Lecture 03

STATIC NETWORK MODELS

Probability & Statistics

Francisco Richter, Martina Boschi and Ernst Wit

Notation

- G: graph;
- V: node set of a graph;
- E: edge set of a graph;
- A: adjacency matrix of a network;
- $d(u)$: degree of a node of a graph;
- $d(u, v)$: distance between two generic nodes of a graph;
- D: diameter of a graph;
- L: average shortest path;
- $\mathcal{G}$: random graph;
- $\mathcal{G}(N, p)$: Erdos-Renyi (ER) Model;
- C: largest connected component of a graph;
- p_C: critical probability for emergence of a giant component in an ER graph;
- $\Pi(k)$: probability that a new node connects to a node with degree k;
- $\bar{d}_G$: average degree of graph G;

Introduction

In today's interconnected world, networks form the backbone of numerous intricate systems. From biological intricacies in organisms and deep-rooted ties in social interactions to vast interdependencies in technological infrastructures, networks are omnipresent. They offer a lens to view and comprehend the complex fabric of interactions that underpin these systems. Given the importance and ubiquity of networks, understanding their essential properties and behaviors becomes a cornerstone for many disciplines, including sociology, computer science, economics, and epidemiology.

Static network models are crucial in this exploration. While dynamic network models unveil the evolution of networks over time, static models freeze a network's snapshot, allowing us to delve deep into its structural properties. Through this lens, we gain profound insights into the network's organization, function, and behavior, even if it is just for a moment in its vast timeline.

Central to the study of these models are a few pivotal concepts and phenomena that have shaped our modern understanding of complex networks. The Erdos-Renyi Graphs, a foundational model, illuminate the properties of purely stochastic systems. On the other hand, the Preferential Attachment Model elucidates the rich-get-richer phenomena often observed in multifarious real-world systems. Meanwhile, the Six Degrees of Separation uncovers the astonishingly short paths within colossal networks, leading to the famous small-world effect. Lastly, the Friendship Paradox offers an intriguing glimpse into the realm of relative connectivity, revealing that most individuals often have fewer friends than their friends, on average.

1 Network Basics

A **network** is a collection of interconnected entities, often represented mathematically by a **graph**. In this book, however, we will use the terms interchangeably.

Definition 1 *A **Graph** G is a collection of interconnected entities called **Nodes** (or Vertices) and the relationships or connections between them, termed as **Edges** (or Links). Each edge connects two nodes and indicates a relationship between them.*

A first mathematical representation of the graph is thus: $G = (V, E)$.

Example 1 *Consider a simple network with 5 nodes and 6 edges:*

$$V = \{1, 2, 3, 4, 5\} \tag{1}$$

$$E = \{(1, 2), (2, 3), (1, 3), (2, 4), (3, 4), (4, 5)\} \tag{2}$$

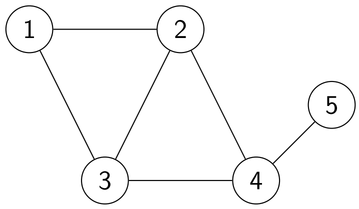


Figure 1: Example network with 5 nodes and 6 edges.

After defining networks in terms of nodes and edges, an additional mathematical and more summarized representation is essential for analysis. The adjacency matrix offers a compact way to depict the relationships within a network.

Definition 2 *The **Adjacency Matrix** A is a square matrix of size $N \times N$ where N is the total number of nodes (**Order** of the network). Entry A_{uv} equals 1 if there is an edge from node u to node v and 0 otherwise.*

Example 2 *The adjacency matrix for the network in the previous example is:*

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

An adjacency matrix is a powerful tool for representing graphs because it encodes all the information about connections between nodes in a systematic manner. The cell A_{uv} represents the edge from node u to node v. If $A_{uv} = 1$, then an edge exists; otherwise, it is zero. This binary encoding simplifies complex relationships into a format easily analyzed computationally.

2 Network Properties

Definition 3 *A **Shortest Path** between two nodes u and v in a graph is a path that has the minimum number of edges (in an unweighted graph) or the minimum sum of edge weights (in a weighted graph) among all possible paths between u and v. The length of this path is denoted as $d(u, v)$.*

Definition 4 *The **Diameter** of a graph is the longest shortest path between any two nodes in the graph. Formally, if $d(u, v)$ is the shortest path between nodes u and v, then the diameter D is defined as:*

$$D = \max_{u,v \in V} d(u, v)$$

Example 3 *Consider the previous network example, where $V = \{1, 2, 3, 4, 5\}$ and $E = \{(1, 2), (2, 3), (1, 3), (2, 4), (3, 4), (4, 5)\}$.*
The shortest paths between all pairs of nodes are:

- *Shortest path from 1 to 2, 3, 4, 5 are 1, 1, 2, 3 respectively.*
- *Shortest path from 2 to 1, 3, 4, 5 are 1, 1, 1, 2 respectively.*
- *Shortest path from 3 to 1, 2, 4, 5 are 1, 1, 1, 2 respectively.*

- *Shortest path from 4 to 1, 2, 3, 5 are 2, 1, 1, 1 respectively.*
- *Shortest path from 5 to 1, 2, 3, 4 are 3, 2, 2, 1 respectively.*

The diameter of this graph is the longest of these shortest paths, which in this case is 3 (from node 1 to node 5).

