From Brains and Friendships to the Stock Market and the Internet

-Sanjukta Krishnagopal

10 July 2018

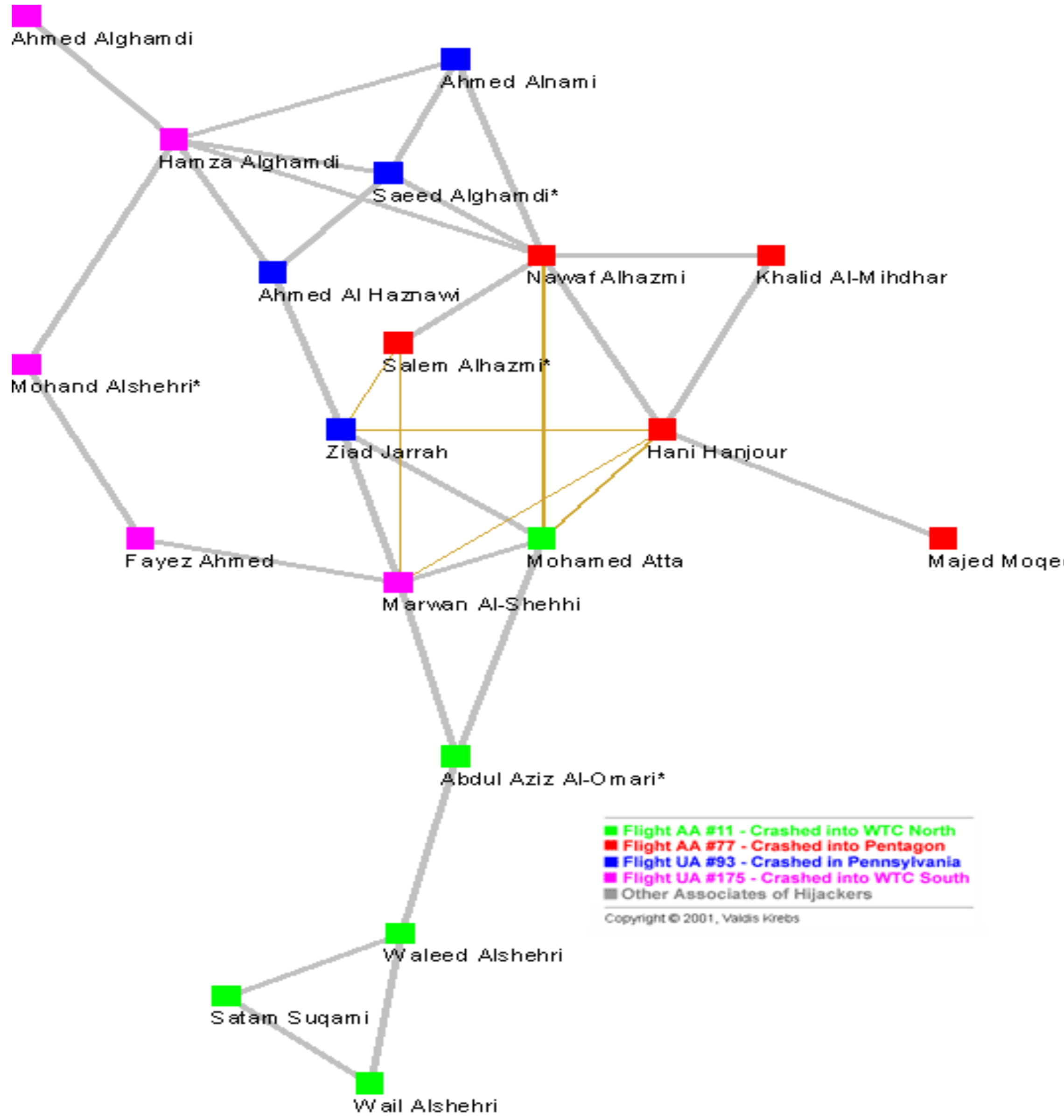
Girls Talk Math Summer Camp


"I think the next century
will be the century
of complexity."

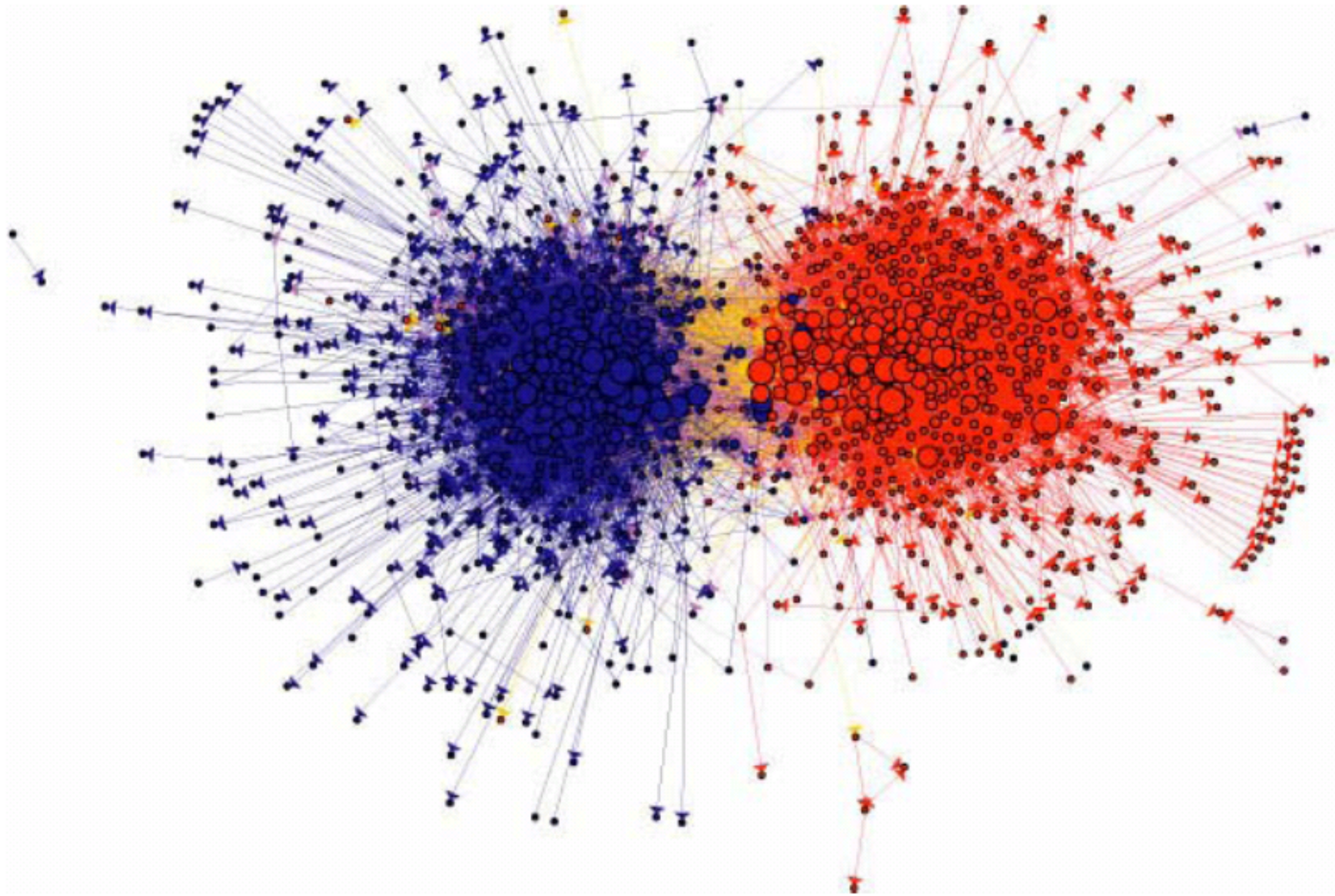
Some real networks

- Social Networks
  - Networks of acquaintances
  - Collaboration networks- actor networks, co-authorship networks
- Information Networks
  - Citation network (directed acyclic)
  - The Web (directed)
- Distribution Networks
  - The Internet
  - Power Grids, Telephone Networks
  - Transportation Networks (roads, railways, pedestrian traffic)
- Biological Networks
  - Protein-Protein Interaction Networks
  - Gene networks
  - Metabolic pathways
  - The Food Web
  - Brain Networks
- Economic Networks
  - Bank networks
- Natural Networks
  - River networks
  - Epidemic

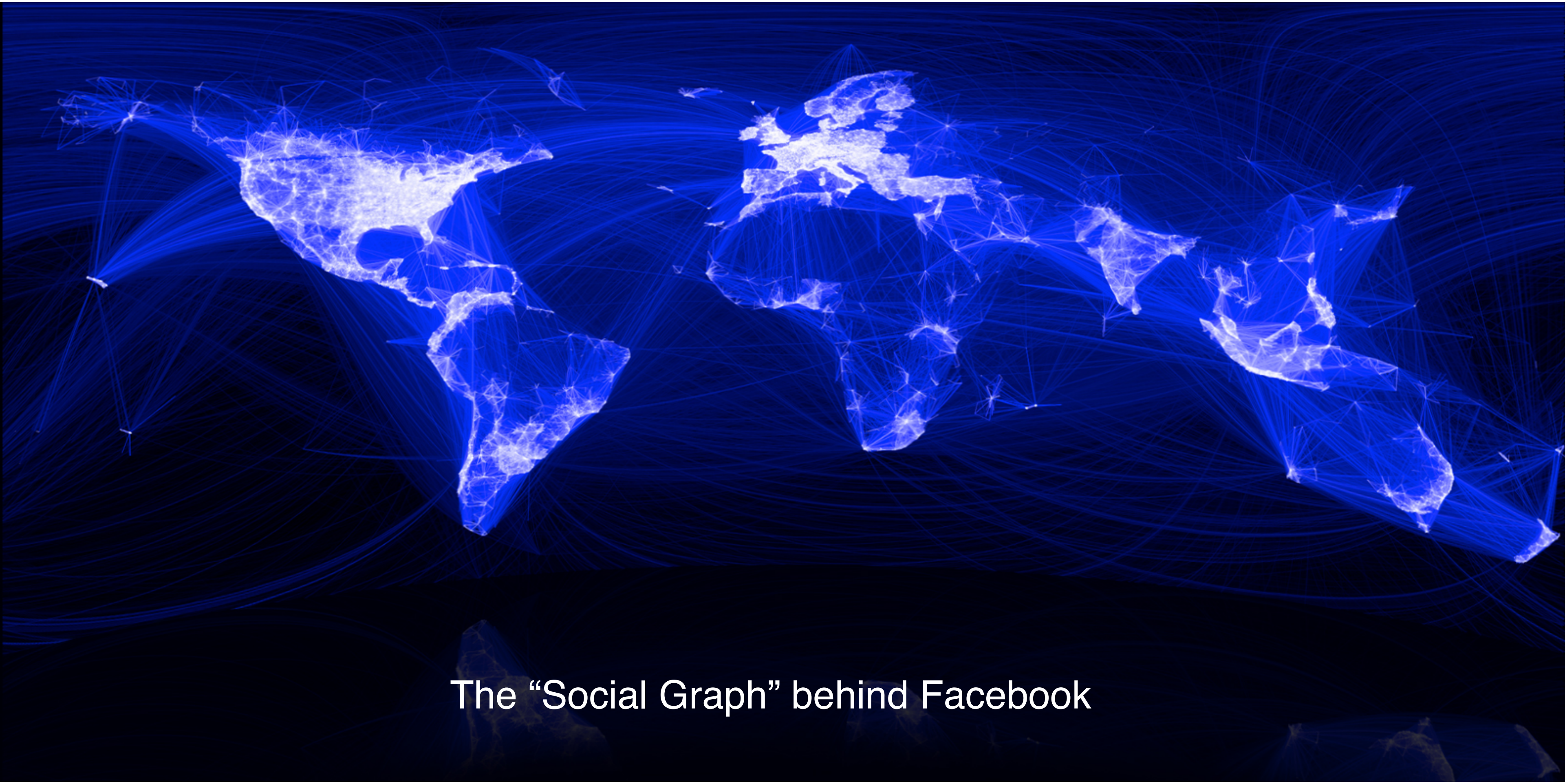
Terrorist Network



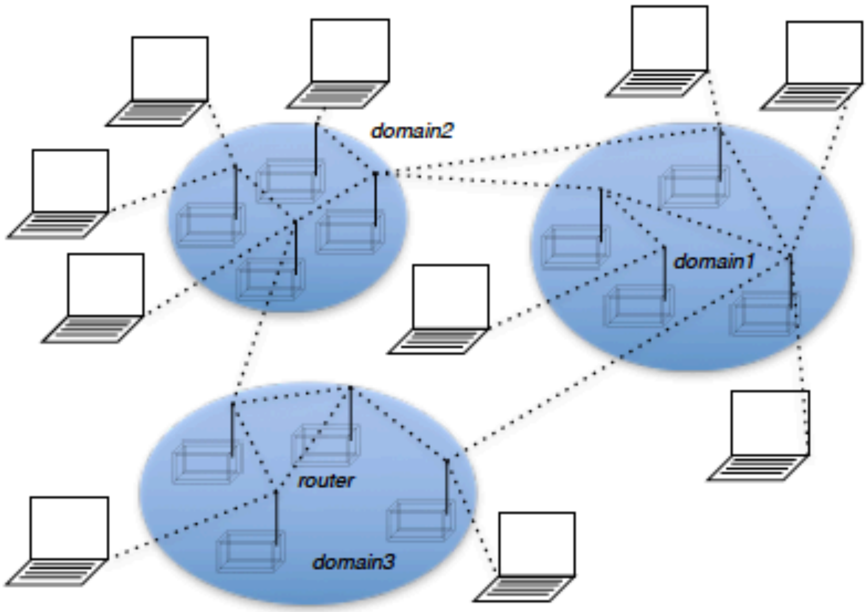
Links among blogs (2004 presidential) election)

