



# How Scale-Free are Biological Networks and Whether it Matters for the Evolutionary Models

---

Raya Khanin

University of Glasgow, Scotland

with Ernst Wit (Lancaster University)



# Is there a Network Biology?

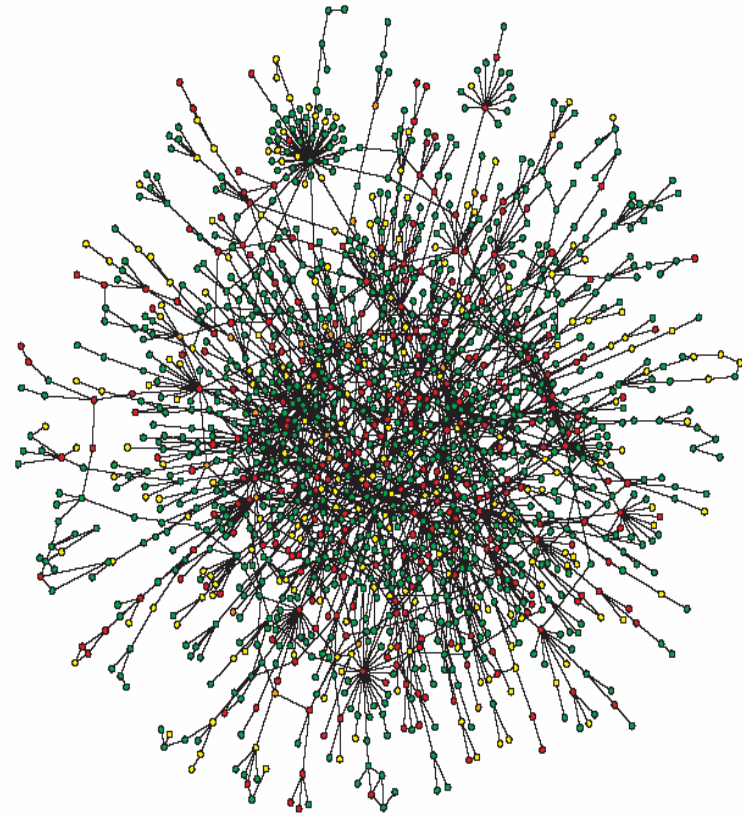
---

## Protein interaction network:

proteins that are connected by physical interactions

**Metabolic network:** metabolic products and substrates that participate in one reaction

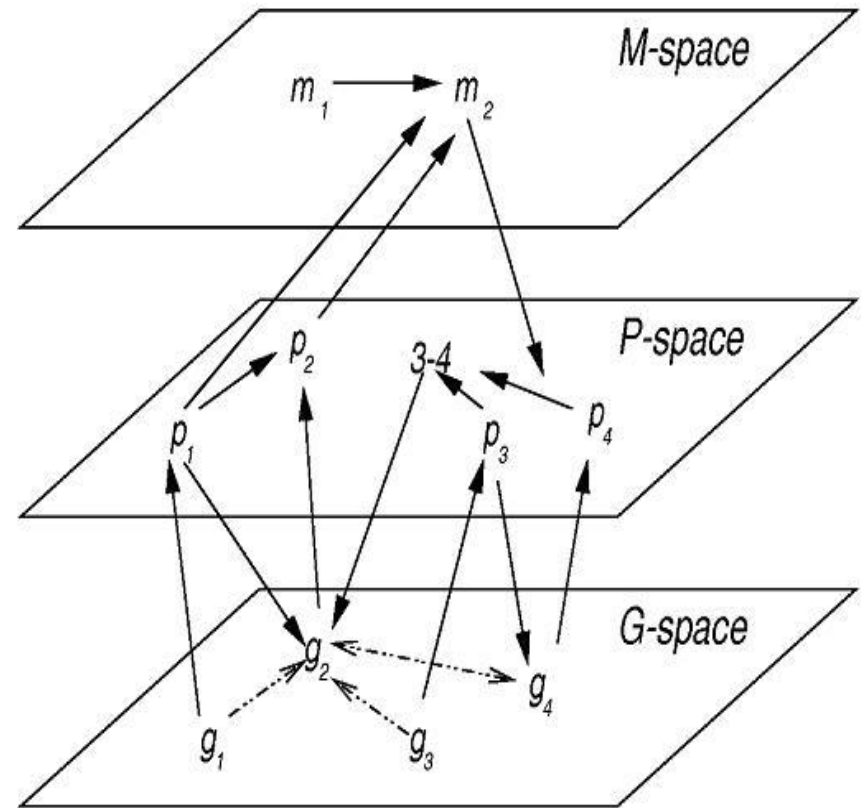
**Gene regulatory network:** two genes are connected if the level of one gene modulates expression of another one by either activation or repression



