ASYMPTOTIC ENUMERATION OF EXTENSIONAL ACYCLIC DIGRAPHS

STEPHAN WAGNER

Abstract. The enumeration of extensional acyclic digraphs, which have the property that the outneighbourhoods are pairwise distinct, was considered in a recent article of Policriti and Tomescu. Several asymptotic questions were left as open problems. In this article, we determine the asymptotic number of such digraphs and show that a number of distributional results can be carried over from ordinary acyclic digraphs. In particular, we consider the distribution of the number of sources, the number of arcs, the maximum rank and the number of vertices of maximum rank, thereby also proving some conjectures made by Policriti and Tomescu. Finally, we study a very similar class of acyclic digraphs and provide analogous distributional results. Extensional acyclic digraphs and Essential acyclic digraphs and Full sets and Asymptotic enumeration and Limit theorems 05C30 and 05A16

1. Introduction

Acyclic digraphs, i.e., digraphs without an oriented cycle, are the directed analogue of trees. Their enumeration is explained very nicely in Section 1.6 of Harary and Palmer's book [8], following the work of Robinson [13]. The counting sequence for labelled acyclic digraphs starts with the terms

$$1, 3, 25, 543, 29281, 3781503, 1138779265, \ldots \tag{1}$$

which is Sloane's A003024 [1]. One sees that the numbers grow quite rapidly, and indeed it turns out that they behave asymptotically like $A \cdot B^n \cdot n! \cdot 2^{\binom{n}{2}}$ for certain constants A and B. The approach that leads to this result will be outlined briefly below, as it will also be relevant for our purposes. It is one of the nicest applications of generating functions and singularity analysis to a graph enumeration problem.

In this paper, we will mostly be concerned with a special subfamily of acyclic digraphs: *extensional acyclic digraphs* (EADs in the following) were studied recently by Policriti and Tomescu [12]. In addition to being acyclic, they have the property that the outneighbourhoods are pairwise distinct. These graphs occur naturally in set theory: a *full* (*transitive*) *set* is a set X with the property that every element of X is also a subset of X. For example, $X = \{\{\}, \{\{\}\}\}$ is such a set. Every full set corresponds bijectively to an acyclic digraph, where the vertices represent sets and the directed edges (arcs) stand for set membership (i.e., an arc from u to v indicates that v is an element of u). It is easy to see that the outneighbourhood of a vertex uniquely characterises the corresponding element of the transitive set, and so the outneighbourhoods of distinct vertices have to be distinct (see the original article of Peddicord [11] for further background on full sets). An example of an EAD is

This material is based upon work supported by the National Research Foundation of South Africa under grant number 70560.

shown in Figure 1: it corresponds to the full set $\{\{\},\{\{\}\},\{\{\{\}\}\},\{\{\},\{\{\}\}\}\}$, and the correspondence between vertices and set elements is indicated as well.

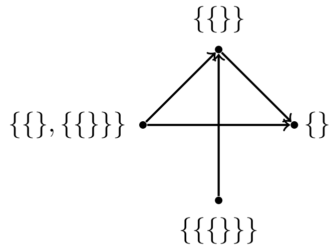

Figure 1. An example of an extensional acyclic digraph.

Policriti and Tomescu provide recursive formulas for determining the number of EADs of given order as well as the number of such digraphs with additional restrictions on the number of sources or the number of vertices of maximum rank (where the rank of a vertex v, in graph-theoretical terms, is the length of a longest directed path starting at v). Moreover, they study *weakly extensional acyclic digraphs* (WEADs), for which the outneighbourhood condition is slightly relaxed: only the nonempty outneighbourhoods have to be pairwise distinct, thus allowing several sinks (corresponding to additional "atoms" in the set construction). At the end of their paper, they remark that the asymptotic enumeration is left as an open problem, and formulate a number of conjectures. In particular, they conjecture (based on numerical values) that the proportion of EADs among all labelled acyclic digraphs converges to a limit, which is roughly 32.6%. This fact will be one of the main results of the present paper. The number of labelled EADs of order 1 to 7 is given by

$$1, 2, 12, 216, 10560, 1297440, 381013920,$$

compare with (1). The numbers in the unlabelled case, $1, 1, 2, 9, \dots$ (Sloane's A001192), which are simply obtained by dividing by the factorials, also enumerate full sets. Aside from determining the asymptotic density of EADs within the class of acyclic digraphs, we will also prove some distributional results: specifically, we consider the distribution of the number of sources and sinks (the latter only in the case of WEADs, which can have several sinks as opposed to EADs), the number of arcs and the maximum rank (length of the longest directed path). Finally, we also briefly show how analogous results can be obtained for *essential acyclic digraphs*, which were studied in [14, 15]. The following table summarises our results:

This is an extended version of the conference paper [16]. It contains more detailed proofs as well as some additional material, in particular Section 8.

2. Preliminaries

In order to understand the methods of the following sections, it is useful to briefly review the approach of Robinson [13] that leads to the aforementioned asymptotic formula for the number of ordinary labelled acyclic digraphs. Let a_n be the number of acyclic digraphs on n labelled vertices. The number of such digraphs for which a

Asymptotic number of EADs and WEADs of order n	Theorem 1 (EADs), Theorem 6 (WEADs)
Limit distribution of the number of sources (discrete distribution)	Theorem 2 (EADs), Theorem 9 (WEADs), Theorem 13 (EssADs)
Central limit theorem for the number of arcs (Gaussian distribution)	Theorem 3 (EADs), Theorem 7 (WEADs), Theorem 11 (EssADs)
Central limit theorem for the maximum rank (= length of the longest directed path, Gaussian distribution)	Theorem 4 (EADs), Theorem 8 (WEADs), Theorem 12 (EssADs)
Limit distribution of the number of vertices of highest rank (discrete distribution)	Theorem 5 (EADs), Theorem 9 (WEADs)

TABLE 1. Summary of results.

