Scale-free networks

Prof. Pascal Frossard

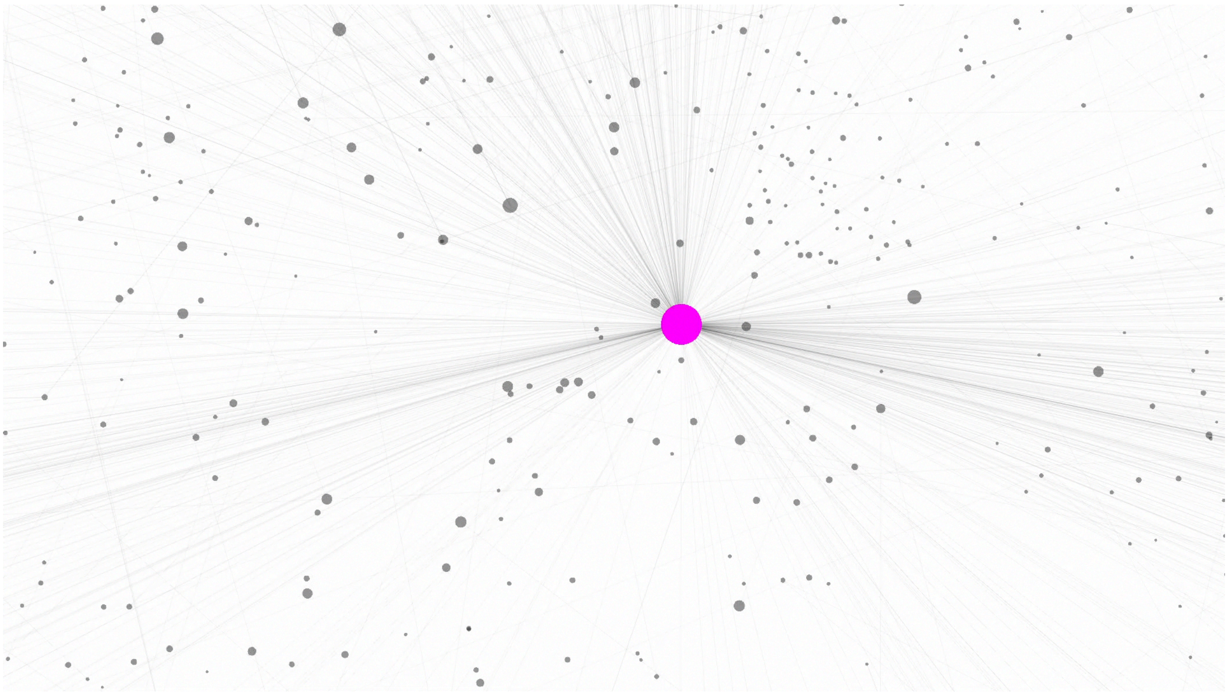
Signal Processing Laboratory (LTS4)

Some slides are taken from Prof. Barabási's class on Network Science (www.BarabasiLab.com)

Outline

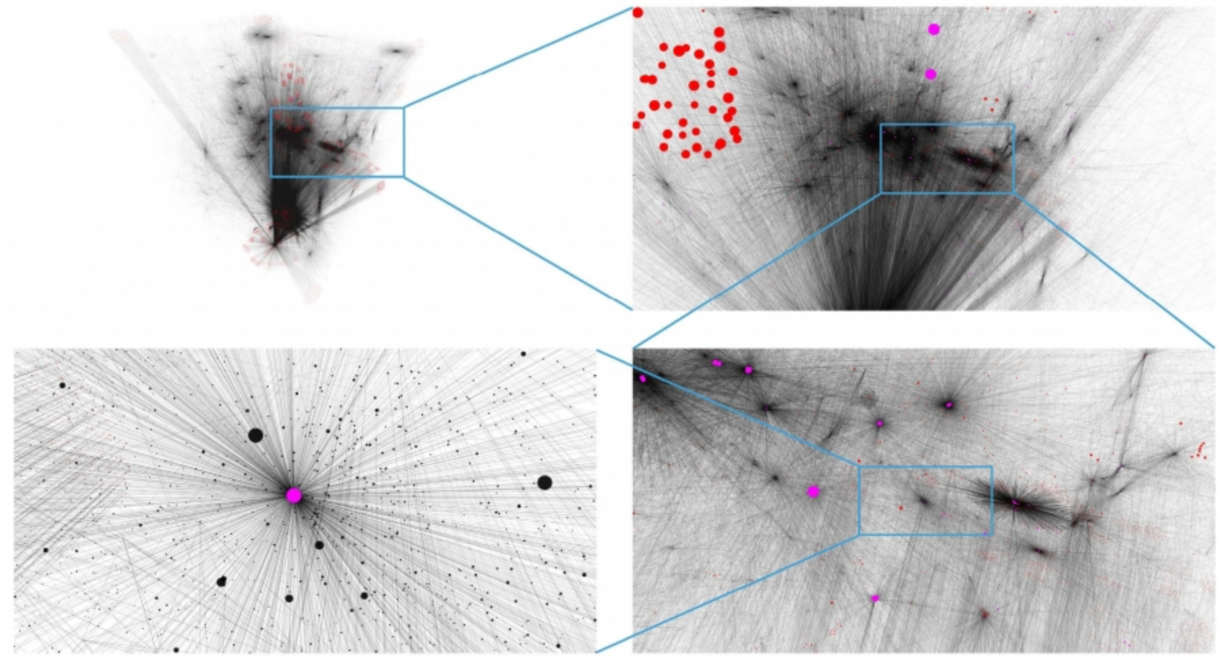
- Power Laws
- Hubs
- The Meaning of Scale-Free
- Ultra-Small Property
- Role of the degree exponent

