COMBINATORICA 4 (1) (1984) 7—19

DIAMETERS OF RANDOM BIPARTITE GRAPHS

Béla BOLLOBÁS and Victor KLEE

Dedicated to Paul Erdős on his seventieth birthday

Received 24 March 1983

0. Introduction

The diameter of a graph is of mathematical interest as one of the most thoroughly studied parameters of graph theory, and of practical interest because of its relationship to the computational complexity of graph algorithms based on breadth-first search. Diameters of random graphs have been studied by Bollobás [3], Klee and Larman [8], Korshunov [11], Moon and Moser [13], and Thanh [15], and of random bipartite graphs by Klee, Larman and Wright [9]. Results on connectedness are also relevant, for a graph is connected if and only if its diameter is finite. Connectedness of random graphs has been studied by Erdős and Rényi [7], and of random bipartite graphs by Bollobás [4] and Klee, Larman and Wright [10].

For $p\in(0, 1)$ and positive integers m and n with $m\leqq n$, let $\mathbf{G}(m, n; p)$ denote the probability space of bipartite graphs in which the first part V consists of m labelled vertices, the second part W (disjoint from V) consists of n labelled vertices, and for each $v\in V$ and $w\in W$ the edge $\{v, w\}$ is present with probability p, independent of the presence of other edges. Thus for each subset Y of $V\times W$ the probability is $p^{|Y|}(1-p)^{mn-|Y|}$ that a random member of $\mathbf{G}(m, n; p)$ has Y as its edge-set.

We seek conditions on the sequences $(m(1), m(2), \dots)$ and $(p(1), p(2), \dots)$ that lead to conclusions about *almost every* member of $\mathbf{G}(m(n), n; p(n))$ — that is, conclusions which hold with a probability $P(n)$ that converges to 1 as $n\to\infty$. Our two main results are the following, dealing respectively with the cases in which m is (is not) much less than n.

Theorem A. *If*

$$n(1-p)^m \to 0 \tag{1}$$

and

$$\frac{n(\log n)^2}{m^2} - 2\log m > \lambda \tag{2}$$

for some constant λ then almost every member G of $\mathbf{G}(m, n; p)$ is such that any two vertices of V have a common neighbor in W (whence G is of diameter $\leqq 4$).

AMS subject classification (1980): 05 C 40; 60 C 05.

Theorem B. *Suppose that for all n,*

$$pn \geqq pm \geqq (\log n)^4 \tag{3}$$

and that k is a fixed positive integer.

If k is odd and

$$p^k m^{(k-1)/2} n^{(k-1)/2} - \log (mn) \to \infty$$

or if k is even and

$$p^k m^{k/2} n^{(k/2)-1} - 2 \log n \to \infty$$

then almost every member of $\mathbf{G}(m, n; p)$ *is of diameter* $\leqq k+1$.

If k is odd and

$$p^k m^{(k-1)/2} n^{(k-1)/2} - \log (mn) \to -\infty$$

or if k is even and

$$p^k m^{k/2} n^{(k/2)-1} - 2 \log n \to -\infty$$

then almost every member of $\mathbf{G}(m, n; p)$ *is of diameter* $\geqq k+2$.

As is shown in Section 1, the growth condition (2) says roughly that $m < < (n \log n)^{1/2}$, while (1) is implied by the assumption that almost no member of $\mathbf{G}(m, n; p)$ has an isolated vertex. Thus Theorem A is in the spirit of the theorem of Pósa [14] and Korshunov [12] asserting that for random graphs, about the same number of edges needed to ensure one property for almost all graphs actually ensures a much stronger property for almost all graphs. (In their case the properties are connectedness and the presence of a Hamilton cycle.)

Note that if (1) holds and $m \geqq (\log n)^2$ then

$$pm - \log n \to \infty \tag{4}$$

for if $pm \leqq \log n + \tau$ for a constant τ then

$$\limsup n(1-p)^m \geqq \limsup n \exp \{(-p-p^2)m\} \geqq e^{-\tau-1} > 0.$$

Though we work with (3) and (4), it appears that our method can be extended to yield sharp or almost sharp results throughout the range (4). However, that would require more complicated proofs. Condition (3) is convenient because it enables us to base the main part of Theorem B's proof directly on a lemma from [3].

1. Theorem A

Before proving Theorem A, we comment on the growth conditions (1) and (2).

