Random graphs, social networks and the internet

Lecture 2

Mariana Olvera-Cravioto

UNC Chapel Hill

molvera@unc.edu

September 30th, 2021

Random graph models

- Some real networks are too big to be analyzed exactly.
- Some may even be constantly changing.
- **Idea:** we can think of our specific real-world graph as just one "typical" element of a larger class.
- If we can show that a property holds for a large class of graphs, it is likely it will hold for our specific graph.
- **Random graphs** are mathematical models that can help us understand large real-world graphs.
- No random graph model can mimic all the properties of a specific real-world graph, so we focus on choosing models that share certain properties that are important to the problem we want to analyze.

Large graph limit

- Random graph models consist of a vertex set $V_n = \{1, 2, \ldots, n\}$ and a set of rules for determining whether a given edge is present or not based on some random events.
- Their mathematical analysis is usually done under the **large graph limit** $n \to \infty$ on a sequence of graphs $\{G(V_n, E_n) : n \geq 1\}$.
- Taking the limit $n \to \infty$ simplifies computations in order for us to identify general properties.
- In practice, establishing results in the large graph limit means that our findings are likely to be true for sufficiently large graphs.

Static vs. evolving models

- Random graph models can be broadly classified into two categories: **static models** and **evolving or growing models**.
- Static models are meant to represent a "snapshot" of a large network.
- In static models $G(V_n, E_n)$ and $G(V_{n+1}, E_{n+1})$ can be totally different.
- Evolving models are meant to describe the growth of a graph as vertices get added to the graph (usually one at a time), so $G(V_n, E_n)$ and $G(V_{n+1}, E_{n+1})$ share most edges.
- In many evolving models edges and vertices never disappear, so $G(V_n, E_n)$ is a subgraph of $G(V_{n+1}, E_{n+1})$.

The Erdős-Rényi random graph

- The simplest model for a random graph is the **Erdős-Rényi model**.
- Consider a graph with vertex set $V_n = \{1, 2, \ldots, n\}$.
- There are a total of $\binom{n}{2}$ possible edges in the graph, and each of them will be chosen to be present or not with a coin flip.
- Suppose you have a coin that lands heads with probability $p \in (0, 1)$.
- For each pair of vertices i and j, toss the coin; if it lands heads, draw an edge between i and j, otherwise do nothing.
- Equivalently, if A denotes the adjacency matrix of the graph, let

$$a_{i,j} = a_{j,i} = 1(\text{coin-flip is a head}), \qquad i \neq j,$$

and set $a_{i,i} = 0$.

Properties of the Erdős-Rényi model

- This is the most studied random graph model there is.
- Some of its connectivity properties are:
  - If $np < 1$ the graph will consists of only small components of size $O(\log n)$.
  - If $np \to c > 1$ the graph will contain a unique *giant* connected component, with all other components of size $O(\log n)$.
  - If $np = 1$ the largest component will have size $O(n^{2/3})$.
  - If $p < (1-\epsilon)n^{-1}\log n$ the graph will most likely be **disconnected**.
  - If $p > (1+\epsilon)n^{-1}\log n$ the graph will most likely be **connected**.
- When the graph is connected, it exhibits the **small-world** property, with typical distance of order $O(\log n)$.

Degree distribution

- To compute the degree distribution we can use binomial probabilities.
- Fix a vertex $i \in V_n$, then its degree is given by

$$D_i = \sum_{j=1}^{n} \chi_{i,j}, \qquad \chi_{i,j} = 1((i,j) \in E_n)$$

- Note that the $\chi_{i,j}$ are independent Bernoulli r.v.s with parameter p.
- Therefore, since all vertices have the same distribution, for all $i \in V_n$,

$$P(D_i = k) = P(D_1 = k) = P(\mathsf{Binomial}(n,p) = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Degree distribution... cont.

- Moreover, if $np \to c$ as $n \to \infty$, we can approximate the binomial as follows:

Degree distribution... cont.

- Moreover, if $np \to c$ as $n \to \infty$, we can approximate the binomial as follows:

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{1}{k!}(np)^k \cdot \frac{(1-p)^n}{(1-p)^k}$$
$$\to 1 \cdot \frac{1}{k!} c^k \cdot \lim_{n\to\infty} e^{n\log(1-p)}, \qquad n \to \infty$$

- To compute the last limit, note that

$$\lim_{n\to\infty} n\log(1-p) = \lim_{n\to\infty} (-np + O(np^2)) = -c$$

- Therefore,

$$\lim_{n\to\infty} P(D_1 = k) = \frac{e^{-c}c^k}{k!}, \qquad k \geq 0,$$

known as a Poisson distribution with mean c.... not **scale-free**.

