Combinatorics, Probability and Computing (1992) 1, 169–180

Random Graph Processes with Degree Restrictions

A. RUCIŃSKI[†] and N. C. WORMALD[‡]

[†] Department of Discrete Mathematics, Adam Mickiewicz University, Matejki 48-49, 60-769 Poznań, Poland

[‡] Department of Mathematics, University of Melbourne, Parkville, VIC 3052, Australia

Received 4 September 1991; revised 29 November 1991

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 bounded above by d. We prove that if $n \to \infty$ with d fixed, then with probability tending to 1, the final result of this process is a graph with $\lfloor nd/2 \rfloor$ edges.

1. Introduction

Random graph processes as defined by Bollobás in [2] provide a dynamic model of a random graph by inserting the edges one by one rather than all simultaneously. In this paper we study a restricted version of this process in which the degrees of the vertices are bounded above. This condition makes analysis more difficult since the resultant probabilistic space is not uniform and the length of the process can vary.

The final stage of such a process is a random graph which is maximal with respect to the given bounds on its degrees. 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 [5]). In practice, the need for such graphs is met by algorithms which are simple but do not generate the graphs uniformly at random. We will essentially be studying an algorithm sharing some of the features of the one given in [8] for generating random graphs with given degrees. These algorithms are not easy to analyse, and this paper instigates an approach by which some of the crucial questions regarding them may be answered, at least when the degrees are bounded or growing only slowly. This hopefully leads to some insight into their behaviour for larger d.

We restrict our study to a process which can be described as follows. We use $d_g(v)$ for the degree of a vertex v in a graph g. Fix $d > 0$ and begin with n isolated vertices, then

[†] Supported by the University of Melbourne and the Australian Research Council. Research partially supported by KBN grant number 2 1087 9101.

[‡] Supported by the Australian Research Council

add edges one by one. The aim is to create an edge-maximal graph g with $\max_v d_g(v) \le d$, so define an *unsaturated* vertex to be one with degree less than d. At each step, choose an available pair of vertices uniformly at random and add the next edge there. By available here, we mean a pair of vertices which are non-adjacent and both unsaturated. (In the simplest implementation of the single step, one actually repeatedly chooses a pair of unsaturated vertices of degree less than d uniformly at random, and accepts the first pair so obtained consisting of non-adjacent vertices.) The process finishes when no more edges can be added; i.e, the set of unsaturated vertices induces a complete subgraph. A graph with this property and with maximum vertex degree at most d is called *d-maximal*. Hence, in a d-maximal graph g on n vertices, there are at most d unsaturated vertices. It follows, as explained in [6], that

$$|E(g)| \ge nd/2 - \lceil (d^2 + 2d)/8 \rceil, \tag{1.1}$$

where $E(g)$ denotes the edge set of g.

For d fixed, we prove in this paper that this process almost surely results in a graph which is d-regular if nd is even, and otherwise has only one unsaturated vertex, which is of degree $d-1$.

More formally, we define a *d-process* to be a sequence $(g_0, g_1, \ldots, g_N) = (g_0^d, g_1^d, \ldots, g_N^d)$ of graphs on the vertex set $[n] = \{1, 2, \ldots, n\}$ such that for some $w \le N = \lfloor nd/2 \rfloor$, the following are satisfied:

(i) $|E(g_i)| = i, \qquad i = 0, \ldots, w,$

(ii) $g_i = g_w, \qquad i = w, \ldots, N,$

(iii) $\emptyset = E(g_0) \subseteq E(g_1) \subseteq \cdots \subseteq E(g_N),$

(iv) g_N is d-maximal.

Property (ii) is included merely for the convenience of having all sequences of equal length.

A *random d-process* is a probabilistic space whose elements are d-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 for which both ends are unsaturated vertices in g_i. Also define

$$a_{i+1} = \binom{u_i}{2} - f_i. \tag{1.2}$$

We assign the probability

$$\prod_{i=1}^{w} \frac{1}{a_i} \tag{1.3}$$

to the d-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 d-maximal, and the process remains static until time N, which is the maximum time a process can possibly run for. We alternatively refer to $g_N = g_N^d$ as g^d. Its edges can be referred to as $e_1, \ldots, e_N$, in the order in which they appear in the process, where $e_{w+1}, \ldots, e_N$ can be left undefined. Note that for fixed d, $N - w$ is bounded above via (1.1).

