Network Science

Jingyuan Sun
ChatGPT Claude
Based on Prof. Laszlo Barabasi's lecture notes
Prof Vassil Alexandrov's taught

Nov, 2024

1 Intro to Network Science

Network Science is the study of complex systems composed of interconnected parts.

1.1 What is Complex?

1. Composed of many interconnected parts: a complex highway system.
2. Characterized by a very intricate arrangement of components: complex machinery.
3. So complicated as to be hard to understand or resolve: a complex problem.

1.2 Characteristics of Network Science

- **Interdisciplinary**: Combines knowledge and methods from multiple fields.
- **Empirical and Data-Driven**: Relies on real-world data to analyze and validate results.
- **Quantitative and Mathematical**: Uses mathematical models to study and predict system behaviors.
- **Computational**: Leverages computational tools to simulate and solve complex problems.

1.3 Applications and Societal Impact

- **Healthcare**: Designing drugs, predicting and tracking the spread of viruses.
- **Security and Cybersecurity**: Preventing terrorism and improving online security.
- **Economics and Management**: Supporting businesses like Amazon and Google with network-based strategies.
- **Social Networks**: Analyzing platforms such as Facebook, Twitter, and LinkedIn.

2 Random Networks

Random networks are a fundamental class of networks in network science, often used as a benchmark for studying complex systems.

2.1 Definition and Degree Distribution

A **random network** consists of N nodes, where each pair of nodes is connected with a probability p. This model is often referred to as the Erdős-Rényi (ER) model.

Mathematically, the degree distribution p_k describes the probability that a randomly chosen node has degree k. For a random network, p_k follows a **binomial distribution**:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \tag{1}$$

where k is the degree of the node, N is the total number of nodes, and p is the probability of a connection between two nodes. why N - 1: In Erdős-Rényi (ER) model, every node's connection most be N - 1(exclude itself).

2.2 Sparse Networks and Poisson Approximation

For large N and small p, where the network is **sparse** (i.e., $k \ll N$), the binomial distribution can be approximated by the **Poisson distribution**:

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \tag{2}$$

where $\langle k \rangle = p(N-1)$ is the average degree of the network.

Explanation: In a random network, most nodes have approximately the same degree, distributed around the mean $\langle k \rangle$. The degree distribution is symmetric and sharply peaked.

2.3 Characteristics of Random Networks

- Random networks are highly homogeneous: most nodes have a similar degree.
- Degree distribution follows a binomial or Poisson distribution for sparse networks.
- They are not robust to random node failure: removing random nodes can quickly break the network apart.

3 Scale-Free Networks

In contrast to random networks, **scale-free networks** exhibit a degree distribution that follows a **power law**. These networks are prevalent in real-world systems such as the internet, biological networks, and social networks.

3.1 Power Law Degree Distribution

The degree distribution p_k of a scale-free network is given by:

$$p_k \sim k^{-\gamma}, \tag{3}$$

where γ is the degree exponent, typically in the range $2 < \gamma < 3$.

Explanation: In scale-free networks, a few nodes (called **hubs**) have very high degrees, while most nodes have relatively few connections. This is in stark contrast to the uniform degree distribution of random networks.

3.2 Hubs and Robustness

The presence of hubs in scale-free networks gives rise to unique properties:

- **Robustness to Random Failures:** Removing random nodes does not significantly affect the network structure because hubs remain intact.
- **Vulnerability to Targeted Attacks:** Removing hubs can fragment the network.

3.3 Comparison with Random Networks

The key difference between scale-free and random networks lies in their degree distributions:

- **Random Networks:** Degree distribution is binomial or Poisson.
- **Scale-Free Networks:** Degree distribution follows a power law.

3.4 Example of the World Wide Web (WWW)

The WWW is modeled as a scale-free network, where a few hubs (highly connected websites) dominate the network. The probability of observing a node with degree k decreases as k increases, following the power law distribution:

$$p_k \sim k^{-\gamma}. \tag{4}$$

Conclusion: The scale-free nature of real-world networks makes them more robust to random failures but more susceptible to targeted attacks. Hubs play a crucial role in maintaining network connectivity.

4 Barabási–Albert Model

4.1 Growth and Preferential Attachment

The Barabási–Albert (BA) model introduces two fundamental mechanisms to explain the emergence of scale-free networks: **Growth** and **Preferential Attachment**. These mechanisms result in networks with a degree distribution following a power law.

Growth Unlike the ER (Erdős–Rényi) model, where the number of nodes N is fixed (static), the BA model assumes that networks expand over time. New nodes are continuously added to the network.

