Introduction to Network Analysis

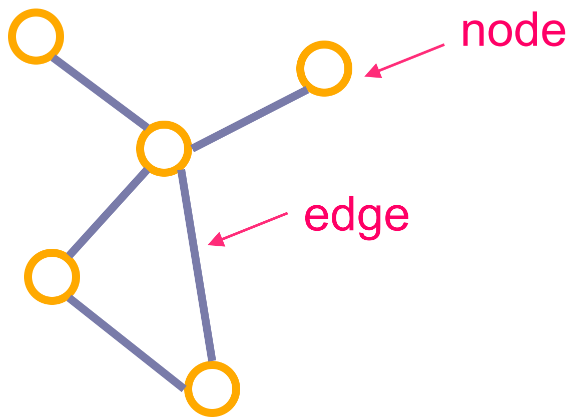
Some materials adapted from Lada Adamic, UMichigan

Outline

- the role of networks in life, nature, and research
- why model networks: structure & dynamics
  - models (structure):
    - Erdos-Renyi random graph
    - Watts-Strogatz small world model
    - Barabasi-Albert scale-free networks
  - implications (dynamics):
    - diffusion of disease and information
    - search by navigating the network
    - resilience
    - IR applications

What are networks?

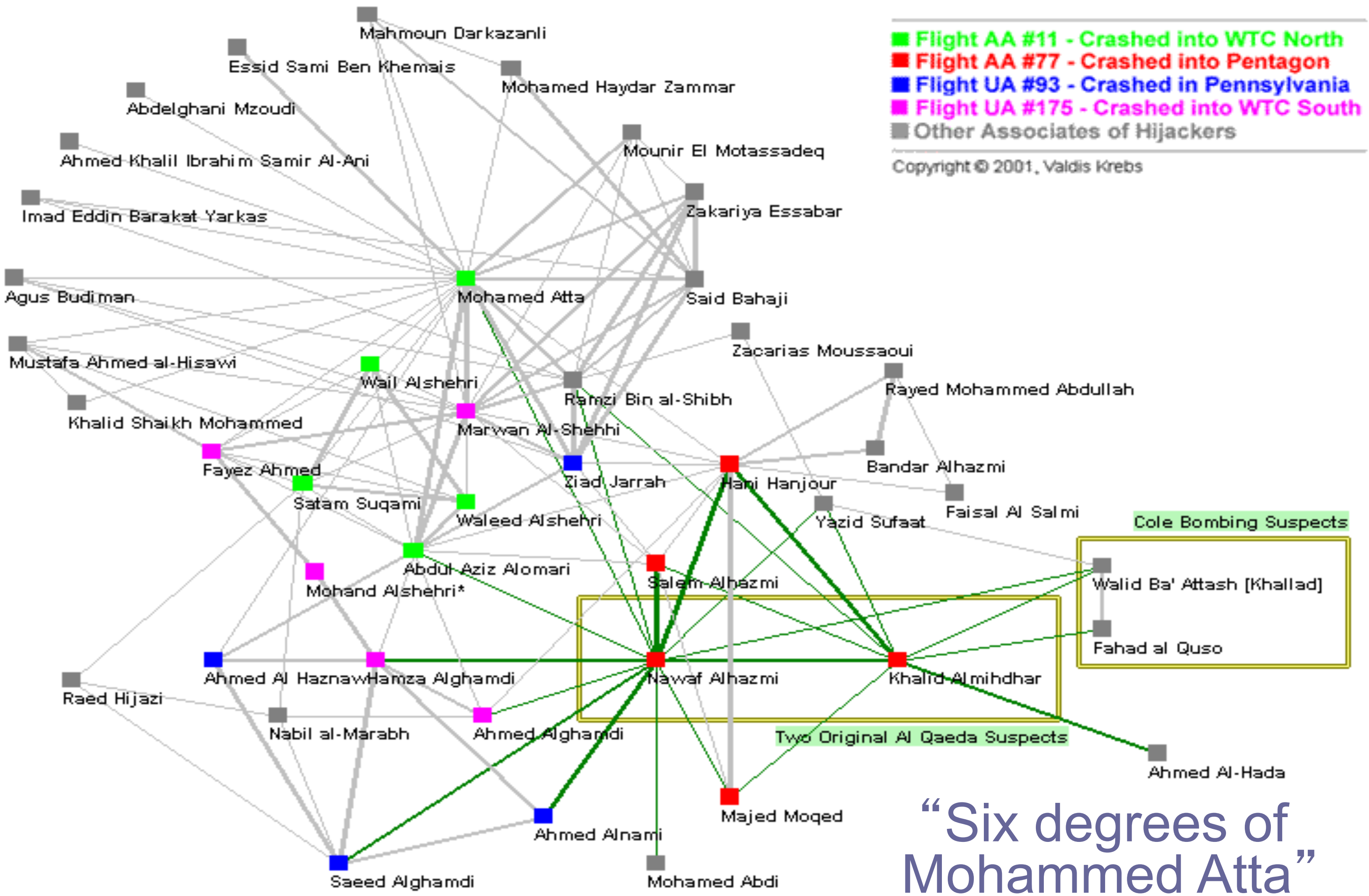
- Networks are collections of points joined by lines.



"Network" ≡ "Graph"

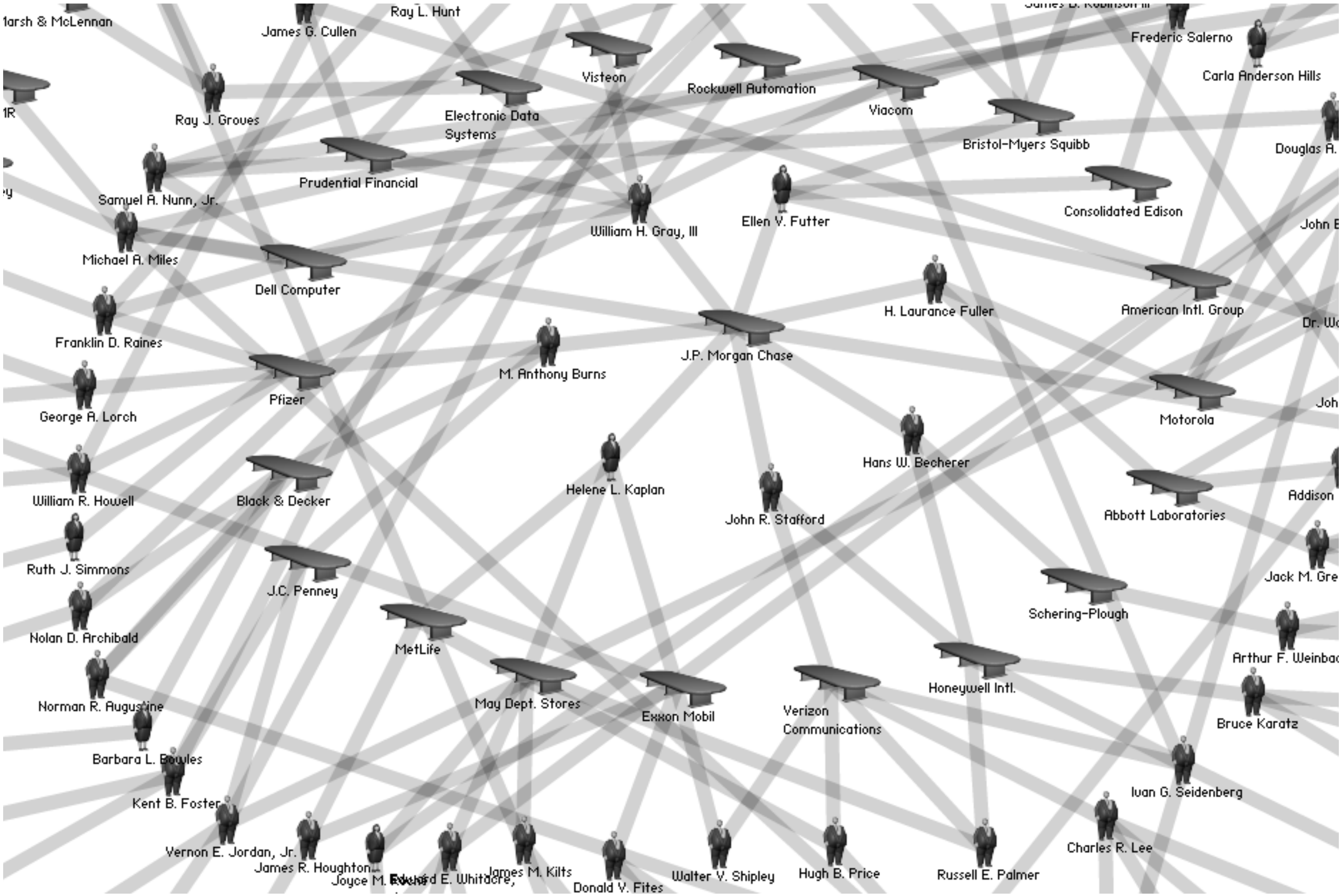
points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

examples: terrorist networks



"Six degrees of Mohammed Atta" Uncloaking Terrorist Networks, by Valdis Krebs

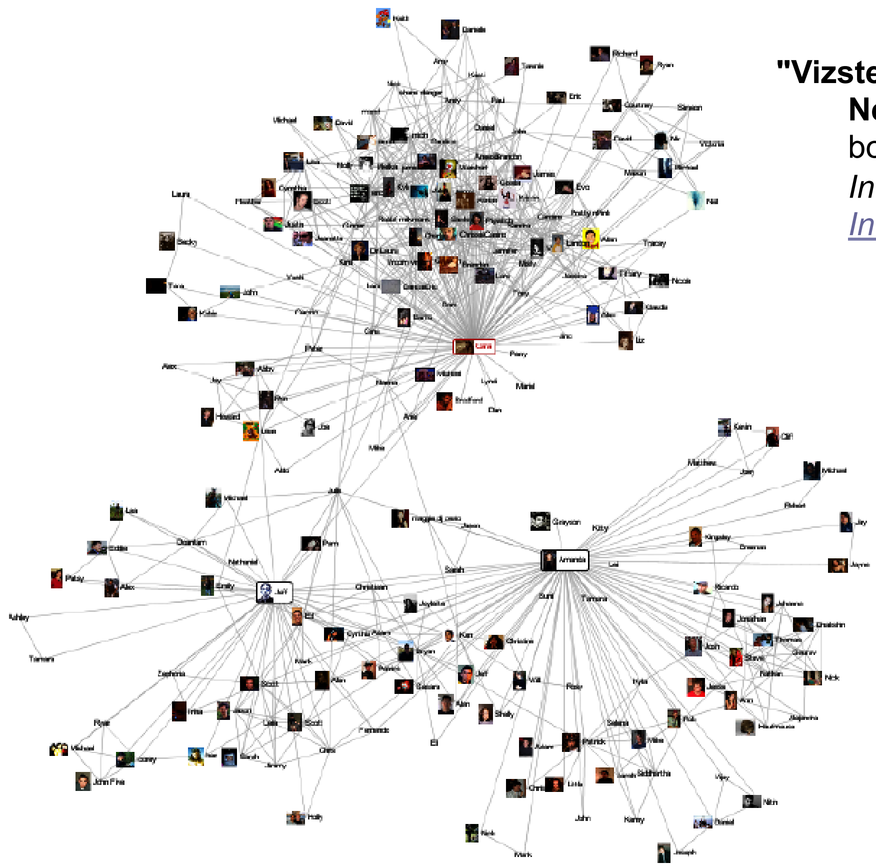
examples: boards of directors



Source: http://theyrule.net

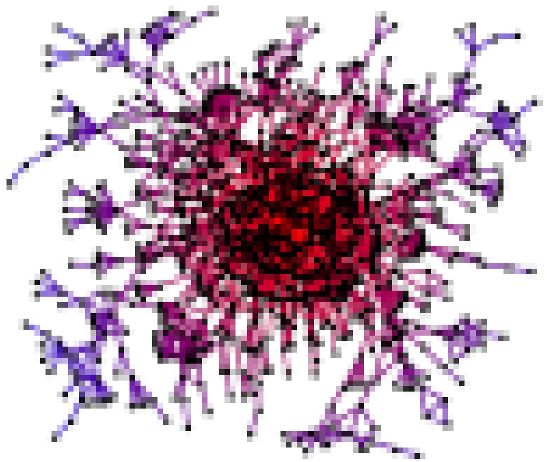
examples: online social networks

- Friendster

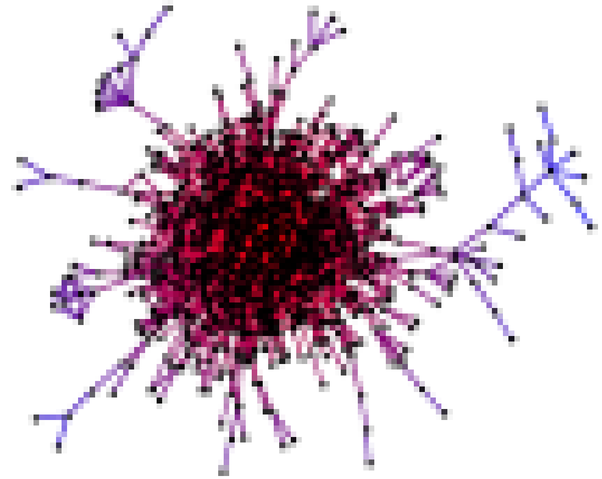


"Vizster: Visualizing Online Social Networks." Jeffrey Heer and danah boyd. *IEEE Symposium on Information Visualization (*[*InfoViz 2005*]*)*.

