The size of the largest strongly connected component of a random digraph with a given degree sequence

Colin Cooper* Alan Frieze†

April 6, 2002

Abstract

We give results on the strong connectivity for spaces of sparse random digraphs specified by degree sequence. A full characterization is provided, in probability, of the fan-in and fan-out of all vertices including the number of vertices with small ($o(n)$) and large (cn) fan-in or fan-out. We also give the size of the giant strongly connected component, if any, and the structure of the bowtie digraph induced by the vertices with large fan-in or fan-out. Our results follow a direct analogy of the extinction probabilities of classical branching processes.

1 Introduction

One of the most important questions in the theory of random graphs concerns the size of the largest component of such a graph. In their formative paper [8], Erdős and Rényi proved a strong dichotomy for the size C_1 of the largest component of the random graph $G_{n,m}$ when $m = cn/2$, c constant. [1] Erdős and Rényi[2] showed that, in $G_{n,m}$, if $c < 1$ then **whp** $C_1 = O(\log n)$ and that if $c > 1$ then $C_1 \sim G(c)n$ for some function $G(c) > 0$. A component of order n is called a *giant component*. When $c = 1$ the situation is more complicated and much effort has gone into an analysis of this case. See e.g. Bollobás [4], Łuczak [12], Łuczak, Pittel and Wierman [13], Janson, Knuth, Łuczak and Pittel [9] and the books by Bollobás [3] and by Janson, Łuczak and Ruciński [10].

Molloy and Reed [14, 15] consider a model of random graphs with a fixed degree sequence. In this model, the number of vertices of degree j is approximately $\lambda_j n$. More precisely, the limiting proportion λ_j of vertices of a given degree j

*Department of Mathematical and Computing Sciences, Goldsmiths College, London SW14 6NW, UK

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213. Supported in part by NSF grant CCR-9818411.

[1]The random graph $G_{n,m}$ has vertex set $[n]$ and m randomly chosen edges.

[2]An event $\mathcal{E}$ occurs with high probability (**whp**) if $\mathbf{Pr}(\mathcal{E})$ tends to 1 as $n \to \infty$.

is fixed, and defines a sequence of non-negative real numbers $\lambda_0, \lambda_1, \dots, \lambda_j, \dots$ where $\lambda_0 + \lambda_1 + \cdots + \lambda_j + \cdots = 1$.

Let $Q = \sum_j j(j-2)\lambda_j$. Molloy and Reed prove that if $Q < 0$ then **whp** a random graph G_λ with the implied degree sequence has maximum component size $C_1 = O(\Delta^2 \log n)$. However, if $Q > 0$ then **whp** G_λ has a unique giant component of size $C_1 \sim c_\lambda n$ for some constant $c_\lambda > 0$, and the second largest component, has size $C_2 = O(\Delta^2 \log n)$. This interesting result has recently proved to be useful in the study of random graphs with degree sequences differing from those seen in the classic theory; for example see Aiello, Chung and Lu [1] in the context of *massive graphs*.

When we consider the connectivity of digraphs we find that much has less has been done. The formative paper in this area, by Karp [11], considers the size of the strongly connected components in the random digraph $D_{n,p}$. See also Uno and Ibaraki [16].

1.1 Definitions, theorems and informal derivation of results

Let $\boldsymbol{l} = (l_{i,j},\ i,j \geq 0)$ be a set of non-negative integers satisfying $\sum_{i,j} l_{i,j} = n$ and $\sum_{i,j} il_{i,j} = \sum_{i,j} jl_{i,j}$ Let $D(\boldsymbol{l})$ be the space of simple digraphs D with vertex set $V = [n]$ and with the following property: The degree sequence of D is fixed and there are $l_{i,j}$ vertices v with in-degree $d^-(v) = i$ and out-degree $d^+(v) = j$. Let D denote a digraph chosen uniformly at random (u.a.r) from the space $D(\boldsymbol{l})$.

Let θn be the number of arcs of D. We assume that $\theta > 0$. Then by counting the total in-degree, and total out-degree respectively, we find

$$\theta n = \sum_{i,j} il_{i,j} = \sum_{i,j} jl_{i,j}.$$

Let

$$p_{i,j} = l_{i,j}/n$$

and let

$$p_j^+ = \sum_i i\frac{l_{i,j}}{\theta n} \text{ and } p_i^- = \sum_j j\frac{l_{i,j}}{\theta n}.$$

Thus p^+ is the distribution of the out-degree of the terminal vertex of a randomly chosen arc. Similarly p^- is the distribution of the in-degree of the initial vertex of a randomly chosen arc. Let

$$d = \sum_{i,j} ij\frac{l_{i,j}}{\theta n} = \sum_j jp_j^+ = \sum_i ip_i^-,$$

be the *average directed degree.*

We will first give an intuitive description of the model and state the main results. Subsequent sections deal with the technical issues and make our description precise.

We will use a sequence of absolute constansts $A_0, A_1, \ldots$ whose exact values are not always specified, but for which precise values can easily be computed.

Definition 1. *Let B^+ be the independent branching process with a single initial node in which the probability distribution of the number of descendants of a node is p^+. Let η^+ be the probability that B^+ continues indefinitely.*

Thus $1-\eta^+$, the extinction probability of B^+, is given by the smallest positive solution of $x = \sum p_j^+ x^j$ and satisfies

$$1-\eta^+ = \sum_{j=0}^{\infty} p_j^+ (1-\eta^+)^j. \tag{1}$$

If $d < 1$ then the smallest positive solution of (1) is at $1-\eta^+ = 1$, whereas if $d > 1$ there is a unique solution with $1-\eta^+ \in (0,1)$.

For a fixed vertex v of D, the *fan-out* of v, $R^+(v)$ is the set of vertices w (including v) reachable from v by a directed (v,w)-path. Similarly, the *fan-in* of v, $R^-(v)$, is the set of vertices u which can reach v by a directed (u,v)-path. We will say that v has a *small fan-out* if $|R^+(v)| \le A_0 \Delta^2 \log n$ and that v has a *small fan-in* if $|R^-(v)| \le A_0 \Delta^2 \log n$. If the fan-out is not small then **whp** it will be of size $A_1 n$, in which case we say it is *large*. Let L^+ denote the set of vertices with a large fan-out and L^- denote the set of vertices with a large fan-in.

If v has out-degree k then by doing a breadth first search from v we see that the arcs of a small $R^+(v)$ will be (approximately) the union of k independent copies of the branching process B^+. Thus a root vertex of out-degree k has probability $(1-\eta^+)^k$ of a finite progeny. Let

$$1-\pi^+ = \sum_{i,k} p_{i,k} (1-\eta^+)^k, \tag{2}$$

then $1-\pi^+$ is (approximately) the probability that a randomly chosen vertex v will have a small fan-out. We should therefore expect that

$$|L^+| \sim \pi^+ n.$$

Similarly, we expect that

$$|L^-| \sim \pi^- n,$$

where

$$1-\pi^- = \sum_{i,k} p_{i,k}(1-\eta^-)^k, \tag{3}$$

and η^- be the smallest positive value satisfying

$$1-\eta^- = \sum_i p_i^-(1-\eta^-)^i. \tag{4}$$

Thus $1-\eta^-$ is the extinction probability of the branching process B^- where the distribution of the number of progeny is given by p^-.

