A phase transition in Barak-Erdős random graphs

Gilles Blanchard[*] & Nicolas Curien[†] & Klara Krause[‡] & Alexander Reisach[§]

Abstract

We study monotone paths in Erdős-Rényi random graphs on numbered vertices. This model appeared in [BT22] where Benjamini & Tzalik established a phase transition at $p = \frac{\log n}{n}$. We refine the critical value to $p = \frac{\log n - \log\log n}{n}$ and identify the critical window of order $\Theta(1/n)$.

1 Introduction

Recall the definition of a Barak-Erdős random graph G_p on the positive integers $\mathbb{N}$: A directed edge (i, j) between any two vertices $i < j \in \mathbb{N}$ exists independently with probability $p \in (0, 1)$. We denote this law by $\mathbb{P}_p$. We say that there is a monotone path from 1 to n, write $1 \nearrow n$, if there exists a path $1 = i_1 \to i_2 \to \cdots \to i_k = n$ in G_p whith $i_1 < i_2 < \cdots < i_k$. We establish a phase transition for the increasing event

$$\{1 \nearrow n \text{ in } G_p\}.$$

Theorem 1.1 (Critical window)**.** *The critical window of the event $\{1 \nearrow n$ in $G_{p_n}\}$ is of order $\Theta(1/n)$ around $\frac{\log n - \log\log n}{n}$. More precisely, if $x \in \mathbb{R}$*

$$\text{for } p_{n,x} = \frac{\log n - \log\log n + x}{n} \qquad \text{we have} \quad \mathbb{P}_{p_{n,x}}(1 \nearrow n) \xrightarrow[n\to\infty]{} 1 - e^{e^{-x}-x}\int_{e^{-x}}^{\infty} \frac{e^{-t}}{t}\mathrm{d}t.$$

Remark 1.2. An alternative expression is

$$\mathbb{P}_{p_{n,x}}(1 \nearrow n) \xrightarrow[n\to\infty]{} \frac{1}{2} + \frac{1}{2}\mathbb{E}\left(\tanh\frac{x - \mathfrak{G}}{2}\right)$$

for a standard Gumbel variable $\mathfrak{G}$ with density $p_{\mathfrak{G}}(z) = \exp(-(z + e^{-z}))$.

The location and width of the phase transition might be guessed using the expected number of monotone paths in G_p. Indeed, a simple counting exercise shows that

$$\mathbb{E}_p(\#\text{ paths } 1 \nearrow n) = \sum_{k=0}^{n-2}\binom{n-2}{k}p^{k+1} = p(1+p)^{n-2}$$

and for $p \equiv p_{n,x}$ the above display converges to e^x as $n \to \infty$.

[*]Université Paris-Saclay. gilles.blanchard@universite-paris-saclay.fr
[†]Université Paris-Saclay. nicolas.curien@gmail.com
[‡]Université Paris-Saclay. klara.krause@universite-paris-saclay.fr
[§]Université Paris-Saclay. alexander.reisach@universite-paris-saclay.fr

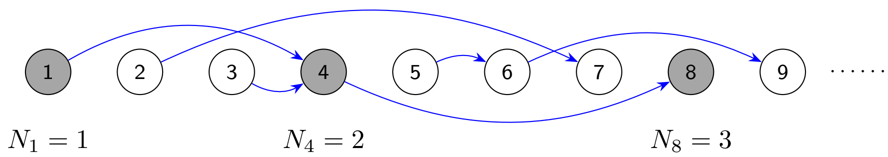

Figure 1: A segment from a sampled graph G_p satisfying $1 \nearrow 4$ and $1 \nearrow 8$.

