Network Science

Lecture 6

17 October, 2024

The Barabási-Albert Model

A Detailed Analysis of Scale-Free Networks

Understanding the structure of complex networks (1)

The Barabási-Albert model represents a key breakthrough in understanding the structure of complex networks:

- It explains why many real-world networks, like the internet or social networks, are scale-free.
- Scale-free networks have a few highly connected nodes, or hubs, and many nodes with fewer connections.
- This model starts with a small number of connected nodes and grows by adding new nodes over time.
- When new nodes are added, they are more likely to connect to highly connected nodes, a process called preferential attachment.
- This process explains the “rich-get-richer” phenomenon seen in many systems, where a few entities become dominant.
- The model reveals how hubs emerge naturally over time, without any central control.

Understanding the structure of complex networks (2)

- Real-world networks, like airline routes or the World Wide Web, show similar hub structures.
- These hubs play a critical role in the network's resilience, ensuring the network stays connected even if most nodes fail.
- The model also explains why these networks are vulnerable to targeted attacks on hubs.
- The average distance between any two nodes in a scale-free network is relatively small, a property known as the "small-world effect."
- This property allows information or diseases to spread quickly across the network.
- In technology, this model helps explain the structure of the internet, where a few websites have most of the links.
- In biology, it shows why certain proteins are more essential than others in cellular networks.

Understanding the structure of complex networks (3)

- Social networks exhibit similar characteristics, with influential individuals (hubs) playing central roles in information flow.
- Mathematically, the degree distribution of a Barabási-Albert network follows a power law, which is a key feature of scale-free networks.
- The model has been extended to explain networks with directed links, weighted links, and evolving topologies.
- Understanding the dynamics of this model helps improve network design and predict network failures.
- Its applications range from epidemiology, where it models the spread of diseases, to economics, where it helps understand market dynamics.
- The Barabási-Albert model is one of the cornerstones of modern network science and has influenced various disciplines.
- Future work involves exploring how networks evolve under different conditions and improving the robustness of these systems against failures.

Hubs represent the most striking difference between a random and a scale-free network. On the World Wide Web, they are websites with an exceptional number of links, like google.com or facebook.com; in the metabolic network they are molecules like ATP or ADP, energy carriers involved in an exceptional number of chemical reactions. The very existence of these hubs and the related scale-free topology raises two fundamental questions:

Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?

Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in real networks?

The first question is particularly puzzling given the fundamental differences in the nature, origin, and scope of the systems that display the scale-free property.

- The *nodes* of the cellular network are metabolites or proteins, while the nodes of the WWW are documents, representing information without a physical manifestation.

- The *links* within the cell are chemical reactions and binding interactions, while the links of the WWW are URLs, or small segments of computer code.

- The *history* of these two systems could not be more different: The cellular network is shaped by 4 billion years of evolution, while the WWW is less than three decades old.

- The *purpose* of the metabolic network is to produce the chemical components the cell needs to stay alive, while the purpose of the WWW is information access and delivery.

To understand why so *different* systems converge to a *similar* architecture we need to first understand the mechanism responsible for the emergence of the scale-free property.

Given the diversity of the systems that display the scale-free property, the explanation must be simple and fundamental.

The answers will change the way we model networks, forcing us to move from describing a network's topology to modeling the evolution of a complex system.

Networks Expand Through the Addition of New Nodes

The random network model assumes that we have a *fixed* number of nodes, N.

Yet, *in real networks the number of nodes continually grows thanks to the addition of new nodes*.

Few examples

1. In 1991 the WWW had a single node, the first webpage build by Tim Berners-Lee, the creator of the Web.

Today the Web has over a trillion documents, an extraordinary number that was reached through the continuous addition of new documents by millions of individuals and institutions

2. The collaboration and the citation network continually expands through the publication of new research papers.

3. The actor network continues to expand through the release of new movies.

4. The protein interaction network may appear to be static, as we inherit our genes (and hence our proteins) from our parents. Yet, it is not: The number of genes grew from a few to the over 20,000 genes present in a human cell over four billion years.

