Math. Proc. Camb. Phil. Soc. (1989), **106**, 459
Printed in Great Britain

On the shape of a random acyclic digraph

BY BRENDAN D. McKAY
Computer Science Department, Australian National University, GPO Box 4, ACT 2601, Australia

(*Received* 9 *March* 1988; *revised* 12 *January* 1989)

Abstract

If D is an acyclic digraph, define the height $h = h(D)$ to be the length of the longest directed path in D. We prove that the values of $h(D)$ over all labelled acyclic digraphs D on n vertices are asymptotically normally distributed with mean Cn and variance $C'n$, where $C \approx 0{\cdot}764334$ and $C' \approx 0{\cdot}145210$. Furthermore, define $V_0(D)$ to be the set of sinks (vertices of out-degree 0) and, for $r \geqslant 1$, define $V_r(D)$ to be the set of vertices v such that the longest directed path from v to $V_0(D)$ has length r. For each $k \geqslant 1$, let $n_k(D)$ be the number of sets $V_i(D)$ which have size k. We prove that, for fixed k, the values of $n_k(D)$ over all labelled acyclic digraphs D on n vertices are asymptotically normally distributed with mean $C_k n$ and variance $C'_k n$, for positive constants C_k and C'_k. Results of Bender and Robinson imply that our claim holds also for unlabelled acyclic digraphs.

1. *Introduction*

By an *acyclic digraph* we mean a simple directed graph without directed cycles. Let $\mathscr{A}_n$ be the set of all labelled acyclic digraphs with n vertices. In this paper we will be concerned with the statistical properties of random members of $\mathscr{A}_n$. If $\mathscr{A} = \bigcup_{n=1}^{\infty} \mathscr{A}_n$ and $f: \mathscr{A} \to \mathbb{R}$, we say that f is asymptotically normal over $\mathscr{A}$ with mean $\mu = \mu(n)$ and variance $\sigma^2 = \sigma^2(n)$ if

$$\lim_{n\to\infty} \sup_x \left| P(n, \mu+\sigma x, f) - \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \right| = 0,$$

where
$$P(n,z,f) = \frac{|\{D \in \mathscr{A}_n \,|\, f(D) \leqslant z\}|}{|\mathscr{A}_n|}.$$

Our fundamental tool for proving the normality of statistics of $\mathscr{A}$ will be theorem 1 of Bender [**1**]. While it should be possible to prove stronger local limit theorems for the same quantities, using theorems 3 or 4 of [**1**], verification of the additional requirements does not appear simple.

Previous results about $\mathscr{A}$ have been obtained by many authors. The generating function and asymptotic value of $A_n = |\mathscr{A}_n|$ were obtained independently by Robinson and Stanley. Define

$$\alpha(z) = \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{q(n)},$$

where $q(n) = n!\,2^{\binom{n}{2}}$. Clearly $\alpha(z)$ is an entire function. It is shown in [**6**] and [**7**] that the zeros of $\alpha(z)$ are all real and positive. Let $\rho \approx 1{\cdot}4880785456\ldots$ be the least zero; the next two are approximately 4·881 and 13·56.

THEOREM 1·1. (i) $\sum_{n=0}^{\infty} \frac{A_n x^n}{q(n)} = \frac{1}{\alpha(x)}$.

(ii) *As* $n \to \infty$, $A_n \sim \frac{q(n)}{\rho^{n+1}\alpha(\rho/2)}$.

Proof. See [6] or [7]. ∎

Theorem 1·1 was extended by Bender, Richmond, Robinson and Wormald [2] to include the dependence on the number of edges.

