Network Science

Class 4: Scale-free property

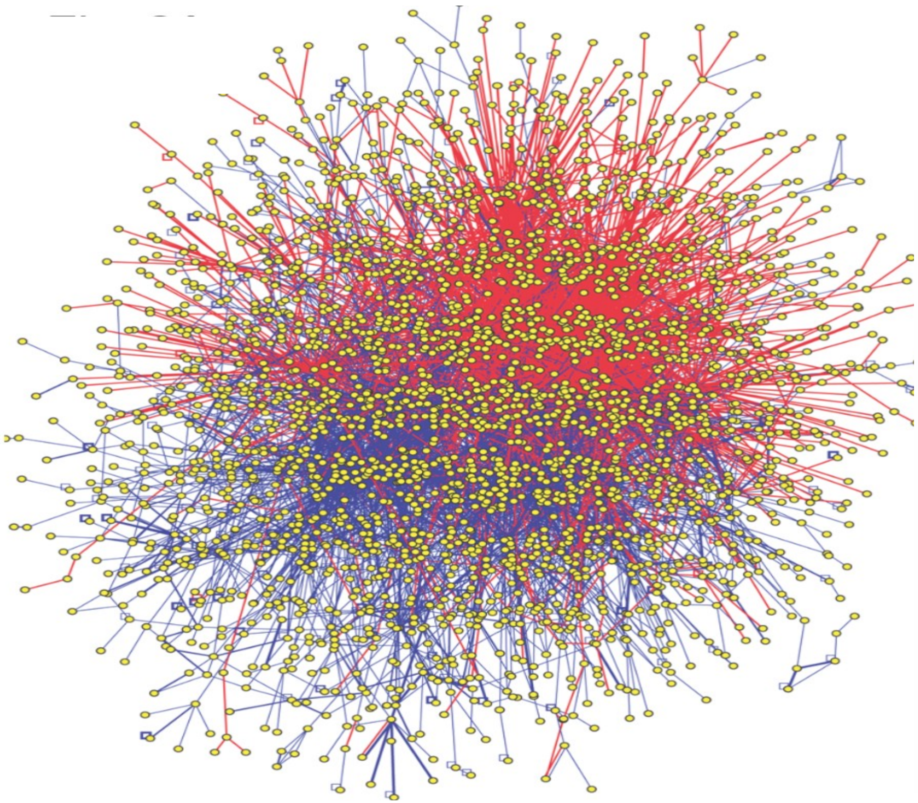
Albert-László Barabási

with Emma K. Towlson, Michael M. Danziger,
Sebastian Ruf and Louis Shekhtman

www.BarabasiLab.com