# Biological networks

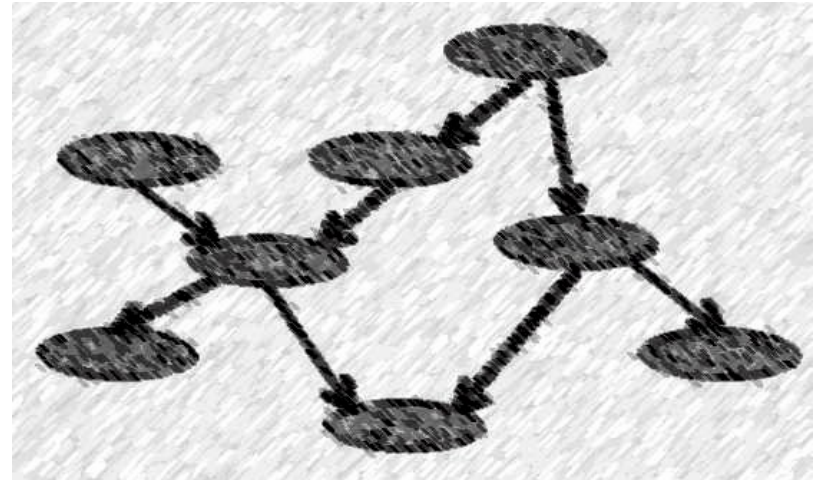
Biological interactions are studied on different levels independently (genes, proteins and metabolites), while they are interconnected and interrelated.

Solid arrows – direct interactions  
Dotted arrows – indirect interactions that occur via intermediates.



# Biological networks as graphs

- **Node** represents a gene, protein or metabolite
- **Edge** represents an association, interaction, co-expression
- **Directed edge** stands for the modulation (regulation) of one node by another: e.g. arrow from gene X to gene Y means gene X affects expression of gene Y





# Common properties of biological networks

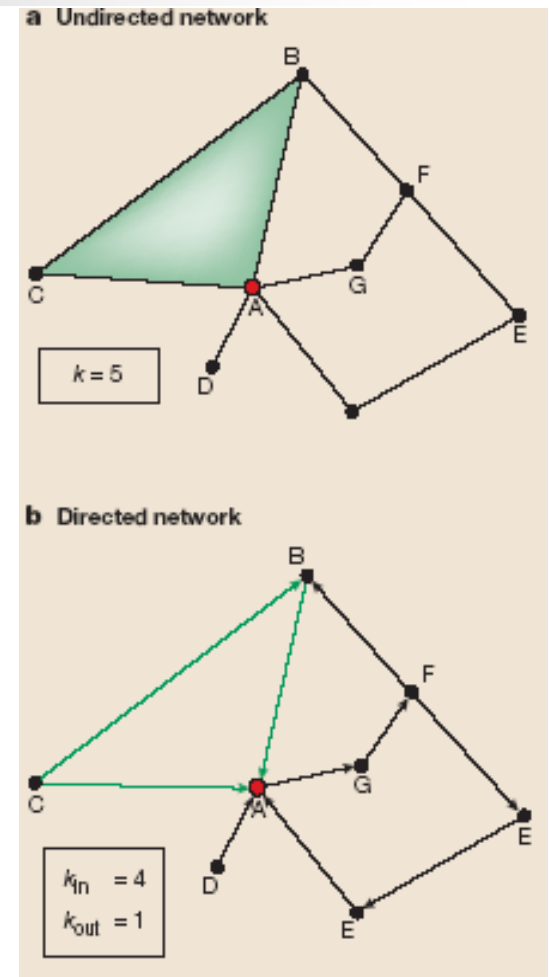
---

- **Small-world property:** short mean path
- There are many nodes with few connections and a few nodes with many connections (**hubs**)
- Hubs are enriched with essential/lethal genes (**lethality and centrality property**; Jeong et al, 2001). This has recently been questioned by A.Wagner, S.Teichmann and others.

