Intro to Random Graphs and Exponential Random Graph Models

Danielle Larcomb

University of Denver

Necessity of Random Graphs

The study of complex networks plays an increasingly important role in the sciences. Some examples:

- Electrical power grids
- Telecommunication networks
- Social relations
- WWW, Internet (pages vs. routers)
- Collaboration of scientists (Erdös Number of Mathematicians)

The structure of these networks affects performance.

Small-World Phenomenon

These networks have the common feature that they are big, really big. Studying their global properties directly is impossible, and so we use random graphs to study local properties a network (probably) has.

Some examples of graphs with "small-world phenomenon":

- 6 degrees of Kevin Bacon
- 8 degrees of Paul Erdös
- 2 webpages are, on average, 16 clicks away from one another

The small-world phenomenon occurs in numerous existing networks, and refers to two different properties:

- 1. Small Distance: between any two pair of nodes, there is a short path
- 2. The Clustering Effect: Two nodes are more likely to be adjacent if they share a common neighbor

A "good" graph model must account for both of these properties in order to be "realistic"

Outline

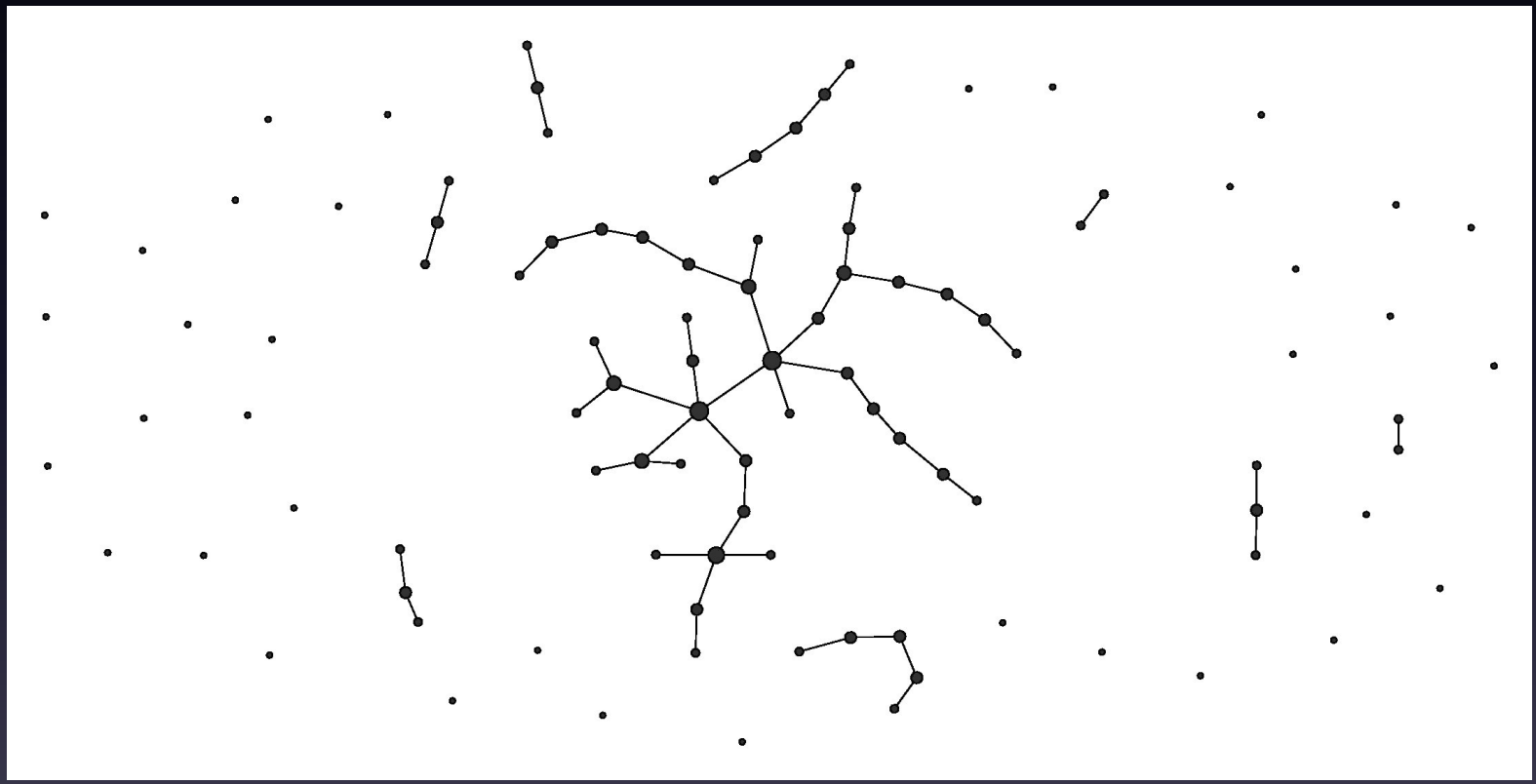
1. Graph Theory Basics
2. What is a Random Graph?
3. The Erdös Rényi Random Graph Model
4. Limitations of the ER Model
5. Exponential Random Graph Model
6. Graph Limits
7. Main Result