Preferential Attachment New nodes prefer to connect to nodes that already have a high degree. Mathematically, the probability $\Pi(k_i)$ that a new node connects to an existing node i with degree k_i is proportional to k_i:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{5}$$

Mechanism The BA model can be summarized as follows:

1. **Start with an initial network** of m_0 nodes.
2. **Growth**: At each time step, a new node is added to the network and connects to m existing nodes.
3. **Preferential Attachment**: The probability of connecting to node i depends on its degree k_i, as defined in Equation (1).

Resulting Degree Distribution The combination of growth and preferential attachment leads to a degree distribution that follows a power law:

$$p_k \sim k^{-\gamma}, \quad \text{where } \gamma \approx 3. \tag{6}$$

This means that a few nodes (hubs) will have a very high degree, while most nodes will have a low degree.

Real-World Examples The BA model explains the emergence of scale-free networks in various real-world systems:

- **World Wide Web (WWW)**: Nodes represent web pages, and edges represent hyperlinks.
- **Citation Network**: Nodes represent papers, and edges represent citations.
- **Actor Network**: Nodes represent actors, and edges represent shared movies.

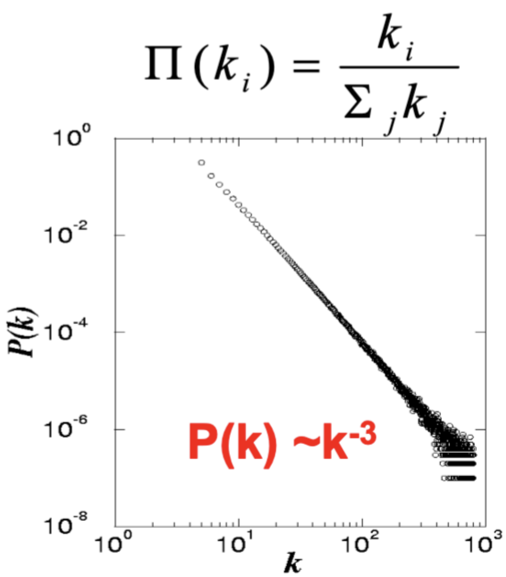

Figure 1: Degree Distribution $\sim\ k^{-3}$

Intuitive Explanation

- **Growth**: Networks naturally expand as new nodes are added.
- **Preferential Attachment**: New nodes prefer to link to well-connected nodes (hubs), which causes these nodes to become even more connected.

This mechanism mimics the "rich get richer" phenomenon observed in many real-world systems.

4.2 Degree Dynamics

In the Barabási–Albert (BA) model, the degree of a node evolves over time following a specific growth law. This results from the combined effects of **growth** and **preferential attachment**. Nodes that join the network earlier accumulate more links over time, leading to a degree distribution with the **first-mover advantage**.

Growth Law and Derivation For a node i, its degree $k_i(t)$ at time t evolves according to the following differential equation:

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A\frac{k_i}{\sum_j k_j}, \tag{7}$$

where:

- $\Pi(k_i)$ is the probability that a new node connects to node i, proportional to its current degree k_i,
- $\sum_j k_j$ is the total degree of all nodes in the network.

Since the total degree in the network at time t is given by:

$$\sum_j k_j = 2mt, \tag{8}$$

the growth equation simplifies to:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \tag{9}$$

Solving the Differential Equation The equation $\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$ is separable. Rearranging:

$$\frac{\partial k_i}{k_i} = \frac{\partial t}{2t}. \tag{10}$$

Integrating both sides from t_i (the time node i joins the network) to t, and assuming $k_i(t_i) = m$ (initial degree when the node joins), we get:

$$\int_{k_i(t_i)=m}^{k_i(t)} \frac{1}{k_i} \, \partial k_i = \int_{t_i}^{t} \frac{1}{2t'} \, \partial t'. \tag{11}$$

The solution is:

$$\ln \frac{k_i(t)}{m} = \frac{1}{2} \ln \frac{t}{t_i}. \tag{12}$$

Exponentiating both sides:

$$\frac{k_i(t)}{m} = \left(\frac{t}{t_i}\right)^{\frac{1}{2}}, \tag{13}$$

which simplifies to:

$$k_i(t) = m \left(\frac{t}{t_i}\right)^{\frac{1}{2}}. \tag{14}$$

Interpretation The degree $k_i(t)$ of a node grows proportionally to $t^{1/2}$. Nodes introduced earlier (t_i small) accumulate more links over time, leading to the **first-mover advantage**. This mechanism explains the emergence of hubs in scale-free networks.

Summary The key results from the degree dynamics in the BA model are:

- The degree $k_i(t)$ of a node increases over time following a power law: $k_i(t) \sim t^{1/2}$.
- Nodes added earlier have a significant advantage, resulting in the formation of **hubs**.
- This growth behavior leads to a scale-free network with a degree distribution that follows a power law.