Zoom into the WWW



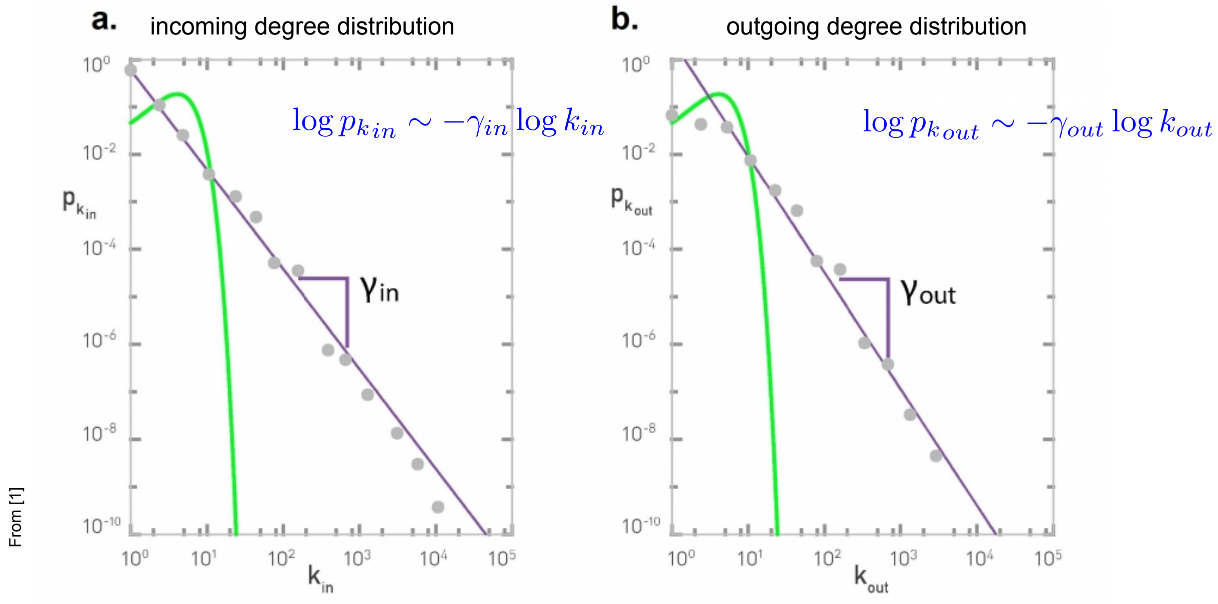
From [1]

The Web is not random



From [1]

WWW: degree distribution



- Degrees do not follow a Poisson distribution (like random networks) but rather a *power law distribution*, of the form $\boxed{p_k \sim k^{-\gamma}}$

Scale-free network

A *scale-free network* is a network whose degree distribution follows a *power law*.

Discrete Formalism

Probability that a node has k links:

$$p_k = Ck^{-\gamma}$$

With normalisation constraints

$$\sum_{k=1}^{\infty} p_k = 1 \quad \text{or} \quad C\sum_{k=1}^{\infty} k^{-\gamma} = 1$$

The parameter *C* becomes

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

Riemann-zeta function

Finally: $p_k = \dfrac{k^{-\gamma}}{\zeta(\gamma)}$

p_0 can be specified separately

Continuum Formalism

Probability of node degree between k_1 and k_2

$$\int_{k_1}^{k_2} p(k)dk \quad \text{with} \quad p(k) = Ck^{-\gamma}$$

With normalisation constraints $\int_{k_{\min}}^{\infty} p(k)dk = 1$

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma}dk} = (\gamma-1)k_{\min}^{\gamma-1}$$

Finally: $\boxed{p(k) = (\gamma-1)k_{\min}^{\gamma-1}k^{-\gamma}}$

Hubs

- Main difference between random and scale-free networks:

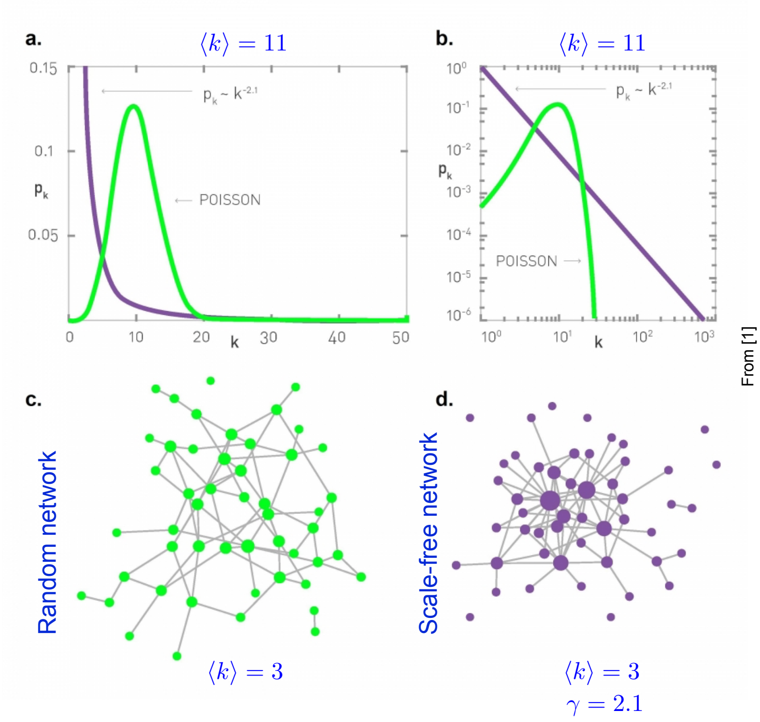
  the tail of degree distribution!

- Scale-free networks have
  - many small degree nodes
  - not so many nodes around $\langle k \rangle$
  - high-degree nodes, aka *hubs*

- WWW example, $\langle k \rangle = 4.6$
  - with Poisson:

  $p_{100} = 10^{-94}$ $\quad N_{k \geq 100} = \simeq 10^{-82}$
  - with power-law, $\gamma_{in} = 2.1$

  $p_{100} = 4 \times 10^{-4}$ $\quad N_{k \geq 100} = 4 \times 10^{9}$



Largest hub

- The expected size of the largest hub is the maximum degree k_{max}
  - it is the *natural cutoff* of the degree distribution p_k