Graph Theory Preliminaries

Some definitions

- $G = (V, E), V = [n], E$ = {(u, v) | there is an edge between u and v}
- The *degree* of a vertex u, d_u, is the number of edges containing u
- A *path* is a sequence of vertices $v_0, v_1, ..., v_k$ such that v_{i-1} is adjacent to v_i for $i = 1, 2, ..., k$, with length k
- If $v_0 = v_k$, then the path is called a *cycle*
- A graph with no cycles is called a *tree*
- A graph is *connected* if any two vertices can be joined by a path
- If a graph is not connected, then we can consider a *connected component*, a subset of the graph that is connected. Thus, a graph is connected if there is only one connected component

Definition Picture



What is a Random Graph?

A **random graph** may be viewed as a random variable defined on a probability space with a probability distribution.

- If we first put all graphs on n vertices in a box and then choose a graph from that box, then the graph we chose was a random graph
- Goal: To be able to say that a random graph (in some model) has a certain property
- Since real life networks are LARGE, we can use random graphs to describe local and probabilistic rules by which vertices are connected to another

Origins: Erdös

"In random graph theory, we consider asymptotic statistical properties of random graphs for n (vertices) approaching infinity"

- What is the probability of $G(n, N)$ ($N \leq \binom{n}{2}$) being completely connected?
- What is the probability that the greatest connected component of $G(n, N)$ should have $n - k$ vertices?
- What is the probability that $G(n, N)$ should consist of exactly $k + 1$ connected components?
- If the edges of a graph with n vertices are chosen successively so that after each step every edge which has not yet been chosen has the same probability to be chosen as the next, and if we continue this process until the graph becomes completely connected, what is the probability that the number of necessary steps V will be equal to a given number L?

Erdös and Rényi: 2 models

$\mathcal{F}(n, m)$: Random graph defined on n vertices, and each graph in $\mathcal{F}(n, m)$ has m edges.

- A graph is chosen uniformly at random from $\mathcal{F}(n, m)$
- Example: In $\mathcal{F}(3, 2)$, each of the three possible graphs on 3 vertices with 2 edges are chosen with probability $\frac{1}{3}$.

$\mathcal{G}(n, p)$: Random graph defined on n vertices, and each edge is chosen independently with probability p. (Edgar Gilbert)

- Each edge is assigned independently of other edges with probability p
- A graph in $\mathcal{G}$ with m edges is chosen with probability $p^m(1-p)^{\binom{n}{2}-m}$

Erdös and Rényi: $\mathcal{G}(n, p)$

- The simplicity of $\mathcal{G}(n, p)$ arises from the basic probability result: $P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n)$ when events A_1, ..., A_n are independent
- This means computing is very simple compared to other models (e.g. $\mathcal{F}(n, m)$)

Example: The probability of a random graph in $\mathcal{G}(n, \frac{1}{2})$ containing a fixed triangle is $\frac{1}{8}$.