If $d<1$ then $\eta^+ = \eta^- = 0$ and hence $\pi^+ = \pi^- = 0$. This suggests that

$$d < 1 \text{ implies } L^+ = L^- = \emptyset.$$

For the number of vertices in a large fan-out we argue as follows: If v has a large fan-out and w has a large fan-in then it is very likely that there will be an arc directed from the fan-out of v to the fan-in of w which of course implies that w would be in the fan-out of v. Conversely, if w has a small fan-in then this is unlikely. Thus we expect that

$$R^+(v) \text{ is large } \Rightarrow R^+(v) \supseteq L^- \text{ and } |R^+(v)\setminus L^-| = o(n) \text{ and so } |R^+(v)| \sim \pi^- n. \tag{5}$$

By analogy, we expect

$$R^-(v) \text{ is large } \Rightarrow R^-(v) \supseteq L^+ \text{ and } |R^-(v)\setminus L^+| = o(n) \text{ and so } |R^-(v)| \approx \pi^+ n. \tag{6}$$

Assume now that $p_0^+, p_0^- > 0$ and $d > 1$. Let us now consider the expected number of arcs in a large fan-in or fan-out. We say that an arc (v,w) has a *large fan-in* (resp. *fan-out*) if v has a large fan-in (resp. w has a large fan-out). By analogy with the vertex case, we would expect that the number of arcs in the fan-out (resp. fan-in) of a vertex with a large fan-out (resp. fan-in) would be close to the number of arcs with a large fan-in (resp. fan-out). To estimate the number of arcs with a large fan-in we need to consider the branching process with progeny distribution p^-. Here we consider each node of the process to correspond to an *arc* of the fan-in our initial root arc. Thus we expect the number of arcs in a large fan-out to be approximately

$$\xi^+ n \text{ where } \frac{\xi^+}{\theta} = \eta^-. \tag{7}$$

Again, there are θn arcs and each has probability $\sim \eta^-$ of having a large fan-in.

Similarly, the number of arcs in a large fan-in is **whp** approximately

$$\xi^- n \text{ where } \frac{\xi^-}{\theta} = \eta^+. \tag{8}$$

Finally, we consider the size of the largest strongly connected component. It follows from (5), (6) that **whp** $L^+ \cap L^-$ is contained in a strongly connected component $\boldsymbol{S}$ of D. In fact $L^+ \cap L^-$ induces a maximal strongly connected **whp**. This is because any vertex in K must have a large fan-in and a large fan-out. The size of this strong component is approximately $(\pi^+ + \pi^- + \psi - 1)n$, where

$$\psi = \sum_{i,j} p_{i,j}(1-\eta^-)^i(1-\eta^+)^j. \tag{9}$$

The RHS of (9) is explained as follows: Choose a random vertex. It has probability $p_{i,j}$ of having in-degree i and out-degree j. The expression $(1-\eta^-)^i(1-\eta^+)^j$ is an estimate of the probability that all of the $i+j$ associated branching processes become extinct i.e. $n\sum_{i,j} p_{i,j}(1-\eta^-)^i(1-\eta^+)^j$ is a good estimate of $|\overline{L^+} \cap \overline{L^-}|$.

The above giant strongly connected component will be unique. Every other strong component will be of size $\leq A_0\Delta^2 \log n$ **whp** since every vertex not in $\boldsymbol{S}$ either has a small fan-out or a small fan-in.

This concludes the intuitive description. Now comes the hard work of making the above discussion precise.

1.2 Main results

The following theorems summarize what will be rigorously verified. Some notes on the structure of the proof of these theorems are given in Section 1.4. The theorem refers to *proper* degree sequences. The definition of proper is deferred until Section 1.3.

Theorem 1. *Let the sequence $(l_{i,j})$ be proper. If $d < 1$, then* **whp** $L^+ = L^- = \emptyset$.

Let the parameters $\eta^+, \pi^+, \xi^+, \eta^-, \pi^-, \xi^-, \psi$ be as defined in (1)-(9).

Theorem 2. *Assume that $(l_{i,j})$ is proper, that p_0^-, $p_0^+ > 0$, and that $d > 1$. Then* **whp**

(i) $|L^+| \sim \pi^+ n$ *and* $|L^-| \sim \pi^- n$.

(ii) *If $v \in L^+$ then $|R^+(v)| \sim \pi^- n$. If $v \in L^-$ then $|R^-(v)| \sim \pi^+ n$.*

(iii) *If* $v \in L^+$ *then* $R^+(v)$ *contains* $\sim \xi^+ n$ *arcs. If* $v \in L^-$ *then* $R^-(v)$ *contains* $\sim \xi^- n$ *arcs.*

(v) *If* $v \in L^+$ *then* $R^+(v) \supseteq L^-$*, and if* $w \in L^-$ *then* $R^-(w) \supseteq L^+$*.*

(vi) *There is a unique giant strongly connected component, with vertex set* $\boldsymbol{S} = L^+ \cap L^-$ *of size* $|\boldsymbol{S}| \sim (\pi^+ + \pi^- + \psi - 1)n$*.*

In fact, in Section 4, we prove rather more than this. In the notation of [6] there is a *bowtie digraph* $\boldsymbol{B} = D[L^+ \cup L^-]$ of expected size $(1-\psi)n$ induced by the union of the vertices with large fan-out or fan-in. This digraph $\boldsymbol{B}$ consists of a maximal strongly connected component $\boldsymbol{S}$ with vertex set $L^+ \cap L^-$ and wings $\boldsymbol{K}^+$ with vertex set $L^+ \cap \overline{L^-}$ and $\boldsymbol{K}^-$ with vertex set $L^- \cap \overline{L^+}$. The wing $\boldsymbol{K}^+$ has arcs directed towards $\boldsymbol{S}$ and the wing $\boldsymbol{K}^-$ has arcs directed away from $\boldsymbol{S}$. The size of any branching in the wings is $\leq A_0 \Delta^2 \log n$. The wings are of expected size asymptotic to $(1 - \pi^+ - \psi)n$ and $(1 - \pi^- - \psi)n$ respectively.

Finally we consider digraphs where there are no vertices of in-degree zero or out-degree zero. The only significant obstruction to the digraph being strongly connected is the existence of small directed cycles consisting entirely of vertices of out-degree 1, or entirely of vertices of in-degree 1.

Theorem 3. *Let* $(l_{i,j})$ *be proper, and let* $\sum l_{i,0} = \sum l_{0,j} = 0$*, then* **whp** *the structure of* D *is as follows:*

(i) *There is a unique giant strongly connected component* $\boldsymbol{S}$ *in* D*, of size* $n - O(\Delta\sqrt{n \log n})$*.*

(ii) *There is a collection* $\boldsymbol{C}$ *of vertex disjoint directed cycles. The vertices of any such cycle are all of out-degree 1 or all of in-degree 1. The subset of* $\boldsymbol{C}$*, consisting of those cycles of in-degree 1 and out-degree 1, has vertex set* $\overline{L^+} \cap \overline{L^-}$*.*

(iii) *Each cycle in* $\boldsymbol{C}$ *is connected to* $\boldsymbol{S}$ *by zero or more directed induced paths, all such paths having the same direction with respect to the given cycle.*

(iv) *Any directed path between two cycles in* $\boldsymbol{C}$ *goes through* $\boldsymbol{S}$*.*

(v) *The expected number of vertices on cycles in* $\boldsymbol{C}$ *is* β *where,*

$$\beta = \log \frac{1}{1 - p_1^-} + \log \frac{1}{1 - p_1^+} - (p_1^+ + p_1^-).$$

(vi) $\lim_{n \to \infty} \mathbf{Pr}(D \textit{ is strongly connected }) = e^{-\beta}$.

1.3 Proper Degree Sequences

Definition 2. *We say the sequence $\boldsymbol{l} = (l_{i,j},\ i,j \geq 0)$ is* proper *if:*

P1: $\theta = (1+o(1))\theta_0$ *and* $d = (1+o(1))d_0$ *where either* $d_0 \leq 1-\epsilon$ *or* $1+\epsilon \leq d_0$ *for absolute constants* θ_0, d_0, ϵ.

P2: $\sum_{i,j} i^2 p_{i,j},\ \ \sum_{i,j} j^2 p_{i,j} \leq A_3$.

