The Annals of Applied Probability
1997, Vol. 7, No. 1, 183–199

RANDOM GRAPH PROCESSES WITH MAXIMUM DEGREE 2

BY A. RUCIŃSKI[1] AND N. C. WORMALD[2]

Adam Mickiewicz University and University of Melbourne

Suppose that a process begins with n isolated vertices, to which edges are added randomly one by one so that the maximum degree of the induced graph is always at most 2. In a previous article, the authors showed that as $n \to \infty$, with probability tending to 1, the result of this process is a graph with n edges. The number of l-cycles in this graph is shown to be asymptotically Poisson ($l \geq 3$), and other aspects of this random graph model are studied.

1. Introduction. A random graph process begins with n vertices, and edges are inserted one at a time at random (see [1]). The authors [4] studied a restricted version of such a process, called a d-process, in which the degrees of the vertices are bounded above by a constant d, and it was shown that with probability tending to 1 as $n \to \infty$, the result of this process is a graph with $\lfloor nd/2 \rfloor$ edges. In the case that nd is even, this is a d-regular graph. Thus, this can be viewed as an algorithm for generating graphs with all degrees equal to d.

Generating graphs with n vertices of given degrees uniformly at random is difficult, and no good algorithm is known in general for degrees much greater than $n^{1/3}$, even for regular graphs (see [3]). In practice, the need for such graphs is met by algorithms which are simple but do not generate the graphs uniformly at random (see, e.g., [6]). However, these algorithms are not easy to analyze, and in [4] we instigated an approach by which some crucial questions regarding these algorithms may be answered. In the present article, we study an algorithm of this general type. We show, in particular, that it produces statistics of fundamental graph properties that differ from those of the uniform distribution. We restrict our attention here to graphs with maximum degree 2. Most problems involving graphs with bounded degrees become trivial when the bound is 2 and are interesting for the bound 3. One theme of this paper is that the problem under consideration already attains substantial complexity when the upper bound is 2.

The results of this paper give some indication of the comparison between nonuniform generation algorithms and uniform generation. Although we only consider the degree 2 case, it is to be hoped that understanding the low-degree case will at least give some idea of what happens for high degrees. It is for high degrees that uniform generation algorithms fail, as mentioned above,

Received March 1995; revised July 1996.

[1]Research supported by the University of Melbourne and the Australian Research Council; research partially supported by KBN Grant 2-1087-91-01.

[2]Research supported by the Australian Research Council.

AMS 1991 *subject classifications*. 05C80, 05C85.

Key words and phrases. Generation algorithms, number of cycles, limiting distributions.

and yet nonuniform methods such as the one studied here can be successful for generation. For example, Connor and Simberloff [2] and Wilson [7] used random $(0, 1)$-matrices with given row and column sums to investigate the distribution of species on a group of islands [where the (i, j)th entry is 1 if the ith species occurs on the jth island]. The basic idea is quite reasonable: to decide whether a pattern of colonization is unusual in some way, one can at least compare with a random $(0, 1)$-matrix with the same row and column sums. Random $(0, 1)$-matrices of roughly the same density do not provide a useful comparison because they would suggest that any pattern with some rare species and some common species (or, dually, some islands with many species and some with few species) is very unlikely. The conclusions of these studies lacked rigor in many ways: for instance, the question of what was the distribution of the matrices generated, and how it affected the statistics being measured, was totally ignored. These questions are very hard to answer, and part of the aim of the present paper is to investigate how far we can answer such questions.

To compare these studies with the present study, we note that $(0, 1)$-matrices are the incidence matrices of bicolored graphs, and so the algorithms of [2] and [7] can be viewed as generating random bicolored graphs with given degrees of the vertices. Even the best uniform generation methods will not cope with graphs as dense as the ones treated there. Instead, two of the methods considered there can be described as follows. Start with all vertices isolated. Randomly select the required number of neighbors of a vertex v_1, then the required number (remaining) of a vertex v_2 and so on. In [2] the vertices $v_1, v_2, \dots$ are in a given initially determined order. In [7], on the other hand, v_i is chosen at each step to be the vertex requiring the greatest number of edges still to be joined to it. It is noted that this seems to lead to higher probability that the algorithm actually terminates with all vertices of the desired specified degrees, rather than getting stuck with one or more vertices requiring extra edges but the deficient vertices already adjacent to each other. In this paper we study a slightly simpler algorithm, in which any two vertices still lacking edges are chosen at random and the edge between them added. It still contains many of the features which cause the difficulty of analysis of all these algorithms.

One of the main statistics studied in [2] and [7] is the number of co-occurrences of two species on two islands. This corresponds to the number of cycles of length 4 in the bipartite graphs. One of the topics of the present study is the number of cycles of given length in the graphs generated.

For this paper, a vertex is *unsaturated* if its degree is less than 2. A graph with maximum vertex degree at most 2 and in which the set of unsaturated vertices induces a complete subgraph is called 2-*maximal*. Formally, we define a 2-*process* to be a sequence $(g_0, g_1, \dots, g_n)$ of graphs on the vertex set $[n] = \{1, 2, \dots, n\}$ such that, for some $w \le n$, the following are satisfied:

1. $|E(g_i)| = i$, $i = 0, \dots, w$;
2. $g_i = g_w$, $i = w, \dots, n$;

3. $\varnothing = E(g_0) \subseteq E(g_1) \subseteq \cdots \subseteq E(g_n)$;
4. g_n is 2-maximal.

Property 2 is included merely for the convenience of having all sequences of equal length. From property 4 it follows that $w = n - 1$ or n.

