i (for $1 \leqslant i \leqslant |E(H)|$) denote the number of columns corresponding to the ith edge of H. Ordering all other types of columns that have at least two 1s, let N_i for $i > |E(H)|$ denote the number of columns of type i. We want to compute

$$\Pr\left\{N_1 > 0 \wedge \cdots \wedge N_{|E(H)|} > 0 \wedge N_{|E(H)|+1} = 0 \wedge \cdots\right\},$$

which, by Lemma 1, is asymptotic to

$$\Pr\{N_1 > 0\}\Pr\{N_2 > 0\}\cdots\Pr\left\{N_{|E(H)|} > 0\right\}\Pr\left\{N_{|E(H)|+1} = 0\right\}\Pr\left\{N_{|E(H)|+2} = 0\right\}\cdots.$$

Now, we notice that by Lemma 2 we have, for $1 \leqslant i \leqslant |E(H)|$,

$$\Pr\{N_i > 0\} \in [\delta, 1-\delta]$$

for some positive constant δ. For $i > |E(H)|$ we have $\Pr\{N_i = 0\} \geqslant \delta$. Thus $\pi(H,\mathscr{C}) \geqslant \delta'$, where δ' is a positive constant (a fixed power of δ).

We can decompose the n vertices of $G(n,m,p)$ into $k = \lfloor n/h \rfloor$ pairwise disjoint sets of h vertices $S_1, S_2, \ldots, S_k$ (plus, perhaps, a few left-overs) and we note that the probability of having a copy of H induced on any of them is at least δ' and these events are mutually independent. Since $k \to \infty$, we have $H \leqslant G(n,m,p)$ with probability tending to 1.

Finally, to prove the last statement (c) of our main theorem, consider the situation when $mp^2 \to \infty$. Then notice that the probability ρ that two vertices v and w are *not* adjacent:

$$\rho = \Pr\{vw \notin E(G)\} = (1-p^2)^m \sim e^{-mp^2},$$

which tends to 0. Thus it is the non-edges that are difficult to form.

Let us see that we may, without any loss of generality, assume that $mp^3 \to 0$. Let Z denote the number of non-edges in $G(n,m,p)$. Thus $EZ = \binom{n}{2}\rho \sim \frac{1}{2}n^2e^{-mp^2}$. Thus if $p = \sqrt{(2\log n + \omega_n)/m}$ (where $\omega_n \to \infty$) then $EZ \to 0$ and with high probability $G(n,m,p)$ is complete. In this case $mp^3 = (2\log n + \omega_n)^{3/2}/\sqrt{m}$, which tends to 0 (unless ω_n is very large).

As before, let $X(H)$ denote the number of copies of H in $G(n,m,p)$. Let $X_2(H) = X(H,E(H))$, that is, the number of copies of H appearing in $G(n,m,p)$ with clique cover consisting of exactly the edges of H. (The subscript '2' refers to the fact that all cliques in this cover have size 2.)

First, we shall prove that if $mp^2 \to \infty$, $mp^3 \to 0$ and H is a fixed graph, then

$$EX(H) \sim EX_2(H) \sim \binom{n}{h}\rho^{|E(\overline{H})|}$$

where $\rho \sim e^{-mp^2}$. To see this let us first consider the clique cover $\mathscr{C} = E(H)$ on a fixed set of h vertices. Let N_i (with $1 \leqslant i \leqslant |E(H)|$) denote the number of columns corresponding

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

to the ith clique (edge) of $\mathscr{C}$. For $|E(H)| < i \leqslant \binom{h}{2}$, let N_i denote the number of columns that generate an edge on the respective *non-edge* pair of vertices of H (columns with 2 ones that are located in particular rows not corresponding to end-points of an edge of H). For $i > \binom{h}{2}$, the N_i denote the number of cliques on 3 or more of the h vertices. We want $N_i > 0$ for $1 \leqslant i \leqslant |E(H)|$ and $N_i = 0$ for $i > |E(H)|$.

Applying Lemmas 1 and 2, we have

$$EX_2(H) \sim \binom{n}{h}\rho^{\binom{h}{2}-|E(H)|} = \binom{n}{h}\rho^{|E(\overline{H})|}.$$

Now if $\mathscr{C}$ is any *other* clique cover of H, then we still have $E(\overline{H}) \cap \mathscr{C} = \emptyset$, but we have some replacement clique(s) in $\mathscr{C}$. For each one that has 3 or more vertices, a term of the form mp^c (with $c \geqslant 3$) replaces mp^2 terms in our expression $EX(H, \mathscr{C})$ and, since $mp^3 \to 0$, we have $EX(H, \mathscr{C}) \ll EX_2(H)$. Thus we can restrict our attention to clique covers of H consisting purely of cliques of size 2, and if $EX_2(H) \to 0$, then with high probability H is not an induced subgraph of $G(n, m, p)$. Further, if $EX_2(L) \to 0$ for any induced subgraph of H, then, again, H is with high probability not an induced subgraph.

Now suppose that $p = \sqrt{(\log n + \omega_n)/(d^*(\overline{H})m)}$. Let L be such that its complement $\overline{L}$ satisfies $\overline{L} \leqslant \overline{H}$ and $d(\overline{L}) = d^*(\overline{H})$, so $d^*(\overline{H}) = E(\overline{L})/V(L)$. Put $\ell = |V(L)|$. Then

$$\begin{aligned} EX_2(L) \asymp n^{\ell}\rho^{|E(\overline{L})|} \sim n^{\ell}\exp\{-mp^2|E(\overline{L})|\} &= n^{\ell} n^{-|E(\overline{L})|/d^*(\overline{H})} e^{-\omega_n E(\overline{L})/d^*(\overline{H})} \\ &= o(1) n^{\ell} n^{-\ell} \to 0. \end{aligned}$$

Thus, with high probability $H \not\leqslant G(n, m, p)$.

On the other hand, suppose $p = \sqrt{(\log n - \omega_n)/(d^*(\overline{H})m)}$. In this case $EX_2(L) \to \infty$ for all $L \leqslant H$. Put $\mu = EX_2(H)$. To use the second moment method, we compute $E[X_2(H)^2]$ and work to show that it is asymptotic to μ^2. We can decompose $E[X_2(H)^2]$ as

$$E[X_2(H)^2] = \sum_{A,B} E[Z_A Z_B],$$

where the sum is over all $h = |V(H)|$ element subsets of $V(G)$, and Z_A is an indicator random variable that is 1 exactly when $G[A]$ is a copy of H. When $A \cap B = \emptyset$ then Z_A and Z_B are independent. There are $\binom{n}{h}\binom{n-h}{h} \sim \binom{n}{h}^2$ such terms, so $\sum_{A,B:|A\cap B|=\emptyset} E[Z_A Z_B] \sim \binom{n}{h}^2 E[Z_A]^2 = \mu^2$. Thus, it remains to show that

$$\sum_{A,B:|A\cap B|=\ell} E[Z_A Z_B] = o(\mu^2),$$

where ℓ is a fixed integer with $1 \leqslant \ell \leqslant h$. There are some $n^{2h-\ell}$ such pairs of sets A and B. The probability we have $Z_A Z_B = 1$ is just ρ^k, where k is the number of non-edges in $G[A \cup B]$. Note that $k = 2|E(\overline{H})| - |E(\overline{L})|$ where $L = G[A \cap B]$. Thus, comparing to $\mu^2 \asymp n^{2h}\rho^{2|E(\overline{H})|}$ we have

$$\frac{n^{2h-\ell}\rho^k}{n^{2h}\rho^{2|E(\overline{H})|}} = \frac{1}{n^{\ell}\rho^{|E(\overline{L})|}} \sim \frac{1}{EX_2(L)} \to 0,$$

as we claimed. □

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

The following result follows from our main theorem.

Corollary 4 (Large α). *For a fixed graph H, there is an $\alpha^* > 0$ such that*

$$\tau_1(H) = 1/(n^{1/2d^*(H)}m^{1/2}) \text{ for all } \alpha \geqslant \alpha^*,$$

where $d^*(H) = \max_{L \leqslant H} |E(L)|/|V(L)|$

Proof. This threshold arises from taking $\mathscr{C} = E(H)$ (*i.e.*, letting $\mathscr{C}$ be the clique cover consisting of exactly the edges of H), and letting S be a set of vertices that induces a subgraph of maximum density in H. What is the asymptotic probability of an edge between two vertices when p is at this threshold? We have $mp^2 = m(\tau_1(H))^2 = n^{-1/d^*(H)}$, which is exactly the probability threshold for the appearance of H as a subgraph in the Erdős–Rényi model $G(n, p)$.