1.1. *If almost no member of* $\mathbf{G}(m, n; p)$ *has an isolated vertex then*

$$n(1-p)^m \to 0. \tag{1}$$

Proof. Let the random variable $I_n = I_n(G)$ denote the number of vertices in W that are isolated in the graph $G \in \mathbf{G}(m, n; p)$. Let q denote the probability that a particular member of W is isolated, so that $q = (1-p)^m$. Then the distribution of I_n is binomial with parameters n and q, whence the kth factorial moment of I_n is $\binom{n}{k} q^k$. If nq converges to a constant $\gamma > 0$ as $n \to \infty$ then $\binom{n}{k} q^k$ converges for each k to γ^k, the kth factorial moment of the Poisson distribution with mean γ. But then I_n converges in distribution to the Poisson distribution, whence $P(I_n = 0) \to e^{-\gamma} < 1$. It follows that if $P(I_n = 0) \to 1$ then $n(1-p)^m \to 0$. ∎

1.2. *The condition that eventually (that is, for all sufficiently large n)*

$$(2_*) \qquad m < (n \log n)^{1/2}$$

is implied by the condition that

$$(2) \qquad \frac{n(\log n)^2}{m^2} - 2 \log m \geqq \lambda$$

for some constant λ, which in turn is implied by the condition that

$$(2^*) \qquad m < (n \log n)^{1/2} \left[1 - \left(\frac{1}{2} + \delta\right) \frac{\log \log n}{\log n}\right]$$

for some fixed $\delta > 0$.

Proof. If (2^*) holds then eventually

$$m^2 < (n \log n)\left[1 - \left(1 + \frac{3\delta}{2}\right) \frac{\log \log n}{\log n}\right],$$

whence eventually

$$\frac{n(\log n)^2}{m^2} - 2 \log m > (\log n)\left[1 + (1+\delta) \frac{\log \log n}{\log n}\right] - (\log n + \log \log n)$$

$$> \delta \log \log n \to \infty$$

and (2) holds. To see that (2) implies (2_*), note that the left side of (2) decreases as m increases, and when $m = (n \log n)^{1/2}$ the left side of (2) becomes

$$\log n - (\log n + \log \log n) \to -\infty. \quad ∎$$

To prove Theorem A, let $P(n)$ denote the probability that in a random member G of $\mathbf{G}(m, n; p)$ there are two vertices in V which have no common neighbor. Since, for each triple of distinct vertices $v, v' \in V$ and $w \in W$, the probability is $1 - p^2$ that v and v' are not both adjacent to w it follows that

$$P(n) \leqq \binom{m}{2} (1-p^2)^n \leqq m^2 e^{-p^2 n}.$$

Thus if the conclusion of Theorem A fails then $\liminf m^2 e^{-p^2 n} > 0$ and, by passing to a subsequence, $2 \log m - p^2 n > -\gamma$ for some constant γ. Then

$$p < \left(\frac{2 \log m + \gamma}{n}\right)^{1/2}$$

and hence eventually

$$n(1-p)^m \geqq \exp\{\log n-(p+p^2)m\}$$

$$\geqq \exp\left\{\log n-\left(\frac{2\log m+\lambda}{n}\right)^{1/2} m-\frac{2\log m+\lambda}{n} m\right\} \geqq \tau,$$

where

$$\tau = \frac{1}{2}\exp\left\{\log n-\left(\frac{2\log m+\lambda}{n}\right)^{1/2} m\right\}.$$

If $\liminf \tau>0$ then $\liminf n(1-p)^m>0$, contradicting (1). If $\liminf \tau=0$ then eventually $\left(\frac{2\log m+\lambda}{n}\right)^{1/2} m>\log n+\mu$, where $\mu=\max\left\{\frac{\gamma-\lambda+1}{2}, 0\right\}$ and λ is the constant of Theorem A. But then

$$2\log m > \frac{(\log n+\mu)^2}{m^2} n-\lambda$$

$$> \frac{n(\log n)^2}{m^2}+2\mu-\lambda > \frac{n(\log n)^2}{m^2}+1-\lambda,$$

contradicting (2). That completes the proof of Theorem A. ▮

2. Lemmas in preparation for Theorem B

When K is a set of finite cardinality k, $\mathbf{A}$ is a family of subsets of K, and $p \in[0, 1]$, let

$$\varphi_{\mathbf{A}}(p) = \sum_{A\in \mathbf{A}} p^{|A|}(1-p)^{k-|A|}.$$

This is just the probability that a member of $\mathbf{A}$ is formed when each point of K is chosen with probability p and the k choices are independent. Though 2.1 below is intuitively obvious, it does require a proof and one has been given by Birnbaum, Esary and Saunders [1]. The short proof below is from [5] and was suggested also by Isaac Namioka.

2.1. *If the family* $\mathbf{A}$ *includes each subset of* K *that contains (resp. is contained in) a member of* $\mathbf{A}$ *then* $\varphi_{\mathbf{A}}(p)$ *is a monotone increasing (resp. decreasing) function of* p.

Proof. For $0\leqq j\leqq k$, let a_j denote the number of sets $A\in\mathbf{A}$ such that $|A|=j$. Note that when $0\leqq j<k$, $(k-j)a_j\leqq(j+1)a_{j+1}$ (resp. $(k-j)a_j\geqq(j+1)a_{j+1}$) for if $|A|=j$ there are (in K) $k-j$ supersets B of A such that $|B|=j+1$, while if $|B|=j+1$ there are $j+1$ subsets A of B such that $|A|=j$.

