and so 24 becomes

$$- \ln \Pr[A^*_{t+dt}] > - \ln \Pr[A^*_t] + \frac{1}{2}\binom{k}{2} n^{-1/2}[f(t) - \frac{2L}{c}]dt \tag{38}$$

As $\Pr[A_0] = 1$, this gives a bound for $-\ln \Pr[A^*_c]$ by taking steps of dt. Letting $dt \to 0$ (or be an "infinitesmal") gives

$$- \ln \Pr[A^*_c] > \frac{1}{2}\binom{k}{2} n^{-1/2} \int_0^c [f(t) - \frac{2L}{c}]dt = 4L\binom{k}{2} n^{-1/2} \tag{39}$$

That is,

$$\Pr[A^*_c] < e^{-2Lk^2 n^{-1/2}} = (n^{-2L\epsilon})^k \tag{40}$$

which is certainly $o(n^{-r})$.

Reintroducing S as a parameter we have $\Pr[\vee_S A^*_c(S)] = o(1)$. Thus, recalling 36,

$$\Pr[\vee_S A_c(S)] \leq \Pr[\vee_S A^*_c(S)] + \Pr[\vee_S W_c(S)] = o(1) \tag{41}$$

But $\neg A_c(S)$ is the event that IND returns Yes and that implies that S is not independent in G_c. Thus $\wedge_S \neg A_c(S)$ implies $\alpha(G_c) \leq k$ giving 2 and hence the Ramsey bound 3.

and if this occurs

$$\sum \binom{X_i^*}{2} = \sum_{u=L}^{U} (Z_u - Z_{u-1}) \binom{u}{2} \tag{35}$$
$$\leq \sum_{u=L}^{U-1} (u-1) Z_u + Z_U \binom{U}{2}$$
$$= O(n^{.5}k) + O(n^{.5}k) = O(n^{.5}k)$$

□

In our case the $\deg(v)$ for $v \notin S$ have precisely these independent distributions X_v. As $n^{.5}k = o(k^2)$

$$\Pr[|VEE| \geq .4 \binom{k}{2} \wedge \neg W] = o(n^{-r})$$

and so $\Pr[W(S) \wedge \neg W] = o(n^{-r})$ and

$$\Pr[\vee_S W(S)] \leq \Pr[W] + \sum_S \Pr[W(S) \wedge \neg W] = o(1) \tag{36}$$

Indeed, we could have strengthened W_t to require $|UNEX| = o(k^2)$.

5.6 Putting It All Together

The probability of generating a particular representation of a tree T is

$$\sim n^{-r} \int_\Gamma e^{-c^2 - y_1^2 - \ldots - y_{2r}^2} dy_1 \cdots y_{2r}$$

The n^{-r} factor is cancelled by the number $\sim n^r$ of possible representations, giving 31.

We have selected D so that $\sum f(T,t)$ over all T with at most D edges is at least $1 - \frac{L}{c}$. Thus the sum over all such T whose root survives is at least $f(t) - \frac{L}{c}$. Conditioning on the birth of e at time t for each such T the probability of generating T as the twintree of e is in the limit $f(T,t)$ and therefore for n sufficiently large the probability of generating one of these T will be at least, say, $f(t) - \frac{2L}{c}$. Combining with the probability of an examinable e being born in $[t, t+dt]$ gives

$$\Pr[\neg A_{t+dt} | A_t^*] > \frac{1}{2} \binom{k}{2} n^{-1/2} [f(t) - \frac{2L}{c}] dt \tag{37}$$

5.5 Wierd Events are Rare

For given n, D, c we write $W(S)$ for the "wierd" event $W(n, S, D, c)$. With $\binom{n}{k} \leq n^k$ different S we will try to bound events with probabiities $o(n^{-k})$. Let G^* denote all e with $x_e \leq c$. Note that the distribution of G^* is precisely that of $G(n,p)$, $p = cn^{-1/2}$. Let W be the event that some vertex of G^* has degree bigger than $2cn^{1/2}$. Elementary estimates give $\Pr[G^*] = o(1)$. Let $W^1(S)$ be the event that G^* has $n^{1/2}\ln^6 n$ edges e in S. Elementary estimates give $\Pr[W^1(S)] = o(n^{-k})$. If $\neg W^1(S)$ then KN has at most $n^{1/2}\ln^6 n$ edges $e \in S$. Then $O(k\ln^{-5} n)$ vertices $i \in S$ can have $\ln^{10} n$ neighbors $j \in S$ in KN so only $O(k^2\ln^{-5} n)$ e in S can be in $UNEX - VEE$.

