On the length of the shortest path in a sparse Barak-Erdős graph

Bastien Mallein* Pavel Tesemnikov†

December 29, 2021

Abstract

We consider an inhomogeneous version of the Barak-Erdős graph, i.e. a directed Erős-Rényi random graph on $\{1, \ldots, n\}$ with no loop. Given f a Riemann-integrable non-negative function on $[0, 1]^2$ and $\gamma > 0$, we define $G(n, f, \gamma)$ as the random graph with vertex set $\{1, \ldots, n\}$ such that for each $i < j$ the directed edge (i, j) is present with probability $p_{i,j}^{(n)} = \frac{f(i/n, j/n)}{n^\gamma}$, independently of any other edge. We denote by L_n the length of the shortest path between vertices 1 and n, and take interest in the asymptotic behaviour of L_n as $n \to \infty$.

1 Introduction

The Barak-Erdős graph is a random directed graph with no loop constructed in the following fashion. Given $n \in \mathbb{N}$ and $p \in (0, 1)$, the Barak-Erős graph $G(n, p)$ is a graph with vertex set $\{1, \ldots, n\}$ such that for each $i < j$, the edge (i, j) from vertex i to vertex j is present with probability p, independently of any other directed edge. This graph is a directed acyclic version of the well-known Erdős-Rényi graph. It can be used to model community food webs in ecology [10], or the task graph for parallel processing in computer sciences [6].

In particular, the length (number of edges) of the longest (directed) path, denoted M_n, has been the subject of multiple studies, as $M_n + 1$ is the number of steps needed to complete the task graph assuming maximal parallelization. Newman [9] proved that $\frac{M_n}{n}$ converges in law to a deterministic function $p \mapsto C(p)$. Increasingly precise bounds were obtained on this function and its generalizations by [3, 2, 4, 7, 8, 5].

In the present article, we take interest in the length L_n of the *shortest* path between vertices 1 and n in this graph, which has been much less studied. It is worth noting that for fixed value of p, one has

$$\mathbf{P}(L_n = 1) = p \quad \text{and} \quad \lim_{n\to\infty} \mathbf{P}(L_n = 2) = 1 - p,$$

*LAGA UMR 7539, Université Sorbonne Paris Nord, and DMA, UMR 8553, École Normale Supérieure, Email: mallein@math.univ-paris13.fr

†Novosibirsk State University, Sobolev Institute of Mathematics and MCA, Email: tesemnikov.p@gmail.com

as with probability $1-(1-p^2)^{n-2}$, there is a vertex $j \in \{2, \dots n-1\}$ connect to both 1 and n, hence L_n is equal to 1 or 2 with high probability. In particular, the length of the shortest path in dense graphs remains tight.

This fact is mentioned in [12], which takes interest in the asymptotic behaviour of $L_n^{(\gamma)}$, the length of the shortest path between vertices 1 and n in a graph with connexion constant $p_n = n^{-\gamma}$, i.e. in the limit of sparse graphs. There, it is shown that for all $k \geq 2$,

$$\lim_{n\to\infty} \mathbf{P}(L_n^{(\gamma)} \leq k) = 0 \quad \text{if } 1 - \frac{1}{k} < \gamma.$$

We extend this result in the present article by obtaining the convergence in distribution of $L_n^{(\gamma)}$ for all $\gamma \in (0,1)$.

We consider here the asymptotic behaviour of the length of the shortest path between 1 and n in a time-inhomogeneous version of the Barak-Erdős graph, defined as follows. Let f be a Riemann-integrable positive function on $[0,1]^2$ and $\gamma \in (0,1)$. For each $n \in \mathbb{N}$ and $i < j$, we set $p_{i,j}^{(n)} = \frac{f(i/n,j/n)}{n^\gamma}$. The time-inhomogeneous sparse Barak-Erdős graph $G(n,f,\gamma)$ is defined as a graph with vertex set $\{1,\dots,n\}$ such that for each $i<j$, the directed edge (i,j) is present with probability $p_{i,j}^{(n)}$.

