

ARE GIANTS IN RANDOM DIGRAPHS ‘ALMOST’ LOCAL?

REMCO VAN DER HOFSTAD AND MANISH PANDEY

ABSTRACT. Recently, the first author showed that the giant in random undirected graphs is ‘almost’ local. This means that, under a necessary and sufficient condition, the limiting proportion of vertices in the giant converges in probability to the survival probability of the local limit. We extend this result to the setting of random digraphs, where connectivity patterns are significantly more subtle. For this, we identify the precise version of local convergence for digraphs that is needed.

We also determine bounds on the number of strongly connected components, and calculate its asymptotics explicitly for locally tree-like digraphs, as well as for other locally converging digraph sequences under the ‘almost-local’ condition for the strong giant. The fact that the number of strongly connected components is *not* local once more exemplifies the delicate nature of strong connectivity in random digraphs.

1. INTRODUCTION AND MAIN RESULTS

1.1. Introduction. In the realm of network and graph theory, the study of directed graphs, also known as digraphs, has gained considerable interest due to their ubiquitous presence in various real-world applications, such as social networks, transportation systems, and information flow analysis. Understanding their connectivity properties is fundamental for gaining insights into the dynamics of complex systems that are modelled using random digraphs. A strongly connected component is called a *strong giant* when its size is asymptotically linear in the number of vertices. The giant size is crucial for the understanding of connectivity in a network. There are other notions for graph components and subsequently giants in digraphs, sometimes called weak giants, which we discuss in more detail in Section 6.

In this paper, we explore the limiting connectivity structure of locally converging digraphs. Specifically, our research focuses on the limiting proportion of vertices in the largest strongly connected component, and the number of connected components, in these graphs.

Local convergence of random graphs, first introduced in [1, 2], is a well-studied notion. It describes what a graph looks like locally, as the number of vertices grows large. It is discussed in detail in [9, Chapter 2] for undirected graphs, and in [9, Section 9.2] for digraphs. When the local limit of a sequence of random graphs/digraphs is known, we are able to predict

properties about the graph that are continuous with respect to the local topology.

For undirected graphs, the number of vertices in the largest connected component, which is also called the giant, is *not* continuous under the local topology, but [8] shows that this quantity is ‘almost’ local. Indeed, the proportion of vertices in the giant converges to the survival probability of the local limit if and only if a certain condition holds. In this paper we show that the giant strongly connected component is also ‘almost’ local for the case of random digraphs.

The difference between the directed and the undirected case is quite delicate. One example to illustrate this subtlety is through the number of connected components. In the undirected case, the asymptotics for the number of connected components can easily be obtained since it is continuous in the local topology. In contrast, the directed case turns out to be much more complex due the stronger, and less local, notion of connectivity. In particular, it is an open question whether it is ‘almost’ local or not, which we answer affirmatively in this paper.

Connectivity properties of random digraphs have attracted substantial attention. The diameter of the directed configuration model is studied in [3], the size of the largest strongly connected component of a random digraph with prescribed degrees in [4], Weak components of the directed configuration model in [5], percolation in simple random digraphs with prescribed degrees in [11], and graph distances in [10].

1.2. Preliminaries. In this section, we discuss different types of connected components in digraphs, and informally introduce the notion of local convergence in directed random graphs.

Let $G = (V(G), E(G))$ be a digraph. We denote the length of the shortest directed path between $u, v \in V(G)$ by $d_G(u, v)$. We define $d_G(u, v) = \infty$ if there is no directed path between u and v . In digraphs, the distance is not symmetric, i.e., $d_G(u, v) \neq d_G(v, u)$.

Definition 1.1 (Vertex-induced subgraph). Let $G = (V(G), E(G))$ be a digraph and $U \subset V(G)$. Then, the digraph induced by the vertex set U , which we denote by $G[U]$, is the digraph obtained by removing all the vertices in $U^c = V(G) \setminus U$, along with the edges incident to these vertices, from G .

◀

Definition 1.2 (Strongly connected component of a vertex). Fix $v \in V(G)$, then we let \mathcal{C}_v be the strongly connected component of v , defined as

$$\mathcal{C}_v := G[\{x \in V(G) : d_G(x, v) < \infty, d_G(v, x) < \infty\}]. \quad (1.1)$$

◀

Definition 1.3 (In- and out-component, weak components of a vertex). Fix $v \in V(G)$, then we let \mathcal{C}_v^- and \mathcal{C}_v^+ to be the in- and out-component of v ,

respectively, defined as

$$\mathcal{C}_v^- := G[\{x \in V(G) : d_G(x, v) < \infty\}], \quad (1.2)$$

$$\mathcal{C}_v^+ := G[\{x \in V(G) : d_G(v, x) < \infty\}]. \quad (1.3)$$

◀

The strongly connected components of a digraph *partitions* the vertex set. Let $(\mathcal{C}_{(i)})_{i \geq 1}$ denote the unique partition formed by the strongly connected components of G_n , ordered in such a way that $|\mathcal{C}_{(1)}| \geq |\mathcal{C}_{(2)}| \geq |\mathcal{C}_{(3)}| \geq \dots$, where $|\cdot|$ denote the size of the vertex set and ties are broken arbitrarily. Thus, $\mathcal{C}_{(1)}$ is the largest strongly connected component(LSCC) of the graph, which we also denote by \mathcal{C}_{\max} .

Local convergence in random digraphs. Local convergence of random digraphs describes what a digraph locally looks like from the perspective of a uniformly chosen vertex, as the number of vertices in the digraph tends to infinity. For example, the sparse Erdős-Rényi random digraph, which is formed by bond percolation on the complete digraph of size n with probability $p = \lambda/n$, locally looks like a Poisson branching process with parameter λ in both directions, as n tends to infinity.

Now let us give the informal definition of local convergence. We write $X_n \xrightarrow{\mathbb{P}} X$ when X_n converges in probability to X .

Definition 1.4 (Forward-backward neighbourhood). *Let G be a digraph and $v \in V(G)$. Then $B_r^{(G)}(v)$ is the rooted digraph with root v , induced by the vertex set $V(B_r^{(G)}(v))$, where*

$$V(B_r^{(G)}(v)) = \{u \in V(G) : \min\{d_G(u, v), d_G(v, u)\} < r\}, \quad (1.4)$$

where a rooted digraph is a digraph with a distinguished vertex, called a root, defined formally in Definition 2.1. ◀

Definition 1.5 (Digraph isomorphism). Two digraphs, G_1 and G_2 are said to be directed isomorphic if there exists a bijection $\phi: V(G_1) \rightarrow V(G_2)$ such that $(v_1, v_2) \in E(G_1)$ precisely when $(\phi(v_1), \phi(v_2)) \in E(G_2)$. This is abbreviated as $G_1 \cong G_2$. ◀

Definition 1.6 (Rooted digraph isomorphism). Two rooted digraphs, (G_1, o_1) and (G_2, o_2) are said to be directed isomorphic if there exists a bijection $\phi: V(G_1) \rightarrow V(G_2)$ such that $(v_1, v_2) \in E(G_1)$ precisely when $(\phi(v_1), \phi(v_2)) \in E(G_2)$. In addition to this, it should also hold that $\phi(o_1) = o_2$. This is abbreviated as $(G_1, o_1) \cong (G_2, o_2)$. ◀

We will rely on two types of local convergence, namely, forward-backward local weak convergence and forward-backward local convergence in probability. Let o_n denote a uniformly chosen vertex from $V(G_n)$. Forward-backward local weak convergence means that

$$\mathbb{P}(B_r^{(G_n)}(o_n) \cong (H, o')) \rightarrow \bar{\mu}(B_r^{(\bar{G})}(o) \cong (H, o')), \quad (1.5)$$

for all rooted digraphs (H, o') . In (1.5), $(\bar{G}, o) \sim \bar{\mu}$ is a *random* rooted graph, which is called the forward-backward local weak limit. For forward-backward local convergence in probability, instead, we require that

$$\frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong (H, o')\}} \xrightarrow{\mathbb{P}} \mu(B_r^{(G)}(o) \cong (H, o')) \quad (1.6)$$

to hold for all rooted digraphs (H, o') . In (1.6), $(G, o) \sim \mu$ is a random rooted graph which is called the local limit in probability (and bear in mind that μ can possibly be a random measure on rooted graphs). Both (1.5) and (1.6) describe the convergence of the proportions of vertices around which the digraph locally looks like a certain specific digraph.

