ISRAEL JOURNAL OF MATHEMATICS, Vol. 59, No. 1, 1987

GRAPHS WHOSE EVERY TRANSITIVE ORIENTATION CONTAINS ALMOST EVERY RELATION

BY

BÉLA BOLLOBÁS† AND GRAHAM BRIGHTWELL

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16, Mill Lane, Cambridge, CB2 1SB, England

ABSTRACT

Given a graph G on n vertices and a total ordering $\prec$ of $V(G)$, the transitive orientation of G associated with $\prec$, denoted $P(G; \prec)$, is the partial order on $V(G)$ defined by setting $x < y$ in $P(G; \prec)$ if there is a path $x = x_1x_2 \cdots x_r = y$ in G such that $x_i \prec x_j$ for $1 \leqq i < j \leqq r$. We investigate graphs G such that every transitive orientation of G contains $\binom{n}{2} - o(n^2)$ relations. We prove that almost every $G_{n,p}$ satisfies this requirement if

$$\frac{pn \log\log\log n}{\log n \log\log n} \to \infty,$$

but almost no $G_{n,p}$ satisfies the condition if $(pn \log\log\log n)/(\log n \log\log n)$ is bounded. We also show that every graph G with n vertices and at most $cn \log n$ edges has some transitive orientation with fewer than $\binom{n}{2} - \delta(c)n^2$ relations.

The following sorting problem was proposed by Rabin (see [7]). Given n objects in some total order unknown to us, we wish to ask a set of questions, all at once, such that no matter what answers we get we can deduce all but $o(n^2)$ of the $\binom{n}{2}$ relations. How many questions will suffice? Here a *question* or *probe* is a pair (a, b) of objects, and the answer reveals whether $a < b$ or $b < a$.

As the probes have to be made simultaneously, these probes form a graph on the set of objects. Thus our problem can be reformulated as follows. Given a graph $G = (V, E)$ of order $n = |G|$ and size $e(G)$, say, consider an acyclic orientation of the edges. Let $G = (V, E)$ be the directed graph obtained in this way and let $C(G)$ be the *acyclic closure* of G: xy is an arc of $C(G)$ if G contains a directed path from x to y. As every acyclic orientation of G is induced by a

† Partially supported by MCS Grant 8104854.
Received November 14, 1986

total order on V, we may assume that E is induced by a total order $\prec$ on V. The arcs of $C(G)$ define a partial order $P(G; \prec)$; we call $P(G; \prec)$ the transitive orientation of G associated with $\prec$.

We are interested in graphs G such that every transitive orientation of G contains many relations. Let $r(G; \prec)$ denote the number of relations in $P(G; \prec)$, and let $t(G) = \max\{\binom{n}{2} - r(G; \prec)\}$, the maximum number of relations of $\prec$ not in $P(G; \prec)$. Obviously if $t(G) = 0$, then $G = K_n$. In this paper, we are looking for graphs G_n of order n and small size $e(G_n)$ such that $t(G_n) = o(n^2)$. Problems of this type have been studied by Bollobás and Rosenfeld [8], Häggkvist and Hell [11, 12, 13], Ajtai, Komlós and Szemerédi [2, 3], Bollobás and Thomason [9], and others, see Bollobás and Hell [7] and Bollobás [5].

After we had submitted this paper, we became aware of papers by Ajtai, Komlós, Steiger and Szemerédi [1], and by Alon, Azar and Vishkin [4] dealing with the same questions as the present paper. In both these papers results are obtained (Theorem 2 of [1] and Proposition 3.5(i) of [4]) which are essentially the same as Theorem 6 and the remark after. Also, both sets of authors prove results (Theorem 1 of [1] and Proposition 3.5(ii) of [4]) which are improved by our Theorem 1, to the extent that our result happens to disprove the Conjecture in [1]. The present authors are currently working on a paper [6] dealing with the general question of how small $t(G)$ can be for graphs G with n vertices and $pn^2/2$ edges, throughout the range of $p = p(n)$.

Before embarking on the results, we recall the concept of a random graph. Let $G_{n,p} \equiv G_p$ denote a random graph on n vertices, with each pair of vertices joined by an edge with probability $p = p(n)$, each pair considered independently. We say that *almost every* (*a.e.*) G_p has property Q, or that Q holds *almost surely*, if the probability that $G_{n,p}$ has Q tends to 1 as $n \to \infty$. For the general theory of random graphs, the reader is referred to Bollobás [5].

In this paper we prove that, if

$$p = \frac{\omega(n) \log n \log\log n}{n \log\log\log n}, \qquad \text{where } \omega(n) \to \infty \text{ as } n \to \infty,$$

then $t(G_{n,p}) = o(n^2)$ for almost every $G_{n,p}$; whereas if

$$p \leqq \frac{c \log n \log\log n}{n \log\log\log n}, \qquad \text{where } c \text{ is a constant},$$

then, for some $\delta = \delta(c) > 0$, $t(G_{n,p}) \geqq \delta n^2$ for almost every $G_{n,p}$.

This latter result implies that, for every constant c, there is a constant $\delta=\delta(c)>0$ such that almost every graph G with n vertices and at most $(cn\log n\log\log n)/(\log\log\log n)$ edges has $t(G)\geqq\delta n^2$. We also show that, for every c, there is an $\varepsilon=\varepsilon(c)>0$ such that every graph G with n vertices and $\leqq cn\log n$ edges has $t(G)\geqq\varepsilon n^2$.

