

BIRTH OF A STRONGLY CONNECTED GIANT IN AN INHOMOGENEOUS RANDOM DIGRAPH

MINDAUGAS BLOZNELIS,* *Vilnius University*

FRIEDRICH GÖTZE,** *Bielefeld University*

JERZY JAWORSKI,*** *Adam Mickiewicz University*

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 on its orientation. For this model, the critical point for the emergence of a giant component is determined via a branching process approach.

Keywords: Inhomogeneous digraph; phase transition; giant component

2010 Mathematics Subject Classification: Primary 05C80

Secondary 90B15; 60J85

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 indegree/outdegree distribution, which are inconsistent with the model $D(n, p)$. Generally, real-life networks are inhomogeneous (see, e.g. [1], [4], [5], and [12]).

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 indegree and outdegree distributions, including power-law distributions.

The main problem considered 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 address this problem, we establish a first-order asymptotic for the number N_1 of vertices in the largest strongly connected component, by showing that $N_1 = \rho n + o_P(n)$.

Received 30 November 2009; revision received 18 November 2011.

* Postal address: Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania.

Email address: mindaugas.bloznelis@mif.vu.lt

** Postal address: Faculty of Mathematics, Bielefeld University, D-33501 Bielefeld, Germany.

Email address: goetze@math.uni-bielefeld.de

*** Postal address: Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland. Email address: jaworski@amu.edu.pl

The fraction $\rho \geq 0$ is expressed in terms of survival probabilities of related branching processes that reflect the statistical properties of a certain neighborhood of a randomly chosen vertex.

Let us mention that mathematically rigorous results on the phase transition in $D(n, p)$ were established in [8] and [9]; see also [11]. More general three-parameter models were studied in [10]. The phase transition in a random digraph with given (nonrandom) indegree and outdegree sequences was shown in [3]. The phase transition behavior of general inhomogeneous random graphs is well understood due to the analysis in the seminal paper by Bollobás *et al.* [2]. This paper has inspired our work on inhomogeneous digraphs. In particular, we adopt the approach developed in [2].

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. The 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.

A digraph D on the vertex set $V = \{v_1, v_2, \dots, v_n\}$ 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 by $\{u \rightarrow v\} \in D$ or just $u \rightarrow v$. 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 starting at w_1 and ending at w_k . The fact that such a path is present in D is denoted by $\{w_1 \rightsquigarrow w_k\} \in D$ or just $w_1 \rightsquigarrow w_k$. If $v \rightsquigarrow w$ and $w \rightsquigarrow v$, then the vertices v and w are said to *communicate*. In this case we write $v \leftrightarrow w$. In addition, we define $v \leftrightarrow v$ 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 ‘ \leftrightarrow ’ is an equivalence relation, it splits the vertex set V into a union of disjoint subsets. 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 characteristic 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 characteristic is the size $N_2 = N_2(D)$ of the second largest SC component.

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 as the number of vertices increases. A similar model of *random graphs* (but not digraphs) has been introduced by Söderberg [13]. Let us introduce some more notation. Let $S = \{s_1, s_2, \dots, s_k\}$ denote the set of types, and let $s(v)$ denote the type of 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 an integer vector $\vec{n} = (n_1, \dots, n_k)$ and a $k \times k$ matrix $\mathbb{P} = \|p_{ij}\|$ with nonnegative 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 the events $\{u \rightarrow v\} \in \mathcal{D}$ are independent and have probabilities $P(u \rightarrow v) = 1 \wedge (p_{ij}n^{-1})$ 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 \bar{n} , we sometimes write $\mathcal{D} = \mathcal{D}_{\mathbb{P}, \bar{n}}$.