Remark 1.1 (Other notions of local convergence of digraphs). Other notions of local convergence of digraphs exist. For example, one can consider digraphs as undirected graphs with marks on the edges that indicate their directions. Then one can define marked local convergence of such marked undirected graphs. This notion is much stronger than forward-backward local convergence. Further, sometimes only one direction is relevant. An example arises in the study of the local limit of the PageRank distribution on a digraph, as studied in [6]. There, the in-component matters, with as extra vertex marks the out-degree of the vertices. We see that convergence of different digraph properties require different local convergence notions. Forward-backward convergence is exactly what we need for the convergence of the strong giant and the number of strongly connected components. ◀

We discuss the definition of forward-backward local convergence more formally in Section 2. We now turn to our main results.

1.3. Main results. Our main results are divided into three subsections, the ‘almost’ locality of the giant, the local limit of the giant, and the number of strongly connected components.

1.3.1. Strong giant is ‘almost’ local. In this section we state a condition for the existence and size of a giant in a sequence of random digraphs that converges locally in probability in the forward-backward sense. We start by showing a generic probabilistic upper bound on the size of the strong giant on random digraph sequences converging locally in probability:

Theorem 1.1 (Upper bound on size of strong giant). *Let $(G_n)_{n \geq 1}$ be a sequence of random digraphs, of size $|V(G_n)| = n$, converging locally in probability in the forward-backward sense to $(G, o) \sim \mu$. Let $\zeta = \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty)$. Then, for all $\varepsilon > 0$,*

$$\mathbb{P} \left(\frac{|\mathcal{C}_{\max}|}{n} \geq \zeta + \varepsilon \right) \rightarrow 0. \quad (1.7)$$

Theorem 1.1 implies that $\zeta = \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty)$ is always an upper bound on the size of the strong giant in locally converging digraph sequences.

Next, we proceed towards a matching lower bound, for which we will state a necessary and sufficient condition. Define

$$N_n^k := \#\{(x, y) \in V(G_n) : |\mathcal{C}_x^-|, |\mathcal{C}_x^+|, |\mathcal{C}_y^-|, |\mathcal{C}_y^+| \geq k, x \not\leftrightarrow y\}, \quad (1.8)$$

which is the number of vertex pairs that have large in- and out-components but are not in the same strongly connected component. In (1.8), $x \not\leftrightarrow y$ means that either x is not connected to y , or y is not connected to x (i.e., x and y are in different strongly connected components). Then, our main result on the ‘almost-local’ nature of the strong giant is as follows:

Theorem 1.2 (Strong giant is ‘almost’ local). *Let $(G_n)_{n \geq 1}$ be a sequence of random digraphs, of size $|V(G_n)| = n$, converging locally in probability in the forward-backward sense to $(G, o) \sim \mu$. Assume that*

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} [N_n^k] = 0. \quad (1.9)$$

Then

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} \zeta = \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty), \quad \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{\mathbb{P}} 0. \quad (1.10)$$

Under the assumption of forward-backward local convergence in probability, the condition in (1.9) is necessary and sufficient for (1.10) to hold.

Remark 1.2 (The limiting value). One might have expected Theorem 1.2 to hold under different conditions with a different limit. Indeed, one may have thought that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o| = \infty)$, which might be true in some cases. However, in many directed random graphs, the local limit is a tree in that \mathcal{C}_o^- and \mathcal{C}_o^+ are disjoint trees μ -almost surely. In such cases, $\mu(|\mathcal{C}_o| = \infty) = 0$, while $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta > 0$ can occur. This exemplifies the subtleties of strongly connected components in the directed setting. It would be of interest to find examples for which $\mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty) = \mu(|\mathcal{C}_o| = \infty)$. ◀

Remark 1.3 (Comparison to undirected setting). The subtleties described in Remark 1.2 are not present in the undirected setting, where $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o| = \infty)$ and \mathcal{C}_o is the undirected connected component in (G, o) . This can be understood by noting that if $u \in B_r^{(G)}(v)$ in the undirected setting, then we know that $u \in \mathcal{C}_v$, while if $u \in B_r^{(G)}(v)$ in the directed setting, then we do not know that $u \in \mathcal{C}_v$. ◀

Remark 1.4 (Applications of Theorem 1.2). Theorem 1.2 can be applied to the directed Erdős-Rényi random graph model, and the directed configuration model with appropriate regularity conditions on the degree distribution. Condition (1.9) for these models can be shown using coupling arguments similar to the ones used in [9, Section 2.3] and [8], respectively. We do not provide more details. ◀

Remark 1.5 (Simplification of condition in Theorem 1.2). It is possible to simplify the conditions in Theorem 1.2. Indeed, define

$$N_n^k(2) = \#\{(x, y) \in V(G_n) : |\mathcal{C}_x^-|, |\mathcal{C}_x^+|, |\mathcal{C}_y^-|, |\mathcal{C}_y^+| \geq k, x \not\leftrightarrow y\}, \quad (1.11)$$

where $x \nrightarrow y$ means that there is no directed path from x to y . In Lemma 3.2 below, we show that $N_n^k(2) \leq N_n^k \leq 2N_n^k(2)$. Thus, one can replace N_n^k by $N_n^k(2)$ in (1.9). \blacktriangleleft

1.3.2. *Local limit of the giant.* In this section, we discuss the local convergence of the giant. Denote the number of vertices v in the giant component satisfying $(d_v^-, d_v^+) = (l, m)$ by $v_1(l, m)$. Then, we have the following asymptotics for $v_1(l, m)$:

Theorem 1.3 (Limiting degree distribution of the giant). *Under the assumptions of Theorem 1.2,*

$$\frac{v_1(l, m)}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty, D_o = (l, m)). \quad (1.12)$$

Further, when $D_{o_n}^-$ and $D_{o_n}^+$ are both uniformly integrable,

$$\frac{|E(\mathcal{C}_{\max})|}{n} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[(D_o^- + D_o^+) \mathbb{1}_{\{|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty\}}]. \quad (1.13)$$

The next theorem tells us what the strong giant and its complement look like locally:

Theorem 1.4 (Local limit of the strong giant and its complement). *Under the assumptions of Theorem 1.2,*

$$\frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty, B_r^{(G)}(o) \cong H_\star), \quad (1.14)$$

and

$$\frac{1}{n} \sum_{v \notin \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| < \infty, B_r^{(G)}(o) \cong H_\star). \quad (1.15)$$