Our first result is as follows.

THEOREM 1. *Suppose*

$$p\geqq\frac{\alpha(n)\log n\log\log n}{n\log\log\log n},\qquad \textit{where } 64\leqq\alpha(n)\leqq\log\log n.$$

Then a.e. G_p *is such that* $t(G_p)\leqq 4n^2/\alpha^{1/2}$.

For $\alpha(n)\to\infty$, this tells us that, if

$$\frac{pn\log\log\log n}{\log n\log\log n}\to\infty,$$

then $t(G_p)$ is almost surely $o(n^2)$. For $\alpha>64$ a fixed constant, we interpret the theorem as saying that, if we want $P(G;<)$ to contain at least εn^2 relations $(\varepsilon>0)$ for every transitive orientation $<$ of G (i.e. $t(G)\leqq\binom{n}{2}-\varepsilon n^2$), then a.e. G_p will do, for

$$p\geqq\frac{\alpha\log n\log\log n}{n\log\log\log n}.$$

As mentioned earlier, weaker versions of Theorem 1 have been proved by Ajtai, Komlós, Steiger and Szemerédi [1], and by Alon, Azar and Vishkin [4]. Essentially, both sets of authors prove that, if

$$p\geqq\frac{\omega(n)\log n\log\log n}{n},$$

and $\omega(n)\to\infty$, then a.e. G_p is such that $t(G_p)=o(n^2)$. In fact, Ajtai, Komlós, Steiger and Szemerédi conjecture that their result is best possible, which Theorem 1 shows to be false. The significance of our additional factor of $\log\log\log n$ is that, at any rate for random graphs, it is best possible, as we shall see in Theorem 5.

The proof of Theorem 1 divides into two parts: first we prove that, for n sufficiently large, every graph G on n vertices satisfying the properties (Q_1) to (Q_4) below has $t(G)\leqq 4n^2/\alpha^{1/2}$; then we show that a.e. G_p satisfies (Q_1) to (Q_4). Let then (Q_1) to (Q_4) be the following properties of graphs of order n.

(Q_1) There is no pair (S, D) of disjoint subsets of $V(G)$ such that

$$|S| = s = \frac{n \log\log\log n}{\alpha^{1/2} \log n}, \qquad |D| = d \equiv \frac{2s}{\alpha^{1/2}} = \frac{2n \log\log\log n}{\alpha \log n},$$

and each vertex of D sends fewer than $\log\log n$ edges to S.

(Q_2) There is no pair (U, V) of disjoint subsets of $V(G)$ such that $|U| = u \leqq n/(\log n)^2$, $|V| = \frac{1}{4}u \log\log n$, and $e(U, V) > u \log\log n$.

(Q_3) There is no pair (U, V) of disjoint subsets of $V(G)$ such that

$$|U| = u \leqq \frac{n \log\log\log n}{\log n \log\log n},$$

$|V| = \frac{1}{2}u \log\log\log n$, and $e(U, V) > u \log\log n$.

(Q_4) There is no pair (Y, Z) of disjoint subsets of $V(G)$ such that

$$|Y| = y = \frac{n \log\log\log n}{\log n \log\log n},$$

$|Z| = z = n/\alpha$, and there are no $Y - Z$ edges.

Here and throughout, we omit integrality signs, which do not affect the argument. For subsets X and Y of $V(G)$, $e(X, Y)$ denotes the number of edges between them. All the properties (Q_i) express the idea that large sets of vertices have about the 'right' number of edges between them.

THEOREM 2. *If G is any graph on n vertices, with n sufficiently large, satisfying (Q_1) to (Q_4), then $t(G) \leqq 4n^2/\alpha^{1/2}$.*

PROOF. Let G be a graph on n vertices satisfying (Q_1) to (Q_4), and take any total ordering $<$ of $V(G)$. We are required to prove that, provided n is sufficiently large,

$$r(G; <) \geqq \binom{n}{2} - 4n^2/\alpha^{1/2}.$$

We may assume without loss of generality that $V[G] = [n] \equiv \{1, 2, \ldots, n\}$, and that $<$ is the standard order on $[n]$.

Let

$$l = \frac{\alpha^{1/2} \log n}{2 \log\log\log n}.$$

We assume for convenience that l is an integer dividing n; it is clear that this does not affect the argument. For $i = 1, 2, \ldots, l$, define

$$A_i = \left\{ m \in \mathbf{N} : \frac{n}{l}(i-1) < m \leqq \frac{ni}{l} \right\}.$$

The A_i are disjoint sets of vertices of size

$$\frac{n}{l} = \frac{2n \log\log\log n}{\alpha^{1/2} \log n}$$

satisfying $A_i < A_j$ for $i < j$. We now define inductively, for each k, a subset B_k of A_k as follows. Set $B_1 = A_1$. Given $B_k \subseteq A_k$ with

$$| B_k | > \frac{2n \log\log\log n}{\alpha^{1/2} \log n}\left(1 - \frac{1}{\alpha^{1/2}}\right) > \frac{n \log\log\log n}{\alpha^{1/2} \log n},$$

let C_{k+1} be the set of vertices in A_{k+1} sending fewer than $\log\log n$ edges to B_k. Since G satisfies (Q_1),

$$| C_{k+1} | < \frac{2n \log\log\log n}{\alpha \log n}.$$

Now set $B_{k+1} = A_{k+1} \setminus C_{k+1}$. We have defined sets B_k such that

$$| B_k | > \frac{2n \log\log\log n}{\alpha^{1/2} \log n}\left(1 - \frac{1}{\alpha^{1/2}}\right),$$

and each vertex of B_{k+1} sends at least $\log\log n$ edges to B_k, for all k.

