SOCIAL NETWORK ANALYTICS

Growing Networks

Growth and Preferential Attachment

Prakash C O

Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Growing Networks

Growth and Preferential Attachment

Prakash C O

Department of Computer Science and Engineering

PES UNIVERSITY ONLINE

Networks Expand Through the Addition of New Nodes

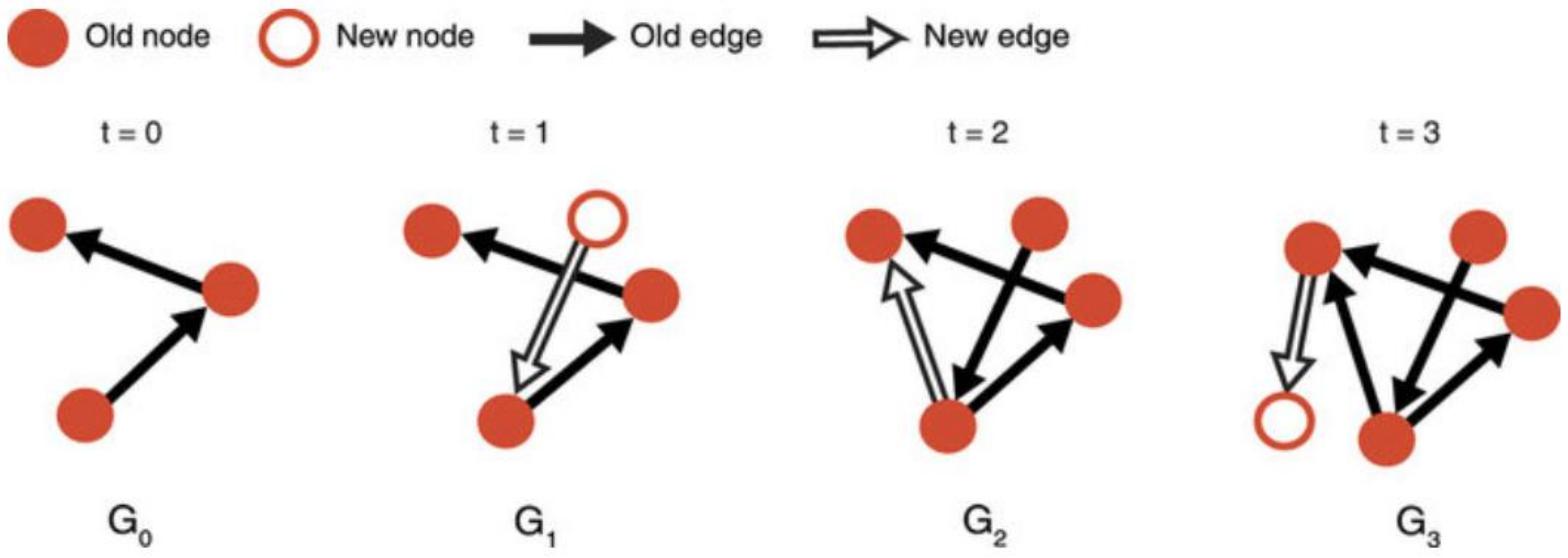
- In general, **networks are not static entities, networks grow/expand, with the continuous addition of new nodes.**

  Examples: Facebook, Twitter, LinkedIn, WWW, Citation Network etc..

- https://www.socialbakers.com/statistics/twitter/profiles/india

- 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.

