Combinatorics, Probability and Computing (2000) 9, 549–572. Printed in the United Kingdom
© 2000 Cambridge University Press

The Maximum Degree of a Random Graph

OLIVER RIORDAN[1] and ALEX SELBY[2]

[1] Trinity College, Cambridge CB2 1TQ, England

[2] Department of Pure Mathematics and Mathematical Statistics, 16 Mill Lane, Cambridge CB2 1SB, England

Received 21 April 1999; revised 30 March 2000

Let $0 < p < 1$, $q = 1 - p$ and b be fixed and let $G \in \mathcal{G}(n,p)$ be a graph on n vertices where each pair of vertices is joined independently with probability p. We show that the probability that every vertex of G has degree at most $pn + b\sqrt{npq}$ is equal to $(c + o(1))^n$, for $c = c(b)$ the root of a certain equation. Surprisingly, $c(0) = 0.6102\ldots$ is greater than $\frac{1}{2}$ and $c(b)$ is independent of p. To obtain these results we consider the complete graph on n vertices with weights on the edges. Taking these weights as independent normal $N(p, pq)$ random variables gives a 'continuous' approximation to $\mathcal{G}(n,p)$ whose degrees are much easier to analyse.

1. Introduction

Let $0 < p < 1$ be fixed and $q = 1 - p$. Throughout the paper we consider a random graph G from the standard model $\mathcal{G}(n,p)$. Thus G has n vertices, and each possible edge is present independently with probability p. The study of the distribution of the maximum degree $d^{\max}(G)$ of such a graph was started by Erdős and Rényi [5]; in the general context of describing as accurately as possible 'almost all' such graphs G, very precise results were given by Bollobás [1] (see also [2]). In fact, these results go much further, giving information about the whole degree sequence. More recently, McKay and Wormald [9] described simple models which approximate the distribution of the entire degree sequence of G, ignoring events of probability $n^{-\omega(n)}$ for some $\omega(n) \to \infty$. It might thus seem that there is nothing more to be said about $d^{\max}(G)$; however, the results described above all apply only to 'typical' random graphs.

Taking $p = \frac{1}{2}$, the probability that a particular vertex of G has degree at most $n/2$ is just over one half. Also, these events are very nearly independent, so it is natural to ask whether the probability that every vertex has degree at most $n/2$, *i.e.*, that $d^{\max}(G) \leqslant n/2$, is roughly 2^{-n}. This question, which we have heard from two different sources, was originally posed by Vera Sós in 1982, but does not seem to have appeared in print.

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

It is immediate that the above probability is at least 2^{-n}; for $x \in V(G)$ the event $\{d(x) \leqslant n/2\}$ is a *down-set*, that is, an event closed under deleting edges from G. Kleitman's Lemma [8] (see also [3], §19) tells us that for down-sets $\mathcal{D}_1$ and $\mathcal{D}_2$ we have $\mathbb{P}(\mathcal{D}_1 \cap \mathcal{D}_2) \geqslant \mathbb{P}(\mathcal{D}_1)\mathbb{P}(\mathcal{D}_2)$. Since the intersection of two down-sets is again a down-set, applying this $n-1$ times gives $\mathbb{P}(d^{\max}(G) \leqslant n/2) \geqslant 2^{-n}$.

The results of [2] and [9] tell us that for many purposes the degrees of G behave roughly as if they were independent. However, these results do not apply to events of the very small probabilities considered here. In fact the dependence does matter, pushing $\mathbb{P}(d^{\max}(G) \leqslant n/2)$ up to c^n for some $c > \frac{1}{2}$. The techniques used here will thus be very different – rather than bounding from above the effect of the dependence, we must estimate it accurately. We do this by 'uncovering' the random graph one vertex at a time in a certain way.

One of the main ideas we shall use is to consider *weighted* random graphs. We can think of $G \in \mathcal{G}(n,p)$ as the graph K_n, where each edge is given an independent random weight, which is 1 with probability p and 0 otherwise. Here we consider instead normally distributed weights with the same mean p and variance pq. As multivariate normal distributions are very simple, the maximum 'degree' of such a weighted random graph is much easier to analyse. Most of the work is to show that the discrete and continuous models behave in roughly the same way.

In fact, the weighted graphs can be thought of as a convenience, which is probably not an essential part of the proof. The original approach was a direct calculation in the discrete case, sketched in the appendix. However, making the final steps rigorous seemed to be rather awkward. Instead we observed that following the first part of an analogous calculation for the continuous case led to a point where the same estimate was an approximation for both the continuous and discrete answers. At this point the calculation could be abandoned, and the result for the continuous case used as an approximation for the discrete one.

The rest of the paper is organized as follows. In Section 2 we state the main result precisely. In Section 3 we give an outline of the proof, sketching a few of the difficulties. In Section 4 we describe the notation we shall use. Section 5 gives the (short) calculation for the continuous case, in fact to rather greater accuracy than needed here. In Section 6 we describe certain 'bad' events and show that they have small probabilities. In Section 7 we give the heart of the proof, postponing the proof of a key approximation lemma to Section 8. Finally, in the appendix we outline our original approach using the calculus of variations.

2. Statement of result

We write $\phi()$ and $\Phi()$ for the standard normal density and distribution functions, respectively.

Theorem 2.1. *Let $0 < p < 1$, $q = 1-p$ and b be fixed. For G a random graph from $\mathcal{G}(n,p)$, let*

$$p_n = \mathbb{P}(d^{\max}(G) \leqslant np + b\sqrt{npq}).$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

Then

$$p_n^{1/n} \to c \quad as \quad n \to \infty,$$

where c is a constant depending only on b, given by

$$c = \max_\theta \{\Phi(b+\theta)e^{-\theta^2/2}\}. \tag{2.1}$$

The above maximum is attained (uniquely) at $\theta_0 = \theta_0(b)$, the root of the equation

$$\theta_0 \Phi(b+\theta_0) = \phi(b+\theta_0). \tag{2.2}$$

If $p = q = \frac{1}{2}$ and $b = 0$ then $\theta_0 = 0.5060544689\ldots$ and $c = 0.6102303991\ldots$.

Remark. If the degrees behaved as independent random variables then $p_n^{1/n}$ would tend to $\lim_{n\to\infty} \mathbb{P}(d_1 \leqslant np + b\sqrt{npq}) = \Phi(b)$, where d_i denotes the degree of the ith vertex. Although $\Phi(b)$ is not the correct limit, it is not hard to see that it must be a lower bound, *i.e.*, that $\liminf p_n^{1/n} \geqslant \Phi(b)$, because the event $d_i \leqslant np + b\sqrt{npq}$ is a down-set.

3. Idea of proof

It is easy to get a result like Theorem 2.1 if, instead of 'weighting' each edge with 0 or 1 with probabilities q and p, which is effectively the case with $\mathcal{G}(n,p)$, we weight each edge with the normal distribution which has the same mean and variance, $N(p,pq)$. These two situations will be referred to as the *discrete* and *continuous* cases. However, to show *directly* that these models give the same result appears to be outside the scope of the usual methods used to prove versions of the central limit theorem.

We instead proceed by considering each vertex in turn and associating some 'state' to each i, $1 \leqslant i \leqslant n$, namely the number (or total weight), D_i, of edges from vertices with index up to i to those with index greater than i. One can think of $(D_i)_{i=1}^n$ as a random process. Roughly speaking, it turns out that D_i provides enough information to work out a suitable approximation to the probability that the degree bound on vertex $i+1$ is satisfied and the $(i+1)$st state is D_{i+1}. To make sense of this we must replace the process (D_i) by a discrete approximation. In fact, for reasons outlined below, we must also remove 'large swings', obtaining a process (D_i'). This is a process which is 'asymptotically' Markovian (in a sense made precise in Lemma 7.2) and which can also 'die' (owing to the bound condition not being satisfied). If we can show that the transition probabilities are similar in the discrete and continuous cases, then this will show that the probabilities of the degree bounds being satisfied are also similar. The exact distribution of D_{i+1}' given D_i' is hard to calculate, but it is possible to calculate its distribution given a certain σ-algebra, $\mathcal{F}_i$. If we call a vertex whose number is no more than i a 'left' vertex and the other vertices 'right' vertices, then $\mathcal{F}_i$ contains the information about any edge from a left vertex to another left vertex and also the information, for each left vertex, about the total weight of edges from that vertex to all right vertices.

Although the transition probabilities are usually similar in the discrete and continuous cases, they can be significantly different when i is too close to 1 or n, or when certain random variables take on extreme values. Coping with the first problem is easy: we merely

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

introduce a cut-off δ_c, and only consider i between $\delta_c n$ and $(1-\delta_c)n$. A simple argument using down-sets shows that working with this restriction and taking the limit $\delta_c \to 0$ does not change the answer. To deal with the second problem we introduce the concept of a *bad* event. This is a subset B of the probability space Ω on which the discrete and continuous random variables can behave in an unpredictable manner. We define four types of these and in each case show that their probability is small compared with c^n, which is essentially the probability we are trying to estimate. In the end this will allow us to disregard the bad events. However, simply dismissing them because their probability is small does not seem to be possible, because it is hard to see the effect of conditioning on them not happening. We therefore proceed in a more roundabout way.

We mention two other difficulties that arise. First, it would be convenient if the probability of $|D_{i+1} - D_i|$ being larger than $M\sqrt{n}$ was small compared with c^n, for some suitably large constant M. This would mean that we would not have to worry about the extremes of this distribution. However, a moment's thought reveals that this is not true, so instead we have to show that the *total* contribution to D_i from such large swings is suitably small. This is the purpose of B_r^2 defined later and explains why we need to work with D_i' rather than D_i.

The other inconvenience arises because we break up the set U on which the degree bounds are satisfied into many small sets, namely the possible paths of the process D_i'. We then compare the continuous and discrete probability of each path, hoping to show there is a small *relative* error. However, although the probability of a bad set is small compared with that of U, it is not necessarily small compared with that of a path. If paths were either wholly good or wholly bad there would not be a problem – we would just restrict ourselves to the good paths. Since this is not the case, on the bad sets we replace the random variables that contribute to D_i' with certain 'ideal' random variables. Only after we have summed the contribution from each path can we then consider the error introduced by the bad sets.

4. Notation used throughout

For convenience we rescale the edge weightings. In the discrete case, we replace the weight on each edge with an independent $(B(1,p)-p)/\sqrt{pq}$. Here and elsewhere in this paper, we shall use $B(n,p)$ to refer to the binomial distribution, $N(\mu,\sigma^2)$ to refer to the normal distribution, and shall allow ourselves the notational convenience of doing arithmetic with them. Thus, for example, the previous expression refers to a random variable which is equal to $\sqrt{q/p}$ with probability p and $-\sqrt{p/q}$ with probability q. If more than one distribution appears in an expression, then these are assumed to be independent copies. For example, we might write $\sum_{i=1}^n N(0,1) = N(0,n)$. The probability measure for the discrete case will be called $\mathbb{P}_d$. In the continuous case, we replace the weight on each edge with an independent $N(0,1)$. The probability measure for this case will be called $\mathbb{P}_c$. In either case, we replace the bound condition on each degree with '$d_i \leqslant b\sqrt{n}$', where d_i is the sum of the weights of the edges incident with the ith vertex. Clearly these three changes taken together make no difference. The symbol $\mathbb{P}$ will be used to mean $\mathbb{P}_c$ or $\mathbb{P}_d$ (*i.e.*, the

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

relevant statement is true in both cases), or it will be used for some auxiliary probability space. The context should make this clear.