There are at most $n/\alpha^{1/2}$ vertices not in $\bigcup B_k$. We claim that, for every vertex x in $\bigcup B_k$, there are fewer than $3n/\alpha^{1/2}$ vertices below x which are in $\bigcup B_k$ and are not $< x$ in $P(G; \prec)$. This will imply that

$$r(G; \prec) \geqq \binom{n}{2} - \frac{n}{\alpha^{1/2}} \cdot n - n \cdot \frac{3n}{\alpha^{1/2}} = \binom{n}{2} - \frac{4n^2}{\alpha^{1/2}},$$

as desired.

Fix k and any vertex x in B_k. We may assume that

$$k \geqq \frac{3 \log n}{2 \log\log\log n},$$

since otherwise there are fewer than $3n/\alpha^{1/2}$ vertices below x in $\bigcup B_k$. For $0 \leqq j < k$, we define

$$D_j = \{ y \in B_{k-j} : y \leqq x \text{ in } P(G; \prec) \}.$$

We claim that $|D_j|$ always grows at essentially the expected rate.

For

$$j \leqq j_0 \equiv \left\lceil \frac{\log n - 2\log\log n}{\log\log\log n - \log 4} \right\rceil,$$

we claim that $|D_j| \geqq (\frac{1}{4}\log\log n)^j$. Indeed, this is true for $j = 0$. Suppose it is true for $j-1$, so that $|D_{j-1}| \geqq (\frac{1}{4}\log\log n)^{j-1}$. Take any subset E_{j-1} of D_{j-1} with

$$|E_{j-1}| = (\tfrac{1}{4}\log\log n)^{j-1} \leqq n/(\log n)^2.$$

There are at least $(\frac{1}{4}\log\log n)^{j-1}\log\log n$ edges from E_{j-1} to B_{k-j}, since $E_{j-1} \subseteq B_{k-j+1}$, and so, since G satisfies (Q_2), E_{j-1} has at least $(\frac{1}{4}\log\log n)^j$ neighbours in B_{k-j}, and these are all in D_j. Therefore $|D_j| \geqq (\frac{1}{4}\log\log n)^j$, as claimed. In particular, we have that $|D_{j_0}| \geqq n/(\log n)^2$. From this point on, we must make do with the slower rate of growth given by (Q_3). Repeating the above argument using (Q_3) in place of (Q_2), we see that, for $j_0 \leqq j \leqq j_0 + j_1$, where

$$j_1 = \left\lceil \frac{\log\log n + \log\log\log\log n - \log\log\log n}{\log\log\log\log n - \log 2} \right\rceil,$$

we have $|D_j| \geqq n/(\log n)^2(\frac{1}{2}\log\log\log n)^{j-j_0}$.

In particular, we have

$$|D_{j_0+j_1}| \geqq \frac{n\log\log\log n}{\log n\log\log n}.$$

This is now large enough so that almost every vertex below $D_{j_0+j_1}$ sends an edge to $D_{j_0+j_1}$. Indeed, since G satisfies (Q_4), at most n/α vertices of G fail to send an edge to $D_{j_0+j_1}$. Hence

$$\begin{aligned} |\{y \in \bigcup B_k : y < x, y \not< x \text{ in } P(G;<)\}| &\leqq \frac{n}{\alpha} + (j_0 + j_1 + 1)\frac{2n\log\log\log n}{\alpha^{1/2}\log n} \\ &\leqq \frac{n}{\alpha^{1/2}}\left(\frac{1}{\alpha^{1/2}} + \frac{5}{2} + \frac{(\log\log n)^2}{\log n}\right) \\ &\leqq \frac{3n}{\alpha^{1/2}}, \quad \text{for } n \text{ sufficiently large.} \end{aligned}$$

This completes the proof of Theorem 2. □

Theorem 2 goes some way towards finding graphs G_n of order n and small size satisfying $t(G_n) = o(n^2)$: all we have to do is find G_n's satisfying (Q_1) to (Q_4). Nevertheless such graphs G_n are not easily constructed. To complete the proof of Theorem 1, we show that, in a certain range of $p(n)$, most random graphs $G_{n,p}$ have the properties.

PROOF OF THEOREM 1. By Theorem 2, all we have to show is that a.e. G_p satisfies (Q_1) to (Q_4). These verifications are entirely routine: in each case it suffices to calculate the expected number of pairs of subsets with the forbidden property, and show that this tends to 0 as $n \to \infty$.

