The Annals of Applied Probability
1995, Vol. 5, No. 4, 1217–1235

DIFFERENTIAL EQUATIONS FOR RANDOM PROCESSES AND RANDOM GRAPHS[1]

BY NICHOLAS C. WORMALD

University of Melbourne

General criteria are given to ensure that in a family of discrete random processes, given parameters exhibit convergence to the solution of a system of differential equations. As one application we consider random graph processes in which the maximum degree is bounded and show that the numbers of vertices of given degree exhibit this convergence as the total number of vertices tends to infinity. Two other applications are to random processes which generate independent sets of vertices in random r-regular graphs. In these cases, we deduce almost sure lower bounds on the size of independent sets of vertices in random r-regular graphs.

1. Introduction. It will sometimes happen that given parameters of a discrete random process are sharply concentrated at almost any given time, in the sense that the variation in the parameters is small compared with the total number of steps in the process. This paper considers processes which generate random graphs and derives conditions under which parameters of the process concentrate around the values of real variables which come from the solution of an associated system of differential equations. When this theory applies, some fundamental features of processes can be deduced merely by solving the differential equations, which can be done numerically if necessary.

A set of vertices of a graph is *independent* if no two vertices in the set are joined by an edge. The technique in this paper formed the foundation for the analysis done in [10] and [11] of the following random graph process, called a random *d-process*, apparently introduced by Paul Erdős. Begin with n isolated vertices. Repeatedly select two vertices at random and, if they both have degree at most $d-1$ and are not already joined by an edge, add an edge between them. The process ends when no more edges can be added, by which time the graph contains at most $\lfloor dn/2 \rfloor$ edges. In [10] it was shown that a random d-process almost surely ends with $\lfloor dn/2 \rfloor$ edges, as $n \to \infty$ with d fixed. A crucial part of the argument involved obtaining an approximate upper bound on the number of vertices of degree 0 at any given time throughout the process. In [11] it was necessary to have corresponding lower bounds in the case of 2-processes, in order to study the distribution of the numbers of short cycles. The approach in all these cases was to use the method of bounded differences, applied to Doob martingales constructed from parameters in a random process, to show that the required parameters almost surely behave

Received February 1993; revised August 1994.
[1]Research supported by the Australian Research Council.
AMS 1991 *subject classification*. 05C80.
Key words and phrases. Random graph, random process, random regular graph, independent set, differential equations, d-process.

in a way which determines their large-scale rate of change fairly precisely. The small-scale behavior is determined essentially by a stochastic difference equation.

In the present paper the theory is presented in a general setting in Section 2, where it is simplified by avoiding the construction of a Doob martingale to bound the probability of large deviations, using supermartingales instead. These results provide sufficient accuracy for the purposes in [10] and thus provide a simplification to the argument there. For the analysis of short cycles in random 2-processes in [11], smaller error terms were required: the results in the present paper would be inadequate as they stand. However, minor modifications would suffice and would again provide a simplification.

The first application here is in Section 3, where it is shown that the numbers of vertices of all possible degrees in a random d-process are approximately determined almost surely as the total number of vertices tends to infinity. The other applications are to greedy algorithms for finding large independent sets of vertices in random regular graphs in Sections 4 and 5. These yield improved lower bounds on the size of independent sets in random regular graphs.

An interesting innovation relating to this theory is presently under development. If the differential equations are attractive in a certain sense, then a large deviation of the parameter from the corresponding real variable will tend to be reduced over time, and the parameter can then be tracked closely until its value gets very small. This was found necessary [15] in order to measure the probability of events which are influenced by the behavior of the process near its natural end.

2. General results. The large-scale behavior of the sum of many "almost independent" variables can often be demonstrated by use of the following lemma. This concerns a problem in large deviations. It can be viewed as a result about random walks similar to the "gambler's ruin" problem as discussed by Feller [3], but with the additional complication that the variables are not necessarily independent. The proof is not given since it follows from exactly the same proof as Azuma's inequality (see [8], Lemma 4.1, or [12], Theorem 3).

LEMMA 1. *Let $Y_0, Y_1, \ldots$ be a supermartingale with respect to a sequence of σ-algebras $\{\mathcal{F}_i\}$ with $\mathcal{F}_0$ empty, and suppose $Y_0 = 0$ and $|Y_{i+1} - Y_i| \leq c$ for $i \geq 0$ always. Then for all $\alpha > 0$,*

$$\mathbf{P}(Y_i \geq \alpha c) \leq \exp(-\alpha^2/2i).$$

For this paper, all random processes are discrete time random processes. Such a process is a probability space Ω which can be conveniently denoted by $(Q_0, Q_1, \ldots)$, where each Q_i takes values in some set S. The elements of Ω are sequences $(q_0, q_1, \ldots)$, where each $q_i \in S$. We use H_t to denote $(Q_0, Q_1, \ldots, Q_t)$, the history of the process up to time t. Uppercase letters are used for the random variables corresponding to the deterministic parameters

denoted by their lowercase counterparts. For a function y defined on histories, the random variable $y(H_t)$ is denoted by Y_t for convenience.

The following theorem is general enough for the purposes in this paper, but it could be easily extended by weakening the hypotheses on the differences of the functions y_i. Its conclusions can also be sharpened in special cases such as is done in [10] and [11].

Consider a sequence Ω_n, $n = 1, 2, \ldots$, of random processes. Thus $q_t = q_t(n)$ and $S = S_n$, but for simplicity the dependence on n is usually dropped from the notation. Asymptotics, denoted by the notations o and O, are for $n \to \infty$, but uniform over all other variables. For a random variable X, we say $X = o(f(n))$ *always* if $\max\{x\colon \mathbf{P}(X = x) \neq 0\} = o(f(n))$. An event occurs *almost surely* if its probability in Ω_n is $1 - o(1)$. We denote by S_n^+ the set of all $h_t = (q_0, \ldots, q_t)$, where each $q_i \in S_n$, $t = 0, 1, \ldots$.

We say that a function $f(u_1, \ldots, u_j)$ satisfies a *Lipschitz condition* on $D \subseteq \mathbb{R}^j$ if a constant $L > 0$ exists with the property that

$$|f(u_1, \ldots, u_j) - f(v_1, \ldots, v_j)| \leq L \sum_{i=1}^{j} |u_i - v_i|$$

for all $(u_1, \ldots, u_j)$ and $(v_1, \ldots, v_j)$ in D. (For the existence of a solution to a set of differential equations, a Lipschitz condition is only needed on a subset of the variables, but here we require it on all of them.)

In the following theorems, note that "uniformly" refers to the convergence implicit in the $o(\ )$ terms. Hypothesis (i) ensures that $Y_t^{(l)}$ does not change too quickly throughout the process, (ii) tells us what we expect the rate of change to be and (iii) ensures that this rate does not change too quickly. For simplicity, we first state a version with a constant bound on the variation of $Y_t^{(l)}$. Then a generalization is proved, showing how this hypothesis can be weakened.