We use upper case letters for the random variables corresponding to the deterministic parameters denoted by their lower case counterparts. Thus, a random d-process is denoted by $(G_0, G_1, \dots, G_N)$. Referring to the earlier informal definition of a random d-process, A_i is the number of pairs of vertices available to be chosen as E_i.

Let e be an available pair of vertices of a graph g_i in the above sense. Then we have

$$\Pr(E_i = e | G_{i-1} = g_{i-1}) = \frac{1}{a_i},$$

and so the formal definition of a random d-process has the properties specified in the earlier informal definition. The fact that this probability is conditional on the past history of the process in a nontrivial way is an important feature of random d-processes. It can be blamed for the difficulties encountered in analysing random d-processes, in comparison with the more commonly studied graph processes. To give some idea of the flavour of these difficulties, we consider the following example.

Throughout this paper, $\mathrm{Ind}(\mathscr{H})$ denotes the indicator function of an event $\mathscr{H}$. For $j \le k$ and given pairs $e_j, \dots, e_k$ of elements of $[n]$, let P denote the conditional probability that

$$E_i = e_i,\ i = j, \dots, k,$$

given that $e_j, \dots, e_k$ are all available pairs of G_{j-1}. Then one might be tempted to write

$$P = \prod_{i=j}^{k} \frac{1}{A_i},$$

but the expression on the right of this equation is a random variable without relevant meaning in this context, and hence the equation must be rejected. However, if we happen to have a lower bound s_i on a_i for $j \le i \le k$, we are justified in writing the inequality

$$P \le \prod_{i=j}^{k} \frac{1}{s_i}.$$

The terminal segment $(G_j, G_{j+1}, \dots, G_N)$ of a random d-process can be viewed as another random d-process for which the degree restrictions are not necessarily uniform, and in which some pairs of vertices (those joined by edges of G_j) are forbidden.

In order to make use of this observation in proofs, we need to introduce a more general graph process. Breaking new ground in the realms of innovative nomenclature, we call this more general process a *generalized d-process*. For this process, each vertex $v \in [n]$ is assigned a natural number $m(v) \le d$ which sets a bound on its degree. (To model the terminal segment above, we would choose $m(v) = d - d_{G_j}(v)$.) Then, v is *unsaturated* in a graph if its degree is strictly less than $m(v)$. Also, at the beginning of this process a subset Φ of forbidden pairs of elements in $[n]$ is distinguished, whose elements are to be avoided when picking an edge. The set Φ viewed as a graph is assumed to have bounded maximum degree. A graph g is *(m, Φ)-maximal* if $E(g) \cap \Phi = \emptyset$, $d_g(v) \le m(v)$ for each $v \in [n]$, and each pair of unsaturated vertices of g is in $E(g) \cup \Phi$.

The rules governing the random generalized process remain the same as in the ordinary d-process: at each stage we choose an available pair uniformly at random and add the next edge there. However, by "available pair" we now mean a pair of unsaturated vertices which is neither already present in the graph nor a member of Φ.

Thus we have, as for d-processes, that the number of available pairs in a graph g_i is

a_{i+1} as determined by (1.2), where u_i is defined as before and f_i is the total number of pairs which are either in Φ or are edges of g_i with both ends unsaturated (or both). In particular, $f_0 = |\Phi|$ provided $m(v) > 0$ for all v. Note that there will be at most $N_0 = \lfloor \frac{1}{2} D_0 \rfloor$ steps in a generalised process, where $D_0 = \sum_v m(v)$ is called the *initial deficit*. Similarly, at any given time i, the *current deficit* is

$$D_i = \sum_v (m(v) - d_{g_i}(v)), \tag{1.4}$$

which is $D_0 - 2i$ provided $i \le w$. The quantity $N_0 - i$ is an upper bound on the time remaining before time w when the process becomes static. Conversely, since Φ has bounded maximum degree, the number of unsaturated vertices in an (m, Φ)-maximal graph is bounded above, and thus, we have $w = N_0 - O(1)$.

For a formal definition of a generalised d-process, we amend the definition of a d-process by replacing N by N_0 throughout, and replacing condition (iv) by

(iv′) g_{N_0} is (m, Φ)-maximal.