given set of k vertices are sources is $2^{k(n-k)}a_{n-k}$, so the inclusion-exclusion principle yields

$$a_n = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} 2^{k(n-k)} a_{n-k}$$

for $n \geq 1$, with initial value $a_0 = 1$. This can be rewritten as

$$\sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} 2^{k(n-k)} a_k = [n = 0],$$

using Iverson's notation $[P] = 1$ if P is true, and $[P] = 0$ otherwise. Divide by $n!2^{\binom{n}{2}}$ to get

$$\sum_{k=0}^{n} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \frac{a_k}{k!} = [n = 0].$$

Let us now introduce the "special" generating functions

$$A(x) = \sum_{n \geq 0} \frac{a_n x^n}{n!} 2^{-\binom{n}{2}}$$

and

$$\phi(x) = \sum_{n \geq 0} \frac{(-1)^n x^n}{n!} 2^{-\binom{n}{2}}.$$

Then the above identity translates to

$$A(x)\phi(x) = 1$$

or

$$A(x) = \frac{1}{\phi(x)}.$$

Note that $\phi(x)$ is an entire function. It is known [13] that its first zero from the origin is $z_0 \approx 1.488079$, and the second is $z_1 \approx 4.881141$. Indeed, the zeros of this function have been studied quite thoroughly on their own right [7].

We deduce that $A(x)$ is meromorphic inside the circle of radius z_1, with a single pole at z_0. So we can apply singularity analysis, see for instance [5, Theorem IV.10]:

if a function $f(x)$ is meromorphic on a disk $|x| \leq R$ with only a single pole at $x = z_0$ and $f(x) \sim \frac{C}{1-x/z_0}$ as $x \to z_0$, then

$$[x^n]f(x) = Cz_0^{-n} + O(R^{-n}).$$

Hence we obtain

$$a_n \sim \frac{n! \cdot 2^{\binom{n}{2}}}{-\phi'(z_0)z_0^{n+1}}.$$

The aim of the following section is to transfer this approach to the asymptotic enumeration of EADs to prove the conjecture of Policriti and Tomescu on the limit proportion of EADs among acyclic digraphs.

3. The number of extensional acyclic digraphs

We try the same approach as outlined in the introduction with the number of extensional acyclic digraphs, which we denote by b_n. Similar ideas were also used in [15] for the asymptotic enumeration of so-called essential acyclic digraphs, to which we will come back later. The inclusion-exclusion argument now yields

$$b_n = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (2^{n-k} - n + k)^{\underline{k}} b_{n-k} \tag{2}$$

with $b_0 = 1$, where $x^{\underline{n}} = x(x-1)\cdots(x-n+1)$ denotes a falling factorial. This recursion was also given by Policriti and Tomescu [12] in their paper. Note that $2^{n-k} - n + k$ is exactly the number of subsets of $n-k$ vertices, without the outneighbourhoods of these vertices (since they cannot be outneighbourhoods of the new vertices). Moreover, we have to take a falling factorial rather than a k-th power, since the new outneighbourhoods have to be pairwise distinct. The recursion (2) can also be written as

$$\frac{b_n}{n!} = \sum_{k=0}^{n-1} (-1)^{n-k-1} \binom{2^k - k}{n-k} \frac{b_k}{k!} + [n = 0]$$

or, which is best for our purposes, as

$$\sum_{k=0}^{n} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2} - \binom{n-k}{2}} \frac{b_k}{k!} \prod_{j=k}^{n-1} (1 - 2^{-k}j) = [n = 0].$$

The product $\prod_{j=k}^{n-1}(1 - 2^{-k}j)$ is "almost 1": we rewrite the equation once again to obtain

$$\sum_{k=0}^{n} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2} - \binom{n-k}{2}} \frac{b_k}{k!}$$
$$= \sum_{k=0}^{n-1} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2} - \binom{n-k}{2}} \frac{b_k}{k!} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k}j)\right) + [n = 0].$$

In terms of the generating function $B(x) = \sum_{n \geq 0} 2^{-\binom{n}{2}} \frac{b_n x^n}{n!}$, this becomes

$$B(x)\phi(x) = 1 + \sum_{n \geq 0} \sum_{k=0}^{n-1} \frac{(-1)^{n-k} b_k}{k!(n-k)!} 2^{-\binom{k}{2} - \binom{n-k}{2}} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k}j)\right) x^n.$$

Now we claim that the sum on the right hand side converges for $|x| < 2z_0$, which implies that it represents a holomorphic function (which we denote by $\psi(x)$) inside this circle. To this end, note that

$$b_k \le a_k \le C_1 k! 2^{\binom{k}{2}} z_0^{-k}$$

for some constant C_1, so that

$$\left| \sum_{k=0}^{n-1} \frac{(-1)^{n-k} b_k}{k!(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k} j)\right) x^n \right| \\ \le C_1 \cdot \sum_{k=0}^{n-1} 2^{-\binom{n-k}{2}} z_0^{-k} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k} j)\right) |x|^n.$$

Now we use the simple inequality

$$\prod_{j=k}^{n-1} (1 - 2^{-k} j) \ge 1 - \sum_{j=k}^{n-1} 2^{-k} j \ge 1 - n^2 2^{-k},$$

which yields

$$\sum_{k=0}^{n-1} 2^{-\binom{n-k}{2}} z_0^{-k} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k} j)\right) |x|^n \\ \le n^2 |x|^n \sum_{k=0}^{n-1} 2^{-\binom{n-k}{2}} (2z_0)^{-k}.$$

The product $2^{-\binom{n-k}{2}} (2z_0)^{-k}$ has its maximum at $k = n - \frac{3}{2} - \log_2(z_0)$, with a value of $C_2(2z_0)^{-n}$ for some constant C_2. Hence we end up with

$$\left| \sum_{k=0}^{n-1} \frac{(-1)^{n-k} b_k}{k!(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \left(1 - \prod_{j=k}^{n-1} (1 - 2^{-k} j)\right) x^n \right| \\ \le C_1 C_2 n^3 \left(\frac{|x|}{2z_0}\right)^n.$$

