Erdős–Rényi Poissonized

Nicolas Curien[*]

Abstract

We introduce a variant of the Erdős–Rényi random graph where the number of vertices is random and follows a Poisson law. A very simple Markov property of the model entails that the Lukasiewicz exploration is made of *independent* Poisson increments. Using a vanilla Poisson counting process, this enables us to give very short proofs of classical results such as the phase transition for the giant component or the connectedness for the standard Erdős–Rényi model.

1 The $G_{\mathrm{Poi}}(\alpha, p)$ model and its exploration

Fix $\alpha > 0$ and let $N \sim \mathcal{P}(\alpha)$ be a random variable following a Poisson law of expectation α. We consider the random graph $G_{\mathrm{Poi}}(\alpha, p)$, which conditionally on N is made of a classical Erdős-Rényi $G(N, p)$ random graph (i.e. N vertices where all $\binom{N}{2}$ edges are independent and present with probability p) that we call the **core**, together with an infinite **stack** of vertices, which are all linked to every vertex of the $G(N, p)$ with probability p independently. *There is no edge between vertices of the stack.* See Figure 1.

Markov property. A step of *exploration* in $G_{\mathrm{Poi}}(\alpha, p)$ is the following: Fix a vertex ρ of the stack and reveal its neighbors $y_1, \dots, y_K$ with $K \geqslant 0$ inside the core. Then, see those vertices $y_1, \dots, y_K$ as new vertices of the stack, in particular erase all possible connections between vertices of the stack. The key lemma is the following:

Lemma 1 (Markov property of $G_{\mathrm{Poi}}(\alpha, p)$). *Let $K \geqslant 0$ be the number of neighbors in the core of a given vertex ρ of the stack in $G_{\mathrm{Poi}}(\alpha, p)$. Then $K \sim \mathcal{P}(\alpha p)$ and conditionally on K, the graph made after removing ρ and placing its K neighbors in the stack has law $G_{\mathrm{Poi}}(\alpha(1-p), p)$.*

Proof. Call $N' = N - K$ the remaining number of vertices after the revelation of the K neighbors of ρ in the $G(N, p)$ part. Then remark the following factorization:

$$
\begin{aligned}
\mathbb{P}(K = k \text{ and } N' = n) &= e^{-\alpha} \frac{\alpha^{n+k}}{(n+k)!} \cdot \binom{n+k}{k} p^k (1-p)^n \\
&= \left(e^{-\alpha(1-p)} \frac{(\alpha(1-p))^n}{n!} \right) \cdot \left(e^{-\alpha p} \frac{(\alpha p)^k}{k!} \right).
\end{aligned}
$$

The statement follows since conditionally on the status of the vertices (being in the stack, or in the remaining part), all possible edges are i.i.d. present with probability p. $\qquad\square$
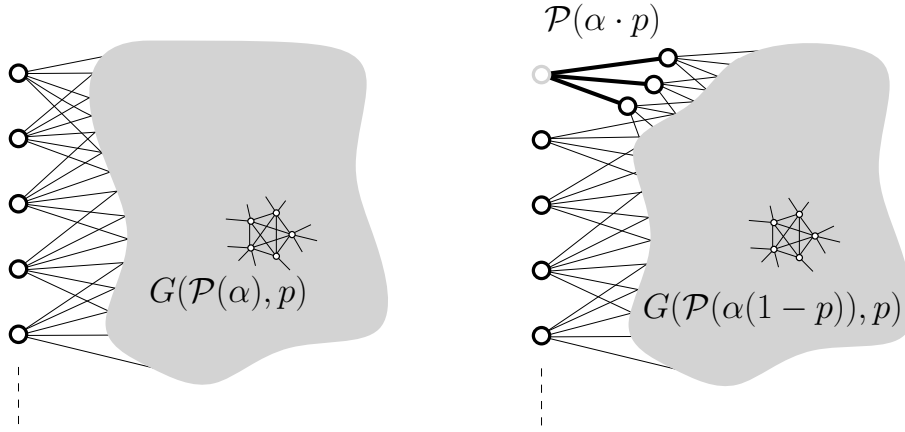
---

[*]Université Paris-Saclay.