A *random* 2-*process* is a probabilistic space whose elements are 2-processes with probabilities assigned as follows. Define u_i to be the number of unsaturated vertices in g_i, and f_i the number of edges whose ends both have degree 1 (isolated edges). Also define

$$a_{i+1} = \binom{u_i}{2} - f_i. \tag{1.1}$$

We assign the probability

$$\prod_{i=1}^{w} \frac{1}{a_i} \tag{1.2}$$

to the 2-process $(g_0, g_1, \ldots, g_n)$.

We think of g_i as being formed at time i. At time $w = w(g_1, \ldots, g_n)$, the graph becomes 2-maximal, and the process remains static until time n, which is the maximum time a process can possibly run for. The edges of g_n can be referred to as $e_1, \ldots, e_n$, in the order in which they appear in the process, where e_n can be left undefined if $w = n - 1$.

We use uppercase letters for the random variables corresponding to the deterministic parameters denoted by their lowercase counterparts. Thus, a random 2-process is denoted by $(G_0, G_1, \ldots, G_n)$, and A_i is the number of pairs of vertices available to be chosen as E_i.

All our asymptotic statements apply to random 2-processes as $n \to \infty$. In particular, a random 2-process has a property Q *almost surely* (a.s.) if $\lim_{n\to\infty} \mathbf{P}(Q) = 1$. A 2-process *saturates* if the final graph g_n is 2-regular. From [4], Theorem 1, a random 2-process almost surely saturates.

The difficulties in analyzing d-processes in general are discussed in [4]. The main idea used there to analyze d-processes is that certain functions of the process should follow long-term trends determined by the expected value of the change in the function for a single step. This gave a differential equation, whose solution approximately bounds above a variable associated with the number of isolated vertices in g_i. In the present context of 2-processes, we show in the next section that this variable is also bounded approximately below by the same function. This enables us to say accurately what the value of A_i is throughout the course of a 2-process (Theorem 2). This in turn gives valuable information for the further investigations of the properties of random 2-processes. The application of differential equations in describing d-processes was given in [9], but there the object was only to obtain $o(1)$ accuracy; here we need more. Cycles are studied in Section 3, and Theorem 2 is used in studying the distribution of the number of cycles of a given length in G_n. The result of major interest here is the following.

THEOREM 1. *Let $l \geq 3$ be fixed. In G_n the number of cycles of length l is asymptotically Poisson. For $l = 3$ the mean converges to*

$$\frac{1}{2}\int_0^\infty \frac{(\log(1+x))^2\,dx}{xe^x} \approx 0.188735349357788830.$$

We acknowledge L. Glasser for providing a formula by which the integral above can be computed efficiently. For $l \geq 4$ we do have a formula for the mean, but it is in the form of an l-fold integral (Theorem 4).

It is Theorem 1 that establishes a fundamental difference between G_n and the 2-regular graphs with the uniform probability distribution, since in the latter case the expected number of triangles is asymptotically $\frac{1}{6}$ (see, e.g., [8]).

2. Numbers of isolated and unsaturated vertices. Let i_j denote the number of vertices of degree 0 in g_j. In this section the distribution of I_j, U_j and A_j is determined sufficiently accurately for $j < n - n^{47/48}$ to establish the results in later sections. This is done by strengthening the argument given in [4] for random d-processes, which only gave an approximate upper bound on I_j, not lower bounds. This strengthening is possible because, in 2-processes, the numbers of isolated and unsaturated vertices determine each other uniquely: by counting vertex degrees, we obtain

$$u_j = 2(n-j) - i_j. \tag{2.1}$$

This will allow us to approximate I_j/n, $j < n - n^{47/48}$, by a function $b(x)$, with $x = j/n$, defined below. Alternatively, we could strengthen the arguments in [9].

The basis for the approximation of I_j/n comes from the following observation, which will be exploited rigorously in the next theorem. If $G_j = g_j$, then the expected decrease in the number of isolated vertices in the next step of the process (i.e., the expected value of $I_j - I_{j+1}$) is equal to twice the probability of hitting two isolated vertices with the randomly added edge e_{j+1}, plus the probability of hitting one isolated vertex. In view of (1.1), this is

$$\frac{2I_j(I_j-1)/2 + I_j(U_j - I_j)}{U_j(U_j-1)/2 - F_j}.$$

Thus, by (2.1) and since the number F_j of isolated edges is $O(n)$, the expected value of $I_{j+1} - I_j$ is approximately

$$\frac{-2I_j}{2n - 2j - I_j}. \tag{2.2}$$

Division of the numerator and denominator by n now suggests the equation

$$b'(x) = \frac{-2b(x)}{2 - 2x - b(x)}, \qquad b(0) = 1. \tag{2.3}$$

Define

$$v(x) = 2 - 2x - b(x). \tag{2.4}$$

Then, since $b(x)$ approximates I_j/n, by (2.1) $v(x)$ approximates U_j/n.

This informal discussion is made more precise in the following, which is the main result in this section.

THEOREM 2. *Let $C_0 > 0$. Then there is a constant C such that for a random 2-process, with probability $1 - o(n^{-C_0})$, we have*

$$\left| I_j - nb\left(\frac{j}{n}\right)\right| < Cn^{11/12}\sqrt{\log n},$$

$$\left| U_j - nv\left(\frac{j}{n}\right)\right| < Cn^{11/12}\sqrt{\log n},$$

$$\left| A_j - \frac{1}{2}n^2 v\left(\frac{j}{n}\right)^2\right| < Cn^{23/12}\log n$$

for all $j = 0, 1, \dots, n - \lfloor n^{47/48}\rfloor$.