This proves that the series indeed converges (absolutely) for $|x| < 2z_0$, which shows that the function ψ in the equation

$$B(x)\phi(x) = 1 + \psi(x)$$

is holomorphic within this region. Hence $B(x)$ is meromorphic inside the open disk $|x| < 2z_0$, except for a simple pole at z_0. We apply singularity analysis again to obtain

$$b_n \sim \frac{n! \cdot 2^{\binom{n}{2}} (1 + \psi(z_0))}{-\phi'(z_0) z_0^{n+1}} = \alpha \cdot \beta^n \cdot n! \cdot 2^{\binom{n}{2}}$$

with $\alpha = -(1 + \psi(z_0))/(z_0 \phi'(z_0)) \approx 0.567952$ and $\beta = z_0^{-1} \approx 0.672008$. Hence the limit $\lim_{n\to\infty} a_n/b_n$ is $1/(1 + \psi(z_0)) \approx 3.065509$. The error term can be described by an additional factor $1 + O(\gamma^{-n})$ for any fixed $\gamma < 2$, since the next singularity has absolute value at least $2z_0$ (note that the second-smallest zero of ϕ is greater than $2z_0$ as well). Let us formulate this as a theorem:

Theorem 1. *The proportion of EADs among all labelled acyclic digraphs converges to*

$$1+\psi(z_0) = 1 + \sum_{n\geq 0}\sum_{k=0}^{n-1} \frac{(-1)^{n-k} b_k}{k!(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \left(1 - \prod_{j=k}^{n-1}(1-2^{-k}j)\right) z_0^n.$$

The numerical value of this limit is approximately $3.065509^{-1} \approx 0.326210$.

Remark 1. *EADs have the nice property that their automorphism group is always trivial, as was shown in* [12]. *This means that the number of unlabelled EADs is simply* $b_n/n!$. *An asymptotic formula follows automatically. Since almost all acyclic digraphs have trivial automorphism group, as shown by Bender and Robinson* [4], *Theorem 1 holds for unlabelled EADs as well.*

4. Extensional acyclic digraphs by number of sources

The limiting distribution of the number of sources is relatively easy to obtain, since the sources also play an essential role in our recursive approach. The inclusion-exclusion principle shows that the number of labelled EADs with exactly ℓ sources, which we denote by $b_{n,\ell}$, is given by

$$b_{n,\ell} = \sum_{k\geq \ell} \binom{k}{\ell} (-1)^{k-\ell} \binom{n}{k} (2^{n-k}-n+k)^{\underline{k}} b_{n-k}.$$

The limit $\lim_{n\to\infty} b_{n,\ell}/b_n$ can now be determined directly from the asymptotic formula for b_n: recall that

$$b_n = \alpha\cdot\beta^n\cdot n!\cdot 2^{\binom{n}{2}}(1+O(\gamma^{-n}))$$

for any fixed $\gamma < 2$. If one combines this with the formula $(2^{n-k}-n+k)^{\underline{k}} = 2^{k(n-k)}(1+O(kn2^{k-n}))$, one obtains

$$b_{n,\ell} \sim \sum_{k\geq \ell} \binom{k}{\ell}(-1)^{k-\ell}\binom{n}{k} 2^{k(n-k)}\cdot\alpha\cdot\beta^{n-k}\cdot(n-k)!\cdot 2^{\binom{n-k}{2}} \sim S_\ell b_n,$$

where the constant S_ℓ is given by

$$\begin{aligned} S_\ell &= \frac{1}{\ell!}\sum_{k\geq\ell}\frac{(-1)^{k-\ell}}{(k-\ell)!}\cdot\beta^{-k}\cdot 2^{-\binom{k}{2}} \\ &= \frac{\beta^{-\ell}}{\ell! 2^{\binom{\ell}{2}}}\sum_{m\geq 0}\frac{(-1)^m\beta^{-m}}{m!}2^{-\binom{m}{2}-m\ell} = \frac{z_0^\ell \phi(2^{-\ell}z_0)}{\ell! 2^{\binom{\ell}{2}}}. \end{aligned}$$

Let us formulate this as a theorem:

Theorem 2. *The distribution of the number of sources in a random EAD converges to a discrete limiting distribution, the limit probability that the number of sources equals* ℓ *is*

$$S_\ell = \frac{z_0^\ell\phi(2^{-\ell}z_0)}{\ell! 2^{\binom{\ell}{2}}}.$$

Numerically,

$$\begin{aligned} &S_1 \approx 0.574362,\ S_2 \approx 0.366214,\ S_3\approx 0.056465,\\ &S_4\approx 0.002902,\ S_5\approx 0.000057,\ldots. \end{aligned}$$

In particular, the ratio between the number of EADs with exactly one source and the number of all EADs tends to a value of approximately $0.574363^{-1} \approx 1.741061$, as conjectured by Policriti and Tomescu. Compare also Table 2, which contains the values of $b_{n,\ell}$ for small n. The same limiting distribution was determined by Liskovets [9] for ordinary acyclic digraphs as well. It is worth pointing out that the limit of the average number of sources is precisely $\beta^{-1} = z_0 \approx 1.488079$, the first zero of the function ϕ, and that moreover the probability generating function of the S_ℓ is

$$\sum_{\ell=1}^{\infty} S_\ell u^\ell = \phi((1-u)z_0),$$

as can be seen by some elementary manipulations. Since $\phi(0) = 1$, we also have $S_\ell \sim z_0^\ell 2^{-\binom{\ell}{2}}/\ell!$, i.e., the probabilities decay very rapidly.

n	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
1	1			
2	2			
3	12			
4	192	24		
5	8160	2400		
6	898560	384480	14400	
7	245145600	126040320	9777600	50400

TABLE 2. Enumeration of labelled EADs with given number of sources.