Let $q=1-p$, whence $\varphi_{\mathbf{A}}(p)=\sum_{j=0}^{k} a_j p^j q^{k-j}$ and thus

$$\varphi_A(p) = \sum_{j=1}^{k} a_j j p^{j-1}q^{k-j}+\sum_{j=0}^{k-1} a_j p^j(j-k)p^{k-j-1}$$

$$= \sum_{j=0}^{k} [a_{j+1}(j+1)-(k-j)a_j]p^j q^{k-j-1} \geqq 0 \quad (\text{resp. } \leqq 0). \ ▮$$

Our use of 2.1 is based on the fact that if K is the edge-set of a graph with vertex-set U, d is a positive integer, and $\mathbf{A}$ is the family of all subsets A of K such that the graph (U, A) is of diameter $\leqq d$ (resp. $\geqq d$), then $\mathbf{A}$ satisfies the first (resp. second) condition in 2.1.

The most essential tool in our proof of Theorem B is the following lemma from [3].

2.2. *Suppose that the random variable X has binomial distribution with parameters a and p. If*

$$0 < \varepsilon < \frac{1}{21}, \quad 0 < p < \frac{1}{3} \quad \textit{and} \quad \varepsilon pa > 40$$

then

$$P(|X-pa| \geqq \varepsilon pa) < \exp\{(-\varepsilon^2 pa/3)/\varepsilon(pa)^{1/2}\}.$$

For each pair x and y of vertices of a graph G, let $\delta_G(x, y)$ denote the smallest number of edges forming a path that joins x and y in G, with $\delta_G(x, y) = \infty$ when there is no such path. This is the G-distance between x and y, and the diameter of G is of course the maximum of $\delta_G(x, y)$ over all pairs (x, y) of vertices of G. For each nonnegative integer r, let $S_r(x, G)$ denote the sphere in G with radius r and center x. That is,

$$S_r(x, G) = \{y: \delta_G(x, y) = r\}.$$

Returning now to our standard notation (V, W, m, n) for discussing labelled bipartite graphs, let x be a fixed but arbitrary member of $V \cup W$, a center with respect to which certain computations will be made. For each $G \in \mathbf{G}(m, n;\, p)$, let

$$S_r(G) = S_r(x, G), \quad s_r = s_r(G) = |S_r(G)|,$$

$$B_r(G) = \bigcup_{i=0}^{\lfloor r/2 \rfloor} S_{r-2i}(G) = \begin{Bmatrix} S_0 \cup S_2 \cup \ldots \cup S_r & \text{when} & r & \text{is even} \\ S_1 \cup S_3 \cup \ldots \cup S_r & \text{when} & r & \text{is odd} \end{Bmatrix},$$

$$b_r = b_r(G) = |B_r(G)|.$$

When $x \in V$ (resp. $x \in W$), let

$$U_r = U_r(x) = V \quad \text{or} \quad W \quad \text{and} \quad l_r = m \quad \text{or} \quad n$$

according as r is even or odd (resp. odd or even). Though m and n enter almost symmetrically in all that follows, some lack of symmetry is imposed by our standing hypothesis that $m \leqq n$.

2.3. *Conditional on specified values of $s_r(G)$ for $r<j$ the distribution of $s_j(G)$ is binomial with parameters $a_j = l_j - b_{j-2}$ and $p_j = 1-(1-p)^{s_{j-1}}$.*

Proof. It suffices to prove the statement with "values of $s_r(G)$" replaced by "sets $S_r(G)$". Of the l_j members of $V \cup W$ that may, by parity considerations, be at distance j from x, there are b_{j-2} at distance $<j$ from x. Consider an arbitrary one of the remaining a_j members of U_j. It fails to be at distance j from x if and only if it fails to be adjacent to any of the s_{j-1} members of U_{j-1} that are at distance $j-1$ from x. This occurs with probability $(1-p)^{s_{j-1}}$. ∎

2.4. *If* (3) *holds and* $\varkappa$ *is an arbitrary real constant then eventually, for a random* $G\in\mathbf{G}(m, n; p)$, *with probability at least* $1-n^{-\varkappa}$ *we have*

$$|s_1-pl_1| \leqq \varepsilon_1 pl_1,$$

where $\varepsilon_1=(\log n)^{-4/3}$.

Proof. The variable s_1 has binomial distribution with parameters $p_1=p$ and $a_1=l_1$. Hence if n is sufficiently large then by Lemma 2.2

$$P(|s_1-pl_1| \geqq \varepsilon_1 pl_1) \leqq \frac{\exp(-\varepsilon_1^2 pl_1/3)}{\varepsilon_1(pl_1)^{1/2}}.$$

Since $pl_1 \geqq (\log n)^4$, this implies that eventually

$$P(|s_1-pl_1| \geqq \varepsilon_1 pl_1) < n^{-\varkappa}. \quad \blacksquare$$

By Lemma 2.1 we may assume that in Theorem B the conditions on p are barely satisfied. To be precise, we shall assume that

$$(5) \qquad -\log\log m \leqq p^k m^{(k-1)/2} n^{(k-1)/2} - \log(mn) \leqq \log\log m$$

if k is odd and

$$(6) \qquad -\log\log m \leqq p^k m^{k/2} n^{(k/2)-1} - 2\log n \leqq \log\log m$$

if k is even.