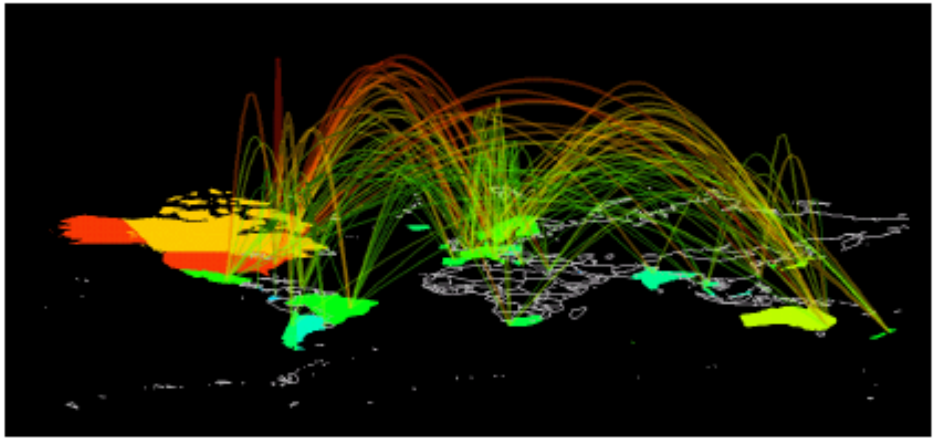


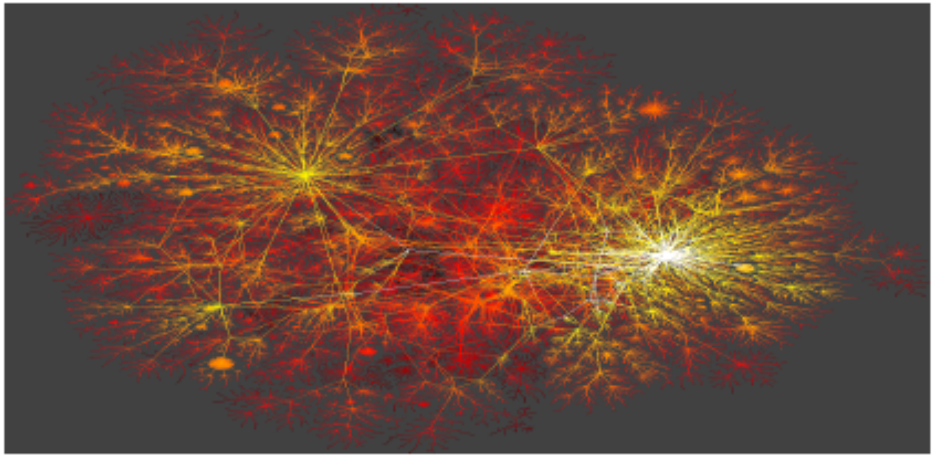
Facebook



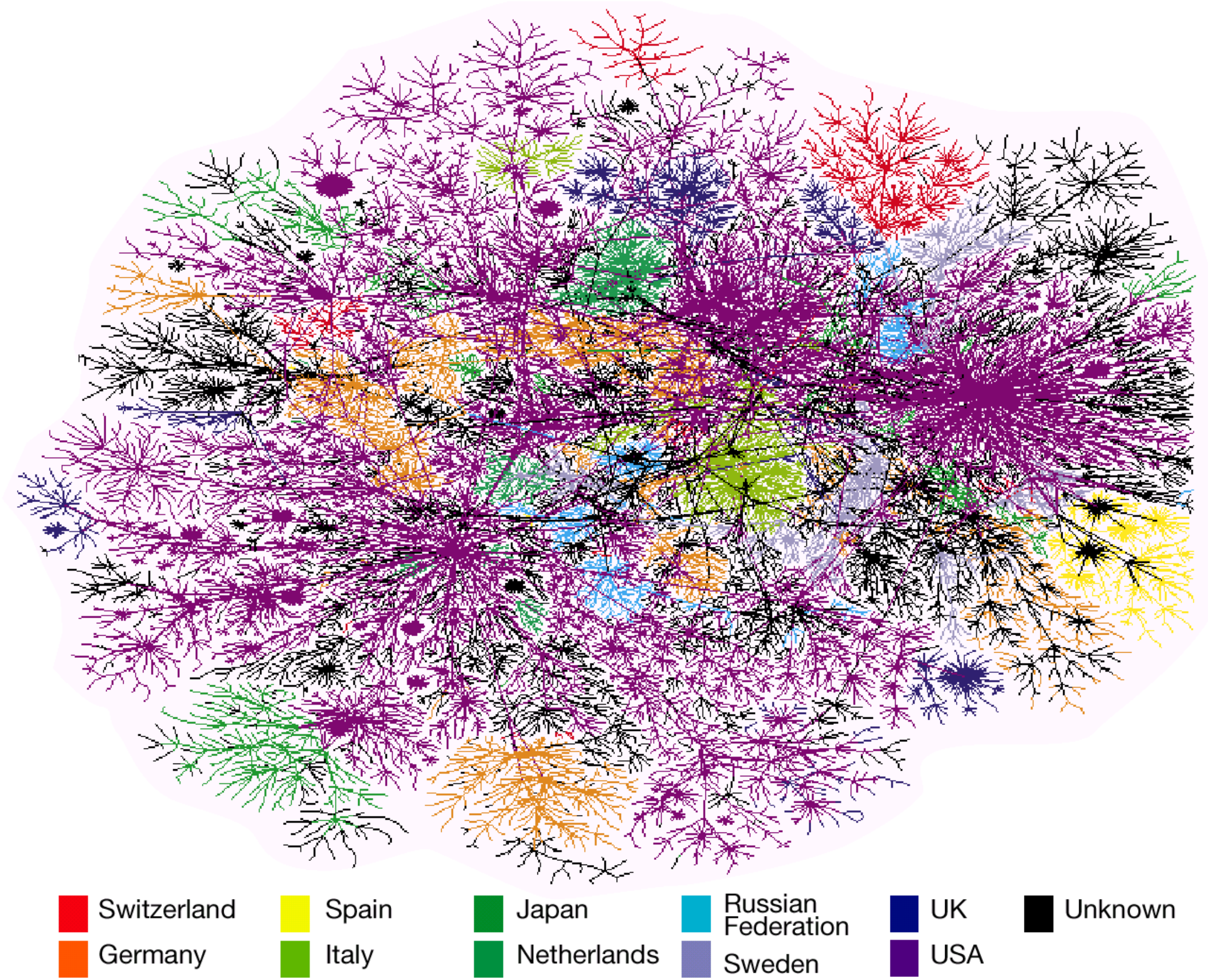
INTERNET



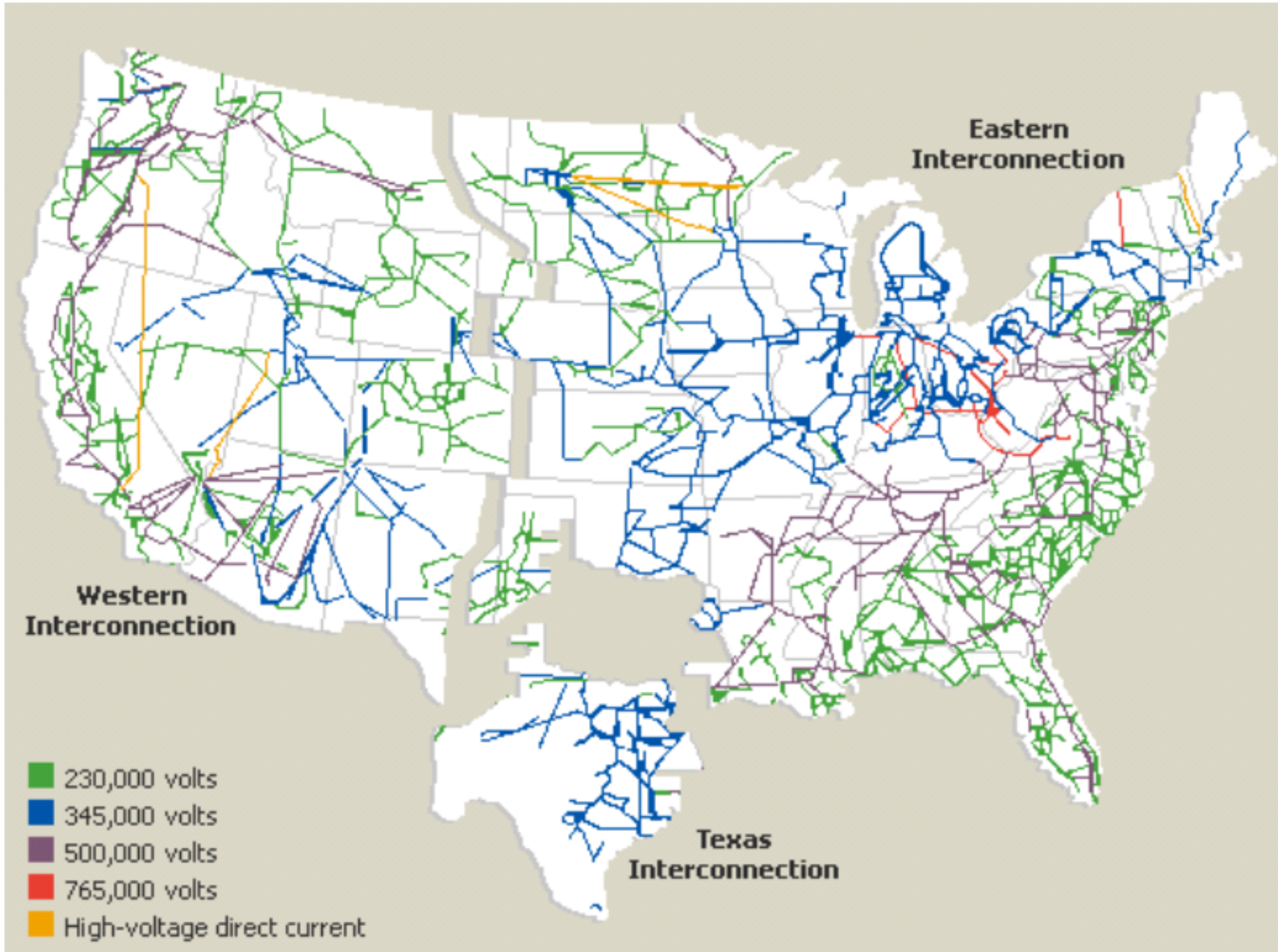




Internet as measured by Hal Burch and Bill Cheswick's Internet Mapping Project.