For a random generalised d-process, the probability associated with each sequence is again given by (1.3).

All our asymptotic statements apply to random d-processes as $D_0 \to \infty$ with d fixed, but uniformly over the parameters $m(v)$ and Φ. In particular, a random d-process has a property Q *almost surely* (*a.s.*) if $\lim_{D_0 \to \infty} \Pr(Q) = 1$. Naturally, if $m(v) \ge 1$ for all v, this can be replaced by $\lim_{n \to \infty} \Pr(Q) = 1$. A process *saturates* if the final deficit, D_N or D_{N_0} as the case may be, is at most 1. Erdős has asked (in a private communication) the following question: "What is the limiting distribution as $n \to \infty$ of the number of unsaturated vertices of G_N?".

Most questions concerning graphs with bounded degrees are trivial when the upper bound is 2 and non-trivial for larger bounds. This is not the case for the question here, although the simpler structure of 2-processes does tend to make computations easier.

For $d = 2$, simulation, and also exact probability calculations up to $n = 500$, did not give any strong suggestion of the correct answer. Exact calculations using the method described in [6, Section 4] show that the probability that a random 2-process saturates is monotonically increasing from $n = 5$ to $n = 500$, at which point it is roughly 0.879, and simulation with $n = 30,000$ suggests a value there of approximately 0.9. Moreover, based on the exact numbers for $n \le 500$, we believe that the probability of not saturating, multiplied by $\log n$, is increasing (though bounded) for all sufficiently large n. (Here and throughout this paper we use log to denote the natural logarithm.) Balińska and Quintas [1] have other results on these processes from simulation.

The main idea we use to analyse 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 gives a differential equation (Section 2) whose solution approximates an upper bound for the function in question almost surely (see Lemma 3.2). To establish the required concentration of the function near its expectation, we use Azuma's martingale inequality. As an answer to Erdős's question we prove the following in Section 3.

Theorem 1.1. *For fixed d, a random generalised d-process almost surely saturates.*

This contrasts with an alternative model of d-maximal graphs where these graphs are given the uniform probability distribution. This alternative model is much more amenable to computations, and in it, as shown in [6], the sets of graphs with 0,1 or 2 unsaturated vertices all have non-zero probability in the limit when nd is even.

The simple structure of 2-processes allows one to apply similar techniques to obtain some other results. In particular, we prove in [7] that in G^2 the number of cycles of length l is asymptotically Poisson and for $l = 3$ the mean converges to

$$\frac{1}{2}\int_0^\infty \frac{(\log(1+x))^2 dx}{xe^x} \approx 0.1887.$$

We note that the above result establishes a fundamental difference between G^2 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 [9] for example).

One natural question on which our methods do not shed light is that of finding the maximum and minimum values of

$$\Pr(G_N^d = G)$$

over all d-maximal graphs G on n vertices, or even over all d-regular graphs G on n vertices. This would give another measure of the difference between G^d and the uniform probability model for regular graphs.

2. A differential equation

In this section we analyse a differential equation and obtain facts associated with its solution, which is used for reference in Section 3.

Let

$$M = \max_v m(v).$$

Here we can assume that $M \geq 2$. Define the functions $b = b(x)$ and $q = q(x)$ by the differential equation

$$b' = \frac{-2b}{D_0/n - 2x - (M-1)b}, \qquad nb(0) = |\{v : m(v) = M\}| \tag{2.1}$$

and

$$q = \frac{2b}{D_0/n - 2x} \tag{2.2}$$

for $0 \leq x < D_0/2n$. Then substituting (2.2) into (2.1) and solving by separating variables gives

$$-\frac{2}{(M-1)q} + \log b(0) - \log q + \frac{D_0}{(M-1)nb(0)} = \log(D_0/2n - x).$$

Since

$$q' = \frac{-(M-1)q^2}{(D_0/n - 2x)(1-(M-1)q/2)}$$

and $q(0) = 2nb(0)/D_0 \leq 2/M$, q is non-increasing and thus bounded above by $2/M$. Therefore we obtain

$$q(x) \sim \frac{-2}{(M-1)\log(D_0/2n - x)} \tag{2.3}$$

as $x \to D_0/2n$. Also we now have $-2 \leq b'(x) < 0$. It is easily checked that $b''(x) > 0$, and hence

$$b(x+\epsilon) - b(x) \geq \epsilon b'(x) \geq -2\epsilon. \tag{2.4}$$

3. Proof of Theorem

We define a *full isolate* to be an isolate satisfying $m(v) = M$, where M was defined in Section 2. Let I_t stand for the number of full isolates in G_t. Our approach to the proof of Theorem 1.1 is initially to study the behaviour of I_t during the process.