Set $\sigma_0=1$ and

$$\sigma_r = \sigma_r(x) = \prod_{j=1}^{r} (pl_j)$$

for $r\geqq 1$. By (5) and (6), if $x\in V$ then

$$(7) \qquad \log n < p\sigma_{k-1} < 3\log n,$$

and if $x\in V\cup W$ and $1\leqq j\leqq k-1$ then

$$(8) \qquad p\sigma_{k-j-1} = \sigma_{k-j}/l_{k-j} \leqq (\sigma_k/l_k)(pm)^{-j} < 3(\log n)^{1-4j}.$$

Set $\varepsilon_r=r(\log n)^{-4/3}=r\varepsilon_1$. We shall be concerned with members G of $\mathbf{G}(m, n; p)$ that satisfy some of the following inequalities involving $s_r(G)$:

$$I_r = I_r(x)\colon\ \sigma_r(1-\varepsilon_r) \leqq s_r \leqq \sigma_r(1+\varepsilon_r).$$

In what follows all inequalities are claimed to hold for all sufficiently large values of n.

2.5. *If* (3) *and* (8) *hold and* x *is an arbitrary real constant then eventually, for a random* $G\in\mathbf{G}(m, n; p)$,

$$P(I_1, I_2, \ldots, I_{k-1} \quad \text{all hold}) > 1-n^{-\varkappa}.$$

Proof. The assertion of the lemma follows if we show that for $1\leqq r\leqq k-1$

$$(9) \qquad P(I_r \quad \text{fails}) < 2^r n^{-\varkappa-1}.$$

In turn, (9) certainly holds for a particular value of r if it holds for all smaller values of r and we have the following bound on a conditional probability:

$$(10) \qquad P(I_r \text{ fails } | I_1, I_2, \ldots, I_{r-1} \text{ hold}) < n^{-\varkappa-1}.$$

Finally, (10) follows if we show that whenever $\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{r-1}$ satisfy $I_1, I_2, \ldots, I_{r-1}$, then

$$(11) \qquad P(I_r \text{ fails } | s_1 = \tilde{s}_1, \ldots, s_{r-1} = \tilde{s}_{r-1}) < n^{-\varkappa-1}.$$

By replacing $\varkappa$ by $\varkappa+1$ in Lemma 2.4 we see that $P(I_1 \text{ fails}) < n^{-\varkappa-1}$.

Suppose now that $2 \leqq r \leqq k-1$ and inequality (9) holds for smaller values of r. Fix integers $\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{r-1}$ satisfying inequalities $I_1, I_2, \ldots, I_{r-1}$ and consider the distribution of s_r conditional on $s_1 = \tilde{s}_1, \ldots, s_{r-1} = \tilde{s}_{r-1}$. This is a binomial distribution with parameters

$$p_r = 1-(1-p)^{s_{r-1}} \quad \text{and} \quad a_r = l_r - b_{r-2}.$$

Our next aim is to estimate these parameters.

By the induction hypothesis

$$p_r = 1-(1-p)^{s_{r-1}} \leqq ps_{r-1} \leqq p\sigma_{r-1}(1+\varepsilon_{r-1})$$

and by (8)

$$p_r \geqq ps_{r-1}\left(1-\frac{ps_{r-1}}{2}\right) \geqq p\sigma_{r-1}(1-\varepsilon_{r-1})(1-p\sigma_{r-1})$$

$$\geqq p\sigma_{r-1}(1-\varepsilon_{r-1})\left(1-3(\log n)^{-3}\right).$$

Hence

$$p_r l_r \geqq \sigma_r(1-\varepsilon_{r-1})\left(1-3(\log n)^{-3}\right).$$

Furthermore, $a_r = l_r - b_{r-2} \leqq l_r$ and since $b_{r-2} \leqq 2\sigma_{r-2}$,

$$a_r \geqq l_r(1-2\sigma_{r-2}/l_r) \geqq l_r\left(1-6(\log n)^{-3}\right).$$

Thus

$$p_r a_r \geqq p_r l_r\left(1-6(\log n)^{-3}\right) \geqq \sigma_r(1-\varepsilon_{r-1})\left(1-3(\log n)^{-3}\right)$$
$$\left(1-6(\log n)^{-3}\right) \geqq \sigma_r/2.$$

Since $r \geqq 2$ this implies

$$(12) \qquad (\log n)^8/2 \leqq \sigma_r/2 \leqq p_r a_r \leqq p_r l_r \leqq p\sigma_{r-1} l_r(1+\varepsilon_{r-1}) = \sigma_r(1+\varepsilon_{r-1}).$$