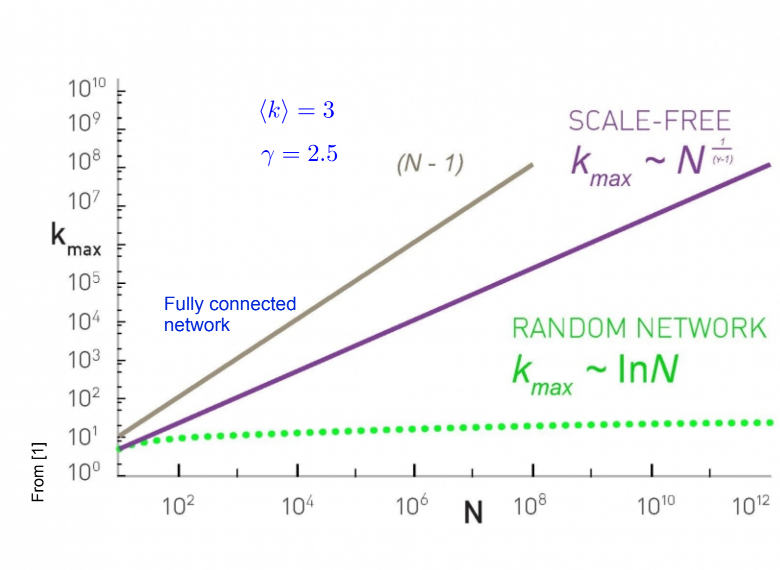
- For scale-free networks, the larger the network, the larger the biggest hub
  - we calculate the maximum degree such as there is at most one node in $[k_{max}, \infty[$
  - this means that

$$\int_{k_{\max}}^{\infty} p(k)dk = \frac{1}{N}$$

  - with $p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$, we can compute $\boxed{k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}}$

- The polynomial dependence on *N* means that there could be orders of magnitude differences between $k_{\min}$ and $k_{\max}$
  - the dependence of $k_{\max}$ on *N* is much slower for random networks

Hubs are big in Scale-free networks



For random networks, we consider an approximation with $p(k) = Ce^{-\lambda k}$

Then, with the same definitions

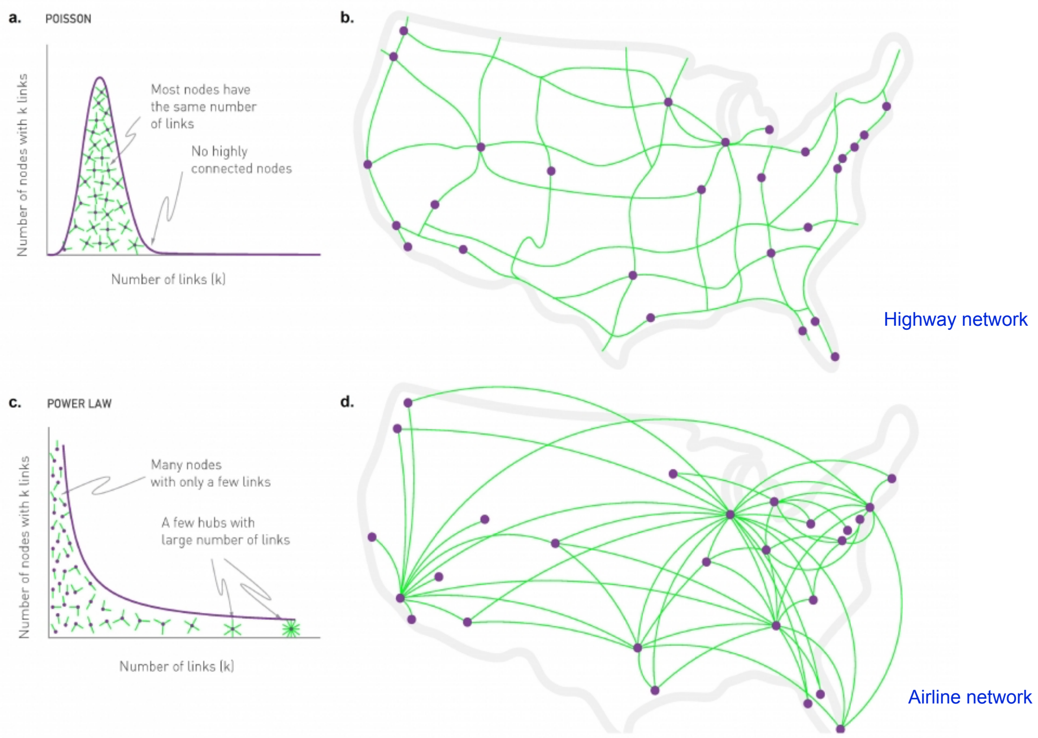
$$\int_{k_{\min}}^{\infty} p(k)dk = 1$$

$$\int_{k_{\max}}^{\infty} p(k)dk = \frac{1}{N}$$

We get $k_{\max} = k_{\min} + \frac{\ln N}{\lambda}$

Hubs in a scale-free network are several orders of magnitude larger than the biggest node in a random network with the same *N* and $\langle k \rangle$!!

Random vs Scale-free networks



From [1]

Why 'Scale-free'?

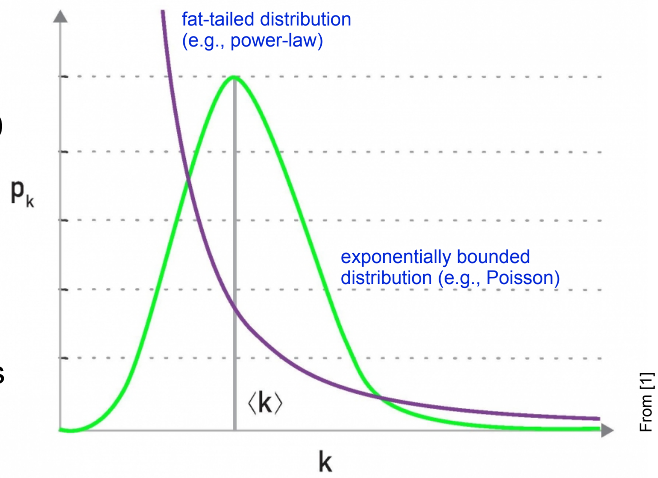
- The n^th moment of the degree distribution: $\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n p_k \approx \int_{k_{\min}}^{\infty} k^n p(k) dk$
  - for $n = 1$ we have the average degree $\langle k \rangle$
  - for $n = 2$ the second moment $\langle k^2 \rangle$ is related to the variance $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$
    σ: the standard deviation
  - for $n = 3$ the third moment $\langle k^3 \rangle$ determines the *skewness*
    symmetry around $\langle k \rangle$