5. The number of arcs

The number of arcs can be analysed by means of a bivariate generating function, as it was done by Bender, Richmond, Robinson and Wormald [3] for ordinary acyclic digraphs: denote the set of EADs with n labelled vertices by $\mathcal{B}_n$. Furthermore, let $B_n(y)$ be the polynomial in which the coefficient of y^r is the number of EADs with n (labelled) vertices and r arcs. For instance, $B_1(y) = 1$, $B_2(y) = 2y$ and $B_3(y) = 6(y^2 + y^3)$. The recursion (2) does not have an exact counterpart, since the number of vertices in the "forbidden" $n-k$ outneighbourhoods is not known when the new set of k sources is added. However, an asymptotic analysis is still possible, and we will be able to prove a central limit theorem. For any $B \in \mathcal{B}_n$, we define $\|B\|$ to be the number of arcs of B and

$$P_{B,k}(y) = \sum_{U_1,U_2,\ldots,U_k} \prod_{j=1}^{k} y^{|U_j|},$$

where the sum is over all k-tuples of pairwise distinct vertex subsets of B none of which is an outneighbourhood in B. Now the inclusion-exclusion argument that led to (2) yields

$$B_n(y) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \sum_{B \in \mathcal{B}_{n-k}} y^{\|B\|} P_{B,k}(y).$$

It is possible to express $P_{B,k}$ in terms of the outdegree sequence of B by means of cycle indices, but for our purpose it is sufficient to work with crude estimates:

assume that y lies in a fixed interval $[\delta, \delta^{-1}]$ around 1. In analogy to the estimate in Section 3, we have

$$(1+y)^{k(n-k)} \geq P_{B,k}(y)$$

and

$$\begin{aligned} P_{B,k}(y) &\geq \prod_{j=n-k}^{n-1} \Big((1+y)^{n-k} - j \max(1,y)^{n-k} \Big) \\ &\geq \Big((1+y)^{n-k} - n \max(1,y)^{n-k} \Big)^k \\ &\geq (1+y)^{k(n-k)} \left(1 - n^2 \left(\frac{\max(1,y)}{1+y} \right)^{n-k} \right) \end{aligned}$$

and thus

$$\Big| 1 - (1+y)^{-k(n-k)} P_{B,k}(y) \Big| \leq n^2 \left(\frac{\max(1,y)}{1+y} \right)^{n-k}.$$

Now define

$$\phi(x,y) = \sum_{n \geq 0} \frac{(-1)^n x^n}{n!} (1+y)^{-\binom{n}{2}}.$$

Analogous reasoning as in the proof of Theorem 1 shows that

$$B(x,y) = \sum_{n \geq 0} \frac{B_n(y) x^n}{n! (1+y)^{\binom{n}{2}}} = \frac{1 + \psi(x,y)}{\phi(x,y)},$$

where

$$\psi(x,y) = \sum_{n \geq 0} \sum_{k=0}^{n-1} \frac{(-1)^{n-k}}{k!(n-k)!} (1+y)^{-\binom{k}{2} - \binom{n-k}{2}} \sum_{B \in \mathcal{B}_k} y^{\|B\|} \left(1 - \frac{P_{B,n-k}(y)}{(1+y)^{k(n-k)}} \right) x^n.$$

As a result of the estimates for $P_{B,k}(y)$ above, we find that the radius of convergence of $\psi(x,y)$ (in x) is strictly greater than that of $B(x,y)$, which is the first zero $\rho(y)$ of the function $\phi(x,y)$. Now singularity analysis yields

$$B_n(y) \sim \frac{n! \cdot (1+y)^{\binom{n}{2}} (1 + \psi(\rho(y), y))}{-\phi_x(\rho(y), y) \rho(y)^{n+1}},$$

from which it is easy to deduce that the limiting distribution of the number of arcs is Gaussian (cf. [5, Theorem IX.8], although the quasi-powers theorem does not apply directly): let the random variable ω_n be the number of arcs in a random EAD of order n. Then the moment generating function of ω_n is

$$\mathbb{E}(e^{\omega_n t}) = \frac{B_n(e^t)}{B_n(1)},$$

and the moment generating function of the renormalised random variable

$$\varpi_n = \frac{\omega_n - \frac{1}{2}\binom{n}{2}}{\frac{1}{2}\binom{n}{2}^{1/2}}$$

is given by

$$\mathbb{E}(e^{\varpi_n t}) = \exp\left(-\binom{n}{2}^{1/2} t \right) \frac{B_n(\exp(y(t)))}{B_n(1)}$$

with $y(t) = 2t\binom{n}{2}^{-1/2}$. One easily finds

$$\left(\frac{1+\exp(y(t))}{2}\right)^{\binom{n}{2}} = \exp\left(\binom{n}{2}^{1/2} t + \frac{t^2}{2} + O\left(\frac{1}{n}\right)\right)$$

as well as

$$\left(\frac{\rho(1)}{\rho(y(t))}\right)^{n+1} = \exp\left(-\frac{2\sqrt{2}\rho'(1)}{\rho(1)}t + O\left(\frac{1}{n}\right)\right),$$

and all other terms in the quotient converge to 1 as $n \to \infty$ for fixed t, hence

$$\lim_{n\to\infty} \mathbb{E}(e^{\varpi_n t}) = \exp\left(-\frac{2\sqrt{2}\rho'(1)}{\rho(1)}t + \frac{t^2}{2}\right),$$

which is the moment generating function of a Gaussian distribution (with mean $-\frac{2\sqrt{2}\rho'(1)}{\rho(1)}$ and variance 1). Thus we have the following theorem:

Theorem 3. *The number of arcs in a random EAD with n vertices is asymptotically normally distributed, with mean* $\sim \frac{1}{2}\binom{n}{2}$ *and variance* $\sim \frac{1}{4}\binom{n}{2}$.

Remark 2. *More precisely, the average number of arcs is*

$$\frac{1}{2}\binom{n}{2} - \frac{\rho'(1)}{\rho(1)}n + C + o(1),$$

which is the same as for arbitrary acyclic digraphs, except for the value of the constant C.