Networks Expand Through the Addition of New Nodes

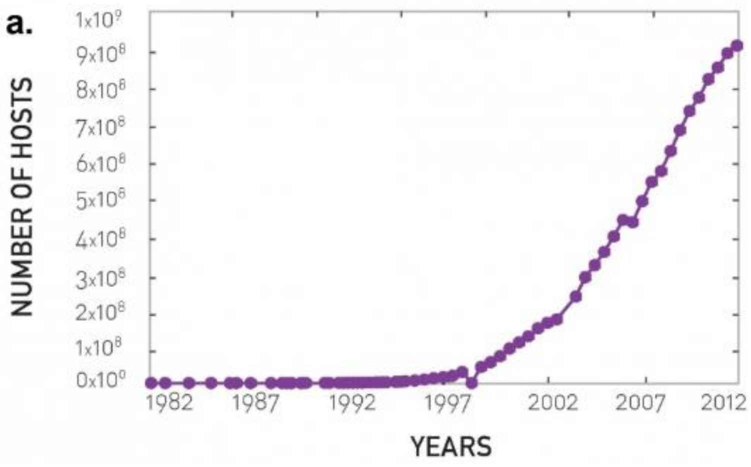


An example of a growing network.: At each time-step, new nodes and edges are added to the network. The number of new nodes and edges at each time-step are free to vary. Note that new edges may emanate from and connect to any old or new nodes. Some examples are: a new edge from a new node to an old node (the network at t = 1), a new edge between existing nodes (the network at t = 2), and a new edge from an existing node to a new node (the network at t = 3).

Networks Expand Through the Addition of New Nodes

Examples:

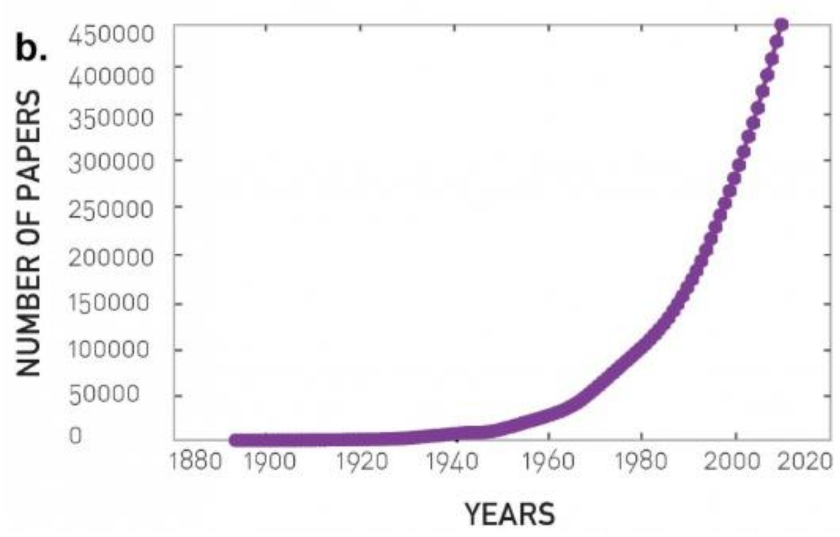
- In 1991 the WWW had a single node, the first webpage built by Tim Berners-Lee, the creator of the Web.
  Today the Web has over a trillion (10^{12}) documents, an extraordinary number that was reached through the continuous addition of new documents by millions of individuals and institutions (Image a).
- http://www.internetworldstats.com/emarketing.htm

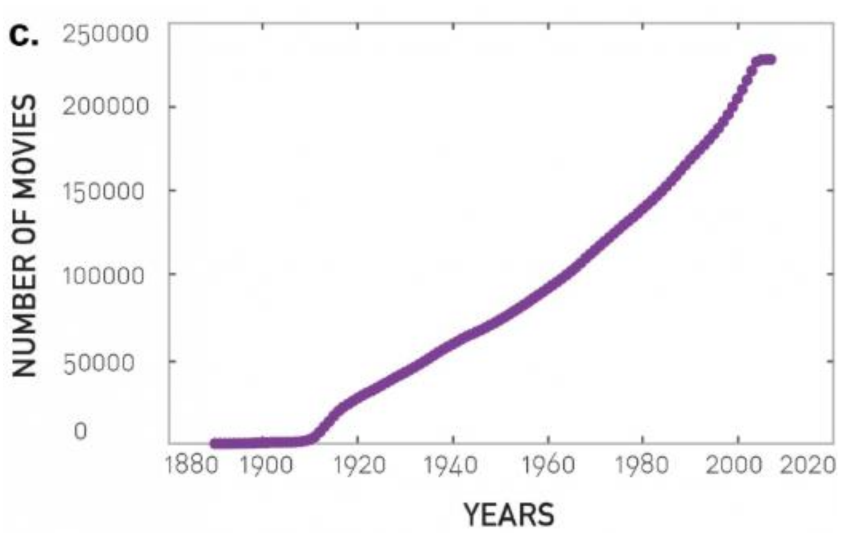


Networks Expand Through the Addition of New Nodes

Examples:

- The collaboration and the citation network continually expands through the publication of new research papers (Image b).
- The actor network continues to expand through the release of new movies (Image c).