- For a scale-free network: $\langle k^n \rangle = \int_{k_{\min}}^{k_{\max}} k^n p(k) dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n - \gamma + 1}$
  - $k_{\min}$ is fixed, $k_{\max}$ depends on *N*
  - limit when $k_{\max} \to \infty$
    - if $n - \gamma + 1 \leq 0$ then $k_{max}^{n-\gamma+1} \to 0$: all moments with $n \leq \gamma - 1$ are finite
    - if $n - \gamma + 1 > 0$ then all moments diverge, i.e., $\langle k^n \rangle \to \infty$
  - for many scale-free networks, $\gamma \in [2, 3]$, then $\langle k \rangle$ is finite, all other moments diverge
  - the degrees are typically in the range $k = \langle k \rangle \pm \sigma_k$, which gets 'unbounded'

Networks with $\gamma < 3$ do not have a meaningful internal scale: they are scale-free!

Lack of an Internal Scale

- For random networks:
  - we have $\sigma_k = \langle k \rangle^{\frac{1}{2}}$
  - hence, the internal scale is $\langle k \rangle$
- For regular lattices
  - all nodes have the same degree: $\sigma = 0$
- Scale-free networks
  - unbounded 'variance'
  - nodes with widely different degrees coexist in the same network
  - divergence of second moment explains some of the most intriguing properties such as robustness to random failures and anomalous spread of viruses.

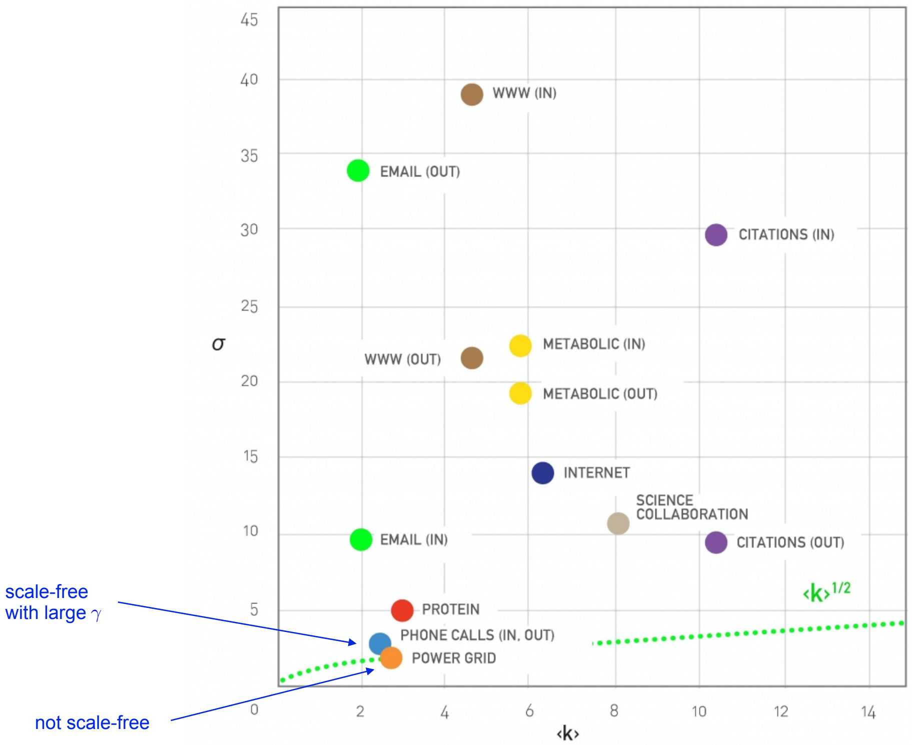


Degree Fluctuations in Real Networks

Network	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

From [1]

Standard Deviation is Large

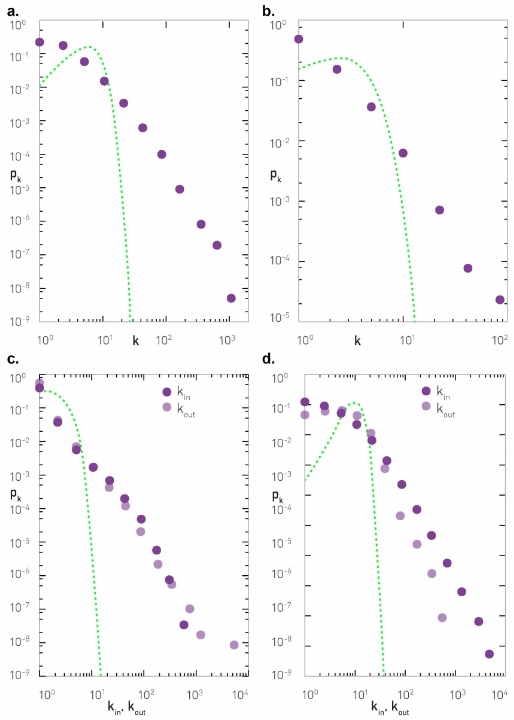


From [1]

Universality of scale-free properties

Internet at the router level

Protein–protein interaction network



Email network

Citation network

From [1]