The proof of this threshold begins by showing that, for any particular $\mathscr{C}$,

$$\max_S \{\tau(H, \mathscr{C}, S), \tau'(H, \mathscr{C}, S)\}$$

may be found by considering only τ' for each S (not τ), as long as α is large enough. When $\mathscr{C} = E(H)$, $\max_S\{\tau'(H, \mathscr{C}, S)\}$ occurs when $S = V(L)$ for $L \leqslant H : |E(H)|/|V(H)| = d^*(H)$, giving $\tau'(H, \mathscr{C}) = 1/(n^{1/2d^*(H)}m^{1/2})$. It can be shown (see [19]) that for any other clique cover there is some choice of S that gives a value of $\tau'(H, \mathscr{C}, S)$ that is greater than the one above when α is large, and so the maximum over $S \subseteq V$ for that clique cover will be at least that big. The justification that there is some such choice of S for each $\tilde{\mathscr{C}} \neq E(H)$ relies on the definition of τ' and the fact that, since $\tilde{\mathscr{C}} \neq E(H)$, there is at least one S for which $|\mathscr{C}\tilde{[S]}|/\Sigma\tilde{\mathscr{C}}[S] < 1/2$. Thus, $\min_{\mathscr{C}} \max_S\{\tau(H, \mathscr{C}, S), \tau'(H, \mathscr{C}, S)\} = 1/(n^{1/2d^*(H)}m^{1/2})$. □

Finally, let us notice that statement (a) of Theorem 3 can also be deduced from a bipartite version of the main result of paper [10]. Indeed, as we pointed out in the Introduction, the binomial bipartite model has a simple equivalence to the random matrix model, and can thus generate graphs from $G(n, m, p)$ according to the transformation from $\mathscr{R}(n, m, p)$ to $\mathscr{G}(n, m, p)$. This equivalence provides a useful relationship between subgraphs of $B(n, m, p)$ and clique covers for subgraphs of $G(n, m, p)$. As a result of this relationship several other results of [10], dealing for example with the distribution of the number of certain classes of strictly balanced subgraphs, can be applied to the study of subgraphs of $G(n, m, p)$ (see [19] for details).

2.2. Example

Let H be a fixed graph on h vertices, such as the graph in Figure 1, and let $X(H)$ denote the number of copies of H induced in $G = G(n, m, p)$.

Now there are a number of possible ways in which H can appear in G. Vertices 1, 2 and 3 form a K_3. In the representation of G (*i.e.*, the matrix R) there may be a column in which vertices 1, 2 and 3 have a common 1, or there may be three columns representing separate edges between the pairs. Thus the first four rows of R look like one of the

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

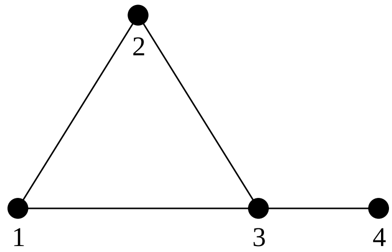

Figure 1 The graph H.

following:

$$\begin{bmatrix} \cdots & 1 & \cdots & 0 & \cdots \\ \cdots & 1 & \cdots & 0 & \cdots \\ \cdots & 1 & \cdots & 1 & \cdots \\ \cdots & 0 & \cdots & 1 & \cdots \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cdots & 1 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots \\ \cdots & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots \\ \cdots & 0 & \cdots & 1 & \cdots & 1 & \cdots & 1 & \cdots \\ \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots \end{bmatrix}.$$

These possibilities correspond exactly to two possible clique covers of the example graph H: $\mathscr{C}_1 = \{\{1,2,3\},\{3,4\}\}$ and $\mathscr{C}_2 = \{\{1,2\},\{2,3\},\{1,3\},\{3,4\}\}$.

Given a clique cover $\mathscr{C}$ of H, let $X(H,\mathscr{C})$ denote the number of copies of H induced in G that are represented by clique cover $\mathscr{C}$. Likewise, let $\pi(H,\mathscr{C})$ denote the probability that H is induced on the first h vertices (in order) of G with clique cover $\mathscr{C}$.

Let us first consider the clique cover $\mathscr{C}_1 = \{\{1,2,3\},\{3,4\}\}$. How could the first four rows of R fail to create this clique cover? We might

(i) be missing a column of the form $[1,1,1,0]^T$,
(ii) be missing a column of the form $[0,0,1,1]^T$, or
(iii) have a 'bad' column such as $[1,0,0,1]^T$ (which would create an unwanted edge between vertices 1 and 4).

We have

$$\pi(H,\mathscr{C}) = \sum_{i,j>0}^{i+j\leqslant m} \binom{m}{i,j,m-i-j} \left[p^3q\right]^i \left[p^2q^2\right]^j \left[q^4+4pq^3\right]^{m-i-j}, \qquad (**)$$

where $q = 1-p$. Note that i records the number of columns of type $[1,1,1,0]^T$ and j records the number of columns of type $[0,0,1,1]^T$. The $\left[q^4+4pq^3\right]^{m-i-j}$ term is for the probability that the remaining $m-i-j$ columns are acceptable, that is, have at most one 1. The right-hand side of $(**)$ simplifies to

$$\left(p^3q+p^2q^2+4pq^3+q^4\right)^m - \left(p^3q+4pq^3+q^4\right)^m - \left(p^2q^2+4pq^3+q^4\right)^m + \left(4pq^3+q^4\right)^m$$

which, provided $mp^2 \to 0$, is asymptotic to m^2p^5. By a similar method, one can show that $\pi(H,\mathscr{C}_2) \sim m^4p^8$.

It is important to note that 1-cliques play a special, exceptional role. Neither $\mathscr{C}_1$ nor $\mathscr{C}_2$ *require* there to be any 1-cliques, but we *allow* them to appear in the representation. Thus we define $\pi(H,\mathscr{C})$ (respectively, $X(H,\mathscr{C})$) to *allow* a 1-clique if it is *not* listed in $\mathscr{C}$, but to

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

require a 1-clique if it *is* listed in $\mathscr{C}$. For example, if $mp \to 0$, we find that

$$\begin{aligned} \pi(H, \mathscr{C}_1) &= m^2 p^5, \\ \pi(H, \mathscr{C}_1 \cup \{\{1\}\}) &= m^3 p^6. \end{aligned}$$

Note that $\mathscr{C}_1$ and $\mathscr{C}_2$ are *not* the only clique covers of H: there are a number of other possibilities. We might add to either $\mathscr{C}_1$ or $\mathscr{C}_2$ some singleton cliques (thereby changing their existence from *permitted* to *required*). Or we might add $\{1,2\}$ to $\mathscr{C}_1$ (changing the status of the clique $\{1,2\}$ from absent to present). Indeed, $\mathscr{C}_1 \cup \mathscr{C}_2$ is a clique cover as well. However, $\mathscr{C}_1$ and $\mathscr{C}_2$ are the only *irreducible* clique covers of H.

We have

$$EX(H, \mathscr{C}_1) \asymp n^4 m^2 p^5 \qquad \text{and} \qquad EX(H, \mathscr{C}_2) \asymp n^4 m^4 p^8.$$

Thus, if

$$p \ll 1/\left(n^{4/5} m^{2/5}\right) \qquad \text{and} \qquad p \ll 1/\left(n^{1/2} m^{1/2}\right),$$

then $E[X(H, \mathscr{C}_1)]$ and $E[X(H, \mathscr{C}_2)]$ tend to 0 as $n \to \infty$. Thus, if we let

$$p = \frac{1}{\omega_n} \min\left\{n^{-4/5} m^{-2/5}, n^{-1/2} m^{-1/2}\right\} = \begin{cases} 1/(\omega_n n^{4/5} m^{2/5}) & \alpha \leqslant 3, \\ 1/(\omega_n n^{1/2} m^{1/2}) & \alpha \geqslant 3, \end{cases}$$

then with high probability $H \nleqslant G(n, m, p)$.

