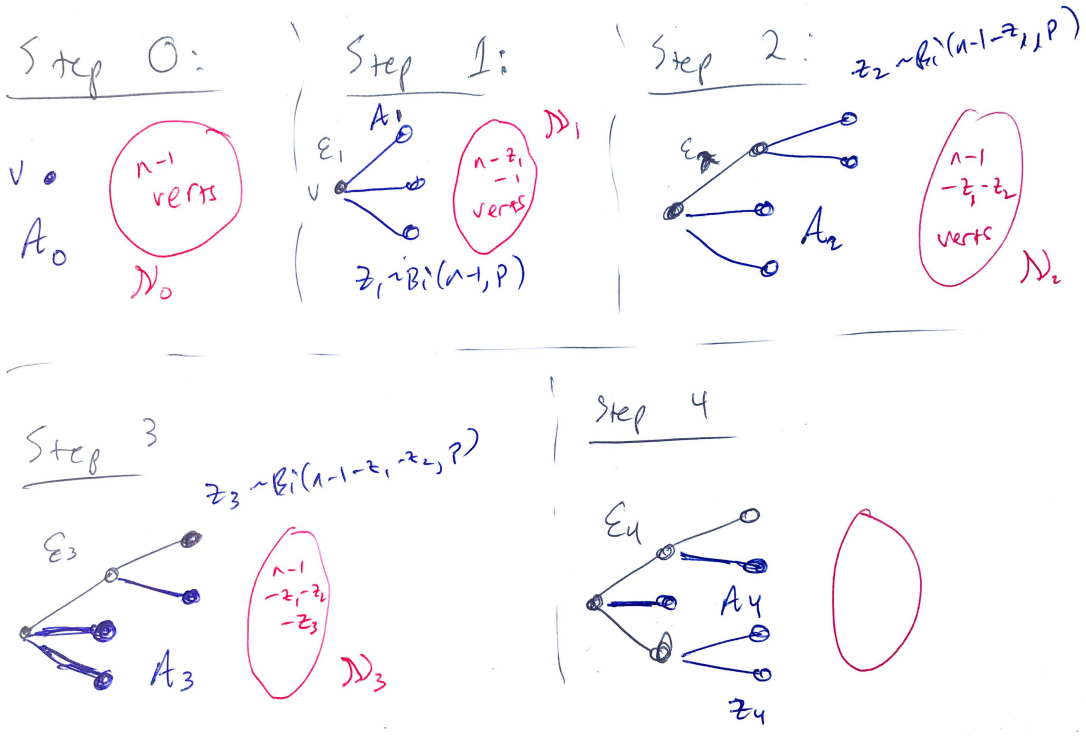
Lecture 11: Components of the Erdős-Rényi random graph: Exploration process

High-level intro for next several lectures: Our goal is to study the *component structure* of $G = G(n,p) \sim \mathrm{ER}(n,p)$. We know something about this from previous lectures: if $p \gg \log(n)/n$, then there is one component with all vertices; and in the "critical" window where $p = (\log(n) + s)/n$, there are a Poisson number of isolated vertices, and then the rest of the vertices are in one component. However, if $p \ll \log(n)/n$, then all we know is that the number of isolated vertices is large, and hence there are many components, but nothing about the sizes of these components. So this is the region we focus on, and, with some foresight for where there will be interesting behaviour, we set $p = \lambda/n$ for some fixed $\lambda > 0$.

For vertex $v \in \{1,\dots,n\}$ of $G(n,p)$, let $\mathcal{C}_v$ be the set of vertices in the component containing v, let $C_v = |\mathcal{C}_v|$ be the number of vertices in the component containing v, and let $C_{\max} = \max_{v=1,\dots,n} C_v$ be the size of the maximal component. We study $C_{\max}$ through the C_v, and will see the interesting "phase transition" that if $\lambda < 1$, then $C_{\max} = \Theta(\log(n))$, while if $\lambda > 1$, then $C_{\max} = \Theta(n)$: so as λ increases from less than 1 to greater than 1, the Erdős-Rényi random graph G goes from having many "small" components, to one "giant" component. Showing this result requires a number of new tools, such as branching process approximations, stochastic domination, and concentration.

The exploration process: We explore $\mathcal{C}_v$ sequentially, by starting at v, then uncovering its neighbours, then new neighbours of those vertices, and so on. The following figure, with notation explained below, gives an example four steps of an exploration process.