Let $W^2(S)$ be the event that $|VEE| \geq \frac{1}{3}\binom{k}{2}$ and $\neg W$. For $v \notin S$ let $\deg(v)$ be the number of neighbors of v in S in G^* and $\deg^-(v)$ the number in KN so that $\deg^-(v) \leq \deg(v)$ and

$$|VEE| \leq \sum_v |VEE(v)| = \sum_v \binom{\deg^-(v)}{2} \tag{33}$$

If $\deg(v) \leq n^{.2}$ then $|VEE(v)| \leq n^{.4}$. But if $v \notin SUL$ then $VEE(v) = \emptyset$ and $|SUL| = n^{1/2+o(1)}$ so the total contribution to $|VEE|$ from these v is $O(n^{.9+o(1)}) = o(k^2)$. Now we consider v with $n^{.2} < \deg(v) \leq 2cn^{1/2}$. We need a technical lemma, very similar to one used by Erdős in his 1961 paper.
Lemma. Let $X_1, \ldots, X_{n-k}$ be indepedent random variables, each with Binomial Distribution $B(k,p)$. Set $X_i^* = X_i$ if $n^{.2} \leq X_i \leq 2cn^{1/2}$, otherwise $X_i^* = 0$. Then

$$\Pr\left[\sum_{i=1}^{n} \binom{X_i^*}{2} = O(n^{.5}k)\right] = 1 - o(n^{-k}) \tag{34}$$

Proof. Set $L = n^{.2}$, $U = 2cn^{.5}$ for convenience. For $L \leq u \leq U$ and all i

$$\Pr[X_i \geq u] \leq \binom{k}{u} p^u \leq \left[\frac{kep}{u}\right]^u < n^{.1u}$$

Let Z_u be the number of i with $X_i \geq u$. As the X_i are independent we crudely bound

$$\Pr[Z_u \geq \frac{30k}{u}] \leq 2^n (n^{-.1u})^{30k/u} < n^{-2k}$$

$$\Pr[Z_u < \frac{30k}{u}, L \leq u \leq U] > 1 - n \cdot n^{-2k} = 1 - o(n^{-k}$$

where $f(T,t)$ is the probability of the branching process of §2 with $c = t$ yielding T.

The proof is similar to that of the previous Claim. Say T has $2r$ edges, labelled $1, \ldots, 2r$. There are $(n - |SUL|)_r \sim n^r$ possible representations of T, fix a particular one. Let edge i be $\{top(i), bot(i)\}$ as before, with edge 0 being e itself, $e = \{top(0), bot(0)\}$. Let , be as before except that all coordinates are in $[0,t]$. For a given $(y_1, \ldots, y_{2r}) \in$, we want the probability of getting this particular T with $x_i \in [y_i, y_i + dy_i]$. From Lemma 3 of §5.3 the conditional probability is $\sim n^{-1/2} dy_i$ of x_i being in the right interval and this remains true for x_i even if we condition on $x_0, \ldots, x_{i-1}$ so the conditional probability of having these edges is $\sim n^{-r} dy_1 \cdots cy_{2r}$.

Now further condition on the values y_i for the tree edges. We still need that $CHECK(e)$ will have no further edges in T. For each of the finite number of pairs u, v of vertices, both involved in the tree T but not an edge of T, the conditional probability that $x_{u,v} < c$ is (Lemma 2 of §5.3) $\leq cn^{-1/2} = o(1)$ so almost surely none of these will affect $CHECK(e)$. For each edge $0 \leq i \leq 2r$ and each vertex v not in the tree nor in S let $B_{v,i}$ be the event that the birthdates of both $\{v, top(i)\}$ and $\{v, bot(i)\}$ are at most y_i.

When can $\{v, top(i)\}$ be in KN ? We must have $top(i)$ a vertex of e as the other vertices are not in SUL. We cannot have both $\{v, top(i)\}, \{v, bot(i)\} \in KN$ because that would mean $i = 0$ (edge e) but $e \in EXAM$. (This is why we require $e \notin VEE$ for being examinable.) From $\neg W_t$ there are only $O(\ln^{10} n)$ vertices v for which any $\{v, top(0)\} \in KN$. (Similarly, $bot(0)$.) With probability $1 - o(1)$ for all such v and all vertices u of the tree for which $\{v, u\} \notin KN$ the birthdate of $\{v, u\}$ is greater than c, so almost surely none of these will affect $CHECK(e)$.

Now consider any other v, there being $\sim n$ of them. Applying the Lemma $\Pr[B_{v,i}] \sim y_i^2 n^{-1}$ given the conditioning. Also $\Pr[B_{v,i} \wedge B_{v,i'}] = O(n^{-3/2})$ for any two edges i, i' so that

$$\Pr[\vee_{i=0}^{2r} B_{v,i}] \sim n^{-1} \sum_{i=0}^{2r} y_i^2$$

Again from the Lemma this holds for v even after further conditioning on $\wedge \neg B_{k',i}$ for any number of other $v' \neq v$ and $0 \leq i \leq 2r$. Thus

$$\ln \Pr[\wedge_v \wedge_i \neg B_{v,i}] \sim \sum_v n^{-1} \sum_i y_i^2 \sim \sum_i y_i^2 \tag{32}$$

But A is independent of D and C_i so

$$\Pr[D|C_iA] = \Pr[D|C_i] \geq \Pr[D]$$

which gives the lower bound. $\Box$

Lemma 2. For all distinct $f, g \in P$

$$\Pr[x_f \leq t|C] \leq \Pr[x_f \leq t] = tn^{-1/2} \tag{29}$$

$$\Pr[x_f, x_g \leq t|C] \leq \Pr[x_f, x_g \leq t] = t^2 n^{-1} \tag{30}$$

Proof. Direct application of FKG Inequality. $\Box$

In our application P is the set of pairs not in KN and not in S. The condition $\min(x_e, x_f) \geq t$ when $e \in KN$ is discarded when $x_e \geq t$ and replaced by $x_f \geq t$ when $x_e < t$. $f \in Q$ when some condition $x_f \geq t$ remains. Pairs $f = \{v, u\}$ and $g = \{v, w\}$ are adjacent in H if some $VEEPROBE(e, v, t_e)$, $e = \{u, w\}$ returned nil.

