Birth of a strongly connected giant in an inhomogeneous random digraph

M. Bloznelis [1], F. Götze, J. Jaworski[2]

Vilnius University, LT-03225 Vilnius Lithuania

Bielefeld University D-33501 Bielefeld Germany

Adam Mickiewicz University 61–614 Poznań Poland

Abstract

We present and investigate a general model for inhomogeneous random digraphs with labeled vertices, where the arcs are generated independently, and the probability of inserting an arc depends on the labels of its endpoints and its orientation. For this model the critical point for the emergence of a giant component is determined via a branching process approach.

key words: inhomogeneous digraph, phase transition, giant component.

2010 Mathematics Subject Classification 05C82, 90B15, 60C05

1 Introduction

Random directed graphs (digraphs) are widely used for modeling networks arising e.g. in physics, biology, social studies and, more recently, bioinformatics, linguistics and the analysis of networks in the internet.

It has been observed that the well studied classical homogeneous random digraph model $D(n,p)$, where arcs are inserted independently and with the same probability p, may not fit real life networks, because the latter often exhibit statistical properties, such as, e.g., power law in-degree/out-degree distribution, that are inconsistent with the model $D(n,p)$. Generally, the real life networks are inhomogeneous (see, e.g., Albert and Barabási [1], Dorogovtsev, Mendes and Samukhin [3], Durrett [4], Newman, Strogatz and Watts [11]).

In this paper we study a very general model of sparse inhomogeneous random digraphs with *independent* arcs. By 'sparse' we mean that the number of arcs does not grow faster than linearly in n, where n is the number of particles (called vertices of the digraph). By inhomogeneity we mean that different arcs are inserted with different probabilities. In particular, our random digraph model is able to produce a wide class of asymptotic in-degree and out-degree distributions including power law distributions.

[1]Research supported by SFB 701 at the University of Bielefeld, Germany

[2]Research supported by the Marie Curie Intra-European Fellowship No. 236845 (RANDOMAPP) within the 7th EU Framework Programme

The main question addressed in the paper is the description of the phase transition in the (strongly connected) cluster size, that is, the emergence of a giant strongly connected component. To answer this question we establish a first order asymptotic for the number N_1 of vertices in the largest strongly connected component, by showing $N_1 = \rho n + o_P(n)$. The fraction $\rho \geq 0$ is expressed in terms of survival probabilities of related branching processes that reflect the statistical properties of a certain neighbourhood of a randomly chosen vertex.
Let us mention that mathematically rigorous results on phase transition in $D(n,p)$ were established in Karp [7] and Łuczak [8], see also Łuczak and Seierstad [10]. More general three parameter models were studied in Łuczak and Cohen [9]. The phase transition in a random digraph with given (non-random) in-degree and out-degree sequences was shown in Cooper and Frieze [5]. We also want to mention the related in-depth study of the phase transition phenomenon in general inhomogeneous random graphs in Bollobás, Janson and Riordan [2], which has inspired our work on inhomogeneous digraphs.

The paper is organized as follows. In Section 2 we define a finite dimensional model of a random inhomogeneous digraph and for this model determine the critical point of the phase transition. Section 3 extends finite dimensional results to inhomogeneous digraphs defined by very general (possibly infinite dimensional) kernels. Proofs are postponed to Section 4.

2 The finite dimensional model

Before presenting our results we briefly recall some relevant facts and notation related to digraphs.
2.1. A digraph D on the vertex set $V = \{v_1, \dots, v_n\}$ is a subset of the set $[V]^2 = \{(u,v),\ u,v \in V\}$ of all ordered pairs of elements of V. Elements of D are called directed edges or arcs. The fact that (u,v) is an element of D is denoted $\{u \to v\} \in D$ or just $u \to v$. More generally, if for some distinct vertices $w_1, \dots w_k \in V$ the collection of arcs $\mathcal{P} = \{(w_1,w_2),\ (w_2,w_3),\ \dots,\ (w_{k-1},w_k)\}$ is a subset of D then $\mathcal{P}$ is called a directed path (*d-path*) starting at w_1 and ending at w_k, and denoted by $v \rightsquigarrow w$. If $v \rightsquigarrow w$ and $w \rightsquigarrow v$ then the vertices v and w are said to *communicate*. In this case we write $v \leftrightsquigarrow w$. In addition, we define that $v \leftrightsquigarrow v$, for every $v \in V$, even in the case where the loop $v \rightsquigarrow v$ is not present in D. A digraph is called *strongly connected* if every pair of its vertices communicates. Since $' \leftrightsquigarrow '$ is an equivalence relation, it splits the vertex set V into a union of disjoint subsets of elements communicating with each other. The subgraph of D induced by such a subset of vertices is called a strongly connected component (*SC-component*).
We are interested in the fast growth of the largest SC-component when the density of random arcs gradually increases in the range $\Theta(n^{-1})$. Here the important parameter is the size $N_1 = N_1(D)$ of the largest SC-component (the number of vertices of the SC-component, which has the largest number of vertices). Another interesting parameter is the size $N_2 = N_2(D)$ of the second largest SC-component.
2.2. We assume that vertices belong to different types and that the probability of an arc depends on the types of its endpoints and the scale parameter n only. In addition, we assume throughout this section that the set of different types is finite and as the number of vertices increases. A similar model of *random graphs* (but not digraphs) has been introduced by Söderberg [12].
Let us introduce some more notation. Let $S = \{s_1, \dots, s_k\}$ denote the set of types, and let $s(v)$ denote the type of a vertex $v \in V$. We write $n = n_1 + n_2 + \dots + n_k$, where each $n_i = \#\{v \in V : s(v) = s_i\}$ denotes the number of vertices of type $s_i \in S$.