We will 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

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

We describe the phase transition in terms of Q and \mathbb{P} using the language of branching processes. Let us consider multitype Galton–Watson processes where particles are of types from S . Given $s \in S$, let $\mathcal{X}(s)$ and $\mathcal{Y}(s)$ denote the Galton–Watson processes 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 means $p_{ij}q_j$ and $p_{ji}q_j$, respectively, $1 \leq i, j \leq k$. 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 nonextinction probabilities of $\mathcal{X}(s)$ and $\mathcal{Y}(s)$, respectively. Write

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

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 and that the matrix $\mathbb{P} = \|p_{ij}\|$ is irreducible. As $n \rightarrow \infty$, we have*

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

and $N_2 = o_{\mathbb{P}}(n)$.

Here, for a sequence of random variables $\{Z_n\}$, we write $Z_n = o_{\mathbb{P}}(n)$ if $\lim_n \mathbb{P}(|Z_n| > \delta n) = 0$ for each $\delta > 0$.

Recall that the matrix $\|p_{ij}\|$ is called *irreducible* if its associated digraph $D_{\mathbb{P}}$ is SC. Here $D_{\mathbb{P}}$ is the digraph on the vertex set S such that $\{s_i \rightarrow s_j\} \in D_{\mathbb{P}}$ whenever $p_{ij} > 0$.

We note that the sum in (2) is positive if and only if at least one of $\mathcal{X}(s)$ or $\mathcal{Y}(s)$ survives with a positive probability for some $s \in S$. This fact follows from the next result, which we prove in Section 4.

Result 1. *In the case where $q_i > 0$, $1 \leq i \leq k$, and the matrix \mathbb{P} is irreducible, the following statements are equivalent:*

- (i) *for all $s \in S$, we have $\rho_{\mathcal{X}}(s) > 0$;*
- (ii) *there exists $s \in S$ with $\rho_{\mathcal{X}}(s) > 0$;*
- (iii) *there exists $s \in S$ with $\rho_{\mathcal{Y}}(s) > 0$;*
- (iv) *for all $s \in S$, we have $\rho_{\mathcal{Y}}(s) > 0$.*

That is, we have the equivalences (i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Leftrightarrow (iv).

Remark 1. In the case where $D_{\mathbb{P}}$ is not SC, Theorem 1 can be applied to its SC components. Let D_1, \dots, D_r denote the SC components of $D_{\mathbb{P}}$. Given $1 \leq m \leq r$, 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 $i \neq j$, any two vertices $v \in V_i$ and $u \in V_j$ do not communicate in \mathcal{D} . Therefore, each SC component of \mathcal{D} is a subgraph

of some \mathcal{D}_m . The asymptotic size of the largest SC component of \mathcal{D}_m is, by Theorem 1, $N_1(\mathcal{D}_m) = \rho_m n + o_P(n)$. Here we define $\rho_m = \sum_{s \in S_m} \rho_X(s) \rho_Y(s) Q(s)$. It follows now that $N_1(\mathcal{D}) = \max_{1 \leq m \leq r} \rho_m n + o_P(n)$. In this way we obtain an extension of (3) to general (not necessarily SC) digraphs $\mathcal{D}_{\mathbb{P}}$.

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, and Q is a probability measure defined on Borel sets of S . Here the matrix $\|p_{ij}\|$ is replaced by a nonnegative $S \times S$ measurable kernel $\kappa(s, t)$, $s, t \in S$.

Let x_1, x_2, \dots be a sequence of random variables with values in S such that the empirical distribution of the first n observations x_1, \dots, x_n approximates the measure Q in probability as $n \rightarrow \infty$. 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 $n \rightarrow \infty$. 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 $\{x\}_1^n = \{x_1, \dots, x_n\}$ with independent arcs having probabilities $P(\{x_i \rightarrow x_j\} \in \mathcal{D}_n) = 1 \wedge (n^{-1} \kappa(x_i, x_j))$, $1 \leq i, j \leq n$. Combining S , Q , and κ we obtain a very large class of inhomogeneous digraphs with independent arcs. Obviously, the model will include digraphs with indegree and outdegree distributions which have power laws.