Now, if any of H's induced subgraphs L has $E[X(L)] \to 0$, then we may also conclude that $\Pr\{H \leqslant G\} \to 0$. The probabilities derived from this, however, do not determine the threshold for H appearing as an induced subgraph of G.

For example, taking $\alpha = 0.2$ and $p = 1/(n^{0.6} m) \gg 1/(n^{4/5} m^{2/5})$, we have $EX(H) \geqslant EX(H, \mathscr{C}_1) \asymp n^4 m^2 p^5 = n^{0.4} \to \infty$. Likewise, one can check that $EX(L) \to \infty$ for all $L \leqslant H$. However, we claim that $\Pr\{H \leqslant G\} \to 0$. Why? Consider vertex 3. In its row there must be at least two 1s: one 1 for the connection to vertex 4, and a second (and possibly third) 1 for the connection to vertices 1 and 2. Let Y denote the number of rows of G's representation matrix with at least two 1s. The expected number of 1s in a given row is mp, which is, in this case, $n^{-0.6}$. The probability that there are two (or more) ones in a given row is asymptotically $(mp)^2 = n^{-1.2}$. Thus $EY \asymp n^{-0.2}$, and therefore with high probability there are no such rows in G's representation. With high probability it is impossible for $H \leqslant G$.

Thus it is not enough that $E(X(L)) \to \infty$ for all $L \leqslant H$ to ensure $H \leqslant G$ with high probability.

What has gone 'wrong'? If we consider the clique cover $\mathscr{C}_1$ or $\mathscr{C}_2$ *restricted* to the one element set $S = \{2\}$ we get a *restricted clique cover* of H induced on S. We denote this by $\mathscr{C}_1[S] = \{\{2\}, \{2\}\}$ (notice that this is a multiset). It is necessary that there be columns in the representation matrix of G corresponding to these restricted clique covers.

Consider the graph of Figure 1. It has two (irreducible) clique covers:

$$\mathscr{C}_1 = \{\{1,2,3\}, \{3,4\}\} \qquad \text{and} \qquad \mathscr{C}_2 = \{\{1,2\}, \{1,3\}, \{2,3\}, \{3,4\}\}.$$

For each we determine $\mathscr{C}[S]$ and $\mathscr{C}'[S]$ for all non-empty subsets of S and then compute $\tau(H, \mathscr{C}, S)$ and $\tau'(H, \mathscr{C}, S)$. These calculations are collected in Table 1.

To compute $\tau(H, \mathscr{C}_1)$ (respectively, $\tau(H, \mathscr{C}_2)$) we need to find the largest entry in the

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Table 1 Computations for $\tau_1(H)$ for the graph in Figure 1.

S	$\mathscr{C}_1[S]$	$\tau(H,\mathscr{C}_1,S)$	$\mathscr{C}'_1[S]$	$\tau'(H,\mathscr{C}_1,S)$
$\{1,2,3,4\}$	$\{123,34\}$	$1/(n^{4/5}m^{2/5})$	$\{123,34\}$	$1/(n^{4/5}m^{2/5})$
$\{1,2,3\}$	$\{123,3\}$	$1/(n^{3/4}m^{1/2})$	$\{123\}$	$1/(nm^{1/3})$
$\{1,2,4\}$	$\{12,4\}$	$1/(nm^{2/3})$	$\{12\}$	$1/(n^{3/2}m^{1/2})$
$\{1,3,4\}$	$\{13,34\}$	$1/(n^{3/4}m^{1/2})$	$\{13,34\}$	$1/(n^{3/4}m^{1/2})$
$\{1,2\}$	$\{12\}$	$1/(nm^{1/2})$	$\{12\}$	$1/(nm^{1/2})$
$\{1,3\}$	$\{13,3\}$	$1/(n^{2/3}m^{2/3})$	$\{13\}$	$1/(nm^{1/2})$
$\{1,4\}$	$\{1,4\}$	$1/(nm)$	$\emptyset$	—
$\{3,4\}$	$\{3,34\}$	$1/(n^{2/3}m^{2/3})$	$\{34\}$	$1/(nm^{1/2})$
$\{1\}$	$\{1\}$	$1/(nm)$	$\emptyset$	—
$\{3\}$	$\{3,3\}$	$1/(n^{1/2}m)$	$\emptyset$	—
$\{4\}$	$\{4\}$	$1/(nm)$	$\emptyset$	—

S	$\mathscr{C}_2[S]$	$\tau(H,\mathscr{C}_2,S)$	$\mathscr{C}'_2[S]$	$\tau'(H,\mathscr{C}_2,S)$
$\{1,2,3,4\}$	$\{12,13,23,34\}$	$1/(n^{1/2}m^{1/2})$	$\{12,13,23,34\}$	$1/(n^{1/2}m^{1/2})$
$\{1,2,3\}$	$\{12,13,23,3\}$	$1/(n^{3/7}m^{4/7})$	$\{12,13,23\}$	$1/(n^{1/2}m^{1/2})$
$\{1,2,4\}$	$\{12,1,2,4\}$	$1/(n^{3/5}m^{4/5})$	$\{12\}$	$1/(n^{3/2}m^{1/2})$
$\{1,3,4\}$	$\{1,13,3,34\}$	$1/(n^{1/2}m^{2/3})$	$\{13,34\}$	$1/(n^{3/4}m^{1/2})$
$\{1,2\}$	$\{12,1,2\}$	$1/(n^{1/2}m^{3/4})$	$\{12\}$	$1/(nm^{1/2})$
$\{1,3\}$	$\{1,13,3,3\}$	$1/(n^{2/5}m^{4/5})$	$\{13\}$	$1/(nm^{1/2})$
$\{1,4\}$	$\{1,1,4\}$	$1/(n^{2/3}m)$	$\emptyset$	—
$\{3,4\}$	$\{3,3,34\}$	$1/(n^{1/2}m^{3/4})$	$\{34\}$	$1/(nm^{1/2})$
$\{1\}$	$\{1,1\}$	$1/(n^{1/2}m)$	$\emptyset$	—
$\{3\}$	$\{3,3,3\}$	$1/(n^{1/3}m)$	$\emptyset$	—
$\{4\}$	$\{4\}$	$1/(nm)$	$\emptyset$	—

upper (respectively, lower) portion of Table 1. Which is largest depends on α, and we achieve the following:

$$\tau(H,\mathscr{C}_1) = \begin{cases} 1/\left(n^{1/2}m\right) & \alpha \leqslant 1/2, \\ 1/\left(n^{4/5}m^{2/5}\right) & 1/2 \leqslant \alpha \leqslant 3, \\ 1/\left(nm^{1/3}\right) & \alpha \geqslant 3, \end{cases}$$

$$\text{and} \quad \tau(H,\mathscr{C}_2) = \begin{cases} 1/\left(n^{1/3}m\right) & \alpha \leqslant 2/9, \\ 1/\left(n^{3/7}m^{4/7}\right) & 2/9 \leqslant \alpha \leqslant 1, \\ 1/\left(n^{1/2}m^{1/2}\right) & \alpha \geqslant 1. \end{cases}$$

Check that for $\alpha \leqslant 3$ we have $\tau(H,\mathscr{C}_1) \leqslant \tau(H,\mathscr{C}_2)$, but for $\alpha \geqslant 3$ we have $\tau(H,\mathscr{C}_1) \geqslant \tau(H,\mathscr{C}_2)$. Thus,

$$\tau_1(H) = \begin{cases} 1/\left(n^{1/2}m\right) & \alpha \leqslant 1/2, \\ 1/\left(n^{4/5}m^{2/5}\right) & 1/2 \leqslant \alpha \leqslant 3, \\ 1/\left(n^{1/2}m^{1/2}\right) & \alpha \geqslant 3. \end{cases}$$

From our main theorem it easily follows that the 'death' threshold for H is

$$\tau_2(H) = \sqrt{\frac{3\log n + \omega_n}{2m}}.$$

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

2.3. The evolutions: thresholds for specific subgraphs

Let us apply our results to some specific subgraphs such as cycles, trees, complete graphs and triangle-free graphs of a given order. In this section we shall present 'birth' thresholds and some 'death' thresholds for induced subgraphs of specified type. The computations can be laborious in some cases. The details of the proof for the complete graph case are provided as an example. For details of the other proofs see [19].