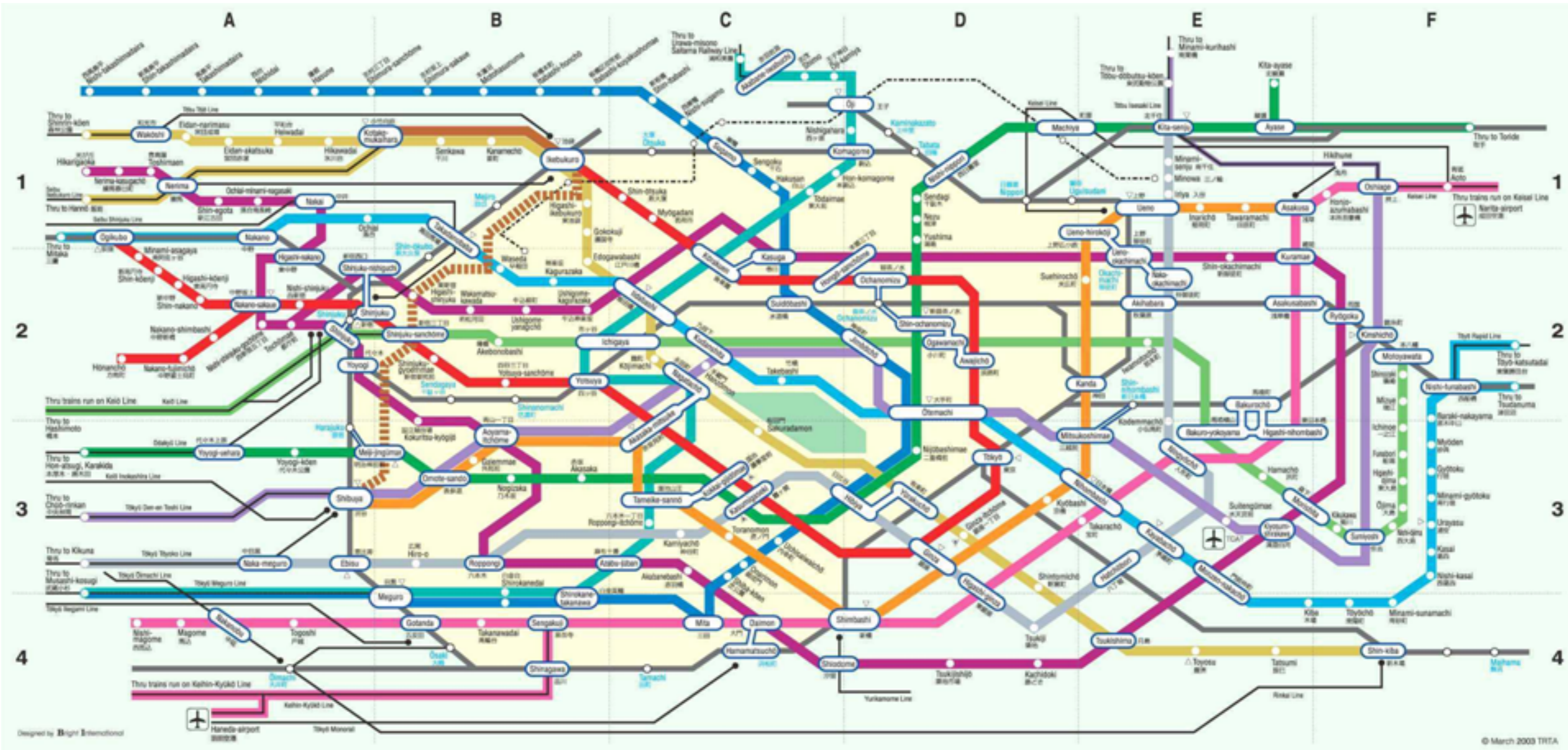
Power networks



transportation networks: airlines

Source: Northwest Airlines WorldTraveler Magazine

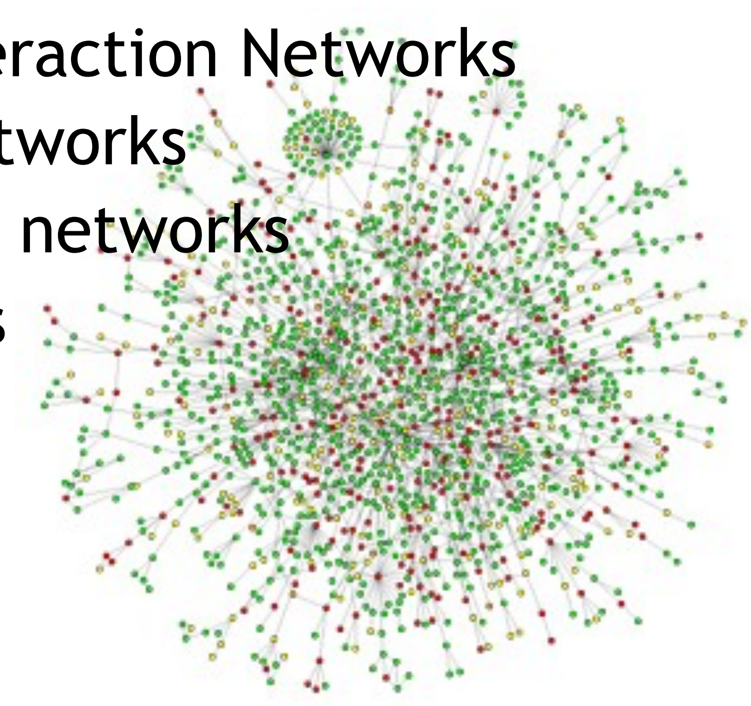
transportation networks: railway maps



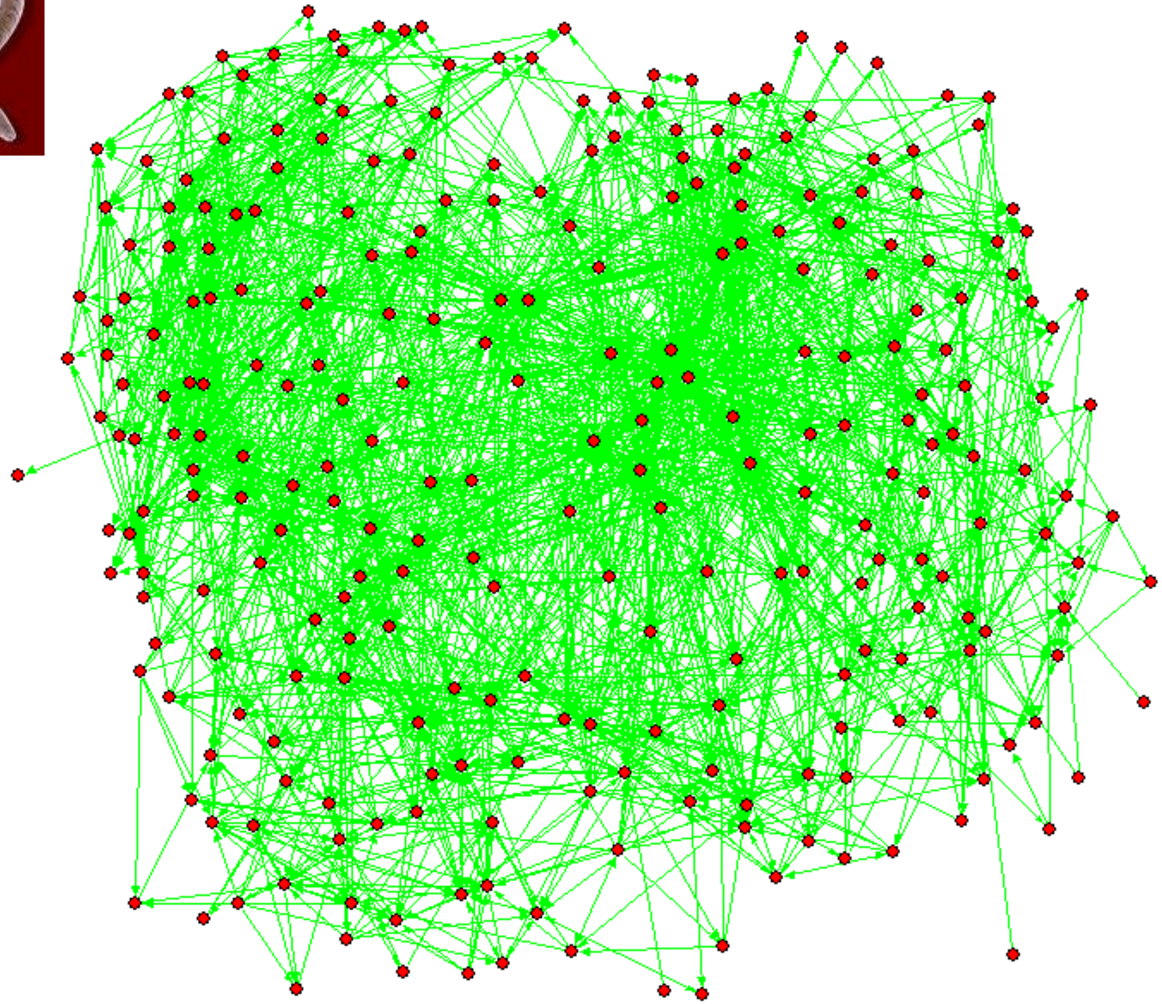
Source: TRTA, March 2003 - Tokyo rail map

Biological networks

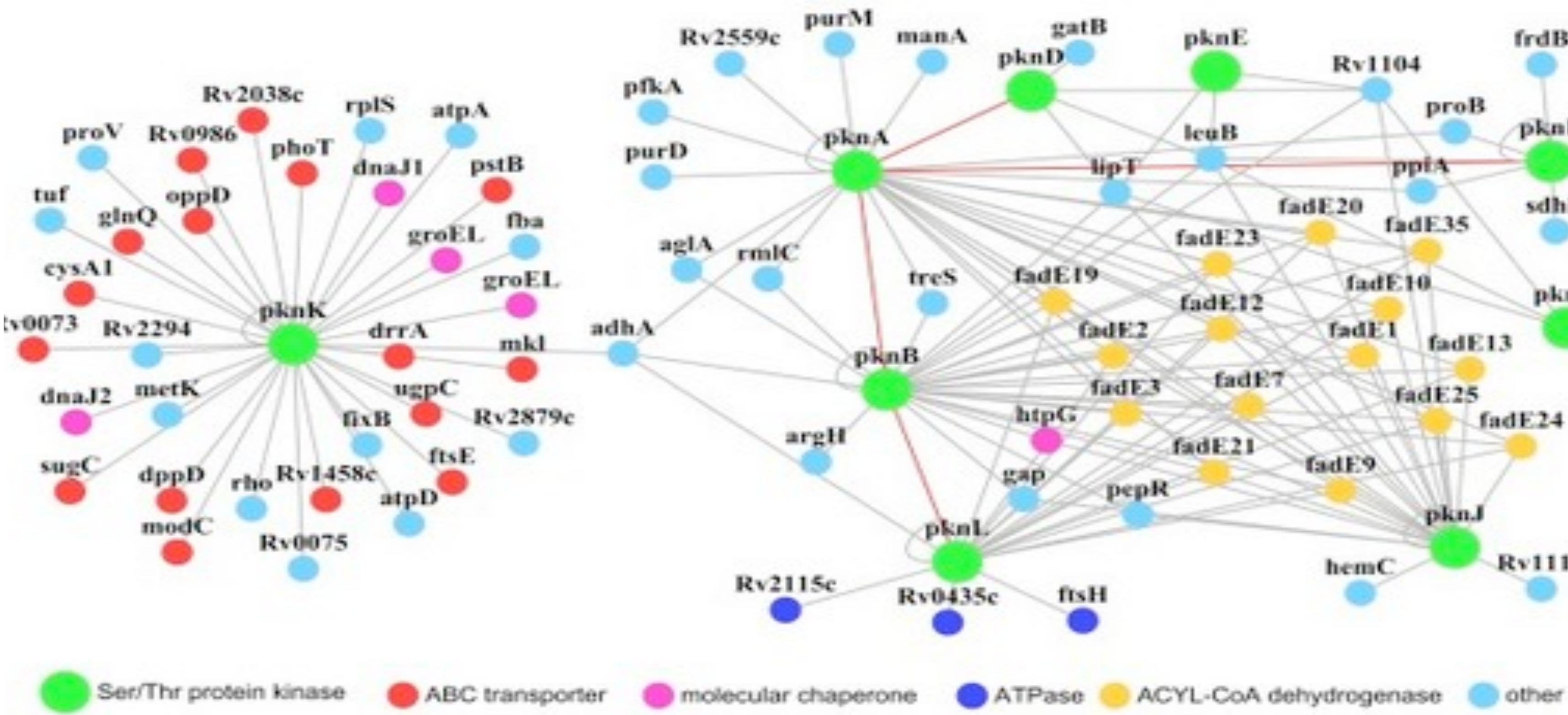
- Biological systems represented as networks
  - Protein-Protein Interaction Networks
  - Gene regulation networks
  - Gene co-expression networks
  - Metabolic pathways
  - The Food Web
  - Neural Networks