The main result of the article is the following estimate on the asymptotic behaviour of L_n for $\gamma = 1 - \frac{1}{k}$.

Theorem 1.1. *Let f be a Riemann-integrable non-negative function on $[0,1]^2$ and $k \in \mathbb{N}$. We fix $\gamma = 1 - \frac{1}{k}$ and we set*

$$c_k(f) = \int_{0<u_1<\ldots<u_{k-1}<1} \prod_{j=0}^{k-1} f(u_j, u_{j+1}) \mathrm{d}u_1 \cdots \mathrm{d}u_{k-1} \in [0,\infty],$$

with $u_0 = 0$ and $u_1 = 1$. Writing L_n for the length of the shortest path in $G(n,f,\gamma)$, we have

$$\lim_{n\to\infty} \mathbf{P}(L_n = k+1) = 1 - \mathbf{P}(L_n = k) = \exp\left(-c_k(f)\right).$$

By coupling arguments, Theorem 1.1 can be extended to describe the convergence in distribution of L_n as $n \to \infty$ for any time-inhomogeneous Barak-Erdős graph.

Corollary 1.2. *Let f be a Riemann-integrable positive function on $[0,1]^2$ and $\gamma > 0$. For $k \geq 2$, we have*

$$\lim_{n\to\infty} \mathbf{P}(L_n \leq k) = \begin{cases} 0 & \text{if } k < \frac{1}{1-\gamma}, \\ e^{-c_k(f)} & \text{if } k = \frac{1}{1-\gamma}, \\ 1 & \text{otherwise.} \end{cases}$$

Observe that for $\gamma = 1$, the Barak-Erdős graph becomes unconnected, so that $L_n = \infty$ with positive probability. In the present article, we do not treat the case $np_n \to \infty$ with $n^{1-\varepsilon} p_n \to 0$ for all $\varepsilon > 0$. However, a phase transition should be observed for the asymptotic behaviour of L_n when $p_n \approx \frac{\log n}{n}$, as the graph becomes disconnected.

1.1 Some examples and applications

A class of inhomogeneous Barak-Erdős graphs previously studied strongly inhomogeneous graph. In this class of graphs, the probability of presence of the edge (i,j) is given by $\theta\frac{(j-i)^\alpha}{n^\beta}$, with $\theta > 0$, $\beta > 0$ and $\alpha \in (-1,\beta)$. This model can be constructed as an inhomogeneous Barak-Erdős graph, setting $f(x,y) = \theta(y-x)^\alpha$ and $\gamma = \beta - \alpha$. Applying Corollary 1.2, we observe that for any $k \geq 2$, if $1 - \frac{1}{k-1} < \beta - \alpha < 1 - \frac{1}{k}$, we have $L_n \to k$ in probability. Additionally, if $\beta - \alpha = 1 - \frac{1}{k}$, we set

$$c = \theta^k \int_{S_k} \prod_{j=1}^{k} t_j^\alpha \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1} = \frac{\theta^k \Gamma(1+\alpha)^k}{\Gamma(k(1+\alpha))},$$

where $S_k = \{(t_1, \dots t_k) : t_1 + \cdots + t_k = 1\}$. We conclude by Theorem 1.1 that L_n converges in distribution to $e^{-c}\delta_k + (1-e^{-c})\delta_{k+1}$ as $n \to \infty$.

Remark that using the coupling given in Proposition 2.3, for a similar model with $\alpha \leq -1$, we can obtain

$$\lim_{n\to\infty} L_n = k \text{ in probability if } \beta - \alpha = \left[1 - \frac{1}{k-1}, 1 - \frac{1}{k}\right).$$

This result is an extension of Tesemnikov's [12] estimates on the length of the shortest path in the inhomogeneous Barak-Erdős graph, setting $\beta = 0$. Outside of the boundary cases, the convergence in probability of L_n to $k \in \mathbb{N}$ can be obtained through first- and second-moment methods, using see Lemma 2.1. We handle the boundary cases by using the Chen-Stein method, showing in Lemma 2.2 that the law of the number of paths of length k is close to a Poisson distribution for n large enough.