Given a integer vector $\overline{n} = (n_1, \ldots, n_k)$ and a $k \times k$ matrix $\mathbb{P} = ||p_{ij}||$ with non-negative entries, define the inhomogeneous random digraph $\mathcal{D}$ on the vertex set V as follows. The set of arcs of $\mathcal{D}$ is drawn at random from $[V]^2$ so that events $\{u \to v\} \in \mathcal{D}$ are independent and have probabilities $\mathbf{P}(u \to v) = 1 \wedge (p_{ij} n^{-1})$, for each $(u, v) \in [V]^2$. Here i and j refer to the types $s_i = s(u)$ and $s_j = s(v)$ of the endpoints u and v. We use the notation $a \wedge b$ and $a \vee b$ for $\min\{a, b\}$ and $\max\{a, b\}$ respectively. In order to stress the dependence of the model on the parameters $\mathbb{P}$ and $\overline{n}$ we sometimes write $\mathcal{D} = \mathcal{D}_{\mathbb{P},\overline{n}}$.
We shall assume that the fraction of vertices of a given type is asymptotically constant. That is, there is a probability distribution Q on the type space S such that for each $q_i = Q(s_i)$ we have

$$q_i > 0 \qquad \text{and} \qquad n_i - q_i n = o(n) \qquad \text{as} \qquad n \to \infty. \tag{1}$$

Clearly, for large n, the typical/statistical characteristics of the digraph $\mathcal{D}_{\mathbb{P},\overline{n}}$ depend solely on the distribution Q and matrix $\mathbb{P} = ||p_{ij}||$. In order to describe the phase transition in $\mathcal{D}_{\mathbb{P},\overline{n}}$ in terms of Q and $\mathbb{P}$ we use the "language" of branching processes.
Let us consider multi-type Galton-Watson processes where particles are of types from S. Given $s \in S$, let $\mathcal{X}(s)$ (respectively $\mathcal{Y}(s)$) denote the Galton-Watson process starting at a particle of type s such that the number of children of type $s_j \in S$ of a particle of type $s_i \in S$ has Poisson distribution with mean $p_{ij} q_j$ (respectively $p_{ji} q_j$), $1 \le i, j \le k$. We write $\mathcal{X} = \{\mathcal{X}(s),\, s \in S\}$ and $\mathcal{Y} = \{\mathcal{Y}(s),\, s \in S\}$. Let $\rho_{\mathcal{X}}(s)$ and $\rho_{\mathcal{Y}}(s)$ denote the non-extinction probability of $\mathcal{X}(s)$ and $\mathcal{Y}(s)$ respectively. Write

$$\rho = \rho_{\mathcal{X}\mathcal{Y}} = \sum_{1 \le i \le k} \rho_{\mathcal{X}}(s_i) \rho_{\mathcal{Y}}(s_i) q_i. \tag{2}$$

We show in Theorem 1 below that a giant SC-component of range n emerges in $\mathcal{D}$ whenever ρ is positive.

Theorem 1. *Assume that (1) holds. Assume that the matrix $\mathbb{P} = ||p_{ij}||$ is irreducible. As $n \to \infty$ we have*

$$N_1(\mathcal{D}_{\mathbb{P},\overline{n}}) = \rho_{\mathcal{X}\mathcal{Y}}\, n + o_P(n) \tag{3}$$

and $N_2 = o_P(n)$.

Here for a sequence of random variables $\{Z_n\}$ we write $Z_n = o_P(n)$ if $\lim_n \mathbf{P}(|Z_n| > \delta n) = 0$ for each $\delta > 0$.
Recall that the matrix $||p_{ij}||$ is called *irreducible* if its associated digraph D_P is strongly connected. Here D_P is the digraph on the vertex set S such that $\{s_i \to s_j\} \in D_P$ whenever $p_{ij} > 0$.
Remark. In the case where D_P is not strongly connected, Theorem 1 can be applied to its strongly connected components. Let $D_1, \ldots, D_r$ denote the strongly connected components of D_P. Given $1 \le m \le r$, let $S_m \subset S$ denote the vertex set of D_m and let $\mathcal{D}_m$ denote the subgraph of the random digraph $\mathcal{D}$ induced by the vertex set $V_m = \{v \in V : s(v) \in S_m\}$. Note that, for $i \ne j$, any two vertices $v \in V_i$ and $u \in V_j$ do not communicate in $\mathcal{D}$. Therefore, each strongly connected component of $\mathcal{D}$ is a subgraph of some $\mathcal{D}_m$. The asymptotic size of the largest strongly connected component of $\mathcal{D}_m$ is, by Theorem 1, $N_1(\mathcal{D}_m) = \rho_m n + o_P(n)$. Here we denote $\rho_m = \sum_{s \in S_m} \rho_{\mathcal{X}}(s) \rho_{\mathcal{Y}}(s) Q(s)$. It follows now that $N_1(\mathcal{D}) = \max_{1 \le m \le r} \rho_m n + o_P(n)$. In this way we obtain an extension of (3) to general (not necessarily strongly connected) digraphs D_P.

3 The general model

The result of Theorem 1 extends to a much more general situation where the type space S is a separable metric space, Q is a probability measure defined on Borel sets of S. Here the matrix $||p_{ij}||$ is replaced by a non-negative $S \times S$ measurable kernel $\kappa(s,t)$, $s,t \in S$.