Erdős-Rényi graphs on numbered vertices $\{1, \ldots, n\}$ appeared as an ingredient to recover point configurations in [BT22]. They prove a sharp threshold at $p = \frac{\log n}{n}$ for the event $1 \nearrow n$. Directed acyclic Erdös-Rényi graphs were introduced in [BE84] and studied as Barak-Erdős graphs since. The articles [FK03], [MR21] and [Fos+24] established maximal path lengths in Barak-Erdős graphs, both for fixed $p \in (0, 1)$ and $p \to 0$.
Numbered vertices can be interpreted as time stamps: each vertex becomes available from its assigned time onward. Related models include random temporal graphs, where monotonicity of paths is considered along edges instead of vertices, see for instance [Ang+18], [Cas+24] and [BKL24]. These works focus on the length of monotone paths and identify several transition phenomena around $p = c\frac{\log n}{n}$ for both longest and shortest such paths.
A vertex order appears naturally if the nodes correspond to events unfolding in time, as for example in graphical models of causality [Pea09]. Erdős-Rényi graphs directed according to a random vertex order are of particular interest to research in causal discovery [SGS01], where they are widely used in the benchmarking of algorithms that learn causal graphs from data.

2 Proof

The central ingredient to prove Theorem 1.1 is the introduction of an exploration process. It indicates the position of having passed a certain number of vertices that are reachable from 1 by a monotone path. We show that, after suitable centering and scaling, this exploration process converges to a Gumbel variable with a deterministic drift in Section 2.1. This allows to compute explicitly the probability that a monotone path $1 \nearrow n$ exists.

Define

$$N_k \coloneqq \# \{1 \leq i \leq k : 1 \nearrow i\}$$

the number of vertices up to k that are reachable from 1 by a monotone path and set $\mathcal{F}_k = \sigma(N_1, \ldots, N_k)$. A new vertex k is reachable from 1 precisely when there is an edge to at least one of the N_{k-1} previously reachable vertices. Hence

$$\mathbb{P}_p\left(N_k = N_{k-1} + 1 | \mathcal{F}_{k-1}\right) = 1 - \mathbb{P}_p\left(N_k = N_{k-1} | \mathcal{F}_{k-1}\right) = 1 - (1-p)^{N_{k-1}}$$

and

$$\mathbb{P}_p(1 \nearrow k) = \mathbb{E}\left(1 - (1-p)^{N_{k-1}}\right). \tag{2.1}$$

The probability that successive gaps between reachable vertices have sizes $k_1, \ldots, k_l \geq 1$ is

$$\mathbb{P}_p\left(\overset{1}{\bullet} \underbrace{\cdots}_{k_1} \bullet \underbrace{\cdots}_{k_2} \bullet \cdots\cdots \bullet \underbrace{\cdots}_{k_l} \bullet\right) = \prod_{i=1}^{l} ((1-p)^i)^{k_i} \left(1 - (1-p)^i\right)$$

where each large bullet marks a vertex reachable from 1 and each group of small dots represents the vertices in the corresponding gap: The i-th gap is a geometric variable with parameter $1 - (1-p)^i$, and the gaps are pairwise independent.

2.1 Exploration process

Fix $p \in (0,1)$ and consider the monotone exploration process $(P_p(a))_{a>0}$ where

$$P_p(a) \coloneqq \inf\{k \geq 1 : N_k = \lfloor a/p \rfloor\}$$

denotes the position of the $\lfloor a/p \rfloor$-th vertex reachable from 1 by a monotone path. Since $P_p(a)$ is obtained by summing the successive gap lengths, we have

$$P_p(a) = \sum_{i=1}^{\lfloor a/p \rfloor} X_i, \quad X_i \sim \text{Geom}\left(1-(1-p)^i\right) \text{ independent.}$$

The exploration process $(P_p(a))_{a>0}$ centered and rescaled, converges, as $p \to 0$, to a standard Gumbel variable $\mathfrak{G}$ with a deterministic drift depending on a.

Proposition 2.1 (Convergence of the monotone exploration process)**.** *We have*

$$\left(pP_p(a) - \log\frac{1}{p}\right)_{a>0} \xrightarrow[p\to 0]{(d)} (\mathfrak{G} + \log(e^a - 1))_{a>0}$$

for the uniform convergence over every compact subset of $(0,\infty)$.

