Spreading Processes and Large Components in Ordered, Directed Random Graphs

Paul Horn[*] Malik Magdon-Ismail [†]

September 21, 2012

Abstract

Order the vertices of a directed random graph $v_1, \dots, v_n$; edge (v_i, v_j) for $i < j$ exists independently with probability p. This random graph model is related to certain spreading processes on networks. We consider the component reachable from v_1 and prove existence of a sharp threshold $p^* = \log n/n$ at which this reachable component transitions from $o(n)$ to $\Omega(n)$.

1 Introduction

In this note we study a random graph model that captures the dynamics of a particular type of spreading process. Consider a set of n ordered vertices $\{v_1, \dots, v_n\}$ with vertex v_1 initially 'infiltrated' (at time step 1). At time steps $2, 3, \dots, n$, vertex v_1 attempts to independently infiltrate, with probability p, each of $v_2, v_3 \dots, v_n$ in turn (one per step). Either v_i gets infiltrated or immunized. If v_i is infected, it attempts to infect $v_{i+1}, \dots, v_n$, also each with probability p; v_i does not attempt to infect $v_1, \dots, v_{i-1}$, however, as prior vertices are already either infiltrated or immunized. At time step i, all infiltrated vertices v_j with $j < i$ are attempting to infiltrate v_i, and v_i gets infiltrated if any one of these attempts succeeds. Intuitively, v_i is more likely to get infiltrated if more vertices are already infiltrated at the time that v_i becomes 'succeptible'. One example of such a contagion process is given in [**?**].

This spreading process is equivalent to the following random model of an ordered, directed graph G: order the vertices $v_1, \dots, v_n$, and for $i < j$, the directed edge (v_i, v_j) exists in G with probability p (independently). Vertex v_i is infected if there is a (directed) path from v_1 to v_i. The question we address is, "What is the size of the set of vertices reachable from v_1?" (the size of the infection). We prove the following sharp result.

Theorem 1. *Let $\mathcal{R}$ be the set of vertices reachable from v_1, and suppose $p = \frac{c \log n}{n} + \xi(n)$, where $\xi(n) = o(\frac{\log n}{n})$ and $c > 0$ is fixed. Then:*

1. *If $c < 1$, then $|\mathcal{R}| = n^{c+o(1)}$, a.a.s.*

[*]Department of Mathematics, Harvard University, `phorn@math.harvard.edu`
[†]Department of Computer Science, Rensselaer Polytechnic Institute, `magdon@cs.rpi.edu`

2. *If* $c = 1$, *then* $|\mathcal{R}| = o(n)$, *a.a.s.*

3. *If* $c > 1$, *then* $|\mathcal{R}| = \left(1 - \frac{1}{c} + o(1)\right) n$, *a.a.s.*

Recall that an event holds a.a.s.(asymptotically almost surely), if it holds with probability $1 - o(1)$; that is it holds with probability tending to one as n tends to infinity. Note that we do not explicitly care whether $\xi(n)$ is positive or negative in the results above.

Similar phase transitions are well known for various graph properties in other random graph models. As shown by Erdős and Rényi in [2], in the $G(n, M)$ model of random graphs, where a graph is chosen independently from all graphs with M edges, there is a similar emergence of a component of size $\Theta(n)$ around $M = \frac{n}{2}$ edges. Likewise, a threshold for connectivity was shown for $M = \frac{n \log n}{2}$ edges. For the more familiar $G(n, p)$ model, where edges are present independenty with probability p, this translates into a threshold at $p = \frac{1}{n}$ for a giant component, and at $p = \frac{\log n}{n}$ for connectivity. A much more comprehensive account of results on properties of random graphs can be found in [1]. Łuczak in [4] and more recently Łuczak and Seierstad in [5], studied the emergence of the giant component in a random directed graphs, in both the directed model where M random edges are present and in the model where edges are present with probability p. Thresholds for strong connectivity were established for random directed graphs by Palásti [6] (for random directed graphs with M edges) and Graham and Pike [3] (for random directed graphs with edge probability p). We are not aware of any results for ordered directed random graphs where edges connect vertices of lower index to higher index.