An other example of interest in the case of Barak-Erdős graphs with exponential density of connexion. Setting $\lambda, \mu \in \mathbb{R}$ and $\gamma > 0$, we consider a Barak-Erdős graph with $p_{i,j}^{(n)} = \frac{e^{\lambda(j-i)/n+\mu}}{n^\gamma}$. In this case, we have

$$c_k(f) = \int_{0<u_1<\ldots<u_{k-1}<1} \prod_{j=0}^{k-1} e^{\lambda(u_{j+1}-u_j)+\mu} \mathrm{d}u_1 \cdots \mathrm{d}u_{k-1} = \frac{e^{\lambda+k\mu}}{k!},$$

and Theorem 1.1 and Corollary 1.2 apply.

2 Proof of the main result

For each $k \leq n$, we denote by $\Gamma_k(n) = \{\rho \in \mathbb{N}^{n+1} : \rho_0 = 1 < \rho_1 < \cdots < \rho_{k-1} < \rho_k = n\}$ the set of increasing paths from 1 to n. As a first step towards estimating the length of the shortest in a time-inhomogeneous Barak-Erdős graph, we compute the mean number of paths of length k.

Lemma 2.1. *Let $\gamma > 0$ and f a Riemann-integrable non-negative function. For $n \in \mathbb{N}$, we write $Z_n(k)$ for the number of paths of length k between 1 and n in $G(n, f, \gamma)$, we have*

$$\mathbf{E}(Z_n(k)) \sim n^{(k-1)-k\gamma} c_k(f).$$

Proof. By linearity of the expectation, we have

$$\begin{aligned}\mathbf{E}(Z_n(k)) &= \sum_{\rho\in\Gamma_k(n)}\prod_{j=0}^{k-1} p^{(n)}_{\rho_j,\rho_{j+1}}\\ &= n^{(k-1)-k\gamma}\frac{1}{n^{k-1}}\sum_{\rho\in\Gamma_k(n)}\prod_{j=0}^{k-1} f\left(\tfrac{\rho_j}{n},\tfrac{\rho_{j+1}}{n}\right).\end{aligned}$$

Then $\lim_{n\to\infty}\frac{1}{n^{k-1}}\sum_{\rho\in\Gamma_k(n)}\prod_{j=0}^{k-1} f\left(\frac{\rho_j}{n},\frac{\rho_{j+1}}{n}\right) = c_k(f)$, by Riemann integration, which completes the proof. □

In particular, we remark that under the assumptions of Theorem 1.1, the mean number of paths of length k in $G(n,f,\gamma)$ converges to $c_k(f)$. Using this observation, we now prove that the number of paths of length k converges to a Poisson variable.

Lemma 2.2. *With the notation and assumptions of Theorem 1.1, we have*

$$\lim_{n\to\infty} Z_n(k) = \mathcal{P}\left(c_k(f)\right) \qquad \textit{in distribution.}$$

Proof. We use the Chen-Stein method [1, 11] to prove the convergence in distribution of $Z_n(k)$. More precisely, we show that for all $j\in\mathbb{N}$ we have

$$\lim_{n\to\infty} j\mathbf{P}(Z_n(k)=j) - \mathbf{E}(Z_n(k))\mathbf{P}(Z_n(k)=j-1) = 0. \tag{2.1}$$

Together with a tightness argument (due to the fact that $\mathbf{E}(Z_n(k))$ converges), it proves that $Z_n(k)$ converges in distribution to a Poisson variable with parameter $\lim_{n\to\infty}\mathbf{E}(Z_n(k)) = c_k(f)$.