If $D \in \mathscr{A}$, let $V_0 = V_0(D)$ be the set of sinks of D. Then, for $k = 1, 2, \ldots$, define $V_k = V_k(D)$ to be the set of sinks of the subgraph of D induced by $V(D)\backslash(V_0 \cup V_1 \cup \ldots \cup V_{k-1})$. Let V_h be the last V_i which is non-empty. It is easy to see that $h = h(D)$ is the length of the longest directed path in D and that $V_0, V_1, \ldots, V_h$ are the sets of the same name defined in the Abstract. We will call $h(D)$ the *height* of D, $(V_0, V_1, \ldots, V_h)$ the *tower* of D, and $V_0, V_1, \ldots, V_h$ the *layers* of D.

The limiting distribution of sizes of $V_0(D)$ was determined by Liskovec [5].

THEOREM 1·2. *For* $n \geq 0, k \geq 1$, *define* $p(n,k)$ *to be the probability that a random* $D \in \mathscr{A}_n$ *has* $|V_0(D)| = k$. *Then, as* $n \to \infty$ *with fixed* k,

$$p(n,k) \to S_k = \frac{\rho^k \alpha(2^{-k}\rho)}{q(k)}. \quad ∎$$

It is proved in [2] that the average height lies eventually in an interval $[c_1 n, c_2 n]$ for some constants c_1 and c_2 with $0 < c_1 < c_2 < 1$. Apart from this, nothing appears to have been previously known about the height, although Liskovec (incorrectly) predicted an average close to n/ρ.

2. *A generating function*

For each $n \geq 1$ and $D \in \mathscr{A}$, let n_i be the number of layers of D which have size i. Define

$$a(x, (y_1, y_2, \ldots)) = \sum_{n\geq 1} \sum_{D \in \mathscr{A}_n} q(n)^{-1} x^n y_1^{n_1(D)} y_2^{n_2(D)} \ldots.$$

THEOREM 2·1.

$$a(x, (y_1, y_2, \ldots)) = \sum_{v_1, v_1, \ldots, v_h} \prod_{i=0}^{h-1} (1 - 2^{-v_i})^{v_{i+1}} \prod_{i=0}^{h} \frac{x^{v_i} y_{v_i}}{q(v_i)},$$

where the sum is over all vectors $(v_0, v_1, \ldots, v_h)$ *such that* $h \geq 0$ *and* $v_i \geq 1$ *for* $0 \leq i \leq h$.

Proof. Consider a particular $(v_0, v_1, \ldots, v_h)$. We will count the number of acyclic digraphs D with height h and $|V_i(D)| = v_i$ for $0 \leq i \leq h$. Let $n = v_0 + v_1 + \ldots + v_h$. The number of ways of assigning vertex labels to the layers $V_0, V_1, \ldots, V_h$ is

$$\binom{n}{v_0, v_1, \ldots, v_h}.$$

Each vertex of layer V_i can be adjacent to any subset of $V_0 \cup V_1 \cup \ldots \cup V_{i-1}$ which contains at least one element of V_{i-1}; the number of possibilities is clearly

$2^{v_0+v_1+\ldots+v_{i-2}}(2^{v_{i-1}}-1)$. Therefore the total number of digraphs corresponding to $(v_0, v_1, \ldots, v_h)$ is

$$\binom{n}{v_0, v_1, \ldots, v_h} \prod_{i=1}^{h} (2^{v_0+\ldots+v_{i-1}})^{v_i}(1-2^{-v_{i-1}})^{v_i},$$

from which the theorem readily follows. ∎

In order to write $a(x, \mathbf{y})$ in a more manageable form, we resort to a matrix representation. Define the infinite matrices

$$\Lambda = \operatorname{diag}\left(xy_1, \frac{x^2y_2}{4}, \ldots, \frac{x^iy_i}{q(i)}, \ldots\right),$$

$$M = (m_{ij}), \quad \text{where } m_{ij} = (1-2^{-i})^j \text{ for } 1 \leqslant i,j < \infty.$$

Let I be the infinite identity matrix. For any matrix M, let $\det(M)$ denote the determinant of M, $\operatorname{adj}(M)$ the adjoint matrix of M, and $\mathscr{S}(M)$ the sum of all the entries of M.