PROOF. We deal in detail with the first inequality, from which the others will follow. There are two results we wish to extract from [4]. First, the inequality [see [4], (3.3)], which in the present context of maximum degree 2 is

$$\mathbf{E}(I_{k+t} - I_k | G_k = g_k) \le \frac{-2ti_k}{2n - 2k - i_k} + \frac{10t^2}{2n - 2k}, \tag{2.5}$$

using (2.1) can easily be strengthened to

$$\mathbf{E}(I_{k+t} - I_k | G_k = g_k) = \frac{-2ti_k}{2(n-k) - i_k} + O\left(\frac{t^2}{n-k}\right) \tag{2.6}$$

for $8 \le t \le u_k$. Here the leading term is just t times the quantity calculated at (2.2). Second, in the more general context of d-processes, we proved [see [4], (3.7)] that if $t^2 = o(n-k)$, then

$$\mathbf{P}\left(\left|I_{k+t} - I_k - \mathbf{E}(I_{k+t} - I_k | G_k)\right| \ge \sqrt{18ct\log n}\,\right) < n^{-c} \tag{2.7}$$

for any $c > 0$. This was done by considering the Doob martingale

$$X_t = \mathbf{E}(X | G_{k+t}), \qquad t = 0, 1, \dots, t_1,$$

where $X = I_{k+t_1} - I_k$, for some fixed t_1. It was shown that, provided $k + t_1$ is not too close to n, the differences $X_{t+1} - X_t$ are bounded, and so Azuma's inequality yields sharp concentration of X_t near 0 and consequently yields (2.7).

Qualitatively speaking, (2.7) pins down the value of I_{k+t} to something close to I_k plus the expected difference given in (2.6). However, the condition $t^2 = o(n-k)$ imposes an upper limit on how far in the process the relationship holds. This restriction can be circumvented by chaining together several applications; that is, we will apply (2.7) to the consecutive terms in a subsequence of $\{I_t\}$. The error in (2.7) increases rather slowly with t and in fact is concave down, so a large step linking two values I_{k_1} and I_{k_2} gives a smaller error than chaining together several small steps. Thus, to minimize the error, t^2 should be made close to $n-k$, thus satisfying the condition above. However, for our purposes, we do not need the minimum possible error, and for computational ease, we will choose t to be approximately $(n-k)n^{-2/3}$.

Define $\bar{k}_0 = 0$ and $\bar{k}_{j+1} = \bar{k}_j + (n-\bar{k}_j)n^{-2/3}$, $j = 1, \dots, s$, $s = \lfloor \frac{1}{48} n^{2/3} \log n \rfloor$, $k_j = \lfloor \bar{k}_j \rfloor$, $\Delta_j = k_{j+1} - k_j$, $j = 1, \dots, s$. Clearly, $\bar{k}_j = n(1-(1-n^{-2/3})^j)$, so $k_s = n - n^{47/48} + O(n^{5/16} \log n)$. Also define

$$\beta(k) = nb\left(\frac{k}{n}\right)$$

and

$$S_j = I_{k_j} - \beta(k_j).$$

Now write

$$S_{j+1} = -T_1 - T_2 + T_3,$$

where T_1 estimates the change in β and T_3 measures the change in I in the following way:

$$\begin{aligned} T_1 &= \beta(k_j + \Delta_j) - \beta(k_j) + \frac{2I_{k_j}\Delta_j}{2n - 2k_j - I_{k_j}}, \\ T_2 &= \beta(k_j) - I_{k_j}, \\ T_3 &= I_{k_j+\Delta_j} - I_{k_j} + \frac{2I_{k_j}\Delta_j}{2n - 2k_j - I_{k_j}}. \end{aligned}$$

We will now bound $|S_{j+1}|$ as a function of $|S_j|$. Throughout this argument we can regard n as fixed, and, for simplicity, write k for k_j and Δ for Δ_j. We have $|T_2| = |S_j|$ and, by (2.6) and (2.7) with $t = \Delta$,

$$\Pr\left(|T_3| \le \sqrt{18c\Delta \log n} + O\left(\frac{\Delta^2}{n-k}\right)\right) > 1 - n^{-c}.$$

At the end of this proof, we show

$$|T_1| \le \frac{4\Delta}{n-k}|S_j| + O\left(\frac{\Delta^2}{n-k}\right). \tag{2.8}$$

Thus, setting $d_1 = O(n^{1/6}\sqrt{\log n})$ and $d_2 = 4n^{-2/3}$, we obtain

$$\mathbf{P}(|S_{j+1}| \le d_1 + (1+d_2)|S_j|, \ \ j = 1, \dots, s) > 1 - sn^{-c}.$$

This iteration allows us to bound $|S_j|$ with high probability by the sequence w_j satisfying $w_0 = 0$, $w_{j+1} = d_1 + (1 + d_2)w_j$; that is,

$$\Pr(|S_j| \le w_j, \ \ j = 1, \dots, s) > 1 - sn^{-c}.$$

Solving the recurrence defining w_j, we obtain

$$w_j = \frac{d_1}{d_2}((1 + d_2)^j - 1) = O(n^{11/12}(\log n)^{1/2}).$$

Since $\Delta_j \le n^{1/3}$ for all j and $I_t \ge I_{t+1} \ge I_t - 2$ for all t, the above approximation remains valid not only for the partition marks k_j but for all $t = 0, \dots, k_s$. Hence, we have the theorem, the second two inequalities following from the first via (1.1), where $f_i \le n$, and (2.1).

It remains to show (2.8). For this, we need some properties of the function $b(x)$ which will be useful also in the next section.

Define

$$q = \frac{b}{1 - x} \tag{2.9}$$

for $0 \le x < 1$. Then substituting (2.9) into (2.3) and solving by separating variables gives

$$-\frac{2}{q} - \log q + 2 = \log(1 - x).$$

Taking into account the fact that q is nonincreasing and therefore bounded above by $q(0) = 1$, we obtain

$$q(x) \sim -\frac{2}{\log(1 - x)} \tag{2.10}$$

as $x \to 1$.

