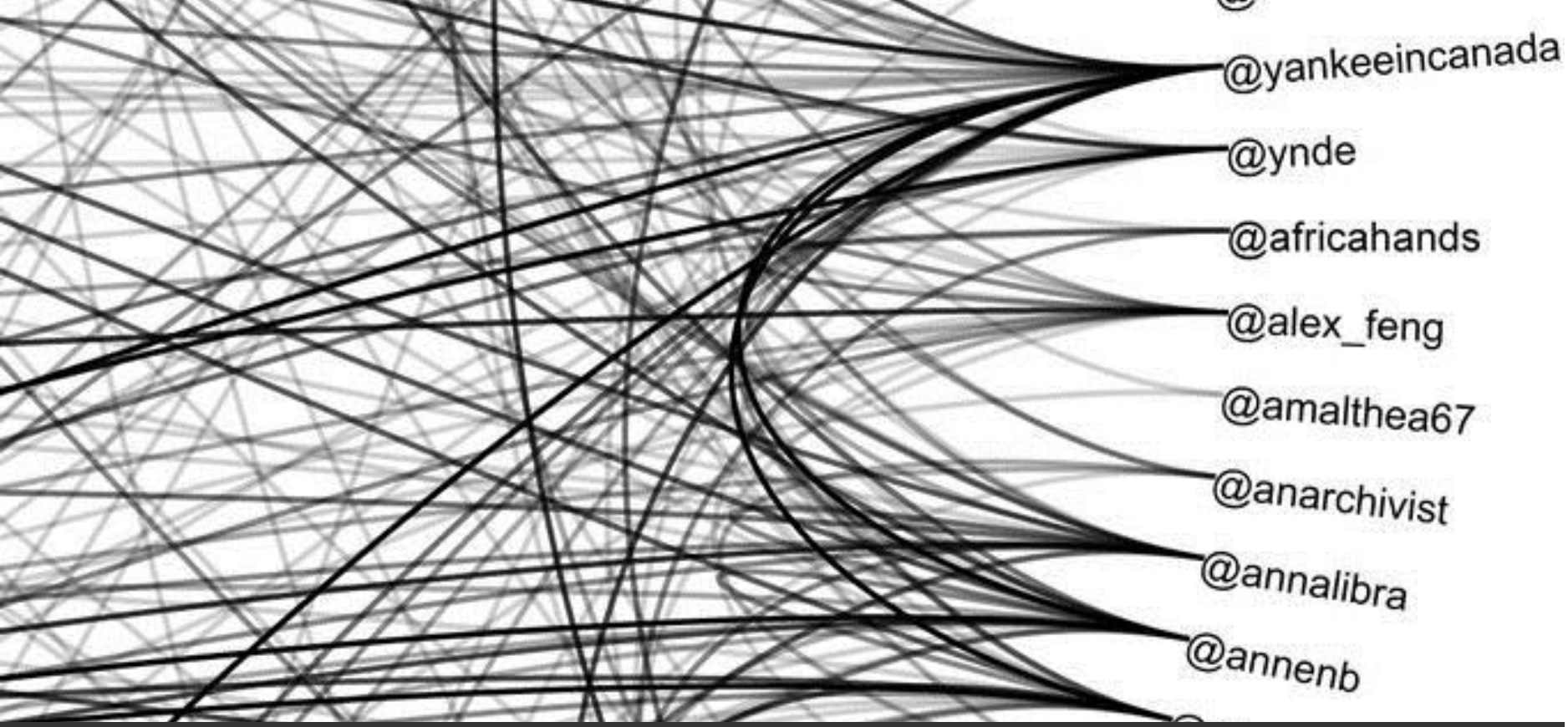


Introduction to networks

Pablo A. Haya Coll // Manuel García-Herranz

pablo.haya@iic.uam.es

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

What is a graph?

- **Graph:** nodes (nodes, vertices, agents, actors, players) and edges (links, ties) connections between nodes

Types of Graph

- Simple Graph:
  - nodes: same type, edges**: symmetric**, **binary**
- Directed Graphs
  - nodes: same type, edges**: asymmetric**
- Weighted Graphs
  - nodes: same type, edges**: numeric**
- Bi-Partite Graph and projections
  - nodes: **different type**