Consequently, if we wish to model these networks, we cannot resort to a static model. Our modeling approach must instead acknowledge that networks are the product of a steady growth process.

In summary, the random network model differs from real networks in two important characteristics:

(A) Growth

Real networks are the result of a growth process that continuously increases N. In contrast the random network model assumes that the number of nodes, N, is fixed.

(B) Preferential Attachment

In real networks new nodes tend to link to the more connected nodes. In contrast nodes in random networks randomly choose their interaction partners.

These two differences between real and random networks, *growth* and *preferential attachment*, play a particularly important role in shaping a network's degree distribution.

THE BARABÁSI-ALBERT MODEL

The recognition that growth and preferential attachment coexist in real networks has inspired a minimal model called the *Barabási-Albert* model, which can generate scale-free networks. Also known as the *BA model* or the *scale-free model*, it is defined as follows:

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps:

(A) Growth

At each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network.

B) Preferential attachment

The probability $\Pi(k)$ that a link of the new node connects to node i depends on the degree k as

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

Preferential attachment is a probabilistic mechanism: A new node is free to connect to *any* node in the network, whether it is a hub or has a single link. The previous equation implies, however, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

While most nodes in the network have only a few links, a few gradually turn into hubs. These hubs are the result of a *rich-gets-richer phenomenon*: Due to preferential attachment new nodes are more likely to connect to the more connected nodes than to the smaller nodes. Hence, the larger nodes will acquire links at the expense of the smaller nodes, eventually becoming hubs.

In summary, the Barabási-Albert model indicates that two simple mechanisms, *growth* and *preferential attachment*, are responsible for the emergence of scale-free networks.

The origin of the power law and the associated hubs is a *rich-gets-richer phenomenon* induced by the coexistence of these two ingredients. To understand the model's behavior and to quantify the emergence of the scale-free property, we need to become familiar with the model's mathematical properties.

DEGREE DYNAMICS

To understand the emergence of the scale-free property, we need to focus on the time evolution of the Barabási-Albert model. We begin by exploring the time-dependent degree of a single node.

In the model an existing node can increase its degree each time a *new* node enters the network. This new node will link to m of the $N(t)$ nodes already present in the system. The probability that one of these links connects to node i.

Let us approximate the degree k_i with a continuous real variable, representing its expectation value over many realizations of the growth process. The rate at which an existing node i acquires links as a result of new nodes connecting to it is

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

The coefficient m describes that each new node arrives with m links. Hence, node i has m chances to be chosen. The sum in the denominator of goes over all nodes in the network except the newly added node, thus $\sum_{j=1}^{N-1} k_j = 2mt - m$

Therefore $\frac{dk_i}{dt} = \frac{k_i}{2t-1}$.

For large t the (-1) term can be neglected in the denominator, obtaining

$$\frac{dk_i}{k_i} = \frac{1}{2}\frac{dt}{t}$$

We obtain $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$.

We call β the *dynamical exponent* and has the value $\beta = \frac{1}{2}$.

The previous equation offers a number of predictions:

1. The degree of each node increases following a power-law with the same dynamical exponent $\beta = 1/2$ Hence all nodes follow the same dynamical law.
2. The growth in the degrees is sublinear (i.e. $\beta < 1$). This is a consequence of the growing nature of the Barabási-Albert model: Each new node has more nodes to link to than the previous node. Hence, with time the existing nodes compete for links with an increasing pool of other nodes.

DEGREE DISTRIBUTION

The distinguishing feature of the networks generated by the Barabási-Albert model is their power-law degree distribution. We can calculate the functional form of p, helping us understand its origin.

A number of analytical tools are available to calculate the degree distribution of the Barabási-Albert network. The simplest is the *continuum theory*. It predicts the degree distribution,

$$p(k) \approx 2m^{1/\beta} k^{-\gamma}$$

with

$$\gamma = \frac{1}{\beta} + 1 = 3.$$

THE ABSENCE OF GROWTH OR PREFERENTIAL ATTACHMENT

The coexistence of growth and preferential attachment in the Barabási-Albert model raises an important question: Are they both necessary for the emergence of the scale-free property? In other words, could we generate a scale-free network with only one of the two ingredients? To address these questions, next we discuss two limiting cases of the model, each containing only one of the two ingredients.