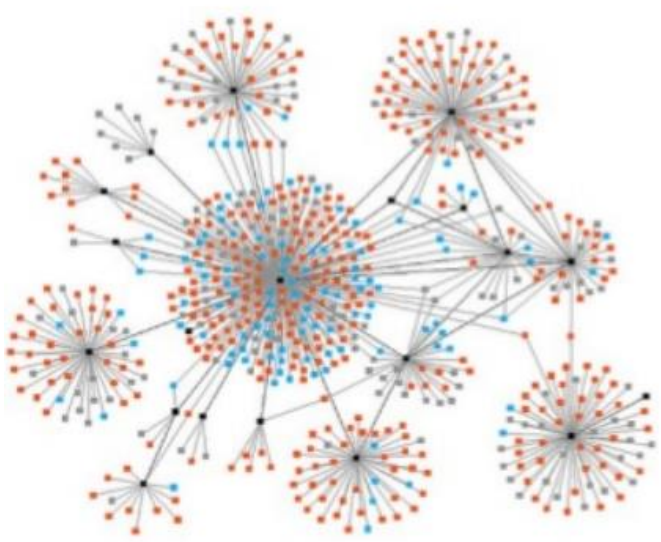




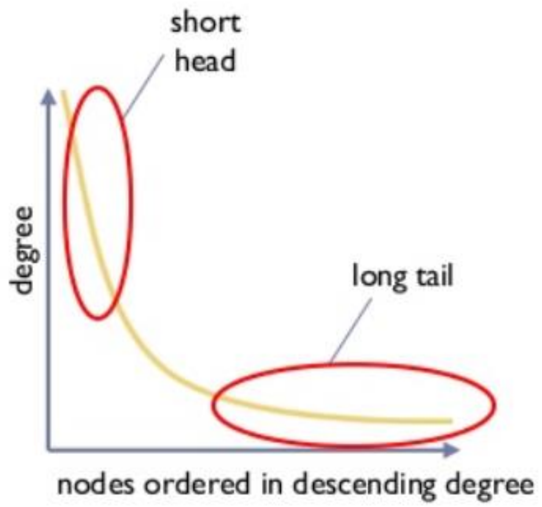
Growth and Preferential Attachment

Nodes Prefer to Link to the More Connected Nodes

- The random network model assumes that we randomly choose the interaction partners of a node.
- **Yet, most real networks new nodes prefer to link to the more connected nodes, a process called preferential attachment.**



Example of network with preferential attachment



Sketch of long-tailed degree distribution

PES UNIVERSITY ONLINE

Nodes Prefer to Link to the More Connected Nodes

Examples 1:

- We are familiar with only a tiny fraction of the trillion or more documents available on the WWW.

- The nodes we know are not entirely random: We all heard about Google and Facebook, but we rarely encounter the billions of less-prominent nodes that populate the Web.

- **As our knowledge is biased towards the more popular Web documents, we are more likely to link to a high-degree node than to a node with only few links.**

- https://en.wikipedia.org/wiki/List_of_most_popular_websites

Nodes Prefer to Link to the More Connected Nodes

Examples 2:

- No scientist can attempt to read the more than a million scientific papers published each year.

- Yet, the more cited is a paper, the more likely that we hear about it and eventually read it.

- **As we cite what we read, our citations are biased towards the more cited publications, representing the high-degree nodes of the citation network.**

Nodes Prefer to Link to the More Connected Nodes

Examples 3:

- The more movies an actor has played in, the more familiar is a casting director with her skills.
- Hence, the higher the degree of an actor in the actor network, the higher are the chances that she will be considered for a new role.

Growth and Preferential Attachment

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

1. **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.

2. **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.