THEOREM 1. *Let a be fixed. For $1 \leq l \leq a$, let $y^{(l)}\colon \bigcup_n S_n^+ \to \mathbb{R}$ and $f_l\colon \mathbb{R}^{a+1} \to \mathbb{R}$, such that for some constant C and all l, $|y^{(l)}(h_t)| < Cn$ for all $h_t \in S_n^+$ for all n. Suppose also that for some function $m = m(n)$:*

(i) *there is a constant C' such that, for all $t < m$ and all l,*

$$|Y_{t+1}^{(l)} - Y_t^{(l)}| < C'$$

always;

(ii) *for all l and uniformly over all $t < m$,*

$$\mathbf{E}(Y_{t+1}^{(l)} - Y_t^{(l)} | H_t) = f_l(t/n, Y_t^{(1)}/n, \ldots, Y_t^{(a)}/n) + o(1)$$

always;

(iii) *for each l the function f_l is continuous and satisfies a Lipschitz condition on D, where D is some bounded connected open set containing the intersection of $\{(t, z^{(1)}, \ldots, z^{(a)})\colon t \geq 0\}$ with some neighborhood of $\{(0, z^{(1)}, \ldots, z^{(a)})\colon \mathbf{P}(Y_0^{(l)} = z^{(l)}n,\ 1 \leq l \leq a) \neq 0$ for some $n\}$.*

Then:

(a) *For* $(0, \hat{z}^{(1)}, \dots, \hat{z}^{(a)}) \in D$ *the system of differential equations*

$$\frac{dz_l}{ds} = f_l(s, z_1, \dots, z_a), \qquad l = 1, \dots, a,$$

has a unique solution in D *for* z_l: $\mathbb{R} \to \mathbb{R}$ *passing through*

$$z_l(0) = \hat{z}^{(l)}, \qquad 1 \le l \le a,$$

and which extends to points arbitrarily close to the boundary of D.

(b) *Almost surely*

$$Y_t^{(l)} = nz_l(t/n) + o(n)$$

uniformly for $0 \le t \le \min\{\sigma n, m\}$ *and for each* l, *where* $z_l(t)$ *is the solution in* (a) *with* $\hat{z}^{(l)} = Y_0^{(l)}/n$, *and* $\sigma = \sigma(n)$ *is the supremum of those* s *to which the solution can be extended.*

NOTES.

1. In applications, we choose m so that the hypotheses of the theorem hold. Often m is close to σn, which may represent some natural boundary.
2. Let $b(n) \to \infty$ as $n \to \infty$. It is immediate from the setting that the theorem remains valid if we replace n by $b(n)$ throughout, except where it is used as an index, namely, in $m(n)$, S_n, $w(n)$, $\lambda(n)$ and $\sigma(n)$. However, this generality is not required in the applications in this paper.
3. If the error term occurring in (ii) can be reduced, at least with high probability, then it is likely that the error $o(n)$ in (b) can be reduced accordingly to show so-called sharp concentration. For example, in the application of this method in [10] it is $o(n^\alpha)$ for a particular $\alpha < 1$. The intention of this theorem is not so much to give a widely applicable result of fullest possible strength, as to give a fairly accurate indication of what situations the method of proof can apply to, besides serving for the present purposes.
4. It will be shown in the proof that the theorem remains valid if the references to "always" in (i) and (ii) are replaced by the restriction to the event that $(t/n, Y_t^{(1)}/n, \dots, Y_t^{(a)}/n) \in D$, and in conclusion (b), the upper bound on t is replaced by the first t for which this event fails.
5. In many applications, the Lipschitz condition on f_l in (iii) prevents us from choosing a domain D which extends to the natural end of the process, which may occur at some time T_0, say. To get around this problem, we can often choose a domain D which the variables will almost surely remain inside until time $T_0 - \varepsilon n$. Then condition (i) ensures that $|Y_t^{(l)} - Y_{T_0}^{(l)}| \le C'\varepsilon n$ for $m \ge t \ge T_0 - \varepsilon n$. Taking $\varepsilon \to 0$ permits us to conclude that the equation in (b) holds for $0 \le t \le \min\{T_0, m\}$ provided $\sigma n \to T_0$ and each z_l is continuous.

In the following we can typically assign any slowly growing function such as $\log n$ to λ, and $w = n^\alpha$ for any $0 < \alpha < 2/3$.

THEOREM 2. *Theorem* 1 *holds if condition* (i) *is weakened to*:

(i′) *for some functions* $w = w(n)$ *and* $\lambda = \lambda(n)$ *with* $\lambda^4 \log n < w < n^{2/3}/\lambda$ *and* $\lambda \to \infty$ *as* $n \to \infty$, *for all* l *and uniformly for all* $t < m$,

$$\mathbf{P}\left(|Y_{t+1}^{(l)} - Y_t^{(l)}| > \frac{\sqrt{w}}{\lambda^2\sqrt{\log n}} \,\middle|\, H_t\right) = o(n^{-3})$$

always on Ω_n.

NOTE. The notes to Theorem 1 still apply [although Note 5 must be modified suitably to take account of the change to condition (i)].

PROOF. There is a unique solution in (a) by a standard result in the theory of first order differential equations. (See Hurewicz [7], Chapter 2, Theorem 11.)

To present the proof, we simplify notation by considering $l = 1$ and refer to $y^{(1)}$, z_1 and f_1 as y, z and f, and so on. The proof for general l is exactly the same. We assume at first that D contains all $(s, z) \subseteq \mathbb{R}^2$ such that $\mathbf{P}(Y_{sn} = zn) \neq 0$ for some $sn = 0, 1, \ldots, m$. Let $0 \leq t \leq m - w$. Condition (ii) gives the expected trend in the rate of change of Y_t at some stage of the process. In order to show that this trend is followed almost surely, we demonstrate concentration of

$$Y_{t+w} - Y_t.$$

Assume first that the inequality in (i′) never holds; that is,

$$|Y_{t+k+1} - Y_{t+k}| \leq \frac{\sqrt{w}}{\lambda^2\sqrt{\log n}} \tag{2.1}$$

always for all k. Then by (ii) for $0 \leq k < w$,

$$\begin{aligned}\mathbf{E}(Y_{t+k+1} - Y_{t+k}|H_{t+k}) &= f((t+k)/n, Y_{t+k}/n) + o(1)\\ &= f(t/n, Y_t/n) + o(1)\end{aligned}$$

by (iii) as $k = o(n)$ and $|Y_{t+k} - Y_t = o(n)|$. Thus there exists a function $g(n) = o(1)$ such that, conditional on H_t,

$$Y_{t+k} - Y_t - kf(t/n, Y_t/n) - kg(n)$$

is a supermartingale in k with respect to the sequence of σ-fields generated by $H_t, \ldots, H_{t+w}$. The differences in this supermartingale are, by (2.1), at most

$$\frac{\sqrt{w}}{\lambda^2\sqrt{\log n}} + f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n) \leq \frac{\sqrt{2w}}{\lambda^2\sqrt{\log n}}.$$

So by Lemma 1,

$$\begin{aligned}&\mathbf{P}\left(Y_{t+w} - Y_t - wf\left(\frac{t}{n}, \frac{Y_t}{n}\right) \geq wg(n) + \frac{2w\alpha}{\lambda^2\sqrt{\log n}} \,\middle|\, H_t\right)\\ &\qquad \leq \exp(-\alpha^2) + o(n^{-1})\end{aligned} \tag{2.2}$$

for all $\alpha > 0$. To avoid assuming (2.1) we can condition on the event that it holds for all relevant k. By (i') this event has probability $1-o(n^{-2})$. Moreover, in the conditional space, (ii) still holds as $|Y| \le Cn$, so (2.2) follows without assuming (2.1).