PLOTTING POWER LAWS

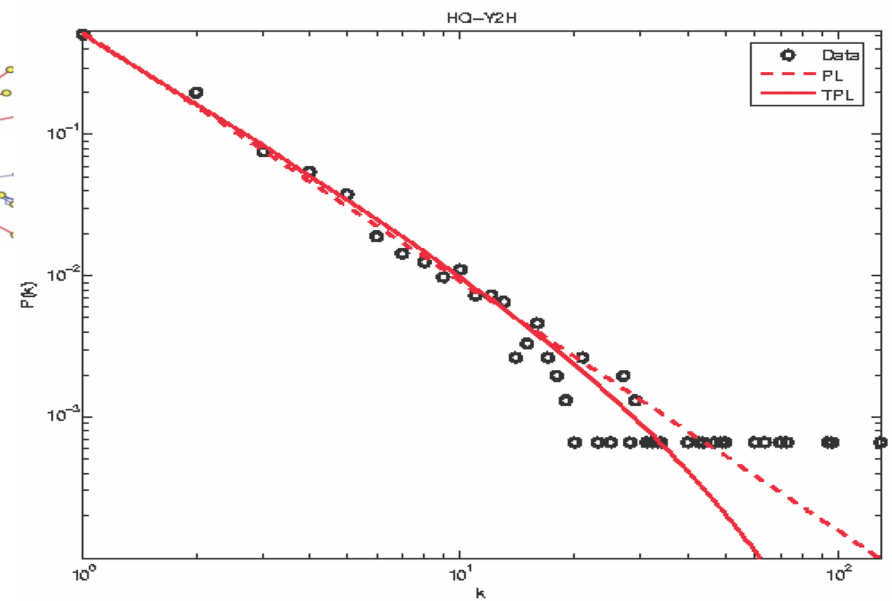
HUMAN INTERACTION NETWORK



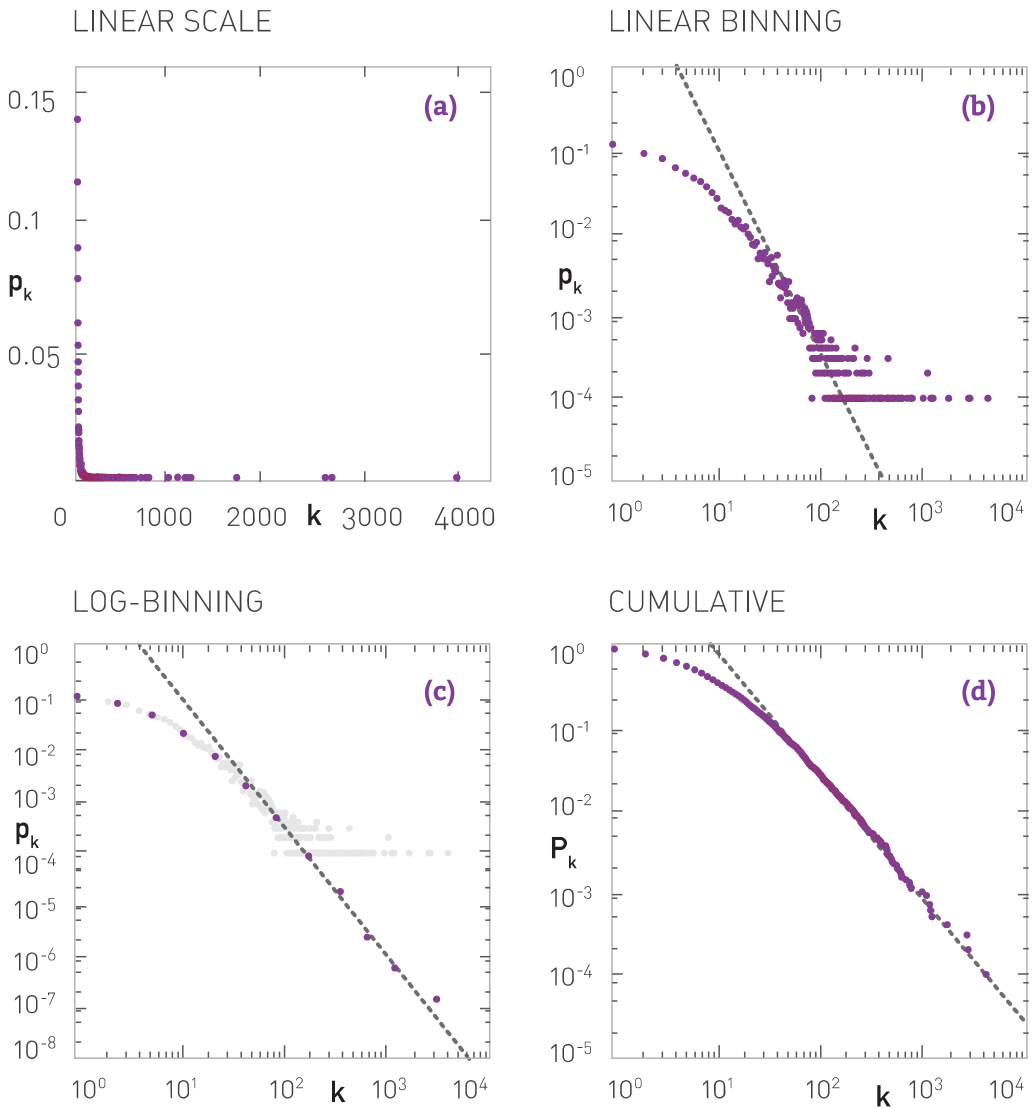
2,800 Y2H interactions

4,100 binary LC interactions

(HPRD, MINT, BIND, DIP, MIPS)