Let $\varepsilon = (\log n)^{-2}$. Then by Lemma 2.2 we have

$$(13) \quad p(|s_r - a_r p_r| \geqq \varepsilon p_r a_r \,|s_1 = \tilde{s}_1, \ldots, s_{r-1} = \tilde{s}_{r-1}) \leqq \frac{\exp(-\varepsilon^2 p_r a_r/3)}{\varepsilon(p_r a_r)^{1/2}} \leqq n^{-\varkappa-1}.$$

Inequality (13) implies (11). Indeed, if $|s_r - p_r a_r| \leqq \varepsilon p_r a_r$ then by (12)

$$s_r \leqq (1+\varepsilon)p_r a_r \leqq \sigma_r(1+\varepsilon_r)$$

and

$$s_r \geqq (1-\varepsilon)p_r a_r \geqq \sigma_r(1-\varepsilon_r).$$

This completes the proof of Lemma 2.5. ∎

Let x_1 and x_2 be fixed but arbitrary distinct vertices in V or in W and for $G\in\mathbf{G}(m, n; p)$ set

$$C_{r-1}(x_i) = S_{r-1}(x_i)-B_{r-1}(x_i),\ c_{r-1}(x_i) = |C_{r-1}(x_i)|,\quad i = 1, 2,$$

$$D_r(x_1, x_2) = U_r-\big(B_{r-2}(x_1)\cup B_{r-2}(x_2)\big),\ d_r = |D_r|,$$

$$E_r(x_1, x_2) = D_r\cap\Gamma\big(C_{r-1}(x_1)\big)\cap\Gamma\big(C_{r-1}(x_2)\big),\ e_r = |E_r|.$$

Here we have suppressed a number of variables, e.g. $B_j(x_i)$ stands for $B_j(x_i, G)$, $C_j(x_i)$ stands for $C_j(x_i, G)$. Furthermore, $U_r=U_r(x_{i-1})=U_r(x_2)$ and, as usual, $\Gamma(X)$ is the set of all vertices joined to one or more vertices in the set X. (Recall that when $x\in V$ (resp. $x\in W$), $U_r(x)$ is V or W according as r is even or odd (resp. odd or even).)

2.6. *Conditional on specified values of $c_{r-1}(x_1)$, $c_{r-1}(x_2)$ and d_r, the variable e_r has binomial distribution with parameters d_r and*

$$\tilde{p}_r = \big(1-(1-p)^{c_{r-1}(x_1)}\big)\big(1-(1-p)^{c_{r-2}(x_2)}\big),$$

Proof. As in Lemma 2.3, it suffices to prove the lemma with "values of $c_{r-1}(x_1)$, $c_{r-1}(x_2)$ and d_r" replaced by "sets $C_{r-1}(x_1)$, $C_{r-1}(x_2)$ and D_r". Consider a member y of D_r. Then $y\in E_r$ iff at least one edge joins y to $C_{r-1}(x_1)$ and at least one edge joins y to $C_{r-1}(x_2)$. These two events are independent and have probabilities

$$1-(1-p)^{c_{r-1}(x_1)} \quad\text{and}\quad 1-(1-p)^{c_{r-1}(x_2)}.$$

Furthermore, the elements of D_r belong to E_r independently of each other. ∎

Let us continue our investigation of the balls with centers x_1 and x_2. For $G\in\mathbf{G}(m, n; p)$ set

$$T_r(x_1, x_2; G) = B_r(x_1, G)\cap B_r(x_2, G),\quad t_r = |T_r|.$$

In the proof of the second part of Theorem B we shall need that most members G of $\mathbf{G}(m, n; p)$ satisfy some of the following inequalities involving $t_r(x_1, x_2; G)$ for $r\geqq 1$:

$$J_r = J_r(x_1, x_2)\colon\ t_r(x_1, x_2; G) \leqq 3^r\sigma_r/(\log n)^2,$$

where $\sigma_r=\sigma_r(x_1)=\sigma_r(x_2)$.

2.7. *If* (3) *and* (8) *hold and $\varkappa$ is an arbitrary real number then eventually, for every r, $1\leqq r\leqq k-1$,*

$$P(J_r\ \textit{holds}) > 1-rn^{-\varkappa}.$$

Proof. Let J_0 be the event that every vertex in V has degree at most $(1+\varepsilon_1)pm$ and every vertex in W has degree at most $(1+\varepsilon_1)pn$.

By replacing k by $k+1$ in Lemmas 2.4 and 2.5 we see that

(14) $$P(J_0) > 1-n^{-\varkappa-1}$$

and

(15) $$P\big(I_1(x_1), \ldots, I_{r-1}(x_1),\ I_1(x_2), \ldots, I_{r-1}(x_2)\ \text{ all hold}\big) > 1-2n^{-\varkappa-1}.$$

We shall prove our lemma by induction on r. Let $1 \leqq r \leqq k-1$ and let us estimate the probability that

$$e_r \geqq \sigma_r/(\log n)^2,$$

conditional on $I_1(x_1), I_2(x_1), \dots, I_{r-1}(x_1)$ and $I_1(x_2), I_2(x_2), \dots, I_{r-1}(x_2)$ all holding.