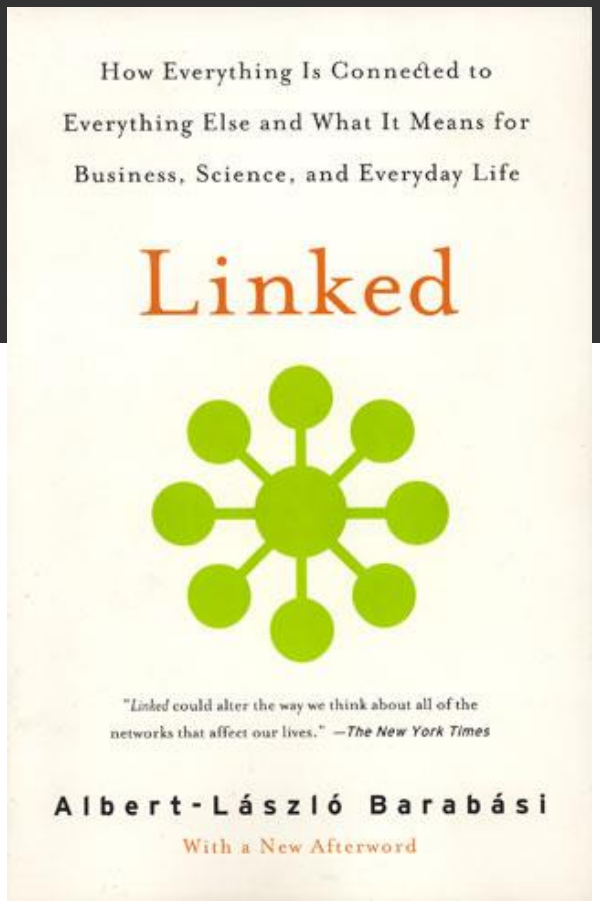
R Scripting

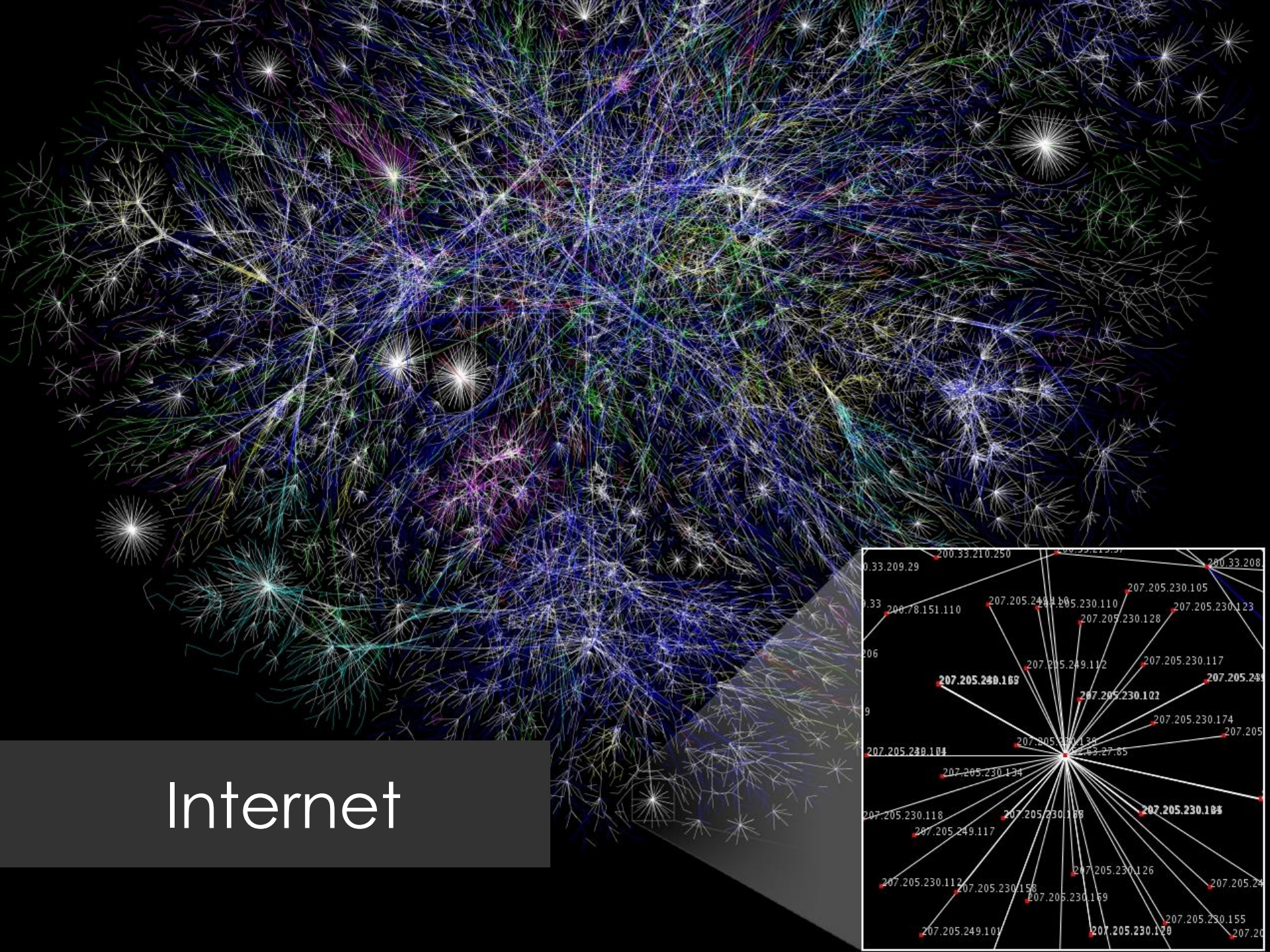
What is a network?

- **Graph:** nodes (nodes, vertices, agents, actors, players) and edges (links, ties) connections between nodes

- **Complex system:** large number of parts, properties of parts are heterogeneously distributed, adaptability, evolvability
  - (1+1 > 2) "The Whole is Greater than the sum of its parts" Aristotle