Claim: If $\neg W_t$ then $\deg(f) \leq 2(D+1)\ln^{10} n$.

Consider the possible $e = \{u, w\}$ above. Suppose $u \in S$. If $w \in S$ then $VEEPROBE$ was called during $CHECK(e)$ and this can occur for at most $\ln^{10} n$ different e by the requirement $e \in EXAM$. If $w \notin S$ then $VEEPROBE$ was called during $CHECK(e')$ with $u \in e'$ as otherwise $CHECK(e')$ would have terminated as soon as it reached the sullied vertex u. There are again at most $\ln^{10} n$ different e' and now at most $D\ln^{10} n$ different e. Now suppose $u \notin S$. A $VEEPROBE$ of an e with vertex u can only occur during that call of $CHECK(e')$ in which u becomes sullied as after its sullied finding it terminates $CHECK$. Thus $VEEPROBE$ of such an e could only be called at most D times. $\Box$

Lemma 3. If $\neg W_t$ and f is a pair not in S with $f \notin KN$ and for which no conclusion $x_f \geq a_f$ can be drawn and if $I \subseteq [0, t]$ is an interval of length u then the conditional (on the Oracle responses up to time t) probability that $x_f \in I$ is asymptotically the unconditional probability $un^{-1/2}$.

Proof. Lemma 2 and the Claim.

5.4 Twintree Probability, Conditionally

Now we return to the conditional probability for $CHECK(e)$ to return Success. For each twintree T with at most D edges let $g(T)$ be the probability $CHECK(e)$ terminates having constructed that twintree.

Claim:

$$\lim_{n\to\infty} g(T) = f(T, t) \tag{31}$$

independent and uniform. Let H be a graph on P. To each $e \in Q$ associate a t_f and to each $\{e,f\} \in H$ associate a t_{ef}. Assume all $t_f, t_{fg} \leq t$. Let C be the condition

$$C = \wedge_{e \in Q}(x_e \geq t_e) \wedge \wedge_{\{e,f\} \in H}(\min(x_e, x_f \geq t_{ef})$$

Let $I \subseteq [0,t]$ be an interval of length u.
Lemma 1. For $f \in P - Q$

$$un^{-1/2}(1 - tn^{-1/2})^{\deg(f)} \leq \Pr[x_f \in I | C] \leq \frac{u}{n^{1/2} - t} \tag{26}$$

Proof: With f fixed let C_d be the conjuntion of the $\deg(f)$ events in C involving f (the dependent conditions) and C_i the conjunction of the other events so that $C = C_d C_i$. Let A denote $x_f \in I$, B denote $x_f > t$. Then

$$\Pr[C|A] \leq \Pr[C_i|A] = \Pr[C_i] = \Pr[C_i|B] = \Pr[C|B], \tag{27}$$

the last as $B \Rightarrow C_d$. Thus

$$\begin{aligned} \Pr[A|C] &\leq \frac{\Pr[A|C]}{\Pr[B|C]} \\ &= \frac{\Pr[A]}{\Pr[B]} \frac{\Pr[C|A]}{\Pr[C|B]} \\ &\leq \frac{\Pr[A]}{\Pr[B]} \end{aligned} \tag{28}$$

giving the upper bound. Let D denote $\wedge x_g > t$, the conjunction over the $\deg(f)$ neighbors g of f. Let C_{id} be the conjunction of the events in C_i involving and such g and C_{ii} the conjunction of all the other events in C_i so that $C_i = C_{ii} C_{id}$. But $D \Rightarrow C_{id}$ so

$$\Pr[C_i|D] = \Pr[C_{ii}|D] = \Pr[C_{ii}] \geq \Pr[C_i]$$

which implies

$$\Pr[D|C_i] \geq \Pr[D]$$

Now we bound

$$\Pr[A|C] = \Pr[A|C_d C_i] \geq \Pr[AC_d|C_i]$$

As A, C_i are independent and $D \Rightarrow C_d$

$$\Pr[A|C] \geq \Pr[A] \Pr[C_d|C_i A] \geq \Pr[A] \Pr[D|C_i A]$$

will go down by at most $\frac{L}{c}cn^{3/2}/2$, small compared to the $10Ln^{3/2}/2$ edges accepted.

We have to be careful of "wierd" events. Let W_t (or, more formally, $W(n,S,D,t)$) be the event that in running IND with $c = t$ either

- There are at least $n^{1/2}\ln^6 n$ edges e in S with $e \in KN$, or
- $|UNEX| \leq \frac{1}{2}\binom{k}{2}$.