Exactly the same argument, but using a submartingale, bounds the lower tail of $Y_{t+w} - Y_t - wf(t/n, Y_t/n)$. Hence with $\alpha = (\log n)^{1/2}\lambda$, we have

$$\mathbf{P}(|Y_{t+w} - Y_t - wf(t/n, Y_t/n)| \ge w(g(n) + 1/\lambda)|H_t) = o(n^{-1}). \tag{2.3}$$

Now define $k_i = iw$, $i = 0, 1, \dots, i_0$, where $i_0 = \min\{\lfloor m/w \rfloor, \lfloor \sigma n/w \rfloor\}$. For some function $\lambda_1 = \lambda_1(n) \to \infty$ as $n \to \infty$, we show by induction that, for each such i,

$$\mathbf{P}(|Y_{k_i} - z(k_i/n)n| \ge B_i) = o(i/n), \tag{2.4}$$

where $B_i = (w/\lambda_1 + Bw^2/n)((1 + Bw/n)^i - 1)n/Bw$ for some $B > 0$.

The induction begins by the fact that $z(0) = Y_0/n$. Write

$$A_1 = Y_{k_i} - z(k_i/n)n,$$
$$A_2 = Y_{k_{i+1}} - Y_{k_i},$$
$$A_3 = z(k_i/n)n - z(k_{i+1}/n)n.$$

The inductive hypothesis (2.4) gives that $|A_1| < B_i$ with probability $1-o(i/n)$. When this inequality holds, by (2.3) we can choose λ_1 so that

$$\left|A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| < \frac{w}{\lambda_1}$$

with probability $1 - o(n^{-1})$. Since z is the solution given in (a) and f satisfies the Lipschitz condition in (iii), we also have

$$|A_3 + wz'(k_i/n)| = O(w^2/n)$$

and so

$$\left|A_3 - \left(-wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right)\right| < \frac{Bw^2 + BwB_i}{n}$$

for n sufficiently large. (This is where B is determined.) Hence

$$|Y_{k_{i+1}} - z(k_{i+1}/n)n| = |A_1 + A_2 + A_3| < B_{i+1}$$

with probability $1-o((i+1)/n)$, and so we have (2.4) by induction. This shows that the equation in (b) is satisfied almost surely at $t = k_i$ for each i up to i_0. Since $k_{i+1} - k_i = w$ and since by (i) the variation in y when t changes by at most w is $o(n)$ with high probability, we have (b).

Finally we note the required modification of the proof if the assumption about D at the start of this proof is false. For $\varepsilon > 0$, define $D' = D'(\varepsilon)$ to be the set of points (s, z) in D of distance at least ε in the z direction from the boundary of D. Now work with D' in place of D (with the corresponding redefinition of σ). Thus we may take $(t/n, Y_t/n) \in D'$ during the inductive step. The assumption that (2.1) holds for all relevant k implies that $((t+k)/n, Y_{t+k}/n)$

does not leave D and so the analysis assuming (2.1) still stands. The probability that (2.1) fails at least once is the same as before (since, provided it has not failed, we are still in D) and so we again have (2.3). The proof goes through as before, since by taking ε' arbitrarily small we can approach arbitrarily close to σn because D is open. This also shows why the claim in Note 4 is valid. □

3. Random graph processes with restricted degrees. Here we show that the numbers of vertices of all possible degrees in a random d-process are approximately determined almost surely, as the total number of vertices increases.

Let Q_t denote the tth edge added in a random d-process, as defined in the Introduction, and let G_t denote the graph induced by the edges $Q_1, \dots, Q_t$. If the process has already stopped at time $t-1$, define $G_t = G_{t-1}$ for convenience. For $0 \le i \le d$ let $Y_t^{(i)}$ denote the number of vertices of degree i in G_t. Then

$$|Y_{t+1}^{(i)} - Y_t^{(i)}| \le 2 \tag{3.1}$$

always. This establishes Theorem 1(i). Furthermore the number of sites available for Q_t is

$$A_t = \binom{n - Y_t^{(d)}}{2} - F_t,$$

where $n - Y_t^{(d)}$ is the number of vertices of degree less than d and F_t denotes the number of edges already present between these vertices. Note that $Y^{(d)} \le 2t/d$ and $F_t \le dn/2$. Thus,

$$A_t = \tfrac{1}{2}(n - Y_t^{(d)})^2 + O(n). \tag{3.2}$$

Each of the A_t available sites is equally likely to be used. So, given G_t and $i < d$, the expected number of vertices of degree i which are changed to degree $i+1$ by the addition of the edge Q_{t+1} is

$$\frac{Y_t^{(i)}(n - Y_t^{(d)} - 1) + O(n)}{A_t}, \tag{3.3}$$

provided $A_t \neq 0$. Here the correction term $O(n)$ is due to those sites already occupied by edges. If we now choose $m = d(n - n^\alpha)/2$, where $\alpha > 1/2$, then $n - Y_t^{(d)} > n^\alpha$ and so, from (3.2), $n = o(A_t)$ and $A_t \sim (n - Y_t^{(d)})^2/2$. As (3.3) contributes negatively to the expected increase of $Y^{(i)}$ for $i < d$ and positively to that of $Y^{(i+1)}$, we now obtain

$$\mathbf{E}(Y_{t+1}^{(i)} - Y_t^{(i)} | H_t) = f_i(t/n, Y_t^{(1)}/n, \dots, Y_t^{(d)}/n) + o(1),$$

where

$$f_i(s, z_0, \dots, z_d) = \frac{2\delta_{i\neq 0} z_{i-1} - 2\delta_{i\neq d} z_i}{1 - z_d}, \tag{3.4}$$

where $\delta_{i\neq j}$ is 0 for $i = j$ and 1 otherwise. This establishes Theorem 1(ii) for these functions f_i. For $\varepsilon > 0$, Theorem 1(iii) is valid for the domain D defined

by $-\varepsilon < s < 1$, $-\varepsilon < z_i < 1+\varepsilon$ for all i and $z_d < 1-\varepsilon$. Thus Theorem 1 applies. Since $Y_t^{(d)} \le 2t/d$, the solution does not leave D until after $t = (1-\varepsilon)\,dn/2$. By (3.1) the value of $Y^{(i)}$ cannot change by more than $\varepsilon\,dn$ after this time. Thus by taking ε arbitrarily small, we obtain the following from Theorem 1(b).