Such a general model, for *random graphs* (not digraphs), was introduced in [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 , $n \geq 2$, 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 multitype Galton–Watson branching processes with type space S . Given $s \in S$, let $\mathcal{X}(s)$ and $\mathcal{Y}(s)$ denote the Galton–Watson processes 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 means $\int_A \kappa(t, u) Q(du)$ and $\int_A \kappa(u, t) Q(du)$, respectively. 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

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

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

$$N_1(\mathcal{D}_n) = \rho_{\mathcal{X}\mathcal{Y}} n + o_P(n) \quad \text{as } n \rightarrow \infty. \tag{4}$$

In order to establish (4), 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. That is, for any measurable $A \subset S$ with $Q(A) \neq 1$ or $Q(A) \neq 0$, the identity $Q \times Q(\{(s, t) \in A \times (S \setminus A): \kappa(s, t) \neq 0\}) = 0$ implies that $Q(A) = 0$ 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

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

as $n \rightarrow \infty$. 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 (4). It can be obtained from Theorem 1 via a finite-dimensional approximation argument similar to that used in the proof of Theorem 3.1 of [2].

4. Proofs

In the proof we will use ideas and techniques developed in [2] and [8].

4.1. 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 directed paths, and let $Y(v)$ denote the set of starting points of directed paths ending at v :

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

Given a function $\omega(n)$ such that $\omega(n) \rightarrow \infty$ and $\omega(n) = o(n)$ as $n \rightarrow \infty$, we say that $v \in V$ is x -big if $|X(v)| \geq \omega(n)$ and that $v \in V$ is y -big if $|Y(v)| \geq \omega(n)$. The sets of x -big and y -big vertices are respectively denoted by $B_x = B_x(\omega)$ and $B_y = B_y(\omega)$. 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 ,

$$n^{-1}|B(\omega)| - \rho_{xy} = o_P(1) \quad \text{as } n \rightarrow \infty. \tag{5}$$

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

$$n^{-1} E|B(\omega)| = \rho_{xy} + o(1), \tag{6}$$

and to establish (5) for at least one such function, say $\omega_0(n) = \ln n$. Indeed, assume that (6) holds. Let ω and ω' be two such functions, and let B and B' denote the corresponding sets of large vertices. As the size of the symmetric difference $B \Delta B' = (B \cup B') \setminus (B \cap B')$ equals $|B(\omega \wedge \omega')| - |B(\omega \vee \omega')| \geq 0$, we obtain, from (6), $E|B \Delta B'| = o(n)$. It follows that $||B| - |B'|| \leq |B \Delta B'| = o_P(n)$. In particular, (5) holds for every ω whenever it is satisfied by at least one such function ω .

Proof of (6). 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. We stop when we have collected at least $\omega(n)$ vertices in the list (hence, $v \in B_x(\omega)$), or when there are no white vertices left in the list (hence, we have explored the 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 $\{w \rightarrow u\} \in \mathcal{D}$ and u was blue when w discovered it during the exploration process. Since the last black vertex adds to the list all its blue neighbors (endpoints of outgoing arcs), the size $|X_\omega(v)|$ cannot exceed $\omega(n)$ by more than the outdegree 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 the direction of arcs ($\{v \rightarrow u\} \in \mathcal{D} \Leftrightarrow \{u \rightarrow v\} \in \mathcal{D}^*$). That is, now the search for neighbors 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 $\{u \rightarrow w\} \in \mathcal{D}$ 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 outdegree in \mathcal{D}^* (indegree in \mathcal{D}) of the vertex last explored.

By a coupling of an exploration process with the approximating Galton–Watson process, Bollobás *et al.* [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 of [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 $n \rightarrow \infty$ uniformly in $v \in V$,

$$\begin{aligned} P(v \in B_x(\omega)) &= \rho_{\mathcal{X}}(s(v)) + o(1), & P(v \in B_y(\omega)) &= \rho_{\mathcal{Y}}(s(v)) + o(1), & (7) \\ P(\Delta_x > \ln n) &= O(n^{-1}), & P(\Delta_y > \ln n) &= O(n^{-1}). & (8) \end{aligned}$$

