Adventures in random graphs: Models, structures and algorithms

Armand M. Makowski

ECE & ISR/HyNet
University of Maryland at College Park
armand@isr.umd.edu

LECTURE 2
Erdős-Renyi graphs $\mathbb{G}(n;p)$

A natural research program

- Absence of isolated nodes and graph connectivity
- Poisson paradigm for the number of isolated nodes and double exponential result for connectivity
- Phase transition: Evolution and size of the giant component
- Diameter of the giant component
- Subgraph containment (e.g., triangle)
- Clustering coefficient (and small world properties)

A blueprint for other classes of random graphs

Bibliography (I): Random graphs

- N. Alon and J.H. Spencer, *The Probabilistic Method* (Second Edition), Wiley-Science Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York (NY) 2000.
- A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Studies in Probability **2**, Oxford University Press, Oxford (UK), 1992.
- B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.
- R. Durrett, *Random Graph Dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge (UK), 2007.

- M. Draief and L. Massoulié, *Epidemics and Rumours in Complex Networks*, Cambridge University Press, Cambridge (UK), 2009.
- S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.

Scalings

A **scaling** is simply a mapping

$$p : \mathbb{N}_0 \rightarrow [0,1] : n \rightarrow p_n$$

Degrees!

Degree rvs

For each $i = 1, 2, \ldots, n$, the **degree** of node i in $\mathbb{G}(n; p)$ is the rv $D_{n,i}(p)$ given by

$$D_{n,i}(p) := \sum_{j=1,\ j \neq i}^{n} \chi_{ij}(p)$$

$$\mathbb{E}\left[D_{n,i}(p)\right] = (n-1)p$$

and

$$D_{n,i}(p) =_{st} \mathrm{Bin}(n-1; p)$$

Poisson convergence

For any scaling $p : \mathbb{N}_0 \to [0,1]$ such that

$$p_n \sim \frac{\lambda}{n}$$

for some $\lambda > 0$, we have the Poisson convergence

$$D_{n,i}(p_n) \Longrightarrow_n \text{Poi}(\lambda), \quad i = 1, 2, \ldots$$

In other words,

$$\lim_{n\to\infty} \mathbb{P}\left[D_{n,i}(p_n) = k\right] = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

and λ provides a natural scale!

Asymptotic mutual independence

For each $I = 1, 2, \ldots$, under the scaling above we have

$$\begin{aligned} & \lim_{n\to\infty} \mathbb{P}\left[D_{n,i}(p_n) = k_i,\ i = 1, \ldots, I\right] \\ = & \prod_{i=1}^{I} \lim_{n\to\infty} \mathbb{P}\left[D_{n,i}(p_n) = k_i\right] \\ = & \prod_{i=1}^{I} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} \end{aligned} \tag{1}$$

with arbitrary $k_1, \ldots, k_I = 0, 1, \ldots$

Easy proof

$$D_{n,i}(p) \text{ vs. } D_{n,j}(p) \quad \text{when} \quad \chi_{ij}(p) = 0$$

Subgraph containment
The case of triangles

Theorem 1 *With scaling $p : \mathbb{N}_0 \to [0,1]$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{G}(n;p_n) \text{ contains a triangle }\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} p_n n = 0 \\ & \\ 1 & \text{if } \lim_{n\to\infty} p_n n = \infty \end{cases} \tag{2}$$

Zero-one law with **critical** scaling $p^{\Delta} : \mathbb{N}_0 \to [0,1]$ given by

$$p^{\Delta}_n = \frac{1}{n}, \quad n = 1, 2, \ldots$$

Generalization available to arbitrary (unlabeled) subgraphs H

Weak zero-one law because strong separation of scales, namely

$$\begin{cases} 0 & \text{if } \lim_{n\to\infty} \frac{p_n}{p_n^\Delta} = 0 \\ \\ 1 & \text{if } \lim_{n\to\infty} \frac{p_n}{p_n^\Delta} = \infty \end{cases}$$

Absence of isolated nodes

Theorem 2 *For any scaling* $p : \mathbb{N}_0 \to [0,1]$ *with deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *determined through*

$$p_n = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots$$

we have

$$\begin{aligned} & \lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{G}(n; p_n) \textit{ contains no isolated nodes }\right] \\ & = \begin{cases} 0 & \textit{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{if } \lim_{n\to\infty} \alpha_n = +\infty \end{cases} \end{aligned} \tag{3}$$