# Network measures

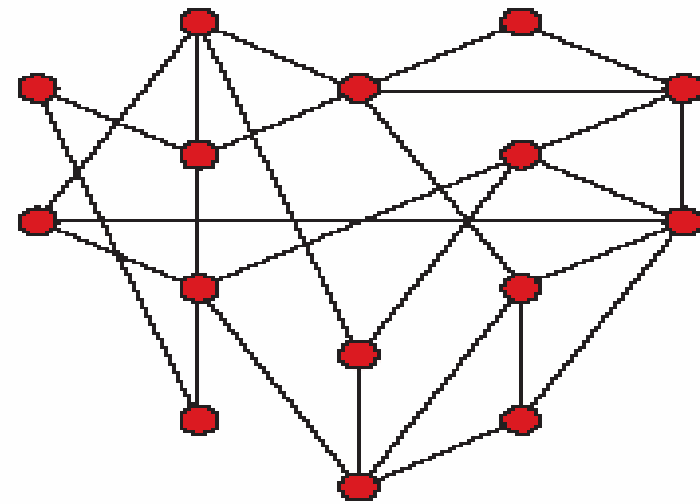
**Degree** (or connectivity) of a node,  $k$ , is the number of links (edges) this node has.

The **degree distribution**,  $P(k)$ , is the probability that a selected node has exactly  $k$  links. Networks are classified by their degree distributions.

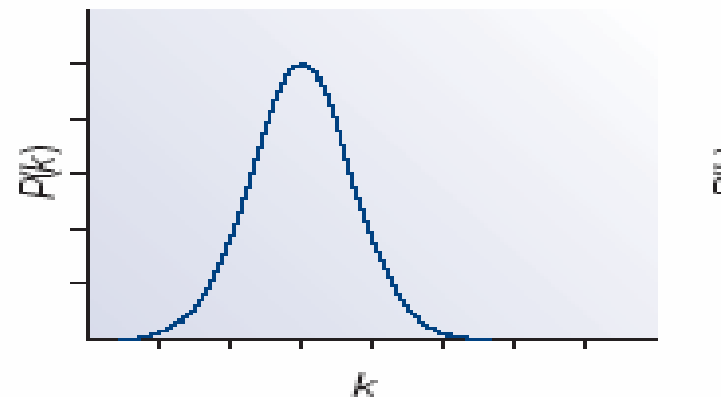


# Random Network

- A fixed number of nodes are connected randomly to each other: start with  $N$  nodes and connect each pair with probability  $p$
- Degrees follow a Poisson distribution: most nodes have roughly the same number of links, approximately equal to the network's average degree,  $\langle k \rangle$ ; nodes that have significantly more or less links than  $\langle k \rangle$  are very rare.



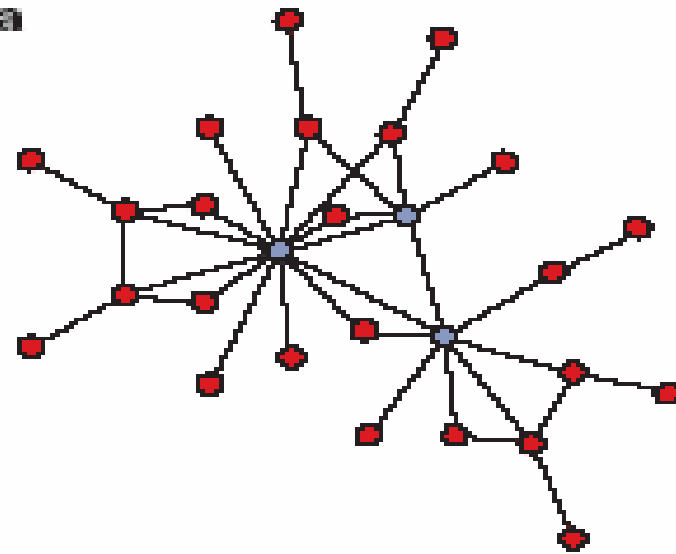
**Ab**



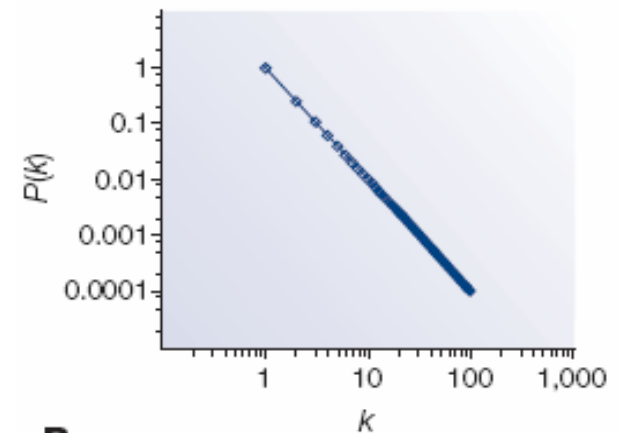
# Biological Networks

- Biological networks have a few nodes with a very large number of links (hubs) and many nodes with only a few links.
- It indicates the absence of a typical node in the network.
- Scale-free networks?