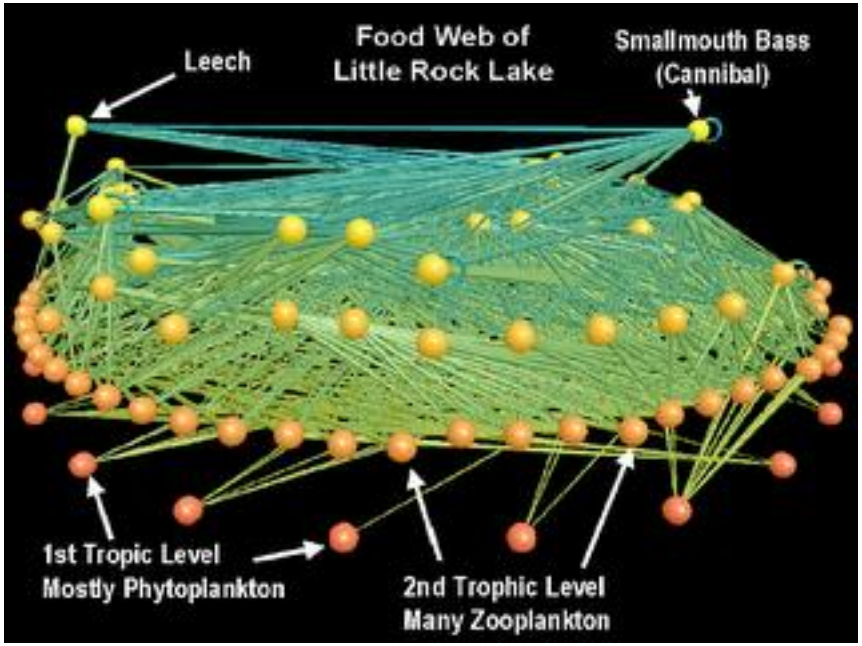
Network types

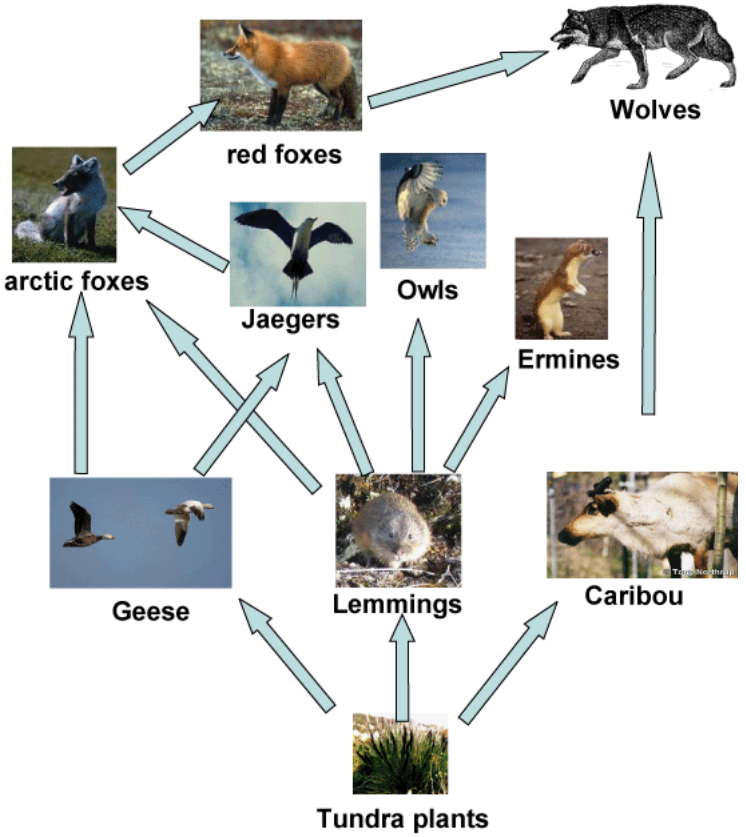



Internet

Transportation networks

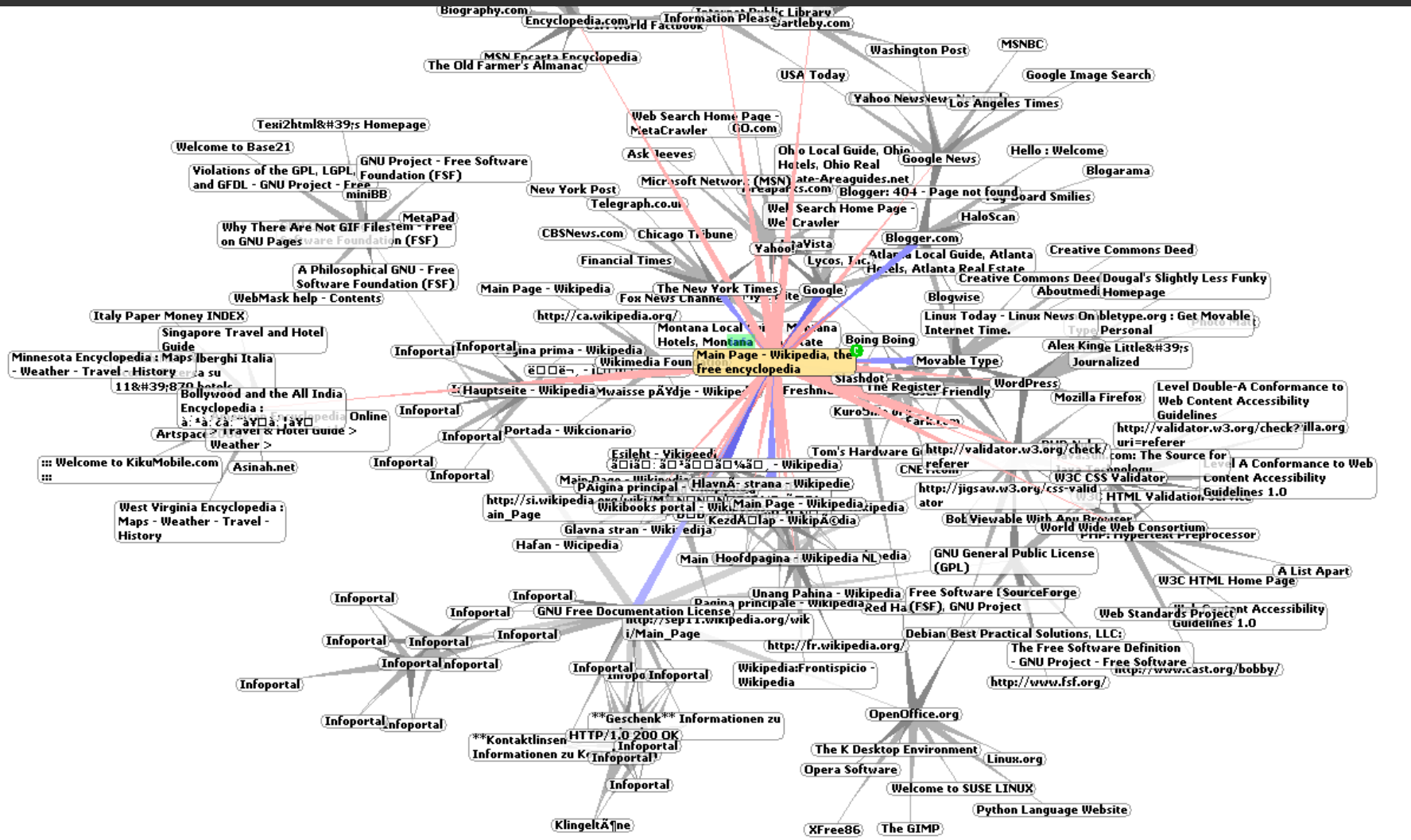
Ecosystems



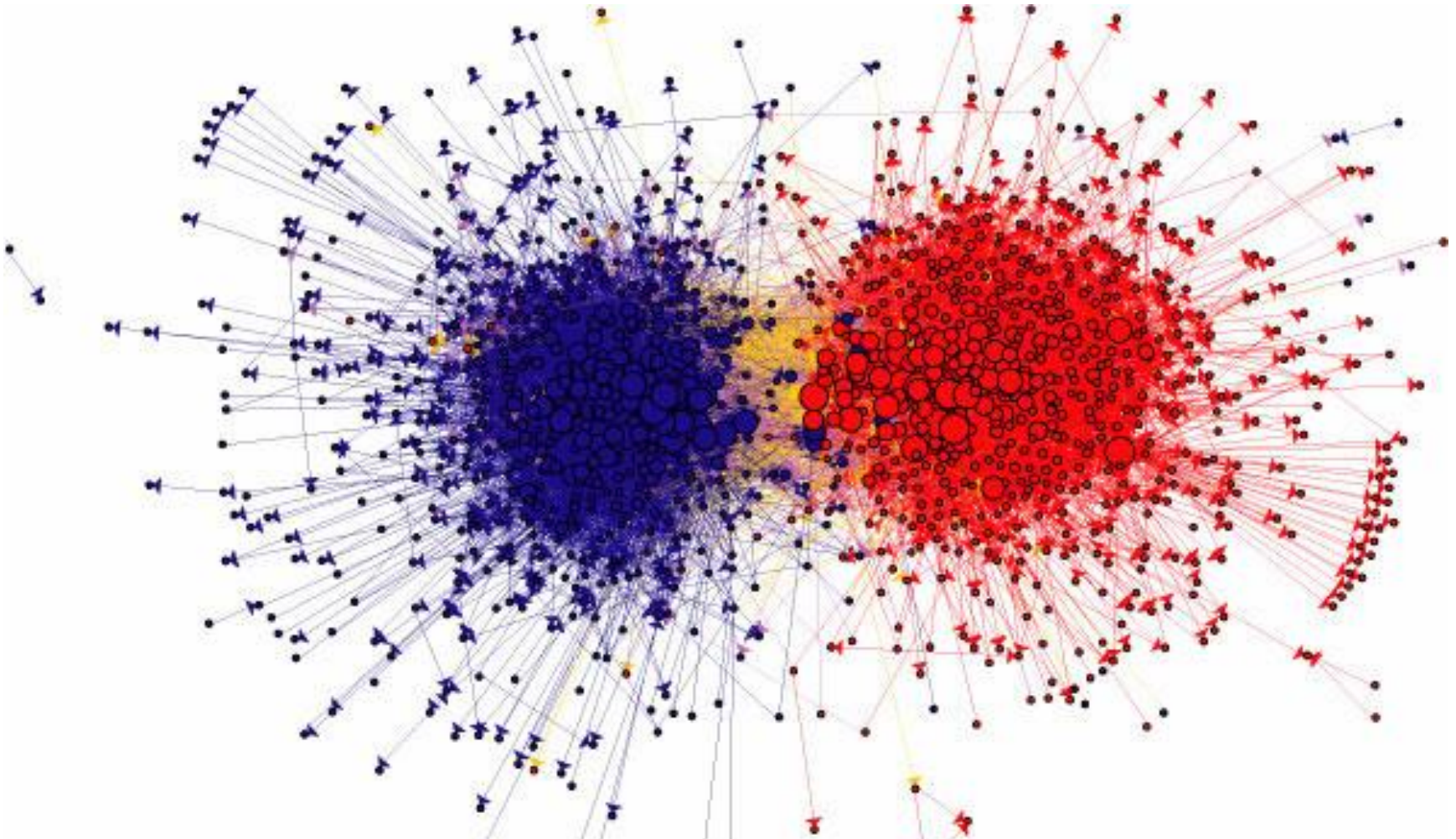


RJ Williams, L Martinez.Simple rules yield complex food webs - Nature, 2000

The World Wide Web



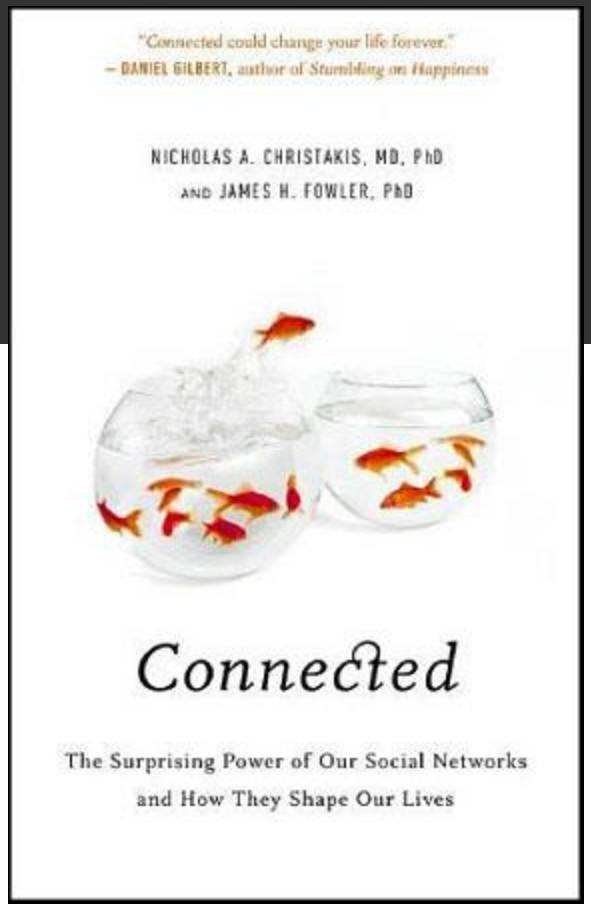