Note that if all these $I_j(x_i)$'s hold then

$$c_{r-1}(x_1) \leqq 2\sigma_{r-1}, \; c_{r-1}(x_2) \leqq 2\sigma_{r-1}$$

and $d_r \leqq l_r$. Consequently by Lemma 2.6 the conditional probability above is at most the probability that a binomial random variable with parameters l_r and $\tilde{p}_r = \left(1-(1-p)^{2\sigma_{r-1}}\right)^2 \leqq 4p^2\sigma_{r-1}^2$ is at least $\sigma_r/(\log n)^2$. Since by (8)

$$l_r \tilde{p}_r \leqq 12\sigma_r/(\log n)^3 \leqq \tfrac{1}{2}\sigma_r/(\log n)^2,$$

by Lemma 2.2 this probability is at most $n^{-\varkappa-1}$. Hence, if n is sufficiently large,

$$(16) \quad P\left(e_r \geqq \sigma_r/(\log n)^2 \,|\, I_1(x_1), \dots, I_{r-1}(x_1), I_1(x_2), \dots, I_{r-1}(x_2) \text{ all hold}\right) \leqq n^{-\varkappa-1}.$$

Let K_r be the event that $I_1(x_1), \dots, I_{r-1}(x_1), I_1(x_2), \dots, I_{r-1}(x_2)$ all hold, J_0 and J_{r-1} hold and $e_r \leqq \sigma_r/(\log n)^2$ holds as well. Then by (14), (15), (16) and the induction hypothesis,

$$P(K_r) \geqq 1-2n^{-\varkappa-1}-n^{-\varkappa-1}-(r-1)n^{-\varkappa}-n^{-\varkappa-1} > 1-rn^{-\varkappa}.$$

To complete the proof we show that if K_r holds then so does J_r. Indeed,

$$T_r(x_1, x_2; G) \subset B_{r-2}(x_1) \cup B_{r-2}(x_2) \cup \Gamma\left(T_{r-1}(x_1, x_2; G)\right) \cup E_r(x_1, x_2).$$

Hence if K_r holds,

$$\begin{aligned} t_r &\leqq b_{r-2}(x_1)+b_{r-2}(x_2)+2l_r t_{r-1}(x_1, x_2)+e_r(x_1, x_2) \\ &\leqq \sigma_r/(\log n)^7+2\sigma_r/(\log n)^2+\sigma_r/(\log n)^2 \leqq 3^r\sigma_r/(\log n)^2. \quad \blacksquare \end{aligned}$$

3. The upper bound in Theorem B

3.1. Theorem. *Suppose that k is a positive integer, $k \geqq 3$, and the integer $m(n)$ and probability $p(n)$ are defined for all positive integers n, with*

$$pn \geqq pm \geqq (\log n)^4.$$

Suppose furthermore that if k is odd then

$$p^k m^{(k-1)/2} n^{(k-1)/2} - \log(mn) \to \infty$$

and if k is even then

$$p^k m^{k/2} n^{(k/2)-1} - 2\log n \to \infty.$$

Then almost every member $G \in \mathbf{G}(m, n; p)$ has diameter at most $k+1$.

Proof. As mentioned earlier, by Lemma 2.1 we may assume that (7) and (8) hold as well. Note that the conditions on p imply that for $x \in V \cup W$ we have

$$p\sigma_{k-1}(1-\varepsilon_{k-1})-\log(l_0 l_k) \to +\infty.$$

Call a vertex $x \in V \cup W$ *remote* if $B_k(x) \neq U_k(x)$, that is if for some vertex $y \in U_k(x)$ we have $d(x, y) \geqq k+2$.

Let $x \in V \cup W$. If $y \in U_k(x)$ and $d(x, y) > k$ then $y \notin B_{k-2}(x)$ and no edge joins x to $S_{k-1}(x)$. Hence

$$P(d(x, y) > k | I_{k-1}(x) \text{ holds}) \leqq (1-p)^{\sigma_{k-1}(1-\varepsilon_{k-1})}$$

and so

$$P(x \text{ is remote} | I_{k-1}(x) \text{ holds}) \leqq l_k(1-p)^{\sigma_{k-1}(1-\varepsilon_{k-1})}.$$

Therefore by Lemma 2.5

$$P(x \text{ is remote}) \leqq l_k(1-p)^{\sigma_{k-1}(1-\varepsilon_{k-1})} + n^{-3}$$

and

$$P(\operatorname{diam} G \geqq k+2) \leqq l_0 l_k (1-p)^{\sigma_{k-1}(1-\varepsilon_{k-1})} + n^{-1}$$

$$\leqq \exp\{\log(l_0 l_k) - p\sigma_{k-1}(1-\varepsilon_{k-1})\} + n^{-1} = o(1). \quad ∎$$

4. The lower bound in Theorem B

Denote by X the *number of remote vertices in* W. Then

$$P(\operatorname{diam} G \geqq k+2) \geqq P(X \geqq 1),$$

so by Lemma 2.1 the lower bound in Theorem B is a consequence of the following assertion.