Let $j\in\mathbb{N}$, we rewrite

$$j\mathbf{P}(Z_n(k)=j) = \mathbf{E}\left(\sum_{\rho\in\Gamma_k(n)} \mathbf{1}_{\{\rho \text{ open}\}}\mathbf{1}_{\{Z_n(k)=j\}}\right), \tag{2.2}$$

where ρ is said to be open if all edges (ρ_i,ρ_{i+1}) are present in the graph. Moreover for all $\rho\in\Gamma_k(n)$, we have

$$\begin{aligned}&|\mathbf{P}(Z_n(k)=j|\rho \text{ open}) - \mathbf{P}(Z_n(k)=j-1)|\\ &\quad\le \mathbf{P}\left(\text{exists a path of length } k \text{ between 1 and } n \text{ sharing an edge with } \rho\right).\end{aligned}$$

Indeed, to construct a graph with same law as $G(n,f,\gamma)$ conditionally on ρ being open, it is enough to add to the graph $G(n,f,\gamma)$ the edges $(\rho_j,\ldots\rho_{j+1})$ for all $1\le j\le n$ if these are not already present. If opening these edges creates new paths, then these path would have to share at least one edge with ρ.

We remark that if there exists a path of length k between 1 and n, there exists $0\le i<j\le k$ and $2\le \ell<k$ such that there exists a path of length ℓ between ρ_i and ρ_j that does not intersect ρ. Writing $Y_{i,j,\ell}$ the number of such paths, with the same method as in Lemma 2.1, we compute

$$\begin{aligned}\mathbf{E}\left(Y_{i,j,\ell}\right) &= \sum_{\rho_i<\overline{\rho}_1<\cdots<\overline{\rho}_{\ell-1}<\rho_j}\prod_{q=0}^{\ell-1} p^{(n)}_{\overline{\rho}_q,\overline{\rho}_{q+1}}\\ &\le n^{-\gamma\ell}\sum_{\overline{\rho}\in\Gamma_k(\ell)}\prod_{j=0}^{\ell-1} f\left(\tfrac{\overline{\rho}_q}{n},\tfrac{\overline{\rho}_{q+1}}{n}\right)\to 0 \text{ as } n\to\infty.\end{aligned}$$

Therefore, by union bound, we deduce that

$$\lim_{n\to\infty} \mathbf{P}(Z_n(k) = j | \rho \text{ open}) - \mathbf{P}(Z_n(k) = j-1) = 0,$$

which then yields by (2.2)

$$\sum_{\rho \in \Gamma_k(n)} \mathbf{P}(Z_n(k) = j \text{ and } \rho \text{ open}) - \mathbf{P}(Z_n(k) = j-1)\, \mathbf{E}(Z_n(k)) = o(\mathbf{E}(Z_n(k))).$$

As $\mathbf{E}(Z_n(k))$ is bounded, we obtain (2.1).

We remark that $\sup_{n\in\mathbb{N}} \mathbf{E}(Z_n(k)) < \infty$, hence $(Z_n(k))$ is tight. Consider any subsequence (n_j) so that $Z_{n_j}(k)$ converges in distribution as $j \to \infty$. Writing Y a random variable with this distribution, we have for all $j \in \mathbb{N}$:

$$j\mathbf{P}(Y = j) = c_k(f)\mathbf{P}(Y = j-1),$$

using that $\mathbf{E}(Z_{n_j}(k)) \to c_k(f)$. Hence $\mathbf{P}(Y = j) = \frac{c_k(f)^j}{j!}\mathbf{P}(Y = 0)$, with $\mathbf{P}(Y > n) \to 0$ as $n \to \infty$. We conclude that Y is a $\mathcal{P}(c_k(f))$ random variable.

As any converging subsequence of $(Z_n(k))$ is converging to $\mathcal{P}(c_k(f))$ in law, we conclude that $Z_n(k)$ converges to $\mathcal{P}(c_k(f))$ in law as $n \to \infty$. □

Before turning to the corollary, we introduce the following coupling estimate, which loosely states that a more connected graph will have a shorter shortest path between 1 and n.

