The Number of Sources and Isolated Vertices in Random Directed Acyclic Graphs

Dimbinaina Ralaivaosaona
University of Stellenbosch, South Africa

Abstract

For a positive integer n and a real number $p \in (0, 1)$, a random directed acyclic digraph $\mathbb{D}_{ac}(n, p)$ is obtained from the binomial random digraph model $\mathbb{D}(n, p)$ conditioned to be acyclic, i.e., directed cycles are forbidden. In the binomial random digraph model $\mathbb{D}(n, p)$, every possible directed edge (excluding loops) occurs independently with probability p. Sources and sinks are among the most natural characteristics of directed acyclic graphs. We investigate the distribution of the number of sources in $\mathbb{D}_{ac}(n, p)$ when p is of the form λ/n, where λ is a fixed positive constant. Because of symmetry, the number of sinks will have the same distribution as the number of sources. Our main motivation is to understand how this distribution changes as we pass through the critical point $p = 1/n$. Since we are in the sparse regime, it makes sense to include the number of isolated vertices as well. In a directed graph an isolated vertex can be regarded as a vertex that is both a source and a sink. We prove asymptotic normality for each of these parameters when $p = \lambda/n$. Our method is based on the analysis of a multivariate generating function from a work of Gessel.

2012 ACM Subject Classification Mathematics of computing → Random graphs; Mathematics of computing → Generating functions

Keywords and phrases Directed acyclic graph, generating function, central limit theorem

Digital Object Identifier 10.4230/LIPIcs.AofA.2022.17

Funding This work is based on the research supported in part by the National Research Foundation of South Africa (grant number: 129413).

Acknowledgements I want to thank the anonymous referees for their helpful comments which helped to improve the presentation of this paper.

1 Introduction

For a positive integer n we consider directed graphs (digraphs) on the vertex set $[n] = \{1, 2, \dots, n\}$ where loops and multi-edges are forbidden. In a digraph, a vertex v is called a source if it has an in-degree zero and a sink if it has an out-degree zero. If we only consider directed acyclic graphs (DAGs), it is well known that a non-empty acyclic digraph has at least one source and one sink. So, we would like to investigate the distributions of the number of these vertices in random DAGs.

The distribution of the number of isolated vertices and its generalisation, the number of vertices of a given degree, in the random (undirected) graphs are well-covered topics in the literature, see for example [1, 2, 6, 18]. It makes sense to extend these results to other graph-like structures. Investigating the number of isolated vertices, sources, and sinks should be the starting point for the case of random DAGs.

The model that we consider in this paper is constructed in the following way: for $p \in (0, 1)$, we first consider the binomial random digraph model $\mathbb{D}(n, p)$, where each of the $n(n-1)$ possible directed edges occurs independently with probability p. Then, the random acyclic digraph $\mathbb{D}_{ac}(n, p)$ is simply $\mathbb{D}(n, p)$ conditioned to be acyclic. Due to limited space, in this paper, we restrict ourselves to the sparse case where $p = \lambda/n$, for which $\lambda > 0$ is fixed. It is known that the model $\mathbb{D}(n, p)$ exhibits a phase transition around the critical point $p = 1/n$. This phase transition has also been analysed in the literature, see for example [7, 10, 11].

For $p = \frac{\lambda}{n}$ an asymptotic formula for the probability that $\mathbb{D}(n,p)$ is acyclic is given in [3] following the approach in [14]: if $\lambda > 0$ is a constant, then

$$\mathbb{P}(\mathbb{D}(n,p) \text{ acyclic}) \sim \begin{cases} (1-\lambda)e^{\lambda} & \text{if } \lambda < 1 \\ C_1 n^{-1/3} & \text{if } \lambda = 1 \\ C_2(\lambda) n^{-1/3} e^{-a(\lambda)n + b(\lambda)n^{1/3}} & \text{if } \lambda > 1, \end{cases} \tag{1}$$

where C_1 is a positive constant, and $C_2(\lambda)$, $a(\lambda)$, and $b(\lambda)$ are positive numbers depending only on $\lambda \geq 1$. These terms are explicitly defined in [3, Sec. 6].

There are a few different random digraph models. For instance, the model considered in [14] is obtained from the random binomial undirected graph $G(n, 2p)$ where a direction is given to each existing edge independently with probability $\frac{1}{2}$. Although, it would also be possible to study such a model with our method, the model in the present paper, when conditioned to be a DAG, gives a more natural generalisation to the uniform random DAG model where a digraph is chosen uniformly at random from the set of all DAGs on the vertex set $[n]$. With our notation the latter model is equivalent to $\mathbb{D}_{ac}(n, \frac{1}{2})$.

The number of sources in the random acyclic digraph model $\mathbb{D}_{ac}(n, \frac{1}{2})$ was already studied by Liskovets [9]. It was shown that the number of sources in $\mathbb{D}_{ac}(n, \frac{1}{2})$ has a discrete limiting distribution as $n \to \infty$. More precisely, if denote by $p(n,k)$ the probability that a uniform random DAG on $[n]$ has exactly k sources, then as $n \to \infty$ we have

$$p(n,k) \sim \frac{\varrho^k \phi(2^{-k}\varrho)}{k! 2^{\binom{k}{2}}}, \quad \text{where} \quad \phi(x) = \sum_{n=0}^{\infty} \frac{(-x)^n}{n! 2^{\binom{n}{2}}}, \tag{2}$$

and $\varrho \approx 1.4880785\ldots$ is the smallest positive solution of the equation $\phi(x) = 0$. The function $\phi(x)$ and its zero ϱ already appeared in earlier results on the enumeration of DAGs, see the work of Robinson [15], Liskovets [8], and Stanley [16]. Returning to the structure of the uniform random DAG $\mathbb{D}_{ac}(n, \frac{1}{2})$, McKay [12] showed that the height is asymptotically normally distributed with mean and variance asymptotically equal to Cn and $C'n$ respectively, where $C \approx 0.764334$ and $C' \approx 0.145210$. It is reasonable to expect that similar results would hold for the number of sources and the height of $\mathbb{D}_{ac}(n,p)$ for fixed $p \in (0,1)$. However, it is not clear how these parameters behave when p tends to zero. Let us now state our result for the number of sources when p is of the form λ/n.