Social media



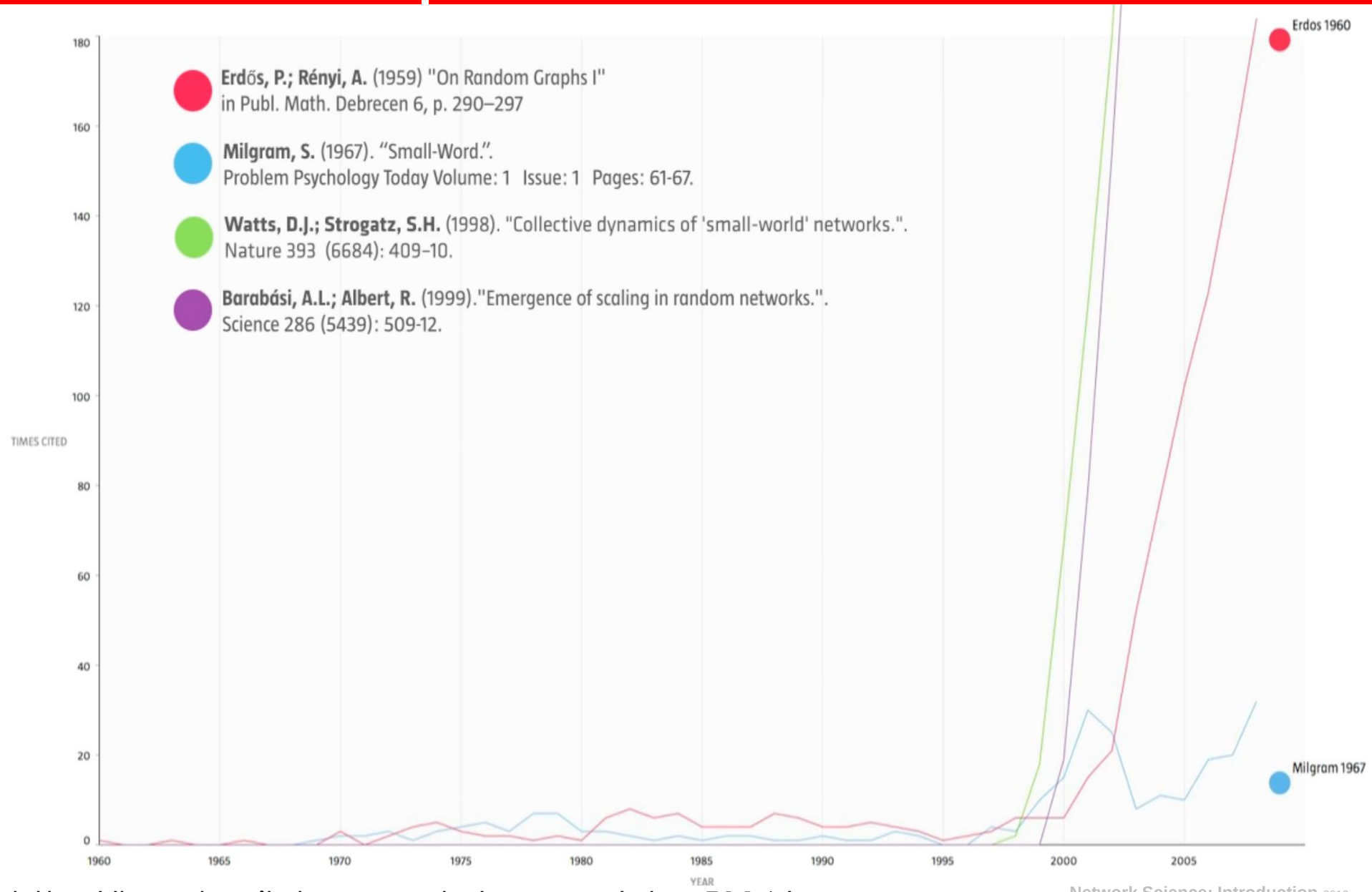
The Political Blogosphere and the 2004 U.S. Election: Divided They Blog

Lada A. Adamic and Natalie Glance, LinkKDD-2005

Social networks



NETWORK SCIENCE | The science of the 21st century



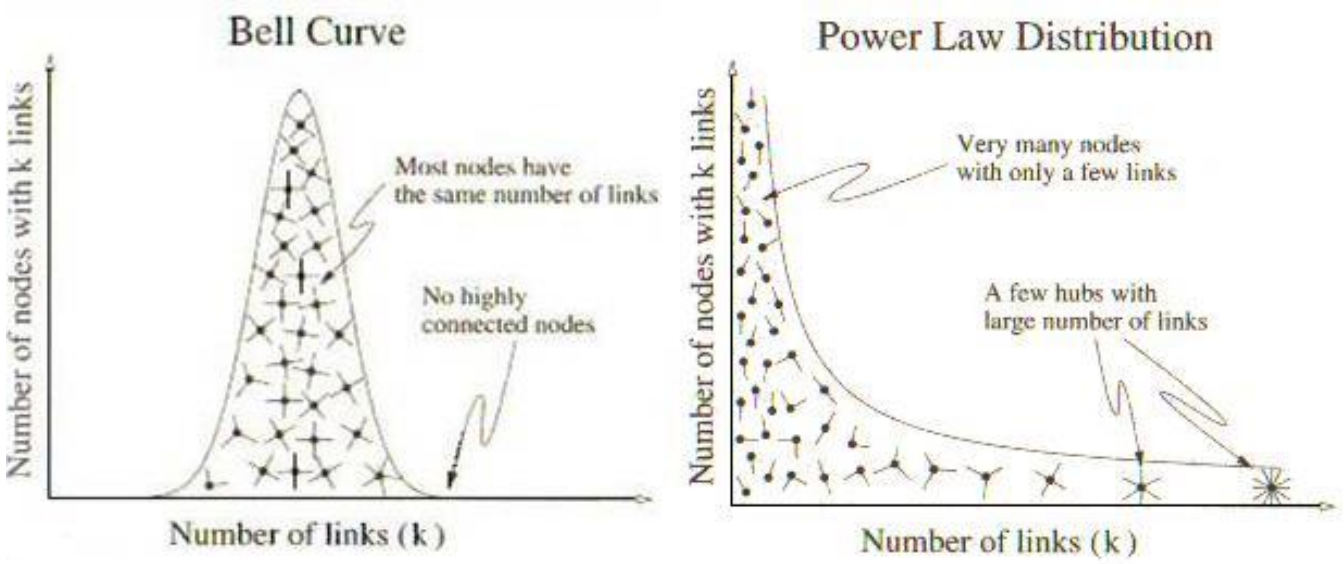
http://barabasilab.neu.edu/courses/phys5116/

Networks properties

How we are?