Let $A_t^* = A_t \wedge \neg W_t$. It will turn out that A_t^* is essentially A_t.

5.2 Birth of an Examinable Edge

Now let $t \in [0,c)$ and let dt be infinitesmal. We bound

$$\Pr[\neg A^*_{t+dt}|A^*_t] > \Pr[\neg A_{t+dt}|A^*_t] \tag{24}$$

and this we shall bound from below.

Indeed, we shall bound from below the probability of $\neg A_{t+dt}$ conditional on any particular history of the procedure IND up to time t satisfying A_t^*. Formally we could modify IND so that $NEXTBORN$ first returns TempHalt if there is no $e \in EXAM$ with $x_e \leq t$ and after TempHalt has been outputted we continue calling $NEXTBORN$ as originally defined. We condition on all the outputs of the Oracle up to TempHalt and the precise values of x_e, $e \in KN$. Because $\neg W_t$

$$|SUL| \leq |KN| + |S| = O(n^{1/2}\ln^6 n) = o(n) \tag{25}$$

Because $\neg W_t$, $NEXTBORN$ is called with $|EXAM| \geq \frac{1}{2}\binom{k}{2}$. Each $e \in EXAM$ has $x_e \in [t, t+dt]$ with probability $dt/[n^{1/2}-t] \sim n^{-1/2}dt$. (For e in S, the Oracle checks x_e only during $NEXTBORN$ – this is the reason for the exception in $FULLPROBE$ – and so the $e \notin KN$ conditionally have x_e independent and uniform in $[t, n^{1/2}]$.) Thus with probability $\sim \frac{1}{2}\binom{k}{2}n^{-1/2}dt$ the call $NEXTBORN$ returns an e with $x_e \in [t, t+dt]$. We fix this e and bound from below the probability that $CHECK(e)$ returns Success. Lets review the conditioning at this point.

- For $f \in KN$ x_f is known. All such $x_f \leq t$, except $x_e \in [t, t+dt]$.
- For some $f = \{i,j\} \in KN$, $v \neq i,j$ it is known that $\min(x_{i,v}, x_{j,v}) > x_f$ since the Oracle responded nil to $VEEPROBE(f, v; x_f)$.

5.3 A General Conditioning Lemma

It will be helpful to consider the conditioning in a slightly more abstract situation. Let $Q \subset P$ be finitie sets. For each $e \in P$ let $x_e \in [0, n^{1/2}]$ be exceeds D we terminate the subprocedure $CHECK$ with output Failure. (This is a critical "give up" aspect of the algorithm. By not probing further the twintree of e we are retaining a relative independence of many of the x_e.)

Now we can describe $CHECK(e)$. Set $T = \{e\}$. At each stage take an $f \in T$ (we can imagine keeping a stack here) for which $FULLPROBE(f)$ has not yet been called and call $FULLPROBE(f)$. (The first call is to $FULLPROBE(e)$.) The procedure may terminate inside $FULLPROBE$ for one of the two reasons above. Otherwise, at some stage all $f \in T$ have had $FULLPROBE(f)$ called. We now give T a twintree structure with e the root and letting g, h be twins of f if they were returned during some $VEEPROBE(f, u, x_f)$ Check wheter the root e survives the twintree T in the sense of §2. If it does $CHECK$ returns Success, if it does not $CHECK$ returns Failure.

When $CHECK$ returns Succes then we terminate $IND(S, D)$ with output Yes. Otherwise we loop back to $NEXTBORN$. This concludes the description of IND.

We claim that if IND returns Yes then S must contain an edge in G_c. Let e be the edge for which $CHECK(e)$ returned Success, so $x_e \leq c$. When the twintree T generated by $CHECK(e)$ is the actual twintree of the relevant history of e then indeed $e \in G_c$. The only way this would not be is if during $CHECK(e)$ a subprocedure $FULLPROBE(f)$ was called where $f = \{v, w\}$ had $v \in S$. In that case $FULLPROBE$ does not check vertices $u \in S$. The relevant history would be different if for some $u \in S$ both $\{u, v\}, \{u, w\}$ were born before f. This could only affect whether f is accepted (and hence whether e is accepted) if both $\{u, v\}, \{u, w\}$ were accepted. But in that case we would have $\{u, v\} \in G_c$, so again S would have an edge in G_c.

For $0 \leq t \leq c$ let A_t be the event (other variables understood) that IND returns Yes by time t – i.e., that some subprocedure $CHECK(e)$ with $x_e \leq t$ has returned Success. If $\neg A_c$ then we have just argued that S is not independent in G_c. Thus

$$\Pr[\neg A_c] > \Pr[S \text{ independent in } G_c] \tag{23}$$

This probability will prove more tractible.

Now to be explicit about D. We pick D so large that for $c' \leq c$ the probability of the random twintree T (of §2) having more than D edges is less than L/c. The intuitive sense here is that the probability of accepting an edge will go down by at most L/c and so the total number of edges accepted

$UNEX$ A set of $e = \{i, j\}$ in S called unexaminable pairs. These consist of:

(a) All $e \in KN$.