Let us show (6) for $\omega(n)$ satisfying $\omega(n) \leq \ln n$. Introduce the events $\mathcal{A}_x(v) = \{|X_\omega(v)| \geq \omega(n)\}$ and $\mathcal{A}_y(v) = \{|Y_\omega(v)| \geq \omega(n)\}$. Since $|X_\omega(v)| \geq \omega(n) \Leftrightarrow v \in B_x(\omega)$ and $|Y_\omega(v)| \geq \omega(n) \Leftrightarrow v \in B_y(\omega)$, we have $P(v \in B(\omega)) = P(\mathcal{A}_x(v) \cap \mathcal{A}_y(v))$. In view of (8), with a high probability, each of the events $\mathcal{A}_x(v)$ and $\mathcal{A}_y(v)$ refer to at most $\omega(n) + \ln n \leq 2 \ln n$ vertices. Therefore, we may expect that, for large n , these events are almost independent and we have (see (7))

$$P(v \in B(\omega)) = P(\mathcal{A}_x(v) \cap \mathcal{A}_y(v)) = \rho_{\mathcal{X}}(s(v))\rho_{\mathcal{Y}}(s(v)) + o(1). \tag{9}$$

We show that (9) holds uniformly in $v \in V$. Then (6) follows from (1), (2), and (9) via the identities

$$E|B(\omega)| = E \sum_{v \in V} \mathbf{1}_{\{v \in B(\omega)\}} = \sum_{v \in V} P(v \in B(\omega)). \tag{10}$$

Let us prove (9). 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 $v \in V$,

$$P(\mathcal{A}) = 1 - o(1) \quad \text{as } n \rightarrow \infty. \tag{11}$$

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

$$P(\overline{\mathcal{A}} \mid X_\omega(v)) \leq \omega(n)(\omega(n) + \ln n)p_*n^{-1}. \tag{12}$$

Here $\overline{\mathcal{A}}$ denotes the complementary event to \mathcal{A} . It follows from (12) that $P(\overline{\mathcal{A}} \cap \mathcal{D}_x) = o(1)$. The latter bound combined with (8) shows (11).

Now we are ready to show (9). 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 (7) applies to the conditional probability $P(\mathcal{A}_y(v) \mid X_\omega(v), \mathcal{A})$ and yields

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

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

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

In view of (8) and (11) we can replace $P(\mathcal{A}_y(v) \cap \mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x)$ by $P(\mathcal{A}_y(v) \cap \mathcal{A}_x(v))$ and $P(\mathcal{A}_x(v) \cap \mathcal{A} \cap \mathcal{D}_x)$ by $P(\mathcal{A}_x(v))$. Therefore, we obtain $P(\mathcal{A}_x(v) \cap \mathcal{A}_y(v)) = \rho_y(s(v)) \times P(\mathcal{A}_x(v)) + o(1)$. Finally, invoking (7) we obtain (9). Thus, we have proved (6) for $\omega(n)$ satisfying the extra condition $\omega(n) \leq \ln n$.

Let us now prove (6) for arbitrary ω (satisfying $\omega(n) \rightarrow \infty$ and $\omega(n) = o(n)$ as $n \rightarrow \infty$). Fix such an ω . We apply (6) to $\omega'(n) = \omega(n) \wedge \ln(n)$, and invoking the inequality $|B(\omega)| \leq |B(\omega')|$, we obtain the upper bound $E|B(\omega)| \leq n\rho_x y + o(n)$. The corresponding lower bound,

$$E|B(\omega)| \geq n\rho_x y + o(n), \tag{13}$$

follows from (10) and the inequality

$$P(v \in B(\omega)) \geq \rho_x(s(v))\rho_y(s(v)) - o(1), \tag{14}$$

which holds uniformly in $v \in V$.

