Network theory: 2

Liubov Tupikina, Marc Santolini (CRI)

Today

1. Network science: new concepts
2. Notebook Hands-on: Python and Gephi

 Break

3. Questions & Answers
4. Projects discussions and Reversed classrooms

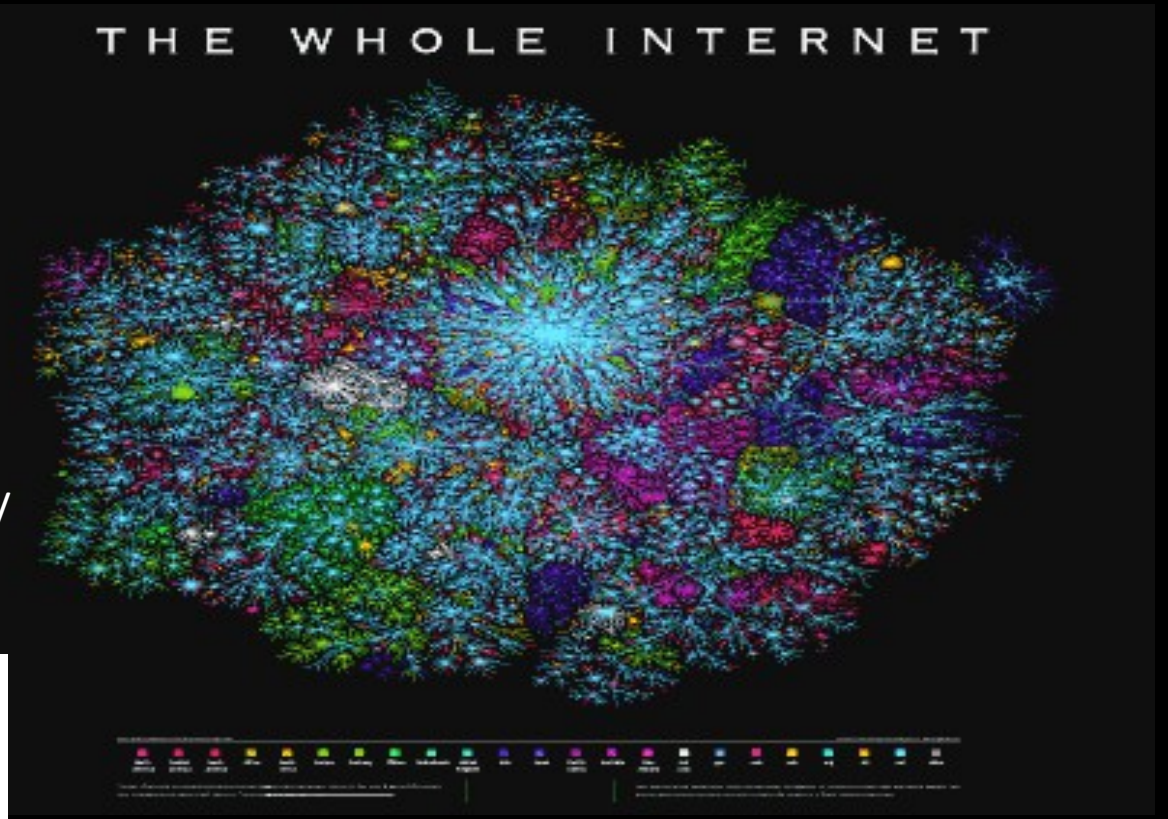
If you have questions?

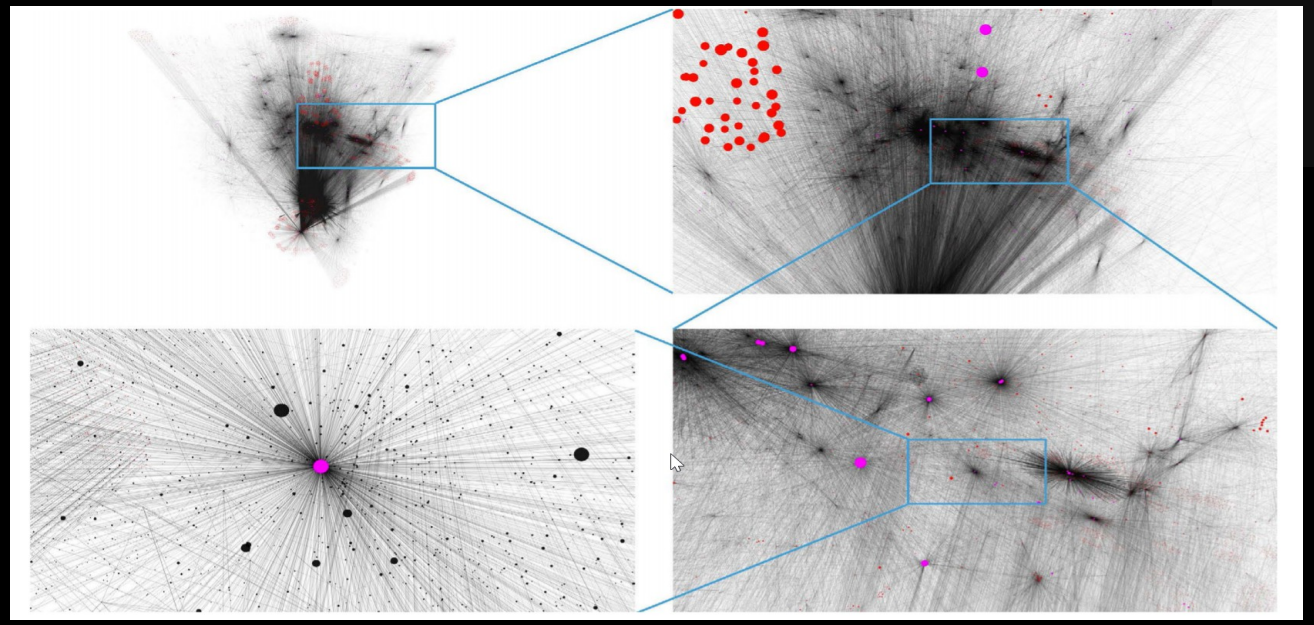
http://networksciencebook.com/

https://www.barabasilab.com

http://networkrepository.com/graph-vis.php

http://www.complexity-explorables.org/explorables/neighbors/

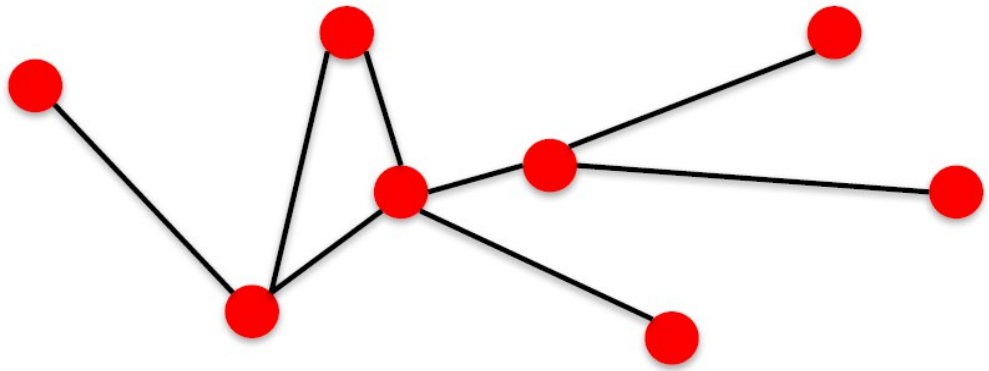




https://snap.stanford.edu/data

https://sites.google.com/a/binghamton.edu/netscied/teaching-learning/network-concepts

What is the network itself?