4.3 Degree Distribution

Notation and Setup. Consider a growing network constructed as follows: at each time step t, a new node is introduced and connected to m existing nodes chosen with probability proportional to their degree. Let $N(t) = t$ represent the total number of nodes at time t. Define $P(k,t)$ as the probability (or fraction of nodes) that a randomly selected node at time t has degree k. In the stationary limit $t \to \infty$, we denote $P(k) = \lim_{t\to\infty} P(k,t)$.

The *degree distribution* of the network is given by the stationary probability $P(k)$ that a node chosen uniformly at random has degree k.

Rate Equations. At each time step, a single new node is added to the network. The number of nodes at time t is $N(t) = t$. The dynamics of $P(k,t)$ can be described by a set of discrete-time rate equations. For $k > m$, we have:

$$(N+1)P(k,t+1) = NP(k,t) + \frac{k-1}{2}P(k-1,t) - \frac{k}{2}P(k,t). \tag{15}$$

$N(t) = t$ is the number of nodes when time = t; $P(k,t)$ means the ratio of the number of nodes which have a k degree when time = t : total nodes. k/2 means preferred connection probability For $k = m$, since each newly introduced node arrives with degree m, the rate equation takes a slightly different form:

$$(N+1)P(m,t+1) = NP(m,t) + 1 - \frac{m}{2}P(m,t). \tag{16}$$

Stationary State. In the stationary limit $t \to \infty$, set $P(k, t+1) = P(k, t) = P(k)$ and $N = t$. Substituting into (15) and (16), we obtain time-independent equations. In particular, (15) reduces to a recursive relation for $k > m$:

$$P(k) = \frac{k-1}{k+2} P(k-1). \tag{17}$$

From (16) and the stationary condition for $k = m$, one can determine the normalization and boundary conditions necessary to solve (17).

[Degree distribution of the Barabási–Albert model] In the Barabási–Albert (BA) model with parameter $m \geq 1$, the stationary degree distribution is:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{18}$$

For large k, this asymptotically behaves as:

$$P(k) \sim k^{-3}. \tag{19}$$

First, we use (16) in the stationary limit to solve for $P(m)$:

$$(N+1)P(m) = NP(m) + 1 - \frac{m}{2} P(m) \implies P(m) = \frac{2}{m+2}.$$

Next, from (17), for $k > m$:

$$P(k) = \frac{k-1}{k+2} P(k-1).$$

Iterating this recursion starting from $k = m+1$, we have:

$$P(m+1) = \frac{m}{m+3} P(m), \quad P(m+2) = \frac{m+1}{m+4} P(m+1), \quad \text{etc.}$$

This telescopes into a product yielding:

$$P(k) = \frac{2}{m+2} \cdot \frac{m(m+1)}{(m+2)(m+3)(m+4)\cdots(k+2)}.$$

Careful simplification shows that the general form for all $k \geq m$ is:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)},$$

after reindexing and simplifying the product terms. As $k \to \infty$, we have $P(k) \sim 2m(m+1)/k^3$, confirming the power-law decay with exponent 3.

The exponent $\gamma = 3$ is independent of m, indicating a universal scaling behavior. Furthermore, the degree distribution $P(k)$ does not depend on time t or the total system size N, which implies that the network reaches a stationary scale-free state. The proportionality constant of the power-law distribution scales as m^2, ensuring that variations in m affect only the prefactor, but not the scaling exponent.

4.4 The Necessity of Growth and Preferential Attachment

In the Barabási–Albert (BA) model, **growth** and **preferential attachment** are two essential mechanisms for generating scale-free networks. Each mechanism contributes uniquely, and their combination is critical to reproducing the power-law degree distribution observed in real-world networks.

1. Growth Alone - **Mechanism**: Growth refers to the continuous addition of new nodes to the network. At each step, new nodes connect to existing nodes with a fixed number of edges. - **Limitation**: Without preferential attachment, new nodes would connect randomly, and the degree distribution would resemble a **Poisson distribution**, as seen in Erdős–Rényi random networks. - This implies that most nodes would have a similar degree, and no "hubs" (high-degree nodes) would emerge. - The network would lack the long-tail distribution characteristic of scale-free networks.

2. Preferential Attachment Alone - **Mechanism**: Preferential attachment ensures that new nodes are more likely to connect to existing nodes with higher degrees. This mechanism embodies the "rich-get-richer" principle. - **Limitation**: Without growth, the network remains static, with no new nodes being added. While preferential attachment can increase the disparity in node degrees, it does not create a dynamic network or expand its size over time. - Real-world networks, such as social media or the World Wide Web, expand continuously with new participants or pages being added.