Lemma 3.1. *For $j = 1$ and 2, and $0 \leq u < v \leq N_0$, let $P(j,u,v)$ be the conditional probability that in a random generalized d-process, the vertices in $[j]$ remain isolated in G_v, given that they were such in G_u. Then, allowing $u = u(n)$, $v = v(n)$,*

$$P(j,u,v) = O\left(\left(\frac{N_0 - v + 1}{N_0 - u}\right)^j\right) \quad \text{as } n \to \infty.$$

Proof. Let F'_t be the number of forbidden pairs (= elements of Φ) in G_t with at least one element in $[j]$. By the definition of a generalised d-process, F'_t is bounded. Letting $\mathscr{H}_t$ be the event that the vertices in $[j]$ are isolated in G_t, we have

$$\begin{aligned} P(j,u,v) &= \Pr(\mathscr{H}_{u+1} \wedge \ldots \wedge \mathscr{H}_v | \mathscr{H}_u) \\ &= \prod_{t=u}^{v-1} \Pr(\mathscr{H}_{t+1} | \mathscr{H}_u \wedge \ldots \wedge \mathscr{H}_t) \\ &= \prod_{t=u}^{v-1} \Pr(\mathscr{H}_{t+1} | \mathscr{H}_t) \\ &= \prod_{t=u}^{v-1} P(j,t,t+1). \end{aligned}$$

Provided $t < N_0 - O(1)$, the process cannot have become static, and so

$$\begin{aligned} P(j,t,t+1) &= \operatorname{Exp}\big(\operatorname{Exp}(\operatorname{Ind}(\mathscr{H}_{t+1})|G_t)|\mathscr{H}_t\big) \\ &= \operatorname{Exp}\left(1 - \frac{j(U_t - j) - F'_t}{A_{t+1}} \middle| \mathscr{H}_t\right) \\ &\leq \exp\left(-\frac{2j}{D_t} + O\left(\frac{1}{D_t^2}\right)\right) \end{aligned}$$

since $D_t/M \le U_t \le D_t$ by (1.4). On the other hand, if $t = N_0 - O(1)$ then $D_t = O(1)$ and so the same conclusion holds. But $D_t = D_0 - 2t \le 2N_0 + 1 - 2t$ and so

$$P(j,u,v) < \exp\left(-\sum_{t=u}^{v-1} \frac{j}{N_0 - t} + O(1)\right). \qquad \square$$

Corollary to Lemma 3.1. *If for some $r_1(n) \to \infty$ there are a.s. $o(r_1(n))$ full isolates at time $N_0 - r_1(n)$, then for some function $\sigma(n)$ tending to infinity, a.s. all full isolates have become non-isolates by time $N_0 - \sigma(n)$.*

Proof. Suppose there are a.s. at most $\tau(n)r_1$ full isolates at time $N_0 - r_1$, where $\tau(n) = o(1)$. Apply Lemma 3.1 with $j = 1$ to each of them, and sum the probabilities. This shows that the expected number of full isolates at time $N_0 - r_2$, $r_2 < r_1$, is $O(\tau(n)r_2)$. Now take $\sigma = r_2 = o(\tau(n)^{-1})$ such that $\sigma \to \infty$. $\square$

Note that the proof of the Corollary actually applies for any prescribed set of isolates, not just the full isolates. The Corollary is one of the key observations to be exploited in the proof of Theorem 1.1. The other is the next lemma, whose proof we postpone.

Lemma 3.2. *Almost surely*

$$I_t \le nb\left(\frac{t}{n}\right) + O(n^{7/8}\sqrt{\log n})$$

for all $t = 0, \ldots, N_0 - \lfloor n^{9/10} \rfloor$.

In [7] we prove that for $d = 2$, $|I_t - nb(t/n)| = O(n^{11/12}\sqrt{\log n})$ for all $t = 0, \ldots, n - \lfloor n^{47/48} \rfloor$ a.s. This fact is then used to derive some structural results on 2-processes. For $d > 2$, however, we do not believe that $nb(t/n)$ is a close approximation to I_t in general.