▶ **Theorem 1.** *Let $S(D)$ denote the number of sources in an acyclic digraph D. Define*

$$\mu(\lambda) = \begin{cases} e^{-\lambda} & \text{if } \lambda < 1 \\ \dfrac{1}{e\lambda} & \text{if } \lambda \geq 1, \end{cases} \quad \text{and} \quad \sigma^2(\lambda) = \begin{cases} e^{-\lambda}(1-e^{-\lambda}) & \text{if } \lambda < 1 \\ \dfrac{e-1}{e^2\lambda} & \text{if } \lambda \geq 1. \end{cases}$$

Then, for a fixed $\lambda > 0$, the expectation of the number of sources in a random acyclic digraph $\mathbb{D}_{ac}(n, \lambda/n)$ satisfies the asymptotic estimate $\mathbb{E}(S(\mathbb{D}_{ac}(n,\lambda/n))) \sim \mu(\lambda)n$ as $n \to \infty$. Moreover, we have

$$\frac{S(\mathbb{D}_{ac}(n,\lambda/n)) - \mu(\lambda)n}{\sqrt{\sigma^2(\lambda)n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Since we are interested in the structure of $\mathbb{D}_{ac}(n,p)$ in the sparse regime, it makes sense to look at the number of isolated vertices. We obtain the following theorem.

▶ **Theorem 2.** *Let $I(D)$ denote the number of isolated vertices in an acyclic digraph D. Define*

$$\mu^*(\lambda) = \begin{cases} e^{-2\lambda} & \text{if } \lambda < 1, \\ \frac{e^{-(\lambda+1)}}{\lambda} & \text{if } \lambda \geq 1, \end{cases}$$

and

$$\sigma^*(\lambda)^2 = \begin{cases} e^{-2\lambda}(1 + (2\lambda - 1)e^{-2\lambda}) & \text{if } \lambda < 1 \\ \lambda^{-1}e^{-(\lambda+1)}(1 + e^{-(\lambda+1)}) & \text{if } \lambda \geq 1. \end{cases}$$

Then, for a fixed $\lambda > 0$, the expectation of the number of isolated vertices in a random acyclic digraph $\mathbb{D}_{ac}(n, \lambda/n)$ satisfies the asymptotic estimate $\mathbb{E}(I(\mathbb{D}_{ac}(n, \lambda/n))) \sim \mu^*(\lambda)n$ as $n \to \infty$. Moreover, we have

$$\frac{I(\mathbb{D}_{ac}(n, \lambda/n)) - \mu^*(\lambda)n}{\sqrt{\sigma^*(\lambda)^2 n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

It is not difficult to show that main terms in asymptotic expansions of the expectations $\mathbb{E}(S(\mathbb{D}(n, \lambda/n)))$ and $\mathbb{E}(I(\mathbb{D}(n, \lambda/n)))$, for the unconditioned random digraph $\mathbb{D}(n, \lambda/n)$, are $e^{-\lambda}n$ and $e^{-2\lambda}n$, respectively for any fixed $\lambda \geq 0$. So, it appears that conditioning on the event that the random digraph is acyclic does not affect these expectations for $\lambda \in (0, 1]$, at least asymptotically. This is, however, not surprising because we know as we see in (1) that $\mathbb{D}(n, \lambda/n)$ is acyclic with positive probability for $\lambda < 1$, and we expect that the random variables $S(\mathbb{D}(n, \lambda/n))$ and $I(\mathbb{D}(n, \lambda/n))$ are concentrated around their expectations. For $\lambda > 1$ these expectations are higher in the random acyclic digraph model. It would be interesting to know what happens when we allow $\lambda \to \infty$ as $n \to \infty$. Extending the above results to such a case seems to be possible but it requires more work; it would certainly be too long for this extended abstract.

Throughout this paper, we adopt the notations and abbreviations of [14] since most of the asymptotic analysis that we need to prove Theorem 1 and Theorem 2 are based on the analytic method developed in that paper. This paper is organised as follows: in Section 2 we present and prove a result of Gessel [4] about a multivariate generating function that includes the number of sources, sinks, and isolated vertices. Section 3 consists of collections of asymptotic results from [14] and some of their consequences. The proofs of Theorem 1 and Theorem 2 are presented in Section 4 and Section 5 respectively.

2 Generating functions

We use the so-called *graphic generating function* for the enumeration of acyclic digraphs. If we denote by $e(D)$ the number of (directed) edges in a digraph D, then we define

$$A_n(y) = \sum_D y^{e(D)},$$

where the sum is taken over all acyclic digraphs on $[n]$. The corresponding graphic generating function is

$$A(x, y) = \sum_{n=0}^{\infty} \frac{A_n(y)x^n}{n!(1+y)^{\binom{n}{2}}}.$$

It turns out that this bivariate generating can be written in the following way:

$$A(x,y) = \frac{1}{\phi(x,y)}, \quad \text{where} \quad \phi(x,y) = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!(1+y)^{\binom{n}{2}}}, \tag{3}$$

see [15]. Observe, that for $y > 0$, the power series in the definition of $\phi(x,y)$ converges for $x \in \mathbb{C}$. Hence, $\phi(x,y)$ can be regarded as an entire function of x for any $y > 0$.