2 A Proof of Theorem 1

Upper bounds: For $i > 1$, let $\mathcal{R}_i$ denote the event that v_i is reachable, and let X_i denote the number of paths to vertex v_i in G. If $\mathcal{P}_i$ denotes the set of all potential paths from v_1 to v_i, then $X_i = \sum_{x \in \mathcal{P}_i} I(x)$ where $I(x)$ is a $\{0, 1\}$ indicator random variable indicating whether the path x exists in G; $I(x) = 1$ if and only if all edges in the path x are present in G. Then,

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_i) = \mathbb{P}(X_i \geq 1) \leq \mathbb{E}[X_i] &= \sum_{x \in \mathcal{P}_i} \mathbb{E}[I(x)] \\
&= \sum_{\ell=0}^{i-2} \sum_{\substack{x \in \mathcal{P}_i \\ |x| = \ell+1}} \mathbb{E}[I(x)] \\
&= \sum_{\ell=0}^{i-2} \binom{i-2}{\ell} p^{\ell+1} = p(1+p)^{i-2} \leq p e^{pi}.
\end{aligned}$$

Let X denote the number of reachable vertices (other than v_1).

$$\mathbb{E}[X] = \sum_{i=2}^{n} \mathbb{P}(\mathcal{R}_i) \leq \sum_{i=1}^{n} p e^{pi} = p \cdot \frac{e^{p(n+1)} - 1}{e^p - 1}.$$

For $p = \frac{c \log n}{n} + \xi(n)$ with $c < 1$,

$$e^{p(n+1)} - 1 = \exp\left(c \log n + o(\log n)\right) - 1 = n^{c+o(1)},$$

and

$$\frac{p}{(e^p-1)} = \left(\sum_{k=1}^{\infty} \frac{p^{k-1}}{k!}\right)^{-1} = 1 + O(p).$$

Thus,

$$\mathbb{E}[X] \le n^{c+o(1)}.$$

Applying Markov's inequality yields that $\mathbb{P}(X > \log(n)\mathbb{E}[X]) = o(1)$, so $X \le \log(n)\mathbb{E}[X] = n^{c+o(1)}$, a.a.s.

Now consider $c > 1$. Let

$$\begin{aligned}
\xi'(n) &:= \frac{3}{c}\max\left\{\frac{n}{\log\log n}, \frac{n^2\xi(n)}{c\log n}\right\} \\
t &:= \frac{n}{c} - \xi'(n)
\end{aligned}$$

Note that by our choice of $\xi'(n)$, and the fact that $\xi(n) = o(\frac{\log n}{n})$, that $\xi'(n) = o(n)$. Then,

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_t) \le pe^{pt} &= p\exp\left(\left(c\frac{\log n}{n} + \xi(n)\right)\left(\frac{n}{c} - \xi'(n)\right)\right) \\
&= p\exp\left(\log(n) + \frac{n\xi(n)}{c} - \frac{c(\log n)\xi'(n)}{n} - \xi(n)\xi'(n)\right) \\
&\le (1+o(1))c\exp\left(\log\log(n) + \frac{n\xi(n)}{c} - \frac{c(\log n)\xi'(n)}{n}\right) = o(1)
\end{aligned}$$

Here, the last inequality comes from the fact that, by our choice of $\xi'(n)$,

$$\frac{c(\log(n))\xi'(n)}{n} - \log\log(n) - \frac{n\xi(n)}{c} \ge \frac{1}{3}\xi'(n).$$

Since pe^{pi} is increasing in i, the expected number of reachable vertices v_i with $i \le t$ is at most $t\mathbb{P}(\mathcal{R}_t) = o(n)$. Applying Markov's inequality, $|\mathcal{R} \cap \{v_1, \ldots, v_t\}| = o(n)$ a.a.s. Thus,

$$|\mathcal{R}| \le n - t + |\mathcal{R} \cap \{v_1, \ldots, v_t\}| = \left(1 - \tfrac{1}{c} + o(1)\right) n \text{ a.a.s.}$$

For $p = \frac{\log n}{n} + \xi(n)$ with $\xi(n) = o\left(\frac{\log n}{n}\right)$, we will write $\xi(n) = \omega(n)\frac{\log n}{n}$, where $\omega(n) \to 0$. Let $t = n \cdot \left(1 - \omega(n) - \frac{1}{\log\log n}\right)$. Then,

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_t) \le pe^{-pt} &= \exp\left[(1+\omega(n))\left(1 - \omega(n) - \tfrac{1}{\log\log n}\right)\log n + \log\left((1+\omega(n))\tfrac{\log n}{n}\right)\right] \\
&= \exp\left[-\omega(n)^2\log n - (1+\omega(n))\frac{\log n}{\log\log n} + \log\log n + \log(1+\omega(n))\right] \\
&= o(1),
\end{aligned}$$

Thus the expected value of $|\mathcal{R} \cap \{v_1, \ldots, v_t\}|$ is $o(n)$ and by Markov's inequality, this is also true a.a.s. Now, since $n - t$ is also $o(n)$, we have that $R = o(n)$ a.a.s. □

To prove the lower bounds, we require a simple lemma similar to Dirichlet's theorem. Let $d(i)$ denote the number of divisors of i and let $d_t(i)$ denote the number of divisors of i that are at most t. Dirichlet's Theorem states that

$$\sum_{i=1}^{k} d(i) = k \log k + (2\gamma - 1)k + O(\sqrt{k}),$$

where γ is Euler's constant. For our purposes, we need a refinement of this result, summing $d_t(i)$.