Example exploration process with $Z_1 = 3$, $Z_2 = 2$, $Z_3 = 0$, and $Z_4 = 2$. Active vertices are in black, explored in blue, and neutral in red.

The exploration process for $\mathcal{C}_v$ is a (random) sequence of partitions of the vertices of G, $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)_{t=0,1,2,\dots}$; we refer to these respectively as active, explored, and neutral. Initially,

$$(\mathcal{A}_0, \mathcal{E}_0, \mathcal{N}_0) = \big(\{v\}, \emptyset, \{1,\dots,n\} \setminus \{v\}\big).$$

Given $(\mathcal{A}_{t-1}, \mathcal{E}_{t-1}, \mathcal{N}_{t-1})$, we generate the tth step according to the following rules.

- If $\mathcal{A}_{t-1} \neq \emptyset$, choose an element of $\mathcal{A}_{t-1}$, call it a_t, and denote the set of vertices in $\mathcal{N}_{t-1}$ that are attached to a_t by

$$\mathcal{Z}_t = \{u \in \mathcal{N}_{t-1} : \{u, a_t\} \text{ is an edge in } G(n,p)\}.$$

  Then set

$$(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t) = \big(\mathcal{A}_{t-1} \setminus \{a_t\} \cup \mathcal{Z}_t,\, \mathcal{E}_{t-1} \cup \{a_t\},\, \mathcal{N}_{t-1} \setminus \mathcal{Z}_t\big).$$

  In other words, at each step, choose an active vertex, label it as explored, and label its neutral edge-neighbours as active.

- If $\mathcal{A}_{t-1} = \emptyset$, then set $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t) = (\mathcal{A}_{t-1}, \mathcal{E}_{t-1}, \mathcal{N}_{t-1})$.

See the picture above for an example exploration process and note that the only edges in the picture are those between the active vertex and *neutral* vertices at that step; there are potentially more edges in G than are given in the picture. The point of the exploration process is that the collection of active vertices over all times, or the final collection of explored vertices, is exactly that of $\mathcal{C}_v$, the component of G containing v. Moreover, the edge structure makes it so we can understand this process in detail by relating it to a *branching process*, which are well-understood objects (we will cover these in more detail soon).

To setup this relationship, we need a bit more notation. Let

$$A_t := |\mathcal{A}_t|, \quad E_t := |\mathcal{E}_t|, \quad N_t := |\mathcal{N}_t|, \quad Z_t := |\mathcal{Z}_t|.$$

Note that E_t is non-decreasing, N_t is non-increasing,

$$A_t + E_t + N_t = n,$$

and, setting

$$\tau_0 := \inf\{t \geq 0 : A_t = 0\},$$

for $t \leq \tau_0$, $E_t = t$ and $A_t + t + N_t = n$. Note also that for $t \geq \tau_0$,

$$(A_t, E_t, N_t) = (A_{\tau_0}, E_{\tau_0}, N_{\tau_0}) = (0, \tau_0, n - \tau_0).$$

We state some key properties in the following lemma.

Lemma 11.1. *In the notation and definitions above*

$$\tau_0 = C_v$$

Proof. $E_{\tau_0} = \tau_0$ is the number of vertices explored by the process, which is the same as C_v. □

The next lemma is crucial to understanding the exploration process.

Lemma 11.2. *In the notation and definitions above, the process $(A_t)_{t\geq 0}$ is a Markov chain given by setting $A_0 = 1$, and for $t \geq 1$, setting*

$$A_t = A_{t-1} + \mathbb{I}[A_{t-1} > 0](-1 + Z_t), \tag{11.1}$$

where, conditional on A_{t-1},

$$Z_t \overset{d}{=} \begin{cases} \mathrm{Bi}\big(n - (t-1) - A_{t-1}, p\big), & A_{t-1} > 0, \\ 0, & A_{t-1} = 0. \end{cases}$$

Proof. The statements under the event $A_{t-1} = 0$ are obvious. If $A_{t-1} > 0$, then the first assertion is because at the next step, we move one active vertex to explored, which accounts for the (-1) in (11.1), and add Z_t vertices from neutral to active. The second assertion is because $Z_t \sim \mathrm{Bi}(N_{t-1}, p)$, but

$$N_{t-1} = n - E_{t-1} - A_{t-1} = n - (t-1) - A_{t-1},$$

since $A_{t-1} > 0$ implies $E_{t-1} = t - 1$. □

We will see in the next lecture that a recursion similar to, but simpler than (11.1) is satisfied for an analogous quantity in branching processes.