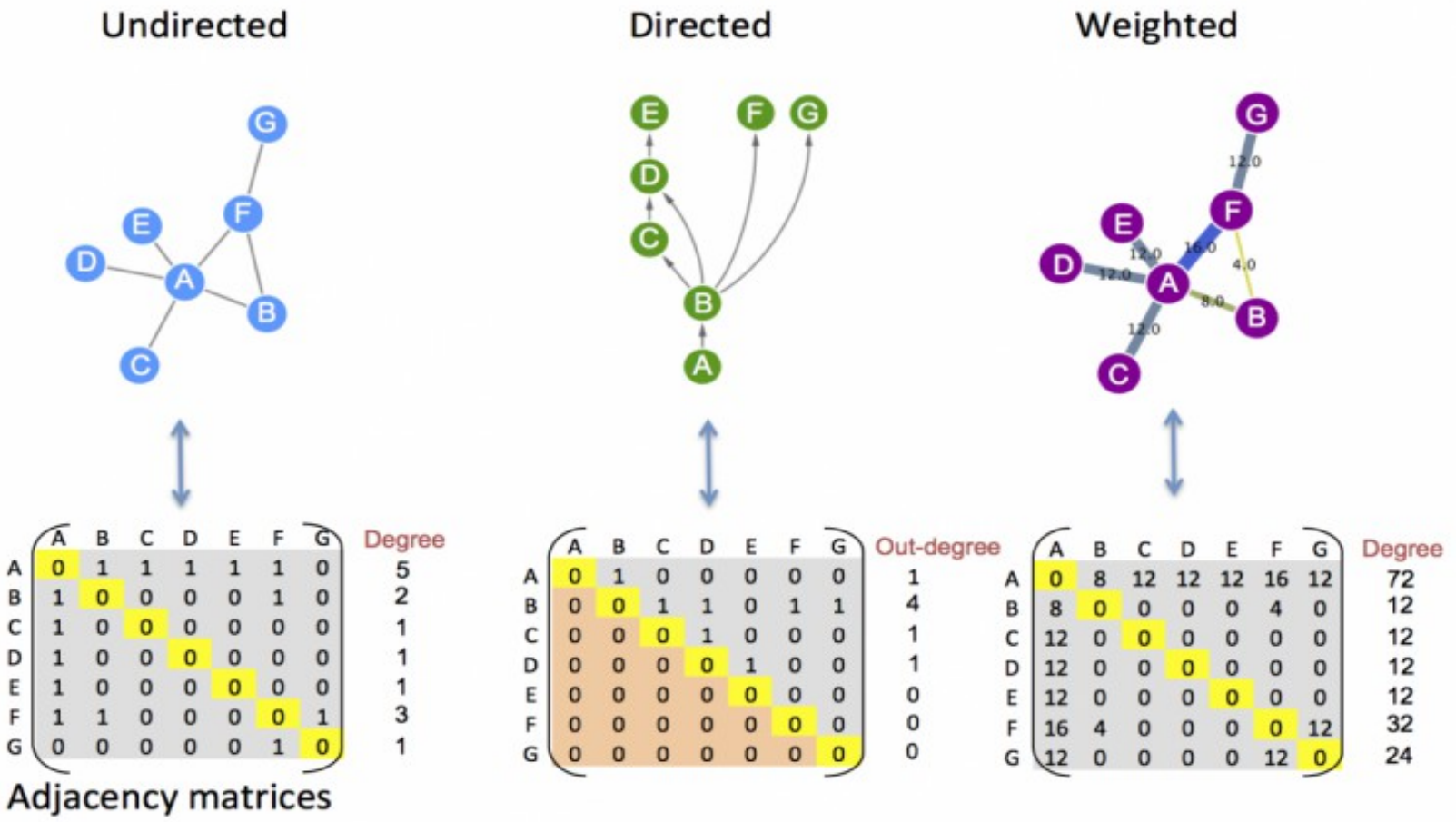


- **components**: nodes, vertices — N
- **interactions**: links, edges — L
- **system**: network, graph — (N,L)

If you want to know more ...

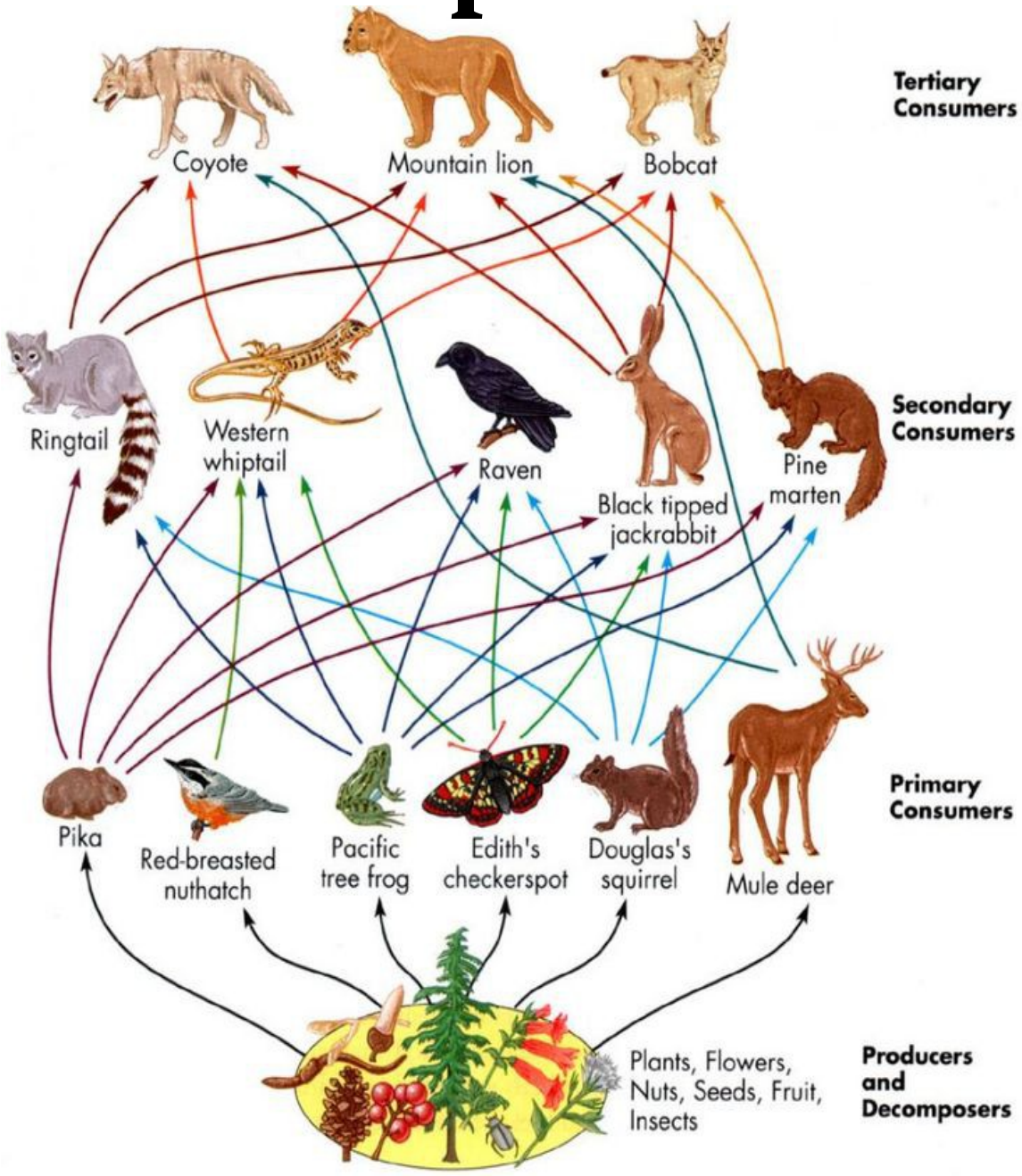
Barabasi book on network science

How to describe a network?

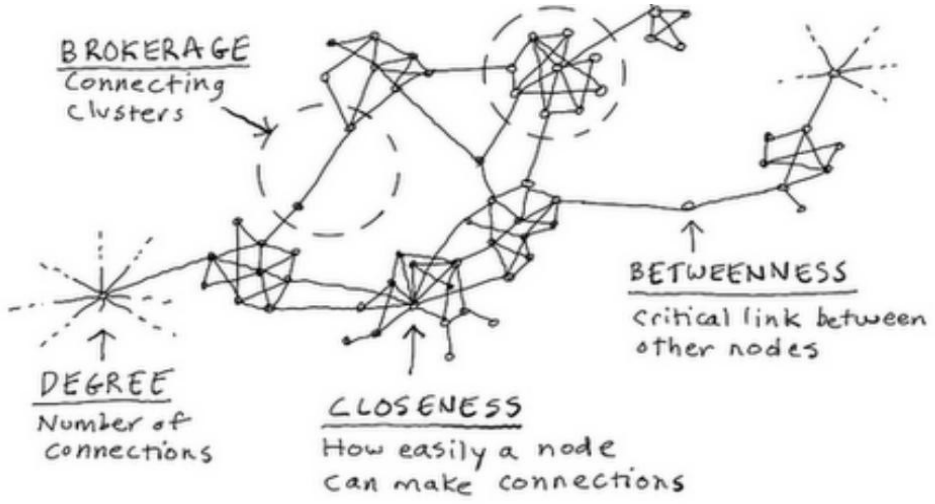


Barabasi book on network science

How to describe importance of nodes?