Rich gets richer

Scale-Free / Power law

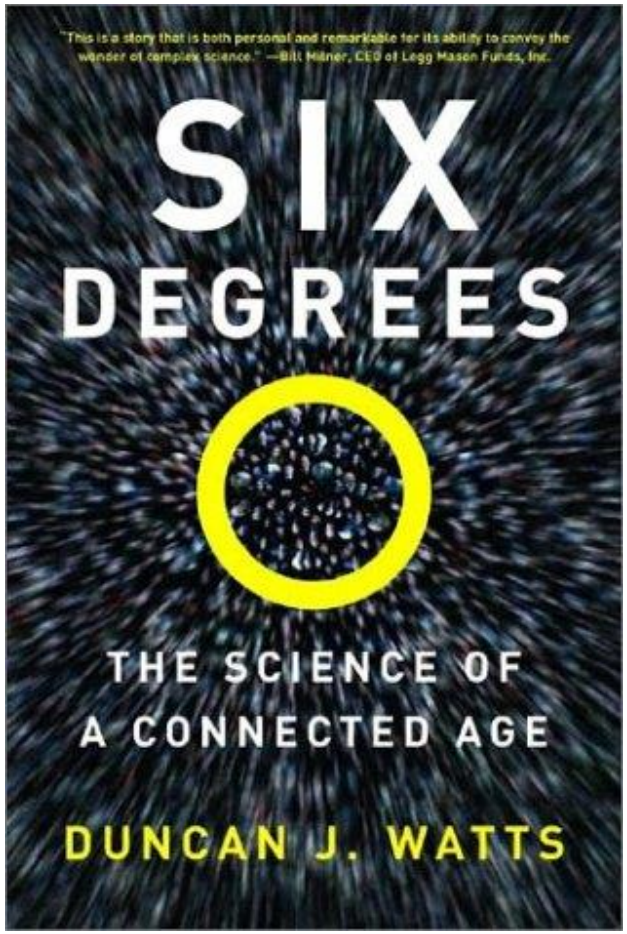


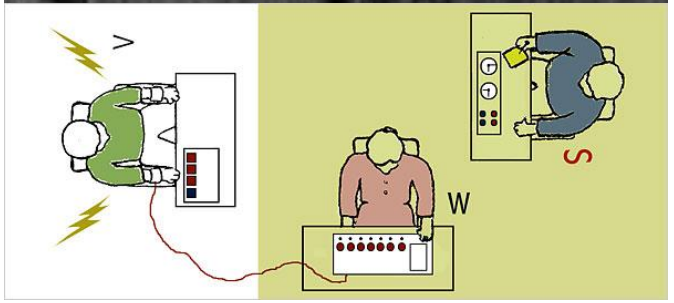
- Networks expand continuously by adding vertices
- New vertices attach preferentially to well connected sites

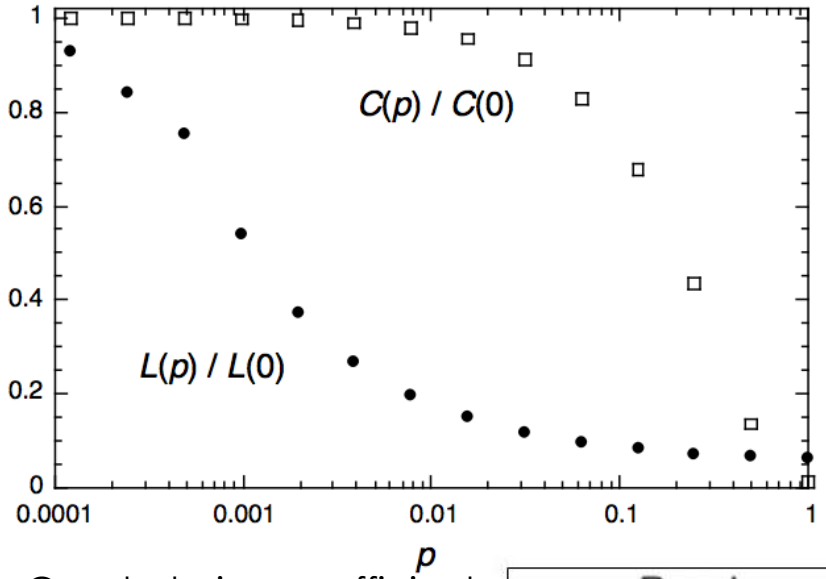
A-L Barabási, et al. Emergence of Scaling in Random Networks. Science 286, 509 (1999)