(Q_1) What is the expected number E of pairs (S, D) contradicting (Q_1)? The number of choices for S and D is at most $\binom{n}{s}\binom{n}{d}$, and the probability that a given vertex sends fewer than log log n edges to S is

$$\sum_{k=0}^{\log\log n - 1} \binom{s}{k} p^k (1-p)^{s-k}.$$

Therefore

$$E \leqq \binom{n}{s}\binom{n}{d}\left[\sum_{k=0}^{\log\log n - 1} \binom{s}{k} p^k(1-p)^{s-k}\right]^d$$

$$\leqq \left(\frac{ne}{s}\right)^s\left(\frac{ne}{d}\right)^d\left[\log\log n\left(\frac{2esp}{\log\log n}\right)^{\log\log n} e^{-ps}\right]^d$$

$$= \left(\frac{e\alpha^{1/2}\log n}{\log\log\log n}\right)^s\left(\frac{e\alpha\log n}{2\log\log\log n}\right)^{2s/\alpha^{1/2}}$$

$$\times\,[\log\log n(2e\alpha^{1/2})^{\log\log n}e^{-\alpha^{1/2}\log\log n}]^{2s/\alpha^{1/2}}$$

$$= \exp\left\{s\left[\log\log n(1+o(1)) + \frac{2}{\alpha^{1/2}}\log\log n(1+o(1))\right.\right.$$

$$\left.\left.+\frac{2}{\alpha^{1/2}}\log\log n\log(2e\alpha^{1/2})(1+o(1)) - 2\log\log n\right]\right\}$$

$$= \exp\left\{2s\log\log n\left[\frac{1}{\alpha^{1/2}}(1+\log(2e\alpha^{1/2})) - 1\right](1+o(1))\right\}$$

$$\to 0 \qquad \text{as } n \to \infty.$$

Therefore a.e. G_p contains no such pair.

(Q_2) For fixed u, what is E_u, the expected number of pairs (U, V) with $|U| = u$ contradicting (Q_2)? The number of choices for U and V is at most

$$\binom{n}{u}\binom{n}{\frac{1}{4}u\log\log n};$$

we can choose the $u \log\log n$ edge from U to V in

$$\binom{\frac{1}{4}u^2\log\log n}{u\log\log n}$$

ways, and the probability that all these edges are in G_p is $p^{u\log\log n}$. Therefore

$$\begin{aligned}E_u &\leqq \binom{n}{u}\binom{n}{\frac{1}{4}u\log\log n}\binom{\frac{1}{4}u^2\log\log n}{u\log\log n}p^{u\log\log n}\\ &\leqq \left(\frac{en}{u}\right)^u\left(\frac{4en}{u\log\log n}\right)^{\frac{1}{4}u\log\log n}\left(\frac{eup}{4}\right)^{\log\log n}\\ &\equiv F_u;\end{aligned}$$

$$\begin{aligned}\frac{F_u}{F_{u-1}} &= \left(\frac{en}{u}\right)\left(\frac{4en}{u\log\log n}\right)^{\frac{1}{4}\log\log n}\left(\frac{eup}{4}\right)^{\log\log n}\left(1+\frac{1}{u-1}\right)^{(u-1)(1-\frac{3}{4}\log\log n)}\\ &\leqq \exp\left\{\log\left(\frac{en}{u}\right)+\frac{\log\log n}{4}\log\left(\frac{4en}{u\log\log n}\right)\right.\\ &\qquad\left.+\log\log n\log\left(\frac{eu\alpha\log n\log\log n}{4n\log\log\log n}\right)-\tfrac{1}{2}\log\log n\right\}\\ &< 1 \qquad \text{if } \tfrac{1}{4}\log\left(\frac{4en}{u\log\log n}\right) < \log\left(\frac{4n\log\log\log n}{eu\alpha\log n\log\log n}\right),\end{aligned}$$

which holds if $u \leqq n/(\log n)^2$. So the expected number of pairs (U, V) with the given properties is

$$\begin{aligned}\sum_{u=1}^{n/(\log n)^2} E_u &\leqq \frac{n}{(\log n)^2}F_1\\ &\leqq \frac{en^2}{(\log n)^2}\left(\frac{4en}{\log\log n}\right)^{\frac{1}{4}\log\log n}\left(\frac{e\alpha\log n\log\log n}{4n\log\log\log n}\right)^{\log\log n}\\ &\to 0 \qquad \text{as } n\to\infty.\end{aligned}$$

Hence almost no G_p contains such a pair (U, V).

(Q_3) The argument here is the same as for (Q_2), and we omit some details. For fixed u, the expected number of such pairs is at most

$$\begin{aligned}
&\binom{n}{u}\binom{n}{\frac{1}{2}u\log\log\log n}\binom{\frac{1}{2}u^2\log\log\log n}{u\log\log n}p^{u\log\log n}\\
&\leqq\left(\frac{en}{u}\right)^{\frac{1}{2}u\log\log\log n}\left(\frac{eu\alpha\log n}{2n}\right)^{u\log\log n}\\
&\equiv G_u;
\end{aligned}$$

$$\frac{G_u}{G_{u-1}}=\left(\frac{en}{u}\right)^{\frac{1}{2}\log\log\log n}\left(\frac{eu\alpha\log n}{2n}\right)^{\log\log n}\left(1+\frac{1}{u-1}\right)^{(u-1)(\frac{1}{2}\log\log\log n-\log\log n)}$$

which is less than 1 if

$$u\leqq\frac{n\log\log\log n}{\log n\log\log n}.$$

So the expected number of pairs (U, V) is at most

$$\begin{aligned}
\frac{n\log\log\log n}{\log n\log\log n}G_1=\exp\{&\log n(1+o(1))+\tfrac{1}{2}\log\log\log n\log n(1+o(1))\\
&-\log\log n\log n(1+o(1))\}
\end{aligned}$$

which tends to 0 as $n\to\infty$.