Distributions in network science

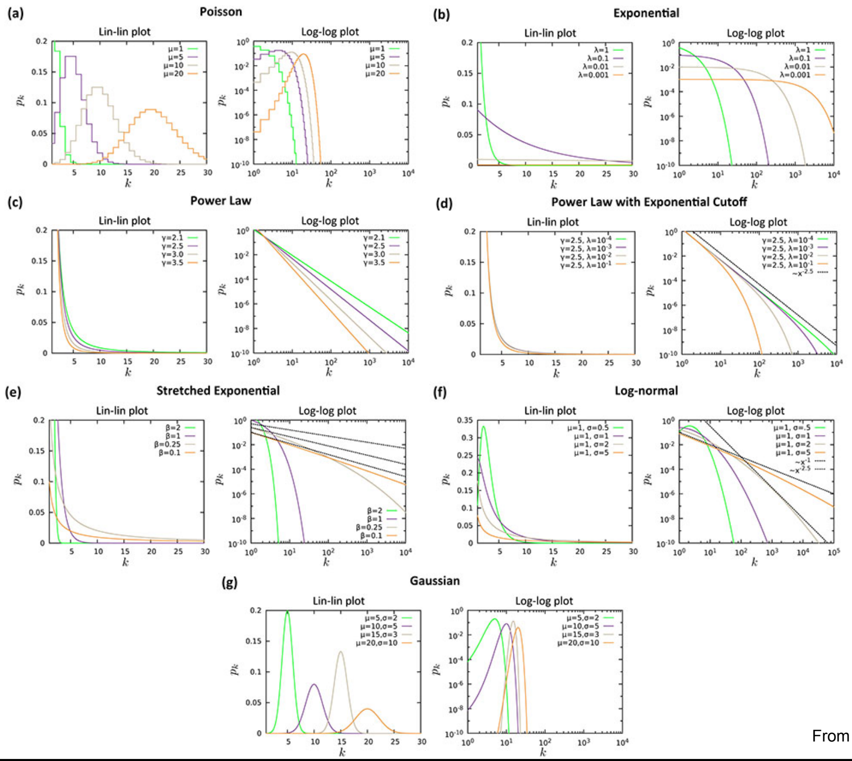
NAME	$p_x/p(x)$	$\langle x \rangle$	$\langle x^2 \rangle$
Poisson (discrete)	$e^{-\mu}\mu^x/x!$	μ	$\mu(1+\mu)$
Exponential (discrete)	$(1-e^{-\lambda})e^{-\lambda x}$	$1/(e^{\lambda}-1)$	$(e^{\lambda}+1)/(e^{\lambda}-1)^2$
Exponential (continuous)	$\lambda e^{-\lambda x}$	$1/\lambda$	$2/\lambda^2$
Power law (discrete)	$x^{-\alpha}/\zeta(\alpha)$	$\begin{cases} \zeta(\alpha-2)/\zeta(\alpha), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \le 1 \end{cases}$	$\begin{cases} \zeta(\alpha-1)/\zeta(\alpha), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \le 2 \end{cases}$
Power law (continuous)	$\alpha x^{-\alpha}$	$\begin{cases} \alpha/(\alpha-1), & \text{if } \alpha > 2 \\ \infty, & \text{if } \alpha \le 1 \end{cases}$	$\begin{cases} \alpha/(\alpha-2), & \text{if } \alpha > 1 \\ \infty, & \text{if } \alpha \le 2 \end{cases}$
Power law with cutoff (continuous)	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)}x^{-\alpha}e^{-\lambda x}$	$\lambda^{-1}\frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}$	$\lambda^{-2}\frac{\Gamma(3-\alpha)}{\Gamma(1-\alpha)}$
Stretched exponential (continuous)	$\beta\lambda^{\beta}x^{\beta-1}e^{-(\lambda x)^{\beta}}$	$\lambda^{-1}\Gamma(1+\beta^{-1})$	$\lambda^{-2}\Gamma(1+2\beta^{-1})$
Log-normal (continuous)	$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-(\ln x-\mu)^2/(2\sigma^2)}$	$e^{\mu+\sigma^2/2}$	$e^{2(\mu+\sigma^2)}$
Normal (continuous)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	$\mu^2+\sigma^2$

Exponentially Bounded Distributions

Fat Tailed Distributions

From [1]

Distributions visualised



From [1]

Ultra-Small Property

- Do hubs actually affect the small-world property?
  - Actually, distances in a scale-free network are smaller than the distances observed in an equivalent random network

- Dependence of the average distance on *N* and γ

$$\langle d \rangle \sim \begin{cases} const. & \gamma = 2 \\ \ln \ln N & 2 < \gamma < 3 \\ \frac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$

Anomalous regime: $k_{\max} \sim N$ and all nodes connect to a central node (*hub and spoke* configuration)

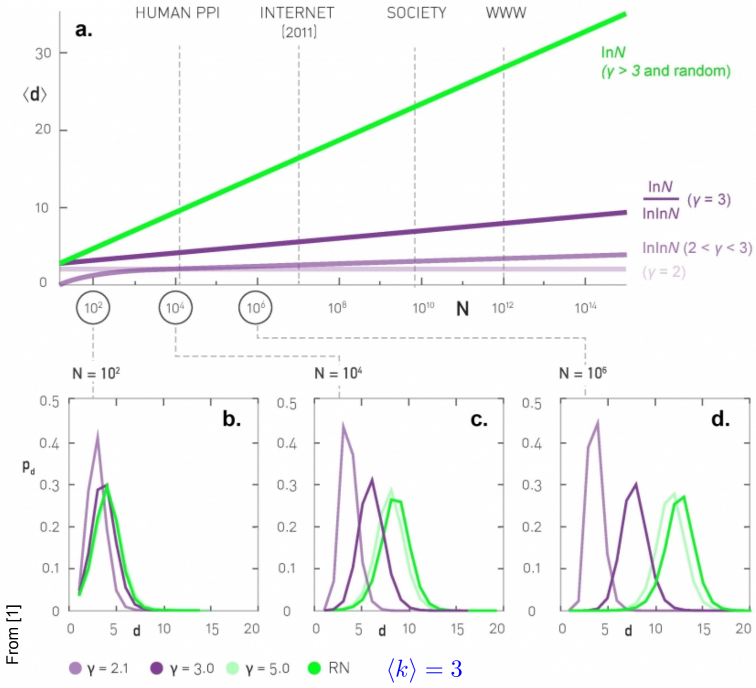
Ultra-small regime: the average distance grows slower than for random networks, as hubs radically reduce the path length

Small world: $\langle k^2 \rangle$ becomes finite and distance follows the small world results from random networks.

Critical point: the second moment does not diverge any more: logarithmic dependence on *N* (as for random networks) but with a correction that shrinks distance wrt random networks.

R. Cohen and S. Havlin. Scale free networks are ultrasmall. Phys. Rev. Lett. 90, 058701, 2003.
B. Bollobás and O. Riordan. The Diameter of a Scale-Free Random Graph. Combinatorica, 24: 5-34, 2004.