Rual et al. Nature 2005; Stelzl et al. Cell 2005



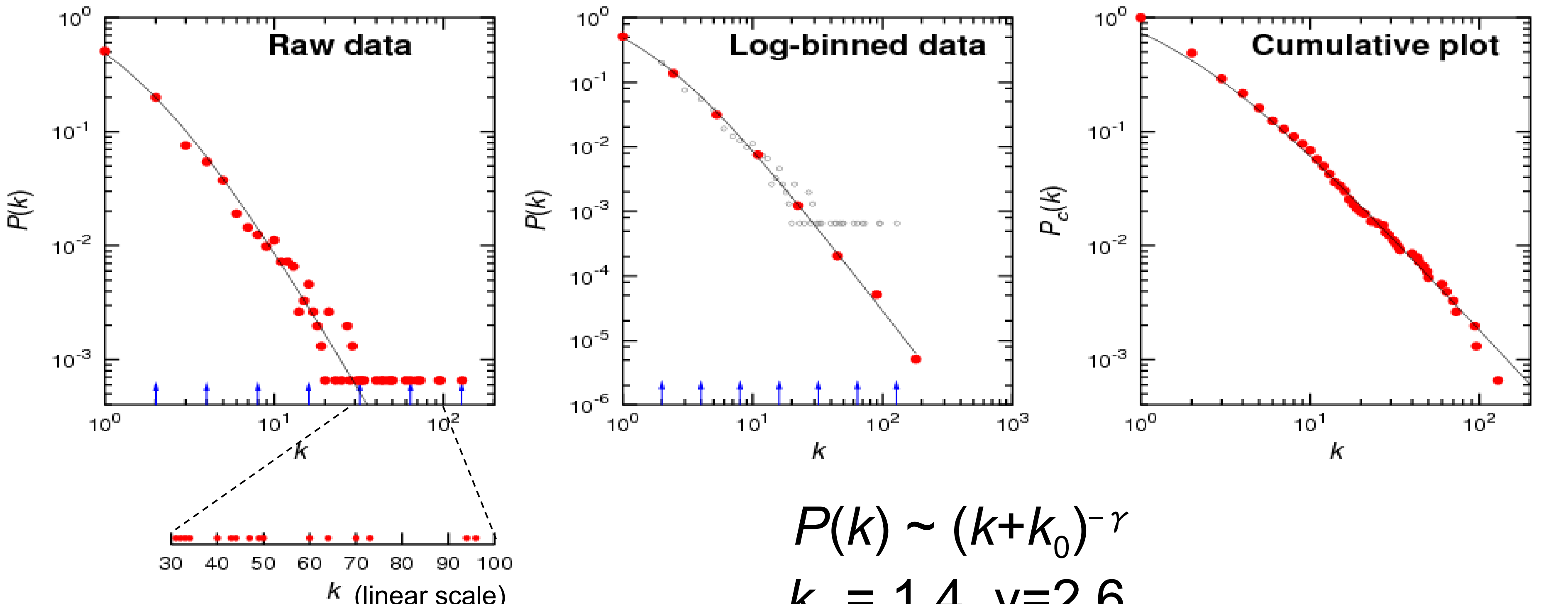
Use a Log-Log Plot

Avoid Linear Binning

Use Logarithmic Binning

Use Cumulative Distribution

HUMAN INTERACTION DATA BY RUAL ET AL.

