The differential equation method for random graph processes and greedy algorithms*

N. C. Wormald†
Department of Mathematics
University of Melbourne
Parkville, VIC 3052, Australia

Contents

*Notes on lectures given at the Summer School on Randomized Algorithms at Antonin, Poland in 1997.
†Supported in part by the Australian Research Council. Email: nick@ms.unimelb.edu.au

1 Introduction

Random graph processes and related discrete random processes are being used increasingly in the analysis of randomised algorithms and the study of random graphs. Some of the important results have recently been obtained by establishing a connection between the process and an associated differential equation[1] or system of differential equations. The solution of the differential equations provides a deterministic approximation for the dynamics of the random process.

This idea of approximation has existed in connection with continuous processes (essentially since the invention of differential equations by Newton for approximation of the motion of bodies in mechanics) before being applied to random graphs, and some results for discrete processes also appeared before. For instance, Kurtz's theorem [31] is applicable to discrete processes. It was used in the analysis of a random greedy matching algorithm by Karp and Sipser [29], which was the first application of this method to random graphs.

This paper is partly a survey of applications of a method or approach, but mainly an exposition of the method and what we can expect from it. The main theorem (Theorem 5.1) attempts to be general-purpose. The general setting is a sequence of random processes indexed by n (which is often the number of vertices in the initial graph of the process), and the aim is to find properties of the random process in the limit as $n \to \infty$. In general, the conclusion one draws after applying the method is that variables defined on a random process are *sharply concentrated*, which informally means that they are asymptotically equal to certain deterministic functions with probability tending to 1 (as $n \to \infty$). These deterministic functions arise as the solutions to a system of ordinary first order differential equations. For some random graph applications the situation is rather delicate and higher accuracy in the approximation is crucial (see Sections 5.2 and 7.2). Higher accuracy can usually be obtained by using ideas from the proof of the main theorem or associated results such as Corollary 4.1.

One of the important features of this approach is that the computation of the approximate behaviour of processes is clearly separated from the proof that the approximation is correct. A good example of this is in Section 3.3.4. In addition, the solution of the differential equation in Section 7.2 gives a suggestion of the crucial point in time at which a random greedy packing algorithm will come to a grinding halt. This feature is hard to glean from other approaches to this problem which use the nibble method of Rödl [48].

This paper gives examples of the various types of results, categorised according to which version of the method is required to obtain them. Some new results are included, in particular, the first application of the differential equation method to packing in hypergraphs in Section 7.2. Also, the main theorem is a little stronger than that in [61], although the proof is almost identical, and the derivation in Section 3.3.3 is new, as is the application to the

[1] What is a title with the first two words "differential equations" doing on a paper which is mainly about graphs? Any reader feeling uncomfortable about this is assured that no real method of solving differential equations is required to read this article except for simple first order ordinary differential equations, and even those only in one or two places. More importantly, the same reader is encouraged to examine the interplay between discrete and continuous mathematics more closely, and to note that obscure methods are not a prerequisite of obtaining interesting results in this area.

process in Section 3.1 and 3.2.

Differential equations have of course turned up in studies of random graphs in other contexts. One appearance particularly related to the present topic occurs in the study by Pittel [43] of the limiting distribution of the number Y_i of tree components of size i in the random graph $\mathcal{G}(n, m)$. A Gaussian distribution result is derived there by considering the dynamics of the process induced on the Y_i by adding a random edge to the graph. A system of ordinary differential equations describes the behaviour of other variables from which the variance and covariance of the Y_i can be estimated. Another example is in Bender et al. [10] which also has $\mathcal{G}(n, m)$ in the background, but there the object of study is the probability of connectedness of the random graph. In that case, the differential equation also gives an indication of trends in the process, but is not used in the same way as in the present notes.

There are naturally some random graph processes to which the differential equation approach has not been applied at all so far. The tree-growing processes such as that studied by Mahmoud et al. [39] provide a good example.

1.1 A brief look at the general method

The basic idea is quite simple. Compute the expected changes in random variables of the process per unit time at time t and, regarding the variables as continuous, write down the differential equations suggested by the expected changes. Then use large deviation theorems to show that with high probability the solution of the differential equations is close to the values of the variables.

As mentioned above, this approach is standard in the study of continuous time processes, such as in the book by Kurtz [32]. (There the main emphasis is convergence in distribution, rather than bounding the probability of very rare events which is regularly required for combinatorial applications.) Conversion to discrete can be done by simple relations between the discrete time and continuous time processes (see Kurtz [32, Proposition 4.5]). Basically, this is because the number of steps which the continuous time random process has taken after a given (long) time is sharply concentrated. More recently, Aldous [2] examined the generalisation of some of the results on the emergence of the giant component in random graphs in the setting of continuous processes. A differential equation result on that type of problem is mentioned in [2, Section 5.2]. Those interested in working with integration and measures may prefer the continuous version of this method. For combinatorialists, especially those familiar with sharp concentration phenomena in the discrete setting, working entirely with discrete processes is no doubt easier.

The theorem in Kurtz [31] used in [29] is stated for discrete processes. Being quite general, it is similar in spirit to Theorem 5.1, but only gives $o(1)$ type bounds on the errors in probability, whereas combinatorial applications often require bounds of the form $O(n^{-C})$ (for some particular constant C).

One of the distinguishing features of the applications considered here, to graph processes, compared to many other applications of differential equations to random processes, is that the vector processes considered are quite often not Markovian. However, they do need to be *close* to Markovian for the method to work, and fortunately this is often the case.

1.2 Graph processes

Most of the random processes we consider here are graph processes. By this we mean a process which begins with some starting graph (usually n isolated vertices) and repeatedly adds edges randomly according to some stochastic rule. (Equivalently, there is a known probability distribution for the next edge or set of edges to be added.) At some point, when there are no more edges which can legally be added, the process becomes constant. This determines the probability space of sequences $G_0, G_1, \ldots$ where G_t is the t^{th} graph in the process. If at time t no more edges can be added, then $G_{i+1} = G_i$ for all $i \geq t$ and we call G_t the *final graph* of the process. We will call the first such t the *natural stopping time* of the process.

As an alternative, the starting graph G_0 can be any given graph and during the process the edges can be deleted at random. The natural stopping time of such a process is again when it becomes constant. Another process will use hypergraphs rather than graphs.

For all our graph processes, G_t will denote the evolving graph (or hypergraph) at time t. Throughout, we use $Y_i(t)$ to denote the number of vertices of degree i in G_t, and with reference to any process in which edges are added, those edge(s) added at time t are denoted by A_t. In general A_t is permitted to be a set of edges. After the natural stopping time, A_t will be empty. We restrict the discussion here to processes that create no loops or multiple edges. It follows that the natural stopping time is finite in each case. If it is fixed at some integer T, we may write $G_0, G_1, \ldots = G_0, G_1, \ldots, G_T$. We will be interested in the behaviour of the process as $n \to \infty$. Thus, we actually consider a sequence of random processes, each with starting graphs of different sizes. The behaviour of variables (such as $Y_i(t)$ which now depends on n) can in many cases be approximated well as $n \to \infty$, and this is the type of result we aim for.

It turns out that with most random processes like this, it is hard to tell what the probability distribution of the final graph is. One exception to this is also perhaps the simplest, called the standard random graph process in Section 2. Another is a process which generates regular graphs uniformly at random. But in general, the lack of knowledge about the precise distribution of the process means that some non-elementary method such as that presented here will be crucial in determining the asymptotic behaviour of some features of the process, and, in particular, properties of the final graph. Even when the distribution of the final graph is "well known", such as with random regular graphs, we can obtain results which have not been obtained by other means by considering the random process generating the graphs. Examples of this occur in [42] and [61].

1.3 Basic notation

1.3.1 Graph Theory

$V(G)$ and $E(G)$ are the vertex set and edge set, respectively, of a graph G. Unless otherwise specified, $V(G) = [n] = \{1, 2, \ldots, n\}$. We use $d(v)$ for a vertex v to mean its degree (usually the graph is understood from context, but if not we may write $d_G(v)$). $\Delta(G)$ denotes the

maximum degree of a vertex of G, and $\delta(G)$ the minimum. There should be no confusion between this and the Kronecker delta function

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

The complete graph on n vertices is as usual denoted by K_n, and the complement of a graph G is denoted by $\overline{G}$.

1.3.2 Probability

We use the notation $\mathbf{P}$ (probability), $\mathbf{E}$ (expectation), $\mathbf{Var}$ (variance) and u.a.r. (uniformly at random). Also, an event $B = B_n$ holds a.a.s. (asymptotically almost surely) if $\lim_{n\to\infty} \mathbf{P}(B_n) = 1$.

$\mathbf{Po}(\lambda)$ denotes the Poisson distribution with expectation λ. Thus, if X has distribution $\mathbf{Po}(\lambda)$ then $\mathbf{P}(X = i) = \lambda^i e^{-i}/i!$ for $i \geq 0$.

For probabilistic methods used in combinatorics, see [4].

1.3.3 Other

For any integer $k \geq 0$,

$$[x]_k = x(x-1)\cdots(x-k+1).$$

The sequence $x_0, x_1, \ldots$ is denoted by $\langle x_i \rangle$.

2 Some random processes and their histories

This is a description of the random processes which are investigated in detail in these notes, together with some background information. These are the ones with which I am most familiar and so represent a biased selection of the processes to which the differential equation method has been profitably applied. Other such processes are mentioned in Section 8.

The first is the classic random graph process, which is mainly included for comparison.

Standard random graph process.

Here $G_0 = \overline{K_n}$, and choose A_t at random from the remaining non-adjacent pairs of vertices. This stops when $t = \binom{n}{2}$. This is a process such that the m'th graph in the sequence is distributed as in $\mathcal{G}(n, m)$, the probability space of (n, m)-graphs with uniform distribution.

Min-degree graph process.

This process does not seem to have appeared elsewhere, and is mainly included to provide a simple but non-trivial example of the use of the differential equation method. $G_0 = \overline{K_n}$. Given G_t, choose a vertex u of minimum degree in G_t u.a.r., then a vertex v

not adjacent to u in G_t u.a.r. Put $A_{t+1} = uv$. The process finishes with the complete graph.

As an example, if G_4 is the example in Figure 1 and $d = 2$, then $\mathbf{P}(A_5 = \{6, 2\}) = 1/16$ since u cannot be 2, $\mathbf{P}(u = 6) = 1/4$, and $\mathbf{P}(v = 2 \mid u = 6) = 1/4$. On the other hand, $\mathbf{P}(A_5 = \{6, 1\}) = 1/8$, since u can be either 1 or 6.

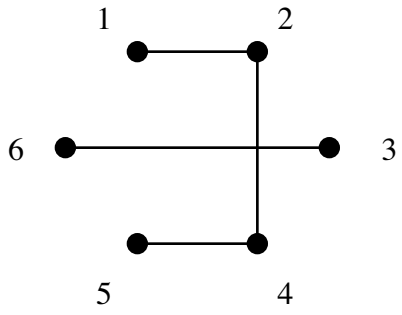

Figure 1: *A graph in a random process*

Degree bounded graph process.

This process has a parameter d, where $d \geq 2$ for non-triviality. $G_0 = \overline{K_n}$. Given G_t, choose u.a.r. a pair of non-adjacent vertices which both have degree strictly less than d. Put $A_{t+1} = uv$. As in [50], for a given d, we call this simply the d-process. The process stops when the graph induced by the vertices of degree less than d is a clique. For example, if G_4 is the example in Figure 1 with $d = 2$, then the next edge added is $\{6, 1\}$ with probability $1/5$ since there are five vacant sites for edges to join two vertices of degree less than 2.

This process has received a lot of attention (Ruciński and Wormald [50, 51] and various papers by Balińska and Quintas, for example [7]), due to an interesting question of Erdős, asking for the asymptotic distribution of the number of vertices of degree less than d in the final graph. This question was settled in [50] using the differential equation approach together with some other arguments. It was shown that a.a.s. the final graph is regular if dn is even, and almost regular, with one vertex of degree $d-1$ and the rest of degree d, otherwise. Such a graph we call *d-saturated*.

Degree bounded star graph processes.

The degree bounded star graph process was introduced by Robalewska [45] (see [46, 47]). It also has a parameter d which determines the maximum degree of the vertices in the graphs generated. It can begin with $G_0 = \overline{K_n}$. Here several edges are added in each step of the process, the idea being that a natural and quick way to generate a d-regular graph is to "fill" the vertices one after another. Given G_t, choose u.a.r. a vertex v of

minimum degree in G_t, and $d - d(v)$ other randomly chosen vertices of degree strictly less than d. The edges from v to these vertices form a star, in fact a $(d - d(v))$-star. G_{t+1} is formed by adding the edges of this star to G_t. For given d we call this simply the star d-process. In this process, up to d edges are added in each step, and so the number of edges in G_t can be as large as td. This process stops when G_t is d-regular or, for some v of minimum degree, there are less than $d - d(v)$ other vertices of degree strictly less than d.

This process is useful for generating random regular graphs, but the distribution is not controlled. Other related random graph generation algorithms were given by Tinhofer [57]. These processes could also be analysed using the differential equation approach.

Process generating random regular graphs

What is the size of the largest independent set (also called stable set) in a d-regular graph? Finding the answer requires solving an NP-complete problem. Here, we are interested in this question when the graphs are chosen u.a.r. Lower bounds on this number were obtained in [61] by studying the standard model (called the configuration or pairing model) for uniformly generating random regular graphs. In this way, the performance of algorithms for finding large independent sets in regular graphs can be studied.

The model (see Bollobás [12] for example) can be described as follows. Take dn points in n buckets labelled $1, 2, \ldots, n$, with d in each bucket, and choose u.a.r. a *pairing* $P = p_1, \ldots, p_{dn/2}$ of the points such that each p_i is an unordered pair of points, and each point is in precisely one pair p_i. If no pair contains two points in the same bucket, and no two pairs contain four points from just two buckets, then we can produce a d-regular graph by joining two distinct vertices i and j if some pair has a point in bucket i and one in bucket j. The d-regular graphs on n vertices are all produced with equal probabilities. In any discussion of this model and algorithms on d-regular graphs in this paper, we assume that dn is even to avoid parity problems.

We redefine this process slightly by specifying that the pairs are chosen sequentially. Actually it is useful to consider a whole variety of processes, each of which produces a pairing chosen u.a.r. These arise from the fact that in choosing the pairs in the random pairing sequentially, the first point in the next random pair can be selected using any rule whatsoever, as long as the second is chosen u.a.r. from the remaining points. For example, one can insist that the next point chosen comes from the lowest-labelled bucket (i.e. vertex) available, or from the vertex containing one of the points in the previous completed pair (if any such ponts are still unpaired). As a consequence of this, for any algorithm being applied to the final random graph, the process for generating the pairs can be determined by the order in which the algorithm queries the edges of the graph, and this order can be defined dynamically. Two special cases of such a process will be treated, arising from two algorithms for finding independent sets of vertices. This was done in [61] to obtain lower bounds on the expected size of a maximal independent set

in a random d-regular graph. One of these algorithms, the *degree-greedy* algorithm, was also studied by Frieze and Suen [21] in the case $d = 3$ (and called MINGREEDY). Their argument relies on a result from Frieze et al. [20] which used a differential equation approach to approximate the number of vertices of degree 0 in the random process.

An approximation algorithm for another graph function, the dominating number, can be studied in the same way for random regular graphs (see Section 8). Undoubtedly, there are still more that can be done in this way.

One can also ask for properties of the evolving graph during the generation of a random regular graph by the method above, if the pairs are chosen u.a.r. at each step. After t pairs have been chosen, plenty of properties can be deduced by direct computation, because all sets of t non-intersecting (i.e. pairwise disjoint) pairs are equally likely. However, for some of the more recalcitrant questions, the differential equation approach may help. Molloy and Reed [42] have examined the size of the largest component in the evolving graph in this way.

Finding the k-core of a random graph

The *k-core* of a graph G is the unique maximal subgraph of minimum degree at least k. We discuss here only the case $k \geq 3$; most aspects of the case $k = 2$ are much simpler. To find the k-core of G one can begin by deleting all the vertices of degree less than k in G. During these deletions, the degrees of some other vertices can fall below k, which are then deleted in a second round, and so on. This is continued until no vertices of degree less than k remain. The resulting graph is the k-core of G. The process discussed here is a slowed-down version of this: beginning with G, choose a random vertex of degree less than k, delete it, and then repeat.

In [44] this random process was applied to a random graph with m edges, to show that a.a.s. in $\mathcal{G}(n, m)$ the k-core first appears at time $m \sim c_1(k)n$ and with size (number of vertices) $\sim c_2(k)n$. Both c_1 and c_2 were determined, as well as a.a.s. the asymptotic size of the k-core for larger m's. The success of this work lay in the choice of the process. The mass deletion process mentioned above takes steps which are seemingly too large to analyse precisely in the limit (though bounds on the threshold of appearance of the k-core were previously obtained in just this way — see Molloy [40]). Other processes, such as the even more slowed-down process of randomly deleting a random edge incident with a vertex of degree less than k, were also considered but don't seem amenable to analysis. An important feature of the chosen process is that at each stage the graph remains random subject to a simple set of random variables which also determine the transition probabilities at each step. No other simple, slow process with this feature has been found for this problem. This was one of the recent big successes of the differential equation method, so quite a lot of attention is devoted to it in these notes.

Greedy packing process

The starting object of this process is some k-uniform hypergraph G_0, so all hyperedges of G_0 contain exactly k vertices. We assume $k \geq 3$. G_{t+1} comes from G_t by selecting

one hyperedge u.a.r., marking its vertices, and deleting the hyperedge as well as all the hyperedges intersecting it. The process stops when no hyperedges remain. The main item of interest is the number of vertices remaining unmarked when the process finishes. This is called *greedy packing* because it greedily selects a set of mutually non-intersecting hyperedges (packing of hyperedges). Hence, unmarked vertices are called *uncovered*. For a good non-trivial example, take G_0 to be the hypergraph whose vertices are the *edges* of the complete graph K_n, and whose edges are the *triangles* of K_n. Then the process just repeatedly deletes triangles from K_n until no triangles remain, and the question is how many edges remain. Joel Spencer has offered \$200 for a proof that the answer is $n^{3/2+o(1)}$ a.a.s.

The interest in this type of question stems from Rödl's solution [48] of an old conjecture of Erdős and Hanani on existence of near-perfect Steiner systems. This was the origin of Rödl's "nibble" method, which was used by several authors to obtain results showing the existence of packings which cover almost all the vertices of hypergraphs. For a *simple* hypergraph; that is, one in which any two vertices are in at most one hyperedge, it can be extended further. The best known result is the recent theorem of Alon et al. [3] that a simple k-uniform, D-regular hypergraph on N vertices has a packing which covers all vertices but at most $O(ND^{-1/(k-1)})$ if $k > 3$ and $O(ND^{-1/2}\ln^{3/2} D)$ if $k = 3$.

However, it is not known how good a packing the greedy algorithm produces. Rödl and Thoma [49], and Spencer [55] show that if the degrees of G_0 are all $D(1 + o(1))$ with $D \to \infty$ and the co-degrees (numbers of common hyperedges in which the pairs of vertices lie) are all $o(D)$ then the greedy algorithm almost surely leaves $o(|V(G)|)$ vertices uncovered. Grable [24] recently improved this to $n^{7/4+o(1)}$ for the above example of triangles in K_n (in which $N = \binom{n}{2}$) and states that his argument generalises to the bound $ND^{-1/2(k-1)+o(1)}$ for k-uniform D-regular simple hypergraphs on N vertices.

Thus far, all the results obtaining upper bounds stronger than $o(N)$ on the number of uncovered vertices assume the initial hypergraph G_0 is simple and regular. It was only very recently that the result of the greedy algorithm was shown to be below $N^{1-\epsilon}$ for some $\epsilon > 0$ [24].

The processes we study here also assume that G_0 is simple, as well as some other conditions on the degrees of the vertices which allows them to vary a little. The conclusion is that the greedy algorithm a.a.s. leaves $o(|V(G)|^{1-\epsilon})$ vertices uncovered. This proof is not very complicated and the value of ϵ which results can be improved with a little work (but some non-trivial modification would be required to equal or better Grable's result).

The approach taken here with greedy packing and differential equations has a lot in common with the Rödl nibble method and adaptations such as used by Grable [24]. In the original nibble method (as applied to this problem), a small number of hyperedges are selected at random, clashing ones (which intersect) are thrown away or otherwise dealt with so that the degrees in the hypergraph do not differ too much, and then this is repeated. In [49] and [24] the greedy algorithm is analysed in roughly this

way. In the present article there is a step in the proof of Theorem 5.1 which uses a supermartingale spanning a number of steps; this is an analogue of a nibble, and the idea is repeated in proving Theorem 7.1 on the packing process. On the other hand, the differential equation method has no need to take nibbles. This is done in Theorem 5.1 for convenience of obtaining a general theorem. Swallowing the process whole tends to give stronger results, as shown in Section 5.2. The relationship between the two approaches is not yet fully explained.

3 Preliminary investigations

In this section we examine some of the processes defined above and make some conclusions based on the non-rigorous assumption that everything goes more or less at the expected rate. The sharp concentration of the variables which is required in these assumptions is discussed rigorously in Section 5.3.

3.1 Min-degree graph process: phase 0

In order to study the structure of the evolving graph G_t in the min-degree graph process, a first step is to find the expected number of vertices of given degree. The process goes through various phases: when the minimum degree is k, the vertices of degree k disappear at a higher rate than they do at any other time. Let us say that at time t the process is in phase k if $\delta(G_t) = k$. That is, recalling $Y_i(t)$ denotes the number of vertices of degree i in G_t, k is minimum such that $Y_k(t) > 0$.

Phase 0 is the easiest to analyse, so this is taken separately in this discussion as a simple example. During this phase, for each new edge A_{t+1} the first vertex u has degree 0, and the second v has

$$\mathbf{P}(d(v) = i) = \frac{Y_i(t)}{n-1} \tag{3.1}$$

since v is free to be any vertex apart from u.

As an aside, note that the vector process $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_{n-1})$ is Markovian; that is, its distribution at time $t+1$ is determined by its value at time t independently of its earlier history. This suggests a quite valid next step in the analysis, namely to compute all the transition probabilities in the Markov chain. This is a straightforward approach, but in many cases it leads to rather complicated formulae before an almost miraculous simplification occurs when computing the expected change in the variables. Instead of taking this course, we prefer to work out the expected changes by considering the contributions from the various possible sources, thereby avoiding the sometimes complicated intermediate formulae.

When the edge A_{t+1} is added, the change in Y_i has contributions from two sources; namely, the change in degree of u and the change in degree of v. The former only affects Y_0 and Y_1, since u changes from degree 0 to degree 1. The latter can affect Y_i in two ways: it can change from $i-1$ to i, increasing Y_i by 1, or from i to $i+1$, decreasing Y_i by 1. Separate these

possibilities by defining indicator variables X_i, for the event $d(v) = i$. Then

$$\begin{aligned} Y_0(t+1) &= Y_0(t) - 1 - X_0, \\ Y_1(t+1) &= Y_1(t) + 1 + X_0 - X_1, \\ Y_i(t+1) &= Y_i(t) + X_{i-1} - X_i \qquad \text{for } i \geq 2. \end{aligned}$$

In other words,

$$Y_i(t+1) = Y_i(t) - \delta_{i0} + \delta_{i1} + X_{i-1} - X_i.$$

Taking expectations, we have by linearity of expectation

$$\begin{aligned} \mathbf{E}(Y_i(t+1) - Y_i(t) \mid G_t) &= -\delta_{i0} + \delta_{i1} + \mathbf{E}X_{i-1} - \mathbf{E}X_i \\ &= -\delta_{i0} + \delta_{i1} + \frac{Y_{i-1}(t) - Y_i(t)}{n-1} \end{aligned} \tag{3.2}$$

in view of (3.1), where $Y_{-1}(t)$ is taken to be 0.

If we now imagine interpolating these variables Y_i, which are defined only at the non-negative integers, by real functions, and *presume* that the changes in the functions are equal to their expected changes, we obtain a system of *differential* equations for $Y_i(t)$, $(i = 0, 1, \ldots)$:

$$Y_i'(t) = -\delta_{i0} + \delta_{i1} + \frac{Y_{i-1}(t) - Y_i(t)}{n-1} \qquad (i \geq 0).$$

It should be emphasised that these differential equations are only *suggested* (at this stage). However, the steps of the process are almost independent, in a sense to be exploited in Section 5.1, where it is shown that the Y_i are indeed concentrated near the values suggested by the solution of these equations.