Corollary 5 (Cycles). *If $h > 3$ then $\tau_1(C_h) = 1/\sqrt{nm}$, while $\tau_2(C_h) = \sqrt{\frac{2\ln n}{(h-3)m}}$.*

Corollary 6 (Complete graphs). *For a complete graph K_h, we have*

$$\tau_1(K_h) = \begin{cases} 1/(nm^{1/h}) & \text{for } \alpha \leqslant 2h/(h-1), \text{ and} \\ 1/(n^{1/(h-1)}m^{1/2}) & \text{for } \alpha \geqslant 2h/(h-1). \end{cases}$$

The above threshold expressions arise from the expectations under the clique covers $\{V(H)\}$ (consisting of a single K_h) and $E(H)$ (the edge cover), respectively. Their validity is proved by a series of three claims. The proofs sometimes use V to refer to $V(H)$.

Claim 1. *For the single-clique cover $\{V(H)\}$,*

$$\max_S \left\{ \tau(K_h, \{V(H)\}, S), \tau'(K_h, \{V(H)\}, S) \right\} = \tau(K_h, \{V(H)\}, V).$$

Proof of Claim 1. The right-hand side of this claim is given by $\tau(K_h, \{V(H)\}, V) = 1/(nm^{1/h})$.

For S of size $2 \leqslant s \leqslant (h-1)$, restricting the clique cover $\{V(H)\}$ to S gives a single s-clique, so

$$\begin{aligned} \tau(K_h, \{V(H)\}, S) &= 1/(nm^{1/s}) < 1/(nm^{1/h}), \text{ and} \\ \tau'(K_h, \{V(H)\}, S) &= 1/(nm^{1/s}) < 1/(nm^{1/h}). \end{aligned}$$

For S of size 1,

$$\begin{aligned} \tau(K_h, \{V(H)\}, S) &= 1/(nm) < 1/(nm^{1/h}), \text{ and} \\ \tau'(K_h, \{V(H)\}, S) &= 0 < 1(nm^{1/h}). \end{aligned}$$

□

Claim 2. *For the edge clique cover $E(H)$,*

$$\max_S \left\{ \tau(K_h, E(H), S), \tau'(K_h, E(H), S) \right\} = \tau(K_h, E(H), V).$$

Proof of Claim 2. The right-hand side of the claim is given by $\tau(K_h, E(H), V) = 1/(n^{1/(h-1)}m^{1/2})$.

Let $S \subset V$, with $s = |S| < h$, and consider the restricted clique covers $E(H)[S]$ and $E(H)'[S]$. The cover $E(H)[S]$ is the edge cover of K_s plus a 1-clique for each of the $(h-s)$ external edges for each vertex of K_s, so

$$\begin{aligned}
\tau(K_h, E(H), S) &= 1/\left(n^{s/(s(s-1)+s(h-s))} m^{(s(s-1)/2+s(h-s))/(s(s-1)+s(h-s))}\right) \\
&= 1/\left(n^{1/(h-1)} m^{(h-(s+1)/2)/(h-1)}\right) \\
&\leqslant 1/\left(n^{1/(h-1)} m^{(h-(h/2))/(h-1)}\right), \text{ since } s \leqslant (h-1), \\
&= 1/\left(n^{1/(h-1)} m^{h/(2(h-1))}\right) \\
&< 1/(n^{1/(h-1)} m^{1/2}).
\end{aligned}$$

The restricted cover $E(H)'[S]$ is just the edge cover of K_s, and hence

$$\tau'(K_h, E(H), S) = 1/(n^{1/(s-1)} m^{1/2}) < 1/(n^{1/(h-1)} m^{1/2}. \qquad \square$$

Claim 3. *For any clique cover $\mathscr{C}$ of K_h in which each clique has size at least two, and for any fixed α,*

$$\tau(K_h, \mathscr{C}, V) \geqslant \tau(K_h, \{V(H)\}, V) \text{ or } \tau(K_h, \mathscr{C}, V) \geqslant \tau(K_h, E(H), V).$$

Proof of Claim 3. Let $\mathscr{C}$ be any clique cover of K_h in which each clique has size at least two, and denote the sizes of the cliques in $\mathscr{C} = \{C_1, C_2, \ldots, C_k\}$ by $r_1, r_2, \ldots, r_k$, respectively. The mechanism of this proof is a comparison of the linear function

$$\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha,$$

for the exponent in $\tau(K_h, \mathscr{C}, V)$, to the functions for the exponents in $\tau(K_h, \{V(H)\}, V)$ and $\tau(K_h, E(H), V)$. These latter functions are given by

$$(*) \quad 1 + \tfrac{1}{h}\alpha \qquad (\text{for } \{V(H)\}), \quad \text{and}$$

$$(**) \quad \tfrac{1}{h-1} + \tfrac{1}{2}\alpha \qquad (\text{for } E(H)).$$

They are plotted in Figure 2, where $x(\alpha)$ represents the function $\left(h/\Sigma\mathscr{C}\right) + \left(|\mathscr{C}|/\Sigma\mathscr{C}\right)\alpha$, with $\mathscr{C}$ being whichever clique cover is of interest.

The comparison will be made in two parts: first for $\alpha \in (0, \frac{2h}{h-1})$, and then for $\alpha \geqslant \frac{2h}{h-1}$. The border point $\alpha = \frac{2h}{h-1}$ is chosen because it is the value of α at which the functions $(*)$ and $(**)$ are equal. Here both function values are $1 + \frac{2}{h-1}$. For all smaller values of α, the function $(*)$ is the greater of the two, and for all larger α, $(**)$ is greater than $(*)$. Thus it is necessary only to compare $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha$ with $(*)$ for $\alpha \in (0, \frac{2h}{h-1})$, and with $(**)$ for $\alpha \geqslant \frac{2h}{h-1}$.

At $\alpha = 0$, $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha = \frac{h}{\Sigma\mathscr{C}}$, which is less than 1 since each vertex is in at least one of the cliques of $\mathscr{C}$. (In order to cover all edges, all end-point pairs must be included somewhere in the cover.)

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

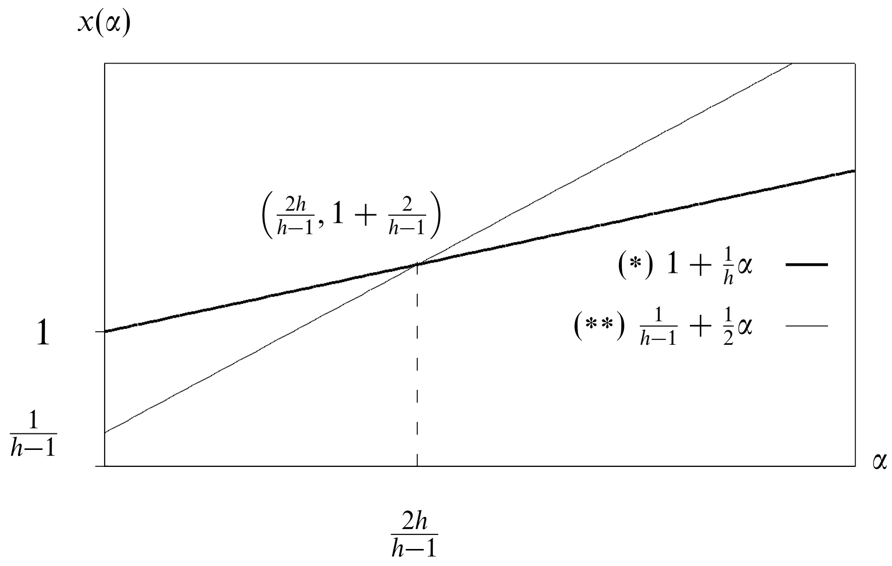

Figure 2 The exponent functions for $\tau(K_h, \{V(H)\}, V)$ and $\tau(K_h, E(H), V)$.

At $\alpha = \frac{2h}{h-1}$,

$$\begin{aligned}
\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha &= \frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\left(\frac{2h}{h-1}\right) \\
&\leqslant \frac{h}{\Sigma\mathscr{C}} + \left(\frac{h+1}{2h} - \frac{h-1}{2\Sigma\mathscr{C}}\right)\left(\frac{2h}{h-1}\right) \text{ (see justification below)} \\
&= 1 + \frac{2}{h-1}.
\end{aligned}$$

The substitution used to obtain the second inequality is justified as follows.