Let $x_1, x_2, \ldots$ be a sequence of random variables with values in S such that the empirical distribution of the first n observations $x_1, \ldots, x_n$ approximates the measure Q in probability as $n \to \infty$. That is, we assume that for each Q-continuous Borel set $A \subset S$ we have $\#\{i \in [1,n] : \, x_i \in A\} n^{-1} = Q(A) + o_P(1)$ as $n \to \infty$. Recall that a Borel set A is called *Q-continuous* whenever its boundary ∂A has zero probability $Q(\partial A) = 0$.

Given n, let $\mathcal{D}_n$ be the random digraph on the vertex set $\{x\}_1^n = \{x_1, \ldots, x_n\}$ with independent arcs having probabilities $\mathbf{P}(\{x_i \to x_j\} \in \mathcal{D}_n) = 1 \wedge (n^{-1}\kappa(x_i, x_j))$, $1 \le i,j \le n$. Combining S, Q and κ we obtain a very large class of inhomogeneous digraphs with independent arcs. Obviously, the model will include digraphs with in-degree and out-degree distributions which have power laws.

Such a general model, for *random graphs* (not digraphs), was introduced in Bollobás, Janson and Riordan [2]. Note that in the case of random graphs it is necessary to assume, in addition, that the kernel κ is symmetric. In the definition of digraphs $\mathcal{D}_n$, $n \ge 2$, we do not require the symmetry of the kernel.

For large n, the phase transition in the digraph $\mathcal{D}_n$ can be described in terms of the survival probabilities of the related multi-type Galton–Watson branching processes with type space S. Given $s \in S$, let $\mathcal{X}(s)$ (respectively $\mathcal{Y}(s)$) denote the Galton-Watson process starting at a particle of type s such that the number of children of types in a subset $A \subset S$ of a particle of type $t \in S$ has Poisson distribution with mean $\int_A \kappa(t,u) Q(du)$ (respectively $\int_A \kappa(u,t) Q(du)$). These numbers are independent for disjoint subsets A and for different particles. The critical point of the emergence of the giant SC-component is determined by the averaged joint survival probability

$$\rho_{\mathcal{X}\mathcal{Y}} = \int_S \rho_{\mathcal{X}}(s) \rho_{\mathcal{Y}}(s) Q(ds) \tag{4}$$

being positive. Here $\rho_{\mathcal{X}}(s)$ and $\rho_{\mathcal{Y}}(s)$ denote the non-extinction probabilities of $\mathcal{X}(s)$ and $\mathcal{Y}(s)$ respectively. In particular, for the general model of an inhomogeneous digraph, (3) reads as follows

$$N_1(\mathcal{D}_n) = \rho_{\mathcal{X}\mathcal{Y}} \, n + o_P(n) \qquad \text{as} \qquad n \to \infty. \tag{5}$$

In order to establish (5) we need to impose further conditions on the kernel κ, like those in [2]. Namely, we need to assume that the kernel κ is irreducible $Q \times Q$ almost everywhere (see Remark after Theorem 1). That is, for any measurable $A \subset S$ with $Q(A) \ne 1$ or 0, the identity $Q \times Q\big(\{(s,t) \in A \times (S \setminus A) : \, \kappa(s,t) \ne 0\}\big) = 0$ implies $Q(A) = 0$ or $Q(S \setminus A) = 0$, see [2]. In addition we assume that κ is continuous almost everywhere on $(S \times S, Q \times Q)$, and the number of arcs in $\mathcal{D}_n$, denoted by $|\mathcal{D}_n|$, satisfies

$$n^{-1}\mathbf{E}|\mathcal{D}_n| \to \int\int_{S \times S} \kappa(s,t) Q(ds) Q(dt) < \infty$$

as $n \to \infty$. Note that here we implicitly assume that κ is integrable, i.e., $\kappa \in L_1(S \times S, Q \times Q)$. We will not give the proof of (5). It can be obtained from Theorem 1 via a finite dimensional approximation argument similar to that of the proof of Theorem 3.1 in [2].

4 Proof

In the proof we shall use ideas and techniques developed in Karp [7] and Bollobás, Janson and Riordan [2].

Proof of Theorem 1. Given a vertex $v \in V$, let $X(v)$ denote the set of vertices that can be reached from v via d-paths, and let $Y(v)$ denote the set of starting points of d-paths ending at v,

$$X(v) = \{u \in V : v \rightsquigarrow u\}, \qquad Y(v) = \{u \in V : u \rightsquigarrow v\}.$$

Given a function $\omega(n)$ such that $\omega(n) \to \infty$ and $\omega(n) = o(n)$ as $n \to \infty$, we say that $v \in V$ is $x-$*big* (respectively $y-$*big*) if $|X(v)| \geq \omega(n)$ (respectively, $|Y(v)| \geq \omega(v)$). The set of $x-$big (respectively $y-$big) vertices is denoted $B_x = B_x(\omega)$ (respectively $B_y = B_y(\omega)$). We write $B = B(\omega) = B_x(\omega) \cap B_y(\omega)$ for the set of vertices which are $x-$big and $y-$big simultaneously. We show that for any such function ω,

$$n^{-1}|B(\omega)| - \rho_{\mathcal{X}\mathcal{Y}} = o_P(1) \qquad \text{as} \quad n \to \infty. \tag{6}$$