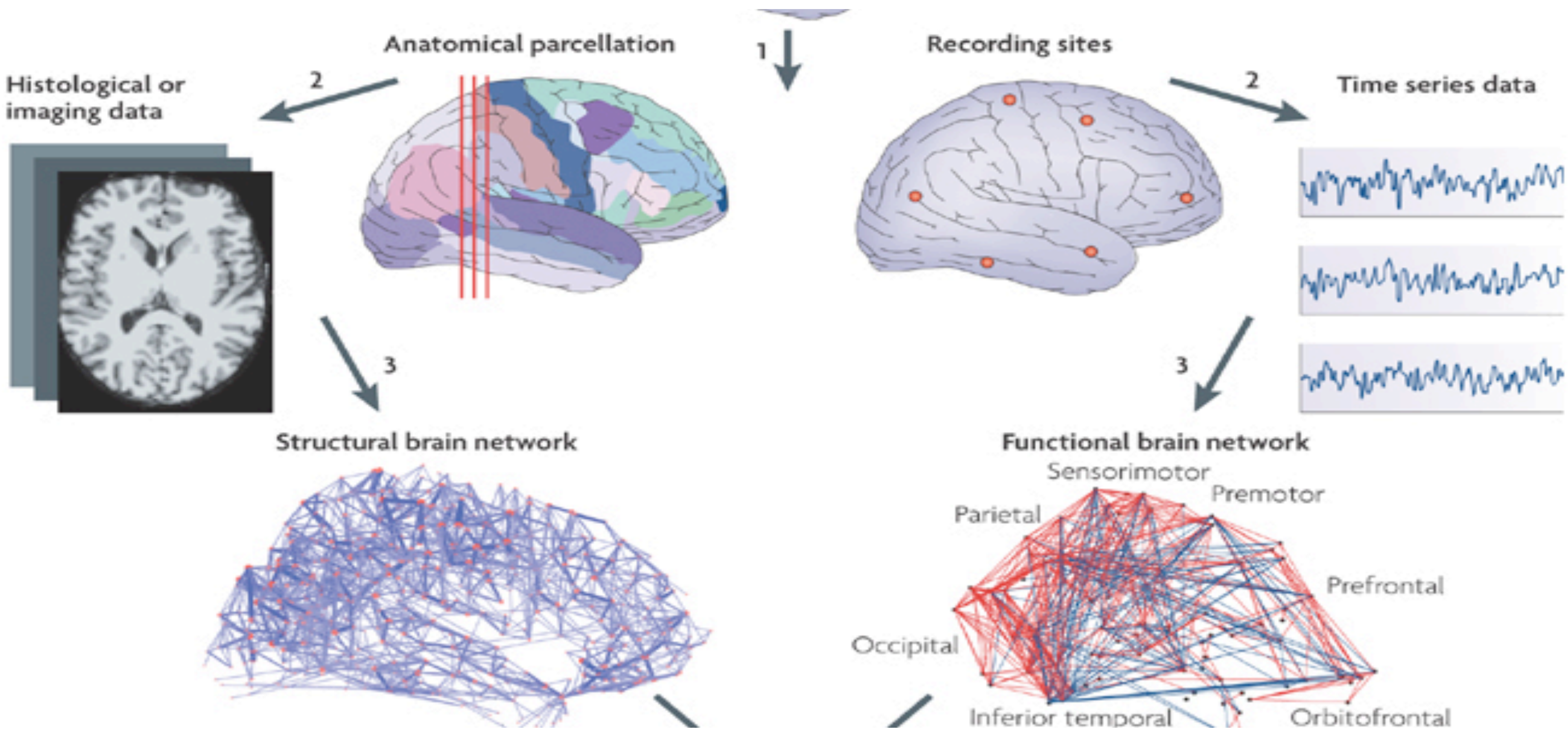
Brain network of C. elegans neurons



Protein interaction networks

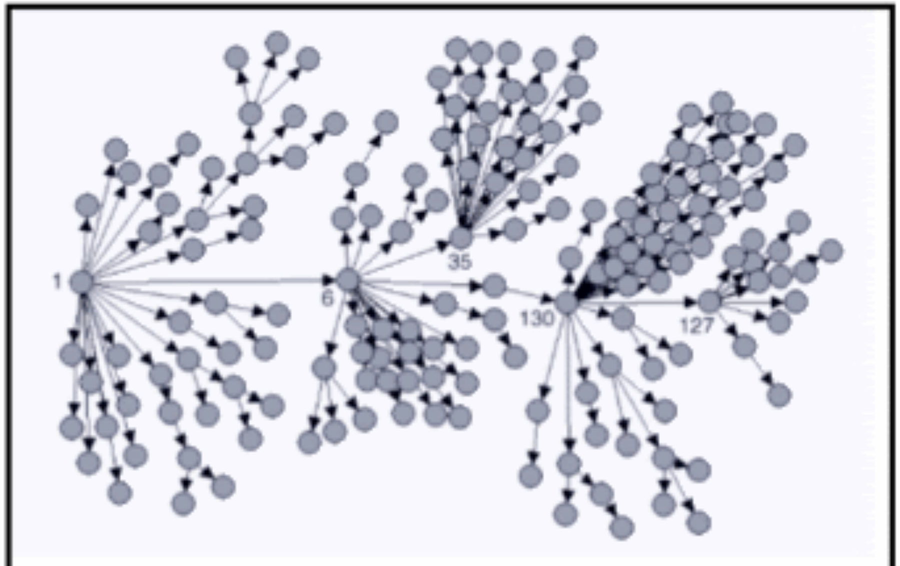


Brain networks



Epidemic networks

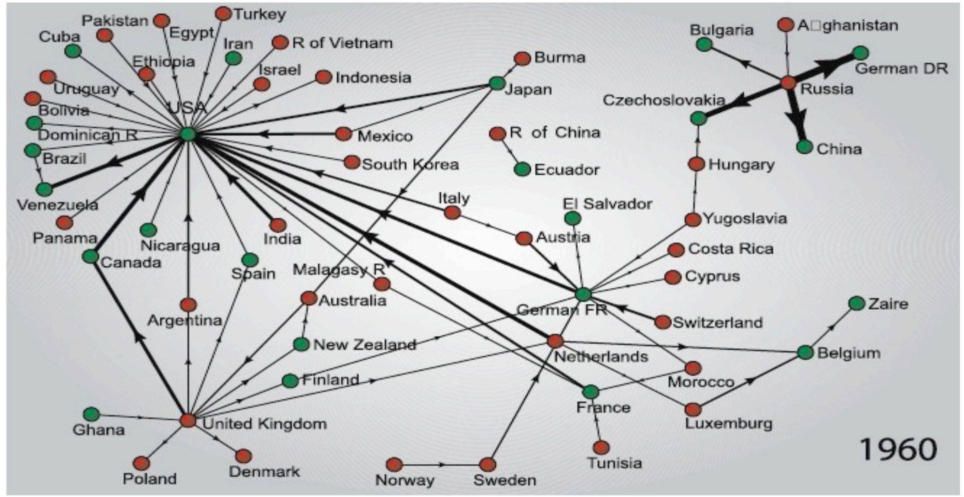
FIGURE 2. Probable cases of severe acute respiratory syndrome, by reported source of infection* — Singapore, February 25–April 30, 2003

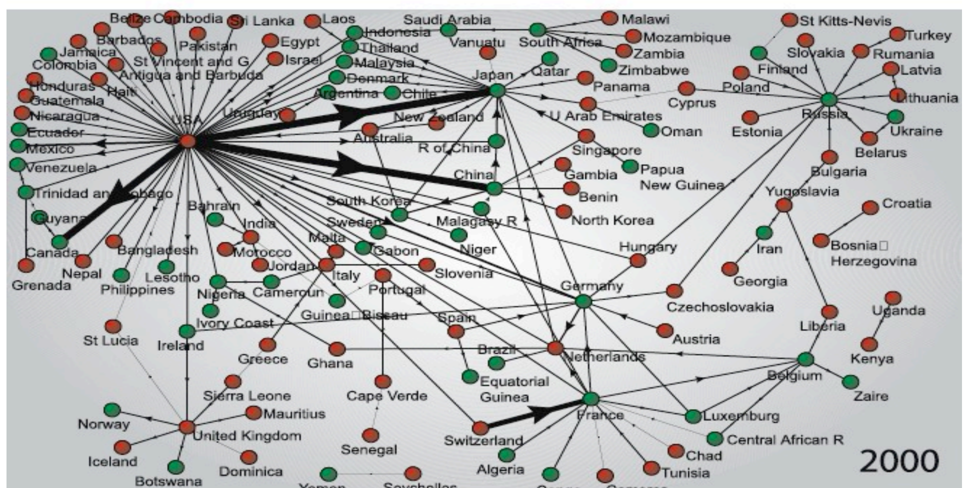


* Patient 1 represents Case 1; Patient 6, Case 2; Patient 35, Case 3; Patient 130, Case 4; and Patient 127, Case 5. Excludes 22 cases with either no or poorly defined direct contacts or who were cases translocated to Singapore and the seven contacts of one of these cases.

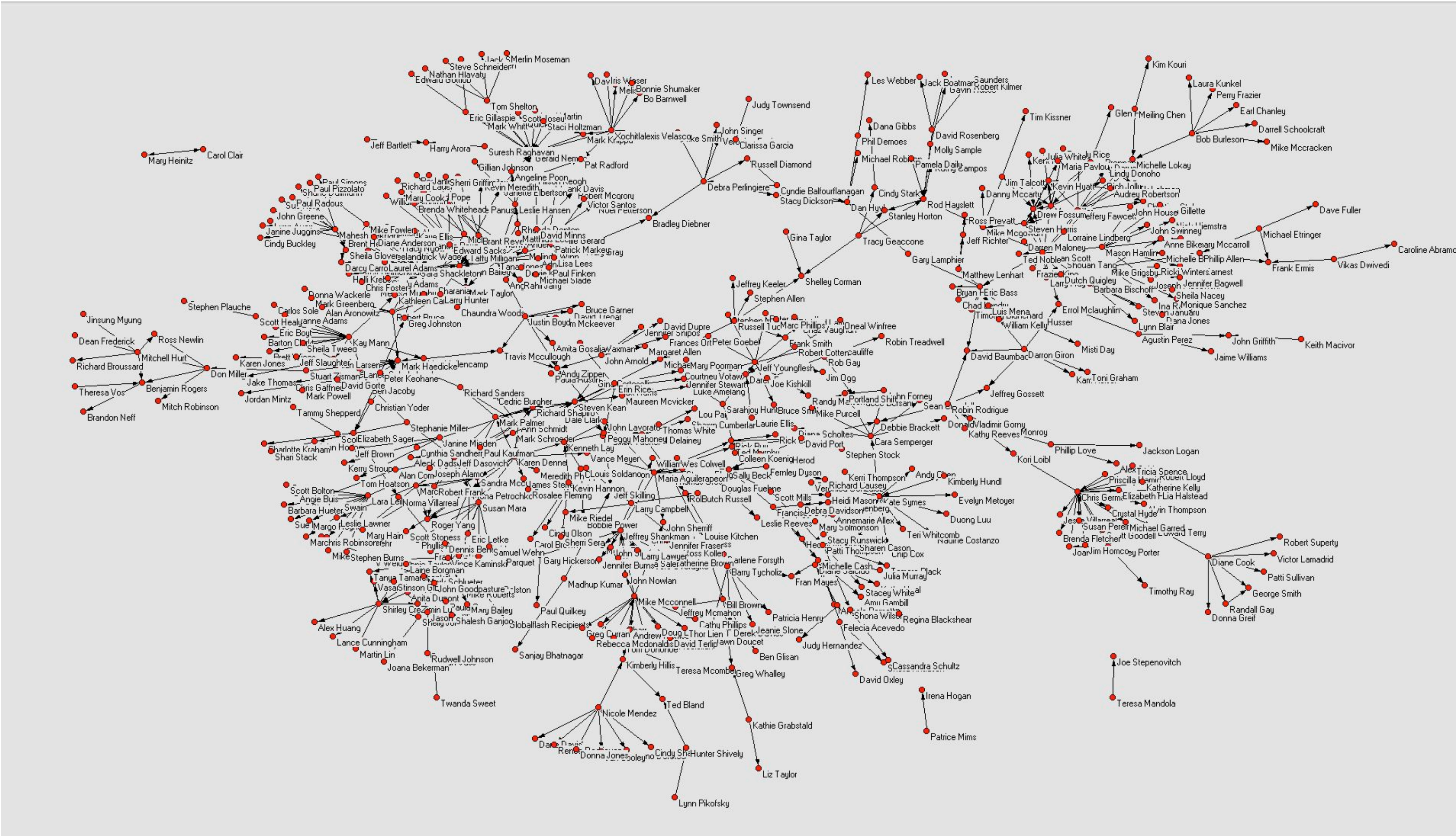
Reference: Bogatti SP. Netdraw 1.0 Network Visualization Software. Harvard, Massachusetts: Analytic Technologies, 2002.

World-trade networks





Email exchanges in a company



Life is complex…

We are surrounded by complex systems, from society- people-people interactions, communication systems to neurons in the brain.

All these systems work together in a seamless fashion.

These systems, random looking at first, upon close inspection display self-organization patterns whose quantification, understanding, prediction and eventually control is the major intellectual challenge.

THE ROLE OF NETWORKS

Behind each system studied in complexity there is an intricate wiring diagram, or a **network**, that defines the interactions between the component.

We will never understand complex system unless we map out and understand the networks behind them.

What is Network Science?

- The study of complex systems that can be represented as (typically dynamic) networks
  - Society, Economy
  - Various biological networks (e.g., metabolism)
  - Brain
  - WWW, Internet, Transportation nets
  - Natural networks (global climate system)
  - Many others

- Relies on:
  - Network data (strong empirical basis)
  - Network models
  - Network algorithms

The roots of Network Science

- Graph Theory
- Statistical Mechanics
- Nonlinear Dynamics
- Games and Learning
- Data mining (“graph mining”) and machine learning
- Algorithms
- Complexity theory

THE CHARACTERISTICS OF NETWORK SCIENCE

Interdisciplinary:

Cell biologists/Computer Scientists/others needs the wiring diagram behind their system, extracting information from incomplete and noisy data sets.

Empirical, Data driven:

Focus is on data and its utility. Not just mathematical models, but tools to test real data. The value addition will be judged by the insights it offers.

Quantitative and Mathematical Nature:

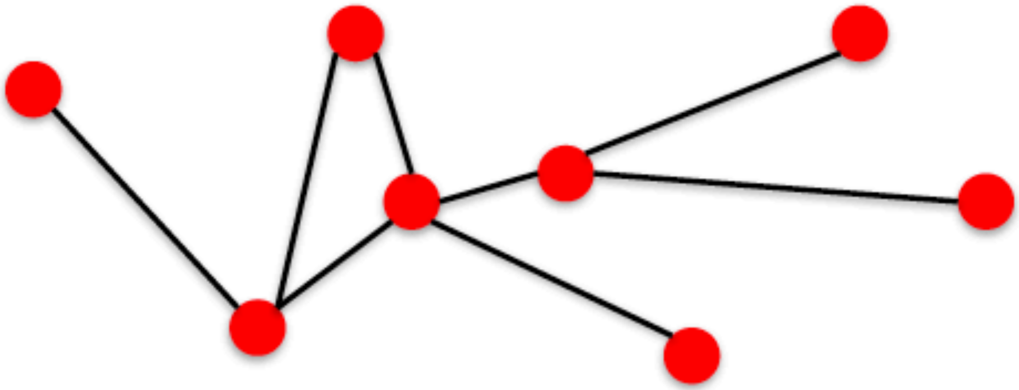
Graph theory; organizing principles from statistical physics, control and information theory, statistics and data mining.

Computational Intensive:

Size of networks giving rise to Big Data. Needs algorithms, database management, data mining, analytics, software tools.

Theory of Network Science

COMPONENTS OF A COMPLEX SYSTEM



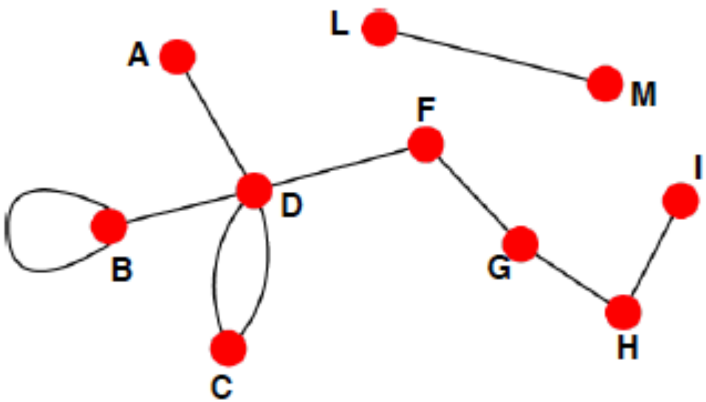
- **components**: nodes, vertices — N
- **interactions**: links, edges — L
- **system**: network, graph — (N,L)

UNDIRECTED VS. DIRECTED NETWORKS

Undirected

Links: undirected (*symmetrical*)

Graph:

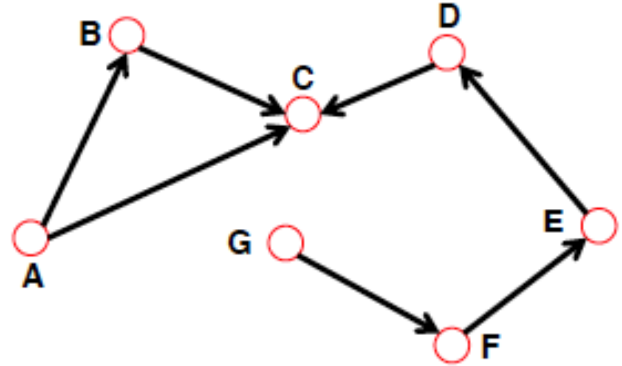


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:



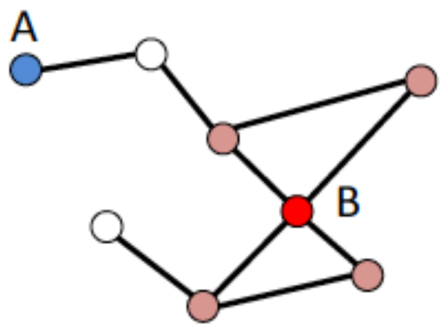
An undirected link is the superposition of two opposite directed links.