examples: Networks of personal homepages



Stanford



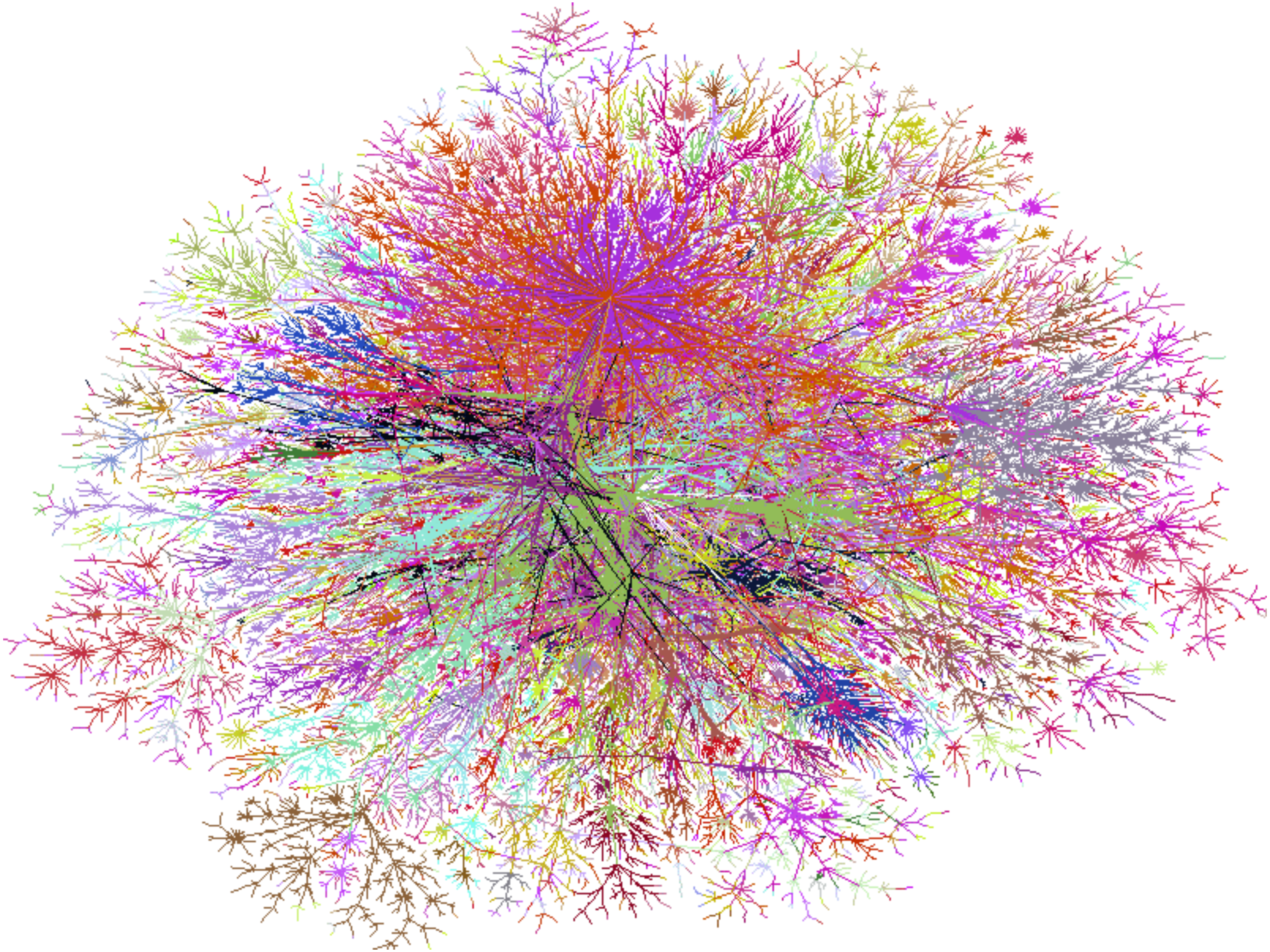
MIT

homophily: what attributes are predictive of friendship?

group cohesion

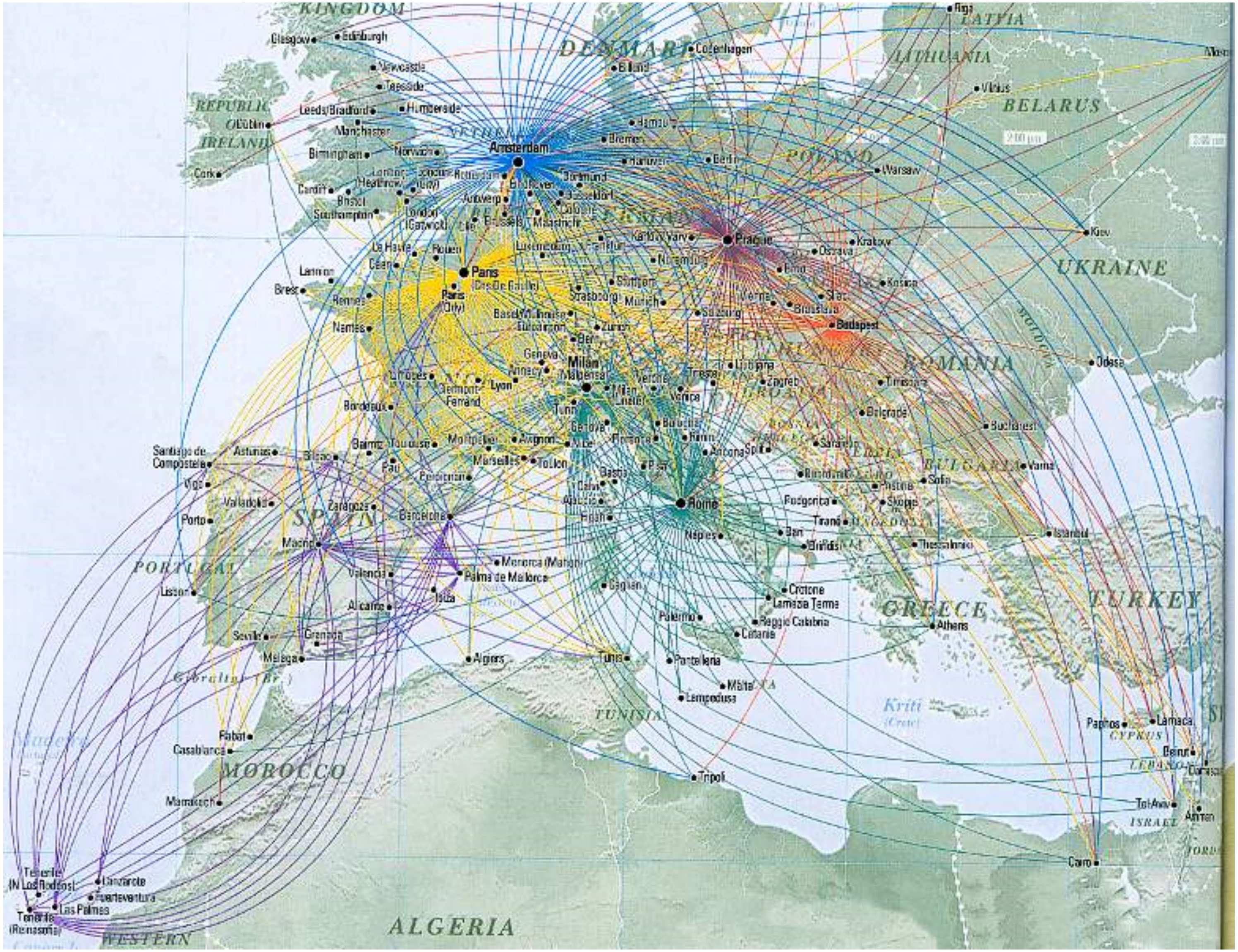
Source: Lada A. Adamic and Eytan Adar, 'Friends and neighbors on the web', *Social Networks*, 25(3):211-230, July 2003.

examples: internet



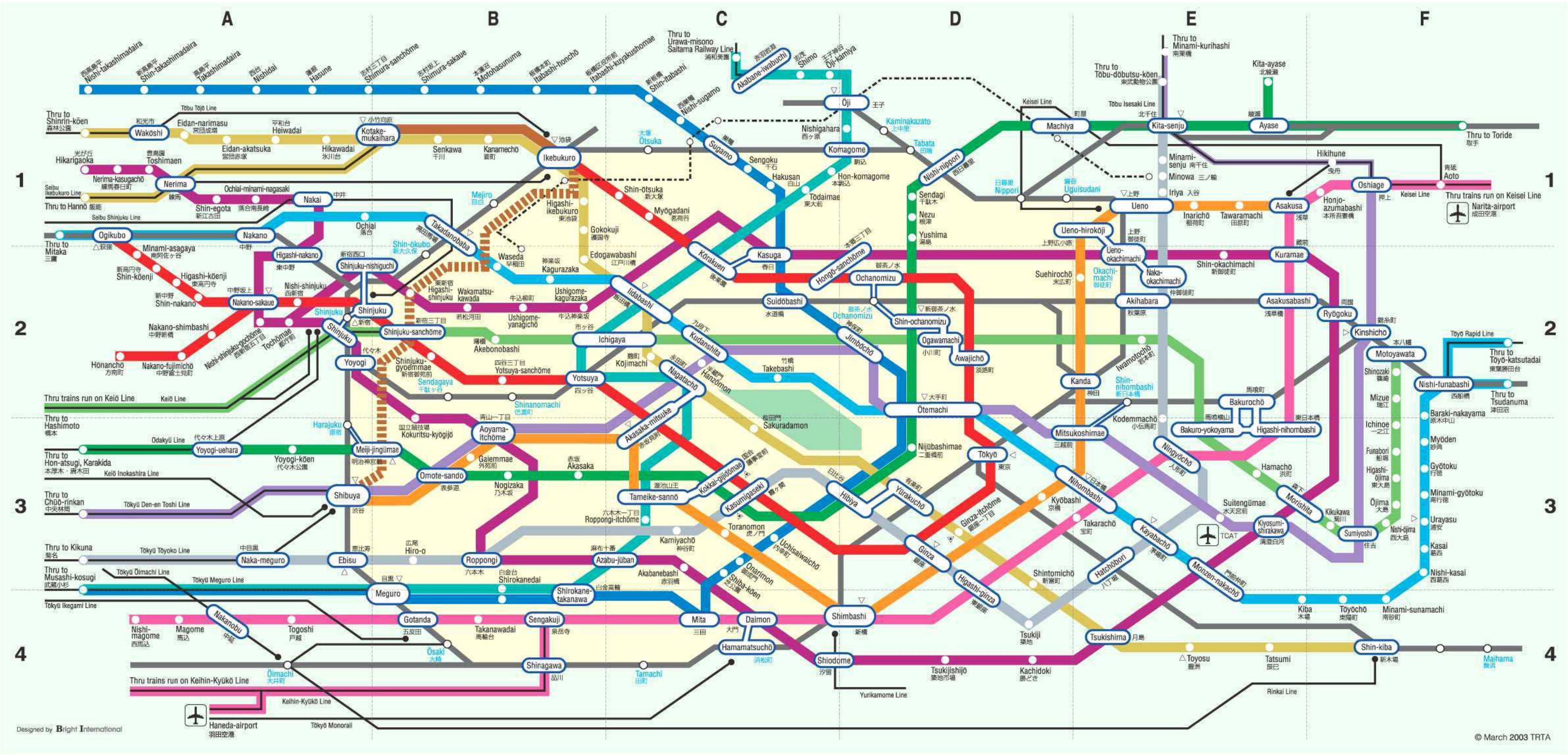
Source: Bill Cheswick http://www.cheswick.com/ches/map/gallery/index.html