1.3.3. *Number of strongly connected components.* In this section, we discuss the number of strongly connected components of digraph sequences having a local limit:

Theorem 1.5 (Limiting number of SCCs). *Under the assumptions of Theorem 1.2, and further assuming that $|\mathcal{C}_{\max}| \xrightarrow{\mathbb{P}} \infty$ when $\zeta = 0$,*

$$\frac{K_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}_\mu \left[\frac{1}{|\mathcal{C}(o)|} \mathbb{1}_{\{|\mathcal{C}^-(o)| < \infty\} \cup \{|\mathcal{C}^+(o)| < \infty\}} \right]. \quad (1.16)$$

Remark 1.6 (Comparison to undirected setting). In the undirected setting, $K_n/n \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[1/|\mathcal{C}(o)|]$ always holds when G_n converges locally in probability, i.e., we do not need to rely on the additional ‘giant-is-almost-local condition’ that is the undirected equivalent of (1.9). This once again shows that strong connectivity in digraphs is a more delicate notion than connectivity in undirected graphs. Having Theorem 1.2 in mind, we can think of

$\mathbb{1}_{\{|\mathcal{C}^-(o)| < \infty\} \cup \{|\mathcal{C}^+(o)| < \infty\}} = 1 - \mathbb{1}_{\{|\mathcal{C}^-(o)| = \infty\} \cap \{|\mathcal{C}^+(o)| = \infty\}}$
as enforcing that o is in the limit of the strong giant. \blacktriangleleft

We next give a simpler limiting result for locally tree-like digraphs, which we first define:

Definition 1.7 (Locally tree-like digraphs). We say that a sequence of digraphs $(G_n)_{n \geq 1}$ is *locally tree-like* when the sequence converges locally in probability to $(G, o) \sim \mu$, which is almost surely a random directed tree. By directed tree we mean that the strongly connected component of each vertex is almost surely the vertex itself i.e., a graph with no directed cycles. Directed configuration models, directed Erdős-Rényi models, etc. are all locally tree-like digraphs. ◀

Theorem 1.6 (Number of SCCs in locally tree-like digraphs). *Let $(G_n)_{n \geq 1}$ be a sequence of random digraphs, of size $|V(G_n)| = n$, converging locally in probability in the forward-backward sense to $(G, o) \sim \mu$, which is a random directed tree. Then, under the assumption of Theorem 1.2,*

$$\frac{K_n}{n} \xrightarrow{\mathbb{P}} 1 - \zeta. \quad (1.17)$$

Remark 1.7 (Locally tree-like digraphs are special). The limiting value for the number of strongly connected components in Theorem 1.5, if computed for locally tree-like digraphs, matches with the expression in Theorem 1.6. Thus, Theorem 1.6 tells us that we do not need additional conditions for locally tree-like digraphs, as in Theorem 1.5. ◀

Remark 1.8 (Upper bound in terms of ζ). The upper bound in Theorem 1.6 holds more generally. Indeed, the bound

$$\frac{K_n}{n} \leq 1 - \frac{|\mathcal{C}_{\max}|}{n} + \frac{1}{n} \quad (1.18)$$

always holds. Thus, under the assumptions of Theorem 1.6, whp for every $\varepsilon > 0$,

$$\frac{K_n}{n} \leq 1 - \zeta + \varepsilon. \quad (1.19)$$

Remark 1.9 (Comparison of proofs between directed and undirected graphs). The complexity of proofs in digraphs in comparison to undirected graphs varies depending on the connectivity property we are investigating. In particular, the proofs for the size of the strong giant are similar to the proofs used for the undirected case in [8]. In contrast, for digraphs, the number of connected components is a non-local quantity, whereas, for undirected graphs, it is local. Thus, the proof becomes more complicated and less general. ◀

2. LOCAL CONVERGENCE IN DIGRAPHS

In this section, we formally define the notion of forward-backward local weak convergence and local convergence in probability for digraphs. There are several versions of local convergence in random digraphs. Amongst these,

the four most commonly discussed are forward, backward, forward-backward, and marked (see Remark 1.1).

We start by defining rooted digraphs:

Definition 2.1 (Rooted digraph). A digraph is $G = (V(G), E(G))$ together with a vertex $o \in V(G)$ is called a rooted digraph. We denote a rooted digraph G with root $o \in V(G)$ by (G, o) . ◀

Next we define the forward-backward metric space over the rooted digraphs:

Definition 2.2 (Metric space). Let \mathcal{G}_\star be the space of all rooted digraphs. Let $(G_1, o_1), (G_2, o_2) \in \mathcal{G}_\star$. Define the metric $d_{\mathcal{G}_\star} : \mathcal{G}_\star \times \mathcal{G}_\star \rightarrow \mathbb{R}_{\geq 0}$ by

$$d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2)) := \frac{1}{R^\star + 1}. \quad (2.1)$$

where R^\star is given by

$$R^\star := \sup\{r \geq 0 \text{ s.t. } B_r^{(G_1)}(o_1) \cong B_r^{(G_2)}(o_2)\},$$

where \cong denotes rooted digraph isomorphism defined in Definition 1.6. ◀

Remark 2.1 (Equivalence classes). Obviously, $R^\star = \infty$ when $(G_1, o_1) \cong (G_2, o_2)$ so that $d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2)) = 0$. Therefore, $d_{\mathcal{G}_\star}$ defines a metric on the equivalence classes of isomorphic digraphs. ◀

Definition 2.3 (Forward-backward local convergence in probability of digraphs). The sequence $(G_n)_{n \geq 1}$ of random digraphs is said to converge locally in probability to the random rooted digraph (G, o) , a random variable taking values in \mathcal{G}_\star having law μ , as $n \rightarrow \infty$, if for every continuous and bounded function $f : \mathcal{G}_\star \rightarrow \mathbb{R}$ in the forward-backward metric,

$$\mathbb{E}_n[f(G_n, o_n)] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[f(G, o)], \quad (2.2)$$

where o_n is a uniformly chosen vertex from $V(G_n)$. ◀

3. GIANTS IN CONVERGING DIGRAPHS: PROOFS OF THEOREMS 1.1 & 1.2

In this section we derive a condition for the existence of a giant in a sequence of random digraphs that converges locally in probability in the forward-backward sense. We start by proving the generic probabilistic upper bound on the size of the strong giant in Theorem 1.1.

Proof of Theorem 1.1. Let $Z_{\geq k}$ be the number of vertices in G_n having in- and out-components size at least k , i.e.,

$$Z_{\geq k} = \sum_{v \in V(G_n)} \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k, |\mathcal{C}_v^+| \geq k\}}. \quad (3.1)$$

Let $\zeta_{\geq k} = \mu(|\mathcal{C}_o^-| \geq k, |\mathcal{C}_o^+| \geq k)$. Then, taking $k \rightarrow \infty$, we have $\zeta_{\geq k} \rightarrow \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty)$.