To include counts on the number of sources, sinks, and isolated vertices, we have to define two related parameters. For an acyclic digraph D, let $S_o(D)$ and $S_i(D)$, respectively, be the number of sources in D that are not isolated vertices and the number of sinks that are not isolated vertices. Furthermore, let $I(D)$ be the number of isolated vertices of D. Then, consider the generating function

$$A_n(y,u_1,u_2,u_3) = \sum_D y^{e(D)} u_1^{S_o(D)} u_2^{S_i(D)} u_3^{I(D)},$$

where the sum is taken over all acyclic digraphs on the vertex set $[n]$. The corresponding graphic generating function is given by

$$A(x,y,u_1,u_2,u_3) = \sum_{n=0}^{\infty} \frac{x^n}{n!(1+y)^{\binom{n}{2}}} A_n(y,u_1,u_2,u_3)x^n.$$

If we denote by $a_{n,m}(k_1,k_1,k_3)$ the number of acyclic digraphs D on the vertex set $[n]$ with $e(D) = m$, $S_o(D) = k_1$, $S_i(D) = k_2$ and $I(D) = k_3$, then the polynomial $A_n(y,u_1,u_2,u_3)$ can be written as follows:

$$A_n(y,u_1,u_2,u_3) = \sum_{m,k_1,k_2,k_3} a_{n,m}(k_1,k_2,k_3) u_1^{k_1} u_2^{k_2} u_3^{k_3} y^m. \tag{4}$$

The next lemma allows us to obtain the joint probability generating function for our parameters in terms of $A_n(y,u_1,u_2,u_3)$.

▶ **Lemma 3.** *We have*

$$\mathbb{E}\left(u_1^{S_o(\mathbb{D}_{ac}(n,p))} u_2^{S_i(\mathbb{D}_{ac}(n,p))} u_3^{I(\mathbb{D}_{ac}(n,p))}\right) = \frac{A_n\left(\frac{p}{1-p}, u_1, u_2, u_3\right)}{A_n\left(\frac{p}{1-p}\right)}.$$

The proof of this lemma is straightforward, so we leave it as an exercise to the reader.

We need to express $A(x,y,u_1,u_2,u_3)$ in more a manageable form in order to obtain any useful estimate of $A_n(y,u_1,u_2,u_3)$ as $n \to \infty$. Fortunately for us, this was already done by Gessel in [4]. Since this result is our main ingredient and some part of the proof was omitted in [4], we give a full proof here for completeness. Before we begin let us state a general property of graphic generating functions. Given infinite sequences $(a_n^{(i)})_n$, for $i \in \{1,2,3\}$, if $f_i(x)$ denotes the graphic generating function associated with the sequence $(a_n^{(i)})_n$, that is

$$f_i(x) = \sum_{n=0}^{\infty} \frac{a_n^{(i)}}{n!(1+y)^{\binom{n}{2}}} x^n,$$

and if $c_n = n!(1+y)^{\binom{n}{2}} \times [x^n]\Big(f_1(x)f_2(x)f_3(x)\Big)$, then the three-term convolution formula is

$$c_n = \sum_{j+k+l=n} (1+y)^{jk+jl+kl} \binom{n}{k,l,j} a_j^{(1)} a_k^{(2)} a_l^{(3)}. \tag{5}$$

▶ **Lemma 4** (Theorem 2 in Gessel [4]). *We have*

$$\sum_{n=0}^{\infty} A_n(y,u_1,u_2,u_1+u_2-1)\frac{x^n}{(1+y)^{\binom{n}{2}}n!} = \frac{\phi((1-u_1)x,y)\phi((1-u_2)x,y)}{\phi(x,y)},$$

and

$$\sum_{n=0}^{\infty} A_n(y,u_1,u_2,u_3)\frac{x^n}{n!} = e^{(u_3-u_1-u_2+1)x}\sum_{n=0}^{\infty} A_n(y,u_1,u_2,u_1+u_2-1)\frac{x^n}{n!}.$$

Proof. Consider the set of quadruples (S_1,S_2,D,E) where S_1 and S_2 are disjoint subsets of $[n]$, D is an acyclic digraph on $[n]\setminus(S_1\cup S_2)$ and E a set of directed edges consisting only of edges from S_1 to $V(D)\cup S_2$, or from $V(D)$ to S_2. Such a quadruple can be interpreted as an acyclic digraph where every vertex in S_1 is a source, and every vertex in S_2 is a sink. If each quadruple is weighted by $y^{|E|+e(D)}u_1^{|S_1|}u_2^{|S_2|}$, then the total weight over all possible quadruples is

$$\sum_{j+k+l=n}(1+y)^{jk+jl+kl}\binom{n}{k,l,j}u_1^j u_2^k A_l(y).$$

On the other hand, consider the set of acyclic digraphs D' on $[n]$ with four distinguished subsets of $[n]$: the first is a subset of the *strictly* sources of D' (no isolated vertices), a subset of the strictly sinks, and a pair of disjoint subsets of the isolated vertices. There is a correspondence between such a set with the set of quadruples (S_1,S_2,D,E): the subset of the strictly sources and the first subset of the isolated vertices of D' form the set S_1, while the subset of the strictly sinks together with the second subset of the isolated vertices form the set S_2, and E is the set edges of D' that go from S_1 to S_2, S_1 to $[n]\setminus(S_1\cup S_2)$, or from $[n]\setminus(S_1\cup S_2)$ to S_2. By taking this correspondence into consideration when calculating the total weight formula above, we deduce that

$$A_n(y,1+u_1,1+u_2,1+u_1+u_2) = \sum_{j+k+l=n}(1+y)^{jk+jl+kl}\binom{n}{k,l,j}u_1^j u_2^k A_l(y).$$

This is a three-term convolution formula, as defined in (5). Therefore, we deduce that

$$A(x,y,1+u_1,1+u_2,1+u_1+u_2) = \phi(-u_1x)\phi(-u_2x)A(x,y).$$

The first equation in the lemma follows by using (3) and by shifting the variables u_1 and u_2 by -1.