Very weak separation of scales: **Very strong** zero-one law with **critical** scaling $p^\star : \mathbb{N}_0 \to [0,1]$ given by

$$p_n^\star = \frac{\log n}{n}, \quad n = 1, 2, \ldots$$

Intermediate separation of scales: **Strong** zero-one law with **critical** scaling $p^\star : \mathbb{N}_0 \to [0,1]$. For any scaling $p : \mathbb{N}_0 \to [0,1]$ such that

$$p_n \sim c\, p_n^\star$$

for some $c \geq 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[\, \mathbb{G}(n; p_n) \text{ contains no isolated nodes } \right] = \begin{cases} 0 & \text{if } 0 \leq c < 1 \\ & \\ 1 & \text{if } 1 < c \end{cases}$$

**Zero-one laws, counting
and the method of
first and second moments**

Basic moment inequalities

For any integrable $\mathbb{N}$-valued rv Z, we have

$$\mathbb{P}[Z > 0] \leq \mathbb{E}[Z]$$

so that

$$1 - \mathbb{E}[Z] \leq \mathbb{P}[Z = 0]$$

For any $\mathbb{N}$-valued rv Z with $\mathbb{E}\left[Z^2\right] < \infty$, we have

$$\frac{\left(\mathbb{E}[Z]\right)^2}{\mathbb{E}[Z^2]} \leq \mathbb{P}[Z > 0]$$

With sequence $\{Z_n,\ n = 1, 2, \ldots\}$

$$1 - \mathbb{E}[Z_n] \le \mathbb{P}[Z_n = 0], \quad n = 1, 2, \ldots$$

so that

$$\lim_{n\to\infty} \mathbb{P}[Z_n = 0] = 1 \quad \textbf{if} \quad \lim_{n\to\infty} \mathbb{E}[Z_n] = 0$$

$$\frac{(\mathbb{E}[Z_n])^2}{\mathbb{E}[Z_n^2]} \le \mathbb{P}[Z_n > 0], \quad n = 1, 2, \ldots$$

so that

$$\lim_{n\to\infty} \mathbb{P}[Z_n = 0] = 0 \quad \textbf{if} \quad \lim_{n\to\infty} \frac{\mathbb{E}[Z_n^2]}{(\mathbb{E}[Z_n])^2} = 1$$

Beware: With $\alpha, \beta > 0$,

$$Z_n = \begin{cases} 0 & \text{w.p. } 1 - n^{-\alpha} \\ \\ n^{\beta} & \text{w.p. } n^{-\alpha} \end{cases}$$

$$\mathbb{P}\left[Z_n = 0\right] = 1 - n^{-\alpha} \quad \text{and} \quad \mathbb{E}\left[Z_n\right] = n^{\beta - \alpha}$$

Although $\lim_{n\to\infty} \mathbb{P}\left[Z_n = 0\right] = 1$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[Z_n\right] = \lim_{n\to\infty} \left(n^{\beta-\alpha}\right) = \begin{cases} 0 & \text{if } \beta < \alpha \\ 1 & \text{if } \beta = \alpha \\ \infty & \text{if } \beta > \alpha \end{cases}$$

The method of first and second moments is often used with **counting** variables of the form

$$Z_n = \sum_{a \in A_n} I_{n,a}$$

with $\{0,1\}$-valued rvs

$$\{I_{n,a},\ a \in A_n\}.$$

$$\mathbb{E}[Z_n] = \sum_{a \in A_n} \mathbb{E}[I_{n,a}]$$

and

$$\begin{aligned}
\mathbb{E}\left[Z_n^2\right] &= \sum_{a \in A_n} \sum_{b \in A_n} \mathbb{E}[I_{n,a} I_{n,b}] \\
&= \sum_{a \in A_n} \mathbb{E}[I_{n,a}] + \sum_{a,b \in A_n:\ a \neq b} \mathbb{E}[I_{n,a} I_{n,b}] \\
&= \mathbb{E}[Z_n] + \sum_{a,b \in A_n:\ a \neq b} \mathbb{E}[I_{n,a} I_{n,b}]
\end{aligned}$$

Therefore

$$\frac{\mathbb{E}\left[Z_n^2\right]}{\left(\mathbb{E}\left[Z_n\right]\right)^2} = \frac{1}{\mathbb{E}\left[Z_n\right]} + \frac{\sum_{a,b \in A_n:\ a \neq b} \mathbb{E}\left[I_{n,a} I_{n,b}\right]}{\left(\mathbb{E}\left[Z_n\right]\right)^2}$$