P3: *Let* $\rho = \max\left(\sum_{i,j} \frac{i^2 j l_{i,j}}{\theta n},\ \sum_{i,j} \frac{j^2 i l_{i,j}}{\theta n}\right)$. *If* $\Delta \to \infty$ *with* n *then* $\rho = o(\Delta)$.

P4: $\Delta \leq n^{1/12}/\log n$.

1.4 Structure of the proof

In Section 2 we establish Theorem 2 (iii) from a gap theorem for the number of arcs in $R^+(v)$ (resp. $R^-(v)$) which establishes (7) (resp. (8)).

In Section 3 we prove Theorem 2 (i). We show that if a vertex of out-degree k has a small fan-out, the number of arcs in this fan-out is well approximated by the number of nodes of k independent copies of B^+. This establishes (1), (2). A similar analysis of small fan-in establishes (4), (3).

In Section 4, we prove Theorem 2 (iv)(b), (v), (vi). We show that **whp** $L^+ \cap L^-$ forms the vertex set $\boldsymbol{S}$ of a maximal strongly connected component.

In Section 5, we give an outline proof of Theorem 3.

2 A gap theorem for the number of arcs of the fan-out $R^+(v)$

Theorem 4. *With probability* $1 - O(1/n^2)$, *for all* $v \in V$, *the number of arcs of* $R^+(v)$ *is either* $\leq A_0 \Delta^2 \log n$ *or equal to* $n\xi^+ n \left(1 + O(\Delta\sqrt{(\log n)/n})\right)$.

2.1 Configuration model

We use the *configuration model* of Bollobás [5]. In this model a vertex $v \in V$ is represented by two sets of *configuration points* W_v^+, W_v^- where $|W_v^+| = d^+(v), |W_v^-| =$

$d^-(v)$. Let $W^+ = \bigcup_v W_v^+$ and $W^- = \bigcup_v W_v^-$. If $a \in W_u^+$ then the underlying vertex of point a is u, and we write $\phi^+(a) = u$ as a shorthand for this. Similarly, if $b \in W_v^-$ we write $\phi^-(b) = v$.

A configuration F is a random matching of W^+ with W^-. Let F be written as

$$F = \{(a,b) : a \in W^+, b \in W^-\}.$$

Let $\boldsymbol{F}$ denote the space of bipartite matchings of W^+ with W^- with the uniform measure.

Associated with F is an underlying multi-digraph $D_F \in M(\boldsymbol{l})$ with vertex set V and arc set

$$A_F = \{(u,v) : \ (a,b) \in F, \ u = \phi^+(a), \ v = \phi^-(b)\}.$$

The uniform measure on $\boldsymbol{F}$ induces a probability measure $\mathbf{Pr}(D_F)$ on multi-digraphs D_F. To justify working with configurations in this paper, we need to argue (i) that if $\boldsymbol{l}$ is proper then

$$\mathbf{Pr}(D_F \text{ is simple}) \geq \chi > 0, \tag{10}$$

where χ is a constant independent of n, and (ii) if D_1, D_2 are two simple digraphs with vertex set V and the same degree sequence d^-, d^+, then $\mathbf{Pr}(D_1) = \mathbf{Pr}(D_2)$

Now it is easily seen that

$$\mathbf{Pr}(D_F = D_1) = \mathbf{Pr}(D_F = D_2) = \frac{1}{(\theta n)!} \prod_{i=1}^{n} d_i^+! \prod_{i=1}^{n} d_i^-!. \tag{11}$$

Let $\lambda(F)$ denote the number of loops in D_F. Let $(x,y) \in F$ be *redundant* if there exists $(a',b') \in F$ such that $\phi(a') = \phi(a), \phi(b') = \phi(b)$ and $a' < a$. Let $m(F)$ denote the number of redundant pairs. The expected number of loops $\lambda(F)$ is d, and the expected number of redundant arcs is bounded above by

$$\mu = \frac{1}{2\theta^2} \left(\sum_{i,j} i^2 p_{i,j} \right) \left(\sum_{i,j} j^2 p_{i,j} \right).$$

We argue next that we can find χ in (10) satisfying

$$\chi \geq \frac{e^{-(d+2\mu)}}{8(d+\mu+1)^2}. \tag{12}$$

Let $A = 2(d+\mu)$. Then

$$\mathbf{Pr}(\lambda(F) + m(F) \geq A) \leq \frac{1}{2}. \tag{13}$$

Next let $\mathcal{M}$ be the set of all possible configurations F and let $\mathcal{M}_{a,b} = \{F \in \mathcal{M} : \lambda(F) = a, \mu(F) = b\}$. We show that for $a + b \leq A$,

$$\frac{|\mathcal{M}_{a-1,b}|}{|\mathcal{M}_{a,b}|} \geq (1 - o(1))\frac{a}{d} \qquad a > 0 \tag{14}$$

$$\frac{|\mathcal{M}_{0,b-1}|}{|\mathcal{M}_{0,b}|} \geq (1 - o(1))\frac{b}{2\mu} \text{ or } \frac{|\mathcal{M}_{0,b}|}{|\mathcal{M}|} \leq \frac{1}{n^{A-b}} \tag{15}$$

To prove (14) we consider the set of pairs $(F, F') \in \mathcal{M}_{a-1,b} \times \mathcal{M}_{a,b}$ such that F' is obtained from F by replacing 2 pairs $(x, y), (z, x')$ by $(x, x'), (z, y)$, where (x, x') is a loop. Observe that each $F \in \mathcal{M}_{a-1,b}$ is in at most $d\theta n$ such pairs and that each $F' \in \mathcal{M}_{a,b}$ is in at least $a(\theta n - o(n))$ such pairs and (14) follows.

To prove (15) we consider the set of pairs $\Pi_b = (F, F') \in \mathcal{M}_{0,b-1} \times \mathcal{M}_{0,b}$ such that F' is obtained from F by replacing 2 pairs $(x', w), (z, y')$ by $(x', y'), (z, w)$, where (x, y) is an existing edge. Observe that each $F \in \mathcal{M}_{0,b-1}$ is in at most

$$\Gamma(F) = \sum_{(x,y)\in F} d^+(\phi(x))d^-(\phi(y))$$

such pairs. Now if F is chosen uniformly from $\mathcal{M}$ then $\mathbf{E}\Gamma(F) = 2\mu\theta n$. Changing $(x, y), (x', y') \in F$ to (x, y'), (x', y) only changes Γ by at most $2\Delta^2$, and we see by the Azuma-Hoeffding martingale inequality that

$$\mathbf{Pr}(|\Gamma(F) - \mathbf{E}\Gamma(F)| \geq t) \leq \exp\left(-\frac{t^2}{4\theta n\Delta^4}\right).$$

Suppose then that $\frac{|\mathcal{M}_{0,b}|}{|\mathcal{M}|} > \frac{1}{n^{A-b}}$. Then putting $t = n/\log n$ into the above we see that for almost all $F' \in \mathcal{M}_{0,b}$ have $\Gamma(F') \leq 2\mu\theta n + o(n)$. Now let $\widehat{\mathcal{M}}_{0,b}$ denote such matchings and let $\widehat{\mathcal{M}}_{0,b-1}$ be those members F of $\mathcal{M}_{0,b-1}$ which occur with a member F' of $\widehat{\mathcal{M}}_{0,b}$ in a pair $(F, F') \in \Pi_b$. Then $\Gamma(F) \leq 2\mu\theta n + o(n) + 2\Delta^2 = 2\mu\theta n + o(n)$. Now each $F' \in \mathcal{M}_{0,b}$ is in at least $b(\theta n - o(n))$ pairs and we deuce that $|\widehat{\mathcal{M}}_{0,b-1}|/|\widehat{\mathcal{M}}_{0,b}| \geq (1 - o(1))b/(2\mu)$ and (15) follows.