(b) All $e = \{i, j\}$ so that $\{i, v\} \in KN$ for at least $\ln^{10} n$ vertices v. (Or similarly for j.)

(c) VEE

$EXAM$ All e in S not in $UNEX$, called examinable pairs.

The importance of $UNEX$ will become more apparent later. Basically we do not want to explore $e \in UNEX$ because they have been too "tarnished" by earlier explorations.

It will be useful to imagine the values x_e hidden from us and that probes are made of an Oracle. Formally such probes are given by the following subprocedures.

$NEXTBORN(K)$ K a set of pairs. The output is that $e \in K$ with x_e minimal. e is added to KN and time t is updated to x_e. Exception: If no $e \in K$ has $x_e \leq c$ we return "nil". (That is, we "stop" at time c.)

$VEEPROBE(e, v, t)$ Here $e = \{i, j\}$, i, j, v are distinct vertices, t real. If

$$\min[x_{i,v}, x_{j,v}] \leq t$$

then output $\{i, v\}, \{j, v\}$ and add these pairs to KN. Otherwise the output is "nil".

Now we describe IND. The outer loop is a call $NEXTBORN(EXAM)$. If the return is nil then IND terminates with output Maybe. Otherwise let $e = \{i, j\}$ be the output. We call a subprocedure $CHECK(e)$. $CHECK$ will keep an auxilliary set variable T, meant to reflect the twintree of e in G_t. Initially set $T = \{e\}$. $CHECK(e)$ will have outputs Success and Failure.

$CHECK$ will have a subprocedure $FULLPROBE(f)$, $f = \{v, w\}$. In $FULLPROBE$ we take all vertices $u \neq v, w$ of the graph in an arbitrary order and call $VEEPROBE(f, u, x_f)$. An exception: if either $v, w \in S$ we do not query for $u \in S$. When the response is nil we go to the next u. Otherwise we add the returned edges $\{u, v\}, \{u, w\}$ to T. (Recall $VEEPROBE$ adds them to KN.) Two important cases: If $v \in SUL$ then we terminate the subprocedure $CHECK$ with output Failure. Also, if the size of T now

5.1 A Modified Dynamic Algorithm

Fix $\epsilon > 0$, arbitrarily small. Fix a particular k-set S with $k = \epsilon n^{1/2}(\ln n)$. Set $L = 3\epsilon^{-1}$ (so that the random graph "works"). Pick c so that (giving ourselves room) $F(c) = 10L$. Our object is to bound the probability that S is independent in G_c.

We modify the dynamic algorithm of our opening paragraph to a procedure we'll call IND. IND has five parameters.

n The number of vertices. Let $V = \{1, \ldots, n\}$ denote the vertex set. The 2-sets $e \subset V$ are called pairs.

S A subset of V. We let k denote the size of S. We'll say e is *in* S if both its vertices are in S.

D A positive integer which plays a key role in telling when to "give up" the search.

c Nonnegative real. The "total time" for IND.

x A function on the pairs. For each e $x_e \in [0, n^{1/2}]$. All probabilities are with the underlying assumption that the x_e are independent and uniform. (We further assume all x_e are distinct, which occurs with probability one.)

The possible outputs of IND are Yes and Maybe. Yes will imply that G_c has an edge in S. Maybe still allows this possibility as we are deliberately not doing a full test. IND will keep these variables.

t Real. The current "time". While formally this is simply another variable we think of IND running dynamically in t. Initially $t = 0$.

KN A set of pairs call "known". These are the pairs or which we "know" x_e. Initially $KN = \emptyset$.

Several auxilliary variables are defined for convenience in terms of KN.

SUL A set of vertices called sullied. $v \in SUL$ if some $\{v, w\} \in KN$.

$VEE(v)$ Defined for $v \notin S$, this is the set of $e = \{i, j\}$ in S for which $\{v, i\}, \{v, j\} \in KN$.

VEE The union of $VEE(v)$ over $v \notin S$.

What is the usual value of Z^f? As $Z^f \geq Z_c$ we've shown that $E[Z^f]$ grows faster than $n^{3/2}$. *We conjecture that* $Z = \Theta(n^{3/2}(\ln n)^{1/2})$ *almost always.* We know that for c fixed $E[Z_c] \sim F(c)n^{3/2}/2$. A simple analysis of 20 gives that