examples: airline networks



Source: Northwest Airlines WorldTraveler Magazine

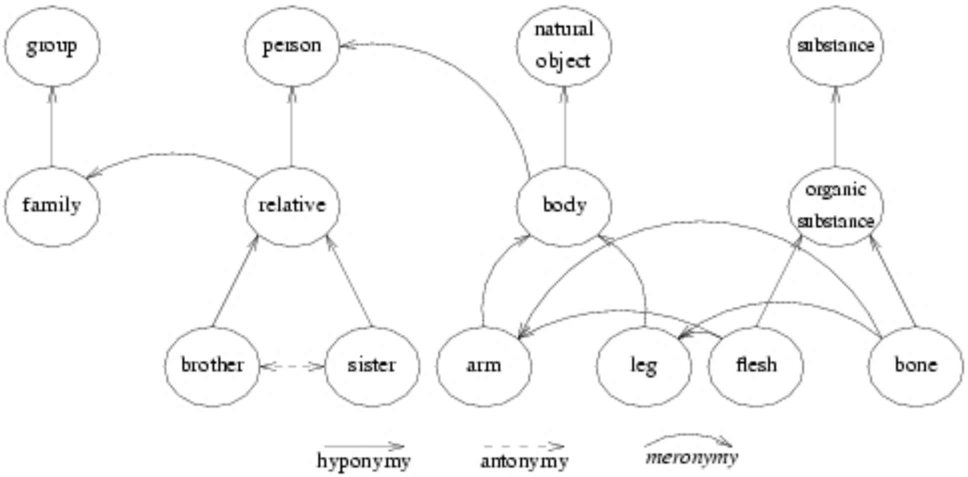
examples: railway networks



Source: TRTA, March 2003 - Tokyo rail map

other examples, e.g. natural language processing

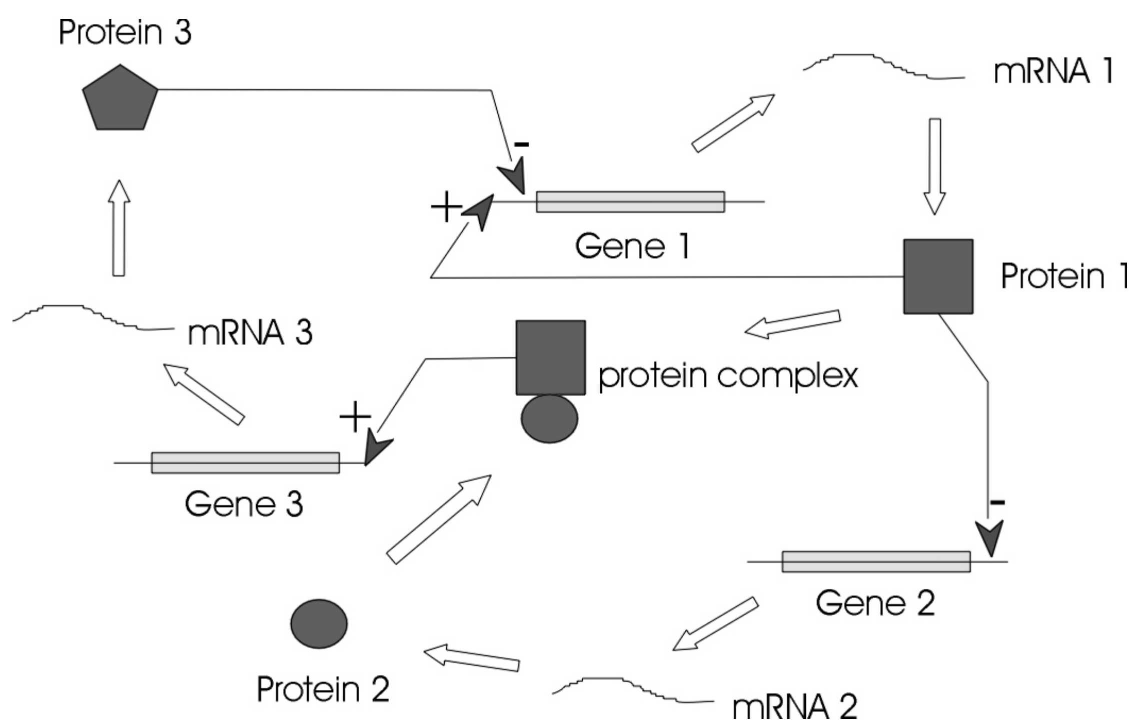
- Wordnet



Source: http://wordnet.princeton.edu/man/wnlicens.7WN

examples: gene regulatory networks

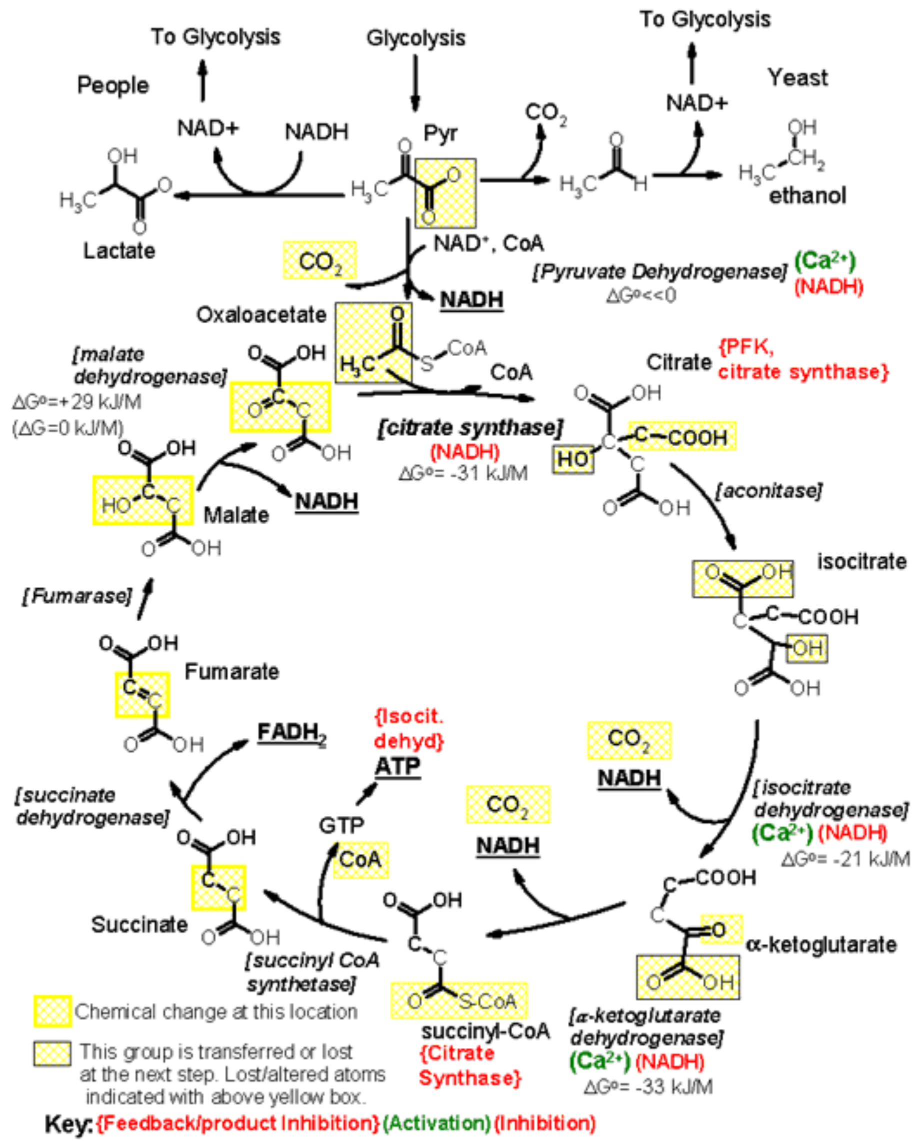
- gene regulatory networks
  - humans have only 30,000 genes, 98% shared with chimps
  - the complexity is in the interaction of genes
  - can we predict what result of the inhibition of one gene will be?



Source: http://www.zaik.uni-koeln.de/bioinformatik/regulatorynets.html.en

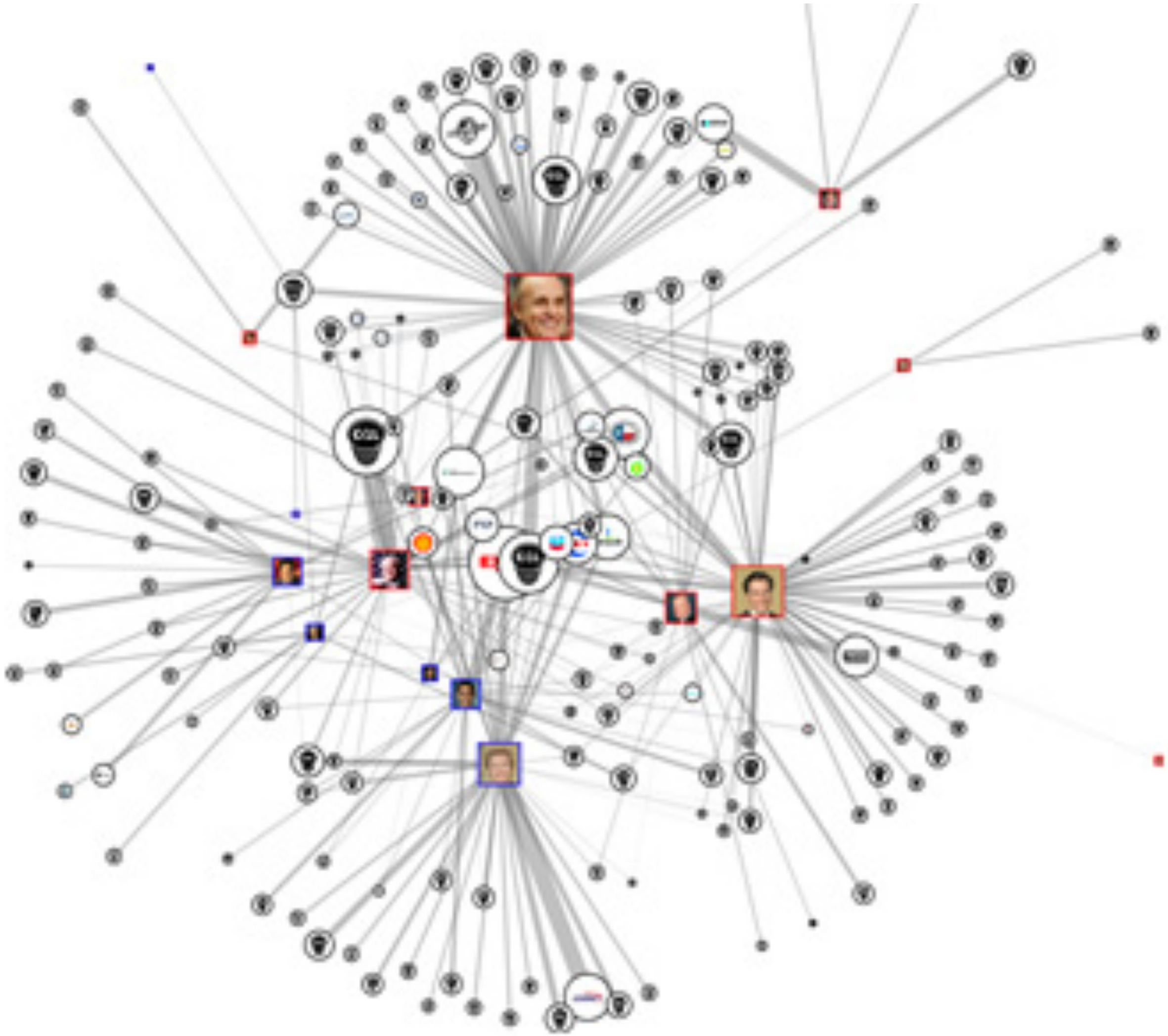
examples: metabolic networks

- Citric acid cycle
- Metabolites participate in chemical reactions



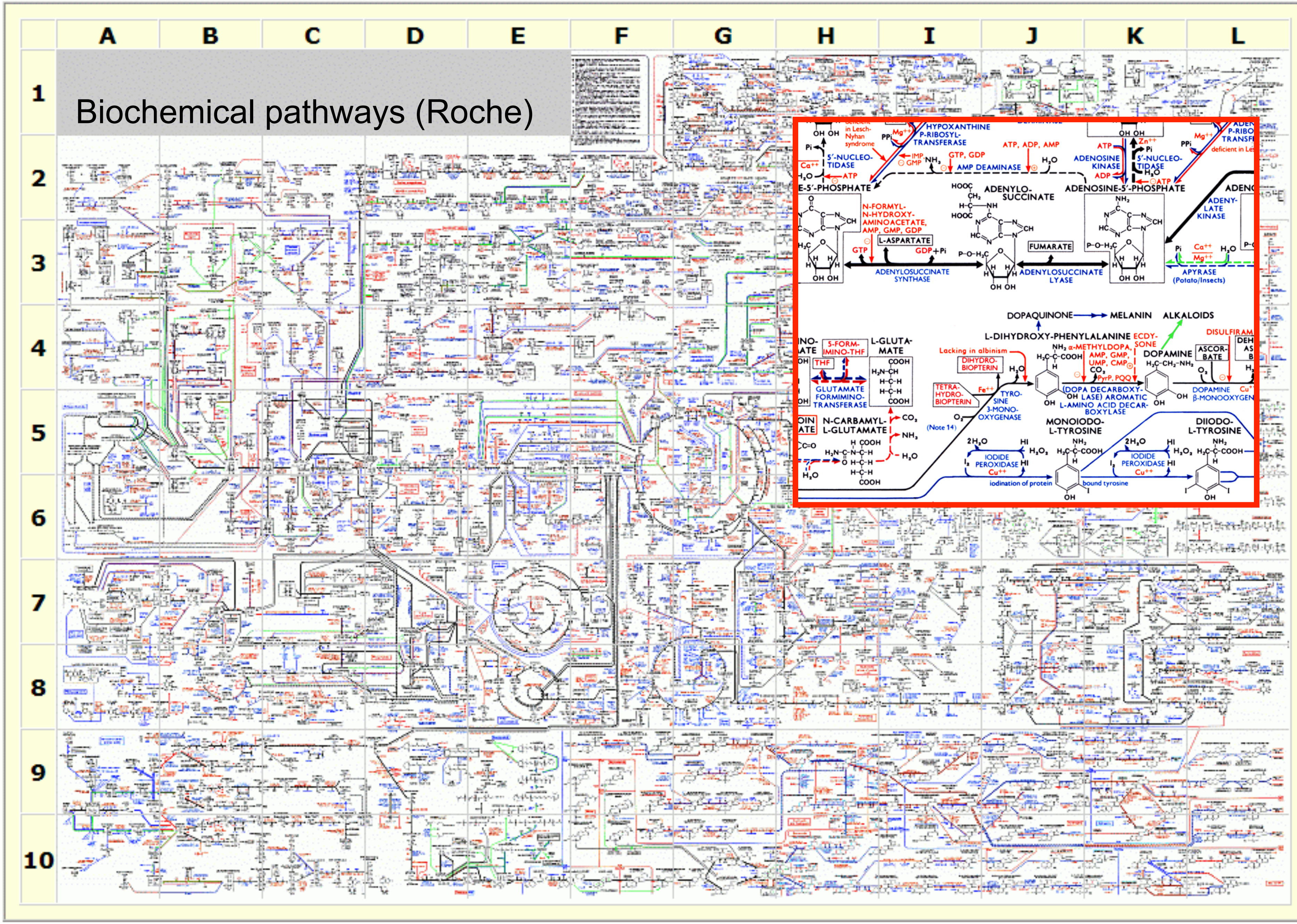
Source: undetermined

Campaign Contributions from Oil Companies



(from http://oilmoney.priceofoil.org/)