It follows from (13) that we can choose $a_0 + b_0 \leq A$ such that

$$\frac{|\mathcal{M}_{a_0,b_0}|}{|\mathcal{M}|} \geq \frac{1}{2(A + 1)^2}.$$

Applying (14) we see that

$$\frac{|\mathcal{M}_{0,b_0}|}{|\mathcal{M}|} \geq \frac{e^{-d}}{2(A + 1)^2}.$$

Applying (15) we obtain

$$\frac{|\mathcal{M}_{0,0}|}{|\mathcal{M}|} \geq \frac{e^{-(d+2\mu)}}{2(A + 1)^2}.$$

This completes the proof of (12) and justifies our use of the configuration model.

2.2 The process $C^+(v)$

The process $C^+(v)$ exposes matching pairs of F, and hence arcs of D_F, one at a time. We use s to count the number of exposed arcs.

We fix a vertex v and generate F, starting with those pairs in F which are the edges used in the construction of $R^+(v)$ when fanning out from v.

Let $M(0) = \emptyset$, and let $M(s) = \{(a_i, b_i),\ i = 1, ..., s\}$ be the partial matching generated by the end of step s. Let $U^+(0) = U^-(0) = \emptyset$, and let $U^+(s) = \{a_i : i = 1, ..., s\}$, $U^-(s) = \{b_i : i = 1, ..., s\}$.

Let $A(0) = \{v\}$ and let $A(s) = \{x \in V :\ \phi^-(b_i) = x,\ 1 \leq i \leq s\}$ be the set of vertices *acquired* by the process up to this point. Thus $A(s) \subseteq R^+(v)$.

Let $I^+(s) \subseteq W^+$ be the out-points of those vertices acquired by the process. Thus $I^+(0) = W_v^+$ and $I^+(s) = \cup_{u \in A(s)} W_u^+$. The elements of $I^+(s)$ are considered to be ordered with respect to the step in which they are acquired. The ordering of elements acquired in the same step is arbitrary. In terms of the digraph D_F each step will either add a new vertex v_s reachable from v or add an arc to a vertex already reached. Let $D_F(v, s)$ denote the subgraph of D_F induced by the arcs $\{\phi^+(a_i), \phi^-(b_i)\}$.

At step $s+1$, an element a_{s+1} of $I^+(s) \setminus U^+(s)$ is chosen and matched with a point b_{s+1} chosen u.a.r from $W^- \setminus U^-(s)$. Then

$$\begin{aligned} M(s+1) &= M(s) \cup \{(a_{s+1}, b_{s+1})\} \\ U^+(s+1) &= U^+(s) \cup \{a_{s+1}\} \\ U^-(s+1) &= U^-(s) \cup \{b_{s+1}\}, \end{aligned}$$

and if $v_{s+1} = \phi^-(b_{s+1})$ then

$$\begin{aligned} A(s+1) &= A(s) \cup \{v_{s+1}\}, \\ I^+(s+1) &= I^+(s) \cup W^+_{v_{s+1}}. \end{aligned}$$

We note that, of course, v_{s+1} may already be an element of $A(s)$.

Define $X(s) = |I^+(s)| - s$ and let

$$\sigma = \min\{s :\ X(s) = 0\}. \tag{16}$$

At time σ we have will have computed the fan-out of v, $R^+(v) = A(\sigma)$ and $R^+(v)$ induces σ arcs in D_F.

The configuration F can then be completed by randomly pairing up the the elements of $W^+ \setminus I^+(\sigma)$ with $W^- \setminus I^-(\sigma)$.

2.3 Branching process lemma

This lemma summarizes some standard results about probability generating functions.

Lemma 1. *Let $f(x) = \sum p_i x^i$ be a probability generating function with expected value d.*

(i) *$f(x)$ is monotone increasing and convex in $[0, 1]$.*

(ii) *If $p_1 \neq 1$, the equation $f(x) = x$ has one solution at $x = 1$ and at most one other solution β in $[0, 1)$.*

(iii) *If $d > 1$ there are exactly two solutions to $f(x) = x$ in $[0, 1]$. If $p_0 = 0$ then 0, 1 are the solutions. If $p_0 > 0$ there is a solution $\beta \in (0, 1)$.*

(iv) *Let $d > 1$ and $p(0) > 0$. Let β be the solution in $(0, 1)$ of $x = f(x)$, then $f'(x) < 1$ for all $x \in [0, \beta]$.*

Proof

(iv) If $d > 1$ there are two solutions, β and 1, to $x = f^-(x)$ in $[0, 1]$. As $f^-(0) = p_0 > 0$ we have that $f^-(x) > x$ for $0 \leq x < \beta$ and $f^-(x) < x$ for $\beta < x < 1$. Thus, at $x = \beta$, the slope of $y = f^-(x)$ at $x = \beta$ is less than the slope of $y = x$. Thus $f'(\beta) < 1$. Moreover $f'(x)$ is an increasing function of x. □

2.4 Approximations for the expected value of $X(s)$

Let the vertex set $V = [n]$ be partitioned into fixed sets $V_{i,j} = \{v : d^-(v) = i,\ d^+(v) = j\}$ and where $l_{i,j} = |V_{i,j}|$.

The set $V \setminus A(s)$, is the set of vertices of D_F not acquired by the process after step s. Let $\delta_{i,j}(s) = |V_{i,j} \setminus A(s)|$ be the number of un-acquired vertices of indegree i and outdegree j. Thus

$$X(s) = \theta n - s - \sum_{i,j} j \delta_{i,j}(s). \tag{17}$$

Let u be a vertex of indegree i. The probability that no point in W_u^- has been matched by the process after step s is $\binom{\theta n - i}{s} / \binom{\theta n}{s}$. This follows because the sequence $S^- = (b_1, ..., b_s)$ has been selected u.a.r without replacement, from $W^- \setminus W_u^-$.

We see that $X(s)$ is a function only of the initial vertex v_0 and the sequence S^-. Thus, for $0 \le s \le \theta n$,

$$\mathbf{E}X(s) = \theta n - s - \sum_{i,j} j\left(l_{i,j} - \mathbf{1}_{v_0 \in V_{i,j}}\right) \frac{\binom{\theta n - i}{s}}{\binom{\theta n}{s}},$$

where $\mathbf{1}_{v_0 \in V_{i,j}}$ is an indicator variable for the event that $v_0 \in V_{i,j}$.

We now give some approximations for $\mathbf{E}X(s)$.

Lemma 2. **(i)** *If $s\Delta = o(n)$ then*

$$\mathbf{E}X(s) = \left((d-1)s + d^+(v_0)\right)(1 + o(1)).$$

(ii) *If $s > \Delta$ then*

$$\mathbf{E}X(s) = \theta n - s - \theta n \sum_i p_i^- \left(1 - \frac{s}{\theta n}\right)^i + O(\Delta).$$

(iii) *Suppose $d > 1$, $p_0^- \ge 0$ and let*

$$g(s) = \theta n \left(1 - \frac{s}{\theta n} - \sum_i p_i^- \left(1 - \frac{s}{\theta n}\right)^i\right).$$

Let $f(x) = \sum p_i^- x^i$, and let $\beta = 1 - \eta^-$ be the solution in $[0, 1)$ to $x = f(x)$.

(a) $1 - f'(\beta) > 0$.

