Week #5: Networks

March 10, 2025

1 Random Networks

Random network models provide a probabilistic framework for studying the structure and dynamics of complex systems. Such models are instrumental in fields ranging from sociology and epidemiology to computer science, as they help us understand phenomena like information diffusion, epidemic spreading, and social connectivity.

A probabilistic model in this context is one that assigns probabilities to different graph structures. To formalize this, we first introduce basic graph notation.

Definition. Graph

A graph G is a pair $G = (V, E)$, where:

- V is a set of nodes (or vertices),
- E is a set of edges, where each edge connects a pair of nodes.

Example.

Consider a graph with five nodes:

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

and the edge set

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

Figure 1 illustrates this graph.

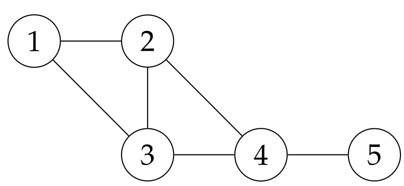

Figure 1: A simple graph with five nodes and six edges.

Graphs can be represented in many ways; one common representation is the adjacency matrix.

Definition. Adjacency Matrix

For a graph $G = (V, E)$ with $N = |V|$ nodes, the *adjacency matrix* A is an $N \times N$ matrix where

$$A_{uv} = \begin{cases} 1 & \text{if there is an edge between nodes } u \text{ and } v, \\ 0 & \text{otherwise.} \end{cases}$$

Example.

For the graph in the previous example, the adjacency matrix is

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

With these fundamentals, we now introduce two important random network models.

The adjacency matrix A not only encodes the structure of a network but also provides a way to analyze connectivity and paths between nodes. The powers of A reveal important information about the number of walks of different lengths between nodes.

Let A be the adjacency matrix of a graph G. The entry $(A^k)_{uv}$ represents the number of walks of length k from node u to node v.

Proof. The proof follows by induction on k.

For $k = 1$, the adjacency matrix A satisfies

$$A_{uv} = 1 \quad \text{if and only if there is an edge between } u \text{ and } v.$$

This directly counts the number of walks of length 1 between nodes.

Suppose that for some k, the matrix A^k counts the number of walks of length k from u to v. Then, for $k + 1$, the matrix multiplication rule gives

$$(A^{k+1})_{uv} = \sum_w (A^k)_{uw} A_{wv}.$$

This equation states that to count the number of walks of length $k + 1$ from u to v, we sum over all intermediate nodes w that have a walk of length k from u to w and a direct edge from w to v, thus completing the proof. □

Example.

Consider the graph with adjacency matrix

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

Computing the square of the adjacency matrix,

$$A^2 = A \cdot A = \begin{pmatrix} 2 & 1 & 1 & 2 & 0 \\ 1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 \\ 2 & 1 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Each entry $(A^2)_{uv}$ counts the number of walks of length 2 between nodes u and v. The entry $(A^2)_{11} = 2$ indicates two distinct walks of length 2 from node 1 to itself.

For $k = 3$, we compute

$$A^3 = A \cdot A^2 = \begin{pmatrix} 2 & 3 & 3 & 1 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 1 & 3 & 3 & 3 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

The entry $(A^3)_{13} = 3$ shows that there are three walks of length 3 between nodes 1 and 3.

Higher powers of A determine connectivity properties. If $(A^k)_{uv} > 0$ for some k, then there exists a walk of length k between u and v. The smallest k such that all entries of A^k are positive gives the graph diameter. This is important for studying how quickly information, diseases, or influence spread in a network.

The structure of a network determines how efficiently information, influence, or even contagion spreads through it. In a social network, a message originating from a single node can propagate to its neighbors, and from there to their neighbors, in a cascading manner. The adjacency matrix powers indicate how long it takes for information to reach different parts of the network. If the graph is connected, then at some power A^k, all nodes will be reachable, meaning the message will have reached the entire network.

In real-world systems, the spreading process is not always deterministic. In the previous examples, the existence of an edge between two nodes indicated a direct connection, but this is not necessarily the case in dynamic systems such as social interactions or wireless communication networks. In such cases, edges may exist probabilistically, meaning that two nodes are connected only with a certain probability. This leads to the concept of random networks, where instead of a fixed adjacency structure, we model networks as probabilistic entities.

Definition. Random Network Model

A random network model is a probabilistic framework in which edges between nodes are assigned according to a stochastic rule. Formally, a random graph is a probability distribution over the set of all possible graphs with a given number of nodes.

Random network models allow us to study network structures emerging from random mechanisms, which is useful in applications such as the modeling of social interactions, epidemiology, and communication networks.

Example.

One of the most fundamental random network models is the Erdős-Rényi model, denoted as $G(n, p)$. In this model, a network is constructed by considering n nodes, and each possible edge between any pair of nodes is included independently with probability p.