Proof of Proposition 2.1. Define $q_i \coloneqq 1-(1-p)^i$. Assume without loss of generality that $p < 1/2$. For any $p < 1/2$ and $i \leq \lfloor a/p \rfloor$ it holds $q_i \leq 1-(1-p)^{\lfloor a/p \rfloor} \leq B(a) < 1$ for some constant $B(a)$ independent of p. Each $P_p(a)$ is the sum of independent geometric variables with expectation $1/q_i$.
We first approximate the rescaled geometric variables q_iX_i by i.i.d. exponential variables via a coupling: Let $(U_i)_i$ be i.i.d. uniform variables on $[0,1]$. Set

$$Y_i = -\log U_i \text{ and } X_i = \inf\left\{k \geq 1 : U_i \geq (1-q_i)^k\right\}$$

so that Y_i are exponential variables with intensity 1 and X_i are geometric variables with mean $1/q_i$. Then $\log\left(\frac{1}{1-q_i}\right)(X_i - 1) \leq Y_i$ which allows the bound

$$|q_iX_i - Y_i| \leq \left|\frac{q_i}{\log(1/(1-q_i))} - 1\right| Y_i + q_i \leq C(a)q_i(Y_i+1) \tag{2.2}$$

for some constant $C(a) > 0$, using that $q_i < B(a) < 1$.

In a second step, we express the exploration process using the exponential variables Y_i and show that the occurring error has converging expectation and vanishing variance, that is

$$pP_p(a) - \log\frac{1}{p} \overset{(d)}{=} \sum_{i=1}^{\lfloor a/p \rfloor} \frac{1}{i} Y_i - \log\frac{1}{p} + R_p(a)$$

with rest term

$$R_p(a) = \sum_{i=1}^{\lfloor a/p \rfloor} p\left(\frac{1}{q_i} - \frac{1}{ip}\right) Y_i + \sum_{i=1}^{\lfloor a/p \rfloor} \frac{p}{q_i}(q_iX_i - Y_i).$$

Lemma 2.2. *As* $p \to 0$*, the rest term satisfies*

$$\mathbb{E}(R_p(a)) = \log(e^a - 1) - \log a + o(1) \text{ and } \operatorname{Var}(R_p(a)) = o(1).$$

In a last step, we use the classical fact that the sum of the rescaled exponential variables, properly centered, converges in law to a standard Gumbel variable,

$$\sum_{i=1}^{\lfloor a/p \rfloor} \frac{1}{i} Y_i - \log \frac{1}{p} \xrightarrow[p \to 0]{(d)} \mathfrak{G} + \log a, \tag{2.3}$$

see e.g. [Cur25, Lemma 11.2]. Therefore

$$pP_p(a) - \log \frac{1}{p} \xrightarrow[p \to 0]{(d)} \mathfrak{G} + \log(e^a - 1)$$

for any fixed $a > 0$. In fact, this is the same Gumbel variable for any two $a_1, a_2 > 0$.
We now conclude the convergence as a process in $a > 0$ via a probabilistic version of Dini's lemma, see [Cur25, Theorem 11.6]: The process

$$\left(pP_p(a) - \log \frac{1}{p} - \mathfrak{G} \right)_{a > 0}$$

is increasing in a and for every fixed $a > 0$ it converges in law to $\log(e^a - 1)$ which is an increasing continuous function. Therefore Dini's lemma implies the convergence as a process

$$\left(pP_p(a) - \log \frac{1}{p} - \mathfrak{G} \right)_{a > 0} \xrightarrow[p \to 0]{(d)} (\log(e^a - 1))_{a > 0}$$

for the topology of uniform convergence over every compact subset of $(0, \infty)$. □

It remains to prove Lemma 2.2.