(Q_4) The expected number of such pairs (Y, Z) is at most

$$\begin{aligned}
\binom{n}{y}\binom{n}{z}(1-p)^{yz}&\leqq(\log n)^{2n/\log n}(e\alpha)^{n/\alpha}e^{-n/2}\\
&\to 0\qquad\text{as } n\to\infty.
\end{aligned}$$

This completes the proof. □

We next prove that, if $(pn\log\log\log n)/(\log n\log\log n)$ is bounded above, the $t(G_p)$ is almost surely at least εn^2, for some $\varepsilon>0$. In fact, we prove a somewhat stronger result, which shows that $t(G)>\varepsilon n^2$ for other classes of graphs G, with about the same number of edges, which are far from random.

THEOREM 3. *Let $\varepsilon>0$, a_1, a_2, $b\geqq 0$ be fixed constants. Suppose (for convenience) that $b(a_1+a_2)\geqq 1$. Suppose that $|G|=n$, and that there is a subset*

Y of $V(G)$ with order $\geqq \varepsilon n$ such that $\Delta(G[Y]) \leqq \log n(\log\log n)^{a_1}$, and furthermore there is a partition of Y into

$$s \leqq \frac{b \log n}{\log\log\log n}$$

classes $Z_1, Z_2, \ldots, Z_s$ such that, for each k, no component of $G[Z_k]$ has order greater than $(\log\log n)^{a_2}$. Then

$$t(G) \geqq \frac{\varepsilon^2}{2b^2(a_1+a_2)^2} n^2 + o(n^2).$$

Here $G[Z]$, for $Z \subseteq V(G)$, denotes the induced subgraph of G with vertex set Z, and $\Delta(H)$ denotes the maximal degree of a graph H.

Before proving Theorem 3, let us see how it can be used. First we note some simpler conditions implying the conditions of Theorem 3.

THEOREM 4. *Suppose $|G| = n$, $e(G) \leqq n \log n(\log\log n)^{c_1}$, and $\chi(G)$, the chromatic number of G, is at most $(c_2 \log n)/(\log\log\log n)$, where c_1 and c_2 are constants with $c_1c_2 \geqq 1$. Let η be any positive real number. Then*

$$t(G) \geqq \frac{(1-\eta)}{2c_1^2c_2^2} n^2,$$

provided n is sufficiently large.

PROOF. At most $\eta n/4$ vertices of G have degree greater than $\log n(\log\log n)^{c_1(1+\eta/4)}$, for n sufficiently large. Let Y be the set obtained by deleting these vertices from $V(G)$. Take a colouring of $G[Y]$ using

$$s \leqq \frac{c_2 \log n}{\log\log\log n}$$

colours, and call the colour classes $Z_1, \ldots, Z_s$. Then G, Y, (Z_i) satisfy the conditions of Theorem 3 with $\varepsilon = 1 - \eta/4$, $a_1 = c_1(1+\eta/4)$, $a_2 = 0$, and $b = c_2$. Hence the result follows. □

THEOREM 5. *Let $c \geqq 1$ be a constant, and suppose that*

$$p \leqq \frac{c \log n \log\log n}{n \log\log\log n}.$$

Then for a.e. G_p,

$$t(G_p) \geqq \frac{(1-\delta)}{2c^2} n^2, \qquad \textit{for any } \delta > 0.$$

PROOF. Certainly almost every G_p has at most $n \log n \log\log n$ edges, and by a result of Bollobás and Thomason [10] (see Chapter XI, Theorem 25, of Bollobás [5]), $\chi(G_p)$ is almost surely less than $((1+\delta/3)c \log n)/(\log\log\log n)$. Hence, by Theorem 4 with $c_1 = 1$, $c_2 = (1+\delta/3)c$ and $\eta = \delta/3$, we have

$$t(G_p) \geqq \frac{(1-\delta)}{2c^2} n^2,$$

almost surely, as required. □

Theorem 4 says that, if the chromatic number of G is small, then $t(G) \geqq \varepsilon n^2$. For random graphs, $\chi(G_p)$ is amost surely small enough, if $(pn \log\log\log n)/(\log n \log\log n)$ is bounded. If we are to produce graphs G with $(cn \log n \log\log n)/(\log\log\log n)$ edges such that $t(G) = o(n^2)$, we have to increase the chromatic number. One way to artificially increase $\chi(G)$ is to partition our n-set into sets of $(c \log n \log\log n)/(\log\log\log n)$ vertices, let each set span a complete graph, and add the edges of a random $G_{n,p}$, with

$$p = \frac{c \log n \log\log n}{n \log\log\log n}.$$

Denote the resulting random graph H_p. In fact, Theorem 3 implies that, in this case also, we almost surely have $t(H_p) \geqq \varepsilon n^2$.