Growth and Preferential Attachment

- **Preferential Attachment:**
  - An extensive body of literature on **the mechanisms of complex network evolution** has been amassed in the time since the subject first began to flourish around the turn of the century.
  - Various mechanisms have been advanced, including preferential attachment, node fitness, node duplication combined with edge duplication and divergence, homophily, topological distance, and node birth/death processes.
  - Among them, preferential attachment and node fitness have garnered special attention.

Preferential Attachment:

- Preferential attachment refers to the fact that new nodes tend to connect to nodes with large degree.
- Preferential attachment means that the more connected a node is, the more likely it is to receive new links.
- Nodes with higher degree have stronger ability to grab links added to the network.

Growth and Preferential Attachment

- **Preferential Attachment: Examples**

1. In social networks, heavily linked nodes represent well-known people with lots of relations.

https://www.socialbakers.com/statistics/twitter/profiles/india

2. When a newcomer enters the community, She/he is more likely to become acquainted with one of those more visible people rather than with a relative unknown.

➢Preferential Attachment: Examples

3. **In World Wide Web,**

   - New pages link preferentially to hubs, i.e. very well known sites such as Google or Wikipedia, rather than to pages that hardly anyone knows.
   - If someone selects a new page to link, the probability of selecting a particular page would be proportional to its degree.
   - This explains the preferential attachment probability rule.

- **Preferential Attachment phenomenon is also known as a "rich-get-richer" rule**, in the sense that links are formed "preferentially" to pages that already have high popularity.

- **Why do we call this a "rich-get-richer" rule?**

  Because the probability that page ℓ experiences an increase in popularity is directly proportional to ℓ's current popularity.

Growing Networks - Summary

- Most networks aren't 'born', they are made.
- Nodes being added over time means that older nodes can have more time to accumulate edges
- Preference for attaching to 'popular' nodes further skews the degree distribution toward a power-law

Assignment: Paper Reading

- **Clustering and preferential attachment in growing networks** - M. E. J. Newman
- **The Structure of Growing Social Networks** - Emily M. Jin, Michelle Girvan, M. E. J. Newman
- **The Link Prediction Problem for Social Networks** - David Liben Nowell, Jon Kleinberg

Assignment: Activity

- https://networkx.github.io/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_prediction.preferential_attachment.html#networkx.algorithms.link_prediction.preferential_attachment

SOCIAL NETWORK ANALYTICS

The Barabási-Albert Model

Prakash C O

Department of Computer Science and Engineering

The Barabási-Albert Model

- In some networks, a few “hubs” have lots of connections, while everybody else only has a few. This model shows one way such networks can arise.

- Such networks can be found in a surprisingly large range of real world situations, ranging from the connections between websites to the collaborations between actors.

- This model generates these networks by a process of “preferential attachment”, in which new network members prefer to make a connection to the more popular existing members.

The Barabási-Albert Model

- The **Barabasi-Albert model** (a.k.a. BA model) introduced in 1998 explains the power-law degree distribution of networks by considering two main ingredients:
    1. **Growth** and
    2. **Preferential Attachment** (Barabasi and Albert 1998).

- The recognition that growth and preferential attachment coexist in real networks has inspired BA model, which can generate scale-free networks.

The Barabási-Albert Model

- The **Barabási–Albert (BA) model** is an algorithm for generating random scale-free networks using a preferential attachment mechanism.

- Many real-world networks such as ***Internet****, the* ***world wide web****,* ***citation networks****, and some* ***social networks*** are thought to be approximately scale-free and certainly contain few nodes (called hubs) with unusually high degree as compared to the other nodes of the network.

