ALMOST ALL CHORDAL GRAPHS SPLIT

E. A. BENDER, L. B. RICHMOND and N. C. WORMALD

Abstract

A *chordal graph* is a graph in which every cycle of length at least 4 has a chord. If G is a random n-vertex labelled chordal graph, the size of the larget clique in about $n/2$ and deletion of this clique almost surely leaves only isolated vertices. This gives the asymptotic number of chordal graphs and information about a variety of things such as the size of the largest clique and connectivity.

1980 *Mathematics subject classification (Amer. Math. Soc.)*: 05 C 99.

A *chordal graph* is a graph in which every cycle of length at least 4 has a chord. They have also been called *triangulated graphs*, *rigid circuit graphs* and *perfect elimination graphs*. See [1], [2], [4], [5] and [6] for further information. A *split graph* is a graph whose vertices can be partitioned into two sets, one of which induces a clique and the other an independent set. This can be viewed as a bipartite graph with added edges joining all the vertices in one part. It is easily seen that every split graph is chordal. We will show that almost all chordal graphs are split.

If G is a labelled chordal graph, let $c(G)$ be the size of the largest clique of G. Let $T(n)$ be the number of n-vertex labelled chordal graphs G and let $T(n, r)$ be the number with $c(G) = r$. All probability statements refer to chordal graphs chosen uniformly at random from the set of all labelled n-vertex chordal graphs. Let $|H|$ be the number of vertices in the graph H.

The following result has been proved with "each clique" replaced by "some clique" by Rose [4]. See also Dirac [2].

LEMMA 1. *For each chordal graph G and each clique R of G there is a sequence*

$$R = G_r, G_{r+1}, \ldots, G_n = G$$

of graphs such that G_{i+1} is obtained from G_i by adjoining a new vertex to one of its cliques. If $u, v \notin R$, $\{u, v\}$ is an edge, and u is adjoined before v, then $C(v) \cap R \subseteq C(u) \cap R$, where $C(w)$ is the clique to which w is adjoined.

PROOF. Suppose the lemma is false and let G be a counterexample with the least number of vertices, say n. Also assume that the connectivity k of G is as large as possible given n. Since the theorem is true for complete graphs, $k \neq n$ and G has a k-cutset K of vertices. By Dirac's Theorem 1 [2], K induces a clique in G. Let C be a component of $G - K$ such that

$$G|_{V(C) \cup K} = H$$

contains R. Since G is minimal and subgraphs of chordal graphs are chordal, the lemma holds for H starting with R. Similarly, the lemma holds for $G - C$ starting with the clique on K. Thus G can be built up by first building H and then building $G - C$, a contradiction.

The last claim in the lemma follows from the fact that $C(v)$ is a clique containing u and all the edges between u and R end in $C(u) \cap R$.

LEMMA 2. *We have*

$$\binom{n}{r}(2^r - r - 1)^{n-r} \leqslant T(n, r) \leqslant \frac{n!(n-r)!}{r!} 2^{r(n-r)},$$

the number of maximal cliques in $T(n, r)$ is at most $n - r + 1$, and $\log_2 T(n) \sim n^2/4$.

PROOF. The lower bound on $T(n, r)$ is proved joining each of $n - r$ vertices to at most $r - 2$ vertices of a complete graph on r vertices. (It is necessary to use $r - 2$ rather than $r - 1$ to avoid overcounting due to more than one r-clique.) We now prove the upper bound. By Lemma 1, all chordal graphs G with $c(G) = r$ are constructed by the following algorithm.

(1) Choose r vertices and let H be the clique on those vertices.

(2) Choose a vertex $v \notin H$ and a (possibly empty) clique $C(v) \subseteq H$ of less than r vertices.

(3) Replace H by H together with v and all possible edges between $C(v)$ and v. (called *adjoining v*.)

(4) If vertices remain, go to Step 2; otherwise set $G = H$.

Selecting the clique in Step 1 gives a factor of $\binom{n}{r}$. Ordering the remaining vertices for Step 2 gives a factor of $(n - r)!$. The clique in Step 2 can be specified by

choosing a subclique of a maximal clique of H. After $k - 1$ vertices have been adjoined, there are at most k maximal cliques and each has at most r vertices, so we obtain a factor of $k2^r$. Multiplying factors gives the upper bound. Since we have $n - r$ vertices to adjoin, there are at most $n - r + 1$ maximal cliques.

We prove the asymptotic claim. Let $r = [n/2]$ and assume that n is large. From the proved bounds,

$$2^{r(n-r)} < T(n, r) < T(n)$$

$$\leqslant n \max_x T(n, x) < n(n!)^2 \max_x 2^{x(n-x)} = n(n!)^2 2^{r(n-r)}.$$

Since $\log(n!) = o(n^2)$, the theorem follows.