Notice that $(G, o) \mapsto \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k, |\mathcal{C}_v^+| \geq k\}}$ is a bounded continuous function in the forward-backward sense of random digraphs, and thus due to the forward-backward local convergence in probability, $Z_{\geq k}/n \xrightarrow{\mathbb{P}} \zeta_{\geq k}$. Therefore, for all $\varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{Z_{\geq k}}{n} - \zeta_{\geq k} \right| > \varepsilon \right) \rightarrow 0. \quad (3.2)$$

Also, $\zeta_{\geq k} \rightarrow \zeta$ as $k \rightarrow \infty$, which means that there exists $K \in \mathbb{N}$ such that $\zeta_{\geq k} \leq \zeta + \varepsilon/2$ for all $k \geq K$. Further, $\{|\mathcal{C}_{\max}| \geq x\} \subseteq \{Z_{\geq k} \geq x\}$, for every $x \geq 1$. Thus,

$$\mathbb{P} \left(\frac{|\mathcal{C}_{\max}|}{n} \geq \zeta + \varepsilon \right) \leq \mathbb{P} \left(\frac{|\mathcal{C}_{\max}|}{n} \geq \zeta_{\geq k} + \frac{\varepsilon}{2} \right) \leq \mathbb{P} \left(\frac{Z_{\geq k}}{n} \geq \zeta_{\geq k} + \frac{\varepsilon}{2} \right) \rightarrow 0. \quad (3.3)$$

This completes the proof. \square

Remark 3.1 (In- and out-components of strongly connected components). Each vertex in a strongly connected component has the same in- and out-component. This is illustrated in Figure 2 for the largest strongly connected component. We denote the in- and out-component associated with $\mathcal{C}_{(i)}$ by $\mathcal{C}_{(i)}^-$ and $\mathcal{C}_{(i)}^+$, respectively. \blacktriangleleft

In the remainder of the proof, the following notation will be essential:

Notation 3.1. Suppose $(X_n)_{n \geq 1}$ and $(X_{n,k})_{n \geq 1, k \geq 1}$ are sequences of random variables. We write $X_{n,k} = o_{k,\mathbb{P}}(X_n)$ when

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|X_{n,k}| > \varepsilon |X_n|) = 0. \quad \blacktriangleleft$$

Lemma 3.1 (Sum of squares of cluster sizes). *Under the assumptions of Theorem 1.2,*

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}} = \zeta^2 + o_{k,\mathbb{P}}(1). \quad (3.4)$$

Proof. By Remark 3.1, we can rewrite $Z_{\geq k}$ in (3.1) as

$$Z_{\geq k} = \sum_{v \in V(G_n)} \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k, |\mathcal{C}_v^+| \geq k\}} = \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}}. \quad (3.5)$$

Squaring (3.5) gives

$$Z_{\geq k}^2 = \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}} + \sum_{i \neq j} |\mathcal{C}_{(i)}| |\mathcal{C}_{(j)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}^-|, |\mathcal{C}_{(i)}^+|, |\mathcal{C}_{(j)}^-|, |\mathcal{C}_{(j)}^+| \geq k\}}. \quad (3.6)$$

The second cross-product term counts the number of pairs of vertices that are not in the same strongly connected component, but have at least k vertices in their in- and out-components. Thus, by (1.8),

$$\frac{Z_{\geq k}^2}{n^2} = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}} + \frac{1}{n^2} N_n^k. \quad (3.7)$$

Fix $\varepsilon > 0$. By the Markov inequality,

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n^2} N_n^k > \varepsilon \right) \leq \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{\varepsilon n^2} \mathbb{E} [N_n^k] = 0, \quad (3.8)$$

Thus, $N_n^k = o_{k, \mathbb{P}}(n^2)$, proving (3.4). This completes the proof. \square

Next we prove Theorem 1.2:

Proof of Theorem 1.2. Due to Theorem 1.1, it is sufficient to show (1.10) for $\zeta > 0$. Thus, we assume that $\zeta > 0$. Observe that

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}} \leq \frac{|\mathcal{C}_{\max}|}{n} \frac{1}{n} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}}, \quad (3.9)$$

which means that

$$\frac{|\mathcal{C}_{\max}|}{n} \geq \frac{\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}}}{\frac{1}{n} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}^-| \geq k, |\mathcal{C}_{(i)}^+| \geq k\}}} = \frac{\zeta^2 + o_{k, \mathbb{P}}(1)}{\zeta_{\geq k} + o_{\mathbb{P}}(1)}. \quad (3.10)$$

Letting $k \rightarrow \infty$ in (3.9), for all $\varepsilon > 0$, we get

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{|\mathcal{C}_{\max}|}{n} \geq \zeta - \varepsilon \right) = 1. \quad (3.11)$$

Equation (3.11), combined with Theorem 1.1, completes the proof of the law of large numbers on the strong giant under the condition (1.9) in Theorem 1.2.

Next we show that (1.9) is also a necessary condition for the giant component to converge to ζ . Suppose that the condition in (1.9) fails. This implies that

$$\limsup_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[N_n^k] = \kappa > 0. \quad (3.12)$$

In turn, this implies that there exists a subsequence $(n_l)_{l \geq 1}$ such that

$$\lim_{l \rightarrow \infty} \limsup_{k \rightarrow \infty} \frac{1}{n_l^2} \mathbb{E}[N_{n_l}^k] = \kappa. \quad (3.13)$$

Thus, along the subsequence $(n_l)_{l \geq 1}$,

$$\begin{aligned} \frac{1}{n_l^2} \mathbb{E} \left[|\mathcal{C}_{\max}|^2 \mathbb{1}_{\{|\mathcal{C}_{\max}^+| \geq k, |\mathcal{C}_{\max}^-| \geq k\}} \right] &\leq \frac{1}{n_l^2} \mathbb{E} \left[\sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}^+| \geq k, |\mathcal{C}_{(i)}^-| \geq k\}} \right] \\ &= \frac{1}{n_l^2} \mathbb{E} [Z_{\geq k}^2 - N_{n_l}^k] \rightarrow \zeta_{\geq k}^2 - \kappa. \end{aligned}$$

Notice that

$$\begin{aligned} \frac{1}{n_l^2} \mathbb{E} [|\mathcal{C}_{\max}|^2] &= \frac{1}{n_l^2} \mathbb{E} \left[|\mathcal{C}_{\max}|^2 \mathbb{1}_{\{|\mathcal{C}_{\max}^+| \geq k, |\mathcal{C}_{\max}^-| \geq k\}} \right] \\ &\quad + \frac{1}{n_l^2} \mathbb{E} \left[|\mathcal{C}_{\max}|^2 \mathbb{1}_{\{|\mathcal{C}_{\max}^+| < k\} \cup \{|\mathcal{C}_{\max}^-| < k\}} \right]. \end{aligned} \quad (3.14)$$

We know that $|\mathcal{C}_{\max}^\pm| \geq |\mathcal{C}_{\max}|$. Thus,

$$\{|\mathcal{C}_{\max}^+| < k\} \cup \{|\mathcal{C}_{\max}^-| < k\} \subseteq \{|\mathcal{C}_{\max}| < k\},$$

so that, in turn,

$$\frac{1}{n_l^2} \mathbb{E} \left[|\mathcal{C}_{\max}|^2 \mathbb{1}_{\{|\mathcal{C}_{\max}^+| < k\} \cup \{|\mathcal{C}_{\max}^-| < k\}} \right] \leq \left(\frac{k}{n_l} \right)^2 \rightarrow 0, \quad (3.15)$$

as $n_l \rightarrow \infty$. This implies

$$\lim_{l \rightarrow \infty} \frac{1}{n_l^2} \mathbb{E}[|\mathcal{C}_{\max}|^2] \leq \zeta^2 - \kappa < \zeta^2. \quad (3.16)$$

We conclude that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$ cannot hold. This is because by the bounded convergence theorem, it would imply that $\mathbb{E} \left[(|\mathcal{C}_{\max}|/n)^2 \right] \rightarrow \zeta^2$ which gives a contradiction with (3.16). \square

Recall the definition of N_n^k in (1.8), and that of $N_n^k(2)$ in (1.11). We now show that both are of the same order of magnitude:

Lemma 3.2 (Relaxing the assumption). $N_n^k(2) \leq N_n^k \leq 2N_n^k(2)$.