THEOREM 3. *Almost surely*

$$Y_t^{(i)} = z_i(t/n)n + o(n)$$

uniformly for $0 \le t \le nd/2$ *and for each* i, *where the* $z_i(t)$ *form the solution to*

$$\frac{dz_i}{ds} = f_i(s, z_0, \ldots, z_t),$$

f_i *given in* (3.4), *with initial conditions* $z_i(0) = 0$ *for* $i > 0$ *and* $z_0 = 1$.

Theorem 3 can be used to determine various quantities asymptotically almost surely, for instance, the maximum number of vertices of degree i occurring during the course of the process.

4. Greedy algorithm for independent vertex sets in random regular graphs. We use the following standard model $\mathscr{G}(r,n)$ for random r-regular graphs on n vertices (where rn is restricted to the even integers), with the uniform probability distribution. Take rn points in n buckets labelled $1, 2, \ldots, n$, with r in each bucket, and choose a random *pairing* $P = p_1, \ldots, p_{rn/2}$ of the points such that $|p_i| = 2$ for all i, each point is in precisely one pair p_i, no pair contains two points in the same bucket and no two pairs contain four points from just two buckets. To get a random graph in $\mathscr{G}(r,n)$, join two distinct vertices i and j if some pair has a point in bucket i and one in bucket j. The conditions on the pairing prevent the formation of loops and multiple edges. Asymptotics in this section will be for $n \to \infty$ with $r \ge 3$ fixed.

The *independence ratio* $\mathscr{I}(G)$ of a graph G with n vertices is the maximum c for which G has an independent set of vertices of cardinality cn. For $G \in \mathscr{G}(r,n)$ let

$$\beta(r) = \sup\{c\colon \mathbf{P}(\mathscr{I}(G) \ge c) = 1 - o(1)\}.$$

As Bollobás ([2], Chapter XI, Corollary 28) shows,

$$\beta(r) \ge \beta_1(r) = \begin{cases} \dfrac{7}{18}, & r = 3, \\ \dfrac{r\log r - r + 1}{(r-1)^2}, & r \ge 4 \end{cases}$$

(see Table 1). These numbers come from lower bounds on the independence ratios of graphs with maximum degree r and large girth, due to Hopkins and Staton [6] and Shearer [13]. The bounds carry over to random regular graphs since the number of cycles of length less than C is almost surely $O(\log n)$ for all C by the results in [16] or Bollobás [1], so such cycles can be destroyed by

TABLE 1
Lower bounds on $\beta(r)$ and, in the last column, upper bounds on $\gamma(r)$

r	$\beta_1(r)$	$\beta_2(r)$	Theorem 4	Theorem 5	$\gamma_1(r)$
3	0.3889	0.4139	0.3750	0.4328	0.4554
4	0.2828	0.3510	0.3333	0.3901	0.4163
5	0.2529	0.3085	0.3016	0.3566	0.3844
6	0.2300	0.2771	0.2764	0.3296	0.3580
7	0.2117	0.2528	0.2558	0.3071	0.3357
8	0.1966	0.2332	0.2386	0.2880	0.3165
9	0.1840	0.2169	0.2240	0.2716	0.2999
10	0.1732	0.2032	0.2113	0.2573	0.2852
11	0.1638	0.1914	0.2003	0.2447	
12	0.1555	0.1811	0.1905	0.2335	
13	0.1482	0.1721	0.1818	0.2234	
14	0.1417	0.1641	0.1739	0.2143	
15	0.1358	0.1569	0.1668	0.2061	
16	0.1305	0.1504	0.1604	0.1985	
17	0.1256	0.1445	0.1545	0.1916	
18	0.1212	0.1391	0.1491	0.1852	
19	0.1171	0.1342	0.1441	0.1793	
20	0.1133	0.1297	0.1395	0.1738	0.1973
50	0.0611	0.0682	0.0748	0.0951	0.1108
100	0.0369	0.0406	0.0447	0.0572	0.0679

deleting a small number of edges. Shearer ([14], Theorem 4) improved these results on graphs with bounded degree and large girth by an iterative formula for a function depending on the degrees of the vertices. For r-regular graphs we call this function $\beta_2(r)$, which is shown in Table 1. We have similarly that $\beta(r) \geq \beta_2(r)$. McKay [9] also claimed without proof the lower bound $\beta(3) \geq \sqrt{2} - 1 = 0.4142\ldots$. These results are improved in Theorem 4 for all but a few small values of r, by an analysis of a greedy algorithm for finding independent sets of vertices. Then in the next section a more sophisticated algorithm is used in Theorem 5 to improve all these results for small r; the method almost certainly gives sharper results for all fixed $r \geq 3$. These data are included in Table 1. On the other hand, Frieze and Luczak [4] proved results for r fixed but arbitrarily large, which say nothing about small r.

Very recently, and independently of the work of the present author, Frieze and Suen [5] obtained $\beta(3) \geq 6\log(3/2) - 2 = 0.43279\ldots$ by analyzing the same algorithm as Theorem 5. Presumably the system of two simultaneous differential equations defining x_1 in Theorem 5 can be solved explicitly in the case $r = 3$ to yield this constant precisely.

Bollobás [2] gave corresponding upper bounds on

$$\gamma(r) = \inf\{c\colon \mathbf{P}(\mathscr{I}(G) \leq c) = 1 - o(1)\}.$$

These were subsequently improved by McKay [9], who gave an upper bound $\gamma_1(r)$ on $\gamma(r)$ with a rather complicated definition (see the last column in

Table 1; the missing values were not given in [9]). Note that the lower bounds in Theorem 5 are almost without exception closer to the upper bounds $\gamma_1(r)$ than to the lower bounds $\beta_2(r)$ or that given in Theorem 4.

The buckets will be called vertices. The pairing referred to must be chosen uniformly at random subject to the constraints given. This can be done by repeatedly choosing an unpaired point (using any rule whatever) and then choosing a partner for this point, to create a new pair. As long as the partner is chosen uniformly at random from the remaining unpaired points, and as long as the process is restarted if a loop or multiple edge is created, the result is a random pairing of the required type. As shown in [17], for instance, the probability of the process having to be restarted is asymptotically a constant. Thus, this possibility can be ignored when proving that statements about the pairing hold with probability $1 - o(1)$, and we henceforth permit the graph formed to contain loops and multiple edges by leaving the pairing unrestricted.

Along with an algorithm for finding an independent set in $G \in \mathscr{G}(r,n)$, we will consider an associated algorithm (called a *generation* algorithm) for simultaneously generating the random graph G while running the algorithm which finds the independent set. At any stage the *degree* of a vertex is the number of points in it which are already in pairs. The rule used for choosing the next unpaired point will in this paper be restricted in the following way. Let u denote the vertex containing the last point selected as unpaired. The next unpaired point is selected randomly from the remaining unpaired points in some vertex v, where $v = u$ if the degree of u is strictly less than r, and otherwise we are free to choose v to be any vertex with degree less than r. We denote the ith vertex chosen as u in this way by v_i. We also leave open the possibility of choosing a vertex of degree r as v_i, in which case of course no unpaired point is selected at this time, but we move on to choosing v_{i+1} immediately. We can postpone the choice of v_i until all the points in v_{i-1} have been paired, and thus the rule for choosing v_i can depend on the shape of G as determined so far at this time.