Lemma 1. $\sum_{i=1}^{k} d_t(i) = k \log \min(t, k) + O(k).$

Proof. For $t > k$ the result follows from Dirichlet's theorem as we may replace $d_t(i)$ with $d(i)$ in the summation. For $t \leq k$,

$$\sum_{i=1}^{k} d_t(i) = k + \left\lfloor \frac{k}{2} \right\rfloor + \left\lfloor \frac{k}{3} \right\rfloor + \cdots + \left\lfloor \frac{k}{t} \right\rfloor \leq k\mathcal{H}_t,$$

where $\mathcal{H}_t$ is the t-th harmonic number. □

Lower bounds: For exposition, assume that we construct our graph on countably many vertices and that we then restrict our attention to the first n vertices. Let X_i denote the index of the i-th reachable vertex (that is not v_1). If $X_i > n$ then $|\mathcal{R}| \leq i$. Set $X_0 = 1$, and for $i \geq 1$, $X_i - X_{i-1}$ is geometrically distributed with parameter $1 - (1-p)^i$. Fix t, and consider $\mathbb{E}[X_t]$:

$$\mathbb{E}[X_t] = \sum_{k=1}^{t} \mathbb{E}[X_k - X_{k-1}] = \sum_{k=1}^{t} \frac{1}{1 - (1-p)^k}.$$

Each term is an infinite geometric series, and so

$$\mathbb{E}[X_t] = \sum_{k=1}^{t} \sum_{j=0}^{\infty} (1-p)^{kj}.$$

As this series is absolutely summable (as $\mathbb{E}[X_t]$ is clearly finite), Fubini's theorem allows us to rearrange terms in the summation to get

$$\mathbb{E}[X_t] = t + \sum_{k=1}^{t} \sum_{j=1}^{\infty} (1-p)^{kj} = t + \sum_{i=1}^{\infty} d_t(i)(1-p)^i.$$

because the term $(1-p)^i$ appears in the original summation (where $i = kj$) once for every divisor i has that is at most t. We now use summation by parts to manipulate the second term:

$$\begin{aligned}
\sum_{i=1}^{\infty} d_t(i)(1-p)^i &= p \sum_{i=1}^{\infty} (1-p)^{i-1} \left(\sum_{\ell=1}^{i} d_t(\ell) \right) \\
&= p \sum_{i=1}^{\infty} (1-p)^{i-1} (i \log(\min\{t, i\}) + O(i)) \\
&\leq p(\log t + O(1)) \sum_{i=1}^{\infty} i(1-p)^{i-1}.
\end{aligned}$$

Since $\sum_{i=1}^{\infty} i(1-p)^{i-1} = 1/p^2$, we have that

$$\mathbb{E}[X_t] = t + \frac{\log t}{p} + O\left(\frac{1}{p}\right). \tag{1}$$

Furthermore, since $X_{k+1} - X_k$ and $X_k - X_{k-1}$ are independent,

$$\begin{aligned}
\mathrm{Var}(X_t) &= \sum_{k=1}^{t} \frac{p}{(1-(1-p)^k)^2} \\
&\le \sqrt{\left(\sum_{k=1}^{t} \frac{p^2}{(1-(1-p)^k)^3}\right)\left(\sum_{k=1}^{t} \frac{1}{(1-(1-p)^k)}\right)} \\
&\le \sqrt{\frac{t}{p}\mathbb{E}[X_t]}. \qquad (2)
\end{aligned}$$

Here, the first inequality follows from an application of Cauchy-Schwarz, and the second from $\frac{p^2}{(1-(1-p)^k)^3} \le \frac{p^2}{p^3} = \frac{1}{p}$.