To show that (14) holds, we first perform a forward exploration starting at the vertex v and obtain the set $X_\omega(v)$. Afterwards, in the digraph induced by the vertex set $V^0 := (V \setminus X_\omega(v)) \cup \{v\}$ we perform a backward exploration starting at v . The set of the thus discovered vertices is denoted by $Y_\omega^0(v)$. 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 (7) applies to the conditional probability $P(|Y_\omega^0(v)| \geq \omega(n) \mid X_\omega(v))$. We have

$$P(|Y_\omega^0(v)| \geq \omega(n) \mid X_\omega(v)) = \rho_y(s(v)) + o(1) \tag{15}$$

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

$$P(|Y(v)| \geq \omega(n) \mid X_\omega(v)) \geq \rho_y(s(v)) + o(1).$$

The latter inequality in combination with the first identity of (7) and the first bound of (8) shows (14). We have now arrived at (13), thus completing the proof of (6).

Proof of (5). We prove (5) for a particular function $\omega_0(n) = \ln n$. Note that (5) follows from (6) and the bound

$$\text{var } |B(\omega_0)| = o(n^2) \tag{16}$$

via Chebyshev’s inequality. In addition, (16) follows from the identities

$$\begin{aligned} \text{var } |B(\omega_0)| &= E|B(\omega_0)|^2 - (E|B(\omega_0)|)^2, \\ E|B(\omega_0)|(|B(\omega_0)| - 1) &= 2E \sum_{\{u,v\} \subset V} \mathbf{1}_{\{v \in B(\omega_0)\}} \mathbf{1}_{\{u \in B(\omega_0)\}}, \end{aligned}$$

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

$$E \mathbf{1}_{\{v \in B(\omega_0)\}} \mathbf{1}_{\{u \in B(\omega_0)\}} = \rho_x(s(v))\rho_y(s(v))\rho_x(s(u))\rho_y(s(u)) + o(1).$$

The latter asymptotic identity is shown in much the same way as (9) above.

Proof of (3). Write $N_1 = N_1(\mathcal{D}_{\mathbb{P}, \bar{n}})$. Note that, for every $\omega(n)$, $N_1 \leq \omega(n) \vee |B(\omega)|$. In combination with (5) this inequality implies the upper bound

$$N_1 n^{-1} \leq \rho_{\mathcal{X}\mathcal{Y}} + o_{\mathbb{P}}(1). \tag{17}$$

Here and below, for a sequence of random variables $\{Z_n\}$, we write $Z_n \leq o_{\mathbb{P}}(1)$ and $Z_n \geq o_{\mathbb{P}}(1)$ if, for every $\delta > 0$, we respectively have $\lim_n \mathbb{P}(Z_n \leq \delta) = 1$ and $\lim_n \mathbb{P}(Z_n \geq -\delta) = 1$.

For $\rho_{\mathcal{X}\mathcal{Y}} = 0$, result (3) follows from (17). For $\rho_{\mathcal{X}\mathcal{Y}} > 0$, result (3) follows from (17) and the lower bound

$$N_1 n^{-1} \geq \rho_{\mathcal{X}\mathcal{Y}} + o_{\mathbb{P}}(1). \tag{18}$$

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

$$p'_{ij} = p_{ij}(1 - \varepsilon), \quad (1 - p''_{ij} n^{-1})(1 - p'_{ij} n^{-1}) = 1 - p_{ij} n^{-1}, \quad 1 \leq i, j \leq k, \tag{19}$$

where $0 < \varepsilon < 1$ 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 $\{v \rightarrow u\} \in \mathcal{D}' \cup \mathcal{D}''$ whenever $\{v \rightarrow u\}$ is present in at least one of the digraphs \mathcal{D}' and \mathcal{D}'' . Note that, by the second equation of (19), 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 multitype Galton–Watson processes with Poisson offspring distributions that approximate the forward and backward explorations of neighborhoods 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). We can show (e.g. by coupling of $\mathcal{X}(s)$ with $\mathcal{X}'(s)$ and $\mathcal{Y}(s)$ with $\mathcal{Y}'(s)$) that $\rho_{\mathcal{X}'}(s) \rightarrow \rho_{\mathcal{X}}(s)$ and $\rho_{\mathcal{Y}'}(s) \rightarrow \rho_{\mathcal{Y}}(s)$ as $\varepsilon \downarrow 0$. In particular, we have

$$\lim_{\varepsilon \rightarrow 0} \rho_{[\varepsilon]} = \rho_{\mathcal{X}\mathcal{Y}}. \tag{20}$$

We are now ready to prove (18). 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 neighborhoods 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 (5) we obtain

$$n^{-1} |B'(\omega_1)| - \rho_{[\varepsilon]} = o_{\mathbb{P}}(1). \tag{21}$$

We will 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 (25) below.