We first consider the following *greedy algorithm* for constructing an independent set of vertices in $G \in \mathscr{G}(r,n)$. Choose the vertices in the set one at a time, each vertex chosen randomly from those not adjacent to any already in the set. This is equivalent to the following generation algorithm for generating the random graph while finding the independent set I. Initially, put $I = \varnothing$. In generating the random graph, choose the vertices $v_1, v_2, \ldots$ as before using the rule that the next vertex v_i is randomly chosen from those currently of degree 0. At this time, add this vertex v_i into I. When no vertices of degree 0 remain, the current set I is outputted, and the rest of the graph generation can go ahead by selecting v_i in any manner.

THEOREM 4. *For $r \geq 3$, the greedy algorithm applied to a random r-regular graph almost surely produces an independent set of vertices of cardinality*

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

PROOF. Define $q_0 = \varnothing$, for $t > 0$, define q_t to be the set of edges incident with the tth vertex chosen in the independent set selected by the greedy algorithm and define g_t to be the graph with vertices $1, 2, \ldots, n$ and edge set $\bigcup_{i=1}^{t} q_i$. For completeness, define $q_t = \varnothing$ for all $t > |I|$. Let $y(g)$ denote the number of vertices of degree 0 in any graph g. It will become evident that the choice of m is immaterial provided it is at least, say, $(\frac{1}{2} - \varepsilon)n$.

We apply Theorem 1 with $l = 1$ and $Y^{(1)} = Y$ and so forth. Part (i) is immediate as $|Y_{t+1} - Y_t| \leq r + 1$ always. For (ii), define

$$f(s, z) = \frac{-rz}{1 - 2s} - 1$$

and

$$D = \{(s, z): -\varepsilon < s < 1/2, \varepsilon < z < 1 - 2s\}$$

for some small $\varepsilon > 0$. We can assume by Note 4 to Theorem 1 that the parameters of the random process are located in D at time t. Thus, we may assume that $Y_t \geq \varepsilon n$. At time t, rt pairs have been chosen and so $rn - 2rt$ points remain. Of these, rY_t are in vertices of degree 0. So of the r partners of the points in the next vertex chosen, the expected number in vertices of degree 0 is $rY_t/(n - 2t) + O(n^{-1})$. The next vertex chosen also reduces y by 1, and so this yields Theorem 1(ii), where the negative sign arises because these numbers represent reductions in the number of vertices of degree 0. Theorem 1(iii) is immediate. Thus the hypotheses of Theorem 1 are satisfied. Note that

$$\frac{dz}{ds} = f(s, z), \qquad z(0) = 1,$$

has the solution

$$z = \frac{(r - 1)(1 - 2s)^{r/2} - (1 - 2s)}{r - 2}.$$

This meets the boundary of D when $z = \varepsilon$, which for small enough ε is arbitrarily close to

$$s = \frac{1}{2} - \frac{1}{2}\left(\frac{1}{r - 1}\right)^{2/(r-2)}.$$

(The "other" boundary $z = 1 - 2s$ is physically impossible to meet since we always have $y \leq n/2 - t$, corresponding to $z \leq \frac{1}{2} - s$.) The theorem follows, since if $y_t < \varepsilon n$, then $y_{t+\varepsilon n} = 0$. □

Theorem 4 apparently improves the result in [2] for $r \geq 3$ and that in [14] for $r \geq 7$ (see Table 1). This claim is not established rigorously since Shearer's result [14] is defined iteratively and hence is difficult to analyze, and in any case all these results are seemingly improved in the next section.

5. Better algorithm for independent vertex sets in random regular graphs. For a further improvement in the lower bound on the independence ratio of random regular graphs, consider the following *neighborly* algorithm. Choose vertices in an independent set I one by one, with the condition that the next vertex is chosen randomly from those with the maximum number of neighbors adjacent to vertices already in I. Equivalently, modify the generation algorithm in the previous section along the following lines. If $v_{i-1} \in I$, choose v_i, v_{i+1} and so on from the vertices neighboring v_{i-1}. Then, when all such neighbors have been used, choose next any vertex of those currently with the largest degree and place it into I. Thus there are two possibilities when selecting the next vertex v_i:

CASE 1. Some neighbor of an element of I has degree less than r. In this case, choose such a vertex at random for v_i. Such vertices v_i are *not* placed into I.

CASE 2. All neighbors of elements of I have degree r. Then v_i is randomly chosen from the vertices of largest degree of those not neighbors of elements of I and is included in I.

When I contains a neighbor of every vertex not in I, it is outputted as an independent set.

Suppose that v_i has just been selected in Case 2. Let d denote the number of pairs selected so far and let R denote the number of points remaining, so that

$$R = rn - 2d.$$

Let $y^{(j)} = y^{(j)}(g)$ denote the number of vertices of degree j in the current graph g, $j = 0, 1, \ldots, r-1$, so that

$$R = ry^{(0)} + (r-1)y^{(1)} + \cdots + y^{(r-1)}.$$

For each point p being selected at random to be paired with a given point in v_i, the probability that p is in a vertex v of degree k is $P = (r-k)y^{(k)}/R$. Then when v is subsequently chosen under Case 1, all $r-1-k$ unpaired points in v will get paired. So each unpaired point in v_i contributes $(r-j)y^{(j)}/R$ to the expected decrease in $y^{(j)}$, due to the possibility of being paired with a vertex of degree j, as well as contributing approximately

$$\sum_{k=0}^{r-2} \frac{(r-k)y^{(k)}}{R}\left(\frac{(r-1-k)(r-j)y^{(j)}}{R}\right)$$

both to the expected decrease in $y^{(j)}$ and the expected increase in $y^{(j+1)}$, for each j, $0 \le j \le r-1$. "Approximately" becomes "asymptotically" if we assume (for the moment) that, for some $\alpha > 0$, $y^{(j)} \ge n^{\alpha}$ for each j, and also $R \ge n^{2\alpha}$.

The effects of these points are all asymptotically equal (as all quantities have values at least n^α and only change by constant amounts while these points are being processed). In addition, if v_i has degree l, then $y^{(l)}$ decreases by 1.

All of this will lead to a condition like Theorem 1(ii). To be precise, if v_i has degree l in the current graph g, then for $0 \le j \le r-1$, the expected increase in $y^{(j)}(g)$ during the selection of v_i and its neighbors is

$$-\delta_{lj} + (r-l)\mu_j + o(1), \tag{5.1}$$

where δ is the Kronecker delta and $\mu_j = \mu_j(g)$ is defined by

$$\mu_j = -\frac{(r-j)y^{(j)}}{R} + \sum_{k=0}^{r-2} \frac{(r-1-k)(r-k)((r+1-j)y^{(j-1)}y^{(k)} - (r-j)y^{(j)}y^{(k)})}{R^2}.$$

If any of the $y^{(j)}$ are less than n^α, the error can be absorbed into the $o(1)$ term in (5.1). So it is valid for all j provided $R > n^{2\alpha}$.