R Scripting

Stanley Milgram



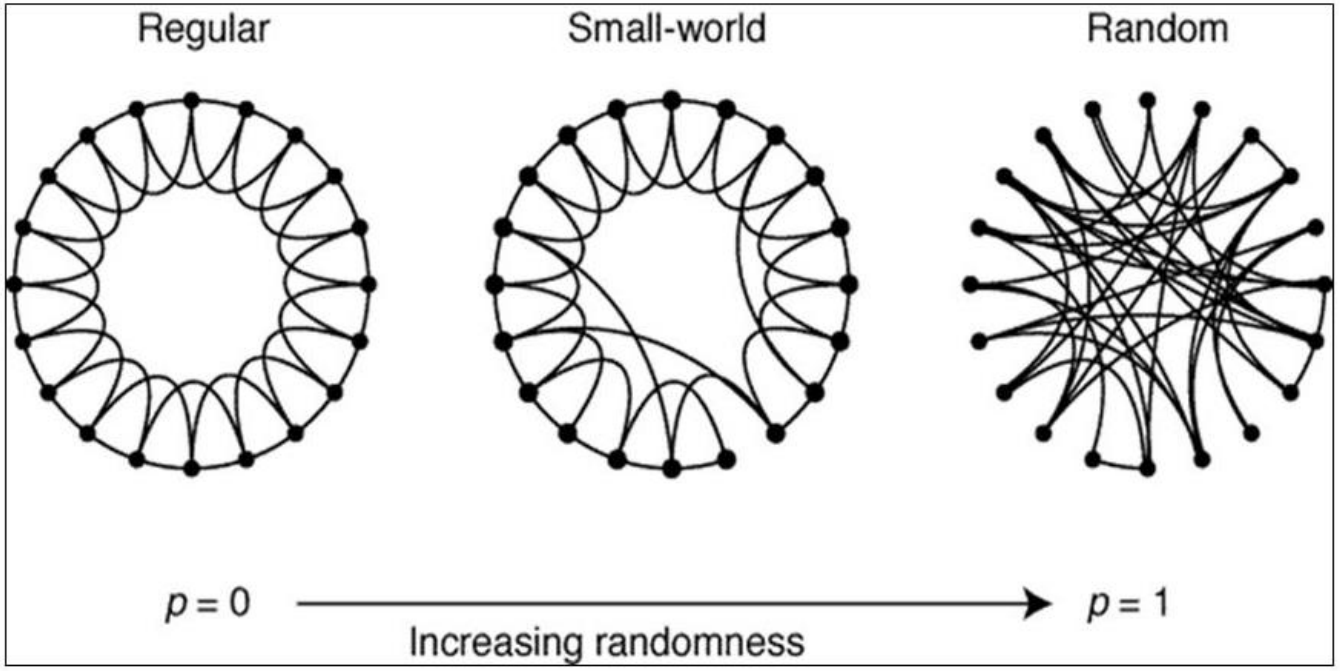




Small World

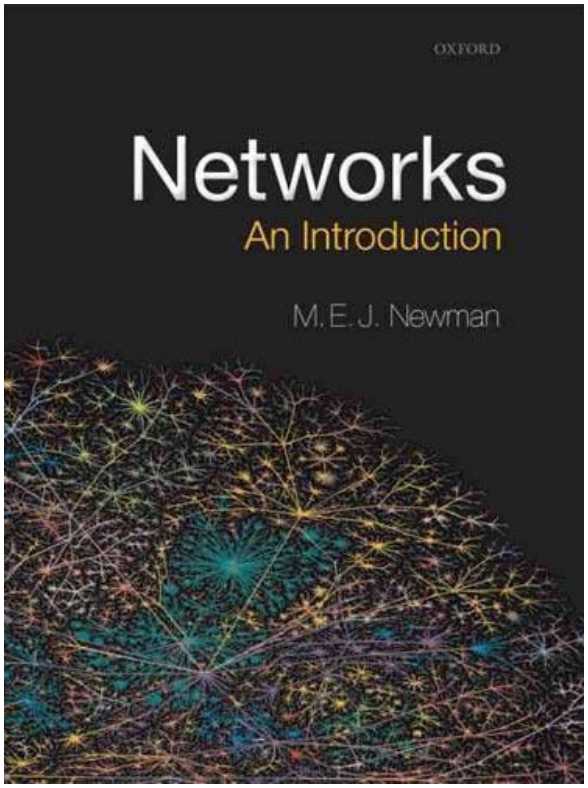
Information/infections travel faster

C = clustering coefficient
L = path length



DJ Watt, SH Strogatz -Collective dynamics of 'small-world' networks - Nature, 1998

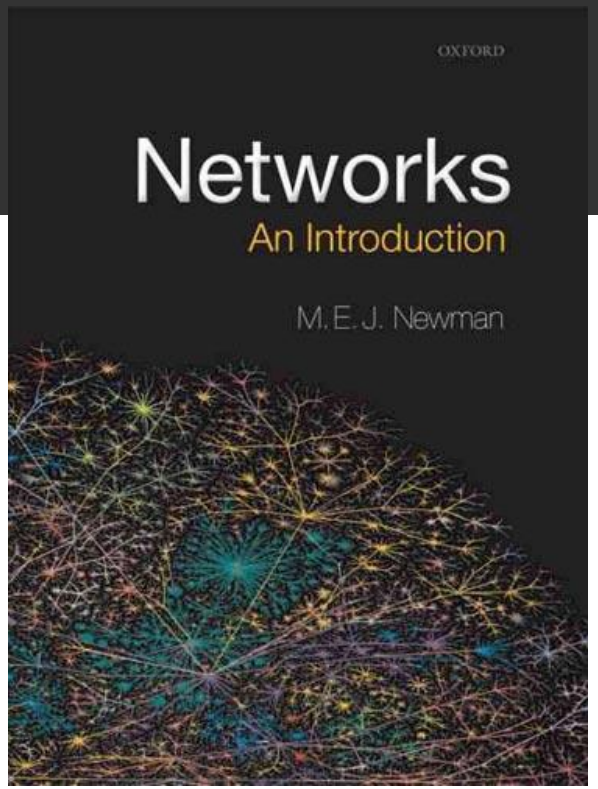
R Scripting



Network metrics

Network metrics

What's important?



CENTRALITY METRICS

DEGREE CENTRALITY

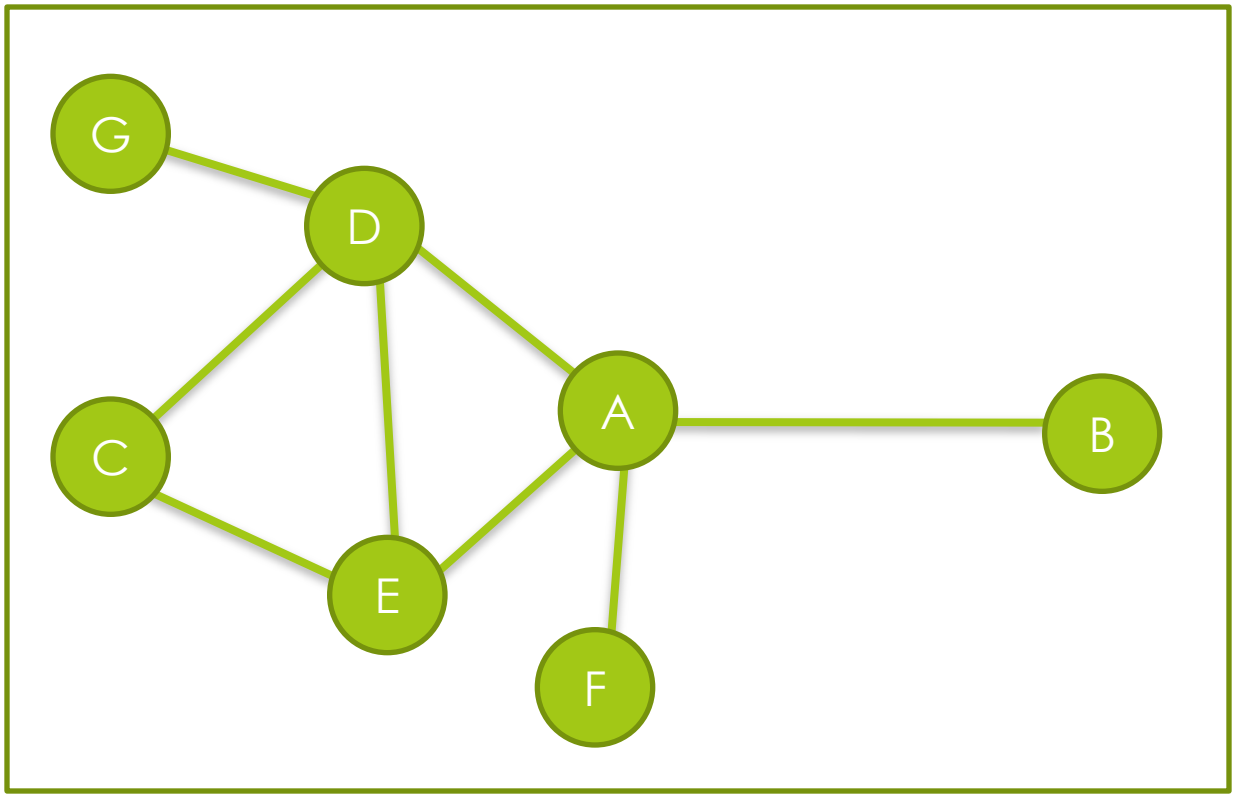
k = number of links

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

where A_{ij} = 1 if nodes *i* and *j* are connected and 0 otherwise

Degree Sequence
(the sequence of degrees for each node in the network)

A = 4
B = 1
C = 2
D = 4
E = 3
F = 1
G = 1



The Oracle of Bacon

http://oracleofbacon.org/

Ronald Reagan (I) has a Miguel Gila number of 3.

Find a different link

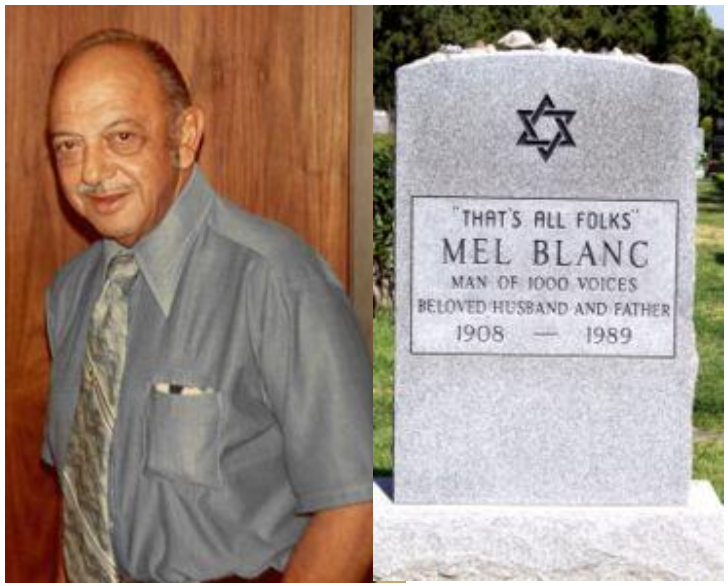
Ronald Reagan (I)
was in
This Is the Army (1943)
with
Gary Merrill
was in
Amarsi male (1969)
with
Fernando Di Leo
was in
Alejandra, mon amour (1979)
with
Miguel Gila

Miguel Gila to Ronald Reagan (I) Find link More options >>

Are the Actors Featured in Most Films the more Central?