We merely sketch the proof of this assertion. The random part of the graph almost surely has chromatic number at most $(2c \log n)/(\log\log\log n)$, and what is more we can take the colour classes each to have order at most $(n \log\log\log n)/(c \log n)$. Consider the various intersections of colour classes with complete graphs. Almost surely none of these intersections contains as many as $\log n$ vertices, and at most $n/(\log n)^2$ of them contain as many as $(\log\log n)^2$. Hence, by deleting at most $n/\log n$ vertices, we obtain a subset Y of $V(G)$, and subsets (Z_i) (the colour classes) satisfying the conditions of Theorem 3 with $a_1 = 1$, $b = 2c$, and $a_2 = 2$. Therefore $t(H_p)$ is almost surely at least εn^2, for some $\varepsilon > 0$.

To find a graph with $(cn \log n \log\log n)/(\log\log\log n)$ edges and $t(G) = o(n^2)$, we seem to need a graph with high chromatic number and yet good "spreading" properties.

Having seen how Theorem 3 can be applied, let us prove it.

PROOF OF THEOREM 3. Let δ be any positive constant, and set $b_0 = 1/(a_1 + a_2 + \delta)$, and

$$s_0 = \frac{b_0 \log n}{\log \log \log n}.$$

Without loss of generality the s_0 largest Z_i's are $Z_1, \ldots, Z_{s_0}$. Let $Y_0 = \bigcup_1^{s_0} Z_i$, and note that

$$|Y_0| \geqq \frac{\varepsilon b_0}{b} n.$$

We give each class Z_i $(1 \leqq i \leqq s_0)$ a fixed ordering $<_i$ of its vertices, and consider orderings of $V(G)$ of the following form. Those vertices not in Y_0 are put at the top; the classes $Z_1, \ldots, Z_{s_0}$ are taken in some order, and then the Z_i are placed in that order below $V(G)\backslash Y_0$, the vertices of Z_i appearing in the given order $<_i$. We claim that, in one of the $s_0!$ orderings $<$ of this form, there are $o(n^2)$ relations in $< |_{Y_0}$. What is more, if each permutation of $Z_1, \ldots, Z_{s_0}$ is given equal probability, we claim that the expected number of relations in $< |_{Y_0}$ is at most $n^{2-\delta/2(a_1+a_2+\delta)}$.

Which paths in $G[Y_0]$ can be chains in such an ordering $<$? The only possibilities are those paths of the form $x_{11} \cdots x_{1j_1} x_{21} \cdots x_{2j_2} \cdots x_{k1} \cdots x_{kj_k}$, where $x_{i1} \cdots x_{ij_i}$ is a chain in some $(Z_{l_i}; <_{l_i})$ for every i, and $k \leqq s_0$. Call such a path a *k-step candidate path*. Call two candidate paths *equivalent* if they have the same $x_{11}, x_{1j_1}, x_{21}, x_{2j_2}, \ldots, x_{k1}, x_{kj_k}$. (Thus equivalent paths differ only inside each Z_i.) Clearly a candidate path P is a chain in $<$ iff every path equivalent to P is a chain in $<$ iff every path equivalent to P is a chain in $<$. Therefore, if we are to count relations in $< |_{Y_0}$, we need only count once for each equivalence class.

How many equivalence classes of k-step candidate paths are there? We have $|Y_0|$ choices for x_{11}. Given x_{i1}, we choose x_{ij_i} from the $\leqq (\log \log n)^{a_2}$ vertices in the same component of $G[Z_{l_i}]$ as x_{i1}. Given x_{ij_i}, we choose $x_{i+1,1}$ from the neighbours of x_{ij_i} in Y_0, and there are at most $\log n(\log \log n)^{a_1}$ of these. Therefore the number of equivalence classes of k-step candidate paths is at most

$$|Y_0| (\log \log n)^{ka_2} [\log n (\log \log n)^{a_1}]^{k-1}.$$

The probability that a given k-step candidate path is a chain in $<$ is just $1/k!$, the probability that the k classes $Z_{l_1}, \ldots, Z_{l_k}$ are in the right order in $<$. Therefore the expected number of relations in $< |_{Y_0}$ is at most

$$\sum_{k=1}^{s_0} |Y_0| (\log n(\log\log n)^{a_1+a_2})^k \frac{1}{k!}$$

$$\leq \frac{\varepsilon b_0}{b} n \frac{b_0 \log n}{\log\log\log n} (eb_0 \log\log\log n(\log\log n)^{a_1+a_2})^{b_0 \log n/\log\log\log n}$$

$$= \exp\left\{\log\left(\frac{\varepsilon b_0^2 \log n}{b \log\log\log n}\right) + \log n\right.$$

$$\left. + \frac{\log n}{(a_1+a_2+\delta)\log\log\log n}((a_1+a_2)\log\log\log n + \log(eb_0 \log\log\log n))\right\}$$

$$= \exp\left\{\log n\left(2 - \frac{\delta}{a_1+a_2+\delta}\right) + o(\log n)\right\}$$

$$\leq n^{2-\frac{1}{2}\delta/(a_1+a_2+\delta)}.$$

Therefore the expected value of $r(G; <)$ is at most

$$\binom{n}{2} - \binom{|Y_0|}{2} + n^{2-\delta/2(a_1+a_2+\delta)};$$

and so

$$t(G) \geq \frac{\varepsilon^2 b_0^2}{2b^2} n^2 + o(n^2) = \frac{\varepsilon^2}{2b^2(a_1+a_2+\delta)^2} n^2 + o(n^2).$$

Since δ was arbitrary,

$$t(G) \geq \frac{\varepsilon^2}{2b^2(a_1+a_2)^2} n^2 + o(n^2),$$

as required. □

Let us now see that, if we restrict the number of edges of G somewhat further than we do in Theorem 5, then *every* G has $t(G) \geq \varepsilon n^2$. Using a somewhat different technique, Alon, Azar, and Vishkin [4] have proved the same result (with a worse constant).