(b) *If ξ^+ is as in (7) then $g(\xi^+ n) = 0$ and provided $h = o(n)$*

$$g(\xi^+ n + h) = -h\left(1 - f'(\beta) + o(1)\right).$$

(c) *The function $g(s)$ has a unique maximum s^* in $(0, \xi^+ n)$ at $s^* = cn$, $c > 0$.*

(iv) *If $d^+(v_0) > 0$ and $d > 1$, then the* unique *solution s_0 to $\mathbf{E}X(s) = 0$ in $0 \le s \le \theta n$ satisfies $s_0 = \xi^+ n + O(\Delta)$.*

Proof (i) Provided $s\Delta = o(n)$

$$\begin{aligned}
\frac{(\theta n - i)_s}{(\theta n)_s} &= \left(1 - \frac{i}{\theta n}\right)\left(1 - \frac{i}{\theta n - 1}\right) \cdots \left(1 - \frac{i}{\theta n - s + 1}\right) && (18)\\
&= 1 - i \sum_{j=0}^{s-1} \frac{1}{\theta n - j} + i^2 \sum_{j<k} \frac{1}{\theta n - j} \frac{1}{\theta n - k} + O\left(\frac{s^3 i^3}{n^3}\right)\\
&= 1 - \frac{si}{\theta n} + O\left(\frac{s^2 i^2}{\theta^2 n^2}\right) && (19)\\
&= 1 - o(1).
\end{aligned}$$

Thus, using (19),

$$\begin{aligned}
\mathbf{E}X(s) &= \theta n - s + d^+(v_0)(1-o(1)) - \sum_{i,j} j l_{i,j} \frac{(\theta n - i)_s}{(\theta n)_s} \\
&= (d-1)s + d^+(v_0)(1-o(1)) + O\left(\frac{\rho s^2}{\theta^2 n}\right) \\
&= ((d-1)s + d^+(v_0))(1-o(1)),
\end{aligned}$$

as $\boldsymbol{l}$ is proper.

(ii) If $s > \Delta$,

$$\begin{aligned}
\frac{(\theta n - i)_s}{(\theta n)_s} &= \prod_{j=0}^{i-1} \frac{\theta n - s - j}{\theta n - j} \\
&= \left(1 - \frac{s}{\theta n}\right)^i \exp\left(-\frac{s}{(\theta n)^2} \frac{i(i-1)}{2} \left(1 + O\left(\tfrac{s}{n}\right)\right)\right).
\end{aligned}$$

This follows because

$$\frac{\theta n - s - j}{\theta n - j} = \frac{\theta n - s}{\theta n} \left(1 - \frac{sj}{(\theta n)^2} \frac{1}{(1 - s/\theta n)(1 - j/\theta n)}\right).$$

Thus

$$\mathbf{E}X(s) = \theta n - s - \theta n \sum_i p_i^- \left(1 - \frac{s}{\theta n}\right)^i + O\left(\frac{\rho s}{\theta^2 n}\right) + O(\Delta).$$

As $\boldsymbol{l}$ is proper, if $\Delta \to \infty$, $\rho = o(\Delta)$ the result follows.

(iii)
(a) See Lemma 1.

(b) Certainly $g(\xi^+ n) = 0$, by definition of β. Thus, for some $\delta \in [0,1]$

$$\begin{aligned}
g(\xi^+ n + h) &= g(\xi^+ n) + h g'(\xi^+ n) + \frac{h^2}{2} g''(\xi^+ n + \delta h)) \\
&= h\left(-1 + \sum i \frac{j l_{i,j}}{\theta n} \left(1 - \frac{\xi^+}{\theta}\right)^{i-1}\right) + O\left(\frac{h^2 \rho}{\theta n}\right) \\
&= -h\left(1 - f'(\beta) + O\left(\frac{h\rho}{n}\right)\right),
\end{aligned}$$

where $f'(\beta) = \sum i p_i^- \beta^{i-1}$. As $0 \le f'(\beta) < 1$ (by Lemma 1), the Taylor expansion above is of order h for any $\beta \in [0,1)$ and $h = o(n)$.

$$
\begin{aligned}
\frac{\partial g(s)}{\partial s} &= -1 + \sum_{i,j} ij \frac{l_{i,j}}{\theta n}\left(1 - \frac{s}{\theta n}\right)^{i-1} \\
&= -1 + f'\left(1 - \frac{s}{\theta n}\right).
\end{aligned}
$$

Now $f'(0) = 0$, $f'(1) = d$ and and $f'(1 - s/\theta n)$ is strictly monotone decreasing in s, so provided $d > 1$, $g(s)$ has a unique maximum at $s^* = cn,\ c > 0$.

(iv) Let $s = \xi^+ n + h$, where $h = o(n)$ then

$$
\mathbf{E}X(\xi^+ n + h) = -h(1 - f'(\beta) + o(1)) + O(\Delta).
$$

□

2.5 Results on the stopping time σ of $C^+(v)$

Theorem 5. *Let $F \in \boldsymbol{F}$. For $v \in V$ let $\sigma(v)$ be the stopping time of of $C^+(v)$ defined in (16). With probability $1 - O\left(\frac{1}{n}\right)$, for all $v \in V$,*

(i) *If $d < 1$ then $\sigma(v) < 6\frac{\Delta^2}{(d-1)^2}\log n$.*

(ii) *If $d > 1$ then*

- **(a)** *Either $\sigma(v) < 6\frac{\Delta^2}{(d-1)^2}\log n$, or*
- **(b)** *$\xi^+ n - \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)} < \sigma(v) < \xi^+ n + \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)}$.*

Proof (i) If $s\Delta = o(n)$ we see from Lemma 2(i) that

$$
\mathbf{E}X(s) = \left(d^+(v_0) + (d-1)s\right)(1 + o(1)),
$$

and so if $d < 1$, $\mathbf{E}X(s) < 0$ when $s > d^+(v_0)/(1-d)$.

Now, $X(s)$ depends only on the sample $S^- = (b_1, \ldots, b_s)$ of length s, which is obtained by independent sampling without replacement from the set W^-. Changing the point b_s selected, can only change $X(s)$ by at most Δ. Applying the Azuma-Hoeffding martingale inequality, we get

$$
\begin{aligned}
\mathbf{Pr}(X(s) > 0) &\le \mathbf{Pr}\left(|\mathbf{E}X(s) - X(s)| \ge |\mathbf{E}X(s)|\right) \\
&\le 2\exp\left(-\frac{(\mathbf{E}X(s))^2}{2s\Delta^2}\right) \\
&\le 2\exp\left(-\frac{(d-1)^2 s}{3\Delta^2}\right).
\end{aligned}
$$

(ii) Let $s_L = \lceil 6\frac{\Delta^2}{(d-1)^2}\log n \rceil$, $s_U = \xi^+ n - \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)}$ and $s_{UU} = \xi^+ n + \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)}$.

We prove that, if $X(s) > 0$ for all $s < s_L$ then with probability $1-o(n^{-3})$ $X(s) > 0$ for $s_L \le s \le s_U$ and $X(s_{UU}) < 0$.

For $\Delta s = o(n)$

$$\mathbf{Pr}(X(s)=0) \le \mathbf{Pr}\left(|\mathbf{E}X(s) - X(s)| \ge \mathbf{E}X(s)\right) \le 2\exp\left(-\frac{(d-1)^2 s}{2\Delta^2}\right),$$

and so the proposition holds for $s_L \le s = o(n/\Delta)$. As $g(s) = \mathbf{E}X(s) - O(\Delta)$ is increasing from $s = 0$ up to $s^* = cn$, we need only consider $s^* \le s \le s_U$. However, using the Azuma-Hoeffding martingale inequality once again we see that

$$\mathbf{Pr}\left(|\mathbf{E}X(s) - X(s)| > 3\Delta\sqrt{n\log n}\right) \le 2\exp\left(-\frac{9n}{2s}\log n\right). \tag{20}$$

If $cn \le s < \xi^+ n - \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)}$ the Lemma 2(iiib) implies $\mathbf{E}X(s) \ge (4-o(1))\Delta\sqrt{n\log n}$ and so (20) implies $\sigma(v)$ is unlikely to be smaller than claimed. A similar argument using the fact that $f(\xi^+ n) = 0$ shows that with high enough probability, $X(s) < 0$ for $s = \xi^+ n + \frac{4\Delta\sqrt{n\log n}}{1-f'(\beta)}$. □

3 The number of vertices with a large fan-in or fan-out.

We make a detailed analysis of the process $C^+(v)$.