Now, suppose that $p = c\frac{\log n}{n} + \xi(n)$ for $c < 1$, and set $t = n^c \exp(-n|\xi(n)| - \log\log(n))$. Then, from (1),

$$\begin{aligned}
\mathbb{E}[X_t] &\le n^c \exp(-n|\xi(n)|) + \frac{c\log n - 2n|\xi(n)| - \log\log n}{n^{-1}(c\log n + n\xi(n))} + O\left(\frac{\log n}{n}\right) \qquad (3) \\
&\le n^c \exp(-n|\xi(n)|) + n - \frac{n^2|\xi(n)| - \log\log n}{(c\log n + n\xi(n))} + O\left(\frac{\log n}{n}\right) \qquad (4) \\
&= n\left(1 - \frac{n|\xi(n)| + \log\log(n)}{(c\log n + n\xi(n))} + o\left(\frac{n|\xi(n)| + \log\log n}{\log n}\right)\right). \qquad (5)
\end{aligned}$$

For n sufficiently large, $\mathbb{E}[X_t] \le n\left(1 - \frac{n|\xi(n)| + \log\log n}{2c\log n}\right)$. Meanwhile, from (2),

$$\mathrm{Var}(X_t) \le (1+o(1))\sqrt{\frac{n^c}{\log n} \cdot (1+o(1))\frac{n}{c\log n} \cdot \mathbb{E}[X_t]} = \frac{n^{\frac{1}{2}(1+c)}}{\log n}\sqrt{\frac{\mathbb{E}[X_t]}{c}} = O\left(\frac{n^{3/2}}{\log n}\right),$$

because $\mathbb{E}[X_t] = O(n)$ and $c < 1$. Chebyshev's inequality asserts that

$$\mathbb{P}\left[|X_t - \mathbb{E}[X_t]| \ge \frac{n^2|\xi(n)| + n\log\log n}{2c\log n}\right] \le \frac{4c^2\log^2 n \cdot \mathrm{Var}(X_t)}{(n^2|\xi(n)| + n(\log\log n))^2} = o(1).$$

Thus, $\mathbb{P}\left[X_t \le \mathbb{E}[X_t] + \frac{n\log\log n}{2c\log n}\right] = 1 - o(1)$. Using (5),

$$\mathbb{P}\left[X_t \le n\left(1 - \frac{\log\log n}{2c\log n} + o\left(\frac{\log\log n}{c\log n}\right)\right)\right] = 1 - o(1),$$

i.e., $X_t < n$ a.a.s. Since $X_t < n$ implies $|\mathcal{R}| \ge t$, we have that $|\mathcal{R}| > n^c \exp(-n|\xi(n)| - \log\log(n)) = n^{c+o(1)}$ a.a.s.

For $c > 1$, take $t = \frac{n\log\log n}{\log n}$. Then, using (1),

$$\mathbb{E}[X_t] \le \frac{n}{c} + o(n).$$

Again, by (2) and because $\mathbb{E}[X_t] = O(n)$, $\text{Var}(X_t) = O(n^{3/2}\sqrt{\log\log n}/\log n) = o(n^{3/2})$. Chebyschev's inequality asserts that

$$\mathbb{P}\left[|X_t - \mathbb{E}[X_t]| \geq n^{3/4}\right] \leq \frac{o(n^{3/2})}{n^{3/2}} = o(1).$$

Hence,

$$\mathbb{P}\left[X_t \leq \mathbb{E}[X_t] + n^{3/4}\right] = 1 - o(1). \tag{6}$$

So, $X_t \leq \frac{n}{c} + o(n)$ a.a.s. We now consider the vertices indexed higher than X_t and show that essentially all of them are reachable. Let Y be the vertices with index higher than X_t which are *not* adjacent to one of the first t reachable vertices in $v_1, \ldots, v_{X_t}$. Then

$$\mathbb{E}[|Y|] = \sum_{j=X_t+1}^{n} (1-p)^t = (n - X_t)(1-p)^t \leq ne^{-pt} = \frac{n}{\log^{c+o(1)} n} = o(n).$$

Applying Markov's inequality, $|Y| = o(n)$ with probability $1-o(1)$. Since the set of vertices indexed above X_t that is not reachable is a subset of Y, $|\mathcal{R}| \geq t + (n - X_t) - |Y|$. Since $|Y|, t$ are $o(n)$ and $X_t = \frac{n}{c} + o(n)$, we have that $|\mathcal{R}| \geq n(1 - \frac{1}{c} + o(1))$ with probability $1 - o(1)$, as desired. □

Acknowledgement. Magdon-Ismail acknowledges that this research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[2] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[3] A. J. Graham and D. A. Pike. A note on thresholds and connectivity in random directed graphs. *Atl. Electron. J. Math.*, 3(1):1–5, 2008.

[4] T. Łuczak. The phase transition in the evolution of random digraphs. *J. Graph Theory*, 14(2):217–223, 1990.

[5] T. Łuczak and T. G. Seierstad. The critical behavior of random digraphs. *Random Structures Algorithms*, 35(3):271–293, 2009.

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