For this purpose it suffices to show that for each ω we have

$$n^{-1}\mathbf{E}|B(\omega)| = \rho_{\mathcal{X}\mathcal{Y}} + o(1) \tag{7}$$

and to establish (6) for at least one such function, say $\omega_0(n) = \ln n$.
Indeed, assume that (7) holds. Let ω, ω' be two such functions and B, B' denote the corresponding sets of large vertices. For the size of the symmetric difference $B \triangle B' = (B \cup B') \setminus (B \cap B')$ equals $|B(\omega \wedge \omega')| - |B(\omega \vee \omega')| \geq 0$, we obtain from (7) that $\mathbf{E}|B \triangle B'| = o(n)$. It follows that $\big||B| - |B'|\big| \leq |B \triangle B'| = o_P(n)$. In particular, (6) holds for every ω whenever it is satisfied by at least one such function ω.
Proof of (7). Forward exploration. Given $v \in V$, we explore the set $X(v)$ as follows. Color all vertices blue. Color v white and put it in the list, which now contains a single white vertex v. Then proceed recursively: choose a white vertex from the list, color it black, reveal all outgoing arcs emerging from this vertex to blue vertices, color these vertices white and add them to the list. Stop when we have collected at least $\omega(n)$ vertices in the list (hence $v \in B_x(\omega)$), or there are no white vertices left in the list (hence we have explored entire set $X(v)$ and $v \notin B_x(\omega)$). Let $X_\omega(v)$ denote the set of vertices collected in the list. Given $u, w \in X(v)$ we say that u is an *f-child* (forward child) of w if $\{w \to u\} \in \mathcal{D}$ and u was blue when w discovered it during the exploration process. Since the last black vertex adds to the list all its blue neighbours (endpoints of outgoing arcs), the size $|X_\omega(v)|$ can not exceed $\omega(n)$ by more than the out-degree Δ_x of the last black vertex.
Backward exploration. We perform the same exploration process starting at v as above, but in the transposed digraph $\mathcal{D}^*$, which is obtained from $\mathcal{D}$ by reversing direction of arcs ($\{v \to u\} \in \mathcal{D} \Leftrightarrow \{u \to v\} \in \mathcal{D}^*$). That is, now the search for neighbours propagates in the reverse direction of the arcs of $\mathcal{D}$. Let $Y_\omega(v)$ be the subset of $Y(v)$ obtained in at most $\omega(n)$ steps of the exploration. Given $u, w \in Y_\omega(v)$, the vertex u is called a $b-$*child* (backward child) of w if $\{u \to w\} \in \mathcal{D}$ and u was blue when w discovered it. Again, we have $|Y_\omega(v)| \leq \omega(n) + \Delta_y$, where now Δ_y is the out-degree in $\mathcal{D}^*$ (in-degree in $\mathcal{D}$) of the vertex last explored.
By a coupling of an exploration process with the approximating Galton-Watson process, Bollobás, Janson and Riordan [2] showed that the fraction of large vertices (the number $|B(\omega)|/n$) converges in probability to the survival probability of the Galton-Watson process, see Lemma 9.6 in [2]. Their results are stated for (undirected) random graphs, but several steps of their

proof extend to random digraphs as well. In particular, by a coupling of the forward exploration process with $\mathcal{X}$ (backward exploration process with $\mathcal{Y}$) we obtain as $n \to \infty$ uniformly in $v \in V$

$$\mathbf{P}(v \in B_x(\omega)) = \rho_{\mathcal{X}}(s(v)) + o(1), \qquad \mathbf{P}(v \in B_y(\omega)) = \rho_{\mathcal{Y}}(s(v)) + o(1), \tag{8}$$
$$\mathbf{P}(\Delta_x > \ln n) = O(n^{-1}), \qquad \mathbf{P}(\Delta_y > \ln n) = O(n^{-1}). \tag{9}$$

Let us show (7) for $\omega(n)$ satisfying $\omega(n) \le \ln n$. Introduce the events $\mathcal{A}_x(v) = \{|X_\omega(v)| \ge \omega(n)\}$ and $\mathcal{A}_y(v) = \{|Y_\omega(v)| \ge \omega(n)\}$. For $|X_\omega(v)| \ge \omega(n) \Leftrightarrow v \in B_x(\omega)$ and $|Y_\omega(v)| \ge \omega(n) \Leftrightarrow v \in B_y(\omega)$ we have $\mathbf{P}(v \in B(\omega)) = \mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A}_y(v))$. In view of (9) with a high probability each of the events $\mathcal{A}_x(v)$ and $\mathcal{A}_y(v)$ refer to at most $\omega(n) + \ln n \le 2\ln n$ vertices. Therefore, we may expect that for large n these events are almost independent and we have (see (8))

$$\mathbf{P}\big(v \in B(\omega)\big) = \mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A}_y(v)) = \rho_{\mathcal{Y}}(s(v))\rho_{\mathcal{X}}(s(v)) + o(1). \tag{10}$$

We show that (10) holds uniformly in $v \in V$. Then (7) follows from (1), (2) and (10) via the identities

$$\mathbf{E}|B(\omega)| = \mathbf{E}\sum_{v\in V} \mathbb{I}_{\{v \in B(\omega)\}} = \sum_{v \in V} \mathbf{P}\big(v \in B(\omega)\big). \tag{11}$$

Let us prove (10). Let $\mathcal{A} = \{X_\omega(v) \cap Y_\omega(v) = v\}$ denote the event that two exploration processes after starting at v do not meet each other in the first $\omega(n)$ steps of the exploration. Observe, that uniformly in $v \in V$

$$\mathbf{P}(\mathcal{A}) = 1 - o(1) \qquad \text{as} \qquad n \to \infty. \tag{12}$$