Let u be a vertex of $V(K_h)$ that appears in the fewest number of cliques of $\mathscr{C}$, and let t be the number of cliques in which u appears. Then

$$\Sigma\mathscr{C} = \sum_{\{i:C_i\supset u\}} r_i + \sum_{\{i:C_i\not\supset u\}} r_i \geqslant [(h-1)+t] + [2(|\mathscr{C}|-t)], \tag{2.1}$$

where $(h-1)$ counts all other vertices aside from u since they must each appear in some clique with u.

For any $v \in V(K_h)$,

$$\begin{aligned}
&\Sigma\mathscr{C} + |\{i : C_i \supset v\}| - (h-1) \\
&\quad\geqslant \Sigma\mathscr{C} + t - (h-1) \text{ (by definition of } t \text{ as a minimum)} \\
&\quad\geqslant [(h-1)+t] + [2(|\mathscr{C}|-t)] + t - (h-1) \text{ (by (2.1))} \\
&\quad= 2|\mathscr{C}|.
\end{aligned}$$

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Taking the sum of both sides of this over all $v \in V(H)$ gives

$$h\Sigma\mathscr{C} + \Sigma\mathscr{C} - h(h-1) \geqslant 2h|\mathscr{C}|.$$

Divided by $2h\Sigma\mathscr{C}$ and rearranged, this yields

$$\frac{|\mathscr{C}|}{\Sigma\mathscr{C}} \leqslant \frac{h+1}{2h} - \frac{h-1}{2\Sigma\mathscr{C}},$$

as desired.

Now, since $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha$ is less than or equal to (∗) at both $\alpha = 0$ and $\alpha = \frac{2h}{h-1}$, and the functions are linear, $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha$ is the minimum function throughout the interval $(0, \frac{2h}{h-1})$.

By the same analysis as above, $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha$ is less than or equal to (∗∗) when $\alpha = \frac{2h}{h-1}$. Its slope is $\frac{|\mathscr{C}|}{\Sigma\mathscr{C}}$. By assumption, this is less than or equal to 1/2, which is the slope of (∗∗). Thus, for all $\alpha \geqslant \frac{2h}{h-1}$, $\frac{h}{\Sigma\mathscr{C}} + \frac{|\mathscr{C}|}{\Sigma\mathscr{C}}\alpha$ must be at most the value of (∗∗). □

This is now enough information to apply the main subgraph theorems, since, for any $\mathscr{C}$ that is not $\{V(H)\}$ nor $E(H)$,

$$\begin{aligned}
&\max_S\{\tau(K_h,\mathscr{C},S),\tau'(K_h,\mathscr{C},S)\}\\
&\quad\geqslant \tau(K_h,\mathscr{C},V) \text{ (so assume } \mathscr{C} \text{ is irreducible)}\\
&\quad\geqslant \max\{\tau(K_h,\{V(H)\},V),\tau(K_h,E(H),V)\} \text{ (by Claim 3)}\\
&\quad= \max_S\{\tau(K_h,\{V(H)\},S),\tau'(K_h,\{V(H)\},S),\tau(K_h,E(H),S),\tau'(K_h,E(H),S)\}\\
&\qquad\quad \text{(by Claims 1 and 2).}
\end{aligned}$$

Hence the minimum $\max_S\{\tau(K_h,\mathscr{C},S),\tau'(K_h,\mathscr{C},S)\}$ is always achieved by $\{V(H)\}$ or $E(H)$, giving

$$\min_{\mathscr{C}}\max_S \tau(K_h,\mathscr{C},S) = \min\{\tau(K_h,\{V(H)\},V),\tau(K_h,E(H),V)\}.$$

This implies that

$$\tau_{K_h,1} = \begin{cases} 1/(nm^{1/h}) & \text{for } \alpha \leqslant 2h/(h-1),\\ 1/(n^{1/(h-1)}m^{1/2}) & \text{for } \alpha \geqslant 2h/(h-1),\end{cases}$$

as desired. □

Corollary 7 (Complete bipartite graphs). *For the complete bipartite graph $K_{h,h}$, the appearance threshold (for all α) is*

$$\tau_{K_{h,h},1} = 1/(n^{1/h}m^{1/2}).$$

For the complete bipartite graph $K_{h,k}$ with $h > k$, the threshold is

$$\tau_{K_{h,k},1} = \begin{cases} 1/(n^{1/h}m) & \text{for } \alpha \leqslant \frac{h-k}{hk},\\ 1/(n^{(h+k)/(2hk)}m^{1/2}) & \text{for } \alpha \geqslant \frac{h-k}{hk}.\end{cases}$$

Only the $h > k$ case will be proved here, since the result for $h = k$ follows similarly.

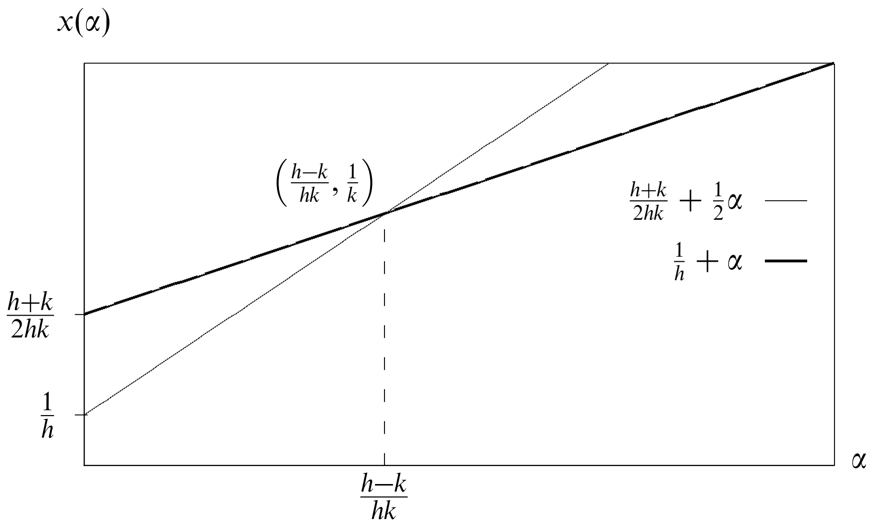

Figure 3 The exponent functions for τ_V and τ_{H_k}.

Since complete bipartite graphs are triangle-free, $E(H)$ is the only clique cover to be considered. The threshold function arises from considering this cover with vertex sets $V(H_k)$ (or any subset of $V(H_k)$), where H_k is the part of size k, and $V(H)$.

Let H_h refer to the vertex set of size h and H_k to the part of size k. It is straightforward to compute that

$$\begin{aligned}
\tau_V = \tau(K_{h,k}, E(H), V(H)) &= 1/\left(n^{\frac{h+k}{2hk}} m^{\frac{1}{2}}\right) \\
\tau_{H_h} := \tau(K_{h,k}, E(H), V(H_h)) &= 1/\left(n^{\frac{1}{k}} m\right) \\
\tau_{H_k} := \tau(K_{h,k}, E(H), V(H_k)) &= 1/\left(n^{\frac{1}{h}} m\right).
\end{aligned}$$

(Note that τ' need not be considered for these particular sets; for $V(H)$ it is the case that $\tau' = \tau$, and H_h and H_k have only 1-cliques so τ' is not applicable.)

The threshold function will be the maximum value of

$$\max\{\tau(K_{h,k}, E(H), S), \tau'(K_{h,k}, E(H), S)\}$$

over all vertex subsets of $V(H_{h,k})$. Since $h > k$, the above results imply that $\tau_{H_h} \leqslant \tau_{H_k}$ always, and τ_{H_h} can never be the maximum. Thus it can be ignored for the remainder of the proof.

Maximizing τ or τ' is the same as minimizing the exponent of n in the function denominator. For convenience, the exponent functions are referred to as $x(\alpha)$ and $x'(\alpha)$ for τ and τ' respectively. These exponents are also functions of the particular vertex set S being considered, and so may sometimes be written as, for instance, $x(\alpha, S)$ to draw attention to this. Figure 3 shows the $x(\alpha)$ functions for the cases $S = V(H)$ and $S = V(H_k)$ computed above.