For $s < A_0\Delta^2\log n$, let

$$q_j^+(s+1) = \mathbf{Pr}(|I^+(s+1)| - |I^+(s)| = j) \qquad \text{for } j \ge 0.$$

Then

$$q_j^+(s) = \frac{\sum_i i l_{i,j} - \sum_i i|V_{i,j} \cap A(s-1)|}{\theta n - (s-1)}, \qquad j \ge 1 \tag{21}$$

$$q_0^+(s) = \frac{\sum_i i l_{i,0} + \sum_{i,j\ge 1} i|V_{i,j} \cap A(s-1)| - (s-1)}{\theta n - (s-1)}. \tag{22}$$

We see immediately that we can simplify this to

$$q_j^+(s) = p_j^+ + O\left(\frac{s\Delta}{n}\right). \tag{23}$$

We note that, once a vertex has branched for the first time, it is assigned an out-degree of zero in subsequent branchings. The value of $q_0^+(s)$ is updated to account for its unused configuration in-points. This covers the case where the fan-out of the vertex is not an arborescence.

Recall that L^+ (resp. L^-) is the set of vertices of D with large ($\geq A_1 n$) fan-out (resp. fan-in).

Theorem 6. *The following results hold* **whp**

(i) *If $d < 1$ then $L^+,\ L^- = \emptyset$.*

(ii) *If $d > 1$ then $\mathbf{E}|L^+| = (1+o(1))\pi^+ n$ and $\mathbf{E}|L^-| = (1+o(1))\pi^- n$.*

Proof (i) This follows from Theorem 5(i).

(ii) Let $B^+(k)$ denote $k = d^+(v)$ independent copies of B^+ rooted at v. We compare the process $C^+(v)$ with $B^+(k)$. We consider both processes to grow in a breadth first manner. As there is a minor problem associated with indices j for which p_j^+ is too small, we consider a modified process in which these probabilities become zero. Let $\zeta = \Delta\sqrt{\log n/n}$, let $J = \{j :\ p_j^+ \leq \zeta\}$ and let $\kappa = \sum_{j\in J} p_j^+ = O(\zeta\Delta)$. We then define

$$\tilde{p}_j^+ = \begin{cases} 0 & j \in J \\ p_j^+/(1-\kappa) & j \notin J, j \leq \Delta \end{cases} \tag{24}$$

Remark 1. *The probability that either $C^+(v)$ or $B^+(k)$ add a node with out-degree $j \in J$ during the first s steps is at most κs.*

So for small s we can "safely" restrict our attention to processes $\widetilde{C}^+(v), \widetilde{B}^+(k)$ which are free from small positive p_j^+. In particular, $\widetilde{B}^+(k)$ is the branching process with root node v of degree k and progeny distribution $\tilde{p}^+$. Similarly $\widetilde{C}^+(v)$ is $C^+(v)$ conditional on $|I^+(s+1)| - |I^+(s)| \notin J$ for $s \geq 0$. Using $\tilde{q}_j^+(s)$ to denote the conditional equivalent of $q_j^+(s)$ we see from (23) and (24) that

$$\tilde{q}_j^+(s) = p_j^+\left(1 + O\left(\frac{s\Delta}{n\zeta} + \Delta\zeta\right)\right). \tag{25}$$

In the case where the branching process has reached extinction after s steps, it forms a tree T with s edges, rooted at v, in which v has out-degree k. The out-degrees of the leaves are zero, and in general the out-degrees of the nodes of T are $k = k_0, k_1, \ldots, k_s$, where nodes are labelled in their order of addition to the branching process. We compare the probabilities of the two processes in terms of these trees. From (25),

$$\begin{aligned}
\mathbf{Pr}(\widetilde{B}^+(k)=T) &= \prod_{i=1}^{s} p^+_{k_j}(1+O(\Delta\zeta)),\\
\mathbf{Pr}(\widetilde{C}^+(v)=T) &= \prod_{i=1}^{s} p^+_{k_j}\left(1+O\left(\frac{i\Delta}{n\zeta}+\Delta\zeta\right)\right).
\end{aligned}$$

Thus

$$\mathbf{Pr}(\widetilde{B}^+(k)=T) = \mathbf{Pr}(\widetilde{C}^+(v)=T)\ \left(1+O\left(\frac{s^2\Delta}{n\zeta}+s\Delta\zeta\right)\right). \tag{26}$$

Let $|B^+(k)|$ denote the number of edges in the tree generated by $B^+(k)$. Let $\rho^+_E(v)$ denote the number of edges in the fan-out of v in the digraph D. Remark 1 and (26) imply that

$$\begin{aligned}
\mathbf{Pr}(\rho^+_E(v)=s) &= O(\kappa s)+(1-O(\kappa s))\,\mathbf{Pr}(|\widetilde{C}^+(v)|=s)\\
&= \mathbf{Pr}(|B^+(k)|=s)+O\left(\frac{s^2\Delta}{n\zeta}+s\Delta\zeta\right).
\end{aligned} \tag{27}$$

Next let $\tau = 6\Delta^2\log n/(d-1)^2$. We will show (below) that

$$\mathbf{Pr}(\tau \le |B^+(k)| < \infty) = O(n^{-1}). \tag{28}$$

Clearly,

$$\mathbf{Pr}(|B^+(k)|<\infty) = (1-\eta^+)^k. \tag{29}$$

It follows from (27), (28) and (29) that

$$\begin{aligned}
\mathbf{Pr}(|R^+(v)|\le \tau+1) &= \sum_{s\le\tau}\mathbf{Pr}(\rho^+_E(v)=s)\\
&= \mathbf{Pr}(|B^+(k)|\le\tau)+O\left(\frac{\tau^3\Delta}{n\zeta}+\tau^2\Delta\zeta\right)\\
&= (1-\eta^+)^k+o(1).
\end{aligned}$$

Part (ii) now follows from Theorem 4 and (2). □

Proof of (28)

We consider a regenerative branching process rooted at vertex v, which permits positive or negative levels of spare nodes. Specifically,

$$\begin{aligned}
Z(0) &= k\\
Z(s+1) &= Z(s)+P(s+1)-1,
\end{aligned}$$

where $P(s+1)$, the progeny at step $s+1$ is an independent random variable with distribution p^+. Thus $\mathbf{E}Z(s)=(d-1)s+k$ and,

$$\begin{aligned}\mathbf{Pr}(Z(s)\leq 0) &\leq \mathbf{Pr}\left(|Z(s)-(d-1)s|\geq (d-1)s\right)\\ &\leq 2\exp\left(-\frac{(d-1)^2 s}{2\Delta^2}\right).\end{aligned}$$

Hence

$$\mathbf{Pr}(\exists s\geq\tau : Z(s)\leq 0)=O(n^{-1})$$

and (28) follows. □

4 Structure of the giant component

If $d<1$ then **whp** any strongly connected component is of size $O(\Delta^2\log n)$. We prove below that if $d>1$ then **whp** D contains a giant strongly connected component.

The condition given in [14], for the existence (**whp**) of a giant component in a graph G with degree sequence $(\lambda_k n)$, is that $Q>0$ where $Q=\sum_k k(k-2)\lambda_k$. In our notation,

$$\begin{aligned}Q &= \sum_k k(k-2)\sum_{i+j=k}\frac{l_{i,j}}{n}\\ &= 4\theta(d-1)+\sum_{i,j}(i-j)^2\frac{l_{i,j}}{n},\end{aligned}$$

after some re-arrangement.

