Asymptotic enumeration of strongly connected digraphs by vertices and edges

Xavier Pérez-Giménez* and Nicholas Wormald†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON, Canada

Abstract

We derive an asymptotic formula for the number of strongly connected digraphs with n vertices and m arcs (directed edges), valid for $m - n \to \infty$ as $n \to \infty$ provided $m = O(n \log n)$. This fills the gap between Wright's results which apply to $m = n + O(1)$, and the long-known threshold for m, above which a random digraph with n vertices and m arcs is likely to be strongly connected.

1 Introduction

One of the most fundamental properties of a directed graph (digraph), and possibly the most useful for communication networks, is that of being *strongly connected*, that is, possessing directed paths both ways between every pair of vertices. It was long ago shown by Moon and Moser [7] that almost all of the 2^{n^2} digraphs with n vertices are strongly connected due to having paths of length 2 between each pair of vertices. So, from the asymptotic enumeration perspective, a more interesting problem is the enumeration of strongly connected digraphs on n vertices with m arcs (i.e. directed edges). In this paper, all digraphs are labelled. Our results cover only *simple* digraphs (i.e. digraphs with no multiple arcs), but unless otherwise stated we allow digraphs to have loops. We also give results for digraphs in which loops are forbidden, which we refer to as *loop-free* digraphs.

Palásti [8] determined the threshold of strong connectivity, as follows. Let α be fixed and define $m(\alpha, n) = \lfloor n \log n + \alpha n \rfloor$. (In this paper, all logarithms are natural.) Then, for a random directed graph having n vertices and m arcs, so that each of the $\binom{n^2}{m}$ possible choices is equiprobable, the probability that the digraph is strongly connected tends to $\exp(-2e^{-\alpha})$ as $n \to \infty$. Multiplying this probability by $\binom{n^2}{m}$ consequently gives an asymptotic formula for the number $S(n, m)$ of strongly connected digraphs with n vertices and m arcs, for such m. This also easily implies that $S(n, m) \sim \binom{n^2}{m}$ if $m = m(\alpha_n, n)$ with $\alpha_n \to \infty$. On the other

*Partially supported by the Province of Ontario under the Post-Doctoral Fellowship (PDF) Program. Current address: Max-Planck-Institut für Informatik, Department 1: Algorithms and Complexity, Saarbrücken, Germany

†Supported by the Canada Research Chairs Program and NSERC.

hand, Wright [12] obtained recurrences for the exact value of $S(n,m)$ when $m = n + O(1)$. (We must require $m \geq n$ to avoid the failure to be strongly connected for trivial reasons.) In this paper, we fill the entire gap between these results, deriving an asymptotic formula for $S(n,m)$, valid for $m - n \to \infty$ as $n \to \infty$ provided $m = O(n \log n)$. Our main result is as follows.

Theorem 1.1. *For $m = m(n) = O(n \log n)$ such that $m - n \to \infty$, the number of strongly connected digraphs with n vertices and m arcs is asymptotic to*

$$\frac{(m-1)!(e^\lambda - 1)^{2n}}{2\pi(1+\lambda-c)\lambda^{2m}} \exp(-\lambda^2/2) \frac{e^\lambda(e^\lambda - 1 - \lambda)^2}{(e^{2\lambda} - e^\lambda - \lambda)(e^\lambda - 1)}, \tag{1}$$

where $c = m/n > 1$ and λ is determined by the equation $c = \lambda e^\lambda/(e^\lambda - 1)$.

Note. In particular, if $c \to 1$ then the expression (1) simplifies asymptotically to

$$\frac{(m-1)!(e^\lambda - 1)^{2n}}{6\pi\lambda^{2m}}, \tag{2}$$

whilst if $c \to \infty$ then (1) is asymptotic to

$$\frac{(m-1)!(e^\lambda - 1)^{2n}}{2\pi\lambda^{2m}} \exp(-\lambda^2/2). \tag{3}$$

Note also that the rate of convergence in the asymptotic formula will depend on the rate of convergence to infinity of $m - n$.

Our result has counterparts for undirected graphs. An asymptotic formula for the number of connected graphs with n vertices and m edges was given for all m such that $m - n \to \infty$ as $n \to \infty$ by Bender, Canfield and McKay [1]. This improved the range of m for which earlier estimates were found, and also the bounds on the error term. A simpler approach to the same problem was given in [11]. This begins by counting connected graphs with no end-vertices, and then considers the number of ways to attach a forest. One of the ways used there to count connected 2-cores was to count connected kernels, which have no vertices of degree 2, and insert vertices of degree 2 into their edges, and another way was based on eliminating isolated cycles by inversion. In the present paper, for the case $m = O(n)$ we use this first of these two alternatives. This has some advantage in providing direct information on properties of the kernel, such as was used in [5] for studying long cycles in the supercritical random graph. In a similar way, we can study the analogous structure for a digraph, which we call its heart. For $m/n \to \infty$ we use a rather different approach to show that random digraphs with all in- and out-degrees at least 1 are strongly connected with high probability.

Our argument requires a formula for the number of digraphs with all in- and outdegrees at least 1 and given number of arcs, which we obtain using the method for counting graphs with given minimum degree developed by Pittel and the second author in [10].

Theorem 1.2. *For $m = m(n) = O(n \log n)$ such that $m - n \to \infty$, the number of digraphs on n vertices and m arcs with all in- and outdegrees at least 1 is asymptotic to*

$$\frac{(m-1)!(e^\lambda - 1)^{2n}}{2\pi(1+\lambda-c)\lambda^{2m}} \exp(-\lambda^2/2),$$

where $c = m/n$ and λ is determined by $c = \lambda e^\lambda/(e^\lambda - 1)$.

Using the same method, we also extend this to digraphs with all outdegrees at least k^+ and all indegrees at least k^-.

Theorem 1.3. *Fix positive integers k^+ and k^-. For $m = O(n \log n)$ such that $m - k^+ n \to +\infty$ and $m - k^- n \to +\infty$, the number of digraphs on n vertices, m arcs, outdegrees at least k^+ and indegrees at least k^- is asymptotic to*

$$\frac{(m-1)!(f_{k^-}(\lambda^-)f_{k^+}(\lambda^+))^n}{2\pi\sqrt{(1+\eta^+-c)(1+\eta^--c)}(\lambda^-\lambda^+)^m}\exp(-\lambda^-\lambda^+/2),$$

where $c = m/n$,

$$f_k(\lambda) = \sum_{i \geq \max\{k,0\}} \lambda^i/i!,$$

λ^+ and λ^- are the unique positive roots of

$$c = \lambda^+ f_{k^+-1}(\lambda^+)/f_{k^+}(\lambda^+), \quad c = \lambda^- f_{k^--1}(\lambda^-)/f_{k^-}(\lambda^-)$$

respectively, and

$$\eta^+ = (\lambda^+)^2 f_{k^+-2}(\lambda^+)/f_{k^+-1}(\lambda^+), \quad \eta^- = (\lambda^-)^2 f_{k^--2}(\lambda^-)/f_{k^--1}(\lambda^-).$$

The results stated so far refer to digraphs that are allowed to have loops but not multiple arcs. In Section 7 we extend these results to the case when loops are forbidden, and obtain the following analogues of Theorems 1.1 and 1.3.

Theorem 1.4. *For $m = O(n \log n)$ such that $m - n \to +\infty$, the number of strongly connected loop-free digraphs with n vertices and m arcs is asymptotic to*

$$\frac{(m-1)!(e^\lambda-1)^{2n}}{2\pi(1+\lambda-c)\lambda^{2m}}\exp\big(-c(1-e^{-\lambda})^2-\lambda^2/2\big)\frac{e^\lambda(e^\lambda-1-\lambda)^2}{(e^{2\lambda}-e^\lambda-\lambda)(e^\lambda-1)},$$

where c and λ are as in Theorem 1.1.

Note that, for Theorem 1.4, the only effect of forbidding loops was to introduce the extra factor $\exp(-c(1-e^{-\lambda})^2)$.

Theorem 1.5. *Fix positive integers k^+ and k^-, and recall the notation of Theorem 1.3. For $m = O(n \log n)$ such that $m - k^+ n \to +\infty$ and $m - k^- n \to +\infty$, the number of loop-free digraphs on n vertices, m arcs, outdegree at least k^+ and indegree at least k^- is asymptotic to*

$$\frac{(m-1)!(f_{k^-}(\lambda^-)f_{k^+}(\lambda^+))^n}{2\pi\sqrt{(1+\eta^+-c)(1+\eta^--c)}(\lambda^-\lambda^+)^m}\exp(-c-\lambda^-\lambda^+/2).$$

For Theorem 1.5, forbidding loops only gave the extra factor e^{-c}.

Cooper and Frieze [3, Theorem 3(vi)] obtained a significant result relevant to this problem, in the form of the asymptotic probability that a random digraph with given degree sequence is strongly connected, under certain assumptions on the degree sequence. It would

be rather straightforward to combine this with our Theorem 1.2, along with properties of degree sequences which we use in our paper, to deduce an asymptotic formula for $S(n,m)$ when $m/n > 1$ is bounded away from 1 and is bounded. For completeness, we derive this case of the formula in a different way, following the same approach as we use for the case $m/n \to 1$, which we consider in Section 5.

Boris Pittel [9] has independently investigated the second approach of [11] mentioned above. Applying it to this problem in the loop-free case, he has independently obtained a formula similar to that in Theorem 1.4 under the stronger restrictions that $m - n \gg n^{2/3}$ and $m = O(n)$, but also including an explicit error estimate. With our method, it would be possible to include explicit error estimates, but with the present argument the bounds for some ranges of m would be rather weak, in particular at the points where we use the method of moments. Our approach uses comparisons between probability spaces which sometimes complicates attempts at sharpening error bounds, but leads to the study of what we call the heart of a digraph, which may prove useful for other distributional results for properties of random digraphs.

2 Basics and notation

2.1 Truncated Poisson distribution

We consider a discrete probability distribution that will be used many times in the argument. Given $\lambda > 0$ and a nonnegative integer k, we say that a random variable (r.v.) Y has a *k-truncated Poisson distribution of parameter λ* (or simply $Y \overset{\mathrm{d}}{\sim} \mathrm{TPo}_k(\lambda)$) if

$$\mathbf{P}(Y = i) = \begin{cases} \dfrac{\lambda^i}{f_k(\lambda)\, i!} & \text{if } i \geq k, \\ 0 & \text{if } 0 \leq i < k, \end{cases}$$

where $f_k(\lambda) = \sum_{i \geq k} \lambda^i / i!$. For later convenience we also define $f_k(\lambda) = e^\lambda$ for integer $k < 0$.

We first give a rough tail bound for a random variable $Y \overset{\mathrm{d}}{\sim} \mathrm{TPo}_k(\lambda)$ for constant k but λ possibly depending on n. Consider constants $A > B > e$, and let p be a constant nonnegative integer. Then for $j \geq \max\{Ae\lambda, k\}$ we have

$$\begin{aligned} \mathbf{E}([Y]_p\, 1_{Y \geq j}) &= \frac{1}{f_k(\lambda)} \left(\sum_{j \leq i < j+p} [i]_p \frac{\lambda^i}{i!} + \sum_{i \geq j+p} [i]_p \frac{\lambda^i}{i!} \right) \leq \frac{1}{f_k(\lambda)} \left(p(j+p)^p \frac{\lambda^j}{j!} + \lambda^p \sum_{i \geq j} \frac{\lambda^i}{i!} \right) \\ &= O\left(\frac{j^p + \lambda^p}{f_k(\lambda)} (e\lambda/j)^j \right) = O\left(B^{-j}\right), \qquad (\text{as } j \to \infty) \end{aligned}$$

where we use $[x]_k$ to denote the falling factorial $x(x-1)\cdots(x-k+1)$ throughout this paper. In particular,

$$\mathbf{P}(Y \geq j) = O\left(B^{-j}\right), \quad \mathbf{E}(Y\, 1_{Y \geq j}) = O\left(B^{-j}\right) \quad \text{and} \quad \mathbf{E}([Y]_2\, 1_{Y \geq j}) = O\left(B^{-j}\right). \tag{4}$$

Our main use of the $\mathrm{TPo}_k(\lambda)$ distribution is to allow us to make computations on the multinomial distribution truncated from below. The following lemma establishes a connection

between these distributions, and will be used throughout the paper, often without an explicit mention. (See for example [2, Section 2] for a proof of this lemma.)

Lemma 2.1. *Distribute $M \geq kN$ distinguishable balls randomly into N distinguishable bins u.a.r. subject to the condition that each bin receives at least $k \geq 1$ balls. Let Y_i be the numbers of balls in bin i. Then the joint distribution of $Y_1, \ldots, Y_N$ is the same as that of N independent copies of $\mathrm{TPo}_k(\lambda)$ for arbitrary $\lambda > 0$ conditional upon $Y_1 + \cdots + Y_N = M$.*

It is easy to see that a variable $Y \overset{\mathrm{d}}{\sim} \mathrm{TPo}_k(\lambda)$ has $\mathbf{E}Y = c$ given by

$$c = \frac{\lambda f_{k-1}(\lambda)}{f_k(\lambda)}. \tag{5}$$

Henceforth, given $c > k$, we assume that λ is set equal to the unique (by [10, Lemma 1]) positive root of this equation. We also define

$$\eta = \frac{\lambda^2 f_{k-2}(\lambda)}{f_{k-1}(\lambda)}. \tag{6}$$

Elementary computations show that, for such choice of λ and η, we have $\mathbf{E}(Y(Y-1)) = \eta c$. More properties of the $\mathrm{TPo}_k(\lambda)$ distribution are given in [10]. It is easy to check that $0 < \lambda \leq c$ in all cases. From [10, Theorem 4(a)] we have the following.

Lemma 2.2. *Let $M = O(N \log N)$ be integer such that $r := M - kN \to \infty$ and put $c = M/N$. Let $Y_1, \ldots, Y_N$ be i.i.d. random variables with $\mathrm{TPo}_k(\lambda)$ distribution, for fixed k, where λ is determined from c in (5), and define η as in (6). Then, as $N \to \infty$,*

$$\mathbf{P}(Y_1 + \cdots + Y_N = M) \sim \frac{1}{\sqrt{2\pi N c(1+\eta-c)}} = \Theta\left(1/\sqrt{r}\right).$$

Throughout the paper, we mostly focus our attention to the case $k = 1$ and simply refer to the $\mathrm{TPo}_1(\lambda)$ distribution as $\mathrm{TPo}(\lambda)$ or simply *truncated Poisson*. In this particular case, (5) can be rewritten as

$$c = \frac{\lambda e^\lambda}{e^\lambda - 1}, \tag{7}$$

and moreover we have $\eta = \lambda$.

On several occasions we use Chernoff bounds for a binomially distributed $\mathrm{Bin}(n, p)$ random variable X in the common form

$$\mathbf{P}(|X - np| > a) < 2e^{-2a^2/n}, \tag{8}$$

or the variation more useful when p is small:

$$\mathbf{P}(|X - np| \geq a) \leq 2e^{-a^2/3np} \quad \text{for} \quad a \leq np \tag{9}$$

(see for example Janson, Łuczak and Ruciński [4, Cor. 2.3]).

We close this subsection with some rather technical lemmas on independent variables with $\mathrm{TPo}_k(\lambda)$ distribution.

Lemma 2.3. *Let $Y_1, \dots, Y_N$ be independent r.v.s with $\mathrm{TPo}_k(\lambda)$ distribution, for fixed k and for $0 < \lambda \le \log N$. Put $C = \mathbf{E}Y_1$. Then for any $t \ge \sqrt{N}\log^2 N$ we have*

$$\mathbf{P}\left(\Big|\sum_{i=1}^{N} Y_i - CN\Big| > t\right) = O\left(e^{-(t^2/8N)^{1/3}}\right),$$

asymptotically as $N \to \infty$.

Proof. Let $Y_{\max} = \max_i\{Y_i\}$. Setting $\Delta = (t^2/8N)^{1/3}$, we have $\mathbf{P}(Y_{\max} > \Delta) \le N\mathbf{P}(Y_1 > \Delta) = O(e^{-\Delta})$ by (4) and since $\Delta = \Omega(\log^{4/3} N)$. Now define

$$W_i = Y_i - C \quad \text{and} \quad W_i^* = W_i\, 1_{Y_i \le \Delta},$$

and again from (4) deduce

$$|\mathbf{E}W_i^*| = |-\mathbf{E}(W_i\, 1_{Y_i > \Delta})| \le \mathbf{E}(Y_i\, 1_{Y_i > \Delta}) = O\left(e^{-\Delta}\right). \tag{10}$$

Moreover, we have $-C \le W_i^* \le \Delta - C$, and then $|W_i^* - \mathbf{E}W_i^*| < \Delta$, so by the Azuma-Hoeffding inequality

$$\mathbf{P}\Big(\Big|\sum_{i=1}^{N}(W_i^* - \mathbf{E}W_i^*)\Big| \ge t/2\Big) \le 2\exp\left(\frac{-t^2}{8\Delta^2 N}\right) = 2e^{-\Delta}. \tag{11}$$

$$\begin{aligned}
\mathbf{P}\left(\Big|\sum_{i=1}^{N} Y_i - CN\Big| > t\right) &\le \mathbf{P}(Y_{\max} > \Delta) + \mathbf{P}\Big(\Big|\sum_{i=1}^{N} W_i^*\Big| > t\Big) \\
&\le O\left(e^{-\Delta}\right) + \mathbf{P}\Big(\Big|\sum_{i=1}^{N}(W_i^* - \mathbf{E}W_i^*)\Big| > t - \Big|\sum_{i=1}^{N}\mathbf{E}W_i^*\Big|\Big) \\
&\le O\left(e^{-\Delta}\right) + \mathbf{P}\Big(\Big|\sum_{i=1}^{N}(W_i^* - \mathbf{E}W_i^*)\Big| > t/2\Big) \\
&= O\left(e^{-\Delta}\right),
\end{aligned}$$

where we used (11) and the fact that $|\sum_{i=1}^{N}\mathbf{E}W_i^*| < t/2$, which follows from (10). □

The following result was essentially shown in [10].