For r in general, the neighborly algorithm will be expected to proceed through several stages. At first, the number of vertices of degree greater than 1 remains extremely small, $\mu_2 < 1$, and so v_i is almost surely of degree 1 or 2. Any vertices of degree 2 created momentarily are chosen continuously until none remains. At some time $(r-2)\mu_2$ reaches 1, stage 1 ends and stage 2 begins: The newly generated vertices of degree 2 regenerate themselves as fast as they are consumed. It seems plausible that later $(r-2)\mu_2$ will again dip below 1, so that the process reverts to a phase like stage 1, or $(r-3)\mu_3$ exceeds 1 and vertices of degree 4 begin appearing in quantity, when we may say that stage 3 has been reached.

For any fixed r, the results in Section 2 are applicable and the size of the independent set of vertices produced can be determined approximately, with probability close to 1. The results are applied separately to each stage to show that all the $y^{(i)}$ are determined approximately. The main complicating factor is the uncertainty of the degree of v_i at the transition between two stages. As it turns out, the results obtained here support the hypothesis that the process almost surely proceeds consecutively through stages 1 to $r-2$, without any significant reversions to previous stages, at the end of which R is approximately 0, and the algorithm terminates soon thereafter.

For the following theorem put

$$\begin{aligned} x_0 &= 0, \\ \hat{z}_{0,0} &= 1, \\ \hat{z}_{0,i} &= 0, \qquad i = 1, \ldots, r-2. \end{aligned}$$

For $j = 1, \dots, r-2$, x_j is determined recursively as follows. First define functions $z_i(x)$, $i = 0, \dots, r-2$, for $x \geq x_{j-1}$, by the system of differential equations

$$\begin{aligned} \frac{dz_i}{dx} &= f_i(x, z_0, \dots, z_{r-2}) = \frac{\delta_{ji}((r-j-1)\tau_{j+1} - 1) + (r-j)\tau_i}{1 + \tau_{j+1}}, \\ z_i(x_{j-1}) &= \hat{z}_{j-1,i}, \qquad i = 0, \dots, j, \\ z_i(x) &\equiv 0, \qquad i > j, \end{aligned} \tag{5.2}$$

where

$$\begin{aligned} \tau_j &= -\frac{(r-j)z_j}{\xi} \\ &\quad + \sum_{k=0}^{r-2} \frac{(r-1-k)(r-k)((r+1-j)z_{j-1}z_k - (r-j)z_jz_k)}{\xi^2}, \\ \xi &= rz_0 + (r-1)z_1 + \cdots + 2z_{r-2} \end{aligned}$$

and z_{-1} is identically 0. Next, if $j \leq r-3$, define x_j to be the smallest nonnegative solution of $(r-j-1)\tau_{j+1} = 1$ and define $\hat{z}_{j,i} = z_i(x_j)$ for $i = 0, \dots, r-2$. Also define x_{r-2} to be the smallest value of $x > x_{r-3}$ for which $\xi = 0$ or $\tau_{r-1} = 1$. (From numerical results for $r \leq 20$, the former would seem to be the case for all $r \geq 3$.)

The conditions in the following theorem are designed to enable numerical verification for fixed r.

THEOREM 5. *Let $r \geq 3$ be fixed. Suppose that*:

(i) *For $j = 1, \dots, r-2$, x_j exists, z_j is strictly positive on the open interval (x_{j-1}, x_j) and*

$$\lim_{x \to x_{j-1}^+} \frac{d\tau_j}{dx} > 0.$$

(ii) *For $j = 1, \dots, r-3$, $z_j(x_j) \neq 0$ and*

$$\lim_{x \to x_j^-} \frac{d\tau_{j+1}}{dx} > 0.$$

Then the cardinality of the independent set of vertices produced by the neighborly algorithm applied to a random r-regular graph is almost surely at least

$$x_{r-2}n + o(n).$$

PROOF. We now formally define stage j of the algorithm inductively for $j = 1, \dots, r-2$ as the period from the end of stage $j-1$ (or at the start of the algorithm, in the case that $j = 1$) until the first time that either $|I| = \lfloor x_j n \rfloor$ or the algorithm has terminated. It will be shown by induction on j that almost surely for $j = 1, \dots, r-2$, when the algorithm begins stage j we have $|I| = x_{j-1}n + o(n)$, $Y^{(i)} = \hat{z}_{j-1,i}n + o(n)$ for $i = 0, \dots, j$, and $Y^{(i)} = O(1)$ for

$i = j+1, \ldots, r-2$. Note that $x_{j-1} < x_j$, since $z_j = z_{j+1} = 0$ and hence $\tau_{j+1} = 0$ at x_{j-1}.

We first examine the dynamics of the algorithm at a point when the variables have values typical for a point early (but not too early) in stage j. Choose $\varepsilon > 0$ arbitrarily small and assume at the point of choosing v_k in Case 2 we have that $Y^{(j)} > \varepsilon n$, $(r-j-1)\mu_{j+1} < 1-\varepsilon$ and $Y^{(j+1)} = c$, where $c \geq 1$ is bounded, that there are no vertices of degree greater than $j+1$ and, moreover, to simplify the discussion, that no such vertices are created. Then one vertex of degree $j+1$ will be chosen for v_k, and by (5.1) the expected value of $Y^{(j+1)}$ at the time of selecting the next vertex in Case 2, conditional on past history, is $(r-j-1)\mu_{j+1}+o(1)$. If at this time $Y^{(j+1)} > 0$, then a vertex of degree $j+1$ is chosen again. Thus, $Y^{(j+1)}$ is determined by a discrete branching process in which the expected number of births (new vertices of degree $j+1$) due to each individual, conditional on past history, is $(r-j-1)\mu_{j+1} + o(1)$, where μ_{j+1} is evaluated at the time of the individual giving birth. At this time the individual itself promptly dies.

Let Z denote the number of vertices in Case 2 of degree $j+1$, beginning with v_k, before the next vertex in Case 2 of degree j, restricted to the next n^{ε} vertices. Then Z is equal to c plus the total number of births throughout the branching process. Let t_i be the time that the ith vertex is chosen during this period. Throughout this period the value of μ_{j+1} can only change by a factor of $(1+O(n^{-1}))$ in one step and so $(r-j-1)\mu_{j+1} < 1-\varepsilon+o(1)$. Hence for large n the random variable $Y_{t_i}^{(j+1)} - c + i\varepsilon/2$ is a supermartingale in i with respect to the history until $Y_{t_i}^{(j+1)}$ becomes 0. Thus from Lemma 1,

$$\mathbf{P}(Z > C\log n) = o(n^{-3}) \tag{5.3}$$

for C a sufficiently large constant.