4.1. Theorem. *Suppose that $k \geqq 3$ is a positive integer, c is a real number, and the integer $m(n)$ and probability $p(n)$ are defined for all sufficiently large positive integers n, with $pn \geqq pm \geqq (\log n)^4$.*

Suppose furthermore that

$$p^k m^{(k-1)/2} n^{(k-1)/2} - \log(mn) = c$$

if k is odd and

$$p^k m^{k/2} n^{(k/2)-1} - 2\log n = c$$

if k is even. Then in distribution X tends to the Poisson distribution with mean e^{-c}. In particular,

$$P(X \geqq 1) \to 1 - e^{-e^{-c}}.$$

Proof. Denote by $E_r(X)$ the rth factorial moment of X, that is set $E_r(X) = E(X(X-1)\dots(X-r+1))$. By the Jordan inequalities (see Comtet [6; p. 195]) the theorem follows if we show that for each r,

$$E_r(X) \to e^{-rc}, \tag{17}$$

the rth factorial moment of the Poisson distribution with mean e^{-c}.

Let $x_1, x_2, \dots, x_r$ be arbitrary but fixed vertices in W and set $l_j = l_j(x_1) = l_j(x_2) = \dots = l_j(x_r)$. Let us estimate the probability P_r that each x_i is a remote vertex. By Lemmas 2.5, 2.6 and 2.7 with probability at least $1 - n^{-r-1}$ there are disjoint sets

$\Gamma_1, \Gamma_2, \ldots, \Gamma_r, \Delta_{k-1}, \Delta_k$ such that for all $1 \leqq i \leqq r$

$$\Gamma_i \subset S_{k-1}(x_i) \subset B_{k-1}(x_i) \subset \Gamma_i \cup \Delta_{k-1} \subset U_{k-1}(x_i),$$

$$\left(\bigcup_{i=1}^{r} B_{k-2}(x_i)\right) \cup \Gamma(\Delta_{k-1}) \subset \Delta_k,$$

$$(1-\gamma)\sigma_{k-1} \leqq |\Gamma_i| \leqq (1+\gamma)\sigma_{k-1},$$

$$|\Delta_{k-1}| \leqq \gamma\sigma_{n-1}$$

and

$$|\Delta_k| \leqq 2\gamma\sigma_{k-1} p l_k < \delta l_n,$$

where $\gamma = (\log n)^{-5/4}$ and $\delta = (\log n)^{-1/5}$.

Given sets $V_i \subset V$ and $W_i \subset W$ with $|V_i| = m_i$ and $|W_i| = n_i$, the probability that there is a vertex in V_i which is joined to no vertex in W_i is $1-\left(1-(1-p)^{n_i}\right)^{m_i}$. Consequently,

$$P_r \leqq \left(1-(1-(1-p)^{(1-\gamma)\sigma_{k-1}})^{l_k}\right)^r = \left(1+o(1)\right)(l_k e^{-p\sigma_{k-1}})^r$$

and

$$P_r \geqq \left(1-(1-(1-p)^{(1+\gamma)\sigma_{k-1}})^{(1-\delta)l_k}\right)^r = \left(1+o(1)\right)(l_k e^{-p\sigma_k})^r.$$

Our choice of p implies that

$$l_0 l_k e^{-p\sigma_{k-1}} = \left(1+o(1)\right)e^{-c},$$

so by the inequalities above,

$$E_r(X) = P_r n(n-1) \ldots (n-r+1) \to e^{-rc}.$$

This proves (17) and so our theorem as well. ∎

The proof of Theorem 4.1 can be refined to show that, under suitable conditions, the number of remote vertices in $V \cup W$ also tends to a Poisson random variable. To be precise, the following theorem can be proved.

4.2. *Suppose that $k \geqq 3$ is a positive integer, λ is a positive real number, and the integer $m(n)$ and probability $p(n)$ are defined for all sufficiently large positive integers n, with $pn \geqq pm \geqq (\log n)^4$.*

Suppose furthermore that

$$m l_k e^{p\sigma_{k-1}} + n l'_k e^{p\sigma'_{k-1}} \to \lambda,$$

where

$$l_k = l_k(x), \quad l'_k = l_k(y), \quad \sigma_{k-1} = \sigma_{k-1}(x)$$

and

$$\sigma'_{k-1} = \sigma_{k-1}(y) \quad \textit{for} \quad x \in V \quad \textit{and} \quad y \in W.$$

Then

$$P(\text{diam } G = k+1) \to e^{-\lambda}$$

and

$$P(\text{diam } G = k+2) \to 1-e^{-\lambda}. \quad ∎$$

5. Consequences for the other model

In working with $\mathbf{G}(m, n; p)$ we have chosen the more tractable of two natural models for random bipartite graphs. The other model is $\mathbf{G}(m, n; E)$, where E is a positive integer $< mn$. The members of $\mathbf{G}(m, n; E)$ are the labelled bipartite graphs that have m vertices in the first part, n vertices in the second part, and precisely E edges, each such graph appearing with probability $\binom{mn}{E}^{-1}$. In order to describe a useful relationship between the two models, we recall that a property Q of graphs is said to be *convex* if, whenever two graphs G_1 and G_2 with the same vertex-set both have property Q, then Q is also possessed by each graph G such that $G_1 \subset G \subset G_2$. Plainly the property of having diameter less than (resp. equal to, greater than) a fixed integer is a convex property of graphs.