As in the proof of the complete graph threshold, the linearity of the exponent functions

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

is used. It will be shown that, at every possible value of α, $x(\alpha)$ and $x'(\alpha)$ for any S is at least as great as for $S = V(H)$ or $S = V(H_k)$. Thus the minimum is always $\min\{x(\alpha, V(H)), x(\alpha, V(H_k))\}$.

Consider any $S \subset V(H_k)$. Let $s = |S|$. Then $x(\alpha) = s/(sh) + 1 = (1/h) + 1$, so this is the same as for $S = V(H_k)$ and need not be considered.

Similarly, if $S \subset V(H_h)$ then $x(\alpha, S) = x(\alpha, H_h)$, so this case may be ignored.

Hence the only interesting case to be considered is that of a set S that has $s_h \geqslant 0$ vertices from part H_h and $s_h \geqslant 0$ vertices from part H_k. Clearly, $s_h \leqslant h$ and $s_k \leqslant k < h$. These facts are used in the proof.

For this S, the exponent functions are

$$\begin{aligned} x(\alpha) &= \frac{s_h + s_k}{s_h k + s_k h} + \left(1 - \frac{s_k s_h}{s_h k + s_k h}\right)\alpha \\ x'(\alpha) &= \frac{s_h + s_k}{2 s_h s_k} + \frac{1}{2}\alpha. \end{aligned}$$

At $\alpha = 0$,

$$\begin{aligned} x &= \frac{s_h + s_k}{s_h k + s_k h} \geqslant \frac{s_h + s_k}{s_h h + s_k h} = \frac{1}{h}, \quad \text{and} \\ x' &= \frac{s_h + s_k}{2 s_h s_k} = \frac{1}{2 s_k} + \frac{1}{2 s_h} \geqslant \frac{1}{2h} + \frac{1}{2h} = \frac{1}{h}, \end{aligned}$$

so both functions are at least as great as the exponent for τ_{H_k}.

At $\alpha = \frac{h-k}{hk}$, which is the value of α at which the exponent functions of τ_V and τ_{H_k} intersect (see Figure 3),

$$\begin{aligned} x &= \frac{s_h + s_k}{2 s_h s_k} + \left(1 - \frac{s_k s_h}{s_h k + s_k h}\right)\left(\frac{h-k}{hk}\right) = \frac{1}{k} + \frac{s_h(k - s_k)(h-k)}{(s_h k + s_k h)(hk)} \geqslant \frac{1}{k}, \quad \text{and} \\ x' &= \frac{s_h + s_k}{s_h k + s_k h} + \frac{1}{2}\left(\frac{h-k}{hk}\right) = \frac{1}{2 s_k} + \frac{1}{2 s_h} + \frac{1}{2k} - \frac{1}{2h} \geqslant \frac{1}{2k} + \frac{1}{2h} + \frac{1}{2k} - \frac{1}{2h} = \frac{1}{k}, \end{aligned}$$

where $(1/k)$ is the value of $x(\alpha, V(H))$ and $x(\alpha, V(H_k))$ at $\alpha = (h-k)/(hk)$. Since all of the functions are linear, S must have $x(\alpha)$ and $x'(\alpha)$ at least as great as $x(\alpha, V(H_k))$ for the entire interval $(0, (h-k)/hk]$.

Moreover, $x(\alpha)$ and $x'(\alpha)$ are at least as great as the function $x(\alpha, V(H))$ at $\alpha = (h-k)/(hk)$. If it is possible to show that the slopes of $x(\alpha)$ and $x'(\alpha)$ are at least that of $x(\alpha, V(H))$, then this will imply that both functions have value at least $x(\alpha, V(H))$ for all $\alpha > (h-k)/(hk)$.

The slope condition turns out to be true, since $x'(\alpha)$ has slope $(1/2)$, which is equal to that of $x(\alpha, V(H))$, and $x(\alpha)$ has slope

$$\left(1 - \frac{s_k s_h}{s_h k + s_k h}\right) \geqslant \left(1 - \frac{s_k s_h}{s_h s_k + s_k s_h}\right) = \frac{1}{2}.$$

Thus

$$\tau_{K_{k,h},1} = \max_S\{\tau(K_{h,k}, E(H), S), \tau'(K_{h,k}, E(H), S)\} = \max\{\tau_V, \tau_{H_k}\}. \qquad \square$$

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Corollary 8 (Triangle-free graphs). *If H is a triangle-free graph with h vertices and maximum degree Δ, then*

$$\tau_1(H) = \begin{cases} 1/(n^{1/\Delta} m^{1-(\varepsilon(Q)/|Q|\Delta)}) & \text{for } \alpha \text{ sufficiently small,} \\ 1/(n^{1/(2d^*(H))} m^{1/2}) & \text{for } \alpha \text{ sufficiently large,} \end{cases}$$

where Q is a subset of $V(H)$ that induces the most edges per vertex with the restriction that all of its vertices have degree Δ in H, and $d^*(H) = \max_{L \leqslant H} |E(L)|/|V(L)|$. □

Note that finding the vertex set Q described above is equivalent to finding a vertex set achieving the maximum average degree for the graph induced on those vertices of H that are of degree Δ.

Corollary 9 (Trees). *If T is a tree with h vertices and maximum degree Δ, then*

$$\tau_1(T) = \begin{cases} 1/(n^{1/\Delta} m^{1-(r-1)/(r\Delta)}) & \text{for } \alpha \text{ sufficiently small,} \\ 1/(n^{h/(2h-2)} m^{1/2}) & \text{for } \alpha \text{ sufficiently large,} \end{cases}$$

where r is the size of a largest subtree induced on vertices all of which have degree Δ in T. □

The proof of the small-α result is an application of the results proved for triangle-free graphs. It requires showing that a set Q of Δ-degree vertices of T that induces the most edges per vertex may be taken to be a set of Δ-degree vertices inducing a largest subtree in T, *i.e.*, that

$$(\text{number of edges in } H[S])/|S| \leqslant (r-1)/r$$

holds for all S containing only vertices of degree Δ, where r is as defined above. The large-α threshold follows from Corollary 4, since $d^*(T) = \frac{h-1}{h}$.

Now consider the evolution of $G(n, m, p)$ as p is increased through a range of various functions for which $mp^2 \to 0$. In which order are the different subgraph thresholds reached? Are the cliques the first fixed graphs to arrive or the last? It turns out that the answer to this question depends on the value of α.

By examining the thresholds given above, one can see that when $\alpha = 3$ the cycle threshold is below that for all cliques on at least 4 vertices. However, when $\alpha = 1/2$ the clique threshold is below the threshold for cycles. This difference will be addressed more fully in a later paper.

3. Average case analysis of gate matrix circuit design

In this section we present an application of our theory to the design of integrated circuits via *gate matrix layouts* (GML).

A gate matrix layout is a stylized design for VLSI circuits [3, 4, 14]. The layout uses a grid-like representation, whose information can be encoded into a matrix of zeroes and ones that is much like the representation matrix introduced in Section 1.

The circuit components to be connected in a GML design are polysilicon lines serving as transistor gates and/or conductors ('gates'), and groups of transistor diffusions that

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

Table 2 Three ways to lay out the same GML circuit design. On the left is the original design. In the middle, we permuted columns 2 and 3. This allows us (right) to combine the first two rows.

1	2	3	4
a	—	a	
	b	—	b
c	—	c	c

1	2	3	4
a	a		
		b	b
c	c	—	c

1	2	3	4
a	a	b	b
c	c	—	c

associate with each other ('nets'). The GML grid is formed by laying all of the gate lines as vertical lines, and then placing horizontal tracks of transistor nets across them. There is a vertical gate line for each discrete input, and a transistor with that input must be placed over that gate. Additionally, all transistors associated in the same net must be placed in a common horizontal track. There may be more than one net in the same track, but only if the nets do not overlap (*i.e.*, their required inputs must be different, and must permit ordering such that one net is incident on all of the required gate lines before the next one in that track begins).