$$F(c) \sim (\ln c)^{1/2} \tag{21}$$

asymptotically as $c \to \infty$. If we "plug in" the final value $c = n^{1/2}$ this would give the conjecture. We emphasize that this is not a valid argument, the limiting relation between $f_n(c)$ and $f(c)$ held only for c a constant, albeit an arbitrarily large one, not for c a function of n. We also note that the results of the next section indicate that, at least to some extent, G_c can be regarded as the random graph $G(n,p)$ with p chosen so that the two models have the same expected number of edges. If this applied to G^f and if the expected number of edges in G^f were $n^{3/2}(\ln n)^{1/2}$ then the simple argument of the next section would give that almost surely $\alpha(G^f) < k$ with $k = \Theta(n^{1/2}(\ln n)^{1/2})$ which would mean $R(3,k) > n$ or, reversing variables. $R(3,k) = \Omega(k^2(\ln k)^{-1})$. This would match the upper bound of Ajtai, Komlós and Szemerédi.
Remark. We've shown G_c has expected size $F(c)n^{3/2}/2$. N. Alon has given an intuitive justification for this. *Suppose* G_c behaved like a random graph with $p = F(c)n^{-1/2}$. By time $c + dc$ an additional $\frac{1}{2}n^{3/2}dc$ pairs are born. The probability that a pair has a common neighbor in $G(n,p)$ is $(1-p^2)^{n-2} \sim \exp[-F(c)^2]$. Thus it would be reasonable to expect $\exp[-F(c)^2]\frac{1}{2}n^{3/2}dc$ pairs to be accepted. This would give $F(c+dc) = F(c) + \exp[-F(c)^2]dc$. Taking dc infinitesmal this gives a differential equation with solution 20.

5 Ramsey $R(3,k)$

Our object here is to show 2. For intuitive guidance in view of 1 lets consider instead of G_c the usual random graph $G \sim G(n,p)$ with $p = Ln^{-1/2}$ Let $k = \epsilon n^{1/2}(\ln n)$. There are $\binom{n}{k} < n^k$ k-sets S and for each

$$\Pr[S \text{ independent}] = (1-p)^{\binom{k}{2}} \sim e^{-pk^2/2} \tag{22}$$

The expected number of independent k-sets is then less than $n^k e^{-pk^2/2} = [ne^{-pk/2}]^k$ which is $o(1)$ for L large. Our object will be to show that 22 is roughly correct for our model G_c. By "roughly correct" we will mean up to a constant factor in the exponent. Such a factor only affects the bound on $R(3,k)$ by a constant factor, and that is not our concern here.

into three rectangles and using Fubini's Theorem

$$E[Z] = 2[F(c+\Delta c) - F(c)] \cdot F(c) + [F(c+\Delta c) - F(c)]^2$$

For c fixed we do asymptotics with $\Delta c \to 0$. As f is nonincreasing the last term is at most $(f(c)\Delta c)^2 = o(\Delta c)$. By continuity (and the fundamental theorem of calculus!)

$$F(c+\Delta c) - F(c) \sim f(c)(\Delta c)$$

so that

$$E[Z] \sim 2f(c)F(c)(\Delta c)$$

Consider Z as A plus the sum over $i \geq 2$ of $i-1$ times the probability Eve has i twinbirths in X, both surviving. Even neglecting the both surviving requirement this sum is $O((\Delta c)^2)$. Thus

$$A \sim 2f(c)F(c)(\Delta c)$$

so that 16 becomes

$$\frac{f(c+\Delta c) - f(c)}{\Delta c} \sim -2f^2(c)F(c)$$

which beomes (in F) the second order differential equation

$$F''(c) = -2(F'(c))^2 F(c) \tag{18}$$

At $c = 0$ we have the initial conditions

$$F(0) = 0, f(0) = F'(0) = 1 \tag{19}$$

Fortuitously (?!) this differential equation has the precise implicit solution

$$c = \int_0^{F(c)} e^{t^2} dt \tag{20}$$

which does indeed have the property that $\lim_{c\to\infty} F(c) = \infty$. This gives 7 and therefore 1.
Remark and Conjecture. Let G^f, Z^f be the final G and its number of edges as defined in our opening paragraph. Note that while the use of independent x_e proved to be a handy analytic tool we could equally well have defined G^f as follows. Randomly order the $\binom{n}{2}$ pairs. Begin with $G = \emptyset$. Add each edge to G if it would not create a triangle. Then G^f in the final value of G.

The n^r factors of 13 asymptotically cancel so 15 giving 11. $\Box$

Now we show 6. Let $\epsilon > 0$ be arbitrarily small and let FIN be a finite family of twintrees so that the branching process yields a $T \in FIN$ with probability at least $1 - \frac{\epsilon}{2}$. (E.g., FIN could be all twintrees with at most some large number D of edges.) Now use 11 to pick n_0 so that for $n > n_0$ and each of the finite number of $T \in FIN$

$$|f_n(T,c) - f(T,c)| < \frac{\epsilon}{2|FIN|}$$

Then $f_n(c)$ is at least the probability that there is a normal relevant history with twintree $T \in FIN$ with the root surviving and that is at least $f(c) - \epsilon$. Also $1 - f_n(c)$ is at least the probability that there is a normal relevant history with twintree $T \in FIN$ with the root not surviving and that is at least $1 - f(c) - \epsilon$. As ϵ was arbitrary this yields 6.

The required uniformity over $c \in [0, C]$ for 6 is easy to check. From 10 given $\epsilon > 0$ we may pick FIN that works for every $c \in [0, C]$ simultaneously. An examination of the proof of 11 gives that the limit is approached uniformly for $c \in [0, C]$.