The nature of the limiting behaviour as $n \to \infty$ can be emphasised by considering scaled versions of the variables which approach fixed functions in the limit. For $i \geq 0$, we define a real function $z_i(x)$ to model the behaviour of $\frac{1}{n}Y_i(xn)$. Then, since $n-1 \approx n$ in the limit, the above differential equations become

$$z_i'(x) = -\delta_{i0} + \delta_{i1} + z_{i-1}(x) - z_i(x) \qquad (i \geq 0) \tag{3.3}$$

where $z_{-1}(t) = 0$ for all t. The initial conditions are $z_0(0) = 1$ and $z_i(0) = 0$ for $i > 0$.

These equations are easily solved one by one, beginning with $z_0'(x) = -1 - z_0(x)$, $z_0(0) = 1$. This is first-order linear; the solution is

$$z_0(x) = 2e^{-x} - 1. \tag{3.4}$$

From here we find $z_1(x) = 2xe^{-x}$, and then in general for $i > 0$

$$z_i(x) = \frac{2x^i}{i!e^x}. \tag{3.5}$$

We will show that (3.4) and (3.5) represent the "shape" of a typical process.

Theorem 3.1 *Take any function $a = a(n)$ for which $1 \le a \le n$. For λ and s satisfying $\lambda = o(\ln 2 - s)$, with probability $1 - O(\frac{a}{\lambda} e^{-n\lambda^3/8})$*

$$Y_l(t) = n z_l(t/n) + O(\lambda n)$$

uniformly for all $0 \le l \le a-1$ and $0 \le t \le sn$.

The proof is in Section 5.3.

By taking for example $a = n$, $\lambda = n^{-1/4}$ and $s = \ln 2 - n^{-1/5}$, we obtain very sharp concentration of all the $Y_i(t)$ near $n z_i(xn)$ until t is within $n^{4/5}$ of $n \ln 2$. From here, since Y_0 decreases in each step by at least 1, Y_0 inevitably reaches 0 within another $O(n^{4/5})$ steps, that is, with $t \sim n \ln 2$, at which point the evolving graph has approximately $n \ln 2$ edges. (The same conclusion comes by applying the theorem with $a = 1$.) When Y_0 reaches 0, phase 1 finishes and phase 2 begins.

3.2 Min-degree graph process: later phases

A similar analysis applies to phase k for $k = 1, 2, \ldots$. There is a complicating factor here however, that there may be edges already present between vertices of degree k and other candidate vertices to join to. We call these *nuisance* edges. Note that one cannot determine, merely from the degree sequence of G_t, the numbers of nuisance edges between the vertices of various degrees. On the other hand, the previous history of the process affects the distribution of these numbers, which are random variables. As a result of this, the vector process $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_{n-1})$ is not Markovian. (This is not a proof, but this statement is easily verified even in the case $k = 1$.)

However, until the graph G_t is fairly dense, these nuisance edges should be quite rare, and so should not affect the general trend. So, for now, we ignore them. Within each phase, the assumption that the random variables Y_i behave as expected leads again to the equations (3.3), but with two differences. This first is that the equations now only apply for Y_i with $i \ge k$, $-\delta_{i0} + \delta_{i1}$ becomes $-\delta_{ik} + \delta_{i(k+1)}$ and Y_{k-1} is taken as 0. The second is that the initial conditions are determined by the previous phase, for which we can take the solutions of the differential equations (3.3). This leads to a rather complicated set of functions z_i, even for phase 1 where the initial conditions are given by (3.5) with $x = \ln 2$. However, the solution can be obtained recursively, and it can be verified that z_i in phase k can be written as $P_{k,i} e^{-x} - \delta_{ik}$ where the $P_{k,i}$ are polynomials in x with $P_{k,i+1} = \int P_{k,i}\, dx$, the constant of integration determined by the initial conditions at the beginning of phase k. For example, in phase 1, $z_1 = 2(1+\ln 2)e^{-x} - 1$. The end of phase 1 is thus represented by $x = \ln 2 + \ln(1+\ln 2)$, so we expect the evolving graph to have approximately $(\ln 2 + \ln(1+\ln 2))n$ edges when phase 2 begins. These results about the process can easily be made precise like Theorem 3.1, and proved inductively phase by phase.

3.3 Degree bounded graph process

Given d, this process (the d-process) adds edges randomly subject to the degree of vertices not exceeding d. Here we take d fixed. Unlike the min-degree graph process, the degree

bounded process is "smooth" in that there are no identifiable phases. However, the process is less "well rounded" at the end: the final graph can in general be any graph on n vertices which is edge-maximal subject to the condition $\Delta \leq d$.

As before, for $0 \leq i \leq d$ let $Y_i(t)$ denote the number of vertices of degree i in G_t. Since vertices of degree d are ineligible for receiving an edge, the number of sites available for A_t is

$$Q(t) = \binom{n - Y_d(t)}{2} - F(t)$$

where $F(t)$ denotes the number of nuisance edges; that is, edges already present between vertices of degree less than d. We will have some control over $Y_d(t)$ but very little control over the random variable $F(t)$. Luckily, it does not affect matters too much since d is fixed: the trivial bound $F(t) \leq dn/2$ gives

$$Q(t) = \frac{1}{2}\left(n - Y_d(t)\right)^2 + O(n). \tag{3.6}$$

However, note that as is the case for the min-degree graph process, the vector process $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_d)$ is not Markovian.

The probability that A_{t+1} occupies any one of the available places is $\dfrac{1}{Q(t)}$. Next, what is the expected change in the variables Y_i in a single step of the process? In order to provide a simple example, we examine $d = 2$ separately and in more detail. The treatment of the general case in Section 3.3.2 does not depend on Section 3.3.1.

3.3.1 The 2-process

In [50] it was the behaviour of the number of vertices of degree 0 that lead to the main theorem, that the d-process almost always results in a d-regular graph (or almost d-regular, in the sense that one vertex has degree $d-1$ instead). Then in [51] various properties of the process were obtained in the case $d = 2$. Underlying this was the behaviour of the basic probability in the process, which is determined by $Q(t)$. In this section we consider the general behaviour of $Q(t)$ when $d = 2$.

Given G_t, the probability that the edge A_{t+1} joins two vertices of degree 0 is just the number of such pairs of vertices divided by the number $Q(t)$ of available places, or $\binom{Y_0(t)}{2}/Q(t)$. The probability it joins a vertex of degree 0 to one of degree 1 is $Y_0(t)Y_1(t)/Q(t)$. Hence

$$\mathbf{E}(Y_0(t+1) - Y_0(t) \mid G_t) = \frac{-2\binom{Y_0(t)}{2} - Y_0(t)Y_1(t)}{Q(t)}. \tag{3.7}$$

But note that the number of edges in G_t is $t = \frac{1}{2}(Y_1(t) + 2Y_2(t))$ by counting degrees, and also

$$n = Y_0(t) + Y_1(t) + Y_2(t). \tag{3.8}$$

Eliminating Y_2 from these, we find

$$Y_1(t) = 2(n - t - Y_0(t)) \tag{3.9}$$

and so, from (3.8) and (3.6),

$$Q(t) = \frac{1}{2}\left(2n-2t-Y_0(t)\right)^2 + O(n) \tag{3.10}$$

$$\sim \frac{1}{2}\left(2n-2t-Y_0(t)\right)^2 \tag{3.11}$$

provided $\frac{n-t}{\sqrt{n}} \to \infty$. Here we used the bound $Y_0(t) \le n-t$, which can be inferred from (3.9). Thus, again using (3.9), (3.7) becomes (after a little algebra and dropping an insignificant term)

$$\mathbf{E}(Y_0(t+1) - Y_0(t) \mid G_t) = \frac{-2Y_0(t)}{2n-2t-Y_0(t)} + o(1) \tag{3.12}$$

provided $\frac{n-t}{\sqrt{n}} \to \infty$. Defining $z(x)$ to model $\frac{1}{n}Y_0(xn)$, as in Section 3.1, suggests the differential equation

$$z'(x) = \frac{-2z(x)}{2-2x-z(x)}. \tag{3.13}$$

The general solution of this is

$$2(1-x) = z(C - \ln z). \tag{3.14}$$

Since $Y_0(0) = n$, the initial condition is $z(0) = 1$, and so

$$z(2 - \ln z) = 2(1-x) \tag{3.15}$$

(but see also Section 3.3.3 for a different way to arrive at this result). The conclusion is that the number of vertices of degree 0 in G_t is a.a.s. close to $nz(t/n)$. This is made precise in Theorem 5.2. The behaviour of $Q(t)$ is also obtained now via (3.11).

In [51], a version of this approximation result was proved, and using this it was shown that the expected number of cycles in the final graph G_n of this process is at most $3 + \log n$. Properties of the number of short cycles were also obtained, as discussed in Section 3.3.4.

More precise results on numbers of cycles and the probability that G_n is hamiltonian were later determined in [58] (see Section 5.4).

3.3.2 The d-process for arbitrary d

In general, the nuisance edges complicate the formula analogous to (3.7) for Y_1, Y_2 etc., so we work with approximations. We need one preliminary inequality. Noting that $dn/2 - t$ is the number of extra edges which G_t would need to make a d-regular graph, which cannot exceed half of the deficiency in all the vertex degrees less than d, we have

$$dn - 2t \le d(n - Y_d(t)). \tag{3.16}$$

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 A_{t+1} is

$$\frac{Y_i(t)(n - Y_d(t) - 1) + O(n)}{Q(t)} = \frac{2Y_i(t)}{n - Y_d(t)} + O\left(\frac{n}{(n-Y_d(t))^2}\right)$$

provided $\frac{dn/2-t}{\sqrt{n}} \to \infty$. Here the error term $O(n)$ on the left is due to those sites already occupied by edges. The error term on the right is from (3.6) and $Y_i(t) < n$. In the following, we assume $t < n(d/2 - \epsilon)$ for some $\epsilon > 0$. Then in view of (3.16), this expression becomes

$$\frac{2Y_i(t)}{n - Y_d(t)} + O\left(n^{-1}\right). \tag{3.17}$$

As (3.17) contributes negatively to the expected increase of $Y_i(t)$, and positively to that of $Y_{i+1}(t)$ for $i < d$, we now obtain for $0 \le i \le d$

$$\mathbf{E}(Y_i(t+1) - Y_i(t) \mid G_t) = \frac{2(1-\delta_{i0})Y_{i-1}(t) - 2(1-\delta_{id})Y_i(t)}{n - Y_d(t)} + O\left(n^{-1}\right). \tag{3.18}$$

Here the Kronecker δ terms ensure that $Y_{i-1}(t)$ is ignored when $i = 0$, as is $Y_i(t)$ when $i = d$.

Introducing $z_i(x)$ to model $\frac{1}{n}Y_i(xn)$ as before, the suggested differential equations are

$$z_i'(x) = \frac{2(1-\delta_{i0})z_{i-1}(x) - 2(1-\delta_{id})z_i(x)}{1 - z_d(x)}, \qquad i = 0, 1, \ldots, d. \tag{3.19}$$

The initial conditions are $z_0(0) = 1$ and $z_i(0) = 0$ for all other i. These equations were given for this process in [61], without a solution. In the following subsection we derive a solution without "solving" them!

Theorem 3.2 *Let z_i, $0 \le i \le d$ be the solutions of (3.19), and $\lambda = o(1)$. Then with probability $1 - O(\frac{1}{\lambda}e^{-n\lambda^3/8})$*

$$Y_l(t) = nz_l(t/n) + O(\lambda n)$$

uniformly for all $0 \le l \le d$ and $0 \le t < n(d/2 - \epsilon)$.

Sharp concentration is shown by taking $\lambda = n^{-1/4}$ for instance.

An improvement of this theorem (see Section 5.4) can be obtained which applies to values of t much closer to the natural end of the process, which occurs by time $nd/2$. A result like this (but with only an upper bound on Y_0) was used together with an easy argument [50, Corollary to Lemma 3.1] to show that a.a.s. $Y_0(t) = 0$ for some t with $nd/2 - t \to \infty$. From this it was shown that the final graph of the d-process is a.a.s. d-saturated.

3.3.3 Probabilistic solution to the differential equations

Some of the justification of the argument in this section is only sketched. Since the argument is merely a device to find a solution to (3.19), this is acceptable. On the other hand, the discussion here also serves as a simple example of an approach which is useful for calculating other things about processes, such as in the next section.

From Theorem 3.2 and (3.6), with probability $1 - O(n^{-K})$ (for any fixed K)

$$Q(t) \sim \frac{1}{2}n^2\alpha^2(t) \tag{3.20}$$

where $\alpha(t) = 1 - z_d(t/n)$. This and the other statements in this section apply only for time $t < n(d/2 - \epsilon)$ for some $\epsilon > 0$. (The process ends anyway by the time $nd/2$.) By (3.16), we can assume throughout that $\alpha(t) \geq \epsilon'$ for some $\epsilon' > 0$. Now "forget" where α came from, and ask what is implied by the knowledge that $Q(t)$ is usually close to some fixed function as given by (3.20). If the concentration is sharp enough, we can determine the expected number of vertices of given degrees, as follows.

Let $u \in [n]$. With knowledge of $Q(t)$ and $Y_d(t)$ we can compute approximately the probability that A_{t+1} is incident with u. If $d(u) = i$, the number of places an edge can go incident with u is between $n - 1 - i - Y_d(t)$ and $n - 1 - Y_d(t)$, and hence approximately $n\alpha(t)$. Thus, defining $\mathcal{H}_t$ to be the event that A_t is incident with u,

$$\mathbf{P}(\mathcal{H}_{t+1}) \sim \frac{n\alpha(t)}{Q(t)} \sim \frac{2}{n\alpha(t)}, \tag{3.21}$$

$$\mathbf{P}(\overline{\mathcal{H}_{t+1}}) \sim 1 - \frac{2}{n\alpha(t)} \sim \exp\left(-\frac{2}{n\alpha(t)}\right). \tag{3.22}$$

The probability that u has degree $k < d$ in G_s can be calculated by classifying processes according to the times $t_1 + 1, \ldots, t_k + 1$ that the edges arrive. The probability of *not* receiving edges at any of the other times is the product of (3.22) over all $0 < t < s$ *apart from* $t = t_1, \ldots, t_k$. Including these extra k factors in this product makes no difference asymptotically, and so the probability of not receiving edges at the undesired times is approximately (see justification below)

$$\prod_{t=0}^{s-1} \exp\left(-\frac{2}{n\alpha(t)}\right) = \exp\left(-\sum_{t=0}^{s-1} \frac{2}{n\alpha(t)}\right) \approx \exp\left(-\int_0^{s/n} \frac{2}{n\alpha(t)}\, dt\right) = e^{-\lambda(x)},$$

where we define

$$\lambda(x) = \int_0^x \frac{2}{n\alpha(t)}\, dt \tag{3.23}$$

and $x = s/n$. The probability that u does receive edges at all of the desired times is the product of (3.21) for $t = t_1, \ldots, t_k$, and so

$$\begin{aligned}
\mathbf{P}(d_{G_s}(u) = k) &\approx e^{-\lambda(x)} \sum_{0 \leq t_1 < \cdots < t_k < s} \prod_{i=1}^{k} \frac{2}{n\alpha(t_i)} \\
&\approx e^{-\lambda(x)} \frac{1}{k!} \sum_{0 \leq t_1 < \cdots < t_k < s} \prod_{i=1}^{k} \frac{2}{n\alpha(t_i)} \\
&= e^{-\lambda(x)} \frac{1}{k!} \prod_{i=1}^{k} \sum_{t=0}^{s-1} \frac{2}{n\alpha(t)} \\
&\approx \frac{\lambda(x)^k e^{-\lambda(x)}}{k!}.
\end{aligned} \tag{3.24}$$

We pause here to note that the requirement in making this argument rigorous is to show that the approximations in (3.21) and (3.22) are valid also when conditioning on the events

that up until time t, the edges incident with u arrived precisely at the prescribed times. Since $\alpha > \epsilon'$, this event turns out to have probability at least $C(\epsilon)n^{-d}$ (strictly, this is shown by using this justification inductively — see [51, Theorem 1] for the details of an argument just like this one) and (3.20) holds with probability at least $1 - O(n^{-d-1})$, so conditioning on such events makes no difference to the conclusion.

Interestingly, (3.24) shows that the distribution of the degree of any given vertex in G_s is asymptotically the same as the distribution of $\min\{X, d\}$ where X is a Poisson variable with mean $\lambda(s/n)$. This determines $Y_k(t)/n$, and hence $z_k(x)$ (approximately), for $i < d$, in terms of $\lambda(x)$. Another equation comes from the fact that the sum of z_i must equal 1 since the total number of vertices is n. Thus we are lead to

$$z_i(x) = \frac{\lambda(x)^i e^{-\lambda(x)}}{i!} \quad (0 \le i < d-1), \qquad z_d(x) = 1 - \sum_{i=0}^{d-1} z_i(x). \tag{3.25}$$

It is easily verified that these functions satisfy the differential equations (3.19) for $i < d$. However, all the variables are expressed in terms of the unknown function $\lambda(x)$. This is determined by the differential equation in (3.19) for $i = d$. However, it is simpler to find another equation, solve for λ, and then check that the differential equation in (3.19) is satisfied. Note that G_s has $s = \frac{1}{2}\sum_{i=1}^{d} Y_i(s)$ edges. So, from the approximate equality of $Y_k(t)/n$ and $z_k(x)$,

$$\sum_{i=0}^{d} i z_i(x) = 2x. \tag{3.26}$$

Eliminating z_d between this and (3.25), we obtain

$$e^{-\lambda(x)} \sum_{i=0}^{d-1} \frac{(d-i)\lambda(x)^i}{i!} = d - 2x.$$

This implicitly determines $\lambda(x)$ and hence all the $z_i(x)$. Finally, it is easy to check that the equation for $z_d'(x)$ is satisfied, by differentiating (3.26) and using the equation for z_d in (3.25).

3.3.4 Distribution of short cycles in the 2-process

In the previous section the vertex degree distribution was obtained purely from the knowledge of Q contained in (3.20). The same methodology was used in [51] to obtain the distribution of short cycles in the final graph G_n of the 2-process. There it was shown that in G_n the number of cycles of length l is asymptotically Poisson for fixed $l \ge 3$. The same method of proof actually gives the fact that the numbers of cycles of given lengths up to some fixed bound are asymptotically independent Poisson. For $l = 3$ the mean was shown to converge to $\frac{1}{2}\int_0^\infty \frac{(\log(1+x))^2}{xe^x}\,dx \approx 0.1887$. This is close to but, significantly, different from the expected number of triangles in a 2-regular graph selected uniformly at random, which is $\frac{1}{6}$ (see Bollobás [11] or Wormald [60]). For $l \ge 3$, the mean was found in the form of a rather intractible n-fold integral.

Underlying the proofs was knowledge of the concentration of Q. The required concentration is easily deduced from the concentration of Y_0 given in Theorem 5.2 of the present paper.

Here is a description of the argument (sketching some of the details, as in the previous section). It uses the method of moments. For this, define the random variable X_l to denote the number of cycles of length l in G_n. One then computes the factorial moments $\mathbf{E}[X_l]_k$ of X_l, checks that they tend towards the factorial moments λ^k of a Poisson variable, and concludes that the variable is asymptotically Poisson, that is

$$\mathbf{P}(X_l = j) \sim \frac{\lambda^j}{j! e^j}. \tag{3.27}$$

See Bender [8], Bollobás [12], Alon, Spencer and Erdős [4] for general descriptions of this method in the combinatorial setting.

We examine only $l = 3$ with $k = 1$ in detail, since the other cases are done the same way. For $\{u, v, w\} \subseteq [n]$, consider the indicator variable W_{uvw} for the event that uvw is a triangle of G_n. Then $X_3 = \sum W_{uvw}$. So by linearity of expectation

$$\mathbf{E}X_3 = \sum \mathbf{E}W_{uvw} = \binom{n}{3}\mathbf{E}W_{123} = \binom{n}{3}\mathbf{P}(W_{123} = 1) \tag{3.28}$$

since by symmetry, each of these $\binom{n}{3}$ indicator variables has the same expectation.

The event $W_{123} = 1$ can be partitioned into the events

$$\mathcal{T}_{q,r,s} = \{A_q = \{1, 2\}\} \wedge \{A_r = \{1, 3\}\} \wedge \{A_s = \{2, 3\}\}$$

where q, r and s are distinct and $1 \le q, r, s \le n$. We first compute the probabilities of these events approximately, assuming

$$Y_0(t) \sim nz(t/n) \tag{3.29}$$

where z is given by (3.15), and then describe the justification of this and the other approximations which use it.

To find the probability of $\mathcal{T}_{q,r,s}$ we need to estimate the probability of not hitting the vertices 1, 2 or 3 except with the edges A_q, A_r and A_s. So consider the complementary probability, of hitting one of these vertices. Take fixed values of q, r and s, and assume without loss of generality $q < r < s$. The number of ways to add an edge to G_t such that the edge is incident with at least one of the vertices $\{1, 2, 3\}$ is

$$\begin{aligned}
&3(n - Y_2(t) - 3) + 3 && \text{if } t < q,\\
&3(n - Y_2(t) - 3) + 2 && \text{if } q \le t < r,\\
&2(n - Y_2(t) - 2) + 1 && \text{if } r \le t < s, \text{ and}\\
&0 && \text{if } t \ge s.
\end{aligned}$$

In view of (3.11) and (3.29), write

$$Q(t) \approx \frac{n^2}{2} v(x)^2 \tag{3.30}$$

where

$$x = \frac{t}{n}, \quad v(x) = 2 - 2x - z(x). \tag{3.31}$$

From (3.9) and (3.29) we have

$$n - Y_2(t) = Y_0(t) + Y_1(t) \approx nv(x).$$

The probability that A_{t+1} is incident with at least one of the vertices 1, 2 or 3 is obtained by dividing the expressions above by $Q(t)$. Neglecting terms which are $O(1/Q(t))$ and using the approximations for $Q(t)$ and $n - Y_2(t)$, this is

$$\frac{6}{nv(x)} \text{ if } t < r, \quad \frac{4}{nv(x)} \text{ if } r \le t < s, \quad \text{and} \quad 0 \text{ if } t \ge s.$$

This leads to the expression

$$\prod_{t=0}^{q-2}\left(1 - \frac{6}{nv(x)}\right) \approx \exp\sum_{t=0}^{q-2}\frac{-6}{nv(x)} \approx \exp\int_0^{q/n}\frac{-6}{v(x)}\,dx$$

for the probability of not hitting $\{1,2,3\}$ with any A_{t+1} for $t < q$. The summation here is a Riemann sum for the integral, and the approximations are justified as long as $v(x)$ is bounded away from 0, or at least does not go to 0 too quickly with n. (This statement is justified in [51] using a sharp concentration result similar to, but actually weaker than, that obtained here in Theorem 5.2. The basic problem is that of estimating probabilities of the conjunctions of events such as $A_q = 12$ and $A_r = 13$, and the justification is similar to that described in the argument in Section 3.3.3. The speed with which $v(x)$ is permitted to go to 0 leads to some complications which we omit from this sketch.) For $q < t < r$ we obtain a similar expression with lower limits q or q/n and upper limits $r - 2$ or r/n, and for $r < t < s$ we get

$$\prod_{t=r}^{s-2}\left(1 - \frac{4}{nv(x)}\right) \sim \exp\sum_{t=r}^{s-2}\frac{-4}{nv(x)} \sim \exp\int_{r/n}^{s/n}\frac{-4}{v(x)}\,dx\,.$$

This is all an estimate of the probability that the process does not join any edges to the vertices $\{1,2,3\}$ at the wrong times. On the other hand,

$$\mathbf{P}(A_q = 12) = \frac{1}{Q(q-1)}, \quad \mathbf{P}(A_r = 13) = \frac{1}{Q(r-1)}, \quad \mathbf{P}(A_s = 23) = \frac{1}{Q(s-1)},$$

where these probabilities are conditional on G_q, G_r and G_s respectively. Estimating each of these using (3.30) and assuming they are asymptotically independent of the probabilities of not hitting at the other times (which also follows from the conditioning argument as mentioned above) leads to

$$\mathbf{P}(\mathcal{T}_{q,r,s}) \sim \frac{8}{n^6(v(\frac{q}{n})v(\frac{r}{n})v(\frac{s}{n}))^2}\exp\left(-6\int_0^{r/n}\frac{dx}{v(x)} - 4\int_{r/n}^{s/n}\frac{dx}{v(x)}\right).$$

From (3.13) and (3.31), $v(x) = \frac{-2z(x)}{z'(x)}$ and so

$$\int \frac{dx}{v(x)} = \int \frac{-z'(x)}{2z(x)}\,dx = \frac{-\ln z(x)}{2} + C\ .$$

Since $z(0) = 1$, this simplifies the exponent to $2\ln(z(\frac{r}{n})) + 4\ln(z(\frac{s}{n}))$ and the result is

$$\begin{aligned}
\mathbf{P}(W_{123} = 1) &= \sum_{1\le q,r,s\le n} \mathbf{P}(\mathcal{T}_{q,r,s}) \\
&= 6 \sum_{1\le q<r<s\le n} \mathbf{P}(\mathcal{T}_{q,r,s}) \\
&\sim 6 \sum_{1\le q<r<s\le n} \frac{8z(\frac{r}{n})(z(\frac{s}{n}))^2}{n^6(v(\frac{q}{n})v(\frac{r}{n})v(\frac{s}{n}))^2} \\
&\sim 48 \int_0^1 \int_{x_1}^1 \int_{x_2}^1 \frac{z(x_2)(z(x_3))^2}{n^3(v(x_1)v(x_2)v(x_3))^2}\,dx_3\,dx_2\,dx_1
\end{aligned}$$

where there is negligible error in replacing the sum by the integral. (The justification of this is omitted in this sketch — this is where it is important how close to 1 x is permitted to get in the above argument, since v goes to 0.)