Figure 1: Discovering the neighborhood of a vertex in the stack in the Poissonized version of the Erdős-Rényi random graph. The stack is made of the white vertices on the left part while the core is represented by the gray part.

Lukasiewicz exploration. In particular, successive explorations in $\mathsf{G}_{\mathrm{Poi}}(\alpha, p)$ yields a sequence of *independent* Poisson random variables with expectation $\alpha p, \alpha p(1-p), \ldots, \alpha p(1-p)^k, \ldots$ whose total sum is just a Poisson variable of parameter $\alpha p \sum_{i \geqslant 0}(1-p)^i = \alpha$, recovering the total number of vertices N as expected. We shall now assume that iteratively, the vertices explored are placed on top of the stack and that we always explore the first vertex of the stack: we get the so-called Lukasiewicz exploration of the graph $\mathsf{G}_{\mathrm{Poi}}(\alpha, p)$, see Figure 2. We encode it in a process $(\mathsf{S}_k : k \geqslant 0)$, the Lukasiewicz walk, defined by $\mathsf{S}_0 = 0$ and where $\Delta \mathsf{S}_k = \mathsf{S}_k - \mathsf{S}_{k-1}$ is the number of neighbors discovered at step k minus one. *Each new minimal record of S thus corresponds to the exploration of the connected component of a new vertex of the initial stack.* Using Lemma 1 we can write *simultaneously* for all $k \geqslant 0$

$$
\begin{aligned}
\mathsf{S}_k &= (\mathcal{P}(\alpha p) - 1) + (\mathcal{P}(\alpha p(1-p)) - 1) + \cdots + (\mathcal{P}(\alpha p(1-p)^{k-1}) - 1) \\
&= \mathcal{N}\left(\alpha p \cdot \sum_{i=0}^{k-1}(1-p)^i\right) - k = \mathcal{N}\left(\alpha(1 - (1-p)^k)\right) - k,
\end{aligned}
\tag{1}
$$

where all the Poisson random variables written above are independent and where $(\mathcal{N}(t) : t \geqslant 0)$ is a standard unit-rate Poisson counting process on $\mathbb{R}_+$. We shall only use the following standard estimates on the Poisson counting process

$$
\frac{\mathcal{N}(t)}{t} \xrightarrow[t \to \infty]{a.s.} 1, \quad \text{and} \quad \frac{(\mathcal{N}(tn) - tn)}{\sqrt{n}} \xrightarrow[n \to \infty]{(d)} (B_t : t \geqslant 0),
\tag{2}
$$

where $(B_t : t \geqslant 0)$ is a standard linear Brownian motion.

2 Phase transition for the giant and Aldous' critical limit

Let us use the Lukasiewicz exploration of the Poissonized version of the Erdős–Rényi random graph to give a straightforward proof of the well-known phase transition for the size of the largest connected component.

2.1 Existence of the giant component

Fix $c > 0$. Let $\alpha = n$ and $p \equiv p_n = \frac{c}{n}$ and denote by $\mathsf{S}^{(n)}$ the resulting Lukasiewicz walk to emphasize the dependence in n. Since we have $(1 - \frac{c}{n})^{[nt]} \to \mathrm{e}^{-ct}$ as $n \to \infty$ uniformly over compact time
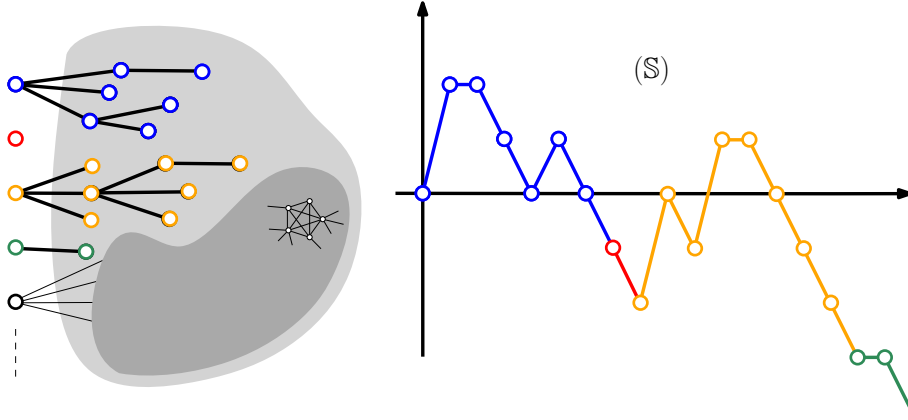
Figure 2: Lukasiewicz exploration of the graph $G_{Poi}(\alpha, p)$ obtained by concatenating the iterative number of neighbors -1.

intervals, using (1) and the law of large numbers (2) we immediately deduce:

Proposition 2 (Fluid limit). *We have the following convergence in probability for the uniform norm over every compact of $\mathbb{R}_+$:*

$$\left(n^{-1} \cdot S^{(n)}_{[nt]}\right)_{t \geqslant 0} \xrightarrow[n \to \infty]{(\mathbb{P})} \left(1 - e^{-ct} - t\right)_{t \geqslant 0}.$$
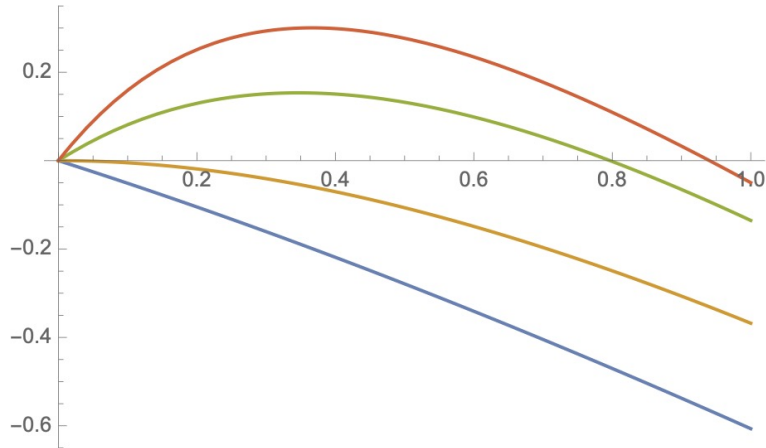

Figure 3: Graphs of the functions $(1 - e^{-ct} - t)_{t \geqslant 0}$ for different of values of c: in blue $c = 1/2$, in orange $c = 1$, in green $c = 2$ and in red $c = 3$.

We write $X_n = o_{\mathbb{P}}(n)$ if the random variable satisfies $X_n/n \to 0$ in probability.

Corollary 3 (Phase transition for $G_{Poi}(n, \frac{c}{n})$). *If $c < 1$ then asymptotically all connected components of the core of $G_{Poi}(n, \frac{c}{n})$ are $o_{\mathbb{P}}(n)$, whereas if $c > 1$ it contains a unique giant component of size $\beta(c)n + o_{\mathbb{P}}(n)$ where $\beta(c)$ is the first positive root of $1 - e^{-ct} - t = 0$, all others component sizes are $o_{\mathbb{P}}(n)$.*

Proof. The size of the connected components in $G_{Poi}(n, \frac{c}{n})$ are given by the lengths of the excursions of $S^{(n)}$ above its running infimum process $\underline{S}^{(n)}_k := \inf\{S^{(n)}_j : 0 \leqslant j \leqslant k\}$. When $c > 1$, fix $\varepsilon \in (0, \beta(c)/2)$ and let us consider

$$I^{(n)} = \inf\left\{i \leqslant \varepsilon n : S^{(n)}_i = \inf_{1 \leqslant j \leqslant \varepsilon n} S^{(n)}_j\right\} \quad \text{and} \quad J^{(n)} = \inf\left\{i \geqslant I^{(n)} : S^{(n)}_i = S^{(n)}_{I^{(n)}} - 1\right\},$$