THEOREM 2·2. $a(x, (y_1, y_2, \ldots)) = \mathscr{S}((I-\Lambda M)^{-1}\Lambda)$.

Proof. It is easily seen that the sum of the terms in Theorem 2·1 corresponding to any particular value of h is $\mathscr{S}((\Lambda M)^h\Lambda)$. Thus

$$a(x, (y_1, y_2, \ldots)) = \mathscr{S}(\Lambda + \Lambda M\Lambda + (\Lambda M)^2\Lambda + \ldots) = \mathscr{S}((I-\Lambda M)^{-1}\Lambda). \quad ∎$$

At this point we should note that we have shown Theorem 2·2 to be correct only in the space of formal generating functions. However, if we write

$$(I-\Lambda M)^{-1} = \operatorname{adj}(I-\Lambda M)/\det(I-\Lambda M),$$

it can be shown (using the fact that the entries of Λ decrease very rapidly down the diagonal) that each entry of $\operatorname{adj}(I-\Lambda M)$, their sum, and $\det(I-\Lambda M)$ are entire functions of x for any fixed and uniformly bounded values of $y_1, y_2, \ldots$. The same is true after any finite number of differentiations by any of the y_i separately or by all the y_i simultaneously.

3. *Extraction of the asymptotics*

Armed with the generating function of the previous section, we can now derive the promised limit theorems.

THEOREM 3·1. *There are constants* $C \approx 0{\cdot}7643344264$ *and* $C' \approx 0{\cdot}1452097407$ *such that the height function h is asymptotically normal over* $\mathscr{A}$ *with mean* Cn *and variance* $C'n$.

Proof. Define

$$H(x, y) = \sum_{n \geqslant 1} \sum_{D \in \mathscr{A}_n} \frac{x^n y^{h(D)}}{q(n)}.$$

By Theorem 2·2,

$$H(x, y) = \frac{c(x, y)}{yd(x, y)},$$

where $\quad c(x, y) = \mathscr{S}(\operatorname{adj}(I-\Lambda M)\Lambda) \quad \text{and} \quad d(x, y) = \det(I-\Lambda M)$

with $\Lambda = \Lambda(x, (y, y, \ldots))$. For reasons noted at the end of the previous section, $c(x, y)$ and $d(x, y)$ are entire functions of x for fixed y, so the singularities of $H(x, y)$

correspond to zeros of $d(x,y)$ which are not cancelled by zeros of $c(x,y)$. ($c(x,y)$ has a zero at $x=0$ which cancels the y in the denominator of H.) Since $H(x,1) = 1/\alpha(x)$, and the least-modulus zero of the entire function $\alpha(x)$ is a simple pole at ρ, we can apply theorem 1 of [1] with $m=0$ and a function $r(s)$ with $r(0)=\rho$. (See [1] for the meanings of m and $r(s)$.) The derivatives $r'(0)$ and $r''(0)$ can be found using formula (3·2) of [1], with $d(x,e^s)$ playing the part of $h(z,e^s)$; this only requires computing the values of the first and second order derivatives of $d(x,y)$ at the point $(x,y)=(\rho,1)$. The most computationally successful approach tried was to compute the necessary derivatives for the leading $k\times k$ minor of $I-\Lambda M$ for increasing k. Convergence to six digits only required $k=5$; $k=10$ gave better than 20 digit accuracy. This computation was done using the symbolic algebra package Maple, although it could be done using conventional numerical methods. ▮

Theorem 3·1 could also be derived directly from the generating function. In our next theorem, we will improve the estimates of the mean and variance. We begin with a lemma of more general application.