Since edges in $\mathcal{F}(n, m)$ are not independently chosen, calculations are more difficult. It is this reason why $\mathcal{G}(n, p)$ has risen to such mathematical fame.

Properties of $\mathcal{G}(n, p)$: Phase Transition

What is a Phase Transition?

$\rightarrow$ Water exhibits a phase transition at 0 degrees Celsius (at standard pressure):

- Below 0 degrees, water exists in solid state
- Above 0 degrees, water exists in a liquid state

For varying p, graphs in $\mathcal{G}(n, p)$ exhibit a phase transition. Below a certain p, graphs in $\mathcal{G}$ are a disjoint union of trees (no cycles). Above this p, graphs in $\mathcal{G}$ have one giant connected component and all others are quite small.

Limitations of Erdös Rényi

- Doesn’t model "small world phenomenon", which we know exists in many realistic networks
- The very thing that makes it easy to deal with, the independence, is what makes it unrealistic
- If $p \geq \frac{log(n)}{n}$, all degrees are similar and close to pn

Exponential Random Graph Model

- ERGMs are able to represent a wide range of common network tendencies by using structural elements from the network incorporated into the model
- For example, if we believe that edges and triangles are an important network feature, then we can include the number of edges and the number of triangles in our model

Intro: ERGM

Consider the set $\mathcal{G}_n$ of all simple graphs G_n on n vertices.

We consider the exponential families,

$$p_\beta(G) = \exp\left(\sum_{i=1}^{k} \beta_i T_i(G) - \psi(\beta)\right), \tag{1}$$

where $\beta = (\beta_1, ..., \beta_k)$ is a vector of real parameters, $T_1, ..., T_k$ are functions on the space of graphs, and ψ is the normalizing constant

- It is very difficult to estimate the parameters in these models
- The normalizing constant is unknown, and very different β values can give essentially the same distribution of graphs

ERGM Cont.

- We want to find a way to evaluate the normalizing constant so that we can determine the parameters $(\beta_1, ..., \beta_k)$ (using MLE, Bayseian Inference)
- Currently, methods for determining ψ only exist for relatively small graphs (using computationally intensive algorithms)

Approach

- The functions of the space, $T_1, ..., T_k$ are known
- The parameters, $\beta = (\beta_1, ..., \beta_k)$, and the normalizing constant, ψ, are unknown
- Using graph limits, we will arrive at an approximation of the normalizing constant, ψ
- The limitation of this approach is that it only applies to *dense* graphs due to the use of graphons (also called graph limits)

A **dense** graph is one in which the graph density,

$$D = \frac{2|E|}{|V|(|V|-1)} \tag{2}$$

is close to 1, and more broadly, D scales like

$$O(|V|^2). \tag{3}$$

Graphons: Motivation and Definitions

Graph Limits allow us to put all simple graphs into the same Probability Space, regardless of the number of vertices (limitation of the E-R model)

- Define $\mathcal{H} = \{H_1, ..., H_k\}$ to be a fixed finite collection of finite simple graphs.
- Define a homomorphism of a finite simple graph H into G as an edge preserving map $V(H) \to V(G)$
- Let $|hom(H, G)|$ be the number of homomorphism of H into G
- We define the homomorphism density to be

$$t(H, G) = \frac{|hom(H, G)|}{|V(G)|^{|V(H)|}} \tag{4}$$

Introduction: Graphons

- Define $h \in \mathcal{W}$, where $\mathcal{W}$ is the set of all symmetric measurable functions from $[0,1]^2$ into $[0,1]$.
- Let G_n be a sequence of dense graphs that become more similar as $n \to \infty$.
- If

$$\lim_{n \to \infty} t(H_i, G_n) = t(H_i, h)$$

for every H_i in the collection $\mathcal{H}$, then the sequence of graphs $\{G_n\}$ converges to $h \in \mathcal{W}$, and h is called a *graphon*, or a *graph limit*.