THEOREM 6. *Let c be a constant, and suppose that $|G| = n$ and $e(G) \leq \frac{1}{2}cn \log n$. If n is sufficiently large, then*

$$t(G) \geq n^2\left(\frac{2}{3c+5}\right).$$

To prove instead that, say,

$$t(G) \geqq n^2\left(\frac{1}{8c} - \frac{1}{128c^2}\right)$$

is fairly simple. We first remove all vertices (at most $n/2$) of degree $\geqq 2c \log n$. Then we take a random ordering $\prec$ of the remaining vertices, and calculate the expected number of chains which span at most δn vertices (i.e. the expected number of chains $x_1 \cdots x_k$ with at most δn vertices between x_1 and x_k in $\prec$), where $\delta = 1/8c$. This is easily seen to be $o(n^2)$. The proof of Theorem 6 is a refinement of this proof. We omit the arithmetical details.

PROOF. Given a graph G with $|G| = n$, and $e(G) \leqq \frac{1}{2}cn \log n$, we define inductively a sequence (G_r) of induced subgraphs of G as follows. Set $G_1 = G$. Suppose that G_r has $n - r + 1$ vertices and $\frac{1}{2}c_r(n - r + 1)\log(n - r + 1)$ edges. If G_r has a vertex of degree $> (\frac{3}{2}c_r + \frac{7}{4})\log(n - r + 1)$, form G_{r+1} by deleting any such vertex. In this case, set

$$c_{r+1} = \frac{2e(G_{r+1})}{(n - r)\log(n - r)}.$$

So G_{r+1} has $n - r$ vertices and $\frac{1}{2}c_r(n - r)\log(n - r)$ edges. If, on the other hand,

$$\Delta(G_r) \leqq \left(\frac{3c_r}{2} + \frac{7}{4}\right)\log(n - r + 1),$$

stop the process.

Let G_0 be the final graph G_r, set $n_0 = |G_0|$, and define c_0 by $e(G_0) = \frac{1}{2}c_0 n_0 \log n_0$. We shall prove that

$$t(G_0) \geqq n_0^2\left(\frac{2}{3c_0 + 5}\right) \geqq n^2\left(\frac{2}{3c + 5}\right),$$

and therefore

$$t(G) \geqq n^2\left(\frac{2}{3c + 5}\right),$$

since we can put the vertices of $G \setminus G_0$ at the top of the ordering.

It is straightforward to check that, provided $n - r \geqq e^{6c_r}$, we have

$$(n - r)^2\left(\frac{2}{3c_{r+1} + 5}\right) > (n - r + 1)^2\left(\frac{2}{3c_r + 5}\right),$$

and thus, for every $r \leqq n - e^{6c}$,

$$(n-r)^2\left(\frac{2}{3c_{r+1}+5}\right) > n^2\left(\frac{2}{3c+5}\right).$$

In particular, we see that n_0 is at least $n(5/(3c+5))^{1/2}$, and so, provided n is sufficiently large,

$$n_0^2\left(\frac{2}{3c_0+5}\right) > n^2\left(\frac{2}{3c+5}\right).$$

We now show that, provided n_0 is sufficiently large,

$$t(G_0) \geqq n_0^2\left(\frac{2}{3c_0+5}\right).$$

We recall that

$$\Delta(G_0) \leqq \left(\frac{3c_0}{2}+\frac{7}{4}\right)\log n_0.$$

Set

$$\varepsilon = \left(\frac{3c_0}{2}+\frac{9}{5}\right)^{-1},$$

and note that

$$\varepsilon\left(\frac{3c_0}{2}+\frac{7}{4}\right) < 1 - \frac{1}{30c+36}.$$

The number of paths of length k in G_0 is at most $n_0[(\frac{3}{2}c_0+\frac{7}{4})\log n_0]^k$. If we take a random ordering $<$ of the vertices of G_0, with each ordering equally likely, the probability that the vertices of a given path $L = x_1x_2\cdots x_{k+1}$ of length k appear in the order given by L, and are all no more than εn_0 higher than x_1 (i.e. $x_1 < x_2 < \cdots < x_{k+1}$, and there are at most $\varepsilon n_0 - 1$ vertices y with $x_1 < y < x_{k+1}$) is at most $\varepsilon^k/k!$. Hence the expected number of chains of length k in $(G_0; <)$ spanning at most εn_0 vertices is at most

$$n_0\left[\left(\frac{3c_0}{2}+\frac{7}{4}\right)\log n_0\right]^k \varepsilon^k/k! \leqq n_0\left[\left(1-\frac{1}{30c+36}\right)\log n_0\right]^k\frac{1}{k!}.$$

This expression is maximised when

$$k = \left(1-\frac{1}{30c+36}\right)\log n_0,$$

and it is less than 1 when $k \geqq 4 \log n_0$. Therefore the expected number of chains in $(G_0; <)$ spanning at most εn_0 vertices is at most

$$\sum_{k=1}^{\varepsilon n_0} n_0 \left[\left(1 - \frac{1}{30c + 36}\right) \log n_0\right]^k \frac{1}{k!}$$

$$\leqq \varepsilon n_0 + 4 \log n_0 \, n_0 \exp\left[\left(1 - \frac{1}{30c + 36}\right) \log n_0\right]$$

$$\leqq n_0^{2-(30c+37)^{-1}},$$

for sufficently large n_0.