(measured in the late 90's)

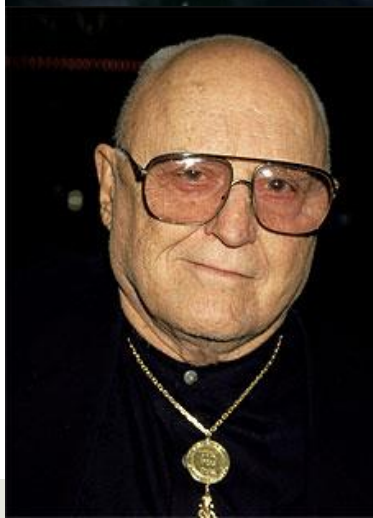
Mel Blanc 759
Tom Byron 679
Marc Wallice 535
Ron Jeremy 500
Peter North 491
TT Boy 449
Tom London 436
Randy West 425
Mike Horner 418
Joey Silvera 410



XXX

A-L Barabasi, "Linked", 2002

Steiger, Rod (2.678695)
Lee, Christopher (I) (2.684104)
Hopper, Dennis (2.698471)
Sutherland, Donald (I) (2.701850)
Keitel, Harvey (2.705573)
Pleasence, Donald (2.707490)
von Sydow, Max (2.708420)
Caine, Michael (I) (2.720621)
Sheen, Martin (2.721361)
Quinn, Anthony (2.722720)
Heston, Charlton (2.722904)
Hackman, Gene (2.725215)
Connery, Sean (2.730801)
Stanton, Harry Dean (2.737575)
Welles, Orson (2.744593)
Mitchum, Robert (2.745206)
Gould, Elliott (2.746082)
Plummer, Christopher (I) (2.746427)
Coburn, James (2.746822)
Borgnine, Ernest (2.747229)