Although a layout that assigns a separate track for each net would be viable, the resulting circuit would undoubtedly be sparser than necessary. Since a chip with reduced surface area is cheaper to produce, we check for non-overlapping nets that can share tracks. We may do this for a given order of gate lines but, in addition, gate lines may be exchanged, carrying their transistors with them, and a different ordering of gates may allow for more nets to be placed in some tracks. See Table 2. Hence, the GML optimization problem is to find a permutation of the order of gates that minimizes the number of tracks required to lay out the circuit.

This problem may be stated equivalently in terms of matrices. We construct a gate matrix M by creating a column for each gate, and a row for each net. A '1' in position (i, j) represents a transistor that is in net i and must be placed on gate j. Roughly, we wish to find a column permutation of M so that the nets of 1s are packed densely in their individual rows (consecutive 1s are desirable).

Independent of the order of columns in M, we may use the matrix to construct the intersection graph G defined by

$$\begin{aligned} V &= \{\text{set of nets}\}, \\ E &= \{(v_i, v_j) : v_i, v_j \text{ are incident on a common gate}\}. \end{aligned}$$

The GML problem is known to be NP-hard [11] in general, but in the case when G is an *interval* graph[3] the problem is equivalent to that of finding a minimum colouring of G. Moreover, in the interval graph case the minimum colouring problem is easily solvable [8]. Thus, one might ask how likely it is that a gate matrix layout will have an interval graph as its associated graph G. We shall return to that question in the next section.

[3] Recall that G is called an *interval graph* provided we can assign to each vertex v of G a real interval I_v so that $vw \in E(G)$ exactly when $I_v \cap I_w \neq \emptyset$.

Consider a related decision problem known in the literature as the k-*GML problem*. Given an $m \times n$ gate matrix M and a fixed k, is there a permutation of columns of M such that the layout is possible in at most k tracks?

It has been recognized by Fellows and Langston (see [7]) that the the gate matrix layout with parameter k is equivalent to the path-width problem with parameter $k-1$, that is, a graph represents a circuit with a k-track layout if and only if G has a path decomposition of width at most $k-1$.

Although the original GML problem is in general very hard, Fellows and Langston [6] proved that the k-GML problem is, surprisingly, solvable in polynomial time (with respect to n). Their argument is based on the results of general theory of graph minors due to Robertson and Seymour (see [16]).

Suppose that matrix M represents the graph G defined as above. Notice that, if G corresponds to a matrix that satisfies the k-GML condition, then a corresponding matrix for any *minor*[4] of G satisfies k-GML also. This means that the family of graphs satisfying k-GML is closed under the minor ordering. Thus, according to Robertson–Seymour theory, the set of *obstructions* to k-GML, that is, minimal graphs under minor ordering that do not satisfy k-GML, must be finite. Since for every fixed graph H the problem of checking if H is a minor of G is solvable in polynomial time, k-GML can be decided in polynomial time as well. In fact, it can be decided in $O(n^2)$ (see [7]). The obstructions are known explicitly only when $k = 1, 2$ and 3. It is trivial to notice that for 1-GML an edge K_2 is such an obstruction. For 2-GML there are two such graphs: a triangle K_3 and subdivided star $K_{1,3}$ given as graph A in Figure 4 in the next subsection. For 3-GML, Kinnersley and Langston [12] were able to list all 110 obstruction graphs. They pointed out also that for $k = 4$ we may expect at least 122 million such graphs.

It is easy to notice, however, that K_k, the complete graph on k vertices, is an obstruction to $(k-1)$-GML for $k \geqslant 2$. Hence, if the gate matrix graph G contains K_k as a subgraph, then we are able at least to give the following 'negative' type result.

For the average case analysis, Corollary 6 implies that in a random gate matrix layout we shall need, with probability tending to 1 *as* $n \to \infty$, *at least* k *tracks if the probability* $p \gg 1/(nm^{1/k})$ *when* $\alpha \leqslant 2k/(k-1)$, *or* $p \gg 1/(n^{1/(k-1)}m^{1/2})$ *for* $\alpha \geqslant 2k/(k-1)$.

What exactly is meant by an average case? In the GML problem statement, the appearance of 0s and 1s in the matrix M is not random, but fixed by the interconnection requirements of the circuit. However, a probabilistic model based on a probability p of having a 0 in any particular component of the matrix is considered in [3].

3.1. When is $G(n, m, p)$ likely to be an interval graph?

In the interest of finding the probability of easy cases for gate matrix layout analysis, we first apply our random intersection graph model and subgraph theorem to study the values of p for which $G(n, m, p)$ will with high probability be an interval graph. This is accomplished by studying the appearance and disappearance thresholds for the family of

[4] Recall that H is a *minor* of G if a graph isomorphic to H can be obtained from G by a series of two operations: taking a subgraph and contracting an arbitrary edge. Then we say that H is less than or equal to a graph G in the *minor order*.

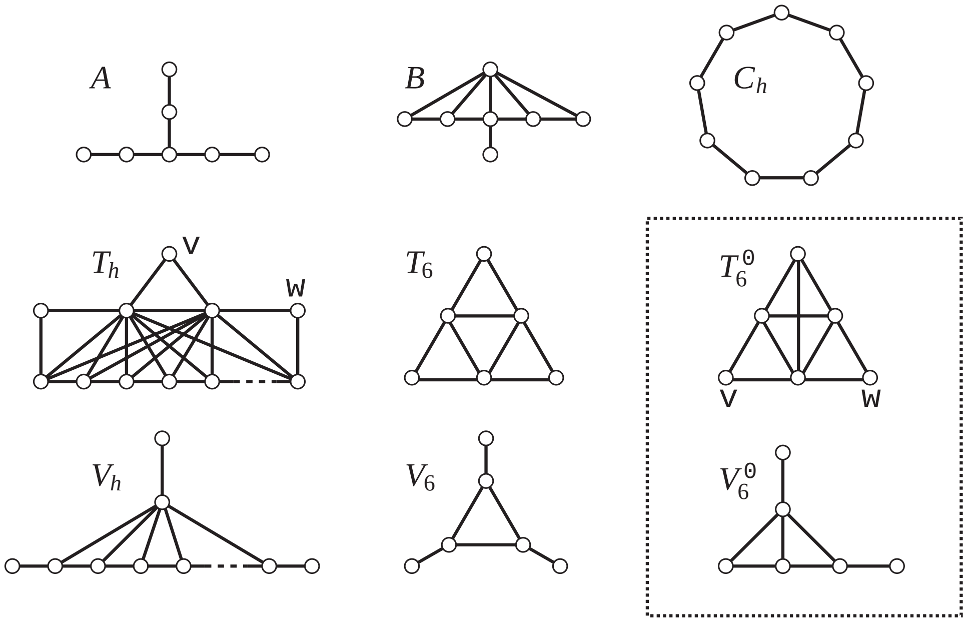

Figure 4 The graphs A, B, C_h ($h \geqslant 4$), T_h ($h \geqslant 6$), and V_h ($h \geqslant 6$) are the minimal non-interval graphs. The graph T'_6 (resp. V'_6) is an induced subgraph of every T_h ($h \geqslant 7$) (resp. V_h ($h \geqslant 7$)).

forbidden subgraphs that characterize the non-interval graphs. These forbidden subgraphs appear in Figure 4. The graphs A, B, C_h ($h \geqslant 4$), T_h ($h \geqslant 6$), and V_h ($h \geqslant 6$) are not interval graphs, but all of their proper induced subgraphs are interval graphs. Indeed, these are the only such graphs (see [13]; see also [8]). Since induced subgraphs of interval graphs are interval, it follows that G is an interval graph if and only if it induces none of the graphs shown in the main part of Figure 4 as subgraphs.

We are thus interested in the range of p that is below all appearance thresholds for forbidden subgraphs, and the range that is above all disappearance thresholds for these subgraphs. The forbidden subgraph family can be partitioned into two individual graphs, A and B, and three infinite families of subgraphs. For each fixed value of α, the appearance threshold for forbidden subgraphs is the minimum of the appearance thresholds for all of the individual forbidden subgraphs. We begin by showing that we need only consider the smallest member of each infinite family, since its appearance threshold is the minimum threshold for its family.

The individual graph thresholds given here result from applications of Theorem 3, and were obtained with the aid of a computer program (to find the maximum and minimum threshold function values depending on α).