4 A Differential Equation

Here we find $f(c)$ as the solution to a differential equation. Consider Eve with birthdate $c + \Delta c$. For Eve to survive she must have no twins both surviving with twinbirthdate $(x, y) \in [0, c]^2$ nor twins both surviving with twinbirthdate $(x, y) \in X$ where we set $X = [0, c + \Delta c]^2 - [0, c]^2$. The Poisson nature of Eve's births make these independent events. Thus

$$f(c + \Delta c) = f(c)(1 - A) \tag{16}$$

where A is the probability Eve does have twins, both surviving, twinbirthdate $(x, y) \in X$. We first bound $0 \le A \le 2\Delta c + (\Delta c)^2$, the latter being an upper bound on the probability Eve has twins with twinbirthdates in this interval. By itself, this implies that f is continuous and nonincreasing. Then f is integrable. We define the integral

$$F(u) = \int_0^u f(t)dt \tag{17}$$

Let Z be the number of Eve's twins with twinbirths in X, both surviving. Then $E[Z]$ is simply the integral of $f(x)f(y)$ over $(x, y) \in X$. Splitting X

Let T have $2r$ edges, label them $1,\ldots,2r$. Let Γ be the set of $(x_1,\ldots,x_{2r})\in[0,c]^{2r}$ such that $x_i<x_j$ whenever edge i lies below edge j in T. Then

$$f(T,c)=\int_\Gamma e^{-c^2-y_1^2-\ldots-y_{2r}^2}dy_1\cdots dy_{2r} \tag{12}$$

Indeed, to generate T with birthdates in the infinitesmal intervals $[y_i,y_i+dy_i]$ there is probability $\prod dy_i$ of having those births, probability $\exp[-c^2]$ for Eve to have no more births and $\exp[-y_i^2]$ for the child of edge i (with birthdate y_i) to have no further children.

Compare this with $f_n(c)$. There are $(n-2)_r$ choices of vertices of G that could generate T. (The vertices of $e=\{i,j\}$ have been fixed but every birth requires a new vertex v.) Fix such a representation of T. Let edge i be represented by the pair $top(i),bot(i)$ of vertices of G and let REP be the of all $r+2$ vertices in the representation (including the vertices of e). Take $(y_1,\ldots,y_{2r})\in\Gamma$. The probability that each edge i in the representation has x_i in the infinitesmal interval $[y_i,y_i+dy_i]$ is $n^{-1/2}dy_i$. This gives

$$f_n(T,c)=(n-2)_r\int_\Gamma A(y_1,\ldots,y_{2r})n^{-r}dy_1\cdots dy_{2r} \tag{13}$$

where A is the probability, conditional on having the edges of T with birthdates y_i, that the relevant history does not contain any more edges. We require the asymptotics of A. With probability $(1-o(1)$ for each $\{u,w\}\subset REP$ that is not an edge we have $x_{u,w}>c$. Now for each $u\notin REP$ and each edge $i=\{top(i),bot(i)\}$ let $B_{u,i}$ be the "bad" event that $x_{u,v}<y_i$ for both $v=top(i)$ and $v=bot(i)$. We'll include the edge e as the case $i=0$. Note that these values (involving a new vertex u) are independent of previous conditionings. Thus

$$A\sim\Pr[\wedge_u\wedge_{i=0}^{2r}\neg B_{u,i}] \tag{14}$$

Clearly $\Pr[B_{u,i}]=y_i^2n^{-1}$ where we interpret $y_0=c$. Fix u and let i range over the $2r+1$ edges. Any two edges i,i' have $\Pr[B_{u,i}\wedge B_{u,i'}]=O(n^{-3/2})$ since even when they overlap in a vertex we are requiring three pairs to have small x-value. As r is fixed the first step of Inclusion-Exclusion gives

$$\Pr[\vee_i B_{u,i}]=(1-o(1))[\sum_i y_i^2]n^{-1}$$

for fixed u. But these events are mutually independent over $u\notin REP$ so

$$A\sim[1-\Pr[\vee_i B_{u,i}]]^{n-(r+2)}\sim e^{-c^2-y_1^2-\ldots-y_{2r}^2} \tag{15}$$

We claim T is finite with probability one. Note that if "Mary" has birthdate a and $b < a$ then the probability Mary has twinbirthdates (x, y) with x in the infinitesmal interval $[b, b + db]$ is $a \cdot db$. Let N_g be the number of children in the g-th generation. Then

$$E[N_g] = 2^g \int^* cx_1 \cdots x_{g-1} dx_1 \cdots dx_g \tag{8}$$

where $\int^*$ is over those $(x_1, \ldots, x_g)$ with $0 < x_g < \ldots < x_1 < c$. Here 2^g represents the choices of birth order and x_i is the birthdate for the i-th generation. This has the precise solution

$$E[N_g] = 2\frac{c^{2g}}{g!} \tag{9}$$

so the total number N of vertices of T has

$$E[N] = 1 + 2\sum_{g=1}^{\infty} \frac{c^{2g}}{g!} = 2e^{c^2} - 1 \tag{10}$$

The finiteness of $E[N]$ gives the claim.

On a twintree T we define bottom-up the notion of a vertex surviving or dying. A childless vertex survives. A vertex dies if and only if it has twins both of whom survive. Now we *define* $f(c)$ to be the probability that the random tree T defined above has its root survie.