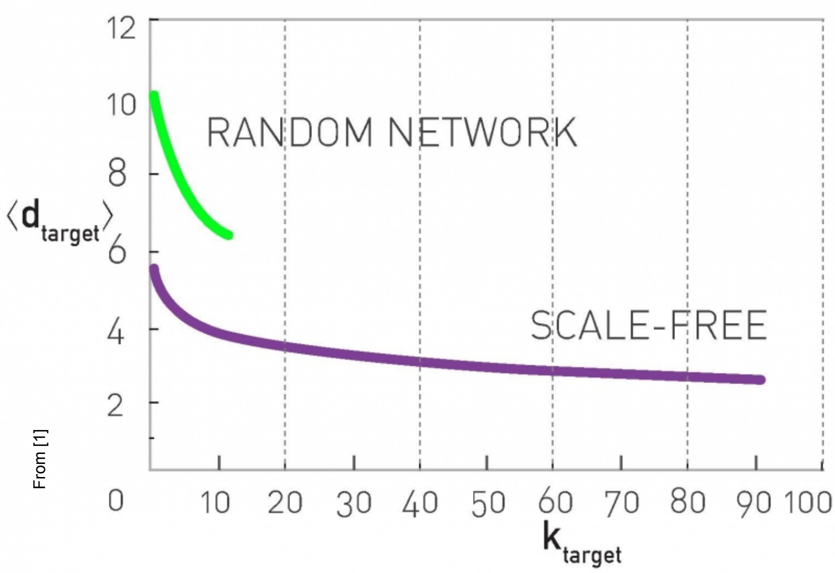
Distances in scale-free networks



- For small networks, all distances are comparable
- For large *N*, differences are important
- For small *N*, path length distribution overlap
- For large *N*, path length distributions are quite different for different γ

Scale-free property shrinks distances!

Always close to hubs



- In general, a node is typically closer to hubs than to less connected nodes.
- This effect is particularly pronounced in scale-free networks
  - There are always short paths linking hubs
  - Many of the shortest paths go through these hubs

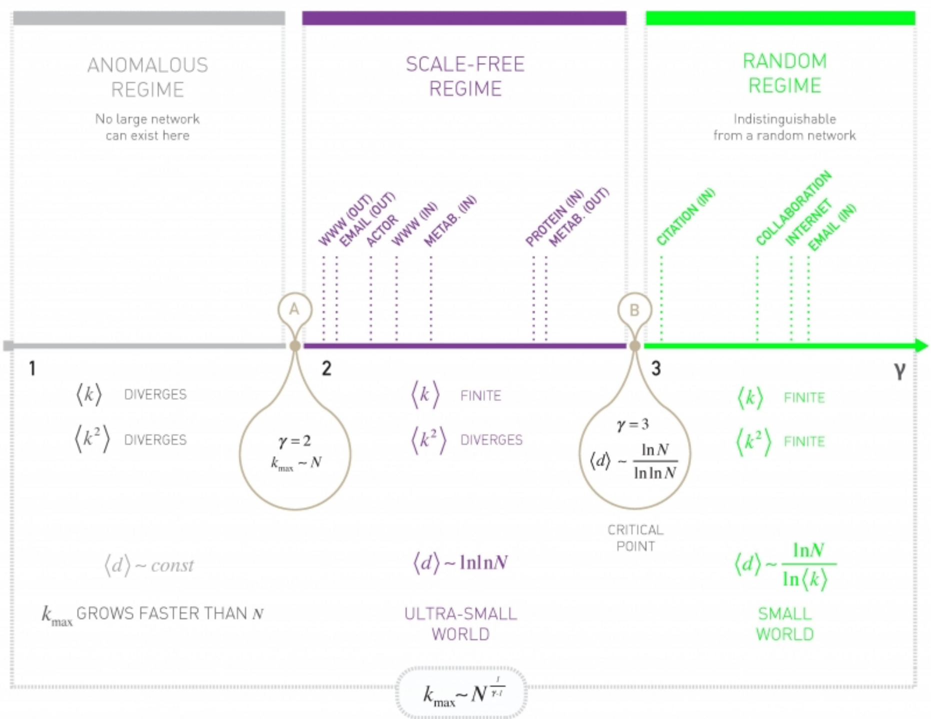
The role of the degree exponent

- Many properties of a scale-free network depend on γ
- Anomalous Regime ($\gamma \leq 2$) — cannot exist without multilinks
  - the number of links to the largest hub grows faster than *N:* out of nodes to connect to
  - the average degree $\langle k \rangle$ diverges for $N \to \infty$
- Scale-Free Regime ($2 < \gamma < 3$) — most interesting regime in practice
  - first moment is the only finite moment as $N \to \infty$: ultra-small world regime
  - the market share of the largest hub decreases as $k_{\max}/N \sim N^{-\frac{\gamma-2}{\gamma-1}}$
- Random Network Regime ($\gamma > 3$) — quite similar to random networks
  - first and second moments are finite: it has similar properties as a random network
  - for large γ , the degree distribution decays fast: hubs are small and less numerous
  - hard to observe a high degree power-law: we need large networks, at least

$$N = \left(\frac{k_{max}}{k_{\min}}\right)^{\gamma-1}$$

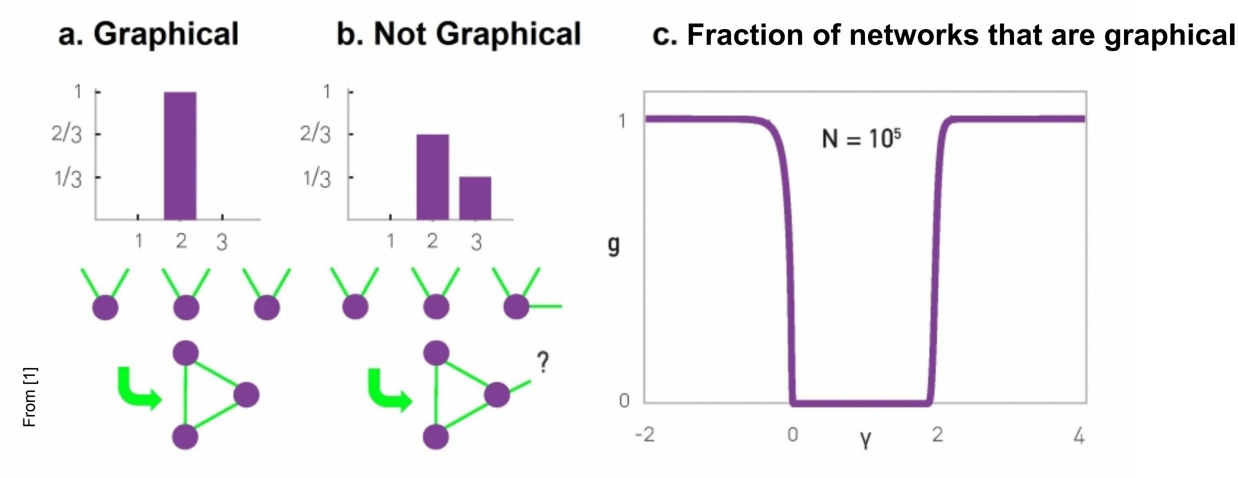
with at least 2-3 orders of magnitude between $k_{\min}$ and $k_{\max}$