We shall use the following notation:
$c = c(b)$ is the constant defined in Theorem 2.1, and
e_{ij} is the weight on the edge from the ith to the jth vertex for $i, j = 1, \ldots, n$, $i \neq j$, so $e_{ij} = e_{ji}$. The following random variables are functions of the e_{ij}:

$$\begin{aligned}
d_{r,i} &= \sum_{j>r} e_{ij} \quad \text{(defined if } i \leqslant r\text{)},\\
D_r &= \sum_{i\leqslant r} d_{r,i} = \sum_{i\leqslant r, j>r} e_{ij},\\
\tau_r &= ((n-r)\sqrt{n})^{-1} D_r \quad \text{(defined if } r < n\text{)},\\
X_r &= n^{-1/2} \sum_{i<r} e_{ir},\\
Y_r &= n^{-1/2} \sum_{j>r} e_{rj},\\
d_r &= \sum_{j\neq r} e_{rj} = \sqrt{n}(X_r + Y_r) = \text{'degree' of } r\text{th vertex}.
\end{aligned}$$

In addition to these, we shall use auxiliary random variables $Z_{1,r}$ and $Z_{2,r}$ where

$$\begin{aligned}
Z_{1,r} &\sim N(0,1),\\
Z_{2,r} &\sim N(0,1),
\end{aligned}$$

and these are independent of each other and the e_{ij}. We shall write Ω for the probability space carrying the e_{ij} and the $Z_{l,r}$. The $Z_{l,r}$ will be used in Section 7 to define the ideal random variables mentioned earlier.

Let $\mathcal{F}_r$ be the σ–algebra generated by $d_{k,i}$ and $Z_{l,k}$ for all i, k, l such that $i \leqslant k \leqslant r$, $l = 1, 2$. Equivalently, $\mathcal{F}_r$ is generated by e_{ij} for $i < j \leqslant r$, $\sum_{j>r} e_{ij}$ for $i \leqslant r$, and $Z_{l,i}$ for $i \leqslant r$, $l = 1, 2$. Note that for $i \leqslant r$, D_i, τ_i, X_i, Y_i, $Z_{l,i}$ and d_i are all $\mathcal{F}_r$-measurable.

As mentioned earlier, we need to introduce a cut-off for r too close to 0 or n. Let $0 < \delta_c < 1/2$ and

$$\begin{aligned}
r_0 &= \lceil \delta_c n \rceil,\\
r_1 &= \lfloor (1-\delta_c) n \rfloor = n - r_0.
\end{aligned}$$

We shall estimate the probability that the degree bounds are satisfied on the vertices from $r_0 + 1$ to r_1 and use a continuity argument to show that when $\delta_c \to 0$ we obtain the correct result.

We shall make use of the following three alternative parameters to r:

$$\begin{aligned}
s &= n - r,\\
\alpha &= r/n,\\
\beta &= s/n.
\end{aligned}$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

These are the events that the upper bounds on the degrees are satisfied:

$$\begin{aligned} U_r &= \{d_r \leqslant b\sqrt{n}\}, \\ U(\delta) &= \bigcap_{\lceil \delta n \rceil < r \leqslant \lfloor (1-\delta)n \rfloor} U_r, \\ U &= U(\delta_{\mathrm{c}}) = \bigcap_{r_0 < r \leqslant r_1} U_r. \end{aligned}$$

Throughout this paper, b, p and, except right at the end of the proof, δ_{c} will be regarded as 'fixed', so no dependence on them will be shown and anything that only depends on them will be thought of as being constant.

We introduce δ_{m} (the 'mesh'), $0 < \delta_{\mathrm{m}} < 1$, which specifies how fine the discrete process D_i' is. It is also used to control the maximum contribution of the large swings in D_i (see Lemma 6.2 below). We shall need another controlling variable $\delta_{\mathrm{s}} > 0$, the 'sub-mesh', to make a two-dimensional version of the form of the central limit theorem we need to use.

In the end the errors that are introduced (prefixed by ϵ) will tend to 0 according to the following mode of convergence:

$$n \to \infty, \text{ then } \delta_{\mathrm{s}} \to 0, \text{ then } \delta_{\mathrm{m}} \to 0. \tag{4.1}$$

5. The continuous result

This first lemma gives the result in the continuous case and so motivates the subsequent approximation arguments.

Lemma 5.1. *Let c and θ_0 be the constants defined by* (2.1) *and* (2.2). *Then*

$$\mathbb{P}_{\mathrm{c}}(d^{\max} \leqslant b\sqrt{n}) \sim K c^n \quad \text{as} \quad n \to \infty,$$

where K is the constant

$$e^{\theta_0(b+\theta_0)}(1 + \theta_0(b + 2\theta_0))^{-1/2}.$$

Proof. In the continuous case the d_i are multivariate normal, so their distribution is determined by their covariance matrix. Now $\mathrm{Var}[d_i] = n-1$ and $\mathrm{Cov}[d_i, d_j] = 1$ if $i \neq j$, so the d_i have the same joint distribution as $Z_i - Z$ where $Z_i \sim N(0, n-2)$ and $Z \sim N(0,1)$ are independent. Conditioning on Z, we see that the desired probability is given by

$$\int_{-\infty}^{\infty} \phi(x)\Phi((b\sqrt{n} + x)/\sqrt{n-2})^n \, \mathrm{d}x = \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{-nf(y)} \, \mathrm{d}y,$$

substituting $x = y\sqrt{n}$, where

$$f(y) = y^2/2 - \log \Phi\left((b + y)\sqrt{\frac{n}{n-2}}\right).$$

Defining $R(x) = \phi(x)/\Phi(x)$, we have that $R'(x) = -(x + R(x))R(x)$. If $x \geqslant 0$ then $R'(x) \leqslant 0$. If $x < 0$ then integrating the inequality $y\phi(y)/x \geqslant \phi(y)$ with respect to y from $-\infty$ to x

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

gives that $R(x)+x \geqslant 0$, and hence that $R'(x) \leqslant 0$ for all x. Now

$$f'(y) = y - R\left((b+y)\sqrt{\frac{n}{n-2}}\right)\sqrt{\frac{n}{n-2}}$$

and

$$f''(y) = 1 - R'\left((b+y)\sqrt{\frac{n}{n-2}}\right)\frac{n}{n-2},$$

so $f''(y) \geqslant 1$. Therefore f has a unique minimum, y_0. We now apply a standard method for asymptotic expansion: that of expanding the exponent about its stationary points (see, *e.g.*, [4], VII, §6). In our case y_0 is not quite fixed but is equal to $\theta_0 + O(n^{-1})$ (since $f''(y) \geqslant 1$), but this does not matter. As $f''(y) \geqslant 1$, the integral dies away sufficiently rapidly outside a small interval (of width $n^{-2/5}$, say) around y_0, and within that interval the integral is asymptotic to

$$\begin{aligned}\sqrt{\frac{n}{2\pi}}\int_{-\infty}^{\infty} e^{-n(f(y_0)+f''(y_0)(y-y_0)^2/2)}\,\mathrm{d}y & \\ = \quad f''(y_0)^{-1/2}e^{-nf(y_0)} & \\ \sim \quad f''(\theta_0)^{-1/2}e^{-nf(\theta_0)}, \quad \text{since } f(\theta_0) = f(y_0) + O(n^{-2}). &\end{aligned}$$

Since $f''(\theta_0) \sim 1 - R'(b+\theta_0) = 1 + \theta_0(b+2\theta_0)$ and $f(\theta_0) = \theta_0^2/2 - \log\Phi((b+\theta_0)(1+n^{-1}+O(n^{-2}))) = \theta_0^2/2 - \log\Phi(b+\theta_0) - \theta_0(b+\theta_0)/n + O(n^{-2})$, the result follows. □

In fact for technical reasons we shall actually require (a crude form of) a slight variant of the above.

Lemma 5.1′. $$\lim_{\delta\to 0}\lim_{n\to\infty} \mathbb{P}_c(U(\delta))^{1/n} = c.$$

Proof. Similarly to the preceding proof, we need to estimate

$$\int_{-\infty}^{\infty} \phi(x)\Phi((b\sqrt{n}+x)/\sqrt{n-2})^{n-2\lceil\delta n\rceil}\,\mathrm{d}x.$$

This is monotonic in δ, so for the purposes of taking the limit as $\delta \to 0$ we may replace $n - 2\lceil\delta n\rceil$ by $(1-\delta)n$. The change from the preceding proof is that

$$f(y) = y^2/2 - (1-\delta)\log\Phi\left((b+y)\sqrt{\frac{n}{n-2}}\right).$$

As before we have $f''(y) \geqslant 1$, which ensures that y_0, the unique minimum of f, depends continuously on δ: in fact that $y_0 = \theta_0 + O(\delta + n^{-1})$. The limit we require is

$$\lim_{\delta\to 0}\lim_{n\to\infty} e^{-f(y_0)},$$

which is equal to c. □

6. The bad events

As mentioned in Section 3, we need to introduce bad sets on which we do not attempt to estimate the random variables. We shall write B_r for the event 'something has gone bad

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

by time r'. The definition of B_r depends on a constant C to be chosen later. In bounding swings we shall use the function $M = M(\delta_{\mathrm{m}}) = C/\delta_{\mathrm{m}}$. We consider four types of bad event, so $B_r = B_r^1 \cup B_r^2 \cup B_r^3 \cup B_r^4$, where

$$\begin{aligned}
B_r^1 &= \left\{ \sum_{i=r_0+1}^{r} (|X_i| + |Y_i|) > Cn \right\}, \\
B_r^2 &= \left\{ \sum_{i=r_0+1}^{r} (|X_i| + |Y_i|) I(|X_i| > M(\delta_{\mathrm{m}}) \text{ or } |Y_i| > M(\delta_{\mathrm{m}})) > \delta_{\mathrm{m}} n \right\}, \\
B_r^3 &= \left\{ (\exists k : r_0 \leqslant k \leqslant r) \sum_{i=1}^{k} d_{k,i}^2 > Cn^2 \right\}, \\
B_r^4 &= \left\{ (\exists k : r_0 \leqslant k \leqslant r) \, |\tau_k| > C \right\}.
\end{aligned}$$

Note that $B_r \subset B_{r+1}$ and $B_r \in \mathcal{F}_r$.

For the statement of the following four lemmas, c_0 is some positive constant less than $c(b)$ chosen once and for all, and $\mathbb{P} = \mathbb{P}_{\mathrm{c}}$ or $\mathbb{P}_{\mathrm{d}}$ (*i.e.*, true in *both* cases).

Lemma 6.1. *For all sufficiently large C, $\mathbb{P}(B_{r_1}^1) < c_0^n$ for all n.*

Lemma 6.2. *For all sufficiently large C, if $M(\delta_{\mathrm{m}}) = C/\delta_{\mathrm{m}}$ then $\mathbb{P}(B_{r_1}^2) < c_0^n$ for all n.*

Lemma 6.3. *For all sufficiently large C, $\mathbb{P}(B_{r_1}^3) < c_0^n$ for all n.*

Lemma 6.4. *For all sufficiently large C, $\mathbb{P}(B_{r_1}^4) < c_0^n$ for all n.*

To prove the four lemmas, we make use of the following sublemma.