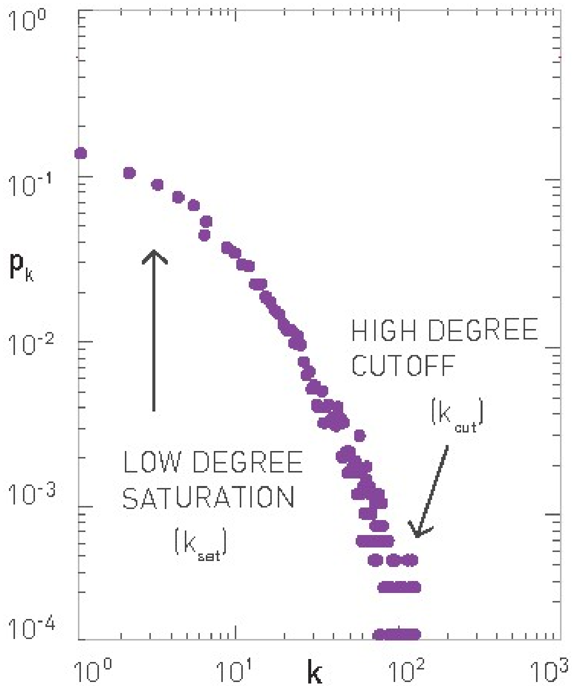


$$P(k) \sim (k+k_0)^{-\gamma}$$

$$k_0 = 1.4,\ \gamma=2.6.$$

COMMON MISCONCEPTIONS

Http://www.nd.edu/~networks

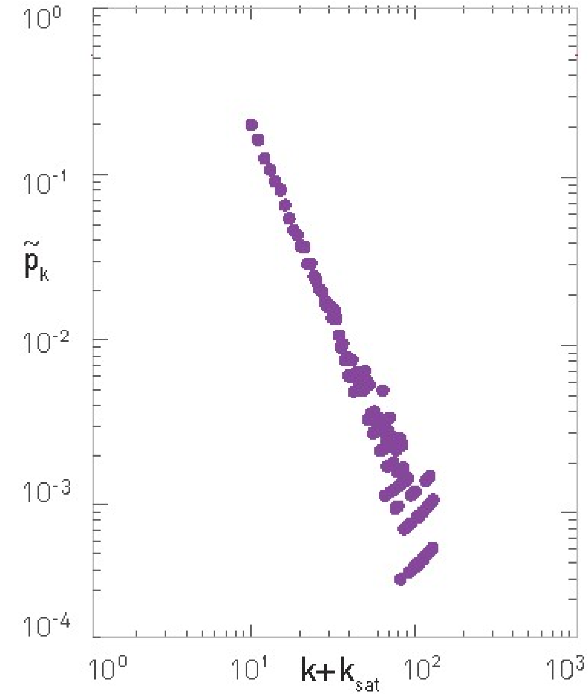


$$p_k = a(k + k_{sat})^{-\gamma} \exp\left(-\frac{k}{k_{cut}}\right)$$

$$\widetilde{p}_k = p_k \exp\left(\frac{k}{k_{cut}}\right)$$

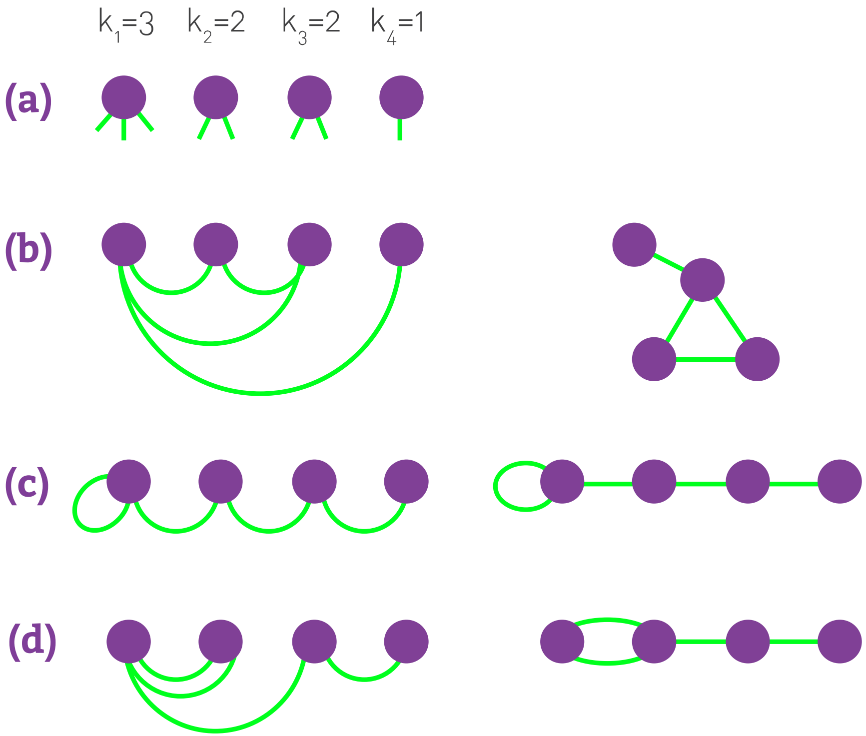
$$\widetilde{k} = k + k_{sat}$$

$$\widetilde{p}_k \sim \widetilde{k}^{-\gamma}$$



Generating networks with a pre-defined p_k

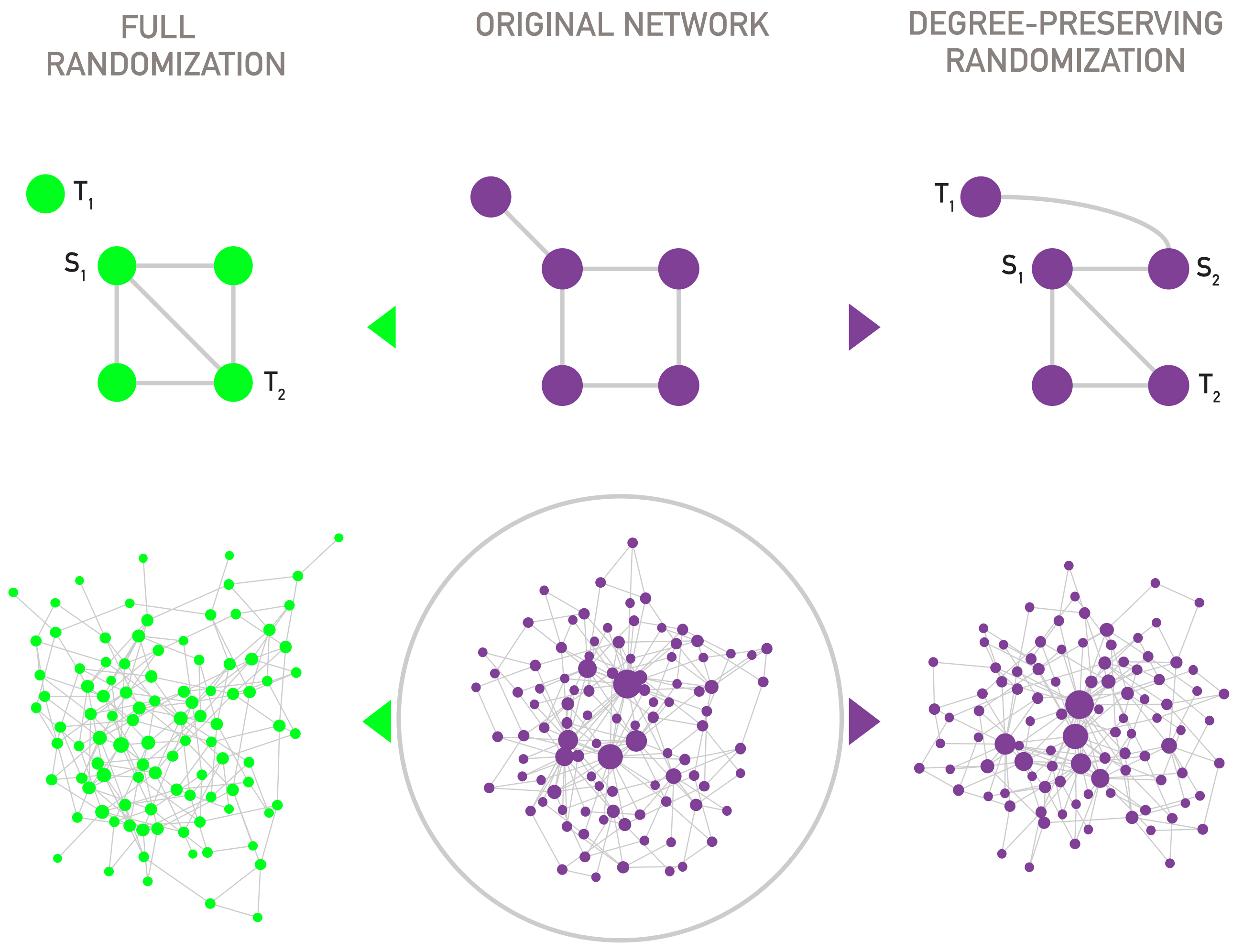
Configuration model



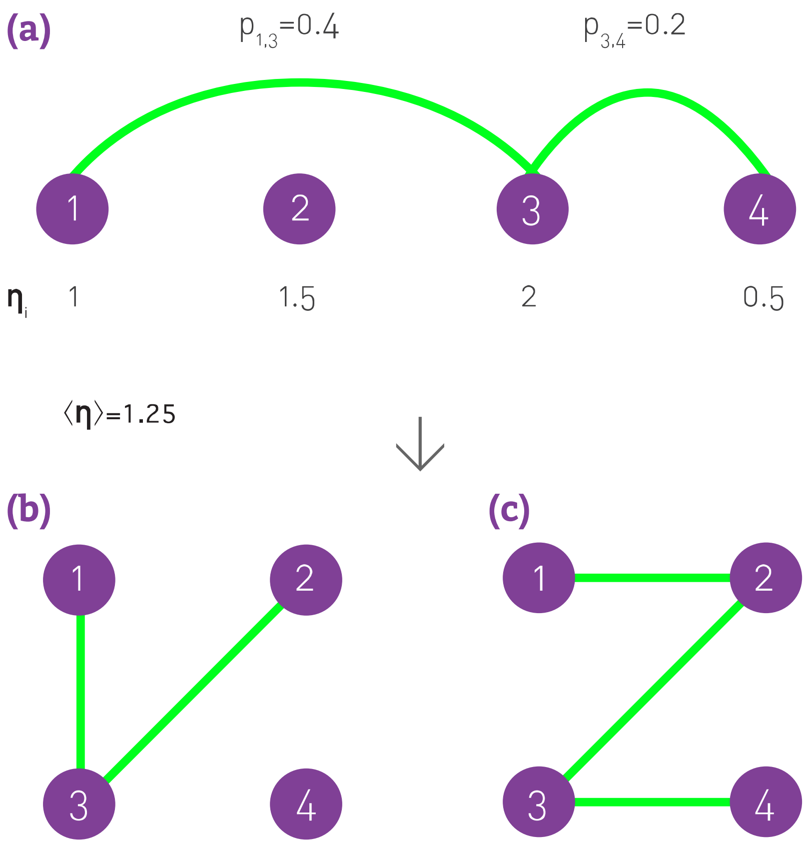
(1) ***Degree sequence****: Assign a degree to each node, represented as stubs or half-links. The degree sequence is either generated analytically from a pre-selected distribution (Box 4.5), or it is extracted from the adjacency matrix of a real network. We must start from an even number of stubs, otherwise we will be left with unpaired stubs.(2)* ***Network assembly****: Randomly select a stub pair and connect them. Then randomly choose another pair from the remaining stubs and connect them. This procedure is repeated until all stubs are paired up. Depending on the order in which the stubs were chosen, we obtain different networks. Some networks include cycles (2a), others self-edges (2b) or multi-edges (2c). Yet, the expected number of self- and multi-edges goes to zero in the limit.*

$$p_{ij} = \frac{k_i k_j}{2L - 1}$$

Degree Preserving randomization



Hidden parameter model



$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}$$

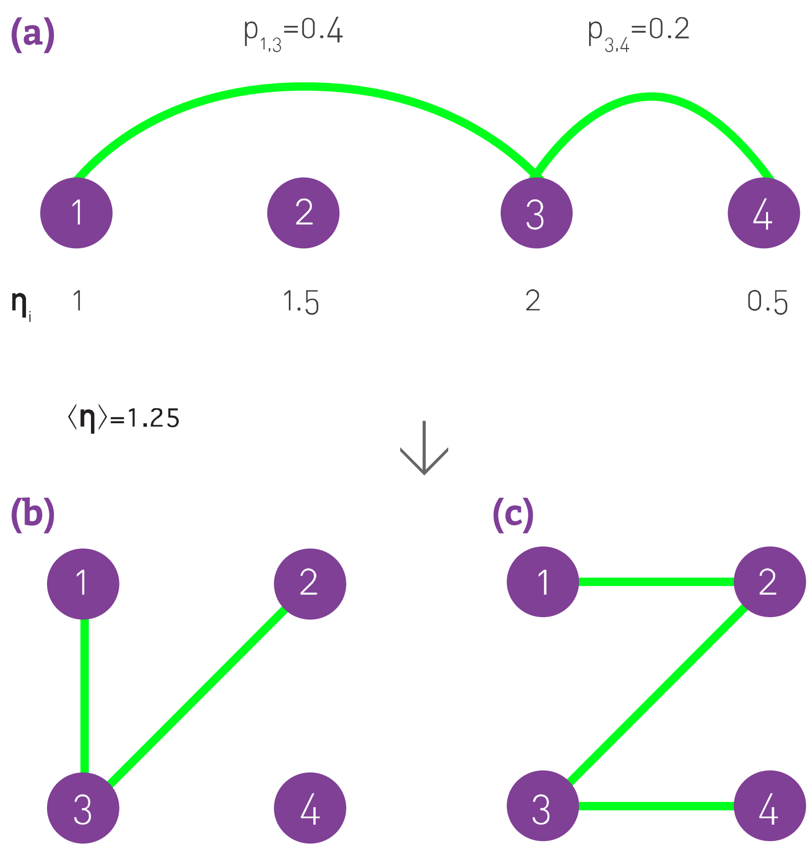
$$p_k = \int \frac{e^{-\eta}\eta^k}{k!} p(\eta) d\eta.$$

$$\{\eta_1, \eta_2, \ldots, \eta_N\}$$

$$p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^k}{k!}.$$

$$\eta_j = \frac{c}{i^\alpha}, i = 1, \ldots, N \qquad p_k \sim k^{-(1+\frac{1}{\alpha})}$$

Hidden parameter model



Start with N isolated nodes and assign to each node a "hidden parameter" η , which can be randomly selected from a ρ(η) distribution. We next connect each node pair with probability

$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle\eta\rangle N}$$

For example, the figure shows the probability to connect nodes (1,3) and (3,4). After connecting the nodes, we end up with

the networks shown in (b) or (c), representing two independent realizations generated by the same hidden parameter sequence (a). The expected number of links in the obtained network is

$$L = \frac{1}{2}\sum_{N}^{i,j} \frac{\eta_i \eta_j}{\langle\eta\rangle N} = \frac{1}{2}\langle\eta\rangle N$$

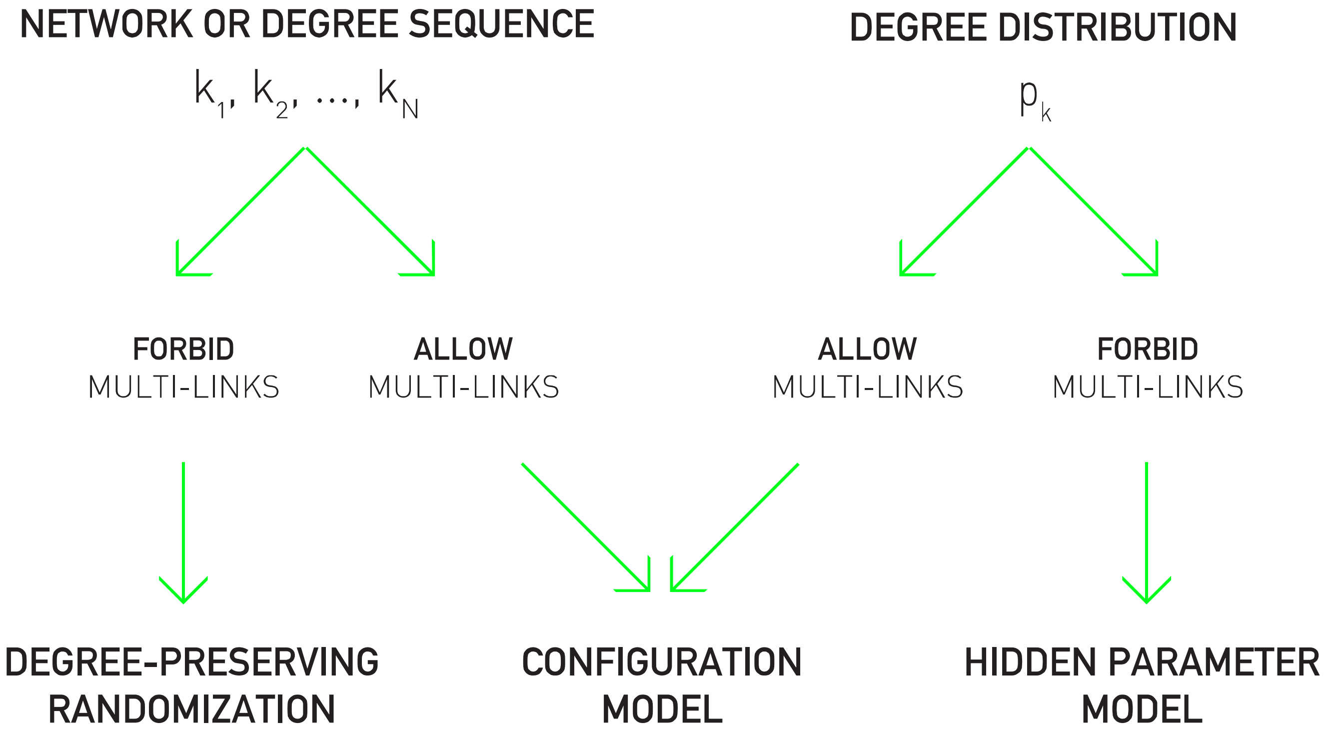
$$\eta_j = \frac{c}{i^{\alpha}}, i = 1, \ldots, N$$

$$p_k = \int \frac{e^{-\eta}\eta^k}{k!} \rho(\eta) d\eta.$$

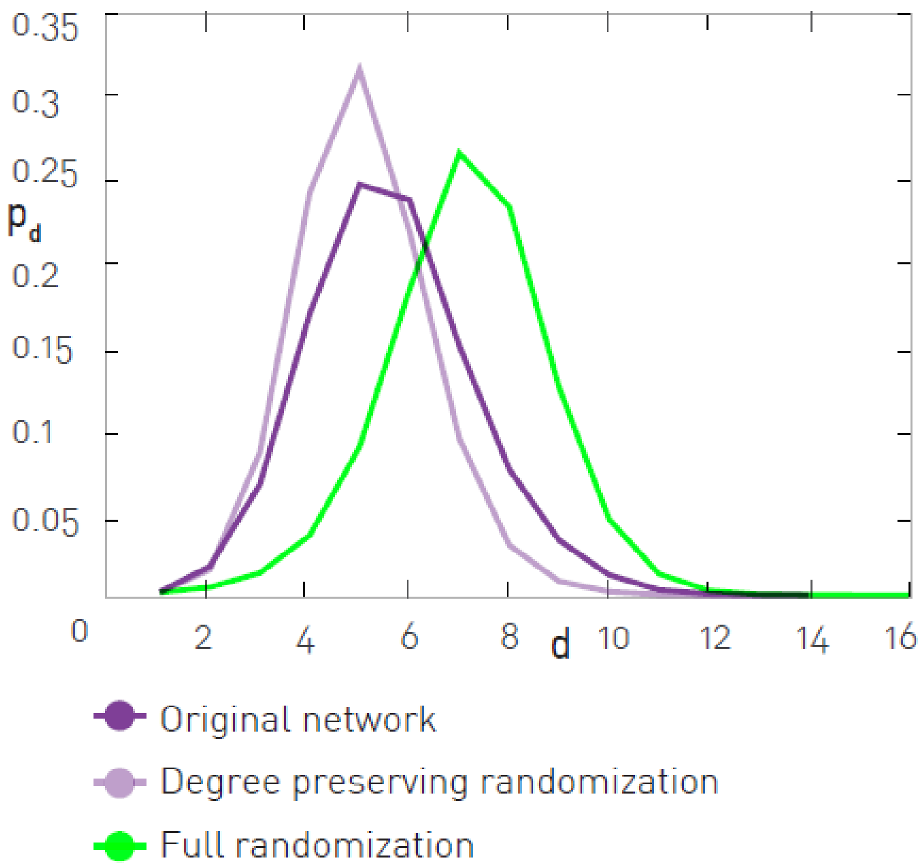
$\{\eta_1, \eta_2, \ldots, \eta_N\}$ $$p_k = \frac{1}{N}\sum_j \frac{e^{-\eta_j}\eta_j^{k}}{k!}.$$

$$p_k \sim k^{-(1+\frac{1}{\alpha})}$$

Decision tree



Case Study: PPI Network Distance Distribution



We have: $\langle d \rangle$=5.61±1.64 (original), $\langle d \rangle$=7.13 ± 1.62 (full randomization), $\langle d \rangle$=5.08 ± 1.34 (degree-preserving randomization).

Something to keep in mind

summary

Section 9

DEGREE DISTRIBUTION

Discrete form:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}.$$

Continuous form:

$$p(k) = (\gamma - 1)k_{min}^{\gamma-1}\, k^{-\gamma}$$

SIZE OF THE LARGEST HUB

$$k_{max} \sim k_{min} N^{\frac{1}{\gamma-1}}.$$

MOMENTS OF p_k for $N \to \infty$

$2 < \gamma < 3$: $\langle k \rangle$ finite, $\langle k^2 \rangle$ diverges.

$\gamma > 3$: $\langle k \rangle$ and $\langle k^2 \rangle$ finite.

DISTANCES

$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2, \\ \dfrac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3, \\ \dfrac{\ln N}{\ln \ln N} & \gamma = 3, \\ \ln N & \gamma > 3. \end{cases}$$

Bounded Networks

We call a network *bounded* if its degree distribution decrease exponentially or faster for high k. As a consequence $<k^2>$ is smaller than $<k>$, implying that we lack significant degree variations. Examples of p_k in this class include the Poisson, Gaussian, or the simple exponential distribution (Table 4.2). The Erdős-Rényi and the Watts-Strogatz networks are the best known network models belonging to this class. Bounded networks lack outliers, consequently most nodes have comparable degrees. Real networks in this class include highway networks and the power grid.

Unbounded Networks

We call a network *unbounded* if its degree distribution has a fat tail in the high-k region. As a consequence $<k^2>$ is much larger than $<k>$, resulting in considerable degree variations. Scale-free networks with a power-law degree distribution (4.1) offer the best known example of networks belonging to this class. Outliers, or exceptionally high-degree nodes, are not only allowed but are expected in these networks. Networks in this class include the WWW, the Internet, the protein interaction networks, and most social and online networks.