3. Combined Mechanisms in the BA Model The combination of growth and preferential attachment overcomes these limitations: 1. **Growth** provides the dynamic expansion observed in real-world networks. New nodes continuously join, increasing the network's size and complexity. 2. **Preferential attachment** introduces heterogeneity into the degree distribution, allowing the emergence of hubs (nodes with exceptionally high degrees). 3. Together, they generate a network with: - Dynamic growth and expanding scale. - A long-tail degree distribution that follows a power law.

4. Real-World Evidence Supporting Both Mechanisms - **Internet**: New websites are continuously created (growth) and tend to link to well-established, highly connected sites like Google or Wikipedia (preferential attachment). - **Citation Networks**: New academic papers are published regularly (growth) and often cite highly cited, influential papers (preferential attachment). - **Social Networks**: New users join platforms like Facebook or Twitter (growth) and are more likely to follow popular influencers or celebrities (preferential attachment).

5. Mathematical Complementarity From a mathematical perspective: 1. **Growth provides the initial conditions**: Growth introduces time-dependent terms into the degree equations (e.g., $k_i(t) = m(t/t_i)^{1/2}$). 2. **Preferential attachment creates heterogeneity**: The selection probability $\Pi(k) \propto k$ ensures that high-degree nodes grow faster, resulting in a non-uniform, power-law degree distribution.

4.5 Measuring Preferential Attachment

In the Barabási–Albert (BA) model, preferential attachment plays a crucial role. To verify this mechanism, it is necessary to measure the attachment probability $\Pi(k)$, which represents the probability of a node with degree k gaining new links over time.

1. Measuring $\Pi(k)$ The rate of degree growth for a node i is proportional to $\Pi(k_i)$:

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i), \tag{20}$$

where $\Pi(k) \sim \frac{\Delta k_i}{\Delta t}$ represents the rate at which nodes with degree k acquire new connections in a fixed time interval Δt.

To measure $\Pi(k)$:

- Track the degree of nodes at time t and $t + \Delta t$.
- Calculate the number of new links Δk for nodes with degree k during Δt.
- Plot $\Delta k/\Delta t$ as a function of k.

2. Reducing Noise To reduce noise in the data, compute the cumulative sum of $\Pi(k)$ over all degrees less than k:

$$\kappa(k) = \sum_{K<k} \Pi(K). \tag{21}$$

This approach yields:

- **No preferential attachment:** $\kappa(k) \sim k$ (linear growth).
- **Linear preferential attachment:** $\kappa(k) \sim k^2$ (quadratic growth).

3. Empirical Results The following networks have been analyzed to measure $\kappa(k)$ and $\Pi(k)$:

- **Citation Network:** Connections based on paper citations.
- **Internet:** Nodes represent websites.
- **Neuroscience Collaboration:** Collaborations in neuroscience research.
- **Actor Collaboration:** Networks of co-acting in movies.

These networks demonstrate $\kappa(k) \sim k^2$, confirming the presence of linear preferential attachment.

4. General Form of $\Pi(k)$ The attachment probability can be expressed generally as:

$$\Pi(k) \approx A + k^{\alpha}, \quad \alpha \leq 1. \tag{22}$$

Linear preferential attachment corresponds to $\alpha = 1$, explaining the emergence of scale-free properties in real-world networks.

4.6 Nonlinear Preferential Attachment

In nonlinear preferential attachment, the probability of a new node connecting to an existing node with degree k is proportional to:

$$\Pi(k) \sim k^{\alpha}, \tag{23}$$

where α is a parameter controlling the dependency of the attachment probability on the degree.

1. Cases Based on α:

- $\alpha = 0$: Random attachment, where the degree distribution follows a simple exponential function.
- $\alpha = 1$: Linear preferential attachment (Barabási–Albert model), generating a scale-free network with degree exponent $\gamma = 3$.
- $0 < \alpha < 1$: **Sublinear Preferential Attachment**. New nodes slightly favor higher-degree nodes, leading to a stretched exponential degree distribution:

  $$p_k \sim k^{-\alpha} \exp\left(-\frac{2\mu(\alpha)}{\langle k\rangle(1-\alpha)} k^{1-\alpha}\right),$$

  and the maximum degree grows logarithmically:

  $$k_{\max} \sim (\ln t)^{1/(1-\alpha)}.$$

- $\alpha > 1$: **Superlinear Preferential Attachment**. New nodes strongly prefer connecting to high-degree nodes, accelerating the "rich-get-richer" process and resulting in a winner-takes-all effect. The maximum degree grows linearly:

  $$k_{\max} \sim t.$$

2. Simulation Results Numerical simulations demonstrate the impact of α on the growth of the maximum degree:

- $\alpha = 0.5$: $k_{\max} \sim (\ln t)^2$.
- $\alpha = 1$: $k_{\max} \sim t^{1/2}$.
- $\alpha = 2.5$: $k_{\max} \sim t$.