Indeed, assume that the set $X_\omega(v)$ is already constructed and now we construct the set $Y_\omega(v)$. Note that conditionally, given $X_\omega(v)$ being of size at most $\omega(n) + \ln n$, each black vertex of $Y_\omega(v)$ discovers at least one b-child in $X_\omega(v)$ with probability at most $|X_\omega(v)|p_* n^{-1} \le (\omega(n) + \ln n)p_* n^{-1}$, where $p_* = \max_{i \le i,j \le k} p_{ij}$. Since there are at most $\omega(n)$ black vertices in $Y_\omega(v)$, we conclude that on the event $\mathcal{D}_x = \{|X_\omega(v)| \le \omega(n) + \ln n\}$ the conditional probability

$$\mathbf{P}\big(\overline{\mathcal{A}} \,\big|\, X_\omega(v)\big) \le \omega(n)(\omega(n) + \ln n)p^* n^{-1}. \tag{13}$$

Here $\overline{\mathcal{A}}$ denotes the complementary event to $\mathcal{A}$. It follows from (13) that $\mathbf{P}(\overline{\mathcal{A}} \cap \mathcal{D}_x) = o(1)$. The latter bound combined with (9) shows (12).
Now we are ready to show (10). Assume again that $X_\omega(v)$ has already been constructed. Now we have to construct the set $Y_\omega(v)$. The vertices of $V' = V \setminus X_\omega(v)$ remain blue. In particular for every i the set V' contains at least $n_i - (\omega(n) + \ln n) = n_i(1 - o(1))$ blue vertices of type s_i. Conditionally on the event $\mathcal{A} \cap \mathcal{D}_x$, the exploration of $Y(v)$ (until we stop it after at most $\omega(n)$ steps) stays within the set V' of size $n(1 - o(1))$. The second identity of (8) applies to the conditional probability $\mathbf{P}\big(\mathcal{A}_y(v) \,\big|\, X_\omega(v), \mathcal{A}\big)$ and yields

$$\mathbf{P}\big(\mathcal{A}_y(v) \,\big|\, X_\omega(v), \mathcal{A}\big) = \rho_{\mathcal{Y}}(s(v)) + o(1)$$

uniformly in $X_\omega(v)$, satisfying the event $\mathcal{D}_x$. Therefore, we have

$$\begin{aligned}\mathbf{P}\big(\mathcal{A}_y(v) \cap \mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x\big) &= \mathbf{P}\big(\mathcal{A}_y(v) \,\big|\, \mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x\big)\mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x) \\ &= \rho_{\mathcal{Y}}(s(v))\,\mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x) + o(1).\end{aligned}$$

In view of (9), (12) we can replace $\mathbf{P}\big(\mathcal{A}_y(v) \cap \mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x\big)$ by $\mathbf{P}(\mathcal{A}_y(v) \cap \mathcal{A}_x(v))$ and $\mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x)$ by $\mathbf{P}(\mathcal{A}_x(v))$. Therefore, we obtain $\mathbf{P}(\mathcal{A}_x(v) \cap \mathcal{A}_y(v)) = \rho_{\mathcal{Y}}(s(v))\mathbf{P}(\mathcal{A}_x(v)) +$

$o(1)$. Finally, invoking (8) we obtain(10). Thus we have proved (7) for $\omega(n)$ satisfying the extra condition $\omega(n) \le \ln n$.
Let us now prove (7) for arbitrary ω (satisfying $\omega(n) \to \infty$ and $\omega(n) = o(n)$ as $n \to \infty$). Fix such an ω. We apply (7) to $\omega'(n) = \omega(n) \wedge \ln(n)$ and invoking the inequality $|B(\omega)| \le |B(\omega')|$, we obtain the upper bound $\mathbf{E}|B(\omega)| \le n\rho_{\mathcal{X}\mathcal{Y}} + o(n)$. The corresponding lower bound

$$\mathbf{E}|B(\omega)| \ge n\rho_{\mathcal{X}\mathcal{Y}} + o(n). \tag{14}$$

follows from (11) and the inequalities, which hold uniformly in $v \in V$,

$$\mathbf{P}(v \in B(\omega)) \ge \rho_{\mathcal{X}}(s(v))\rho_{\mathcal{Y}}(s(v)) - o(1). \tag{15}$$

To show that (15) holds, we first perform a forward exploration starting at v and obtain the set $X_\omega(v)$. Afterwards, in the digraph induced by the vertex set $V^0 := \big(V \setminus X_\omega(v)\big) \cup \{v\}$ we perform a backward exploration starting at v. The set of the thus discovered vertices is denoted by $Y^0_\omega(v)$. For a set V^0 containing at least $n_i - (\omega(n) + \ln n) = n_i(1 - o(1))$ vertices of every type $s_i \in S$, the approximation of the distribution of the backward exploration process by the distribution of the Galton-Watson process $\mathcal{Y}(v)$ remains valid. That is, the second identity of (8) applies to the conditional probability $\mathbf{P}\big(|Y^0_\omega(v)| \ge \omega(n) \,\big|\, X_\omega(v)\big)$. We have

$$\mathbf{P}\big(|Y^0_\omega(v)| \ge \omega(n) \,\big|\, X_\omega(v)\big) = \rho_{\mathcal{Y}}(s(v)) + o(1), \tag{16}$$

uniformly v and in $X_\omega(v)$ satisfying $|X_\omega(v)| \le \omega(n) + \ln n$. Since $Y(v)$ contains $Y^0_\omega(v)$ as a subset, we obtain

$$\mathbf{P}\big(|Y(v)| \ge \omega(n) \,\big|\, X_\omega(v)\big) \ge \rho_{\mathcal{Y}}(s(v)) + o(1).$$