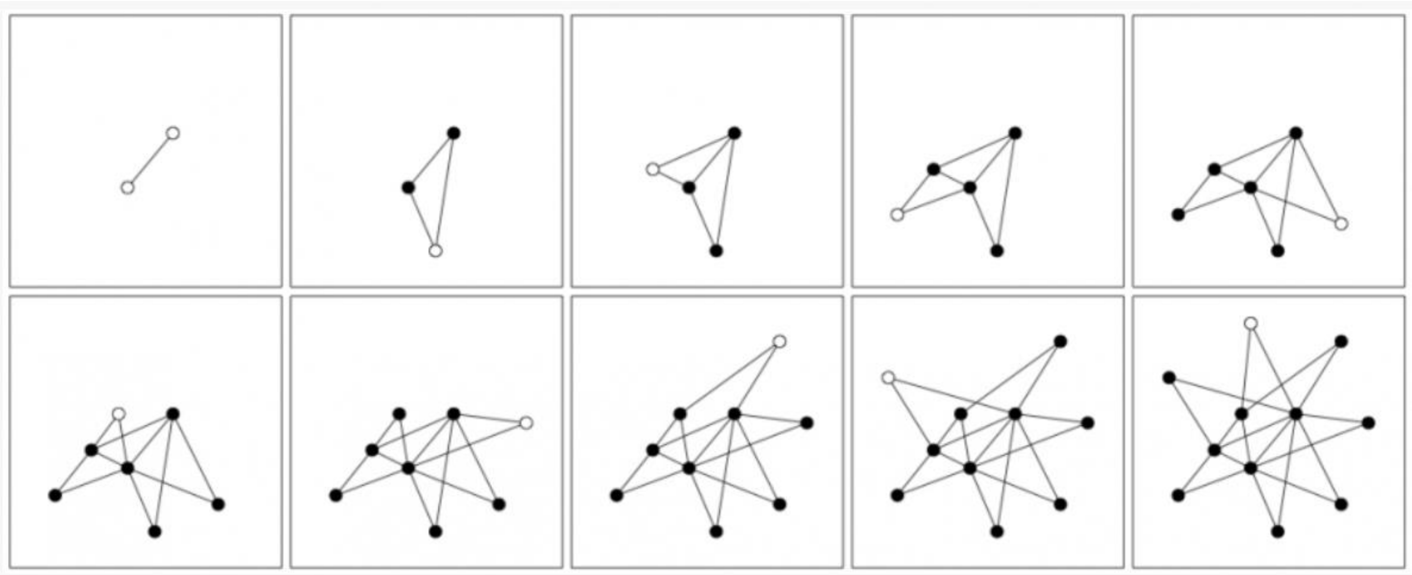
  The BA model tries to explain the existence of such nodes in real networks.

The Barabási-Albert Model: Algorithm

- We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link.

  1. Growth:

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



The Barabási-Albert Model: Algorithm

2. Preferential attachment:

The probability *Π*(k_i) that a link of the new node connects to existing node *i* depends on the degree k_i as

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

Degree of node i

Sum of the degree of all nodes in the network

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.

Equation above 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.