Now by (3.28), since $\binom{n}{3} \sim n^3/6$,

$$\mathbf{E}X_3 \sim 8 \int_0^1 \int_{x_1}^1 \int_{x_2}^1 \frac{z(x_2)(z(x_3))^2}{(v(x_1)v(x_2)v(x_3))^2}\,dx_3\,dx_2\,dx_1.$$

Set

$$y_i = v(x_i)/z(x_i).$$

Then by (3.31), (3.15) and (3.13)

$$y_i = 1 - \log z(x_i), \qquad dy_i = \frac{2dx_i}{v(x_i)}.$$

With this change of variables, and since $y(z(0)) = y(1) = 1$ and $\lim_{x\to 1} y(z(x)) = \lim_{z\to 0} y(z) = \infty$,

$$\mathbf{E}X_3 \sim 8 \int_1^\infty \int_{y_1}^\infty \int_{y_2}^\infty \frac{e^{y_1-y_3}}{y_1y_2y_3}\,dy_3\,dy_2\,dy_1\ .$$

where the factor $z(x_3)/z(x_1)$ was replaced by $e^{y_1-y_3}$ by the definition of y_i. This integral becomes

$$\int_1^\infty \int_{y_1}^\infty \frac{e^{y_1-y_3}(\log y_1 - \log y_3)}{y_1y_3}\,dy_3\,dy_1$$

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

$$\begin{aligned}
x &= y_3 - y_1, \\
y &= \log y_3 - \log y_1
\end{aligned}$$

gives

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

which gives the above-mentioned integral formula for λ in (3.27) when $l = 3$.

To establish the fact that X_3 is asymptotically Poisson, we can show that its factorial moments, for example $\mathbf{E}(X_3(X_3-1))$, behave as required for (3.27). The argument is very similar to that for $\mathbf{E}X_3$, so we do not include the details here.

3.4 Degree bounded star graph process

3.4.1 The star d-process in general

This process resembles the min-degree graph process, since it progresses through phases separated by the times at which the minimum degree of G_t increases. Again, phase i denotes the set of t for which $\delta(G_t) = i$. We will take d to be fixed.

An important simplifying feature of the star d-process compared to the ones above concerns the nature of the vector process $\mathbf{Y}(t) = (Y_0(t), \ldots, Y_d(t))$. The source of this simplification is the fact that there cannot be any nuisance edges joining a newly selected vertex and the other vertices of degree less than d. This is because at each step, at least one end of every edge has degree d. The vertices of degree less than d hence form an independent set in G_t, and so it is clear that the probability distribution of $\mathbf{Y}(t+1)$ is determined entirely from $\mathbf{Y}(t)$. It follows that the vector process $\mathbf{Y}(t)$ is Markov. The advantage of this is that in the analysis of the expected change of the variables in one step conditional on $\mathbf{Y}(t)$, there is no need for an error term of the sort occuring in (3.17). However, the differential equations associated with this process unfortunately seem to have more complicated solutions than for the d-process, even for $d = 2$.

We can argue as in Section 3.3.2. If the process is in phase k at time t, when the star is added to form G_{t+1} the centre of the star changes from degree k to degree d. In addition, the expected number of other vertices changing from degree i to degree $i+1$ is $(d-k)(Y_i - O(1))/(n - Y_d(t) - 1)$ for $i \le d-1$. This is because the other vertices in the star are chosen u.a.r. from the $n - Y_d(t) - 1$ vertices of degree less than d. The $O(1)$ error term accounts for the fact that during the process of adding these $d-k$ edges one by one, the value of Y_i and of the denominator can change by $O(1)$. (However, as mentioned above, since $\mathbf{Y}(t)$ is Markov, it is possible to compute this error term explicitly. At this point we include such an error term for minor simplification of the formulas and argument rather than by necessity as it was for d-processes.) So provided $n - Y_d(t) \to \infty$, the expected increase in Y_i in this step is

$$\delta_{id} - \delta_{ik} + \frac{(d-k)(Y_{i-1}(t) - Y_i(t))}{n - Y_d(t)} + o(1).$$

Following [45, 47], write $Z_i = \sum_{j=0}^{i-1} Y_j$, the number of vertices of degree strictly less than i. (This seems to give slightly simpler equations than considering Y_i.) The expression above now becomes

$$-1 - \frac{(d-k)(Z_i(t) - Z_{i-1}(t))}{Z_d(t)}.$$

Writing $z(x) = \frac{1}{n}Z(xn)$, the differential equations suggested are

$$z_i'(x) = -1 - \frac{(d-k)(z_i(x) - z_{i-1}(x))}{z_d(x)} \tag{3.32}$$

for $i = k+1, \ldots, d$, where $z_0, \ldots, z_k$ are defined to be 0. As z_i corresponds to Z_i, the initial conditions for phase 0 of the process are $z_i = 1$ for all $i > 0$.

A result like Theorems 3.1 and 3.2 can now be stated, which says that the Z_i are a.a.s. close to the functions $nz_i(t/n)$. The validity extends until the time that Z_k becomes 0, which must be close to x_k where x_k is the first positive solution of $z_k(x) = 0$ (see the proof of Theorem 3.2 in Section 5.3 for a similar argument). Since phase 0 finishes when Y_0 first becomes 0, and Y_0 decreases by at least 1 in each step in phase 0, we can conclude that phase 1 will a.a.s. begin approximately when $t \approx nx_0$. Then during phase 1 the same differential equations will govern the process, but with $k = 1$, and the initial conditions are the values of the solution to the first system at $x = x_0$. So note that the functions z_i will behave differently here, and the derivatives are not continuous at the phase transition. The finish of phase 1 is determined by x_1 which is defined in terms of the phase 1 equations. And so on for the later phases. Without actually solving the differential equations (3.32) in [47] it is shown that the solutions are such that for $0 < k \leq d-1$, $z_{k+1}(x)$ is strictly positive when phase k begins (which is when $z_{k-1}(x)$ first reaches 0 in phase $k-1$). Hence a.a.s. the process itself exhibits all the phases up to $d-1$; that is, at the end of each phase k there is at least a constant times n vertices of degree $k+1$.

Star d-processes have a particular appeal as an object of study, due to the fact that the last phase $(d-1)$ consists of the addition of a uniformly chosen random matching to an independent set of vertices of degree $d-1$ (assuming the number of such vertices is even!). Such a random matching is for the most part much easier to analyse than say the last constant times n edges of the random d-process. For example, in Section 3.3.4, the last part of the process causes the most trouble, where we have less information about the conditional distribution of the changes in the degree sequence for each step.

3.4.2 The star 2-process

Robalewska [46] was able to obtain a good deal of information in the case $d = 2$. In this case for phase 0 the differential equations (3.32) become $z_1' = -1 - 2z_1/z_2$ and $z_2' = -3 + 2z_1/z_2$. The solution for z_1 is given implicitly by the equation

$$-\frac{1}{2}\ln(2v^2 + 3v + 2) - \frac{5}{\sqrt{7}}\arctan\frac{\sqrt{7}(v+1)}{1-v} = \ln(2x-1),$$

where $v = z_1/(2x-1)$, and from this that the first positive solution of $z_1 = 0$ is at

$$x_0 = \frac{1}{2}\left(1 - \frac{1}{\sqrt{2}}\exp\left(-\frac{5}{\sqrt{7}}\arctan\sqrt{7}\right)\right) \approx 0.4640\,. \tag{3.33}$$

So from the sharp concentration demonstrated in the previous section for star d-processes in general, the random star 2-process a.a.s. has a transition from phase 0 to phase 1 (i.e.,

all vertices of degree 0 disappear) at $t \sim x_0 n \approx 0.4640n$. Upon summing the degrees of the vertices of G_t one obtains the relation $Z_1 + Z_2 = 2(n - 2t)$, from which it follows (using the sharp concentration) that $z_1 + z_2 = 2(1 - 2x)$ (which can alternatively be derived from, but is not immediately suggested by, the differential equations (3.32)). This means that z_2 and Z_2 are now determined in phase 0 (a.a.s., in the case of Z_2). On the other hand, in phase 1 the number of vertices of degree 1 decreases by exactly 2 at every step. Equivalently, $z_2(x) = 2(1 - x_0 - x)$. Since $z_2 = 0$ in this phase corresponds to the end of the whole star 2-process, this must occur when $t \approx 1 - x_0 \approx 0.5360$.

In [46] several features of the star 2-processes were studied using the above results on the concentration of the variables Z_i. These include the distribution of the number of paths of length $k \leq 4$ at time t, and similarly for cycles. It was shown for instance that the number of cycles in the evolving graph at the end of phase 0 is asymptotically Poisson $\mathbf{Po}(\lambda_0)$ with

$$\lambda_0 = -\frac{1}{4}\ln 2 + \frac{3}{2\sqrt{7}}\arctan\sqrt{7} \approx 0.5124 \tag{3.34}$$

and the number in the final graph of the process is asymptotically Poisson $\mathbf{Po}(\lambda)$ with

$$\lambda = \frac{1}{2}(\gamma + \ln n) - \frac{1}{\sqrt{7}}\arctan\sqrt{7},$$

where $\gamma \approx 0.5772$ is Euler's constant. (Note that the term $\frac{1}{2\gamma}$ in [46, Theorem 3.1] should be $\frac{\gamma}{2}$. Also, the claimed bound $O(1/\ln n)$ on the total variation distance between this variable and $\mathbf{Po}(\lambda)$ applies only for the number of cycles formed in phase 1; the addition of the number of cycles in phase 0 could conceivably increase this. However, it does follow from the argument there that the expected number of cycles is equal to $\lambda + o(1)$.) As in Section 3.3.4 for 2-processes, the numbers of cycles of given fixed length in the final graph of the random star 2-process are asymptotically Poisson, and Robalewska obtained the expected numbers of 3-, 4- and 5-cycles as single or double integrals. The expected number of triangles is approximately 0.0797, which is very different from that in d-processes or random 2-regular graphs with the uniform distribution (see Section 3.3.4). These results about short cycles were obtained using an argument which is quite different from that in Section 3.3.4 and [51], but which uses results from yet another application of the differential equation method. It runs along the following lines. First, the auxiliary variable giving the number of paths of length l is shown, at the end of phase 0, to be a.a.s. $c_l n + o(n)$ where c_l is a constant depending on l. This is proved using a system of differential equations for variables counting the numbers of paths of length l ($l = 2, \ldots, k$) at time t. The number of cycles of length j is determined from these numbers and from the number of ways to join up paths of various lengths into a j-cycle. Using the uniformity of the matching placed in phase 1, the expectation can be determined, as well as the asymptotic Poisson property (using the method of moments).

Another result from [46] is that the probability that the final graph is hamiltonian is asymptotically $\frac{1}{2}e^{-\lambda_0}\sqrt{\pi}/\sqrt{n(1-2x_0)}$ with λ_0 and x_0 as in (3.34,3.33). This was obtained by computing the probability of creating no cycles until the last edge of the whole process. Unlike the paths and cycles, this requires only knowledge of the concentration of the Z_i, and

no auxiliary variables. The argument required is less delicate than the analogous computation for the random (non-star) 2-process.

3.5 Simple greedy algorithm for independent sets

This is the first of the two algorithms we consider for finding independent sets in random d-regular graphs. The analysis uses the pairing process described in Section 2. Note that at any stage of selecting the pairs, there is a corresponding graph whose edges correspond to the pairs, and which we refer to as the *evolving* graph.

This algorithm is the following: given a random d-regular graph G, choose the vertices in an independent set I consecutively, each vertex chosen randomly from those not already in the set, nor adjacent to a vertex in the set. In order to satisfy its description, this algorithm must find all vertices adjacent to I as each new vertex is added to I, which requires *probing* all the edges adjacent to vertices in I.

We need to relate the algorithm to the random pairing described in Section 2, in which we are free to choose at each step any unpaired point, and then randomly select its mate for a pair. This can be done using what is called the *method of deferred decisions* by Knuth et al. [30]. Randomly selecting the mate of an unpaired point is called *exposing* the pair. When the algorithm calls for probing all edges incident with a vertex v added to the independent set, the pairing process can expose all the unexposed pairs containing points in (the bucket corresponding to) v. When the algorithm chooses another vertex v, the pairing process exposes all the points in that vertex, and so on. (In this respect, this process is the same as the star d-process, which exposes a star at each step, but the probabilities are different.) Thus, if we incorporate the random choice of v into the pairing process, then this process simulates the greedy independent set algorithm on a random d-regular graph. The pairing process generates the edges of the random graph just as the algorithm requires them. When no more vertices with no paired points are left, i.e. the graph being generated has minimum degree at least 1, the algorithm stops, so the rest of the pairing process is irrelevant.

An appropriate set of random variables has to be specified. Also, there is no need to have the basic time parameter for these variables the same as the canonical one in the definition of the pairing process. The size of the independent set produced by the algorithm is equal to the number of times the pairing process selects a new vertex for pairing all its points, before the number of vertices with no paired points drops to 0. So keeping track of the number of vertices of degree 0 in the evolving graph, and the number of vertices added to I, will suffice to determine all that is required. However, if we choose the basic time variable t to denote the number of pairs chosen, then these variables behave somewhat erratically, since the first pair chosen after putting a new vertex v into I *always* reduces the number of vertices of degree 0, whereas other pairs *sometimes* reduce this number. So instead let the time t denote the number of vertices in the set I, and let $Y(t)$ denote the number of vertices of degree 0 in the evolving graph after all the edges incident with these t vertices have been generated. Then the size of the independent set found is just the first t for which $Y(t) = 0$.

The number of points in the vertices of degree 0 at time t is $dY(t)$, and there are $nd - 2td$ unmatched points in total (since d pairs are chosen for each unit of time). So the probability

that a random point is in a vertex of degree 0 at time t is $\frac{dY(t)}{nd-2dt} = \frac{Y(t)}{n-2t}$. This probability changes a little during the next step, when the new vertex of degree 0 is chosen, as soon as the first point is eliminated from it (and the edge to this point is about to be exposed). When this happens, $Y(t)$ immediately decreases by 1, and the other point in the exposed edge is chosen randomly from the $nd - 2dt - 1$ remaining points. Nevertheless, for each of the d edges exposed in this step, the probability of joining to a vertex of degree 0 (previously) is $\frac{Y(t)}{n-2t} + O(\frac{1}{n-2t-2})$ provided $2t+2 < n$. Thus, using P_t to denote the pairing generated by time t, and noting $Y(t) \leq n - 2t$,

$$\mathbf{E}(Y(t+1) - Y(t) \mid P_t) = -1 - \frac{dY(t)}{n-2t} + O(\frac{1}{n-2t-2}) \tag{3.35}$$

for $2t + 2 < n$.

Writing $Y(t) = nz(t/n)$, the differential equation suggested is

$$z'(x) = -1 - \frac{dz(x)}{1-2x} \tag{3.36}$$

with initial condition $z(0) = 1$. This is a first-order equation with solution

$$z(x) = \frac{(d-1)(1-2x)^{d/2} - (1-2x)}{d-2}.$$

The smallest positive solution of $z(x) = 0$ is

$$x_0 = \frac{1}{2} - \frac{1}{2}\left(\frac{1}{d-1}\right)^{2/(d-2)}.$$

Using the same method of proof as Theorems 3.1 and 3.2, it follows that the independent set which the greedy algorithm finds in a random d-regular graph a.a.s. has size asymptotically $\frac{n}{2} - \frac{n}{2}\left(\frac{1}{d-1}\right)^{2/(d-2)}$.

This is not as large as the one found using the degree-greedy algorithm (see Section 7.1).

4 Bounding large deviations

To show that the random variables in a process usually approximate the solution of differential equations as derived in Section 3, we use large deviation inequalities. These inequalities are often used to give an upper bound on the probability that a random variable deviates very far from its expected value. In probabilistic applications to discrete mathematics, abundant use is made of Markov's inequality,

$$\mathbf{P}(X \geq \alpha \mathbf{E}X) \leq \frac{1}{\alpha} \tag{4.1}$$

where X is a non-negative random variable and $\alpha > 1$.

Although Markov's inequality is the simplest, some of the most powerful inequalities of this type come relatively easily by applying Markov's to a simple transform of the random variable in question. In a typical situation with a random process, the aim is to show that the random variable Y_t of interest is sharply concentrated. In fact, $Y_t - Y_0$ is the sum of the differences $Y_i - Y_{i-1}$, $1 \leq i \leq t$. If one is lucky, the differences $Y_i - Y_{i-1}$ are independent, in which case the Chernoff bounds (see [16] and [4]) are very useful and are often close to the true values. When the differences are not independent but there is a large degree of independence, results can often be obtained by making use of analogous bounds given for martingales (defined below) by Hoeffding [26], which are also useful in the independent case. These were surveyed by McDiarmid [34] and Bollobás [12] and new applications are continually occurring (for a physically close example, see Srinivasan [56]). The martingale is often related to $\langle Y_t \rangle$ in some indirect, occasionally ingenious, way, designed so that concentration of the elements in the martingale sequence implies concentration of Y_t.

In using these inequalities it is helpful to realise that they tend to be close to sharp. Hence they usually give bounds less than 1 on the probability as long as the deviations are approximately as large as the standard deviation, or bigger.

4.1 Martingales and supermartingales

A *martingale* with respect to a random process $G_0, G_1, \ldots$, with G_0 fixed, is a sequence $X_0, X_1, \ldots$ of random variables defined on the random process such that[2]

$$\mathbf{E}(X_{i+1} \mid G_0, G_1, \ldots, G_i) = X_i \quad (i = 0, 1, \ldots) \ .$$

In most applications in the random graph literature, and in the present article, the martingale satisfies the property that $\mathbf{E}(X_{i+1} \mid G_0, G_1, \ldots, G_i) = \mathbf{E}(X_{i+1} \mid G_i)$, so the condition above becomes

$$\mathbf{E}(X_{i+1} \mid G_i) = X_i. \tag{4.2}$$

Also, in most of these applications, the random process is of finite length. This can also be viewed as an infinite process which becomes constant by some predetermined time. Similarly, the restriction of any associated martingale to an initial sequence can also be defined to be a martingale.

The martingale above can be written as the sequence of partial sums of the sequence $\langle Y_i \rangle = \langle X_i - X_{i-1} \rangle$, in which case $\langle Y_i \rangle$ is called a *martingale difference sequence*. The Hoeffding and Chernoff bounds rely on the same type of large number effect that causes the distribution of $\sum_{i=1}^{t} Y_i$ to be close to normal, and so they work best when no small subset of the Y_i dominates the sum. This is most conveniently ensured by hypothesising that there is some uniform upper bound on the differences $|X_i - X_{i-1}|$, hence the term "bounded difference inequality".

We begin with a simple example of a martingale defined on a graph process in a very natural way. Consider a process $G_0, G_1, \ldots$ where G_0 is any given graph with $n \geq 3$ vertices,

[2]Also $\mathbf{E}X_i$ should be finite for all i.

and G_{i+1} comes from G_i by the following operation. Choose three vertices at random from G_i. If they induce a complete graph K_3 or its complement, do nothing. If there is only one edge joining them, with probability $\frac{2}{3}$ delete it and otherwise add the other two edges to make a K_3. If there are two edges joining them, with probability $\frac{1}{3}$ delete them and otherwise add the third edge. Put $X_i = |E(G_i)|$. If we condition on G_i and additionally on the outcome of the random selection of three vertices, the expected increase in the number of edges is 0. Consequently, it is also 0 conditioning only on G_i, and so (4.2) is immediately satisfied; that is, $\{X_i\}$ is a martingale with respect to $\{G_i\}$.

Since this process is such a simple one, some questions about it can be answered easily. For instance, note that with probability 1 the process eventually becomes constant. It is a Markov chain whose states are the n-vertex graphs, and the only absorbing states are the complete graph and its complement. We call the absorbing state reached the final graph. What is the probability p that the final graph G is the complete graph? This is easily found from the observation that for a martingale $\mathbf{E}X_{i+1} = \mathbf{E}X_0$ for all $i \geq 0$, and so

$$|E(G_0)| = \mathbf{E}|E(G_0)| = \mathbf{E}|E(G_i)| \to p\binom{n}{2}$$

as $n \to \infty$, from which $p = |E(G_0)|/\binom{n}{2}$.

A natural follow-up question is on the rate of convergence: how long does it usually take to reach the final graph? This is much harder to answer, but a good indication comes from the bounded difference inequalities of Hoeffding [26], such as the following one which is often called Azuma's inequality [6].

Lemma 4.1 *Let $X_0, X_1, \ldots, X_t$ be a martingale such that $|X_i - X_{i-1}| \leq c_i$, $1 \leq i \leq t$, for constants c_i. Then for any $\alpha > 0$*

$$\mathbf{P}(|X_t - X_0| \geq \alpha) \leq 2\exp\left(-\frac{\alpha^2}{2\sum c_i^2}\right).$$

This is often applied with $\alpha >> \sqrt{t}$ and the c_i all small non-zero integers. In the martingale discussed above, $c_i = 2$ for all i. Hence $\mathbf{P}(|X_j - X_0| \geq \alpha) \leq 2e^{-\alpha^2/8j}$ $(\alpha > 0)$. From this, the expected time taken for the process to reach an absorbing state is at least $C\alpha^2$, where $\alpha = \min\{|E(G_0)|, \binom{n}{2} - |E(G_0)|\}$, and C is a constant.

Unfortunately, natural martingales of the type in this example are not often useful. Instead, one often applies the above inequality to the sequence of variables

$$Y_i = \mathbf{E}(X_n \mid G_0, G_1, \ldots, G_i)$$

for $i \leq n$. Often called Doob's martingale process, this is always a martingale for any variables X_i determined by $G_0, G_1, \ldots, G_i$. Establishing some bound on the differences in such Y_i is required for Lemma 4.1. There is a broad class of graph processes for which such bounds have been shown to be automatic (see McDiarmid [34, §3] and Alon, Spencer and Erdős [4]). However, it can be rather tricky, such as in [50], where the Doob process was originally used

to establish the relevance of differential equations such as that occurring in Section 3.3.1 (also used in [51]).

Instead of using the Doob martingale, for our purposes here it is easier to obtain results which are just as strong using supermartingales. A *supermartingale* with respect to $G_0, G_1, \ldots$ is a sequence $X_0, X_1, \ldots$ of random variables defined on $G_0, G_1, \ldots$ such that

$$\mathbf{E}(X_{i+1} \mid G_0, G_1, \ldots, G_i) \leq X_i \quad (i = 0, 1, \ldots).$$

We will use the following simple supermartingale inequality analogous to Lemma 4.1. This lemma appeared in [61], but it is nothing more than the standard technique for estimating large deviations, which involves applying Markov's inequality to the moment generating function of X_i. The proof comes immediately, from a standard proof of Lemma 4.1 as given in [34]. We give the proof here because the method and its adaptations are very powerful for obtaining Hoeffding and Chernoff [16] type inequalities.

Lemma 4.2 *Let $X_0, X_1, \ldots, X_t$ be a supermartingale where $X_0 = 0$ and $|X_i - X_{i-1}| \leq c_i$ for $i \geq 1$ and constants c_i. Then for all $\alpha > 0$,*

$$\mathbf{P}(X_t \geq \alpha) \leq \exp\left(-\frac{\alpha^2}{2\sum c_i^2}\right).$$

Proof. Take any $h > 0$ (whose value will be optimised later). Using Markov's inequality for the second step,

$$\mathbf{P}(X_t \geq \alpha) = \mathbf{P}(e^{hX_t} \geq e^{h\alpha}) \leq e^{-h\alpha}\mathbf{E}(e^{hX_t}). \tag{4.3}$$

But

$$\begin{aligned} \mathbf{E}(e^{hX_t}) &= \mathbf{E}\left(e^{hX_{t-1}}e^{h(X_t - X_{t-1})}\right) \\ &= \mathbf{E}(e^{hX_{t-1}}\mathbf{E}(e^{h(X_t - X_{t-1})} \mid G_0, \ldots, G_{t-1})) \end{aligned} \tag{4.4}$$

since[3] $\mathbf{E}(AB) = \mathbf{E}(A\mathbf{E}(B \mid C))$ for any random variables A, B and C with A a function of C.

Since e^{hx} is convex on $[-c_t, c_t]$, we have for $x \in [-c_t, c_t]$

$$e^{hx} \leq \frac{1}{2}(e^{hc_t} + e^{-hc_t}) + \frac{x}{2c_t}(e^{hc_t} - e^{-hc_t}).$$

So

$$\begin{aligned} &\mathbf{E}(e^{h(X_t - X_{t-1})} \mid G_0, \ldots, G_{t-1}) \\ &\leq \mathbf{E}\left(\frac{1}{2}(e^{hc_t} + e^{-hc_t}) + \frac{X_t - X_{t-1}}{2c_t}(e^{hc_t} - e^{-hc_t}) \mid G_0, \ldots, G_{t-1}\right) \\ &\leq \frac{1}{2}(e^{hc_t} + e^{-hc_t}) \qquad \text{as } X_0, X_1, \ldots \text{ is a supermartingale} \\ &= \cosh(hc_t) = \sum_{k=0}^{\infty} \frac{(c_t h)^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{c_t^{2k} h^{2k}}{2^k k!} = e^{\frac{1}{2}c_t^2 h^2}. \end{aligned} \tag{4.5}$$

[3]See, for example, Grimmett and Stirzaker [25, Section 3.7]). Technically, in this application we are not conditioning on a variable C but a sub-σ-field of the σ-field of our random process. This works the same way since we can always define a random variable with distinct values on the blocks of the partition of the underlying set induced by the sub-σ-field.

So (4.4) now becomes

$$\mathbf{E}(e^{hX_t}) \leq \mathbf{E}(e^{hX_{t-1}})e^{\frac{1}{2}c_t^2h^2} \leq e^{\frac{1}{2}\sum_{i=1}^{t} c_i^2h^2}$$

by induction. So from (4.3),

$$\mathbf{P}(X_t \geq \alpha) \leq e^{-h\alpha}e^{\frac{1}{2}\sum c_i^2h^2}.$$

From this the best bound is obtained by putting $h = \frac{\alpha}{\sum c_i^2}$, which gives the lemma. ∎

A *submartingale* is defined like a supermartingale, but with $\mathbf{E}(X_{i+1} \mid G_0, G_1, \ldots, G_i) \geq X_i$. Since this is equivalent to $-X_i$ being a supermartingale, statements like Lemma 4.2 also provide upper bounds on the lower tails of submartingales. Since a martingale is both a submartingale and a supermartingale, Lemma 4.1 follows.

4.2 Use of stopping times

The concept of a *stopping time* with respect to a random process is almost indispensible. This is any random variable T with values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ such that[4] one can determine whether $T = n$ from knowledge of the process up to time n. The name can be misleading, since a process does not necessarily *stop* when it reaches a stopping time.

The key result we require of stopping times is the following from [25, Section 12.4]. It says that if a supermartingale $\langle X_i \rangle$ is stopped at a stopping time (i.e., $\langle X_i \rangle$ becomes static for all time after the stopping time) then the result is a supermartingale. In the case of processes of finite length, this is a special case of the Optional Sampling Theorem. We denote $\min\{i, T\}$ by $i \wedge T$.

Theorem 4.1 *If, with respect to some process, $\langle X_i \rangle$ is a supermartingale and T is a stopping time, then $\langle X_{i \wedge T} \rangle$ is also a supermartingale with respect to the same process.*

Many applications of Lemma 4.2 to processes will benefit slightly from the use of stopping times. A common situation is that the expected increase of a variable X_i, per step of the process, can be bounded above. It will then follow that the variable is unlikely to increase much faster than at the rate suggested by that upper bound. This turns out to be such an important application that we state it as a separate result. (This contains Lemma 4.2 in the case $b = 0$.) Stopping times help by showing that the bound given by Lemma 4.2 for the deviation of X_t applies with the same probability for all of the X_i, $i \leq t$ (which appears to give something for free):

[4]Technically, T is defined with respect to a filtration; that is, a nested sequence of σ-fields $\{\mathcal{F}_i\}$, with $\mathcal{F}_0$ trivial. In the case of random graph processes, $\{\mathcal{F}_i\}$ is just the σ-field generated by the graph subsequences $G_0, G_1, \ldots, G_i$, which is equivalent to the partition of the underlying set of the probability space induced by all possible outcomes of the graph sequence up to time i. The random variable T is a stopping time if it has values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ and the event $\{T = n\}$ is measurable with respect to $\mathcal{F}_n$.

Corollary 4.1 *Let $G_0, G_1, \ldots, G_t$ be a random process and X_i a random variable determined*[5] *by $G_0, G_1, \ldots, G_i$, $0 \le i \le t$. Suppose that for some real b and constants c_i, $\mathbf{E}(X_i - X_{i-1} \mid G_0, G_1, \ldots, G_{i-1}) < b$ and $|X_i - X_{i-1} - b| \le c_i$ for all $1 \le i \le t$. Then for all $\alpha > 0$,*

$$\mathbf{P}\left(\exists i \ (0 \le i \le t) \ : \ X_i - X_0 \ge ib + \alpha\right) \le \exp\left(-\frac{\alpha^2}{2\sum c_j^2}\right).$$

Proof. Note that $Y_i = X_i - X_0 - ib$ defines a supermartingale whose differences are bounded above by c_i. Let T be the stopping time which is defined as the first time that $Y_i \ge \alpha$. By Theorem 4.1, $\langle Y_{i \wedge T} \rangle$ is a supermartingale. Applying Lemma 4.2 to this supermartingale at time t gives the corollary. ∎

Note. Now that we have stopping times under our belt, we can give a slightly slicker proof of Lemma 4.2 and Corollary 4.1. From (4.5)

$$\mathbf{E}(e^{hX_i} \mid G_0, \ldots, G_{i-1}) \le e^{hX_{i-1}} e^{\frac{1}{2}c_i^2 h^2}.$$

It follows that

$$Y_i = e^{hX_i} e^{-\frac{1}{2}s_i h^2} \qquad \text{where} \qquad s_i = \sum_{j=1}^{i} c_j^2$$

defines a supermartingale (with respect to the same process as $\langle X_i \rangle$). Define the stopping time T_0 to be $\min\{i \ : \ X_i > \alpha\}$, and $T = t \wedge T_0$. By Theorem 4.1, $\langle Y_T \rangle$ is a supermartingale. So $\mathbf{E}Y_T \le \mathbf{E}Y_0 = 1$; that is,

$$\mathbf{E}(e^{hX_T} \le e^{\frac{1}{2}s_T h^2}) \le e^{\frac{1}{2}s_t h^2}.$$

So

$$\mathbf{P}(X_i - X_0 \ge \alpha \ \text{ for some } \ 0 \le i \le t) \le e^{\frac{1}{2}s_t h^2 - h\alpha}.$$

As before put $h = \frac{\alpha}{s_t}$, to obtain Corollary 4.1 in the case $b = 0$. (The general case is similar, again by considering $Y_i = X_i - X_0 - ib$.)

Besides being used to prove the main theorem, Corollary 4.1 can be especially useful when a process is nearing its completion, or nearing the transition to another phase, and the trends in some of the relevant variables become erratic, which is typical if the variables are discrete and their values become small. One of the variables, or a function of them, can have its behaviour bounded at least in one direction by bounding the expected differences. Examples occur in establishing the phase transitions in the degree-greedy algorithm for independent sets in Section 7.1 and also the final part of the process finding the k-core in Theorem 6.4.

A simple application is given next.

4.3 Large components in d-processes

One of the central themes in the study of random graphs is the *phase transition*, being the time at which the largest component of the evolving random graph $\mathcal{G}(n, m)$ grows from "very

[5] To be precise, X_t is measurable with respect to the σ-field generated by $G_0, G_1, \ldots, G_i$.

small" to "very large". "Very small" can be defined as smaller than any fixed power of n, and "very large" can be larger than a constant times n. See Janson et al. [28] for many interesting results on this. (This usage of "phase transition" is a little different from the phases of processes in the present paper.)

Consider the degree bounded graph process (d-process) with $d \geq 3$, and define $l(G)$ to be the number of vertices in the largest component of G_t. An unsolved problem is to show that this process exhibits a phase transition in the above sense. In particular, does there exist $c > 0$ such that for all $\epsilon > 0$,

$$\mathbf{P}(l(G_t) < n^{0.1}) \to 1 \text{ for } t < n(c-\epsilon) \quad \text{and} \quad \mathbf{P}(l(G_t) > n^{0.9}) \to 1 \text{ for } t > n(c+\epsilon).$$

This occurs at $c = \frac{1}{2}$ in $\mathcal{G}(n,m)$. In [52] it is shown that if the phase transition occurs in the random d-process, $d \geq 3$, then $c \leq (\frac{d}{2} - \frac{1}{6})n$. This was done by showing a version of the following which is slightly stronger (and for general d). The argument is restricted here to $d = 3$ for simplicity, and without trying to get the best constants. In any case, the argument here cannot reduce t_0 below n, whereas the phase transition no doubt occurs well before this time.

Proposition 4.1 *For $t_0 = \lceil \frac{5}{4}n \rceil$ in the random 3-process, $l(G_{t_0}) \geq \frac{n}{100}$ a.a.s.*

Proof. This is based on the fact that if there is no large component then there is a good chance of joining two components. Every time this happens, the number of components goes down. This cannot happen more than $n-1$ times!

If $l(G_t) \geq \frac{n}{100}$ at time $t > 0$, we say that the process is *successful* at time t, and let $S(t)$ be the indicator variable for this event.

Consider $G(t)$ in the case that the 3-process is *not* successful at time $t \leq \frac{5n}{4}$. All components have at most $\frac{n}{100}$ vertices, and by counting edges, the number of vertices of degree 3 is at most $\frac{5n}{6}$. It follows that when the next edge A_{t+1} is chosen, there are at least $\frac{n}{6} - 1$ choices for each of its ends u and v. We can regard u as being chosen first. Then since all components have at most $\frac{n}{100}$ vertices, the probability that v is in the same component as u (this is conditional on G_t) is at most $(\frac{n}{100} - 1)/(\frac{n}{6} - 1) < \frac{6}{100}$. Thus for $G(t)$ such that $S(t) = 0$,

$$\mathbf{P}(D(t+1) = D(t) - 1 \mid G_t) > \frac{94}{100}$$

where $D(t)$ denotes the number of components of G_t. The only other possibility is $D(t+1) = D(t)$. Thus

$$\mathbf{E}(D(t+1) - D(t) \mid G_t) \leq -0.94\,.$$

Defining

$$Y(t) = \begin{cases} D(t) & S(t-1) = 0 \\ Y(t-1) - 0.94t & \text{otherwise,} \end{cases}$$

it follows that for all $G(t)$, $\mathbf{E}(Y(t+1) - Y(t) \mid G_t) \leq 0.94$. Moreover, the differences satisfy $|Y(i) - Y(i-1) + 0.94| \leq 0.94$. Applying Corollary 4.1 (with $c = 1$ for convenience, and

noting $Y(0) = n$),

$$\mathbf{P}(Y(t) \geq n - 0.94t + n^{\frac{2}{3}}) \leq \exp(-n^{\frac{4}{3}}/(2t)) \leq \exp(-2n^{\frac{1}{3}}/5) \to 0 \tag{4.6}$$

for $t \leq \frac{5}{4}n$. But if the process is unsuccessful at time $t_0 - 1$, $Y(t_0) = D(t) > 0$ always, whereas (4.6) implies that $Y(t_0) < -\frac{1}{6}n$ a.a.s. Thus, a.a.s. the process is successful by time t_0, and the proposition follows. ∎

4.4 Bounded differences with a tail

The large deviation inequalities in this article use the method of bounded differences. However, it is a mistake to think that the method only applies when there are relatively small sure bounds (i.e., bounds which always hold) on the differences as in Corollary 4.1. Occasionally, the only sure bound on the differences is too large to be useful, and yet the differences are small with high probability. Shamir and Spencer [53] were apparently the first to exploit this feature in connection with graph theory, and the consequences of their proof were stated more generally in McKay and Wormald [38, Lemma 3.1]. The method of proof there uses the technique as if the differences were all small, but with an extra twist to cope with the occasional large differences. Godbole and Hitczenko [23] give a survey of results of this type. We call this situation "differences with a tail" to distinguish it from the case where the sure upper bounds on differences are just as useful as the almost sure ones.

Although the large deviation inequalities given so far seem to require sure upper bounds, no separate lemma of that type is needed for differences with a tail, as they can be taken care of essentially by stopping the process appropriately (as in the proof of Theorem 5.1). The processes for independent sets in regular graphs (Section 7.1) and for packing use analysis involving differences with a tail and require this aspect of Theorem 5.1.

5 Proving approximation by differential equations

Having large deviation inequalities as in the previous section, there are various ways to go about showing, for a variable $Y(t)$ in a random process, that $Y(t) \approx f(t)$ for some function f arising from the solution of one or more differential equations. A quite general situation is treated first, in which the theorem usually serves to show the approximation holds with probability $1 - o(e^{-n^{\epsilon}})$ for some $\epsilon > 0$. Additionally, it usually shows the error in the approximation is quite small (typically $o(n^{-\epsilon'})$, but has a Lipschitz condition which often causes this accuracy to be lost near the end of the process. However, in such situations the technique used in the proof is often good enough to recover this accuracy to a large extent.

For convenience, the general theorem breaks time up into little pieces and considers a supermartingale and submartingale in each piece. Another approach is to transform the random variable into something "close" to a martingale throughout the whole process (or almost all of it) so that the large deviation inequality can be applied just once. This tends to produce better bounds since, for example, with $c_i = c$ for each i, the deviation inequalities produce small probabilities when the deviation α is a constant times the square root of the

number of steps. So one long supermartingale is better than many shorter ones. The same effect occurs if one considers bounding the total number of heads, X, occurring in n fair coin tosses by summing bounds on the deviations due to $\sqrt{n}$ successive runs of $\sqrt{n}$ coin tosses each. In each case the standard deviation is $Cn^{1/4}$, so the sum of these bounds will be of the order of $Cn^{3/4}$. On the other hand, the standard deviation of X is $\frac{1}{2}\sqrt{n}$, and by considering the process as a whole, one begins to get meaningful bounds for deviations of this order of magnitude. We call this the *wholistic* approach. It is not so easy to write a general theorem using this principle, so instead we give an example in Section 5.2.

5.1 General-purpose theorem

For the main theorem we lay a fairly general setting. The random processes are any discrete time random processes. Such a process is a probability space which can be conveniently denoted by $(Q_0, Q_1, \ldots)$ where each Q_i takes values in some set S. The elements of the space 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.

Consider a sequence of random processes indexed by n, $n = 1, 2, \ldots$. 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. $S^{(n)+}$ denotes 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

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

for all $(u_1, \ldots, u_j)$ and $(v_1, \ldots, v_j)$ in D. We call L a *Lipschitz constant* for f, and note that $\max_{1 \leq i \leq j} |u_i - v_i|$ is the distance between $(u_1, \ldots, u_j)$ and $(v_1, \ldots, v_j)$ in the ℓ^∞ metric. (For the existence of a solution to a set of differential equations, a Lipschitz condition is only needed on the variables after the first, but for our purposes here we demand it on all of them. Note that ℓ^1 distance was used in [61], but the ℓ^∞ distance fits a little more nicely into the proof. Both distances give equivalent theorems.)

We scale both variable values and time by a factor of n because for many combinatorial applications this gives them a fixed limiting distribution. This is convenient when considering the solution of the corresponding differential equations: there is only one set of equations rather than different equations for each n. Whenever the need for different scalings of variables and time arises, time can be pre-scaling as in the proof of Theorem 7.1.

For variables $Y_1, \ldots, Y_a$ defined on the components of the process, and $D \subseteq \mathbb{R}^{a+1}$, define the *stopping time* $T_D(Y_1, \ldots, Y_a)$ to be the minimum t such that $(t/n, Y_1(t)/n, \ldots, Y_a(t)/n) \notin D$. This is written as T_D when $Y_1, \ldots, Y_a$ are understood from the context.

The following theorem is a generalisation of the two theorems in [61], and has stronger conclusions. Although only very few changes are required in the proof, we include the full proof here since some of the details skimmed in the proof in [61] require more careful examination in the present setting.

In the statement of the theorem, "uniformly" refers to the convergence implicit in the $o()$ terms. Hypothesis (i) ensures that $Y_l(t)$ 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 in time (or as the values of the variables change).

The main differences between this theorem and the first theorem in [61] are that the probability that the variables of the process are not well approximated by the solution of the differential equation can now be made very small, and the error of approximation can be made small, if the errors in hypotheses (i) and (ii) are small. Also, the note after the theorem gives a version with the number of variables unbounded. There does not seem such a great need for this version so the main theorem is stated without this option to avoid further complexity.

Theorem 5.1 *For $1 \le l \le a$, where a is fixed, let $y_l : S^{(n)+} \to \mathbb{R}$ and $f_l : \mathbb{R}^{a+1} \to \mathbb{R}$, such that for some constant C_0 and all l, $|y_l(h_t)| < C_0 n$ for all $h_t \in S^{(n)+}$ for all n. Let $Y_l(t)$ denote the random counterpart of $y_l(h_t)$. Assume the following three conditions hold, where in (ii) and (iii) D is some bounded connected open set containing the closure of*

$$\{(0, z_1, \ldots, z_a) : \mathbf{P}(Y_l(0) = z_l n, 1 \le l \le a) \ne 0 \text{ for some } n\} .$$

(i) (Boundedness hypothesis.) For some functions $\beta = \beta(n) \ge 1$ and $\gamma = \gamma(n)$, the probability that

$$\max_{1 \le l \le a} |Y_l(t+1) - Y_l(t)| \le \beta,$$

conditional upon H_t, is at least $1 - \gamma$ for $t < T_D$.

(ii) (Trend hypothesis.) For some function $\lambda_1 = \lambda_1(n) = o(1)$, for all $l \le a$

$$| \mathbf{E}(Y_l(t+1) - Y_l(t) \mid H_t) - f_l(t/n, Y_1(t)/n, \ldots, Y_a(t)/n) | \le \lambda_1$$

for $t < T_D$.

(iii) (Lipschitz hypothesis.) Each function f_l is continuous, and satisfies a Lipschitz condition, on

$$D \cap \{(t, z_1, \ldots, z_a) : t \ge 0\},$$

with the same Lipschitz constant for each l.

Then the following are true.

(a) For $(0, \hat{z}_1, \ldots, \hat{z}_a) \in D$ the system of differential equations

$$\frac{dz_l}{dx} = f_l(x, z_1, \ldots, z_a), \qquad l = 1, \ldots, a$$

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

$$z_l(0) = \hat{z}_l,$$

$1 \le l \le a$, and which extends to points arbitrarily close to the boundary of D;

(b) Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large constant C, with probability $1 - O(n\gamma + \frac{\beta}{\lambda}\exp(-\frac{n\lambda^3}{\beta^3}))$,

$$Y_l(t) = nz_l(t/n) + O(\lambda n) \tag{5.1}$$

uniformly for $0 \le t \le \sigma n$ and for each l, where $z_l(x)$ is the solution in (a) with $\hat{z}_l = \frac{1}{n}Y_l(0)$, and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within ℓ^∞-distance $C\lambda$ of the boundary of D.

Note. A version of the theorem also holds with the number a of variables a function of n, the domain D a function of n but with all Lipschitz constants uniformly bounded, and with the probability in (b) replaced by

$$1 - O\left(an\gamma + \frac{a\beta}{\lambda}\exp\left(-\frac{n\lambda^3}{\beta^3}\right)\right),$$

provided each function f_l depends only on x and $z_1, \ldots, z_l$. This last condition is to avoid complicated issues around the solutions of infinite sets of differential equations: in this case, the solution of the infinite system is defined to be the set of functions solving all the finite systems obtained for each fixed l by restricting to the equations for the derivatives of $z_1, \ldots, z_l$.
Proof. There is a unique solution in (a) by a standard result in the theory of first order differential equations. (See Hurewicz [27, Chapter 2, Theorem 11].)

To present the proof, we simplify notation by considering first $a = 1$ and refer to y_1, z_1 and f_1 as y, z and f, and so on. We also initially take $\gamma = 0$, so that the inequality in the boundedness hypothesis is deterministically true. The modification for general a and γ is referred to at the end.

Taking $\lambda > \lambda_1$ as in (b), define

$$w = \left\lceil \frac{n\lambda}{\beta} \right\rceil \tag{5.2}$$

and let $t \ge 0$. Note immediately that if $\beta/\lambda > n^{1/3}$ there is nothing to prove because the probability in the conclusion is not restricted. Hence $w \ge n^{2/3}$. Similarly, we can assume $\lambda < 1$ by the assumption $\lambda = o(1)$, or by noting the approximation in the conclusion is otherwise trivially satisfied. The trend hypothesis 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

$$Y(t + w) - Y(t).$$

We will assume for now that for a sufficiently large positive constant C, $(t/n, y(t)/n)$ is ℓ^∞-distance at least $C\lambda$ from the boundary of D. The size of C required is determined from the Lipschitz constant in the Lipschitz hypothesis; wherever this hypothesis is used in the following argument, we need to know that the point is still inside D. For $0 \le k < w$, noting that $k\beta/n = O(\lambda)$, we have by the trend hypothesis

$$\begin{aligned} \mathbf{E}(Y(t + k + 1) - Y(t + k) \mid H_{t+k}) &= f((t + k)/n, Y(t + k)/n) + O(\lambda_1) \\ &= f(t/n, Y(t)/n) + O(\lambda_1 + k\beta/n) \end{aligned} \tag{5.3}$$

by the Lipschitz hypothesis as $|Y(t+k) - Y(t)| \leq k\beta$ by the boundedness hypothesis. Thus there exists a function

$$g(n) = O(\lambda_1 + w\beta/n) = O(\lambda) \tag{5.4}$$

such that conditional on H_t,

$$Y(t+k) - Y(t) - kf(t/n, Y(t)/n) - kg(n)$$

$(k = 0, \ldots, w)$ 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 the boundedness hypothesis, at most

$$\beta + |f(t/n, Y(t)/n) + g(n)| \leq \beta + O(1) \leq \kappa\beta \tag{5.5}$$

for some constant $\kappa > 0$ (using the boundedness of D and again the Lipschitz hypothesis to make sure the variables do not leave D, and the fact that β is bounded below). So by Lemma 4.2 (and replacing $\alpha^2/2w\kappa^2\beta^2$ by a new α),

$$\mathbf{P}\left(Y(t+w) - Y(t) - wf(t/n, Y(t)/n) \geq wg(n) + \kappa\beta\sqrt{2w\alpha} \mid H_t\right) \leq e^{-\alpha} \tag{5.6}$$

for all $\alpha > 0$. We will now set

$$\alpha = \frac{n\lambda^3}{\beta^3}. \tag{5.7}$$

The lower tail of $Y(t+w) - Y(t) - wf(t/n, Y(t)/n)$ can be bounded by exactly the same argument, but using a submartingale (or, as an alternative, negate the function and use a supermartingale). Hence (increasing κ by a constant factor) we have

$$\mathbf{P}\left(|Y(t+w) - Y(t) - wf(t/n, Y(t)/n)| \geq wg(n) + \kappa\beta\sqrt{w\alpha} \mid H_t\right) \leq 2e^{-\alpha}. \tag{5.8}$$

Now define $k_i = iw$, $i = 0, 1, \ldots, i_0$ where $i_0 = \lfloor \sigma n/w \rfloor$. We next show by induction that for each such i,

$$\mathbf{P}(|Y(k_j) - z(k_j/n)n| \geq B_j \text{ for some } j \leq i) = O(ie^{-\alpha}) \tag{5.9}$$

where

$$B_j = Bw\left(\lambda + \frac{w}{n}\right)\left(\left(1 + \frac{Bw}{n}\right)^j - 1\right)\frac{n}{Bw}$$

and B is a constant determined below.

