Social Network Analysis

#4 Degree centrality

MIME.

What is centrality?

Centrality

From Wikipedia, the free encyclopedia

In graph theory and network analysis, indicators of **centrality** identify the most important vertices within a graph. Applications include identifying the most influential person(s) in a social network, key infrastructure nodes in the Internet or urban networks, and super-spreaders of disease. Centrality concepts were first developed in social network analysis, and many of the terms used to measure centrality reflect their sociological origin.[1]

Degree centrality [edit]

Main article: Degree (graph theory)

PageRank centrality [edit]

Main article: PageRank

Betweenness centrality [edit]

Main article: Betweenness centrality

Eigenvector centrality [edit]

Main article: Eigenvector centrality

Closeness centrality [edit]

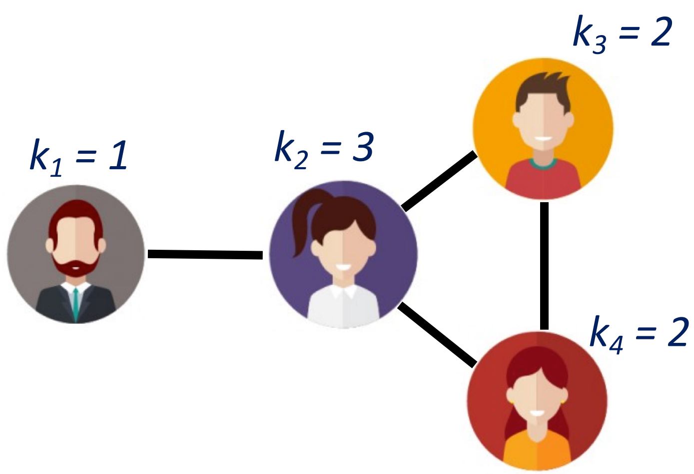
Main article: Closeness centrality

MiME.

Degree centrality

Degree (undirected)

- The **degree** k_i of node i in an **undirected** networks is
  the # of links i has to other nodes, or
  the # of nodes i is linked to

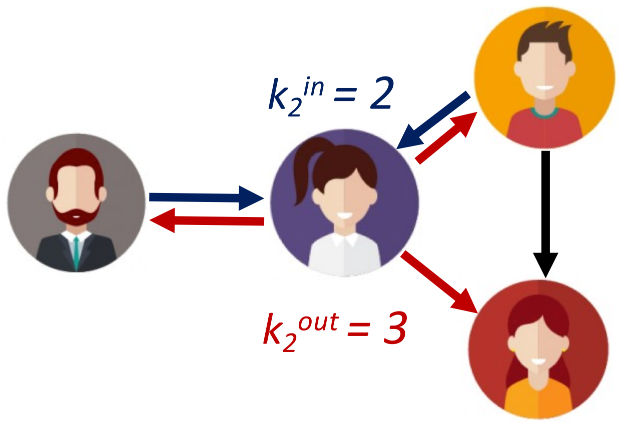


The **average** degree is

$$\langle k \rangle = \sum k_i / N = (1+3+2+2)/4$$
$$= 2$$

Degree (directed)

- For directed networks we distinguish between
  - in-degree k_i^{in} = # of entering links
  - out-degree k_i^{out} = # of exiting links



The average degree is

$$\langle k \rangle = \sum k_i^{out} / N = (1+3+2+0)/4$$
$$= \sum k_i^{in} / N = (1+2+1+2)/4$$
$$= 3/2$$

Meaning

- A social-capital measure of cohesion
- In-degree = importance as an Authority
- Out-degree = importance as a Hub

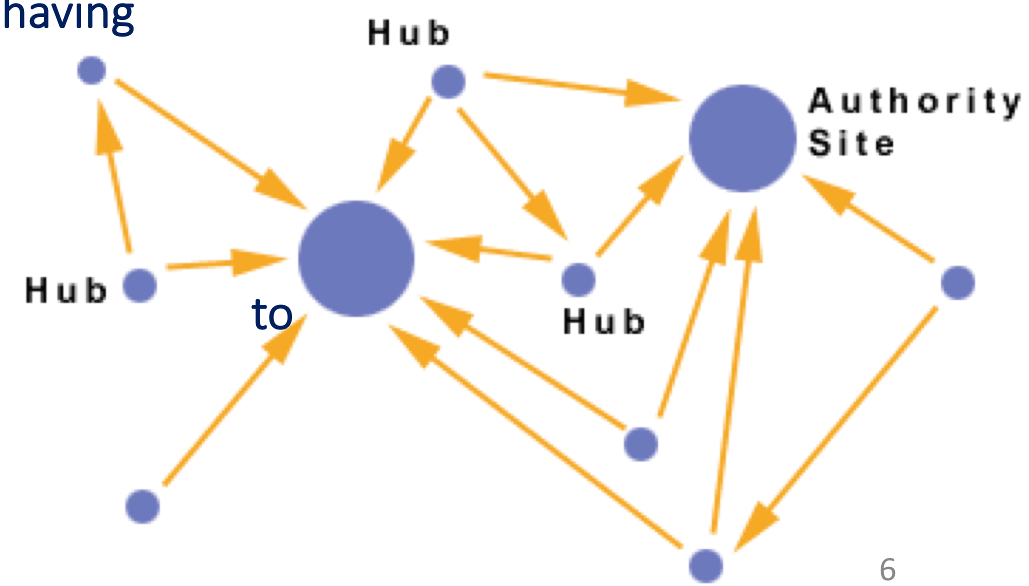
In www:

- Authorities (quality as a content provider)

nodes that contain useful information, or having a high number of edges pointing to them (e.g., course homepages)

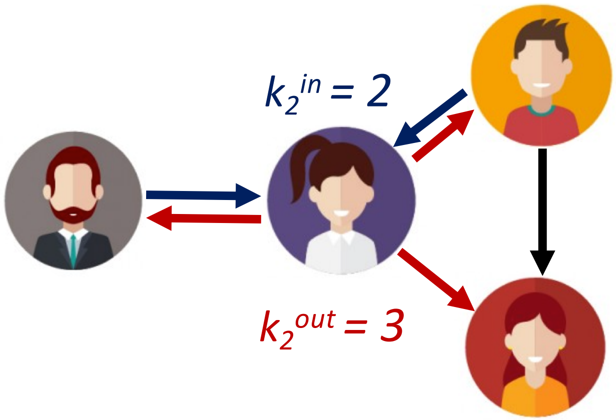
- Hubs (quality as an expert)

trustworthy nodes, or nodes that link many authorities (e.g., course bulletin)



Adjacency matrix & degree

- The in (out) degree can be obtained by summing the adjacency matrix over rows (columns)



no self-loops in this case!!!

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \begin{array}{l} \\ k_2^{in} = 2 \\ \\ \\ \end{array}$$

$k_2^{out} = 3$

- It works also with (positive) weights! ☺

Average degree

The # of nodes is *N*

The # of edges is *L*

Different measures for L !!!

Undirected

0	1	0	0	0	0	0	0
1	0	1	1	0	0	1	0
0	1	0	1	1	0	1	0
0	1	1	0	1	0	0	0
0	0	1	1	0	1	0	0
0	0	0	0	1	0	1	1
0	1	1	0	0	1	0	1
0	0	0	0	0	1	1	0

Directed

0	1	0	0	0	0	0	0
1	0	1	1	0	0	1	0
0	1	0	1	1	0	1	0
0	1	1	0	1	0	0	0
0	0	1	1	0	1	0	0
0	0	0	0	1	0	1	1
0	1	1	0	0	1	0	1
0	0	0	0	0	1	1	0

The average degree is

- Undirected $\langle k \rangle = 2L/N$
- Directed $\langle k \rangle = L/N$

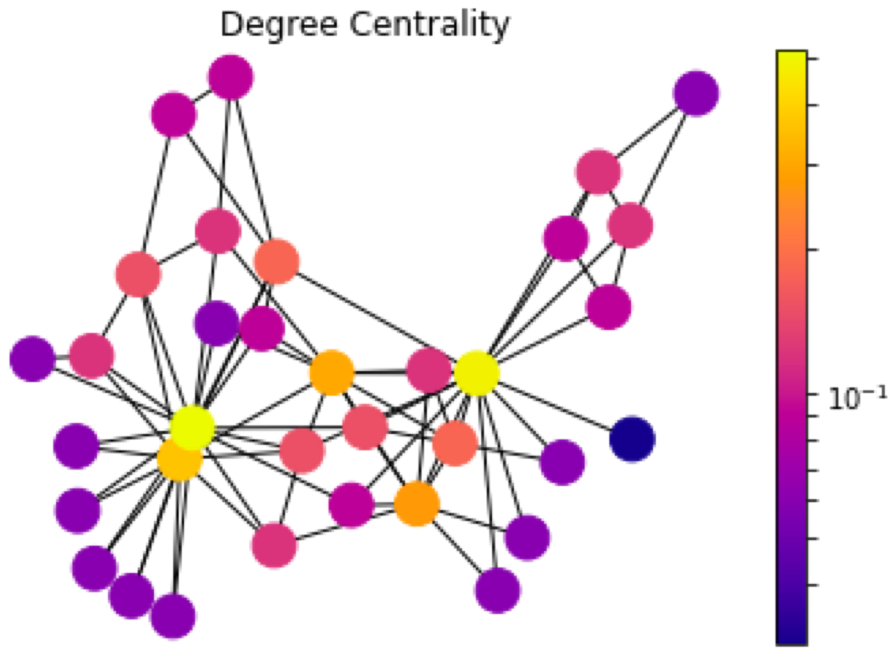
Real networks are sparse

- The maximum (average) degree is $N-1$
- In real networks $<k> \ll N-1$

NETWORK	N	L	$\langle k \rangle$
Internet	192,244	609,066	6.34
WWW	325,729	1,497,134	4.60
Power Grid	4,941	6,594	2.67
Mobile Phone Calls	36,595	91,826	2.51
Email	57,194	103,731	1.81
Science Collaboration	23,133	93,439	8.08
Actor Network	702,388	29,397,908	83.71
Citation Network	449,673	4,689,479	10.43
E. Coli Metabolism	1,039	5,802	5.58
Protein Interactions	2,018	2,930	2.90

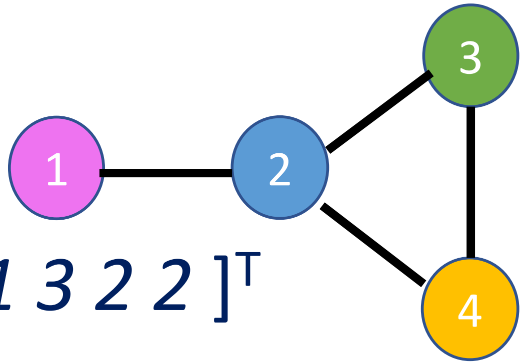
Illustrate degree centrality

Graphical representations

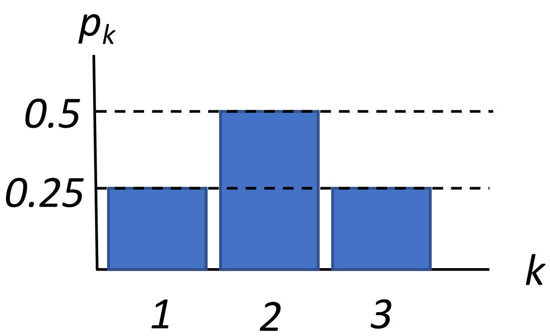


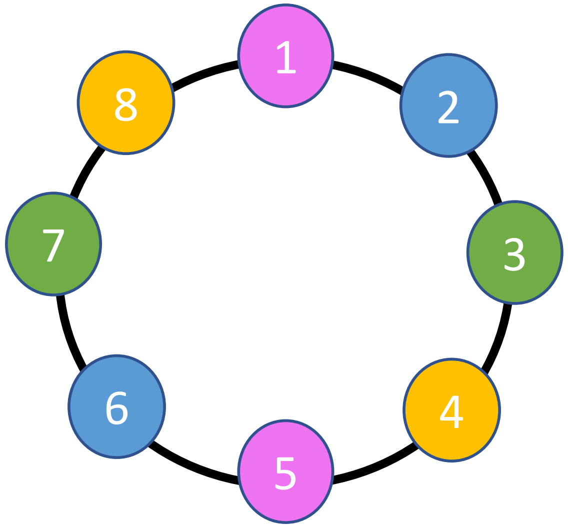
Degree distribution p_k

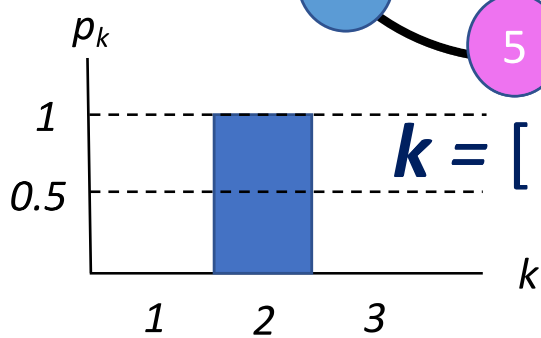
- ✓ a <u>probability</u> distribution where p_k is the fraction of nodes that have degree exactly equal to k
- ✓ p_k = # of nodes with degree k, divided by N



$\mathbf{k} = [\ 1\ 3\ 2\ 2\ ]^T$







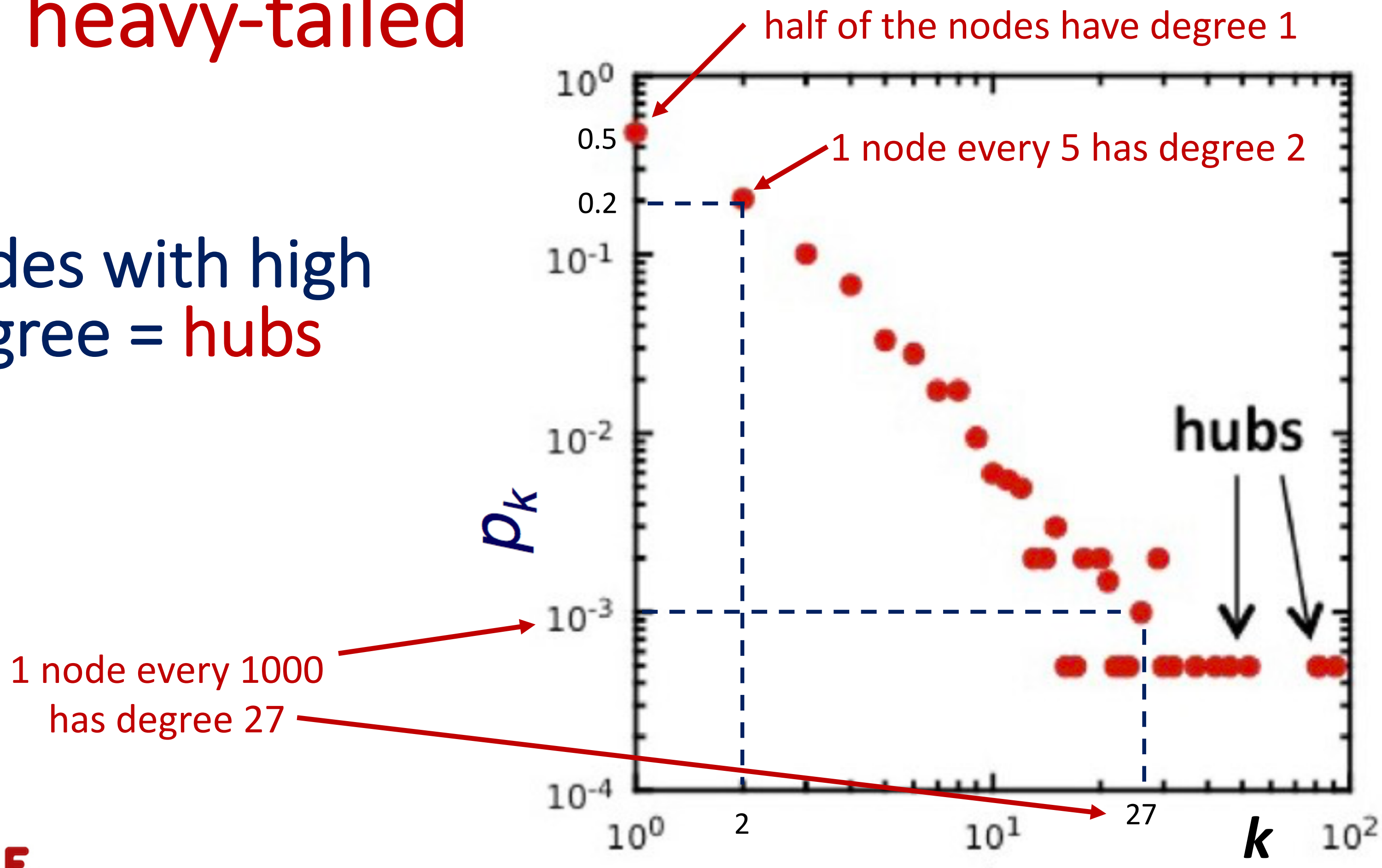
$\mathbf{k} = [\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ ]^T$

MIME.

Log-log plot

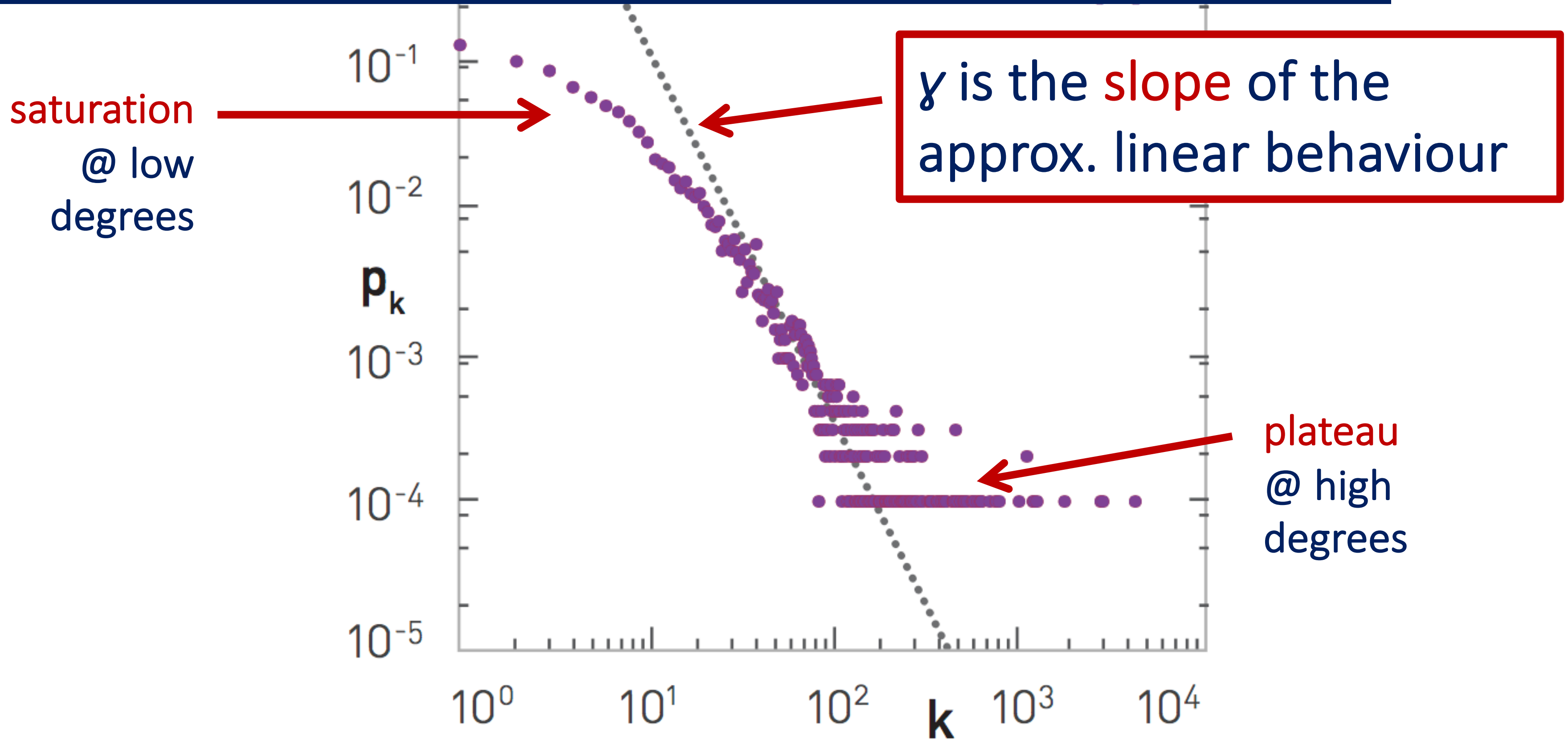
- In real world (large) networks, degree distribution is typically heavy-tailed

nodes with high degree = hubs



Scale-free networks

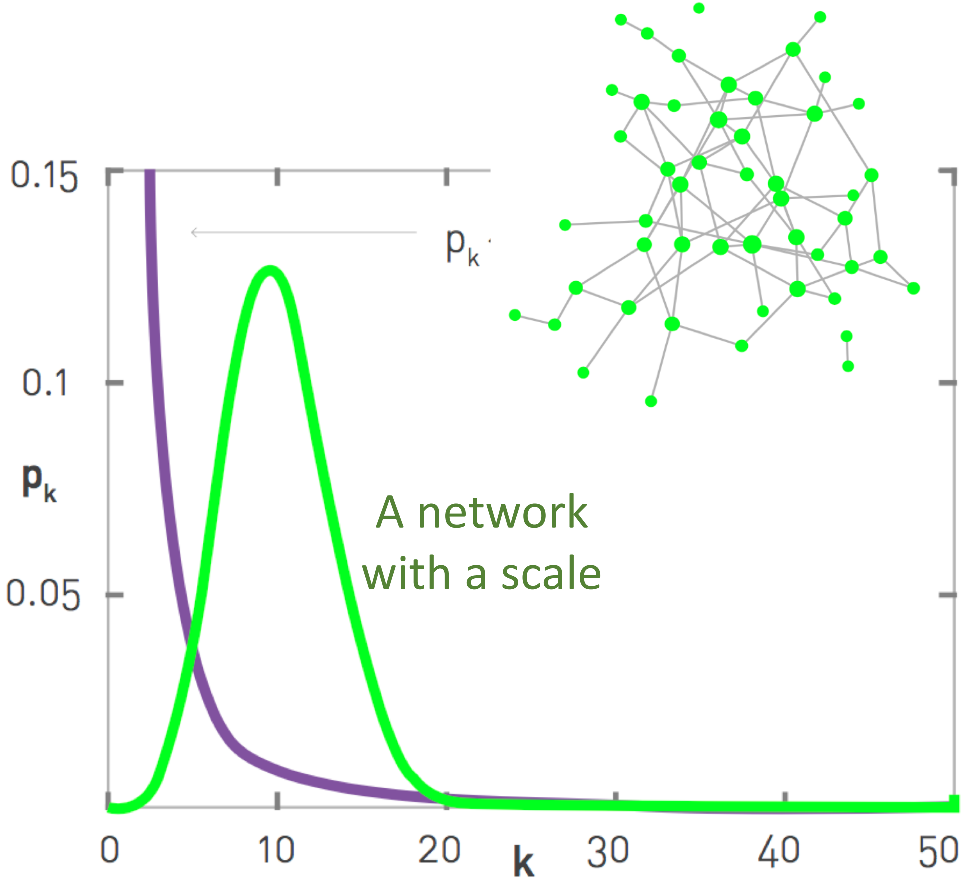
The power-law

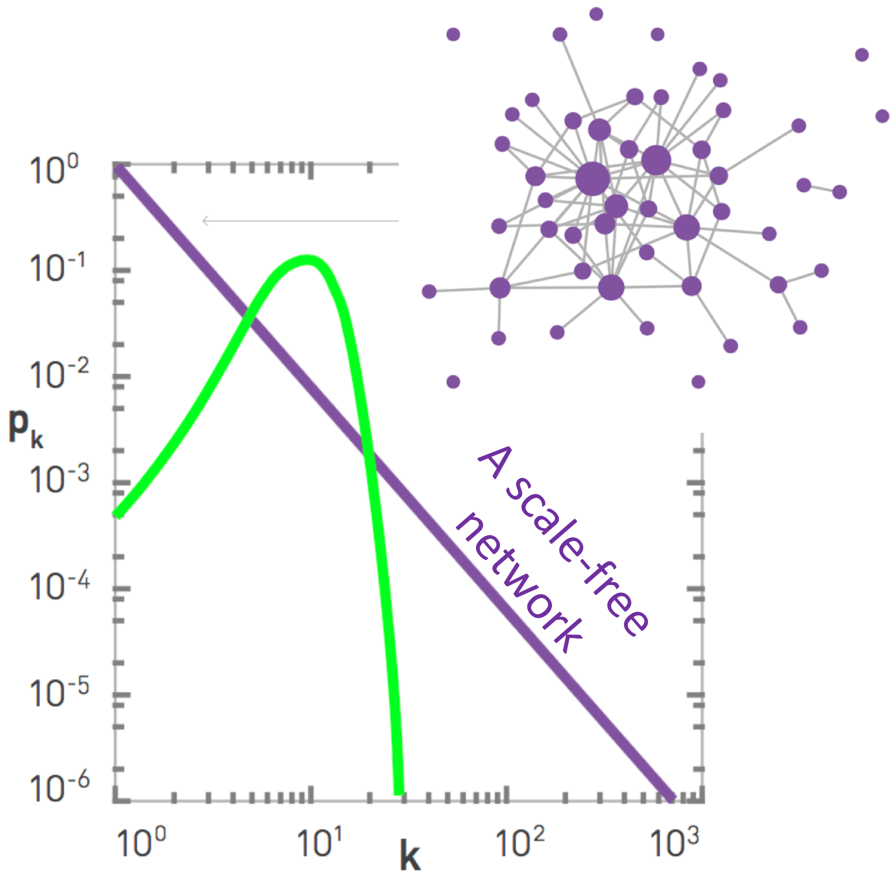


The degree distribution follows a power-law if

$\boxed{p_k = C\,k^{-\gamma}}$ → $\ln(p_k) = c - \gamma \cdot \ln(k)$

Scale-free networks



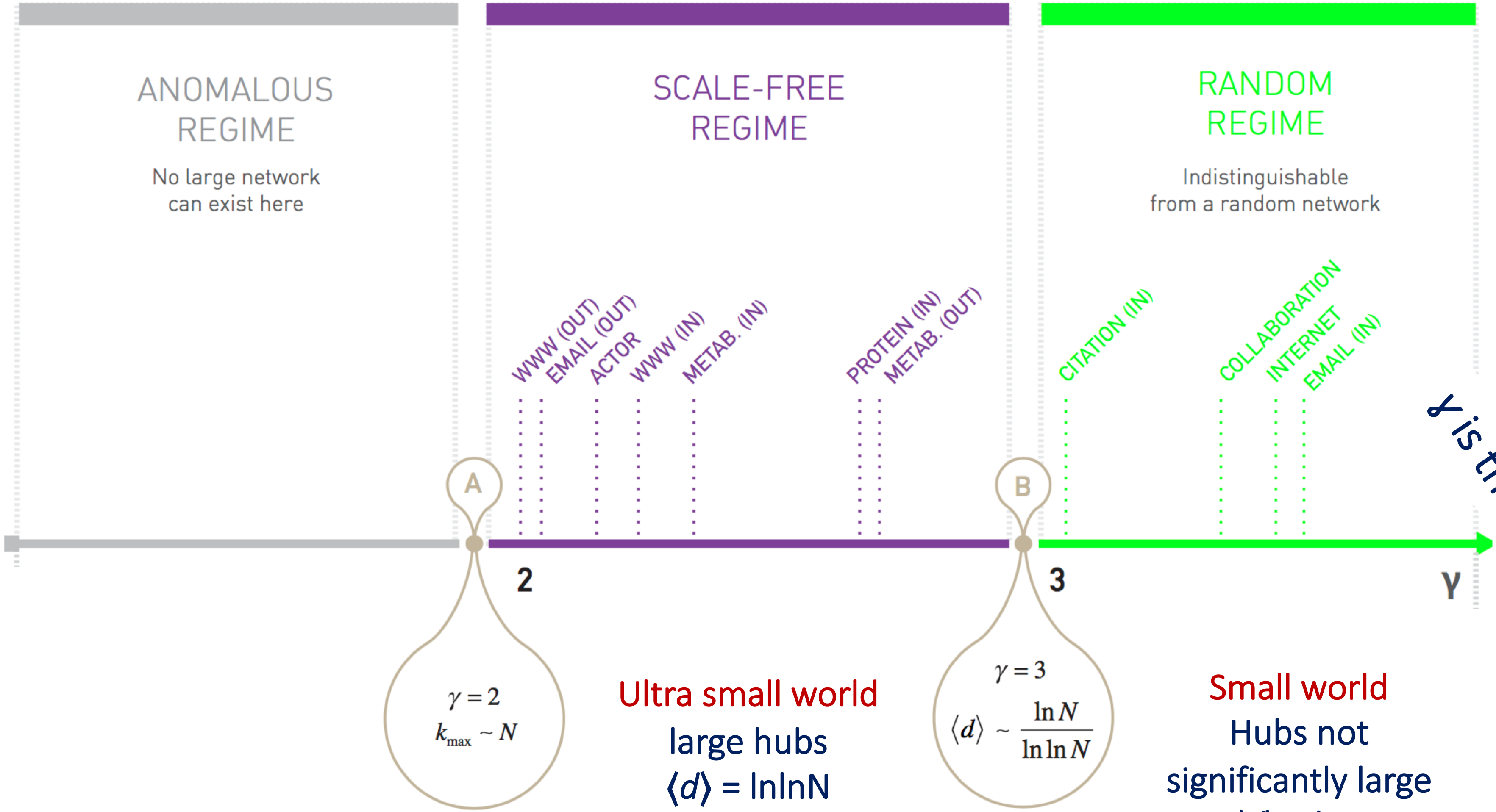


- ➢ Randomly wired network
- ➢ Has smaller hubs
- ➢ Needs a linear plot

- ➢ Power-law network
- ➢ Has big hubs
- ➢ Needs a log-log plot

The scale-free regime

A.L. Barabási, Network science, http://barabasi.com/networksciencebook



γ is the slope

Ultra small world
large hubs
$\langle d \rangle = \ln\ln N$

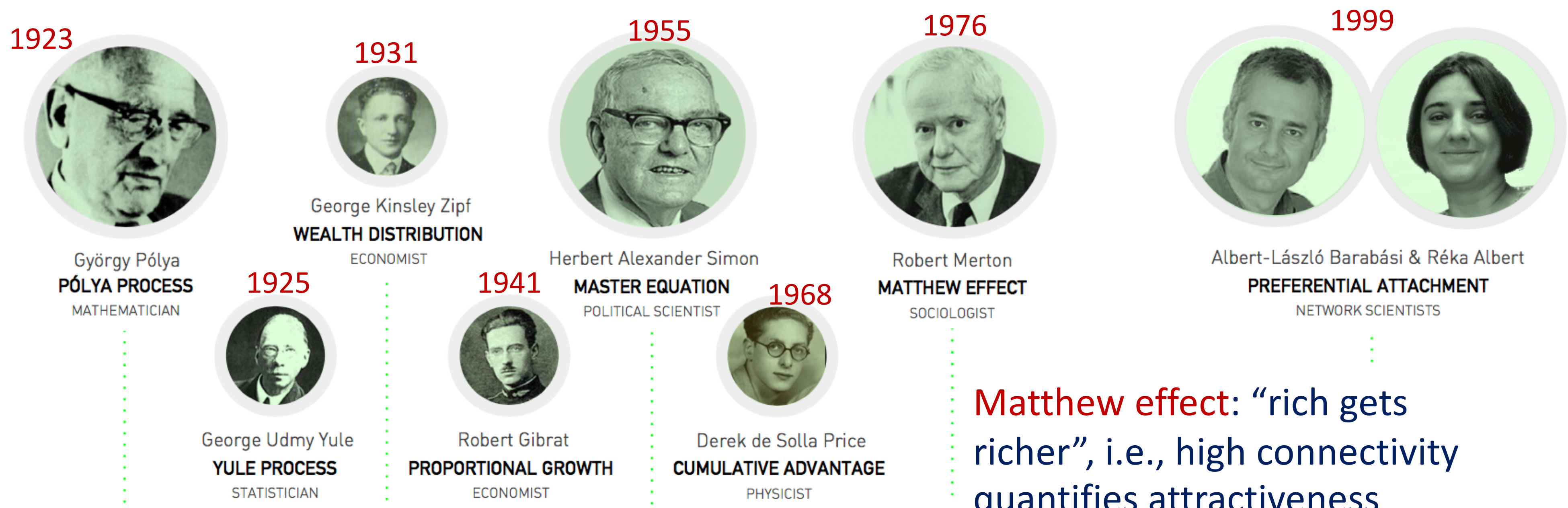
Small world
Hubs not significantly large
$\langle d \rangle = \ln N$

e.g. in www $N=7 \cdot 10^9$, $\ln(N)=22.7$, $\ln(\ln(N))=3.12$ (very small)

Why a power law?

#1. Preferential attachment

- (new) nodes tend to link to the more connected nodes (e.g., think of www)
- This idea has a long history



Matthew effect: “rich gets richer”, i.e., high connectivity quantifies attractiveness

MIME

Explaining preferential attachment

- Citation network

  researchers decide what papers to read and cite by "copying" references from papers they have read → papers with more citations are more likely to be cited

- Social network

  the more acquaintances an individual has, the higher the chancer of getting new friends, i.e., we "copy" the friends of friends → difficult to get friends if you have none

This is called the copying model

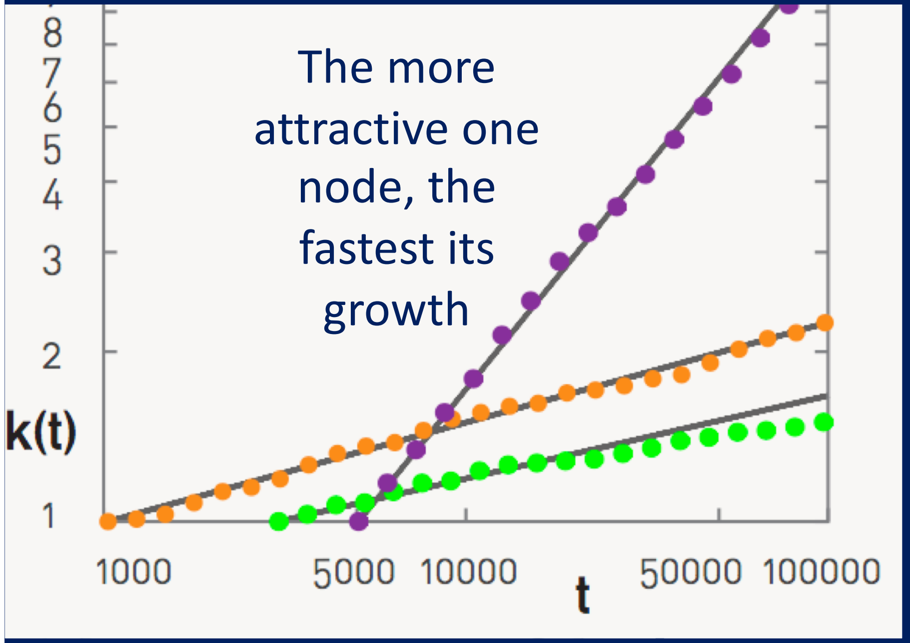
Why a power law?

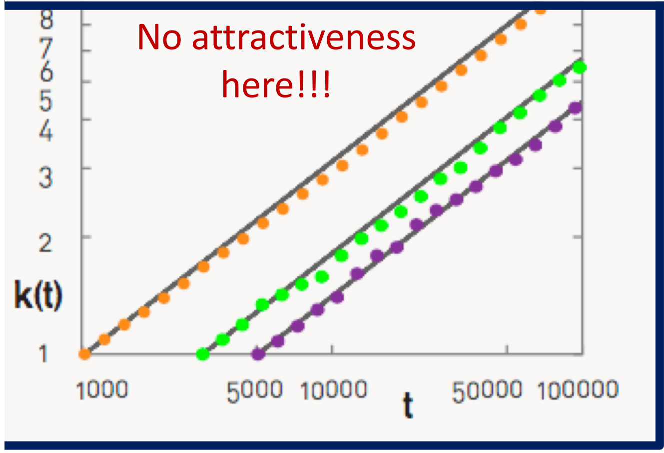
#2. Attractiveness

- There is an innate ability of a node to attract links (just a quality assessment of the individual)
- Otherwise oldest nodes would have an inherent advantage and cannot be defeated (*first mover's advantage*), which is in contrast with intuition and evidence

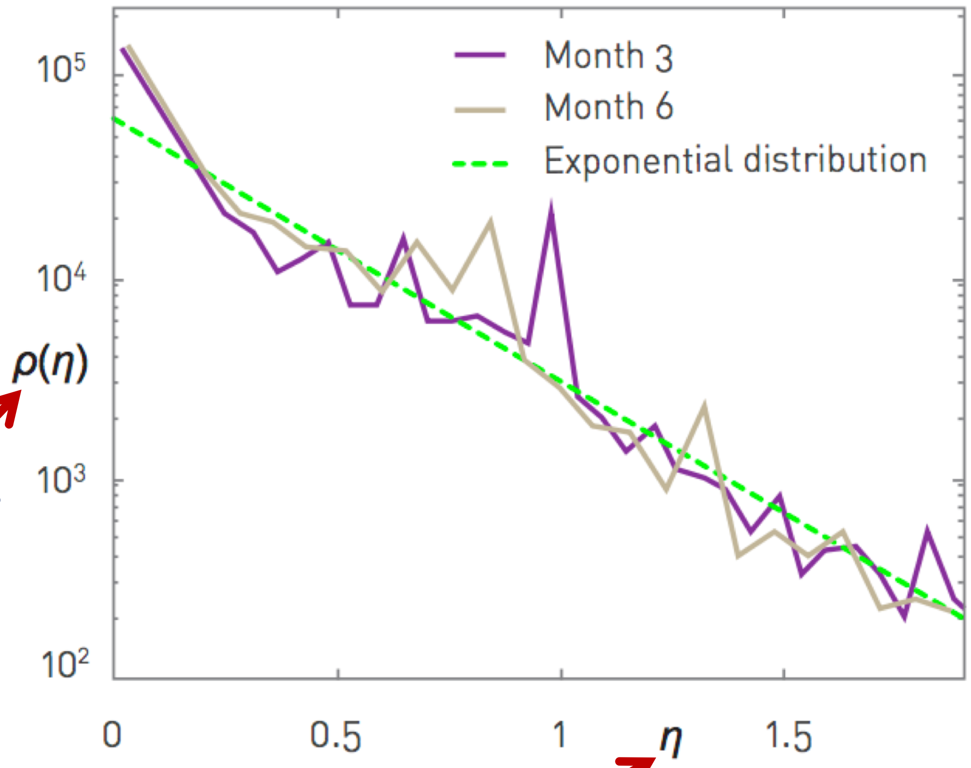
e.g., Altavista [1990] → Google [2000] → Facebook [2011] → Instagram [202?]

Measuring attractiveness





Attractiveness distribution function (probability)



Attractiveness

MIME

Bianconi-Barabási (mathematical) model

The model:

- **Growth** – at time step N a new node is added with **attractiveness** η
- Attractiveness is a random number drawn from a given **distribution** $\rho(\eta)$
- **Preferential attachment** - probability of linking to node i is proportional to both the degree and attractiveness, i.e., $p_i = C\, k_i \eta_\mathrm{i}$

Today take-aways

- Degree, degree distribution, loglog plot
- Authorities and hubs
- Power law, scale-free networks
- Slope, Ultra-small-world regime
- Preferential attachment, attractiveness