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

Bastien Mallein, Pavel Tesemnikov

To cite this version:

Bastien Mallein, Pavel Tesemnikov. On the length of the shortest path in a sparse Barak-Erdős graph. Statistics and Probability Letters, 2022, 190, pp.ID 109634. hal-03514819v2
On the length of the shortest path in a sparse Barak-Erdős graph

Bastien Mallein*  Pavel Tesemnikov†

July 4, 2022

Abstract

We consider an inhomogeneous version of the Barak-Erdős graph, i.e. a directed Erdős-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 \( p^{(\gamma)}_{ij} = \frac{f(i/n, j/n)}{n^\gamma} \), 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, introduced by Barak and Erdős in [1], is a random directed graph with no loop constructed as follows. Given \( n \in \mathbb{N} \) and \( p \in (0, 1) \), the Barak-Erdős 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, introduced by Erdős and Rényi in [4]. It can be used to model community food webs in ecology [14], or the task graph for parallel processing in computer sciences [8].

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 [15] proved that \( M_n \) converges in law to a deterministic function \( p \mapsto C(p) \). Increasingly precise bounds were obtained on this function and its generalizations by [5, 3, 6, 12, 13, 7].

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

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

*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 \(1 - (1 - \rho^2)^n - 2\), there is a vertex \(j \in \{2, \ldots n-1\}\) connected 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 [17], 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 \(k \geq 2\),

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

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 space-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 space-inhomogeneous sparse Barak-Erdős graph \(G(n,f,\gamma)\) is defined as a graph with vertex set \(\{1, \ldots , n\}\) such that for each \(i < j\), the directed edge \((i,j)\) is present with probability \(p_{i,j}^{(n)}\).

We remark that for \(\gamma = 0\), the sequence \(G(n,f,0)\) is a sequence of graphs converging to a graphon (a scaling limit of random graphs introduced by Lovász and Szegedy [11]). More generally, putting a mass \(n^\gamma\) to each edge of the graph \(G(n,f,\gamma)\), the sequence of weighted graphs converges to the graphon \(f\). We refer to [10] for definitions and results linked to graphons, and to [9] for a short self-contained introduction.

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

\[
c_k(f) = \int_{0 < u_1 < \ldots < u_{k-1} < 1} \prod_{j=0}^{k-1} f(u_j, u_{j+1}) \, du_1 \cdots du_{k-1} \in [0, \infty],
\]

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

\[
\lim_{n \to \infty} \Pr(L_n = k + 1) = 1 - \Pr(L_n = k) = \exp(-c_k(f)).
\]

By coupling arguments, Theorem 1.1 can be extended to describe the convergence in distribution of \(L_n\) as \(n \to \infty\) for any space-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

\[
\lim_{n \to \infty} \Pr(L_n \leq k) = \begin{cases} 0 & \text{if } k < \frac{1}{1-\gamma}, \\ 1 - e^{-c_k(f)} & \text{if } k = \frac{1}{1-\gamma}, \\ 1 & \text{otherwise}. \end{cases}
\]
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 are strongly inhomogeneous graphs. In this class of graphs, the probability of presence of the edge \((i, j)\) is given by \( \theta \frac{j-i}{n^\gamma} \), 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. Similarly, if \( \beta - \alpha = 1 - \frac{1}{k} \), we set \( c_k = \theta \frac{k!}{\Gamma(k(1+\alpha))} \), where \( S_k = \{(t_1, \ldots, t_k) \in [0, 1]^k : t_1 + \cdots + t_k = 1\} \). We conclude by Theorem 1.1 that \( L_n \) converges in distribution to \( e^{-c_k}k^{\delta_k} + (1 - e^{-c_k})k^{\delta_k+1} \) 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 [17] 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, with using 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.

Another 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 \( p_{i,j}^{(n)} = e^{\lambda(i-j)/n+\mu} \). In this case, we have

\[
c_k(f) = \int_{0<u_1<\cdots<u_{k-1}<1} \prod_{j=0}^{k-1} e^{\lambda(u_{j+1}-u_j)+\mu} du_1 \cdots du_{k-1} = \frac{e^{\lambda+k\mu}}{(k-1)!},
\]

and Theorem 1.1 and Corollary 1.2 apply.

2 Proof of the main result

For each \( k \leq n \), we denote by

\[
\Gamma_k(n) = \{\rho \in \mathbb{N}^{k+1} : \rho_0 = 1 < \rho_1 < \cdots < \rho_{k-1} < \rho_k = n\}
\]
the set of increasing paths of length $k$ from 1 to $n$. As a first step towards estimating the length of the shortest path in a space-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

$$E(Z_n(k)) \sim n^{(k-1) - k\gamma} c_k(f) \quad \text{as } n \to \infty.$$ 

**Proof.** By linearity of the expectation, we have

$$E(Z_n(k)) = \sum_{\rho \in \Gamma_k(n)} \prod_{j=0}^{k-1} f\left(\frac{\rho_j}{n}, \frac{\rho_{j+1}}{n}\right).$$

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-distributed random variable.