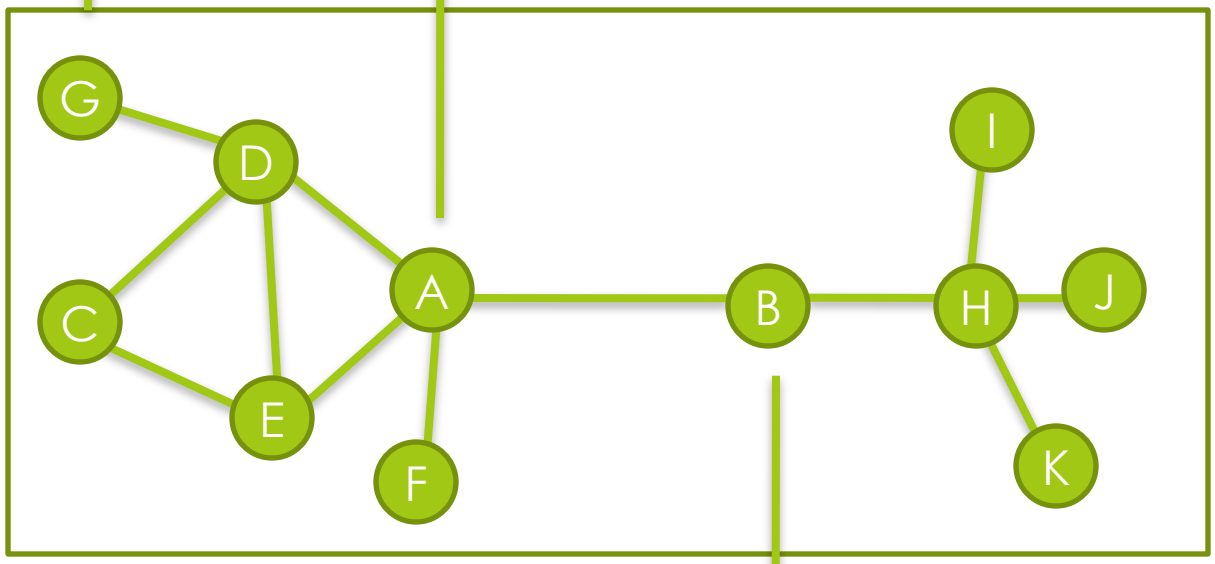


CLOSENESS CENTRALITY

C = average distance to neighbors

C(G)=1/10 (1*1+2*3+3*2+4*1+5*3)
C(G)=3.2

C(A)=1/10 (1*4+2*3+3*3)
C(A)=1.9

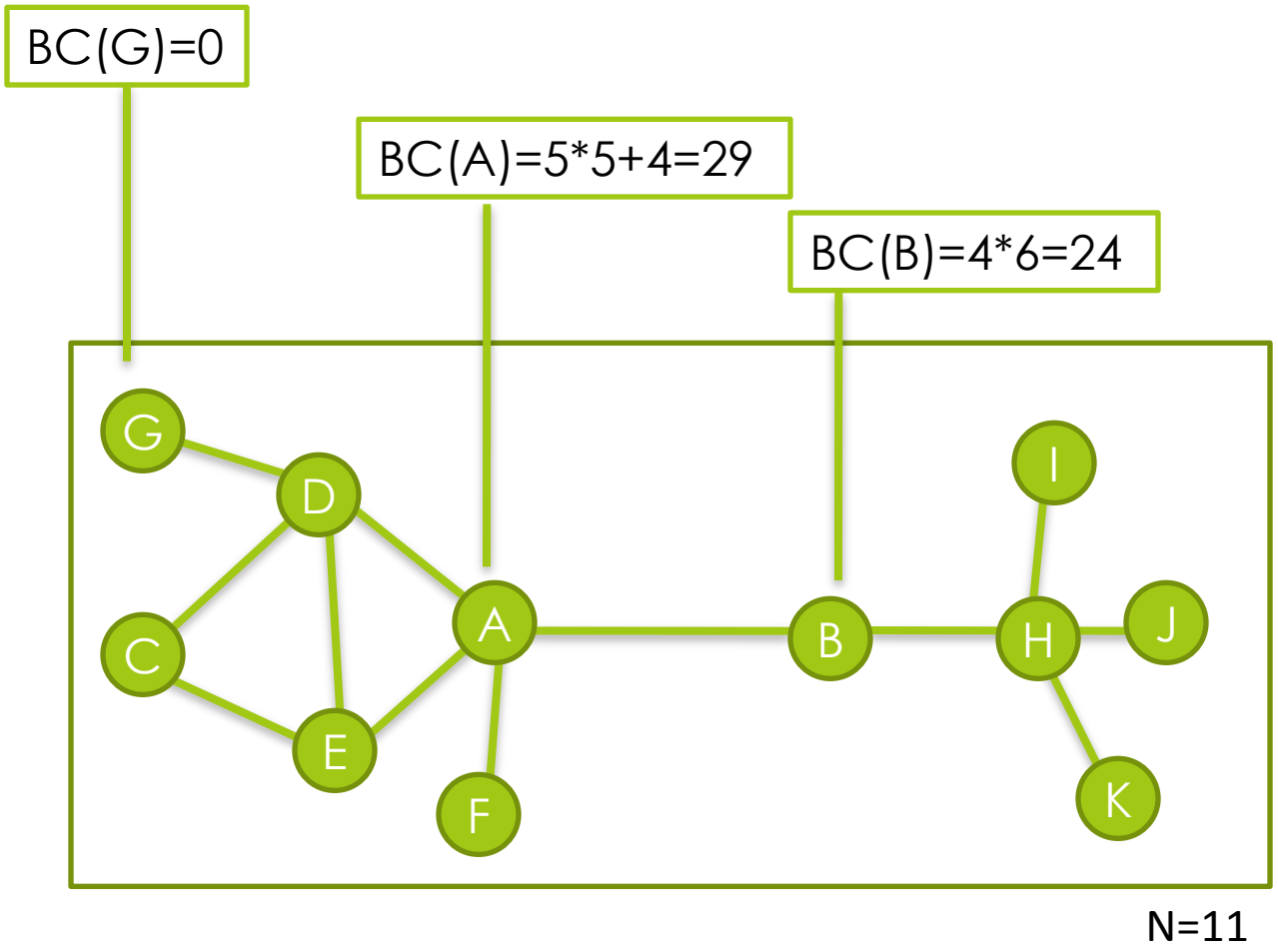


N=11

C(B)=1/10 (1*2+2*6+3*2)
C(B)=2

BETWENNESS CENTRALITY

BC= number of shortest paths that go through a node.



EIGENVECTOR CENTRALITY

Consider the Adjacency Matrix A_{ij} = 1 if node *i* is connected to node *j* and 0 otherwise. Now, measure the centrality of a node, as the sum over the centralities of all nodes….

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x_j$$

This is equivalent to eigenvalue problem:

$A\mathbf{x}=\lambda\mathbf{x}$

Then the eigenvector centrality of node (i) is defined as:

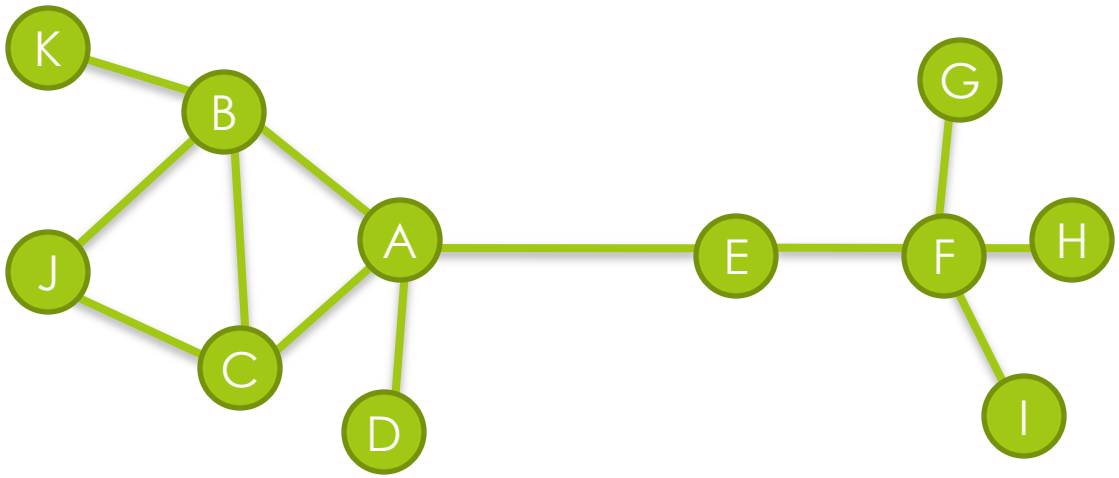
x_i

where λ is the largest eigenvalue associated with **A** and x is its associated eigenvector.

Xi can be large either because a vertex has many neighbors or because it has important neighbors or both

PAGE RANK

PR=Probability that a random walker with interspersed Jumps would visit that node.
PR=Each page votes for its neighbors.



PR(A)=PR(B)/4 + PR(C)/3 + PR(D)+PR(E)/2
A random surfer eventually stops clicking

PR(X)=(1-d)/N + d(ΣPR(y)/k(y))

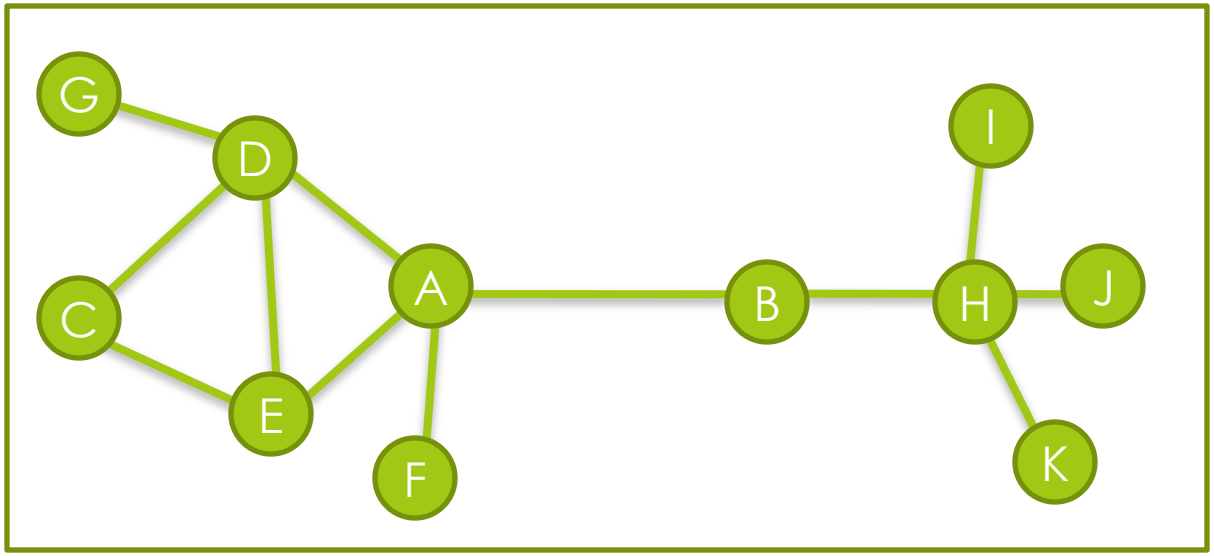
CLUSTERING METRICS

- Measure the density of a group of nodes in a Network

Clustering Coefficient, Transitivity

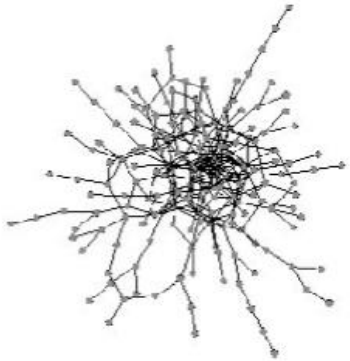
$$C_i = 2\Delta / k(k-1)$$

$C_A = 2/12 = 1/6$ $C_C = 2/2 = 1$ $C_E = 4/6 = 2/3$

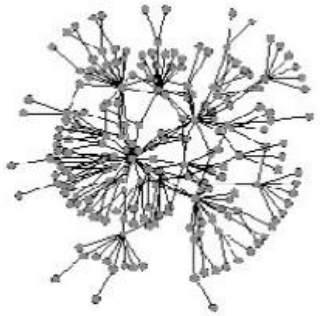


Connectivity

- Assortative: pairing of equals



- Disassortative: pairing of different



- Transitivity: probability that two neighbors of i are connected between them

R Scripting