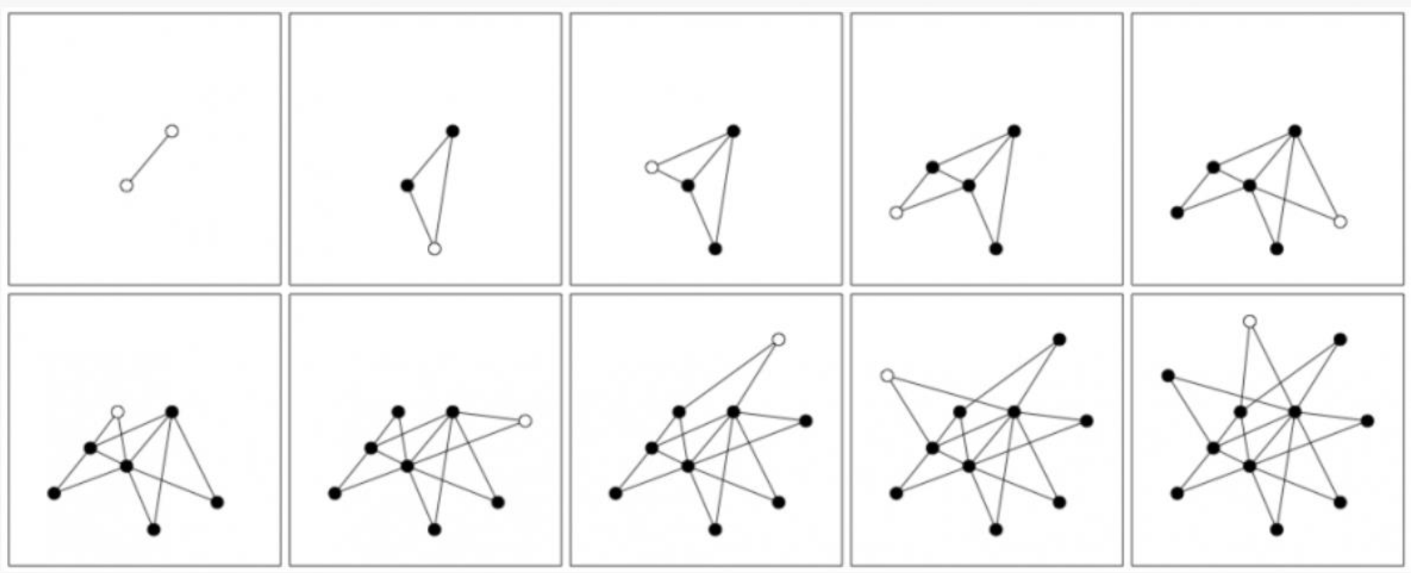
The Barabási-Albert Model

- **Evolution of the Barabási-Albert Model**

  The sequence of images shows nine subsequent steps of the BA model.

  Empty circles mark the newly added node to the network, which decides

  where to connect its two links (*m*=2) using preferential attachment



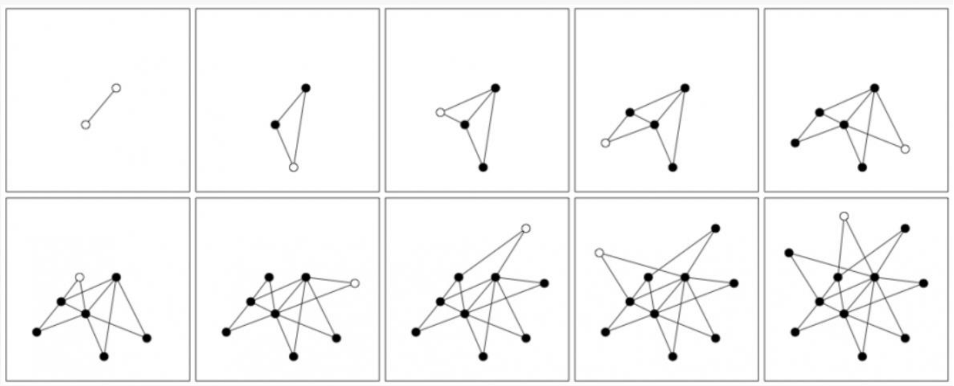
The Barabási-Albert Model

- **Evolution of the Barabási-Albert Model**

  As image below indicate, 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.