Lemma 6.5. *Let Z_m denote $m^{-1/2}$ multiplied by the sum of m independent copies of the edge-weighting random variable $e_{1,2}$. Thus $Z_m \sim (B(m,p) - mp)/\sqrt{mpq}$ in the discrete case and $Z_m \sim N(0,1)$ in the continuous case. Then*

$$\mathbb{P}(|Z_m| \geqslant x) \leqslant 2e^{-2pqx^2} \quad (x \geqslant 0), \tag{6.1}$$

$$\mathbb{E}[e^{\theta|Z_m|}] \leqslant 2e^{\theta^2/(8pq)}. \tag{6.2}$$

Proof. In the continuous case (6.1) follows from the fact that $1-\Phi(x) \leqslant e^{-x^2/2}/(x\sqrt{2\pi}) \leqslant e^{-x^2/2}$ for $x \geqslant 1/\sqrt{2\pi}$, and $e^{-x^2/2} \geqslant \frac{1}{2}$ for $0 \leqslant x \leqslant 1/\sqrt{2\pi}$. Also (6.2) is clear since $\mathbb{E}[e^{\theta|Z_m|}] \leqslant \mathbb{E}[e^{\theta Z_m} + e^{-\theta Z_m}] = 2e^{\theta^2/2}$. In the discrete case we use a standard large deviation-type argument. Let $\mu(\theta) = \mathbb{E}[e^{\theta Z_m}] = (pe^{\theta q/\sqrt{mpq}} + qe^{-\theta p/\sqrt{mpq}})^m$. Markov's inequality implies that $\mathbb{P}(Z_m \geqslant x) \leqslant \inf_{\theta \geqslant 0}[e^{-\theta x}\mu(\theta)]$. For $x \geqslant 0$, $\inf_{\theta \geqslant 0}[e^{-\theta x}\mu(\theta)] = \inf_{\theta}[e^{-\theta x}\mu(\theta)]$, and for all x, $\inf_{\theta}[e^{-\theta x}\mu(\theta)] = \exp(-mH(x\sqrt{pq/m}))$, where $H(x) = (p+x)\log(1+x/p) + (q-x)\log(1-x/q)$. Differentiating twice shows that $H(x) \geqslant 2x^2$, establishing (6.1). To prove (6.2) note that $\log \mu()$, being the logarithm of a moment generating function, is convex, so we can

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

re-use the inequality $\inf_\theta[e^{-\theta x}\mu(\theta)] \leqslant e^{-2pqx^2}$ by writing $\mu(\theta) = \sup_x[e^{\theta x}\inf_{\theta'}[e^{-\theta' x}\mu(\theta')]] \leqslant \sup_x[e^{\theta x - 2pqx^2}] = e^{\theta^2/(8pq)}$. □

Proof of Lemma 6.1. We have

$$\begin{aligned}
\mathbb{P}(B^1_{r_1}) &= \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}(|X_i|+|Y_i|) > Cn\right)\\
&\leqslant \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}|X_i| > \frac{1}{2}Cn\right) + \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}|Y_i| > \frac{1}{2}Cn\right)\\
&= 2\mathbb{P}\left(\sum_{i=r_0+1}^{r_1}|X_i| > \frac{1}{2}Cn\right)\\
&\leqslant 2\exp\left(-\frac{1}{2}Cn\theta\right)\mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\theta|X_i|\right)\right], \quad \text{for any } \theta.
\end{aligned}$$

The X_i are independent with distribution $((i-1)/n)^{1/2}Z_{i-1}$ as in Lemma 6.5. Thus

$$\mathbb{E}[\exp(\theta|X_i|)] \leqslant 2\exp(\theta^2(i-1)/(8npq)) \leqslant 2\exp(\theta^2/(8pq))$$

and

$$\mathbb{P}(B^1_{r_1}) \leqslant \exp\left((1+1/n)\log 2 - \frac{1}{2}C\theta + \theta^2/(8pq)\right)^n,$$

and, choosing $\theta = 2Cpq$,

$$\mathbb{P}(B^1_{r_1}) \leqslant \exp(2\log 2 - C^2pq/2)^n.$$

The quantity inside the exponential can be made as small as desired by choosing C sufficiently large. □

Proof of Lemma 6.2. Writing M for $M(\delta_{\mathrm{m}}) = C/\delta_{\mathrm{m}}$, we have

$$\begin{aligned}
&(|X_i|+|Y_i|)I(|X_i| > M \text{ or } |Y_i| > M)\\
&\quad\leqslant (|X_i|+|Y_i|)(I(|X_i| > M) + I(|Y_i| > M))\\
&\quad\leqslant 2|X_i|I(|X_i| > M) + 2|Y_i|I(|Y_i| > M).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{P}(B^2_{r_1}) &= \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}(|X_i|+|Y_i|)I(|X_i| > M \text{ or } |Y_i| > M) > \delta_{\mathrm{m}}n\right)\\
&\leqslant \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}2|X_i|I(|X_i| > M) > \frac{\delta_{\mathrm{m}}n}{2}\right) + \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}2|Y_i|I(|Y_i| > M) > \frac{\delta_{\mathrm{m}}n}{2}\right)\\
&= 2\mathbb{P}\left(\sum_{i=r_0+1}^{r_1}|X_i|I(|X_i| > M) > \frac{\delta_{\mathrm{m}}n}{4}\right).
\end{aligned}$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

Let $(a_k)_{k=1}^{\infty}$ be a sequence of positive numbers whose sum is 1. Then

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=r_0+1}^{r_1}|X_i|I(|X_i|>M)>\frac{\delta_{\mathrm{m}}n}{4}\right)\\
&\leqslant \mathbb{P}\left(\sum_{i=r_0+1}^{r_1}\sum_{k=1}^{\infty}2MI(|X_i|>kM)>\frac{\delta_{\mathrm{m}}n}{4}\sum_{k=1}^{\infty}a_k\right)\\
&\leqslant \mathbb{P}\left((\exists k)\sum_{i=r_0+1}^{r_1}2MI(|X_i|>kM)>\frac{\delta_{\mathrm{m}}na_k}{4}\right)\\
&\leqslant \sum_{k=1}^{\infty}\mathbb{P}\left(\sum_{i=r_0+1}^{r_1}I(|X_i|>kM)>\frac{\delta_{\mathrm{m}}na_k}{8M}\right).
\end{aligned}$$

Using (6.1), we see that in the continuous and discrete cases

$$\begin{aligned}
\mathbb{P}(|X_i|>kM) &\leqslant 2\exp(-2k^2M^2pqn/(i-1))\\
&\leqslant 2\exp(-2k^2M^2pq)=p_1, \quad \text{say,}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=r_0+1}^{r_1}I(|X_i|>kM)>\frac{\delta_{\mathrm{m}}na_k}{8M}\right) &\leqslant 2^n p_1^{\delta_{\mathrm{m}}na_k/(8M)}\\
&= 2^{(1+\delta_{\mathrm{m}}a_k/(8M))n}\exp\bigl(-nk^2M\delta_{\mathrm{m}}a_kpq/4\bigr).
\end{aligned}$$

Choosing a_k proportional to $k^{-3/2}$, this becomes (for some positive constants A, B, C'),

$$\begin{aligned}
&\exp\Bigl((A+B\delta_{\mathrm{m}}k^{-3/2}/M-C'M\delta_{\mathrm{m}}\sqrt{k})n\Bigr)\\
&\leqslant \exp\Bigl((A+B/(M\delta_{\mathrm{m}})-C'M\delta_{\mathrm{m}}\sqrt{k})n\Bigr), \quad \text{since } \delta_{\mathrm{m}}<1,\\
&= \exp\Bigl((A+B/C-C'C\sqrt{k})n\Bigr), \quad \text{as } C=M\delta_{\mathrm{m}}.
\end{aligned}$$

If $D\geqslant 1$ then $\sum_{k=1}^{\infty}e^{-D\sqrt{k}}\leqslant 5e^{-D}$. Thus, taking C sufficiently large,

$$\begin{aligned}
\mathbb{P}(B_{r_1}^2) &\leqslant \sum_{k=1}^{\infty}\exp\Bigl((A+B/C-C'C\sqrt{k})n\Bigr)\\
&\leqslant 5\exp(A+B/C-C'C)^n\\
&< c_0^n,
\end{aligned}$$

completing the proof of the lemma. □

Proof of Lemma 6.3. For a given k, the $d_{k,i}$ $(1\leqslant i\leqslant k)$ are independent and $d_{k,i}$ has the distribution $(n-k)^{1/2}Z_{n-k}$ as in Lemma 6.5. Thus, for any $\theta\geqslant 0$,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{k}d_{k,i}^2>x\right) &\leqslant e^{-\theta x}\prod_{i=1}^{k}\mathbb{E}[e^{\theta d_{k,i}^2}]\\
&= e^{-\theta x}\prod_{i=1}^{k}\left(1+\int_0^{\infty}\theta e^{\theta y}\mathbb{P}(d_{k,i}^2\geqslant y)\,\mathrm{d}y\right)
\end{aligned}$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

$$\begin{aligned} &\leqslant e^{-\theta x}\left(1+\int_0^\infty 2\theta\exp(\theta y-2pqy/n)\,\mathrm{d}y\right)^k, \quad \text{using (6.1)},\\ &= e^{-\theta x}\left(\frac{1+\theta n/(2pq)}{1-\theta n/(2pq)}\right)^k, \quad \text{if } \theta n/(2pq)<1. \end{aligned}$$

Choosing $\theta = pq/n$,

$$\begin{aligned} \mathbb{P}(B^3_{r_1}) &\leqslant n\exp(-Cpqn)3^n\\ &\leqslant \exp(\log 3+\log n/n-Cpq)^n, \end{aligned}$$

and the lemma follows. □

Proof of Lemma 6.4. Take $k \leqslant r_1$. Since τ_k has the distribution $\sqrt{k/(n(n-k))}Z_{k(n-k)}$ as in Lemma 6.5,

$$\begin{aligned} \mathbb{P}(|\tau_k|>x) &\leqslant 2\exp(-2pqx^2n(n-k)/k)\\ &\leqslant 2\exp(-2pqx^2n\delta_{\mathrm{c}}). \end{aligned}$$

Thus

$$\mathbb{P}(B^4_{r_1}) \leqslant 2n\exp(-2pqC^2n\delta_{\mathrm{c}}),$$

and the lemma follows. □

For the rest of the paper we shall assume that C is chosen large enough to meet the requirements of all four lemmas. We shall write M for $M(\delta_{\mathrm{m}}) = C/\delta_{\mathrm{m}}$ for brevity.

7. The process (D'_r)

As discussed earlier, we shall work with (D'_r), a discrete approximation to (D_r) with the large swings removed. Each D'_r will take values in $\delta_{\mathrm{m}}\sqrt{n}\mathbb{Z}$.

We see that D_r can be written in terms of X_i, Y_i by

$$D_r = \sqrt{n}\sum_{i\leqslant r}(Y_i - X_i),$$

and we shall define D'_r in an analogous way.

Let $D'_{r_0} = \lfloor D_{r_0}/(\delta_{\mathrm{m}}\sqrt{n})\rfloor\delta_{\mathrm{m}}\sqrt{n}$, and for $r \geqslant r_0$ define inductively

$$\begin{aligned} \tau'_r &= (s\sqrt{n})^{-1}D'_r,\\ X_{I,r+1} &= \tau'_r + \sqrt{\alpha}Z_{1,r+1} \quad (\text{so } X_{I,r+1}\sim N(\tau'_r,\alpha)),\\ Y_{I,r+1} &= \sqrt{\beta}Z_{2,r+1} \quad (\text{so } Y_{I,r+1}\sim N(0,\beta)),\\ X''_{r+1} &= X_{r+1}I(B^{\mathrm{c}}_r) + X_{I,r+1}I(B_r),\\ Y''_{r+1} &= Y_{r+1}I(B^{\mathrm{c}}_r) + Y_{I,r+1}I(B_r),\\ D''_{r+1} &= D'_r + \lfloor (Y''_{r+1}-X''_{r+1})/\delta_{\mathrm{m}}\rfloor\delta_{\mathrm{m}}\sqrt{n},\\ E_{r+1} &= \{|D''_{r+1}-D'_r|/\sqrt{n}\leqslant 2M+\delta_{\mathrm{m}},\ \ |X''_{r+1}+Y''_{r+1}|\leqslant 2M\}\\ &= \{|\lfloor (Y''_{r+1}-X''_{r+1})/\delta_{\mathrm{m}}\rfloor|\delta_{\mathrm{m}}\leqslant 2M+\delta_{\mathrm{m}},\ \ |X''_{r+1}+Y''_{r+1}|\leqslant 2M\},\\ X'_{r+1} &= X''_{r+1}I(E_{r+1}), \end{aligned}$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

$$\begin{aligned} Y'_{r+1} &= Y''_{r+1} I(E_{r+1}), \\ D'_{r+1} &= D'_r + \lfloor (Y'_{r+1} - X'_{r+1})/\delta_{\mathrm{m}} \rfloor \delta_{\mathrm{m}} \sqrt{n} \\ &= D''_{r+1} I(E_{r+1}) + D'_r I(E^{\mathrm{c}}_{r+1}). \end{aligned}$$

These definitions are somewhat cumbersome partly because of the need to make a 'clean' cut-off in terms of the discretized variables X'_r and Y'_r. That is why we first define the double primed quantities, which are what the variables would be if there were no cut-off, and then define the primed quantities in terms of these.