Proof of Lemma 2.2. For the expectation notice that $\mathbb{E}(q_i X_i) = 1 = \mathbb{E}(Y_i)$ so that

$$\begin{aligned}
\mathbb{E}(R_p(a)) = \mathbb{E}\left(\sum_{i=1}^{\lfloor a/p \rfloor} p \left(\frac{1}{q_i} - \frac{1}{ip} \right) Y_i \right) &= p \sum_{i=1}^{\lfloor a/p \rfloor} \frac{1}{q_i} - \frac{1}{ip} \\
&= p \sum_{i=1}^{\lfloor a/p \rfloor} \left(\frac{1}{1 - e^{-ip}} - \frac{1}{ip} \right) + p \sum_{i=1}^{\lfloor a/p \rfloor} \left(\frac{1}{q_i} - \frac{1}{1 - e^{-ip}} \right) \\
&\leq \int_0^a \left(\frac{1}{1 - e^{-s}} - \frac{1}{s} \right) \mathrm{d}s + p \sum_{i=1}^{\lfloor a/p \rfloor} \frac{C(a)}{i} \\
&= \log(e^a - 1) - \log a + o(1)
\end{aligned}$$

for some constant $C(a) > 0$ depending on a. For the variance write

$$Z_i = \left(\frac{1}{q_i} - \frac{1}{ip} \right) Y_i + \frac{1}{q_i} (q_i X_i - Y_i) \leq c_1(a) Y_i + c_2(a)$$

for some constants $c_1(a) > 0$ and $c_2(a) > 0$, using the bound $\left(\frac{1}{q_i} - \frac{1}{ip} \right) \leq \frac{a}{2(1 - e^{-a})}$ for $i \leq \lfloor a/p \rfloor$ and (2.2). Then, as $Y_i \sim \mathrm{Exp}(1)$ for all i, this yields

$$\mathrm{Var}(Z_i) \leq \mathbb{E}\left(Z_i^2 \right) = \mathcal{O}(1).$$

Independence of the Z_i thus yields

$$\mathrm{Var}\left(R_p(a) \right) = p^2 \sum_{i=1}^{\lfloor a/p \rfloor} \mathrm{Var}(Z_i) = o(1).$$

□

2.2 Critical window

We now choose p dependent on n and let $n \to \infty$. More precisely, let $b > 0$ and let $p = p_n$ satisfy

$$n = \frac{\log(b/p_n)}{p_n}. \tag{2.4}$$

Proof of Theorem 1.1. For each $n \in \mathbb{N}$ define the random variable $\mathfrak{X}_n$ as the number of vertices smaller than n that are reachable from 1 by a monotone path, scaled by p_n, that is

$$P_{p_n}(\mathfrak{X}_n) = n \quad \text{with} \quad \lfloor \mathfrak{X}_n/p_n \rfloor = N_n.$$

Then, by (2.1)

$$\mathbb{P}_{p_n}(1 \nearrow n) = \mathbb{E}\left(1 - (1-p_n)^{\lfloor \mathfrak{X}_n/p_n \rfloor - 1}\right) = \mathbb{E}\left(1 - e^{-\mathfrak{X}_n}(1 + o(1))\right). \tag{2.5}$$

On the other hand, by the convergence in law from Proposition 2.1 and the Skorokhod representation theorem there is a probability space with a variable $\tilde{\mathfrak{G}}$ and a process $\tilde{P}_{p_n}$ inducing variables $\tilde{\mathfrak{X}}_n$ with the same law as $\mathfrak{G}$, P_{p_n} and $\mathfrak{X}_n$ respectively such that almost-sure-convergence

$$np_n - \log\frac{1}{p_n} - \log\left(e^{\tilde{\mathfrak{X}}_n} - 1\right) \xrightarrow[n\to\infty]{} \tilde{\mathfrak{G}} \quad \text{a.s.}$$

holds. Therefore

$$1 - e^{-\tilde{\mathfrak{X}}_n} \xrightarrow[n\to\infty]{} 1 - \frac{1}{1 + be^{-\tilde{\mathfrak{G}}}} \quad \text{a.s.} \tag{2.6}$$

and this convergence holds in law for $\mathfrak{X}$ and $\mathfrak{G}$. Combine (2.5) and (2.6) to compute

$$\mathbb{P}_{p_n}(1 \nearrow n) = 1 - \mathbb{E}\left(\frac{1}{1 + be^{-\mathfrak{G}}}\right) + o(1) = 1 - \frac{1}{b}e^{1/b}\int_{1/b}^{\infty}\frac{e^{-t}}{t}\mathrm{d}t + o(1). \tag{2.7}$$