LEMMA 3. *If $f(n)/\sqrt{n \log n} \to \infty$ as $n \to \infty$ and $A_1 < 1$, then for all sufficiently large n,*

$$Pr\{|c(G) - n/2| > f(n)\} \leqslant \exp(-A_1 f(n)^2).$$

PROOF. From the previous lemma, $T(n, r) \leqslant (n!)^2 2^{n^2/4} 2^{-(r-n/2)^2}$. Thus

$$Pr\{|c(G) - n/2| > f(n)\} = T(n)^{-1} \sum_{|r-n/2| > f(n)} T(n, r)$$

$$< T(n)^{-1} n(n!)^2 2^{n^2/4} 2^{-f(n)^2}$$

$$\leqslant \exp(O(n \log n) - f(n)^2).$$

LEMMA 4. *Let $s(G)$ be the maximum possible size of the largest clique in $G - R$, the graph obtained from G by deleting a maximum clique R. If $f(n)/\sqrt{n \log n} \to \infty$ as $n \to \infty$ and $A_2 < 1/4$, then for all sufficiently large n,*

$$Pr\{s(G) > f(n)\} \leqslant \exp(-A_2 f(n)^2).$$

PROOF. We will bound $T(n, r, s)$, the number of chordal graphs on n vertices with $c(G) = r$ and $s(G) = s$. Rather than use the algorithm, we create R and $G - R$ separately and then connect them. The number of choices for the vertex labels of R is $\binom{n}{r}$. Since G is chordal, so is $G - R$. Hence there are $T(n - r, s)$ choices for the graphs $G - R$ given R. How many ways can R and $G - R$ be connected? Let S be an s-vertex clique in $G - R$. If $v \in S$, let $R(v)$ be the vertices in R to which v is joined. It follows from the latter half of Lemma 1 that for some ordering $v_1, \ldots, v_s$ of S, $R(v_i) \subseteq R(v_{i+1})$. Hence S can be connected to R in at most $s!(s + 1)^r$ ways. The remaining $n - r - s$ vertices of $G - R$ can each be

connected to R in at most 2^r ways. Thus

$$T(n, r, s) \leqslant \binom{n}{r} T(n-r, s) s! (s+1)^r 2^{r(n-r-s)}$$

$$\leqslant \binom{n}{r} \frac{(n-r)!(n-r-s)!}{s!} 2^{s(n-r-s)} s! (s+1)^r 2^{r(n-r-s)}$$

$$\leqslant n^{3n} 2^{(r+s)(n-r-s)}.$$

We now sum over $s > f(n)$ and over all r with $|n/2 - r| < f(n)/2$. Thus $r + s > (n + f(n))/2$ in the sum. The sum can be shown to be small by the same method used in proving Lemma 3, and the neglected range gives a small contribution by Lemma 3. The lemma follows.

THEOREM. *If* $\alpha > \frac{1}{2}\sqrt{3}$, *n is sufficiently large and G is a random n-vertex chordal graph, then*

$$Pr\{ G \text{ is a split graph} \} > 1 - \alpha^n.$$

PROOF. Let $f(n) = n^{2/3}$. By Lemmas 3 and 4, the fraction of n-vertex chordal graphs which do not satisfy $|n/2 - r| < f(n)$ and $s < f(n)$, where $r = c(G)$ and $s = s(G)$, is $o(\alpha^n)$. Hence we neglect them. Note that $n = (1 + o(1))^r$ and $n^{f(n)} = (1 + o(1))^r$, facts we shall find useful at times in the proof.

We will give a refined version of the algorithm for generating chordal graphs that eliminates many of the duplicate ways to produce a graph. We will begin by picking out a maximal independent set of edges B in $G - R$, and denote the set of ends of edges in B by U. Note that $Z = (G - R) - B$ consists only of isolated vertices. We will adjoin the vertices of Z to H as soon as possible, respecting the lexicographic order (i.e., increasing numerical order of labels) of Z when we can. Since no pair of vertices in Z are joined by an edge, each is joined to the last U vertex preceding it, and the vertices of Z adjoined between the adjoining of any two vertices of U are in lexicographic order. Here "between" has the obvious meaning before the first vertex in U is adjoined or after the last. It will be useful to distinguish X, the set of vertices $u \in U$ with $C(u)$ large. Here is the modified algorithm.

(1) Choose r vertices and let $H = R$ be the clique on those vertices.

(2) Choose an ordered set U of $2b$ of the remaining vertices. Let Z be the remaining vertices and let X be a subset of U containing x vertices.

(3) Pair the vertices of U to specify a set B of b edges.

(4) Order the vertices for adjoining subject to the conditions that all vertices in U are in the same order as in Step 2 and that all vertices in Z between two vertices in U are in lexicographic order.