Proof. For $x, y \in V(G_n)$ we define the events $A_{x,y}$ and $B_{x,y}$ as

$$A_{x,y} = [|\mathcal{C}_x^-|, |\mathcal{C}_x^+|, |\mathcal{C}_y^-|, |\mathcal{C}_y^+| \geq k, x \rightarrow y], \quad (3.17)$$

$$B_{x,y} = [|\mathcal{C}_x^-|, |\mathcal{C}_x^+|, |\mathcal{C}_y^-|, |\mathcal{C}_y^+| \geq k, x \leftarrow y]. \quad (3.18)$$

Then,

$$N_n^k = \sum_{(x,y) \in V(G_n)^2} \mathbb{1}_{A_{x,y} \cup B_{x,y}}, \quad N_n^k(2) = \sum_{(x,y) \in V(G_n)^2} \mathbb{1}_{A_{x,y}} = \sum_{(x,y) \in V(G_n)^2} \mathbb{1}_{B_{x,y}}. \quad (3.19)$$

Using the fact $\mathbb{1}_A \leq \mathbb{1}_{A \cup B} \leq \mathbb{1}_A + \mathbb{1}_B$, we get

$$N_n^k(2) = \sum_{(x,y) \in V(G_n)^2} \mathbb{1}_{A_{x,y}} \leq N_n^k \leq \sum_{(x,y) \in V(G_n)^2} (\mathbb{1}_{A_{x,y}} + \mathbb{1}_{B_{x,y}}) = 2N_n^k(2), \quad (3.20)$$

which completes the proof. \square

4. LOCAL LIMIT OF THE GIANT: PROOF OF THEOREMS 1.3 & 1.4

In this section we prove local limit properties of the giant component:

Proof of Theorem 1.3. When $\zeta = 0$, there is nothing to prove, so we may again assume that $\zeta > 0$. For $A \subseteq \mathbb{N}^2$, we define

$$Z_{A, \geq k} = \sum_{v \in V(G_n)} \mathbb{1}_{\{|\mathcal{C}_v^-|, |\mathcal{C}_v^+| \geq k, (d_v^-, d_v^+) \in A\}}. \quad (4.1)$$

Assuming that G_n converges locally in probability in the forward-backward sense, we get

$$\frac{Z_{A, \geq k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-|, |\mathcal{C}_o^+| \geq k, (d_o^-, d_o^+) \in A). \quad (4.2)$$

Since $|\mathcal{C}_{\max}| > k$ with high probability (it is here that we use that $\zeta > 0$), we have

$$\frac{1}{n} \sum_{(l, m) \in A} v_1(l, m) \leq \frac{Z_{A, \geq k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-|, |\mathcal{C}_o^+| \geq k, (d_o^-, d_o^+) \in A). \quad (4.3)$$

Define

$$\mu_k(A) = \mu(|\mathcal{C}_o^-|, |\mathcal{C}_o^+| \geq k, (d_o^-, d_o^+) \in A). \quad (4.4)$$

Applying (4.3) for $A = \{(l, m)\}^c$, we get

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} |\mathcal{C}_{\max} - v_1(l, m)| \leq \mu_k(A) + \frac{\varepsilon}{2} \right) = 1. \quad (4.5)$$

We argue by contradiction. Suppose that for some pair $(l, m) \in \mathbb{N}^2$,

$$\liminf_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} v_1(l, m) \leq \mu_k(\{l, m\}) - \varepsilon \right) = \kappa > 0. \quad (4.6)$$

Then along the subsequence $(n_j)_{j \geq 1}$, where the above liminf is attained, with a non-zero probability the following holds:

$$\frac{1}{n} |\mathcal{C}_{\max}| = \frac{1}{n} (|\mathcal{C}_{\max}| - v_1(l, m)) + \frac{1}{n} v_1(l, m) \leq \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty) - \frac{\varepsilon}{2}, \quad (4.7)$$

which contradicts Theorem 1.2. Thus, (1.12) holds.

For the next part, we just observe that

$$\frac{|E(\mathcal{C}_{\max})|}{n} = \frac{1}{2n} \sum_{k, l} (k + l) v_1(k, l). \quad (4.8)$$

Thus, for any natural number N .

$$\frac{|E(\mathcal{C}_{\max})|}{n} = \frac{1}{2n} \sum_{k+l \leq N} (k + l) v_1(k, l) + \frac{1}{2n} \sum_{k+l > N} (k + l) v_1(k, l). \quad (4.9)$$

From (1.12) and the dominated convergence theorem, we conclude that

$$\begin{aligned} \frac{1}{2n} \sum_{k+l \leq N} (k + l) v_1(k, l) &\xrightarrow{\mathbb{P}} \frac{1}{2} \sum_{k+l \leq N} (k + l) \mu(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty, D_o = (k, l)) \\ &= \frac{1}{2} \mathbb{E}_\mu \left[(D_o^- + D_o^+) \mathbb{1}_{\{D_o^- + D_o^+ < N\}} \right]. \end{aligned} \quad (4.10)$$

For the second term in (4.9), we bound $v_1(k, l)$ by the total number of vertices with in-degree k and out-degree l , which we denote by $n_{k,l}$, to get

$$\begin{aligned} \frac{1}{2n} \sum_{k+l>N} (k+l)v_1(k, l) &\leq \frac{1}{2} \sum_{k+l>N} (k+l) \frac{n_{k,l}}{n} \\ &= \frac{1}{2} \mathbb{E} \left[(D_{o_n}^- + D_{o_n}^+) \mathbb{1}_{\{D_{o_n}^- + D_{o_n}^+ > N\}} \mid G_n \right]. \end{aligned} \quad (4.11)$$

By uniform integrability of $D_{o_n}^-$ and $D_{o_n}^+$,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[(D_{o_n}^- + D_{o_n}^+) \mathbb{1}_{\{D_{o_n}^- + D_{o_n}^+ > N\}} \right] = 0. \quad (4.12)$$

Thus, using the Markov inequality we can conclude that for each $\varepsilon > 0$, there exists a $N = N(\varepsilon) < \infty$ such that

$$\mathbb{P} \left(\mathbb{E} \left[(D_{o_n}^- + D_{o_n}^+) \mathbb{1}_{\{D_{o_n}^- + D_{o_n}^+ > N\}} \mid G_n \right] > \varepsilon \right) \rightarrow 0. \quad (4.13)$$

This completes the proof for the second part of the theorem. \square

Proof of Theorem 1.4. When $\zeta = 0$, there is nothing to prove, so we may again assume that $\zeta > 0$. The proof is a minor modification of that of Theorem 1.3. Let \mathcal{G}_\star be the space of all rooted digraphs. By forward-backward local convergence in probability of $(G_n)_{n \geq 1}$, for every $\mathcal{H}_\star \subseteq \mathcal{G}_\star$,

$$\frac{1}{n} \sum_{v \in V(G_n)} \mathbb{1}_{\{B_r^{(G_n)}(v) \in \mathcal{H}_\star\}} \xrightarrow{\mathbb{P}} \mu(B_r^{(G)}(o) \in \mathcal{H}_\star). \quad (4.14)$$