The function $f(b) = \frac{1}{b}e^{1/b}\int_{1/b}^{\infty}\frac{1}{t}e^{-t}\mathrm{d}t$ is continuous on $(0,\infty)$ with asymptotics

$$f(b) \xrightarrow{b\to\infty} 0 \quad \text{and} \quad f(b) \xrightarrow{b\to 0} 1.$$

So for any $b > 0$ the choice p_n in (2.4) yields a non-trivial probability for $\{1 \nearrow n\}$. It remains to determine the matching value $b = b_x$ when choosing $p_{n,x}$ as in the theorem. From (2.4) we have

$$b_x = p_{n,x}e^{np_{n,x}} = \left(\frac{\log n - \log\log n}{n} + \frac{x}{n}\right)\frac{n}{\log n}e^x = \left(1 + \frac{x - \log\log n}{\log n}\right)e^x \sim e^x.$$

Plug this into equation (2.7). □

Remark 2.3 (Number of paths)**.** Denote by M_a^+ the number of paths $1 \nearrow P_p(a)$, conditioned to be positive. We conjecture that, as $p \to 0$, M_a^+ converges to a geometric random variable with parameter e^{-a}. Heuristically, it satisfies the recursive relation

$$M_a^+ \overset{(d)}{=} \sum_{i=1}^{W} M_{aU_i}^+$$

where W is a $\mathrm{Poisson}(a)$ random variable conditioned to be positive, and $(U_i)_{i\geq 1}$ are i.i.d $\mathrm{Unif}[0,1]$ variables.

References

[Ang+18] Omer Angel et al. "Long Monotone Trails in Random Edge-Labellings of Random Graphs". In: *Combinatorics, Probability and Computing* 29 (2018), pp. 22–30.

[BE84] Amnon B Barak and Paul Erdös. "On the maximal number of strongly independent vertices in a random acyclic directed graph". In: *SIAM Journal on Algebraic Discrete Methods* 5.4 (1984), pp. 508–514.

[BT22] Itai Benjamini and Elad Tzalik. "Determining a Points Configuration on the Line from a Subset of the Pairwise Distances". In: (Aug. 2022). DOI: 10.48550/arXiv.2208.13855.

[BKL24] Nicolas Broutin, Nina Kamčev and Gábor Lugosi. "Increasing paths in random temporal graphs". In: *Annals of Applied Probability* 34.6 (Dec. 2024), pp. 5498–5521. DOI: 10.1214/24-AAP2097.

[Cas+24] Arnaud Casteigts et al. "Sharp Thresholds in Random Simple Temporal Graphs". In: *SIAM Journal on Computing* 53.2 (2024), pp. 346–388. DOI: 10.1137/22M1511916.

[Cur25] Nicolas Curien. *A Random Walk Among Random Graphs*. Vol. 31. Cours Spécialisés. Paris: Société Mathématique de France, 2025.

[FK03] Sergey Foss and Takis Konstantopoulos. "Extended renovation theory and limit theorems for stochastic ordered graphs". In: *Markov Process. Related Fields* 9.3 (2003), pp. 413–468.

[Fos+24] Sergey Foss et al. "Last passage percolation and limit theorems in Barak-Erdős directed random graphs and related models". In: *Probab. Surv.* 21 (2024), pp. 67–170. DOI: 10.1214/24-PS28.

[MR21] Bastien Mallein and Sanjay Ramassamy. "Barak–Erdős graphs and the infinite-bin model". In: *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*. Vol. 57. 4. Institut Henri Poincaré. 2021, pp. 1940–1967.

[Pea09] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

[SGS01] Peter Spirtes, Clark Glymour and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2001.