Poisson vs. scale-free

- A Poisson distribution is a **light-tailed**, i.e., its tail decreases exponentially fast.
- Poisson random variables are all close to their mean.
- A scale-free distribution is **heavy-tailed**, i.e.,

$$\sum_{k=0}^{\infty} e^{\epsilon k} P(D = k) = \infty$$

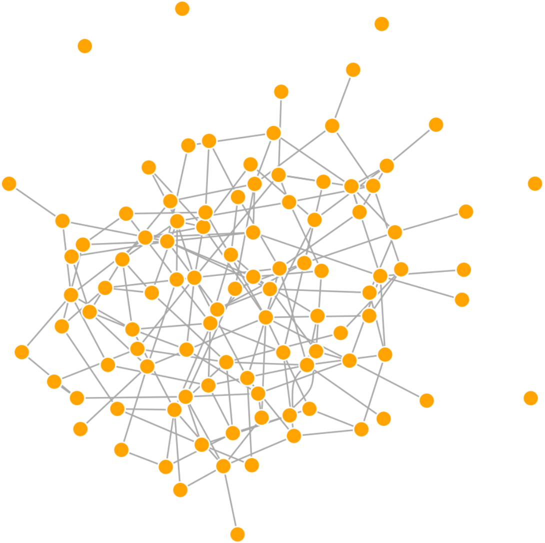
  for all $\epsilon > 0$.
- Heavy-tailed random variables can take extremely large values.
- In particular, for any $k \geq 1$,

$$\lim_{m \to \infty} P(D > k + m | D > m) = 1$$

  which can be interpreted as:

"Given that D is large, most likely it is huge."

An Erdős-Rényi graph



Inhomogeneous random graphs

- Erdős-Rényi graphs are quite **homogeneous**, i.e., all the vertices have degrees close to their common mean.
- Real-world networks are often scale-free.
- We can create random graphs that have inhomogeneous degrees by allowing the edge probabilities to vary from vertex to vertex.
- To each vertex $i \in V_n$ assign a value $w_i \geq 0$, and define the edge probability

$$p_{i,j}^{(n)} := P((i,j) \in E_n) = \frac{w_i w_j}{l_n} \wedge 1, \qquad i \neq j,$$

  where $l_n = w_1 + \cdots + w_n$.
- The adjacency matrix of the graph is given by:

$$a_{i,j} = \begin{cases} 1, & \text{with probability } p_{i,j}^{(n)}, \\ 0 & \text{with probability } 1 - p_{i,j}^{(n)}. \end{cases}$$

Inhomogeneous random graphs... cont.

- Each edge is determined independently of other edges.
- This choice of edge probabilities corresponds to the **Chung-Lu model**.
- The expected degree of vertex $i \in V_n$ is:

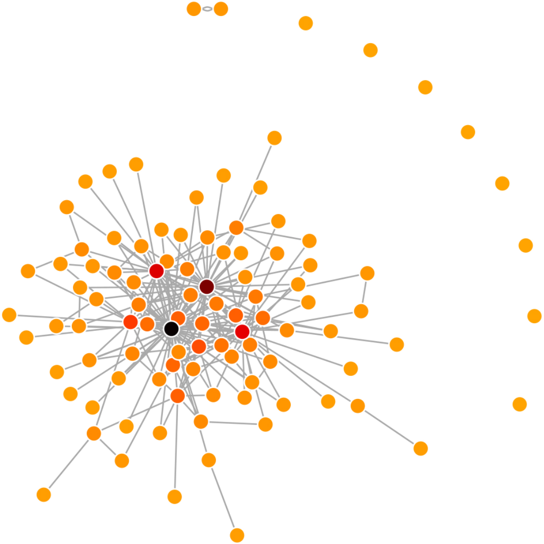
$$E[D_i] = \sum_{j=1}^{n} p_{i,j}^{(n)} \approx w_i$$

- If we let

$$F(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1(w_i \leq x),$$

  then the degree distribution "looks" like F.
- If we set $w_i = p$ for all $i \in V_n$ we get an Erdős-Rényi model.
- Scale-free graphs can be obtained by choosing F to be a power-law distribution.

An inhomogeneous random graph



Graphs with communities

- Inhomogeneous random graphs can be **scale-free** and will have the **small-world** property.
- However, they do not have community structure.
- Suppose we want to generate a graph with K communities.
- To each vertex $i \in V_n$ assign a community label $c_i \in \{1, 2, \dots K\}$.
- Now sample edges independently using edge probabilities of the form:

$$p_{i,j}^{(n)} = P((i,j) \in E_n) = \frac{\kappa(c_i, c_j)}{n}, \qquad i \neq j,$$

  where $\kappa : \{1, \dots, K\} \times \{1, \dots, K\} \to [0, \infty)$.
- The size of community $k \in \{1, \dots, K\}$ is $n\pi_{n,k} = \sum_{i=1}^{n} 1(c_i = k)$.