3

so that $\mathbb{S}^{(n)}$ performs an excursion above is running infimum over the time interval $[I^{(n)}, J^{(n)}]$. By the property of the Lukasiewicz exploration, this excursion coincides with the exploration of a connected component $\mathcal{C}^{(n)}$ of the stack of size $J^{(n)} - I^{(n)}$ and the above proposition entails that $I^{(n)}/n \to 0$ and $J^{(n)}/n \to \beta(c)$ in probability as $n \to \infty$. In particular we have $|\mathcal{C}^{(n)}| = \beta(c)n + o_{\mathbb{P}}(n)$ as desired. This component is further split into $\Delta\mathbb{S}_{I^{(n)}} + 1$ components inside the core: Consider then

$$\tilde{I}^{(n)} = \inf\left\{ I^{(n)} + 1 \leqslant i \leqslant \varepsilon n : \mathbb{S}_i^{(n)} = \inf_{I^{(n)}+1 \leqslant j \leqslant \varepsilon n} \mathbb{S}_j^{(n)} \right\} \quad \text{and} \quad \tilde{J}^{(n)} = \inf\left\{ i \geqslant \tilde{I}^{(n)} : \mathbb{S}_i^{(n)} = \mathbb{S}_{\tilde{I}^{(n)}}^{(n)} - 1 \right\},$$

then the excursion time $[\tilde{I}^{(n)}, \tilde{J}^{(n)}]$ corresponds to the exploration of a subcomponent of $\mathcal{C}^{(n)}$ inside the core (after removing the vertex of the stack it contains), and similarly we deduce from the above that $\tilde{I}^{(n)}/n \to 0$ and $\tilde{J}^{(n)}/n \to \beta(c)$ in probability as $n \to \infty$, thus showing the existence of the giant component inside the core. The control of the other excursions above $\mathbb{S}^{(n)}$ (and of all excursions in the case $c < 1$) is done similarly using the convergence of Proposition 2. A slight difficulty is that that the convergence over every compact of $\mathbb{R}_+$ is not sufficient to prevent from unexpected deviations of the process $\mathbb{S}^{(n)}$ at large values k_n such that $k_n/n \to \infty$. But notice that the number of vertices left in the core after the first $[xn]$ steps of exploration has law $\mathcal{P}(n(1 - c/n)^{[xn]})$ whose expectation is $\sim ne^{-cx}$ as $n \to \infty$. By choosing x large, we can ensure that with high probability all components of size comparable to n are described by the $[xn]$ steps of $\mathbb{S}^{(n)}$. $\qquad\square$

Back to the $G(n, p)$ model. The analogous statements in the case of $G(n, p)$ can be deduced from the above results. The key idea for the depoissonization being that $\mathcal{P}(n)$ is concentrated around the value n with $\sqrt{n}$ fluctuations. Indeed, if we let $N_- \sim \mathcal{P}(n - n^{7/12})$ and $N_+ \sim \mathcal{P}(n + n^{7/12})$ then we have the natural inclusions

$$G(N_-, p) \subset G(n, p) \subset G(N_+, p), \tag{3}$$

which hold with high probability when $n \to \infty$ because $\mathbb{P}(\mathcal{P}(n - n^{7/12}) \leqslant n) \to 1$ and $\mathbb{P}(\mathcal{P}(n + n^{7/12}) \geqslant n) \to 1$ as $n \to \infty$ (notice that we only coupled the cores of G_{Poi} and not their stacks). By Corollary 3, with high probability, both random graphs on the left-hand side and right-hand side of the last display have a unique component of size $\approx \beta(c)n$, all others being of size negligible in front of n. A moment of thought enables to deduce that the same is true for the graph $G(n, p)$ sandwhiched between those two, and this is a classical result of Gilbert–Erdős–Rényi [3, 4].

2.2 Refined estimates

Let us turn to refined estimates on the clusters size still in the case $\alpha = n$ and $p \equiv p_n = \frac{c}{n}$ for $c > 0$.

CLT for the giant. In this regime, for fixed $k \geqslant 0$ we have $n(1 - (1 - \frac{c}{n})^k) \to ck$ as $n \to \infty$ so using (1) we deduce the following convergence in distribution

$$\left(\mathbb{S}_k^{(n)} : k \geqslant 0 \right) \xrightarrow[n \to \infty]{} \left(\mathcal{N}(c \cdot k) - k : k \geqslant 0 \right).$$

