Preferential Attachment

IV124

Josef Spurný & Eva Výtvarová

Faculty of Informatics, Masaryk University

March 31, 2023

Models and Networks

model *(network)*	clusters	small diameter	hubs
Grid	yes	no	no
Erdös-Rényi *(random)*	no	yes	no
Watts-Strogatz *(small world)*	yes	yes	no
Barabási-Albert *(scale-free)*	no	yes	yes

Many real-world networks contain hubs:

- protein-protein interaction, gene expression, metabolic networks
- human communication (phone calls, emails...)
- human interaction (science / movie cooperation, wealth distribution...)
- www, internet, power grids

Generating Random Networks with Given p_k

General approach:

- Generate a random network based on a given degree distribution p_k.
- Allows for the creation of surrogate data for a real network.
- Does not reveal anything about the origin of the network's structure itself.

Three main variants:

- Degree-preserving randomization
- Generative models
  - Configuration model
  - Hidden variable model

Degree-Preserving Randomization

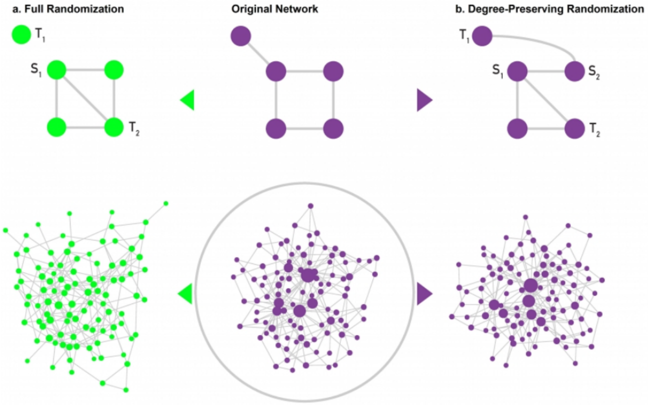
Procedure:

- Input an existing network
- Randomly select a pair of edges and swap them
- Multilinks are forbidden
- Repeat until all links are swapped at least once
- $(i, j), (k, l) \rightarrow (i, l), (k, j)$

Properties:

- Preserves the size, density, and p_k of the network.
- Other parameter-dependent properties are lost.

Degree-Preserving randomization



[1]

[1] Barabási: Network Science Book

Configuration Model

Procedure:

- Input a set of nodes with given degrees (obtained from adjacency matrix)
- Links are cut in a half such that they remain stubs
- Randomly connect pairs of stubs to get links

Properties:

- Probability of an edge between nodes i and j: $\frac{k_i k_j}{2|E|-1} \implies$ prefers connecting high-degree nodes
- Leads to the creation of loops and multiple edges
- Degree of nodes is preserved, network is random

Hidden variable model

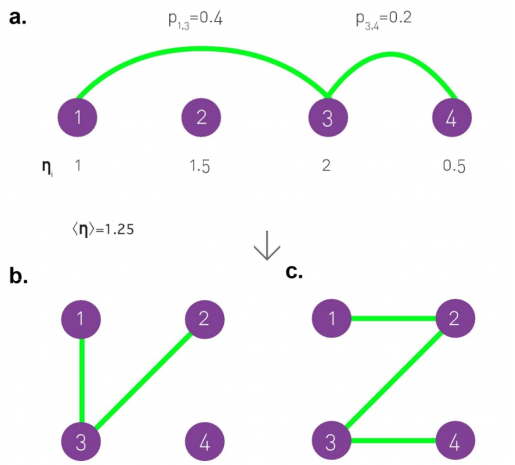
Procedure:

- start with isolated nodes
- assign each node a parameter value η_i from the distribution $\rho(\eta)$
- add edge (i, j) with probability $\frac{\eta_i \eta_j}{\langle\eta\rangle N}$
- e.g., for scale-free networks: $\eta_i = c/i^\alpha, i = 1, \ldots, N$
  - results in a network with $p_k \approx k^{-(1+\frac{1}{\alpha})}$

Properties:

- does not create multi-edges and loops
- flexible regarding the desired p_k

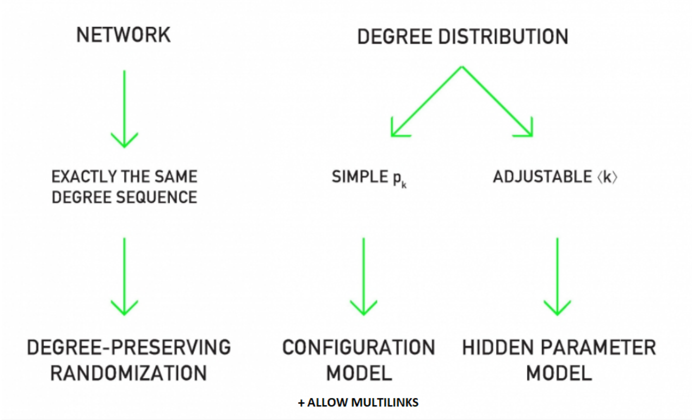
Hidden variable model



2

[2] Barabási: Network Science Book

Which model to choose?



[3]

[3] Barabási: Network Science Book

Growth models

Motivation:

- we are interested in the principles behind the scale-free nature of networks shared across vastly different systems

Observation:

- real networks often form through gradual evolution (adding nodes)
- citation network, WWW, …

Preferential attachment

Intuition:

- newly arriving nodes are more likely to be connected to popular nodes with high k_i
- *rich get richer effect*

General procedure:

- iteratively add a node with a given number of edges
- the probability of connecting to an existing node j depends on k_j

Barabási-Albert Model

Procedure

- each new node arrives with m edges
- the probability of attaching to node i is given by:

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

Resulting degree distribution:

$$p(k) \approx 2m^2 k^{-3}$$

Netlogo demo

...

Nonlinear preferential attachment

In general, $\Pi(k) \sim k^{\alpha}$

Sublinear ($\alpha < 1$)

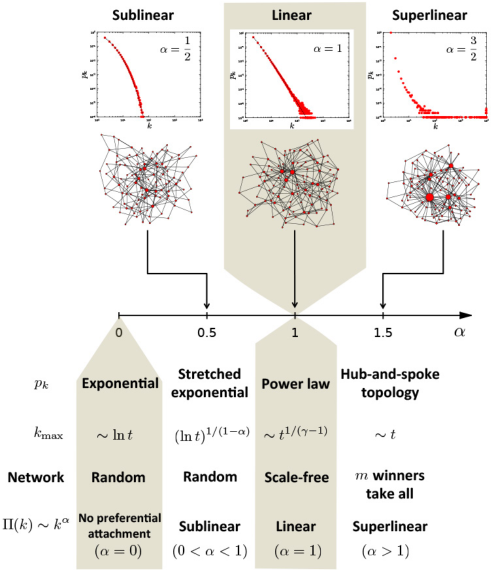
- Not enough to create hubs: random network

Linear ($\alpha \approx 1$)

- Scale-free network

Superlinear ($\alpha > 1$)

- The tendency of rich-get-richer dominates
- Winner-takes-all: star topology



[4]

[4] Barabási: Network Science Book

Bianconi-Barabási Model

Motivation:

- BA model favors older nodes (rich-get-richer)
- However, in real-world, older nodes are not necessary the richest: Myspace vs. Facebook; Yahoo vs. Google...
- capable nodes can surpass existing dominant hubs – add the *fitness* parameter to the model

Bianconi-Barabási Model

Procedure

- in each step, add a node with m edges and fitness η from a given distribution $\rho(\eta)$
- the probability of connecting the new incoming node to node i is given by:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Bianconi-Barabási Model

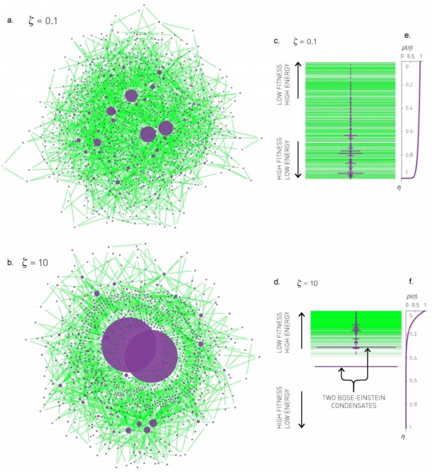
Properties:

- Even a small difference in node fitness leads to large differences in degree
- If node fitness η is identical for all nodes, the model reduces to BA
- For a uniform distribution of $\rho(\eta)$, we get a scale-free network
- Node "age" is not the main determining factor for the resulting degree

Netlogo demo

...