γ - dependent properties



From [1]

No scale-free networks for $\gamma < 2$?

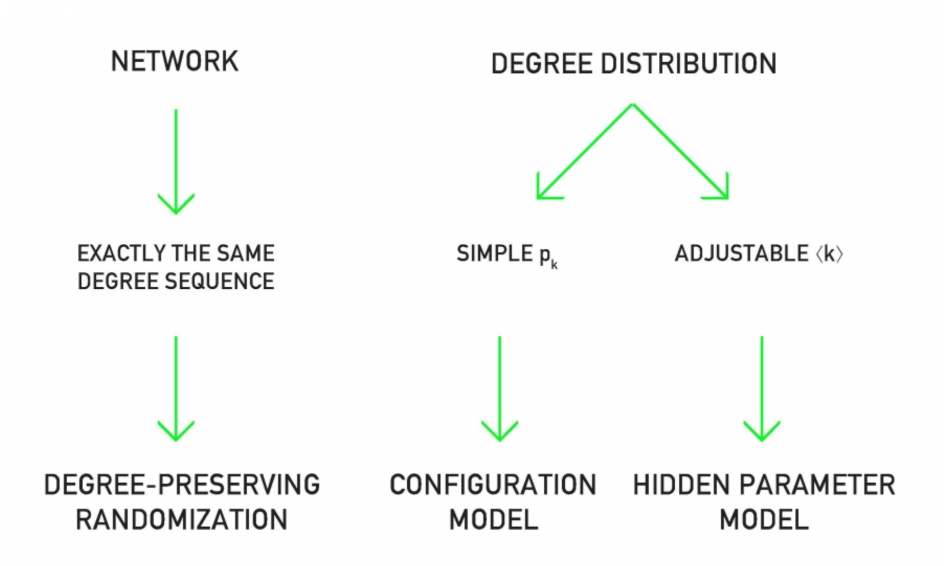


- A degree sequence that can be turned into simple graph (i.e., a graph lacking multi-links or self-loops) is called graphical - it can be verified by a specific algorithm
- If we apply the algorithm to scale-free networks we find that the number of graphical degree sequences drops to zero for $\gamma < 2$ (the largest hub grows faster than *N* !)

I. Charo Del Genio, G. Thilo, and K.E. Bassler. All scale-free networks are sparse. Phys. Rev. Lett. 107:178701, 10 2011.

Scale-free networks generation

- Analysis or comparisons might require to generate scale-free networks
- Different methods are available, depending on requirements and constraints

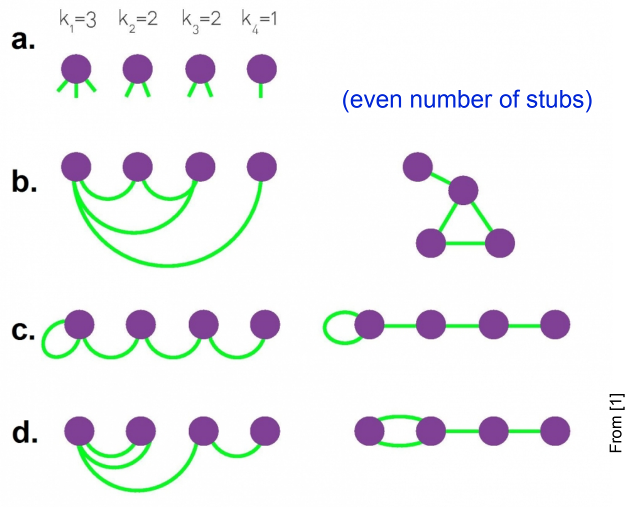


Configuration model

- Generating a network with a pre-defined degree distribution
  - pick a random pair of stubs and connect them
  - repeat the process iteratively
- Probability of connecting *i* and *j*

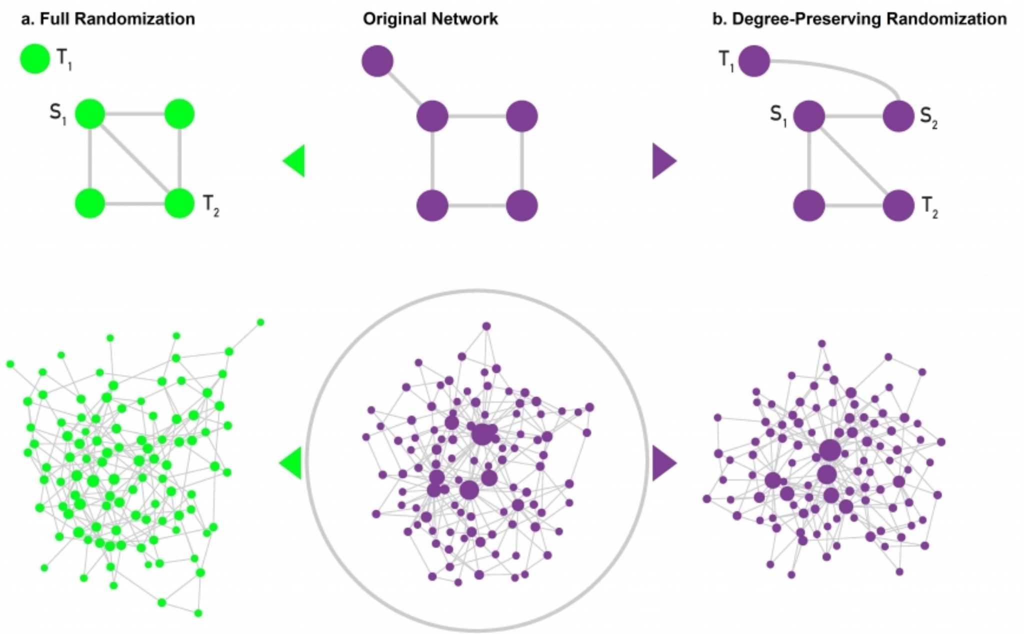
$$p_{ij} = \frac{k_i k_j}{2L-1}$$

- Self-loops and multi-links are possible
  - their probability does to 0 for large *N*