Biochemical pathways (Roche)

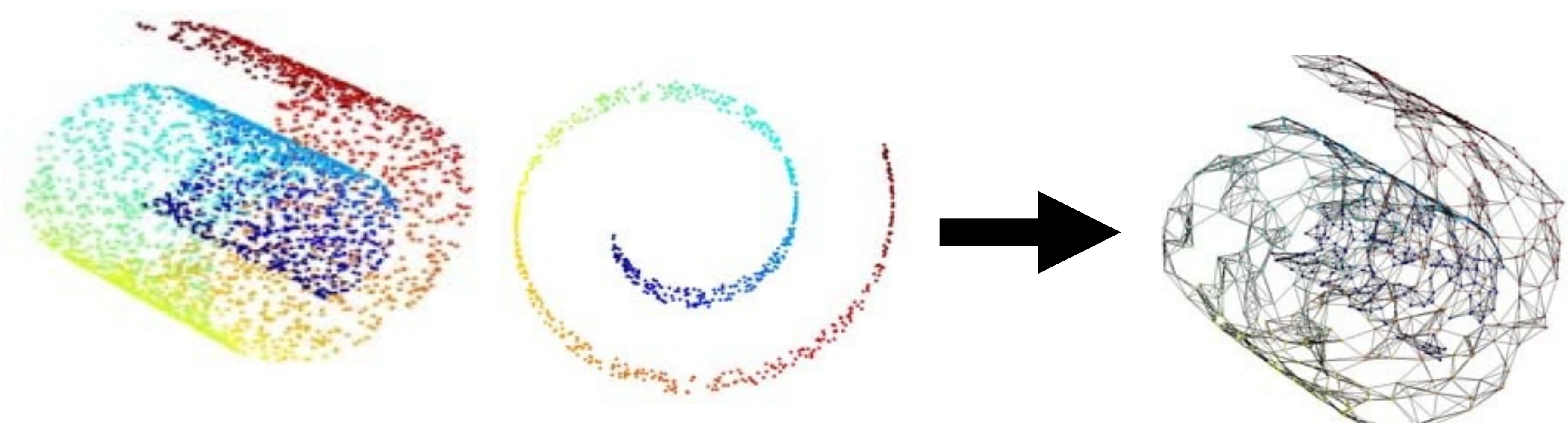


Source: Roche Applied Science, http://www.expasy.org/cgi-bin/show_thumbnails.pl

Creating a network from a surface

- Sample points from the surface
- Connect each point to the k closest points as measured by Euclidean distance

$$dist(a,b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2}$$



Creating a network from data

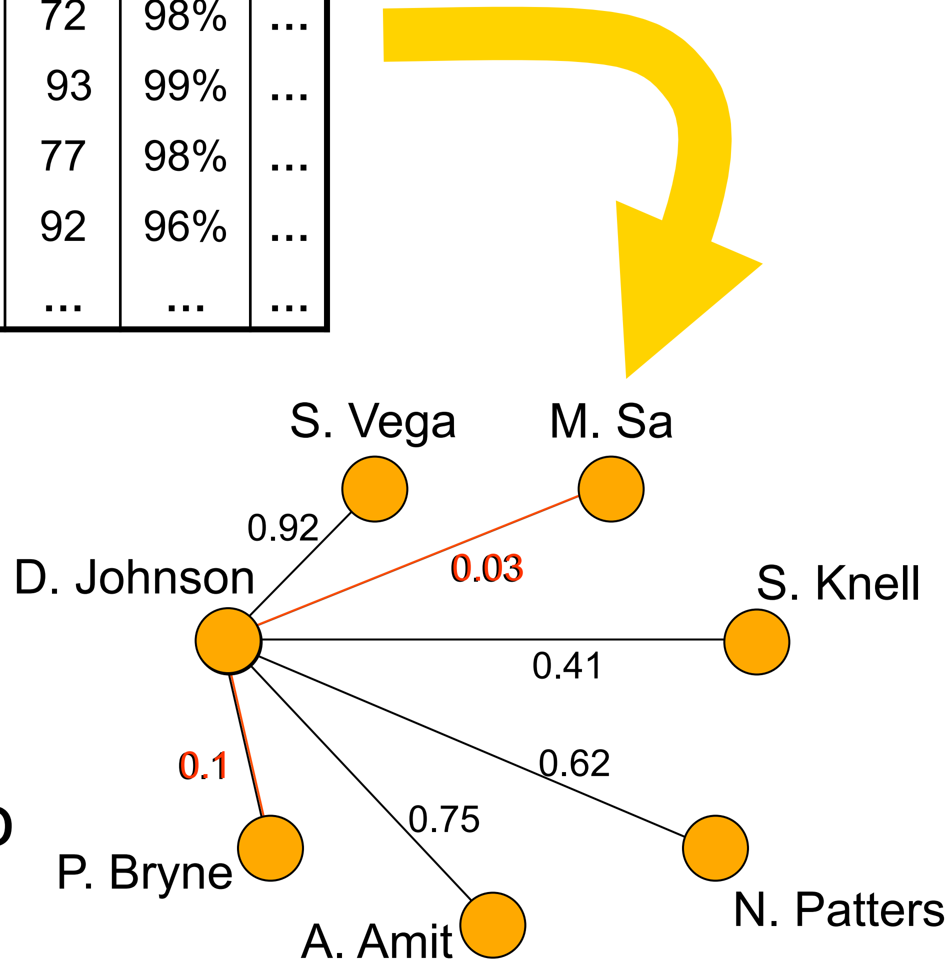
Medical Patients

Name	Age	Weight	Height	HR	SBP	DBP	SpO_2	…
D. Johnson	32	153	70	82	134	72	98%	…
S. Knell	47	169	65	130	169	93	99%	…
P. Bryne	42	128	61	102	129	77	98%	…
A. Amit	39	191	68	121	143	92	96%	…
…	…	…	…	…	…	…	…	…

1.) Measure the distance between pairs

$$dist(a,b) = \sqrt{\sum_i w_i (a_i - b_i)^2}$$

2.) Connect each patient to its k nearest neighbors



Creating a network from data

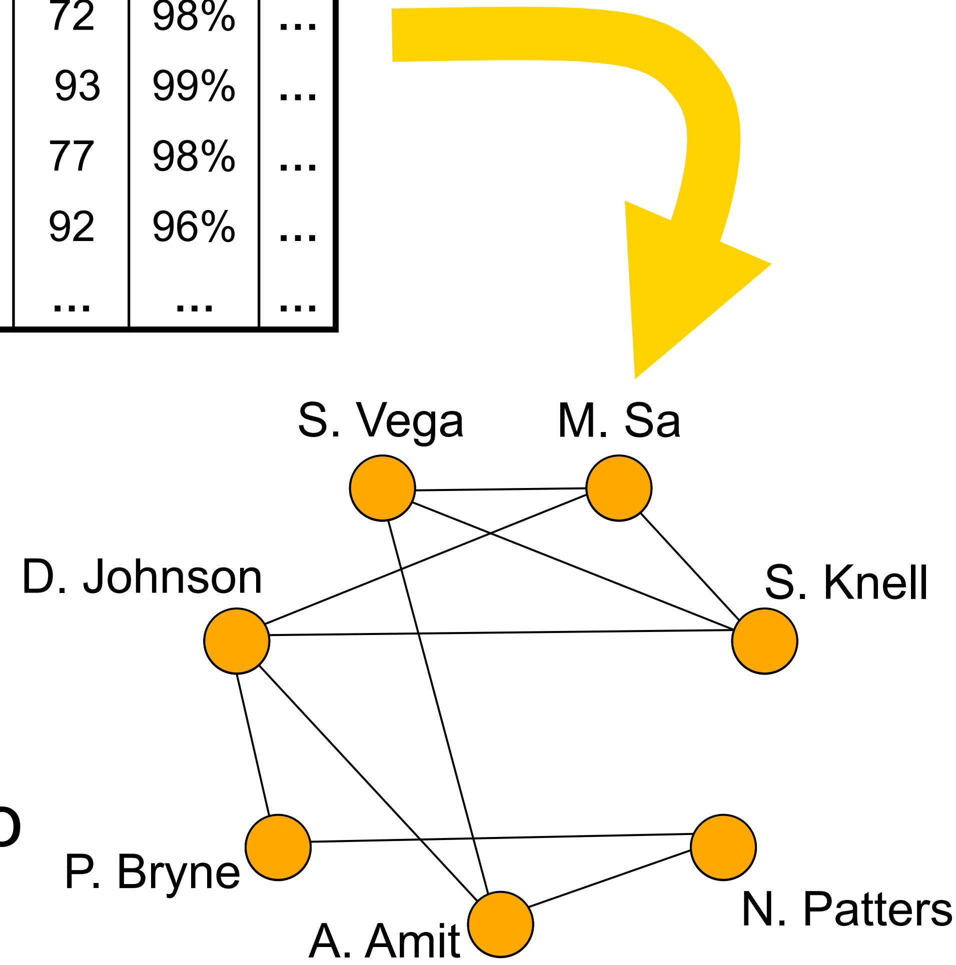
Medical Patients

Name	Age	Weight	Height	HR	SBP	DBP	SpO_2	…
D. Johnson	32	153	70	82	134	72	98%	…
S. Knell	47	169	65	130	169	93	99%	…
P. Bryne	42	128	61	102	129	77	98%	…
A. Amit	39	191	68	121	143	92	96%	…
…	…	…	…	…	…	…	…	…

1.) Measure the distance between pairs

$$dist(a,b) = \sqrt{\sum_i w_i (a_i - b_i)^2}$$

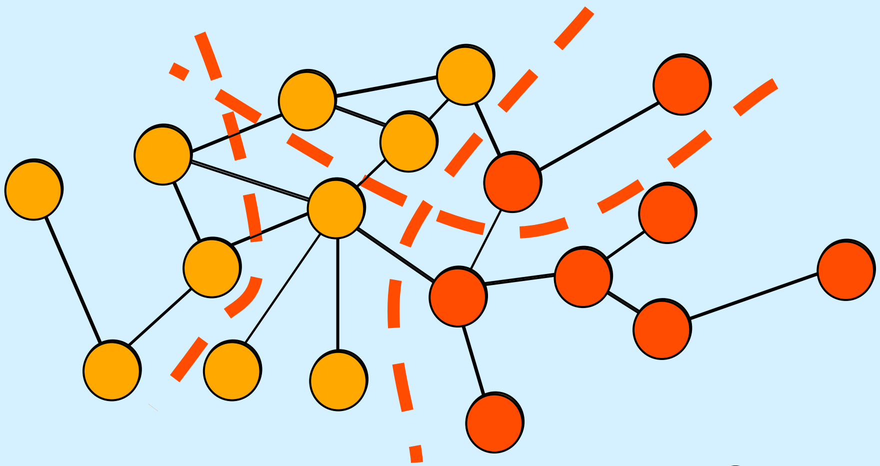
2.) Connect each patient to its k nearest neighbors



Graph partitioning

Goal: Partition the graph into multiple groups (*clusters*)

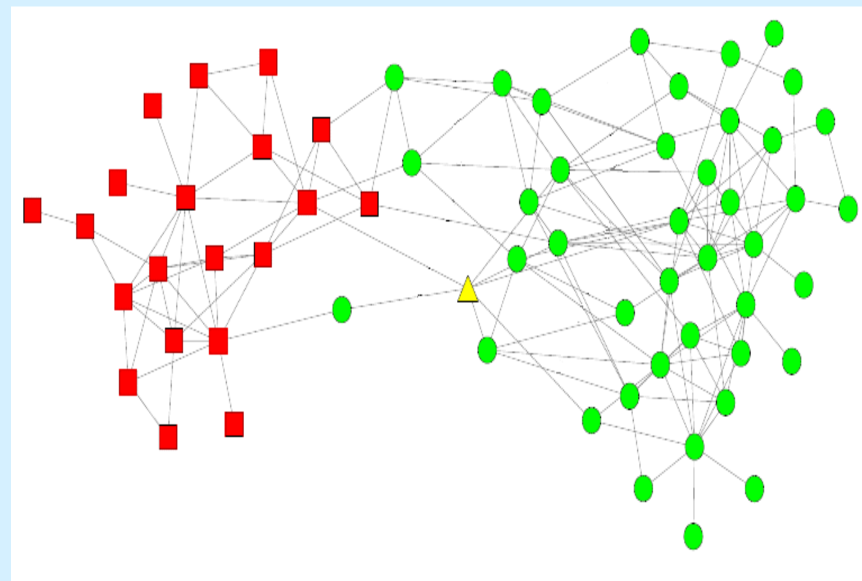
Find the two clusters



But where do we cut?