For the second equation, by removing all isolated vertices, each acyclic digraph D on $[n]$ can be associated with a pair (S,D') where S is a subset of $[n]$ (the set of isolated vertices of D) and D' is an acyclic digraph on $[n]\setminus S$ which has no isolated vertices. Thus, we obtain

$$A_n(y,u_1,u_2,u_3) = \sum_{j=0}^{n}\binom{n}{j}u_3^j A_{n-j}(y,u_1,u_2,0).$$

This is the standard two-term convolution formula for the product of two exponential generating functions. So, we obtain

$$\sum_{n=0}^{\infty} A_n(y,u_1,u_2,u_3)\frac{x^n}{n!} = e^{u_3x}\sum_{n=0}^{\infty} A_n(y,u_1,u_2,0)\frac{x^n}{n!}.$$

In particular, we have

$$\sum_{n=0}^{\infty} A_n(y, u_1, u_2, u_1 + u_2 - 1)\frac{x^n}{n!} = e^{(u_1+u_2-1)x} \sum_{n=0}^{\infty} A_n(y, u_1, u_2, 0)\frac{x^n}{n!}.$$

Eliminating the series involving $A_n(y, u_1, u_2, 0)$ gives the formula in the lemma. ◀

3 Asymptotic analysis

The results in this section are mainly drawn from [14] or are consequences of the results in [14]. So, let us first summarise the notations and abbreviations that we used in [14]. Throughout, y is a positive number that tends to zero as $n \to \infty$. In our case, in view of Lemma 3, $y = p/(1-p)$ where $p = \lambda/n$. Moreover, it is convenient to use the abbreviation $\alpha = \log(1 + y)$ and $\beta = \sqrt{1 + y}$. So, $y \sim \alpha$ as $n \to \infty$. In particular, $\alpha \to 0^+$ as $n \to \infty$. In addition, for a complex number x, we define $w = w(x) = W_0(-x\alpha\beta)$, where W_0 is the principal branch of the Lambert W function. It is useful to keep these notations and abbreviations in mind when reading the rest of this paper.

3.1 Estimates of $\phi(x, y)$

We begin by providing asymptotic estimates for $\phi(x, y)$ where $y \to 0^+$ and x is complex. We obtain the next lemma with a minor modification of a similar result in [14], the reader can also consult [3] which contains more details.

▶ **Lemma 5.** *The function $\phi(x, y)$ satisfies the following asymptotic formulas as $y \to 0^+$, both estimates are uniform in x:*

- *If $x = o(\alpha^{-1})$, then*

$$\phi(x, y) \sim e^{\frac{1}{2\alpha}(w^2+2w)}. \tag{6}$$

- *If $x = \mathcal{O}(\alpha^{-1})$ and $w = w(x)$ is bounded away from -1, then*

$$\phi(x, y) \sim 2^{5/6}\pi^{1/2}w^{-1/3}\alpha^{-1/6}\mathrm{Ai}(R)e^{\frac{2}{3}R^{3/2}+\frac{1}{2\alpha}(w^2+2w)}, \tag{7}$$

where $R = 2^{-2/3}(1 + w)^2 w^{-4/3}\alpha^{-2/3}$, and $\mathrm{Ai}(z)$ is the Airy function.

3.2 Coefficient extraction

We are going to need estimates of $A_m(y)$ for a certain range of values of m close to n. For our purposes, it suffices to consider m to be of the form $n + o(n^{2/3})$. By the Cauchy integral formula, we have

$$\frac{A_m(y)}{m!(1+y)^{\binom{m}{2}}} = \frac{1}{2\pi i}\oint_{|x|=\rho} \frac{1}{\phi(x, y)x^{m+1}}dx. \tag{8}$$

The value of $\rho > 0$ will be chosen depending on the sign of $\lambda - 1$.

▶ **Lemma 6.** *Let λ be a fixed value in the interval $(0, 1)$. If $m = n+\mathcal{O}(n^{2/3-\epsilon})$ for a constant $\epsilon > 0$, then as $n \to \infty$,*

$$A_m(y) \sim m!(1 + y)^{\binom{m}{2}} \frac{1 - \alpha m}{\sqrt{2\pi m}}\rho^{-m}e^{-\frac{1}{2\alpha}(\alpha^2m^2-2\alpha m)},$$

where $\rho = \frac{m}{\beta}e^{-\alpha m}$.

Proof. With the abbreviation $w(x) = W_0(-x\alpha\beta)$, let

$$h(t) = -\frac{1}{2\alpha}\left(w^2(\rho e^{it}) + 2w(\rho e^{it})\right).$$

Hence, if $\alpha\rho$ is bounded away from e^{-1}, then by the second part of Lemma 5, the Equation (8) yields

$$\frac{A_m(y)}{m!(1+y)^{\binom{m}{2}}} = (1+o(1))\rho^{-m}\int_{-\pi}^{\pi} P(\alpha, w(\rho e^{it}))e^{h(t)-imt}\,dt,$$

where $P(\alpha, w(\rho e^{it})) = 2^{-11/6}\pi^{-3/2}w^{1/3}\alpha^{1/6}\mathrm{Ai}(R)^{-1}e^{-\frac{2}{3}R^{3/2}}$ and $R = 2^{-2/3}(1+w)^2w^{-4/3}\alpha^{-2/3}$. Since, in our case $\rho = \mathcal{O}(ne^{-\lambda})$, $w(\rho e^{it})$ is bounded away from -1. Hence, $|R| \to \infty$ as $\alpha \to 0^+$ for $t \in [-\pi, \pi]$. Recall the following well known asymptotic formula for the Airy function: for any $\varepsilon > 0$,

$$\mathrm{Ai}(z) \sim \frac{e^{-\frac{2}{3}z^{3/2}}}{2\sqrt{\pi}z^{1/4}} \quad \text{as } |z| \to \infty, \text{ and } |\mathrm{Arg}(z)| \le \pi - \varepsilon, \tag{9}$$

see [13, (9.7.5)]. We can prove from this that $P(\alpha, w(\rho e^{it}))$ is bounded uniformly for $t \in [-\pi, \pi]$. In fact, skipping the details, the estimate of $P(\alpha, w(\rho e^{it}))$ simplifies to

$$P(\alpha, w(\rho e^{it})) \sim \frac{1}{2\pi}(1 + w(\rho e^{it}))^{1/2},$$

uniformly for $t \in [-\pi, \pi]$.