See [2] for details

Degree-preserving randomisation



Generation of a random network

Generation of a network with similar degree distribution as reference one

Hidden parameter model

- Generation of scale-free networks with a predefined degree distribution and no self-loop or multi-links
  - start with isolated nodes, assign hidden parameter η_i chosen from $p(\eta)$ or given by $\{\eta_i\}$ leading resp. to

$$p_k = \int \frac{e^{-\eta}\eta^k}{k!}\rho(\eta)d\eta \quad \text{or} \quad p_k = \frac{1}{N}\sum_j \frac{e^{-\eta_j}\eta_j^k}{k!}$$

$$\langle k \rangle = \langle \eta \rangle$$

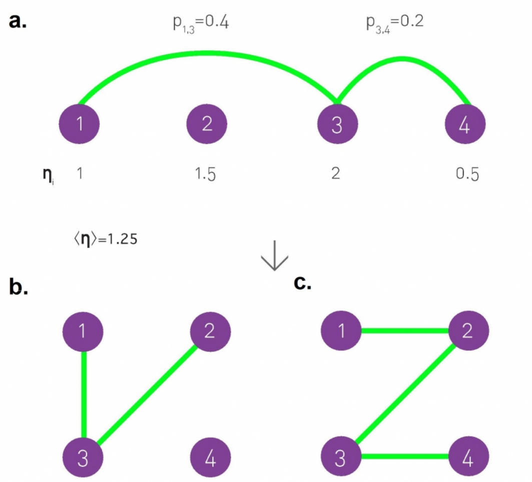
  - connect each node with probability

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

  - to have a scale-free network, use

$$\eta_i = \frac{c}{i^\alpha},\ i = 1, ..., N \longrightarrow p_k \sim k^{-(1+\frac{1}{\alpha})}$$

tune γ through α



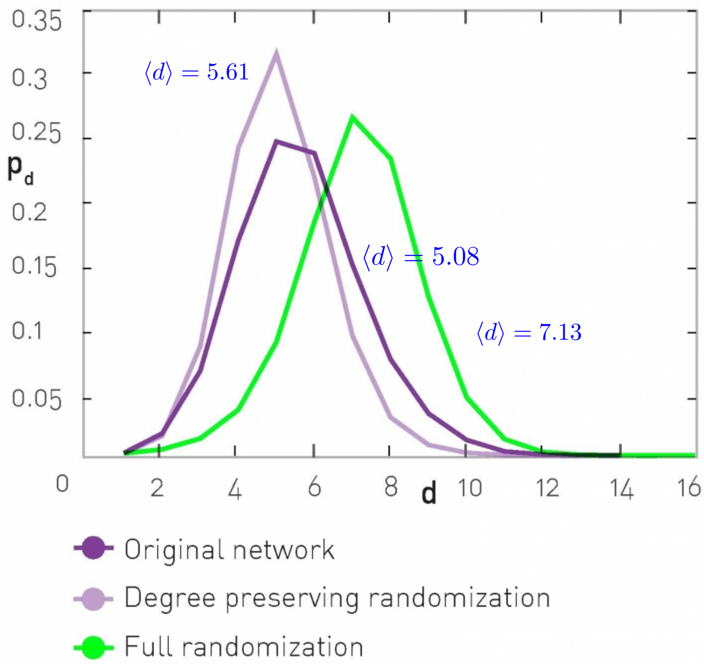
Sample network results

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

B. Söderberg. General formalism for inhomogeneous random graphs. Phys. Rev. E 66: 066121, 2002.
M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. Phys. Rev. E 68: 036112, 2003.
G. Caldarelli, I. A. Capocci, P. De Los Rios, and M.A. Muñoz. Scale- Free Networks from Varying Vertex Intrinsic Fitness. Phys. Rev. Lett. 89: 258702, 2002.

Testing the small-world property

Degree-preserving randomisation preserves the (small) expected degree value



Full randomisation increases the expected degree value: loss of scale-free property of the original network

Summary: scale-free networks

At a glance: Scale-free networks

- Degree distribution

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)} \quad \text{or} \quad p(k) = (\gamma - 1) k_{\min}^{\gamma-1} k^{-\gamma}$$

- Size of the largest hub

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

- Moments of p_k for $N \to \infty$
  - $2 < \gamma \leq 3:$ $\langle k \rangle$ finite, $\langle k^2 \rangle$ diverges
  - $\gamma > 3:$ $\langle k \rangle$ and $\langle k^2 \rangle$ finite
- Distance

$$\langle d \rangle \sim \begin{cases} const. & \gamma = 2 \\ \ln \ln N & 2 < \gamma < 3 \\ \frac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$

- Exponentially bounded networks (fast decreasing degree distribution for high *k*) lack outliers
  - such as highway networks or power grids: most nodes have comparable degrees
- Scale-free networks (fat tailed degree distributions) have hubs that change the system's behaviour
  - Many networks of scientific and practical interest, from the WWW to the subcellular networks or social ones
  - Outliers are expected in these networks
- Power-law rarely seen in 'pure' form:
  - If $\langle k^2 \rangle$ is large: scale-free behaviour; if $\langle k^2 \rangle$ is small and comparable to $\langle k \rangle$: random network approximation holds

References

[1] Network Science, by Albert-László Barabási, 2016 - Chapters 4 and 5

[2] Networks: An Introduction, by M. Newman, 2010