One law if

$$\lim_{n\to\infty} \mathbb{E}\left[Z_n\right] = 0$$

Zero law if

$$\lim_{n\to\infty} \mathbb{E}\left[Z_n\right] = \infty$$

and

$$\lim_{n\to\infty} \frac{\sum_{a,b \in A_n:\ a \neq b} \mathbb{E}\left[I_{n,a} I_{n,b}\right]}{\left(\mathbb{E}\left[Z_n\right]\right)^2} = 1.$$

The exchangeable case

For each $n = 1, 2, \ldots$, the $\{0,1\}$-valued rvs $\{I_{n,a},\ a \in I_n\}$ are **exchangeable** rvs.

Then,

$$\mathbb{E}\left[Z_n\right] = |A_n|\mathbb{E}\left[I_{n,1}\right]$$

and

$$\sum\nolimits_{a,b \in A_n:\ a \neq b} \mathbb{E}\left[I_{n,a} I_{n,b}\right] = |A_n|\left(|A_n| - 1\right) \cdot \mathbb{E}\left[I_{n,1} I_{n,2}\right],$$

whence

$$\frac{\mathbb{E}\left[Z_n^2\right]}{\left(\mathbb{E}\left[Z_n\right]\right)^2} = \frac{1}{\mathbb{E}\left[Z_n\right]} + \frac{\left(|A_n| - 1\right)}{|A_n|} \cdot \frac{\mathbb{E}\left[I_{n,1} I_{n,2}\right]}{\left(\mathbb{E}\left[I_{n,1}\right]\right)^2}$$

With

$$\lim_{n\to\infty} |A_n| = \infty$$

One law if

$$\lim_{n\to\infty} |A_n| \cdot \mathbb{E}\left[I_{n,1}\right] = 0$$

Zero law if

$$\lim_{n\to\infty} |A_n| \cdot \mathbb{E}\left[I_{n,1}\right] = \infty$$

and

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[I_{n,1} I_{n,2}\right]}{\left(\mathbb{E}\left[I_{n,1}\right]\right)^2} = 1$$

Proof of Theorem 2

$$Z_n = \text{Number of isolated nodes in } \mathbb{G}(n; p_n)$$

with

$$I_{n,a} = \prod_{i=1,\ i\neq a}^{n} \chi_{ai}(p_n)$$

so that

$$[\ \mathbb{G}(n; p_n) \text{ contains no isolated nodes }] = [Z_n = 0]$$

$$|A_n| = n - 1$$

$$\mathbb{E}[I_{n,a}] = (1 - p_n)^{n-1}$$

and

$$\mathbb{E}[I_{n,a} I_{n,b}] = (1 - p_n)^{2n-3}$$

Proof of Theorem 1

$$Z_n = \text{Number of unlabeled triangles in } \mathbb{G}(n; p_n)$$

with

$$I_{n,a} = \mathbf{1}\,[\text{Triangle } a = (ijk) \text{ exists in } \mathbb{G}(n; p_n)\ ]$$

Identically distributed but not exchangeable!

$$A_n = \text{Collection of all unlabeled triangles on } \{1, \ldots, n\}$$

so

$$|A_n| = \binom{n}{3} \sim \frac{n^3}{6}$$

$$\mathbb{E}\left[I_{n,a}\right] = p_n^3$$

but

$$\mathbb{E}\left[I_{n,a}I_{n,b}\right] = ??$$

as it depends on the pair a, b, and easy modifications available!

More Poisson convergence in Theorem 2

Theorem 3 *For any scaling* $p : \mathbb{N}_0 \to [0,1]$ *with deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *determined through*

$$p_n = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots$$

we have

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{G}(n; p_n) \textit{ contains no isolated nodes }\right] = e^{-e^{-c}}$$

whenever $\lim_{n\to\infty} \alpha_n = c$ *for some scalar* c *in* $\mathbb{R}$.

Double exponential result – Theorem 3 implies Theorem 2!