In the branching process, call the c original vertices of degree $j+1$ "first generation," their children "second generation" and so on. Then the expected number of second generation individuals is $c((r-j-1)\mu_{j+1}+o(1))$ and, in general, of kth generation is $c((r-j-1)\mu_{j+1}+o(1))^{k-1}$, where, since μ_{j+1} only changes slowly as mentioned above, we can take for its value that at the point of choosing v_k. It follows that

$$\mathbf{E}(Z) = c/(1-(r-j-1)\mu_{j+1}) + o(1).$$

In this discussion the creation of vertices of degree greater than $j+1$ in Case 2 has been ignored, but this possibility is so unlikely and has such a minor effect that it clearly does not affect the assertion.

Now assume that at some point

$$Y^{(j)} > \varepsilon n, \qquad (r-j-1)\mu_{j+1} < 1-\varepsilon, \qquad Y^{(i)} = 0 \quad \text{for all } i > j. \tag{5.4}$$

We conclude by arguing as for the derivation of the above equation for $\mathbf{E}(Z)$ and using (5.1) that, after processing a vertex of degree j in Case 2, the expected number of vertices of degree $j+1$ in Case 2 before the next one of

degree j is

$$\frac{(r-j)\mu_{j+1}}{1-(r-j-1)\mu_{j+1}} + o(1),$$

where the slowly changing value of μ_{j+1} can be taken at the beginning of this sequence. Hence, again using (5.1), the expected increase in $Y^{(i)}$ from adding one vertex of degree j to I up until just before adding the next of degree j is

$$\begin{aligned} &-\delta_{ji} + (r-j)\mu_i + \frac{(r-j)\mu_{j+1}(r-j-1)\mu_i}{1-(r-j-1)\mu_{j+1}} + o(1) \\ &\quad = -\delta_{ji} + \frac{(r-j)\mu_i}{1-(r-j-1)\mu_{j+1}} + o(1). \end{aligned} \tag{5.5}$$

The rest of the proof involves:

(a) showing that the assumptions (5.4) for $j = 1$ become true at some point early in stage 1;

(b) establishing a similar statement for $1 < j < r-2$ in stage j using the inductive hypothesis;

(c) applying Theorem 2 throughout the bulk of stage j for each j;

(d) examining the very end of stage j to validate the inductive hypothesis.

(a) For the first part, note that $(r-1)\mu_1 - 1 = (r-1)^2 - 1 \geq 3$ initially. This value can change only a little for the first $\varepsilon' n$ vertices added to I, provided $\varepsilon' > 0$ is arbitrarily small. Hence for large n the random variable $2i - Y^{(1)}$, evaluated when the ith vertex is added during this period, determines a supermartingale. Thus at the end of this time, $Y^{(1)} > \varepsilon n$ for some $\varepsilon > 0$ almost surely. Also, similar to the derivation of (5.3) we have for some constant C that $Y^{(2)} < C \log n$ and $Y^{(i)} < C$ for all $i > 1$ with probability $1 - o(n^{-3})$ (noting $\mu_i < \varepsilon$ here for $i > 1$). With probability $1 - o(1)$ it follows that at some point within $C' \log n$ steps thereafter, we have the assumptions (5.4) for $j = 1$.

(b) Second, the early section of stage j is handled by much the same argument. Quantities need only to be computed approximately here. Let $0 \leq i \leq r-2$. Per vertex in Case 2 of degree j, the expected increase in $Y^{(i)}$ is given by (5.1) with j replaced by i and l by j. During the first steps in stage j, this is a.s. close to $-\delta_{ji} + (r-j)\tau_i$, since by the inductive hypothesis, the value of μ_i at the start of stage j is a.s. equal to $\tau_i + o(1)$ at x_{j-1}. This value in turn is approximately equal to $f_i(s, z_0, \ldots, z_j)$ since τ_{j+1} is very close to 0. This again is equal to dz_i/dx evaluated close to x_{j-1}. By the argument in (a), this implies that the approximate average change in μ_j per vertex in Case 2 of degree j is almost surely close to $d\tau_j/dx$ evaluated at $x = x_{j-1}^+$. This is strictly positive by hypothesis (i). A similar argument applies to vertices of degree $j-1$ added to I in Case 2: again μ_j is a.s. expected to increase by (ii). To be precise, we conclude that, for ε' sufficiently small, after the first $\varepsilon' n$ steps in stage j, the increase in μ_j is almost surely at least c for some $c = c(\varepsilon') > 0$. In this discussion, vertices in Case 2 of degree greater than j have been ignored; they do not affect the result since they are relatively rare as $\mu_i \approx 0$ for $i > j$. Again

the argument in (a) allows us to conclude that at some point in the first $\varepsilon' n$ steps of stage j, the assumptions (5.4) hold for ε sufficiently small.

(c) Third, to apply Theorem 2, consider any time when the assumptions (5.4) hold, and denote by t_{i-1} the time that the ith vertex of degree j in Case 2 after this time is selected. Define a random process $(Q_0, Q_1, \ldots)$ by letting Q_0 contain all edges already present at time t_0, and in general letting Q_k be the set of edges added by the algorithm after t_{k-1} up to time t_k (where $Q_k = \varnothing$ if t_k is undefined). Define G_k to be the graph with edge set $\bigcup_{i=0}^{k} Q_i$, and let $Y_k^{(i)} = Y^{(i)}(G_k)$ for $i = 0, \ldots, r-2$. Let Y_k denote the cardinality of I at time t_k.

We apply Theorem 2 to the variables $Y^{(0)}, \ldots, Y^{(j)}$ and Y. Define D to be the set of all $(s, z_0, \ldots, z_j, \rho)$ for which $(r-j-1)\tau_{j+1} < 1-\varepsilon$, $\xi > \varepsilon$, $-\varepsilon < s < 1$, $-\varepsilon < z_i < 1$ for all i, $-\varepsilon < \rho < 1$ and $z_j > \varepsilon$. Put $w = n^{1/4}$, $m = n$ and $\lambda = \log n$.