In order to apply the saddle point method, we need to evaluate the first few derivatives of h. We have

$$h'(t) = -\frac{i}{\alpha}w(\rho e^t), \quad h''(t) = \frac{1}{\alpha}\frac{w(\rho e^t)}{1 + w(\rho e^t)}, \text{ and } \quad h^{(3)}(t) = \frac{i}{\alpha}\frac{w(\rho e^t)}{(1 + w(\rho e^t))^3}.$$

The saddle point equation is $h'(0) - im = 0$ which is equivalent to $w(\rho) = -\alpha m$ or $\rho = \frac{m}{\beta}e^{-\alpha m}$. Moreover, $h''(0) = \frac{-m}{1-\alpha m}$ which is of order α^{-1} under our assumptions, and $h^{(3)}(t) = \mathcal{O}(\alpha^{-1})$. The standard saddle point method applies in this case: the integral on the right-hand side of is divided into two parts

$$\int_{-\alpha^{-c}}^{\alpha^c} P(\alpha, w(\rho e^{it}))e^{h(t)-imt}\,dt + \int_{|t|\ge\alpha^c} P(\alpha, w(\rho e^{it}))e^{h(t)-imt}dt$$

where c is a fixed number in the interval $(1/3, 1/2)$. When we apply the saddle point method, we call the first integral the local integral and the rest the tail. By Taylor expanding $h(t) - imt$ and $P(\alpha, w(\rho e^{it}))$ for $|t| \le \alpha^c$, we deduce the following estimate for the local integral:

$$\int_{-\alpha^{-c}}^{\alpha^c} P(\alpha, w(\rho e^{it}))e^{h(t)-imt}\,dt = (1+o(1))P(\alpha, w(\rho))e^{h(0)}\int_{-\alpha^{-c}}^{\alpha^c} e^{\frac{1}{2}h''(0)t^2}dt. \tag{10}$$

The integral on the right-hand side can be approximated by a Gaussian integral, by extending its range to $(-\infty, \infty)$, with an error term smaller than any power of α. Hence, we deduce that

$$\int_{-\alpha^{-c}}^{\alpha^c} P(\alpha, w(\rho e^{it}))e^{h(t)-imt}\,dt = (1+o(1))P(\alpha, w(\rho))e^{h(0)}\sqrt{\frac{2\pi}{-h''(0)}}.$$

On the other hand, it is not difficult to show that $\mathrm{Re}(h(t) - h(0))$ is negative for $\alpha^c \leq |t| \leq \pi$ (for small enough α). In fact, we can show that there exists a positive constant $C > 0$ (independent of α and t) such that $\mathrm{Re}(h(t) - h(0)) \leq -C\alpha^{2c-1}$ for $\alpha^c \leq |t| \leq \pi$. This is enough to prove that the tail integral is much smaller than the local integral, and therefore, it can be neglected. Thus, we we obtain

$$\frac{A_m(y)}{m!(1+y)^{\binom{m}{2}}} \sim P(\alpha, w(\rho))e^{h(0)}\sqrt{\frac{2\pi}{-h''(0)}}.$$

Expressing everything in terms of m gives us the estimate in the statement of the lemma. ◀

Next, we consider the critical case.

▶ **Lemma 7.** *If $\lambda = 1$ and $m = n + \mathcal{O}(n^{2/3-\epsilon})$ for some constant $\epsilon > 0$, then as $n \to \infty$ we have*

$$A_m(y) \sim m!(1+y)^{\binom{m}{2}} \frac{\alpha^{2/3} Ai(0)}{2\pi\phi(\rho, y)\rho^m} \int_{-\infty}^{\infty} \frac{1}{Ai(-2^{1/3}it)}\, dt,$$

where $\rho = \frac{1}{e\alpha\beta}$.

Proof. This is a direct application of [14, Lemma 9, Eq. (27)]. Since $p = \frac{1}{n}$, $y = \frac{p}{1-p}$ and $\alpha = \log(1+y)$. We can show that if $m = n + \mathcal{O}(n^{2/3-\epsilon})$, then $m = \alpha^{-1} + o(\alpha^{-2/3})$. So, [14, Lemma 9, Eq. (27)] applies with $b = 0$. ◀

Finally, for the super-critical case, we have the following result:

▶ **Lemma 8.** *If $\lambda > 1$ and $m = n + \mathcal{O}(n^{2/3-\epsilon})$ for some constant $\epsilon > 0$ then as $n \to \infty$ we have*

$$A_m(y) = -m!(1+y)^{\binom{m}{2}} \frac{1}{\varrho_1(y)^{m+1}\phi_x(\varrho_1(y), y)} + O\left(\frac{\alpha^{2/3}}{|\phi(\rho, y)|\rho^m}\right).$$

where $\varrho_1(y)$ is the smallest x-solution of the equation $\phi(x, y) = 0$, $\phi_x(\varrho_1(y), y)$ is $\partial_x\phi(x, y)|_{x=\varrho_1(y)}$, and ρ is of the form

$$\rho = \frac{1}{e}y^{-1} - \frac{b}{2^{1/3}e}\,y^{-1/3}.$$

The constant b can be any fixed number in the interval (a_2, a_1), where a_j is the zero of the Airy function $\mathrm{Ai}(z)$ that is j-th closest to 0.

Proof. Once again m is of the form $m = \alpha^{-1} + o(\alpha^{-2/3})$, so the argument in [14, Sec. 3.2.3] remains valid. Recall from [14, Theorem 1] that the j-th zero of $\phi(x, y)$ satisfies the asymptotic formula

$$\varrho_j(y) = \frac{1}{e}y^{-1} - \frac{a_j}{2^{1/3}e}\,y^{-1/3} - \frac{1}{6e} + O(y^{1/3}), \quad \text{as } y \to 0^+, \tag{11}$$

where a_j is the zero of the Airy function $\mathrm{Ai}(z)$ that is j-th closest to 0. So, the choice of ρ guarantees that the circle $|x| = \rho$ contains only one pole of $A(x, y)$. ◀

4 The number of sources

We are now ready to prove Theorem 1. This section consists entirely of the proof of Theorem 1. We shall begin with the estimate of the average.