Identify the different parts of the rabbit

Social network of 62 dolphins
(Lusseau et al., 2003)

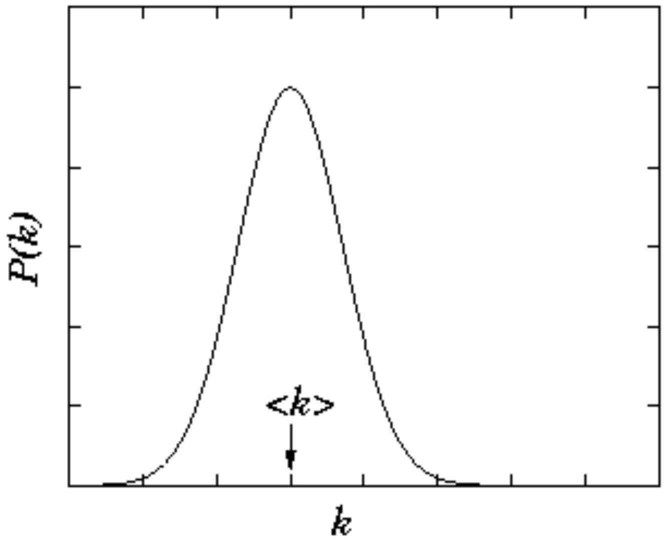


Predict how the community will split when △ departs

modeling networks: random networks

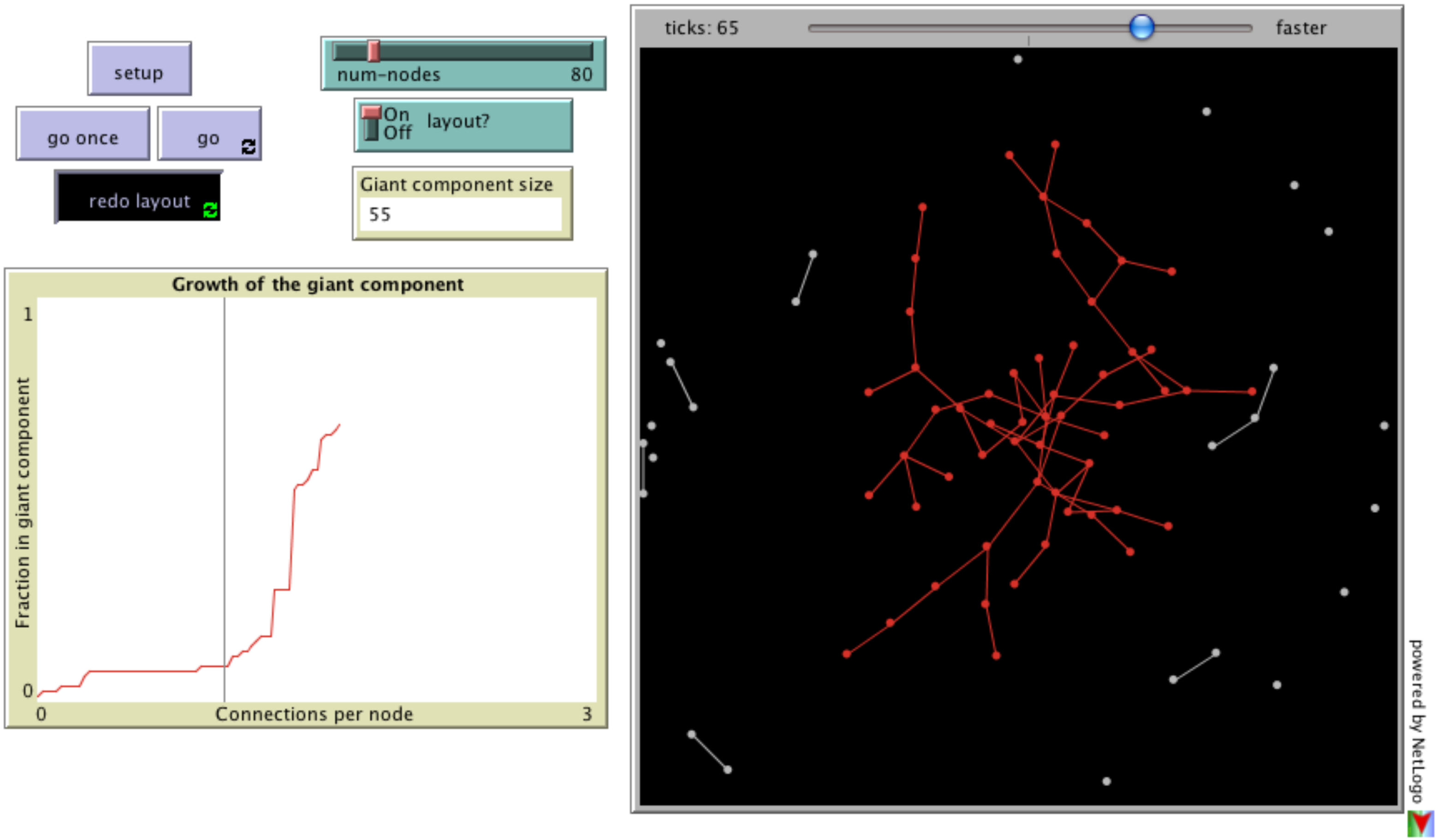
- Nodes connected at random
- Number of edges incident on each node is Poisson distributed

Poisson distribution



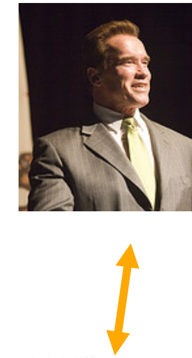
Erdos-Renyi random graphs

- What happens to the size of the giant component as the density of the network increases?

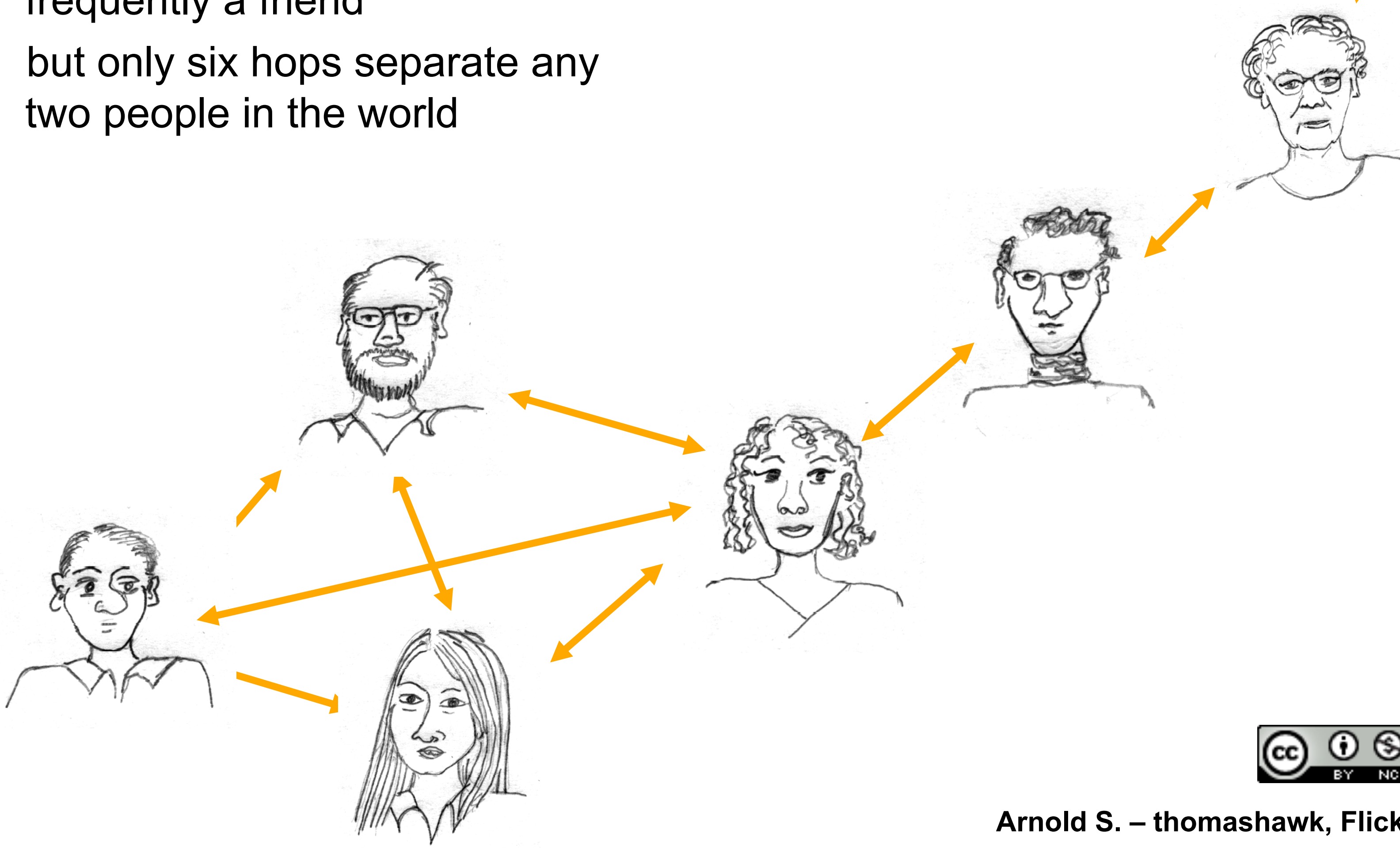


http://ccl.northwestern.edu/netlogo/models/run.cgi?GiantComponent.884.534

modeling networks: small worlds

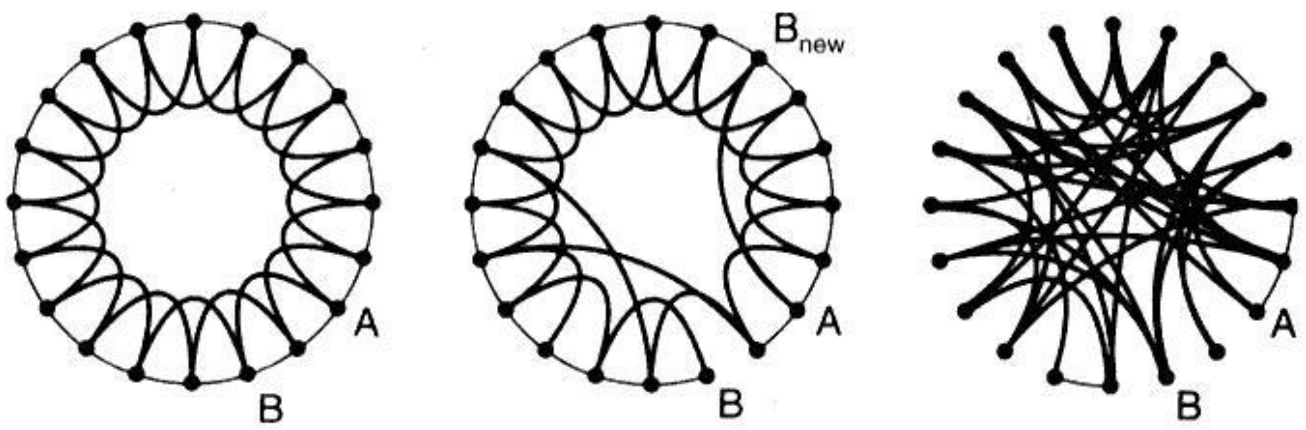


- Small worlds
  - a friend of a friend is also frequently a friend
  - but only six hops separate any two people in the world



Small world models

- Duncan Watts and Steven Strogatz
  - a few random links in an otherwise structured graph make the network a small world: the average shortest path is short



regular lattice:
my friend’s friend is always my friend

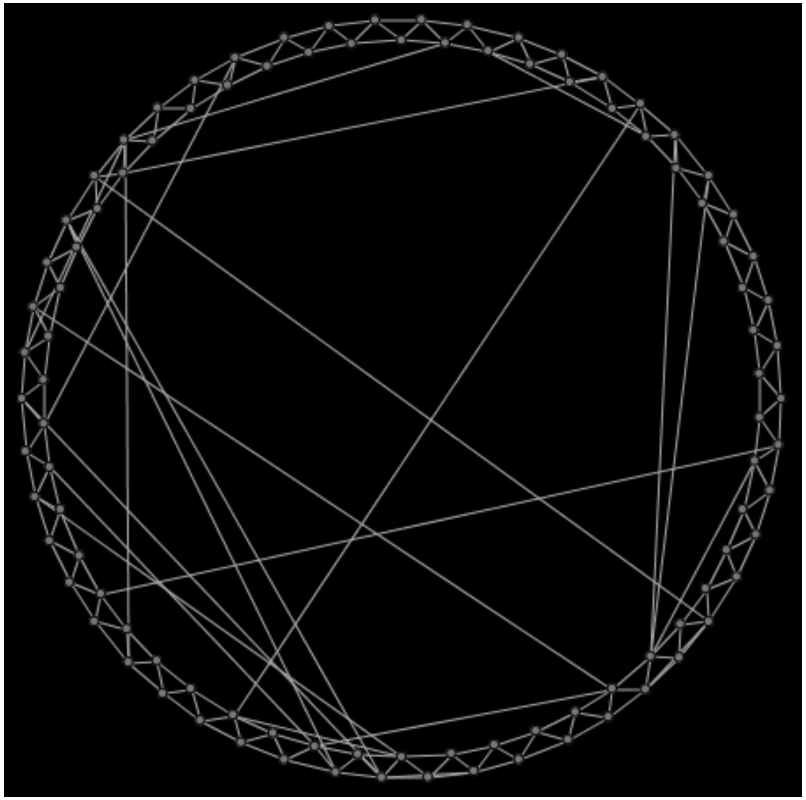
small world:
mostly structured with a few random connections

random graph:
all connections random

Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Watts Strogatz Small World Model