The definitions of E_r and B_r^2 mean that the difference $|\tau_r - \tau'_r|$ is controlled as follows.

Outside E_r either $|X''_r - Y''_r| > 2M$ or $|X''_r + Y''_r| > 2M$, which implies that $|X''_r| > M$ or $|Y''_r| > M$. Thus, outside B_r (not necessarily outside E_r any more), if $r_0 \leqslant r \leqslant r_1$,

$$\begin{aligned} |D_r - D'_r| &= \left| D_{r_0} - D'_{r_0} + \sqrt{n} \sum_{i=r_0+1}^{r} (Y_i - X_i - \lfloor (Y'_i - X'_i)/\delta_{\mathrm{m}} \rfloor \delta_{\mathrm{m}}) \right| \\ &\leqslant \delta_{\mathrm{m}} \sqrt{n} + \sqrt{n} \sum_{i=r_0+1}^{r} \left(|Y_i - X_i - (Y'_i - X'_i)| + \delta_{\mathrm{m}} \right) \\ &= (1 + r - r_0) \delta_{\mathrm{m}} \sqrt{n} + \sqrt{n} \sum_{i=r_0+1}^{r} |Y_i - X_i| I(E^{\mathrm{c}}_i) \\ &\leqslant n^{3/2} \delta_{\mathrm{m}} + \sqrt{n} \sum_{i=r_0+1}^{r} |Y_i - X_i| I(|X_i| > M \text{ or } |Y_i| > M) \\ &\leqslant n^{3/2} (\delta_{\mathrm{m}} + \delta_{\mathrm{m}}), \quad \text{using } B_r^2 \subset B_r. \end{aligned}$$

So, outside B_r, if $r_0 \leqslant r \leqslant r_1$ then

$$|\tau_r - \tau'_r| \leqslant 2\delta_{\mathrm{m}}/\delta_{\mathrm{c}}. \tag{7.1}$$

By analogy with $d_r = \sqrt{n}(X_r + Y_r)$, let $d''_r = \sqrt{n}(X''_r + Y''_r)$. For $r < n$ and $m \in \mathbb{Z}$ define the event $S_{r,m}$ by

$$\begin{aligned} S_{r,m} &= \{ D'_r = m\delta_{\mathrm{m}} \sqrt{n} \text{ and } (\forall i : r_0 < i \leqslant r)\, d''_i \leqslant b\sqrt{n} \} \\ &= \{ \tau'_r = m\delta_{\mathrm{m}}/s \text{ and } (\forall i : r_0 < i \leqslant r)\, d''_i \leqslant b\sqrt{n} \}. \end{aligned} \tag{7.2}$$

Lemma 7.1. *Let $X \sim (B(N,p) - Np)/\sqrt{Npq}$, $Y \sim N(0,1)$, and let A, δ be positive constants. There exists a function $G = G(p, A, \delta, N)$ such that $G = O(N^{1/4})$ and for all a satisfying $|a| \leqslant AN^{1/4}$, δ' satisfying $\delta/2 \leqslant \delta' \leqslant 2\delta$, and $N \geqslant 16/(\delta^4 p^2 q^2)$ we have*

$$e^{-G} \leqslant \mathbb{P}(X \in [a, a + \delta' N^{-1/4}))/\mathbb{P}(Y \in [a, a + \delta' N^{-1/4})) \leqslant e^G.$$

Proof. $\mathbb{P}(X \in [a, a + \delta' N^{-1/4}))$ is given by the sum of

$$\binom{N}{i} p^i q^{N-i},$$

where i ranges from $i_0 = \lceil Np + a\sqrt{Npq} \rceil$ to $i_1 = \lceil Np + (a + \delta' N^{-1/4})\sqrt{Npq} - 1 \rceil$. However,

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

if we compare a summand at i with one at $i+j>i$, we see that the ratio is given by

$$R_{ij}=\frac{(i+j)\dots(i+1)}{(N-i)\dots(N-i-j+1)}\left(\frac{q}{p}\right)^j.$$

Since $|i-Np|\leqslant(AN^{3/4}+2\delta N^{1/4})\sqrt{pq}$, it follows that R_{ij} is between $(1-BN^{-1/4})^j$ and $(1+BN^{-1/4})^j$ for some constant $B=B(p,A,\delta)$. We also have $j\leqslant 2\delta N^{1/4}\sqrt{pq}$, so R_{ij} is bounded above and below, that is, $1/D\leqslant R_{ij}\leqslant D$ for some $D=D(p,A,\delta)$. Therefore, at the expense of a factor of $D2\delta N^{1/4}\sqrt{pq}$, which is much finer than the final estimate we require, we may replace the sum by its first term. Stirling's formula tells us that the first term is asymptotic (uniformly over allowable a) to

$$\frac{1}{\sqrt{2\pi pqN}}e^{-H(x_0)N},$$

where $H(x)=(p+x)\log(1+x/p)+(q-x)\log(1-x/q)$ and $x_0=i_0/N-p$. Now $H(x)=x^2/(2pq)+O(x^3)$ and $|x_0-aN^{-1/2}\sqrt{pq}|\leqslant N^{-1}$, so $|x_0|\leqslant AN^{-1/4}+N^{-1}$ and $|H(x_0)N-a^2/2|\leqslant EN^{1/4}$ for some $E=E(p,A)$.

We turn to $\mathbb{P}(Y\in[a,a+\delta' N^{-1/4}))$ which is given by the integral of $\phi(x)$ from a to $a+\delta' N^{-1/4}$. If $x,x+y\in[a,a+\delta' N^{-1/4}]$, then $\phi(x)/\phi(x+y)$ is between $e^{-|xy|-y^2/2}$ and $e^{|xy|+y^2/2}$. Now $|xy|+y^2/2$ is bounded, so we may replace the integral by $\phi(a)$ at the expense of a factor of $(N^{-1/4}\delta/2)^{-1}$ multiplied by the maximum value of $e^{|xy|+y^2/2}$.

Combining the above estimates, we see that G can be chosen to be $EN^{1/4}+F\log N$ for some $F=F(p,A,\delta)$. □

We now arrive at the main lemma.

Lemma 7.2. *There exist ϵ_0, ϵ_1 depending only on n, δ_{m}, δ_{s} such that $\epsilon_0,\epsilon_1\to 0$ as* (4.1)*, and if $r_0\leqslant r<r_1$ and $|m|\leqslant Cn/\delta_{\mathrm{m}}$ then*

$$\begin{aligned}\sup_{S_{r,m}}\mathbb{P}_{\mathrm{d}}(S_{r+1,m+\Delta m}\mid\mathcal{F}_r)&\leqslant\inf_{S_{r,m}}e^{\epsilon_0+\epsilon_1\delta_{\mathrm{m}}|\Delta m|}\mathbb{P}_{\mathrm{c}}(S_{r+1,m+\Delta m}\mid\mathcal{F}_r),\\ \sup_{S_{r,m}}\mathbb{P}_{\mathrm{c}}(S_{r+1,m+\Delta m}\mid\mathcal{F}_r)&\leqslant\inf_{S_{r,m}}e^{\epsilon_0+\epsilon_1\delta_{\mathrm{m}}|\Delta m|}\mathbb{P}_{\mathrm{d}}(S_{r+1,m+\Delta m}\mid\mathcal{F}_r).\end{aligned}\tag{7.3}$$

(*Both probabilities in* (7.3) *will be zero if the increment Δm is outside a certain range.*)

Remark. In the above statement and throughout this paper the conditional probability $\mathbb{P}(.\mid\mathcal{F}_r)$ is being used to mean a particular 'good version' of something that would in general only be defined up to a.e. equivalence. This is explained in the proof of Lemma 7.2.

We postpone the proof of Lemma 7.2 and look at its consequences.

Lemma 7.3. *For $0<\delta_{\mathrm{c}}<\frac{1}{2}$,*

$$\left(\frac{\mathbb{P}_{\mathrm{d}}(U)}{\mathbb{P}_{\mathrm{c}}(U)}\right)^{1/n}\to 1\quad\text{as } n\to\infty.$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

Proof. For this proof, we assume that $\delta_{\mathrm{m}} \leqslant \delta_{\mathrm{c}}{}^2 C/2$. This ensures that for $r_0 \leqslant r \leqslant r_1$ we have $s(C/\delta_{\mathrm{m}} + 2/\delta_{\mathrm{c}}) \leqslant Cn/\delta_{\mathrm{m}}$, and makes the notation easier, because now, for $r_0 \leqslant r \leqslant r_1$, if $S_{r,m} \not\subset B_r$ then $|m| \leqslant Cn/\delta_{\mathrm{m}}$ using (7.1), (7.2) and the definition of B_r^4.

Let $\mathbf{m} = (m_{r_0}, m_{r_0+1}, \dots, m_{r_1})$ be a sequence of integers.

Let $T_{r,\mathbf{m}} = S_{r_0,m_{r_0}} \cap S_{r_0+1,m_{r_0+1}} \cap \ldots \cap S_{r,m_r}$. Note that $T_{r,\mathbf{m}} \in \mathcal{F}_r$, and for a given r the $T_{r,\mathbf{m}}$ provide a partition of $\{(\forall i : r_0 < i \leqslant r)\, d_i'' \leqslant b\sqrt{n}\} \subset \Omega$.

If $(\forall r)\, |m_r| \leqslant Cn/\delta_{\mathrm{m}}$, we have

$$\begin{aligned}\mathbb{P}_{\mathrm{c}}(T_{r+1,\mathbf{m}} \mid S_{r_0,m_{r_0}}) &= \mathbb{P}_{\mathrm{c}}(T_{r,\mathbf{m}} \mid S_{r_0,m_{r_0}})\mathbb{P}_{\mathrm{c}}(T_{r+1,\mathbf{m}} \mid T_{r,\mathbf{m}}) \\ &\leqslant \mathbb{P}_{\mathrm{c}}(T_{r,\mathbf{m}} \mid S_{r_0,m_{r_0}})e^{\epsilon_0+\epsilon_1\delta_{\mathrm{m}}|m_{r+1}-m_r|}\mathbb{P}_{\mathrm{d}}(T_{r+1,\mathbf{m}} \mid T_{r,\mathbf{m}}),\end{aligned}$$

by Lemma 7.2, and similarly with $\mathbb{P}_{\mathrm{c}}$ and $\mathbb{P}_{\mathrm{d}}$ exchanged. So, by induction,

$$\begin{aligned}\mathbb{P}_{\mathrm{d}}(T_{r_1,\mathbf{m}} \mid S_{r_0,m_{r_0}}) &\leqslant e^{B(\mathbf{m})}\mathbb{P}_{\mathrm{c}}(T_{r_1,\mathbf{m}} \mid S_{r_0,m_{r_0}}), \\ \mathbb{P}_{\mathrm{c}}(T_{r_1,\mathbf{m}} \mid S_{r_0,m_{r_0}}) &\leqslant e^{B(\mathbf{m})}\mathbb{P}_{\mathrm{d}}(T_{r_1,\mathbf{m}} \mid S_{r_0,m_{r_0}}),\end{aligned} \tag{7.4}$$

where

$$B(\mathbf{m}) = \epsilon_0(r_1 - r_0) + \epsilon_1\delta_{\mathrm{m}} \sum_{r=r_0}^{r_1-1} |m_{r+1} - m_r|.$$