We generate the digraph \mathcal{D}'' in k steps so that $\mathcal{D}'' = \mathcal{D}_1 \cup \dots \cup \mathcal{D}_k$. Here \mathcal{D}_i , $1 \leq i \leq k$, are independent copies of $\mathcal{D}_{\|p^*_{ij}\|, \bar{n}}$, 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 $\{s_i \rightarrow s_j\} \in D_{\mathbb{P}}$, identities (19) imply that

$$p^*_{ij} \geq k^{-1} \varepsilon p_{ij} \geq p_0, \quad p_0 := k^{-1} \varepsilon \min\{p_{ij} : p_{ij} > 0\} > 0. \tag{22}$$

For $i = 1, \dots, k$, define $X^i(v) = \{u \in V : \{v \rightsquigarrow u\} \in \mathcal{D}' \cup \mathcal{D}_1 \cup \dots \cup \mathcal{D}_i\}$ and write $V(s_i) = \{v \in V : s(v) = s_i\}$. We will 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

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

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

$$P(\mathcal{H}) = 1 - o(1) \quad \text{as } n \rightarrow \infty. \tag{23}$$

Observe, that the event \mathcal{H} depends on the random graph $\mathcal{D}^* = \mathcal{D}' \cup \mathcal{D}_1 \dots \cup \mathcal{D}_{k-1}$ 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

$$P(N_1 \geq |B'(\omega_1)| \mid \mathcal{D}^*) = 1 - o(1). \tag{24}$$

This bound in combination with (21) and (23) implies the lower bound

$$N_1 n^{-1} \geq \rho_{[\varepsilon]} + o_p(1). \tag{25}$$

Letting $\varepsilon \downarrow 0$ from (20) we obtain (18). We complete the proof of (3) by first showing (24) and then (23).

Proof of (24). Given \mathcal{D}^* , define the events

$$\mathcal{A}_{uv} = \{\text{there exists } x \in X^{k-1}(u) \text{ and } y \in Y'_{\omega_1}(v) \text{ such that } \{x \rightarrow y\} \in \mathcal{D}_k\}$$

for $u, v \in B'(\omega_1)$. Let $\bar{\mathcal{A}}_{uv}$ denote the event complement to \mathcal{A}_{uv} . Introduce the sum

$$S = \sum_{\{u,v\} \subset B'(\omega_1)} (\mathbf{1}_{\bar{\mathcal{A}}_{uv}} + \mathbf{1}_{\bar{\mathcal{A}}_{vu}})$$

which (given \mathcal{D}^*) 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}^*) 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 that it is the r th set $Y_r := Y'_{\omega_1}(v) \cap V(s_r)$. Since $D_{\mathbb{P}}$ is SC, $\{s_j \rightarrow s_r\} \in D_{\mathbb{P}}$ 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

$$P(\bar{\mathcal{A}}_{uv} \mid \mathcal{D}^*) \leq (1 - p_{j_r}^* n^{-1})^{|X_j| \times |Y_r|} \leq (1 - p_0 n^{-1})^{|X_j| \times |Y_r|} \leq c' n^{-4}. \tag{26}$$

Here $c' > 0$ denotes a constant depending only on $\|p_{ij}\|$ and ε . It follows from (26), by Chebyshev's inequality, that, given the event \mathcal{H} , we have

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

Since $N_1 \geq |B'(\omega_1)| - S$, the latter bound implies (24).

