MUNI
FI

Hubs, Rich Club, Scale Free Network

IV124

Josef Spurný & Eva Výtvarová

Faculty of Informatics, Masaryk University

March 31, 2023

Hubs

Short definition: nodes with high degree

What does high mean:

- reminder: binomial degree distribution in random network
- hubs: far to the right from the expected distribution
- *far* means, for example, at least one standard deviation from the mean

Hubs: A ZOO of Complex Networks

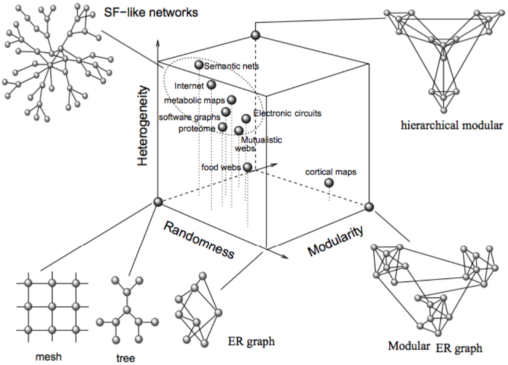


FIG. 3 A zoo of complex networks. In this qualitative space, three relevant characteristics are included: randomness, heterogeneity and modularity. The first introduces the amount of randomness involved in the process of network's building. The second measures how diverse is the link distribution and the third would measure how modular is the architecture. The position of different examples are only a visual guide. The domain of highly heterogeneous, random hierarchical networks appears much more occupied than others. Scale-free like networks belong to this domain. [1]

[1] https://noduslabs.com/radar/types-networks-random-small-world-scale-free/

Hubs: Motivation

Why do we observe hubs?

- network structure is a result of self-organization
- real-world systems have limited resources
- maintaining links is costly
- hubs allow coordination, faster spreading...
- hubs are found across multiple scales - fractal (scale-free) distribution
- scale-free distribution is a result of optimization process in systems with finite resources[2]

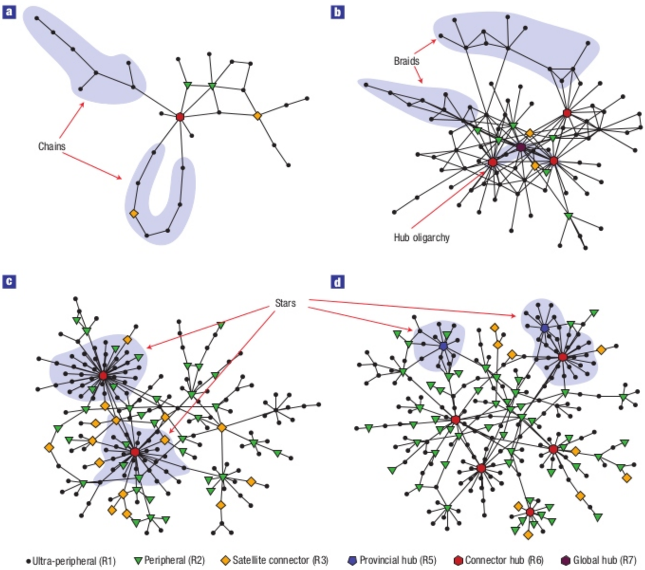
[2] Csermely, P. (2006), pp. 20. Weak links.

Hubs: Motivation

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



Guimera et al. (2007) doi:10.1038/nphys489

Hubs in Detail

If the network has a community structure, we can distinguish between:

- global hubs
- provincial hubs within modules
- connector hubs connecting multiple modules
- peripheral hubs
- satellite connectors

We quantify this using the so-called participation coefficient.

Participation Coefficient[3]

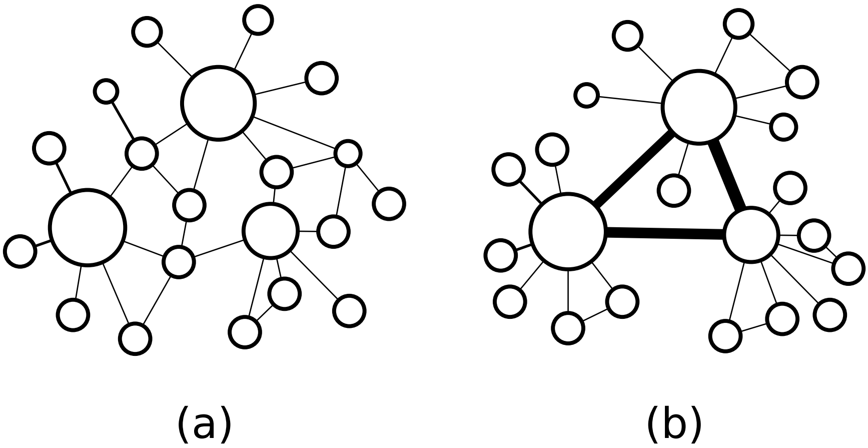
$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{\kappa_{is}}{k_i} \right)^2$$

- N_M is the total number of modules
- κ_{is} is the number of edges from node i to module s

- $P \leq 0.3$ provincial hubs
- $0.3 < P \leq 0.75$ connectors
- $0.75 < P$ global hubs

[3] Guimera, Amaral (2008). Functional cartography of complex metabolic networks.

Rich-club



Rich-club (k-core)

Rich club is a result of network assortativeness (homophily)

Describes whether
dominant nodes form a tightly interconnected core.

$$\varphi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}$$

- $E_{>k}$ is the number of edges between $N_{>k}$ nodes with degree greater than k
- it represents the fraction of edges between these nodes out of all possible edges

Rich-club: Normalization

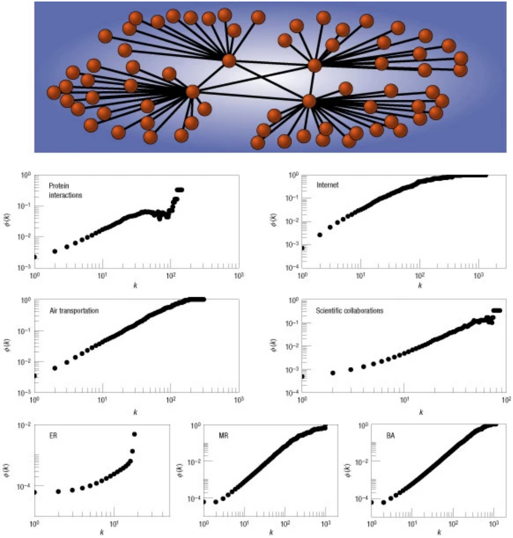
Null model:

$$\varphi_{un}(k) \sim \frac{k^2}{\langle k \rangle N}$$

Normalized RC:

$$\rho_{unc}(k) = \frac{\varphi(k)}{\varphi_{un}(k)}$$

Rich-club examples[4]



[4] Colizza et al. (2006) doi:10.1038/nphys209

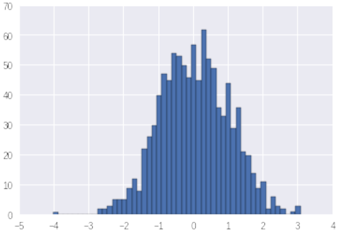
Detour

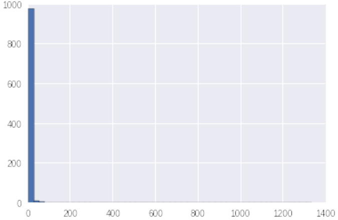
Project proposals...

Hubs are nodes with a surprisingly high degree. What does the surprisingly high degree mean?

Hubs and node degree distribution

random network vs. scale-free network



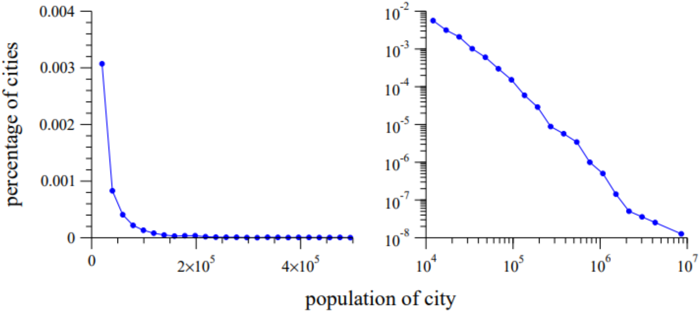


binomial distribution
Gaussian distribution
Poisson distribution
normal distribution

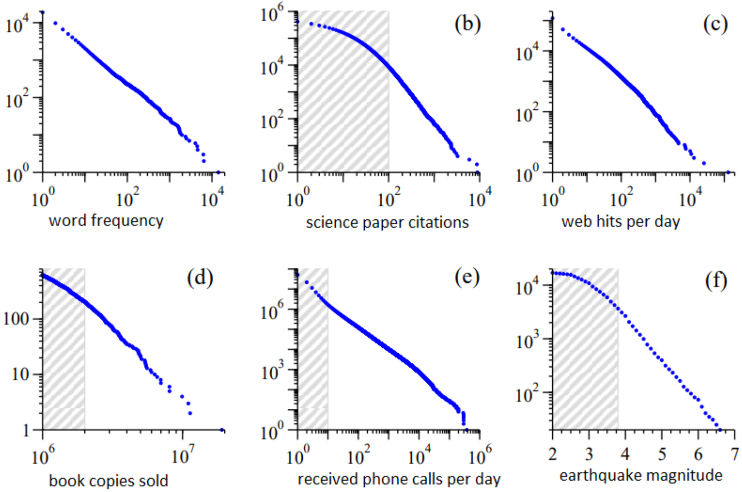
power-law distribution
Pareto distribution
heavy-tailed distribution
fat-tailed distribution

Logarithmic representation of distribution

On a normal histogram, we don't see much, but a log-log representation is much more interesting.

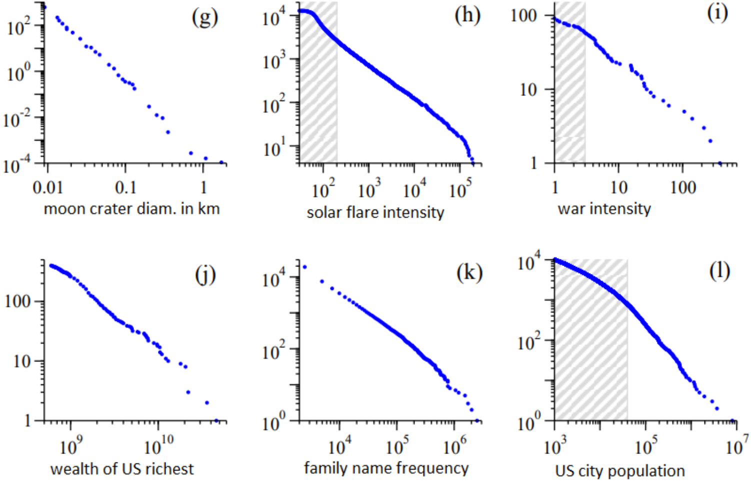


Power law distribution[5]

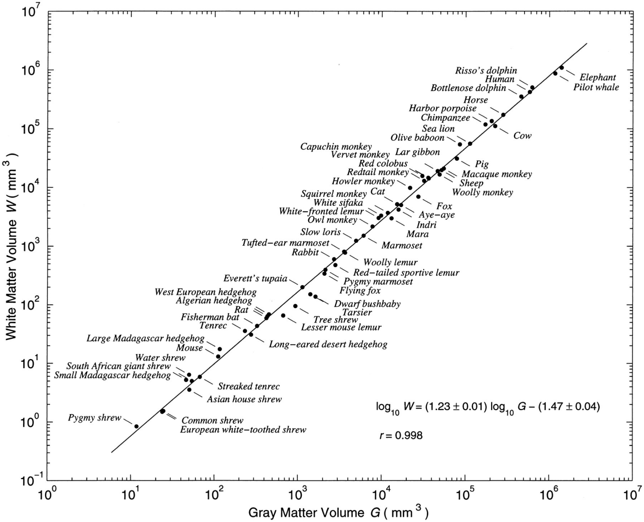


[5] Newman, M. E. (2005) DOI:10.1080/00107510500052444

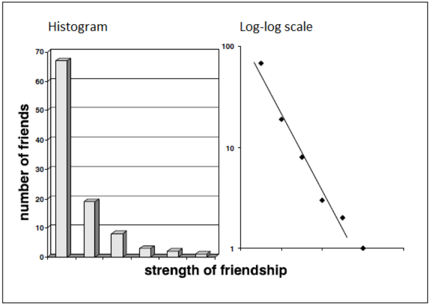
Power law distribution



Power law distribution



From graph to equation: Power law



$$\log(a^k) = k \cdot \log(a)$$
$$\log(ab) = \log(a) + \log(b)$$

The relationship can be described as $y = kx + q$:

$$\log(P) = \log(c) - \gamma \log(D)$$

After exponentiation :

$$P = cD^{-\gamma}$$

where

- P is the probability to meet a good friend or acquaintance
- c is constant
- γ is scaling exponent

Power law: variance

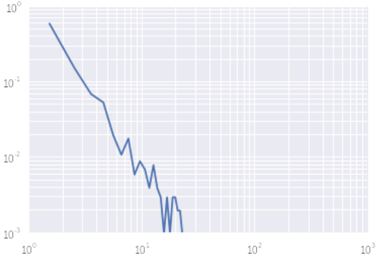
The second moment (variance) is generally infinite $\Longrightarrow$ as the sample (network) size increases, so will the maximum value.

	$\overline{k}$	σ
yeast protein-protein interaction network	2.9	4.88
E. Coli metabolic network	5.58	20.79
WWW network	4.60	30.27

Let's compute the scaling exponent: Log-log representation[6]

Problem:

- as the degree increases, fewer samples are available, hence noise increases.

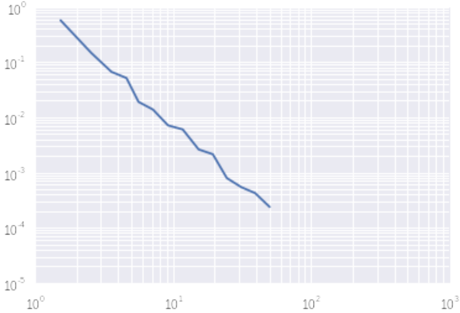


[6] Note: always use logarithm with base 10.

Let's compute the scaling exponent: Log-log representation

Solution:

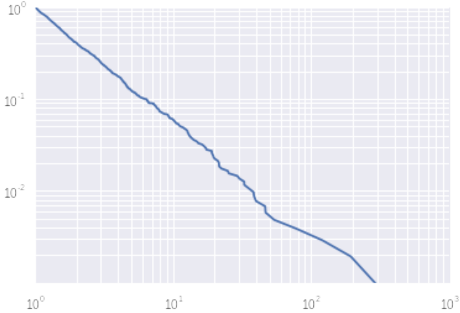
- logarithmic bins



Let’s compute the scaling exponent: Log-log representation

Solution:

- complementary cumulative distribution function
- $P_{\geq}(x) = x^{1-\gamma}$



Scale-free networks

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically.

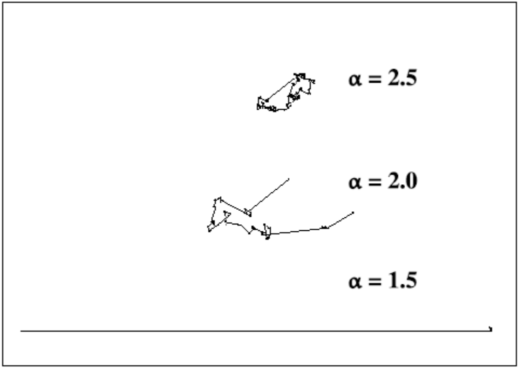
Many systems seem to be in the regime of $2 < \gamma < 3$. Is there any reason for that?

Example: Scale-free Search Strategy

- When social insects seek food, they frequently make cost-efficient small trips
- From time to time, they make longer jumps
- Rarely, they make very long journeys
- This search strategy is called Lévy flight and follows the power law $P = cL^{-\alpha}$, where L is length of a trip
- Why it is the best strategy?
  - It minimizes probability to return to the same site (disadvantage of random search)
  - It maximizes chance to end in new location (disadvantage of grid search)

  $\Longrightarrow$ best survival strategy

Lévy flight



Lévy flight search patterns for 1000 steps. Values for $\alpha = 2$ proven to be optimal[7].

[7] Csermely (2006). Weak links, pp. 29

Classes of scale-free networks

Anomalous regime $\gamma \leq 2$

- the degree of the largest hub grows faster than the size of the network
- such a network is not asymptotically possible without loops

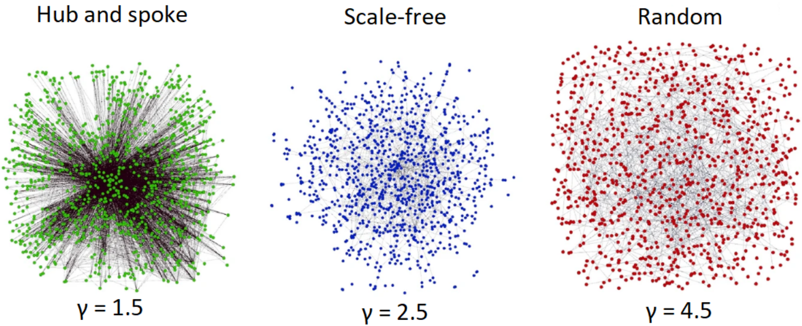
Scale-free regime $2 < \gamma < 3$

- the first moment of the distribution (mean) is finite, other moments diverge (variance, skewness, ...)
- specific behavior of dynamic processes (diffusion)
- robustness against random failure
- vulnerable to targeted attack (unlike a random network)

Random network regime $\gamma > 3$

- probability of large hubs decreases too fast
- hard to distinguish from a random network

Classes of scale-free networks



Empirical estimation of exponent

Motivation:

- important for null models, e.g. for the study of dynamic processes

Procedure:

- start with log-log transformation, fit a line
- use logarithmic intervals or cumulative distribution to remove noise
- apply the method of least squares, maximum likelihood, etc. in your favorite statistical software

Example in Excel

...

MUNI
FACULTY
OF INFORMATICS