Lemma 2.4. *Let $Y_1, \dots, Y_N$ be independent r.v.s with $\mathrm{TPo}_k(\lambda)$ distribution, for fixed k and for $0 < \lambda \le \log N$. Put $C = \mathbf{E}(Y_1(Y_1 - 1))$. Then*

$$\mathbf{P}\left(\Big|\sum_{i=1}^{N} Y_i(Y_i - 1) - CN\Big| > 4N^{1/2}\log^8 N\right) = O\big(\exp(-\log^3 N)\big),$$

asymptotically as $N \to \infty$.

Proof. The statement in the lemma comes directly from equation (33) in [10], considering (16), (22), (28), (29) and Lemmas 1 and 2 of that paper. See also the proof of Lemma 2.3 which uses the same method in full detail. □

We will use the following for $k = 1, 2$.

Lemma 2.5. *Let $k \geq 1$ be an integer, and let $Y_1, \ldots, Y_N$ be independent $\mathrm{TPo}_k(\lambda)$ r.v.s. Consider N bins, place Y_i balls in bin i ($i = 1, \ldots, N$), and then select each ball independently with probability $q \leq 1/2$ where $Nq \geq \log^2 N$. Then the number X of bins containing at least one selected ball satisfies*

$$\mathbf{P}(|X - \mathbf{E}X| > \sqrt{\mathbf{E}X} \log N) = e^{-\Omega(\log^2 N)}$$

asymptotically as $N \to \infty$, and moreover

$$\mathbf{E}X/n > kq(1 - (k-1)/4 + (2^{-k}/k)\mathbf{P}(Y_1 \geq k+1)).$$

Proof. Let q' be the probability that a bin contains at least one selected ball. We have

$$\begin{aligned} 1 - q' &< (1-q)^k \mathbf{P}(Y_i = k) + (1-q)^{k+1}\mathbf{P}(Y_i \geq k+1) \\ &= (1-q)^k - q(1-q)^k \mathbf{P}(Y_i \geq k+1). \end{aligned}$$

Using the elementary bound $(1-q)^k \leq 1 - kq + \binom{k}{2}q^2$ and the fact that $q \leq 1/2$, we obtain

$$\begin{aligned} q' &> kq(1 - (k-1)q/2) + q(1-q)^k \mathbf{P}(Y_i \geq k+1) \\ &\geq kq(1 - (k-1)/4 + (2^{-k}/k)\mathbf{P}(Y_i \geq k+1)), \end{aligned} \tag{12}$$

and trivially $q' \geq q$ in any case. Since $X \overset{\mathrm{d}}{\sim} \mathrm{Bin}(N, q')$, it follows by (9) that

$$\mathbf{P}\big(|X - Nq'| > \sqrt{Nq'} \log N\big) < 2e^{-\log^2 N/3}. \tag{13}$$

□

2.2 Probability spaces of digraphs and degree sequences

Let $\mathcal{G}(n, m)$ be the set of digraphs on n labelled vertices and m arcs. In our definition of digraph we allow loops but not multiple arcs (in particular, each vertex has at most one loop). It is a simple matter to adjust our arguments for loop-free digraphs (see Section 7). For a given digraph in $\mathcal{G}(n, m)$, let $\vec{d^+} = (d_1^+, \ldots, d_n^+)$ and $\vec{d^-} = (d_1^-, \ldots, d_n^-)$ denote respectively the sequences of out- and indegrees of the vertices. The degree of vertex i is defined to be the tuple $d_i = (d_i^+, d_i^-)$, so the joint in- and outdegree sequences can be represented by $\vec{d} = (d_1, \ldots, d_n)$. For feasibility, it is necessary that

$$\sum_{i=1}^n d_i^+ = \sum_{i=1}^n d_i^- = m. \tag{14}$$

Let $c = m/n$ and assume that $c > 1$ throughout the article, though m and hence c are functions of n. Let $\mathcal{G}_{1,1}(n, m)$ be the set of digraphs in $\mathcal{G}(n, m)$ such that $d_i^+, d_i^- \geq 1$ for all $i \in \{1, \ldots, n\}$. (Note that this is a necessary condition for strong connectedness when $n > 1$.) Elements of $\mathcal{G}_{1,1}(n, m)$ we call $(1,1)$-*dicores* or simply *dicores*. We also write $\mathcal{G}(n, m)$ and $\mathcal{G}_{1,1}(n, m)$ to denote the corresponding uniform probability spaces. We define $r = m - n = (c-1)n$ and

assume $r \to \infty$. We distinguish three subcases: very sparse, with $r = o(n)$ or equivalently $c \to 1$; moderately sparse, with $r = \Theta(n)$; and a denser case, with $c \to \infty$ but $c = O(\log n)$. (All logarithms are natural unless otherwise specified.)

Let $\mathcal{D}$ be the set of sequences $\vec{d} = (d_1, \dots, d_n)$, with $d_i = (d_i^+, d_i^-)$ for $i \in \{1, \dots, n\}$, where the $2n$ entries d_i^+ and d_i^- are positive integers. Let $\widehat{\mathcal{D}}$ be the subset of sequences in $\mathcal{D}$ satisfying the total sum conditions (14). Note that $\widehat{\mathcal{D}}$ coincides with the set of all possible degree sequences of dicores in $\mathcal{G}_{1,1}(n, m)$. Given any $\vec{d} \in \widehat{\mathcal{D}}$, let $\mathcal{G}(\vec{d})$ denote the set (and also the corresponding uniform probability space) of digraphs with degree sequence $\vec{d}$. Also consider the usual directed pairing model $\mathcal{P}(\vec{d})$, defined as follows. Take n bins, where the i-th bin contains points of two types, namely d_i^+ *out-points* and d_i^- *in-points*, and consider a random matching of the m out-points with the m in-points. Each element in $\mathcal{P}(\vec{d})$ corresponds to a multidigraph in the obvious way, and the restriction to *simple* digraphs (i.e. with no multiple arcs) generated this way is uniform.

In order to study the distribution of degree sequences of $\mathcal{G}_{1,1}(n, m)$, it will prove useful to turn the sets $\mathcal{D}$ and $\widehat{\mathcal{D}}$ into suitable probability spaces, as follows. Random degree sequences $\vec{d} \in \mathcal{D}$ are chosen by taking the $2n$ entries d_i^+ and d_i^- as independent copies of $\mathrm{TPo}(\lambda)$. Let Σ be the event in $\mathcal{D}$ that (14) holds, and define $\widehat{\mathcal{D}}$ to be the corresponding conditional probability space. Moreover, let $\mathcal{P}_{1,1}(n, m)$ be the probability space of random pairings in $\mathcal{P}(\vec{d})$ where the degree sequence $\vec{d}$ is drawn from the distribution of $\widehat{\mathcal{D}}$ defined above. Each pairing in $\mathcal{P}_{1,1}(n, m)$ corresponds to a multidigraph, and as will become apparent later the restriction of $\mathcal{P}_{1,1}(n, m)$ to simple digraphs generates elements of $\mathcal{G}_{1,1}(n, m)$ uniformly.

We also need the notation $d_{\max}^+ = \max\{d_i^+ : 1 \le i \le n\}$ and $d_{\max}^- = \max\{d_i^- : 1 \le i \le n\}$.

3 Asymptotic enumeration of dicores

Here we prove Theorems 1.2 and 1.3 by adapting the main argument of [10]. Before that, we need some lemmata. The following result is an immediate consequence of Theorem 4.6 in [6] by McKay (we just need to use the standard interpretation of digraphs with loops as bipartite graphs).

Lemma 3.1 (McKay)**.** *Let $\vec{d} \in \widehat{\mathcal{D}}$ be a sequence of degrees and suppose that $d_{\max}^+, d_{\max}^- \le \Delta$ for some $\Delta = o(m^{1/4})$. Then the probability that a random element of $\mathcal{P}(\vec{d})$ has no multiple arcs is*

$$\exp\left(-\frac{1}{2m^2}\sum_{i,j=1}^{n} d_i^+(d_i^+ - 1)d_j^-(d_j^- - 1) + O\left(\frac{\Delta^4}{m}\right)\right),$$

uniformly for all $\vec{d}$.

The following technical result estimates the probability that a degree sequence in $\mathcal{D}$ satisfies (14), and averages the probability that a random pairing is simple over any subset of degree sequences with that property. Here λ and c are defined as in Theorem 1.2.

Lemma 3.2. *Assume that $m - n \to \infty$ and $m = O(n \log n)$.*

(a) $\mathbf{P}_{\mathcal{D}}(\Sigma) \sim \dfrac{1}{2\pi nc(1+\lambda-c)} = \Theta(1/(m-n))$.

Moreover, if S is the event that a random pairing in $\mathcal{P}(\vec{d})$ or $\mathcal{P}_{1,1}(n,m)$ is simple, then

(b) $\mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(S) = \mathbf{E}_{\widehat{\mathcal{D}}}\big(\mathbf{P}_{\mathcal{P}(\vec{d})}(S)\big) \sim e^{-\lambda^2/2}$;

(c) for any r.v. X on $\widehat{\mathcal{D}}$ satisfying $|X| \le x$ for some fixed constant $x \in \mathbb{R}$,

$$\mathbf{E}_{\widehat{\mathcal{D}}}\big(\mathbf{P}_{\mathcal{P}(\vec{d})}(S)\cdot X\big) = (1+o(1))\, e^{-\lambda^2/2}\, \mathbf{E}_{\widehat{\mathcal{D}}}X + O\left(e^{-\log^3 n}\right).$$

Proof. From Lemma 2.2, the independent events $\sum_i d_i^+ = m$ and $\sum_i d_i^- = m$ each have probability $(1+o(1))/\sqrt{2\pi nc(1+\lambda-c)}$, which gives (a). Note that (b) follows from (c) by setting $X = 1$, since the bound on m implies $\lambda = O(\log n)$, so it only remains to prove (c). For this, we follow the proof of [10, Theorem 4(b)] almost exactly.

We require some definitions. Let $F = F(\vec{d}) = \mathbf{P}_{\mathcal{P}(\vec{d})}(S)$ and

$$\tilde{F} = \exp\left(-\frac{1}{2}D^+D^-\right),$$

where

$$D^+ = \frac{1}{m}\sum_{i=1}^{n} d_i^+(d_i^+-1) \qquad \text{and} \qquad D^- = \frac{1}{m}\sum_{j=1}^{n} d_j^-(d_j^--1).$$

We set $\Delta = \log^3 n$, and let $\mathcal{B}_1$ denote the 'bad' event that $d^+_{\max} > \Delta$ or $d^-_{\max} > \Delta$. From (4) we obtain $\mathbf{P}_{\mathcal{D}}(\mathcal{B}_1) \le 2n\mathbf{P}(Y > \Delta) = O(nB^{-\Delta})$. Then, we use the result from (a) to deduce that $\mathbf{P}_{\widehat{\mathcal{D}}}(\mathcal{B}_1) \le \mathbf{P}_{\mathcal{D}}(\mathcal{B}_1)/\mathbf{P}_{\mathcal{D}}(\Sigma) = O(n^2c(1+\lambda-c)B^{-\Delta}) = O\big(e^{-\log^3 n}\big)$.

In view of Lemma 3.1 and bearing in mind that $0 \le F, \tilde{F} \le 1$ and $|X| \le x$, we can write

$$\begin{aligned}
\mathbf{E}_{\widehat{\mathcal{D}}}(FX) &= \mathbf{E}_{\widehat{\mathcal{D}}}(FX\,1_{\mathcal{B}_1}) + \mathbf{E}_{\widehat{\mathcal{D}}}(FX\,1_{\overline{\mathcal{B}_1}}) \\
&= O(\mathbf{P}_{\widehat{\mathcal{D}}}(\mathcal{B}_1)) + (1+O(\Delta^4/m))\mathbf{E}_{\widehat{\mathcal{D}}}(\tilde{F}X\,1_{\overline{\mathcal{B}_1}}) \\
&= O\big(e^{-\log^3 n}\big) + (1+O(\Delta^4/m))\mathbf{E}_{\widehat{\mathcal{D}}}(\tilde{F}X\,1_{\overline{\mathcal{B}_1}}). \qquad (15)
\end{aligned}$$

Simple computations show that $\mathbf{E}D^+ = \mathbf{E}D^- = \lambda$ (with D^+ and D^- independent). Set $t = 8n^{-1/2}\log^9 n$, and define $\mathcal{B}_2$ to be the 'bad' event that $|D^+D^-/2 - \lambda^2/2| > t$. Whenever $\mathcal{B}_2$ does not hold, we have $\tilde{F} = \exp(-\lambda^2/2 + O(t)) = (1+O(t))\exp(-\lambda^2/2)$, so

$$\begin{aligned}
\mathbf{E}_{\widehat{\mathcal{D}}}(\tilde{F}X\,1_{\overline{\mathcal{B}_1}}) &= \mathbf{E}_{\widehat{\mathcal{D}}}(\tilde{F}X\,1_{\overline{\mathcal{B}_1}\wedge\mathcal{B}_2}) + \mathbf{E}_{\widehat{\mathcal{D}}}(\tilde{F}X\,1_{\overline{\mathcal{B}_1}\wedge\overline{\mathcal{B}_2}}) \\
&= O(\mathbf{P}_{\widehat{\mathcal{D}}}(\mathcal{B}_2)) + (1+O(t))\, e^{-\lambda^2/2}\, \mathbf{E}_{\widehat{\mathcal{D}}}(X). \qquad (16)
\end{aligned}$$

It only remains to bound $\mathbf{P}_{\widehat{\mathcal{D}}}(\mathcal{B}_2)$. Set $s = t/(2\log n) = 4n^{-1/2}\log^8 n$, and note that if $|D^+ - \lambda| \le s$ and $|D^- - \lambda| \le s$ then

$$|D^+D^-/2 - \lambda^2/2| \le \frac{1}{2}(|D^+-\lambda||D^--\lambda| + \lambda|D^+-\lambda| + \lambda|D^--\lambda|) \le \frac{s^2+2s\log n}{2} \le t.$$

Therefore, by Lemma 2.4,

$$\mathbf{P}_{\widehat{\mathcal{D}}}(\mathcal{B}_2) \le \mathbf{P}_{\widehat{\mathcal{D}}}(|D^+ - \lambda| > s) + \mathbf{P}_{\widehat{\mathcal{D}}}(|D^- - \lambda| > s) = O\big(e^{-\log^3 n}\big). \tag{17}$$

Part (c) in the statement follows by combining (15), (16) and (17). □

Now we are in good shape to prove the theorem.

Proof of Theorem 1.2. Observe that $|\mathcal{P}(\vec{d})| = m!$, and that each simple digraph with degree sequence $\vec{d}$ comes from exactly $\prod_{i=1}^n d_i^+! d_i^-!$ different pairings in $\mathcal{P}(\vec{d})$. Thus

$$|\mathcal{G}(\vec{d})| = \frac{m!\mathbf{P}_{\mathcal{P}(\vec{d})}(S)}{\prod_{i=1}^n d_i^+! d_i^-!},$$

where S denotes the event that a random pairing in $\mathcal{P}(\vec{d})$ has no multiple arcs. Define

$$Q = \sum_{\vec{d}\in\widehat{\mathcal{D}}} \prod_{i=1}^n \frac{1}{d_i^+! d_i^-!} = \frac{(e^\lambda - 1)^{2n}}{\lambda^{2m}} \mathbf{P}_{\mathcal{D}}(\Sigma).$$

Therefore, summing over all degree sequences, we can write

$$\begin{aligned}
|\mathcal{G}_{1,1}(n,m)| &= \sum_{\vec{d}\in\widehat{\mathcal{D}}} \frac{m!\mathbf{P}_{\mathcal{P}(\vec{d})}(S)}{\prod_{i=1}^n d_i^+! d_i^-!} \\
&= m!Q\mathbf{E}_{\widehat{\mathcal{D}}}\left(\mathbf{P}_{\mathcal{P}(\vec{d})}(S)\right) \\
&= m!\frac{(e^\lambda - 1)^{2n}}{\lambda^{2m}} \mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(S)\mathbf{P}_{\mathcal{D}}(\Sigma) \\
&\sim \frac{m!(e^\lambda - 1)^{2n}}{2\pi m(1+\lambda-c)\lambda^{2m}} \exp(-\lambda^2/2),
\end{aligned}$$

where we used Lemma 3.2. □

In addition, the computations in the proof of Theorem 1.2 give the following.