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

$$|Y(k_{i+1}) - z(k_{i+1}/n)n| = |A_1 + A_2 + A_3 + A_4|$$

where

$$\begin{aligned}
A_1 &= Y(k_i) - z(k_i/n)n, \\
A_2 &= Y(k_{i+1}) - Y(k_i) - wf(k_i/n, Y(k_i)/n), \\
A_3 &= wz'(k_i/n) + z(k_i/n)n - z(k_{i+1}/n)n, \\
A_4 &= wf(k_i/n, Y(k_i)/n) - wz'(k_i/n).
\end{aligned}$$

The inductive hypothesis (5.9) gives that

$$|A_1| < B_i, \tag{5.10}$$

together with the similar statements for smaller i, all hold with probability $1-O(ie^{-\alpha})$. When this is true, by (5.8) and using (5.2), (5.4) and (5.7) we have with probability $1-O(\exp(-\alpha))$

$$|A_2| < B'w\lambda$$

for a suitable universal constant B'. (This is the point at which the assumption included above, about the scaled variables not approaching within distance $C\lambda$ of the boundary of D, is justified. It follows from the inductive hypothesis using the fact, seen below, that $B_i = O(n\lambda)$.) Since z is the solution given in (a) and f satisfies the Lipschitz hypothesis we have

$$|(w/n)z'(k_i/n) - (z(k_{i+1}/n) - z(k_i/n))| = O(w^2/n^2) \le B''w^2/n^2$$

(for n sufficiently large) for a suitable constant B'', and so

$$|A_3| \le \frac{B''w^2}{n}.$$

Again using the same two facts, as well as (5.10), we obtain

$$|A_4| \le \frac{B''wB_i}{n}$$

(redefining B'' appropriately). Set $B = \max\{B', B''\}$. Summing the bounds on the A_i now gives

$$|Y(k_{i+1}) - z(k_{i+1}/n)n| < B_{i+1}$$

with probability $1 - o((i+1)e^{-\alpha})$; that is, we have (5.9) by induction for $i \le i_0$.

Note that

$$B_i = O(n\lambda + w) = O(n\lambda)$$

since β is bounded below. Also for any $t \le \sigma n$, put $i = \lfloor t/w \rfloor$. From time k_i to t the change in Y and z is at most $w\beta = O(n\lambda)$. Hence from (5.9), with probability $1 - O((n/w)e^{-\alpha})$

$$|Y(t) - z(t/n)n| = O(\lambda n).$$

The theorem now follows (for the case $a = 1$) from (5.2) and (5.7).

Finally, the modifications required for arbitrary a and γ are very simple. The modification for arbitrary γ is just to alter the process leading up to (5.5), by conditioning on the event that the inequality in (i) holds at each step. This alters the expected change in Y given in (ii), but since $Y \le C_0 n$ it can only alter it by at most $C_0 n\gamma$. So just replace λ_1 by $\lambda_1 + C_0\gamma$ in the rest of the proof, and note that the probability that any of these events fails to hold throughout the process is $O(n\gamma)$, so subtract this from the probability that (5.1) holds. The result follows.

Next for $a > 1$: the inductive hypothesis (5.9) is modified to

$$\mathbf{P}(|Y_l(k_j) - z_l(k_j/n)n| \geq B_j \text{ for some } j \leq i) = O(aie^{-\alpha})$$

for all $l \leq a$. The proof of the same statement for $i+1$ is modified only in that the statement has to be verified for each of a variables, and so the probability of failure is multiplied by a. The theorem follows when a is fixed, and the Note after the theorem follows for a arbitrary. ∎

5.2 The wholistic approach

This is a way of applying the large deviation inequalities alternative to that in Section 5.1. The general idea is to transform the random variable to something close to a martingale. This seems to be difficult to carry out effectively in general without knowing something about the solution of the associated differential equations.

Suppose that Y is one of the variables under examination and the solution to the differential equations gives the function $z(x)$, so that $Y(t)$ is to be approximated by $nz(t/n)$. A simple idea is to consider the random variable $Y(t) - nz(t/n) + f(t)$ where f is some slowly growing function. This should be a supermartingale if f is sufficiently large and positive, and a submartingale if f is large negative. Even this can give a significant improvement over Theorem 5.1. How well the approximation works depends on the stability of the differential equations, and how badly the Lipschitz condition fails. One particular improvement on this idea was first made, at least in a graph process setting, in [44]. This uses the general solution of the differential equations to find a function which is constant along all trajectories (as opposed to the one above, which is only 0 along the trajectory whose initial conditions correspond to those of the process). This function will be close to a martingale, and the trick of adding and subtracting a small function can be used to create a true submartingale and supermartingale.

For an illustration, consider the number of isolated vertices in the random 2-process, denoted by Y_0. As well as demonstrating the "wholistic" approach described at the start of Section 5, this will also show how to cope with the problem often encountered near the end of processes where the Lipschitz hypothesis fails (see Section 5.4 for more discussion on this.) The argument here is in a form which generalises easily to the multivariate case.

Theorem 5.2 *Take* $0 < \delta < \min\{3\epsilon, \frac{1}{3} + \frac{1}{2}\epsilon\}$. *For sufficiently large* κ', *with probability at least* $1 - \exp\left(-n^{2/3+\epsilon-2\delta}/\kappa'\right)$

$$n - t = Y_0(t)\left(1 + O(n^{-\delta}) + \frac{1}{2}\ln\frac{n}{Y_0(t)}\right) \tag{5.11}$$

for all $0 \leq t < \lfloor n - n^{2/3+\epsilon} \rfloor$.

Proof. Recall that $Y_0(t)$ is approximated by $nz(\frac{t}{n})$ where z satisfies the associated differential equation (3.13) whose general solution is (3.14). This can be put in the form

$$\frac{2(x-1)}{z} - \ln z = C. \tag{5.12}$$

It is a little simpler now to work with unscaled variables, so note that the unscaled differential equation is

$$Y_0' = \frac{-2Y_0}{2n-2t-Y_0}. \tag{5.13}$$

Defining

$$H(\mathbf{w}) = \frac{2(t-n)}{Y_0} - \ln\frac{Y_0}{n}, \qquad \mathbf{w} = (t, Y_0)$$

in view of (5.12), the general solution of (5.13) is

$$H(\mathbf{w}) = C.$$

We will take $\mathbf{w} = \mathbf{w}_t = (t, Y_0(t))$, and consider the sequence of random variables $\langle H(\mathbf{w}_t)\rangle$.

By counting vertex degrees in the 2-process,

$$Y_0 \le n - t \tag{5.14}$$

always. Given this, note that the second order partial derivatives of $H(\mathbf{w})$ with respect to t and Y_0 are $O(1/Y_0^2)$. Therefore

$$H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) = (\mathbf{w}_{t+1} - \mathbf{w}_t)\cdot \text{ grad } H(\mathbf{w}_t) + O\left(\frac{1}{Y_0(t)^2}\right) \tag{5.15}$$

where $\cdot$ denotes the scalar product and grad $H = (H_t, H_{Y_0})$, provided $Y_0(t+1) \ge 1$. Note that

$$\begin{aligned}\mathbf{E}(\mathbf{w}_{t+1} - \mathbf{w}_t \mid G_t) &= (1, \mathbf{E}(Y_0(t+1) - Y_0(t) \mid G_t)) \\ &= \left(1, \frac{-2Y_0(t)}{2n-2t-Y_0(t)} + O\left(\frac{n}{(n-t)^2}\right)\right)\end{aligned}$$

from (3.7), (3.9) and (3.10), (again using (5.14) in estimating the error) provided the error term is $o(1)$; that is, provided $n - t >> \sqrt{n}$. Also observe that

$$\left(1, \frac{-2Y_0(t)}{2n-2t-Y_0(t)}\right)\cdot \text{ grad } H(\mathbf{w}_t) = 0$$

as $H(\mathbf{w})$ is constant along every trajectory $\mathbf{w}$ of (5.13). So, noting that

$$\text{grad } H(\mathbf{w}) = \left(\frac{O(1)}{Y_0}, \frac{O(1)}{Y_0} + \frac{O(n-t)}{Y_0^2}\right) \tag{5.16}$$

and taking the expectation of (5.15) conditional on G_t we obtain

$$\mathbf{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) = O\left(\frac{n}{Y_0(t)^2(n-t)}\right) \tag{5.17}$$

assuming $3 \le Y_0(t) \le n - t$ (which ensures $Y_0(t+1) \ge 1$) and $\sqrt{n} = o(n-t)$.

There is the problem that $H(\mathbf{w})$ is undefined if the process hits $Y_0 = 0$. The simple remedy is to stop the sequence $H(\mathbf{w}_t)$ as soon as it hits $Y_0 \le 2$. More than that, since $H(\mathbf{w}_0) = -2$, introduce the stopping time

$$T = \min\{t \,:\, |H(\mathbf{w}_t) + 2| \ge n^{-\delta}\}$$

and note from the definition of H that $Y_0 > (n-t)/\ln n$ for $t < T$. Then from (5.17),

$$\mathbf{E}(H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T}) \mid G_t) = O\left(\frac{n}{Y_0(t)^2(n-t)}\right) = O\left(\frac{n(\ln n)^2}{(n-t)^3}\right)$$

for $0 \le t \le n - n^{1/2+\epsilon}$. From (5.16) and the fact that Y_0 and t change by at most 2 per step of the process, we also have

$$|H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T})| = O\left(\frac{n-t}{Y_0(t)^2}\right) = O\left(\frac{(\ln n)^2}{n-t}\right).$$

Put $t_0 = n - n^{2/3+\epsilon}$ and apply Corollary 4.1 to the sequences $\langle H(\mathbf{w}_{t\wedge T})\rangle$, $t \le t_0$, and symmetrically $\langle -H(\mathbf{w}_{t\wedge T})\rangle$, with $\alpha = \frac{1}{2}n^{-\delta}$, $b = \kappa(\ln n)^2 n^{-3\epsilon}$ and $c_j = \kappa(\ln n)^2/(n-j)$, for some sufficiently large fixed κ. Noting that

$$\sum_{j=1}^{t_0} c_j^2 = O(n^{-2/3-\epsilon})(\ln n)^2$$

and since $\delta < 3\epsilon$, the conclusion is

$$\mathbf{P}\left(|H(\mathbf{w}_{i\wedge T}) - H(\mathbf{w}_0)| \ge \frac{2}{3}n^{-\delta} \quad \text{for some } 0 \le i \le t_0\right) \le \exp\left(-n^{2/3+\epsilon-2\delta}/\kappa'\right)$$

for κ' sufficiently large. As $H(\mathbf{w}_0) = -2$, this implies from the definition of T that $T < t_0$ with the same low probability. But $T \ge t_0$ implies (5.11) as required. ∎

Note. It is possible to get good almost sure bounds on Y_0 even much closer to the end of the process. (See Section 5.4.)

5.3 Simple applications

Here the sharp concentration results in Section 3 are obtained as applications of Theorem 5.1.

Proof of Theorem 3.1. Note that G_t is determined by the history H_t. In view of (3.2), we can set $f_l = -\delta_{i0} + \delta_{i1} + z_{i-1} - z_i$ for $(i \ge 0)$ where z_{-1} is identically 0, and let D be defined

by $z_0 > 0$ and (say) $-1 < x < 3$ and $-1 < z_i < 2$ for all $i > 0$. Then the trend hypothesis of Theorem 5.1 holds with

$$\lambda_1 = \max \frac{|Y_{i-1}(t) - Y_i(t)|}{n(n-1)} \leq \frac{1}{n-1}$$

since $0 \leq Y_i \leq n$ always. The boundedness hypothesis holds with $\beta = 2$ and $\gamma = 0$, and we can take a a function of n by the Note after the theorem. The conclusion (b) gives Theorem 3.1, since the relation between λ and s ensures that $z_0(s/n) >> \lambda$. ∎

Note that the method in Section 5.2 can be applied to Y_0 for example to get a sharper result for that particular variable, and with some work can potentially give sharper results for all the variables.

For degree bounded graph processes, as for the later phases of the min-degree process, the random vector under consideration is not Markovian (due to the nuisance edges). Nevertheless, Theorem 5.1 applies. For the next theorem the domain D also has to be chosen carefully to avoid the singularity in the functions f_l as x approaches $nd/2$.

Proof of Theorem 3.2. This is very similar to the previous proof, but using (3.18) for the trend hypothesis. The choice of the domain D is the only problem. There is a singularity in the functions f_i on the right in (3.19). So choose a domain D with $z_d > \epsilon'$ for some fixed ϵ', and Theorem 5.1 can be applied. The statement of Theorem 3.2 follows, except only for all $0 \leq t < n\hat{x}$ where $\hat{x}$ is the least x for which $z_d(x) < \epsilon' + C\lambda$. But since Y_d is well approximated by $nz_d(t/n)$ inside D, by (3.16) it is impossible that $\hat{x} < d/2$. (The solution to the differential equations in Section 3.3.3 can be used alternatively at this point.) The theorem follows. ∎

The verification of the statements about random star d-processes (Section 3.4) and the simple greedy algorithm for independent sets (Section 3.5) are similar.

5.4 Obtaining higher precision

As mentioned in Section 3.3.2, Theorem 3.2 can easily be improved, in the sense that good approximation by the solutions of the differential equations can be shown to hold much closer to the end of the process. This applies to basically all the other applications of Theorem 5.1 in which the function giving the expected trends in the process becomes highly non-Lipschitz near the very end of the natural process.

One remedy is to use the wholistic approach, but it is hard to say how generally that type of argument applies. A quite general approach is to rework the proof of Theorem 5.1 in any particular situation, using different Lipschitz constants depending on what part of the domain D the process has reached. This is straightforward enough given the inductive nature of the proof. (For a similar argument, see the proof of Theorem 7.1.) In this way one can easily extend the result of Theorem 5.1 to time $n^{1-\epsilon}$ from the end of the process. This gives sufficient accuracy to approach within $n^{1-\epsilon}$ steps of the end of the process in the case of d-processes, for instance; a version of this argument was used in [50] and gave sufficient accuracy for the arguments there about saturation of processes.

Moreover, in the case of 2-processes, the differential equation has a stability property that enables this approach to be pushed to the extreme: we can use this idea to obtain almost sure bounds on the number of isolated vertices and isolated edges in the random 2-process until only $O(\log^4 n)$ steps from the end, the limiting probability that the final graph of 2-process is connected (Hamilton cycle) is obtained in [58]. Other properties such as the expected number of cycles follow similarly. There is no doubt that such stability can be taken advantage of in other situations as well, but no examples seem to be in the random graph or random algorithm literature.

6 Difficulties with transition probabilities

In this section we will estimate the size of the k-core of a random graph. This study has two difficulties not met in this paper until now. One is that the transition probabilities are very hard to compute, and so we spend a good deal of effort estimating them. The second is that, partly as a result of the first difficulty, some possible states of the process are very undesirable because the subsequent transition probabilities have not been estimated. We show how to ignore such states even if they cannot be conveniently described in terms of a fixed domain D. This only requires that the process avoids such states with high probability.

6.1 Eliminating undesirable states

We will use the following variation of Theorem 5.1 in which the possibility exists that the trend and boundedness hypotheses become invalid, due to the process reaching some "unusual" state. If this occurs with a sufficiently low probability, the conclusion of the theorem is similar. In general, we can deduce that a.a.s. the approximation by differential equations is valid until one of the "unusual" states is reached. The "usual" states are determined by a set $\hat{D}$.

Theorem 6.1 *For any set $\hat{D} = \hat{D}(n) \subseteq \mathbb{R}^{a+1}$, define the stopping time $T_{\hat{D}} = T_{\hat{D}(n)}(Y_1, \ldots, Y_a)$ to be the minimum t such that $(t/n, Y_1(t)/n, \ldots, Y_a(t)/n) \notin \hat{D}$. Assume that the first two hypotheses of Theorem 5.1 apply only with the restricted range $t < T_{\hat{D}}$ of t. Then the conclusions of the theorem hold as before, except with $0 \le t \le \sigma n$ replaced by $0 \le t \le \min\{\sigma n, T_{\hat{D}}\}$.*

Proof. For $1 \le j \le a$ define random variables $\hat{Y}_j$ by

$$\hat{Y}_j(t+1) = \begin{cases} Y_j(t+1) & \text{if } t < T_{\hat{D}} \\ Y_l(t) + f_j(t/n, Y_1(t)/n, \ldots, Y_l(t)/n) & \text{otherwise} \end{cases}$$

for all $t + 1 \ge 0$. Then the $\hat{Y}_j$ satisfy the hypotheses of Theorem 5.1, and so the theorem follows as $\hat{Y}_j(t) = Y_j(t)$ for $0 \le t < T_{\hat{D}}$. ∎

6.2 The k-core of a random graph

As mentioned in Section 2, the process considered here starts with a graph G_0 on n vertices and repeats the operation of deleting a random vertex of degree less than k. Here $k \geq 3$ is fixed, and G_0 is a *random* graph in $\mathcal{G}(n, m(0))$. (Here $m(0)$ is used as it denotes the first element of the sequence of random numbers of edges in the deletion algorithm.) The object is to determine the distribution of the number of vertices remaining in the final graph. In this way we can capture the point at which the k-core appears in a random graph in $\mathcal{G}(n, m(0))$. Throughout this discussion, we will be assuming that $m(0)$ is bounded above and below by positive constants times n.

Define a *light* vertex to be one of degree strictly less than k. All other vertices are *heavy*. Instead of actually deleting vertices, it is more convenient to delete edges only. Thus, each step of the process deletes all the edges incident with a randomly chosen non-isolated light vertex x_t from the current graph G_t to produce G_{t+1}. Introduce the random vector

$$\mathbf{w}(t) = (v_0(t), v_1(t), \ldots, v_{k-1}(t), m(t)),$$

where $v_i(t)$ denotes the random number of vertices of degree i in G_t, $0 \leq i \leq k-1$, and $m(t) = |E(G_t)|$. As usual, n is fixed for the process and we are interested in the asymptotic distribution of the variables as $n \to \infty$. Some useful notation is

$$v = \sum_{i=1}^{k-1} v_i, \quad \bar{v} = n - \sum_{i=0}^{k-1} v_i, \quad s = \sum_{i=1}^{k-1} i v_i \quad \bar{s} = 2m - s \tag{6.1}$$

so that v is the number of non-isolated light vertices, $\bar{v}$ the number of heavy vertices, and s and $\bar{s}$ are total degrees of these sets of vertices. The process ends when $\bar{v} = n - v_0$, at which point there are no non-isolated light vertices. In order to examine this point, we will show sharp concentration of each of the entries in $\mathbf{w}(t)$ for appropriate t.

Since G_0 is random in $\mathcal{G}(n, m(0))$, $\mathbf{w}(0)$ is a random vector. The first important observation is that conditional upon $\mathbf{w}(0)$, the distribution of G_0 must be uniform in the set $G(\mathbf{w}(0))$, where $G(\mathbf{w})$ is the set of graphs with variables given by $\mathbf{w}$. The next important observation is that, conditional upon $\mathbf{w}(t)$, G_t is uniformly distributed in $G(\mathbf{w}(t))$. This follows by induction, since, given $\mathbf{w}(t)$ and $\mathbf{w}(t+1)$, and a graph G in $G(\mathbf{w}(t+1))$, the number of graphs $G' \in G(\mathbf{w}(t))$ such that G comes from G' by deleting the edges incident with a non-isolated light vertex is independent of G. (Just consider the reverse operation: select as x_t any vertex of degree 0 in G. The vectors $\mathbf{w}(t+1)$ and $\mathbf{w}(t)$ determine the degree of x_t in G', the degrees of the light vertices which must join to x_t, and also the number of heavy vertices so joining. All these vertices can be chosen in any way from the sets of vertices of the required degrees. The number of ways of making this multiple choice is determined precisely by $\mathbf{w}(t+1)$.) Thus, conditioning on $\mathbf{w}(t)$ and $\mathbf{w}(t+1)$, the graphs in $G(\mathbf{w}(t+1))$ occur uniformly at random. Hence the same is true conditional upon $\mathbf{w}(t+1)$ alone.

This means that the process $\langle \mathbf{w}(t) \rangle$, $t \geq 0$, is a Markov chain.

6.2.1 The transition probabilities

A little thought reveals that the transition probabilities of $\langle \mathbf{w}(t) \rangle$ can be determined from the numbers of graphs with given degree sequence. The present section examines this rather intriguing problem, resulting in the differential equations (6.21). The section can be omitted if only the consequences of the differential equations are sought.

Before giving the details on how the computation works, we turn to examination of the problem of estimating the required numbers of graphs. No "nice" formula for them is known. Fortunately, an asymptotic formula suffices for the current purposes, but it takes a lot of work to get into a usable form! The most general result in this direction for graphs of low degrees is the one in the following theorem, appearing in [37]. Define $g(\mathbf{d})$ to be the number of graphs with degree sequence $\mathbf{d} = (d_1, \dots, d_n)$.

Theorem 6.2 *For $r > 0$, define $M_r = \sum_{1 \le j \le n} [d_j]_r$ (in particular, $M_1 = 2\mu = \sum_j d_j$), and $d_{\max} = \max_{1 \le j \le n} d_j$. If $M_1 \to \infty$ and $d_{\max} = o(M_1^{1/3})$ as $n \to \infty$ then*

$$g(\mathbf{d}) = \frac{(M_1 - 1)!!}{\prod_{j=1}^n d_j!} \exp\left(-\frac{M_2}{2M_1} - \frac{M_2^2}{4M_1^2} - \frac{M_2^2 M_3}{2M_1^4} + \frac{M_2^4}{4M_1^5} + \frac{M_3^2}{6M_1^3} + O\left(\frac{d_{\max}^3}{M_1} \right) \right). \tag{6.2}$$

Here $n!!$ denotes $1 \cdot 3 \cdots n$ for an odd integer n.

Previously, McKay [35] obtained a weaker version of this result, with the restriction $d_{\max} = o(M_1^{1/4})$ and with the error term $O(d_{\max}^4 / M_1)$. In this case the error term absorbs the last three terms in the exponent, giving

$$g(\mathbf{d}) = \frac{(M_1 - 1)!! e^{-\lambda/2 - \lambda^2/4 + o(d_{\max}^4/M_1)}}{\prod_{j=1}^n d_j!} \tag{6.3}$$

where $\lambda = \frac{M_2}{M_1}$, which was obtained even earlier by Bender and Canfield [9] in the case that $d_{\max}$ is bounded, and Bollobás [11] for $d_{\max}$ growing very slowly. Bollobás explicitly used the random pairing described in Section 2 and used so profitably in Sections 3.5 and 7.1, further analysis of which yielded (6.2) and (6.3), whereas Bender and Canfield analysed an equivalent model.

The cardinality of $G(\mathbf{w})$ can now be found asymptotically, at least for all the significant values of $\mathbf{w}$. To simplify notation the dependence on t is not displayed when it is clear from context. Recalling (6.1),

$$|G(\mathbf{w})| = \frac{n!}{\bar{v}! \prod_{i=0}^{k-1} v_i!} \sum_{\mathbf{d} \in A} g(\mathbf{d}), \tag{6.4}$$

where $A = A(\mathbf{w})$ is the set of all nonnegative n-tuples $\mathbf{d}$ such that for $0 \le i \le k-1$, the number of d_j which equal i is exactly v_i, $\sum_{j=1}^n d_j = 2m$, and $i < j$ whenever $d_i < d_j < k$ or $d_i < k \le d_j$. (These ordering conditions just ensure that the entries of $\mathbf{d}$ occur in non-decreasing order except that those which are at least k are not ordered amongst themselves.) Note that A will be nonempty for large n provided $2m \ge k\bar{v} + \sum_{i=0}^{k-1} i v_i$.

Before using (6.4) and (6.3) to estimate $|G(\mathbf{w})|$, we need to argue that the upper bound $d_{\max} = o(M_1^{1/4})$ causes no problems. This hinges on the fact that the number $m(0)$ of edges

at the start of the process is only $O(n)$, and consequently $\Delta(G_0)$ is a.a.s. small. One way to do this (as in [44]) is to begin by using an idea in [17] to relax the process generating $G \in \mathcal{G}(n, m(0))$ so that multiple edges and loops are permitted. A bound on the probabilities of having large degrees in the multigraph process is quite easy to obtain since (if it is defined appropriately) the vertex degrees are distributed as Poisson variables. The probability a graph has a vertex of large degree is then bounded above by the probability the multigraph has the same property, divided by the probability that the multigraph process produces a graph. (In this sketch the details are omitted.) In this way, or others, it can be shown that when $m(0) = O(n)$, for fixed $b \in (0, \frac{1}{3})$,

$$\mathbf{P}(\Delta(G) \geq n^b) = O(e^{-n^b}) \tag{6.5}$$

for $G \in \mathcal{G}(n, m(0))$. Taking $b = \frac{1}{4}$ say, this shows that with very high probability the initial vector $\mathbf{w}(0)$ is suitable for application of the asymptotic formulae, and even satisfies the stronger upper bound on $d_{\max}$ required in (6.3). However, it may be that the process starts with such a $\mathbf{w}(0)$, but later (undesirably) reaches a state where the probability of high degree vertices in a random graph in $G(\mathbf{w})$ is quite large. The solution to this difficulty is to artificially stop the process if any $\mathbf{w}(t)$ is reached such that

$$\mathbf{P}(\Delta(G) \geq n^b) \leq \exp(-n^{b/2}) \quad \text{for a random } G \in G(\mathbf{w}(t)) \tag{6.6}$$

is not true.