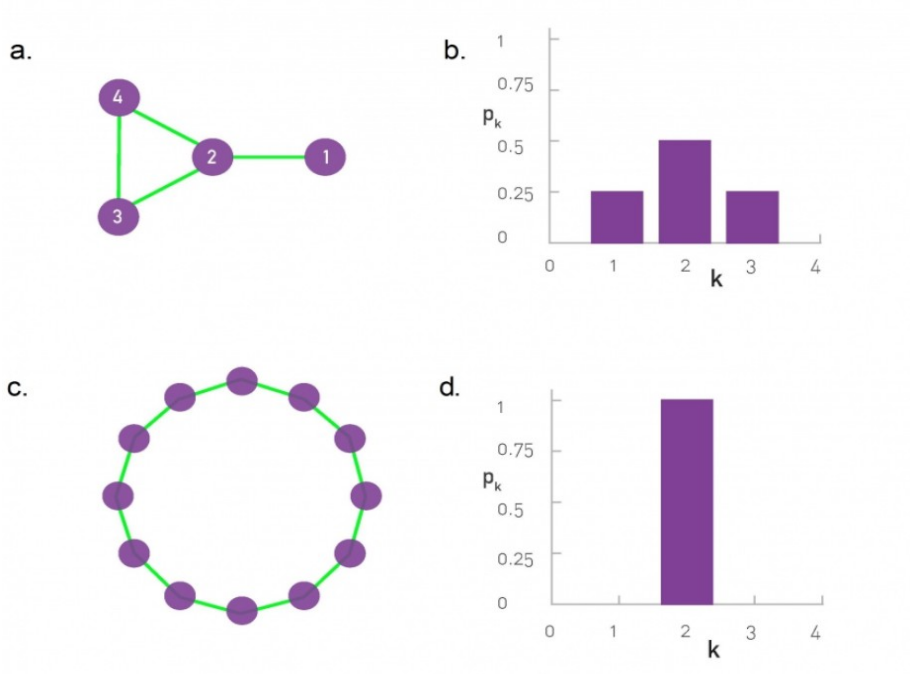


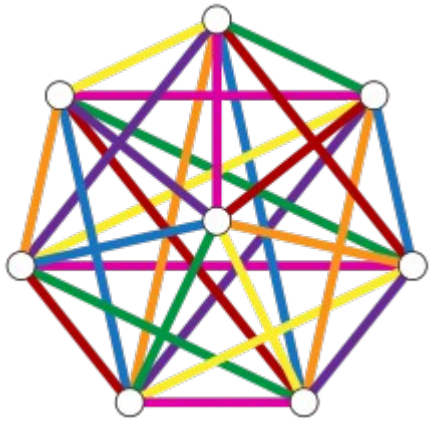
Network measures:
How to describe a network?

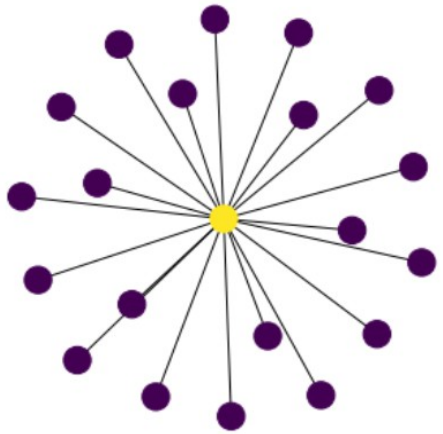


How to describe a network?

Degree of node A is the number of nodes adjacent with the node A.
Plotting degree distribution for a network can give important insight on its properties.





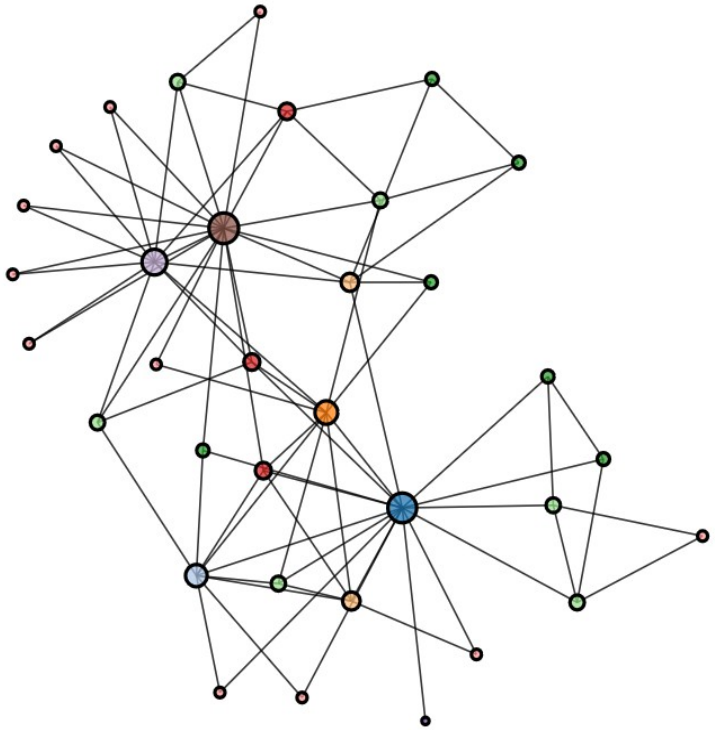


?

Barabasi book on network science

How to describe a real network?

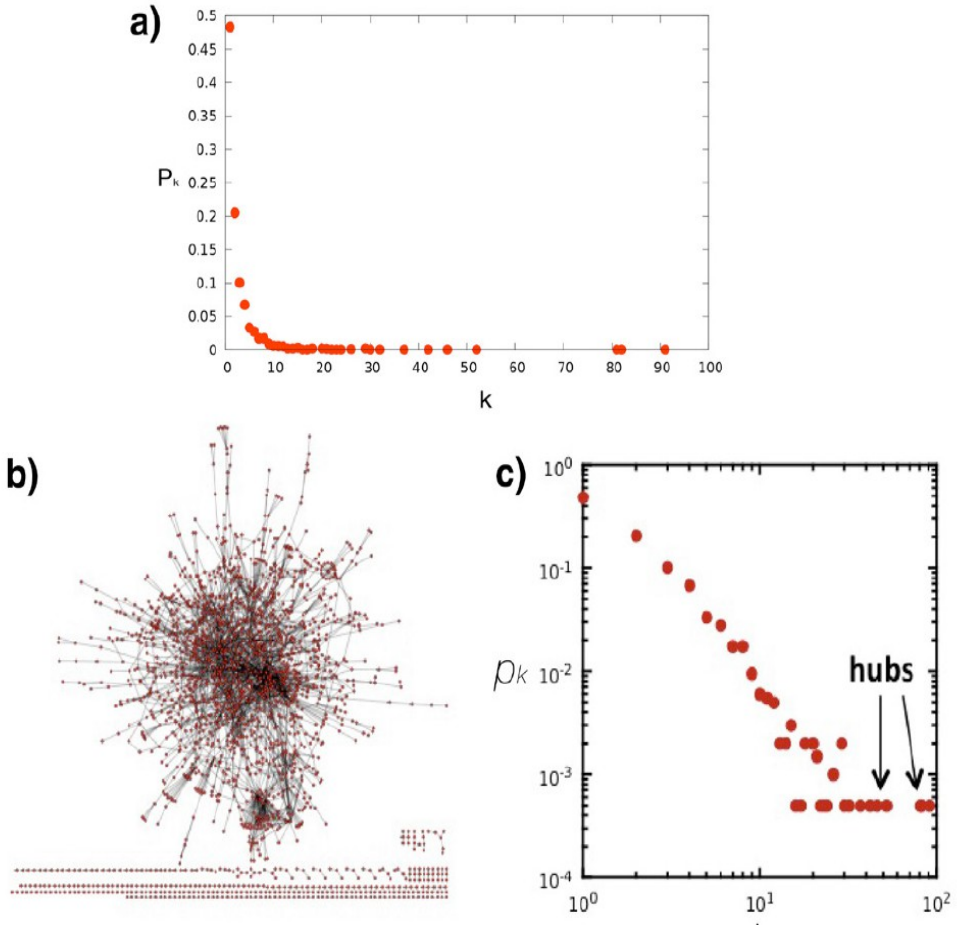
What is degree distribution of larger network? (Liuba -hands on session)
How to visualise a network in a nice way? (Marc - visualisation)