**Ba**



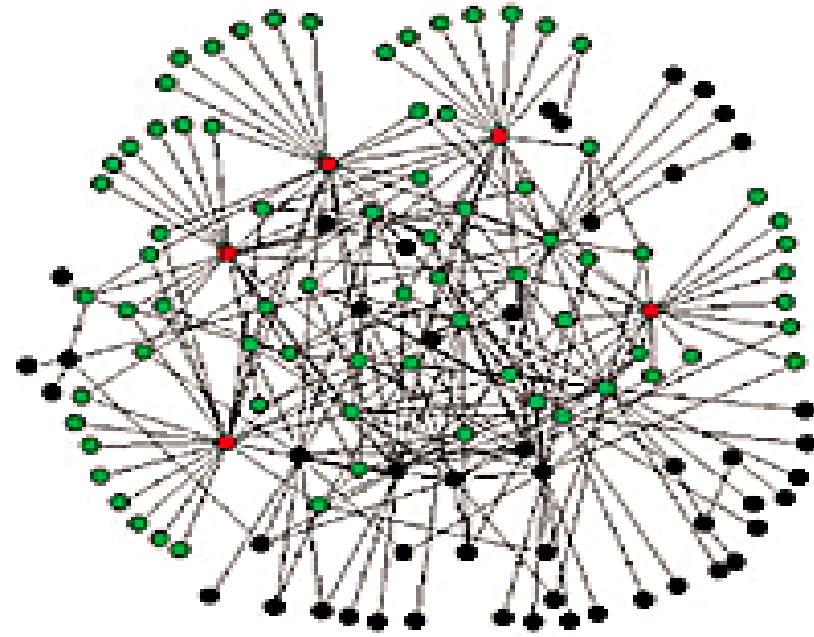
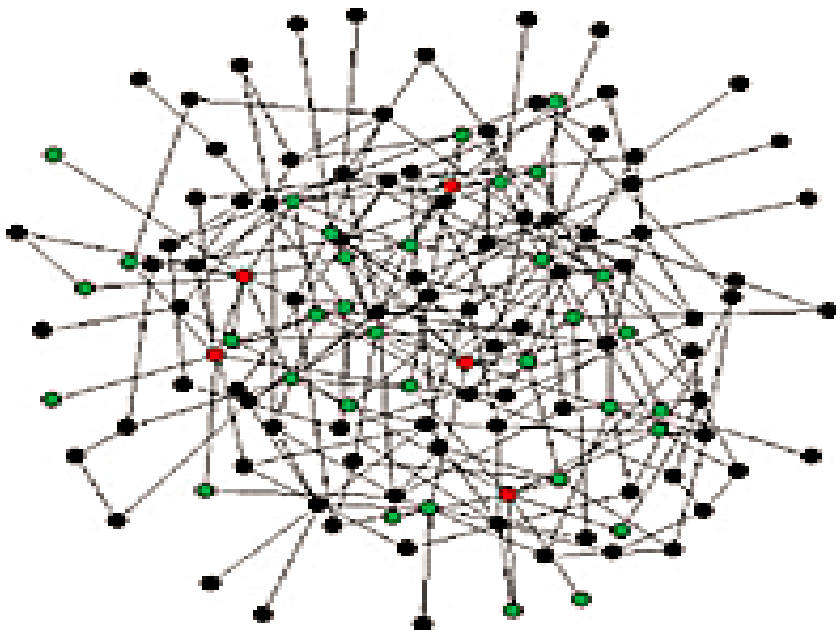
**Bb**





# Comparing Random and Scale-free connectivity distributions

- In the random network, the five nodes with the most links (in red) are connected to only 27% of all nodes (green). In the scale-free network, the five most connected nodes (red) are connected to 60% of all nodes (green) (Barabasi and Oltvai, Nature, 2004)





# Biological scale-free networks

---

- Metabolic networks (Jeong et al, 2000; Fell and Wagner, 2000; Ma and Zeng, 2003)
- Protein-protein interactions (Jeong et al, 2000)
- Protein domain networks (Rzhetsky and Gomez, 2001; Wuchty, 2001)
- Gene interactions (Tong et al, 2004)
- Gene regulatory networks (Guelzim et al, 2002)
- Gene co-expression networks (Bergmann et al, 2004)



# Other biological “scale-free” examples

---

- Distribution of gene expression levels (Ueda et al, 2004)
- and spot intensities on microarrays (Hoyle et al, 2002)
- Frequency of occurrence of generalized parts in genomes of different organisms (Luscombe et al, 2002)



# Examples of other scale-free networks

---

Network of citations between scientific papers

Network of collaborations, social networks

Electrical power grids, airline traffic routes, railway networks

World Wide Web; other communication and computer networks

Internet: number of links per page, number of clicks per page

*"These laws, applying equally well to the cell and the ecosystem, demonstrate how unavoidable nature's laws are and how deeply self-organization shapes the world around us." (A.Barabasi, 2002).*



# Google: scale-free networks

---

"Scale-free networks are everywhere. They can be seen in ... terrorist networks. What this means for counter-terrorism experts?"