Knowing the shortest paths and diameters in a network has a wide range of practical applications. For instance, in social networks, the diameter can provide insights into how quickly information may spread across the network. In transportation networks, identifying the shortest paths is crucial for optimizing travel routes. Understanding these properties is integral for network resilience, efficiency, and information dissemination.

3 Graph Models

Random graphs are a foundational construct in the mathematical treatment of network theory, offering researchers a framework for understanding the probabilistic interactions within complex networks.

Definition 5 (Random Graph) *A* ***Random Graph*** $\mathcal{G}$ *is a graph in which the presence or absence of an edge between any two distinct nodes is determined by a probabilistic rule.*

Random graph models describe the probability distributions on the graph.
Among various random graph models, the Erdos-Renyi model stands out due to its simplicity and foundational role in network theory.

Definition 6 (Erdos-Renyi Model) *The* ***Erdos-Renyi Model****, denoted as* $\mathcal{G}(N, p)$*, is defined as a random graph consisting of* N *nodes, where each potential edge between distinct nodes* u *and* v *is included with probability* p*, independently of the other edges.*

The probability of an edge forming between any two nodes u and v in the Erdos-Renyi model $\mathcal{G}(N, p)$ is:

$$P(\text{Edge between } u \text{ and } v) = p$$

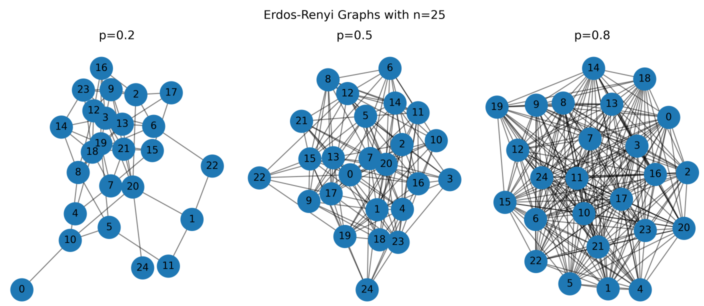


Figure 2: Variations of Erdos-Renyi graphs with $n = 25$ nodes and varying p.

Definition 7 (Giant Connected Component) *A* ***Giant Connected Component*** *in a graph* G *is a connected component that includes a substantial fraction of the entire set of nodes in the graph. In the Erdos-Renyi model* $\mathcal{G}(N, p)$*, a giant connected component emerges when the edge probability* p *surpasses a critical value* p_C*.*

The critical probability p_C for the emergence of a giant component in an Erdos-Renyi graph $\mathcal{G}(N, p)$ is approximately:

$$p_C \approx \frac{\log(N)}{N}$$

Experiment 1 (Empirical Estimation of p_C) *To empirically estimate the critical probability* p_C *at which a giant connected component forms in an Erdos-Renyi graph* $\mathcal{G}(N, p)$*, we will perform the following experiment:*

1. *Initialize* $N = 1000$ *nodes.*
2. *Vary* p *from 0 to 1 in increments of 0.01.*

3. *For each p, generate a random graph $\mathcal{G}(N, p)$.*

4. *Identify the largest connected component C in G.*

5. *Record the size $|C|$ of the largest connected component.*

6. *Plot $|C|$ as a function of p.*

7. *Observe the value of p where $|C|$ starts to dramatically increase. This value is an empirical estimate of p_C.*

The Erdos-Renyi model, despite its simplistic assumptions, serves as a key reference model in the realm of network theory. Its clear framework for randomness offers a foundation for the study of more complex networks, making it a cornerstone in disciplines such as computer science, biology, and social sciences.

The Preferential Attachment Model, popularized by Albert-László Barabási and Réka Albert, is grounded in the adage "the rich get richer." Nodes are more likely to link to nodes that already have many connections. This dynamic leads to "hubs" or nodes with significantly higher connectivity than others.
Mathematically, the probability $\Pi(k)$ that a new node connects to a node with k connections is proportional to k.