Remark 3. *It should be possible to obtain a stronger local limit theorem as in* [3], *valid if the number of arcs lies between* $\epsilon\binom{n}{2}$ *and* $(1-\epsilon)\binom{n}{2}$ *for some fixed* ϵ, *but the details may be quite intricate.*

Remark 4. *A simple heuristic argument explains why the number of arcs essentially follows a binomial distribution (thus in the limit a normal distribution): in the recursive construction that led us to* (2), *edges are included almost independently and with probabiliy almost equal to* 1/2. *This explains why the total number of arcs is essentially the sum of* $\binom{n}{2}$ *independent Bernoulli variables.*

6. The maximum rank in an extensional acyclic digraph

The *rank* of a set is recursively defined by $\mathrm{rk}(\{\}) = 0$ and $\mathrm{rk}(S) = 1+\sup_{x\in S}\mathrm{rk}(x)$. In the correspondence between full sets and EADs, the rank of an element of the full set corresponds to the length of the longest directed path starting at the associated vertex in the EAD, see Figure 2.

The maximum rank of a set is thus precisely equivalent to what McKay [10] defines as the *height* of an acyclic digraph. In his paper, McKay proves a central limit theorem for the height of random acyclic digraphs, which remains true for EADs, as we will see in this section. Again, only a few modifications of the argument are necessary. Let $d_{n,r,h}$ denote the number of EADs with n labelled vertices of which r have highest rank h. Since such an EAD is obtained by adding r sources to an EAD with k vertices of highest rank $h-1$ for some k such that each of the new sources has an arc to one of the old vertices of highest rank, we have the recursion

$$d_{n,r,h} = \sum_{k\geq 1}\binom{n}{r}((2^k-1)2^{n-r-k})^{\underline{r}}\, d_{n-r,k,h-1}$$

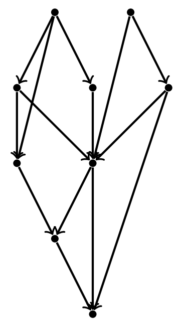

FIGURE 2. An EAD in which the vertices are arranged by rank.

with initial values $d_{n,r,0} = 1$ if $n = r = 1$ and $d_{n,r,0} = 0$ otherwise. Let us now define the generating functions

$$D_r(x,y) = \sum_{n\geq 1}\sum_{h\geq 0} \frac{d_{n,r,h}x^n y^h}{n!2^{\binom{n}{2}}}.$$

Then the recursion above translates to the functional equation

$$D_r(x,y) = \frac{x^r y}{r!2^{\binom{r}{2}}} \sum_{k\geq 1}(1-2^{-k})^r (D_k(x,y) - Q_{k,r}(x,y)) + x\cdot[r=1],$$

where

$$Q_{k,r}(x,y) = \sum_{n\geq r+1}\sum_{h\geq 1} \frac{d_{n-r,k,h-1}x^{n-r}y^{h-1}}{(n-r)!2^{\binom{n-r}{2}}} \left(1 - \prod_{j=0}^{r-1}\left(1 - \frac{j}{2^k-1}\cdot 2^{-(n-r-k)}\right)\right).$$

Now set

$$R_r(x,y) = -\frac{x^r y}{r!2^{\binom{r}{2}}} \sum_{k\geq 1}(1-2^{-k})^r Q_{k,r}(x,y) + x\cdot[r=1], \tag{3}$$

define the infinite matrix $M = M(x,y)$ by its entries $m_{ij} = x^i y(1-2^{-j})^i/(i!2^{\binom{i}{2}})$, and let $D = D(x,y)$ and $R = R(x,y)$ be the infinite vectors whose entries are $D_1(x,y), D_2(x,y),\dots$ and $R_1(x,y), R_2(x,y),\dots$ respectively. Then we have

$$D = MD + R \tag{4}$$

or

$$D = (I-M)^{-1}\cdot R,$$

where I stands for the infinite identity matrix. As it was pointed out by McKay in [10], the rapid convergence of the entries m_{ij} of M as $i\to\infty$ implies that the formal inverse $(I-M)^{-1} = \operatorname{adj}(I-M)/\det(I-M)$ is indeed well-defined. Since we will have to estimate the product with the vector R (in the case of ordinary EADs, this vector is simply the first unit vector $(1,0,0,\dots)^T$), let us treat this infinite matrix inverse in more detail. A priori, the determinant exists as a formal power series: when it is expanded, the number of terms of the form $cx^N y^M$ in this determinant is $M!q(N,M)$, where $q(N,M)$ is the number of partitions of N into M distinct terms. Each of them has a coefficient of the form

$$\pm \prod_{m=1}^{M} \frac{(1-2^{-j_m})^{i_m}}{i_m! 2^{\binom{i_m}{2}}}$$

with $i_1 < i_2 < \cdots < i_M$ and $i_1 + i_2 + \cdots + i_M = N$. It follows from these conditions that $M \leq \sqrt{2N}$, and thus

$$\sum_{m=1}^{M} \binom{i_m}{2} = \frac{1}{2} \sum_{m=1}^{M} i_m^2 - \frac{N}{2} \geq \frac{1}{2M} \Big(\sum_{m=1}^{M} i_m \Big)^2 - \frac{N}{2} \geq \frac{N^{3/2}}{2\sqrt{2}} - \frac{N}{2}.$$

Therefore, each coefficient can be estimated as

$$\left| \pm \prod_{m=1}^{M} \frac{(1-2^{-j_m})^{i_m}}{i_m! 2^{\binom{i_m}{2}}} \right| \leq 2^{-\sum_{m=1}^{M} \binom{i_m}{2}} = O\Big(\exp(-C_3 N^{3/2}) \Big)$$

for a constant $C_3 > 0$. On the other hand, $\log(M!q(N,M)) = O(\sqrt{N} \log N)$. Hence the coefficients of the formal power series that defines $\det(I - M)$ decay so rapidly that it represents a function that is analytic in x and y in the entire complex plane. The same is true for all entries of $\operatorname{adj}(I - M)$ (which are themselves similar determinants), and the entries are even uniformly bounded if x and y are restricted to compact sets.