Now $S_{r_0,m} = \{m\delta_{\mathrm{m}}\sqrt{n} \leqslant D_{r_0} < (m+1)\delta_{\mathrm{m}}\sqrt{n}\}$. We seek to prove that $\mathbb{P}_{\mathrm{c}}(S_{r_0,m})$ and $\mathbb{P}_{\mathrm{d}}(S_{r_0,m})$ are similar. If $N = r_0 r_1$ then $D_{r_0}/\sqrt{N}$ has the distribution Z_N in the sense of Lemma 6.5 and so can act as X or Y in Lemma 7.1. The condition $\omega \in S_{r_0,m}$ corresponds to $D_{r_0}/\sqrt{N} \in [m\delta_{\mathrm{m}}\sqrt{n/N}, (m+1)\delta_{\mathrm{m}}\sqrt{n/N})$, so to apply Lemma 7.1 we need to know the maximum value of $|m|\delta_{\mathrm{m}}\sqrt{n/N}N^{-1/4}$. If $|m| \leqslant Cn/\delta_{\mathrm{m}}$ then $|m|\delta_{\mathrm{m}}\sqrt{n/N}N^{-1/4} \leqslant C(n^2/N)^{3/4} \leqslant C(\delta_{\mathrm{c}}(1-\delta_{\mathrm{c}}))^{-3/4} = A$, say. The length of the above interval is $\delta_{\mathrm{m}}\sqrt{n/N}$ and $\lim_{n\to\infty} N^{1/4}\delta_{\mathrm{m}}\sqrt{n/N} = \delta_{\mathrm{m}}(\delta_{\mathrm{c}}(1-\delta_{\mathrm{c}}))^{-1/4} = \delta$, say. Therefore Lemma 7.1 may be applied to $D_{r_0}/\sqrt{N}$ with the above N, A and δ. The outcome is that for $|m| \leqslant Cn/\delta_{\mathrm{m}}$,

$$\mathbb{P}_{\mathrm{c}}(S_{r_0,m})/\mathbb{P}_{\mathrm{d}}(S_{r_0,m}) = e^{o(n)}, \tag{7.5}$$

where the $o(n)$ does not depend on m.

Let $m_r = D_r'/(\delta_{\mathrm{m}}\sqrt{n})$, so that if $\omega \in S_{r,m}$ for some m then $m = m_r(\omega)$. Then $|m_{r+1} - m_r| = |\lfloor (Y_{r+1}' - X_{r+1}')/\delta_{\mathrm{m}} \rfloor| \leqslant |\lfloor (Y_{r+1}'' - X_{r+1}'')/\delta_{\mathrm{m}} \rfloor|$, so $|m_{r+1} - m_r|\delta_{\mathrm{m}} \leqslant |Y_{r+1}''| + |X_{r+1}''| + \delta_{\mathrm{m}}$. On $B_{r_1}^{\mathrm{c}}$, $X_i = X_i''$ and $Y_i = Y_i''$ for $i \leqslant r_1$. Thus, on $B_{r_1}^{\mathrm{c}}$,

$$\begin{aligned}|m_{r_0}| &\leqslant Cn/\delta_{\mathrm{m}}, \quad \text{and} \\ B(\mathbf{m}) &\leqslant \epsilon_0(r_1 - r_0) + \epsilon_1 \sum_{r=r_0}^{r_1-1} \left(|Y_{r+1}| + |X_{r+1}| + \delta_{\mathrm{m}}\right) \\ &\leqslant \epsilon_0(r_1 - r_0) + \epsilon_1(C + \delta_{\mathrm{m}})n, \quad \text{using } B_{r_1}^1 \subset B_{r_1} \\ &\leqslant (\epsilon_0 + \epsilon_1(C + \delta_{\mathrm{m}}))n.\end{aligned}$$

Regarding $\mathbf{m} = (m_{r_0}, m_{r_0+1}, \dots, m_{r_1})$ as any sequence of integers once more, let T be the union of $T_{r_1,\mathbf{m}}$ over all $\mathbf{m}$ satisfying $(\forall r)\, |m_r| \leqslant Cn/\delta_{\mathrm{m}}$ and $B(\mathbf{m}) \leqslant (\epsilon_0 + \epsilon_1(C + \delta_{\mathrm{m}}))n$. Informally, T is the union of all paths on which the d_i'' are at most $b\sqrt{n}$, where D_i' does not become too large, and on which D_i' does not wiggle too much as i varies. All other

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

paths $T_{r_1,\mathbf{m}}$ are contained in the bad event B_{r_1}, which has very small probability. Thus $U'' \setminus B_{r_1} \subset T$, where $U'' = \{(\forall i : r_0 < i \leqslant r_1)\, d''_i \leqslant b\sqrt{n}\}$ is the event that the modified degrees d''_i are small. We only modified the degrees on the bad event, that is, $d''_i = d_i$ on $B^{\mathrm{c}}_{r_1}$ for $i \leqslant r_1$. Thus $U \setminus B_{r_1} = U'' \setminus B_{r_1} \subset T$. Also $T \subset U''$, so $T \setminus B_{r_1} \subset U'' \setminus B_{r_1} = U \setminus B_{r_1}$. Therefore the symmetric difference $T \,\Delta\, U$ is contained in B_{r_1}, and Lemmas 6.1, 6.2, 6.3, and 6.4 imply that $\mathbb{P}_{\mathrm{c}}(T \,\Delta\, U) \leqslant 4c_0^n$ and $\mathbb{P}_{\mathrm{d}}(T \,\Delta\, U) \leqslant 4c_0^n$. Using (7.4) and (7.5), we have that $\mathbb{P}_{\mathrm{d}}(T) \leqslant \mathbb{P}_{\mathrm{c}}(T)e^E$ and $\mathbb{P}_{\mathrm{d}}(T) \geqslant \mathbb{P}_{\mathrm{c}}(T)e^{-E}$ where $E = (\epsilon_0 + \epsilon_1(C + \delta_{\mathrm{m}}))n + o(n)$. Thus

$$\begin{aligned}
\mathbb{P}_{\mathrm{d}}(U) &\leqslant (\mathbb{P}_{\mathrm{c}}(U) + \mathbb{P}_{\mathrm{c}}(T \setminus U))e^E + \mathbb{P}_{\mathrm{d}}(U \setminus T) \leqslant (\mathbb{P}_{\mathrm{c}}(U) + 4c_0^n)e^E + 4c_0^n,\\
\mathbb{P}_{\mathrm{d}}(U) &\geqslant (\mathbb{P}_{\mathrm{c}}(U) - \mathbb{P}_{\mathrm{c}}(U \setminus T))e^{-E} - \mathbb{P}_{\mathrm{d}}(T \setminus U) \geqslant (\mathbb{P}_{\mathrm{c}}(U) - 4c_0^n)e^{-E} - 4c_0^n.
\end{aligned}$$

Lemma 5.1′ implies that $\mathbb{P}_{\mathrm{c}}(U)^{1/n}$ converges to a limit c_1 as $n \to \infty$. Since $U(\delta)$ is decreasing in δ, we have that $c_1 \geqslant c > c_0$. Therefore for sufficiently small $\epsilon_0 + \epsilon_1(C + \delta_{\mathrm{m}})$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}_{\mathrm{d}}(U)^{1/n} &\leqslant c_1 \lim_{n\to\infty} e^{\epsilon_0+\epsilon_1(C+\delta_{\mathrm{m}})},\\
\liminf_{n\to\infty} \mathbb{P}_{\mathrm{d}}(U)^{1/n} &\geqslant c_1 \lim_{n\to\infty} e^{-(\epsilon_0+\epsilon_1(C+\delta_{\mathrm{m}}))}.
\end{aligned}$$

Letting $\delta_{\mathrm{s}} \to 0$ then $\delta_{\mathrm{m}} \to 0$ shows that $\mathbb{P}_{\mathrm{d}}(U)^{1/n} \to c_1$ as $n \to \infty$. Since c_1 was also the limit of $\mathbb{P}_{\mathrm{c}}(U)^{1/n}$ as $n \to \infty$, the lemma is proved. □

Theorem 2.1 now follows as an easy corollary.

Proof of Theorem 2.1. The events U_i and the event U are down-sets when considered as subsets of $\mathcal{G}(n,p)$. This implies that $p_n \geqslant \mathbb{P}_{\mathrm{d}}(U)\mathbb{P}_{\mathrm{d}}(U_1)^{2r_0}$. The central limit theorem tells us that $\mathbb{P}_{\mathrm{d}}(U_1) \to \Phi(b)$ as $n \to \infty$, and we also have the trivial bound $\mathbb{P}_{\mathrm{d}}(U) \geqslant p_n$. Taking nth roots, letting $n \to \infty$ then $\delta_{\mathrm{c}} \to 0$ and applying Lemmas 5.1′ and 7.3, we arrive at Theorem 2.1. □

8. Proof of main lemma

We aim to prove that on $S_{r,m}$,

$$\begin{gathered}
e^{-(\epsilon_0+\epsilon_1\delta_{\mathrm{m}}|\Delta m|)/2} P_I \leqslant \mathbb{P}(S_{r+1,m+\Delta m} \mid \mathcal{F}_r) \leqslant e^{(\epsilon_0+\epsilon_1\delta_{\mathrm{m}}|\Delta m|)/2} P_I,\\
\text{where } \epsilon_0 = \epsilon_0(n,\delta_{\mathrm{m}},\delta_{\mathrm{s}}),\ \epsilon_1 = \epsilon_1(n,\delta_{\mathrm{m}},\delta_{\mathrm{s}}) \text{ and } \epsilon_0,\epsilon_1 \to 0 \text{ as (4.1)},
\end{gathered} \tag{8.1}$$

for $\mathbb{P} = \mathbb{P}_{\mathrm{c}}$ and $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$, for some 'ideal' $P_I = P_I(n,r,m,\Delta m)$ which does not depend on $\omega \in S_{r,m}$. This will establish (7.3).

We shall divide into two cases in three different ways:

(i) $\omega \in S_{r,m} \setminus B_r$ ('good') and $\omega \in S_{r,m} \cap B_r$ ('bad'),
(ii) $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$ ('discrete') and $\mathbb{P} = \mathbb{P}_{\mathrm{c}}$ ('continuous'), and
(iii) $\Delta m \neq 0$ and $\Delta m = 0$.

To make it completely clear what combination of cases we are referring to, we shall mark each major change with a new subheading.

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

All cases

For brevity we shall define $X = X''_{r+1}$, $Y = Y''_{r+1}$. Thus, in the good case $X = X_{r+1}$, $Y = Y_{r+1}$, and in the bad case $X = X_{I,r+1}$, $Y = Y_{I,r+1}$.

Note that if $|\Delta m| > 2M/\delta_{\mathrm{m}}+1$ then $S_{r,m} \cap S_{r+1,m+\Delta m} \subset E^{\mathrm{c}}_{r+1}$, and so $\mathbb{P}(S_{r+1,m+\Delta m} \mid \mathcal{F}_r) = 0$ on $S_{r,m}$ both for $\mathbb{P} = \mathbb{P}_{\mathrm{c}}$ and for $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$. We shall thus assume throughout this section that $|\Delta m| \leqslant 2M/\delta_{\mathrm{m}} + 1$.

As mentioned in the remark following the statement of Lemma 7.2, we have to interpret $\mathbb{P}(. \mid \mathcal{F}_r)$ to be a particular version of the conditional probability. Given $S_{r,m}$, whether or not we are in $S_{r+1,m+\Delta m}$ is determined by X and Y. It will thus be convenient to speak of 'the distribution of X and Y given $\mathcal{F}_r$', for example, to make sense of Lemma 8.2. In other words, we would like $V \mapsto \mathbb{P}((X,Y) \in V \mid \mathcal{F}_r)$ to be a probability measure for (almost all) $\omega \in S_{r,m}$. It might seem there could be a technical difficulty with this, since for each V the conditional expectation $\mathbb{P}((X,Y) \in V \mid \mathcal{F}_r)$ is only defined up to a.e. equivalence, and there are in general uncountably many possible V. However, there is no problem here, as will be seen when we come to consider the three possibilities (a) good, discrete, (b) good, continuous, and (c) bad.