Global guerrillas (network organization, infrastructure disruption, and the emerging marketplace of violence)

[http://globalguerrillas.typepad.com/globalguerrillas/2004/05/scalefree\\_terror.html](http://globalguerrillas.typepad.com/globalguerrillas/2004/05/scalefree_terror.html)

Scale|free : <http://www.scalefree.info> (Tools for a connected, distributed, and backed up existence: Social Networking, Communities of Practice and Knowledge Management)

# A unifying paradigm?



THE INTERNET, mapped on the opposite page, is a scale-free network in that some sites (starbursts and detail above) have a seemingly unlimited number of connections to other sites. This map, made on February 6, 2003, traces the shortest routes from a test Web site to about 100,000 others, using like colors for similar Web addresses.

## Scale-Free Networks

Scientists have recently discovered that various complex systems have an underlying architecture governed by shared organizing principles. This insight has important implications for a host of applications, from drug development to Internet security

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

*The concept of scale-free network has emerged as a powerful unifying paradigm in the study of complex networks in biology, physics and social sciences.*

(Barabasi and Oltvai, Nature Rev., 2004)



# Philosophical impact?

---

- *These findings have considerable philosophical impact: the power-law connectivity distribution seen in scale-free networks seems to emerge as one of the very few universal mathematical laws of life.*

Wolf, Karev and Koonin, 2002



# Why such an interest in scale-freeness of biological networks?

---

- One-parametric distribution
- Presence of hubs and a large number of nodes with a few connections has been cited as the most characteristic feature of scale-free biological networks
- Scale-free networks have small-world property
- Robust to random breakdowns





# Not unique to scale-free networks

---

- These are vital properties of biological networks
- However, they are not unique to scale-free networks
- Scale-free networks are just one class of small-world networks
- Any network with so-called broad-tail (or broad-scale) distribution has a few hubs



# A unique property of scale-free networks: invariance to changes in scale

---

The term scale-free refers to a system defined by a functional form  $f(x)$  that remains unchanged within a multiplicative factor under a rescaling of the independent variable  $x$ . Effectively, this means **power-law forms**, since these are the only solution to

$$f(ax) = bf(x) \text{ for all } x$$

Network is defined as **scale-free** if a randomly picked node has  $k$  connections with other nodes with a probability that follows a power-law

$$p(k) \sim k^{-\gamma}$$



# Invariance under re-scaling

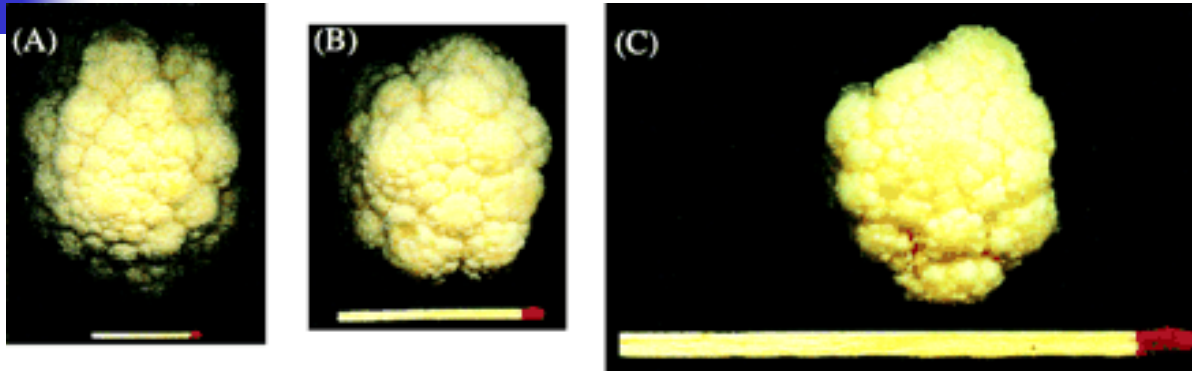
---

The ratio of two connectivities in a scale-free network is invariant under rescaling:

$$\frac{p(k_1)}{p(k_2)} = \frac{p(\alpha k_1)}{p(\alpha k_2)} = F\left(\frac{k_1}{k_2}\right)$$

This implies that scale-free networks are **self-similar**, i.e. any part of the network is statistically similar to the whole network and parameters are assumed to be independent of the system size.