Simple consequence of

Theorem 4 *For any scaling* $p : \mathbb{N}_0 \to [0,1]$ *with deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *determined through*

$$p_n = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots$$

we have

$$Z_n \Longrightarrow_n \text{Poi}(e^{-c})$$

whenever $\lim_{n\to\infty} \alpha_n = c$ *for some scalar* c *in* $\mathbb{R}$*.*

$$[\ \mathbb{G}(n; p_n) \text{ contains no isolated nodes }] = [Z_n = 0]$$

Connectivity

Theorem 5 *For any scaling* $p : \mathbb{N}_0 \to [0,1]$ *with deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *determined through*

$$p_n = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots$$

we have

$$\begin{aligned} & \lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{G}(n; p_n) \textit{ is connected }\right] \\ & = \begin{cases} 0 & \textit{if } \lim_{n\to\infty} \alpha_n = -\infty \\ & \\ 1 & \textit{if } \lim_{n\to\infty} \alpha_n = +\infty \end{cases} \end{aligned} \tag{4}$$

Basic facts

With $0 \leq p \leq 1$ and all $n = 1, 2, \ldots$, set

$$I_n(p) = [\ \mathbb{G}(n;p) \text{ contains no isolated nodes }\ ]$$

and

$$C_n(p) = [\ \mathbb{G}(n;p) \text{ is connected }\ ]$$

Because $C_n(p) \subseteq I_n(p)$, we note that

$$\mathbb{P}\left[C_n(p)\right] \leq \mathbb{P}\left[I_n(p)\right]$$

and

$$\mathbb{P}\left[I_n(p)\right] = \mathbb{P}\left[C_n(p)\right] + \mathbb{P}\left[I_n(p) \cap C_n(p)^c\right]$$

The ER argument for connectivity

Consider a scaling $p : \mathbb{N}_0 \to [0,1]$.

The zero law of Theorem 2 already implies the zero law of Theorem 5

Under the appropriate conditions, by the one law in Theorem 2, we have $\lim_{n\to\infty} \mathbb{P}[I_n(p_n)] = 1$ so that

$$\lim_{n\to\infty} \mathbb{P}[C_n(p_n)] = 1$$

if and only if

$$\mathbb{P}[I_n(p_n) \cap C_n(p_n)^c] = 0$$

Of connectivity and spanning trees

Union bound arguments and exchangeability yield

$$\begin{aligned}
&\mathbb{P}\left[I_n(p_n) \cap C_n(p_n)^c\right] \\
&\leq \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \sum_{S \in \mathcal{N}_r} \mathbb{P}\left[\begin{array}{c} S \text{ is a connected subgraph of } \mathbb{G}(n;p_n) \\ \text{and } S \text{ is isolated from } S^c \text{ in } \mathbb{G}(n;p_n) \end{array}\right] \\
&= \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \sum_{S \in \mathcal{N}_r} \mathbb{P}\left[\begin{array}{c} S \text{ is a connected subgraph} \\ \text{of } \mathbb{G}(n;p_n) \end{array}\right] \cdot (1-p_n)^{r(n-r)} \\
&= \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} (1-p_n)^{r(n-r)} \cdot \mathbb{P}\left[\begin{array}{c} \{1,\ldots,r\} \text{ is a connected} \\ \text{subgraph of } \mathbb{G}(n;p_n) \end{array}\right]
\end{aligned}$$

Let $\mathcal{T}_r$ denote the collection of all labelled trees on $\{1, \ldots, r\}$. By Cayley's Theorem,

$$|\mathcal{T}_r| = r^{r-2}.$$

Another union bound argument gives

$$\begin{aligned} \mathbb{P}\left[\begin{array}{c} \{1,\ldots,r\} \text{ is a connected} \\ \text{subgraph of } \mathbb{G}(n;p_n) \end{array}\right] &\leq \sum_{T \in \mathcal{T}_r} \mathbb{P}\left[T \subseteq \mathbb{G}(n;p_n)\right] \\ &= r^{r-2} p_n^{r-1} \end{aligned}$$

Combining

$$
\begin{aligned}
\mathbb{P}\left[I_n(p_n) \cap C_n(p_n)^c\right] &\le \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} (1-p_n)^{r(n-r)} \cdot r^{r-2} p_n^{r-1} \\
&\le \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} r^{r-2} \cdot p_n^{r-1} e^{-p_n r(n-r)} \\
&\le \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{en}{r}\right)^r r^{r-2} \cdot p_n^{r-1} e^{-p_n r(n-r)} \\
&= \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{r^2} (en)^r \cdot p_n^{r-1} e^{-p_n r(n-r)}
\end{aligned}
$$

since

$$(1-x) \le e^{-x}, \quad 0 \le x \le 1$$

On the diameter

Emergence of the giant component