The latter inequality in combination with the first identity of (8) and the first bound of (9) shows (15). We now have arrived at (14), thus completing the proof of (7).
Proof of (6). We prove (6) for a particular function $\omega_0(n) = \ln n$. Note that (6) follows from (7) and the bound

$$\mathbf{Var}|B(\omega_0)| = o(n^2) \tag{17}$$

via Chebyshev's inequality. In addition, (17) follows from the identities

$$\mathbf{Var}|B(\omega_0)| = \mathbf{E}|B(\omega_0)|^2 - (\mathbf{E}|B(\omega_0)|)^2,$$
$$\mathbf{E}|B(\omega_0)|(|B(\omega_0)| - 1) = 2\mathbf{E}\sum_{\{u,v\}\subset V} \mathbb{I}_{\{v\in B(\omega_0)\}}\mathbb{I}_{\{u\in B(\omega_0)\}}$$

combined with the identity, which holds uniformly in $\{u, v\} \subset V$,

$$\mathbf{E}\mathbb{I}_{\{v\in B(\omega_0)\}}\mathbb{I}_{\{u\in B(\omega_0)\}} = \rho_{\mathcal{X}}(s(v))\rho_{\mathcal{Y}}(s(v))\rho_{\mathcal{X}}(s(u))\rho_{\mathcal{Y}}(s(u)) + o(1).$$

The latter asymptotic identity is shown in much the same way as (10) above.
Proof of (3). Write $N_1 = N_1(\mathcal{D}_{\mathbb{P},\overline{n}})$. Note that for every $\omega(n)$ we have $N_1 \le \omega(n) \vee |B(\omega)|$. In combination with (6) this inequality implies the upper bound

$$N_1 n^{-1} \le \rho_{\mathcal{X}\mathcal{Y}} + o_P(1). \tag{18}$$

Here and below for a sequence of random variables $\{Z_n\}$ we write $Z_n \le o_P(1)$ (respectively $Z_n \ge o_P(1)$) if for every $\delta > 0$ we have $\lim_n \mathbf{P}(Z_n \le \delta) = 1$ (respectively $\lim_n \mathbf{P}(Z_n \ge -\delta) = 1$).

For $\rho_{\mathcal{X}\mathcal{Y}} = 0$ the result (3) follows from (18). For $\rho_{\mathcal{X}\mathcal{Y}} > 0$ the result (3) follows from (18) and the lower bound

$$N_1 n^{-1} \geq \rho_{\mathcal{X}\mathcal{Y}} + o_P(1). \tag{19}$$

In order to show this lower bound we generate the digraph $\mathcal{D}$ in two steps. Firstly we generate a digraph $\mathcal{D}' = \mathcal{D}_{||p'_{ij}||,\overline{n}}$ and then, on the top of it, we generate another digraph $\mathcal{D}'' = \mathcal{D}_{||p''_{ij}||,\overline{n}}$ independently of $\mathcal{D}'$. Here the numbers p'_{ij} and p''_{ij} are defined by the equations

$$p'_{ij} = p_{ij}(1-\varepsilon), \qquad (1 - p''_{ij}n^{-1})(1 - p'_{ij}n^{-1}) = 1 - p_{ij}n^{-1}, \qquad 1 \leq i, j \leq k. \tag{20}$$

Here $0 < \varepsilon < 1$ is fixed and we assume that n is so large that all $p_{ij}n^{-1} < 1$. The union $\mathcal{D}' \cup \mathcal{D}''$ is the digraph on the vertex set V such that, for each $(u,v) \in [V]^2$, we have $\{v \to u\} \in \mathcal{D}' \cup \mathcal{D}''$ whenever $\{v \to u\}$ is present in at least one of the digraphs $\mathcal{D}'$ and $\mathcal{D}''$. Note that, by the second equation of (20), the random digraphs $\mathcal{D}' \cup \mathcal{D}''$ and $\mathcal{D}$ have the same probability distribution.
Let $\mathcal{X}' = \{\mathcal{X}'(s),\, s \in S\}$ and $\mathcal{Y}' = \{\mathcal{Y}'(s),\, s \in S\}$ be the multi-type Galton-Watson processes with Poisson offspring distributions that approximate the forward and backward explorations of neighbourhoods of vertices in $\mathcal{D}'$. They are defined in the same way as $\mathcal{X}$ and $\mathcal{Y}$ above, but with respect to the matrix $||p'_{ij}||$. Let $\rho_{[\varepsilon]} = \rho_{\mathcal{X}'\mathcal{Y}'}$ be defined by (2). One can show (e.g., by coupling of $\mathcal{X}(s)$ with $\mathcal{X}'(s)$ and $\mathcal{Y}(s)$ with $\mathcal{Y}'(s)$) that $\rho'_{\mathcal{X}}(s) \to \rho_{\mathcal{X}}(s)$ and $\rho'_{\mathcal{Y}}(s) \to \rho_{\mathcal{Y}}(s)$ as $\varepsilon \downarrow 0$. In particular, we have

$$\lim_{\varepsilon \to 0} \rho_{[\varepsilon]} = \rho_{\mathcal{X}\mathcal{Y}}. \tag{21}$$