Proof of Theorem 1.1. This is by induction on d. Beginning with $d = 1$, we consider a random generalised 1-process. Applying Lemma 3.1 with $j = 2$ we see that a.s. for $t = N_0 - \lfloor N_0^{1/3} \rfloor$, each pair of unsaturated vertices in Φ will be intersected by at least one edge of G_t. Thus at time t the unsaturated vertices (i.e. those still of degree 0) a.s. form an independent set in Φ viewed as a graph. From such a state, the process is forced to saturate.

Now consider arbitrary $d > 1$ and apply Lemma 3.2. Setting $k = N_0 - \lfloor n^{9/10} \rfloor$, we have, by (2.3),

$$I_k \le nb(k/n) + O(n^{7/8}(\log n)^{1/2}) \sim n\left(D_0/2n - k/n\right) q(k/n) \sim \frac{20n^{9/10}}{(M-1)\log n}.$$

Thus by the Corollary to Lemma 3.1, a.s. there are no full isolates at time $t = N_0 - \sigma$ for some $\sigma \to \infty$. If this is the case, the remaining part of our random d-process can now be viewed as a random generalized $(d-1)$-process on the unsaturated vertices of G_t, with Φ being the set of all edges of G_t with both endpoints unsaturated, and with $m(v) = d - d_{G_t}(v)$ for all vertices v. Note that now $M \le d - 1$, and the initial deficit of the new process equals the current deficit of the original process at time t. By induction, this $(d-1)$-process saturates a.s., implying the theorem. □

It turns out that our method of proof actually gives a bound $O(1/\log n)$ on the probability that a random 2-process does not saturate, which from the experimental data is of the correct order. Note also that our proof actually establishes Theorem 1.1 for increasing $d = d(n)$, but for this to be true we need $d = o(\log\log\cdots\log n)$ where log is iterated $d-1$ times. We believe, however, that for large d our results are far from optimal. We conjecture that a random d-process saturates almost surely as long as $d = o(n)$ and that $1/\log n$ is the correct order of magnitude for the probability of not saturating for all bounded d.

Proof of Lemma 3.2. Let $F_t^{(l)}$ be the number of forbidden pairs containing l full isolates of G_t, $l = 1, 2$. It follows from (1.4) that for each generalised d-process

$$\frac{D_0 - 2k}{M} \le u_k \le D_0 - 2k - (M-1)i_k. \tag{3.1}$$

(Note that for $d = 2$ the above right hand side inequality becomes an equation. This is why $nb(t/n)$ approximates I_t so well for 2-processes.) We have

$$\begin{aligned}\mathrm{Exp}(I_{t+1} - I_t | G_t) &= \frac{-2I_t(U_t - 1) + 4F_t^{(2)} + 2F_t^{(1)}}{U_t(U_t-1) - 2F_t} \\ &\le \frac{-2I_t}{U_t} + \frac{2\bar{d}}{U_t - 1}\end{aligned}$$

where $\bar{d}$ is a bound on the maximum degree in Φ. Assume that $0 \le t < t_1$ and $4\bar{d} \le t_1 \le \frac{1}{4}u_k - \frac{1}{2}$. Since

$$U_t \ge U_{t+1} \ge U_t - 2, \qquad I_t \ge I_{t+1} \ge I_t - 2 \tag{3.2}$$

we have that $\mathrm{Exp}(I_{k+t+1} - I_{k+t} | G_k = g_k)$ is equal to

$$\begin{aligned}&\sum_{g_{k+t}} \Pr\,(G_{k+t} = g_{k+t} | G_k = g_k)\mathrm{Exp}(I_{k+t+1} - I_{k+t} | G_k = g_k \wedge G_{k+t} = g_{k+t}) \\ &\quad= \sum \Pr(G_{k+t} = g_{k+t} | G_k = g_k)\mathrm{Exp}(I_{k+t+1} - I_{k+t} | G_{k+t} = g_{k+t}) \\ &\quad\le \sum \Pr(G_{k+t} = g_{k+t} | G_k = g_k)\left(\frac{-2i_{k+t}}{u_{k+t}} + \frac{2\bar{d}}{u_{k+t} - 1}\right) \\ &\quad\le \frac{-2i_k + 4t_1 + 2\bar{d}}{u_k} + \frac{2\bar{d}(2t_1 + 1)}{u_k(u_k - 2t_1 - 1)} \\ &\quad\le \frac{-2i_k + 5t_1}{u_k}.\end{aligned}$$