LEMMA 3·1. *Let $F(x,y)=\sum_{n,k\geqslant 0} a_{kn}x^n y^k$, where each a_{kn} is real and non-negative. Suppose that there are real numbers $0<\rho<\rho'$ such that $d(x,y)=1/F(x,y)$ satisfies the following conditions:*

(i) *$d(x,y)$ and all its derivatives of order three or less are analytic in x for $|x|\leqslant\rho'$ and $y=1$;*

(ii) *$d(x,1)$ is non-zero for $|x|\leqslant\rho'$ except for a simple zero at $x=\rho$.*

For each n, define a discrete random variable X_n with probability generating function formed by normalizing the coefficient of x^n in $F(x,y)$. Then, as $n\to\infty$, X_n has mean $An+A'+O((\rho/\rho')^n)$ and variance $Bn+B'+O((\rho/\rho')^n)$, where A, A', B and B' are given by the following expressions, in which subscripts denote differentiation and missing arguments are $(\rho,1)$:

$$A=\frac{d_y}{\rho d_x};$$

$$A'=A+\frac{d_y d_{xx}}{d_x^2}-\frac{d_{xy}}{d_x};$$

$$B=\frac{d_y}{\rho d_x}-\frac{2d_y d_{xy}}{\rho d_x^2}+\frac{d_y^2}{\rho^2 d_x^2}+\frac{d_{yy}}{\rho d_x}+\frac{d_y^2 d_{xx}}{\rho d_x^3};$$

$$B'=B+\frac{2d_y d_{xxy}}{d_x^2}-\frac{d_y^2 d_{xxx}}{d_x^3}+\frac{d_y d_{xx}}{d_x^2}-\frac{d_{xy}}{d_x}+\frac{2d_y^2 d_{xx}^2}{d_x^4}+\frac{d_{xx}d_{yy}}{d_x^2}+\frac{d_{xy}^2}{d_x^2}-\frac{d_{xyy}}{d_x}-\frac{4d_y d_{xx} d_{xy}}{d_x^3}.$$

Proof. The mean is derived as theorem C of [4]. The variance can be derived in precisely the same way. Higher moments could be computed similarly. To reduce the probability of error, these calculations were performed using the symbolic algebra package Maple. ▮

When Lemma 3·1 is applied to the function $H(x,y)$ of Theorem 3·1, the following result is obtained.

THEOREM 3·2. *Let $\mu(n)$ and $\sigma^2(n)$ be respectively the mean and variance of the height*

Table 1

k	C_k	C'_k	S_k
1	0·5478347055	0·4909736925	0·5743623733
2	0·1979192508	0·1124338637	0·3662136732
3	0·0180023617	0·0172276787	0·0564645435
4	$5·70873 \times 10^{-4}$	$5·70569 \times 10^{-4}$	$2·90231 \times 10^{-3}$
5	$7·19708 \times 10^{-6}$	$7·19721 \times 10^{-6}$	$5·66517 \times 10^{-5}$
6	$3·85217 \times 10^{-8}$	$3·85217 \times 10^{-8}$	$4·49589 \times 10^{-7}$
7	$9·06982 \times 10^{-11}$	$9·06982 \times 10^{-11}$	$1·51097 \times 10^{-9}$
8	$9·58495 \times 10^{-14}$	$9·58495 \times 10^{-14}$	$2·20861 \times 10^{-12}$
9	$4·60293 \times 10^{-17}$	$4·60293 \times 10^{-17}$	$1·43063 \times 10^{-15}$

$h(D)$ over $D \in \mathscr{A}_n$. Then, as $n \to \infty$, $\mu(n) = Cn + E + o(3^{-n})$ and $\sigma^2(n) = C'n + E' + o(3^{-n})$, where

$$C \approx 0·7643344264, \quad E \approx -0·8688716771,$$

$$C' \approx 0·1452097407, \quad E' \approx -0·0989894976. \quad \blacksquare$$

THEOREM 3·3. *Fix $k \geqslant 1$. For $D \in \mathscr{A}$, define $n_k(D)$ to be the number of layers of D with size k. Then there are positive constants C_k and C'_k such that the function n_k is asymptotically normal over $\mathscr{A}$ with mean $C_k n$ and variance $C'_k n$.*