Proof of (23). Let $P'(\cdot) = P(\cdot \mid \mathcal{D}')$ denote the conditional probability given \mathcal{D}' . In order to prove (23), we show that, for each $v \in B'(\omega_1)$ and $1 \leq i \leq k$,

$$P'(|X^{k-1}(v) \cap V(s_i)| < n(\ln n)^{-1}\kappa) = O(n^{-4}). \tag{27}$$

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

$$s_r = t_0 \rightarrow t_1 \rightarrow \dots \rightarrow t_j = s_i \tag{28}$$

in $D_{\mathbb{P}}$. Define $x_m = k^{-1}(p_0/4)^m Q(t_1) \cdots Q(t_m)$. We claim that, for $1 \leq m \leq j$,

$$P(|X^m(v) \cap V(t_m)| < n(\ln n)^{-1}x_m) = O(n^{-4}). \tag{29}$$

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 $x \leq x_i$, bound (29) (for $m = j$) implies (27). We show (29) for $m = 1, 2, \dots$. 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

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

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

$$p_1 \geq 1 - (1 - p_0n^{-1})^{|X^0|} \geq 1 - (1 - p_0n^{-1})^{\omega_1(n)/k}.$$

Indeed, p_0n^{-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 shows 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 $E|X^1| \geq 2x_1n(\ln n)^{-1}(1 + o(1))$. Now, an application of the Chernoff bound (see, e.g. [7]) to the binomial random variable $|X^1|$ shows (29) for $m = 1$. Next, given the event $\{|X^1(v) \cap V(t_1)| \geq n(\ln n)^{-1}x_1\}$, which is independent of \mathcal{D}_2 , we show that

$$P(|X^2(v) \cap V(t_2)| < n(\ln n)^{-1}x_2 \mid \mathcal{D}', \mathcal{D}_1) = O(n^{-4})$$

in much the same way. That is, we show that the 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 $2x_2n(\ln n)^{-1}$. In this way, proceeding along the path (28) 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 (29).

In the very last step of the proof of Theorem 1 we show that $N_2 = o_{\mathbb{P}}(n)$. Indeed, this bound follows immediately from (5) and (18) via the simple inequality $N_1 + N_2 \leq 2\omega(n) + |B(\omega)|$.

4.2. Proof of Result 1

The implications (i) \Rightarrow (ii) and (iv) \Rightarrow (iii) are obvious. Here we show that (ii) \Rightarrow (i) and (ii) \Rightarrow (iii). The remaining implications (iii) \Rightarrow (iv) and (iii) \Rightarrow (ii) are shown in the same way.

Assume that $s \in S$ and $\rho_{\mathcal{X}}(s) > 0$. We first show that $\rho_{\mathcal{X}}(\bar{s}) > 0$ for each $\bar{s} \in S \setminus \{s\}$. Since $D_{\mathbb{P}}$ is SC, there exists a sequence $\bar{s} = s_0 \rightarrow s_1 \cdots \rightarrow s_r = s$ of types such that $p_{i-1 i} > 0$ for $i = 1, \dots, r$. Therefore, with a positive probability, particles of type s are among descendants of $\mathcal{X}(\bar{s})$. But, each such particle has a positive survival probability $\rho_{\mathcal{X}}(s) > 0$. Hence, $\rho_{\mathcal{X}}(\bar{s}) > 0$.

Now we show (iii). We replace q_1, \dots, q_k by rationals $q_1^* = h_1/h, \dots, q_k^* = h_k/h$ such that $0 < q_i^* \leq q_i, 1 \leq i \leq k$. Here h_i and h are integers. Let \mathcal{X}^* and \mathcal{Y}^* be the Galton–Watson processes defined in the same way as \mathcal{X} and \mathcal{Y} , but with q_1, \dots, q_k replaced by q_1^*, \dots, q_k^* . We choose q_i^* sufficiently close to $q_i, 1 \leq i \leq k$, so that $\rho_{\mathcal{X}^*}(s) > 0$. Let $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ be multitype Galton–Watson processes with type spaces $\tilde{S} = \{s_{11}, \dots, s_{1h_1}, \dots, s_{k1}, \dots, s_{kh_k}\}$, where the number $\tilde{\mathcal{X}}_{s_{jr}}(s_{it})$ of children of type s_{jr} of a particle of type s_{it} has Poisson distribution with means $p_{ij}h^{-1}$ and $p_{ji}h^{-1}$, respectively. We assume that numbers of children of different types