Let us now estimate $R_r(x,y)$ in a similar way as the functions before. Suppose for now that x and y are positive and real and note that all coefficients of $Q_{k,r}(x,y)$ are positive. Since

$$\prod_{j=0}^{r-1} \Big(1 - \frac{j}{2^k - 1} \cdot 2^{-(n-r-k)} \Big) \geq 1 - \sum_{j=0}^{r-1} \frac{j}{2^k-1} \cdot 2^{-(n-r-k)} \geq 1 - r(r-1)2^{-(n-r)},$$

we get

$$\begin{aligned} \sum_{k \geq 1} (1-2^{-k})^r Q_{k,r}(x,y) &\leq r(r-1) \sum_{k \geq 1} \sum_{n \geq r+1} \sum_{h \geq 1} \frac{d_{n-r,k,h-1} x^{n-r} y^{h-1}}{(n-r)! 2^{\binom{n-r}{2}}} 2^{-(n-r)} \\ &= r(r-1) \sum_{k \geq 1} \sum_{n \geq 1} \sum_{h \geq 1} \frac{d_{n,k,h-1}}{n! 2^{\binom{n}{2}}} (x/2)^n y^{h-1} \\ &\leq r(r-1) \sum_{n \geq 1} \frac{a_n}{n! 2^{\binom{n}{2}}} \Big(\frac{x}{2} \max(1,y) \Big)^n, \end{aligned}$$

which means that the sum in the definition of $R_r(x,y)$ converges absolutely provided that $|x| \max(1,|y|) < 2z_0$, and thus it represents an analytic function if this holds. As it was mentioned earlier, the entries of $\operatorname{adj}(I - M)$ are analytic and uniformly bounded if $|x|$ and $|y|$ are. The factor $r!2^{\binom{r}{r}}$ in the denominator of (3) thus guarantees that all entries of the product $\operatorname{adj}(I - M) \cdot R$ are still analytic in x if y is in some fixed region around 1 and $|x| < 2z_0 / \max(1,|y|)$.

Hence we can write

$$D_r(x,y) = \frac{s_r(x,y)}{t(x,y)}$$

for any $r \geq 1$, where $t(x,y) = \det(I - M)$ and $s_r(x,y)$ is analytic in an open circle that contains the smallest zero $\tau(y)$ of $t(x,y)$ (in particular, note that $\tau(1)$ has to be z_0), provided that y is restricted to a suitable region around 1. Moreover, the total generating function

$$D(x,y) = \sum_{r \geq 1} D_r(x,y) = \frac{s(x,y)}{t(x,y)}$$

can also be written as such a fraction. Now one can use standard results on perturbation of meromorphic singularities (see [5, Theorem IX.9]) to obtain the following result:

Theorem 4. *The maximum rank of a vertex in a random EAD with n vertices is asymptotically normally distributed, with mean $\sim \mu_r n$ and variance $\sim \sigma_r n$, where*

$$\mu_r = -\frac{\tau'(1)}{\tau(1)} \approx 0.764334 \quad \text{and} \quad \sigma_r = \left(\frac{\tau'(1)}{\tau(1)}\right)^2 - \frac{\tau''(1)+\tau'(1)}{\tau(1)} \approx 0.145210.$$

This limiting distribution is the same as for ordinary acyclic digraphs, as determined in the aforementioned paper of McKay [10]. Moreover, we find that the asymptotic proportion of EADs with r vertices of maximum rank is $s_r(z_0,1)/s(z_0,1)$. We have the following theorem:

Theorem 5. *The distribution of the number of vertices of maximum rank in a random EAD converges to a discrete limiting distribution, the limit probability that the number of vertices of maximum rank equals r is*

$$H_r = s_r(z_0,1)/s(z_0,1).$$

Numerically,

$$H_1 \approx 0.815221,\ H_2 \approx 0.171843,\ H_3 \approx 0.012571,$$
$$H_4 \approx 0.000361,\ H_5 \approx 0.000004, \dots.$$

Compare the values in Table 3. In particular, the ratio between the number of EADs with a unique vertex of maximum rank and the number of all EADs with n vertices converges to a a limit whose numerical value is $0.815221^{-1} \approx 1.22666$, which was also conjectured by Policriti and Tomescu.

Remark 5. *Comparing residues in* (4) *at $x = z_0$ ($y = 1$), we see that the vector whose entries are $H_1, H_2, \dots$ is an eigenvector of the infinite matrix $M(z_0,1)$. The rapid decay of the entries of $M(z_0,1)$ along columns implies that the probabilities $H_1, H_2, \dots$ decrease equally rapidly.*

n	$r=1$	$r=2$	$r=3$	$r=4$
1	1			
2	2			
3	12			
4	192	24		
5	9120	1440		
6	1082880	208800	5760	
7	314979840	62657280	3366720	10080

Table 3. Enumeration of labelled EADs with given number of vertices of maximum rank.

7. Weakly extensional acyclic digraphs

Let us now see how the asymptotic number of WEADs can be determined along the same lines. We denote by $c_{n,s}$ the number of labelled WEADs with s sinks. Then clearly $c_{n,s} = 0$ for $n < s$, $c_{s,s} = 1$, and by analogous reasoning to equation (2)

$$c_{n,s} = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (2^{n-k} - n + k + s - 1)^{\underline{k}} c_{n-k,s}.$$

It is not hard to modify the proof of Theorem 1 to obtain

Theorem 6. *For any given s, the proportion of labelled WEADs with s sinks among all labelled acyclic digraphs converges to a limit W_s.*

The exact value of the limit W_s is determined by the convergent series

$$W_s = \frac{z_0^s}{s!2^{\binom{s}{2}}} + \sum_{n \geq s} \sum_{k=s}^{n-1} \frac{(-1)^{n-k}}{(n-k)!} \cdot 2^{-\binom{k}{2} - \binom{n-k}{2}} \frac{c_{k,s}}{k!} \left(1 - \prod_{j=k-s+1}^{n-s} (1 - 2^{-k} j)\right) z_0^n.$$