MODEL A

To test the role of preferential attachment we keep the growing character of the network (ingredient A) and eliminate preferential attachment (ingredient B). Hence, *Model A* starts with m_0 nodes and evolves following these steps:

(A)Growth

At each time step we add a new node with $m(\leq m_0)$ links that connect to m nodes added earlier.

(B)Preferential Attachment

The probability that a new node links to a node with degree k_i is $\Pi(k_i) = \frac{1}{(m_0 + t - 1)}$.

That is, $\Pi(k_i)$ is independent of k_i, indicating that new nodes choose randomly the nodes they link to.

The continuum theory predicts that for Model A $k_i(t)$ increases logarithmically with time

$$k_i(t) = m \ln\left(e \frac{m_0 + t - 1}{m_0 + t_i - 1}\right)$$

a much slower growth than the power law increase. Consequently the degree distribution follows an exponential

$$p(k) = \frac{e}{m} \exp\left(-\frac{k}{m}\right).$$

An exponential function decays much faster than a power law, hence it does not support hubs. Therefore the lack of preferential attachment eliminates the network's scale-free character and the hubs. Indeed, as all nodes acquire links with equal probabilty, we lack a rich-get-richer process and no clear winner can emerge.

MODEL B

To test the role of growth next we keep preferential attachment (ingredient B) and eliminate growth (ingredient A). Hence, *Model B* starts with N nodes and evolves following this step:

(B) Preferential Attachment

At each time step a node is selected randomly and connected to node i is chosen with probability $\Pi(k)$. As $\Pi(0)=0$ nodes with $k=0$ are assumed to have $k=1$, otherwise they can not acquire links.

In Model B the number of nodes remains constant during the network's evolution, while the number of links increases linearly with time. As a result for large t the degree of each node also increases linearly with time.

$$k_i(t) \approx \frac{2}{N} t .$$

Indeed, in each time step we add a new link, without changing the number of nodes.

At early times, when there are only a few links in the network (i.e. $L \ll N$), each new link connects previously unconnected nodes. In this stage the model's evolution is indistinguishable from the Barabási-Albert model with m=1. Numerical simulations show that in this regime the model develops a degree distribution with a power-law tail.

Yet, p_k is not stationary. Indeed, after a transient period the node degrees converge to the average degree and the degree develops a peak. For $t \to N(N\text{-}1)/2$ the network becomes a complete graph in which all nodes have degree $k_{max}=N\text{-}1$, hence $p_k = \delta(N\text{-}1)$.

In summary, the absence of preferential attachment leads to a growing network with a stationary but exponential degree distribution.

In contrast the absence of growth leads to the loss of stationarity, forcing the network to converge to a complete graph.

This failure of Models A and B to reproduce the empirically observed scale-free distribution indicates that growth and preferential attachment are simultaneously needed for the emergence of the scale-free property.

MEASURING PREFERENTIAL ATTACHMENT

We showed that growth and preferential attachment are jointly responsible for the scale-free property. The presence of growth in real systems is obvious: All large networks have reached their current size by adding new nodes. But to convince ourselves that preferential attachment is also present in real networks, we need to detect it experimentally. We show how to detect preferential attachment by measuring the $\Pi(k)$ function in real networks.

Preferential attachment relies on **two** distinct hypotheses:

Hypothesis 1

The likelihood to connect to a node depends on that node's degree k. This is in contrast with the random network model, for which $\Pi(k)$ is independent of k.

Hypothesis 2

The functional form of $\Pi(k)$ is linear in k.

Both hypotheses can be tested by measuring $\Pi(k)$. We can determine $\Pi(k)$ for systems for which we know the time at which each node joined the network, or we have at least two network maps collected at not too distant moments in time.

In summary, we can detect the presence (or absence) of preferential attachment in real networks. The measurements show that the attachment probability depends on the node degree. We also find that while in some systems preferential attachment is linear, in others it can be sublinear.

NON-LINEAR PREFERENTIAL ATTACHMENT

The observation of sublinear preferential attachment raises an important question: What is the impact of this nonlinearity on the network topology? To answer this we replace the linear preferential attachment and calculate the degree distribution of the obtained *nonlinear Barabási-Albert model*.

The behavior for $\alpha=0$ is clear: In the absence of preferential attachment we are back to Model A. Consequently the degree distribution follows the exponential.

For $\alpha = 1$ we recover the Barabási-Albert model, obtaining a scale-free network with degree distribution.

THE ORIGINS OF PREFERENTIAL ATTACHMENT

Given the key role preferential attachment plays in the evolution of real networks, we must ask, where does it come from? The question can be broken to two narrower issues:

Why does $\Pi(k)$ depend on k?

Why is the dependence of $\Pi(k)$ linear in k?

In the past decade we witnessed the emergence of two philosophically different answers to these questions. The first views preferential attachment as the interplay between random events and some structural property of a network.

These mechanisms do not require global knowledge of the network but rely on random events, hence we will call them *local* or *random* mechanisms. The second assumes that each new node or link balances conflicting needs, hence they are preceeded by a cost-benefit analysis. These models assume familiarity with the whole network and rely on optimization principles, prompting us to call them *global* or *optimized* mechanisms.We discuss both approaches.

LOCAL MECHANISMS

The Barabási-Albert model postulates the presence of preferential attachment. Yet, as we show below, we can build models that generate scale- free networks apparently without preferential attachment. They work by *generating* preferential attachment. We discuss two such models and derive $\Pi(k)$ for them, allowing us to understand the origins of preferential attachment.

Link Selection Model

The *link selection model* offers perhaps the simplest example of a local mechanism that generates a scale-free network without preferential attachment. It is defined as follows:

- *Growth*: At each time step we add a new node to the network.

Link Selection: We select a link at random and connect the new node to one of the two nodes at the two ends of the selected link.

The model requires no knowledge about the overall network topology, hence it is inherently local and random. Unlike the Barabási-Albert model, it lacks a built-in $\Pi(k)$ function. Yet, next we show that it generates preferential attachment.

- We start by writing the probability q_k that the node at the end of a randomly chosen link has degree k as $q_k = Ckp_k$.

It is the probability that a new node connects to a node with degree k. The fact that the bias in is linear in k indicates that the link selection model builds a scale-free network by generating linear preferential attachment.

Copying Model

While the link selection model offers the simplest mechanism for preferential attachment, it is neither the first nor the most popular in the class of models that rely on local mechanisms. That distinction goes to the *copying model*. The model mimics a simple phenomena: The authors of a new webpage tend to borrow links from other webpages on related topics. It is defined as follows:

The probability of selecting a particular node in step (i) is $1/N$. Step (ii) is equivalent with selecting a node linked to a randomly selected link. The probability of selecting a degree-k node through this copying step (ii) is $k/2L$ for undirected networks. Combining (i) and (ii), the likelihood that a new node connects to a degree-k node follows

$$\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k,$$

which, being linear in k, predicts a linear preferential attachment.

The popularity of the copying model lies in its relevance to real systems:

- *Social Networks*: The more acquaintances an individual has, the higher is the chance that she will be introduced to new individuals by her existing acquaintances. In other words, we "copy" the friends of our friends. Consequently without friends, it is difficult to make new friends.
- *Citation Networks*: No scientist can be familiar with all papers published on a certain topic. Authors decide what to read and cite by "copying" references from the papers they have read. Consequently papers with more citations are more likely to be studied and cited again.
- *Protein Interactions*: Gene duplication, responsible for the emergence of new genes in a cell, can be mapped into the copying model, explaining the scale-free nature of protein interaction networks.

Taken together, we find that both the link selection model and the copying model generate a linear preferential attachment through random linking.

A longstanding assumption of economics is that humans make rational decisions, balancing cost against benefits. In other words, each individual aims to maximize its personal advantage. This is the starting point of rational choice theory in economics and it is a hypothesis central to modern political science, sociology, and philosophy. As we show below, such rational decisions can lead to preferential attachment.

Consider the Internet, whose nodes are routers connected via cables. Establishing a new Internet connection between two routers requires us to lay down a new cable between them. As this is costly, each new link is preceded by a careful cost-benefit analysis. Each new router (node) will choose its link to balance access to good network performance (i.e. proper bandwith) with the cost of laying down a new cable (i.e. physical distance). This can be a conflicting desire, as the closest node may not offer the best network performance.

DIAMETER AND CLUSTERING COEFFICIENT

To complete the characterization of the Barabási-Albert model we discuss the behavior of the network diameter and the clustering coefficient.

Diameter

The network diameter, representing the maximum distance in the Barabási-Albert network, follows for $m > 1$ and large N

$$\langle d \rangle \sim \frac{\ln N}{\ln\ln N}.$$

Therefore the diameter grows slower than $\ln N$, making the distances in the Barabási-Albert model smaller than the distances observed in a random graph of similar size. The difference is particularly relevant for large N.

Note that the average distance $\langle d \rangle$ scales in a similar fashion. Indeed, for small N the $\ln N$ term captures the scaling of $\langle d \rangle$ with N, but for large $N(\geq 10^4)$ the impact of the logarithmic correction $\ln \ln N$ becomes noticeable.

Clustering coefficient

The clustering coefficient of the Barabási-Albert model

$$\langle C \rangle \sim \frac{(\ln N)^2}{N}$$

The prediction is quite different from the $1/N$ dependence obtained for the random network model. The difference comes in the $(\ln N)^2$ term, that increases the clustering coefficient for large N. Consequently the Barabási-Albert network is locally more clustered than a random network.

SUMMARY

The most important message of the Barabási-Albert model is that network structure and evolution are inseparable. Indeed, in the Erdős-Rényi, Watts-Strogatz, the configuration and the hidden parameter models the role of the modeler is to cleverly place the links between a *fixed number of nodes*. Returning to our earlier analogy, the networks generated by these models relate to real networks like a photo of a painting relates to the painting itself: It may look like the real one, but the process of generating a photo is drastically different from the process of painting the original painting. The aim of the Barabási-Albert model is to capture the processes that assemble a network in the first place. Hence, it aims to paint the painting again, coming as close as possible to the original brush strokes. Consequently, the modeling philosophy behind the model is simple: *to understand the topology of a complex system, we need to describe how it came into being*.

Random networks, the configuration and the hidden parameter models will continue to play an important role as we explore how certain network characteristics deviate from our expectations. Yet, if we want to explain the origin of a particular network property, we will have to use models that capture the system's genesis.

The Barabási-Albert model raises a fundamental question: Is the combination of growth and preferential attachment the real reason why networks are scale-free? We offered a *necessary* and *sufficient* argument to address this question. First, we showed that growth and preferential attachment are jointly needed to generate scale-free networks, hence if one of them is absent, either the scale-free property or stationarity is lost. Second, we showed that if they are both present, they do lead to scale-free networks. This argument leaves one possibility open, however: Do these two mechanisms explain the scale-free nature of *all* networks? Could there be some real networks that are scale-free thanks to some completely different mechanism? The answer is provided: we did encountered the link selection, the copying and the optimization models that do not have a preferential attachment function built into them, yet they do lead to a scale-free network.

We showed that they do so by generating a linear $\Pi(k)$. This finding underscores a more general pattern: To date all known models and real systems that are scale-free have been found to have preferential attachment. Hence the basic mechanisms of the Barabási-Albert model appear to capture the origin of their scale-free topology.

The Barabási-Albert model is unable to describe many characteristics of real systems:

- The model predicts γ=3 while the degree exponent of real networks varies between 2 and 5
- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.
- Many processes observed in networks, from linking to already existing nodes to the disappearance of links and nodes, are absent from the model.
- The model does not allow us to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage.

- While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality it is not designed to capture the details of any particular real network. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property. Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal.

We can show, these limitations can be systematically resolved.