To apply Theorem 6.1 later, denote by $\hat{D}$ the set of all $(t, \mathbf{w}(t))$ for which (6.6) is true. Regardless of D (which we have not yet defined), note that $\mathbf{P}(T_{\hat{D}} < T_D)$ is bounded above by the probability that the process ever reaches one of the states outside $\hat{D}$. Since

$$\mathbf{P}(\Delta(G_0) \geq n^b) = O(\exp(-n^b))$$

and

$$\mathbf{P}(\Delta(G_t) \geq n^b \mid (t, \mathbf{w}(t)) \notin \hat{D}) > \exp(-n^{b/2}),$$

it follows that

$$\mathbf{P}((t, \mathbf{w}(t)) \notin \hat{D}) = O(\exp(-n^{b/2})),$$

and so the probability that $\mathbf{w}$ ever, for any t, leaves $\hat{D}$, is $O(m \exp(-n^{b/2}))$. Thus

$$\mathbf{P}(T_{\hat{D}} < T_D) = O(m(0) \exp(-n^{b/2})). \tag{6.7}$$

To estimate the summation in (6.4), note that the vertex degrees below k are fixed and so, provided $m > \epsilon_1 n$ for some $\epsilon_1 > 0$, (6.3) becomes

$$\sum_{\mathbf{d} \in A} g(\mathbf{d}) = \frac{(2m-1)!!(1 + o(d_{\max}^4/M_1))}{\prod_{i=1}^{k-1} i!^{v_i}} F(v_0, \ldots, v_{k-1}, \bar{s}) \tag{6.8}$$

where

$$\begin{aligned} F(v_0,\dots,v_{k-1},\bar{s}) &= \sum_{d_1+\cdots+d_{\bar{v}}=\bar{s}} \frac{e^{-\lambda/2-\lambda^2/4}}{\prod_{j=1}^{\bar{v}} d_j!}, \\ \lambda=\lambda(v_0,\dots,v_{k-1},d_1,\dots,d_{\bar{v}}) &= \frac{\sum_{i=1}^{k-1} i(i-1)v_i+\sum_{j=1}^{\bar{v}} d_j(d_j-1)}{\sum_{i=1}^{k-1} iv_i+\sum_{j=1}^{\bar{v}} d_j}. \end{aligned}$$

(We will stop the process when $m(t)$ falls below $\epsilon_1 n$ for some small $\epsilon_1 > 0$.) So estimation of F is all that is required. Apart from the exponential factor, which can be shown to be sharply concentrated for the relevant sequences $d_1,\dots,d_{\bar{v}}$, F can be related to the probability that the sum of a set of independent random variables is equal to $\bar{s}$, where each variable is identically distributed as Poisson conditioned on being at least k. This was done in [44], and the probability was found asymptotically by choosing the expectations of the variables so as to maximise that probability and using a local limit theorem for the sum of i.i.d. (independent identically distributed) random variables (see [19, Chapter XVI]). The result is (6.13) below, which requires some definitions first.

Let $Z(z)$ be a random variable with distribution $\mathbf{Po}(z)$, where the expectation z will be specified shortly, and note that

$$\mathbf{P}(Z(z)\geq k)=e^{-z}e_k(z)$$

where

$$e_k(z)=\sum_{i\geq k}\frac{z^i}{i!}. \tag{6.9}$$

Also let $X(z)$ be a random variable with the distribution of $Z(z)$ conditioned upon $Z(z)\geq k$, so that

$$\mathbf{P}(X(z)=i)=\frac{\mathbf{P}(Z(z)=i)}{\mathbf{P}(Z(z)\geq k)},\qquad i\geq k. \tag{6.10}$$

Now choose the value of $z=z(\bar{v},\bar{s},k)$ so that

$$\bar{v}\mathbf{E}X(z)=\bar{s}, \tag{6.11}$$

(with $\bar{s}$ defined in (6.1)) which is determined uniquely since $\mathbf{E}X(z)$ is strictly increasing with z. It turns out that if $d_1,\dots,d_{\bar{v}}$ are i.i.d. copies of $X(z)$ then λ is sharply concentrated near

$$\bar{\lambda}=\frac{\sum_{i=1}^{k-1} i(i-1)v_i+\bar{v}\mathbf{E}(X(z)(X(z)-1))}{\sum_{i=1}^{k-1} iv_i+\bar{v}\mathbf{E}X(z)}. \tag{6.12}$$

Interpreting this appropriately, the formula which results is

$$F(v_0,\dots,v_{k-1},\bar{s})=(1+O(n^{-1/2}\log n))\frac{(e_k(z))^{\bar{v}}\exp(-\bar{\lambda}/2-\bar{\lambda}^2/4)}{z^s\sqrt{\bar{v}2\pi\mathbf{Var}X(z)}} \tag{6.13}$$

provided

$$m > \epsilon_1 n, \quad \bar{v} \geq \epsilon_1 n \quad \text{and} \quad \bar{s} \geq (k + \epsilon_1)\bar{v}\,. \tag{6.14}$$

(Again, we can stop the process if these conditions fail to be met, by defining the domain D appropriately.) From (6.4), (6.8) and (6.13), if (6.6) and (6.14) hold then

$$|G(\mathbf{w})| = \left(1 + O\left(\frac{1}{n^{1-4b}} + \frac{\log n}{\sqrt{n}}\right)\right) \frac{n!(2m-1)!!(e_k(z))^{\bar{v}} e^{-\bar{\lambda}/2 - \bar{\lambda}^2/4}}{z^{\bar{s}}\sqrt{\bar{v}2\pi \mathbf{Var} X(z)}\bar{v}! \prod_{i=0}^{k-1} v_i! i!^{v_i}}\,. \tag{6.15}$$

(Actually, in [44], the error exponent $1 - 4b$ is improved to $1 - 3b$ by arguing, similar to the derivation of (6.6), that the likely degree sequences encountered have M_2, M_3 and M_4 all $O(n)$, in which case the error in (6.2) truly absorbs these terms. This gives a slightly sharper result in the end. A possibly simpler argument for this improvement is that if we retain these terms and use (6.2) in place of (6.3), the fact that the terms vary slowly — see the treatment of $\bar{\lambda}$ below — shows that they have no asymptotic effect.)

At last, we can return to consideration of the transition probabilities in the Markov chain $\langle \mathbf{w}(t) \rangle$, $t \geq 0$. (In this process, unlike some of the earlier ones such as in Section 3.1, there does not seem to be any way to calculate the expected changes in the variables without essentially computing all the transition probabilities of the Markov chain. There is no obvious way, for instance, to separate the expected change into separate expectations for each deleted edge, due to the presence of factors depending on the degree of the chosen vertex x_t.) Denote $\mathbf{w}(t)$ by $\mathbf{w} = (v_0, v_1, \ldots, v_{k-1}, m)$ and consider the conditional probability (given $\mathbf{w}(t)$) that $\mathbf{w}(t+1)$ is equal to some vector $\mathbf{w}' = (v_0', v_1', \ldots, v_{k-1}', m')$. Suppose that the vertex x_t, whose incident edges are deleted from G_t, has degree j, and is adjacent in G_t to exactly u_i vertices of degree i for $1 \leq i \leq k$ and to u_{k+1} vertices of degree at least $k+1$. Also set $u_0 = -1$; then the following equations are true:

$$\begin{aligned} v_i' - v_i &= -u_i + u_{i+1} - \delta_{ij}, \qquad 0 \leq i \leq k-1, \\ \bar{v}' - \bar{v} &= -u_k, \\ u_{k+1} &= j - \sum_{i=1}^{k} u_i\,, \end{aligned} \tag{6.16}$$

where $\bar{v}'$ is for $\mathbf{w}'$ as $\bar{v}$ is for $\mathbf{w}$.

We approach the transition probability computation by counting the possibilities for $G \in G(\mathbf{w}(t))$ with the vertex x_t distinguished, and of degree j. The total number of these graphs is just $v_j|G(\mathbf{w})|$. By the uniformity of the distribution of G, the probability that the neighbours of the randomly chosen x_t (conditional on $d(x_t) = j$) are in accordance with the u_i can be computed by counting such graphs. That is, we count the graphs $G' \in G(\mathbf{w}')$ after deletion of the edges incident with x_t, with the vertex x_t, and its neighbours, distinguished, since from such G' we can deduce G, and vice versa. Note that x_t is now of degree 0 (and specially distinguished), and of the other distinguished vertices there are u_i of degree $i-1$ for $1 \leq i \leq k$,

and u_{k+1} of degree at least k. Such vertices can be distinguished in precisely

$$f(\mathbf{w},\mathbf{w}') = v_0' \prod_{i=0}^{k} \binom{v_i' - \delta_{i0}}{u_{i+1}}$$

ways, with v_k' denoting $\bar{v}'$. Using (6.16), this can be written as

$$\begin{aligned} f(\mathbf{w},\mathbf{w}') &= \prod_{i=0}^{k} \frac{v_i'!}{(v_i' - u_{i+1} - \delta_{i0})! u_{i+1}!} \\ &= \frac{v_k'!}{(v_k - u_k - u_{k+1})! u_{k+1}!} \prod_{i=0}^{k-1} \frac{v_i'!}{(v_i - u_i - \delta_{i0} - \delta_{ij})! u_{i+1}!} \\ &= \frac{\bar{v}! v_j}{(v_k - u_k - u_{k+1})! u_k! u_{k+1}!} \prod_{i=0}^{k} \frac{v_i'!}{v_i!} \prod_{i=1}^{k-1} \binom{v_i - \delta_{ij}}{u_i}, \end{aligned}$$

and we have from the above argument that

$$\mathbf{P}(\mathbf{w}(t+1) = \mathbf{w}' \mid \mathbf{w}(t) = \mathbf{w} \wedge d(x_t) = j) = \frac{f(\mathbf{w},\mathbf{w}')|G(\mathbf{w}')|}{v_j |G(\mathbf{w})|}.$$

Since x_t is chosen uniformly at random from the v non-isolated light vertices,

$$\mathbf{P}(d(x_t) = j \mid \mathbf{w}(t) = \mathbf{w}) = \frac{v_j}{v}.$$

Thus

$$\mathbf{P}(\mathbf{w}(t+1) = \mathbf{w}' \mid \mathbf{w}(t) = \mathbf{w}) = \frac{f(\mathbf{w},\mathbf{w}')|G(\mathbf{w}')|}{v|G(\mathbf{w})|}\,. \tag{6.17}$$

Next, we estimate this ratio using (6.15), taking various parts of the ratio separately, and writing z', $\bar{s}'$ and $\bar{\lambda}'$ for $\mathbf{w}'$ as z, $\bar{s}$ and $\bar{\lambda}$ are for $\mathbf{w}$ (and recalling that $\bar{v} = v_k$):

$$\frac{(2m'-1)!!}{(2m-1)!!} = (1 + O(n^{-1}))\frac{1}{(2m)^j},$$

$$\frac{f(\mathbf{w},\mathbf{w}')\bar{v}! \prod_{i=0}^{k-1} v_i! i!^{v_i}}{v(\bar{v}')! \prod_{i=0}^{k-1} v_i'! i!^{v_i'}} = \binom{v_k - u_k}{u_{k+1}} \frac{v_j j!}{v(k-1)!^{u_k}} \prod_{i=1}^{k-1} i^{u_i} \prod_{i=1}^{k} \binom{v_i - \delta_{ij}}{u_i}.$$

Moreover, from (6.14)

$$\frac{\sqrt{\bar{v} 2\pi \mathbf{Var} X(z)}}{\sqrt{\bar{v}' 2\pi \mathbf{Var} X(z')}} = 1 + O(n^{-1})$$

and

$$\frac{e^{-\bar{\lambda}'/2 - (\bar{\lambda}')^2/4}}{e^{-\bar{\lambda}/2 - \bar{\lambda}^2/4}} = 1 + O(n^{-1}).$$

For the last remaining factors in (6.17), a fairly simple but technical argument leading to [44, (4.30)] shows that regardless of the degree of x_t, z and z' are almost equal; to be precise,

$$\frac{z^{\bar{s}}(e_k(z'))^{\bar{v}'}}{(z')^{\bar{s}'}(e_k(z))^{\bar{v}}} = (1+O(n^{-1}))z^{\bar{s}-\bar{s}'}(e_k(z))^{\bar{v}'-\bar{v}} = (1+O(n^{-1}))\left(\frac{z^k}{e_k(z)}\right)^{u_k} z^{u_{k+1}}$$

since $\bar{s}$ is the total degree of heavy vertices and so $\bar{s}-\bar{s}' = ku_k + u_{k+1}$. From (6.14), the binomials and m factors can be estimated by

$$\frac{1}{(2m)^{u_i}}\binom{v_r - C}{u_i} = \left(\frac{v_r}{2m}\right)^{u_i}\frac{1}{u_i!} + O\left(\frac{v_r^{u_i-1}}{m^{u_i}}\right) = \left(\frac{v_r}{2m}\right)^{u_i}\frac{1}{u_i!} + O\left(\frac{1}{n}\right)$$

for any $C > 0$, with $r = i \le k$ or $r+1 = k+1 = i$.

Recalling that the total number of edges deleted is

$$d(x_t) = m - m' = \sum_{i=1}^{k+1} u_i, \tag{6.18}$$

and not forgetting the error term in (6.15), (6.17) now gives the transition probability asymptotically in the very nice form

$$\mathbf{P}(\mathbf{w}(t+1) = \mathbf{w}' \mid \mathbf{w}(t) = \mathbf{w}) = \frac{v_j}{v}\binom{j}{u_1,\ldots,u_{k+1}}p_1^{u_1}p_2^{u_2}\cdots p_{k+1}^{u_{k+1}} + O\left(\frac{1}{n^{1-4b}} + \frac{\log n}{\sqrt{n}}\right)$$

where

$$p_i = \frac{iv_i}{2m} \quad (i = 1,\ldots,k-1), \quad p_k = \frac{\bar{v}z^k}{2m(k-1)!e_k(z)}, \quad p_{k+1} = \frac{\bar{v}z}{2m}$$

and $\binom{j}{u_1,\ldots,u_{k+1}} = \frac{j!}{u_1!\cdots u_{k+1}!}$ is the multinomial coefficient. Thus, conditional upon $\mathbf{w}$ and $d(x_t) = j$, the vector $(u_1,\ldots,u_{k+1})$ has asymptotically the multinomial distribution resulting from j independent trials with probabilities $p_1,\ldots,p_{k+1}$. In the multinomial distribution, u_i has expected value jp_i for each i, $1 \le i \le k+1$. Hence

$$\begin{aligned}\mathbf{E}(u_i \mid \mathbf{w}(t) = \mathbf{w}) &= \sum_{j=1}^{k-1}\frac{v_j}{v}jp_i + O\left(\frac{1}{n^{1-4b}} + \frac{\log n}{\sqrt{n}}\right)\\ &= \frac{sp_i}{v} + O\left(\frac{1}{n^{1-4b}} + \frac{\log n}{\sqrt{n}}\right).\end{aligned} \tag{6.19}$$

This determines the expected value of $v_i(t+1)-v_i(t)$ via (6.16) and of $m(t+1)-m(t)$ via (6.18), and hence of $\mathbf{E}(\mathbf{w}(t+1)-\mathbf{w}(t) \mid \mathbf{w}(t))$, to the same accuracy. This gives us enough information to present the appropriate differential equations: writing

$$\nu_i(x) = \frac{v_i(xn)}{n}, \quad \nu(x) = \frac{v(xn)}{n}, \quad \mu(x) = \frac{m(xn)}{n}, \quad \bar{\nu}(x) = \frac{\bar{v}(xn)}{n}, \quad \sigma = \frac{s}{n} \tag{6.20}$$

the equations are (recalling $u_0 = -1$ and noting that $\mathbf{E}(\delta_{ij} \mid \mathbf{w}) = v_i/v$ and that the p_i must sum to 1)

$$\begin{aligned}
\frac{d\nu_i}{dx} &= \delta_{i0} + \frac{(i+1)\nu_{i+1}\sigma}{2\mu\nu} - \frac{i\nu_i\sigma}{2\mu\nu} - \frac{\nu_i}{\nu} \qquad (0 \le i < k-1) \\
\frac{d\nu_i}{dx} &= \frac{z^k\bar{\nu}\sigma}{2\mu\nu(k-1)!e_k(z)} - \frac{i\nu_i\sigma}{2\mu\nu} - \frac{\nu_i}{\nu} \qquad (i = k-1) \\
\frac{d\mu}{dx} &= -\frac{\sigma}{\nu}
\end{aligned} \tag{6.21}$$

where z can now be defined (analogous to (6.11)) by

$$\bar{\nu}\mathbf{E}X(z) = 2\mu - \sigma \tag{6.22}$$

(recall that $X(z)$ is distributed as $\mathbf{Po}(z)$ conditioned on being at least k, as in (6.10)), and e_k is still as in (6.9). The other variables (σ, ν and z) appearing here are simple functions of the ν_i and μ, determined by (6.1) and (6.11). Note that the initial conditions relevant for the process are determined by the random graph $G_0 \in \mathcal{G}(n, m(0))$.

6.2.2 Dealing with difficult differential equations

Although we have not been able to solve the differential equations (6.21) for all the variables, it is possible to wring information on the size of the k-core from them.

It is shown in [44, pp. 135] that the system (6.21) implies

$$\frac{z^2}{\mu} = C_1, \qquad \frac{e^z\bar{\nu}}{e_k(z)} = C_2 \tag{6.23}$$

for some constants C_1 and C_2 which will depend on the initial conditions. We will be considering the implications of this when $m = cn/2$. (Here c can be thought of as an approximately constant function of n, or merely any function of n, bounded above and bounded away from 0, such that our conclusions will depend on what ranges c lies in.)

The sort of large deviation argument leading to (6.5) can also be used to show that for $\mathbf{w} = \mathbf{w}(0)$ (which is determined by a random $G \in \mathcal{G}(n, m(0))$) the probability of the event

$$|\bar{\nu}(0) - e^{-c}e_k(c)| + |2\mu(0) - \sigma(0) - c\pi_k(c)| < n^{-\epsilon_1} \tag{6.24}$$

is $1 - O(e^{-n^\epsilon})$, where

$$\pi_k(z) = \mathbf{P}(Z(z) \ge k-1) \tag{6.25}$$

for $Z(z)$ as before distributed as $\mathbf{Po}(z)$, where ϵ and ϵ_1 are sufficiently small positive constants. Specific values can be computed; this was done in the argument in [44], but the present argument is different and we are not paying attention to the actual values.) Note also that the probability generating function of $X = X(z)$ in (6.10) is $g(y) = \mathbf{E}y^X$ and so

$$\mathbf{E}X(z) = g'(1) = \frac{z\pi_k(z)e^z}{e_k(z)}\,. \tag{6.26}$$

Restrict to G_0 satisfying (6.24) and apply Theorem 6.1 with the variables $Y_1, \ldots, Y_a$ being the $v_i(t)$ and $m(t)$. Since the degrees of at most k vertices change in any one step of the process, and m changes by at most k, we can take $\beta = 2k$ and $\gamma = 0$. D is the domain suggested by (6.14) together with $v > \epsilon_2 n$ (where ϵ_2 will be chosen much smaller than ϵ_1); that is, considering (6.20), D is the set of $(\nu_0, \ldots, \nu_{k-1}, \mu, x)$ such that

$$\mu > \epsilon_1, \quad \bar{\nu} > \epsilon_1, \quad 2\mu - \sigma > (k + \epsilon_1)\bar{\nu}, \quad \nu > \epsilon_2. \tag{6.27}$$

The set $\hat{D}$ was defined just after (6.6) to eliminate the extreme degree sequences for which we could not compute the trends. Equation (6.19) now establishes the trend hypothesis where the functions f_i are those on the right hand side of equations (6.21). Since we are not aiming for best possible results here, take $\lambda = n^{-\epsilon}$. Note that the lower bound $\nu > \epsilon_1$ is crucially required for the Lipschitz hypothesis. From this theorem, and noting (6.7) which guarantees that the extreme degree sequences outside $\hat{D}$ are almost never encountered, we conclude that with probability $1 - O(e^{-n^\epsilon})$,

$$\frac{v_i(t)}{n} = \nu_i(t/n) + O(n^{-\epsilon}) \quad (0 \le i < k), \qquad \frac{m(t)}{n} = \mu(t/n) + O(n^{-\epsilon}) \tag{6.28}$$

as long as the inequalities in (6.27) are satisfied with some slack (that is, with ϵ_1 replaced by $C\epsilon_1$ for some sufficiently large $C > 0$).

To state the implications for the k-core, we define

$$\gamma_k = \inf_{\lambda > 0} \left\{ \frac{\lambda}{\pi_k(\lambda)} \right\}$$

where π_k is defined in (6.25).

Theorem 6.3 *Choose ϵ_0 arbitrarily small. If $c < \gamma_k - \delta$ for fixed $\delta > 0$ and $m = cn/2$, then the probability that $G \in \mathcal{G}(n, m)$ has a k-core with at least $\epsilon_0 n$ vertices is $O(e^{-n^\epsilon})$ for some $\epsilon > 0$.*

Note. Łuczak [33] showed that a.a.s. in the standard random graph process, every k-core which is nonempty has at least $0.0002n$ vertices. Thus, if we choose ϵ_0 below this small constant, the theorem shows the threshold of appearance of a k-core is greater than $\gamma_k - \delta$ for all $\delta > 0$.

Proof. First, a word about the strategy of choosing ϵ_1 and ϵ_2. It is helpful to consider ϵ_1 to be a very small fraction of ϵ_0 (at least, we take it less than ϵ_0 and reserve the right to choose ϵ_1 smaller later) and also of δ. Then ϵ_2 is much smaller again. Then if the process (or more correctly, the vector of scaled variables of the process) exits D, the remaining number of edges is much smaller than both δn and $\epsilon_0 n$, and if it exits at $\nu = \epsilon_2$, the number of heavy vertices remaining is much larger than the number of light vertices.

We can clearly assume that the concentration of the initial variables at the start of the process given in (6.24) holds, as well as the approximations to the variables throughout the process given in (6.28) (by choosing ϵ in the present theorem sufficiently smaller than those

occurring earlier). So consider the solution of the differential equations (6.21) with some initial conditions such that $\bar{\nu}(0)$ and $\sigma(0)$ satisfy (6.24). Then using the definition of z (6.22), and (6.26) and (6.24), we find

$$\mathbf{E}X(z(0)) = \mathbf{E}X(c) + O(n^{-\epsilon_1})$$

and hence

$$z(0) = c + o(1) \tag{6.29}$$

(since $\mathbf{E}X(z)$ is clearly increasing with z).

We first show that for sufficiently small and well chosen ϵ_1 and ϵ_2 the solution cannot exit D at the boundary where $2\mu - \sigma = (k+\epsilon_1)\bar{\nu}$ or where $\nu = \epsilon_2$, and then deduce from this that the k-core is a.a.s. quite small.

Suppose to the contrary, firstly, that the solution exits D with $2\mu(x) - \sigma(x) \le (k+\epsilon_1)\bar{\nu}(x)$. (In fact, equality must hold, by continuity.) With $X(z)$ as in (6.10), we find using (6.22)

$$k + \epsilon_1 \ge \frac{2\mu(x) - \sigma(x)}{\bar{\nu}(x)} = \mathbf{E}X(z(x)) = k + \mathbf{E}(X(z(x)) - k) > k + Cz(x)$$

for some positive constant C independent of ϵ_1. (The last is easily verified in the case of z bounded, which is all we need here.) Thus $z(x) < \epsilon_1/C$. But from the second equation in (6.23),

$$\bar{\nu}(x) = \frac{e_k(z(x))e^{z(0)}\bar{\nu}(0)}{e^{z(x)}e_k(z(0))}.$$

From this, (6.29) and (6.24) give

$$\bar{\nu}(x) = \frac{e_k(z(x))}{e^{z(x)}} = \frac{e^{-z(x)}z(x)^k}{k!}(1 + O(z(x))) = O(\epsilon_1^k) < \epsilon_1$$

for ϵ_1 sufficiently small, as $z(x) < \epsilon_1/C$. This means the solution has already exited D, a contradiction.

Next, suppose the solution exits D with $\nu(x) = \epsilon_2$. Then $\mu > \epsilon_1$ and so

$$2\mu - \sigma > 2\mu - k\epsilon_2 = 2\mu(1 + O(\epsilon_1/\epsilon_2)). \tag{6.30}$$

Combining (6.22), (6.26) and (6.23) shows

$$\frac{2\mu - \sigma}{z\pi_k(z)} = \frac{\bar{\nu}e^z}{e_k(z)} = C_2.$$

Hence by the definition of γ_k,

$$\begin{aligned}
\gamma_k &\le \frac{z(x)}{\pi_k(z(x))} \\
&= \frac{z^2(x)}{2\mu(x)} \times \frac{2\mu(x) - \sigma(x)}{z(x)\pi_k(z(x))}(1 + O(\epsilon_1/\epsilon_2)) \qquad \text{by (6.30)} \\
&= \frac{z^2(0)}{2\mu(0)} \times \frac{2\mu(0) - \sigma(0)}{z(0)\pi_k(z(0))}(1 + O(\epsilon_1/\epsilon_2)) \qquad \text{by (6.23)} \\
&= c(1 + O(\epsilon_1/\epsilon_2))
\end{aligned}$$

by (6.24) and (6.29), which contradicts $c < \gamma_k - \delta$ for suitably chosen $\epsilon_1 << \epsilon_2$. So the solution cannot exit D on this boundary.

Thus, it exits with $\mu(x) = \epsilon_1$ or $\bar{\nu}(x) = \epsilon_1$, and up to this point (i.e., for all smaller x) has $\nu > \epsilon_2$. By (6.28), the random process has $v(t) > n(\epsilon_2 - o(1))$ light vertices for all $t \leq xn$, and so does not terminate before this time. But at this time (again by (6.28)) there are at most $n(\epsilon_1 + o(1))$ heavy vertices left, so the k-core is no larger than this. Making sure that ϵ_1 is smaller than ϵ_0 now completes the proof. ∎

Whenever c is a little larger than γ_k, $c = \lambda/(\pi_k(\lambda))$ has two roots for λ. Denote the larger one by $\lambda_k(c)$. The following theorem follows from the results in [44]. As mentioned before, the approach to the proof there is a little different from the one here, but still the present approach is sufficient for the following.

Theorem 6.4 *Choose ϵ_0 arbitrarily small. If $c > \gamma_k + \delta$ for some fixed $\delta > 0$ and $m = cn/2$. Then with probability $1 - O(e^{-n^\epsilon})$ for some $\epsilon > 0$, $G \in \mathcal{G}(n, m)$ has a k-core with*

$$ne^{\lambda_k(c)}e_k(\lambda_k(c)) + o(n)$$

vertices.

Proof sketch. It can be shown that in this case the differential equation solution must have $\nu(x) = 0$ for some x when $\bar{\nu}$ is approximately $e^{\lambda_k(c)}e_k(\lambda_k(c))$. The random deletion process almost surely follows the differential equation at least approximately until it leaves a domain D like the one in Theorem 6.3. From this point, the Lipschitz condition fails, and in fact when the number v of light vertices drops to a very small value, the expected change in the number of light vertices per deletion step will start to vary significantly due to lack of concentration of the numbers v_i. However, it can still be shown that the expected change in $s(t)$ in a single step is negative. Then Corollary 4.1 is sufficient to show that the process finishes in at most Cn steps for some C which can be made arbitrarily small. (Actually, the relative rate of change of $s(t)$ with respect to $m(t)$ will continue to be roughly constant until v hits 0, but this is not strictly needed for the proof.) ∎

Note. In [44], the threshold was determined more precisely (with error n^{-C} rather than just δ), and a value for the exponent ϵ in the probability was determined. In addition, it was shown that a.a.s. in the standard random graph process, the very edge, say the m'th, which creates a nonempty k-core produces one of "full" size. To be precise, a.a.s. for all m such that G_m has a nonempty k-core, its size is $ne^{\lambda_k(2m/n)}e_k(\lambda_k(2m/n)) + o(n^{-\zeta})$ for $\zeta > 0$ given in [44]. The higher accuracy was obtained by using the method of proof of Theorem 5.2, applied to the general integrals of the differential equations given by (6.23).

7 Differences with a tail

Sometimes the upper bound on the differences in variables is too large to be useful, but the differences are with high probability much smaller than the upper bound, as discussed in Section 4.4. In this case Theorem 5.1 will be used with non-zero γ.

7.1 The degree-greedy algorithm for independent sets

Here we consider the second algorithm for finding independent sets in random regular graphs (see Section 2). The results given here were obtained in [61] (where it was called the *neighbourly* algorithm), and also in the case $d = 3$ in [21] (called MINGREEDY). The analysis uses the pairing process, as for the simple greedy algorithm in Section 3.5.

The degree-greedy algorithm is as follows. Given a random d-regular graph G,

1. Set $I := \emptyset$;
2. If $V(G) = \emptyset$, stop. Otherwise, select a vertex v u.a.r. from the vertices of G of minimum degree;
3. Set $I := I \cup \{v\}$ and delete v and all its neighbours from G;
4. Return to step 2.

An equivalent way to express this algorithm is as follows. A vertex of degree strictly less than d is called *unsaturated*. Choose the vertices in an independent set I consecutively, each vertex chosen randomly from those which, in the graph of edges so far probed, are of maximum degree subject to being unsaturated. After adding a vertex v to I, probe all its incident edges as well as all those of its neighbours. Thus, the edges probed in this algorithm are the ones deleted in the one above.

As with the simple greedy algorithm, this one can be incorporated as part of a pairing process generating a random regular graph, by exposing pairs when probed.

The pairing process incorporating the degree-greedy algorithm.

> First add a random vertex to an empty set I. When v is added, random mates are selected for all remaining unpaired points in (the bucket corresponding to) v, and then for all other points in the vertices adjacent to v. Then the next vertex v is selected for I, which is done randomly from those currently of distance at least 2 from I, and of maximum degree subject to this constraint. The process continues until all vertices have distance at most 1 from vertices in I.

In the following discussion, this pairing process is regarded as the algorithm. Note that the pairing is completed (that is, all edges are exposed) at or before the step in which the last vertex is added to I.

Before setting up the differential equations, the typical behaviour of the algorithm must be considered. For d in general, this behaviour has some interesting features (which have not yet been fully determined, as explained below). The current degree of the vertex v added at each step clearly affects the size of the independent set ultimately produced, and the typical value of its degree can change throughout the process. At first, the number of unsaturated vertices of degree greater than 1 remains bounded, and the vertices added to I are almost all of degree 1. But when $|I| > \epsilon n$, vertices of degree 2 are created with positive probability when adding the edges from v or its neighbours. Still, any such vertices created are chosen in

the next few steps of the algorithm until none remain. This is called phase 1 of the algorithm. But at some time the newly generated vertices of degree 2 begin to regenerate themselves as fast as they are consumed. Shortly after this, the degree 2 vertices are created faster than they are consumed, and we say that the algorithm is in phase 2. At this time, all but a bounded number of unsaturated vertices have degree at most 2. The transition between phases is rather blurred, and before getting technical we do not have a precise definition of the phases; roughly speaking, if most of the vertices being added are of degree k or $k+1$ then all but a bounded number of unsaturated vertices have degree at most k, and the algorithm is in phase k. This keeps going until phase $d-2$, when the vertices being chosen have degree $d-2$, and any vertices of degree $d-1$ created are immediately chosen next.

It would be possible that a phase $d-1$ occurs, in which vertices of degree $d-1$ are created faster than they are destroyed. It would similarly be possible that at some time in phase k the vertices of degree k are, perhaps through some unlikely events, or perhaps even a.a.s., used up prematurely, and the process reverts to an earlier phase. Indeed, these scenarios are clearly possible in any particular run of the algorithm, and can even be highly likely *if* the process starts not with an empty pairing but with a carefully selected initial pairing. However, the solution of the differential equations given below shows that, at least for $d \leq 100$, the process a.a.s. passes through phases $1, 2, \ldots, d-2$ with no reversions and without reaching a phase $d-1$. I conjecture that this is the case for all $d \geq 3$. These statements will be made precise by defining phases in terms of the differential equations governing the process.

We are at liberty to define the basic variable t of the differential equations arbitrarily. For this process, a convenient choice of t is *not* the number of pairs added to the growing graph, nor is it (as in the simple greedy algorithm) the number of vertices added to I. Instead, the definition of a unit time interval depends on the phase k of the algorithm, and both are defined inductively. We begin after the first vertex has been put into I and all edges incident with its neighbours have been determined. Here $t = 1$ and $k = 1$.

So now assume that the algorithm has reached some general point in phase k. We shall assume that there are many unsaturated vertices of degree at most k, and that, for a period of time at least, each individual vertex added to I a.a.s. has degree either k or $k+1$. (The precise statement proved in [61] is that this is true at least after $o(n)$ vertices have been added to I in phase k. Except near the end of phase k, the unsaturated vertices of degree at least $k+2$ are created only rarely, because only a bounded number of unsaturated vertices have degree $k+1$ and so it is unlikely that they are selected when the pairs are exposed.) Define a *clutch* to be a set of vertices added consecutively to I beginning with one of degree k and ending just before the next vertex of degree k. Increment time by 1 for each clutch of vertices, and so define G_t to be the graph evolved after t clutches. Now the expected changes to G_t in one unit of time can be calculated. Define $Y_i(t)$ as usual, and define $S(t)$ to be the number of unpaired points remaining at time t. This is determined by the deficiencies of the unsaturated vertices, and so

$$S(t) = \sum_{i=0}^{d-1} (d-i) Y_i(t).$$

Condition on G_t and consider the process while the next clutch of vertices is added to I.

Computation here is only approximate, ignoring the changing values of the variables during this segment of the algorithm, and making various other assumptions which can also be justified. Also we assume that $Y_i(t) = 0$ for $i > k$. When a point is selected at random to be paired with another, the probability it lies in a vertex of degree j is $(d-j)Y_j(t)/S(t)$. Thus, when a vertex of degree k is added to I, the expected number of vertices of degree j hit during exposure of its $d-k$ remaining edges is

$$\frac{(d-k)(d-j)Y_j(t)}{S(t)},$$

and the expected number of vertices of degree i hit in turn when the remaining $d-j-1$ edges from *each* of these vertices are exposed is

$$\frac{(d-k)(d-j)Y_j(t)}{S(t)} \times \frac{(d-j-1)(d-i)Y_i(t)}{S(t)}.$$

The former change from degree j to degree d, and so contribute only to a reduction in Y_j, whilst the latter change from degree i to degree $i+1$, assuming (at least a.a.s.) that the vertices hit during taking the clutch are all distinct. So they contribute to both a reduction in Y_i and an increase in Y_{i+1}. This is summed over all j up to $d-2$, to show that the expected increase in Y_i when *one vertex* v of degree k (not a clutch, yet) is added to I is

$$-\delta_{ik} + (d-k)\mu_i \tag{7.1}$$

where $\mu_i = \mu_i(t)$ is given by

$$\mu_i = -\frac{(d-i)Y_i(t)}{S(t)} + \sum_{j=0}^{d-2} \frac{(d-j-1)(d-j)((d+1-i)Y_{i-1}(t) - (d-i)Y_i(t))Y_j(t)}{S(t)^2}\ .$$

Here the term $-\delta_{ik}$ stands for the loss of the vertex v itself in the case $i = k$. On the other hand, if v has degree $k+1$, the result is

$$(d-k-1)\mu_i \tag{7.2}$$

in place of (7.1), for all $i \le k$. (This will not be needed for $i = k+1$.)

The expected number of vertices of degree $k+1$ in a clutch can be found as follows. Beginning with $Y_{k+1}(t) = 0$, the expected number created when the first vertex of the clutch and all its neighbours' edges are exposed is, from (7.1), $(d-k)\mu_{k+1}$. Consider this as the first generation of a birth-death process in which the individuals are the vertices of degree $k+1$, each giving birth to a number of children (essentially independently of the others) with expected number $(d-k-1)\mu_{k+1}$ by (7.1). Then the expected number in the j^{th} generation is

$$(d-k)\mu_{k+1}((d-k-1)\mu_{k+1})^{j-1},$$

and the expected total number of births in the whole birth-death process is

$$\frac{(d-k)\mu_{k+1}}{1-(d-k-1)\mu_{k+1}}. \tag{7.3}$$

(During phase k the denominator will be positive; only near the transition to phase $k+1$ will it approach 0.) The births represent vertices of degree $k+1$ in the clutch. For $i \le k$, the expected increase in Y_i when the clutch is added is obtained by multiplying this by (7.2), and adding (7.1), for the first vertex of the clutch. This simplifies to

$$\mathbf{E}(Y_i(t+1) - Y_i(t) \mid G_t) = -\delta_{ik} + \frac{(d-k)\mu_i}{1-(d-k-1)\mu_{k+1}}\,.$$

As in the other examples, this forms the basis for the differential equations. Write $Y_i(t) = nz_i(t/n)$, $\mu_i(t) = n\tau_i(t/n)$ and $S(t) = n\xi(t/n)$. The system of differential equations suggested is

$$z_i' = -\delta_{ik} + \frac{(d-k)\tau_i}{1-(d-k-1)\tau_{k+1}} \quad (i \le k); \quad z_i = 0 \quad (i > k)\,. \tag{7.4}$$

Here differentiation is with respect to x, where xn represents the number of clutches, and from the definitions of μ and S

$$\tau_i = -\frac{(d-i)z_i}{\xi} + \sum_{j=0}^{d-2} \frac{(d-j-1)(d-j)((d+1-i)z_{i-1} - (d-i)z_i)z_j}{\xi^2}$$

and

$$\xi = \sum_{i=0}^{d-1}(d-i)z_i.$$

Along the way one must compute the number $Y(t)$ of vertices in I at time t. The expected increase in Y at time t is the expected size of a clutch, which is the expected number of births given in (7.3), plus 1 for v. So writing $Y(t) = nz(t/n)$, the differential equation suggested for z is

$$z' = 1 + \frac{(d-k)\tau_{k+1}}{1-(d-k-1)\tau_{k+1}} = \frac{1+\tau_{k+1}}{1-(d-k-1)\tau_{k+1}}\,.$$

So far, differentiation has been with respect to x. However, since x does not appear other than as a parameter in any of these equations, we can compute the ratio $\frac{dz_i}{dz} = \frac{z_i'(x)}{z'(x)}$ using (7.4) and obtain

$$z_i' = \frac{\delta_{ik}((d-k-1)\tau_{k+1} - 1) + (d-k)\tau_i}{1+\tau_{k+1}} \quad (i \le k); \quad z_i = 0 \quad (i > k) \tag{7.5}$$

where differentiation is with respect to z, and all functions can be taken as functions of z, which represents the cardinality of I (scaled by $\frac{1}{n}$).

For the first phase, $k = 1$ and the initial conditions are

$$z_0(0) = 1, \quad z_i(0) = 0 \quad (i > 0).$$

The technical definition of the phases can now be given, inductively. Phase k refers to all of the process corresponding to values of z between $z^{(k-1)}$ and $z^{(k)}$. On this interval, the functions z_i are given by (7.5) with initial conditions given by the final values in the previous

phase, at $z^{(k-1)}$. The final part of the definition is that $z^{(0)} = 0$ and $z^{(k)}$ is the first solution for $z \geq z^{(k-1)}$ of $(d-k-1)\tau_{k+1} = 1$. (This is when the expected number of births in a clutch would tend to infinity.)

Theorem 5.1 can be applied to the process within each phase. This is similar to the applications in Section 5.3 except that γ is now non-zero, because there is no upper bound on the number of vertices in a clutch. However, as long as the expected number of births (7.3) in the birth-death process is bounded above, the probability of getting say n^ϵ births is $O(n^{-K})$ for any fixed K. (This comes from a standard argument, which can go like the following. If (7.3) is bounded then the expected number of new vertices of degree $k+1$ found when one such vertex and all its neighbours are exposed is at most $1-\epsilon$ for some $\epsilon > 0$. Thus, the number of such vertices in the current graph has an expected decrease at each step. Then an argument like the proof of Corollary 4.1 with $b = -\epsilon$ shows that with very high probability the number of these vertices in the current graph reaches 0 before too long.) Thus Theorem 5.1 applies with $\gamma = n^{-K}$ and $\beta = n^\epsilon$ and an appropriate domain D. By this argument, it was shown in [61] that during each phase, $Y_i(t) = nz_i(t/n) + o(n)$ a.a.s. It follows that the size of the independent set at the end of the algorithm is asymptotic to $nz^{(k)}$, where k is the last phase. The solutions of the systems of differential equations were computed for many small values of d, and as mentioned above, in all cases the system passes through phases $1, 2, \ldots, d-2$ and then finishes. (Some numerical analysis needed to be done to check that with the approximate solutions found, at the end of each phase k the number of vertices of degree $k+1$ does indeed begin to grow. That is, the birth rate, per individual, in the birth-death process generating vertices of degree $k+1$ must grow above 1. This was done by examining the derivative of the function determining the birth rate. This derivative was shown to be positive near the phase transition. In this way, there is no problem caused by the fact that the sharp concentration result only applies in a domain which is exited near the end of a phase. Alternative scenarios, where say the maximum degree unsaturated vertices reverted to degree $k-1$ or less, were ruled out by the numerical computation in these cases, but have not been ruled out theoretically in general.) Thus, in all these cases, the size of the independent set found by this algorithm is a.a.s. asymptotic to $nz^{(d-2)}$.

The almost sure lower bounds on maximum independent set size of a random d-regular graph ($d \geq 3$) given by this algorithm are closer to the best known upper bounds (McKay [36]) than to the lower bounds given by the simple greedy algorithm in Section 3.5, and seem to exceed the best lower bounds obtained by other means (Shearer [54]) for all d. The following table shows the bounds given by the simple greedy and degree-greedy algorithms for small d, as well as the upper bounds from [36].

d	simple greedy	degree-greedy	upper bound
3	0.3750	0.4328	0.4554
4	0.3333	0.3901	0.4163
5	0.3016	0.3566	0.3844
6	0.2764	0.3296	0.3580
7	0.2558	0.3071	0.3357
8	0.2386	0.2880	0.3165
9	0.2240	0.2716	0.2999
10	0.2113	0.2573	0.2852

Finally, for $d = 3$ the explicit solution $z^{(1)} = 6\ln\frac{3}{2} - 2$ can be derived. This value was found in [21] basically using the same idea as in Section 3.3.1 for 2-processes; that is, making use of the sum of degrees condition to eliminate one variable. (In [21] the behaviour of Y_0 was determined in [20] using approximation by differential equations.)

Eliminating only one variable does not seem to help much for $d > 3$, so fix on $d = 3$. Then there is only one phase, so $k = 1$ and the equation for z_0' in (7.5) is

$$\frac{dz_0}{dz} = \frac{2\tau_0}{1+\tau_2}, \tag{7.6}$$

$$\tau_0 = \frac{-3z_0(\xi + 6z_0 + 2z_1)}{\xi^2}, \tag{7.7}$$

$$\tau_2 = \frac{2z_1(6z_0 + 2z_1)}{\xi^2}. \tag{7.8}$$

Since Y_2 is virtually 0 during phase 1 (this observation relies on the differential equation approximation being valid; $z_2 = 0$ in phase 1), the edges added during this process reach new vertices which have degree 0 with probability $3Y_0/(3Y_0 + 2Y_1)$. So for each edge added this is the expected decrease in Y_0. (We also require the fact that vertices of degee 0 are usually not added to the independent set, except for the very first vertex; this comes either from the observation that a random cubic graph is almost always connected (see [12] or [59]), or as a corollary of the fact that the number of vertices of degree 1 grows away from 0 during this phase, which also follows from the validity of the differential equation approximation.) On the other hand, with every edge added, ξ decreases by 2. It follows that $\frac{d\xi}{dz_0} = \frac{2\xi}{3z_0}$, since $n\xi$ and nz_0 approximate $3Y_0 + 2Y_1$ and Y_0. Solving this equation with initial condition $\xi = 3$ when $z_0 = 1$ gives $\xi = 3z_0^{2/3}$. Substituting this and $z_1 = \frac{1}{2}\xi - \frac{3}{2}z_0$ into (7.6) and solving (Maple or Mathematica helps!) with initial condition $z_0 = 1$ when $z = 0$ gives z as the following function of z_0:

$$6\ln 3 - 2 - 2\ln(z_0 + 8) - 4\ln(2 + z_0^{1/3}) + 2\ln(z_0^{2/3} - 2z_0^{1/3} + 4) + 3z_0^{1/3} - \frac{3}{2}z_0^{2/3} + \frac{1}{2}z_0.$$

Putting $z_0 = 0$ to find the end of phase 1 yields $z^{(1)} = 6\ln\frac{3}{2} - 2$ as required.

7.2 Greedy packing

Let $G_0 = (V, E)$ be a k-uniform simple hypergraph and consider the greedy packing process. Each step deletes a randomly chosen hyperedge and all the vertices it contains. Thus all hyperedges containing those deleted vertices are also deleted. We are interested mainly in how many vertices remain (of degree 0) at the end of the process.

Let $\nu = |V|$ and $\mu = |E|$. The variables we will consider are the degrees of the vertices: let $Y_i(t)$ denote the degree of vertex i in G_t. We assume at the outset that

$$|Y_i(0) - r| \leq \delta \tag{7.9}$$

for all i, where δ and r are functions of n. We only obtain useful results here when $\delta << r << \nu$. (Other cases can also be treated, but to simplify this discussion if we concentrate on the case that the differences have tails.) For asymptotics these variables all go to ∞. Throughout, $k \geq 3$ will be fixed.

Since each vertex degree is only likely to change very occasionally, we scale time by a factor of

$$M = \frac{\nu}{r}.$$

That is, redefine G_{t+1} to be the graph obtained after M hyperedges are chosen and deleted *starting with* G_t. We call these *deletion operations*, to distinguish them from the steps of the process (each of which contains M deletion operations). We assume for convenience that M is an integer; this assumption does not alter the outcome of the argument. This particular scaling is chosen just to conveniently conform to the setting of Theorem 5.1, in which the range of the variables is roughly of the same order of magnitude as the length of the process. Note here that the process must end after at most ν/k hyperedges are deleted, since the k vertices in each hyperedge disappear forever. Thus, the length of the process in scaled time is at most r/k.

Continuation of the process depends on some of the remaining vertices having non-zero degree, which leads to the idea of keeping track of the vertex degrees during the process. To gain some idea of what we might expect the truth to be, regarding the number of vertices surviving to the end, it is reasonable (but has not been proved yet) to suppose that at any time until near the end of the process, the graph is "almost" random given the number of edges. This cannot hold forever, but let us be optimistic and expect the process to keep going until its vertex degrees become roughly constant, from which time we might suppose that some significant proportion of the surviving vertices will survive to the end of the process.

We first run through a trial calculation using the uniform randomness assumption, since a more accurate version is exactly what we need later. In the trial calculation, we can pretend that all vertices have the same degree, say $z(t/r)r$, at time t. We write x for t/r throughout.

There are altogether exactly $\nu - kMt$ surviving vertices remaining in the graph G_t. The number of hyperedges is the sum of degrees divided by k, namely

$$\frac{(\nu - kMt)rz}{k} = \frac{\nu(1 - kx)rz}{k}. \tag{7.10}$$

We can also estimate the change in the degree of a surviving vertex per step. Each vertex is in rz hyperedges, each of which has $k-1$ other vertices and hence is deleted if any of its intersecting rz hyperedges is the chosen one. The (estimated) probability of reducing the degree in one deletion operation is therefore the product of these three numbers divided by the number of hyperedges given above; that is, $\frac{k(k-1)rz}{\nu(1-kx)}$. Multiplying by $M = \nu/r$, we obtain the expected degree reduction in one step, conditional on the vertex surviving:

$$\frac{k(k-1)z}{(1-kx)}. \tag{7.11}$$

Using the usual differential equation philosophy, the suggested equation is

$$z'(x) = -\frac{k(k-1)z}{(1-kx)} \tag{7.12}$$

with initial condition $z(0) = 1$ since z is the degree scaled by $1/r$. The solution is

$$z(x) = (1-kx)^{k-1}. \tag{7.13}$$

The degrees drop to about 1 when z is $1/r$, that is, the number of surviving vertices, $\nu(1-kx)$, is

$$\frac{\nu}{r^{\frac{1}{k-1}}}. \tag{7.14}$$

In the example in the introduction, where the vertices of G_0 are edges of the complete graph K_n, and the edges are the triangles of K_n, the process greedily deletes the edges of triangles until no triangles remain. Here $k = 3$, $\nu = \binom{n}{2}$ and $r = n-1$, so (7.14) suggests that the number of edges remaining at the end of the process is about $n^{3/2}$, in line with Spencer's conjecture.

We now take a rigourous look at the general situation. As mentioned in Section 2, we are not aiming here for the best possible value of ϵ which this method will deliver.

Theorem 7.1 *Let G be a k-uniform hypergraph with ν vertices. Assume $\nu < r^C$ for some constant C, $\delta = O(r^{1/3})$ and $r = o(\nu)$. Also assume (7.9) holds. Then for any $\epsilon_0 < \frac{1}{9k(k-1)+3}$, a.a.s. at most ν/r^{ϵ_0} vertices remain at the end of the greedy packing process applied to G.*

Proof. Instead of trying to force the vertex degrees into a differential equation with ν variables, it seems to be no loss to use just one variable, whose differential equation will be (7.13). Theorem 5.1 does not apply directly so we have to rework the proof slightly.

First examine a generic situation: suppose that at time $t = rx$ the degrees of all the vertices are approximately equal, and in fact the degree Y_i of each vertex i satisfies the inequality

$$|Y_i - rz| < \xi \tag{7.15}$$

for some $\xi = \xi_x$ where

$$0 < \log^2 r < \xi = o(rz). \tag{7.16}$$

Here and in the following we use (7.13) for the definition of z. We will find upper and lower bounds on the expected change in Y_i during one step for each vertex i, as well as a bound on the tail of the distribution, conditional on the vertex not being deleted during this step.

First consider the number of hyperedges present during the next step of the process (i.e., the next M deletion operations). There are altogether exactly $\nu - kMt$ undeleted vertices remaining in the graph G_t. We can then replace the estimate (7.10) of the number of hyperedges at the start of this step by the more accurate

$$(rz + \phi\xi)(\nu - kMt)/k = \frac{rz\nu(1-kx)}{k}(1 + \phi\xi/rz)$$

where ϕ denotes a function whose absolute value is bounded above by 1, different at each occurrence. At each of the M deletion operations throughout this step, k vertices are deleted of degree $O(rz)$, so $O(krz) = O(rz)$ hyperedges are lost, accumulating to at most $O(Mrz) = O(\nu z)$ throughout the whole step. So the number of hyperedges is always

$$\frac{\nu(1-kx)rz}{k}\left(1 + \phi\xi/rz + O\left(\frac{1}{r(1-kx)}\right)\right) \tag{7.17}$$

throughout this whole step.

The next task is to bound with high probability the change in a vertex degree in this step. Later we compute the expected change, focussing on vertex 1. We assume this vertex is not deleted in this step. By (7.15), the argument above estimating the probability that the current degree of vertex 1, i.e. $d(1)$, is changed in one deletion operation is valid to within a constant factor, giving the rough upper bound on this probability as

$$p = O\left(\frac{rz}{\nu(1-kx)}\right) = O\left(\frac{(1-kx)^{k-2}}{M}\right)$$

by (7.13).

Let us call a deletion operation where $d(1)$ changes "successful" and let s denote the number of these. For a simple large deviation inequality on s, just compute $\mathbf{E}\binom{s}{b}$ where $b = \lfloor \log r \rfloor$. This can alternatively be expressed as the sum, over all b-subsets of the M operations, of the probability that all operations in the subset are successful. Note the indicator variables for the successful operations are *not* independent here, but nevertheless the probability that a given b-subset is successful is bounded above by p^b using a chain of *conditional* probabilities. Thus, by linearity of expectation and since $\log \nu = O(\log r)$,

$$\mathbf{E}\binom{s}{b} \le \binom{M}{b}p^b = \left(\frac{O(1)}{b}\right)^b = O(\nu^{-3}).$$

Hence using Markov's inequality (4.1) with $\alpha = C\nu^3$,

$$\mathbf{P}(s \ge b) = \mathbf{P}\left(\binom{s}{b} \ge 1\right) = O(\nu^{-3}).$$

Note that since the hypergraph is simple, $d(1)$ can decrease by at most k for every deletion operation, so $Y_1(t) - Y_1(t+1) < kb$ if $s < b$. Thus with $\beta = kb$ and $\gamma = O(\nu^{-3})$ we obtain the boundedness hypothesis of Theorem 5.1 *for this particular step*. Applying this argument to all vertices gives

$$\mathbf{P}\left(\max_{1\le i\le \nu}\{Y_i(t) - Y_i(t+1)\} \ge \beta\right) \le \gamma \tag{7.18}$$

with

$$\beta = k\lfloor \log r \rfloor, \qquad \gamma = O(\nu^{-2}).$$

Now we assume $\gamma = 0$; on reflection at the end of the argument it will be seen that such a small γ has no effect on the conclusion.

We now turn to estimating the expected decrease in $d(1)$ during these M operations. By linearity of expectation, this is

$$\mathbf{E}(Y_1(t) - Y_1(t+1)) = \sum_{j=1}^{M} \mathbf{E}\Delta_j$$

where Δ_j is the decrease in the j'th deletion operation. Compute $\mathbf{E}\Delta_j$ as

$$\mathbf{E}(\Delta_j \mid A)\mathbf{P}A + \mathbf{E}(\Delta_j \mid \overline{A})\mathbf{P}\overline{A}$$

where A is the event that already in this step, up to this operation, at least one vertex has decreased its degree by at least β, and $\overline{A}$ is the complement of A. The greatest possible decrease is clearly $O(r)$, and $\mathbf{P}A = O(\nu^{-2})$ by (7.18). So $\mathbf{E}(\Delta_j \mid A)\mathbf{P}A = O(\nu^{-1})$. For $\mathbf{E}(\Delta_j \mid \overline{A})$, we just have to run through the earlier estimates more carefully, knowing that the degrees have decreased by at most β since the start of this step. Thus, from (7.15), each vertex now has degree

$$rz + \phi\xi + O(\beta). \tag{7.19}$$

In the argument leading to (7.11), the degree of vertices was estimated three times, twice in the numerator and once in the denominator coming from (7.10). Hence, without multiplying by M, we have

$$\begin{aligned}\mathbf{E}(\Delta_j \mid \overline{A}) &= \frac{k(k-1)z}{M(1-kx)} \times \frac{(1+\phi\xi/rz + O(\beta/rz))^2}{1+\phi\xi/rz + O(\beta/rz)} \\ &= \frac{k(k-1)z}{M(1-kx)}\left(1 + \frac{(3\phi + o(1))\xi}{rz}\right),\end{aligned}$$

being careful with ϕ's (which can be different at different occurrences so the identical-looking factors do not cancel), and using (7.16). Thus

$$\mathbf{E}(Y_1(t) - Y_1(t+1)) = \frac{k(k-1)z}{(1-kx)}\left(1 + \frac{(3\phi + o(1))\xi}{rz}\right) = f(x) + \phi\lambda_x$$

in place of the trend hypothesis, where by (7.13)

$$f(x) = k(k-1)(1-kx)^{k-2}, \qquad \lambda_x = \frac{(3k(k-1)+o(1))\xi}{r(1-kx)}.$$

Hence, looking a little later in the process and guided by (5.3), assuming the bounds (7.15) on vertex degrees, we have

$$\begin{aligned}\mathbf{E}(Y_1(t+j+1) - Y_1(t+j) \mid H_{t+j}) &= f(x+j/r) + \phi\lambda_{x+j/r} \\ &= f(x) + O(w/r)(1-kx)^{k-3} + \phi\lambda_x(1+o(1))\end{aligned}$$

provided $j \le w = o(r(1-kx))$.

We follow the rest of the proof of Theorem 5.1, but this time Y is the degree of any particular vertex, as above. The same argument is applied to each of the vertices separately, in order to pass from time t to $t+w$. The conclusion of the induction will be almost sure bounds on *all* the degrees of the vertices at time $t+w$. It is now simpler to have the length w of the supermartingale a function of i; we change w to

$$w_i = r^{1-\epsilon}(1-kx_i) \tag{7.20}$$

(actually, rounded to an integer) for any $\epsilon < 1/3$ and where $x_i = k_i/r$ is the current value of t/r at the start of the supermartingale, and inductively $k_{i+1} = k_i + w_i$. For later reference, since each w_i covers $1 - r^{-\epsilon}$ of the distance from k_i to r/k, we have

$$k_i \approx \frac{r}{k}\left(1 - (1-r^{-\epsilon})^i\right) \tag{7.21}$$

where the approximation is only due to integer rounding at each step. The estimate above now permits

$$g(r) = r^{-\epsilon} + \phi\lambda_x(1+o(1))$$

in place of (5.4). Then we conclude (5.6) as before. In (5.7), just take $\alpha = 3\log\nu = O(\log r)$ so that the probability is small enough to be able to multipy by ν, for the number of variables, and r, for the number of values of i (at most). Now apply the rest of the argument of the proof of Theorem 5.1, starting with $B_i = \xi_i$. Then from the supermartingale argument,

$$\begin{aligned}|A_2| < w_i g(n) + \sqrt{w_i}(\log r)^{O(1)} &= O(r^{1-2\epsilon}) + \phi r^{1-\epsilon}(3k(k-1)+o(1))\xi_i \\ &= O(r^{1-2\epsilon}) + rTB_i\end{aligned}$$

where $T = \phi r^{-\epsilon}(3k(k-1)+o(1))$. As before $A_1 = B_i$, $A_3 = O(w_i^2/r) = O(r^{1-2\epsilon})$ and this time $A_4 = 0$ since $z = f'(x)$ by the definition of z.

We can thus define

$$B_{i+1} = O(r^{1-2\epsilon}) + B_i(1+T)$$

and $B_0 = r^{1/3}$ is in agreement with the bounds on the degrees of the vertices given initially. Solving the recurrence, define

$$B_s = \sum_{i=1}^{s} O(r^{1-2\epsilon})(1+T)^s = O(sr^{1-2\epsilon})e^{sT}$$

and the bounds (7.15) are established inductively on all the vertex degrees, with $\xi_i = B_i$ (provided (7.16) holds). The probability of failure of the bounds is easily $o(1)$ if $s = O(r \log r)$. Choosing $s = s_0 = \epsilon_1 r^\epsilon \ln r / 3k(k-1)$, we obtain

$$B_{s_0} = O(r^{1-\epsilon+\epsilon_1}).$$

Now from (7.21),

$$k_{s_0} \approx \frac{r}{k} - r^{1-\epsilon_1/3k(k-1)},$$

which represents $x = \frac{1}{k} - r^{-\epsilon_1/3k(k-1)}$, and so at this time in the process, from (7.13), $rz = k^{k-1} r^{1-\epsilon_1/3k(k-1)}$. Thus, if $\epsilon_1(1 + 1/3k(k-1)) < \epsilon$, we have $B_{s_0} = o(rz)$ and thus a.a.s. the vertices all have degree at least 0. So the process a.a.s. finishes later than this. Since $\epsilon < 1/3$, this means we can choose any $\epsilon_1 < \frac{k(k-1)}{3k(k-1)+1}$. The number of vertices remaining is $k\nu/r$ times the remaining number of process steps, $r/k - k_{s_0}$, and is thus ν/r^{ϵ_0} for any $\epsilon_0 < \frac{1}{9k(k-1)+3}$, as claimed. ∎

Postscript. The reason that Theorem 5.1 does not help when applied directly to the unscaled packing process is that the variables only change very slowly: their expected changes are much smaller than the (constant) upper bound on their maximum change in one deletion operation. But the errors in approximation in the theorem are determined only by the upper bound on the differences, and the number of steps. After scaling time the differences are effectively multiplied by only $\log \nu$ but the number of steps is divided by a power of ν.

Note that there is another way to deal with the problem of slowly changing variables: use the philosophy of Theorem 5.1 or similar, but apply a different type of martingale inequality such as in McDiarmid [34, Theorem(6.1)] specially adapted to the case that the expectation of the changes is much smaller than the maximum change. To prove such inequalities one can adapt the proof of Lemma 4.2 to the case that the differences are 0 with high probability. The main difference stems from a different estimation of the exponential e^{hx}. It is often that a good choice of h has hx very small, so useful results are obtained by expanding the exponential using Taylor series. (See the proofs of Theorem (6.1) and Corollary (5.2) in [34].) The results obtained from doing this are not very much different from that above.

On the other hand, the simple wholistic approach referred to at the start of Section 5.2 does significantly help in the present problem. But this still does not seem to imply anything better than the result in [24].

8 Other processes

In this closing section we describe some of the other graph processes and greedy algorithms to which the differential equation method has been applied.

Dominating number of random cubic graphs

The dominating number of a graph is the size of a smallest set S of vertices such that every vertex not in S is adjacent to at least one vertex in S. Molloy and Reed [41] used the differential equation method to obtain $0.3126n$ as an almost sure upper bound on the

dominating number of a random n-vertex 3-regular graph by analysing a greedy algorithm which walks along a Hamilton cycle. This has similarities to the degree-greedy algorithm for independent sets, and relies on results of Robinson and Wormald which imply that for proving almost sure bounds, one can assume that a random 3-regular graph is formed from a Hamilton cycle plus a random matching. Duckworth and Wormald [18] have recently found that a more direct analogue of the degree-greedy approach performs better (again using the differential equation method).

The pairing process for random graphs with given degrees

Molloy and Reed analysed this process, which generates random graphs with given degrees u.a.r., to show that the size of (number of vertices in) the largest component is sharply concentrated near a value which they determined. The variables of interest are the numbers $Y_i(t)$ of vertices of degree i still containing i unpaired points after t pairs have been chosen. The main theorem of [61] was applied separately to each Y_i to show that these numbers are sharply concentrated, and established from this the asymptotic concentration and value of the size of the largest component in the final graph.

Greedy list-colouring process

Finding the chromatic number of an arbitrary graph also requires solving an NP-complete problem. A natural heuristic for k-colouring a graph can be described as follows. Assume there are k colours available. Assign colours to the vertices one by one (in a way to be described). For each vertex maintain a list of the colours available; that is, not appearing on neighbours already coloured. The simple criterion for choosing the next vertex to be coloured is to choose it u.a.r. from those with the shortest lists. The colour is also chosen u.a.r. from those on the list. If no list ever becomes empty, the heuristic successfully finds a k-colouring of the graph. In this case the process is a process of graphs whose vertices have associated lists. (A coloured vertex is equivalent to a vertex with a list containing only one colour.)

Achlioptas and Molloy [1] found the threshold in $\mathcal{G}(n,p)$ (the common model of random graphs) for almost sure success of this process using the differential equation approach. The variables of interest here are the numbers of vertices with a list of length i, for each $i \leq k$. They found they could solve the resulting system of differential equations for $k = 3$, and hence determined the threshold in $\mathcal{G}(n,p)$ at which the greedy list-colouring heuristic becomes almost surely successful. For larger k, inequalities were obtained on the solutions.

Karp-Sipser algorithm

A simple algorithm for finding a matching in a random graph in $\mathcal{G}(n,p)$ is to repeat the step of selecting a random edge X to add to the matching and then deleting the vertices of the edge (and all incident edges). If the edge is selected u.a.r. this gives a simple greedy algorithm. Karp and Sipser [29] modified this as follows: in the steps where vertices of degree 1 exist, choose one u.a.r. and let X be its incident edge. They showed using a differential equation method that that for $p = O(n^{-1})$ this algorithm a.a.s. finds a matching which has cardinality within $o(n)$ of the maximum matching in the graph. Aronson et al. [5] gave a more precise analysis using a wholistic approach similar to that in [44], and amongst other things improved the $o(n)$ to $O(n^{1/5}\log^K n)$ for a particular K.

Greedy heuristics for random instances of k-sat

Consider a random logic formula in conjunctive normal form with n variables and exactly k literals per clause. Finding a truth assignment satisfying such a formula is NP-hard, so heuristic algorithms are of interest. A class of greedy heuristics uses the repeated step of randomly choosing a literal x, deleting all clauses containing x, and deleting all occurrences of $\bar{x}$ from other clauses. If any clauses become empty the heuristic fails. Various rules for selecting x (which give various heuristics) have been studied. In a particular version studied by Chao and Franco [14, 15] and Frieze and Suen [22], x is chosen from a randomly selected clause of smallest size. This is then very similar to the degree-greedy algorithm for independent sets. It even exhibits just the same types of phases. (Other variations resemble the Karp-Sipser algorithm above, for matchings.) These papers use the differential equation method to study the probability the heuristic succeeds on a random formula with cn clauses. In [14, 15], an appropriate system of differential equations is derived based on variables which count clauses containing i literals (analogous to vertices of degree i in graph processes). There, arguments concerning sums of binomials are used in place of supermartingale inequalities to obtain some rough estimates. In [22], the case $k = 3$ is examined for this heuristic and more precise results are obtained using a piecewise approach. This is very similar to a special case of Theorem 5.1, but with an argument about sums of nearly independent and nearly equal binomial variables playing the role of a supermartingale inequality. An argument analogous to that in Section 7.1 will no doubt give precise results for arbitrary k.

ACKNOWLEDGMENTS

My appreciation goes to Michał Karonski and Hans Jürgen Prömel for arranging the Summer School. I am greatly indebted to my wife, Hania, for her encouragement and for invaluable assistance with revising the early draft of these notes. Thanks are also due to Ian Wanless for careful a reading.

References

[1] D. Achlioptas and M. Molloy, The analysis of a list-coloring algorithm on a random graph, *FOCS 97*, 204–212.

[2] D.J. Aldous, Emergence of the giant component in special Marcus-Lushnikov processes, *Random Structures and Algorithms* **12** (1998), 179–196.

[3] N. Alon, J.H. Kim and J. Spencer, Nearly perfect matchings in regular simple hypergraphs, *Israel J. Math.* **100** (1997), 171–187.

[4] N. Alon, J.H. Spencer and P. Erdős, *The Probabilistic Method.* Wiley, New York, 1992.

[5] J. Aronson, A. Frieze and B.G. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser revisited, *Random Structures and Algorithms* **12** (1998), 111–177.

[6] K. Azuma, Weighted sums of certain dependent random variables, *Tôkuku Math. J.* **19** (1967), 357–367.

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

[8] E.A. Bender, Asymptotic Methods in Enumeration, *SIAM Review 16* (1974), 485–515.

[9] E. A. Bender and E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences, *J. Combin. Theory Ser. A* **24** (1978), 296–307.

[10] E.A. Bender, E.R. Canfield and B.D. McKay, The asymptotic number of labeled connected graphs with a given number of vertices and edges, *Random Structures and Algorithms* **1** (1990), 127–169.

[11] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *Eur. J. Combinat.* **1** (1980), 311–316.

[12] B. Bollobás, *Random Graphs*. Academic Press, London, 1985.

[13] B. Bollobás, Martingales, isoperimetric inequalities and random graphs, in *Combinatorics*, Eger (Hungary) 1987. *Coll. Math. Soc. J. Bolyai, 52, Akad. Kiadó*, Budapest (1988), 113–139.

[14] M. T. Chao and J. Franco, Probabilistic analysis of two heuristics for the 3-satisfiability problem, *SIAM J. Comput.* **15** (1986), 1106–1118.

[15] M. T. Chao and J. Franco, Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k-satisfiability problem, *Information Sciences* **51** (1990), 289–314.

[16] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23** (1952), 493–507.

[17] V. Chvátal, Almost all graphs with $1.44n$ edges are 3-colourable, *Random Structures and Algorithms* **2** (1991), 11–28.

[18] W. Duckworth and N.C. Wormald, Minimum dominating sets in random cubic graphs (in preparation).

[19] W. Feller, *An Introduction to Probability Theory and Its Applications, II (2nd Edition).* Wiley, New York, 1971.

[20] A.M. Frieze, A.J. Radcliffe and S. Suen, Analysis of a simple matching algorithm on random cubic graphs. *SODA 1993*, 344–351.

[21] A.M. Frieze and S. Suen, On the independence number of random cubic graphs, *Random Structures and Algorithms* **5** (1994), 649–664.

[22] A.M. Frieze and S. Suen, Analysis of two simple heuristics on a random instance of k-SAT, *J. Algorithms* **20** (1996), 312–355.

[23] A.P. Godbole and P. Hitczenko,Beyond the method of bounded differences. *Microsurveys in discrete probability* (Princeton, NJ, 1997), 43–58, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., **41**, Amer. Math. Soc., Providence, RI, 1998.

[24] D. Grable, On random greedy triangle packing, *Elec. J. Combinat.* **4** (1997), #R11.

[25] G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes*, 2nd edn, Clarendon, Oxford (1992).

[26] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Am. Stat. Assoc.* **58** (1963), 13–30.

[27] W. Hurewicz, *Lectures on Ordinary Differential Equations*, M.I.T. Press, Cambridge Massachusetts (1958).

[28] S. Janson, D.E. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, *Random Structures and Algorithms* **4** (1993), 233–358.

[29] R.M. Karp and M. Sipser, Maximum matchings in sparse random graphs, *Proceedings of the Twenty-Second Annual IEEE Symposium on Foundations of Computing* (1981), 364–375.

[30] D. Knuth, R. Motwhani and B. Pittel, Stable husbands, *Random Structures and Algorithms* **1** (1990), 1–14.

[31] T.G. Kurtz, Solutions of ordinary differential equations as limits of pure Markov jump processes, *J. Appl. Probab.* **7** (1970), 49–58.

[32] T.G. Kurtz, *Approximation of Population Processes*, SIAM, Philadelphia (1981).

[33] T. Łuczak, Size and connectivity of the k-core of a random graph, *Discrete Math.* **91** (1991), 61–68.

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

[35] B. D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combinatoria* **19A** (1985), 15–25.

[36] B.D. McKay, Independent sets in regular graphs of high girth, *Proceedings of the Australia-Singapore Joint Conference on Information Processing and Combinatorial Mathematics, Singapore, 1986. Ars Combinatoria* **23A** (1987), 179–185.

[37] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica* **11** (1991), 369–382.

[38] B.D. McKay and N.C. Wormald, The degree sequence of a random graph. I. The models, *Random Structures and Algorithms* **11** (1997), 97–117.

[39] H. M. Mahmoud, R. T. Smythe and J. Szymański, On the structure of random plane-oriented recursive trees and their branches, *Random Structures and Algorithms* **4** (1993), 151–176.

[40] M. Molloy, M. Math. Thesis, University of Waterloo (1992).

[41] M. Molloy and B. Reed, The dominating number of a random cubic graph, *Random Structures and Algorithms* **7** (1995), 209–221.

[42] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combinatorics, Probability and Computing* **7** (1998), 295–305.

[43] B. Pittel, On tree census and the giant component in sparse random graphs, *Random Structures and Algorithms* **1** (1990), 311–342.

[44] B. Pittel, J. Spencer and N.C. Wormald, Sudden emergence of a giant k-core in a random graph, *J. Combinatorial Theory, Series B* **67** (1996), 111–151.

[45] H. Robalewska, *Stochastic Processes and Random Graphs*, PhD Thesis, University of Melbourne, 1995.

[46] H. Robalewska, Random graphs generated by the star 2-process, *Random Structures and Algorithms* **11** (1997), 125–149.

[47] H. Robalewska and N.C. Wormald, Random star d-processes (to appear).

[48] V. Rödl, On a packing and covering problem, *Eur. J. Combinat.* **5** (1985), 69–78.

[49] V. Rödl and L. Thoma, Asymptotic packing and the random greedy algorithm, *Random Structures and Algorithms* **8** (1996), 161–177.

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

[51] A. Ruciński and N.C. Wormald, Random graph processes with maximum degree 2, *Annals of Applied Probability* **7** (1997), 183–199.

[52] A. Ruciński and N.C. Wormald, The probability that a random d-process results in a connected graph, *J. Austral. Math. Soc.* (submitted).

[53] E. Shamir and J. Spencer, Sharp concentration of the chromatic number on random graphs $G_{n,p}$, *Combinatorica* **7** (1987), 121–129.

[54] J.B. Shearer, A note on the independence number of triangle-free graphs, II, *J. Combinatorial Theory, Series B* **53** (1991), 300–307.

[55] J. Spencer, Asymptotic packing via a branching process, *Random Structures and Algorithms* **7** (1995), 167–172.

[56] A. Srinivasan, Approximation algorithms via randomised rounding: a survey, *This Volume*.

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

[58] A. Telcs and N.C. Wormald, Hamiltonicity of random graph processes with maximum degree 2 (in preparation).

[59] N.C. Wormald, The asymptotic connectivity of labelled regular graphs, *J. Combin. Theory Ser. B* **31** (1981), 156–167.

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

[61] N.C. Wormald, Differential equations for random processes and random graphs, *Annals of Applied Probability* **5** (1995), 1217–1235.