Corollary 3.3. *The elements in $\mathcal{G}_{1,1}(n,m)$ can be uniformly generated by restricting the probability space $\mathcal{P}_{1,1}(n,m)$ to simple pairings and considering the corresponding digraph.*

Proof. A dicore G in $\mathcal{G}_{1,1}(n,m)$ with degree sequence $\vec{d}$ comes from exactly $\prod_{i=1}^n d_i^+! d_i^-!$ different pairings. Each of these pairings must be simple and has probability

$$\frac{\lambda^{2m}/(e^\lambda - 1)^{2n}}{m!\prod_{i=1}^n d_i^+! d_i^-!} \big(\mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(S)\big)^{-1} \tag{18}$$

in the space $\mathcal{P}_{1,1}(n,m)$ conditional upon the event S of being simple. The product of (18) times $\prod_{i=1}^n d_i^+! d_i^-!$ does not depend on the particular $\vec{d}$, and therefore the distribution of G when generated from simple pairings is uniform. □

Finally, we can extend the concept of dicore defined in Section 2 as follows. Given $k = (k^+, k^-)$ where k^+ and k_- are positive integer constants, a *k-dicore* is an element of $\mathcal{G}(n,m)$ with a degree sequence satisfying $d_i^+ \geq k^+$ and $d_i^- \geq k^-$, for all $i \in \{1,\dots,n\}$. Let $\mathcal{G}_k(n,m)$ denote both the set of k-dicores and the corresponding uniform probability space.

In order to study the degree sequences of $\mathcal{G}_k(n,m)$, we need some definitions. Let λ^+ and η^+ (resp., λ^- and η^-) be obtained from (5) and (6) after replacing k, λ and η by k^+, λ^+ and η^+ (resp., by k^-, λ^- and η^-). Define the set of degree sequences $\mathcal{D}_k$ analogously to $\mathcal{D}$, with the extra condition that $d_i^+ \geq k^+$ and $d_i^- \geq k^-$, for all $i \in \{1,\dots,n\}$, and similarly let $\widehat{\mathcal{D}}_k$ be the subset of sequences in $\mathcal{D}_k$ satisfying (14). Moreover, we endow $\mathcal{D}_k$ with a probability distribution by selecting the d_i^+ and the d_i^- independently according to the $\mathrm{TPo}_{k^+}(\lambda^+)$ and the $\mathrm{TPo}_{k^-}(\lambda^-)$ distributions, respectively. The $\widehat{\mathcal{D}}_k$ space is simply $\mathcal{D}_k$ conditional upon (14). Furthermore, we define $\mathcal{P}_k(n,m)$ as we did for $\mathcal{P}_{1,1}(n,m)$ but randomising the degree sequence $\vec{d}$ according to the distribution of $\widehat{\mathcal{D}}_k$ defined above.

Now we are in good shape to extend the argument in the proof of Theorem 1.2 to general k-dicores.

Proof of Theorem 1.3. The proof is straightforward by going along the same steps as the proof of Theorem 1.2, but replacing $\mathcal{D}$, $\widehat{\mathcal{D}}$ and $\mathcal{P}_{1,1}(n,m)$ by $\mathcal{D}_k$, $\widehat{\mathcal{D}}_k$ and $\mathcal{P}_k(n,m)$, and considering the distributions $\mathrm{TPo}_{k^+}(\lambda^+)$ or $\mathrm{TPo}_{k^-}(\lambda^-)$ instead of $\mathrm{TPo}(\lambda)$ when appropriate. The key part is extending Lemma 3.2 to the new setting, which is also straightforward. The extended statement is as follows. Assume that $m - k^+n \to \infty$, $m - k^-n \to \infty$ and $m = O(n \log n)$. Then

(a) $\mathbf{P}_{\mathcal{D}_k}(\Sigma) \sim \dfrac{1}{2\pi nc\sqrt{(1+\eta^+-c)(1+\eta^--c)}}$.

Moreover, if S is the event that a random pairing in $\mathcal{P}(\vec{d})$ or $\mathcal{P}_k(n,m)$ is simple, then

(b) $\mathbf{P}_{\mathcal{P}_k(n,m)}(S) = \mathbf{E}_{\widehat{\mathcal{D}}_k}\big(\mathbf{P}_{\mathcal{P}(\vec{d})}(S)\big) \sim e^{-\lambda^+\lambda^-/2}$;

(c) for any r.v. X on $\widehat{\mathcal{D}}_k$ satisfying $|X| \leq x$ for some fixed constant $x \in \mathbb{R}$,

$$\mathbf{E}_{\widehat{\mathcal{D}}_k}\big(\mathbf{P}_{\mathcal{P}(\vec{d})}(S)\cdot X\big) = (1+o(1))\, e^{-\lambda^+\lambda^-/2}\, \mathbf{E}_{\widehat{\mathcal{D}}_k} X + O\left(e^{-\log^3 n}\right). \qquad \square$$

4 Moderately sparse case: c bounded

In this section we will prove Theorem 1.1 for the case that $c = m/n$ is bounded and also bounded away from 1.

A *sink-set* in a digraph G is a non-empty proper subset S of vertices such that the out-set of S is a subset of S. That is, no arc in G goes from S to $V(G) \setminus S$. A set of vertices is a *source-set* if its complement is a sink-set. A sink-set in a digraph with minimum outdegree at least 1 is *plain* if its vertices all have outdegree exactly 1, and is otherwise *complex*. Plain and complex source-sets are defined analogously by replacing outdegree by indegree. Observe that a digraph G is strongly connected iff it has no sink-set (and equivalently no source-set).

We use the term *s-set* to denote sets of vertices which are a sink-set or a source-set and, with a slight abuse of notation, also identify s-sets with their corresponding induced digraphs.

Note that a digraph G with m arcs is not strongly connected iff it has an s-set with at most $m/2$ arcs (if there is an s-set S with more than $m/2$ arcs, then consider the s-set $V(G) \setminus S$ instead). We first show that a.a.s. any complex s-set of $\mathcal{G}_{1,1}(n,m)$ must contain more than $m/2$ arcs. This will then allow us to restrict our attention to the existence of plain s-sets in order to determine the probability that $\mathcal{G}_{1,1}(n,m)$ is strongly connected with $o(1)$ error.

Proposition 4.1. *Suppose that $c = m/n$ is bounded and bounded away from 1. A digraph in $\mathcal{G}_{1,1}(n,m)$ a.a.s. has no complex s-set containing at most $m/2$ arcs.*

Proof. It is presumably possible to analyse $\mathcal{G}_{1,1}(n,m)$ or $\mathcal{P}_{1,1}(n,m)$ directly to achieve the desired result, by an expectation argument similar to that commonly used for connectivity of graphs. However, the expectation itself seems to be difficult to analyse. Instead we introduce another probability space, by partitioning according to the indegree sequence and to the multiset of outdegrees. More precisely, we will consider slices of $\mathcal{P}_{1,1}(n,m)$ with indegree sequence $\vec{d^-}$ and outdegree sequence being a permutation of $\vec{d^+}$, for each $\vec{d} \in \mathcal{D}$.

One could argue by partitioning according to the joint values of $\vec{d^-}$ and $\vec{d^+}$, but certain nasty combinations of in- and outdegrees, in which the vertices of outdegree 1 all have large indegree, are likely to cause trouble, and rather ad-hoc arguments may be required to bound the troublesome cases (see e.g. the approach in [3]). It is conceivable that allowing permutations of the outdegree sequence instead helps to explain a little more of the structure of the typical digraph in $\mathcal{G}(n,m)$.

To facilitate calculations of probabilities, for each $\vec{d} \in \widehat{\mathcal{D}}$ we introduce a probability space, $\mathcal{P}'(\vec{d})$, which is similar to common models (called pairing or configuration models) for random graphs or digraphs with given degree sequence. Consider two sets of points $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_m\}$, with A partitioned into nonempty sets (which we call *bins*) A_i, $i = 1, \ldots, n$ (corresponding to the vertices of the digraph) with $|A_i| = d_i^+$ for each i, and similarly B partitioned into nonempty sets (bins) B_i, $i = 1, \ldots, n$ with $|B_i| = d_i^-$ for each i. We write $\alpha(a_i) = j$ if $a_i \in A_j$, and $\beta(b_i) = j$ if $b_i \in B_j$. A random element of $\mathcal{P}'(\vec{d})$ is a random bijection $\phi : A \to B$ together with a random permutation σ of $[n]$, such that the pair (ϕ, σ) is chosen u.a.r. Each element in $\mathcal{P}'(\vec{d})$ can be mapped in a natural way to a pairing in $\mathcal{P}_{1,1}(n,m)$, obtained by identifying points in $A_{\sigma(j)}$ and points in B_j with out-points and in-points of bin j. This corresponds in turn to a multidigraph M which has an arc (u,v) for each point $a_i \in A_{\sigma(u)}$ such that $\phi(a_i) \in B_v$, or equivalently the arc (multi)set is $\{\sigma^{-1}(\alpha(a_i))\beta(\phi(a_i)) : 1 \le i \le m\}$. Observe that M has indegree sequence $\vec{d^-}$ and an outdegree sequence which is a random permutation of $\vec{d^+}$. As usual, all graph theory statements referred to an element in $\mathcal{P}'(\vec{d})$ should be understood in terms of the corresponding multidigraph.

Define U to be the event, defined on any relevant probability spaces, that there is a complex proper sink-set containing at most $m/2$ arcs. Ultimately, we will do the calculations in the space $\mathcal{P}'(\vec{d})$ with $\vec{d}$ randomised according to its distribution in the space $\widehat{\mathcal{D}}$. Call this space $\mathcal{P}'_{1,1}(n,m)$. Averaging over $\vec{d}$ makes computations a little easier than arguing about its typical values. In fact, observe that the distribution of a random degree sequence $\vec{d} \in \widehat{\mathcal{D}}$ stays invariant if we randomly permute the entries of the outdegree sequence $\vec{d^+}$. Hence, we deduce

that

$$\mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(U) = \mathbf{E}_{\widehat{\mathcal{D}}}\left(\mathbf{P}_{\mathcal{P}(\vec{d})}(U)\right) = \mathbf{E}_{\widehat{\mathcal{D}}}\left(\mathbf{P}_{\mathcal{P}'(\vec{d})}(U)\right) = \mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U). \tag{19}$$

Thus, in view of Corollary 3.3 and Lemma 3.2, we have

$$\mathbf{P}_{\mathcal{G}_{1,1}(n,m)}(U) = \mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(U \mid S) \leq (1+o(1))e^{\lambda^2/2}\,\mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(U) = O(\mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U)). \tag{20}$$

Therefore, we only need to show that $\mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U) = o(1)$ in order to prove the theorem statement for complex sink-sets. The result extends immediately to complex source-sets by considering the converse digraph.

The remainder of the proof consists of bounding the probability that an element of $\mathcal{P}'_{1,1}(n,m)$ has a complex sink-set with at most $m/2$ arcs. Observe that, if S is a complex sink-set and v_0 is a vertex in S with outdegree strictly greater than 1 (there must exist at least one of these because S is complex), then the set $S' \subseteq S$ of vertices reachable by directed paths from v_0, including v_0 of course, is also a complex sink-set. This follows easily from the facts that every arc out from a vertex of S' must join to a vertex in S', all vertices in S' have outdegree at least 1, and S' has a vertex v_0 with outdegree at least 2. Therefore, we only need to consider complex sink-sets which are precisely the set of vertices reachable from some vertex v_0.

Given a vertex v_0, the following algorithm will terminate with S being the set of vertices reachable from v_0. The algorithm works by maintaining a set S of bins A_i corresponding to vertices reachable from v_0, and investigating the vertices reachable from S. It does this by looking at the points in bins in S. The set T contains precisely such points which have not yet been investigated.

Algorithm

Let v_0 be the initially chosen vertex. Start with $S = \{v_0\}$, $T = A_{\sigma(v_0)}$, and repeat the following until T is empty. Pick $a_i \in T$, add to S the vertex $v = \beta(\phi(a_i))$ (if it is not already there), delete a_i from T and, if v was not already in S, add all elements in $A_{\sigma(v)}$ to T.

If the algorithm terminates with S being a complex sink-set containing at most half of the arcs of M, we say that it terminates *properly*, and otherwise improperly. We complete the proof of the theorem by showing that the probability that there exists a vertex v_0 such that the algorithm terminates properly, when begun from v_0, is $o(1)$.

As is common in analysing algorithms like this, we will make use of the fact that, conditioning on any set of values of a uniformly random permutation, the remaining values are still distributed uniformly at random. Thus, the algorithm can be performed simultaneously with the generation of the random bijection ϕ and permutation σ. At the start, ϕ and σ entirely undetermined and we can choose $\phi(a_i)$ at random from the unused points of B at each step of the algorithm. Similarly, we may choose $\sigma(v_0)$ initially, and then $\sigma(v)$ at each step where the vertex v was not already in S, randomly from the indices i of the unused bins A_i. Thus, we may initially choose u.a.r. a permutation $\phi_1, \ldots, \phi_m$ of B, and independently a permutation $\sigma_1, \ldots, \sigma_n$ of $[n]$ u.a.r., and use ϕ_1 for the first value of ϕ called for in the algorithm, ϕ_2 for the second, and so on, and similarly for σ. Set $K_k = \{\phi_1, \ldots, \phi_k\}$ and $J_s = \{\sigma_1, \ldots, \sigma_s\}$. Since

the ϕ_i and σ_i are pre-chosen randomly, it follows that, for given k and s,

$$J_s \subseteq [n] \text{ and } K_k \subseteq B \text{ are independent and u.a.r.} \tag{21}$$

In particular, the joint distribution of J_s and K_k does not not depend on the algorithm, which is the important feature that simplifies analysis.

Now define

$$\hat{k} = \sum_{j \in J_s} |A_j|, \tag{22}$$

and let U_{v_0} denote the event that the algorithm terminates properly, and let k be the (random) number of steps taken by the algorithm at termination (i.e. arcs between vertices in S) and s the cardinality of S. In particular, S is a complex sink-set with at most $m/2$ arcs. In the event U_{v_0}, since the termination condition implies that T is empty, it follows that

$$\hat{k} = k. \tag{23}$$

Also define

$$\hat{s} = |\{u \,:\, u = v_0 \text{ or } K_k \cap B_u \neq \emptyset\}|. \tag{24}$$

Note that at each step, since $\beta(\phi(a_i))$ is added to S, we have $S = \{v_0\} \cup \{u \,:\, K_k \cap B_u \neq \emptyset\}$ and hence

$$\hat{s} = s. \tag{25}$$

Moreover, the fact that S is complex is equivalent to the condition that there are more arcs chosen than vertices in S, and so $k > s$. Hence, an upper bound on $\mathbf{P}(U_{v_0})$ is the probability that (23) and (25) hold for some k and s with $k \leq m/2$ and $s < k$.

Denote the event that (25) holds, given k and s, with $\hat{s}$ generated according to (24) given (21), by $H^-_{k,s}$, and similarly the event that (23) holds, given k and s, with $\hat{k}$ generated according to (22), by $H^+_{k,s}$. Also put $H_{k,s} = H^+_{k,s} \wedge H^-_{k,s}$. We will prove that

$$\mathbf{P}\Bigg(\bigcup_{k \leq m/2} \bigcup_{s<k} H_{k,s} \Bigg) = o(n^{-1}). \tag{26}$$

Then $\mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U_{v_0}) = o(n^{-1})$, and the result follows by taking the union bound $\mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U) \leq \sum_{v_0} \mathbf{P}_{\mathcal{P}'_{1,1}(n,m)}(U_{v_0})$ and from (20).

It only remains to show (26), for which we split k into two intervals.

Case 1. $\log^4 n < k \leq m/2$.