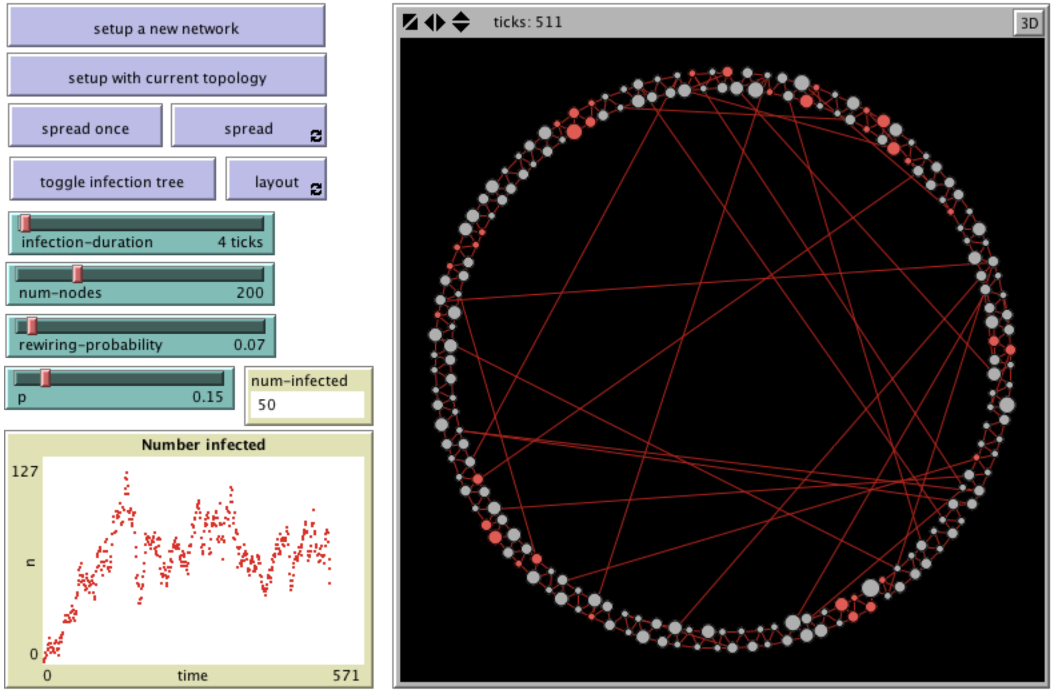
- As you rewire more and more of the links and random, what happens to the clustering coefficient and average shortest path relative to their values for the regular lattice?



http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldWS.html

SIS models and small worlds

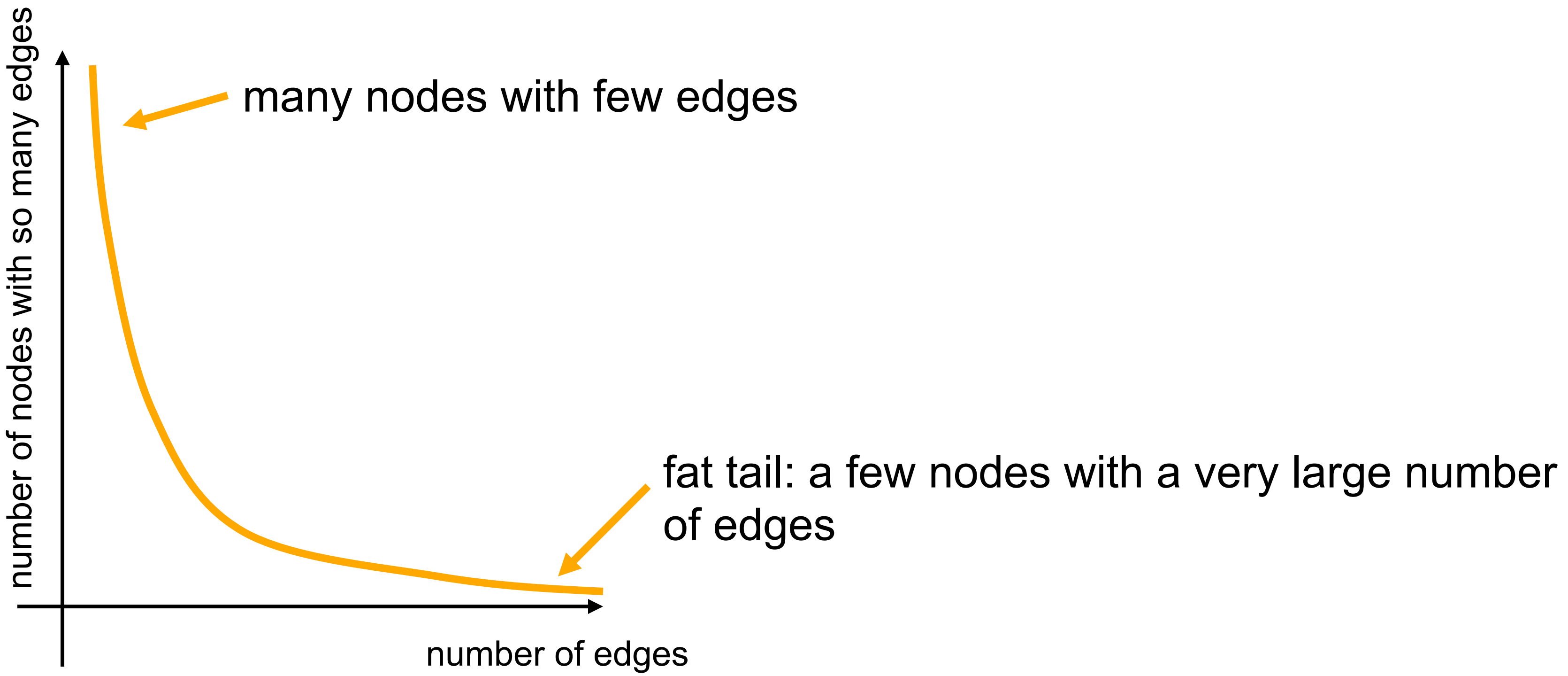
- SIS model: nodes return to “susceptible” state after being infected
- What is the role of random shortcuts in diffusion?



http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldWS.html

modeling networks: power law networks

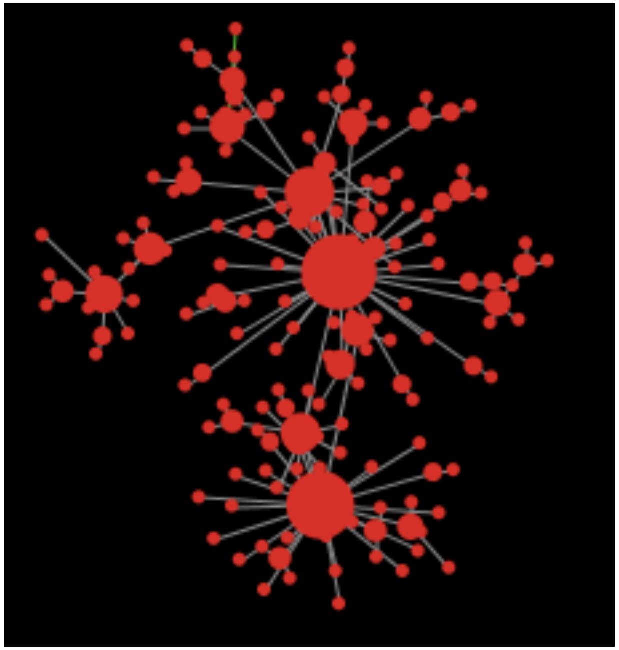
- Many real world networks contain hubs: highly connected nodes
- Usually the distribution of edges is extremely skewed



no “typical” number of edges

network growth & resulting structure

- random attachment: new node picks any existing node to attach to
- preferential attachment: new node picks from existing nodes according to their degrees



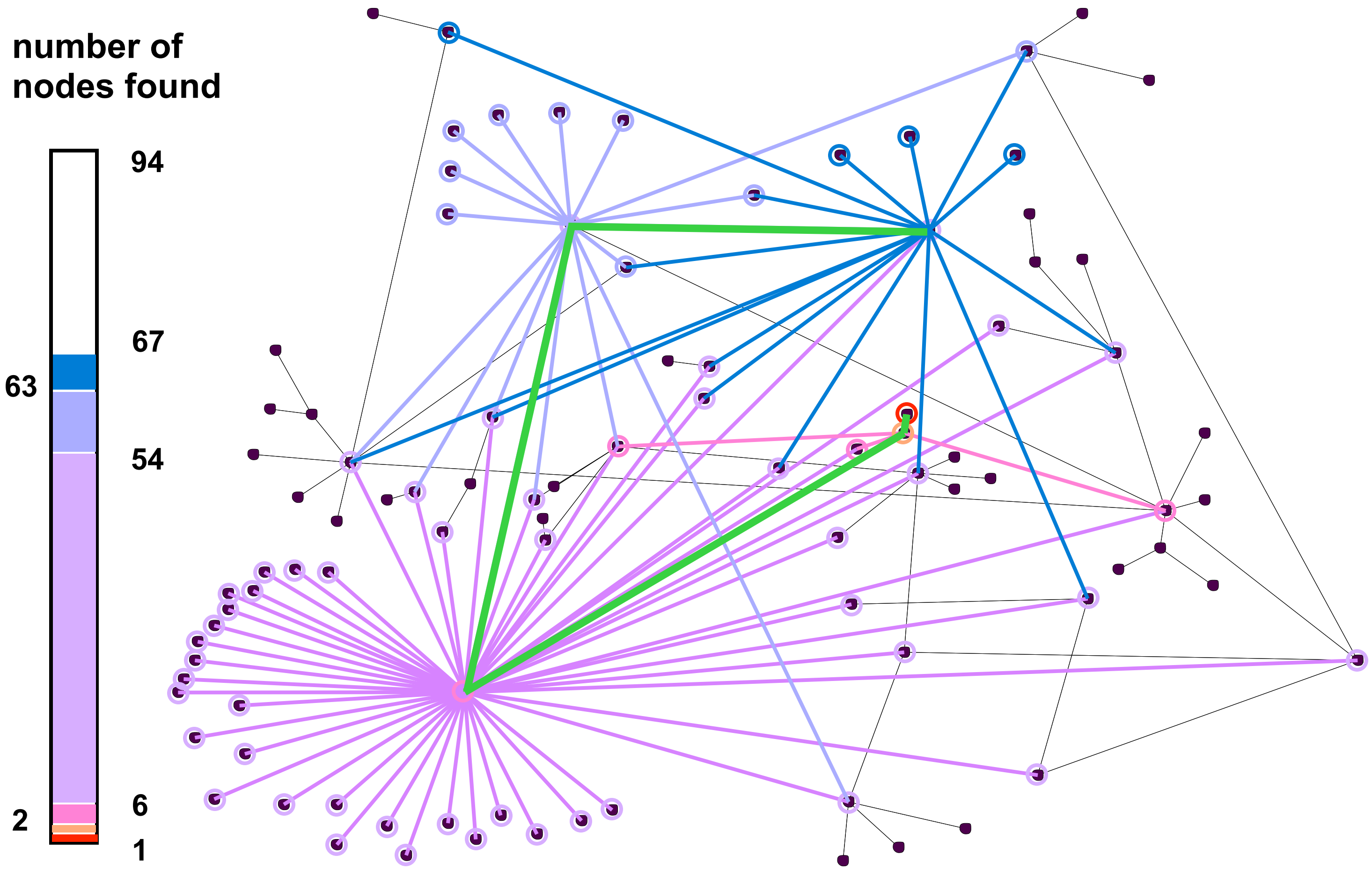
http://projects.si.umich.edu/netlearn/NetLogo4/RAndPrefAttachment.html

What implications does this have?

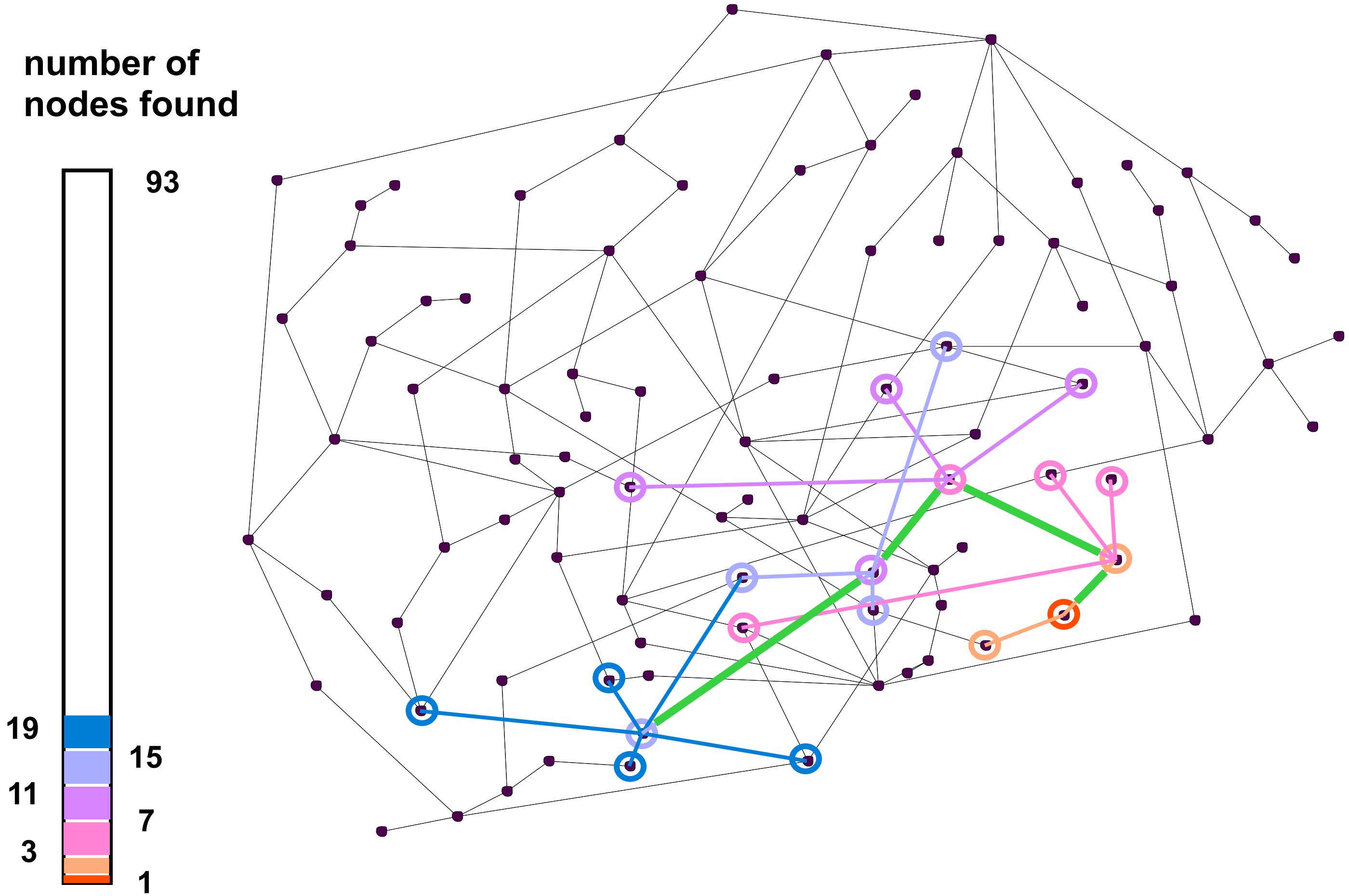
- Robustness
- Search
- Spread of disease
- Opinion formation
- Spread of computer viruses
- Gossip

How do we search?