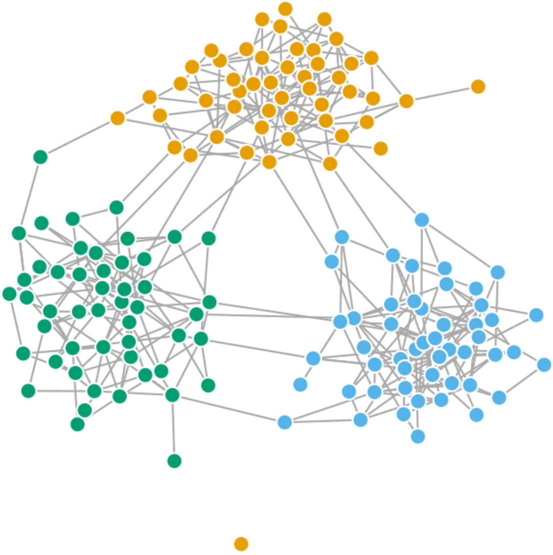
Graphs with communities... cont.

- This construction is known as a **stochastic block model**.
- In order to create communities we choose $\kappa(c_i, c_j)$ be "large" for $i = j$, and "small" for $i \neq j$.
- The expected degree of a vertex in community $m \in \{1, \ldots, K\}$ is:

$$E[D_i | c_i = m] = \sum_{j=1}^{n} \frac{\kappa(m, c_j)}{n} = \sum_{r=1}^{K} \kappa(m, r)\pi_{n,r}$$

- Stochastic block models are homogeneous within each community, but can have different expected degree from one community to another.

A stochastic block model



Graphs with clustering

- The **global clustering coefficient** of a graph is

$$\frac{\text{number of triangles}}{\text{number of open wedges}}$$

- Inhomogeneous random graphs do not have significant clustering.
- In fact, inhomogeneous random graphs are **locally tree-like**.
- They have "long" cycles of length $O(\log n)$.
- The clustering coefficient in the models we have seen converges to zero as $n \to \infty$.
- Real-world graphs often have positive clustering coefficients, especially social networks.

Graphs with clustering... cont.

- To construct a graph with non-negligible clustering, we start by generating a **bipartite graph** with vertex sets $V_n = \{1, \dots, n\}$ and $\mathcal{A}_m = \{a_1, \dots, a_m\}$, $n, m \geq 1$.
- To each vertex $i \in V_n$ assign a value $w_i \geq 0$ and define

$$p_i = \frac{\gamma w_i}{n} \wedge 1,$$

  where $\gamma > 0$ is a fixed parameter.
- Next, for each $i \in V_n$ toss a coin that lands heads with probability p_i with each of the vertices in A_m, and draw an edge if it is a head.
- Let $N(i) \subseteq \mathcal{A}_m$ be the set of neighbors of i.
- We will now construct a new graph $G(V_n, E_n)$, with adjacency matrix A by setting:

$$a_{i,j} = 1(N(i) \cap N(j) \neq \varnothing)$$

Graphs with clustering... cont.

- This model is called a **random intersection graph**.
- Let $F(x) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} 1(w_i \leq x)$ be the weight distribution, and assume it has finite mean.
- If we choose $m = \lfloor \beta n \rfloor$, the degree of vertex $i \in V_n$ in $G(V_n, E_n)$ will have (approximately) the distribution of

$$\text{Poisson}(\beta\gamma w_i) + \text{Poisson}(\gamma),$$

  with the two Poisson r.v.s independent of each other.
- As with inhomogeneous random graphs, we can obtain the scale-free property by choosing F to be a power-law distribution.
- The parameters β, γ can be used to tune the clustering coefficient to cover the entire range $(0, 1)$, with small values of $\beta\gamma$ producing higher clustering.

An intersection graph

The Albert-Barabási model

- All the random graph models we have seen so far are **static**.
- Static models do not explain how graphs grow.
- **Evolving** models propose a mechanism for choosing how a new vertex will connect to the existing graph.
- Vertices are labeled in the order in which they arrive to the graph.
- One of the most famous evolving random graph models is the **Albert-Barabási graph** or **preferential attachment model**.
- This model assumes that an incoming vertex will choose a vertex to connect to with probability proportional to its degree.
- In other words, newcomers "prefer" to attach to high degree vertices.

The Albert-Barabási model... cont.

- The model starts with one vertex that has a self-loop.
- At each time step, a new vertex arrives and connects by drawing one edge either to itself, or to an existing vertex.
- Let $D_i(k)$ be the degree of vertex i after k vertices have arrived.
- When vertex $k+1$ arrives it attaches to vertex i with probability:

$$p_i(k) = \begin{cases} \frac{D_i(k)}{2k+1}, & i = 1, \ldots, k, \\ \frac{1}{2k+1}, & i = k+1. \end{cases}$$

- This model produces **scale-free** graphs with degree distribution:

$$P_k(n) = \frac{1}{n} \sum_{i=1}^{n} 1(D_i(n) = k) \approx 4k^{-3}$$

for large n.

Thank you for your attention.