Numerical values of these constants can now be obtained quite easily, since the sum converges exponentially:

$$W_1 \approx 0.326210,\ W_2 \approx 0.283213,\ W_3 \approx 0.049917,$$
$$W_4 \approx 0.002732,\ W_5 \approx 0.000055, \ldots.$$

In particular, the total number of WEADs whose sinks are labelled with distinct labels from the set $\{0, 1\}$ is given by $2(c_{n,1} + c_{n,2}) \sim (2 + 2W_2/W_1)c_{n,1}$. The limit ratio $2 + 2W_2/W_1 \approx 3.736383$ was conjectured by Policriti and Tomescu, albeit for WEADs with exactly one element of highest rank. It is clear, however, that the distribution of the number of vertices of highest rank (as given in Theorem 5) is the same for WEADs as for ordinary EADs (see Theorem 9 below). Altogether, we find that the proportion of WEADs among all acyclic digraphs is in the limit $\sum_{s=1}^{\infty} W_s \approx 0.662127$. Table 4 shows a few explicit values of $c_{n,s}$.

n	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	1						
2	2	1					
3	12	9	1				
4	216	180	28	1			
5	10560	9060	1540	75	1		
6	1297440	1122480	195720	10350	186	1	
7	381013920	330445080	58053240	3144750	61194	441	1

Table 4. Enumeration of labelled WEADs with given number of sinks.

If the sinks of a WEAD receive additional labels, then the automorphism group is necessarily trivial again. Hence the number of unlabelled WEADs in which the s sinks bear one of t distinct labels is given by $t^{\underline{s}} c_{n,s}/n!$. Asymptotically, however, it is not necessary to endow the sinks with labels: as almost every acyclic digraph has trivial automorphism group and WEADs with a given number of sinks form a positive proportion, we can infer that Theorem 6 also holds for unlabelled WEADs.

All distributional results derived for EADs in the previous sections can be proved for WEADs in the same way quite easily. Explicitly, we have:

Theorem 7. *For any fixed s, the number of arcs in a random WEAD with s sinks and n vertices is asymptotically normally distributed, with mean $\sim \frac{1}{2}\binom{n}{2}$ and variance $\sim \frac{1}{4}\binom{n}{2}$.*

Theorem 8. *For any fixed s, the maximum rank of a vertex in a random WEAD with s sinks and n vertices is asymptotically normally distributed, with mean $\sim \mu_r n$ and variance $\sim \sigma_r n$, where the constants are the same as in Theorem 4.*

Theorem 9. *For any fixed s, the number of sources and the number of vertices of maximum rank in a random WEAD with s sinks and n vertices asymptotically follow the same discrete distributions as for EADs (see Theorem 2 and Theorem 5).*

8. Essential acyclic digraphs

Essential acyclic digraphs (in the following EssADs) are another very interesting class of digraphs whose definition is similar to that of EADs. They originate from the study of Bayesian networks [2], and their enumeration has been studied by Steinsky in two papers [14, 15]. An acyclic digraph is called essential if there is no pair of two vertices u and v such that the inneighbourhood of u is the union of the inneighbourhood of v and the set $\{v\}$; in other words, if there is no arc from a vertex v to a vertex u such that u has the same inneighbours as v (except for v itself).

For the recursion, it is convenient now to add sinks with each step. If e_n denotes the number of EssADs of order n, then one has the recursion

$$e_n = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (2^{n-k} - n + k)^k \, e_{n-k}$$

in analogy to (2), since each of k newly added sinks has $2^{n-k} - n + k$ possible inneighbourhoods ($n - k$ choices are forbidden by the aforementioned condition), and they do not necessarily have to be distinct, in contrast to EADs. The following was proven in [15]:

Theorem 10. *The proportion of EssADs among all labelled acyclic digraphs converges to a limit whose numerical value is $\approx 13.651798^{-1} \approx 0.073250$.*

The precise value of the constant is

$$1 + \sum_{n \geq 0} \sum_{k=0}^{n-1} \frac{(-1)^{n-k} e_k}{k!(n-k)!} 2^{-\binom{k}{2} - \binom{n-k}{2}} \Big(1 - (1 - k2^{-k})^{n-k}\Big) z_0^n,$$

and the proof is indeed very similar to the proof of our first main theorem (Theorem 1). The number of arcs and the height of EssADs can be treated in the same way as for EADs. Let us state the following theorems without proof, since their proofs are almost identical to those of Theorem 3 and Theorem 4 respectively.

Theorem 11. *The number of arcs in a random EssAD with n vertices is asymptotically normally distributed, with mean $\sim \frac{1}{2}\binom{n}{2}$ and variance $\sim \frac{1}{4}\binom{n}{2}$.*

Theorem 12. *The length of a longest directed path in a random EssAD with n vertices is asymptotically normally distributed, with mean $\sim \mu_r n$ and variance $\sim \sigma_r n$, where the constants are the same as in Theorem 4.*

The number of sources, however, is somewhat different. While the number of sinks of a random EssAD follows the same discrete distribution as for all acyclic digraphs (which, by symmetry, is the same as the distribution of the number of sources, which is in turn the same as the distribution of the number of sources of a random EAD, see Section 4), this is not the case for the number of sources of a random EssAD. In order to determine the distribution, we have to make use of a bivariate generating function, in which u marks the number of sources (cf. [6], where the problem of counting acyclic digraphs by sources and sinks is studied in detail). Let $e_n(u)$ be the polynomial in which the coefficient of u^k is the number of EssADs of order n with precisely k sources. Then we have

$$e_n(u) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (2^{n-k} - n + k + u - 1)^k \, e_{n-k}(u),$$

since a newly added sink is only also a source if its inneighbourhood is empty (which is one of the $2^{n-k} - n + k$ possible choices). This can be rewritten as

$$\sum_{k=0}^{n} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \frac{e_k(u)}{k!} = \sum_{k=0}^{n-1} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \frac{e_k(u)}{k!} \Big(1 - (1 - 2^{-k}(k+1-u))^{n-k}\Big) + [n=0].$$