Directed links :
URLs on the www
phone calls
metabolic reactions

Degree, Average Degree and Degree Distribution

NODE DEGREES

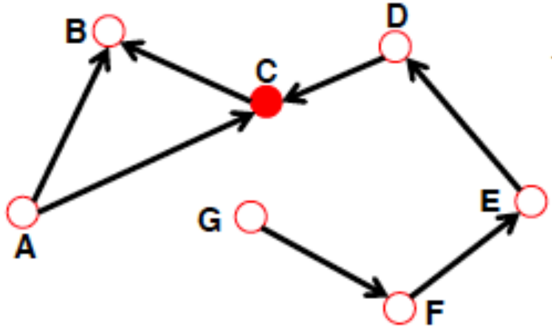
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed



In *directed networks* we can define an in-degree and out-degree.

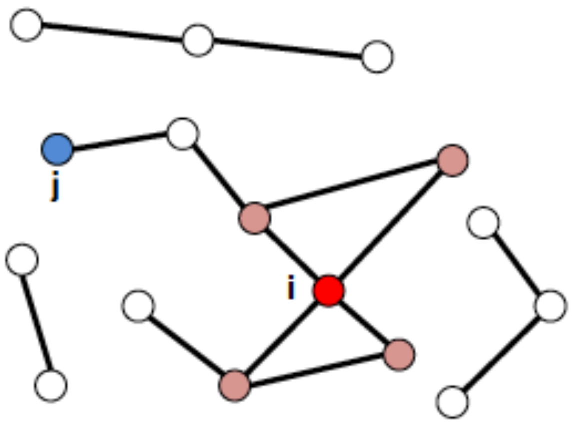
The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \quad k_C^{out} = 1 \quad k_C = 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

AVERAGE DEGREE

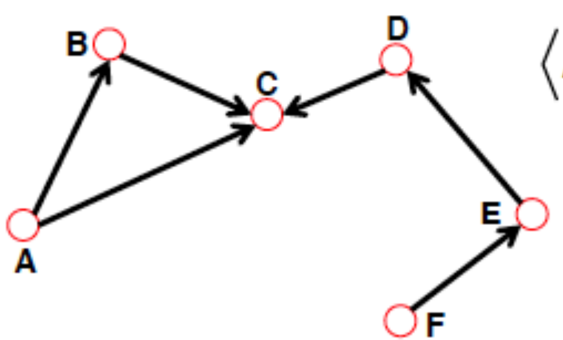
Undirected



$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

Directed

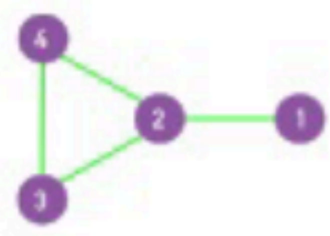


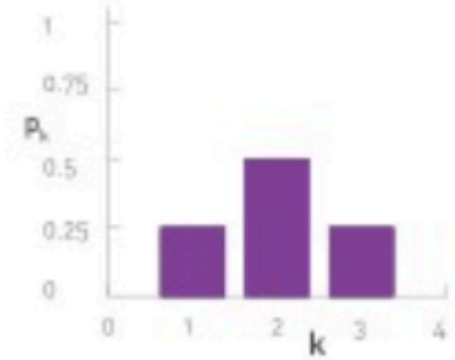
$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Degree distribution

P(k): probability that a
randomly chosen node
has degree *k*

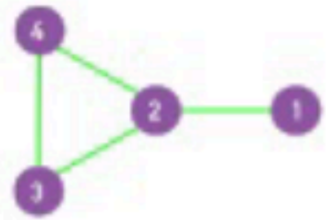


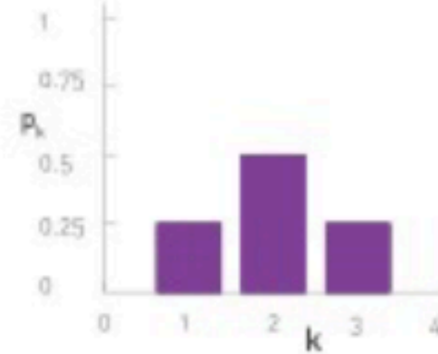


DEGREE DISTRIBUTION

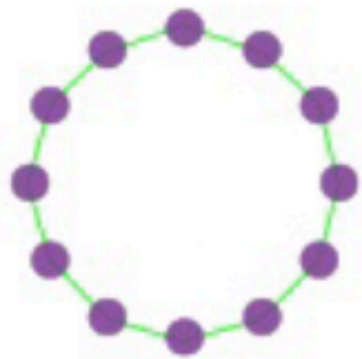
Degree distribution

P(k): probability that a randomly chosen node has degree *k*





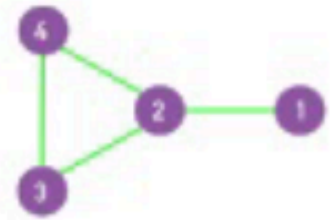
What is the degree distribution?

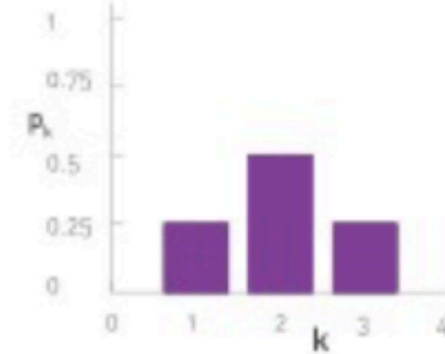


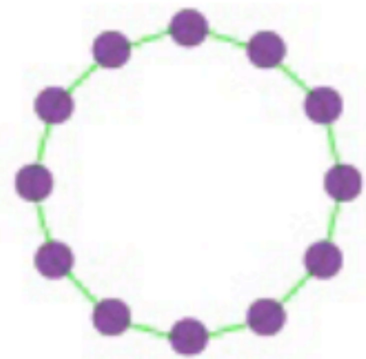
DEGREE DISTRIBUTION

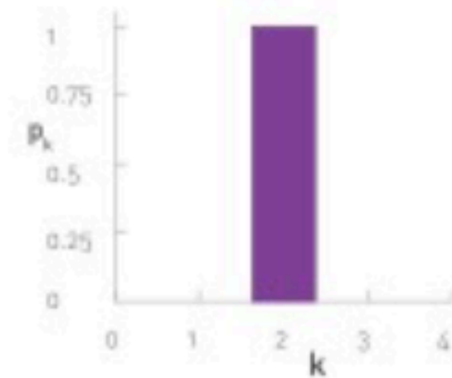
Degree distribution

P(k): probability that a
 randomly chosen node
has degree *k*







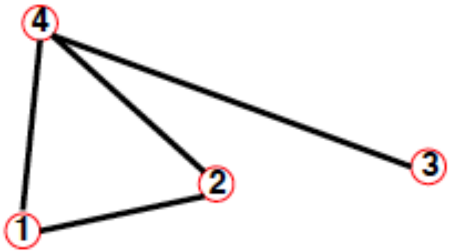


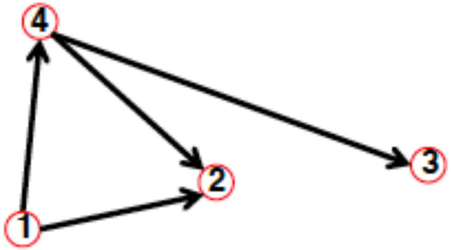
Average Degree

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
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-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Adjacency matrix

ADJACENCY MATRIX





A_{ij}=1 if there is a link between node *i* and *j*

A_{ij}=0 if nodes *i* and *j* are not connected to each other.

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Note that for a directed graph (right) the matrix is not symmetric.

$A_{ij} = 1$ if there is a link pointing from node *j* and *i*

$A_{ij} = 0$ if there is no link pointing from *j* to *i*.

For which network is the adjacency matrix symmetric?

A. Directed

B. Undirected

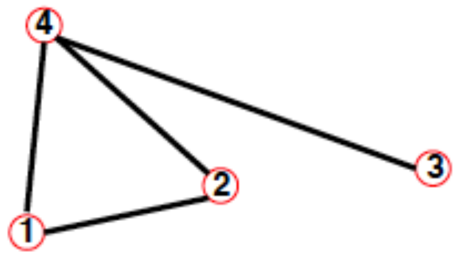
For which network is the adjacency matrix symmetric?

A. Directed

B. Undirected

ADJACENCY MATRIX AND NODE DEGREES

Undirected



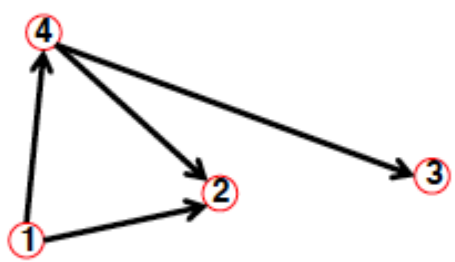
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

Directed



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

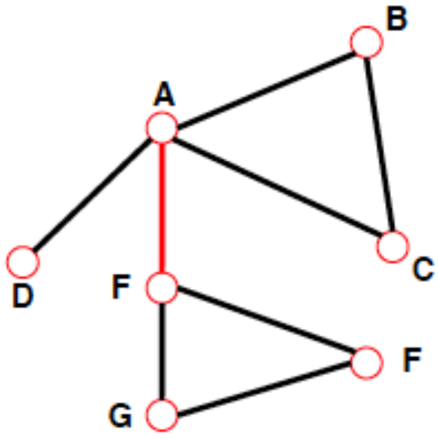
$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

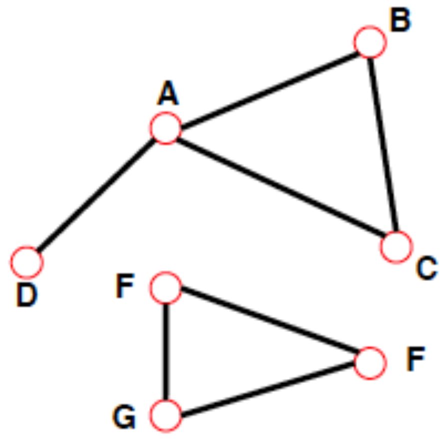
$$k_j^{out} = \sum_{i=1}^{N} A_{ij}$$

CONNECTEDNESS

CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.

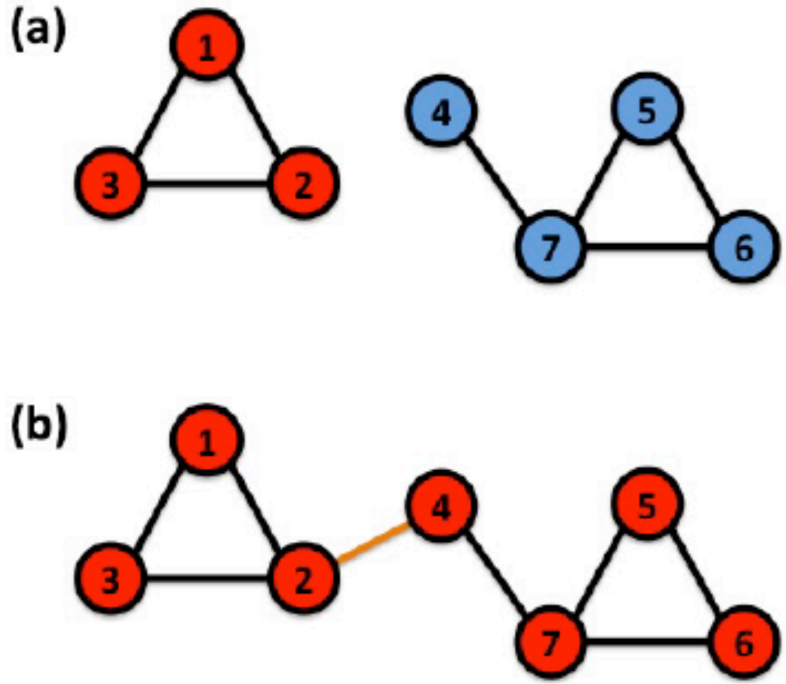




Largest Component:
Giant Component

The rest: **Isolates**

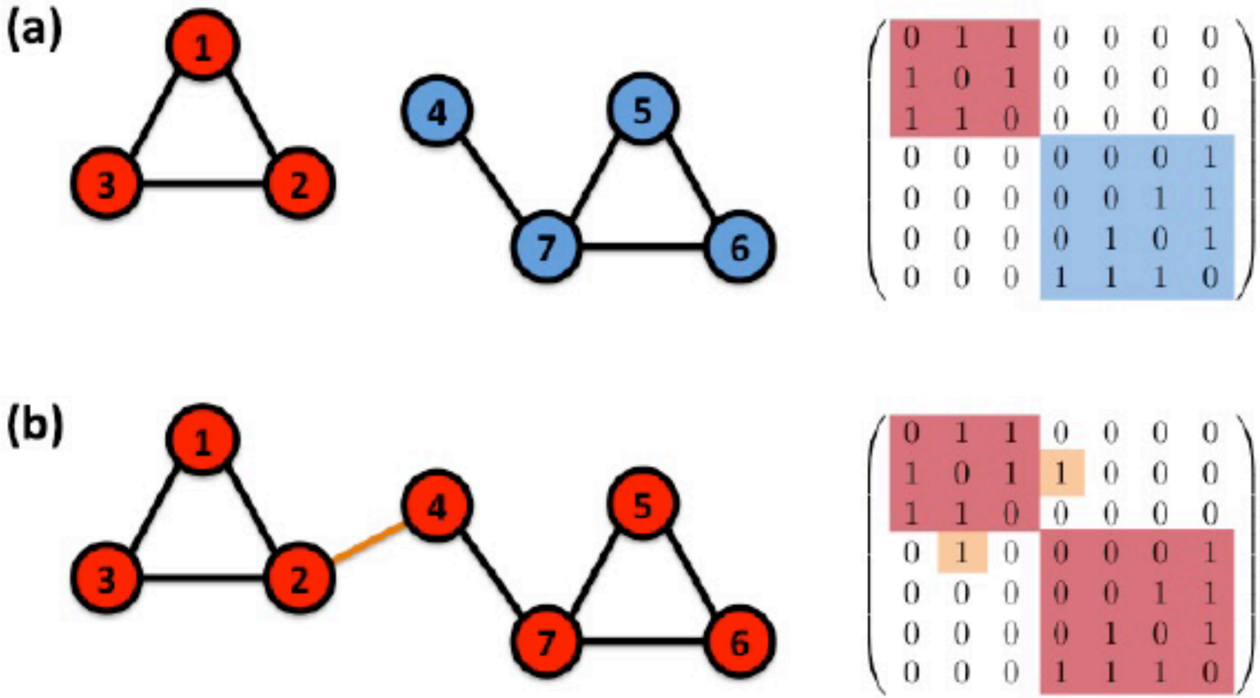
Bridge: if we erase it, the graph becomes disconnected.