4.1 Estimate of the average

When considering the number of sources, the corresponding generating function is $A(x, y, u, 1, u)$, where the variable u indicates the number of sources. By Lemma 4,

$$A(x, y, u, 1, u) = \frac{\phi((1-u)x, y)}{\phi(x, y)}.$$

Differentiating this once with respect to u, yields

$$\partial_u A(x, y, u, 1, u)|_{u=1} = \frac{x}{\phi(x, y)} = xA(x, y).$$

Using the formula in Lemma 3, the coefficient of x^n in $\partial_u A(x, y, u, 1, u)|_{u=1}$ is

$$\frac{\partial_u A_n(y, u, 1, u)|_{u=1}}{n!(1+y)^{\binom{n}{2}}} = \mathbb{E}(S(\mathbb{D}_{ac}(n, p)))A_n(y).$$

We know that the coefficient $[x^n](xA(x, y)) = \frac{A_{n-1}(y)}{n!(1+y)^{\binom{n}{2}}}$. Hence, we obtain the following exact formula for $\mathbb{E}(S(\mathbb{D}_{ac}(n, p)))$:

$$\mathbb{E}(S(\mathbb{D}_{ac}(n, p))) = n(1+y)^{n-1}\frac{A_{n-1}(y)}{A_n(y)}. \tag{12}$$

This formula has a simple combinatorial explanation; the term $(1+y)^{n-1}A_{n-1}(y)$ is the generating function for the acyclic digraphs on $[n]$ with one marked source. Lemmas 6–8 can then be used to estimate $A_{n-1}(y)$ and $A_n(y)$ for $y = \frac{p}{1-p} = \frac{\lambda}{n-\lambda}$, and the estimate of $\mathbb{E}(S(\mathbb{D}_{ac}(n, p)))$ in Theorem 1 follows easily. The calculations were done in SageMath [17].

4.2 Asymptotic normality

For the rest of this section we shall slightly abuse notation and simply abbreviate $A_n(y, u, 1, u)$ by $A_n(y, u)$. This should not create confusion as this notation does not appear anywhere else in the paper. To prove the central limit theorem, we need to estimate $A_n(y, u)$ when u is a complex number of the form $1 + \mathcal{O}(n^{-1/2})$. It is convenient to write $v = u - 1$. So, by definition, we have

$$\frac{A_n(y, u)}{n!(1+y)^{\binom{n}{2}}} = \frac{1}{2\pi i}\oint_{|x|=\rho} \frac{\phi(-vx, y)}{\phi(x, y)x^{m+1}}dx.$$

We can then apply the saddle point method to estimate the integral on the right-hand side just as we did in Section 3 but with the extra term $\phi(-vx, y)$ in the integrand. Again the three cases $\lambda < 1$, $\lambda = 1$, and $\lambda > 1$ must be separated. The result is given in the next lemma.

▶ **Lemma 9.** *We have*

$$\frac{A_n(y, u)}{A_n(y)} \sim e^{g(n,\lambda,v)} \quad as \quad n \to \infty$$

uniformly $v = \mathcal{O}(n^{-1/2})$, *where*

$$g(n, \lambda, v) = \begin{cases} ne^{-\lambda}(v - \frac{1}{2}e^{-\lambda}v^2) & if \quad \lambda < 1 \\ \frac{n}{e\lambda}(v - \frac{1}{2e}v^2) & if \quad \lambda \geq 1. \end{cases}$$

Proof (Sketch). We sketch a proof which relies heavily on methods in [14]. However, the approach is very similar to the estimate of the integral in Equation (8). We will be using the same choices of ρ as in Subsection 3.2, depending on the sign of $\lambda - 1$. Hence, in all the cases, if $x = v\rho e^{it}$ where $t \in [-\pi, \pi]$, then $w(-vx) = W_0(v\alpha\beta) = o(1)$ for $v = \mathcal{O}(\sqrt{\alpha})$. Thus, we may use (6) of Lemma 5 to estimate $\phi(-vx, y)$. By Taylor approximation

$$w(-v\rho e^{it})^2 + 2w(-v\rho e^{it}) = 2\alpha v\rho\beta e^{it} - v^2\alpha^2\beta\rho^2 e^{2it} + \mathcal{O}(\alpha^{3/2}).$$

Recalling that $\beta = e^{\alpha/2}$, we can substitute β in the above estimate by $1 + \mathcal{O}(\alpha)$. Therefore, we obtain the following estimate of $\phi(-vx, y)$:

$$\phi(-vx, y) \sim e^{v\rho e^{it} - \frac{1}{2}v^2\alpha\rho^2 e^{2it}} \tag{13}$$

uniformly $v = \mathcal{O}(\sqrt{\alpha})$. Noting that the second term in the exponent is a bounded term, it does not have much effect on the application of the method. However, the first term does affect the local integrals. Skipping the details, the results are summarised as follows:

- For $\lambda < 1$ we choose $\rho = \frac{n}{\beta}e^{-\alpha n}$. When calculating the local integral in the current case, instead of the Gaussian integral (10), we have

  $$P(\alpha, w(\rho))e^{v\rho - \frac{1}{2}v^2\alpha\rho^2} e^{h(0)} \int_{-\alpha^c}^{\alpha^c} e^{v\rho i t + \frac{1}{2}h''(0)t^2}\, dt,$$

  where P and h are as defined in the proof of Lemma 6. The range of the integral can be extended to $(-\infty, \infty)$ with an error term smaller than any power of α. This leads to

  $$A_n(y, u) \sim A_n(y)e^{v\rho - \frac{1}{2}v^2\rho^2\frac{\alpha}{\lambda}}.$$

- For $\lambda = 1$, we choose $\rho = \frac{1}{e\alpha\beta}$. The estimate of the integral is based on [14, Lemma 9, Eq. (27)]. There, the length of the range of the local integral is much shorter, $\mathcal{O}(\alpha^c)$, where $c \in (1/2, 2/3)$. For t in that range, the estimate in (13) simplifies further to

  $$\phi(-v\rho e^{it}, y) \sim e^{v\rho - \frac{1}{2}v^2\alpha\rho^2}.$$

  Therefore, we get

  $$A_n(y, u) \sim A_n(y)e^{v\rho - \frac{1}{2}v^2\alpha\rho^2}$$

- For $\lambda > 1$ we choose $\rho = \frac{1}{e}y^{-1} - \frac{b}{2^{1/3}e}\, y^{-1/3}$ exactly as in Lemma 8. The Cauchy integral formula is used to obtain the main term, and [14, Lemma 9] to estimate the error. We get

  $$A_n(y, u) \sim A_n(y)e^{v\varrho_1(y) - \frac{1}{2}v^2\alpha\varrho_1(y)^2}$$

Expressing everything in terms n and v gives the estimate in the lemma. ◀

The central limit theorem in Theorem 1 follows directly from Lemma 9 using Hwang's quasi-power theorem [5].

5 The number of isolated vertices

5.1 Estimate of the average

The generating function in this case is $A(x, y, 1, 1, u)$, i.e., the variable u indicates the number of isolated vertices. Let us abbreviate $A_n(y, 1, 1, u)$ by $A_n^*(y, u)$. Hence, we have

$$\sum_{n=0}^{\infty} A_n^*(y, u)\frac{x^n}{n!} = e^{(u-1)x}\sum_{n=0}^{\infty} A_n(y)\frac{x^n}{n!} \tag{14}$$

Differentiating once with respect to u and substituting $u = 1$, we get $\partial_u A_n^*(y,u)|_{u=1} = nA_{n-1}(y)$. Therefore, we get

$$\mathbb{E}(I(\mathbb{D}_{ac}(n,p))) = n\frac{A_{n-1}(y)}{A_n(y)}.$$

Comparing the latter with the expression of $\mathbb{E}(S(\mathbb{D}_{ac}(n,p)))$ that we obtained in the previous section, we see that the only difference is the term $(1+y)^{n-1}$. The combinatorial interpretation is that the term $A_{n-1}(y)$ is the generating function for the acyclic digraphs on $[n]$ with one marked isolated vertex. The estimate of the mean in Theorem 2 follows from easily the above formula.

5.2 Asymptotic normality

We estimate the quotient $\frac{A_n^*(y,u)}{A_n(y)}$ when u is a complex number of the form $1 + \mathcal{O}(n^{-1/2})$. Again, we write $v = u - 1$. We begin with the follow observation:

▶ **Lemma 10.** *Let $a \neq 0$ be a fixed real constant, then the following estimate holds uniformly for $v = \mathcal{O}(n^{-1/2})$ as $n \to \infty$:*

$$(1 - av\mu^*(\lambda))^{-n/a} \sim \sum_{0\le j<n^{3/5}} \binom{n}{j} (v\mu^*(\lambda))^j e^{(a+1)\frac{j^2}{2n}}, \tag{15}$$

and

$$e^{nv\mu^*(\lambda)} \sim \sum_{0\le j<n^{3/5}} \binom{n}{j} (v\mu^*(\lambda))^j e^{\frac{j^2}{2n}}.$$

Proof. First we take the binomial expansion of the term on the left-hand side of (15), we obtain

$$(1 - av\mu^*(\lambda))^{-n/a} = \sum_{j=0}^{\infty} \binom{-\frac{n}{a}}{j} (-a)^j (v\mu^*(\lambda))^j. \tag{16}$$

Then, we simplify the summand of the above series as follows:

$$\begin{aligned}
\binom{-\frac{n}{a}}{j}(-a)^j &= \frac{n(n+a)(n+2a)\cdots(n+a(j-1))}{j!} \\
&= \binom{n}{j}\frac{(1+\frac{a}{n})(1+\frac{2a}{n})\cdots(1+\frac{(j-1)a}{n})}{(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{(j-1)}{n})} \\
&= \binom{n}{j} e^{(a+1)\frac{j^2}{2n}+\mathcal{O}(\frac{j}{n})}.
\end{aligned}$$

Thus, the series on the right-hand side of (16) is asymptotically equal to

$$\sum_{0\le j<n^{3/5}} \binom{n}{j} (v\mu^*(\lambda))^j e^{(a+1)\frac{j^2}{2n}} + \sum_{j\ge n^{3/5}} \binom{-\frac{n}{a}}{j} (-a)^j (v\mu^*(\lambda))^j.$$

We need to show that the contribution from the second term is significantly smaller. To that end, observe that

$$(1 - av\mu^*(\lambda))^{-n/a} = e^{v\mu^*(\lambda)n+\mathcal{O}(1)},$$

for $v = \mathcal{O}(n^{-1/2})$, which gives us the order magnitude of the main term. On the other hand, the bound $v = \mathcal{O}(n^{-1/2})$ and the Stirling's formula yield

$$\sum_{j \geq n^{3/5}} \binom{-\frac{n}{a}}{j} (-a)^j (v\mu^*(\lambda))^j = \mathcal{O}\left(e^{-(\frac{1}{10}+o(1))n^{3/5}\log(n)}\right)$$

which is asymptotically much smaller than the main term $(1 - av\mu^*(\lambda))^{-n/a}$.