are independent random variables. Note that an observer, who can read only the first index of the type of a particle (e.g. the index i of s_{ij}), does not distinguish between the processes \mathcal{X}^* and $\tilde{\mathcal{X}}$ (respectively between \mathcal{Y}^* and $\tilde{\mathcal{Y}}$). Hence, the observer will tell us that $\rho_{\tilde{\mathcal{X}}}(s_{it}) > 0$ for each $1 \leq t \leq h_i$. Here $s_i = s$. Now we are going to apply the well-known fact from the theory of multitype Galton–Watson processes (see, e.g. Theorem 7.1 of [6]) that the maximal positive eigenvalue $\lambda_{\tilde{\mathcal{X}}}$ of the matrix $M_{\tilde{\mathcal{X}}} = \|E \tilde{X}_{\tilde{s}}(\tilde{u})\|_{\tilde{s}, \tilde{u} \in \tilde{S}}$ is larger than 1 if and only if $\rho_{\tilde{\mathcal{X}}}(\tilde{s}) > 0$ for some $\tilde{s} \in \tilde{S}$. Hence, $\rho_{\tilde{\mathcal{X}}}(s_{it}) > 0 \Rightarrow \lambda_{\tilde{\mathcal{X}}} > 1$. Furthermore, since the matrix $M_{\tilde{\mathcal{Y}}}$ is the transpose of $M_{\tilde{\mathcal{X}}}$, both matrices have the same spectrum. In particular, $\lambda_{\tilde{\mathcal{Y}}} = \lambda_{\tilde{\mathcal{X}}} > 1$. Finally, the abovementioned result implies that $\rho_{\tilde{\mathcal{Y}}}(\tilde{u}) > 0$ for some $\tilde{u} \in \tilde{S}$, and the observer will tell us that $\rho_{\mathcal{Y}^*}(u) > 0$ for some $u \in S$.

Acknowledgements

We thank the anonymous referee for pointing out the problem addressed in Result 1. M. Bloznelis would also like to thank Colin Cooper for valuable and inspiring discussions, and acknowledge the support by SFB 701 at the University of Bielefeld, Germany. J. Jaworski acknowledges support from the Marie Curie Intra-European Fellowship No. 236845 (RAN-DOMAPP) within the 7th EU Framework Programme and by the National Science Centre (DEC-2011/01/B/ST1/03943).

References

- [1] ALBERT, R. AND BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74**, 47–97.
- [2] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31**, 3–122.
- [3] COOPER, C. AND FRIEZE, A. (2004). The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics Prob. Comput.* **13**, 319–337.
- [4] DOROGOVTSSEV, S. N., MENDES, J. F. F. AND SAMUKHIN, A. N. (2001). Giant strongly connected component of directed networks. *Phys. Rev. E* **64**, 025101(R), 4 pp.
- [5] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge University Press.
- [6] HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.
- [7] JANSON, S., ŁUCZAK, T. AND RUCIŃSKI, A. (2001). *Random Graphs*. John Wiley, New York.
- [8] KARP, R. M. (1990). The transitive closure of a random digraph. *Random Structures Algorithms* **1**, 73–93.
- [9] ŁUCZAK, T. (1990). The phase transition in the evolution of random digraphs. *J. Graph Theory* **14**, 217–223.
- [10] ŁUCZAK, T. AND COHEN, J. E. (1992). Giant components in three-parameter random directed graphs. *Adv. Appl. Prob.* **24**, 845–857.
- [11] ŁUCZAK, T. AND SEIERSTAD, T. G. (2009). The critical behavior of random digraphs. *Random Structures Algorithms* **35**, 271–293.
- [12] NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118, 17 pp.
- [13] SÖDERBERG, B. (2002). General formalism for inhomogeneous random graphs. *Phys. Rev. E* **66**, 066121, 6 pp.