Interpretation of Graphons

- $h(x, y)$ denotes the probability of an edge between x and y

Example For the Erdös Rényi model $\mathcal{G}(n, p)$, the limit function is $h(x, y) = p$.
Since we use graphons, this method is only useful when $p \nrightarrow 0$ as $n \to \infty$. In other words, the method works only for dense Erdös Rényi graphs. The natural notion for the graphon limit of the Erdös Rényi graphs with fixed parameter p is the edge density of the graphs, which converges to p.

Every graphon has a sequence of dense simple graphs converging to it

- A sequence of dense simple graphs converging means that there is some graphon in $\mathcal{W}$ that the sequence converges to. Conversely, for every $h \in \mathcal{W}$, there is a graph sequence $\{G_n\}$ with h as its graph limit.
- A finite simple graph G on $\{1, 2, ..., n\}$ can be represented as a graph limit, $f^G : [0, 1]^2 \to [0, 1]$, by defining

$$f^G(x, y) = \begin{cases} 1 & \text{if } (\lceil nx \rceil, \lceil ny \rceil) \text{ is an edge in } G, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

- Therefore, $t(H, f^G) = t(H, G)$ for every simple graph H and the constant sequence $\{G, G, ..., G\}$ converges to the graph limit f^G.

This allows all simple graphs, no matter the number of vertices, to be expressed as elements of the same probability space, $\mathcal{W}$.

Background of Main Result

- Convergence in $\mathcal{W}$ is determined by a metric on $\mathcal{W}$ called the cut distance, denoted d_*
- Two graphons $f \sim g$ if $f(x,y) = h_\sigma(x,y) = h(\sigma x, \sigma y)$ for some measure preserving bijection σ of $[0,1]$. We may think of σ as a relabeling of the vertices of a finite graph and consider the resulting quotient space $\widetilde{\mathcal{W}} = \mathcal{W}/\sim$.
- The cut distance on $\mathcal{W}$ induces a metric on the quotient space and we have a metric space $\left(\widetilde{\mathcal{W}}, \delta_*\right)$
- Facts about the space $\widetilde{\mathcal{W}}$:
  - $\widetilde{\mathcal{W}}$ is compact
  - The homomorphism density $t(H,\cdot)$ is continuous for any finite simple graph H

The Main Result

If $T : \widetilde{\mathcal{W}} \to \mathbb{R}$ is a bounded continuous function, then

$$\psi = \lim_{n\to\infty} \psi_n = \sup_{\widetilde{h}\in\widetilde{\mathcal{W}}} (T(\widetilde{h}) - I(\widetilde{h})) \tag{6}$$

where

$$I(u) = \frac{1}{2}(u \log u + (1-u)\log(1-u)) \tag{7}$$

$I : [0,1] \to \mathbb{R}$.
We can extend the function I to $\widetilde{\mathcal{W}}$ as

$$I(\tilde{h}) = \int\int_{[0,1]^2} I(h(x,y))dxdy \tag{8}$$

where h is a representative of the equivalence class $\tilde{h}$.

Consequences of Main Result

- Saying: there is a graph that is smaller that has representative features of my bigger graph, and we are taking that ψ that we can calculate because it is smaller, and then applying it to the larger graph
- We use this smaller graph's normalizing constant as an approximation of the ψ that we can't calculate for the larger graph

- Much easier to calculate probabilities from this space
- ERGMs good for modeling the clustering effect
- The use of graphons means we need graphs to be dense, and not all complex networks are dense

References

- http://www.math.ucsd.edu/ fan/wp/randomg.pdf
- http://www.di.ens.fr/ bouillar/PACS/randomGraphs.pdf
- http://arxiv.org/abs/1102.2650
- Wikipedia
- http://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf