COSC 421/COSC 521

Network Models: Preferential Attachment

Dr. Ifeoma Adaji

I respectfully acknowledge that UBC Okanagan is situated on the traditional, unceded and ancestral territory of the [Syilx Okanagan Nation](#)

iClicker Question

Larger random networks have Poisson distribution

A. True

B. False

iClicker Question

During the evolution of a random network, at the subcritical point, which one of these is correct?

A. *$p < 1/n$*
B. *$c = 1$*
C. *There is at least one giant component*

iClicker Question

During the evolution of a random network, at the supercritical point, which one of these is **not** correct?

A. $c > 1$

B. $p > 1/n$

C. This stage has the most relevance to real networks

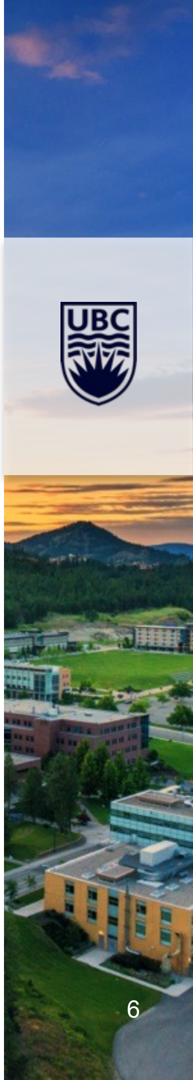
D. First appearance of a giant component that looks like a network happens at this stage

E. None of the above is wrong

Objectives

At the end of today’s class, you should be able to:

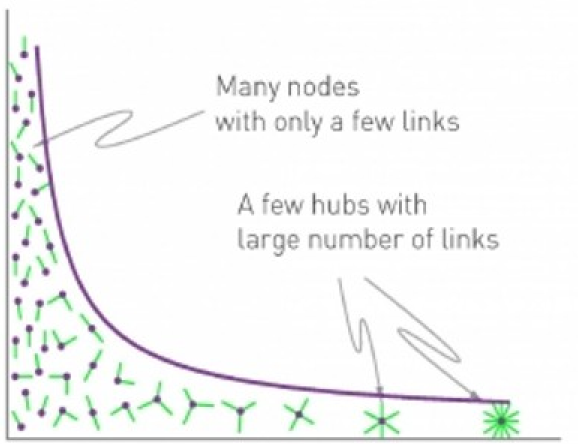
- Describe the Barabási-Albert model of network formation
- Identify the ingredients/factors needed for the Barabási-Albert model
- Describe what happens in the absence of these factors



Recall - Degree Distribution

- The histograms of many networks look like this

Fraction $P(k)$ of nodes with degree K



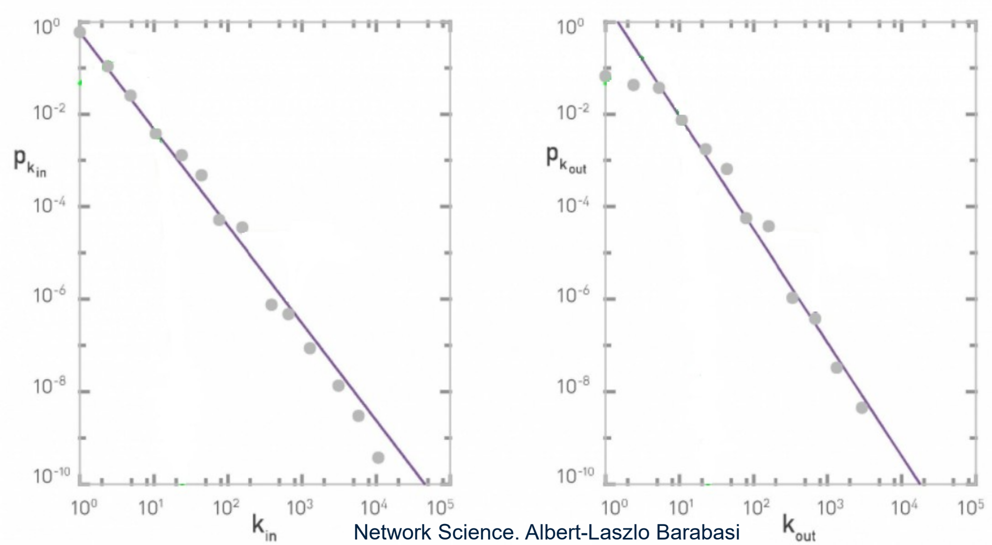
Degree K

- Most of the nodes in a network have low degree
- There is a significant long "tail" distribution
    - nodes with significantly high degrees

Network Science. Albert-Laszlo Barabasi

Recall - Power Law

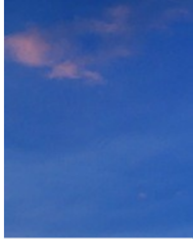
- The degree distribution of many networks follow the power law distribution
- Example, in-degree and out-degree distribution of the WWW mapped in 1999
  - Shown on double logarithmic axis (log-log plot)
- Can be written as $p_k = C\,k^{-\alpha}$
  - C is a constant determined by the normalization condition
  - α is the exponent of the power law; typically between 2 and 3



Network Science. Albert-Laszlo Barabasi

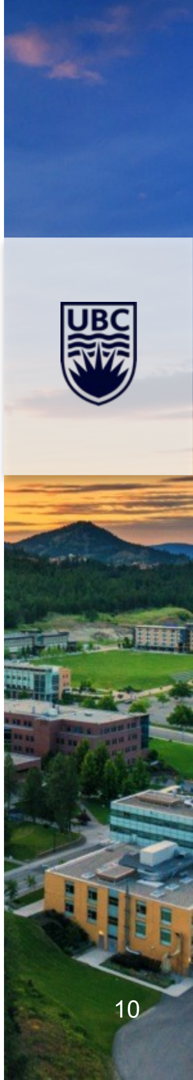
Power Law

- Occurrence of power law in empirical data is often considered a potential indicator of interesting underlying processes
- A natural question to ask is **how can we create a network to have such distribution?**
- This question was considered in the 1970s by Derek John de Solla Price, a British Physicist
  - He proposed a model of network formation that gives rise to power-law degree distributions
- Price's interest was in citation networks



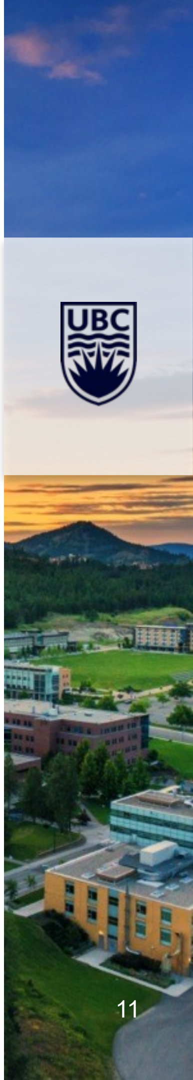
Power Law

- Price was inspired by work of Herbert Simon
    - Power law in economic data such as distribution of people's wealth
    - Simon's explanation for wealth distribution is based on idea that people who already have a lot of money gain more at a rate proportional to how much they currently have
- Price adapted Simon's method in the context of networks
- Price named Simon's mechanism cumulative advantage; today it's known as **preferential attachment**



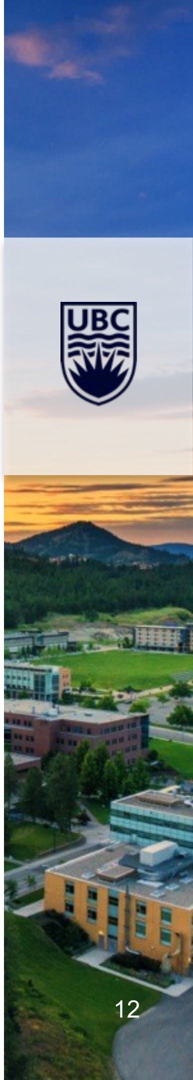
Preferential Attachment

- The more connected a node is, the more likely it is to receive new edges
- Rich get richer



Preferential Attachment

- There are several models for generating random scale-free networks using preferential attachment
- The Barabási-Albert Model
  - Tries to explain the existence of hubs in real networks such as WWW, and social networks using growth and preferential attachment
- Price’s Model
  - Based on citation networks
  - The way existing papers get new citation is proportional to the number of existing citations the paper already has; papers with many citations should be cited more than papers with few citations
- These models enable us generate scale free networks that are similar to real world networks

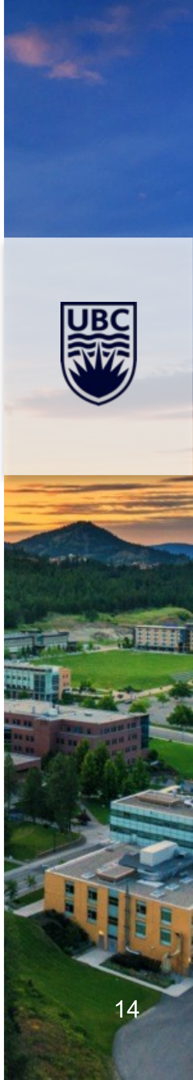


The Barabási-Albert Model of Preferential Attachment

The Barabási-Albert model

The random network model differs from real networks in two important characteristics. The BA model indicates that these 2 mechanisms are responsible for the emergence of scale-free networks

- **Growth:** While the random network model assumes that the number of nodes is fixed, real networks are the result of a growth process that continuously increases
- **Preferential Attachment:** While nodes in random networks randomly choose their interaction partner, in most real networks new nodes prefer to link to the more connected nodes
    - Ex. In WWW, new pages link to hubs/well known sites such as Google
    - The probability of selecting a random page on the WWW is proportional to its degree



The Barabási-Albert model

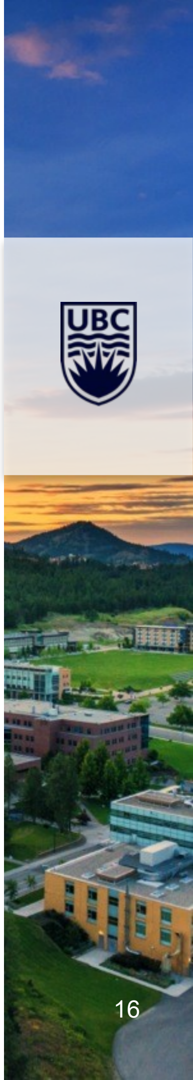
We start with n_0 nodes, links between them are chosen arbitrarily, each node has at least one link

The network develops following two steps or factors:

- Growth
  - At each timestep we add a new node with m ($\leq n_0$) links that connect the new node to n nodes already in the network
- Preferential Attachment
  - The probability $\Pi(k)$ that an edge of the new node connects to node i depends on the degree k_i as $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$
  - Preferential attachment is a probabilistic mechanism. A new node can connect to *any* node in the network, whether it is a hub or has a single link
  - Equation above suggests that if a new node has a choice between a degree-two and a degree-four node, it is twice likely to connect with degree-four node
  - Thus, heavily linked nodes tend to quickly accumulate more links
  - New nodes have a preference to attach to heavily linked nodes

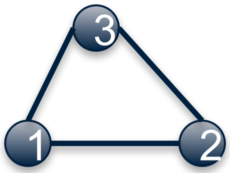
Basic Barabási-Albert model

- Start with an initial set of n_0 fully connected nodes, e.g. $n_0 = 3$
- Add new nodes one by one, each one with exactly m edges
  - At each time step, add a new node with m ($\leq n_0$) edges
- Each new edge connects to an existing node in proportion to the number of edges that node already has
  - Preferential attachment
- After t timesteps the Barabási-Albert model generates a network with $n = t + n_0$ nodes and $n_0 + mt$ edges

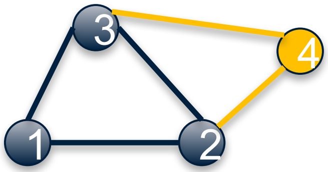


Basic Barabási-Albert model

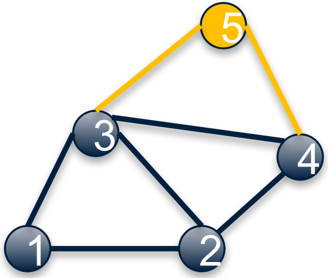
- Start with an initial set of n_0 fully connected nodes, e.g. $n_0 = 3$
- Each node has an **equal number** of edges (2)
    - Probability of choosing any node is 1/3
    - [1,1,2,2,3,3]



- Add a new node, and it will have m edges, lets make $m = 2$
    - m has to be less than or equal to n_0 (in this case 3)
    - Select 2 random elements from the array, say 2 and 3
        - New array = [1,1,2,2,2,3,3,3,4,4]
- Probability of selecting 1,2,3 or 4 are: 1/5, 3/10,3/10 and 1/5

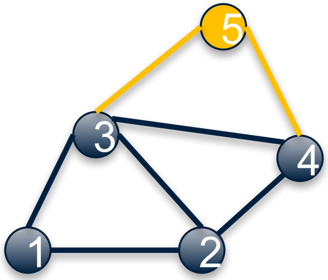


- Add a new node based on preferential attachment
    - New array = [1,1,2,2,2,3,3,3,3,4,4,4,5,5]
    - Etc.



Basic Barabási-Albert model

After t timesteps the Barabási-Albert model generates a network with $n = t + n_0$ nodes and $n_0 + mt$ edges



$t = 2$

$n_0 = 3$

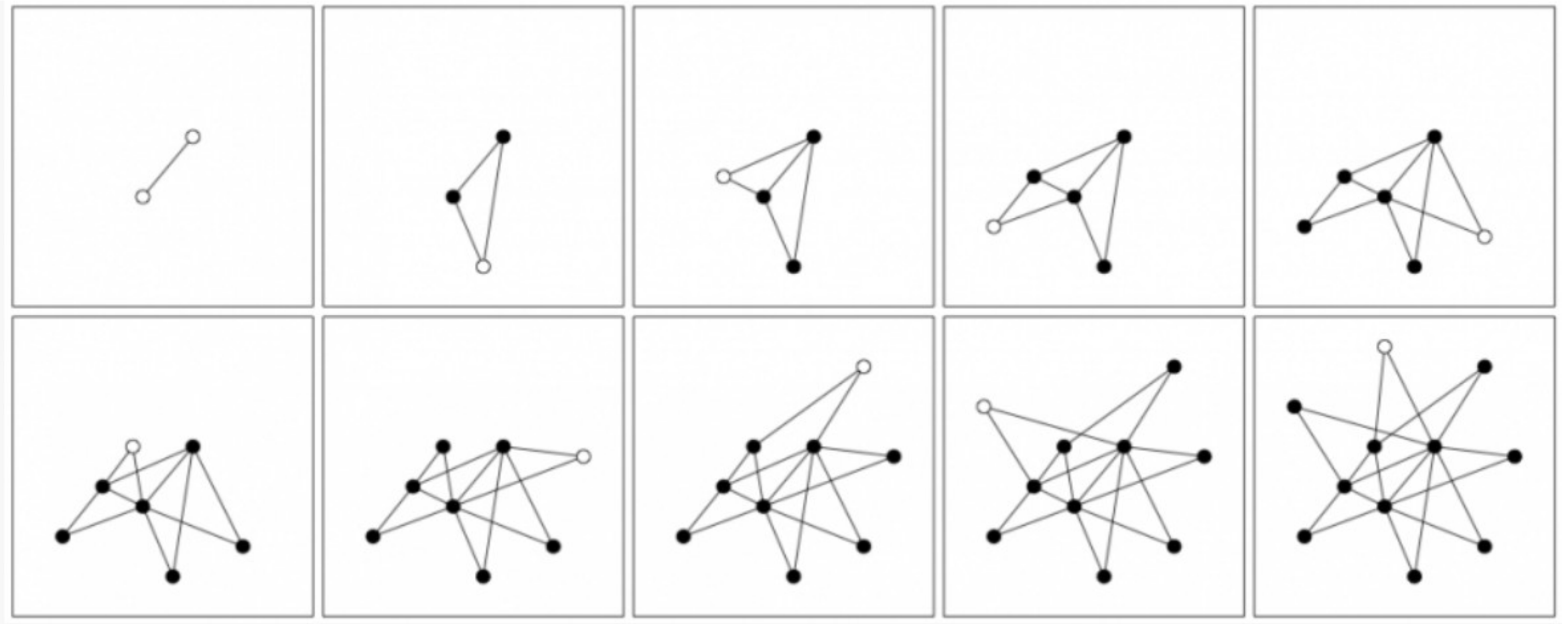
$m = 2$

Nodes $= t + n_0 = 2 + 3 = 5$

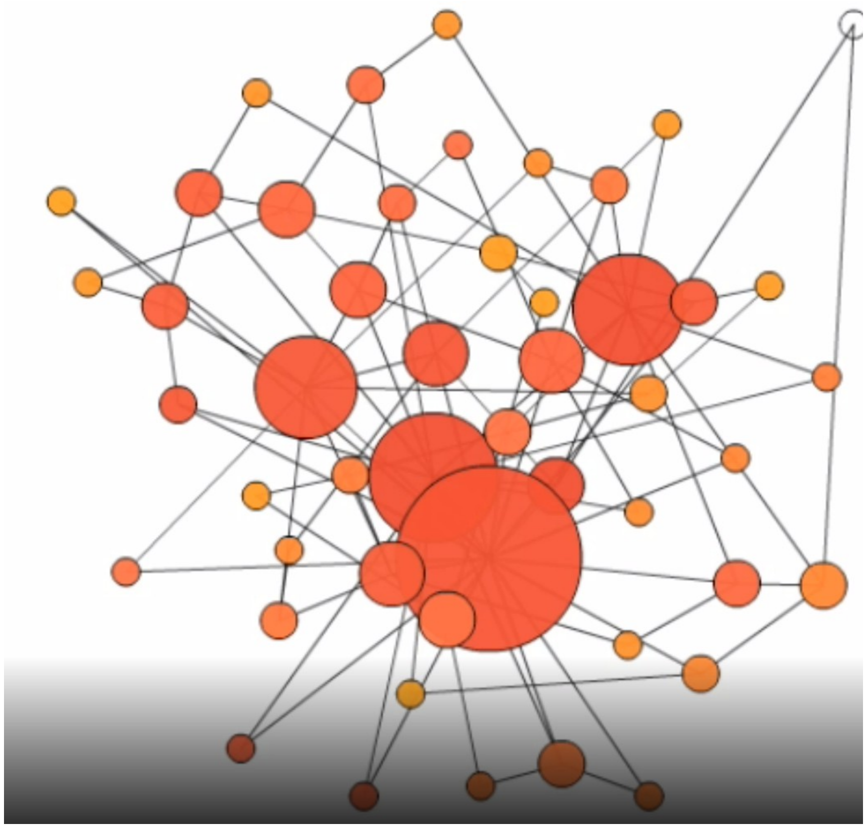
Edges $= n_0 + mt = 3 + 2*2 = 7$

The Barabási-Albert model

- 9 subsequent steps of the Barabási-Albert model
- Newly added node to the network are indicated by the unshaded circles
- Preferential attachment is used to decide where to connect the 2 edges of each new node (*m*=2)



After a while



Degree Dynamics

- The rate at which an existing node *i* acquires links as a result of new nodes connecting to it is: $$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$ $$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$
- Coefficient *m* indicates that each new node arrives with *m* edges
  - Node *i* has *m* chances to be chosen (since every new node has *m* edges)
  - Sum in denominator (above) goes over all nodes except new nodes we are adding, thus: $\sum_{j=1}^{N-1} k_j = 2mt - m$ Sum of degree of all nodes = 2 * # edges
- Therefore
  - $\frac{dk_i}{dt} = \frac{k_i}{2t-1}$ -1 can be neglected for large t, so $\frac{dk_i}{k_i} = \frac{1}{2}\frac{dt}{t}$
  - By integrating above equation and using the fact that $k_i(t_i)=m$, meaning that node *i* joined the network at time t_i with *m* links, we obtain:

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$$ β = dynamical exponent $$\beta = \frac{1}{2}$$

Degree Dynamics

Equation $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$ offers some predictions:

- The degree of each node increases following a power-law with the same dynamical exponent β =1/2
- Each new node has more nodes to link to than the previous node.
  - With time the existing nodes compete for links with an increasing pool of other nodes
- The earlier node *i* was added, the higher is its degree $k_i(t)$. Hence, hubs are large because they arrived earlier, a phenomenon called *first-mover advantage* in marketing and business

The Barabási-Albert model

- While most nodes in the network have only a few links, a few gradually turn into hubs
- 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 node
- Larger nodes will likely acquire links at the expense of the smaller nodes, eventually becoming hubs
- The older are richer
  - Nodes accumulate links as time goes on, which gives older nodes an advantage since newer nodes are going to attach preferentially
- The rate at which the node *i* acquires new links is given by the derivative of $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$
  - which is $\frac{dk_i(t)}{dt} = \frac{m}{2}\frac{1}{\sqrt{t_i t}}$
  - This indicates that in each time step, older nodes acquire more links

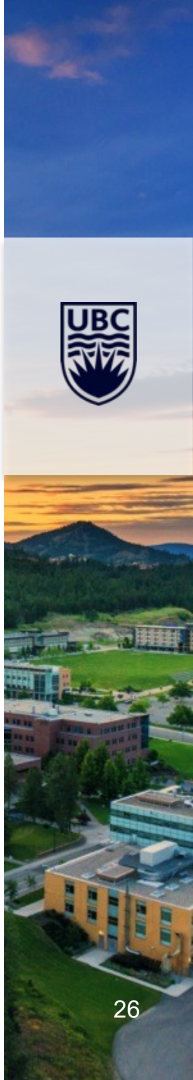
The Barabási-Albert model

- In summary, it shows that in real networks nodes arrive one after the other
- This generates a competition for links during which the older nodes have an advantage over the younger ones, eventually turning into hubs
- Which of the 2 factors (growth or preferential attachment) is more important for a scale free network?
- Which do you think can exist without the other?

Absence of Growth or Preferential Attachment

Absence of Growth or Preferential Attachment

- Are both necessary for the emergence of the scale-free property?
- Can we generate a scale-free network with only growth or with only preferential attachment?



Model A – Role of Preferential Attachment

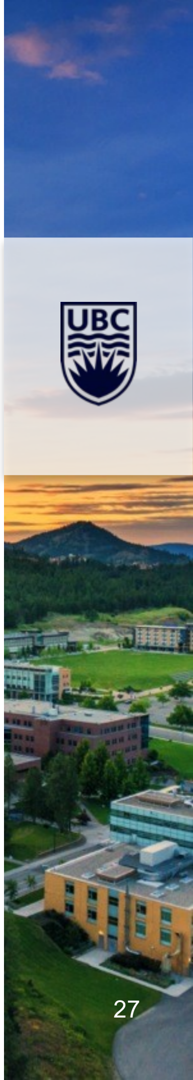
Test the role of preferential attachment; **keep growing the network and remove preferential attachment**

Network starts with n_0 nodes and evolves following these steps:

- **Growth:**
  - At each time step we add a new node with m ($\leq n_0$) edges that connect to n nodes added earlier
- **Preferential Attachment:**
  - The probability that a new node links to a node with degree k_i is

$$\Pi(k_i) = \frac{1}{(n_0+t-1)}$$

    - ***$\Pi(k_i)$* is independent of *k_i***
    - new nodes choose randomly the nodes they link to
- Network grows but degree distribution does not follow the power law distribution



Model B – Role of Growth

Test the role of growth; keep the preferential attachment

Network starts with N nodes and evolves following these steps:

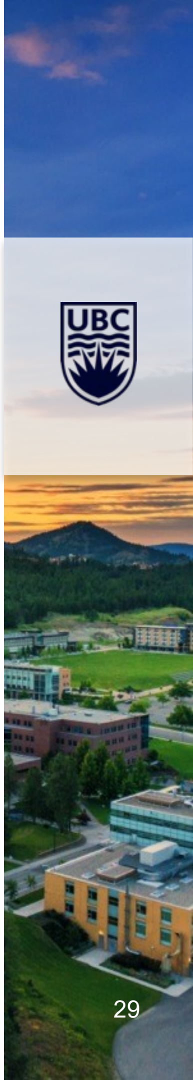
- **Preferential Attachment:**
    - At each time step a node is selected randomly and connected to node i with degree k_i already present in the network
        - i is chosen with the probability $\Pi(k)$
        - nodes with k=0 are assumed to have k=1, otherwise they can not acquire links (since $\Pi(0)$ will be 0)
    - The number of nodes remains constant during the network's evolution, while the number of links increases linearly with time
        - For large t the degree of each node also increases linearly with time

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

- The absence of growth forces the network to converge to a complete graph

Absence of Growth or Preferential Attachment

- Models A and B fail to reproduce the scale-free distribution of scale-free networks
- Indication that both **growth** and **preferential attachment** are **simultaneously** needed for the emergence of the scale-free property



Measuring Preferential Attachment

Measuring Preferential Attachment

- We showed that growth and preferential attachment are jointly responsible for the scale free property
- The presence of growth is obvious as all real world networks grow by adding new nodes
- We need to detect preferential treatment experimentally
- Can do that by measuring *$\Pi(k)$* in real networks

Measuring Preferential Attachment

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*

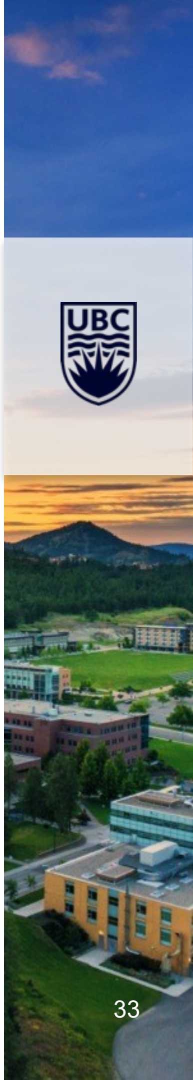
Hypothesis 2

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

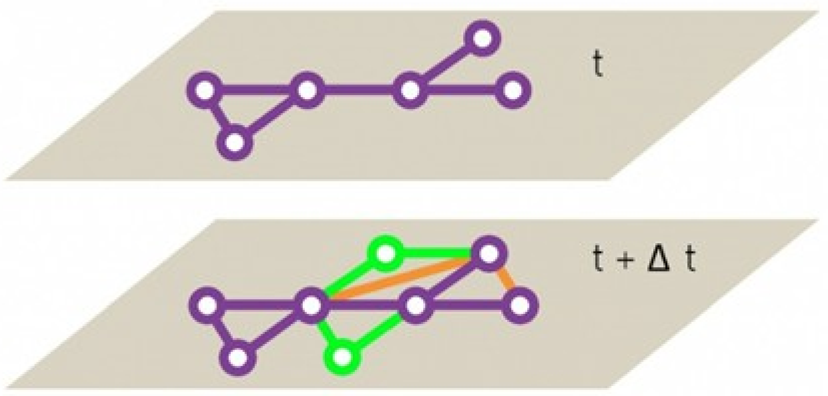
Both hypotheses can be tested by measuring $\Pi(k)$.

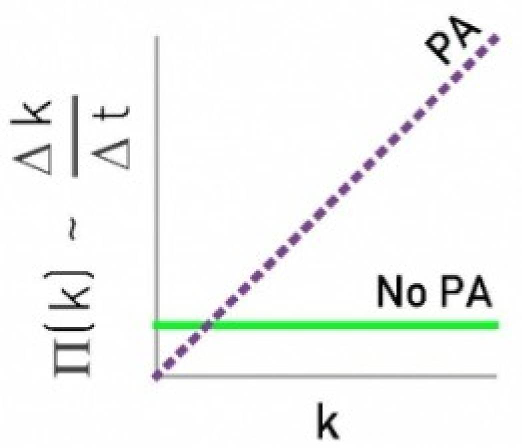
Measuring Preferential Attachment

- We can determine $\Pi(k)$ for systems for which we know the time each node joined the network
- For a network with 2 maps
  - first taken at time t and second at time $t + \Delta t$
- For nodes that changed their degree during the Δt time frame we measure $\Delta k_i = k_i(t+\Delta t) - k_i(t)$
- Relative change $\Delta k_i/\Delta t$ should follow $\frac{\Delta k_i}{\Delta t} \sim \Pi(k_i)$
  - providing the functional form of preferential attachment
- In the presence of preferential attachment $\Delta k/\Delta t$ will depend linearly on a node's degree at time t



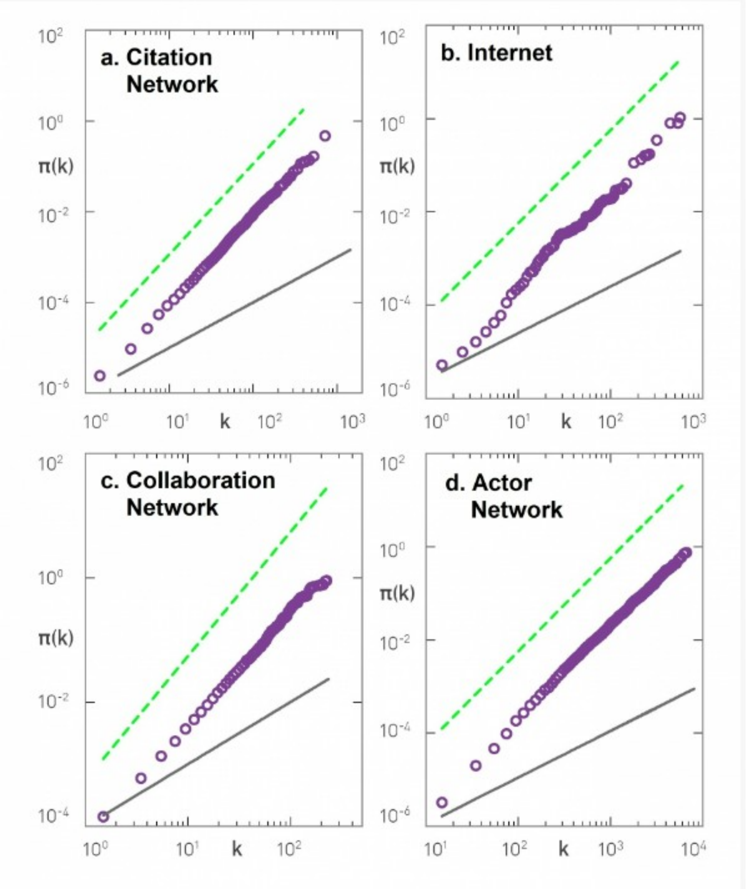
Measuring Preferential Attachment





Albert-Laszio Barabasi, Network Science 2016

Measuring Preferential Attachment

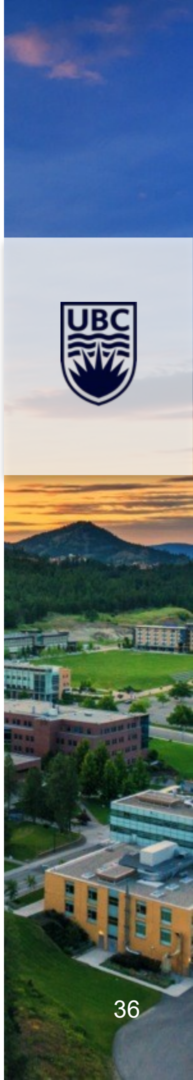


- Real systems
- Dashed line = linear preferential attachment
- Continuous lines = absence of preferential attachment

Albert-Laszio Barabasi, Network Science 2016

Summary

- Described the Barabási-Albert model of network formation
- Identified growth and preferential attachment as the ingredients/factors needed for the Barabási-Albert model
- Described growth and preferential attachment
- Described what happens in the absence of these factors
- Measured preferential attachment



References

- Newman, Mark. Networks, Second Edition, 2018. Oxford University Press
- Albert-Laszio Barabasi, Network Science 2016