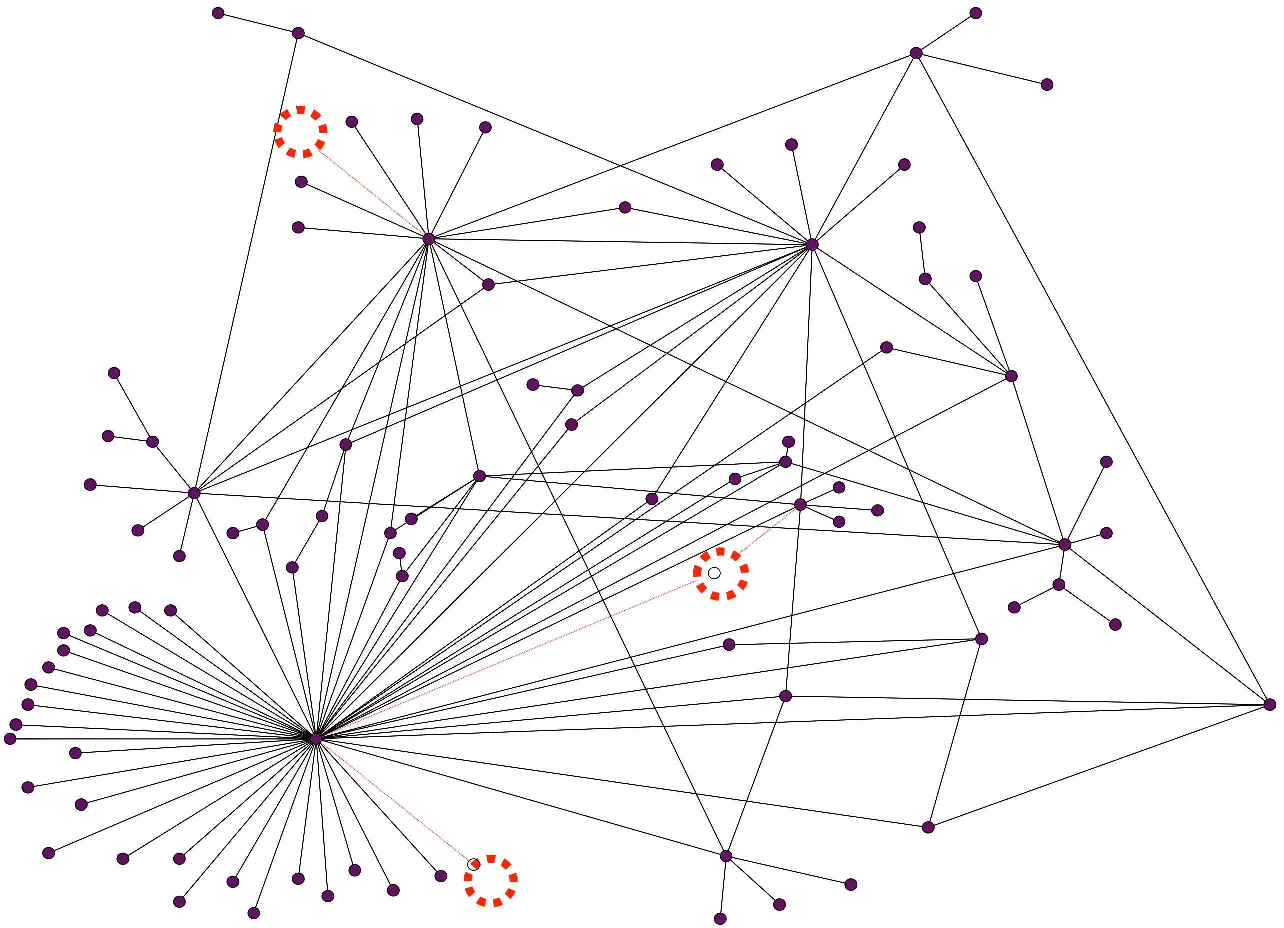
power-law graph



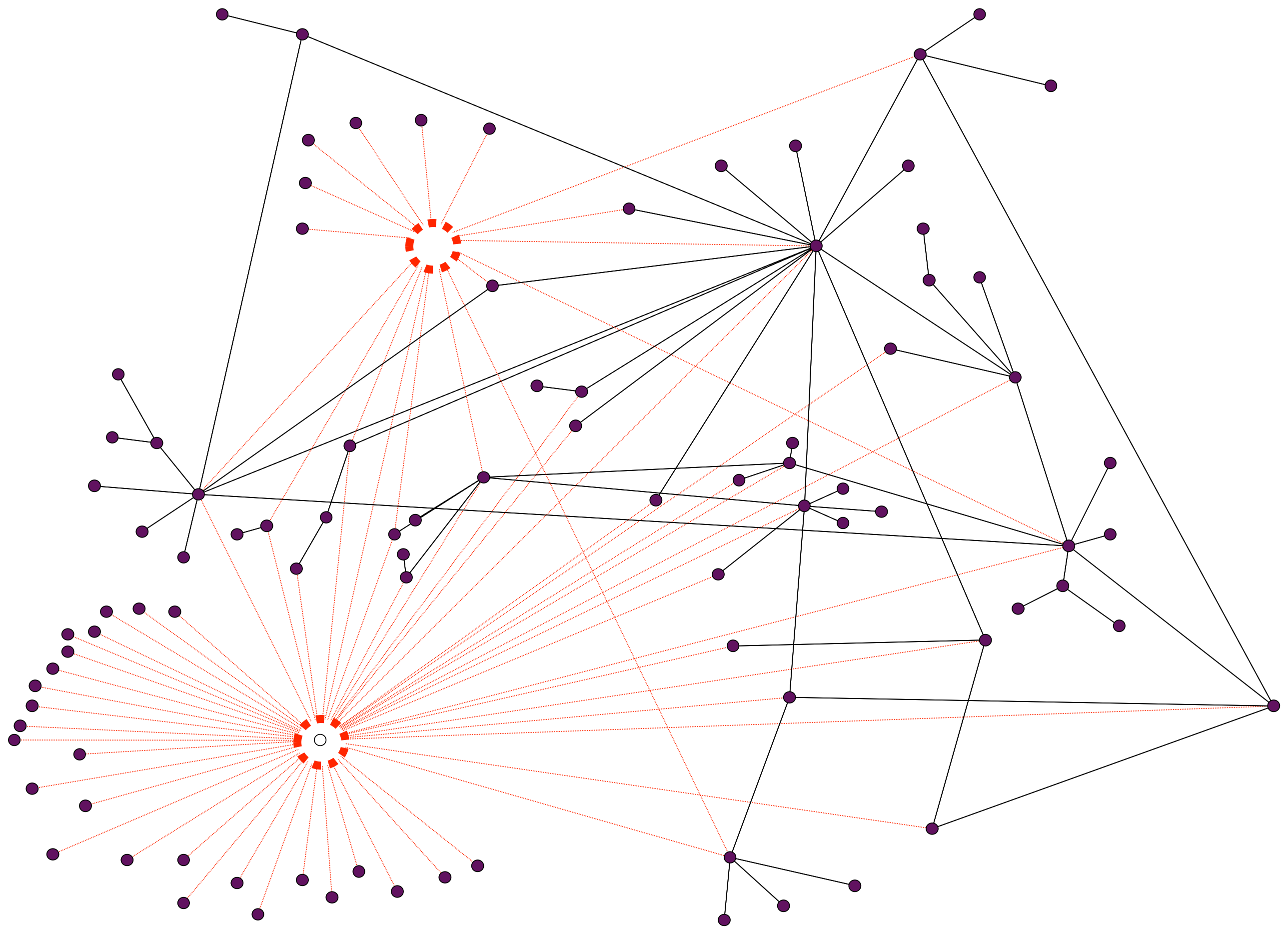
Poisson graph



Power-law networks are robust to random breakdown

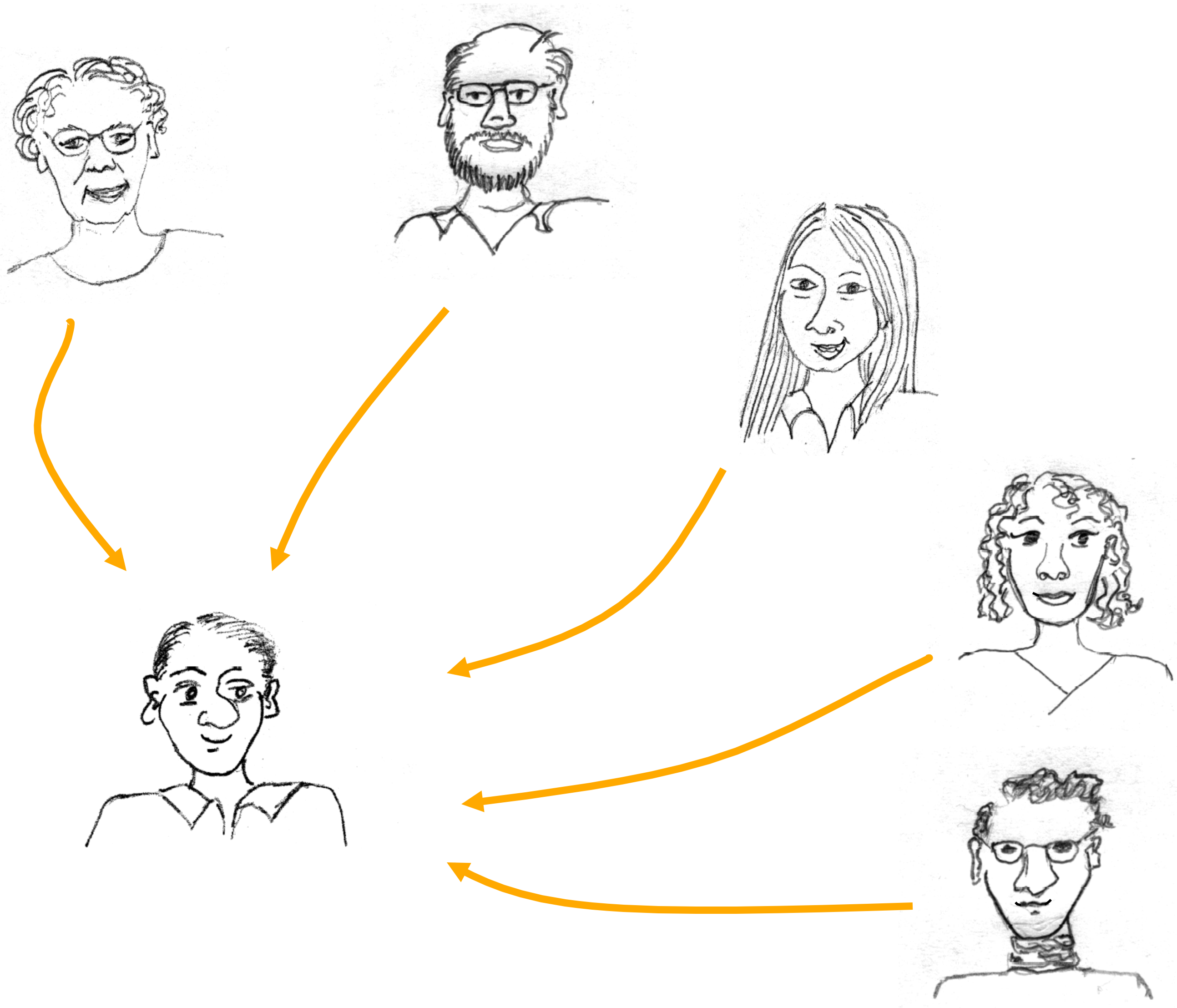


But are especially vulnerable to targeted attack

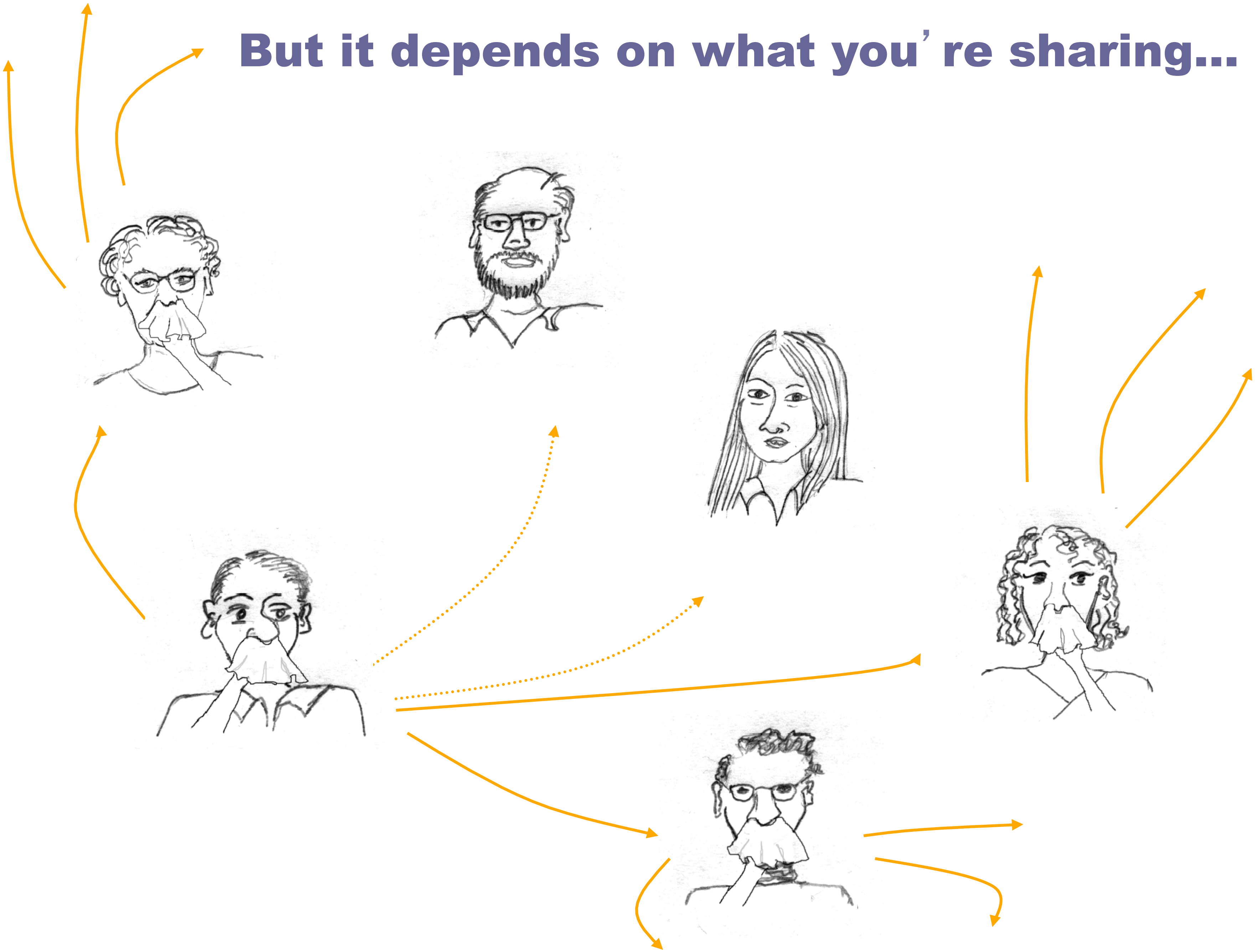


- Targeting and removing hubs can quickly break up the network

In social networks, it' s nice to be a hub



But it depends on what you’re sharing...

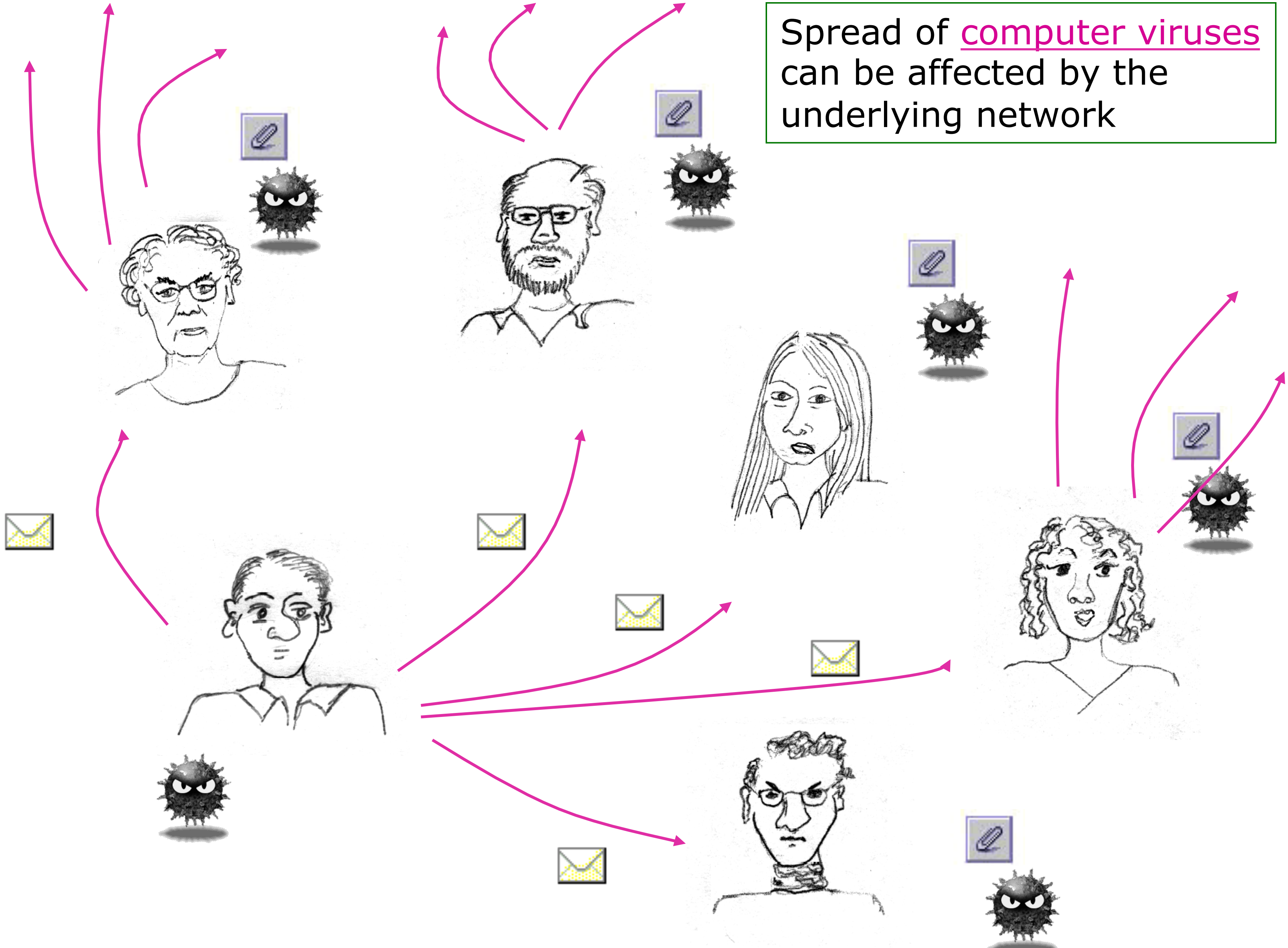


The role of hubs in epidemics

- In a power-law network, a virus can persist no matter how low its infectiousness

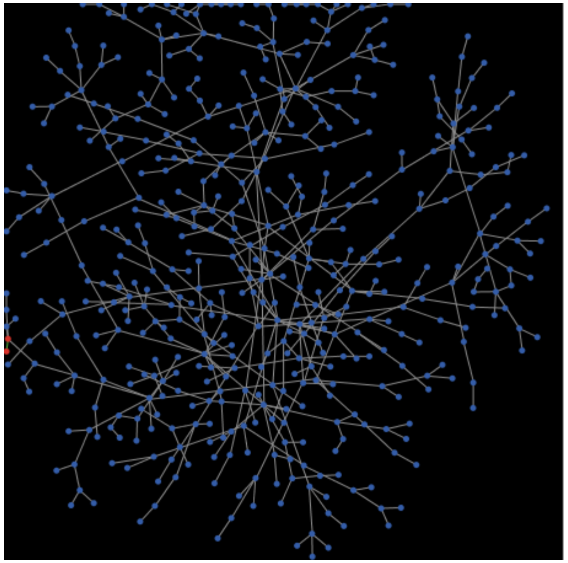
- Many real world networks do exhibit power-laws:
  - needle sharing
  - sexual contacts
  - email networks

Spread of computer viruses can be affected by the underlying network

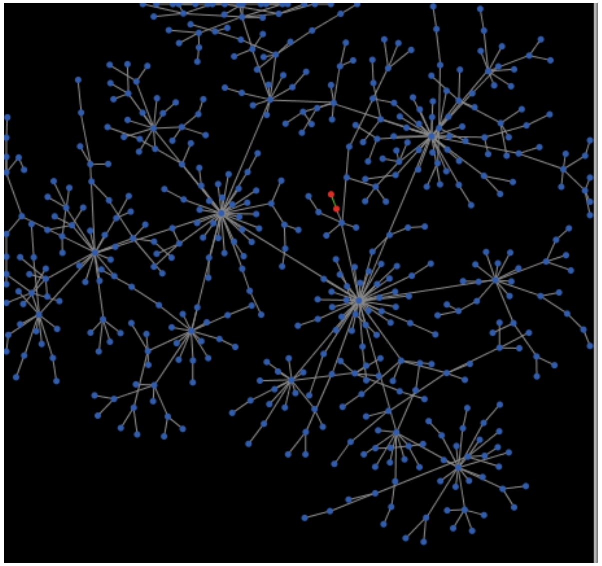


SI models & network structure