Proposition 2.3. *Let G_n, $\overline{G}_n$ be two inhomogeneous Barak-Erdős graphs such that an edge between i and j is present with probability $p_{i,j}^{(n)}$ and $\overline{p}_{i,j}^{(n)}$ respectively. If $p_{i,j}^{(n)} \leq \overline{p}_{i,j}^{(n)}$ for any i and j, then there exists a coupling between G_n and $\overline{G}_n$ such that $L_n \geq \overline{L}_n$.*

Proof. We assume $\overline{G}_n$ to be constructed on some probability space. Take any existing edge (i, j) of $\overline{G}_n$ and do the following procedure: chosen edge is stayed in graph with probability $p_{i,j}^{(n)}/\overline{p}_{i,j}^{(n)}$ and deleted with remained probability. This procedure creates a random graph distributed exactly as G_n and is a subgraph of $\overline{G}_n$. Therefore, as no new edge was added, the length of the shortest path cannot have decrease. □

Proof of Corollary 1.2. We assume first that $k < \frac{1}{1-\gamma}$. Then, by Lemma 2.1, we have

$$\lim_{n\to\infty} \sum_{j=1}^{k} \mathbf{E}(Z_n(j)) = 0,$$

therefore $\mathbf{P}(L_n \leq k) \to 0$ by Markov inequality.

The case $k = \frac{1}{1-\gamma}$ is covered by Theorem 1.1.

Finally, if $k > \frac{1}{1-\gamma}$, then for all $A > 0$, the Barak-Erdős graph $G(n, f, \gamma)$ can be coupled with $G(n, Af, \frac{k-1}{k})$ for n large enough, using Proposition 2.3. Therefore

$$\liminf_{n\to\infty} \mathbf{P}(L_n \leq k) \geq 1 - e^{-A^k c_k(f)},$$

using Theorem 1.1 and that $c_k(Af) = A^k c_k(f)$. As f is positive, $c_k(f)$ is positive, and letting $A \to \infty$ we conclude that $\mathbf{P}(L_n \leq k) \to 1$. □

Acknowledgements. The research of both authors was supported by joint grant (19-51-15001) of the Russian Foundation for Basic Research and from the French CNRS' PRC grant.

References

[1] L. H. Y. Chen. Poisson Approximation for Dependent Trials. *The Annals of Probability*, 3(3):534 – 545, 1975.

[2] K. Chernysh and S. Ramassamy. Coupling any number of balls in the infinite-bin model. *J. Appl. Probab.*, 54(2):540–549, 2017.

[3] S. Foss and T. Konstantopoulos. Extended renovation theory and limit theorems for stochastic ordered graphs. *Markov Process. Relat. Fields*, 9(3):413–468, 2003.

[4] S. Foss and T. Konstantopoulos. Limiting properties of random graph models with vertex and edge weights. *J. Stat. Phys.*, 173(3-4):626–643, 2018.

[5] S. Foss, T. Konstantopoulos, B. Mallein, and S. Ramassamy. Estimation of the last passage percolation constant in a charged complete directed acyclic graph via perfect simulation. arXiv:2110.01559, 2021.

[6] E. Gelenbe, R. Nelson, T. Philips, and A. Tantawi. An Approximation of the Processing Time for a Random Graph Model of Parallel Computation. In *Proceedings of 1986 ACM Fall Joint Computer Conference*, ACM '86, page 691–697, Washington, DC, USA, 1986. IEEE Computer Society Press.

[7] B. Mallein and S. Ramassamy. Two-sided infinite-bin models and analyticity for Barak-Erdős graphs. *Bernoulli*, 25(4B):3479–3495, 2019.

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

[9] C. M. Newman. Chain lengths in certain random directed graphs. *Random Struct. Algorithms*, 3(3):243–253, 1992.

[10] C. M. Newman and J. E. Cohen. A Stochastic Theory of Community Food Webs IV: Theory of Food Chain Lengths in Large Webs. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 228(1252):355–377, 1986.

[11] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970, 2, 583-602 (1972)., 1972.

[12] P. Tesemnikov. On the asymptotics for the minimal distance between extreme vertices in a generalised Barak-Erdős graph. *Sib. Èlektron. Mat. Izv.*, 15:1556–1565, 2018.