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

$$\lim_{n \to \infty} Z_n(k) = \mathcal{P}(c_k(f)) \quad \text{in distribution.}$$

**Proof.** We use the Chen-Stein method [2, 16] to prove the convergence in distribution of $Z_n(k)$. More precisely, we show that for all $j \in \mathbb{N}$ we have

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

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

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

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

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

$$|P(Z_n(k) = j|\rho \text{ open}) - P(Z_n(k) = j - 1)| \leq P\left( \text{there exists a path of length } k \text{ between 1 and } n \right),$$

where $\rho$ is distinct from $\rho$ that shares an edge with $\rho$. 

4
Indeed, to construct a graph with same law as $G(n, f, \gamma)$ conditionally on $\rho$ being open, it is enough to add to the graph $G(n, f, \gamma)$ the edges $(\rho_i, \rho_{i+1})$ for all $1 \leq i \leq k$ if these are not already present. If opening these edges creates new paths, then these paths would have to share at least one edge with $\rho$.

We remark that if there exists a path of length $k$ between 1 and $n$ distinct from $\rho$ and that shares an edge with $\rho$, then there exists $0 \leq i_1 < i_2 \leq k$ and $2 \leq \ell < k$ such that there exists a path of length $\ell$ between $\rho_{i_1}$ and $\rho_{i_2}$ that does not intersect $\rho$. Writing $Y_{i_1, i_2, \ell}$ the number of such paths, with the same method as in Lemma 2.1, we compute

$$E(Y_{i_1, i_2, \ell}) = \sum_{\rho_{i_1} < \rho_{i_2} < \cdots < \rho_{i_2+1}} \prod_{q=0}^{\ell-1} P_{n, \rho_{i_q+1}}^{(n)}$$

$$\leq n^{-\gamma}\ell \sum_{\rho_{i_1} < \rho_{i_2} < \cdots < \rho_{i_2+1}} \prod_{q=0}^{\ell-1} f\left(\frac{\rho_i}{n}, \frac{\rho_{i+1}}{n}\right) \to 0 \text{ as } n \to \infty,$$

using that there are at most $n^{\ell-1}$ paths of length $\ell$ between $\rho_{i_1}$ and $\rho_{i_2}$. Therefore, by union bound, we deduce that

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

which implies, by (2.2), that

$$\sum_{\rho \in \Gamma_2(n)} P(Z_n(k) = j \text{ and } \rho \text{ open}) - P(Z_n(k) = j - 1) E(Z_n(k)) = o(E(Z_n(k))).$$

As $E(Z_n(k))$ is bounded, we obtain (2.1).

We remark that $\sup_{n \in \mathbb{N}} E(Z_n(k)) < \infty$, 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 $j \in \mathbb{N}$:

$$jP(Y = j) = c_k(f)P(Y = j - 1),$$

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

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

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

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

therefore $P(L_n \leq k) \to 0$ 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 $G(n, Af, \frac{f}{k+1})$ for $n$ large enough, using Proposition 2.3. Therefore

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

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 $P(L_n \leq k) \to 1$. 

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] Amnon B. Barak and Paul Erdős. On the maximal number of strongly independent vertices in a random acyclic directed graph. *SIAM J. Algebraic Discrete Methods*, 5:508–514, 1984.

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

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

[4] Pál Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci., Ser. A*, 5:17–61, 1960.

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

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

[7] Sergey Foss, Takis Konstantopoulos, Bastien Mallein, and Sanjay Ramassamy. Estimation of the last passage percolation constant in a charged complete directed acyclic graph via perfect simulation, 2021.

[8] 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.
[9] Daniel Glasscock. What is . . . a graphon? Notices Am. Math. Soc., 62(1):46–48, 2015.

[10] László Lovász. Large networks and graph limits, volume 60 of Colloq. Publ., Am. Math. Soc. Providence, RI: American Mathematical Society (AMS), 2012.

[11] László Lovász and Balázs Szegedy. Limits of dense graph sequences. J. Comb. Theory, Ser. B, 96(6):933–957, 2006.

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

[13] Bastien Mallein and Sanjay Ramassamy. Barak–Erdős graphs and the infinite-bin model. Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 57(4):1940 – 1967, 2021.

[14] 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.

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

[16] Charles 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.

[17] Pavel’ Igor’evich Tesemnikov. On the asymptotics for the minimal distance between extreme vertices in a generalised Barak-Erdős graph. Sib. Élektrown. Mat. Izv., 15:1556–1565, 2018.