3 The Relevant History

Here we show 6. Fix $e = \{i, j\}$ and $c > 0$, condition on $x_e = c$, and consider $f_n(c)$. Define the *relevant history* of e to be a set T of edges defined as follows. $e \in T$. If $\{u, l\} \in T$ and $x_{u,v}, x_{l,v} < x_{u,l}$ then $\{u, v\}, \{l, v\} \in T$. We can find T by a breadth first search, we search an edge $\{u, l\}$ already in T by checking whether any v satisfy the condition and if so adding those edges to T. We call the relevant history *normal* if every time such a v is found it is a vertex that has not yet appeared in any of the edges of T. When the relevant history is normal we give T a twintree structure, letting $\{u, v\}, \{l, v\}$ be twins of $\{u, l\}$, with $\{u, v\}$ the firstborn if and only if $u < l$. For any twintree T let $f(T, c)$ be the probability that the branching process of §2 gives T and let $f_n(T, c)$ be the probability that the relevant history of e is normal with twintree T.

Claim:

$$\lim_{n\to\infty} f_n(T, c) = f(T, c) \tag{11}$$

improving Paul Erdős's classic 1961 lower bound on $R(3,k)$.

Fix a pair $e = \{i,j\}$. We say e survives at time c if there is no $k \neq i,j$ with $\{i,k\},\{j,k\} \in G_c$. Let $f_n(c)$ be the probability that e survives at time c given $x_e = c$. This is independent of the particular e. In an infinitesmal time range c to $c+dc$ there is probability $n^{3/2}dc/2$ that some edge e is born and probability $n^{3/2}f_n(c)dc/2$ that an edge is accepted. Thus

$$E[Z_c] = \frac{n^{3/2}}{2}F_n(c) \tag{4}$$

where we define

$$F_n(c) = \int_0^c f_n(t)dt \tag{5}$$

We shall give an explicit function $f(c)$ so that

$$\lim_{n\to\infty} f_n(c) = f(c) \tag{6}$$

and further the limit is uniform in that for every $C, \epsilon > 0$ there exists n_0 so that $|f_n(c) - f(c)| < \epsilon$ for all $n > n_0$ and all $0 \le c \le C$. We'll further show, by explicit integration, that

$$\int_0^\infty f(c) = \infty \tag{7}$$

Lets show that this implies 1. Pick C so that $\int_0^C f(c)dc > L+1$. Pick n_0 so that for $n > n_0$ and $0 \le c \le C$ we have $|f_n(c) - f(c)| < C^{-1}$. Then

$$F_n(C) = \int_0^C f_n(c)dc > \int_0^C f(c)dc - 1 > L$$

2 A Branching Process

To define $f(c)$ we consider a branching process beginning with a root "Eve" with birthdate c. Eve gives birth to ordered twins, with birthdates x,y. The set of "twinbirthdates" (x,y) is given by a Poisson distribution with unit density over $[0,c]\times[0,c]$. That is, for any $0 \le x,y < c$ and dx, dy infinitesmal Eve has probability $dx \cdot dy$ of having a birth (x',y') with $x' \in [x, x+dx]$, $y' \in [y+dy]$. A child with birthdate a then has children (always twins) independently by the same process, twinbirthdates $(x,y) \in [0,a] \times [0,a]$. These children in turn may have children, and so on. Let T be the random tree so generated. We'll call T a twintree, in adddition to root, mother and daughter it contains the relation twin.

Maximal TriangleFree Graphs and Ramsey $R(3,k)$

Joel Spencer

March 1, 1995

1 Results

We describe a random dynamic algorithm that creates a graph G on a vertex set $V = \{1, \ldots, n\}$. The 2-sets $e \subset V$ are called *pairs*. To each pair e assign, independently and uniformly, a real $x_e \in [0, n^{1/2}]$. (We further assume the x_e are distinct, this occurs with probability one.) We call x_e the birthtime of e. Begin at time zero with G empty. Let time increase. When an edge e is born add it to G if and only if that does not create a triangle in G. If e is added to G we say e is accepted, otherwise rejected. Let G_c be G at time $t = c$ and G^f be the final G, at time $t = n^{1/2}$. Let Z_c, Z^f be the number of edges of G_c, G^f respectively. All these are random variables, dependent on the choices of the x_e. We will show:

- For all L there exist c, n_0 so that for $n > n_0$

$$E[Z_c] \geq L\frac{n^{3/2}}{2} \tag{1}$$

- For all $\epsilon > 0$ there exist c, n_0 so that for $n > n_0$

$$\Pr[\alpha(G_c) \geq \epsilon n^{1/2}(\ln n)] < 1 \tag{2}$$

In particular, there exists a graph $G = G_c$ which is trianglefree and has no independent set of size $\epsilon n^{1/2}(\ln n)$. That is, the Ramsey Function $R(3,k) > n$ for $k = \epsilon n^{1/2}(\ln n)$. Reversing, for all $M > 0$ if k is sufficiently large then

$$R(3,k) > M\frac{k^2}{\ln^2 k} \tag{3}$$