Definition 8 *The **Degree** $d(u)$ of a node u represents the number of edges connected to that node.*

Definition 9 (Preferential Attachment Model) *The **Preferential Attachment Model** is a foundational concept often used to explain how real-world networks evolve to exhibit a scale-free degree distribution. Proposed by Barabási and Albert in 1999, this model argues that networks grow by the principle of "the rich get richer," whereby new nodes are more likely to connect to already well-connected nodes.*

The central equation governing the Preferential Attachment Model is:

$$\Pi(k) = P(\text{Edge between new node and } u \text{ with } d(u) = k) = \frac{k}{\sum_{v \in V} d(v)}$$

where $\Pi(k)$ is the probability that a new node will connect to a node with degree k, and the sum runs over all nodes v in the network.
This results in a **scale-free network**, characterized by a **power-law degree distribution**.
The Preferential Attachment Model provides a basis for understanding how highly connected "hubs" emerge in networks. These hubs play a critical role in the network's overall structure and resilience, often dominating processes like information spread or failure propagation.

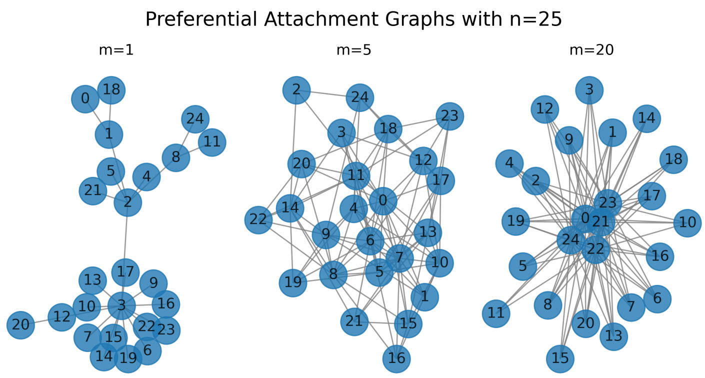


Figure 3: Three variations of the Erdos-Renyi graph with $N = 25$ nodes and edge probabilities $p = 0.2, 0.5, 0.8$.

4 Social Networks Analysis

Online social networks have become an integral part of our daily lives. Platforms like Twitter, Facebook, and Instagram have millions of users who are constantly connecting,

forming both weak and strong ties. These networks evolve dynamically, with users adding friends based on existing connections or sometimes even at random.
To study the structure and evolution of these networks, we turn to graph theory. Consider an initial scenario where a new social platform is launched and the first 100 users join, forming connections at random. This can be represented using the Erdos-Renyi model $\mathcal{G}(N, p)$, where N is the number of users and p is the probability that a pair of users form a connection.
After this initial phase, as the platform gains popularity, users start joining influenced by their friends or by existing popular users. This is where the Preferential Attachment Model comes into play. New users are more likely to connect to popular users, causing them to gain even more connections.
To simulate this scenario:

1. Start with the Erdos-Renyi model, $G(100, 0.05)$, representing the initial 100 users.

2. As more users join, they form connections based on the Preferential Attachment Model. For simplicity, let's assume each new user forms connections with 5 existing users, and we add 900 new users this way.

Mathematically, the probability $\Pi(k)$ for a new user to connect to an existing user with k connections is given by:

$$\Pi(k) = \frac{k}{\sum_{vinV} d(v)}$$

By combining the Erdos-Renyi and Preferential Attachment models, we get a representation of an online social network. Analyzing this combined network can provide insights such as the average path length between users, which is indicative of the "small-world" phenomenon seen in real-world networks.

4.1 Friendship Paradox

The Friendship Paradox can be mathematically described by the skewed degree distributions often observed in social networks. Let N be the total number of nodes in the graph, and let $d(v_i)$ be the degree of the i^{th} node. The average degree $\bar{d}_G$ is calculated as:

$$\bar{d}_G = \frac{1}{N} \sum_{i=1}^{N} d(v_i)$$

However, the probability of selecting a person (or node) as a friend increases with the number of friends that person has. When you randomly pick an edge, you are more likely to land on a node with a higher degree. The average degree of a node reached through a random edge $\bar{d}_{\text{friends}}$ is:

$$\bar{d}_{\text{friends}} = \frac{\sum_{i=1}^{N} d(v_i)^2}{\sum_{i=1}^{N} d(v_i)}$$

Due to the squared term in the numerator, $\bar{d}_{\text{friends}}$ will generally be greater than $\bar{d}_G$, thus confirming the Friendship Paradox.
The Friendship Paradox has practical implications in real-world networks, such as social media platforms. The majority of users typically have fewer followers than those they follow, a statistical trend that has significant applications in viral marketing strategies and information dissemination.