As given in Corollary 5, the appearance threshold for a cycle on h vertices, where h is greater than or equal to 4, is

$$\tau_1(C_h) = 1/\left(n^{\frac{1}{2}} m^{\frac{1}{2}}\right).$$

So we need only consider C_4 for this family, with threshold as above.

https://doi.org/10.1017/S0963548398003459 Published online by Cambridge University Press

For the family of graphs T_h, $h \geqslant 6$, we compare all larger members to T_6. Consider the graph T_6' in Figure 4. It is an induced subgraph of each member of the T_h-family for which $h \geqslant 7$ (vertices v and w correspond as indicated). One checks that the appearance threshold for T_6' is greater than or equal to that of T_6, for all α (see [19]). Thus, it suffices (for the T-family) to consider only T_6, whose threshold is

$$\tau_1(T_6) = \begin{cases} 1/\left(n^{\frac{1}{2}}m^{\frac{1}{2}}\right) & \text{for } \alpha \leqslant 1, \\ 1/\left(n^{\frac{2}{3}}m^{\frac{1}{3}}\right) & \text{for } 1 \leqslant \alpha \leqslant 2, \\ 1/\left(n^{\frac{1}{3}}m^{\frac{1}{2}}\right) & \text{for } \alpha \geqslant 2. \end{cases}$$

Now consider the graph V_6' in Figure 4, noting that it is an induced subgraph of V_h for all $h \geqslant 7$. Again, one checks that the appearance threshold for V_6' is always at least that of V_6, so the appearance thresholds of V_h, $h \geqslant 7$ are greater than or equal to that of V_6. It thus suffices (for the V-family) to consider only V_6, whose threshold is

$$\tau_1(V_6) = \begin{cases} 1/\left(n^{\frac{1}{2}}m^{\frac{2}{3}}\right) & \text{for } \alpha \leqslant \frac{3}{4}, \\ 1/\left(n^{\frac{2}{3}}m^{\frac{4}{9}}\right) & \text{for } \frac{3}{4} \leqslant \alpha \leqslant 3, \\ 1/\left(n^{\frac{1}{2}}m^{\frac{1}{2}}\right) & \text{for } \alpha \geqslant 3. \end{cases}$$

For the two individual forbidden subgraphs, A and B, we have

$$\tau_1(A) = \begin{cases} 1/\left(n^{\frac{1}{3}}m\right) & \text{for } \alpha \leqslant \frac{1}{3}, \\ 1/\left(n^{\frac{4}{9}}m^{\frac{2}{3}}\right) & \text{for } \frac{1}{3} \leqslant \alpha \leqslant \frac{5}{6}, \\ 1/\left(n^{\frac{7}{12}}m^{\frac{1}{2}}\right) & \text{for } \alpha \geqslant \frac{5}{6}, \end{cases}$$

and

$$\tau_1(B) = \begin{cases} 1/\left(n^{\frac{1}{3}}m^{\frac{5}{6}}\right) & \text{for } \alpha \leqslant \frac{1}{5}, \\ 1/\left(n^{\frac{3}{8}}m^{\frac{5}{8}}\right) & \text{for } \frac{1}{5} \leqslant \alpha \leqslant \frac{1}{3}, \\ 1/\left(n^{\frac{5}{12}}m^{\frac{1}{2}}\right) & \text{for } \frac{1}{3} \leqslant \alpha \leqslant \frac{5}{6}, \\ 1/\left(n^{\frac{5}{11}}m^{\frac{5}{11}}\right) & \text{for } \frac{5}{6} \leqslant \alpha \leqslant \frac{6}{5}, \\ 1/\left(n^{\frac{1}{2}}m^{\frac{5}{12}}\right) & \text{for } \frac{6}{5} \leqslant \alpha \leqslant 2, \\ 1/\left(n^{\frac{1}{3}}m^{\frac{1}{2}}\right) & \text{for } \alpha \geqslant 2. \end{cases}$$

Taking the minimum of all of these thresholds, we find that the appearance threshold for the forbidden subgraphs is

$$\tau_1\text{ (forbidden subgraph set)} = \begin{cases} 1/\left(n^{\frac{1}{2}}m^{\frac{2}{3}}\right) & \text{for } \alpha \leqslant \frac{3}{4}, \\ 1/\left(n^{\frac{2}{3}}m^{\frac{4}{9}}\right) & \text{for } \frac{3}{4} \leqslant \alpha \leqslant \frac{3}{2}, \\ 1/\left(n^{\frac{7}{12}}m^{\frac{1}{2}}\right) & \text{for } \alpha \geqslant \frac{3}{2}. \end{cases}$$

This means that for $p = \tau/\omega$, with τ as above, $G(n,m,p)$ is with high probability an interval graph. If $p = \omega\tau$ (and $mp^2 \not\to \infty$), then G is with high probability not an interval graph.

When $mp^2 \to \infty$ we approach the end of the evolution and we consider the threshold when none of the forbidden subgraphs appear. This problem is much simpler. Consider the forbidden subgraph C_4. By Theorem 3 we have $\tau_2(C_4) = \sqrt{2\log n/m}$, which is the same as the threshold for $G(n,m,p)$ becoming a complete graph! Thus, in order to have probability tending to 1 that $G(n,m,p)$ an interval graph (for 'large' p), we need to choose p large enough that $G(n,m,p)$ is almost surely complete, that is, $p = \sqrt{\frac{2\log n+\omega}{m}}$.

References

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

[2] Das, S. Expected number of tracks for gate matrix, Dept. of Computer Science, University of North Texas.

[3] Das, S., Deo, N., and Prasad, S. (1989) *Gate Matrix Layout Revisited: Algorithmic Performance and Probabilistic Analysis*, Vol. 405 of *Lecture Notes in Computer Science*, Springer, pp. 280–290.

[4] Deo, N., Krishnamoorthy, M., and Langston, M. (1987) Exact and approximate solutions for the gate matrix layout problem, *IEEE Trans. Computer-Aided Design* **CAD-6** 79–84.

[5] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *MTA Mat. Kut. Int. Kozl.* **5** 17–61.

[6] Fellows, M. and Langston, M. (1987/88) Nonconstructive advances in polynomial-time complexity. *Information Processing Letters* **26** 157–162.

[7] Fellows, M. and Langston, M. (1994) On search, decision, and the efficiency of polynomial-time algorithms. *J. Computer and System Sciences* **49** 769–779.

[8] Golumbic, M. (1980) *Algorithmic Graph Theory and Perfect Graphs*, Academic.

[9] Karoński, M. (1995) Random graphs. In *Handbook of Combinatorics* (R. L. Graham, M. Grötschel and L. Lovász, eds), Vol. 1, Elsevier, pp. 351–380.

[10] Karoński, M. and Ruciński, A. (1989) Small subgraphs of k-partite random graphs. In *Combinatorial Mathematics: Proc. Third International Conference, New York, 1985*, 555, New York Academy of Sciences, pp. 230–240.

[11] Kashiwabara, T. and Fujisawa, T. (1979) An NP-complete problem on interval graphs. *Proc. IEEE Symposium on Circuits and Systems*, 82–83.

[12] Kinnersley, N. and Langston, M. (1994) Obstruction set isolation for the gate matrix layout problem. *Discrete Appl. Math.* **54** 169–213

[13] Lekkerkerker, C. B. and Boland, J. C. (1962) Representation of a finite graph by a set of intervals on the real line. *Fund. Math.* **51** 45–64.

[14] Lopez, A., and Law, H.-F. (1980) A dense gate matrix layout method for MOS VLSI. *IEEE Trans. Electron Devices* **ED-27** 1671–1675.

[15] Marczewski, E. (1945) Sur deux propriétés des classes d'ensembles. *Fund. Math.* **33** 303–307.

[16] Robertson, N. and Seymour, P. D. (1985) Graph minors – A Survey. In *Surveys in Combinatorics* (I. Anderson, ed.), Cambridge University Press, Cambridge, pp. 153–171.
[17] Scheinerman, E. A. (1988) Random interval graphs. *Combinatorica* **8** 357–371.
[18] Scheinerman, E. A. (1990) An evolution of interval graphs. *Discrete Math.* **82** 287–302.
[19] Singer, K. (1995) *Random Intersection Graphs*, PhD thesis, Johns Hopkins University.