(5) For the next vertex v to be adjoined choose a (possibly empty) clique $C(v) \subseteq H$ of less than r vertices subject to the following constraints.

   (a) If $v \in U$, $\{u, v\} \in B$, and u has been adjoined, then $u \in C(v)$.

(b) If $v \in X$, then $|C(v)| \geqslant r - 2f(n)$.

(c) If $v \in U - X$, then $|C(v)| \leqslant r - 2f(n)$.

(d) If $v \in Z$, the last vertex in U that was adjoined to H is in $C(v)$ and $C(v) \cap Z = \varnothing$.

(6) Replace H by H together with v and all edges between $C(v)$ and v.

(7) If vertices remain, go to Step 5; otherwise set $G = H$.

There are $\binom{n}{r}$ choices at Step 1, less than $n^{2b}(2b)^x = (1 + o(1))^{br}$ at Step 2, and $(2b)!/b!2^b < n^{2b} = (1 + o(1))^{br}$ at Step 3. Hence there are at most

$$(1) \qquad\qquad \binom{n}{r}(1 + o(1))^{br}$$

choices for the first three steps.

We now turn to Steps 4 and 5. These are best handled together, for each vertex $v \notin R$. Step 4 can be viewed as associating each vertex $v \in Z$ with the last vertex of U adjoined before it (or with $\varnothing$ if v is adjoined before any vertex in U). As argued in Lemma 2, there are less than $n2^r$ choices for $C(v)$. This is too crude an estimate for many cases. We begin by improving the estimate for vertices in U. By the above view of Step 4, there are no choices at Step 4 for U. A vertex in X can be adjoined in at most

$$(2) \qquad\qquad n \sum_{j \leqslant 2f(n)} \binom{r}{j} < n^2 r^{2f(n)} = (1 + o(1))^r$$

ways. Suppose $\{u, v\} \in B$. We bound $N(u, v)$, the number of ways to choose $C(u)$ and $C(v)$.

Case 1. If $u, v \in X$, then $N(u, v) = (1 + o(1))^r$ by (2).

Case 2. If $u \in X$ and $v \in U - X$, then $N(u, v) \leqslant (1 + o(1))^r n2^r = (1 + o(1))^r 2^r$ by previous observations.

Case 3. Suppose that $u, v \in U - X$ and u is adjoined first. Choose a maximal clique M, of which there are less than n and none have more than r vertices. Choose a subclique $C(u)$ of M. By Lemma 1, $C(v) \cap R \subseteq C(u)$. Thus, given M, we can specify $C(u)$ and $C(v) \cap R$ in $3^{|M|} \leqslant 3^r$ ways. The remainder of $C(v)$ is a clique outside R and so can be specified in less than $n2^s = (1 + o(1))^r$ ways. Thus $N(u, v) \leqslant 3^r(1 + o(1))^r$.

Synthesis. Let α be the number of edges in B with both ends in X and β the number with one end in X. Then $x = 2\alpha + \beta$. The total contribution from adjoining the vertices in U is at most

$$(3) \qquad (1 + o(1))^{br} 2^{\beta r} 3^{r(b-\alpha-\beta)} = 3^{br}(2/3)^{\beta r}(1/\sqrt{3})^{2\alpha r}(1 + o(1))^{br}$$

$$\leqslant (3 + o(1))^{br}(2/3)^{rx}.$$

Consider $v \in Z$.

Case 1. If we associate v with $\varnothing$ for the ordering of Step 4, then there is one choice for when to adjoin v and at most 2^r choices for $C(v)$ since $C(v) \subseteq R$.

Case 2. How many ways can we have $C(v) \cap (U - X) \neq 0$? there are less than n choices at Step 4 for ordering v. Suppose that $u \in C(v) \cap (U - X) \neq \varnothing$. By Lemma 1, $C(v) \cap R \subseteq C(u)$. Also, $|C(u)| < r - 2f(n)$. Since there are less than n choices for u, there are less than $n2^{r-2f(n)}$ choices for $C(v) \cap R$. The remainder of $C(v)$ is a clique outside R and so can be specified in less than $n2^s$ ways. Thus there are at most $n^2 2^{r-f(n)}$ choices for Steps 4 and 5.

Case 3. How many ways can we have $\varnothing \neq C(v) \cap U \subseteq X$? By Step 5(d), Step 4 must associate v with some vertex $u \in X$. This can be done in x ways. Thus $C(v) \subseteq C(u) \cup u$ and $u \in C(v)$. Hence there are at most $x2^{|C(u)|} \leqslant x2^{r-1}$ ways to carry out Steps 4 and 5.