First observe that $Y^{(i)}$ and Y are bounded above by Cn. For part (i′) of Theorem 2 and (ii) of Theorem 1, by the note after Theorem 2 we can assume that $(t/n, Y^{(0)}/n, \ldots, Y^{(j)}/n, Y/n)$ lies inside D. Here $(r-j-1)\mu_{j+1} < 1-\varepsilon$. As in the argument leading to (5.3), the creation of vertices of degree at least $j+2$ has negligible effect, and (5.3) holds. Thus the number of vertices added to G_k to get G_{k+1} is at most $C \log n$ with probability $1 - o(n^{-3})$, so Theorem 2 (i′) is satisfied. Similarly, we have Theorem 1(ii) for the variables $Y^{(i)}$, where the appropriate f_i is determined from (5.5) (this happens to be independent of s):

$$f_i(s, z_0, \ldots, z_j) = -\delta_{ji} + \frac{(r-j)\tau_i}{1-(r-j-1)\tau_{j+1}}. \tag{5.6}$$

Here and in the following equation for f, any references to z_i in τ for $i > j$ should be taken as 0. Between time t_{k-1} and t_k the expected increase in $|I|$ is given by the expected number of vertices of degree $j+1$, from (5.3), together with the one of degree j. Thus

$$\mathbf{E}(Y_{t+1} - Y_t | H_t) = f(t/n, Y_t^{(0)}/n, \ldots, Y_t^{(j)}/n) + o(1),$$

where

$$f(s, z_0, \ldots, z_j) = 1 + \frac{(r-j)\mu_{j+1}}{1-(r-j-1)\mu_{j+1}} = \frac{1+\mu_{j+1}}{1-(r-j-1)\mu_{j+1}}. \tag{5.7}$$

Hence Theorem 1(ii) holds. Theorem 1(iii) is immediate, and we conclude that Theorem 2 applies. The derivatives in the differential equations in Theorem 1(a) are given for dz_i/ds, $i = 0, \ldots, j$, by (5.6) and for dx/ds, where x is the real variable corresponding to Y, by (5.7). Replacing dz_i/ds by $(dz_i/dx)(dx/ds)$ allows us to eliminate s from this system and gives the system defined in (5.2).

We conclude from Theorem 1(b) that for all i until the solution leaves D, $Y_t^{(i)} = Z_i(t)n + o(n)$, where Z_i is determined by the solution to the differential

equations in (5.2) with initial conditions $Z_i(0) = Y_0^{(i)}/n$. By choosing ε and ε' sufficiently small, these initial values can be made to approach arbitrarily close to the values at the beginning of stage j, which by the inductive hypothesis are arbitrarily close to $\hat{z}_{j-1,i}$. For $i < j$, since the only negative terms in the defining relation of τ_i are due to z_i, it is clear that z_i stays positive in the solution in D. By hypothesis (i), z_j also stays positive in D. Hence the solution to (5.2) leaves D when $(r-j-1)\tau_{j+1} = 1-\varepsilon$. By taking ε arbitrarily small, this can be made to be arbitrarily close to $x = x_j$. Thus, in view of the stability of the differential equations implied by the Lipschitz condition in D, the value of Y_i/n when there are εn steps remaining in stage j is almost surely arbitrarily close to $\hat{z}_{j,i}$ for arbitrarily small ε. Similarly, $|I|/n$ is arbitrarily close to x_j.

(d) The final task to complete the induction is to bridge the gap from when the parameters leave D in stage j to the end of stage j. In this period the variables $Y^{(i)}$ can alter in value by at most $C\varepsilon n$ for some constant C. As ε can be made arbitrarily small, we conclude that almost surely $Y^{(i)} = \hat{z}_{j,i}n + o(n)$ for $i \le j+1$. For $i \ge j+2$ the argument for the early part of stage $j+1$ shows that $Y^{(i)}$ remains at most $C \log n$ almost surely. The claim that it is $O(1)$ at the start of stage $j+1$ follows from the fact that $\hat{z}_{j,i-1} = 0$ and so $Y^{(i-1)} = o(n)$. Thus the probability that vertices of degree i are created on any given step is $o(1)$, and the expected number created is $o(1)$. This completes the induction and therefore the proof. □

To use Theorem 5 numerically in a strictly rigorous way would require analysis of all errors occurring in the required numerical solution of the differential equations, both those due to the approximation of the solution method and those due to the inherent inaccuracy of floating point arithmetic. Such an analysis is fairly simple but very tedious, since to be correct it depends on some of the precise details of the implementation, and the system is so well behaved that there are no surprises. So we do not give any such analysis in this paper, but merely report the results which were obtained by a second-order Runge–Kutta method which was repeated for ever-decreasing step sizes down to approximately 10^{-6} and gave well-behaved results, clearly to at least four decimal places. These are in Table 1. In obtaining these results, the hypotheses of Theorem 5 had to be verified numerically, which was done by checking that quantities required to be positive were greater than the largest possible error which could have occurred in the calculations. The condition that z_j has to be positive on (x_{j-1}, x_j) cannot be checked in this manner near x_{j-1} since z_j approaches 0 there. However, the derivative of z_j at that point is positive, which suffices.

For all $r \le 20$, stages 1 to $r-2$ are passed through consecutively, and the value of ξ becomes very small at the end of stage $r-2$, indicating that virtually all the vertices are used up at that time. Thus, in these cases Theorem 5 actually gives the asymptotic size of the independent set found by the algorithm, not merely a lower bound. We conjecture that this is the case for all $r \ge 3$.

Acknowledgments. Thanks are due Brendan McKay for suggesting the problem (in Section 4) for which the author originally developed this technique, and to Fima Klebaner for useful discussions on deviations.

REFERENCES

[1] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1** 311–316.
[2] BOLLOBÁS, B. (1985). *Random Graphs*. Academic Press, London.
[3] FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*. Wiley, New York.
[4] FRIEZE, A. M. and ŁUCZAK, T. (1992). On the independence numbers of random regular graphs. *J. Combin. Theory Ser. B* **54** 123–132.
[5] FRIEZE, A. M. and SUEN, S. (1994). On the independence number of random cubic graphs. *Random Structures and Algorithms* **5** 649–664.
[6] HOPKINS, G. and STATON, W. (1982). Girth and independence ratio. *Canad. Math. Bull.* **25** 179–186.
[7] HUREWICZ, W. (1958). *Lectures on Ordinary Differential Equations*. MIT Press.
[8] MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics 1989* (J. Siemons, ed.) 148–188. Cambridge Univ. Press.
[9] MCKAY, B. D. (1987). Independent sets in regular graphs of high girth. *Proceedings of the Australia–Singapore Joint Conference on Information Processing and Combinatorial Mathematics, Singapore, 1986. Ars Combin.* **23A** 179–185.
[10] RUCIŃSKI, A. and WORMALD, N. C. (1992). Random graph processes with degree restrictions. *Combinatorics, Probability and Computing* **1** 169–180.
[11] RUCIŃSKI, A. and WORMALD, N. C. (1995). Random graph processes with maximum degree 2. Unpublished manuscript.
[12] SHAMIR, E. and SPENCER, J. (1987). Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica* **7** 121–129.
[13] SHEARER, J. B. (1983). A note in the independence number of triangle-free graphs. *Discrete Math.* **46** 83–87.
[14] SHEARER, J. B. (1991). A note in the independence number of triangle-free graphs, II. *J. Combin. Theory Ser. B* **53** 300–307.
[15] TELCS, A. and WORMALD, N. C. (1995). Hamiltonicity of random graph processes with maximum degree 2. Unpublished manuscript.
[16] WORMALD, N. C. (1981). The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** 168–182.
[17] WORMALD, N. C. (1984). Generating random regular graphs. *J. Algorithms* **5** 247–280.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MELBOURNE
PARKVILLE, VIC 3052
AUSTRALIA