The second estimate is done in a similar manner. Notice that

$$\binom{n}{j}(v\mu^*(\lambda))^j e^{\frac{j^2}{2n}} = (1+o(1))\frac{(nv\mu^*(\lambda))^j}{j!}$$

uniformly for $j < n^{3/5}$. Summing over $j < n^{3/5}$ and using the truncation argument above completes the proof of the lemma. ◀

▶ **Lemma 11.** *We have*

$$\frac{A_n^*(y,u)}{A_n(y)} = (1+o(1))e^{\mu^*(\lambda)vn + \frac{1}{2}a(\lambda)\mu^*(\lambda)^2 v^2 n},$$

as $n \to \infty$, uniformly for $v = \mathcal{O}(n^{-1/2})$, where

$$a(\lambda) = \begin{cases} 2\lambda - 1 & \textit{if } \lambda < 1 \\ \lambda & \textit{if } \lambda \geq 1. \end{cases}$$

Proof. If we extract the coefficient of x^n from the left and right sides of Equation (14) and dividing both by $A_n(y)$, we obtain

$$\frac{A_n^*(y,u)}{A_n(y)} = \sum_{j=0}^{n} \binom{n}{j} v^j \frac{A_{n-j}(y)}{A_n(y)}.$$

It is easy to show that $A_{n-j}(y) \leq A_n(y)$ for $y \geq 0$ from that fact that any acyclic digraph on $n - j$ vertices can be extended to an acyclic digraph on n vertices with same number of edges by simply adding j isolated vertices. This implies that the quotient $\frac{A_{n-j}(y)}{A_n(y)}$ is bounded above by 1, and the same argument we used in the proof of the previous lemma yields

$$\frac{A_n^*(y,u)}{A_n(y)} \sim \sum_{0 \leq j < n^{3/5}} \binom{n}{j} v^j \frac{A_{n-j}(y)}{A_n(y)},$$

as $n \to \infty$ uniformly for $v = \mathcal{O}(n^{-1/2})$. By making use of Lemmas 6–8, we obtain

$$\frac{A_{n-j}(y)}{A_n(y)} = (1+o(1))\mu^*(\lambda)^j \times \begin{cases} e^{\lambda j^2/n} & \text{if } \lambda < 1 \\ e^{\frac{1}{2}(\lambda+1)j^2/n} & \text{if } \lambda \geq 1 \end{cases}$$

uniformly for $0 \leq j \leq n^{3/5}$. Putting two latter estimates together and applying Lemma 10 for the appropriate values of a complete the proof the lemma. ◀

The central limit theorem in Theorem 2 follows easily from Lemma 11.

6 Conclusion

We consider a random DAG model $\mathbb{D}_{ac}(n, p)$ on the vertex set $[n] = \{1, 2, \ldots, n\}$ which naturally generalises the uniform random DAG model on $[n]$. We established that if p is of the form λ/n, where $\lambda > 0$ is fixed, then the number of sources, sinks and isolated vertices are all asymptotically normal with means proportional to n. For further investigation, it would be interesting to know the typical shape of such a random DAG. We could look at the distribution of the height, which was considered by McKay [12] for the case $p = \frac{1}{2}$. We expect the height to be much smaller than n for $p = \lambda/n$, but it seems that even an estimate of the expectation of this parameter would require a significant amount of work. Alternatively, we could also investigate the number of vertices at a given level (the level of a vertex v in a DAG is the length of the longest directed path from a source to v). McKay also considered this parameter for $p = \frac{1}{2}$ in [12]. The case where p is tending to zero seems to be more challenging. It is not even clear if the graphic generating functions for these parameters can be written in forms that we can analyse. Hence, these problems are left for future work.

References

1. A. D. Barbour. Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.*, 92(2):349–359, 1982. doi:10.1017/S0305004100059995.
2. A. D. Barbour, MichałKaroński, and Andrzej Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989. doi:10.1016/0095-8956(89)90014-2.
3. É. de Panafieu, S. Dovgal, D. Ralaivaosaona, V. Rasendrahasina, and S. Wagner. The birth of the strong components, 2020. arXiv:2009.12127.
4. I. M. Gessel. Counting acyclic digraphs by sources and sinks. *Discrete Mathematics*, 160(1):253–258, 1996.
5. H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19(3):329–343, 1998. doi:10.1006/eujc.1997.0179.
6. M. Karoński and A. Ruciński. Poisson convergence and semi-induced properties of random graphs. *Math. Proc. Cambridge Philos. Soc.*, 101(2):291–300, 1987. doi:10.1017/S0305004100066664.
7. R. M. Karp. The transitive closure of a random digraph. *Random Structures & Algorithms*, 1(1):73–93, 1990.
8. V. A. Liskovec. On the enumeration of strongly connected directed graphs. *Dokl. Akad. Nauk BSSR*, 17:1077–1080, 1163, 1973.
9. V. A. Liskovets. On the number of maximal vertices of a random acyclic digraph. *Theory of Probability & Its Applications*, 20(2):401–409, 1976.
10. T. Łuczak. The phase transition in the evolution of random digraphs. *Journal of Graph Theory*, 14:217–223, 1990.
11. T. Łuczak and T. Seierstad. The critical behavior of random digraphs. *Random Structures & Algorithms*, 35:271–293, 2009.
12. B. D. Mckay. On the shape of a random acyclic digraph. *Mathematical Proceedings of the Cambridge Philosophical Society*, 106(3):459–465, 1989. doi:10.1017/S0305004100068195.
13. F. W. Olver, D. W. Lozier, Ronald F. Boisvert, and C. W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, USA, 1st edition, 2010.
14. D. Ralaivaosaona, V. Rasendrahasina, and S. Wagner. On the probability that a random digraph is acyclic. In *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020)*, volume 159 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 25:1–25:18, 2020.
15. R. W. Robinson. Counting labelled acyclic digraphs. In *New Directions in Graph Theory (Ed. F. Harary)*. New York: Academic Press, 1973.

16 R. P. Stanley. Acyclic orientations of graphs. *Discrete Mathematics*, 5(2):171–178, 1973.
17 The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 8.9)*, 2019. https://www.sagemath.org.
18 P. Zbigniew. On the number of vertices of given degree in a random graph. *J. Graph Theory*, 8(1):167–170, 1984. doi:10.1002/jgt.3190080119.