We first bound probabilities in the distribution of $\hat{s}$ as determined by (24). Recall the truncated Poisson distribution as defined in Section 2. Let $\Omega = \Omega(n, c, q)$ denote the probability space in which there are n bins B_i with $|B_i| = d^-_i$, where $d^-_1, \ldots, d^-_n$ are independent random variables each with the distribution of $\mathrm{TPo}(\lambda)$, and such that a random subset T of the points in the bins is chosen by including each point independently with probability

$$q = k/m.$$

Let $\hat{s}$ be the number of the bins that are either occupied by at least one point of T or happen to be the bin v_0. It follows from Lemma 2.5 that

$$\mathbf{P}_\Omega\big(|\hat{s} - nq'| > \sqrt{nq'}\log n\big) = o(n^{-3}), \tag{27}$$

where q' is the probability that a bin contains some point of T. Note that $q' > q(1+\epsilon)$ for some positive constant ϵ that can be determined from (12). Now define E to be the event in Ω that the total content of bins is $\sum_{i=1}^n d_i^- = m$ and that exactly k points are chosen in T. Observe that, in the probability space Ω conditional upon E, the number $\hat{s}$ of bins containing at least one element of T is distributed as in the definition of $H_{k,s}^-$. From Lemma 2.2, $\sum_{i=1}^n d_i^- = m$ holds with probability $\Theta(n^{-1/2})$, and—conditional on that—the event $|T| = k$ has probability $\Theta(k^{-1/2})$, since $|T| \overset{\mathrm{d}}{\sim} \mathrm{Bin}(m, k/m)$. Hence, $\mathbf{P}_\Omega(E) = \Omega(n^{-1})$ and by (27)

$$\mathbf{P}\left(\bigcup_{|s-nq'|>\sqrt{nq'}\log n} H_{k,s}^-\right) = \mathbf{P}_\Omega(|\hat{s} - nq'| > \sqrt{nq'}\log n \mid E) = o(n^{-2}). \tag{28}$$

The next (and simpler) step is to define $\Omega' = \Omega'(n, c, s)$ to be the probability space in which there are n bins A_i with $|A_i| = d_i^+$ all independent random variables each with the distribution of $\mathrm{TPo}(\lambda)$, and such that a uniformly random set of s of the bins is chosen. Assume that s lies in the range $|s - nq'| \le \sqrt{nq'}\log n$, and in particular $s = \Omega(\log^4 n)$. Let $\hat{k}$ be the total number of points in the selected bins. From Lemma 2.3 we obtain the tail bound

$$\mathbf{P}_{\Omega'}\left(\hat{k} \le \frac{cs}{1+\epsilon/2}\right) \le e^{-\Theta(s^{1/3})} = o(n^{-4}). \tag{29}$$

Now let E' be the event in Ω' that $\sum_{i=1}^n d_i^+ = m$, and recall from Lemma 2.2 that $\mathbf{P}_{\Omega'}(E') = \Theta(n^{-1/2})$. Observe that in Ω' conditional upon E' the distribution of $\hat{k}$ is the same as the one in the definition of $H_{k,s}^+$. Moreover, the fact that $|s - nq'| \le \sqrt{nq'}\log n$ implies that $k \le cs/(1+\epsilon+o(1))$. In view of all that and from (29), we obtain that for $|s-nq'| \le \sqrt{nq'}\log n$

$$\mathbf{P}(H_{k,s}^+) \le \mathbf{P}_{\Omega'}\left(\hat{k} \le \frac{cs}{1+\epsilon+o(1)} \mid E'\right) = o(n^{-3}). \tag{30}$$

Taking the union bound over all s such that $|s-nq'| \le \sqrt{nq'}\log n$, combining it with (28) and summing over all k between $\log^4 n$ and $m/2$ completes the proof of (26) for this range of k.

Case 2. $k \le \log^4 n$.

For $\vec{d} \in \widehat{\mathcal{D}}$, the event $d_{\max}^- < \log^2 n$, or equivalently $|B_i| < \log^2 n$ for each i, holds with probability $1 - o(n^{-1})$. This follows readily from bounding the probability of the complement in $\mathcal{D}$ and then conditioning upon $\sum_i d_i^- = m$ (see (4) and Lemma 2.2). Since $\mathcal{P}'_{1,1}(n, m)$ is just $\mathcal{P}'(\vec{d})$ with $\vec{d}$ distributed as in $\widehat{\mathcal{D}}$, we may focus on $\mathcal{P}'(\vec{d})$ for a particular $\vec{d}$ satisfying the above property. Note that $\hat{s} = s < k$ according to (24) if the random choice $\{\phi_1, \dots, \phi_k\}$ of elements of B determines at most $k-2$ bins other than v_0. This has probability $O(k^4(\log^2 n/m)^2)$. Hence, in $\mathcal{P}'(\vec{d})$

$$\sum_{k\le\log^4 n} \mathbf{P}\left(\bigcup_{s<k} H_{k,s}^-\right) = \sum_{k\le\log^4 n} \mathbf{P}(\hat{s} \le k-1) = O(\log^{24} n/n^2) = o(1/n). \qquad \square$$

Next consider plain s-sets of $\mathcal{G}_{1,1}(n,m)$.

Proposition 4.2. *Suppose that $c = m/n$ is bounded and bounded away from 1. The probability that a digraph in $\mathcal{G}_{1,1}(n,m)$ has no plain s-set is asymptotic to*

$$\frac{e^\lambda(e^\lambda - 1 - \lambda)^2}{(e^{2\lambda} - e^\lambda - \lambda)(e^\lambda - 1)}, \tag{31}$$

with λ determined by the equation $c = \lambda e^\lambda/(e^\lambda - 1)$.

Proof. The simplest sink-sets or source-sets are those whose vertices induce a directed cycle. Call them *sink-cycles* or *source-cycles* accordingly. An *s-cycle* is just a set of vertices which is either a sink-cycle or a source-cycle. Observe that each plain s-set must contain at least one s-cycle, so we can restrict our attention to s-cycles.

For any natural $k \geq 1$, let C_k be the (random) number of s-cycles of order at most k. Let D be the number of double arcs. Define

$$\mu_k = \sum_{j=1}^{k} \frac{2(c/e^\lambda)^j - (c/e^{2\lambda})^j}{j}.$$

Easy computations show that $2(c/e^\lambda)^j > (c/e^{2\lambda})^j$, so that there are no cancellations in any term of the definition of μ_k. We first claim that, for constant k, $\mathbf{E}_{\mathcal{P}_{1,1}(n,m)}C_k \sim \mu_k$, $\mathbf{E}_{\mathcal{P}_{1,1}(n,m)}D \sim \lambda^2/2$, and moreover C_k and D are asymptotically jointly independent Poisson. Elementary calculations show that

$$\mu = \lim_{k\to\infty} \mu_k = \log\left(\frac{(e^{2\lambda} - e^\lambda - \lambda)(e^\lambda - 1)}{e^\lambda(e^\lambda - 1 - \lambda)^2}\right).$$

On the other hand, we claim that the probability there is an s-cycle of order greater than k can be bounded by some function f_k such that $\lim_{k\to\infty} f_k = 0$. In view of all this, setting S to be the event that $\mathcal{P}_{1,1}(n,m)$ has no multiple arcs and V to be the event in $\mathcal{G}_{1,1}(n,m)$ or $\mathcal{P}_{1,1}(n,m)$ that there are no s-cycles, we get

$$\mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(V \cap S) \sim e^{-\mu-\lambda^2/2}.$$

Then the proof of the result follows immediately from Lemma 3.2(b) and the fact that $\mathbf{P}_{\mathcal{G}_{1,1}(n,m)}(V) = \mathbf{P}_{\mathcal{P}_{1,1}(n,m)}(V \mid S)$.

Now we proceed to verify the claims we made about C_k, D and the expected number of "long" s-cycles. To make the computations easier, we generate the elements of $\mathcal{P}_{1,1}(n,m)$ using a slight variation of the $\mathcal{P}'_{1,1}(n,m)$ model in which the out-points $a_1,\ldots,a_m$ (resp. in-points $b_1,\ldots,b_m$) are assigned independently and u.a.r. to the out-bins $A_1,\ldots,A_n$ (resp. in-bins $B_1,\ldots,B_n$) conditional upon each bin receiving at least one point (note that the degree sequence thus obtained is distributed as in $\widehat{\mathcal{D}}$). In addition to that, a random bijection ϕ of the out- and in-points, and a random permutation of the labels of the out-bins are chosen as before independently and u.a.r. (alternatively we may consider σ to be a random bijection of the out- and in-bins).

First we wish to compute the joint factorial moments of C_k and D, for constant k. We shall index all possible s-cycles of length at most k by their *position* (i.e. the vertices they use in cyclic order). More precisely, a position c of length ℓ is defined to be a tuple $i_1, \ldots, i_\ell$ of distinct elements in $\{1, \ldots, n\}$ given in cyclic order. We say that a random element of $\mathcal{P}'_{1,1}(n, m)$ has an s-cycle at position c, if it has an s-cycle on vertices $v_1, \ldots, v_\ell$ where each vertex v_j corresponds to the bins $A_{\sigma(i_j)}$ and B_{i_j}. Note that the length of an s-cycle occurring at position c is also the length of c.

Let r be a constant natural number, and fix a tuple of distinct positions $c_1, \ldots, c_r$ of lengths $\ell_1 \leq k, \ldots, \ell_r \leq k$ such that the sets of indices defining any two different positions c_{j_1} and c_{j_2} are pairwise disjoint. Let $X_{c_1,\ldots,c_r}$ be the indicator function for the event that there is an s-cycle at each position c_i. We compute the asymptotic probability that this event holds for fixed $\ell_1, \ldots, \ell_r$. The probability that the out-bins are assigned to the corresponding in-bins is

$$1/[n]_{\ell_1+\cdots+\ell_r} \sim 1/n^{\ell_1+\cdots+\ell_r}. \tag{32}$$

Condition on this, and note that the degrees of the bins and the matching of the points occur independently from that. Now we claim that the probability that the right s-cycles occur at $c_1, \ldots, c_r$ is asymptotic to

$$\prod_i (2a^{\ell_i}/n^{\ell_i} - a^{2\ell_i}/m^{\ell_i}), \tag{33}$$

where $a = \lambda/(e^\lambda - 1) = c/e^\lambda$. Observe that the events of having a sink-cycle or having a source-cycle at c_i are not disjoint, so the probability of the union is the sum of probabilities minus the probability of having both a sink- and a source-cycle at c_i. Thus, in order to estimate (33), we can specify for each c_i one of the three former events (sink-cycle, source-cycle or both). More precisely, given $r_1+r_2+r_3 = r$ and relabelling $c_1, \ldots, c_r$ to $c_1, \ldots, c_{r_1}, c'_1, \ldots, c'_{r_2}, c''_1, \ldots, c''_{r_3}$ and $\ell_1, \ldots, \ell_r$ to $\ell_1, \ldots, \ell_{r_1}, \ell'_1, \ldots, \ell'_{r_2}, \ell''_1, \ldots, \ell''_{r_3}$, we shall compute w.l.o.g. the following probability: having a sink-cycle at $c_1, \ldots, c_{r_1}$ (and possibly a source-cycle too); having a source-cycle at $c'_1, \ldots, c'_{r_2}$ (and possibly a sink-cycle too); or having both a sink- and a source-cycle at $c''_1, \ldots, c''_{r_3}$. We shall see that this probability is asymptotic to

$$\frac{a^{\ell_1+\cdots+\ell_{r_1}+\ell'_1+\cdots+\ell'_{r_2}+2(\ell''_1+\cdots+\ell''_{r_3})}}{n^{\ell_1+\cdots+\ell_{r_1}+\ell'_1+\cdots+\ell'_{r_2}} m^{\ell''_1+\cdots+\ell''_{r_3}}}, \tag{34}$$

and this leads to (33) by easy inclusion-exclusion.

To arrive at (34), we require that $\ell_1 + \cdots + \ell_{r_1} + \ell''_1 + \cdots + \ell''_{r_3}$ specific out-bins determined by $c_1 + \cdots + c_{r_1} + c''_1 + \cdots + c''_{r_3}$ contain exactly one point each. By symmetry, the probability of this is $(\mathbf{E}[N]_{\ell_1+\cdots+\ell_{r_1}+\ell''_1+\cdots+\ell''_{r_3}})/[n]_{\ell_1+\cdots+\ell_{r_1}+\ell''_1+\cdots+\ell''_{r_3}}$, where N is the number of out-bins with only one point. N is concentrated around an by [2, Lemma 1] since the distribution of balls in bins is truncated multinomial. Hence $\mathbf{E}[N]_{\ell_1+\cdots+\ell_{r_1}+\ell''_1+\cdots+\ell''_{r_3}} \sim (an)^{\ell_1+\cdots+\ell_{r_1}+\ell''_1+\cdots+\ell''_{r_3}}$ and the probability is $(1+o(1))a^{\ell_1+\cdots+\ell_{r_1}+\ell''_1+\cdots+\ell''_{r_3}}$. Analogously, we need that some specific $\ell'_2 + \cdots + \ell'_{r_2} + \ell''_1 + \cdots + \ell''_{r_3}$ in-bins contain only one point, which is independent from the previous and has probability $(1+o(1))a^{\ell'_2+\cdots+\ell'_{r_2}+\ell''_1+\cdots+\ell''_{r_3}}$. Conditional upon all this, we need that for each $c_1, \ldots, c_{r_1}$, the only point in each out-bin is matched to some point in the corresponding in-bin; for each $c'_1, \ldots, c'_{r_2}$, the only point in each in-bin is matched to some point in the corresponding out-bin; and for each $c''_1, \ldots, c''_{r_3}$, the only point in each out-bin is matched to the only point in the corresponding in-bin. Observe that the number of

points in those out-bins that have not been exposed remains independent truncated Poisson conditional to fixed sum $m-\ell_1-\cdots-\ell_{r_1}-\ell''_1-\cdots-\ell''_{r_3}$. An analogous thing happens for in-bins that were not exposed and the sum of their degrees is $m-\ell'_1-\cdots-\ell'_{r_2}-\ell''_1-\cdots-\ell''_{r_3}$. The probability of matching the in- and out-points appropriately for s-cycles at $c''_1,\ldots,c''_{r_3}$ is $1/[m]_{\ell''_1+\cdots+\ell''_{r_3}} \sim 1/m^{\ell''_1+\cdots+\ell''_{r_3}}$. We condition on that and on the event that no out-point corresponding to $c_1,\ldots,c_{r_1}$ is matched to any in-point corresponding to $c'_1,\ldots,c'_{r_2}$ (which happens with probability $1+o(1)$). This makes the construction of the remaining sink-cycles independent from that of source-cycles. For the sink-cycles, we have to match the only point in each out-bin with some point in the corresponding in-bin. By symmetry, the first matching has probability $1/(n-\ell'_1-\cdots-\ell'_{r_2}-\ell''_1-\cdots-\ell''_{r_3}) \sim 1/n$. Conditional to some matchings being exposed, the probability that the next out-point is matched to a point in an in-bin which contains one matched point already is $O(1/n)$ since there is negative correlation between the events that two out-points are matched to in-points in the same bin (condition on any given degree). Thus that out-point is matched to some point in an unexposed in-bin with probability $1+o(1)$ and conditional to that, again by symmetry, chooses the right in-bin with probability $(1+o(1))/n$. This gives a probability $(1+o(1))/n^{\ell_1+\cdots+\ell_{r_1}}$ for having the matchings required for the sink-cycles. Analogously, the source-cycles give a $(1+o(1))/n^{\ell'_1+\cdots+\ell'_{r_2}}$ factor, and this establishes the estimate (34).

So far we were dealing with fixed cycle tuples $c_1,\ldots,c_r$. Recall that C_k is the random number of s-cycles occurring, and k is fixed. To compute the r-th factorial moment it suffices to multiply (32) and (33) by the number of ways of choosing r different $c_1,\ldots,c_r$, which is

$$\sum_{\ell_1,\ldots,\ell_r\in\{1,\ldots,k\}} \frac{([n]_{\ell_1+\cdots+\ell_r})^2}{\ell_1\cdots\ell_r}.$$

Hence,

$$\mathbf{E}[C_k]_r \sim \left(\sum_{\ell=1}^{k} \frac{2a^{\ell_i}-(a^2/c)^{\ell_i}}{\ell_i}\right)^r = {\mu_k}^r.$$

Let D be the number of double arcs occurring. Recall that the out-points are placed in the out-bins (and the in-points in the in-bins) uniformly at random and independently conditional upon getting at least one point in each bin. We index double arcs according to their position, where each position j is a set of two different out-points in the same out-bin along with a set of two different in-points in the same in-bin. Let Z be the number of positions for double arcs. We have the trivial bound $Z\le m^4$. Combining together Lemmas 2.1, 2.4 and 2.2, we have that $|Z-(\lambda cn/2)^2|<n^{1.6}$ with probability $1-O(n^{1/2}e^{-\log^3 n})$. Hence, $\mathbf{E}[Z]_s\sim(\lambda cn/2)^{2s}$. We say that there is a double arc at j if the out-points are matched to the in-points in any of the two possible ways. Fix s different positions $j_1,\ldots,j_s$. The probability of having double arcs at $j_1,\ldots,j_s$ is is $2^s/[m]_{2s}$. Therefore,

$$\mathbf{E}[D]_s=\mathbf{E}[Z]_m 2^s/[m]_{2s}\sim(\lambda^2/2)^s.$$

In order to compute joint moments of C_k and D, we condition to s-cycles happening at some fixed positions $c_1,\ldots,c_r$ and we specify the type of each cycle (sink-, source- or both) in the same fashion we used in the previous computations (sink does not exclude source and vice-versa). To make computations easier we also condition on the particular in-points and

out-points matched to create the s-cycles. Conditional upon all this, we compute $\mathbf{E}[D]_s$. The same computations we did before are still valid if applied to the in- and out-points that were not used in the construction of the s-cycles, and this yields the the same asymptotic value $(\lambda^2/2)^s$. So

$$\mathbf{E}[C_k]_r[D]_s \sim \left(\sum_{\ell=1}^{k} \frac{2a^{\ell_i} - (a^2/c)^{\ell_i}}{\ell_i}\right)^r (\lambda^2/2)^s$$

The claim that the distributions are asymptotically independent Poisson now follows by the standard method of moments.