We are now ready to prove (19). Fix the function $\omega_1(n) = n/\ln n$. Given $v \in V$, let $X'_{\omega_1}(v)$ and $Y'_{\omega_1}(v)$ denote the neighbourhoods of v discovered by the forward and backward explorations performed in $\mathcal{D}'$. Let $B'(\omega_1)$ denote the set of large vertices of $\mathcal{D}'$. From (6) we obtain

$$n^{-1}|B'(\omega_1)| - \rho_{[\varepsilon]} = o_P(1). \tag{22}$$

We shall show that with high probability every pair of vertices from $B'(\omega_1)$ communicate in $\mathcal{D}$. This will imply that N_1 is at least as large as $|B'(\omega_1)|$, and, as a consequence, we then obtain the lower bound (26).
We generate the digraph $\mathcal{D}''$ in k steps so that $\mathcal{D}'' = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_k$. Here $\mathcal{D}_i$, $1 \leq i \leq k$ are independent copies of $\mathcal{D}_{||p^*_{ij}||,\overline{n}}$, where the matrix $||p^*_{ij}||$ is defined by the equations $(1 - p^*_{ij}n^{-1})^k = 1 - p''_{ij}n^{-1}$, $1 \leq i, j \leq k$. Note that for $\{s_i \to s_j\} \in D_P$ identities (20) imply

$$p^*_{ij} \geq k^{-1}\varepsilon p_{ij} \geq p_0, \qquad p_0 := k^{-1}\varepsilon \min\{p_{ij} : p_{ij} > 0\} > 0. \tag{23}$$

For $i = 1, \ldots, k$ denote $X^i(v) = \{u \in V : \{v \rightsquigarrow u\} \in \mathcal{D}' \cup \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_i\}$ and write $V(s_i) = \{v \in V : s(v) = s_i\}$. We shall show that with high probability every set $X^{k-1}(v)$, $v \in B'(\omega_1)$, contains at least $\Theta(n(\ln n)^{-1})$ vertices of each type. More precisely, the event

$$\mathcal{H} = \cap_{v \in B'(\omega_1)} \cap_{1 \leq i \leq k} \big\{|X^{k-1}(v) \cap V(s_i)| \geq n(\ln n)^{-1}\varkappa\big\},$$

where $\varkappa := k^{-1}(p_0/4)^{k-1}q_1 \times \cdots \times q_k$, has probability

$$\mathbf{P}\big(\mathcal{H}\big) = 1 - o(1) \qquad \text{as} \qquad n \to \infty. \tag{24}$$

Observe, that the event $\mathcal{H}$ depends on the random graph $\mathcal{D}^\star = \mathcal{D}' \cup \mathcal{D}_1 \cdots \cup \mathcal{D}_{k-1}$ and is independent of $\mathcal{D}_k$. We show that on the event $\mathcal{H}$ the conditional probability, which now refers to $\mathcal{D}_k$, satisfies

$$\mathbf{P}\Big(N_1 \geq |B'(\omega_1)| \,\Big|\, \mathcal{D}^\star\Big) = 1 - o(1). \tag{25}$$

This bound in combination with (24) and (22) implies the lower bound

$$N_1 n^{-1} \geq \rho_{[\varepsilon]} + o_P(1). \tag{26}$$

Letting $\varepsilon \downarrow 0$ from (21) we obtain (19). We complete the proof of (3) by first showing (25) and then (24).

Proof of (25). Given $\mathcal{D}^\star$, define the events

$$\mathcal{A}_{uv} = \big\{\exists \;\; x \in X^{k-1}(u), \;\; \exists \;\; y \in Y'_{\omega_1}(v) \;\; \text{such that} \;\; \{x \to y\} \in \mathcal{D}_k\big\}, \qquad u, v \in B'(\omega_1).$$

Let $\overline{A}_{uv}$ denote the event complement to $\mathcal{A}_{uv}$. Introduce the sum

$$S = \sum_{\{u,v\} \subset B'(\omega_1)} \mathbb{I}_{\overline{\mathcal{A}}_{uv}}$$

which (given $\mathcal{D}^\star$) is at least as large as the number of pairs $\{u, v\} \in B'(\omega_1)$ that do not communicate in $\mathcal{D}$. We claim that on the event $\mathcal{H}$, we have $S = 0$ with high (conditional given $\mathcal{D}^\star$) probability. Indeed, the largest of the sets $Y'_{\omega_1}(v) \cap V(s_i)$, $1 \leq i \leq k$ is of size at least $\omega_1(n)/k$. Assume it is the r-th set $Y_r := Y'_{\omega_1}(v) \cap V(s_r)$. Since D_P is strongly connected, $\{s_j \to s_r\} \in D_P$ for some s_j. Given the event $\mathcal{H}$, the set $X_j := X^{k-1}(v) \cap V(s_j)$ is of size at least $\Theta(n(\ln n)^{-1})$. Therefore, we have

$$\mathbf{P}\big(\overline{\mathcal{A}}_{uv} \,\big|\, \mathcal{D}^\star\big) \leq (1 - p^*_{jr} n^{-1})^{|X_j| \times |Y_r|} \leq (1 - p_0 n^{-1})^{|X_j| \times |Y_r|} \leq c' n^{-4}. \tag{27}$$

Here $c' > 0$ denotes a constant depending only on $||p_{ij}||$ and ε. It follows from (27) by Chebyshev's inequality that given the event $\mathcal{H}$ we have

$$1 - \mathbf{P}(S = 0 | \mathcal{D}^\star) = \mathbf{P}(S \geq 1 | \mathcal{D}^\star) \leq \mathbf{E}(S | \mathcal{D}^\star) \leq c' n^{-2}.$$

Since $N_1 \geq |B'(\omega_1)| - S$, the latter bound implies (25).