The following is a bipartite analogue of part (ii) of Theorem 8 on p. 133 of [2], and its proof is the same; $\lfloor t \rfloor$ denotes the integer floor of t.

5.1. *Suppose that $m(n) \leqq n$ and $0 < p(n) < 1$ for all n, and that $pmn \to \infty$ and $(1-p)mn \to \infty$ so $n \to \infty$. If Q is a convex property of graphs and almost every member of $\mathbf{G}(m, n; p)$ has property Q, then so has almost every member of $\mathbf{G}(m, n; \lfloor pmn \rfloor)$.* ▮

It was conjectured in [9] that if r is a fixed positive integer and the functions $m(n)$ and $E(n)$ are such that

$$\frac{E^{2r-2}}{m^{r-2}n^r} \to 0 \quad \text{and} \quad \frac{E^{2r}}{m^r n^{r+1}} - \log n \to \infty,$$

then almost every member of $\mathbf{G}(m, n; E)$ is of diameter $2r$ or $2r+1$. The following related result is a consequence of our Theorem B.

5.2. Theorem. *Suppose that $E(n) \geqq n(\log n)^4$ for all n, and that k is a fixed positive integer. If k is odd and*

$$\frac{E^k}{m^{(k+1)/2} n^{(k+1)/2}} - \log mn \to \infty$$

or k is even and

$$\frac{E^k}{m^{k/2} n^{(k/2)+1}} - 2 \log n \to \infty$$

then almost every member of $\mathbf{G}(m, n; E)$ is of diameter $\leqq k+1$. If the limits are $-\infty$ in the respective cases then almost every member of $\mathbf{G}(m, n; E)$ is of diameter $\geqq k+2$.

Proof. Let $p(n) = E(n)/mn$, whence (3) holds. Plainly $pmn \to \infty$, and if $(1-p)mn$ does not converge to ∞ an easy direct argument shows almost every member of $\mathbf{G}(m, n; E)$ is of diameter 3. Hence the stated conclusions follow from 5.1 and Theorem B. ▮

References

[1] Z. W. BIRNBAUM, J. D. ESARY and S. C. SAUNDERS, Multicomponent systems and structures and their reliability, *Technometrics* **3** (1961), 55—77.
[2] B. BOLLOBÁS, *Graph Theory*, Springer-Verlag, New York, 1979.
[3] B. BOLLOBÁS, The diameter of random graphs, *Trans. Amer. Math. Soc.* **267** (1981), 41—52.
[4] B. BOLLOBÁS, Counting coloured graphs of high connectivity, *Canad. J. Math.* **33** (1981), 476—484.
[5] B. BOLLOBÁS, *Random Graphs*, in preparation.
[6] L. COMTET, *Advanced Combinatorics*, D. Reidel Publ. Co. Dordrecht, Holland, 1974.
[7] P. ERDŐS and A. RÉNYI, On the strength of connectedness of a random graph, *Acta Math. Acad Sci. Hungar.* **12** (1961), 261—267.
[8] V. L. KLEE and D. G. LARMAN, Diameters of random graphs, *Canad. J. Math.* **33** (1981), 618—640.
[9] V. L. KLEE, D. G. LARMAN and E. M. WRIGHT, The diameter of almost all bipartite graphs, *Studia Sci. Math. Hungar.* **15** (1980), 39—43.
[10] V. L. KLEE, D. G. LARMAN and E. M. WRIGHT, The proportion of labelled bipartite graphs which are connected, *J. London Math. Soc.* **24** (1981), 397—404.
[11] A. D. KORSHUNOV, On the diameter of random graphs, *Soviet Mat. Dokl.* **12** (1971), 302—305.
[12] A. D. KORSHUNOV, Solution of a problem of Erdős and Rényi on Hamilton cycles in nonoriented graphs, *Soviet Mat. Dokl.* **17** (1976), 760—764.
[13] J. W. MOON and L. MOSER, Almost all (0, 1) matrices are primitive, *Studia Sci. Math. Hungar.* **1** (1966), 153—156.
[14] L. PÓSA, Hamiltonian circuits in random graphs, *Discrete Math.* **14** (1976), 359—364.
[15] THANH LE CONG, Estimates of some parameters of finite graphs with applications (in Russian), *Elektr. Inform. Kyber.* **13** (1977), 505—521.

Béla Bollobás

Department of Mathematics
Trinity College
Cambridge CB2 1TQ, England

Victor Klee

Department of Mathematics
University of Washington
GN—50, Seattle WA 98 195, USA