Also we now have $-2 \le b'(x) < 0$. It is easily checked that $b''(x) > 0$, and, hence,

$$b(x + \varepsilon) - b(x) \ge \varepsilon b'(x) \ge -2\varepsilon \tag{2.11}$$

for all ε sufficiently small. Note also that

$$b(x) \le 1 - x \tag{2.12}$$

since $q(x) \le 1$ for $0 \le x < 1$. Define

$$h(x, y) = \frac{-2y}{2 - 2x - y}, \qquad 0 \le x < 1, \ 0 \le y \le 1.$$

Then, for $x_0 \le x$, $y \le 1 - x$ and $y_0 \le 1 - x_0$,

$$\begin{aligned} |h(x, y) - h(x_0, y)| &\le (x - x_0) \max_{u \in [x_0, x]} \left| \frac{\partial h}{\partial u}(u, y) \right| \\ &= (x - x_0) \max_u \frac{4y}{(2 - 2u - y)^2} \le \frac{4(x - x_0)}{1 - x}, \end{aligned}$$

because $y \le 1 - x \le 1 - u$. Also

$$|h(x_0, y) - h(x_0, y_0)| \le |y - y_0| \max_v \left| \frac{\partial h}{\partial v}(x_0, v) \right| \le |y - y_0| \frac{4}{1 - x_0},$$

where the maximum is over all v between y_0 and y inclusively. (The order of y and y_0 is immaterial.) Thus,

$$|h(x, y) - h(x_0, y_0)| \le 4\left(\frac{x - x_0}{1 - x} + \frac{|y - y_0|}{1 - x_0} \right). \tag{2.13}$$

By definition,

$$\begin{aligned} |T_1| &= n \left| \int_{k/n}^{(k+\Delta)/n} \left(b'(x) + \frac{2I_k/n}{2 - 2k/n - I_k/n} \right) dx \right| \\ &\le \Delta \max_x \left| b'(x) - h\left(\frac{k}{n}, \frac{I_k}{n} \right) \right|, \end{aligned} \tag{2.14}$$

where $k/n \le x \le (k + \Delta)/n$.

Note that $b'(x) = h(x, b(x))$. By (2.12) and the fact that $I_k \le n - k$, the assumptions under which inequality (2.13) holds are satisfied with $x_0 = k/n$, $y_0 = I_k/n$ and $y = b(x)$. Thus, by (2.13), we get

$$\left| h(x, b(x)) - h\left(\frac{k}{n}, \frac{I_k}{n} \right) \right| \le \frac{4\Delta}{n(1 - x)} + \left| b(x) - \frac{I_k}{n} \right| \frac{4n}{n - k}.$$

Since b is decreasing, we have

$$\left| b(x) - \frac{I_k}{n} \right| \le \frac{1}{n}|S_j| + b\left(\frac{k}{n} \right) - b(x)$$

and, by (2.11),

$$b\left(\frac{k}{n} \right) - b(x) \le b\left(\frac{k}{n} \right) - b\left(\frac{k}{n} + \frac{\Delta}{n} \right) \le \frac{2\Delta}{n}.$$

Substituting these inequalities into (2.14) and noting that $x \le (k + \Delta)/n = (k + o(k))/n$, we obtain (2.8). □

By a different choice of Δ_j, we can vary the exponents in Theorem 2, and there is no guarantee that our proof gives the optimal values.

3. Cycles. Throughout this section we let $(G_1, \ldots, G_n)$ be a random 2-process, and put $G = G_n$. The following elementary bound gives some information on cycles in G.

THEOREM 3. *We have* $\mathbf{E}X(G) \le 3 + \log n$.

PROOF. Let $X(G_i)$ denote the total number of cycles in G_i, and $K(G_i)$ the number of components of G_i which are paths of length at least 2. Writing $Y_j = X(G_{j+1}) - X(G_j)$, we have $Y_j \in \{0, 1\}$. Given G_j, the conditional probability of the event $Y_j = 1$ is at most

$$\frac{2K(G_j)}{U_j(U_j - 1) - 2F_j} \leq \frac{1}{U_j - 2} \leq \frac{1}{n - j - 2}$$

(provided $j \leq n - 3$) as $F_j + K(G_j) \leq \frac{1}{2}U_j$. Hence,

$$\mathbf{E}X(G) = \mathbf{E}\sum Y_j \leq 2 + \sum_{j=0}^{n-3} \frac{1}{n - j - 2},$$

and the theorem follows. $\square$

For $l \geq 3$ let X_l denote the number of cycles of length l contained in G. Unfortunately, we do not have nice answers to many of the natural questions on the joint or individual distributions of the X_l, but the results of the previous section do permit many functions to be given explicitly in terms of integrals. Note that Theorem 3 gives an upper bound on the expected value of the sum of the X_l.

PROOF OF THEOREM 1. To avoid confronting l-fold integrals immediately, we give details in the case $l = 3$ before considering the more general case. We concentrate on finding the asymptotic value of $\mathbf{E}X_3$.

Let C be as in Theorem 2, for $C_0 = 6$. Also let

$$p_3 = \mathbf{P}(\text{vertices 1, 2 and 3 form a triangle in } G).$$

We now have

$$\mathbf{E}X_3 = \binom{n}{3} p_3.$$

For $j = 1, \dots, n$, define

$$\mathscr{H}_j = \begin{cases} \{E_j \cap \{1, 2, 3\} = \varnothing\}, & \text{if } j \notin \{r, s, t\}, \\ \{E_r = \{1, 2\}\}, & \text{if } j = r, \\ \{E_s = \{2, 3\}\}, & \text{if } j = s, \\ \{E_t = \{1, 3\}\}, & \text{if } j = t. \end{cases}$$

Let $\mathscr{B}_j$ be the event $\mathscr{H}_1 \wedge \cdots \wedge \mathscr{H}_{j-1}$, and $\mathscr{T}(r, s, t)$ the event $\mathscr{H}_r \wedge \mathscr{H}_s \wedge \mathscr{H}_t$. Thus, $\mathscr{T}(r, s, t)$ is the event that the edges of a triangle with vertices 1, 2 and 3 are added at times r, s and t in a given order, and $\mathscr{B}_j$ is the event that the steps before the jth edge is added do not rule $\mathscr{T}(r, s, t)$. We have

$$p_3 = 6 \sum_{1 \leq r < s < t \leq n} p_{r,s,t}, \tag{3.1}$$

where

$$\begin{aligned} p_{r,s,t} &= \mathbf{P}(\mathscr{T}(r,s,t)) \\ &= \mathbf{P}(\mathscr{B}_{t+1}) \\ &= \prod_{j=1}^{t} P_j, \qquad (3.2)\\ P_j &= \mathbf{P}(\mathscr{H}_j | \mathscr{B}_j). \end{aligned}$$

Note that, for any g_{j-1},

$$\mathbf{P}(\mathscr{H}_j | \mathscr{B}_j \wedge \{G_{j-1} = g_{j-1}\}) = 1 - \frac{z_j(u_{j-1})}{a_j} \tag{3.3}$$

for $1 \le j < t$, $j \ne r, s, t$, where

$$z_j(x) = \begin{cases} 3(x-3)+3, & \text{if } j < r, \\ 3(x-3)+2, & \text{if } r < j < s, \\ 2(x-2)+1, & \text{if } s < j < t, \end{cases}$$

provided the conditional probability is well defined. Here $z_j(u_{j-1})$ gives the number of available edges of g_{j-1} incident with 1, 2 or 3.

Choose $47/48 < \beta < \alpha < 1$ and rewrite (3.1) as

$$p_3 = 6(S_1 + S_2 + S_3), \tag{3.4}$$

where S_1 contains those terms with $t \le n - n^\beta$, S_2 contains those terms with $t > n - n^\beta$ and $r \le n - n^\alpha$ and S_3 contains the rest. We examine S_2 first because it is simplest.

Since $i_j \le u_j$ we have from (2.1) that

$$n - j \le u_j \le 2(n-j)$$

for all j. Also, from (1.1) we get

$$a_j \le \binom{u_{j-1}}{2},$$

and so (3.3) is bounded above by $1 - 6/u_{j-1} + O(u_{j-1}^{-2})$ for $u_{j-1} \ge 3$ and $j < s$.

It follows that

$$P_r = O((n-r)^{-2}), \qquad P_s = O((n-s)^{-2}), \qquad P_t = O((n-t)^{-2}),$$

$$\log P_j \le \frac{-3}{n-j+1} + O\left(\frac{1}{(n-j)^2}\right)$$

for $j < r$ or $r < j < s$, and, similarly,

$$\log P_j \le \frac{-2}{n-j+1} + O\left(\frac{1}{(n-j)^2}\right)$$

for $s < j < t$. Thus, from (3.2),

$$\begin{aligned} p_{r,s,t} &= O\left(\frac{1}{(n-r)^2}\frac{1}{(n-s)^2}\frac{1}{(n-t)^2}\left(\frac{n-s}{n}\right)^3\left(\frac{n-t}{n-s}\right)^2\right) \\ &= O\left(\frac{1}{n^3}\frac{1}{(n-r)^2}\frac{1}{(n-s)}\right), \end{aligned} \tag{3.5}$$

and so

$$\begin{aligned} S_2 &= O\left(\frac{n^\beta}{n^3}\sum_{i>n^\alpha}\frac{1}{i^2}\sum_{j=1}^{i-1}\frac{1}{j}\right) \\ &= O(n^{-3+\beta-\alpha}\log n) \\ &= o(n^{-3}). \end{aligned} \tag{3.6}$$

Write $\mathscr{K}_j$ for the event that I_j satisfies the inequality given by Theorem 2, with C as chosen at the start of this proof. Thus, by Theorem 2,

$$\mathbf{P}(\mathscr{K}_1 \wedge \cdots \wedge \mathscr{K}_{\lfloor n-n^\beta \rfloor}) \ge n - o(n^{-6}),$$

and A_j and U_j satisfy similar inequalities.

Put $t_0 = \lfloor n - n^\alpha \rfloor$ and note that

$$\begin{aligned} S_3 &\le \sum_{t_0<r<s<t\le n} \mathbf{P}(\mathscr{B}_{t_0} \wedge \mathscr{T}(r,s,t)) \\ &\le \sum_{t_0<r<s<t\le n} [\mathbf{P}(\mathscr{B}_{t_0} \wedge \mathscr{T}(r,s,t) | \mathscr{K}_{t_0}) + 1 - \mathbf{P}(\mathscr{K}_{t_0})] \\ &= \sum_{t_0<r<s<t\le n} [\mathbf{P}(\mathscr{T}(r,s,t)|\mathscr{B}_{t_0} \wedge \mathscr{K}_{t_0})\mathbf{P}(\mathscr{B}_{t_0}|\mathscr{K}_{t_0}) + o(n^{-6})] \end{aligned} \tag{3.7}$$

by Theorem 2.

Using the argument leading to (3.5), we get

$$\begin{aligned} \mathbf{P}(\mathscr{T}(r,s,t)|\mathscr{B}_{t_0} \wedge \mathscr{K}_{t_0}) &= O(\mathbf{P}(\mathscr{T}(r,s,t)|\mathscr{B}_{t_0})) \\ &= O\left(\frac{1}{(n-t_0)^3}\frac{1}{(n-r)^2}\frac{1}{(n-s)}\right). \end{aligned}$$

Note that $\mathscr{B}_{t_0}$ is the event that r, s and t are all isolated in G_{t_0}. Thus, by symmetry of the vertices,

$$\mathbf{P}(\mathscr{B}_{t_0}|I_{t_0} = i) = \frac{\binom{n-3}{i-3}}{\binom{n}{i}} < \frac{i^3}{n^3}.$$

So, by Theorem 2, (2.9) and (2.10), we obtain, on summing over all i in the defining range for $\mathscr{K}_{t_0}$,

$$\begin{aligned}\mathbf{P}(\mathscr{B}_{t_0}|\mathscr{K}_{t_0}) &= \sum_i \mathbf{P}(I_{t_0}=i|\mathscr{K}_{t_0})\mathbf{P}(\mathscr{B}_{t_0}|I_{t_0}=i)\\ &= O((n-t_0)^3/(\log n)^3)\\ &= O(n^{3\alpha-3}/(\log n)^3).\end{aligned}$$

Thus, since the number of terms in (3.7) is $O(n^{3\alpha})$, it gives

$$S_3 = O\left(\frac{1}{n^3(\log n)^2}\right). \tag{3.8}$$

To examine S_1, we need to repeat the calculation leading to (3.5) more accurately. The method is similar, but the use of Theorem 2 complicates matters.

Suppose that $t \le n - n^\beta$. We wish to approximate

$$P_j = \mathbf{P}(\mathscr{H}_j|\mathscr{B}_j)$$

by

$$\mathbf{P}(\mathscr{H}_j|\mathscr{B}_j \wedge \mathscr{K}_{j-1})$$

in order to take advantage of Theorem 2. Thus, we need a lower bound on $\mathbf{P}(\mathscr{B}_j)$. To do this, we use induction on j. The actual inductive statement which we prove is that, for $j = 0, 1, \dots, n - \lfloor n^{47/48} \rfloor$,

$$P_j = \begin{cases} (2+o(1))\Big/\left(nv\left(\dfrac{j}{n}\right)\right)^2, & \text{if } j = r, s \text{ or } t,\\ 1 - \dfrac{6}{nv(j/n)} + o(n^{-1}), & \text{if } j < r \text{ or } r < j < s,\\ 1 - \dfrac{4}{nv(j/n)} + o(n^{-1}), & \text{if } s < j < t.\end{cases} \tag{3.9}$$

Here $o(\ )$ is uniform over j. Assume this is true for all numbers less than j. If $j \le r$, then

$$\begin{aligned}\mathbf{P}(\mathscr{B}_j) &= \prod_{k=1}^{j-1} P_k\\ &= \prod_{k=1}^{j-1} \exp\left(-\frac{6}{nv(k/n)} + o(n^{-1})\right)\\ &= \exp\left(o(1) - 6\sum_{k=1}^{j-1}\frac{1}{nv(k/n)}\right)\\ &\sim \exp\left(-6\int_0^{j/n}\frac{dx}{v(x)}\right).\end{aligned} \tag{3.10}$$

To justify the accuracy of the latter integral, we note that v is strictly decreasing and $v(j/n) > n^{-1/48}$. Similarly, if $r < j \le s$, then the resultant formula for $\mathbf{P}(\mathscr{B}_j)$ is equal to (3.10) multiplied by $2/(nv(r/n))^2$, whereas if $s < j \le t$, then

$$\mathbf{P}(\mathscr{B}_j) \sim \frac{4}{n^4 v(r/n)^2 v(s/n)^2} \exp\left(-6\int_0^{s/n} \frac{dx}{v(x)} - 4\int_{s/n}^{j/n} \frac{dx}{v(x)}\right). \tag{3.11}$$

Note from (2.4) that

$$v = -\frac{2b}{b'} \tag{3.12}$$

and

$$v = b(1 - \log b). \tag{3.13}$$

Hence, we have

$$\int_{c_0}^{c_1} \frac{dx}{v(x)} = \int_{c_0}^{c_1} \frac{-b'(x)\,dx}{2b(x)} = \frac{1}{2}(\log b(c_0) - \log b(c_1)) \tag{3.14}$$

and $\log b(0) = 0$. Thus, (3.10) reduces to a ratio of small powers of b. Since $v \le 2$, the formula for $r < j \le s$ is (3.10) divided by $O(n^2)$, and, similarly, (3.11) is a product of small powers of b divided by $O(n^4)$. For all j under consideration we have by (2.10) that $b > C/(n^{1/48}\log n)$. Hence,

$$\mathbf{P}(\mathscr{B}_j) > n^{-5} \tag{3.15}$$

for n sufficiently large. This holds for $j < n - \lfloor n^{47/48} \rfloor$.

Before making use of these calculations in proving (3.9), we note that

$$\begin{aligned}\mathbf{P}(\mathscr{H}_j \wedge \mathscr{B}_j) &= \mathbf{P}(\mathscr{H}_j \wedge \mathscr{B}_j \wedge \mathscr{K}_{j-1}) + O(1 - \mathbf{P}(\mathscr{K}_{j-1}))\\ &= \mathbf{P}(\mathscr{H}_j \wedge \mathscr{B}_j \wedge \mathscr{K}_{j-1}) + o(n^{-6})\end{aligned}$$

and, similarly,

$$\mathbf{P}(\mathscr{B}_j) = \mathbf{P}(\mathscr{B}_j \wedge \mathscr{K}_{j-1}) + o(n^{-6}).$$

By (3.13) and the choice of C_0, we now get

$$\begin{aligned}P_j &= \mathbf{P}(\mathscr{H}_j | \mathscr{B}_j)\\ &= \mathbf{P}(\mathscr{H}_j | \mathscr{B}_j \wedge \mathscr{K}_{j-1})(1 + o(n^{-1}))\\ &= \mathbf{E}(\mathbf{E}(\mathbf{I}(\mathscr{H}_j)|G_{j-1})|\mathscr{B}_j \wedge \mathscr{K}_{j-1})(1 + o(n^{-1})),\end{aligned}$$

where $\mathbf{I}(\mathscr{H})$ denotes the indicator function of an event $\mathscr{H}$. For $j < r$, the outer expectation here becomes

$$\begin{aligned}\mathbf{E}\left(1 - \frac{3(U_{j-1} - 3) + 3}{A_j}\,\middle|\, \mathscr{B}_j \wedge \mathscr{K}_{j-1}\right) &= 1 - \frac{3nv(j/n) + o(n^{12/13})}{n^2 v(j/n)^2/2 + o(n^{25/13})}\\ &= 1 - \frac{6}{nv(j/n)} + o(n^{-1}).\end{aligned}$$

Similar computations apply to the values of j falling into the other intervals. This completes the inductive proof of (3.9).

Putting $j = t$, arguing as for (3.10) and (3.11) and using (3.12) gives

$$\mathbf{P}(\mathscr{B}_{t+1}) \sim \frac{8b(s/n)b(t/n)^2}{n^6 v(r/n)^2 v(s/n)^2 v(t/n)^2}. \tag{3.16}$$

Note that S_1 is the sum of this quantity over $1 \leq r < s < t \leq n - n^{\beta}$. Thus, using (3.1), (3.2), (3.4), (3.6) and (3.8), we now get

$$p_3 = o(n^{-3}) + \frac{48}{n^6} \sum_{1 \leq r < s < t \leq n - n^{\beta}} \frac{b(s/n)b(t/n)^2}{v(r/n)^2 v(s/n)^2 v(t/n)^2}.$$

Thus, since $\mathbf{E}X_3 \sim n^3 p_3/6$, we have

$$\mathbf{E}X_3 \sim 8 \int_0^{\mu} \int_{x_1}^{\mu} \int_{x_2}^{\mu} \frac{b(x_2)b(x_3)^2}{v(x_1)^2 v(x_2)^2 v(x_3)^2} \, dx_3 \, dx_2 \, dx_1,$$

where $\mu = 1 - n^{-1/48}$. The justification for approximating the sum by the integral becomes clear after the following changes of variable.

Set

$$y_i = 1 - \log b(x_i).$$

Then, by (3.12) and (3.13), we have

$$y_i = \frac{v(x_i)}{b(x_i)}, \qquad dy_i = \frac{2dx_i}{v(x_i)}.$$

Thus,

$$\mathbf{E}X_3 \sim \int_1^{\mu_1} \int_{y_1}^{\mu_1} \int_{y_2}^{\mu_1} \frac{\exp(y_1 - y_3)}{y_1 y_2 y_3} \, dy_3 \, dy_2 \, dy_1,$$

where $\mu_1 = 1 - \log b(\mu)$. It is easy to verify that the integral is bounded and that the upper limits can be replaced by ∞. Hence,

$$\begin{aligned} \mathbf{E}X_3 &\sim \int_1^{\infty} \int_{y_1}^{\infty} \int_{y_2}^{\infty} \frac{\exp(y_1 - y_3)}{y_1 y_2 y_3} \, dy_3 \, dy_2 \, dy_1 \\ &= \int_1^{\infty} \int_{y_1}^{\infty} \frac{\exp(y_1 - y_3)(\log y_1 - \log y_3)}{y_1 y_3} \, dy_3 \, dy_1, \end{aligned} \tag{3.17}$$

upon reversing the order of the second and third integrals. Making the substitutions

$$x = y_3 - y_1,$$
$$y = \log y_3 - \log y_1$$

gives

$$\mathbf{E}X_3 \sim \int_0^\infty \int_0^{\log(x+1)} \frac{e^{-x}y}{x}\,dy\,dx$$
$$\sim \frac{1}{2}\int_0^\infty \frac{(\log(1+x))^2\,dx}{xe^x}.$$

To establish the fact that X_3 is asymptotically Poisson, we show that its factorial moments behave correctly. First consider $\mathbf{E}(X_3(X_3-1))$. We do not give all the details since the argument is similar to that for $\mathbf{E}X_3$. In particular, C_0 must be rechosen.

We have

$$\mathbf{E}(X_3(X_3-1)) = \binom{n}{3}\binom{n-3}{3}p_{3,3},$$

where

$$p_{3,3} = 36 \sum_{1\le r<s<t\le n}\ \sum_{1\le r'<s'<t'\le n} \mathbf{P}(\mathscr{T}(r,s,t)\wedge \mathscr{T}'(r',s',t'))$$

and $\mathscr{T}'(r',s',t')$ is the event that $E_{r'}=\{4,5\}$, $E_{s'}=\{5,6\}$ and $E_{t'}=\{4,6\}$. Of course, terms in this sum in which $\{r',s',t'\}\cap\{r,s,t\}\neq\varnothing$ contribute 0. We can analyze this in the same way that we examined $\mathbf{E}X_3$, modifying the definition of $\mathscr{H}_j$ in the obvious way. Note that in place of the factor 6 multiplying the integral in (3.10), there is the factor 12 if, in addition, $k<r'$, whereas if say $s<k\le r'$ and $k\le t$ we obtain instead of (3.11)

$$\mathbf{P}(\mathscr{B}_k) \sim \frac{4}{n^4 v(r/n)^2 v(s/n)^2}\exp\left(-12\int_0^{s/n}\frac{dx}{v(x)} - 10\int_{s/n}^{k/n}\frac{dx}{v(x)}\right),$$

which is asymptotic to (3.11) multiplied by

$$\exp\left(-6\int_0^{k/n}\frac{dx}{v(x)}\right).$$

In this way the effects of the two triangles separate into two factors, and, thus, for $\mathbf{P}(\mathscr{B}_{t+1})$ we have the product of the function of r, s and t given on the right-hand side of (3.14), together with the same function of r', s' and t'. Hence, the sextuple summation above separates into the product of two triple summations, and we get

$$\mathbf{E}(X_3(X_3-1)) \sim (\mathbf{E}X_3)^2.$$

For similar reasons the ith factorial moment of X_3 is asymptotic to the ith power of $\mathbf{E}X_3$, $i\ge 2$, and so we deduce that X_3 is asymptotically Poisson. In the same way it is readily verified that X_l is asymptotically Poisson, $l\ge 4$. □