Proof. Define

$$N_k(x, y) = \sum_{n \geqslant 1} \sum_{D \in \mathscr{A}_n} \frac{x^n y^{n_k(D)}}{q(n)}.$$

Then

$$N_k(x, y) = \frac{e(x, y)}{f(x, y)},$$

where

$$e(x, y) = \mathscr{S}(\operatorname{adj}(I - \Lambda M)\Lambda), \quad f(x, y) = \det(I - \Lambda M),$$

with $\Lambda = \Lambda(x, (1, 1, \ldots, y, 1, 1, \ldots))$, where the '$y$' is in the kth entry of the second argument. The proof now proceeds exactly as that of Theorem 3·1. $\blacksquare$

It has been proved by Bender and Robinson [3] that almost all members of $\mathscr{A}$ have trivial automorphism groups and are weakly connected. It follows that Theorems 3·1 and 3·3 also hold for the classes of labelled weakly connected acyclic digraphs, unlabelled acyclic digraphs, and unlabelled weakly connected acyclic digraphs.

Some approximate values of C_k and C'_k are given in Table 1. As checks, note that $\Sigma_{k=0}^{\infty} C'_k = C$ and $\Sigma_{k=0}^{\infty} kC_k = 1$. We also give some approximate values of S_k, as defined in Theorem 1·2. It is worth noting here that the values of C_k given in Table 1 demonstrate that typical layers do not have the same distribution of sizes as does the bottom layer, disproving a conjecture made by Liskovec [5]. Similarly, the average height is quite different from n/ρ.

Finally, we present some counts of acyclic digraphs by height. The number shown at position (n, h) in Table 2 is $a_{n,h}/((h+1)!\,2^{\binom{h}{2}})$, where $a_{n,h}$ is the number of labelled acyclic digraphs of order n and height h. This number is easily shown to be an integer.

Table 2

h	n = 1	2	3	4	5	6	7	8	9	10	11
0	1	1	1	1	1	1	1	1	1	1	1
1		1	6	43	420	5821	113946	3142903	121796520	6631278361	507233141646
2			1	22	610	25355	1634801	165904172	26641660900	6765756939985	2713383541189651
3				1	70	6050	805700	175178976	63165264960	37991918437600	38228553440862400
4					1	204	50288	19044480	11971739136	12852596994048	23795321191223296
5						1	560	371840	376406016	633769500672	1843282177228800
6							1	1472	2528256	6563880960	28281544507392
7								1	3744	16135680	104255078400
8									1	9280	97996800
9										1	22528
10											1

REFERENCES

[1] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combin. Theory Ser. A* **15** (1973), 91–111.

[2] E. A. Bender, L. B. Richmond, R. W. Robinson and N. C. Wormald. The asymptotic number of acyclic digraphs. I. *Combinatorica* **6** (1986), 15–22.

[3] E. A. Bender and R. W. Robinson. The asymptotic number of acyclic digraphs. II. *J. Combin. Theory Ser. B* **44** (1988), 363–369.

[4] B. E. Eichinger, D. M. Jackson and B. D. McKay. Generating function methods for macromolecules at surfaces. I. One molecule at a plane surface. *J. Chem. Phys.* **85** (1986), 5299–5305.

[5] V. A. Liskovec. The number of maximal vertices of a random acyclic digraph. *Teor. Veroyatnost. i Primenen.* **20** (1975), 412–421; *Theory Probab. Appl.* **20** (1975), 401–409.

[6] R. W. Robinson. Counting labelled acyclic digraphs. In *New Directions in Graph Theory* (Academic Press, 1973), pp. 239–273.

[7] R. P. Stanley. Acyclic orientations of graphs. *Discrete Math.* **5** (1973), 171–178.