Synthesis. Combining the above, we see that there are less than

(4)
$$2^{r(n-r-2b)}\left(1 + x/2 + n^2 2^{-f(n)}\right)^{n-r-2b}$$
$$\leqslant 2^{r(n-r-2b)}(1 + x/2)^{n-r}(1 + o(1))$$

ways to order and adjoin the $n - r - 2b$ vertices that lie in Z, where $(1 + o(1))$ comes from $n \log(1 + n^2 2^{-f(n)}) = o(1)$.

By (1), (3) and (4), the number of graphs constructed by the algorithm is at most

(5)
$$\binom{n}{r}2^{r(n-r)}\left(\frac{3}{4} + o(1)\right)^{br}(2/3)^{rx}(1 + x/2)^{n-r}(1 + o(1)).$$

Summing over $x \leqslant 2b$ is simple: The maximum of $t(x) = (2/3)^{rx}(1 + x/2)^{n-r}$ occurs at

$$2 + x = \frac{n - r}{r \log(3/2)} \leqslant \frac{1 + o(1)}{\log(3/2)} = 2.47.$$

Since $t(0)$ and $t(1)$ are both $(1 + o(1))^r$, we see that (5) summed on x is bounded by

$$\binom{n}{r}2^{r(n-r)}\left(\frac{3}{4} + o(1)\right)^{br}.$$

Summing this on $b \geqslant 1$, we get

$$\binom{n}{r}2^{r(n-r)}\left(\frac{3}{4} + o(1)\right)^r.$$

Summing this on r with $|r - n/2| \le f(n)$ and using the comment at the beginning of this proof, we obtain

$$T(n)\,Pr\{b > 0\} \le \left(\frac{3}{4} + o(1)\right)^{n/2} \sum_r \binom{n}{r} 2^{r(n-r)} + o(\alpha^n T(n)).$$

By Lemma 2 the sum is $O(T(n))$.

Since the structure of typical chordal graphs is so simple, we can easily derive a variety of corollaries such as the following. Here n denotes the number of vertices in each graph.

COROLLARY 1. *The number of chordal graphs, the number of split graphs, twice the number of triangle-free graphs, and twice the number of bipartite graphs are all asymptotic to*

$$\sum_r \binom{n}{r} 2^{r(n-r)} \sim \begin{cases} \binom{n}{\frac{n}{2}} 2^{n^2/4} \sum_{k=-\infty}^{\infty} 2^{-k^2}, & \text{if } n \text{ is even;} \\[2mm] \binom{n}{\frac{n-1}{2}} 2^{n^2/4} \sum_{k=-\infty}^{\infty} 2^{-(k+1/2)^2}, & \text{if } n \text{ is odd.} \end{cases}$$

PROOF. Erdös, Kleitman and Rothschild obtained asymptotics for the number of triangle-free graphs [3]. The rest of the claims follow from the theorem and the fact that the number of split graphs is twice the number of bipartite.

COROLLARY 2. *If $f(n) \to \infty$, then $Pr\{|c(G) - n/2| > f(n)\} \to 0$ as $n \to \infty$.*

PROOF. This amounts to estimating the sums in Corollary 1 for $|k| \ge f(n)$.

COROLLARY 3. *Let $k(G)$ denote the connectivity of G. If $f(n)/\sqrt{n \log n} \to \infty$ as $n \to \infty$,*

$$Pr\{|k(G) - n/4| > f(n)\} \to 0 \quad as\ n \to \infty.$$

PROOF. It suffices to consider split graphs. The connectivity of a split graph is the minimum vertex degree. This is easily seen to be about $n/4$.

References

[1] P. Buneman, 'A characterisation of rigid circuit graphs', *Discrete Math.* **9** (1974), 205–212.

[2] G. A. Dirac, 'On rigid circuit graphs', *Abh. Math. Sem. Univ. Hamburg* **25** (1961), 71–76.

[3] P. Erdös, D. J. Kleitman and B. L. Rothschild, *Asymptotic enumeration of K_n-free graphs*, Colloquio Internazionale sulle Teorie Combinatorie (Roma, 1973), Tomo II. Atti dei Convegni Lincei, No. 17 (1976), 19–27.

[4] F. Gavril, 'The intersection graphs of subtrees in trees are exactly the chordal graphs', *J. Combin. Theory Ser. B* **16** (1974), 47–56.

[5] D. J. Rose, 'Triangulated graphs and the elimination process', *J. Math. Anal. Appl.* **32** (1970), 597–609.

[6] D. J. Rose, R. E. Tarjan and G. S. Lueker, 'Algorithmic aspects of vertex elimination on graphs', *SIAM J. Comput.* **5** (1976), 266–283.

Department of Mathematics
University of California at San Diego
La Jolla, California 92093
USA

Department of Combinatorics
and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Department of Mathematics,
  Statistics and Computer Science
University of Newcastle
New South Wales 2308
Australia