In $(G_0; <)$, there are $\varepsilon n_0^2 - \frac{1}{2}\varepsilon^2 n_0^2$ pairs of vertices at most εn_0 apart in $<$ and, in some ordering $<$, at most $n_0^{2-(30c+37)^{-1}}$ of these pairs are related in $P(G_0; <)$. Hence

$$t(G_0) \geqq \varepsilon n_0^2 - \tfrac{1}{2}\varepsilon^2 n_0^2 - n_0^{2-(30c+37)^{-1}}.$$

It is straightforward to check that

$$\varepsilon - \tfrac{1}{2}\varepsilon^2 > \frac{2}{3c_0 + 5},$$

and hence

$$t(G_0) \geqq n_0^2 \left(\frac{2}{3c_0 + 5}\right).$$

This completes the proof. □

If G is close to regular, the expected number of relations missing in a randomly chosen ordering $<$ is almost exactly $n^2(1/c - 1/2c^2)$, so Theorem 6 is almost the best result that any such "averaging" argument can give. However, the method involved is still fairly crude, and we make no attempt to choose a "good" ordering, beyond putting all vertices of high degree at the top.

For small c, one can prove a stronger result: if G is a graph with n vertices and fewer than $\frac{1}{2}cn \log n$ edges, then $t(G) \geqq \frac{1}{2}n^2(1 - c/2)^2(1 - \varepsilon)$, for any positive ε and n sufficiently large.

Theorem 5 implies (see Chapter II, Theorem 2 of Bollobás [5]) that *almost every* graph G with n vertices and at most $(cn \log n \log\log n)/(\log\log\log n)$ edges has $t(G) \geqq \delta(c)n^2$. Theorem 6 says that *every* graph G with n vertices and at most $cn \log n$ edges has $t(G) \geqq \varepsilon(c)n^2$. This suggests the following question. Is there a function $\omega(n) \to \infty$ such that every G with n vertices and at most

$\omega(n)n \log n$ edges has $t(G) \geqq \varepsilon n^2$, for some fixed positive ε? Theorem 1 implies that such a function $\omega(n)$ must be such that $(\omega(n) \log\log\log n)/(\log\log n)$ is bounded. We conjecture that there is such an $\omega(n)$, and furthermore that we can take

$$\omega(n) = \frac{c \log\log n}{\log\log\log n}.$$

CONJECTURE. *Every graph with n vertices and at most*

$$(cn \log n \log\log n)/(\log\log\log n)$$

edges has $t(G) \geqq \delta(c)n^2$, *for some* $\delta(c) > 0$.

REFERENCES

1. M. Ajtai, J. Komlós, W. Steiger and E. Szemerédi, *Almost sorting in one round*, preprint.

2. M. Ajtai, J. Komlós and E. Szemerédi, *An $O(n \log n)$ sorting network*, ACM Symposium on Theory of Computing **15** (1983), 1–9.

3. M. Ajtai, J. Komlós and E. Szemerédi, *Sorting in $c \log n$ parallel steps*, Combinatorica **3** (1983), 1–19.

4. N. Alon, Y. Azar and U. Vishkin, *Tight complexity bounds for parallel comparison sorting*, Proc. 27th Annal Symp. on Foundations of Computer Science, Toronto, IEEE, 1986, pp. 502–510.

5. B. Bollobás, *Random Graphs*, Academic Press, London, 1985, xv + 447pp.

6. B. Bollobás and G. Brightwell, *Transitive orientations of graphs*, submitted.

7. B. Bollobás and P. Hell, *Sorting and graphs*, in *Graphs and Order* (I. Rival, ed.), NATO ASI Series, Reidel, Dordrecht, Boston, Lancaster, 1984, pp. 169–184.

8. B. Bollobás and M. Rosenfeld, *Sorting in one round*, Isr. J. Math. **38** (1981), 154–160.

9. B. Bollobás and A. Thomason, *Parallel sorting*, Discrete Appl. Math. **6** (1983), 1–11.

10. B. Bollobás and A. G. Thomason, *Random graphs of small order*, in *Random Graphs*, Ann Discrete Math. (1985), 47–97.

11. R. Häggkvist and P. Hell, *Graphs and parallel comparison algorithms*, Congr. Num. **29** (1980), 497–509.

12. R. Häggkvist and P. Hell, *Parallel sorting with constant time for comparisons*, SIAM J. Comput. **10** (1981), 465–472.

13. R. Häggkvist and P. Hell, *Sorting and merging in rounds*, SIAM J. Alg. Discrete Methods **3** (1982), 465–473.