➢ The Barabási-Albert model Illustration in NetLogo

1. http://ccl.northwestern.edu/netlogo/

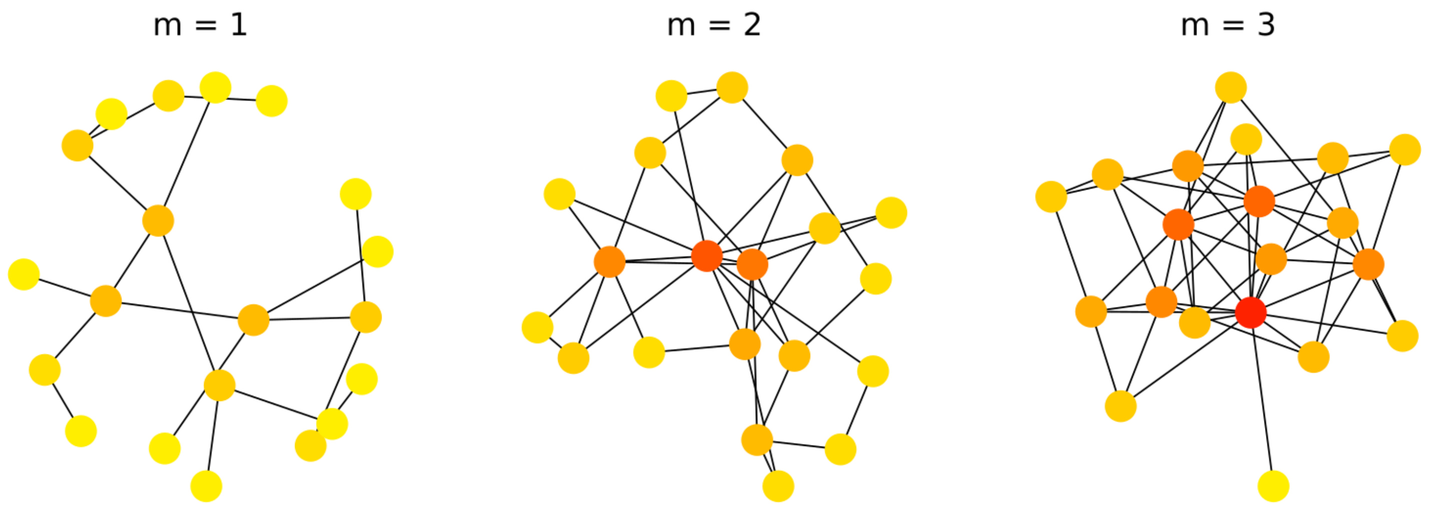
    Go to File/Model Library/Networks/Preferential Attachment

    Note: m=1 is fixed

2. http://ladamic.com/netlearn/NetLogo501/**RAndPrefAttachment.html**

    Try when m=1, m=2 and m=3 (NetLogo 5.3.1)

The Barabási-Albert Model



- Display of three graphs generated with the Barabasi-Albert (BA) model. Each has 20 nodes and a parameter of attachment m as specified. The color of each node is dependent upon its degree (same scale for each graph).

The Barabási-Albert Model

➢ **The Barabási-Albert model Illustration in NetworkX**

1. BarabasiAlbert.py

The Barabási-Albert Model

Properties:

- **Degree distribution**
  - The degree distribution resulting from the BA model is scale free, in particular, it is a power law degree distribution.
- **Average path length**
  - The average path length of the BA model increases approximately logarithmically with the size of the network. The actual form has a double logarithmic correction and goes as

$$\ell \sim \frac{\ln N}{\ln \ln N}.$$

  - The BA model has a systematically shorter average path length than a random graph.

The Barabási-Albert Model

Properties:

- **How robust is a BA Model?**
  - When removing/attacking randomly nodes in the BA network, any fraction of nodes will not break the network.
  - On the other hand removing/attacking only a relatively small fraction of highest degree nodes will cause the network to collapse.

The Barabási-Albert Model

Barabasi and Albert shows that

- Making a network grow with new nodes that
  - Enter the network in successive times
  - Attach preferentially to nodes that already have many links
- Lead to a network structure that is
  - Small world
  - Clustered and
  - Power-law: the distribution of link on the network nodes obeys to the power law distribution!

Quiz:

- Relative to the random growth model, the degree distribution in the preferential attachment model
  1. resembles a power-law distribution less
  2. resembles a power-law distribution more

References

- Social Network Analysis: **Lada Adamic,** University of Michigan.
- Wikipedia – Current Literature
- http://networksciencebook.com/chapter/5#growth

THANK YOU

Prakash C O
Department of Computer Science and Engineering
coprakasha@pes.edu
+91 98 8059 1946

Growing Networks

- **In general, networks are not static entities, networks grow/expand, with the continuous addition of new nodes.**
  - The Web, the Internet, acquaintances, the scientific literature, etc.
  - Thus, edges are added in a network with time.
- **The probability that a new node connect to another existing node may depend on the characteristics of the existing node.**
  - This is not simply a random process of independent node additions.
  - But there could be “preferences” in adding an edge to a node
  - E.g.,. Google, a well known and reliable Internet router, a cool guy who knows many people, a famous scientist, all of these could attract more link…

Quiz:

- How could one make the growth model more realistic for social networks?
  1. old nodes die
  2. some nodes are more sociable
  3. friendships vane over time
  4. all of the above