Consequently, by (3.1),

$$\operatorname{Exp}(I_{k+t_1} - I_k | G_k = g_k) \le \frac{-2t_1 i_k + 5t_1^2}{u_k}$$

$$\text{(3.3)} \qquad \le \frac{-2t_1 i_k}{D_0 - 2k - (M-1)i_k} + \frac{5t_1^2 M}{D_0 - 2k}.$$

In order to make use of (3.3) we demonstrate sharp concentration of

$$X = I_{k+t_1} - I_k.$$

Define the Doob martingale

$$X_t = \operatorname{Exp}(X | G_k, E_{k+1}, \ldots, E_{k+t}),$$

$0 \le t \le t_1$. Then $X_0 = \operatorname{Exp}(X|G_k)$ and $X_{t_1} = X$.

We now wish to bound $|X_t - X_{t-1}|$ for $t \ge 1$. Fix a d-process $\pi_0 = (g_0, \ldots, g_{N_0})$ and write $\operatorname{Exp}(e)$ for $\operatorname{Exp}(X|G_{k+t} = g_{k+t-1} \cup \{e\})$. Then

$$\text{(3.4)} \qquad |X_t(\pi_0) - X_{t-1}(\pi_0)| \le \sum_e \Pr(E_{k+t} = e)|\operatorname{Exp}(e_{k+t}) - \operatorname{Exp}(e)|.$$

Let $S(e)$ denote the set of sequences $\vec{\eta} = (\eta_{k+t}, \ldots, \eta_{k+t_1})$ of pairs of vertices for which $\eta_{k+t} = e$. Turn $S(e)$ into a probabilistic space by assigning to each $\vec{\eta} \in S(e)$ the probability

$$\Pr_e(\vec{\eta}) = \Pr(E_{k+t+1}, \ldots, E_{k+t_1} = \eta_{k+t+1}, \ldots, \eta_{k+t_1} | G_{k+t} = g_{k+t-1} \cup \{e\}),$$

where the probability on the right is for d-processes. Fix e and e', and let S_0 (respectively S_0') denote the subset of $S(e)$ $(S(e'))$ containing those sequences $(\eta_{k+t}, \ldots, \eta_{k+t_1})$ for which none of $(\eta_{k+t+1}, \ldots, \eta_{k+t_1})$ are adjacent to either e or e'. The probability that E_j is adjacent to e or e' is $O(u_k(\pi_0)^{-1})$ for $k \le j \le k + t_1$ by (1.2) and (3.2). Thus

$$\text{(3.5)} \qquad \Pr_e(S(e) \setminus S_0) + \Pr_{e'}(S(e') \setminus S_0') = O\left(\frac{t_1}{u_k(\pi_0)}\right).$$

Next, for $\vec{\eta} \in S_0$ define

$$\sigma(\vec{\eta}) = (e', \eta_{k+t+1}, \ldots, \eta_{k+t_1}) \in S_0'.$$

We say that a generalised d-process is *consistent* with $\vec{\eta}$ if $G_{k+t-1} = g_{k+t-1}$ and $E_{k+t}, \ldots, E_{k+t_1} = \eta_{k+t}, \ldots, \eta_{k+t_1}$. Let π and π' be consistent with $\vec{\eta}$ and $\sigma(\vec{\eta})$ respectively. Then for $k + t + 1 \le j \le k + t_1$, we have

$$|u_j(\pi) - u_j(\pi')| \le 2$$

and hence

$$\text{(3.6)} \qquad \frac{\Pr_e(\vec{\eta})}{\Pr_{e'}(\sigma(\vec{\eta}))} = 1 + O\left(\frac{t_1}{u_k(\pi_0)}\right).$$