When $c > 1$, using (2) we easily deduce the convergence in law of the time $I^{(n)}$ defined in the course of the proof of Corollary 3 towards $\mathcal{I} := \text{argmin}\{\mathcal{N}(c \cdot k) - k : k \geqslant 0\}$. In particular, the number of connected components of G_{Poi} explored before finding the giant converges to $K_c = \mathcal{N}(c \cdot \mathcal{I}) - \mathcal{I}$ which by a simple application of Markov property is a geometric random variable (whose success

parameter is a posteriori $\beta(c)$ by Corollary 3). Similarly, the time $J^{(n)}$ can be further estimated. Put $k = [\beta(c)n + x\sqrt{n}]$ for $x \in \mathbb{R}$ in a compact interval using (2) and notice that as $n \to \infty$ we have

$$\mathsf{S}_k^{(n)} = \mathcal{N}\Big(\underbrace{n\Big(1 - (1 - \tfrac{c}{n})^{[\beta(c)n + x\sqrt{n}]}\Big)}_{=\beta(c)n + xc(1-\beta(c))\sqrt{n} + o(n^{1/2})} \Big) - k \underset{(2)}{\approx} \sqrt{n}\Big(B_{\beta(c)} + x(c(1 - \beta(c)) - 1) + o_{\mathbb{P}}(1)\Big),$$

where $(B_t : t \geqslant 0)$ is the Brownian motion appearing in (2). Combining those two arguments we easily deduce that

$$\frac{J^{(n)} - \beta(c)n}{\sqrt{n}} \xrightarrow[n\to\infty]{(d)} \frac{B_{\beta c}}{1 - c^*}, \qquad \text{where } c^* = c(1 - \beta(c)), \tag{4}$$

which together with the convergence of $I^{(n)}$ establishes the central limit theorem for the size of the giant component in $\mathsf{G}_{\mathrm{Poi}}$.

Remark. In the case of the Erdős–Rényi $G(n, \tfrac{c}{n})$ for $c > 1$ the central limit theorem of the size of the giant makes appear a variance $\beta(c)(1 - \beta(c))/(1 - c^*)^2$, see [5, 2]. The additional $(1 - \beta(c))$ factor can easily be explained if we conditioned our model to have a core of size $n + o(\sqrt{n})$. We suspect that it is possible to derive the fixed-size Erdős–Rényi case from the above fact using soft arguments.

Critical case and Aldous's limit. In the critical case $p = \tfrac{1}{n}$ we can give a analog of a result of Aldous in the case of the standard fixed-size Erdős-Rényi [1] indicating that the size of the clusters in the near critical regime is of order $n^{2/3}$:

Proposition 4 (Near critical case)**.** *Fix $\lambda \in \mathbb{R}$. For $p \equiv p_n = \tfrac{1}{n} + \tfrac{\lambda}{n^{4/3}}$ with $\lambda \in \mathbb{R}$, the Lukasiewicz walk of $\mathsf{G}_{\mathrm{Poi}}(n, p_n)$ satisfies*

$$\Big(n^{-1/3} \cdot \mathsf{S}_{[n^{2/3}t]}^{(n)}\Big)_{t \geqslant 0} \xrightarrow[n\to\infty]{(d)} \Big(B_t + \lambda t - \frac{t^2}{2}\Big)_{t \geqslant 0}.$$

Proof. Putting $k = [n^{2/3}t]$ for $t \in [0, A]$ in a compact time interval in the equation (1) yields to

$$\mathsf{S}_{[n^{2/3}t]}^{(n)} = \mathcal{N}\Big(\underbrace{n\Big(1 - (1 - \tfrac{1}{n} - \tfrac{\lambda}{n^{4/3}})^{[n^{2/3}t]}\Big)}_{=tn^{2/3} + \lambda t n^{1/3} - \frac{t^2}{2} n^{1/3} + o(n^{1/3})} \Big) - [n^{2/3}t] \underset{(2)}{\approx} n^{1/3}\Big(B_t + \lambda t - \frac{t^2}{2}\Big). \qquad \square$$

Remark. We suspect it is possible to recover the result of Aldous on the fixed-size $G(n, p)$ using the above proposition and rather soft arguments. We however did not pursue this goal in this short note.