If $2M \geqslant b$ and $\Delta m \neq 0$, then on $S_{r,m}$ the conditional probability $\mathbb{P}(S_{r+1,m+\Delta m} \mid \mathcal{F}_r)$ is given by $\mathbb{P}((X,Y) \in R \mid \mathcal{F}_r)$, where

$$R = \{(x,y) : \Delta m \delta_{\mathrm{m}} \leqslant y - x < (\Delta m + 1)\delta_{\mathrm{m}},\ -2M \leqslant x + y \leqslant b\}.$$

The case $\Delta m = 0$ is special because X'_{r+1} and Y'_{r+1} are set to zero if they would otherwise become too large. Thus $S_{r+1,m}$ encompasses both the 'genuine' set $S_{r+1,m} \cap E_{r+1}$ (analogous to $S_{r+1,m+\Delta m}$ for $\Delta m \neq 0$) and the large deviation set $S_{r+1,m} \setminus E_{r+1}$. It would be nice to be able to treat E^{c}_{r+1} like the other bad sets, but recall from the remarks made in Section 3 that we have to allow our process to make some large jumps – we merely restrict their total contribution.

To prove Lemma 7.2 we shall thus proceed as follows.

First we shall approximate X and Y by normal random variables, if necessary. Unfortunately the mean of one of these variables depends on $\omega \in S_{r,m}$, which P_I is not allowed to do. We should like to show that one can approximate (X,Y) sufficiently well by a pair $(X^\star, Y^\star)$ of independent $N(\tau'_r, \alpha)$, $N(0,\beta)$ random variables so that $\mathbb{P}((X,Y) \in R \mid \mathcal{F}_r)$ is close to $\mathbb{P}((X^\star, Y^\star) \in R \mid \mathcal{F}_r)$. However, because of the requirement of uniformity in the limits involved, we cannot do this. Instead we shall estimate $\mathbb{P}((X,Y) \in R \mid \mathcal{F}_r)$ itself. It is thus convenient to take for P_I not the exact value of $\mathbb{P}((X^\star, Y^\star) \in R \mid \mathcal{F}_r)$, but rather a simplification of this.

Let $w' = \Delta m \delta_{\mathrm{m}}$, so w' and $\tau'_r = m\delta_{\mathrm{m}}/s$ are constant over $S_{r,m}$, and let

$$x = (b/2 - (\beta - \alpha)w'/2 - \beta\tau'_r)/\sqrt{\alpha\beta}.$$

We shall take $P_I = \delta_{\mathrm{m}} \Phi(x) \phi(w' + \tau'_r)$: this is the approximation to $\mathbb{P}((X^\star, Y^\star) \in R \mid \mathcal{F}_r)$ obtained by ignoring the condition $-2M \leqslant x + y$, and treating the width of R as small.

We start by approximating X and Y by suitable normal random variables. This is only needed in the good case, as in the bad case X and Y are *defined* to have the distribution we want. In the good continuous case X and Y are already normal and we just have to adjust the variances slightly; most of the work is in the good discrete case. For this

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

we quote XVI.5, Theorem 2 of [6], which is a version of a theorem originally due to Berry-Esséen – a central limit theorem with error bounds.

Theorem 8.1. *Let the W_k be independent variables such that*

$$\mathbb{E}[W_k] = 0, \quad \mathbb{E}[W_k^2] = \sigma_k^2, \quad \mathbb{E}[|W_k^3|] = \rho_k.$$

Put

$$s_r^2 = \sigma_1^2 + \cdots + \sigma_r^2, \quad r_r = \rho_1 + \cdots + \rho_r,$$

and let $Z_r = (W_1 + \cdots + W_r)/s_r$. Then, for all x and r,

$$|\mathbb{P}(Z_r \leqslant x) - \Phi(x)| \leqslant 6r_r/s_r^3.$$

Using this result we are ready to prove the following lemma.

Lemma 8.2. *There exist constants C_1 and C_2 such that, if $|\Delta m| \leqslant 2M/\delta_{\mathrm{m}} + 1$, then X and Y are independent given $\mathcal{F}_r$ and*

$$|\mathbb{P}((X - \tau_r)/\sqrt{\alpha} \leqslant x \mid \mathcal{F}_r) - \Phi(x)| \leqslant C_1/\sqrt{n} \tag{8.2}$$

and

$$|\mathbb{P}(Y/\sqrt{\beta} \leqslant x \mid \mathcal{F}_r) - \Phi(x)| \leqslant C_2/\sqrt{n} \tag{8.3}$$

hold on $S_{r,m} \cap B_r^{\mathrm{c}}$, for $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$ and for $\mathbb{P} = \mathbb{P}_{\mathrm{c}}$.

Proof. We consider the discrete and continuous cases separately.

The good, discrete case

The event B_r is in the sub-σ-algebra $\mathcal{F}_r'$ of $\mathcal{F}_r$ generated by the $d_{i,j}$ that contribute to generating $\mathcal{F}_r$ (*i.e.*, B_r does not depend on any $Z_{i,j}$). Since we are in the good case, X and Y also do not depend on any $Z_{i,j}$ and so conditioning X,Y on $\mathcal{F}_r$ is equivalent to conditioning on $\mathcal{F}_r'$. Now $\mathcal{F}_r'$ is generated by a countable partition so there is no problem defining a good version of the conditional probability in the good, discrete case.

Given $\mathcal{F}_r$ we know $d_{r,i}$. In the original, unrescaled language of 'number of edges', $d_{r,i}$ would be the total number of edges from vertex i to vertices greater than r. By symmetry, the probability, given $\mathcal{F}_r$, of an edge between vertex i and vertex $r+1$ would be $d_{r,i}/s$. Therefore, returning to the rescaled notation, if we define

$$\begin{aligned} p_i &= p + d_{r,i}\sqrt{pq}/s, \\ W_i &\sim \frac{1}{\sqrt{pq}}(B(1, p_i) - p_i), \end{aligned}$$

then

$$\tau_r = \frac{1}{\sqrt{npq}} \sum_{i=1}^{r} (p_i - p),$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

and the distributions of X and Y given $\mathcal{F}_r$ are independently

$$\begin{aligned} X &\sim \frac{1}{\sqrt{n}}\sum_{i=1}^{r} W_i + \tau_r, \\ Y &\sim \frac{1}{\sqrt{npq}}(B(s-1,p) - (s-1)p). \end{aligned}$$

Now

$$\begin{aligned} \mathbb{E}[W_i] &= 0, \\ \mathrm{Var}[W_i] &= (pq)^{-1}p_iq_i, \\ \mathbb{E}[|W_i|^3] &= (pq)^{-3/2}p_iq_i(p_i^2+q_i^2) \leqslant (pq)^{-3/2}/8 = C_3, \text{ say,} \end{aligned}$$

so

$$\mathrm{Var}[X \mid \mathcal{F}_r] = \frac{1}{n}s_r^2,$$

where

$$\begin{aligned} s_r^2 &= (pq)^{-1}\sum_{i=1}^{r} p_iq_i, \\ &= r + \frac{q-p}{\sqrt{pq}}\tau_r\sqrt{n} - \frac{1}{s^2}\sum_{i=1}^{r} d_{r,i}^2. \end{aligned}$$

Since we are outside B_r, we have $|\tau_r| \leqslant C$ and $\sum_{i=1}^{r} d_{r,i}^2 \leqslant Cn^2$, so

$$|s_r^2 - r| \leqslant C_4\sqrt{n}. \tag{8.4}$$

Applying Theorem 8.1 tells us that, for all x, on B_r^{c}

$$|\mathbb{P}_{\mathrm{d}}((X-\tau_r)\sqrt{n}/s_r \leqslant x \mid \mathcal{F}_r) - \Phi(x)| \leqslant 6C_3rs_r^{-3} \leqslant C_5/\sqrt{n}. \tag{8.5}$$

Now $x\phi(x)$ is bounded, so

$$|\Phi(x\sqrt{r}/s_r) - \Phi(x)| \leqslant C_6/\sqrt{n}. \tag{8.6}$$

Combining (8.4), (8.5) and (8.6) proves (8.2) for $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$.

Theorem 8.1 also gives (8.3) for $\mathbb{P} = \mathbb{P}_{\mathrm{d}}$.

The good, continuous case

Let $i \leqslant r$. Given $d_{r,i}$, it can be checked that $e_{i,r+1} \sim N(d_{r,i}/s, (s-1)/s)$, *i.e.*, that if $Z_i \sim N(d_{r,i}/s, (s-1)/s)$ then $\mathbb{P}(Z_i \in V)$ is a version of the conditional expectation $\mathbb{P}(e_{i,r+1} \in V \mid d_{r,i})$. Independence allows us to combine these facts over i. Since $e_{r+1,j}$ is independent of $\mathcal{F}_r$ for $j > r+1$ we have that, given $\mathcal{F}_r$, outside B_r, X and Y are independent and

$$\begin{aligned} X &\sim \frac{1}{\sqrt{n}}\sum_{i=1}^{r} Z_i = N(\tau_r, (1-1/s)\alpha), \\ Y &\sim \frac{1}{\sqrt{n}}\sum_{i=r+2}^{n} N(0,1) = N(0,(s-1)/n) = N(0, \beta - 1/n). \end{aligned}$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

Using the fact that $x\phi(x)$ is bounded it follows that, replacing C_1 and C_2 by larger constants if necessary, (8.2) and (8.3) also hold for $\mathbb{P} = \mathbb{P}_c$. □

We are now ready to prove Lemma 7.2. The proof splits into three cases: the heart of the proof is the good case with $\Delta m \neq 0$. The bad case with $\Delta m \neq 0$ is much easier, as it has been 'cooked' to work. For $\Delta m = 0$ the only difference is that we must also consider the event $S_{r+1,m} \setminus E_{r+1}$. However, it will be easy to show that the probability of this event is small compared to P_I.

Proof of Lemma 7.2. We first consider the most difficult case.

The good case, $\Delta m \neq 0$
Let $\Delta m \in \mathbb{Z}$ with

$$0 < |\Delta m| \leqslant 2M/\delta_m + 1. \tag{8.7}$$

As described above, we define the rectangle $R \subset \mathbb{R}^2$ by

$$R = \{(x, y) : \Delta m \delta_m \leqslant y - x < (\Delta m + 1)\delta_m,\ -2M \leqslant x + y \leqslant b\}.$$

Recall that $D'_{r+1} = D'_r + I(E_{r+1})\lfloor (Y - X)/\delta_m \rfloor \delta_m \sqrt{n}$. Since $\Delta m \neq 0$, $S_{r,m} \cap S_{r+1,m+\Delta m} \subset E_{r+1}$. Thus, assuming $2M \geqslant b$, on all of $S_{r,m}$ we have

$$\mathbb{P}(S_{r+1,m+\Delta m} \mid \mathcal{F}_r) = \mathbb{P}((X, Y) \in R \mid \mathcal{F}_r). \tag{8.8}$$

Let $X_I \sim N(\tau_r, \alpha)$, $Y_I \sim N(0, \beta)$. We should like to approximate $\mathbb{P}((X, Y) \in R \mid \mathcal{F}_r)$ by $\mathbb{P}((X_I, Y_I) \in R)$ but we need to be careful about uniformity. That is, the approximation may be a good one for $n \geqslant n_0$, but this is only useful if the same n_0 works for all allowable r, m, and Δm.