Thus, (1.15) follows from (1.14). Therefore, we only need to show (1.14). Define

$$\begin{aligned} Z_{\geq k, \mathcal{H}_\star} &= \frac{1}{n} \sum_{v \in V(G_n)} \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k, |\mathcal{C}_v^+| \geq k, B_r^{(G_n)}(v) \in \mathcal{H}_\star\}} \\ &\xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-| \geq k, |\mathcal{C}_o^+| \geq k, B_r^{(G)}(o) \in \mathcal{H}_\star). \end{aligned} \quad (4.15)$$

Because of Theorem 1.2, we have that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta > 0$. Thus, on the high probability event $\{|\mathcal{C}_{\max}| \geq k\}$,

$$\frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \in \mathcal{H}_\star\}} \leq Z_{\geq k, \mathcal{H}_\star} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}_o^-| \geq k, |\mathcal{C}_o^+| \geq k, B_r^{(G)}(o) \in \mathcal{H}_\star). \quad (4.16)$$

Define

$$\mu_k(\mathcal{H}_\star) = \mu(|\mathcal{C}_o^-| \geq k, |\mathcal{C}_o^+| \geq k, B_r^{(G)}(o) \in \mathcal{H}_\star). \quad (4.17)$$

Then, by (4.16),

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \in \mathcal{H}_\star\}} \leq \mu_k(\mathcal{H}_\star) + \frac{\varepsilon}{2} \right) = 1. \quad (4.18)$$

Choosing $\mathcal{H}_\star = \{H_\star\}$, gives the upper bound. For, $\mathcal{H}_\star = \{H_\star\}^c$ in (4.18),

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \left[|\mathcal{C}_{\max}| - \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \right] \leq \mu_k(\mathcal{H}_\star) + \frac{\varepsilon}{2} \right) = 1. \quad (4.19)$$

We again argue by contradiction. For this, we assume that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \leq \mu_k(\{H_\star\}) - \frac{\varepsilon}{2} \right) = \kappa > 0. \quad (4.20)$$

Then, with asymptotic probability κ ,

$$\begin{aligned} \frac{|\mathcal{C}_{\max}|}{n} &= \frac{1}{n} \left[|\mathcal{C}_{\max}| - \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \right] + \frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \mathbb{1}_{\{B_r^{(G_n)}(v) \cong H_\star\}} \\ &\leq \zeta_{\geq k} - \varepsilon/2 \leq \zeta - \varepsilon/2. \end{aligned} \quad (4.21)$$

This contradicts Theorem 1.2. Thus, (4.20) cannot hold. This completes the proof. \square

5. NUMBER OF STRONG COMPONENTS: PROOF OF THEOREMS 1.5 & 1.6

In this section, we prove results concerning the number of strongly connected components of random digraph sequences having a local limit.

5.1. Number of strong components is ‘almost’ local: Proof of Theorem 1.5. In this section, we prove Theorem 1.5. We prove a lower bound that holds more generally, and an upper bound that holds under the assumptions stated in Theorem 1.5.

We first use that

$$K_n = \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|}. \quad (5.1)$$

Indeed, if one sums the right-hand side for the vertices as partitioned by the strongly connected components, one obtains

$$\sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} = \sum_{1 \leq i \leq K_n} \sum_{v \in \mathcal{C}_{(i)}} \frac{1}{|\mathcal{C}_{(i)}|} = \sum_{1 \leq i \leq K_n} 1 = K_n. \quad (5.2)$$

Equation (5.1) is the starting point of our analysis. However, we emphasize that, unlike in the undirected case, $h(G, o) = 1/|\mathcal{C}_v|$ is *not* a bounded and continuous function in the forward-backward sense.

For the lower bound, we fix k and bound

$$K_n \geq \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{|\mathcal{C}_v^-| < k\} \cup \{|\mathcal{C}_v^+| < k\}}. \quad (5.3)$$

For each $k < \infty$, the function

$$h(G, o) = \frac{1}{|\mathcal{C}_o|} \mathbb{1}_{\{|\mathcal{C}_o^-| < k\} \cup \{|\mathcal{C}_o^+| < k\}}$$

is a bounded continuous function. Indeed, any $u \in \mathcal{C}_o$ is on a directed cycle starting and ending at o . When $|\mathcal{C}_o^-| < k$ or $|\mathcal{C}_o^+| < k$, there cannot be any cycle longer than $k - 1$, so that we can determine $h(G, o)$ on the basis of $B_k^{(G)}(o)$ (recall Definition 1.4).

We conclude that

$$\frac{1}{n} \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{|\mathcal{C}_v^-| < k\} \cup \{|\mathcal{C}_v^+| < k\}} \xrightarrow{\mathbb{P}} \mathbb{E}_\mu \left[\frac{1}{|\mathcal{C}_o|} \mathbb{1}_{\{|\mathcal{C}_o^-| < k\} \cup \{|\mathcal{C}_o^+| < k\}} \right]. \quad (5.4)$$

Since, for $k \rightarrow \infty$,

$$\mathbb{E}_\mu \left[\frac{1}{|\mathcal{C}_o|} \mathbb{1}_{\{|\mathcal{C}_o^-| < k\} \cup \{|\mathcal{C}_o^+| < k\}} \right] \rightarrow \mathbb{E}_\mu \left[\frac{1}{|\mathcal{C}_o|} \mathbb{1}_{\{|\mathcal{C}_o^-| < \infty\} \cup \{|\mathcal{C}_o^+| < \infty\}} \right], \quad (5.5)$$

we obtain that

$$\frac{K_n}{n} \geq \mathbb{E}_\mu \left[\frac{1}{|\mathcal{C}_o|} \mathbb{1}_{\{|\mathcal{C}_o^-| < \infty\} \cup \{|\mathcal{C}_o^+| < \infty\}} \right] + o_{\mathbb{P}}(1), \quad (5.6)$$

as required.

For the upper bound, we again fix k and now write

$$K_n = \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{|\mathcal{C}_v^-| < k\} \cup \{|\mathcal{C}_v^+| < k\}} + \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}}. \quad (5.7)$$

It suffices to prove that the second term in (5.7) is $o_{k, \mathbb{P}}(n)$, which is what we will do now. For this, we make the further split

$$\begin{aligned} & \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}} \\ &= \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} [\mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}} - \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}}] + \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}}. \end{aligned} \quad (5.8)$$

We note that

$$\frac{1}{n} \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}} = \frac{1}{n} \sum_{v \in \mathcal{C}_{\max}} \frac{1}{|\mathcal{C}_{\max}|} = \frac{1}{n} \rightarrow 0. \quad (5.9)$$

Further, since $|\mathcal{C}_{\max}| \xrightarrow{\mathbb{P}} \infty$ by assumption, we have that $|\mathcal{C}_{\max}| \geq k$ whp, so that whp also

$$\mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}} - \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}} \geq 0.$$

As a result, using that $|\mathcal{C}_v| \geq 1$, on the high probability event that $|\mathcal{C}_{\max}| \geq k$,

$$\begin{aligned} & \frac{1}{n} \sum_{v \in V(G_n)} \frac{1}{|\mathcal{C}_v|} [\mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}} - \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}}] \\ & \leq \frac{1}{n} \sum_{v \in V(G_n)} [\mathbb{1}_{\{|\mathcal{C}_v^-| \geq k\} \cap \{|\mathcal{C}_v^+| \geq k\}} - \mathbb{1}_{\{v \in \mathcal{C}_{\max}\}}] \\ & = \frac{1}{n} [Z_{\geq k} - |\mathcal{C}_{\max}|]. \end{aligned} \quad (5.10)$$

By Theorem 1.1 and forward-backward local convergence in probability of G_n ,

$$\frac{1}{n} [Z_{\geq k} - |\mathcal{C}_{\max}|] \xrightarrow{\mathbb{P}} \zeta_{\geq k} - \zeta, \quad (5.11)$$

which vanishes as $k \rightarrow \infty$. We conclude that also the left-hand side of (5.10) is $o_{k, \mathbb{P}}(1)$, which, together with (5.9), completes the proof. \square

5.2. Isolated vertices in digraphs and locally tree-like digraphs. In this section, we investigate strongly isolated vertices, which are of interest because each such vertex contributes a unique strongly connected component. In locally tree-like graphs these are the major contributors when counting the number of strongly connected components, as we will show in this section.