3 Connectedness

As another application of our Poissonization technique, let us give a short proof of the sharp phase transition for connectedness in the fixed-size Erdős–Rényi [3, 4]:

Proposition 5. *For $c \in \mathbb{R}$ we have*

$$\mathbb{P}\Big(G\Big(n, \frac{\log n + c}{n}\Big) \text{ is connected} \Big) \xrightarrow[n\to\infty]{} e^{-e^{-c}}.$$

Proof. Let $p \equiv p_n = \frac{\log n + c}{n}$. We shall first prove the convergence of the proposition when the number N of vertices of the Erdős–Rényi graph is random and distributed according to one plus a Poisson

law of expectation n. Connectedness of this graph is equivalent to the fact that inside $\mathrm{G}_{\mathrm{Poi}}(n, p)$ all vertices of the stack, except the first one, have a trivial connected component. This is the case if and only if the Lukasiewicz walk $(\mathsf{S}^{(n)})$ starts with a (large) excursion and once it has reached level -1, it makes only jumps of -1. Using (1) and (2), it is easy to see that the first hitting time $\tau_{-1}^{(n)}$ of -1 by the process $\mathsf{S}^{(n)}$ is concentrated around n and more precisely using similar calculations as in (4) shows that

$$\frac{\tau_{-1}^{(n)} - n}{\sqrt{n}} \xrightarrow[n\to\infty]{(d)} B_1. \tag{5}$$

Besides, since the increments of $\mathsf{S}^{(n)}$ are Poisson and independent, by the Markov property of the exploration we have that

$$
\begin{aligned}
\mathbb{P}(\rho \cup G(N, p) \text{ is connected}) \;&=\; \mathbb{P}\left(\Delta \mathsf{S}_i^{(n)} = -1, \forall i \geqslant \tau_{-1}^{(n)}\right) = \mathbb{E}\left[\mathbb{P}(\Delta \mathsf{S}_i^{(n)} = -1, \forall i \geqslant \tau_{-1}^{(n)} \mid \tau_{-1}^{(n)})\right] \\
&=\; \mathbb{E}\left[\mathbb{P}\left(\mathcal{P}(n(1 - p_n)^{\tau_{-1}^{(n)}}) = 0\right)\right] = \mathbb{E}\left[\exp\left(-n\left(1 - \frac{\log n + c}{n}\right)^{\tau_{-1}^{(n)}}\right)\right],
\end{aligned}
$$

and by (5) the random variable inside the expectation converges in probability to $\mathrm{e}^{-\mathrm{e}^{-c}}$ since $\tau_{-1}^{(n)} = n + O_{\mathbb{P}}(\sqrt{n})$. The desired statement follows. To come back to the fixed-size $G(n, p)$ model, notice that the function $\phi(n, p) = \mathbb{P}(G(n, p) \text{ is connected})$ is increasing in p for n fixed, but the monotonicity in n is not clear. However, the natural inclusion $G(n, p) \subset G(n + 1, p)$ enables to write

$$
\begin{aligned}
\phi(n + 1, p) \;&\geqslant\; \mathbb{P}(G(n, p) \text{ is connected and the } (n+1)th \text{ vertex is connected to one of the first } n \text{ vertices}) \\
&\geqslant\; \phi(n, p) - \mathrm{P}(\mathrm{Bin}(n, p) = 0) \geqslant \phi(n, p) - \mathrm{e}^{-np}.
\end{aligned}
$$

Recalling that $p_n = \frac{\log n + c}{n}$ we deduce that if $0 \leqslant k_n = o(n)$ then we have $\phi(n + k_n, p_n) \geqslant \phi(n, p_n) + o(1)$. With the notation of (3) this shows that

$$\mathbb{P}(G(N_-, p_n) \text{ is connected}) - o(1) \leqslant \mathbb{P}(G(n, p_n) \text{ is connected}) \leqslant \mathbb{P}(G(N_+, p_n) \text{ is connected}) + o(1),$$

and by sandwhiching, the middle term does converge to $\mathrm{e}^{-\mathrm{e}^{-c}}$. $\qquad\square$

References

[1] D. Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Probab., (1997), pp. 812–854.

[2] B. Bollobás and O. Riordan, *Asymptotic normality of the size of the giant component via a random walk*, Journal of Combinatorial Theory, Series B, 102 (2012), pp. 53–61.

[3] P. Erdős and A. Rényi, *On random graphs i*, Publ. math. debrecen, 6 (1959), p. 18.

[4] P. Erdős and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci, 5 (1960), pp. 17–60.

[5] B. Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures & Algorithms, 1 (1990), pp. 311–342.