Let R' be the largest subset of R that is a union of small squares of the form $[u\delta_s, (u+1)\delta_s) \times [v\delta_s, (v+1)\delta_s)$ where $u, v \in \mathbb{Z}$. Also let R'' be the smallest covering of R by such small squares.

Consider a small square $Q = [u_0, u_1) \times [v_0, v_1)$ in $R'' \supset R'$. On $S_{r,m} \setminus B_r$,

$$\begin{aligned} \mathbb{P}((X, Y) \in Q \mid \mathcal{F}_r) &= \mathbb{P}(X \in [u_0, u_1), Y \in [v_0, v_1) \mid \mathcal{F}_r) \\ &= (\mathbb{P}(X_I \in [u_0, u_1)) + \delta_x)(\mathbb{P}(Y_I \in [v_0, v_1)) + \delta_y), \end{aligned}$$

where $|\delta_x| \leqslant 2C_1/\sqrt{n}$ and $|\delta_y| \leqslant 2C_2/\sqrt{n}$ by Lemma 8.2. For given δ_m, δ_s, the variables u_0, u_1, v_0, v_1, and the means of X_I and Y_I are bounded, and the variances of X_I and Y_I are bounded and bounded away from 0, so $\mathbb{P}(X_I \in [u_0, u_1))$ and $\mathbb{P}(Y_I \in [v_0, v_1))$ are bounded below by x_0 and y_0 say, where x_0 and y_0 are independent of $r, m, \Delta m$. This gives us that on $S_{r,m} \setminus B_r$,

$$\begin{aligned} \mathbb{P}((X, Y) \in R' \mid \mathcal{F}_r) &\geqslant e^{-\epsilon_2}\mathbb{P}((X_I, Y_I) \in R'), \\ \mathbb{P}((X, Y) \in R'' \mid \mathcal{F}_r) &\leqslant e^{\epsilon_2}\mathbb{P}((X_I, Y_I) \in R''), \end{aligned} \tag{8.9}$$

where ϵ_2 depends on n, δ_m, δ_s only and tends to 0 as $n \to \infty$, and hence as (4.1).

To prove that the relative error between $\mathbb{P}((X_I, Y_I) \in R')$ and $\mathbb{P}((X_I, Y_I) \in R'')$ tends to

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

0, it will be more convenient to complete an explicit calculation. Let

$$R^{\pm} = \{(x,y) : \Delta m\delta_{\mathrm{m}} \mp 2\delta_{\mathrm{s}} \leqslant y - x < (\Delta m+1)\delta_{\mathrm{m}} \pm 2\delta_{\mathrm{s}}, \\ -2M \mp 2\delta_{\mathrm{s}} \leqslant x + y \leqslant b \pm 2\delta_{\mathrm{s}}\},$$

so

$$R^{-} \subset R' \subset R \subset R'' \subset R^{+}. \tag{8.10}$$

Let $D = Y_I - X_I$, $S = \beta X_I + \alpha Y_I$. These are multivariate normal with zero covariance, and hence independent. Note that $D \sim N(-\tau_r, 1)$, $S \sim N(\beta\tau_r, \alpha\beta)$ and $X_I + Y_I = 2S + (\beta-\alpha)D$. For an infinitesimal interval,

$$\begin{aligned} &\mathbb{P}(-2M \mp 2\delta_{\mathrm{s}} \leqslant X_I + Y_I \leqslant b \pm 2\delta_{\mathrm{s}} \mid Y_I - X_I \in [w, w + \mathrm{d}w))/\,\mathrm{d}w \\ &= \mathbb{P}(-2M \mp 2\delta_{\mathrm{s}} \leqslant 2S + (\beta-\alpha)w \leqslant b \pm 2\delta_{\mathrm{s}}) \\ &= \Phi_1^{\pm}(w) - \Phi_2^{\pm}(w), \end{aligned}$$

where

$$\begin{aligned} \Phi_1^{\pm}(w) &= \Phi((b/2 \pm \delta_{\mathrm{s}} - (\beta-\alpha)w/2 - \beta\tau_r)/\sqrt{\alpha\beta}), \\ \Phi_2^{\pm}(w) &= \Phi((-M \mp \delta_{\mathrm{s}} - (\beta-\alpha)w/2 - \beta\tau_r)/\sqrt{\alpha\beta}), \end{aligned}$$

so

$$\mathbb{P}((X_I, Y_I) \in R^{\pm}) = \int_{\Delta m\delta_{\mathrm{m}} \mp 2\delta_{\mathrm{s}}}^{(\Delta m+1)\delta_{\mathrm{m}} \pm 2\delta_{\mathrm{s}}} (\Phi_1^{\pm}(w) - \Phi_2^{\pm}(w))\phi(w+\tau_r)\,\mathrm{d}w. \tag{8.11}$$

It is easy to check that for $L > 0$, $\Phi(x-L)/\Phi(x)$ is increasing in x (this follows from the fact that $R'(x) \leqslant 0$, which was shown in the proof of Lemma 5.1). Therefore, to work out the largest possible value of $\Phi_2^{\pm}(w)/\Phi_1^{\pm}(w)$ we need to consider the largest values of $|w|$ and τ_r. For $w \in [\Delta m\delta_{\mathrm{m}} \mp 2\delta_{\mathrm{s}}, (\Delta m+1)\delta_{\mathrm{m}} \pm 2\delta_{\mathrm{s}})$, we have $|w| \leqslant 2M + 2\delta_{\mathrm{m}} + 2\delta_{\mathrm{s}}$ by (8.7). Since we are considering the good case, from the definition of B_r^4 we have $|\tau_r| \leqslant C$, so

$$\begin{aligned} \frac{\Phi_2^{\pm}(w)}{\Phi_1^{\pm}(w)} &\leqslant \frac{\Phi((-M \mp \delta_{\mathrm{s}} + (1-2\delta_{\mathrm{c}})(M + \delta_{\mathrm{m}} + \delta_{\mathrm{s}}) + C)/\sqrt{\alpha\beta})}{\Phi((b/2 \pm \delta_{\mathrm{s}} + (1-2\delta_{\mathrm{c}})(M + \delta_{\mathrm{m}} + \delta_{\mathrm{s}}) + C)/\sqrt{\alpha\beta})} \\ &= \frac{\Phi((-2\delta_{\mathrm{c}}M + f^{\pm}(\delta_{\mathrm{m}}, \delta_{\mathrm{s}}))/\sqrt{\alpha\beta})}{\Phi(((1-2\delta_{\mathrm{c}})M + g^{\pm}(\delta_{\mathrm{m}}, \delta_{\mathrm{s}}))/\sqrt{\alpha\beta})} \\ &\qquad \text{for some bounded functions } f^{\pm}, g^{\pm} \\ &\leqslant \frac{\Phi(2(-2\delta_{\mathrm{c}}M + f^{\pm}(\delta_{\mathrm{m}}, \delta_{\mathrm{s}})))}{\Phi(2((1-2\delta_{\mathrm{c}})M + g^{\pm}(\delta_{\mathrm{m}}, \delta_{\mathrm{s}})))} \\ &\qquad \text{for sufficiently large } M, \text{ since } \alpha\beta \leqslant 1/4. \end{aligned}$$

The above bound tends to 0 as $M \to \infty$ (*i.e.*, as $\delta_{\mathrm{m}} \to 0$) and is independent of $r, m, \Delta m$. Thus we may ignore $\Phi_2^{\pm}(w)$ and write

$$\begin{aligned} \mathbb{P}((X_I, Y_I) \in R^{-}) &\geqslant e^{-\epsilon_3}\delta_{\mathrm{m}}\Phi_1^{-}(w^{-})\phi(w^{-} + \tau_r), \\ \mathbb{P}((X_I, Y_I) \in R^{+}) &\leqslant e^{\epsilon_3}\delta_{\mathrm{m}}\Phi_1^{+}(w^{+})\phi(w^{+} + \tau_r), \end{aligned} \tag{8.12}$$

for some $w^{\pm}$ between $\Delta m\delta_{\mathrm{m}} \mp 2\delta_{\mathrm{s}}$ and $(\Delta m+1)\delta_{\mathrm{m}} \pm 2\delta_{\mathrm{s}}$, where ϵ_3 is some function of δ_{m},

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

δ_s only, which tends to 0 as (4.1). Note that we have changed the length of the integration interval from $\delta_m \pm 4\delta_s$ to δ_m by absorbing $-\log(1-4\delta_s/\delta_m)$ into ϵ_3.

We need to examine how much the right-hand side of (8.12) can vary as ω varies over $S_{r,m} \setminus B_r$.

Let $R(x) = \phi(x)/\Phi(x)$ as before. Obviously $R(x) \geqslant 0$. On the other hand, integrating the inequality $(y/R(x)+1)^2\phi(y) \geqslant 0$ with respect to y from $-\infty$ to x gives that $-R'(x) = (x+R(x))R(x) \leqslant 1$. For $x \geqslant 0$ we have $R(x) \leqslant 2\phi(0) \leqslant 1$, so $0 \leqslant R(x) \leqslant |x|+1$ for all x.

Since $R(x) = \frac{d}{dx}\log\Phi(x)$, we have that

$$\begin{aligned} e^{-\lambda|\Delta x|} &\leqslant \Phi(x+\Delta x)/\Phi(x) \leqslant e^{\lambda|\Delta x|}, \\ e^{-\lambda|\Delta x|} &\leqslant \phi(x+\Delta x)/\phi(x) \leqslant e^{\lambda|\Delta x|}, \end{aligned} \tag{8.13}$$

where $\lambda = |x| + |\Delta x| + 1$. (The inequalities for $\phi()$ are obvious.)

Recall that $w' = \Delta m \delta_m$, so w' and $\tau'_r = m\delta_m/s$ are constant over $S_{r,m}$, and $|w^{\pm} - w'| \leqslant \delta_m + 2\delta_s$. This inequality together with (7.1) and (8.13) allows us to estimate the cost of replacing $w^{\pm}$ by w', τ_r by τ'_r and δ_s by 0 in the right-hand side of (8.12) after $\Phi_1^{\pm}()$ has been replaced by its definition in terms of $\Phi()$. More precisely, let

$$x^{\pm} = (b/2 \pm \delta_s - (\beta-\alpha)w^{\pm}/2 - \beta\tau_r)/\sqrt{\alpha\beta},$$

so $\Phi_1^{\pm}(w^{\pm}) = \Phi(x^{\pm})$, and recall that

$$x = (b/2 - (\beta-\alpha)w'/2 - \beta\tau'_r)/\sqrt{\alpha\beta}.$$

Then

$$|x^{\pm} - x| \leqslant (\delta_s + (\delta_m + 2\delta_s) + 2\delta_m/\delta_c)/\sqrt{\delta_c(1-\delta_c)} = \epsilon_4,$$

say, with $\epsilon_4 \to 0$ as (4.1). Also

$$\begin{aligned} |x| &\leqslant (|b|/2 + \delta_m|\Delta m| + C + 2\delta_m/\delta_c)/\sqrt{\delta_c(1-\delta_c)} \\ &\leqslant C_7 + C_8\delta_m|\Delta m| \end{aligned}$$

for some constants C_7, C_8, and $\lambda = |x| + |x^{\pm} - x| + 1$ is bounded by $C_9 + C_8\delta_m|\Delta m|$ for some C_9.

Turning to the $\phi(.)$ term of (8.12), ϵ_4 is also an upper bound for $|w^{\pm} + \tau_r - (w' + \tau'_r)|$, while $C_7 + C_8\delta_m|\Delta m|$ is also an upper bound for $|w' + \tau'_r|$.

Recall that

$$P_I = \delta_m \Phi(x)\phi(w' + \tau'_r).$$

Combining (8.8), (8.9), (8.10), (8.12) and (8.13) we see that (8.1) holds on $S_{r,m} \setminus B_r$ with

$$\begin{aligned} \epsilon_0 &= 2(\epsilon_2 + \epsilon_3 + 2C_9\epsilon_4), \\ \epsilon_1 &= 4C_8\epsilon_4, \end{aligned}$$

completing the proof of the lemma in this case.