Definition 5.1 (Strongly isolated vertices). A vertex $v \in V(G)$ is said to be *strongly isolated* if $\mathcal{C}_v = \{v\}$. We denote the set of strongly isolated vertices of G by $V_1(G)$, and denote $\alpha_1(n) = |V_1(G)|$.

We next study $\alpha_1(n)$:

Lemma 5.1 (Upper bound for the proportion of strongly isolated vertices). *Under the assumptions of Theorem 1.2, and for all $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha_1(n) \leq 1 - \zeta + \varepsilon) = 1, \quad (5.12)$$

where $\zeta = \mathbb{P}(|\mathcal{C}_o^-| = |\mathcal{C}_o^+| = \infty)$.

Proof. The number of isolated vertices cannot exceed the number of vertices outside the giant component, since the vertices in the giant cannot be isolated. This gives $\alpha_1(n) \leq n - |\mathcal{C}_{\max}|$. Dividing by n on both sides gives the required result. \square

Definition 5.2 (In- and out-neighbourhood of a fixed finite radius). Let G be a digraph and $v \in V(G)$, then we let $B_k^-(v)$ and $B_k^+(v)$ to be the subgraphs induced by the vertices in the k radius in- and out-neighbourhood of v , respectively, defined as

$$B_k^-(v) = G[\{u \in V(G) \text{ s.t. } d_G(u, v) < k\}], \quad (5.13)$$

$$B_k^+(v) = G[\{u \in V(G) \text{ s.t. } d_G(v, u) < k\}]. \quad (5.14)$$

Similarly, we let $\partial B_k^-(v)$ and $\partial B_k^+(v)$ to be the boundary of k radius in- and out-neighbourhoods of v , respectively, defined as

$$\partial B_k^-(v) = \{u \in V(G) \text{ s.t. } d_G(u, v) = k\}, \quad (5.15)$$

$$\partial B_k^+(v) = \{u \in V(G) \text{ s.t. } d_G(v, u) = k\}. \quad (5.16)$$

◀

We next turn to locally tree-like digraphs:

Lemma 5.2 (Proportion of strongly isolated vertices in locally tree-like digraphs). *Suppose that G_n is a sequence of random digraphs such that G_n converges forward-backward locally in probability to $(G, o) \sim \mu$, which is almost surely a random directed tree. Then, under the assumption of Theorem 1.2,*

$$\alpha_1(n) \xrightarrow{\mathbb{P}} 1 - \zeta. \quad (5.17)$$

Proof of Lemma 5.2. Fix a finite $k \in \mathbb{N}$. Then

$$\begin{aligned} \alpha_1(n) &= \frac{1}{n} \sum_{v \in V(G_n)} \mathbb{1}_{\{|\mathcal{C}_v|=1\}} \\ &\geq \frac{1}{n} \sum_{v \in V(G_n)} \mathbb{1}_{\{|B_k^-(v) \cap B_k^+(v)|=1, |\partial B_k^-(v)| |\partial B_k^+(v)|=0\}} \\ &\xrightarrow{\mathbb{P}} \mu(|B_k^-(o) \cap B_k^+(o)| = 1, |\partial B_k^-(o)| |\partial B_k^+(o)| = 0). \end{aligned} \quad (5.18)$$

If the local limit is a directed tree, then $\mu(|B_k^-(o) \cap B_k^+(o)| = 1) = 1$, since otherwise there will be a directed cycle in the local limit graph. Thus, we can write

$$\begin{aligned} \mu(|B_k^-(o) \cap B_k^+(o)| = 1, |\partial B_k^-(o)| |\partial B_k^+(o)| = 0) \\ = \mu(|\partial B_k^-(o)| |\partial B_k^+(o)| = 0). \end{aligned} \quad (5.19)$$

Also, we know that, as $k \rightarrow \infty$,

$$\mu(|\partial B_k^-(o)| |\partial B_k^+(o)| = 0) \rightarrow \mu(\{|\mathcal{C}_o^-| < \infty\} \cup \{|\mathcal{C}_o^+| < \infty\}) = 1 - \zeta. \quad (5.20)$$

This completes the proof. \square

Proof of Theorem 1.6. By the definition of $\alpha_1(n)$ in Definition 5.1,

$$\frac{K_n}{n} \geq \alpha_1(n), \quad (5.21)$$

Also,

$$K_n \leq 1 + n - |\mathcal{C}_{\max}|. \quad (5.22)$$

From (5.21) and (5.22),

$$\alpha_1(n) \leq \frac{K_n}{n} \leq 1 - \frac{|\mathcal{C}_{\max}|}{n} + \frac{1}{n}. \quad (5.23)$$

Lemma 5.2 and Theorem 1.2 then complete the proof. \square

6. DISCUSSION

6.1. Discussion results and literature. The main results in this paper concern the proportion of vertices in the LSCC and the number of connected components. We prove that these properties are ‘almost local’. Why are these connectivity properties important? Apart from indicating the connectivity of the graph, they are also used in the comparisons of centrality measures (see, e.g., [14]). This investigates how much a graph breaks down in terms of its connectivity, when we remove the most central nodes. We next list some open problems of our work.

In Theorem 1.5, we prove that (1.9) is a necessary and sufficient condition for $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta = \mu(|\mathcal{C}_o^+| = |\mathcal{C}_o^-| = \infty)$ to hold. Of course, $\zeta \geq \zeta' \equiv \mu(|\mathcal{C}_o| = \infty)$. What is a necessary and sufficient condition for $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta'$ to hold? These conditions should really be different, since, for example, for locally tree-like digraphs, $\zeta' = 0$, while $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta = \mu(|\mathcal{C}_o^+| = |\mathcal{C}_o^-| = \infty) > 0$ can hold.

Currently, to describe the number of strongly connected components in Theorem 1.5, we rely on the necessary and sufficient condition (1.9) for the proportion of vertices in the giant being equal to $\zeta = \mu(|\mathcal{C}_o^+| = |\mathcal{C}_o^-| = \infty)$. What is the necessary and sufficient condition for the convergence $K_n/n \xrightarrow{\mathbb{P}} \alpha$ to hold, and is it possible that α takes a different form than that on the right-hand side of (1.16)?

6.2. Large weak components. In this section, we discuss weak giants. The definition of a strong component in a digraph is well accepted. In contrast, the notion of a weak component has been used differently in the literature, as we discuss next.