(a)
1
2
3
4
5
6
7
(b)
1
2
3
4
5
6
7

CONNECTIVITY OF UNDIRECTED GRAPHS

Adjacency Matrix

The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

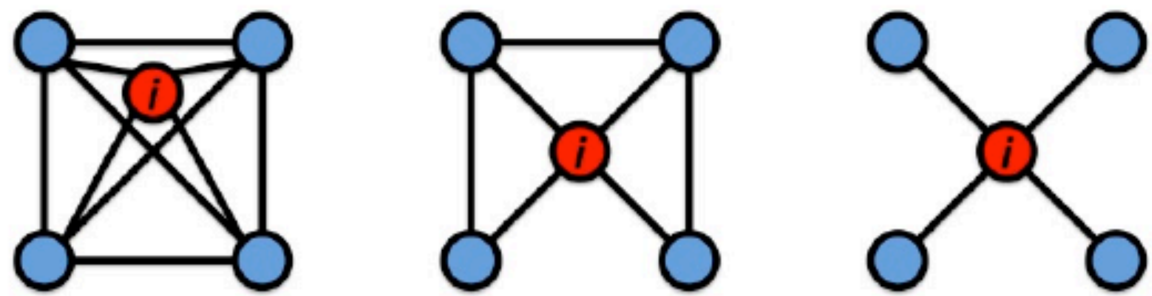


Clustering coefficient

CLUSTERING COEFFICIENT

✱ Clustering coefficient:

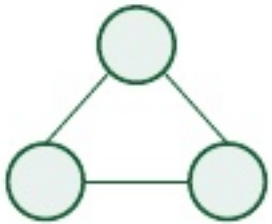
what fraction of your neighbors are connected?



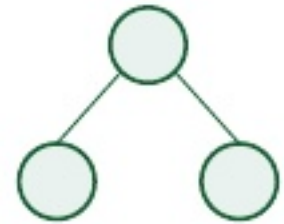
$C_i = 1$ $\quad$ $C_i = 1/2$ $\quad$ $C_i = 0$

Watts & Strogatz, Nature 1998.

Global clustering coefficient



Closed triplet



Triplet

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}$$

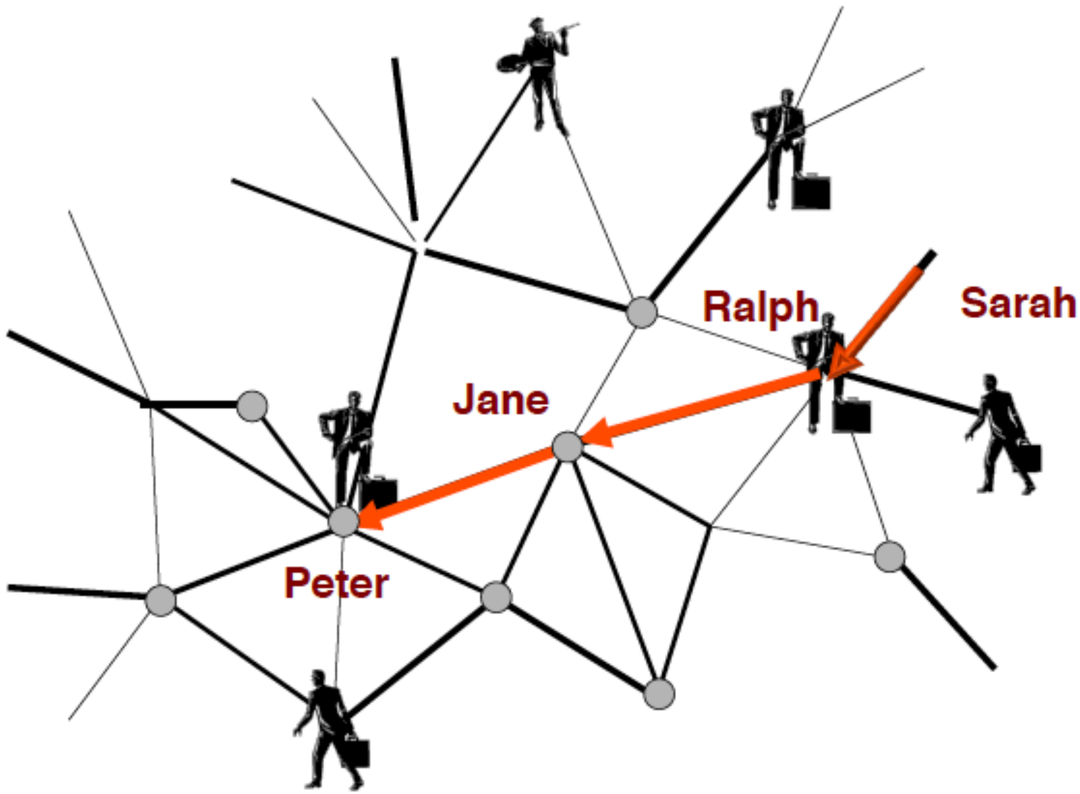
RANDOM NETWORK MODEL

Definition:

A **random graph** is a graph of N nodes where each pair of nodes is connected by probability **p**.

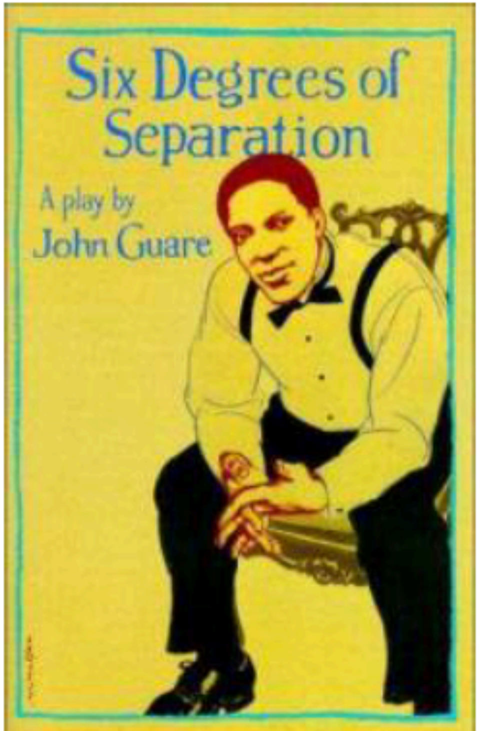
Small worlds

SIX DEGREES small worlds

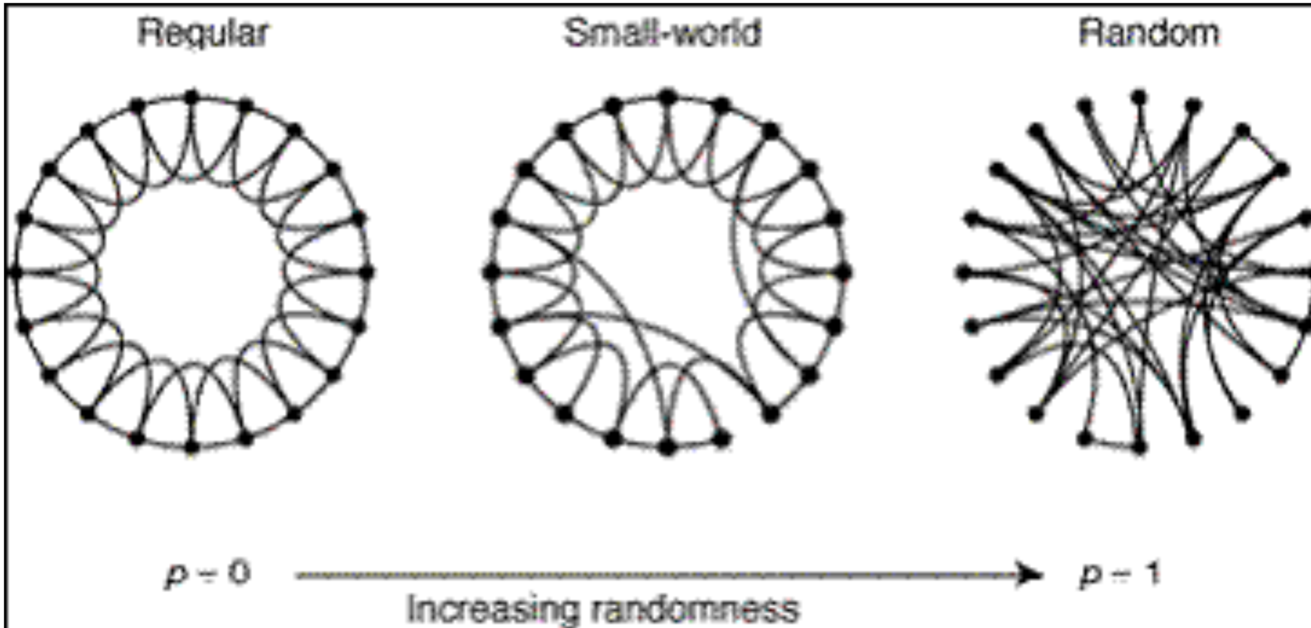


Frigyes Karinthy, 1929
Stanley Milgram, 1967

SIX DEGREES | 1991: John Guare



"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice.... It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."



Regular
Small-world
Random
p = 0
p = 1
Increasing randomness

Real networks are not random

Then what are they?

The random network model differs from real networks in two important characteristics:

Growth: While the random network model assumes that the number of nodes is fixed (time invariant), real networks are the result of a growth process that continuously increases.

Preferential Attachment: While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

Barabási & Albert, *Science* **286,** 509 (1999)

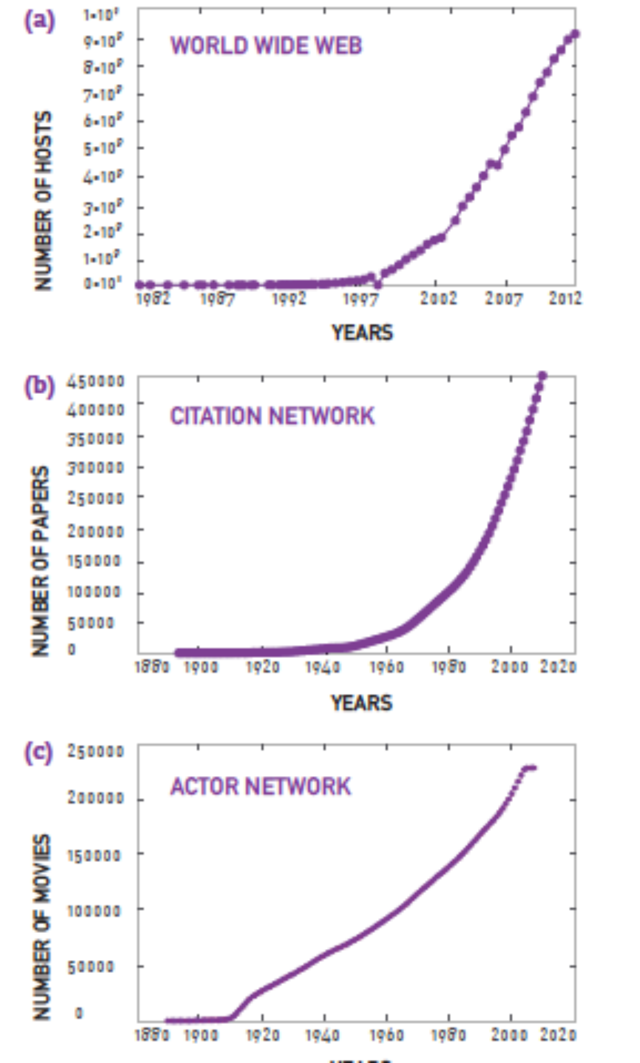
Barabasi Albert (BA) Model

BA MODEL: Growth

Previously:

the number of nodes, N, is fixed (static models)

NETWORKS EXPAND THROUGH THE ADDITION OF NEW NODES



Barabási & Albert, *Science* **286,** 509 (1999)

BA MODEL: Preferential attachment

Previously: links are added randomly to the network

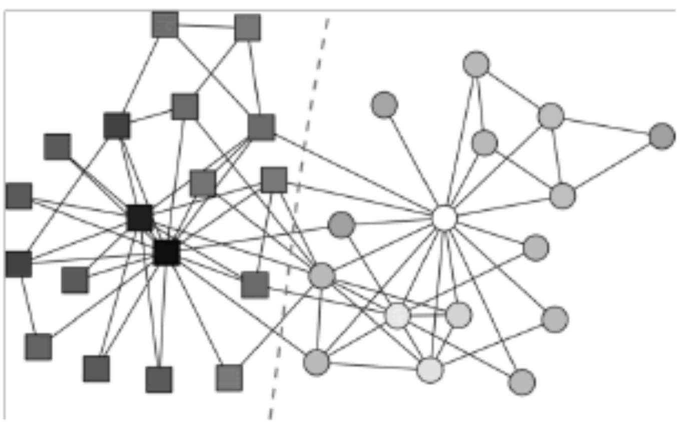
NEW NODES PREFER TO CONNECT TO THE MORE CONNECTED NODES

Barabási & Albert, *Science* **286**, 509 (1999)

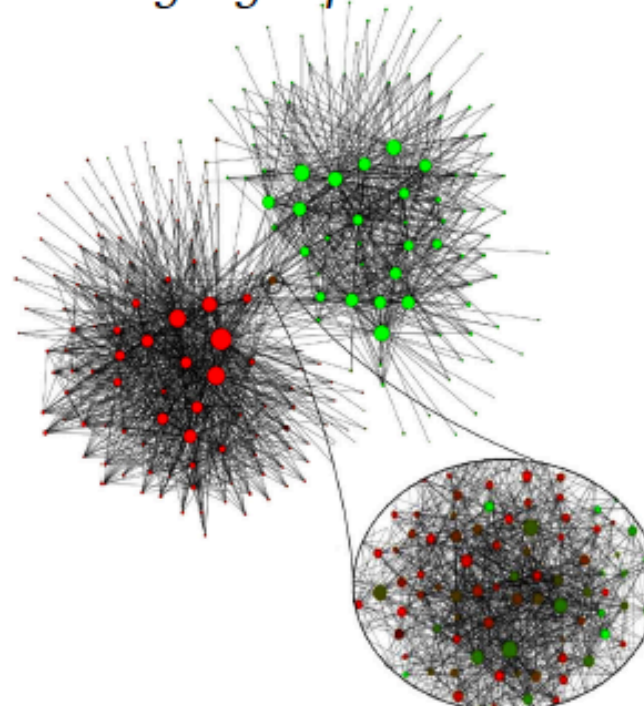
Community Detection in Networks

Examples of communities

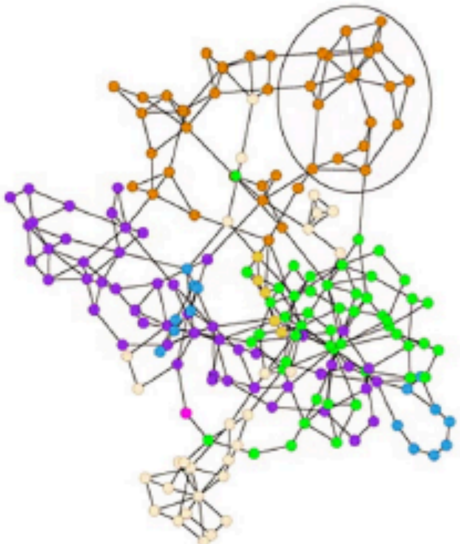
→ *Karate Club:*
Breakup of the club

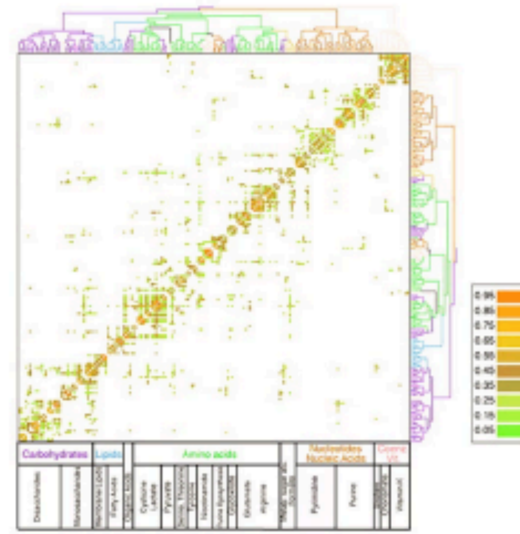


→ *Belgian Phone Data:*
Language spoken



Biological Networks





E. Ravasz et al., *Science* 297 (2002).

A.-L. Barabási, *Network Science: Communities.*

Modularity

➢ Real networks are fragmented into groups or modules

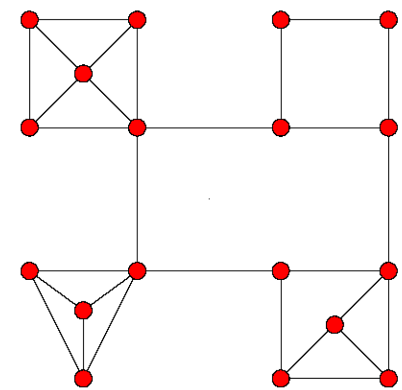
❖ Society: Granovetter, M. S. (1973) ; Girvan, M., & Newman, M.E.J. (2001); Watts, D. J., Dodds, P. S., & Newman, M. E. J. (2002).

❖ WWW: Flake, G. W., Lawrence, S., & Giles. C. L. (2000).

❖ Biology: Hartwell, L.-H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999).

❖ Internet: Vasquez, Pastor-Satorras, Vespignani(2001).

The more separable fragments (communities) in the network - the higher it's modularity



→ Modularity is a measure associated to a partition

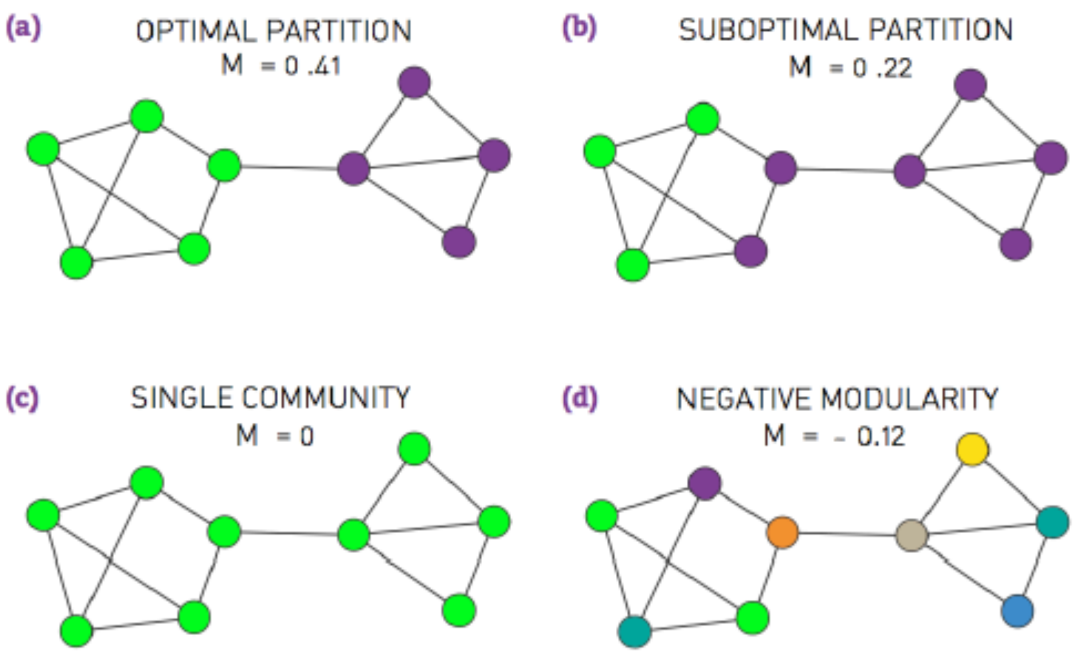
MEJ Newman, *PNAS* 103 (2006).

A.-L. Barabási, *Network Science: Communities.*

Modularity

Which partition $\{C_c,\ c = 1,\ n_c\}$ **?**

Community detection through modularity maximizing algorithms



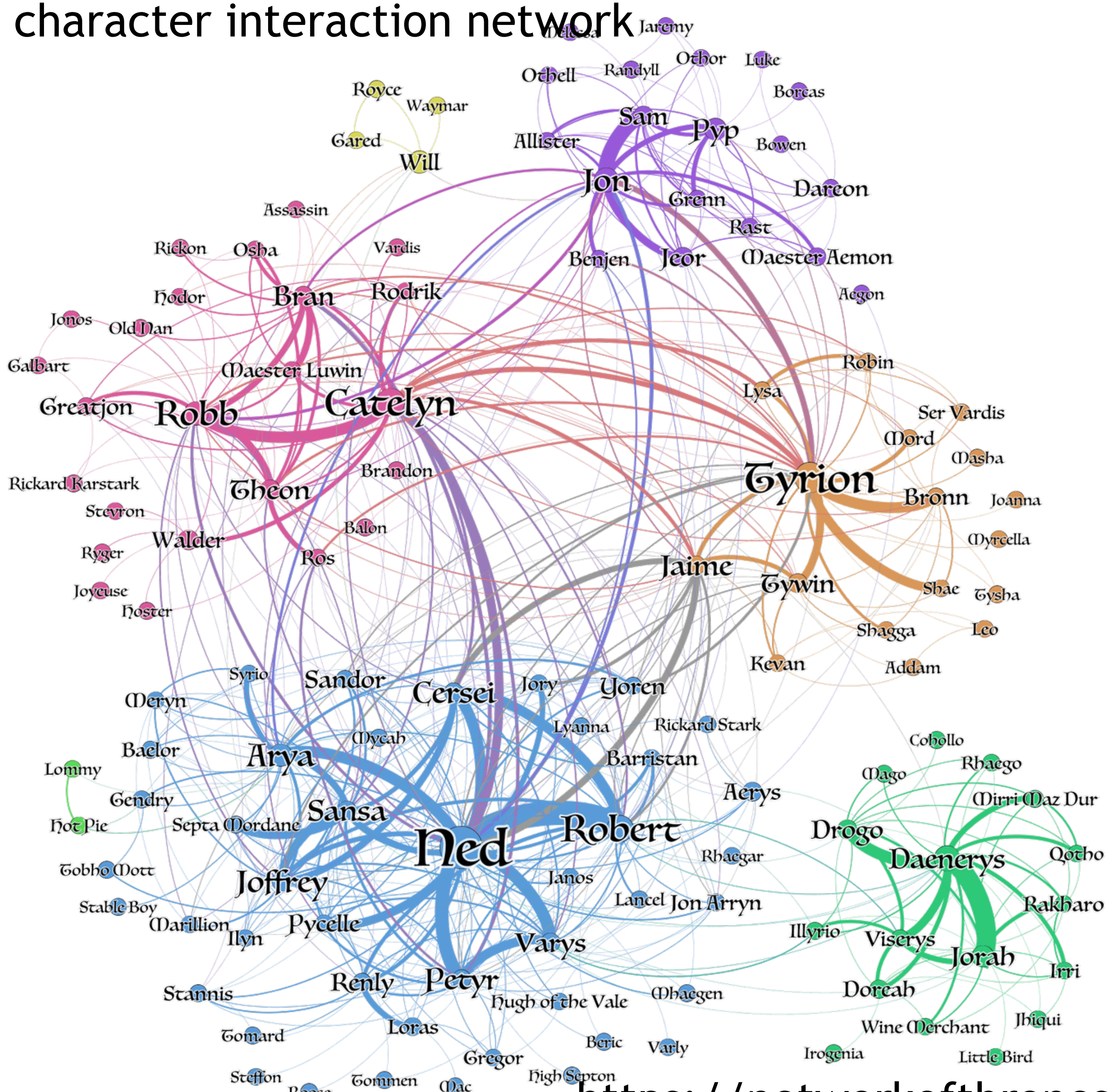
A.-L. Barabási, *Network Science: Communities.*

Check out

https://mrpandey.github.io/d3graphTheory/index.html

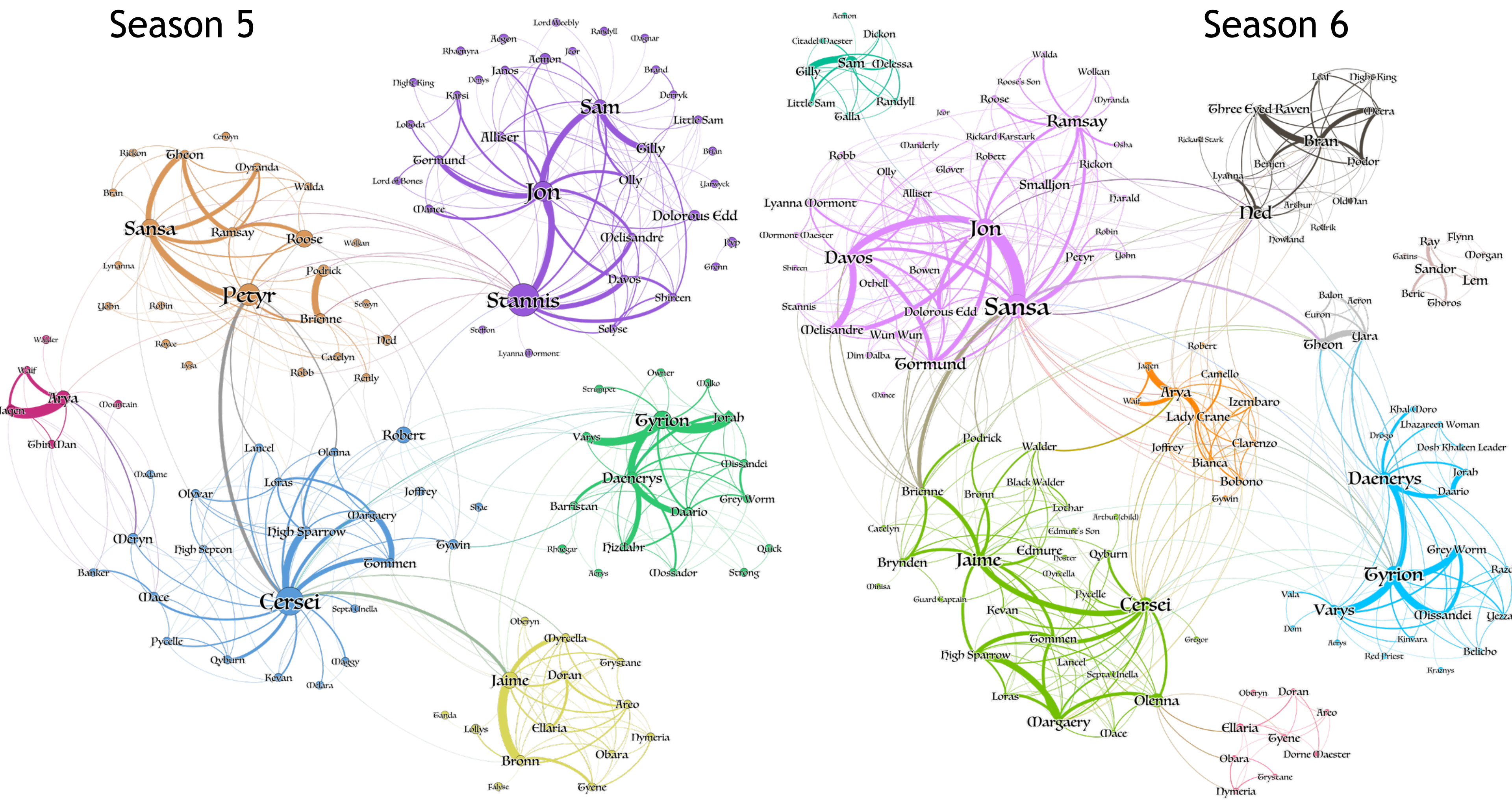
Also…

Game of Thrones character interaction network
Season 1



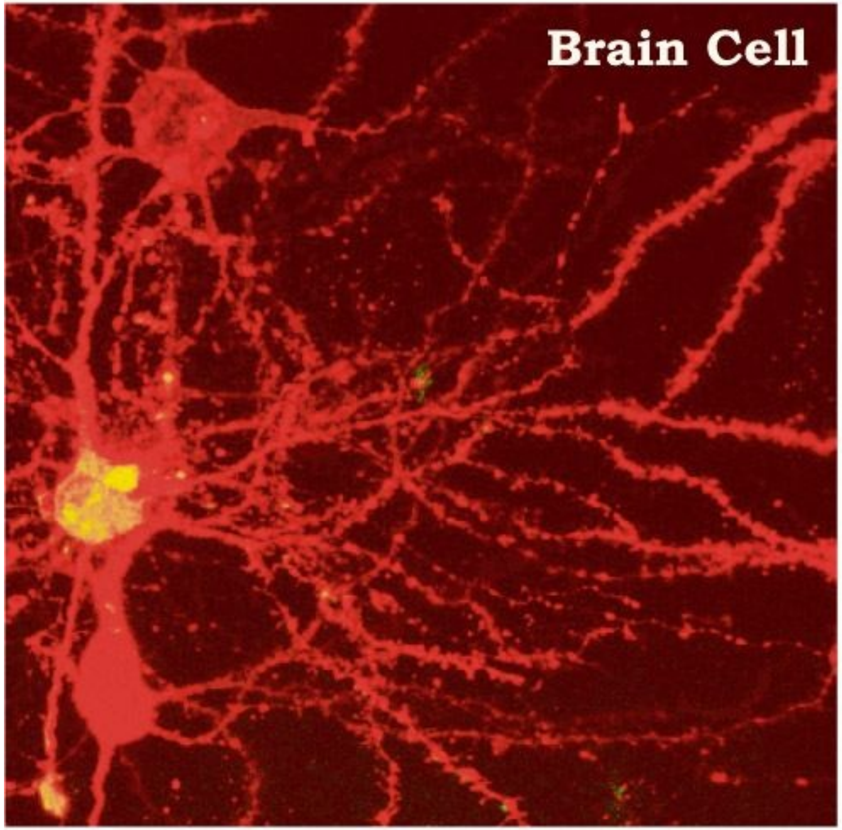
https://networkofthrones.wordpress.com/

Game of Thrones character interaction network



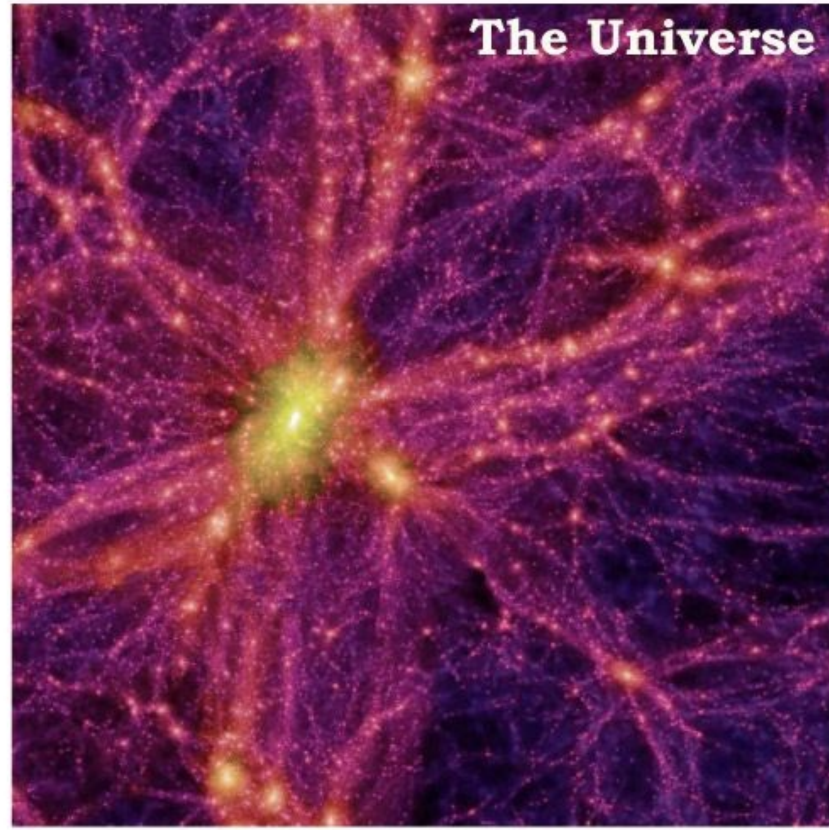
https://networkofthrones.wordpress.com/

One is only micrometers wide. The other is billions of light-years across. One shows neurons in a mouse brain. The other is a simulated image of the universe. Together they suggest the surprisingly similar patterns found in vastly different natural phenomena. *DAVID CONSTANTINE*



Mark Miller

Mark Miller, a doctoral student at Brandeis University, is researching how particular types of neurons in the brain are connected to one another. By staining thin slices of a mouse's brain, he can identify the connections visually. The image above shows three neuron cells on the left (two red and one yellow) and their connections.



Virgo Consortium

An international group of astrophysicists used a computer simulation last year to recreate how the universe grew and evolved. The simulation image above is a snapshot of the present universe that features a large cluster of galaxies (bright yellow) surrounded by thousands of stars, galaxies and dark matter (web).

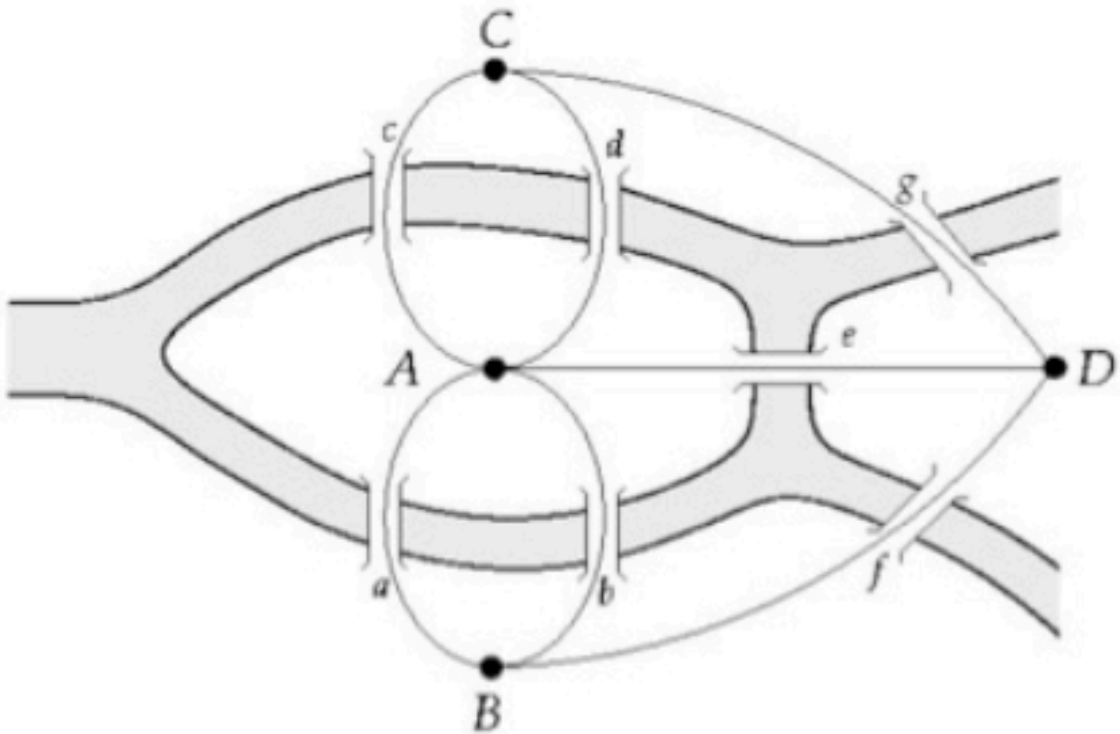
Source: Mark Miller, Brandeis University; Virgo Consortium for Cosmological Supercomputer Simulations; www.visualcomplexity.com

The New York Times

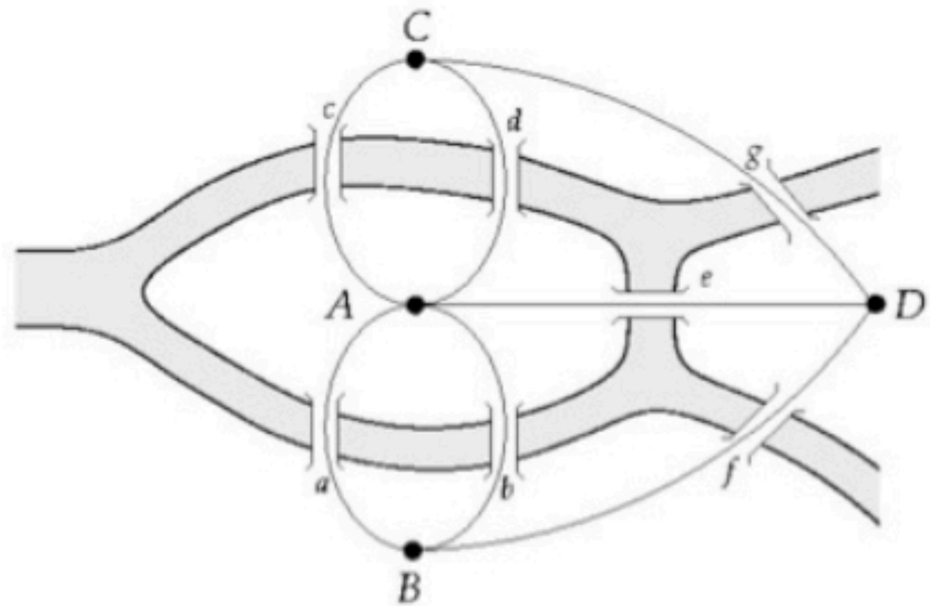

thank you
danke
謝謝
ngiyabonga
teşekkür ederim
рахмат
Баярлалаа
спасибо
mersi
barka
welalin
tack
spas
kia ora
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
mahalo
gracias
tapadh leat
хвала
asante
manana
obrigada
tenki
murakoze
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
köszönöm
hvala
grazie
bayarlalaa
enkosi
bedankt
dziękuję
akun dankon
mochchakkeram
djiere dieuf
tau
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpay
go raibh maith agat
sobodi
dėkuji
obrigado
sagolun
sukriya
kop khun krap
taiku
arigatô
takk
dakujem
trugarez
ありがとう
tanemirt
mèsi
didi madloba
najis tuke
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
merce
мерси
kam sah hamnida
rahmat
감사합니다
xiexie
ευχαριστώ
merci

THE BRIDGES OF KONIGSBERG

Can one walk across the seven bridges and never cross the same bridge twice?


C
c
d
g
e
A
D
a
b
f
B

THE BRIDGES OF KONIGSBERG



Can one walk across the seven bridges and never cross the same bridge twice?

1735: Euler's theorem:

(a) If a graph has more than two nodes of odd degree, there is no path.

(b) If a graph is connected and has no odd degree nodes, it has at least one path.

References:

For slides:

https://www.barabasilab.com/course