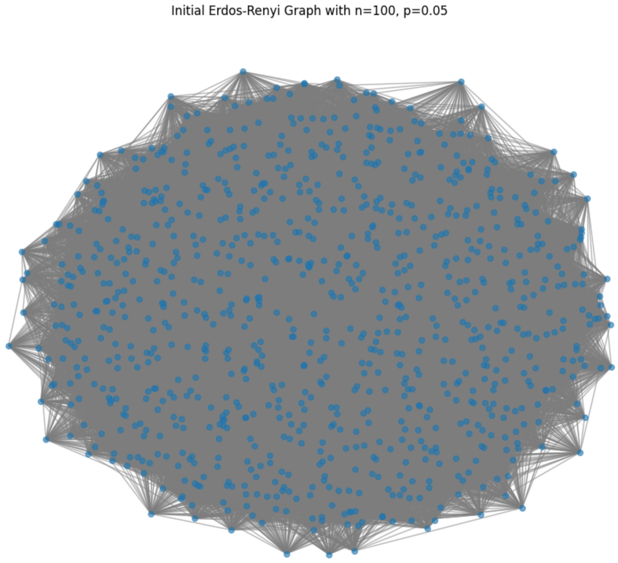
Example 4 *In the simulation experiment, we can calculate the degree distribution of the final network. On Figure 5 we see that most nodes are not connected but that there is a cluster of highly connected nodes (influencers).*

4.2 Six Degrees of Separation

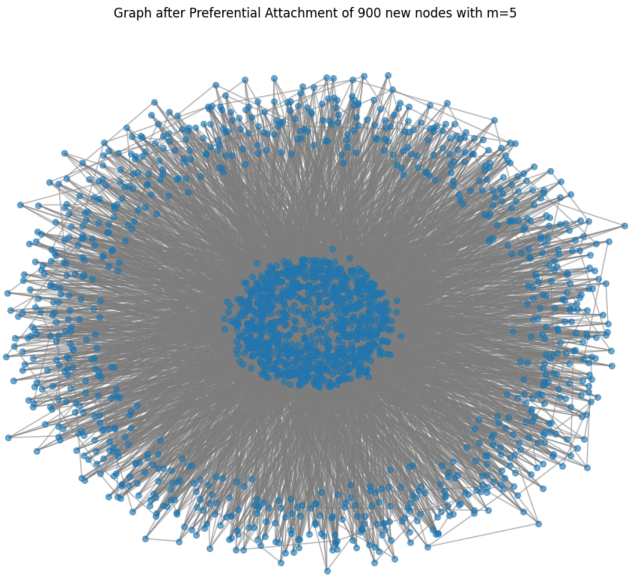
The Six Degrees of Separation can be understood in terms of "shortest paths" in a graph. Given a graph $G = (V, E)$, the average shortest path L is calculated as:

$$L = \frac{2}{N(N-1)} \sum_{u \neq v} d(u, v)$$

where $d(u, v)$ represents the shortest distance between nodes u and v.



(a) Initial Erdos-Renyi Network



(b) After Preferential Attachment

Figure 4: Comparison of the initial Erdos-Renyi network with the evolved network post preferential attachment. These graphs provide a visual representation of the growth and evolution of connections as the network matures.

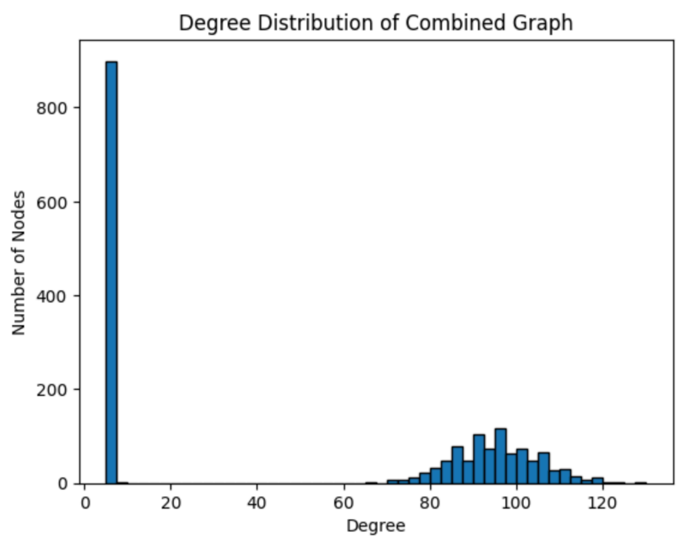


Figure 5: Degree distribution of the combined graph.

> In many real-world networks, especially social networks, it has been observed that L is much smaller than the network size N, exhibiting the 'small-world' property.
> The Six Degrees of Separation posits that any two individuals are, on average, separated by at most six social connections, leading to the idea of a "small world". Originating from a short story by Frigyes Karinthy and later buttressed by experiments like that of Stanley Milgram, it underscores the compactness of social networks.