The method of proof of Theorem 1 gives an asymptotic value of $\mathbf{E}X_l$ for $l\ge 4$. The statement of this, in the following theorem, requires some development. First, consider the formation of an l-cycle on $[l]$, in the course of a random 2-process. There are $(l-1)!/2$ possibilities for the l-cycle, but we can choose just one, say $1\ 2\ \cdots\ l\ 1$. As the edges of this cycle come in, the exponential coefficients in formulas analogous to (3.10) and (3.11) keep changing in a pattern determined by the number of saturated vertices in that cycle. In the case $l = 3$, all six orderings of the appearances of the edges in the triangle gave the same exponential coefficients. However, for $l \geq 4$, the ordering is of significance. We associate the $l!$ possible orderings with the elements σ of the symmetric group S_l of order l. Denote by $\sigma^*(i)$ the number of new vertices of degree 2 created in the l-cycle when the ith edge is added according to σ. When applying the argument in the proof of Theorem 1, we obtain the following as the analogue of (3.15). Note that in the sequence $1-\sigma^*(1), \ldots, 1-\sigma^*(l)$, all partial sums are positive except the total sum, which is 0. In addition, all such sequences are realizable in this context.

THEOREM 4. *For $l \geq 3$, $\mathbf{E}X_l$ is asymptotic to*

$$\frac{1}{2l}\sum_{\sigma\in S_l}\int_1^\infty dx_1\int_{x_1}^\infty dx_2\cdots\int_{x_{l-1}}^\infty dx_l\frac{\exp\Big((1-\sigma^*(1))x_1+\cdots+(1-\sigma^*(l))x_l\Big)}{x_1\cdots x_l}.$$

4. Related matters. Some other questions are easily answered from the results we have obtained. For instance, the maximum number of vertices of degree 1 occurring throughout a 2-process is seen from Theorem 2, (2.1), (2.3), (2.9) and the equation following it to be approximately n/e almost surely, occurring approximately at time $t = n(1 - 3/2e)$.

We acknowledge P. Erdös for contributing most of the questions in the following list. Following the spirit of this paper, we ask only for the limiting or asymptotic behavior as $n \to \infty$.

1. When does the first cycle appear?
2. What is the maximum number of isolated edges throughout the process?
3. How much time remains when the last vertex of degree 0 disappears?
4. What is the distribution of the length of the longest cycle of g_n?
5. What is the distribution (or even just the expectation) of the number of cycles of g_n?
6. What is the asymptotic probability that g_n is a cycle of length n?
7. How close is $\mathbf{E}X_l$ in the limit to $1/(2l)$, which is the expected number of l-cycles in a 2-regular graph chosen uniformly at random [8]?

Since the first version of this article, answers were found to questions 1, 5 and 6, to appear in a forthcoming article [5]. The answer to question 6 is slightly different from the corresponding probability that a random 2-regular graph (with the uniform distribution) is Hamiltonian, which is asymptotically $e^{3/4}\sqrt{\pi}/2\sqrt{n}$.

Methods similar to those in the present paper will probably suffice to answer question 2. Presumably an answer to question 7 requires evaluation of the integrals occurring in Theorem 4. In order to do this, one would presumably need to get some insight into the distribution of the sequence σ^* for a random permutation σ. This leads to the study of a random 2-process performed on an underlying graph which is an l-cycle, rather than on the complete graph as for ordinary 2-processes.

REFERENCES

[1] BOLLOBÁS, B. (1985). *Random Graphs*. Academic Press, London.

[2] CONNOR, E. F. and SIMBERLOFF, D. (1979). The assembly of species communities: chance or competition? *Ecology* **60** 1132–1140.

[3] MCKAY, B. D. and WORMALD, N. C. (1991). Uniform generation of random regular graphs of moderate degree. *J. Algorithms* **11** 52–67.

[4] RUCIŃSKI, A. and WORMALD, N. C. (1992). Random graph processes with degree restrictions. *Combinatorics, Probability and Computing* **1** 169–180.

[5] TELCS, A. and WORMALD, N. C. (1996). Hamiltonicity of random 2-processes. Unpublished manuscript.

[6] TINHOFER, G. (1979). On the generation of random graphs with given properties and known distribution. *Appl. Comput. Sci., Ber. Prakt. Inf.* **13** 265–297.

[7] WILSON, J. B. (1987). Methods for detecting non-randomness in species co-occurrences: a contribution. *Oecologia* **73** 579–582.

[8] WORMALD, N. C. (1981). The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** 168–182.

[9] WORMALD, N. C. (1995). Differential equations for random processes and random graphs. *Ann. Appl. Probab.* **5** 1217–1235.

DEPARTMENT OF DISCRETE MATHEMATICS
ADAM MICKIEWICZ UNIVERSITY
MATEJKI 48-49
60-769 POZNAŃ
POLAND

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MELBOURNE
PARKVILLE, VIC 3052
AUSTRALIA
E-MAIL: nick@maths.mu.oz.au