# Scale-free Networks

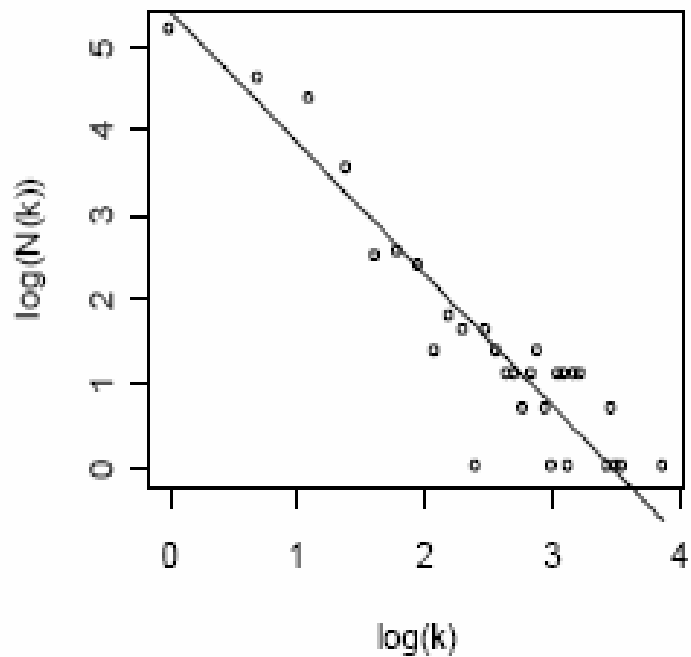


Scale-free (self-similarity) properties of a common cauliflower plant: it is virtually impossible to determine whether one is looking at a photograph of a complete vegetable or its part, unless an additional scale-dependent object (a match) is added. (A) Complete vegetable; (B) a small segment of the same vegetable; (C) small part of the segment shown in B. The same match was used in all three photographs to provide a sense of scale for an otherwise scale-free structure (Gomez et al, 2001).

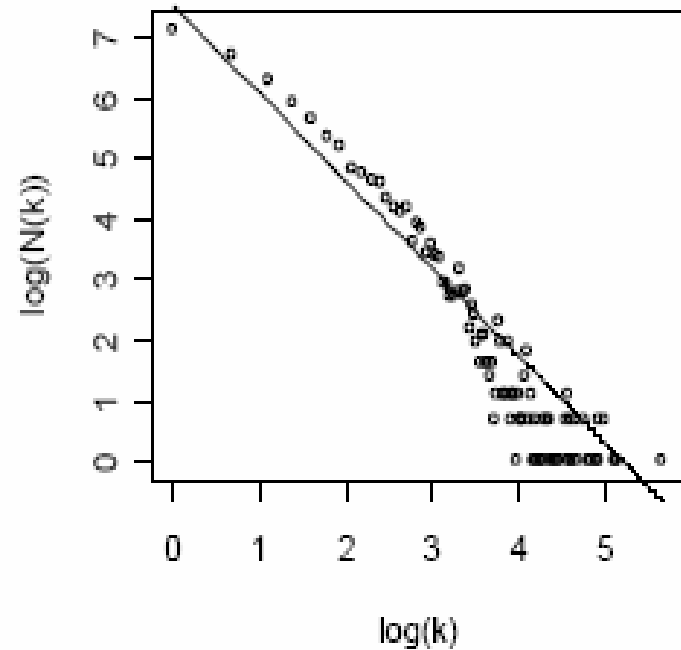
Note: vegetable was purchased in Sloan supermarket in Manhattan's Upper West Side.

# Fitting straight line to the data

Guelzim et al, 2002



Tong et al, 2004





# Fitting straight line to the data

---

- To find indication of a power-law

$$P(k) = \frac{k^{-\gamma}}{\xi(\gamma)}$$

the data is usually graphically fitted  
by a straight line on a log-log scale

$$\log(P(k)) = \log C - \gamma \log(k)$$

- Log contract data
- Fit to a few points is not particularly good
- Fitting a line does not make the points follow it!