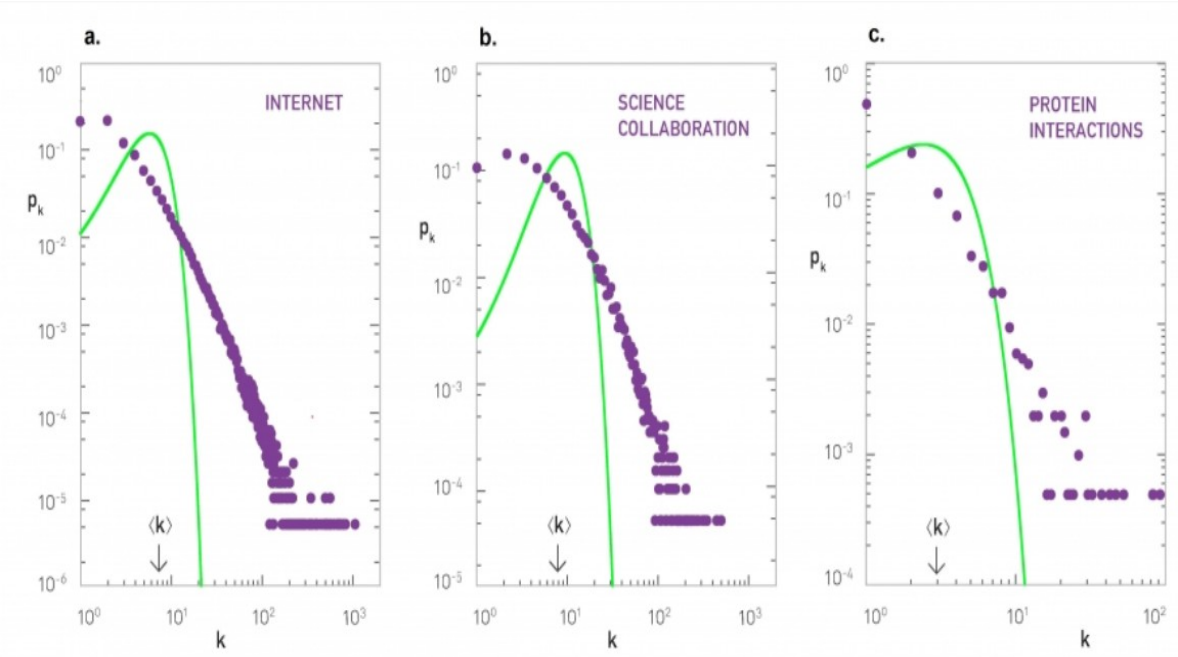


http://networkrepository.com/soc-karate.php

Network Data Statistics	
Nodes	34
Edges	78
Density	0.139037
Maximum degree	17
Minimum degree	1
Average degree	4
Assortativity	-0.475613
Number of triangles	135
Average number of triangles	3
Maximum number of triangles	18
Average clustering coefficient	0.570638
Fraction of closed triangles	0.255682
Maximum k-core	5
Lower bound of Maximum Clique	5

Scale-free networks (lecture 1)

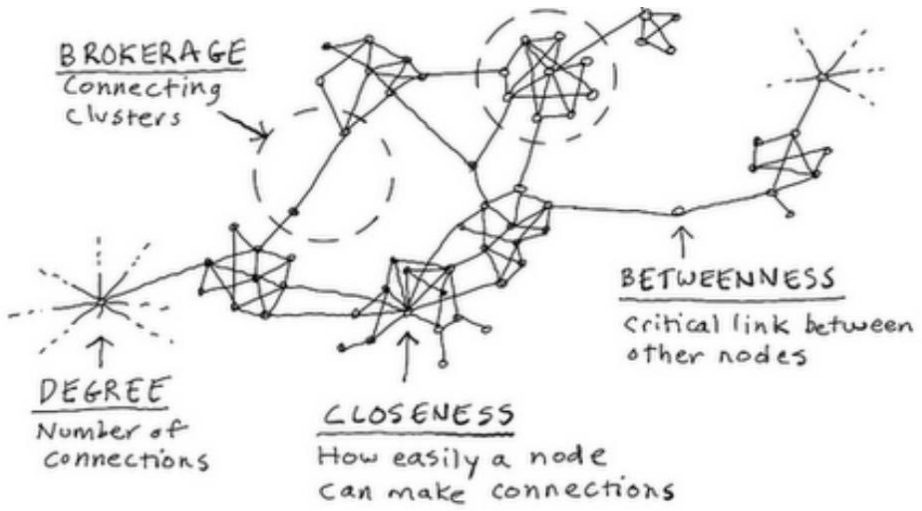




"Scale free networks can be found in many places"... at the same time
"Be careful with power-laws"

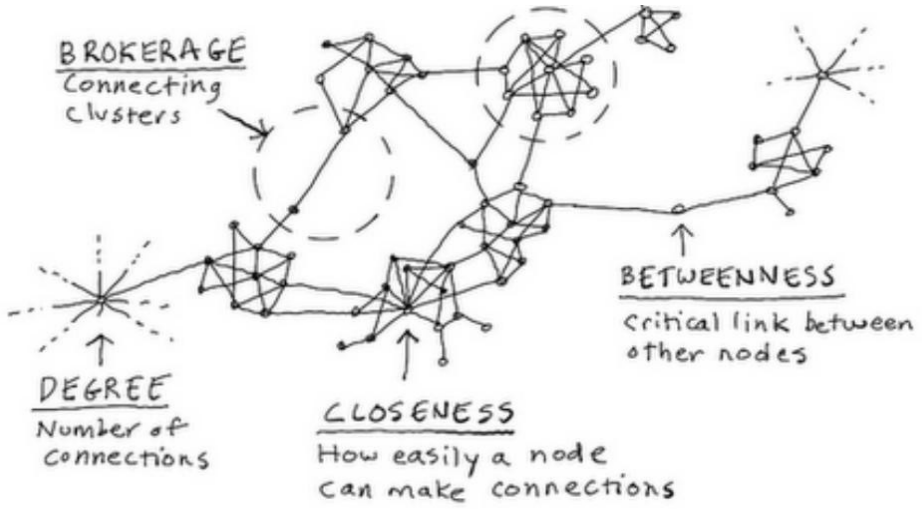
Cytoscape software alternative to Gephi for biological networks
Barabasi book

Network measures: clustering



$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk}}{k_i(k_i - 1)}$$

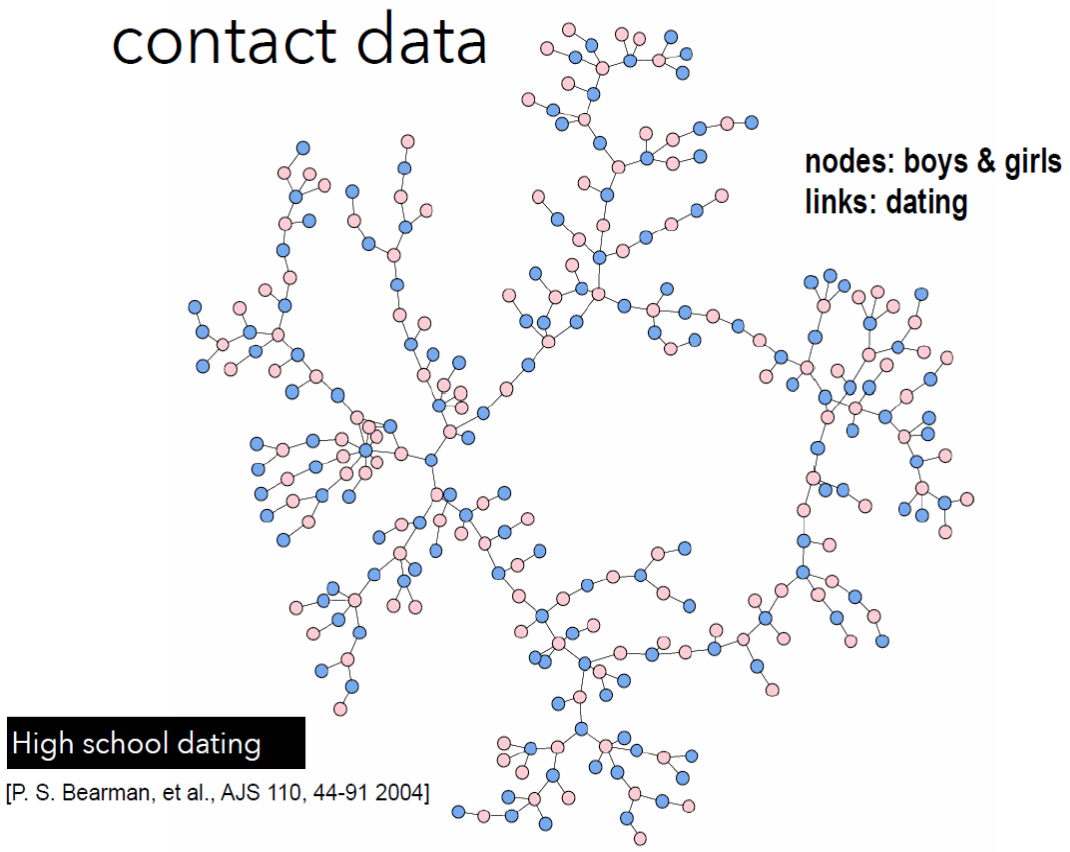
Network measures: betweenness centrality



$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Additional notions:
Paths in a network
Shortest paths
Random walks on a network

For your projects:
How to analyze networks with network measures?



Network	Nodes	Links	Directed / Undirected	N	L	‹K›
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,133	93,437	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43

Hands-on part

Python
Gephi

Q&A Group discussions
Projects and reversed classrooms

Hands-on part

Idea:
Each group gets one paper
Each paper → each topic

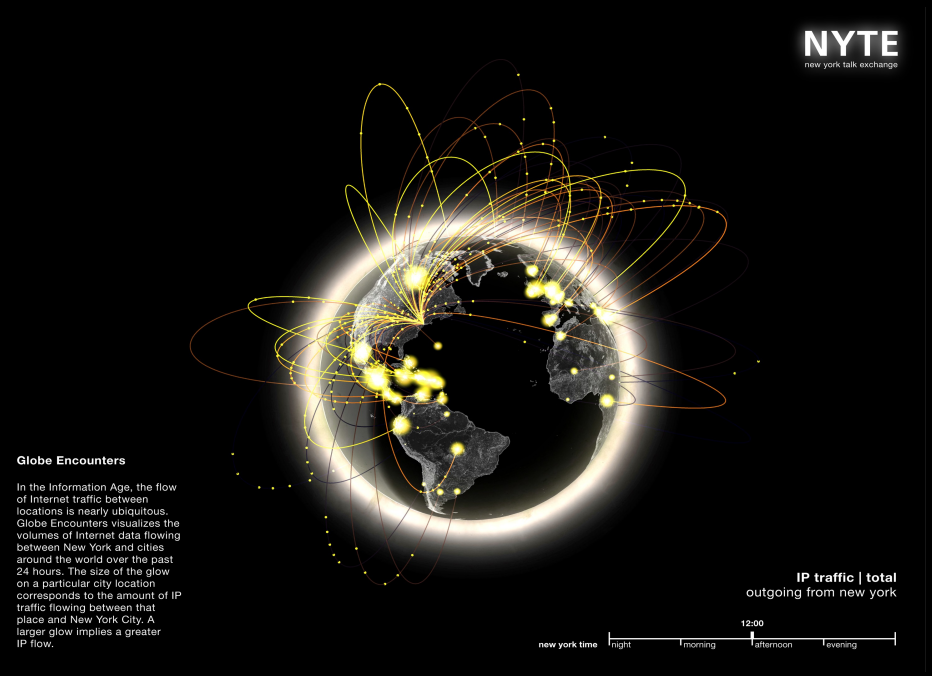
Internet and WWW:
Data collected in 1999. Ref: Albert, R., Jeong, H., & Barabasi, A. L. (1999). Internet: Diameter of the world-wide web. Nature, 401(6749), 130-131

Citation network:
Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007)

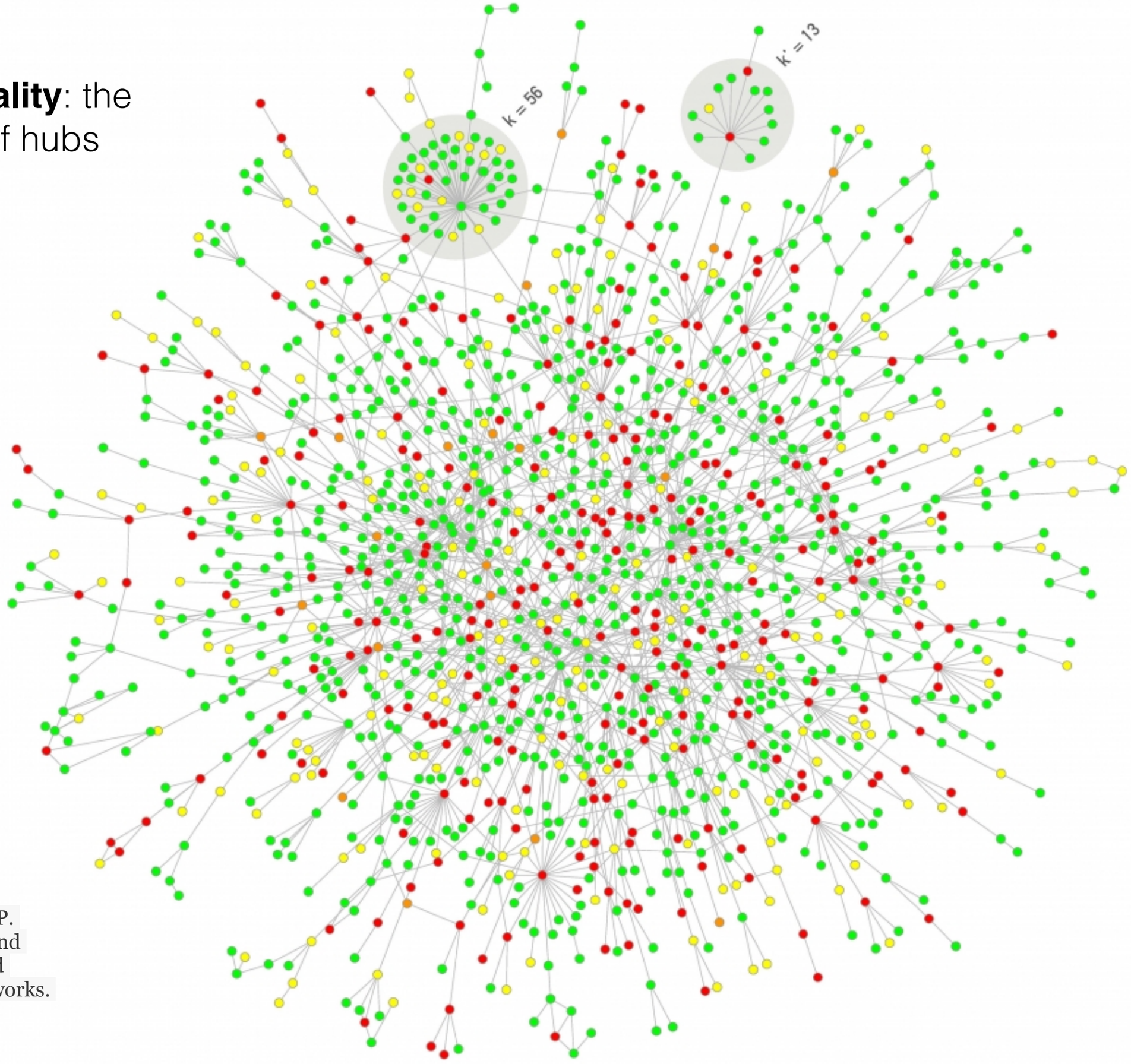
Neural networks:
Classes of small-world networks L. A. N. Amaral (2000)

Ecological network
Ecological networks Montoya Sole (2000)

Human networks:
Beamer et al.



centrality-lethality: the importance of hubs



H. Jeong, B. Tombor, S. P. Mason, A.-L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks. Nature 411: 41-42, 2001.

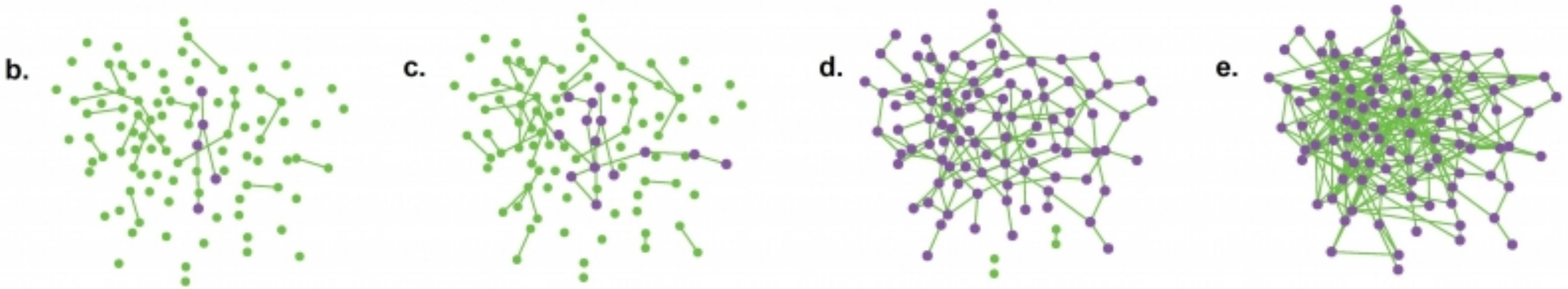
many metrics…

Table 2: Definitions of network science terms and variables.