It only remains to bound the probability of existence of s-cycles of length greater than k by some function f_k such that $\lim_{k\to\infty} f_k = 0$. It is enough to deal with sink-cycles, since the result for source-cycles follows by considering the converse digraph. Take a length $\ell > k$. We now condition on the number N as defined above. We can choose $([n]_\ell)^2/\ell$ different positions for such a cycle. For each of these, the probability that the bins are matched the right way is $1/[n]_\ell$ (regardless of N). The probability that each of the ℓ out-bins contains exactly one point is at most $(N/n)^\ell$. The probability that each of the ℓ out-points is matched to a point in the corresponding in-bin is at most $1/[n]_\ell$ (since conditional upon i matched pairs, the probability of the next matched pair is $1/(n-i)$ times the probability of hitting a point in an in-bin not previously hit). This is again regardless of N.

Putting this together, this expectation is at most

$$\sum_{\ell>k} (N/n)^\ell/\ell.$$

this tends to 0 for large k provided $N/n < (a+1)/2$. The probability that N is larger than this is $o(1)$ by the concentration mentioned above. □

From Theorem 1.2, Proposition 4.1 and Proposition 4.2 we immediately obtain Theorem 1.1 for the case c is bounded and bounded away from 1.

5 Very sparse case: $c \to 1$

Here we prove Theorem 1.1 for the case $c \to 1$. Thus $r = m - n = o(n)$, and we assume $r = m - n \to \infty$. We define the directed graph analogue of the kernel of a graph as follows. A cycle component of a digraph is a connected component which is simply a directed cycle. A digraph with each in-and outdegree at least 1 and with no cycle components is called a *preheart*. The *heart* of a preheart G is the multidigraph $H(G)$ obtained from G by repeatedly choosing a vertex v of in- and outdegrees both 1, deleting v and its two incident arcs uv and vw, and inserting the arc uw. The condition that G contains no isolated cycle ensures that the heart is always a multidigraph. The vertices of $H(G)$ are just the vertices of G of total degree at least 3.

Note that a digraph is strongly connected iff it is an isolated cycle or a preheart with strongly connected heart. Thus we may use ideas similar to those in Section 4 to study the heart, as a key step to enumerate strongly connected digraphs. Connectivity properties of the heart can be easier to prove than for just the $(1,1)$-dicore. In the dicore, some complex s-sets

can involve many vertices of in- and outdegree 1 and just a few other vertices. We will focus on the heart and also use randomisation of the in- and outdegree sequences, as in the $\mathcal{P}(\vec{d})$ and $\mathcal{P}'(\vec{d})$ models in Section 4.

Consider any given degree sequence $\vec{d} \in \widehat{\mathcal{D}}$, and let $T = T(\vec{d}) = \{i : d_i^+ + d_i^- \geq 3\}$. We put $n' = |T|$ and $m' = \sum_{i \in T} d_i^+ = \sum_{i \in T} d_i^-$, and note that $m - n = m' - n'$. For simplicity of presentation, renumber the vertices if necessary so that $T = [n']$.

Let $\mathcal{H}(\vec{d})$ be the probability space of *heart configurations* generated as follows. For each $i \in T$ consider a bin containing labelled points of two types, namely d_i^+ *out-points* and d_i^- *in-points*, and then choose a random matching of the in-points with the out-points (there are m' of each kind). Note that each heart configuration in $\mathcal{H}(\vec{d})$ corresponds to a multidigraph on vertex set T obtained in a natural way by identifying bins with vertices and adding an arc (u, v) for each out-point in u matched to an in-point in v.

Moreover, given a heart configuration H, we construct a *preheart configuration* Q by taking an assignment of $[n] \setminus T$ to the arcs of H (i.e. the pairs of matched up points), such that the numbers assigned to each arc are given a linear ordering. Denote this assignment, including the linear orderings, by f. Let $\mathcal{Q}(\vec{d})$ be the probability space of random preheart configurations created by taking $H \in \mathcal{H}(\vec{d})$ and choosing f u.a.r. Note that each $Q \in \mathcal{Q}(\vec{d})$ corresponds to a multidigraph with n vertices, m arcs and degree sequence $\vec{d}$. Henceforth, any graph terminology referring to a heart or preheart configuration should be interpreted in terms of the corresponding multidigraph.

Lemma 5.1. *The digraphs generated from the restriction of $\mathcal{Q}(\vec{d})$ to simple preheart configurations (i.e. with no multiple arcs) are uniformly distributed.*

Proof. Each simple digraph comes from $\prod_{i=1}^n d_i^+! d_i^-!$ different preheart configurations. □

As will become apparent later in the argument, it turns out that the degree sequence distribution induced by the uniform probability space of all prehearts on n vertices and m arcs is close in some sense to $\widehat{\mathcal{D}}$. This motivates considering the probability spaces $\mathcal{H}(n, m)$ and $\mathcal{Q}(n, m)$, defined by choosing a random element from $\mathcal{H}(\vec{d})$ and $\mathcal{Q}(\vec{d})$ respectively, where the degree sequence $\vec{d}$ is also random and distributed as in $\widehat{\mathcal{D}}$.

Given a degree sequence $\vec{d} \in \widehat{\mathcal{D}}$, we distinguish four kinds of vertices depending on whether their in- and outdegree are equal to 1, or larger. For $i, j \in \{1, 2\}$, let $N_{i,j}$ be the set of vertices with indegree of type i and outdegree of type j (type 1 means 1 and type 2 means greater than 1). Let $\boldsymbol{a} = (a_{1,1}, a_{1,2}, a_{2,1}, a_{2,2})$, where $a_{i,j} = |N_{i,j}|$. This $\boldsymbol{a}$ is of course a function of $\vec{d}$. Observe that any $\boldsymbol{a}$ which is *feasible* (i.e. occurs in $\widehat{\mathcal{D}}$ with nonzero probability) satisfies $a_{1,1} + a_{1,2} + a_{2,1} + a_{2,2} = n$, $1 \leq a_{1,2} + a_{2,2} \leq r$ and $1 \leq a_{2,1} + a_{2,2} \leq r$. Conversely, it is easy to check that, for sufficiently large n, any nonnegative tuple $\boldsymbol{a}$ satisfying the above conditions is feasible. Note also that $n' = a_{1,2} + a_{2,1} + a_{2,2}$ and $m' = r + a_{1,2} + a_{2,1} + a_{2,2}$.

We will want to condition on "typical" values of $\boldsymbol{a}$. Denote by Γ the event that

$$|a_{1,2} - r| \leq \sqrt{r} \log r, \quad |a_{2,1} - r| \leq \sqrt{r} \log r \quad \text{and} \quad a_{2,2} \leq \max\{2r^2/n, \sqrt{r}\}.$$

Note in particular that Γ implies

$$n' \sim 2r \qquad m' \sim 3n'/2 \sim 3r, \qquad a_{1,2} \sim a_{2,1} \sim r, \qquad a_{2,2} = o(r). \tag{35}$$

We next show something somewhat stronger than $\mathbf{P}_{\widehat{\mathcal{D}}}(\Gamma) = 1 - o(1)$.

Lemma 5.2.

$$\mathbf{E}_{\widehat{\mathcal{D}}}(m'(1 - 1_\Gamma)) = o(1).$$

Proof. First we observe that m' is deterministically at most $3r$ in $\widehat{\mathcal{D}}$. This upper bound is immediate from the fact that the underlying undirected graph of the heart has n' vertices, $m' = n' + r$ edges and average degree $2m'/n' \geq 3$. Hence, by Lemma 3.2(a), it suffices to bound the probability that Γ fails by $o(1/r^2)$ in $\mathcal{D}$. Here, $a_{1,2}$, $a_{2,1}$ and $a_{2,2}$ are binomially distributed with expectations r, r and r^2/n respectively. (Note that $r \to \infty$, but r^2/n need not be large.) Hence, standard bounds (if r grows very slowly, (9) does not suffice, but in any case we can simply consider ratios of consecutive binomial probabilities) shows that the conditions on $a_{1,2}$ and $a_{2,1}$ in the definition of Γ hold with probability $1 - o(1/r^2)$. A similar argument ensures that $a_{2,2}$ has the required concentration with probability $1 - o(1/r^2)$, but the analysis is split into two cases. If $r \leq n^{3/5}$, then $r^2/n \leq r^{1/3}$ and we easily bound the probability that $a_{2,2} > \sqrt{r}$, for instance by comparing with a binomial with mean $r^{1/3}$. On the other hand, if $r > n^{3/5}$, we bound the probability that $a_{2,2} > 2r^2/n$ using (9). □

This result allows us to condition on feasible $\boldsymbol{a}$ satisfying Γ. In fact, for any given feasible tuple $\boldsymbol{a}$, we denote by $\mathcal{H}(\boldsymbol{a})$ and $\mathcal{Q}(\boldsymbol{a})$ respectively the probability spaces $\mathcal{H}(n, m)$ and $\mathcal{Q}(n, m)$ conditional on having that particular $\boldsymbol{a}$.

Lemma 5.3. *Let $\boldsymbol{a}$ be any feasible tuple satisfying Γ. Then a random heart configuration in $\mathcal{H}(\boldsymbol{a})$ a.a.s. has no complex s-set of at most $m'/2$ arcs.*

Proof. The argument shares many features with the proof of Proposition 4.1, in particular using auxiliary randomisation to simplify computations. For each $\vec{d} \in \widehat{\mathcal{D}}$ recall the definition of T, the relevant set of vertices for the heart configuration, and the quantities m' and n' above, and for consistency let N' denote $T = [n']$. Consider, analogous to the definition of $\mathcal{P}'(\vec{d})$, two sets of points $A = \{a_1, \ldots, a_{m'}\}$ and $B = \{b_1, \ldots, b_{m'}\}$, partitioned respectively into bins $A_1, \ldots, A_{n'}$ and bins $B_1, \ldots, B_{n'}$, with $|A_i| = d_i^+$ and $|B_i| = d_i^-$ for each $i \in N'$. We write $\alpha(a_i) = j$ if $a_i \in A_j$, and $\beta(b_i) = j$ if $b_i \in B_j$. Define the probability space $\mathcal{H}'(\vec{d})$ to be a random bijection $\phi : A \to B$ chosen u.a.r. together with two random permutations σ and τ of N', chosen independently of ϕ and of each other and u.a.r. subject to the conditions that $d^+_{\sigma(i)} = 1$ whenever $d_i^+ = 1$, and $d^-_{\tau(i)} = 1$ whenever $d_i^- = 1$. We need an appropriate randomisation of the degrees. Thus, consider the probability space $\mathcal{H}'(\boldsymbol{a})$, whose elements are selected at random from $\mathcal{H}'(\vec{d})$ with $\vec{d}$ a random member of $\widehat{\mathcal{D}}$ but conditional on the particular value of the vector $\boldsymbol{a} = \boldsymbol{a}(\vec{d})$.

Observe that each element H' in $\mathcal{H}'(\boldsymbol{a})$ corresponds in a natural way to an element H in $\mathcal{H}(\boldsymbol{a})$, obtained by identifying the points in $A_{\sigma(j)}$ and those in $B_{\tau(j)}$ with the out-points and in-points, respectively, of bin (vertex) j (in the same way that elements in $\mathcal{H}'(\vec{d})$ can be mapped to elements in $\mathcal{H}(\vec{d})$). Moreover, the H obtained this way has the same distribution as in $\mathcal{H}(\boldsymbol{a})$, since the distribution of the degree sequence and thus $\boldsymbol{a}$ stay invariant after permuting the indices of the vertices in N' by σ and τ (so it does not matter if we condition to a particular $\boldsymbol{a}$ before or after applying σ and τ). Hence, setting U to be the event in $\mathcal{H}(\boldsymbol{a})$

or $\mathcal{H}'(\boldsymbol{a})$ that there is a complex sink-set containing at most $m'/2$ arcs, we have

$$\mathbf{P}_{\mathcal{H}(\boldsymbol{a})}(U) = \mathbf{P}_{\mathcal{H}'(\boldsymbol{a})}(U).$$

Henceforth we can do all calculations in $\mathcal{H}'(\boldsymbol{a})$, which simplifies the analysis as $\mathcal{P}'_{1,1}(n,m)$ did in Section 4.

By the same argument as in the proof of Proposition 4.1, in order to bound the probability of U, we can restrict our attention to complex sink-sets whose vertices are all reachable from some vertex v_0. If the set of vertices reachable from vertex v_0 is a complex sink-set then essentially the same algorithm as in Section 4 will terminate with S being such a sink-set. We restate the algorithm in the current setting as follows:

Start with $S = \{v_0\}$, $R = A_{\sigma(v_0)}$, and repeat the following until R is empty. Pick $i \in R$, add to S the vertex v such that $\phi(a_i) \in B_{\tau(v)}$ (if it is not already there), delete i from R and, if v was not already in S, add all elements in $A_{\sigma(v)}$ to R.

As in the proof of Proposition 4.1, the algorithm can be performed simultaneously with the generation of the random bijection ϕ and permutations σ and τ, piecemeal at each step of the algorithm.

We need some notation to describe the generation of σ and τ. Let

$$N_2^+ = N_{2,1} \cup N_{2,2} = \{i \in N' : d_i^+ > 1\}, \quad N_2^- = N_{1,2} \cup N_{2,2} = \{i \in N' : d_i^- > 1\},$$
$$N_1^+ = N' \setminus N_2^+, \quad N_1^- = N' \setminus N_2^-.$$

Also, at the start generate u.a.r. random permutations $\hat{\sigma}_j$ of N_j^+ and $\hat{\tau}_j$ of N_j^- $(j = 1, 2)$, and $\hat{\phi}$ of B, which we will view precisely as orderings of these sets (as in the proof of Proposition 4.1). Initially, let $\sigma(v_0)$ be the first element of $\hat{\tau}_j$, where j is determined by $v_0 \in N_j^+$. At each step, $\phi(a_i)$ is defined to be the next element of B in the ordering $\hat{\phi}$. At each step where $\tau^{-1}(\beta(\phi(a_i)))$ has not yet been determined, choose v to be the next unused member of N_j^- in the ordering $\hat{\tau}_j$. Set $\tau(v) = \beta(\phi(a_i))$. Then, if $\sigma(v)$ is not yet determined, define it as the next member of N_j^+ (where $v \in N_j^+$ determines j) in the ordering $\hat{\sigma}_j$.

At any given stage, when k points $\phi(a_i)$ have been chosen so far, let $K \subseteq B$ denote the set of these points, which must be the first k points of $\hat{\phi}$. (This corresponds to the set K_k in the proof of Proposition 4.1; we suppress the indices such as k for simplicity.) Also let J^+ denote the set of values $\sigma(v)$ determined so far (note this is precisely $\{\sigma(v) : v \in S\}$), and somewhat asymmetrically, define J^- to be the set of vertices whose image under τ has been determined. Then $J^- = S$ if v_0 was chosen at some stage as v, and otherwise $J^- = S \setminus v_0$. Define the following random sets referring to a step after which precisely $k < m'$ arcs have been exposed:

$$\begin{aligned} J_1^+ &= J^+ \cap N_1^+, \\ J_2^+ &= J^+ \cap N_2^+, \\ J_1^- &= J^- \cap N_1^-, \\ J_2^- &= J^- \cap N_2^-, \end{aligned}$$

and put $t_1^+ = |J_1^+|$ and so on. Then at each step of the algorithm, conditional upon having given cardinalities that can feasibly occur, the permutations $\hat{\sigma}$ etc. determine these sets, and

ensure that each of these sets occurs u.a.r. as subsets of N_1^+, N_2^+, N_1^- and N_2^- respectively, and the same property holds for K as a subset of points in B with cardinality k. Furthermore, all these sets occur jointly independently of each other. For $\mathbf{t} = (t_1^+, t_2^+, t_1^-, t_2^-)$, let $\Omega(k, \mathbf{t})$ denote the probability space of such independently chosen sets, K and the J_i^+ etc., with these cardinalities. Next define

$$\begin{aligned}
\hat{k} &= \sum_{j \in J^+} |A_j|, && (36)\\
\hat{t}_1^+ &= |(J^- \cup \{v_0\}) \cap N_1^+|, && (37)\\
\hat{t}_2^+ &= |(J^- \cup \{v_0\}) \cap N_2^+|, && (38)\\
\hat{t}_1^- &= |i \in N_1^- \,:\, K \cap B_i \neq \emptyset|, && (39)\\
\hat{t}_2^- &= |i \in N_2^- \,:\, K \cap B_i \neq \emptyset|. && (40)
\end{aligned}$$

By the form of the algorithm, at each iteration, precisely after the point when a new image of σ is exposed, we have that $t_1^+ + t_2^+ = t_1^- + t_2^-$ and also

$$\hat{t}_i^+ = t_i^+ \text{ and } \hat{t}_i^- = t_i^-, \quad i = 1, 2. \tag{41}$$

Moreover, in the event U_{v_0} that the algorithm terminates with S being a sink-set, we have

$$\hat{k} = k \tag{42}$$

and

$$k > t_1^+ + t_2^+ \tag{43}$$

if it is complex.

Thus, setting $F_{k,\mathbf{t}}$ to be the event that the tuple $\mathbf{t}$ occurs in the algorithm after k arcs are exposed, and H the event that (41) and (42) hold, we have by the union bound

$$\mathbf{P}_{\mathcal{H}'(\boldsymbol{a})}(U_{v_0}) \subseteq \sum_{k \leq m'/2} \sum_{t_1^+ + t_2^+ = t_1^- + t_2^- < k} \mathbf{P}_{\mathcal{H}'(\boldsymbol{a})}(F_{k,\mathbf{t}} \cap H). \tag{44}$$

Note that H is also defined in the space $\Omega(k, \mathbf{t})$.

We now note that, using the earlier observation that motivated defining $\Omega(k, \mathbf{t})$,

$$\mathbf{P}_{\mathcal{H}'(\boldsymbol{a})}(F_{k,\mathbf{t}} \cap H) \leq \mathbf{P}_{\mathcal{H}'(\boldsymbol{a})}(H \mid F_{k,\mathbf{t}}) = \mathbf{P}_{\Omega(k,\mathbf{t})}(H).$$

Thus it suffices to show

$$\sum_{k \leq m'/2} \sum_{t_1^+ + t_2^+ = t_1^- + t_2^- < k} \mathbf{P}_{\Omega(k,\mathbf{t})}(H) = o(1/n'), \tag{45}$$

as the lemma follows from this using the argument in Proposition 4.1 from (26) onwards.

Conditional on the values of k and $\mathbf{t}$, the random variables $\hat{k}$ etc. depend only on the random permutations $\hat{\phi}$ etc., and in particular the distribution of $\hat{k}$ only depends on t_1^+, t_2^+ and $\vec{d^+}$; the distributions of $\hat{t}_1^-$ and $\hat{t}_2^-$ only depend on k and $\vec{d^-}$; the distributions of $\hat{t}_1^+$ and $\hat{t}_2^+$ only depend on t_1^-, t_2^- and $\vec{d^+}$.

Case 1. $\log^4 n' < k \leq m'/2$

Let $g = 1/1000$. Let E_1 be the event that $|\hat{t}_1^-/(k/3) - 1| \leq g$ and $\hat{t}_2^-/(k/2) - 1 > -g$. Let E_2 be the event that $|\hat{t}_1^+/t_2^- - 1| \leq g$ and $|\hat{t}_2^+/t_1^- - 1| \leq g$. Let E_3 be the event that $|\hat{k}/(t_1^+ + 2t_2^+) - 1| \leq g$. Given any fixed values for k and $\mathbf{t}$ with $k > \log^4 n'$, if both (41) and (42) hold then, clearly, at least one of E_1, E_2 and E_3 must fail for n' sufficiently large. We claim that each of $\overline{E}_1$, $E_1 \setminus E_2$, and $(E_1 \cap E_2) \setminus E_3$ have probability $o((n')^{-5})$ in the spaces $\Omega(k, \mathbf{t})$ occurring in (45). Thus $\mathbf{P}_{\Omega(k,\mathbf{t})}(H) = o((n')^{-5})$ in all cases, yielding (45) by summing over k and the constrained $\mathbf{t}$.

To verify the claims about the E_i, the same type of argument as in Case 1 in the proof of Proposition 4.1 suffices. For instance, regarding E_1, recall that K is a random subset of the points in B of cardinality k. We can instead assume that the points of B are independently choosen with probability k/m', and condition later on obtaining precisely k points, which holds with probability $\Theta(1/\sqrt{k})$. Therefore, it is enough to show that E_1 has probability $1-o((n')^{-6})$ in the unconditional probability space where elements of B are chosen independently. Noting by (35) that $|N_1^-| \sim r$, $|N_2^-| \sim r$ and $m' = \Theta(n')$, we have by (9) that the probability that the number t_1^- of points chosen in N_1^- satisfies $|t_1^-/(k/3) - 1| > g$ is $o(1/(n')^6)$. Similarly, from Lemma 2.5 (applied to $|N_2^-|$ copies of $\mathrm{TPo}_2(\lambda)$ with $q = k/m'$), the probability that $|t_2^-/(k/2) - 1| > g$ is $o(1/(n')^6)$. For E_2, note that $\hat{t}_1^+ = |V \cap N_1^+|$, where V denotes the set of the first t_2^- elements of $\hat{\tau}_2$. Since V is a random subset of N_2^-, and since by (35) $|N_2^- \setminus N_1^+| = o(|N_2^-|)$, we have $|\hat{t}_1^+/t_2^- - 1| \leq g$ with probability $o((n')^{-5})$ provided say $t_2^- > \log^2 n'$. This is guaranteed by E_1. The other statement in E_2 works exactly the same, and thus the probability of $E_1 \setminus E_2$ is $o((n')^{-5})$. Finally, for E_3, conditional on $\boldsymbol{a}$, we just consider the fixed number $r + a_{2,1} + a_{2,2}$ of balls thrown randomly into the $a_{2,1} + a_{2,2}$ bins conditional on at least two in each bin, (one ball in all other bins) and argue as for (29) to deduce that when t_2^+ bins are selected u.a.r., with high probability they contain approximately $2t_2^+$ balls.

Case 2. $k \leq \log^4 n'$

The argument for Case 2 in the proof of Proposition 4.1 applies almost directly to the current setting, with of course $\mathcal{P}'(\vec{d})$ and $\mathcal{P}'_{1,1}(n, m)$ replaced by $\mathcal{H}'(\vec{d})$ and $\mathcal{H}'(\boldsymbol{a})$. The only twist is that we have to show that, conditional upon $\boldsymbol{a}$, the indegree sequence has maximum less than $\log^2 n$ with probability $1 - o(1/n)$. Such a sequence can be generated by putting $r + n' - a_{2,1}$ elements randomly into $a_{1,2} + a_{2,2}$ bins subject to each bin receiving at least two balls. By (35) $r + n' - a_{2,1} = 2(a_{1,2} + a_{2,2}) + o(r)$ and so the required property follows easily. □

Lemma 5.4. *Let $\boldsymbol{a}$ be any feasible tuple satisfying Γ. Then a random preheart configuration in $\mathcal{Q}(\boldsymbol{a})$ is simple and strongly connected with probability $1/9 + o(1)$.*

Proof. A preheart configuration $Q \in \mathcal{Q}(\boldsymbol{a})$ is strongly connected iff its underlying heart configuration $H = H(Q)$ is. Note moreover that H is distributed as in $\mathcal{H}(\boldsymbol{a})$, by construction.

Recall the definition of s-cycle from the proof of Proposition 4.2, and note that if H has no complex s-set of at most $m'/2$ arcs, then strong connectedness of H is equivalent to H having no s-cycles. Thus, in view of Lemma 5.3, we only need to show that a heart configuration in $\mathcal{H}(\boldsymbol{a})$ has no s-cycles with probability $1/9 + o(1)$, and that when inserting $m - m'$ vertices in

the arcs in order to generate a preheart configuration $Q \in \mathcal{Q}(\boldsymbol{a})$ we get a simple digraph a.a.s.

Since Γ holds, we have (35). We first claim that this implies that a.a.s. the number S of pairs of points that lie in the same in-bin is $O(r)$. Let $n_2 := a_{2,1} + a_{2,2}$ which must be $r - o(r)$. We have a distribution of $r + n_2$ points into n_2 in-bins chosen u.a.r. conditional upon each bin receiving at least two points. If $r - n_2 = o(\log r)$ say, immediately $S \le r + O(\log^2 r) = O(r)$. If on the other hand $r - n_2 \to \infty$ (but recall it is $o(r)$), then this multinomial distribution can be approached by n_2 independent 2-truncated Poissons conditional upon having sum $r + n_2$ (see Lemma 2.1). Combining Lemmas 2.4 and 2.2, we deduce that $S = O(r)$ with probability $1 - O((r-n_2)^{1/2} e^{-\log^3 r}) = 1 - o(1)$.

The same holds for out-bins, so we may assume that the number of ways of choosing a set $\{a_1, a_2\}$ of out-points in the same bin and a set $\{b_1, b_2\}$ of in-points in the same bin is $O(r^2)$. The probability that a_1, a_2 are matched to b_1, b_2 thus creating a double arc is $O(1/r^2)$. The probability that a given double arc in H gets no vertex inserted during the construction of Q is $(m'-2)(m'-1)/(m-2)(m-1) = O(r^2/n^2) = o(1)$. Combining these conclusions, the expected number of double arcs in H that get no vertex inserted during the construction of Q is $o(1)$, and therefore Q is simple a.a.s.

Let

$$\mu_k = \sum_{j \ge 1}^{k} \frac{2}{j} \left(\frac{2}{3}\right)^j \quad \text{and} \quad \mu = \lim_{k\to\infty} \mu_k = 2 \log \frac{1}{1 - 2/3} = \log 9.$$

The number of s-cycles of order at most k in $\mathcal{H}(\boldsymbol{a})$ is asymptotically Poisson of mean μ_k. This follows from estimating the factorial moments of this number of s-cycles in a similar way as in the proof of Proposition 4.2. The present case is simpler in two ways: firstly, there are no sets of vertices which are both a sink-cycle and a source-cycle, since this would imply having isolated cycles consisting of vertices of degree $(1,1)$. Secondly, the fact that the number of bins with degree exactly $(1,2)$ and the number of bins with degree exactly $(2,1)$ are each concentrated around r, and that the number of points in bins with higher degrees is negligible makes calculations much simpler than those in the proof of Proposition 4.2. As before, the probability of having some s-cycles of order greater than k can easily be bounded by some f_k such that $\lim_{k\to\infty} f_k = 0$. Therefore, the probability of having no s-cycles is $e^{-\mu} + o(1)$ as required. □

Finally, we proceed to prove Theorem 1.1 for the case $c \to 1$. Denote by $K(n,m)$ the number of strongly connected digraphs with n vertices and m arcs. Given any degree sequence $\vec{d} \in \widehat{\mathcal{D}}$, there are exactly $m!(m'/m)$ preheart configurations in $\mathcal{Q}(\vec{d})$. Thus, in view of Lemma 5.1 and setting A to be the event simple and strongly connected, we can write

$$\begin{aligned} K(n,m) &= \sum_{\vec{d} \in \widehat{\mathcal{D}}} \frac{m!(m'/m)\mathbf{P}_{\mathcal{Q}(\vec{d})}(A)}{\prod_{i=1}^n d_i^+! d_i^-!} \\ &= (m-1)! \mathbf{E}_{\widehat{\mathcal{D}}} \left(m' \mathbf{P}_{\mathcal{Q}(\vec{d})}(A) \right) \frac{(e^\lambda - 1)^{2n}}{\lambda^{2m}} \mathbf{P}_{\mathcal{D}}(\Sigma) \\ &\sim \frac{(m-1)!}{2\pi(m-n)} \frac{(e^\lambda - 1)^{2n}}{\lambda^{2m}} \mathbf{E}_{\widehat{\mathcal{D}}} \left(m' \mathbf{P}_{\mathcal{Q}(\vec{d})}(A) \right), \end{aligned} \tag{46}$$

since $\mathbf{P}_{\mathcal{D}}(\Sigma) \sim \frac{1}{2\pi nc(1+\lambda-c)} \sim \frac{1}{2\pi(m-n)}$ by Lemma 3.2.

To estimate $\mathbf{E}_{\widehat{\mathcal{D}}}\left(m'\mathbf{P}_{\mathcal{Q}(\vec{d})}(A)\right)$ we will restrict ourselves to the event Γ. If Γ holds, then (35) gives $m' \sim 3n'/2 \sim 3(m-n)$. From Lemmata 5.3 and 5.4, for any $\boldsymbol{a}$ satisfying Γ, we have

$$\mathbf{P}_{\mathcal{Q}(\boldsymbol{a})}(A) \sim \frac{1}{9}.$$

Moreover, from Lemma 5.2, we have that $\mathbf{E}_{\widehat{\mathcal{D}}}(m'(1-1_\Gamma)) = o(1)$ and in particular $\mathbf{P}(\Gamma) = 1 - o(1)$. Therefore,

$$\begin{aligned}\mathbf{E}_{\widehat{\mathcal{D}}}\left(m'\mathbf{P}_{\mathcal{Q}(\vec{d})}(A)\right) &= \mathbf{E}_{\widehat{\mathcal{D}}}\left(m'1_\Gamma\mathbf{P}_{\mathcal{Q}(\vec{d})}(A)\right) + \mathbf{E}_{\widehat{\mathcal{D}}}\left(m'(1-1_\Gamma)\mathbf{P}_{\mathcal{Q}(\vec{d})}(A)\right)\\ &= (1+o(1))3(m-n)\mathbf{P}_{\mathcal{Q}(n,m)}(A \mid \Gamma) + o(1)\\ &\sim (m-n)/3.\end{aligned}$$

Combining this with (46), we obtain (2) and thus complete the proof of the theorem.

6 Denser case: $c \to \infty$

In this section, we treat the case that $c \to \infty$ with $c = O(\log n)$. For such c, it follows easily from (7) that

$$c = \lambda + o(1). \tag{47}$$

Our goal is to obtain the asymptotic number of strongly connected digraphs in this case, and therefore complete the proof of Theorem 1.1. The main result in this section is the following.

Proposition 6.1. *For $c := m/n \to \infty$ with $c = O(\log n)$, a random digraph in $\mathcal{G}_{1,1}(n,m)$ is a.a.s. strongly connected.*

This result, combined with Theorem 1.2, gives the asymptotic number of strongly connected digraphs in the case that $c \to \infty$ with $c = O(\log n)$, which by (47) is asymptotic to (3), and thus the proof of Theorem 1.1 is complete.

Proof. As explained at the start of Section 4, it suffices to show that a.a.s. there are no sink-sets. As before, we let s be the cardinality of a hypothetical sink-set. By duality it suffices to consider only $s \le n/2$.

Let $G \in \mathcal{G}_{1,1}(n,m)$. We consider two cases. Let K be fixed, and chosen sufficiently large as determined by the argument in Case 2 below.

Case 1: $1 \le s \le c^K$

Let N_1 be the number of vertices of outdegree 1 in G. Recalling the relation (5) between c and λ, define $f = \lambda/(e^\lambda - 1) + n^{-1/3}$. (Note that f is a function of c.) The probability that one truncated Poisson r.v. equals 1 is $\lambda/(e^\lambda - 1)$. Hence, by (8), in the space $\mathcal{D}$ with N_1 interpreted in the natural way, we have

$$\mathbf{P}(N_1 \ge fn) = e^{-\Omega(\log^3 n)}.$$

Then the same conclusion holds in $\widehat{\mathcal{D}}$ by Lemma 3.2(a). This also transfers to the random graph $G \in \mathcal{G}_{1,1}(n,m)$ by Lemma 3.2(b) and Corollary 3.3, which show that probabilities multiply by at most $e^{O(\log^2 n)}$.

We will condition on $N_1 = n_1$ where $n_1 < fn$, and consider the set $\mathcal{N} = \{n_1 : n_1 < fn,\ \mathbf{P}(N_1 = n_1) > 1/n^2\}$. Then $\mathbf{P}(N_1 \notin \mathcal{N}) = O(1/n) = o(1)$. Let H denote the event of having a sink-set of size at most c^K. We will show that

$$\max_{n_1 \in \mathcal{N}} \mathbf{P}(H \mid N_1 = n_1) = o(1), \tag{48}$$

which implies the result immediately for this case.

Let us denote by Δ the maximum degree (in- or out-) of $G \in \mathcal{G}_{1,1}(n,m)$. It helps to consider separately the event J that $\Delta < \log^2 n$. By (4) and Lemma 2.2, we have $1 - \mathbf{P}(J) = o(1/n^2)$. So, letting X_s denote the number of sink-sets of cardinality s, equation (48) follows if we show

$$\sum_{1 \le s \le c^K} \mathbf{E}(X_s \wedge 1_J \mid N_1 = n_1) = o(1) \tag{49}$$

for all $n_1 \in \mathcal{N}$.

By symmetry, we can assume the n_1 vertices of outdegree 1 are specified in advance, so we may work in the restricted model, $\hat{\mathcal{G}}_{1,1}(n,m,n_1)$, in which $V(G)$ is partitioned into two sets of vertices, n_1 in a set A all of outdegree 1, and the rest in a set B all of higher outdegrees. This is equivalent to $\mathcal{G}_{1,1}(n,m)$ conditioned on the set of vertices of outdegree 1 being precisely A. We use $\mathbf{E}^*$ to denote expectation in this probability space.

Let us fix a set of vertices S with $|S| = s$ and $|S \cap A| = i$ (also put $R = V \setminus S$). We will bound the probability $p(s,i,q)$ that $G \in J$, S is a sink-set of G, and the set Q of arcs with both ends in S satisfies $|Q| = q$. Note that $p(s,i,q)$ does not depend on our particular choice of S. The vertices in B have outdegree at least 2, so S being a sink-set implies $q \ge 2s - i$. Also, we only consider $q \le \Delta s = (\log n)^{O(1)}$, as required by event J. By vertex symmetry,

$$\mathbf{E}^*(X_s \wedge 1_J) = \sum_{i=0}^{s} \binom{n_1}{i} \binom{n-n_1}{s-i} \sum_{q=2s-i}^{\Delta s} p(s,i,q) \le n^s \sum_{\substack{0 \le i \le s \\ 2s-i \le q \le \Delta s}} f^i p(s,i,q). \tag{50}$$

We bound $p(s,i,q)$ using a switching technique. Take any digraph $G \in J$ with a sink-set S as above, choose a set Q' of arcs with both ends *not* in S and with $|Q'| = q$, match up the arcs in Q with those in Q' in any manner, delete all arcs in Q and Q', and for each matched pair $uv \in Q$ and $u'v' \in Q'$, add the arcs uv' and $u'v$. We call this operation a *switching*. The number of ways it can be performed on G without creating any multiple arcs depends on Δ, the maximum degree. For each arc $uv \in Q$, there are at most Δ arcs wv and hence at most Δ^2 arcs wx excluded from choice as $u'v'$ due to causing double arcs. The same bound applies to the number of exclusions of the form wx coming from arcs ux. Also, at least $m - 2s\Delta$ arcs have both ends in B. Hence the number of valid switchings is at least $(m - O(\Delta^2 + s\Delta))^q$. Performing such a switching produces some digraph G' with the same (in,out)-degree sequence as G. How many such switchings can produce the same digraph G'? Assume G has r arcs directed from R to S, so G' has $r + q$ such arcs. Choose q of these and pair them up with the

q arcs of G' from S to R, to reverse the switching. This gives an upper bound (some reverse switchings may be invalid) of say $(r+q)^q$ digraphs G which can produce G'.

At this point, we may deduce that the contribution to $p(s,i,q)$ from digraphs G with the given values of r and Δ is at most

$$\frac{(r+q)^q}{(m-O(\Delta^2+s\Delta))^q} = \left(\frac{O(1)(r+q)}{m}\right)^q, \tag{51}$$

since $G \in J$ and $\Delta < \log^2 n$. Note that the contribution to $p(s,i,q)$ from all r such that $r \le c^3s^3$ or $r \le 18q$ is

$$\left(\frac{(c+q)^{O(1)}}{m}\right)^q. \tag{52}$$

To eliminate the influence of unusually large values of r will require a more elaborate argument. Assume $r > \max\{c^3s^3, 18q\}$. There must be some v of G' in S adjacent from k vertices in R, for some $k > 8c$, since otherwise we would have $r \le 8cs$. For such a vertex v, perform an additional switching to G': choose $k - \lfloor 4c \rfloor$ arcs $u_1v, \dots, u_{k-\lfloor 4c\rfloor}v$, $u_i \in R$, and replace them by arcs u_iw_i with each $w_i \in R$ (without producing multiple arcs). This produces a digraph G_1 having A as its set of vertices of outdegree 1. The number of ways of performing this switching is at least (omitting floor functions for simplicity) $\binom{k}{k-4c}(n-s-\Delta)^{k-4c}$ since each vertex u_i has outdegree at most Δ. Each possible G_1 is produced in at most $\binom{m}{k-4c}s$ ways, so the number of G' divided by the number of G_1 is at most

$$\frac{\binom{m}{k-4c}s}{\binom{k}{k-4c}(n-s-\Delta)^{k-4c}} \le s\left(\frac{ec/k}{1-s/n-\Delta/n}\right)^{k-4c} \le s(2ec/k)^{k/2} \tag{53}$$

(for large enough n), bounding the upper binomial above by $(em/(k-4c))^{k-4c}$ and the lower one below by $(k/(k-4c))^{k-4c}$, and using $k > 8c$ and $s+\Delta = (\log n)^{O(1)}$.

If any vertex of G_1 in S is still adjacent from more than $8c$ vertices in R, we may repeat the previous step, to obtain G_2, G_3 and so on, up to some graph G'' with the property that all vertices in S are adjacent from at most $8c$ vertices in R. Suppose that G'' is obtained by applying the switching to j vertices $v_1, \dots, v_j$ of G' in S which are adjacent from $k_1, \dots, k_j$ vertices in R respectively. Let $\tilde{k} = \sum_{h=1}^{j} k_h$. We must have $r \le \tilde{k} + 8cs$, since each vertex in $S \setminus \{v_1, \dots, v_j\}$ contributes to r at most $8c$. In view of (53), the factors of the j switchings multiply to give

$$s^j(2ec)^{\tilde{k}/2}\prod_{h=1}^{j} k_h^{-k_h/2} \le s^j(2ec)^{\tilde{k}/2}(\tilde{k}/j)^{-\tilde{k}/2}$$

by log-convexity of x^x. The worst case is $j = s$ which shows that the number of possible G' is $s^s\big(O(cs/\tilde{k})\big)^{\tilde{k}/2}$ times the number of G'', where $\tilde{k} \ge r - 8cs$. Note that $G'' \in \hat{\mathcal{G}}_{1,1}(n,m,n_1)$, since it has the same outdegree sequence as G. Thus, the contribution to $p(s,i,q)$ from G with given $r > \max\{c^3s^3, 18q\}$ is

$$\frac{\big(O(r+q)\big)^q}{m^q}\, s^s \left(\frac{O(cs)}{r-8cs}\right)^{(r-8cs)/2} = \frac{(O(r))^q}{m^q}\, r^q \left(O(r^{-2/3})\right)^{6q} = \left(\frac{O(1)}{r^2m}\right)^q \tag{54}$$

where we used the facts that $s \le q < r$, $cs < r^{1/3}$ and $(r-8cs)/2 \ge r/3 \ge 6q$ (for large enough n). Summing this over $r > \max\{c^3s^3, 18q\}$ and since $q \ge s \ge 1$, we obtain $(o(1)/m)^q$. In view of this and of (52), (50) gives

$$\mathbf{E}^*(X_s \wedge 1_J) \le n^s \sum_{\substack{0\le i\le s\\ 2s-i\le q\le \Delta s}} f^i((c+q)^{O(1)}/m)^q.$$

So this is at most

$$n^s \sum_{0\le i\le s} f^i((c+2s-i)^{O(1)}/m)^{2s-i} \le (c^{O(1)}n/m)^s \sum_{0\le i\le s} f^i(c^{O(1)}/m)^{s-i},$$

which is at most $\big(c^{O(1)}\max\{f,1/m\}\big)^s$. Summing this over $s \in \{1,\dots,c^K\}$ gives $o(1)$, which proves (49) as desired.

Case 2: $c^K < s \le n/2$

Let $N^+_{\le 3}$ be the number of vertices of outdegree at most 3 in G, and let $h = c^3/(e^c-1)+c^3/n$. Then $\mathbf{E}N^+_{\le 3} \le nh/2$, and by comparing with a binomial r.v. with expected value $nh/2$, and using (9), we have

$$\mathbf{P}_{\mathcal{D}}(N^+_{\le 3} \ge hn) = e^{-\Omega(hn)} = o(m^{-1}e^{-c^2}).$$

So by Lemma 3.2(a,b), we deduce that $N^+_{\le 3} < hn$ a.a.s. in $\mathcal{G}_{1,1}(n,m)$, and it suffices to prove that, conditional on this event, a.a.s. there are no sink-sets with cardinality s in the range under consideration.

Henceforth, we consider $\mathcal{P}'_{1,1}(n,m)$ (defined near the start of Section 4) conditional upon a fixed outdegree sequence satisfying $N^+_{\le 3} = n^+_{\le 3}$ for some $n^+_{\le 3} < hn$. Again by Lemma 3.2(b), it is enough to show that if R is the event that there exists a sink-set of size s satisfying $c^K \le s \le n/2$,

$$\mathbf{P}(R) = o(e^{-c^2}). \tag{55}$$

We note that our usual approach to proving properties of the indegree sequence would be to work with a sequence of independent truncated Poisson r.v.s, prove what we want, and then condition on the sum. However, the last step increases probabilities of bad events in a manner unacceptable for the present argument. To avoid this, we define an auxiliary sequence $\hat{d}^-_1,\dots,\hat{d}^-_n$ of independent copies of $\mathrm{Bin}(\hat{n},\hat{c}/n)$, where $\hat{n} = (1+\delta)n$ and $\hat{c} = (1+\delta)c$, and we set

$$\delta = \epsilon/8,\ \epsilon = 0.1.$$

This sequence will be used to stochastically dominate some random variables defined on the indegree sequence of $\mathcal{P}'_{1,1}(n,m)$.

We next define some events that hold with high probability for the degree sequence. Let $\Delta = \lfloor 5\log n/\log\log n + c^2\rfloor$ (so in particular $\Delta \le \log^3 n$, which suffices for most of our argument). This Δ turns out to be a typical bound on the maximum indegree. Let

$$p_j = \mathbf{P}(\mathrm{TPo}(\lambda) = j) = \frac{\lambda^j}{(e^\lambda-1)j!},$$
$$\hat{p}_j = \mathbf{P}(\mathrm{Bin}(\hat{n},\hat{c}/n) = j) = \binom{\hat{n}}{j}(\hat{c}/n)^j(1-\hat{c}/n)^{\hat{n}-j},$$

and set

$$j_0 = \min\{j \geq 1 : np_j \geq \log^{10} n\}, \quad j_3 = \max\{j : np_j \geq \log^{10} n\}. \tag{56}$$

We note that j_0 and j_3 are well defined since for instance we have $np_{\lfloor\lambda\rfloor} = \Omega(n/\sqrt{\log n})$. Define the interval $I = \{j_0, \ldots, j_3\}$ and let $I' = \{1, \ldots, \Delta\} \setminus I$. Informally speaking, I is the set of typical indegrees and I' the set of rare indegrees. Let V and V' be the set of vertices with indegrees in I and I', respectively. Define H to be the event that the following properties hold: $d^-_{\max} \leq \Delta$ (so $\{V, V'\}$ is a partition of the set of vertices); there exists a permutation σ of $\{1, \ldots, n\}$ with the property that $d^-_i \leq 1 + \hat{d}^-_{\sigma(i)}$ for each $i \in V$; and moreover $|V'| = o(\log^{13} n)$. (The '+1' in the inequality $d^-_i \leq 1 + \hat{d}^-_{\sigma(i)}$ is to make it easier for our argument to cope with the fact that the Poisson variable is truncated at 1, whereas the binomial is not.)

We make several claims whose proofs are postponed. The first is the following.

Claim 1: $\mathbf{P}(H) = 1 - o(e^{-c^2})$.

The rest of the proof consists of showing that $\mathbf{E}(X1_H) = O(0.93^s)$. This, together with Claim 1, gives (55) and we are done.

Given a set S of vertices ($|S| = s$), let $d^-(S)$ and $d^+(S)$ denote respectively the sum of the in- and the outdegrees of the vertices in S. (Recall that $d^+(S)$ is given since we are conditioning on a particular outdegree sequence.) We may generate $\mathcal{P}'_{1,1}(n, m)$ by specifying the random bijection ϕ last, which shows that

$$\mathbf{P}\big(S \text{ is sink-set} \mid d^-(S) = t\big) \leq \left(\frac{t}{m}\right)^{d^+(S)} \leq \left(\frac{t}{m}\right)^{4s-3i},$$

where i is the number of vertices in S of outdegree at most 3. Since the outdegree sequence was fixed and $n^+_{\leq 3} < hn$, the number of sets S of size s with parameter i is

$$\binom{n^+_{\leq 3}}{i}\binom{n - n^+_{\leq 3}}{s-i} \leq \binom{s}{i}\frac{(hn)^i n^{s-i}}{s!} \leq \left(\frac{2en}{s}\right)^s h^i.$$

Putting these together, we can bound the expected number X of sink-sets of size s restricted to the event H by distinguishing cases according to the size of $d^-(S)$:

$$\mathbf{E}(X1_H) \leq \sum_{i=0}^{s}\left(\frac{2en}{s}\right)^s h^i \left[\left(\frac{(1+\epsilon)cs}{m}\right)^{4s-3i} + \sum_{t \geq (1+\epsilon)cs}^{\Delta s \wedge m}\left(\frac{t}{m}\right)^{4s-3i}\mathbf{P}\big((d^-(S) = t) \wedge H\big)\right]. \tag{57}$$

We need some care in treating the terms with $t \geq (1+\epsilon)cs$. Let $t_1 = t(1+\epsilon/2)/(1+\epsilon)$ and $t_2 = t\epsilon/\big(2(1+\epsilon)\big)$. Note that $t_1 + t_2 = t$ with $t_1 \geq (1+\epsilon/2)cs$ and $t_2 \geq (\epsilon/2)cs$.

Let S_I and $S_{I'}$ be the subsets in S with degrees in I and I' respectively.

Claim 2: $\mathbf{P}\big((d^-(S_I) \geq t_1) \wedge H\big) \leq e^{-\Omega(t)}$.

Claim 3: $\mathbf{P}\big((d^-(S_{I'}) \geq t_2) \wedge H\big) = e^{-\Omega(t \log\log n)}$.

From these claims, it immediately follows that $\mathbf{P}\big((d^-(S) = t) \wedge H\big) = e^{-\Omega(t)}$, and with (57)

in mind we note that

$$\sum_{t\geq(1+\epsilon)cs}^{\Delta s\wedge m}\left(\frac{t}{m}\right)^{4s-3i}e^{-\Omega(t)}=O\left(\left(\frac{(1+\epsilon)cs}{m}\right)^{4s-3i}e^{-\Omega(cs)}\right)=o\left(\left(\frac{(1+\epsilon)cs}{m}\right)^{4s-3i}\right),$$

so then from (57) and using $m=cn$,

$$\mathbf{E}(X1_H)\leq(1+o(1))\sum_{i=0}^{s}\left(\frac{2en}{s}\right)^s h^i\left(\frac{(1+\epsilon)s}{n}\right)^{4s-3i}. \tag{58}$$

For $i\leq s/100$ (recalling $\epsilon=0.1$ and $s\leq n/2$),

$$\begin{aligned}\left(\frac{2en}{s}\right)^s h^i\left(\frac{(1+\epsilon)s}{n}\right)^{4s-3i} &\leq (2e(1+\epsilon))^s\left(\frac{(1+\epsilon)s}{n}\right)^{3s-3i}\\ &< (2e(1.1))^s\left(\frac{1.1}{2}\right)^{2.97s}<0.92^s,\end{aligned}$$

whilst for $i\geq s/100$,

$$\left(\frac{2en}{s}\right)^s h^i\left(\frac{(1+\epsilon)s}{n}\right)^{4s-3i}\leq\left(\frac{2en}{s}\right)^s h^{s/100}\left(\frac{1.1s}{n}\right)^s\leq\left(2.2eh^{1/100}\right)^s<0.92^s.$$

Thus, (58) gives $\mathbf{E}(X1_H)=O(s0.92^s)\leq 0.93^s$, as desired. It only remains to prove Claims 1–3.

Proof of Claim 1. If $Y\overset{\mathrm{d}}{\sim}\mathrm{TPo}(\lambda)$, then

$$\mathbf{P}(Y\geq\Delta)=O(\mathbf{P}(Y=\Delta))=O((e\lambda/\Delta)^{\Delta})=O(n^{-5}e^{-c^2}).$$

Thus the statement holds for the first part of H by taking union bound and using Lemma 3.2(a).

For the second part, consider a sequence $d_1^-,\dots,d_n^-$ of independent truncated Poisson r.v.s, and a sequence $\hat{d}_1^-,\dots,\hat{d}_n^-$ of independent binomial r.v.s, as follows:

$$d_i^-\overset{\mathrm{d}}{\sim}\mathrm{TPo}(\lambda),\qquad \hat{d}_i^-\overset{\mathrm{d}}{\sim}\mathrm{Bin}(\hat{n},\hat{c}/n),$$

where $\hat{n}=(1+\delta)n$ and $\hat{c}=(1+\delta)c$ (recalling $\delta=\epsilon/8$). For convenience, when $d_i^-=j$ we call it a vertex of indegree j, even though the sequence is randomly generated in the absence of any graph. We use a similar convention for $\hat{d}_i^-$.

Recalling the definitions of p_j and $\hat{p}_j$ above (56), we have that for $1\leq j\leq\Delta$

$$\mathbf{P}(d_i^-=j)=p_j\sim e^{-\lambda}\frac{\lambda^j}{j!}\quad\text{and}\quad\mathbf{P}(\hat{d}_i^-=j)=\hat{p}_j\sim e^{-(1+\delta)^2c}\frac{((1+\delta)^2c)^j}{j!}.$$

Let Y_j and $\hat{Y}_j$ be the numbers of vertices of indegree at least j for each of the models, and similarly, Z_j and $\hat{Z}_j$ the numbers of vertices of indegree at most j. We have

$$\mathbf{E}Y_j=n\mathbf{P}(d_i^-\geq j),\quad\mathbf{E}\hat{Y}_j=n\mathbf{P}(\hat{d}_i^-\geq j),\quad\mathbf{E}Z_j=n\mathbf{P}(d_i^-\leq j),\quad\mathbf{E}\hat{Z}_j=n\mathbf{P}(\hat{d}_i^-\leq j).$$

Recall the definition of j_0 and j_3 in (56), and let $j_1 = c - \sqrt{c}/100$ and $j_2 = (1 + 3\delta/2)c$. It is straightforward to check that $1 \leq j_0 \leq j_1 \leq j_2 \leq j_3 \leq \Delta$. If $j_2 \leq j \leq j_3$, we easily verify that $\mathbf{P}(d_i^- \geq j) = \Theta(p_j)$, and also that $p_j = o(\hat{p}_j)$ (by considering the ratios p_{j+1}/p_j and $p_j/\hat{p}_j$). Hence, we have that $\mathbf{P}(d_i^- \geq j) = o(\mathbf{P}(\hat{d}_i^- \geq j))$. If $j_1 \leq j \leq j_2$, we have that $\mathbf{P}(d_i^- \geq j) \leq 3/4$ (since $\mathrm{TPo}(\lambda)$ is asymptotically normal with mean λ and variance λ, and truncation has a negligible effect on this), and clearly $\mathbf{P}(\hat{d}_i^- \geq j) \sim 1$ (for similar reasons or using second moment method). Therefore for $j_1 \leq j \leq j_3$ we have that $\mathbf{E}Y_j \leq (4/5)\mathbf{E}\hat{Y}_j$. Moreover, note that Y_j and $\hat{Y}_j$ are binomially distributed, and for j in this range we have that $\mathbf{E}Y_j \geq np_j \geq \log^{10} n$. Hence, by (9) and taking a union bound, the probability that $|Y_j/\mathbf{E}Y_j - 1| > 1/10$ or $|\hat{Y}_j/\mathbf{E}\hat{Y}_j - 1| > 1/10$ for some j in the range $j_1 \leq j \leq j_3$ is $e^{-\Omega(\log^{10} n)}$. In particular, this implies that $Y_j \leq \hat{Y}_j$ for all $j \in [j_1, j_3]$ with probability $1 - e^{-\Omega(\log^{10} n)}$.

On the other hand, if $j_0 \leq j \leq j_1$, we easily verify that $\mathbf{P}(\hat{d}_i^- \leq j) = \Theta(\hat{p}_j)$, and also that $\hat{p}_j = o(p_j)$ (considering the ratios $\hat{p}_{j-1}/\hat{p}_j$ and $p_j/\hat{p}_j$). Therefore, $\mathbf{E}\hat{Z}_j = o(\mathbf{E}Z_j)$. Similarly as before, Z_j and $\hat{Z}_j$ are binomially distributed and $\mathbf{E}Z_j \geq np_j \geq \log^{10} n$. Using (9) again, we conclude that $\hat{Z}_j \leq Z_j$ for all $j \in [j_0, j_1]$ with probability $1 - e^{-\Omega(\log^{10} n)}$. Here we distinguish the two cases $\mathbf{E}\hat{Z}_j \geq (\log^{10} n)/2$ and $\mathbf{E}\hat{Z}_j < (\log^{10} n)/2$, and for the second case use stochastic domination of $\hat{Z}_j$ by a binomial r.v. of expectation $(\log^{10} n)/2$.

Summarising, we have that

$$Y_j \leq \hat{Y}_j, \qquad \forall j \in [j_0 + 1, j_3] \tag{59}$$

with probability $1 - e^{-\Omega(\log^{10} n)}$, where we used that $Z_j + Y_{j+1} = \hat{Z}_j + \hat{Y}_{j+1} = n$. However, what we really want is a suitable modification of (59) that holds for the range $j \in [j_0, j_3]$ and incorporates the '+1' shift in the definition of H. To do this, we distinguish two cases. If $j_0 > 1$, then it is straightforward to verify that $p_{j_0-1} n \geq \log^8 n$, so the same argument as before but changing j_0 to $j_0 - 1$ shows that $Y_j \leq \hat{Y}_j \leq \hat{Y}_{j-1}$ for all $j \in [j_0, j_3]$ with probability $1 - e^{-\Omega(\log^8 n)}$. Otherwise if $j_0 = 1$, we trivially deduce from (59) that $Y_j \leq \hat{Y}_{j-1}$ for all $j \in [j_0, j_3]$ with probability $1 - e^{-\Omega(\log^{10} n)}$. Putting everything together, we conclude that

$$Y_j \leq \hat{Y}_{j-1}, \qquad \forall j \in [j_0, j_3]$$

with probability $1 - e^{-\Omega(\log^8 n)}$. Thus, if this last inequality holds, then we can rearrange $\{1, \dots, n\}$ by some permutation σ in such a way that $d_i^- \leq 1 + \hat{d}_{\sigma(i)}^-$ for all $i \in I$.

Conditioning on the truncated Poisson r.v.s of the sequence $d_1^-, \dots, d_n^-$ having fixed sum m only multiplies the probability of failure by $O(\sqrt{m}) = o(n)$. The claim follows immediately for the second part of event H.

Finally for the last part, we look again at the sequence $d_1^-, \dots, d_n^-$ of independent truncated Poisson r.v.s. Observe that the number $|V'|$ of vertices in the sequence with indegrees in I' is binomially distributed with expectation

$$n \sum_{j \in I'} p_j \leq \Delta \log^{10} n = O(\log^{12} n).$$

We can use stochastic domination of $|V'|$ by a binomial r.v. of expectation $\log^{25/2} n$ and (9) to deduce that $|V'| \leq 2\log^{25/2} n$ with probability $1 - e^{-\Omega(\log^{25/2} n)}$. The statement follows as before by conditioning on the sum of the $d_1^-, \dots, d_n^-$ to be m.

Proof of Claim 2. Since we are restricting the probability space to the event H and the choice of S is uniformly randomised, we can bound the probability in question by replacing $\mathbf{P}\big((d^-(S_I) \geq t_1) \wedge H\big)$ by $\mathbf{P}(\hat{d}^-(S) + s \geq t_1)$, where $\hat{d}^-(S) = \sum_{i \in S} \hat{d}_i^-$ (informally speaking, $\hat{d}^-(S) + s$ is the total indegree of S after having replaced the original indegree sequence by $\hat{d}_1^- + 1, \dots, \hat{d}_n^- + 1$.) In this model $\hat{d}^-(S) \stackrel{\mathrm{d}}{\sim} \mathrm{Bin}(s\hat{n}, \hat{c}/n)$, and it is immediate to verify (using standard deviation bounds on binomials; see also (9)) that $\mathbf{P}(\hat{d}^-(S) \geq t_1 - s) \leq e^{-\Omega(t)}$, since $t_1/\mathbf{E}\hat{d}^-(S) \geq (1+\epsilon/2)/(1+\delta)^2 > 1$ and $s = o(t_1)$.

Proof of Claim 3. Recall that V' is the set of vertices with degrees in I', and that H implies that $|V'| \leq \log^{13} n$. So the contribution of these to $d^-(S)$ is at most $\log^{16} n$. Thus, if $s > \log^{16} n$, we have $\mathbf{P}(d^-(S_{I'}) \geq t_2) = 0$, since $t_2 \geq \epsilon cs/2 > \log^{16} n$.

So we may assume that $s \leq \log^{16} n$. In this case $c \leq \log^{16/K} n \leq \log^{1/5} n$ (if say $K \geq 100$). Thus, c^2 is negligible in the definition of Δ and we have $\Delta \sim 5 \log n / \log\log n$. (Here we must be precise, since the trivial bound $\Delta \leq \log^3 n$ is not enough for this part of the argument.) Let us fix any $r \leq \log^{13} n$ and restrict to the event that $|V'| = r$. Then we may use a model in which elements of S are chosen independently, each with probability s/n, and condition on the size s being achieved. Before conditioning, the number $Z = |V' \cap S|$ satisfies $Z \stackrel{\mathrm{d}}{\sim} \mathrm{Bin}(|V'|, s/n)$, and conditioning on the size s multiplies any probabilities by $O(s^{-1/2})$. Note that $\mathbf{E}Z = o(1)$ and thus, by elementary consideration of the binomial distribution, $\mathbf{P}(Z \geq j) = O(\mathbf{P}(Z = j))$. Hence

$$\begin{aligned}\mathbf{P}(d^-(V' \cap S) \geq t_2 \wedge H) &\leq \mathbf{P}(Z \geq t_2/\Delta) = O(\mathbf{P}(Z = t_2/\Delta)) = \\ &= O\left(\binom{|V'|}{t_2/\Delta}(s/n)^{t_2/\Delta}\right) = O\left(\left(\frac{e|V'|s\Delta}{t_2 n}\right)^{t_2/\Delta}\right) = e^{-\Omega(t \log\log n)}. \quad \square\end{aligned}$$

7 Loop-free case

This section treats the case that digraphs are not permitted to have loops. We prove Theorems 1.4 and 1.5, which are analogues of Theorems 1.1 and 1.3. To prove these theorems, we need the following result, which is similar to Lemmas 2.4 and 2.3.

Lemma 7.1. *Let $Y_1^+, \dots, Y_N^+, Y_1^-, \dots, Y_N^-$ be independent r.v.s with $\mathrm{TPo}_k(\lambda)$ distribution, for fixed k and for $0 < \lambda \leq \log N$. Let $c = \mathbf{E}Y_1^+$. Then for any $t \geq \sqrt{N}\log^3 N$ we have*

$$\mathbf{P}\left(\Big|\sum_{i=1}^N Y_i^+ Y_i^- - c^2 N\Big| > t\right) = O\left(e^{-(t^2/8N)^{1/5}}\right),$$

asymptotically as $N \to \infty$.

Proof. The argument is almost identical to that of Lemma 2.3, so we just state the main differences. Here, we redefine $\Delta = (t^2/8N)^{1/5}$, $Y_{\max} = \max_{1 \leq i \leq N}\{Y_i^+, Y_i^-\}$, $W_i = Y_i^+ Y_i^- - c^2$ and $W_i^* = W_i 1_{E_i}$, where E_i is the event that $Y_i^+ \leq \Delta$ and $Y_i^- \leq \Delta$. Note that $\Delta = \Omega(\log^{6/5} N)$, and that $\lambda \leq c \leq (1 + o(1))\log N$. It only remains to find appropriate bounds on $\mathbf{P}(Y_{\max} > \Delta)$, $|\mathbf{E}W_i^*|$ and $|W_i^* - \mathbf{E}W_i^*|$, and then apply the same steps as in the proof of

Lemma 2.3. The bound $\mathbf{P}(Y_{\max} > \Delta) = O(e^{-\Delta})$ is obtained analogously. In view of (4) and the fact that $\mathbf{E}((c - Y_i^-)\, 1_{Y_i^- > \Delta}) < 0$, we easily deduce

$$|\mathbf{E}W_i^*| = \mathbf{E}\left((c + Y_i^+)\, 1_{Y_i^+ \le \Delta}\right) \mathbf{E}\left((Y_i^- - c)\, 1_{Y_i^- > \Delta}\right) \le 2c\mathbf{E}\left(Y_i^-\, 1_{Y_i^- > \Delta}\right) = O(e^{-\Delta}).$$

Finally, we have $k^2 - c^2 \le W_i^* \le \Delta^2 - c^2$, and therefore $|W_i^* - \mathbf{E}W_i^*| < \Delta^2$. □

Proof of Theorem 1.5. After extending Lemma 3.2 to the loop-free case, the proof is identical to that of Theorem 1.3. So we just describe this extension of Lemma 3.2, which requires inserting an e^{-c} factor in the asymptotic expressions in parts (b) and (c). The main adjustment in the proof is to redefine $\tilde{F} = \exp(-D_0 - D^+D^-/2)$, where $D_0 = \frac{1}{m}\sum_{i=1}^n d_i^+ d_i^-$. Instead of Lemma 3.1, we use a version which excludes loops. Again, we can use [6, Theorem 4.6] with digraphs loop-free digraphs interpreted as bipartite graphs with a specific perfect matching being forbidden. Under the same conditions as Lemma 3.1, this implies that the probability that a random element of $\mathcal{P}(\vec{d})$ has no loops and no multiple arcs is

$$\exp\left(-\frac{1}{m}\sum_{i=1}^n d_i^+ d_i^- - \frac{1}{2m^2}\sum_{i,j=1}^n d_i^+(d_i^+ - 1)d_j^-(d_j^- - 1) + O\left(\frac{\Delta^4}{m}\right)\right),$$

uniformly for all $\vec{d}$. $\mathcal{B}_2$ is the event that $|D_0 - c| > t$ or $|D^+D^-/2 - \lambda^+\lambda^-/2| > t$. We bound the probability that $|D_0 - c| > t$ using Lemma 7.1 (which shows that mD_0 has expectation close to $c^2 n$). □

Proof of Theorem 1.4. Again, we only need to point out how to change the proof of Theorem 1.1. For the case that c is bounded and bounded away from 1, we simply extend Propositions 4.1 and 4.2 in Section 4 to the loop-free case, and combine them with Theorem 1.5. Proposition 4.1 implies its own extension in this new setting, since the probability of an s-set when conditioning on loop-free digraphs can only increase by the inverse of the probability of having no loops, which is $\Theta(1)$ by comparing Theorems 1.5 and 1.3. Proposition 4.2 is extended as follows.

Proposition 7.2. *Suppose that $c = m/n$ is bounded and bounded away from 1. The probability that a digraph in $\mathcal{G}_{1,1}(n, m)$ has no plain s-set is asymptotic to*

$$e^{c(2/e^\lambda - 1/e^{2\lambda})} \frac{e^\lambda(e^\lambda - 1 - \lambda)^2}{(e^{2\lambda} - e^\lambda - \lambda)(e^\lambda - 1)}, \tag{60}$$

with λ determined by the equation $c = \lambda e^\lambda/(e^\lambda - 1)$.

Proof. The argument is almost identical to that of Proposition 4.2. We sketch the main differences. C_k is again the number of s-cycles of order at most k but we exclude s-cycles of order 1 since they will be regarded as loops. D is redefined to be the number of loops and double arcs. We have

$$\mu_k = \sum_{j=2}^k \frac{2(c/e^\lambda)^j - (c/e^{2\lambda})^j}{j},$$

and

$$\mu = \lim_{k\to\infty} \mu_k = \log\left(\frac{(e^{2\lambda} - e^{\lambda} - \lambda)(e^{\lambda} - 1)}{e^{\lambda}(e^{\lambda} - 1 - \lambda)^2}\right) - c(2/e^{\lambda} - 1/e^{2\lambda}).$$

The rest of the argument consists in bounding the probability of having s-cycles of order greater than k for large k and showing that C_k and D are asymptotically jointly independent Poisson with expectations $\mathbf{E}_{\mathcal{P}_{1,1}(n,m)}C_k \sim \mu_k$ and $\mathbf{E}_{\mathcal{P}_{1,1}(n,m)}D \sim c + \lambda^2/2$. □

The formula (2), for the very sparse case ($c \to 1$) of Theorem 1.1, remains unchanged: in the proof of Lemma 5.4 one can easily see that the expected number of loops that get no vertices inserted while creating the preheart from the heart is $o(1)$ using an approach similar to the one for double arcs.

Finally, for the denser case ($c \to \infty$ with $c = O(\log n)$) it suffices to verify that Proposition 6.1 is still valid if loops are not permitted. Actually, the argument in Section 6 works for this setting with only the following trivial modifications. Note that for the first case in the proof ($s \le c^K$), the initial switchings do not create or destroy loops. The additional switchings can be performed in at least $\binom{k}{k-4c}(n - s - \Delta - 1)^{k-4c}$ ways without creating loops (which only requires replacing Δ by $\Delta + 1$) and the resulting bounds obtained in Section 6 are unaffected. The argument for the second case of the proof ($c^K < s \le n/2$) remains valid with the only difference that we have to additionally condition on having no loops. The extra effect of forbidding loops gives an additional asymptotic e^{-c} factor to the probability in Lemma 3.2(b) (see the proof of Theorem 1.5 for the extension of Lemma 3.2 to loop-free digraphs). Since $e^{-c^2} = o(e^{-c-\lambda^2/2})$, showing (55) still suffices in the loop-free context. □

Acknowledgements This paper relies heavily on results obtained in [10, 11] by Boris Pittel and the second author. We gratefully acknowledge discussions in 2001 with Boris on the subject of the present paper. We also thank the referees for pointing out many improvements to the presentation.

References

[1] E.A. Bender, E. R. Canfield and B.D. McKay, The asymptotic number of labeled connected graphs with a given number of vertices and edges, *Random Structures Algorithms* **1** (1990), 127–169.

[2] J. Cain and N. Wormald, Encores on cores, *Electronic Journal of Combinatorics* **13** (2006), Research Paper 81, 13 pp.

[3] C. Cooper and A.M. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence, *Combinatorics, Probability & Computing* **13** (2004), 319–338.

[4] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley, New York, 2000.

[5] G. Kemkes and N. Wormald, An improved upper bound on the length of the longest cycle of a post-critical random graph (submitted).

[6] B.D. McKay, Asymptotics for 0–1 matrices with prescribed line sums. in *Enumeration and Design* (D.M. Jackson and S.A. Vanstone, eds.), pp. 225–238, Academic Press, Toronto (1984).

[7] J.W. Moon and L. Moser, Almost all $(0,1)$ matrices are primitive, *Studia Sci. Math. Hungar.* **1** (1966), 153–156.

[8] I. Palásti, On the strong connectedness of directed random graphs, *Studia Sci. Math. Hungar.* **1** (1966), 205–214.

[9] B. Pittel, Counting strongly-connected, sparsely edged directed graphs, preprint.

[10] B. Pittel and N.C. Wormald, Asymptotic enumeration of sparse graphs with a minimum degree constraint, *J. Combinatorial Theory, Series A* **101** (2003), 249–263.

[11] B. Pittel and N.C. Wormald, Counting connected graphs inside-out, *J. Combinatorial Theory, Series B* **93** (2005), 127–172.

[12] E.M. Wright, Formulae for the number of sparsely-edged strong labelled digraphs, *Quart. J. Math. Oxford Ser. (2)* **28** (1977), 363–367.