We introduce the generating function $E(x,u) = \sum_{n \geq 0} 2^{-\binom{n}{2}} \frac{e_n(u) x^n}{n!}$ as in Section 3 and obtain

$$E(x,u)\phi(x) = 1 + H(x,u)$$

with

$$H(x,u) = \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \frac{(-1)^{n-k}}{(n-k)!} 2^{-\binom{k}{2}-\binom{n-k}{2}} \frac{e_k(u)}{k!} \Big(1 - (1 - 2^{-k}(k+1-u))^{n-k}\Big) x^n$$

Assume that $0 < u < 2$; we use the inequality

$$e_k(u) \leq a_k \max(1,u)^k \leq C_1 k! 2^{\binom{k}{2}} z_0^{-k} \max(1,u)^k$$

as well as

$$(1 - 2^{-k}(k+1-u))^{n-k} \geq 1 - (n-k)(k+1-u)2^{-k} \geq 1 - n^2 2^{-k}$$

and estimate the coefficients of $H(x,u)$ in the same way as in Section 3 to find that it converges absolutely for $|x| < 2z_0/\max(1,u)$ and is thus holomorphic in this region. By singularity analysis, we find

$$\lim_{n\to\infty} \frac{e_n(u)}{a_n} = 1 + H(z_0,u)$$

for all $0 < u < 2$, which means that the distribution of the number of sources converges to the discrete distribution whose probability generating function is given by $(1 + H(z_0,u))/(1 + H(z_0,1))$. The probabilities can again be determined numerically:

Theorem 13. *The distribution of the number of sources in a random EAD converges to a discrete limiting distribution, the limit probability that the number of sources equals ℓ is*

$$\tilde{S}_\ell = [u^\ell]\frac{1 + H(z_0, u)}{1 + H(z_0, 1)}.$$

Numerically,

$$\tilde{S}_1 = 0,\ \tilde{S}_2 \approx 0.700275,\ \tilde{S}_3 \approx 0.276627,$$
$$\tilde{S}_4 \approx 0.022528,\ \tilde{S}_5 \approx 0.000564,\ \tilde{S}_6 \approx 0.000005, \ldots.$$

Note that an EssAD of order > 1 cannot have a single source, which is easy to see from the definition: arrange the vertices by rank, as in Figure 2. If there was only a single source v at rank k, one of its outneighbours at rank $k - 1$ would have v as its only inneighbour, thus violating the definition of an EssAD.

9. Conclusion

We have seen that roughly 32.6% of all (labelled or unlabelled) acyclic digraphs are extensional, and more generally, we were able to determine the proportion of WEADs with a fixed number of sinks among all acylic digraphs. The distribution of the number of sources, the number of arcs and the maximum rank are the same for EADs as for the family of all acyclic digraphs, meaning that they have essentially the same shape. Analogous results hold for the very similar class of EssADs.

Things might be quite different if other subfamilies of acyclic digraphs are studied. For instance, Policriti and Tomescu mention the problem of counting *slim* EADs, which have the additional property that removal of an arbitrary arc yields an acyclic digraph which is no longer extensional. It seems to be a quite challenging problem to determine the asymptotic number of such EADs.

Acknowledgment

I would like to thank Alexandru Tomescu for telling me about this interesting problem.

References

[1] The On-Line Encyclopedia of Integer Sequences (2012). Published electronically at `http://oeis.org`
[2] Andersson, S.A., Madigan, D., Perlman, M.D.: A characterization of Markov equivalence classes for acyclic digraphs. Ann. Statist. **25**(2), 505–541 (1997)
[3] Bender, E.A., Richmond, L.B., Robinson, R.W., Wormald, N.C.: The asymptotic number of acyclic digraphs. I. Combinatorica **6**(1), 15–22 (1986)
[4] Bender, E.A., Robinson, R.W.: The asymptotic number of acyclic digraphs. II. J. Combin. Theory Ser. B **44**(3), 363–369 (1988)
[5] Flajolet, P., Sedgewick, R.: Analytic combinatorics. Cambridge University Press, Cambridge (2009)
[6] Gessel, I.M.: Counting acyclic digraphs by sources and sinks. Discrete Math. **160**(1-3), 253–258 (1996)
[7] Grabner, P.J., Steinsky, B.: Asymptotic behaviour of the poles of a special generating function for acyclic digraphs. Aequationes Math. **70**(3), 268–278 (2005)
[8] Harary, F., Palmer, E.M.: Graphical enumeration. Academic Press, New York (1973)
[9] Liskovets, V.A.: The number of maximal vertices of a random acyclic digraph. Teor. Verojatnost. i Primenen. **20**(2), 412–421 (1975)
[10] McKay, B.D.: On the shape of a random acyclic digraph. Math. Proc. Cambridge Philos. Soc. **106**(3), 459–465 (1989)

[11] Peddicord, R.: The number of full sets with n elements. Proc. Amer. Math. Soc. **13**, 825–828 (1962)
[12] Policriti, A., Tomescu, A.I.: Counting extensional acyclic digraphs. Information Processing Letters **111**(16), 787–791 (2011)
[13] Robinson, R.W.: Counting labeled acyclic digraphs. In: New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich., 1971), pp. 239–273. Academic Press, New York (1973)
[14] Steinsky, B.: Enumeration of labelled chain graphs and labelled essential directed acyclic graphs. Discrete Math. **270**(1-3), 267–278 (2003)
[15] Steinsky, B.: Asymptotic behaviour of the number of labelled essential acyclic digraphs and labelled chain graphs. Graphs Combin. **20**(3), 399–411 (2004)
[16] Wagner, S.: Asymptotic enumeration of extensional acyclic digraphs. In: Proceedings of the ANALCO12 Meeting on Analytic Algorithmics and Combinatorics, Kyoto, January 16, 2012, pp. 1–8 (2012)

Stephan Wagner, Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
E-mail address: `swagner@sun.ac.za`