Term/variable	Definition
N	number of nodes, N, in graph
E	number of edges, E, in graph
network density	ratio of the number of edges to the maximum number of possible edges $\frac{2E}{N(N-1)}$
distance, $d(n_i, n_j)$	shortest path between node i and node j $d(n_i, n_j)$ where $n_i, n_j \in N$
average shortest path length, L	average length of shortest path between pairs of nodes $L = \frac{1}{N(N-1)} \bullet \sum_{i \neq j} d\left(n_i, n_j\right)$
diameter, D	largest shortest path between nodes $D = \max_{n_i \in N, n_j \in N} d\left(n_i, n_j\right)$
closeness centrality	inverse of the sum of the length of the shortest paths between node i and all other nodes in the graph $C_i = \frac{1}{\sum_j d\left(n_i, n_j\right)}$
degree, k_i	number of edges attached to node i
average degree, $\langle k \rangle$	average number of edges per node in network $\langle k \rangle = \frac{1}{N} \sum_{n=i}^{N} k_i$
local clustering coefficient, c_i	number of edges between the neighbors of node i divided by the maximum number of edges between those neighbors $c_i = \frac{2\lvert e_{jk} \rvert}{k_i(k_i - 1)}$ where $n_j, n_k \in N_i,\ e_{jk} \in E$
average clustering coefficient, $\langle C \rangle$	average clustering coefficient of nodes in the network $\langle C \rangle = \frac{1}{N} \sum_{n=i}^{N} c_i$
modularity, Q	proportion of edges that fall within subgroups of nodes minus the expected proportion if edges were randomly distributed, range $[-1, 1]$
average efficiency, E_G	measure of how efficiently information is exchanged in the network $E_G = \frac{1}{n(n-1)} \sum_{i \neq j \in N} \frac{1}{d\left(n_i, n_j\right)}$
largest connected component	largest group of nodes in the network that are connected to each other in a single component
degree distribution, $P(k)$	probability distribution of node degrees in the network
γ	power-law exponent for the degree distribution
Small world structure	network with short average path lengths and relatively high clustering coefficient (relative to a random graph with similar density)
scale-free network	network with a degree distribution that is power-law distributed

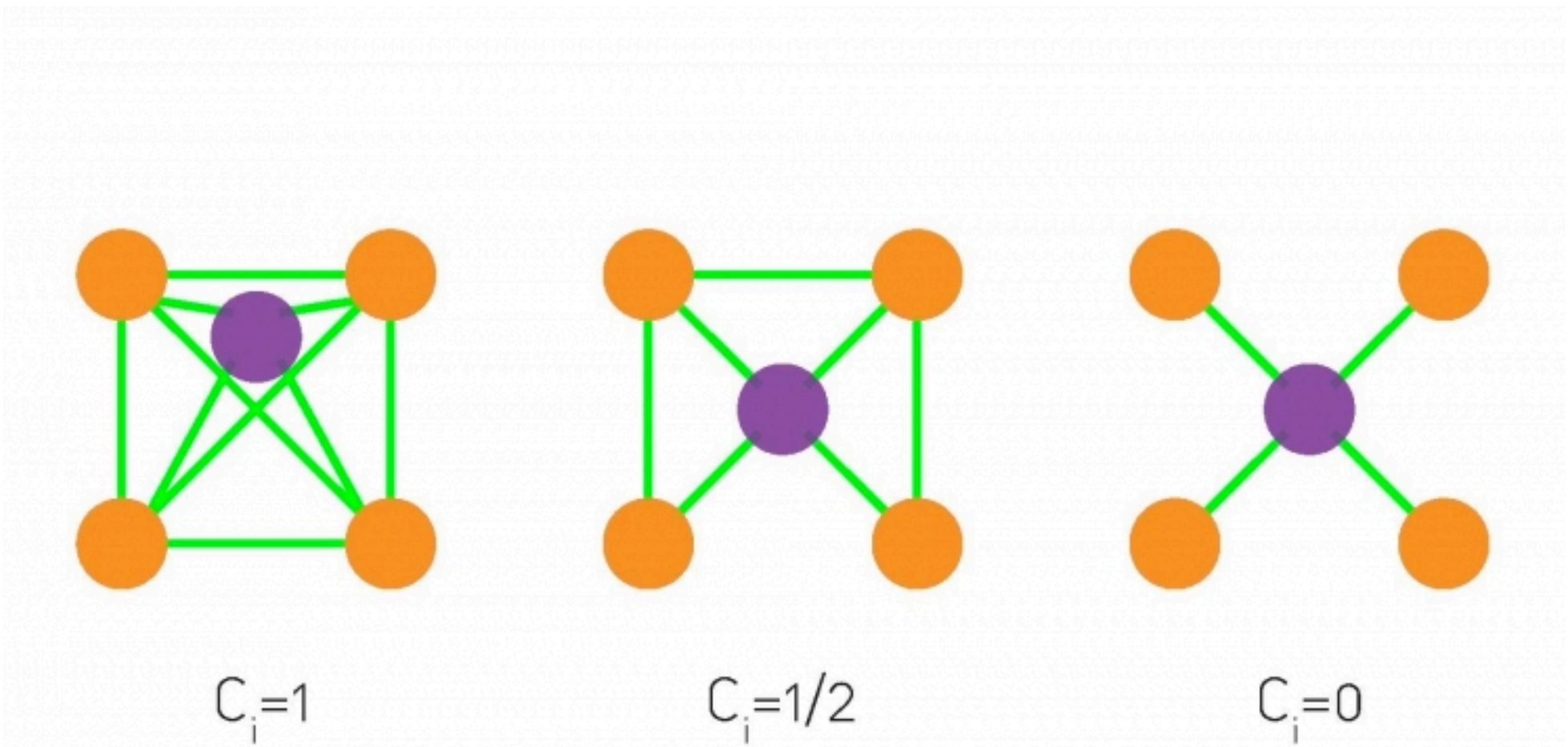
example from Siew et al, Complexity 2019

Density

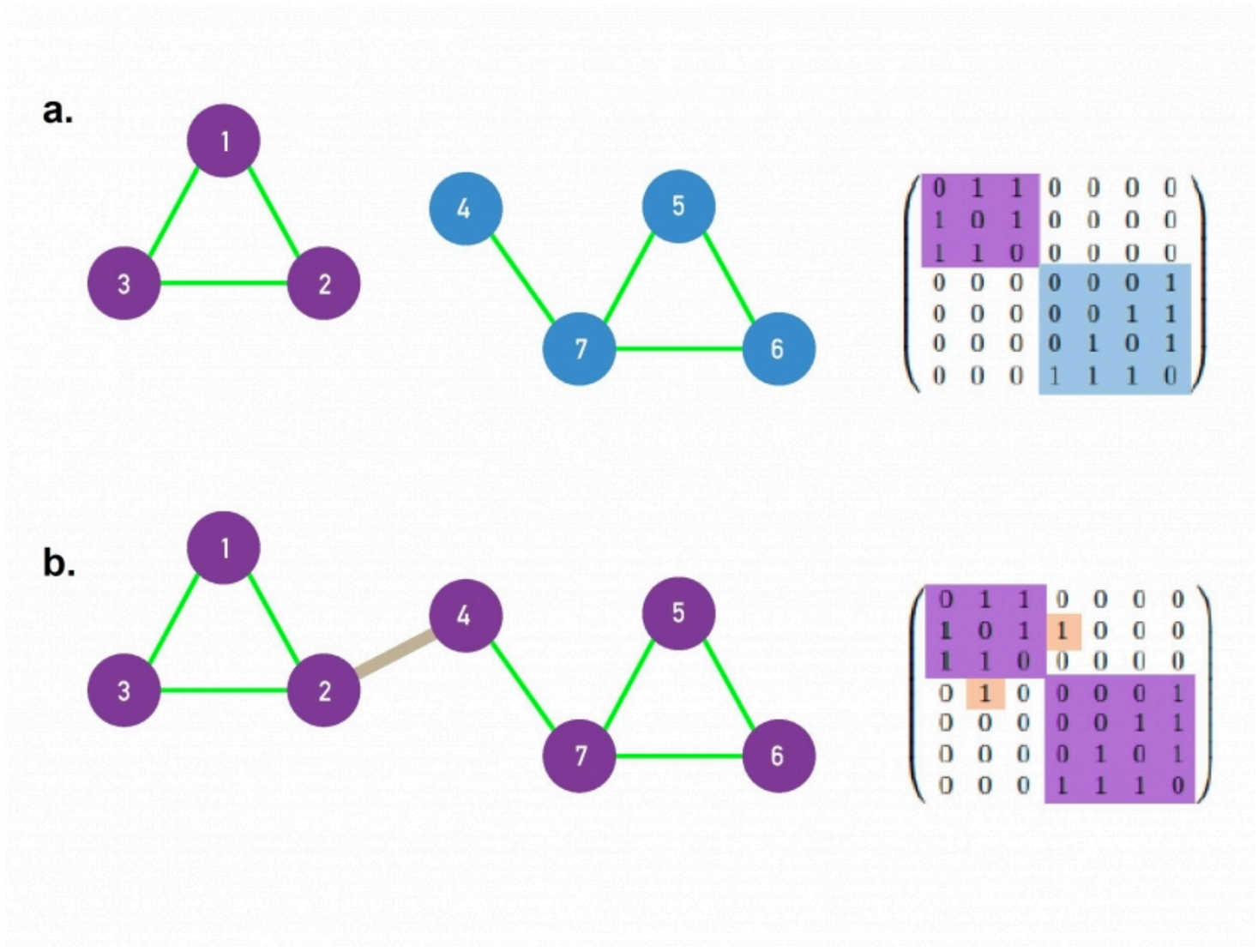


low **high**

CLUSTERING COEFFICIENT



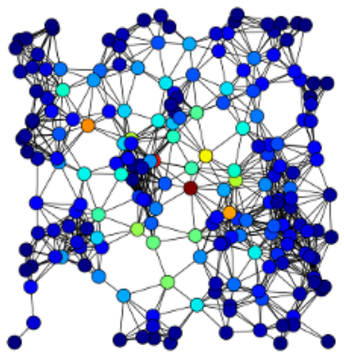
CONNECTED COMPONENTS

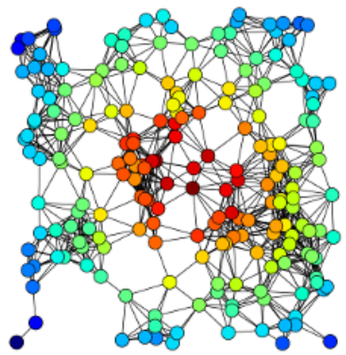


CENTRALITIES

https://en.wikipedia.org/wiki/Centrality

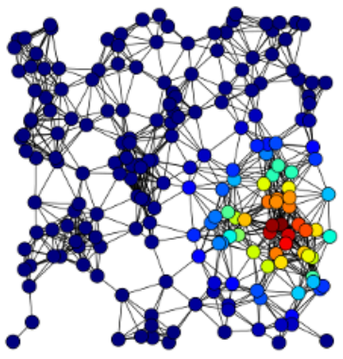
Betweenness centrality

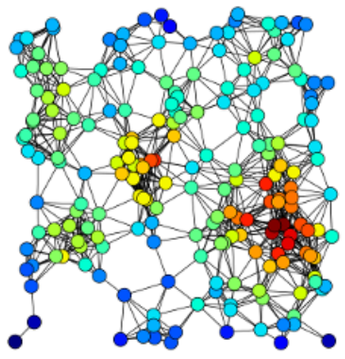




Closeness centrality

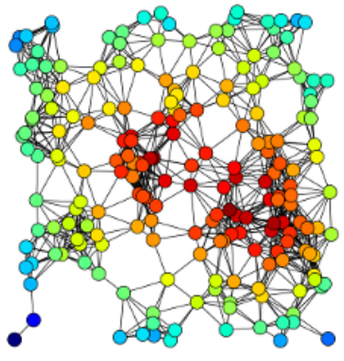
Eigenvector centrality