# Estimating the power exponent by ML

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)} \quad \zeta(\gamma) = \sum_{j=1}^{\infty} j^{-\gamma}$$

- Observed values:  $x_i$  is a connectivity of node  $i$

$$L(\gamma | x) = \prod_{i=1}^N x_i^{-\gamma} / \zeta(\gamma)$$

- the likelihood function for  $N$  observed connectivities

$$l(\gamma | x) = -\gamma \sum_{i=1}^N \log x_i - N \log \zeta(\gamma)$$

- the log-likelihood is maximized wrt  $\gamma$



# Goodness-of-fit

---

$$E_k = Nk^{-\gamma} / \zeta(\gamma)$$

- $E(k)$  - expected(k) with estimated  $\gamma$ ;  $O(k)$  – observed(k)
- Consider

$$T = \sum_{k=1}^{k^*} (O_k - E_k)^2 / E_k \sim \chi_{k^*-2}^2$$

(approximately) under  $H_0$ : network is scale-free

- Pool for connectivity values over  $k^*$ , for which  $E(k) < 5$ . The chi-squared statistic is approximately chi-squared distributed with  $k^*-2$  dof. The p-value, calculated by the exceedence probability of a chi-squared distribution, is the probability a network has such connectivities if they were drawn from the power-law distribution



# Exponents of the power-law and “scale-free” p-values

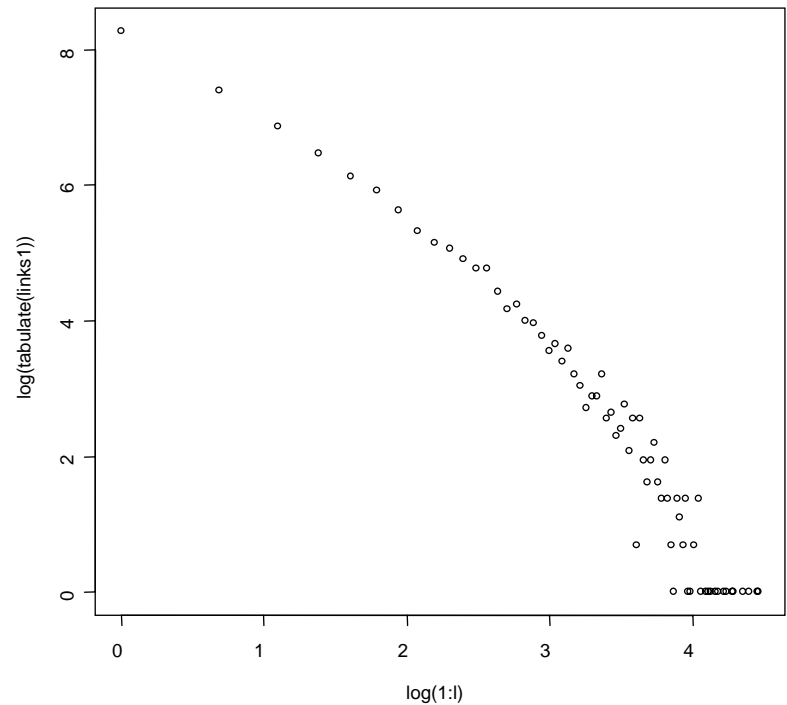
Datasets	$\hat{\gamma}$	$p$ -value
Uetz	2.05	0.0004
Schwikowski	1.87	$<10^{-4}$
Ito	2.02	$<10^{-4}$
Li	2.31	0.0014
Rain	1.61	$<10^{-4}$
Giot	1.53	$<10^{-4}$
Tong	1.45	$<10^{-4}$
Lee	2.03	$<10^{-4}$
Guelzim	1.56	$<10^{-4}$
Spellman/Cho	1.10(0.96)	$<10^{-4}$ ( $<10^{-4}$ )

P-value summarizes the evidence against  $H_0$  that the network is scale-free, suggesting that the networks are NOT consistent with power-law distributions.

# Truncated power-law

$$p(k) \sim k^{-\gamma} \exp(-k / k_c)$$

$k_c$  is the cut-off, s. t. the number of connections is less than expected for pure scale-free networks for  $k > k_c$  and the behaviour is approximately scale-free within the range  $1 \leq k < k_c$



# Parameters of the truncated power-law and “truncated power-law” p-values

Datasets	$\hat{\gamma}$	$\hat{k}_c$	$P$ -value
Uetz	1.60	8.7	0.370
Schwikowski	1.26	6.2	0.105
Ito	1.79	26	$<10^{-4}$
Li	2.10	19.5	0.018
Rain	1.12	11.5	0.200
Giot	1.09	20	0.001
Tong	0.96	23.7	$<10^{-4}$
Lee	1.96	294	$<10^{-4}$
Guelzim	1.18	15	$<10^{-4}$
Spellman/Cho	1.07(0.78)	73(99)	0.700 (0.100)



## Exponents of the power-law and “scale-free” p-values under relaxing the independence assumption

Datasets	$\hat{\gamma}$	$p$ -value
Uetz	1.96	$<10^{-3}$
Schwikowski	1.79	$<10^{-3}$
Ito	1.98	$<10^{-3}$
Li	2.26	$<10^{-3}$
Rain	1.52	$<10^{-3}$
Giot	1.48	$<10^{-3}$
Tong	1.39	$<10^{-3}$
Lee	2.03	0.007
Guelzim	1.52	$<10^{-3}$
Spellman/Cho	1.08(0.94)	$<10^{-3}$ ( $<10^{-3}$ )

# Scale-free or not: why is it important?

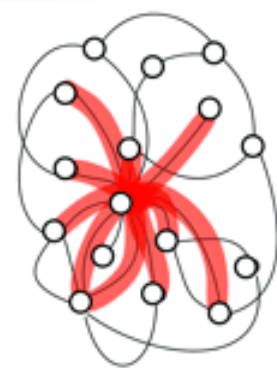
- Network architecture



Regular



Random



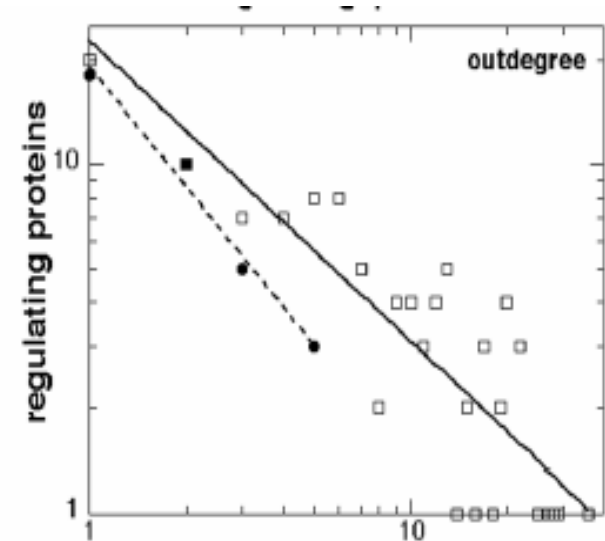
Scale-free

- Evolutionary models

- Scalability (self-similarity) criterion does not hold; one has to exercise caution while applying the features of an already studied part of the network to investigate properties of the unknown part of the same network, or of other networks that seem similar


# Example of gene network

- Guelzim and co-authors (Nature, 2002) conclude that connectivity distributions of gene transcriptional network for yeast and E-coli are scale-free, and thus *"bacterial and fungal genetic networks are free of characteristic scale with respect to the distributions of both regulating and departing connections"*.



Number of regulated genes per regulating protein

We found, however, that the scale-free property does not even hold and therefore such biological conclusions are invalid



# Evolutionary models that yield scale-free networks

---

- **Preferential attachment model** (Barabasi and Albert)
- **Duplication and Divergence models**, whereby a newly created node inherits all or part of the connections of the parent node (Bhan et al, 2002; Rzhetsky and Gomez, 2001; van Noort et al, 2004)
- **Divergent “Big Bang” model** with evolutionary drift (Dokholyan et al, 2002)
- **Birth-Death-Innovation model** (Karev et al, 2002; 2004)
- Parameters and assumptions of these models are adjusted to yield scale-free networks.



# Evolutionary models

---

- Plausible evolutionary phenomena such as evolutionary drift has been shown to contradict the power-law (Przytycka and Yu, 2004)
- Model for link dynamics and gene duplications (Berg et al, 2004) results in broad connectivity distribution





# Qualitative properties of biological networks still hold!

---

- Existence of hubs
- Many nodes with a few connections
- Small-world property
- High clustering of nodes
- And various other measures



# Alternative non-scale free distributions

---

$$p(k) \sim (k+k_0)^{-\gamma} \exp(-(k+k_0)/k_c)$$

- generalized truncated power-law

$$p(k) \sim k^{-(\gamma+1)} \exp(-(k/k_c)^{-\gamma})$$

- stretched exponential distribution
- geometric random graph: a geometric graph with  $n$  independently and uniformly distributed points in a metric space (Przulj, Bioinformatics, 2007)
- Stumpf and colleagues (2005) applied formal statistical model selection methods to determine which functional form best describes networks. And it is **not the power-law!**



# Sampling of PPI networks

---

- PPI networks are sampled only partially!
- Sampling of networks with different topologies (random, exponential, power-law, truncated normal) result in scale-free networks (Han et al, Nature Biotech, 2005).
- On the contrary, subnets of scale-free networks are not scale-free (Stumpf et al, PNAS, 2005).



# Sampling of PPI networks

---

- PPI Networks contain lots of non-specific interactions; interesting non-random topological features of these graphs need not have any evolutionary origin (Deeds et al, PNAS, 2006).
- Observed topology of subnets cannot be confidently extrapolated to complete networks!



# Is there Network Biology?

---

Get the Topology right!