Formally, let $V = \{1, 2, \ldots, n\}$ be a set of n nodes. The edge set E is generated randomly such that for each pair (u, v) with $u \neq v$,

$$P((u, v) \in E) = p.$$

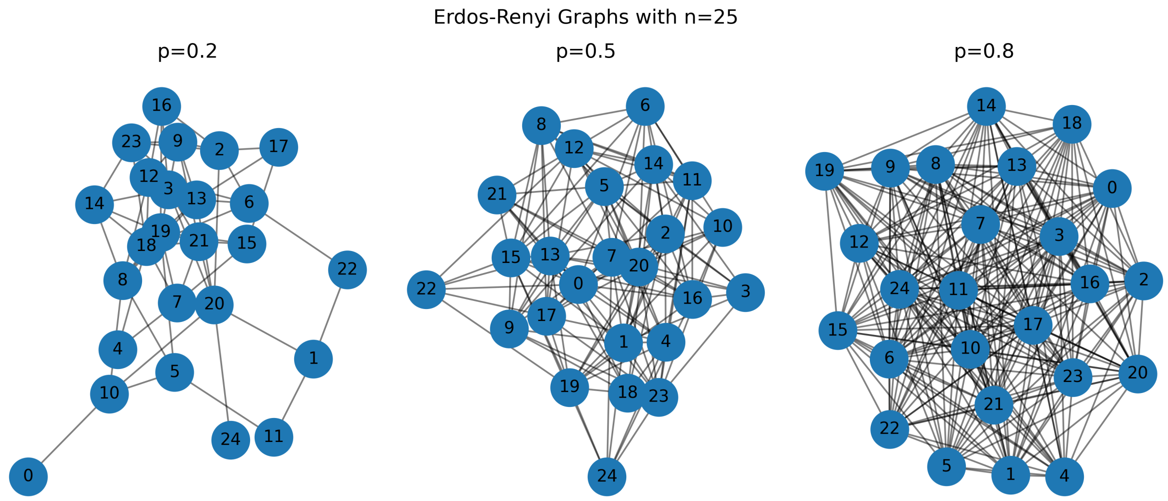

Figure 2: Variations of Erdős-Rényi graphs with $N = 25$ nodes for different values of p.

The Erdős-Rényi model provides a simple yet powerful way to analyze properties such as connectivity, clustering, and the emergence of giant components. When p is small, the network consists mostly of small disconnected components, whereas when p is large, most nodes become part of a single connected component.

Random network models describe the presence or absence of edges probabilistically, but in many real-world systems, the interactions between nodes are not simply about whether an edge exists. Instead, relationships may involve dependencies between random variables associated with the nodes, where one random variable influences another in a structured way.

Definition. Structural Equation Model (SEM)

A Structural Equation Model consists of a set of observed random variables $X_1, X_2, \ldots, X_n$ related through a system of equations:

$$X_i = f_i(X_{\text{pa}(i)}, \epsilon_i),$$

where:

- $\text{pa}(i)$ denotes the set of parent nodes influencing X_i,
- f_i is a function describing the relationship between X_i and its parents,
- ϵ_i is a noise term accounting for randomness in the system.

Structural equation models are particularly useful in describing causal relationships. Instead of assuming that edges are simply indicators of presence, they represent direct influences between variables.

Example.

Consider a network where three variables interact:

$$X_1 = \epsilon_1, \quad X_2 = aX_1 + \epsilon_2, \quad X_3 = bX_1 + cX_2 + \epsilon_3.$$

This system describes a process where X_1 influences X_2, and both X_1 and X_2 influence X_3. If we interpret this as a directed graph, we obtain:

$$X_1 \to X_2, \quad X_1 \to X_3, \quad X_2 \to X_3.$$

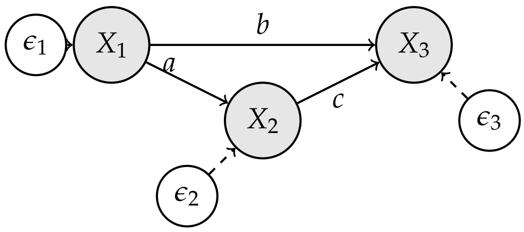

Figure 3: A simple Structural Equation Model (SEM) with three observed variables X_1, X_2, X_3 influenced by noise terms $\epsilon_1, \epsilon_2, \epsilon_3$.

In matrix notation, this system can be written as:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ b & c & 0 \end{pmatrix}, \quad \mathbf{X} = \mathbf{B}\mathbf{X} + \boldsymbol{\epsilon}.$$

Solving for $\mathbf{X}$, we obtain:

$$\mathbf{X} = (I - \mathbf{B})^{-1}\boldsymbol{\epsilon}.$$

Then, we have seen how networks and probability can be used to represent different types of interaction and relatedness; soon we will see that can be used as statics objects representing dynamical systems, but before that let's take a look at fundamentals dependence probability theory.