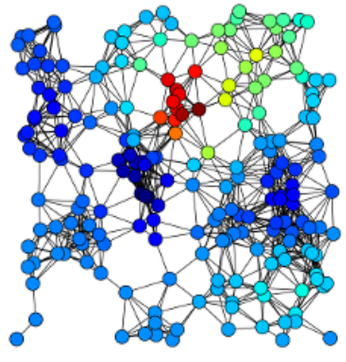




Degree centrality

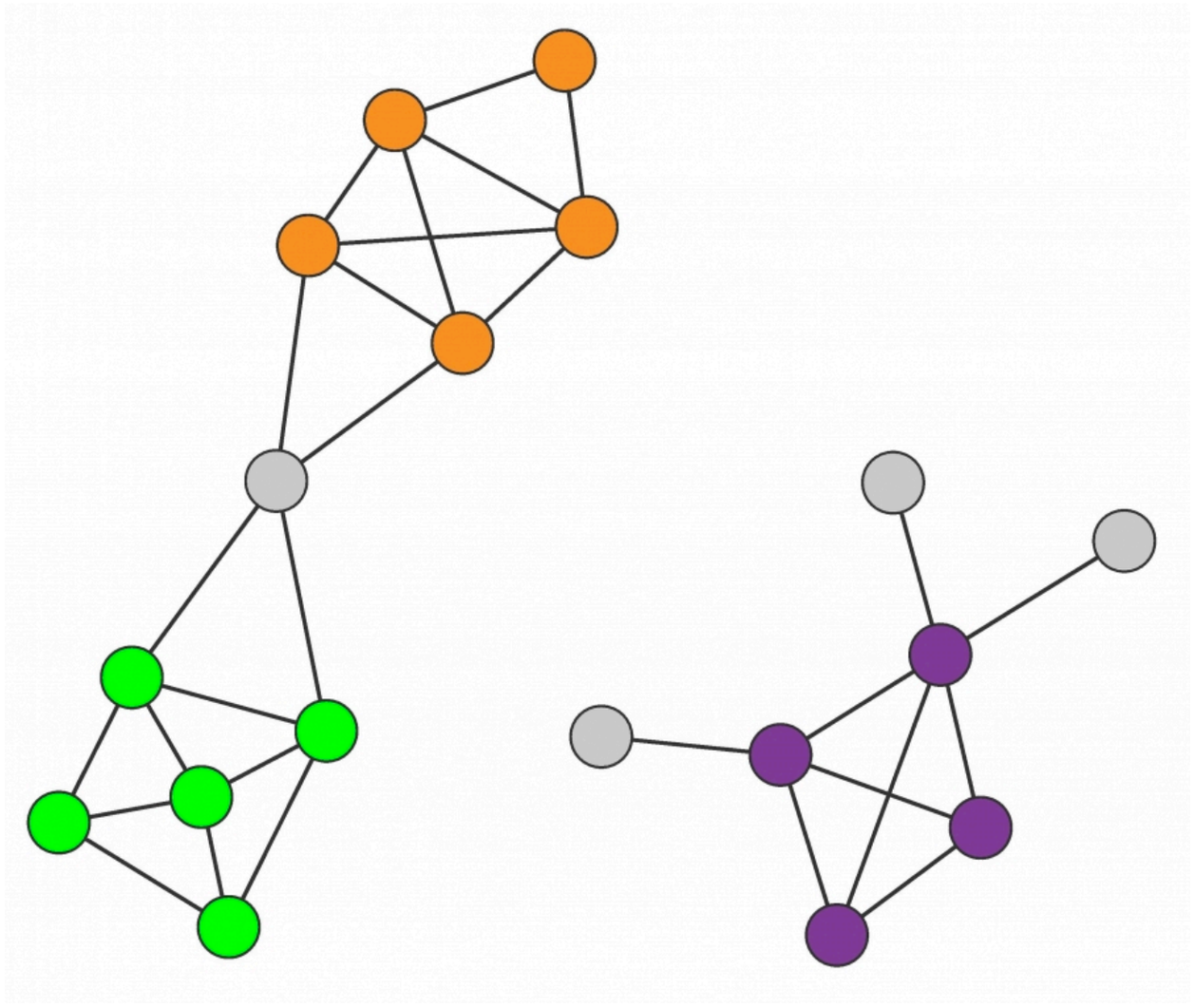
Harmonic centrality



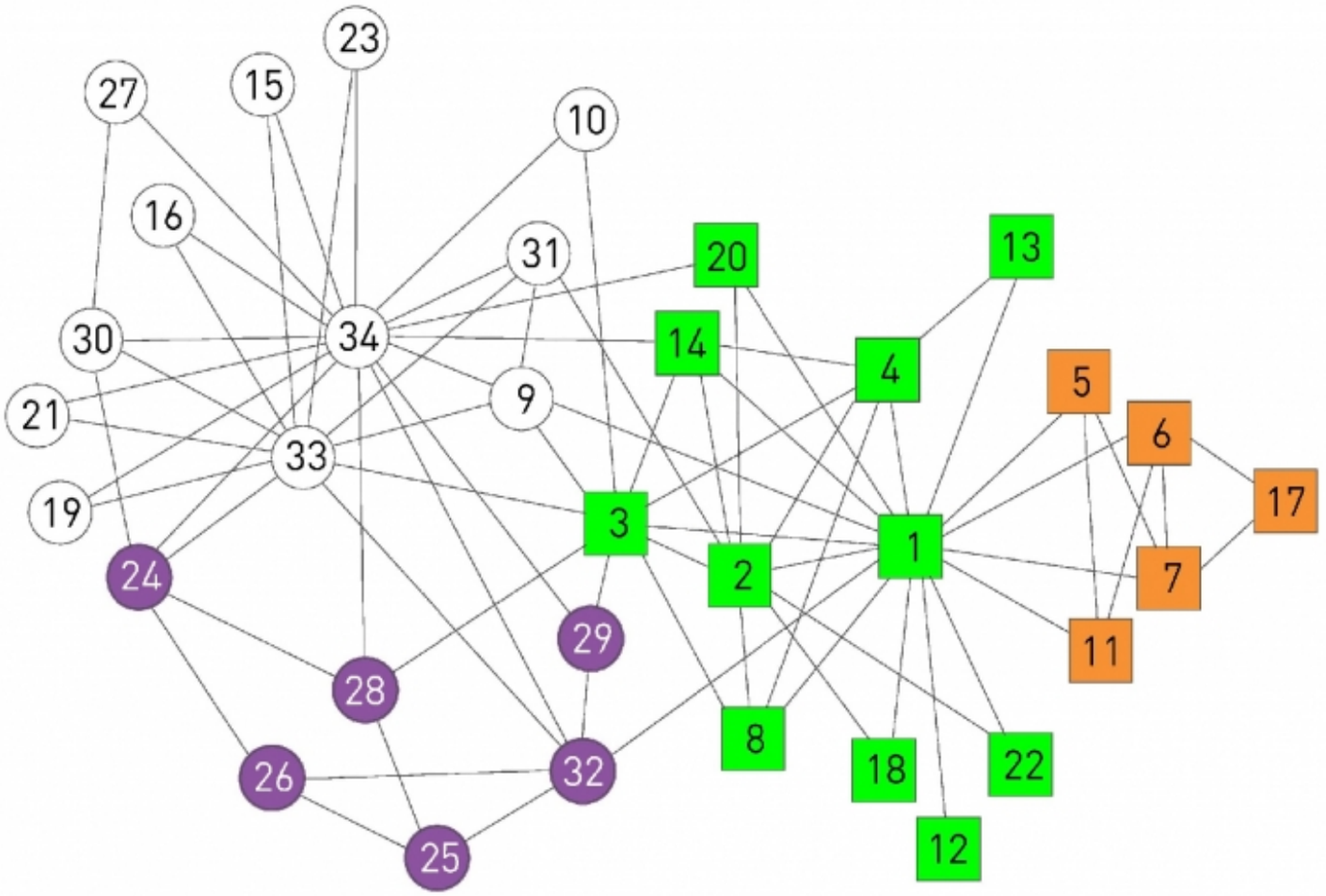


Katz centrality

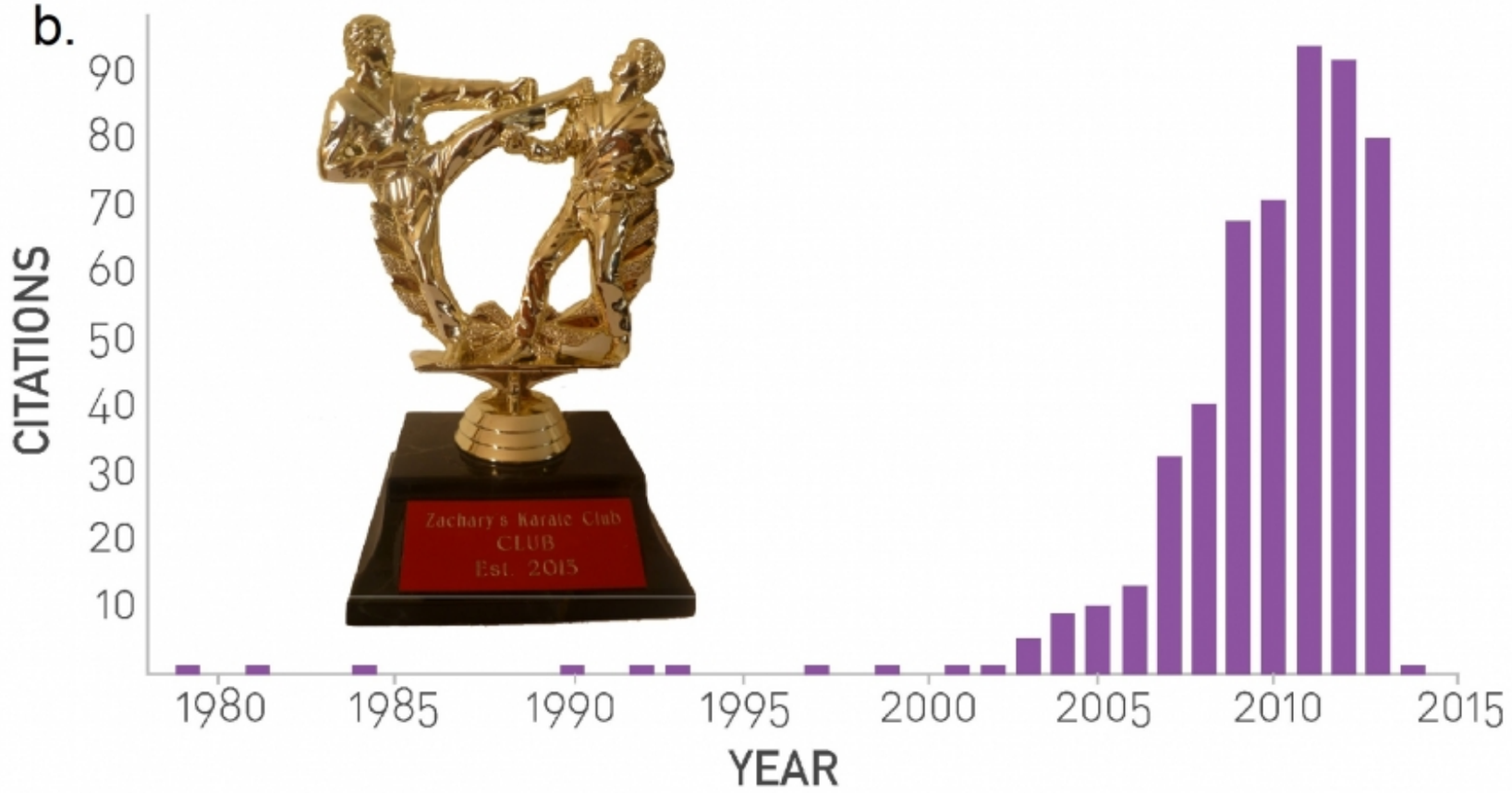
COMMUNITIES



a.


23
27
15
10
16
31
20
13
30
34
14
4
5
21
9
6
33
17
19
3
1
24
2
7
29
11
28
8
26
32
18
22
25
12

b.


CITATIONS
90
80
70
60
50
40
30
20
10
Zachary's Karate Club
CLUB
Est. 2015
1980
1985
1990
1995
2000
2005
2010
2015
YEAR