Thus if $d>1$ then $Q>0$, and the underlying graph G of the digraph D has a giant component **whp**. We note that the converse is untrue. For example, let $|V_{0,\Delta}|=n/2$, $|V_{\Delta,0}|=n/2$. Then $Q=\Delta(\Delta-2)>0$, so for $\Delta>2$ **whp** there is a giant component in G. However, there can never be a giant strongly connected component in D.

Lemma 3. *Let $v\in V$ and let $u\notin R^+(v)$, then*

(i) $\mathbf{Pr}(R^-(v)$ *is small* $\mid R^+(v)$ *is small* $)=\mathbf{Pr}(R^-(v)$ *is small* $)+O\left(\frac{\Delta^5\log^2 n}{n}\right)$.

(ii) $\mathbf{Pr}(R^+(u)$ *is small* $\mid R^+(v)$ *is small* $)=\mathbf{Pr}(R^+(u)$ *is small* $)+O\left(\frac{\Delta^5\log^2 n}{n}\right)$.

Proof We give a proof of part (i), the proof of part (ii) is similar. Let $R^+(v)$ be small, so that $\sigma = O(\Delta^2 \log n)$. Given the matching $M(\sigma)$ determined by step σ, the remaining configuration $\widehat{F} = F \setminus M(\sigma)$ is random. We now start a process $\widehat{C}^-(v)$ with $\widehat{W}^+ = W^+ \setminus U^+(\sigma)$ and $\widehat{W}^- = W^- \setminus U^-(\sigma)$. Let τ be the step at which the construction of $R^-(v)$ halts. The results of the previous sections still hold, provided that at no step $t \leq \tau$ we hit a vertex with a configuration point in $U^-(\sigma)$. Let $\widehat{U}^-(\tau)$ be the points of $\widehat{W}^-$ used in the construction of $R^-(v)$ and let $S = \cup_{x \in R^+(v)} W_x^-$ be the points we wish to avoid. Thus

$$\mathbf{Pr}(S \cap U^-(\tau) \neq \emptyset) = \frac{O(\Delta^2 \log n \times \Delta)}{\theta n} O(\Delta^2 \log n).$$

□

Corollary 7. **(i)** $\mathbf{E}(|\overline{L^+} \cap \overline{L^-}|) \sim \psi n$,

(ii) $\mathbf{E}(|\overline{L^+} \cap L^-|) \sim (1 - \pi^- - \psi)n$,

(iii) $\mathbf{E}(|L^+ \cap \overline{L^-}|) \sim (1 - \pi^+ \psi)n$,

(iv) $\mathbf{E}(|L^+ \cap L^-|) \sim (\pi^+ + \pi^- + \psi - 1)n$.

(v) $|L^+|$, $|\overline{L^+}|$, $|L^-|$, $|\overline{L^-}|$ *and the quantities in* **(i)**-**(iv)** *above, are concentrated within* $\sqrt{n}\Delta^{5/2}(\log n)^2$ *of their expected value, with probability* $1 - O(1/\log^2 n)$.

Proof We note that Corollary 7 **(v)** follows from Lemma 3 and the Chebychev inequality. □

Lemma 4. **whp** *the following conditions hold simultaneously for all* $u, v \in V$.

(i) $|R^+(u)|, |R^+(v)| > A_0 \frac{\Delta^2}{(d-1)^2} \log n$ *implies* $R^+(u) \cap R^+(v) \neq \emptyset$.

(ii) $|R^+(u)|, |R^-(v)| > A_0 \frac{\Delta^2}{(d-1)^2} \log n$ *implies* $R^+(u) \cap R^-(v) \neq \emptyset$.

Proof Let $R^+(u)$ be large, and let $J^-(s)$ be the unpaired "in-points" of $A(s)$ immediately after step s of $C^+(u)$. Thus

$$J^-(s) = \left(\cup_{x \in A(s)} W_x^-\right) \setminus U^-(s).$$

The expected value of $J^-(s)$ is

$$\mathbf{E}J^-(s) = \sum_{i,j} i l_{i,j} \left(1 - \frac{\binom{\theta n - i}{s}}{\binom{\theta n}{s}}\right) - s.$$

Let $q_i^- = \sum_j \frac{i l_{i,j}}{\theta n}$. As $R^+(u)$ is large, and so **whp** $\sigma \sim \xi^+ n$ then, from Lemma 2(ii) we can write

$$\mathbf{E}J^-(\sigma) = \theta n(1 + o(1)) \left(1 - \frac{\xi^+}{\theta} - \sum_i q_i^- \left(1 - \frac{\xi^+}{\theta}\right)^i\right).$$

We now prove that $\mathbf{E}J^-(\sigma) \sim cn$ for some $c > 0$.

We note that $0 \le q_i^- \le 1$ and $\sum_i q_i^- = \sum_j p_j^+ = 1$. Thus (q_i^-) is a probability distribution with expected value $d > 1$. As $q_0^- = 0$ the only positive solutions of $x = \sum_i q_i^- x^i$ are 0 and 1. Moreover $\sum_i q_i^- x^i$ is convex, so $x > \sum_i q_i^- x^i$ for $x \in (0,1)$.

Given $M(\sigma)$ determined by halting $C^+(u)$ after step σ, we can start $\widehat{C}^+(v)$ on the configuration $\widehat{F}$ with $\widehat{W}^+ = W^+ \setminus U^+(\sigma)$ and $\widehat{W}^- = W^- \setminus U^-(\sigma)$, and consider $\widehat{R}^+(v)$.

Possibly $\widehat{R}^+(v)$ is completed in $s = O(\Delta^2 \log n)$ steps. This implies either $\widehat{R}^+(v) = R^+(v)$ is small or $R^+(u) \cap (R^+(v) \setminus \widehat{R}^+(v))$ is nonempty.

Suppose now that $\widehat{R}^+(v)$ is large. At any step there are at most $\theta n - \sigma$ available in-points. Thus the probability, that at step s, there is a matching arc $(\widehat{a}_s, \widehat{b}_s)$ with $\widehat{b}_s \in J^-(\sigma)$ stochastically dominates the corresponding probability for a binomial $B(s, c/\theta)$ random variable. Thus when $s = (2\theta/c)\log n$

$$\mathbf{Pr}(\widehat{U}^-(s) \cap J^-(\sigma) = \emptyset) \le \left(1 - \frac{c}{\theta}\right)^s = O\left(n^{-2}\right).$$

$\square$

Let $D[X]$ denote the sub-digraph of D induced by the vertex set X.

Corollary 8. (i) *Let* $\boldsymbol{S} = L^+ \cap L^-$, *then* **whp** $|\boldsymbol{S}| \sim (\pi^+ + \pi^- + \psi - 1)n$.

(ii) $D[\boldsymbol{S}]$ *is a maximal strongly connected component.*

(iii) (a) *If* $u, v \in \boldsymbol{S}$ *then* $R^+(u) = R^+(v)$.

(b) *Let* $R^+(\boldsymbol{S})$ *be the fan-out of (any vertex of)* $\boldsymbol{S}$.
Let $K^+(v) = R^+(v) \setminus R^+(\boldsymbol{S})$.
whp *for all* $v \in L^+ \setminus L^-$, $|K^+(v)| = O(\Delta^2 \log n)$.

(iv) whp *for all* $v \in L^+$, $R^+(v) \supseteq L^-$.

Proof (i) follows from Corollary 7.

(ii) For all $u, v \in \boldsymbol{S}$, we have $R^+(u) \cap R^-(v) \ne \emptyset$ by Lemma 4(ii). Hence there is a directed (u,v)-path, and it follows that $\boldsymbol{S}$ is strongly connected. Let $w \in V$ and suppose there is a directed path from $\boldsymbol{S}$ to w, and from w to $\boldsymbol{S}$. Thus $R^+(w)$ and $R^-(w)$ are large, so $w \in \boldsymbol{S}$.

(iii) $K^+(v)$ is $\widehat{R}^+(v)$ in the proof of Lemma 4.

(iv) This follows from Lemma 4(ii). $\square$

In the notation of [6] there is a *bowtie digraph* $\boldsymbol{B} = D[L^+ \cup L^-]$ induced by the union of the vertices with large fan-out or fan-in. This digraph $\boldsymbol{B}$ consists of a maximal strongly connected component $\boldsymbol{S}$ with vertex set $L^+ \cap L^-$ and wings $\boldsymbol{K}^+$ with vertex set $L^+ \cap \overline{L^-}$ and $\boldsymbol{K}^-$ with vertex set $L^- \cap \overline{L^+}$. The wing $\boldsymbol{K}^+$ consists of directed paths from $L^+ \cap \overline{L^-}$ terminating at vertices of $\boldsymbol{S}$. Similarly the wing $\boldsymbol{K}^-$ consists of paths directed away from $\boldsymbol{S}$ and passing through vertices of $L^- \cap \overline{L^+}$.

Each vertex v of $\boldsymbol{K}^+$ has a small out-branching $K^+(v)$ of which the sub-branching $K^+(v) \cap \boldsymbol{K}^+$ points from v towards $\boldsymbol{S}$. Similarly, each vertex v of $\boldsymbol{K}^-$ has a small in-branching $K^-(v)$ of which the sub-branching $K^-(v) \cap \boldsymbol{K}^-$ points away from $\boldsymbol{S}$ towards v. The maximum size of the branchings $K^+(v),\ K^-(v)$ is $O(\Delta^2 \log n)$.

For a vertex $v \in L^+$, $R^+(v) = K^+(v) \cup \boldsymbol{S} \cup \boldsymbol{K}^-$. Very possibly $K^+(v) \setminus \boldsymbol{K}^+ \neq \emptyset$. In fact these vertices comprise a substantial part of $V(C_1) - V(\boldsymbol{B}) = \boldsymbol{M}$, the vertices of the giant component C_1 which do not lie within the bow-tie $\boldsymbol{B}$.

For sequences satisfying the conditions of [14], we can estimate the size of $\boldsymbol{M}$. Let $l_k = \sum_{i+j=k} l_{i,j}$, and let γ be the smallest non-negative solution of

$$x = \sum_{k \geq 1} \frac{k l_k}{2\theta n} x^{k-1}.$$

From [14], the size of the giant component C_1 is asymptotic to αn where

$$\alpha = 1 - \sum_{\kappa \geq 1} \frac{l_k}{n} \gamma^k.$$

Thus we have the following corollary.

Corollary 9. *For sequences* (l_k) *satisfying [14],* **whp** *the size of* $\boldsymbol{M}$ *is* $(1 + o(1))(\alpha - 1 + \psi)n$.

More information on the k-cores of the giant is given in [7].

5 The case where $p_0^- = p_0^+ = 0$: Proof of Theorem 3

We first prove a lemma which shows that **whp** small fan-outs (resp. fan-ins) are at most unicyclic.

Lemma 5. Whp *for all* $v \in V$*, the number of edges induced by* $R^+(v)$ *during the first* $O(\Delta^2 \log^2 n)$ *steps of* $C^+(v)$ *is at most* $|R^+(v)|$.

Proof Let $A(s)$ be the vertices of $R^+(v)$ acquired by the end of step s of the process $C^+(v)$. Consider the matching edge (a_{s+1}, b_{s+1}) selected at step $s+1$. Let $J(s) = \left(\cup_{u \in A(s)} W^-(u)\right) - U^-(s)$ be the set of unused configuration points of $A(s)$ in W^-. The probability that an element of $J(s)$ is chosen as b_{s+1} is $O(\Delta s/\theta n)$.

The probability that there exists a vertex v such that this event occurs twice during the first $O(\Delta^2 \log^2 n)$ steps of $C^+(v)$ is at most

$$O\left(n(\Delta^2 \log^2 n)^2 \left(\frac{\Delta^3 \log^2 n}{\theta n}\right)^2\right) = o(1).$$

□

Because the minimum in-degree and out-degree of the digraph is at least 1, at least one of the following is true for every vertex v. The vertex v is either on a directed cycle C, or on a directed path P which originates in a directed cycle C_1, and leads to a directed cycle C_2.

If $v \in \overline{L^+} \cap \overline{L^-}$ then v must be on an isolated directed cycle C. For, if not, the edge density of the fan-out of a vertex of C_1 contradicts Lemma 5.

Let $v \in \overline{L^+} \cap L^-$, so that $v \in \boldsymbol{K}^-$. Suppose first that v lies on a cycle C. If there is an edge incident with and directed away from C, this leads to a cycle C' contradicting Lemma 5. Next, let v be on a path P originating at a vertex $w \in \boldsymbol{S}$. P terminates in a cycle C. If any vertex of P, except w, has out-degree at least 2, we obtain a contradiction to Lemma 5.

We now consider the expected number $m^+(k)$ of subsets of vertices of out-degree 1, forming directed cycles of length $k \geq 2$.

$$\begin{aligned}
m^+(k) &= \sum_{\sum f_i = k} \frac{(k-1)!}{(\theta n)_k} \prod_{i \geq 1} \binom{l_{i,1}}{f_i} i^{f_i} \\
&= \frac{1}{k} \sum_{\sum f_i = k} \binom{k}{f_1, ..., f_i, ...} \frac{1}{(\theta n)_k} \prod_{i \geq 1} i^{f_i} (l_{i,1})_{f_i} \\
&= \left(1 + O\left(\frac{k^2}{n}\right)\right) \frac{1}{k} \sum_{\sum f_i = k} \binom{k}{f_1, ..., f_i, ...} \prod \left(\frac{i l_{i,1}}{\theta n}\right)^{f_i} \\
&= \left(1 + O\left(\frac{k^2}{n}\right)\right) \frac{1}{k} \left(p_1^+\right)^k .
\end{aligned}$$

Theorem 3 (i),(ii) follow from standard techniques. Theorem 3 (iii) follows by applying the methods of [2]

References

[1] W. Aiello, F.R. Chung and L. Lu, A random graph model for massive graphs, *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (2000) 171-180.

[2] R. Arratia, L. Goldstein and L. Gordon, Two moments suffice for Poisson approximations: the Chen–Stein method, *Annals of Probability* 17 (1989) 9-25.

[3] B.Bollobás, *Random Graphs,* Academic Press, 1985. (second edition 2001).

[4] B.Bollobás, The evolution of random graphs, *Transactions of the American Mathematical Society* 286 (1984) 257-274.

[5] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal on Combinatorics* 1 (1980) 311-316.

[6] A. Broder, R. Kumar, F.Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener. *Graph structure in the web.* `http://gatekeeper.dec.com/pub/DEC/SRC/publications/stata/www9.htm`

[7] C. Cooper, The size of the cores of a random graph with a given degree sequence, to be submitted, (2002).

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61

[9] S. Janson, D.E. Knuth, T. Łuczak and B.G. Pittel, The birth of the giant component, *Random Structures and Algorithms* 4 (1993) 233-358.

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

[11] R.M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* 1 (1990) 73-94.

[12] T. Łuczak, Components behavior near the critical point of the random graph process, *Random Structures and Algorithms* 1 (1990) 287-310.

[13] T. Łuczak, B.G. Pittel and J.C. Wierman, The structure of a random graph near the point of the phase transition, *Transactions of the American Mathematical Society* 341 (1994) 721-748.

[14] Molloy and B.A. Reed, A Critical Point for Random Graphs with a Given Degree Sequence, *Random Structures and Algorithms* 6 (1995) 161-180.

[15] M. Molloy and B.A. Reed, The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence, *Combinatorics, Probability and Computing* 7 (1998) 295-306.

[16] Y. Uno and T. Ibaraki, Reachability Problems of Random Digraphs, *Graphs and Networks*, Vol.E81-A No.12 p.2694.