In- and out-components. One could consider weak components to be in- and the out-components. For example, in [4], the authors study the largest in- and out-component sizes for the directed configuration model. We define the sizes of the largest in- and out-components \mathcal{I}_{\max} and \mathcal{O}_{\max} to be

$$\mathcal{I}_{\max} = \max_{v \in V(G_n)} |\mathcal{C}_v^-| \quad \text{and} \quad \mathcal{O}_{\max} = \max_{v \in V(G_n)} |\mathcal{C}_v^+|. \quad (6.1)$$

The strong giant having linear size implies that its in- and out-components also have linear size, but the converse is not necessarily true. We visualize this in Figure 1. The big question is whether the maximal in- and out-

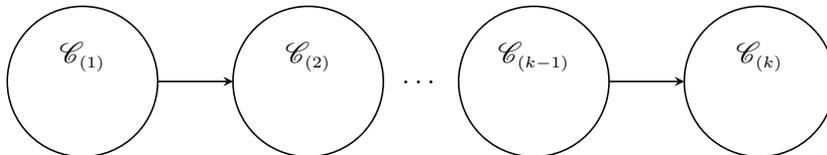


FIGURE 1. A linear-sized giant in- or out-component can exist despite the fact that there is no linear-sized strong giant.

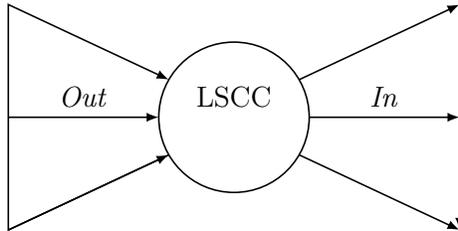


FIGURE 2. The ‘bow-tie’ figure used to describe the connectivity structure of large directed networks.

components are ‘almost’ local or not, and if so, under what condition. In particular, a natural question is to find a condition that guarantees that the proportion of vertices in \mathcal{I}_{\max} is asymptotically equal to the proportion of vertices in the out-component of \mathcal{C}_{\max} . See Figure 2, where LSCC stands for the largest strongly connected component. This means that the condition for the existence of weak giants should be weaker than the conditions given in Theorem 1.2.

For the in- and out-components, we believe it to be possible that $|\mathcal{C}_{\max}^+|/n \xrightarrow{\mathbb{P}} \alpha \leq \mu(|\mathcal{C}_o^-| = \infty)$ and vice-versa for the in-component. We state the following open problem:

Open Problem 6.1. *Find the necessary and sufficient conditions for*

$$\frac{|\mathcal{C}_{\max}^+|}{n} \xrightarrow{\mathbb{P}} \alpha^+ \quad \text{and} \quad \frac{|\mathcal{C}_{\max}^-|}{n} \xrightarrow{\mathbb{P}} \alpha^- \tag{6.2}$$

to hold.

When $|\mathcal{C}_{\max}^+|/n \xrightarrow{\mathbb{P}} \alpha^+$, then one can easily deduce that $\alpha^+ \leq \mu(|\mathcal{C}_o^-| = \infty)$. It is not clear when $\alpha^+ = \mu(|\mathcal{C}_o^-| = \infty)$ holds. One difficulty to prove such a result is that, unlike the strongly connected components, the in- and out-components do not *partition* the vertex set.

Ignoring the directions of edges. One can also consider weak components to consist of the set of vertices that can be reached by recursively following all edges regardless of their orientations [5, 13]. This corresponds to the connected components in the underlying undirected graph, i.e., the graph obtained after removing the edge directions. These types of components find applications in real-world problems, including epidemiology, data mining, communication networks, world wide web structure, etc. In this case, the results in [8] can be used to give a necessary and sufficient condition for the proportion of vertices in the giant to be the survival probability of the local limit.

Weak components as a relation. The 1972 paper by Graham, Knuth, and Motzkin [12, 7] introduced a rather different notion of weak components. Formally, it can be defined by the following 4 symmetric relations on the vertex set of a digraph:

- (1) For any two vertices u and v in $V(G)$, $u \iff v$ if both vertices are in the same strongly connected component, i.e., $\max\{d_G(u, v), d_G(v, u)\} < \infty$. In other words, there exists a path from v to u and vice versa.
- (2) For any two vertices u and v in $V(G)$, $u \parallel v$ if $\min\{d_G(u, v), d_G(v, u)\} = \infty$, i.e., there is no path from v to u and from u to v .
- (3) For any two vertices u and v in $V(G)$, $u \approx v$ if either $u \iff v$ or $u \parallel v$.
- (4) We define R to be the transitive closure of \approx . This means that if there is a chain $(u = v_0, v_1, \dots, v_k = v)$ such that $u \approx v_1 \approx \dots \approx v_{k-1} \approx v$, then uRv .

The relation R is an equivalence relation and partitions the vertex set into the weak components. This concept of weak component is not a local quantity as a small perturbation in the graph can easily dramatically change the components in the graph. In particular, the addition of *one* isolated vertex can change the definition of all these weak components. This suggests that, in terms of local convergence, this notion is too sensitive to local changes. Yet, Pacault [15] gives an algorithm for computing weak components.

REFERENCES

1. David Aldous and J. Michael Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. MR 2023650
2. Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. MR 1873300
3. Xing Shi Cai and Guillem Perarnau, *The diameter of the directed configuration model*, Ann. Inst. Henri Poincaré Probab. Stat. **59** (2023), no. 1, 244–270. MR 4533728
4. Colin Cooper and Alan Frieze, *The size of the largest strongly connected component of a random digraph with a given degree sequence*, Combin. Probab. Comput. **13** (2004), no. 3, 319–337. MR 2056402
5. Matthew Coulson and Guillem Perarnau, *Weak components of the directed configuration model*, Extended Abstracts EuroComb 2021: European Conference on Combinatorics, Graph Theory and Applications, Springer, 2021, pp. 682–687.
6. Alessandro Garavaglia, Remco van der Hofstad, and Nelly Litvak, *Local weak convergence for PageRank*, Ann. Appl. Probab. **30** (2020), no. 1, 40–79. MR 4068306
7. Ronald L Graham, Donald E Knuth, and TS Motzkin, *Complements and transitive closures.*, Discret. Math. **2** (1972), no. 1, 17–29.
8. Remco van der Hofstad, *The giant in random graphs is almost local*, arXiv:2103.11733 [math.PR], 2021.

9. Remco van der Hofstad, *Random graphs and complex networks. Vol. 2*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2024. MR 3617364
10. Pim van der Hoorn and M. Olvera-Cravioto, *Typical distances in the directed configuration model*, *Ann. Appl. Probab.* **28** (2018), no. 3, 1739–1792. MR 3809476
11. Femke van Ieperen and Ivan Kryven, *Percolation in simple directed random graphs with a given degree distribution*, *Probability in the Engineering and Informational Sciences* (2023), 1–22.
12. Donald E. Knuth, *The art of computer programming*, *Sorting and searching* **3** (1999), no. 426-458, 4.
13. Ivan Kryven, *Emergence of the giant weak component in directed random graphs with arbitrary degree distributions*, *Physical Review E* **94** (2016), no. 1, 012315.
14. Decebal Constantin Mocanu, Georgios Exarchakos, and Antonio Liotta, *Decentralized dynamic understanding of hidden relations in complex networks*, *Scientific Reports* **8** (2018), no. 1, 1571.
15. Jean Francois Pacault, *Computing the weak components of a directed graph*, *SIAM Journal on Computing* **3** (1974), no. 1, 56–61.