Proof of (24). Let $\mathbf{P}'(\cdot) = \mathbf{P}(\cdot | \mathcal{D}')$ denote the conditional probability given $\mathcal{D}'$. In order to prove (24), we show that, for each $v \in B'(\omega_1)$ and $1 \leq i \leq k$,

$$\mathbf{P}'\big(|X^{k-1}(v) \cap V(s_i)| < n(\ln n)^{-1} \varkappa\big) = O(n^{-4}). \tag{28}$$

Fix $v \in B'(\omega_1)$ and i. The largest of sets $X'_{\omega_1}(v) \cap V(s_j)$, $1 \leq j \leq k$ is of size at least $\omega_1(n)/k$. Assume it is the r-th set $X^0 := X'_{\omega_1}(v) \cap V(s_r)$. Since D_P is strongly connected we find a shortest path

$$s_r = t_0 \to t_1 \to \cdots \to t_j = s_i \tag{29}$$

in D_P. Denote $\varkappa_m = k^{-1}(p_0/4)^m Q(t_1) \cdots Q(t_m)$. We claim that, for $1 \leq m \leq j$,

$$\mathbf{P}'\big(|X^m(v) \cap V(t_m)| < n(\ln n)^{-1} \varkappa_m\big) = O(n^{-4}). \tag{30}$$

Note that, in view of the inclusions $X'_{\omega_1}(v) = X^0(v) \subset X^1(v) \subset \cdots \subset X^{k-1}(v)$ and the inequalities $\varkappa \leq \varkappa_i$, the bound (30) (for $m = j$) implies (28). We show (30), for $m = 1, 2, \ldots$. Let $m = 1$. Let X^1 denote the set of endpoints in $V(t_1)$ of arcs in $\mathcal{D}_1$ that start at vertices in X^0. We have

$$|X^1(v) \cap V(t_1)| \geq |X^1| = \sum_{w \in V(t_1)} \mathbb{I}_{\{w \in X^1\}}$$

The right-hand sum has binomial distribution $Bi(|V(t_1)|, p_1)$ with $|V(t_1)| = nQ(t_1) + o(n)$ trials and success probability

$$p_1 \geq 1 - (1 - p_0 n^{-1})^{|X^0|} \geq 1 - (1 - p_0 n^{-1})^{\omega_1(n)/k}.$$

Indeed, $p_0 n^{-1}$ is the smallest probability of arcs in $\mathcal{D}_1$, and there are at least $|X^0| \geq \omega_1(n)/k$ vertices in $X^0(v)$ that "try" to send an arc to a given vertex $w \in V(t_1)$. A simple analysis show that $p_1 = \Theta(ln^{-1}(n))$. In particular, for large n we have $p_1 \geq p_0/(2k \ln n)$. Therefore, we obtain $\mathbf{E}|X^1| \geq 2\varkappa_1 n(\ln n)^{-1}(1 + o(1))$. Now, an application of Chernoff bound (see, e.g., Janson, Łuczak and Ruciński [6]) to the binomial random variable $|X^1|$ shows (30) for $m = 1$. Next, given the event $\{|X^1(v) \cap V(t_1)| \geq n(\ln n)^{-1}\varkappa_1\}$, which is independent of $\mathcal{D}_2$, we show

$$\mathbf{P}\Big(|X^2(v) \cap V(t_2)| < n(\ln n)^{-1}\varkappa_2 \,\Big|\, \mathcal{D}', \mathcal{D}_1\Big) = O(n^{-4}).$$

in much the same way. That is, we show that number of different endpoints in $V(t_2)$ of arcs in $\mathcal{D}_2$ that start in $X^1(v) \cap V(t_1)$ is strongly concentrated around $2\varkappa_2 n(\ln n)^{-1}$. In this way, proceeding along the path (29) until the last endpoint s_j and using arcs from $\mathcal{D}_1$, $\mathcal{D}_2$ etc. for the successive steps respectively, we arrive at (30).
In the very last step of the proof we show $N_2 = o_P(n)$. Indeed, this bound follows immediately from (6), (19) via the simple inequality $N_1 + N_2 \leq 2\omega(n) + |B(\omega)|$. □

Acknowledgement. Mindaugas Bloznelis would like to thank Colin Cooper for valuable and inspiring discussion.

References

[1] R. Albert, A.-L.Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (2002), 47–97

[2] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogenious random graphs, Random Structures Algorithms 31 (2007), 3–122

[3] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, Giant strongly connected component of directed networks, Phys. Rev. E 64 (2001), 025101(R)

[4] R. Durrett, Random Graph Dynamics, Cambridge University Press, New York, 2007

[5] C. Cooper, A. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence, Comb. Probab. Comput. 13 (2004) 319–337

[6] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, Wiley, New York, 2001

[7] R. M. Karp, The transitive closure of a random digraph, Random Structures Algorithms 1 (1990), 73–93

[8] T. Łuczak, The phase transition in the evolution of random digraphs, J. Graph Theory 14 (1990), 217–223.

[9] T. Łuczak, J. E. Cohen, Giant components in three-parameter random directed graphs, Adv. Appl. Probab. 24 (1992), 845–857

[10] T. Łuczak, T. G. Seierstad, The critical behavior of random digraphs, Random Structures Algorithms 35 (2009), 271–293

[11] M. E. J. Newman, S. H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications Phys. Rev. E 64 (2001), 026118

[12] B. Söderberg, General formalism for inhomogeneous random graphs, Phys. Rev. E 66 (2002), 066121