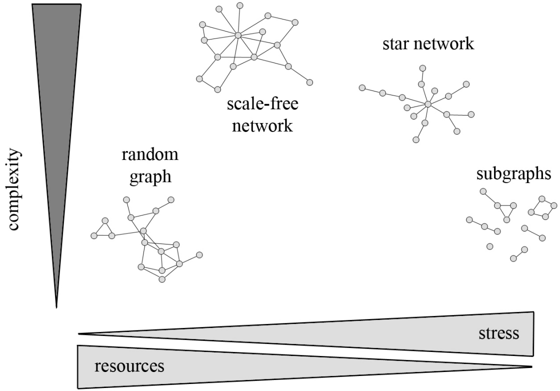
Bose-Einstein Condensation



5

[5] Barabási: Network Science Book

Topological Phase Transition



[6]

[6] Csermely (2006). Weak links, pp 75.

What Do We Know So Far?

Summary + Going Above and Beyond

Node Centralities

Degree

Number of links connected to the node. In directed network, we distinguish in-degree and out-degree.

Eigenvector centrality

Self-referential measure of centrality – node has high eigenvector centrality if it connects to other nodes that have high eigenvector centrality.

Homophily

Assortativity

A positive assortativity coefficient indicates that nodes tend to link to other nodes with the same or similar degree.

Disassortativity

A negative assortativity coefficient indicates that nodes tend to link to other nodes with different degree.

Rich Club Coefficient

Measures the extent to which well-connected nodes also connect to each other.

Communities

Local Clustering

How close are nodes' neighbours to be a complete graph (clique).

Community structure

A subset of network that maximizes within-group links a minimizes between-group links.

Modularity

A statistic which denotes to what extent the network may be divided into groups.

Paths

Path

A sequence of linked nodes that never visits a single node more than once (as opposed to walks which allow this).

Betweenness centrality

Fraction of all shortest paths in the network that contain a given node. High BC = many shortest paths through the node. Similarly, edge betweenness centrality may be calculated.

Closeness centrality

Measures how short the shortest paths are from selected node to all other nodes.

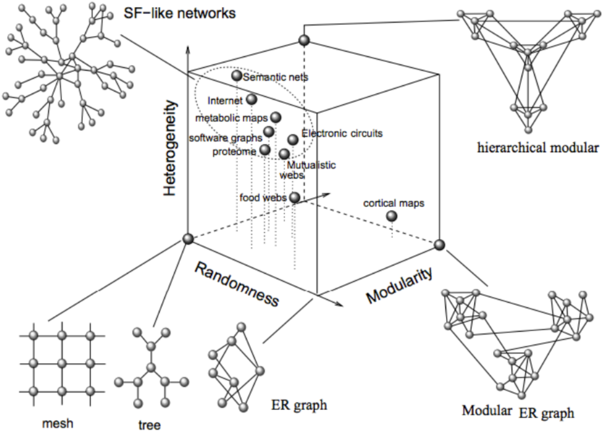
Basic network characteristics

- Avg./min./max. degree & degree distribution
- Avg. path length & BC or Closeness centrality distribution
- Connectedness – number of components & size of giant component
- Modularity & modularity classes (communities)
- Comparison to known network models

Models and Networks

model class	model	observation
static	Erdös-Rényi	giant component
	Watts-Strogatz	small worlds
generative	configuration model	loops and multilinks
	hidden parameter model	adjustable p_k
growing	Barabási-Albert	rich-get-richer; scale-free
	Bianconi-Barabási	winner-takes-all

A ZOO of Complex Networks



[7]

---

[7] Types of Networks

MUNI
FACULTY
OF INFORMATICS