The bad case

Here we have a good version of the conditional probability because in general for a bounded measurable f, and for vector-valued random variables $\mathbf{A}$, $\mathbf{B}$, if $\mathbf{A}$ is independent of $\mathbf{B}$ then $\mathbf{b} \mapsto \mathbb{E}[f(\mathbf{A},\mathbf{b})]$ is a version of $\mathbb{E}[f(\mathbf{A},\mathbf{B}) \mid \sigma(\mathbf{B})]$. We apply this with $\mathbf{A} =$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

$(Z_{1,r+1}, Z_{2,r+1})$, $\mathbf{B}$ = 'the collection of random variables generating $\mathcal{F}_r$', and $f(\mathbf{a}, \mathbf{b}) = I((\mathbf{a}, \mathbf{b}) \in V)$. In effect we treat τ'_r as a constant given $\mathcal{F}_r$, and (X, Y) will have the distribution $(N(\tau'_r, \alpha), N(0, \beta))$.

The bad case, $\Delta m \neq 0$

This is relatively easy compared with the good case, because it has been 'cooked' to work. In this case $(X, Y) = (X_{I,r+1}, Y_{I,r+1})$, which has exactly the distribution of $(X^\star, Y^\star)$. A similar argument to the one leading to (8.11), applied directly to R, shows that on $S_{r,m} \cap B_r$,

$$\begin{aligned} \mathbb{P}(S_{r+1,m+\Delta m} \mid \mathcal{F}_r) &= \int_{w'}^{w'+\delta_{\mathrm{m}}} (\Phi_1(w) - \Phi_2(w))\phi(w + \tau'_r)\,\mathrm{d}w, \text{ where} \\ \Phi_1(w) &= \Phi((b/2 - (\beta - \alpha)w/2 - \beta\tau'_r)/\sqrt{\alpha\beta}), \\ \Phi_2(w) &= \Phi((-M - (\beta - \alpha)w/2 - \beta\tau'_r)/\sqrt{\alpha\beta}), \\ w' &= \Delta m \delta_{\mathrm{m}}. \end{aligned}$$

There is no problem repeating the previous argument which led to the elimination of the $\Phi_2()$ term and the approximation of the integral by the same P_I. The reason is that the only places 'goodness' was used in the previous argument were:

(i) in getting the approximating normal formulas (8.2), (8.3),
(ii) in the bounding of τ_r, and
(iii) in the approximation of τ'_r by τ_r.

In the bad case we start off with X and Y being normally distributed with the mean of X already τ'_r and the other mean and variances correct, so we have no need of (i) or (iii). Instead of (ii), we use the condition at the start of the lemma, $|m| \leqslant Cn/\delta_{\mathrm{m}}$, to bound $|\tau'_r|$ by C/δ_{c}.

Therefore, increasing ϵ_0 and ϵ_1 if necessary, (8.1) also holds on $S_{r,m} \cap B_r$.

The case $\Delta m = 0$

The above reasoning which led to (8.1) is nearly valid if $\Delta m = 0$: it gives an estimate of $\mathbb{P}(S_{r+1,m} \cap E_{r+1} \mid \mathcal{F}_r)$ rather than $\mathbb{P}(S_{r+1,m} \mid \mathcal{F}_r)$. To prove that (8.1) is still true as it stands (with possibly increased ϵ_0, ϵ_1), it is therefore sufficient to show that $\mathbb{P}(\{d''_{r+1} \leqslant b\sqrt{n}\} \setminus E_{r+1} \mid \mathcal{F}_r)$ is small compared with $\mathbb{P}(S_{r+1,m} \cap E_{r+1} \mid \mathcal{F}_r)$, or equivalently that $\mathbb{P}(\{d''_{r+1} \leqslant b\sqrt{n}\} \setminus E_{r+1} \mid \mathcal{F}_r)$ is small compared with P_I at $\Delta m = 0$. The latter is given by

$$P_I = \delta_{\mathrm{m}} \Phi((b/2 - \beta\tau'_r)/\sqrt{\alpha\beta})\phi(\tau'_r).$$

By assumption $|\tau'_r|$ is bounded by C/δ_{c}, and $\delta_{\mathrm{c}} \leqslant \alpha, \beta \leqslant 1 - \delta_{\mathrm{c}}$, so P_I is bounded below by $C_{10}\delta_{\mathrm{m}}$ for some constant $C_{10} > 0$.

Now

$$\begin{aligned} E_{r+1} &= \{|\lfloor (Y - X)/\delta_{\mathrm{m}} \rfloor \delta_{\mathrm{m}}| \leqslant 2M + \delta_{\mathrm{m}}, \ \ |X + Y| \leqslant 2M\} \\ &\supset \{|Y - X| \leqslant 2M, \ \ |X + Y| \leqslant 2M\}, \end{aligned}$$

and

$$\{d''_{r+1} \leqslant b\sqrt{n}\} = \{X + Y \leqslant b\},$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

so

$$\{d''_{r+1} \leqslant b\sqrt{n}\} \setminus E_{r+1} \subset \{X \leqslant |b|/2 - M\} \cup \{Y \leqslant |b|/2 - M\},$$

if $2M \geqslant |b|$. Let

$$\epsilon = (\mathbb{P}(X \leqslant |b|/2 - M \mid \mathcal{F}_r) + \mathbb{P}(Y \leqslant |b|/2 - M \mid \mathcal{F}_r))/(C_{10}\delta_{\mathrm{m}}).$$

In the good case $|\tau_r| \leqslant C$ and, if $M \geqslant |b|/2 + C$, then (8.2) and (8.3) give that

$$\epsilon \leqslant (\Phi(|b|/2 - M + C) + \Phi(|b|/2 - M) + (C_1 + C_2)/\sqrt{n})/(C_{10}\delta_{\mathrm{m}}).$$

Recalling that $M = C/\delta_{\mathrm{m}}$, we see that this bound depends only on n, δ_{m} and tends to 0 as (4.1). (As a technicality, we are allowing ϵ_0 to take the value ∞, which may be necessary before ϵ has become sufficiently small.)

In the bad case $X \sim N(\tau'_r, \alpha)$ and $Y \sim N(0, \beta)$, so

$$\epsilon \leqslant (\Phi(|b|/2 - M + C/\delta_{\mathrm{c}}) + \Phi(|b|/2 - M))/(C_{10}\delta_{\mathrm{m}}),$$

and again this upper bound depends only on n, δ_{m} and tends to 0 as (4.1).

Thus $\mathbb{P}(\{d''_{r+1} \leqslant b\sqrt{n}\} \setminus E_{r+1} \mid \mathcal{F}_r)$ is indeed small compared with P_I and hence with $\mathbb{P}(S_{r+1,m} \cap E_{r+1} \mid \mathcal{F}_r)$, completing the proof of the lemma. □

Concluding remark. Having completed the proof of Lemma 7.2, and so Theorem 2.1, we finish by pointing out an easy generalization. The argument given did not depend on the precise distribution of the edge weights. We could thus draw the same conclusion assuming only that the e_{ij} are independent and identically distributed with $\mathbb{E}[|e_{ij}|^3]$ finite.

9. Appendix

The method of proof used in this paper grew out of a slightly different approach, which we sketch here. The vertices are still considered in order, keeping track of D_r (or τ_r) which is our state, but instead of saying that the discrete probability is similar to that derived from normal variables, we attempt to calculate it directly in the different continuum limit in which α is considered as a real variable. The authors are not sure whether this can be made into a rigorous argument – rather than complete this line of reasoning we decided that it was simpler to use similar methods to prove that it gave the same answer as in the continuous (normal) case.

As seen in the previous section, the distributions of X_r and Y_r are approximately $N(\tau_r, \alpha)$ and $N(0, \beta)$ respectively. We consider the vertices in 'blocks' of size $n\Delta\alpha$ for some small $\Delta\alpha$, and assume that the distributions of X_r and Y_r are approximately independent and identically distributed as r varies over a block. The suffix 'r' is now dropped in preparation for making the continuum limit. Thus $\Delta(n\beta\tau)$, the change in $n\beta\tau = n^{-1/2}D$ from r to $r + n\Delta\alpha$, is given by the sum of $n\Delta\alpha$ copies of Z, the distribution of $Y - X$ conditioned on $X + Y \leqslant b$. Let $\dot{\tau} = \Delta\tau/\Delta\alpha$ and

$$L(\alpha, \tau, \dot{\tau}) = \lim_{n\to\infty} \left[(n\Delta\alpha)^{-1} \log \mathbb{P}\left(\sum_{n\Delta\alpha} Z \approx \Delta(n\beta\tau) \right) \right] + \log \mathbb{P}(X + Y \leqslant b),$$

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press

for some suitable interpretation of '$\approx$'. A standard result on large deviations (see, *e.g.*, [7], §5.11) implies that $L = \min_\theta \log f(\theta)$ where

$$f(\theta) = \mathbb{E}[e^{-\theta(Y-X+\tau-\beta\dot{\tau})} I(X+Y \leqslant b)],$$

and a simple calculation gives

$$f(\theta) = e^{\theta^2/2+\beta\theta\dot{\tau}} \Phi(b + (\beta - \alpha)\theta - \tau).$$

The logarithm of the nth root of the probability of a path $\tau : [0,1] \to \mathbb{R}$ is then given by $\int_0^1 L(\alpha, \tau, \dot{\tau})\, \mathrm{d}\alpha$. We expect the sum over all paths of the probability of a path to be dominated by the sum over paths very close to the most likely path, and in fact that

$$\log \lim_{n\to\infty} p_n^{1/n} = \max_\tau \int_0^1 L(\alpha, \tau, \dot{\tau})\, \mathrm{d}\alpha,$$

where the maximum is taken over paths such that $\tau(0) = 0$.

In summary, $\log(c)$ would be given in this method by the extremal value of

$$\int_0^1 \left(\frac{\theta^2}{2} + \log \Phi(b + (1-2\alpha)\theta - \tau) + (1-\alpha)\theta \frac{\mathrm{d}\tau}{\mathrm{d}\alpha} \right) \mathrm{d}\alpha \tag{9.1}$$

over $\tau, \theta : [0,1] \to \mathbb{R}$ with $\tau(0) = 0$ (minimized over θ and maximized over τ). We can write down and solve the Euler–Lagrange equations for this problem. It turns out that the minimum is attained at $\theta(\alpha) \equiv \theta_0$ and $\tau(\alpha) = -2\theta_0\alpha$, where θ_0 is the constant defined by (2.2). Substituting this in (9.1) gives the same constant c as in Theorem 2.1.

References

[1] Bollobás, B. (1981) Degree sequences of random graphs. *Discrete Math.* **33** 1–19.
[2] Bollobás, B. (1985) *Random Graphs*, Academic Press, London/New York.
[3] Bollobás, B. (1986) *Combinatorics*, Cambridge University Press, Cambridge/New York.
[4] Courant, R. and Hilbert, D. (1953) *Methods of Mathematical Physics*, Vol. I, Interscience Publishers, New York/London.
[5] Erdős, P. and Rényi, A. (1959) On random graphs, I. *Publ. Math. Debrecen* **6** 290–297.
[6] Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd edn, Wiley, New York/London/Sydney.
[7] Grimmett, G. R. and Stirzaker, D. R. (1992) *Probability and Random Processes*, 2nd edn, Clarendon Press, Oxford/New York.
[8] Kleitman, D. J. (1966) Families of non-disjoint subsets. *J. Combin. Theory* **1** 153–155.
[9] McKay, B. D. and Wormald, N. C. (1997) The degree sequence of a random graph, I: The models. *Random Structures and Algorithms* **11** 97–118.

https://doi.org/10.1017/S0963548300004491 Published online by Cambridge University Press