Define $X(\tilde{\eta}) = X(\pi)$ for any π consistent with $\tilde{\eta}$. We now have

$$\begin{aligned}
|\mathrm{Exp}(e) - \mathrm{Exp}(e')| &= \left| \sum_{\tilde{\eta}\in S(e)} \mathrm{Pr}_e(\tilde{\eta})X(\tilde{\eta}) - \sum_{\tilde{\eta}\in S(e')} \mathrm{Pr}_{e'}(\tilde{\eta})X(\tilde{\eta}) \right| \\
&\le \left| \sum_{\tilde{\eta}\in S_0} \mathrm{Pr}_e(\tilde{\eta})X(\tilde{\eta}) - \mathrm{Pr}_{e'}(\sigma(\tilde{\eta}))X(\sigma(\tilde{\eta})) \right| \\
&\quad + (\mathrm{Pr}_e(S(e)\setminus S_0) + \mathrm{Pr}_{e'}(S(e')\setminus S_0')) \max_{\tilde{\eta}\in S(e)\cup S(e')} X(\tilde{\eta}) \\
&\le \sum_{\tilde{\eta}\in S_0} \Big(|X(\tilde{\eta}) - X(\sigma(\tilde{\eta}))| \mathrm{Pr}_e(\tilde{\eta}) \\
&\quad + X(\sigma(\tilde{\eta}))|\mathrm{Pr}_e(\tilde{\eta}) - \mathrm{Pr}_{e'}(\sigma(\tilde{\eta}))| \Big) + O\left(\frac{t_1^2}{u_k}\right) \\
&\le 2 + O\left(\frac{t_1^2}{D_0 - 2k}\right).
\end{aligned}$$

Here we used (3.5), (3.6), (3.1) and the facts that $|X(\tilde{\eta}) - X(\sigma(\tilde{\eta}))| \le 2$ for $\tilde{\eta} \in S_0$ and

$$\max_{\tilde{\eta}\in S(e)\cup S(e')} X(\tilde{\eta}) \le 2t_1$$

by (3.2).

It now follows from (3.4) and Azuma's inequality (see for example Bollobás [3] or McDiarmid [4]) that

$$\mathrm{Pr}\left(|X_{t_1} - X_0| \ge C\sqrt{2ct_1 \log n} \right) < n^{-c}$$

for any $c > 0$, where $C = 2 + O(t_1^2/(D_0 - 2k))$. If $t_1^2 = o(N_0 - k)$ then $C < 3$ and hence

$$\mathrm{Pr}\left(|I_{k+t_1} - I_k - \mathrm{Exp}(I_{k+t_1} - I_k | G_k)| \ge \sqrt{18ct_1 \log n} \right) < n^{-c} \tag{3.7}$$

for any $c > 0$. Thus, in view of (3.3), the function $b = b(x)$ as defined in (2.1) should be an approximate upper bound for I_k/n, where $x = k/n$.

To justify this approximation we will partition the interval $[0, N_0]$ as follows. Let $k_j = j\Delta$, $\Delta = \lfloor n^{1/4} \rfloor$, $j = 0, \ldots, s$, where s is chosen so that $|k_s - (N_0 - n^{9/10})| = O(n^{1/4})$. Note that $s < n^{3/4}$. Now we shall prove by induction on j that

$$\mathrm{Pr}(I_{k_j} \le \beta(k_j) + 2\Delta + jR) = 1 - O(jn^{-c}) \tag{3.8}$$

where $\beta(k) = nb(k/n)$ and $R = 5n^{-2/5} + \sqrt{18c\Delta \log n}$ and c is as in (3.7). This is trivially true for $j = 0$.

Throughout the induction we can regard n as being fixed. Let us fix j and set $k = k_j$, and assume that (3.8) holds. Noting that $k_{j+1} = k + \Delta$, let $S = I_{k+\Delta} - \beta(k+\Delta) = -T_1 - T_2 - T_3$, where

$$\begin{aligned}
T_1 &= \beta(k+\Delta) - \beta(k) + \frac{2I_k\Delta}{D_0 - 2k - (M-1)I_k}, \\
T_2 &= \beta(k) - I_k, \\
T_3 &= I_{k+\Delta} - I_k + \frac{2I_k\Delta}{D_0 - 2k - (M-1)I_k}.
\end{aligned}$$

Define the events

$$\begin{aligned} \mathscr{A} &: S \le 2\Delta + (j+1)R, \\ \mathscr{A}_1 &: -T_2 \le 2\Delta + jR, \\ \mathscr{A}_2 &: T_3 \le R, \\ \mathscr{B} &: I_k > \beta(k). \end{aligned}$$

By the inductive hypothesis,

$$\Pr(\mathscr{A}_1) = 1 - O(jn^{-c}).$$

Observe that, given $\mathscr{B}$,

$$\begin{aligned} T_1 &\ge \beta(k+\Delta) - \beta(k) + \frac{2\beta(k)\Delta}{D_0 - 2k - (M-1)\beta(k)} \\ &= \beta(k+\Delta) - \beta(k) - b'(k/n)\Delta \\ &= 0 \end{aligned}$$

by (2.1), and (2.4) with $\epsilon = \Delta/n$. Thus $\mathscr{A}_1 \wedge \mathscr{A}_2 \wedge \mathscr{B}$ implies $\mathscr{A} \wedge \mathscr{B}$.

Note also that the complement of $\mathscr{B}$ implies

$$I_{k+\Delta} \le I_k \le \beta(k) \le \beta(k+\Delta) + 2\Delta,$$

which gives $\mathscr{A}$.

Hence,

$$\begin{aligned} \Pr(\mathscr{A}) &= 1 - \Pr(\mathscr{B}) + \Pr(\mathscr{A} \wedge \mathscr{B}) \\ &\ge 1 - \Pr(\mathscr{B}) + \Pr(\mathscr{A}_1 \wedge \mathscr{A}_2 \wedge \mathscr{B}) \\ &\ge -\Pr(\mathscr{B}) + \Pr(\mathscr{A}_1) + \Pr(\mathscr{A}_2 \wedge \mathscr{B}) \\ &= 1 + \Pr(\mathscr{A}_2 \wedge \mathscr{B}) - \Pr(\mathscr{B}) - O(jn^{-c}) \\ &= 1 - O((j+1)n^{-c}) \end{aligned}$$

where the last step follows from $\Pr(\mathscr{A}_2) = 1 - O(n^{-c})$, which is true by (3.3) and (3.7) with $t_1 = \Delta$. This completes the inductive proof of (3.8) and Lemma 3.2 follows, since neither I_k nor $nb(k/n)$ can change by more than 2 on passing from k to $k+1$. □

4. Open Problems

1 How big can $d = d(n)$ be so that a random d-process still saturates almost surely? We think that it does so as long as $d = o(n)$ and that for $d = cn$ the limit probability of saturating depends on the constant c.

2 For fixed d, what is the rate of decay of the probability of not saturating? We are convinced that $\frac{1}{\log n}$ is correct regardless of the value of d.

3 For $d \ge 3$, is there a function $f(x)$ which approximates I_t/n? We know that $b(x)$ is good for $d = 2$ but we doubt if it is for higher d.

4 When does the last vertex of degree k, $0 \leq k \leq d-2$, disappear and what is the largest number of vertices of degree k, $1 \leq k \leq d-1$, present at one time during the process? Only the case $d = 2$ is within our grasp and the key is again the function $b(x)$.

5 For $d \geq 3$, does the random d-process almost surely result in a connected graph? We conjecture that it does. In [7] we show that this is false for $d = 2$, by studying the distribution of cycles. However, we can still only conjecture that the random 2-process a.s. results in a disconnected graph.

References

[1] Balińska, K. T. and Quintas, L. V. (1990) The sequential generation of random f-graphs. Line maximal 2-, 3-, and 4-graphs, *Computers & Chemistry* **14** 323-328.

[2] Bollobás, B. (1985) *Random graphs*, Academic Press, London.

[3] Bollobás, B. (1990) Sharp concentration of measure phenomena in the theory of random graphs, *Proc. of Random Gaphs '87*, Karoński et al. eds., Wiley.

[4] McDiarmid, C. (1989) On the method of bounded differences, *Surveys in Combinatorics 1989* (invited papers of the 12th British Combinatorial Conference), J. Siemons ed. 148-188.

[5] McKay, B. D. and Wormald, N. C. (1991) Uniform generation of random regular graphs of moderate degree, *J. Algorithms* **11**, 52-67.

[6] Ruciński, A. (1990) Maximal graphs with bounded maximum degree: Structure, asymptotic enumeration, randomness, *Proc. III of the 7th Fischland Colloquium, Rostock Math. Kolloq.* **41** 47-58.

[7] Ruciński, A. and Wormald, N. C. (in preparation) A nontrivial random 2-regular graph.

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

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