- Will random or preferential attachment lead to faster diffusion?



random growth

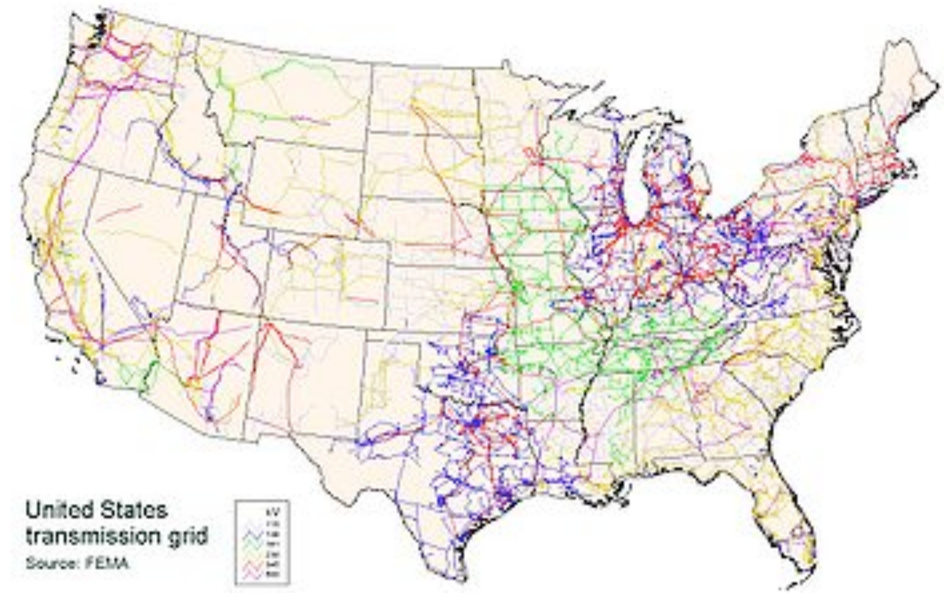


preferential growth

http://projects.si.umich.edu/netlearn/NetLogo4/BADiffusion.html

resilience:
power grids and cascading failures

- Vast system of electricity generation, transmission & distribution is essentially
  a single network



- Power flows through
  all paths from source to sink
  (flow calculations are
  important for other networks,
  even social ones)

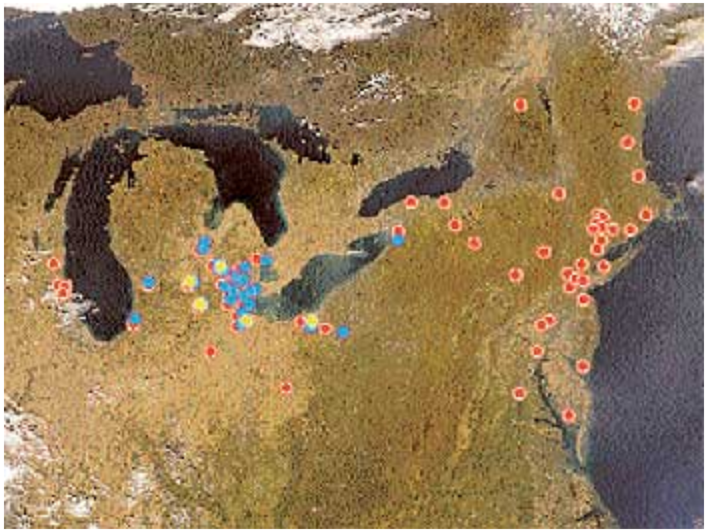
- All AC lines within an
  interconnect must be in sync

- If frequency varies too much (as line approaches capacity), a circuit breaker takes the generator out of the system

- Larger flows are sent to neighboring parts of the grid – triggering a cascading failure

Source: .wikipedia.org/wiki/File:UnitedStatesPowerGrid.jpg

Cascading failures



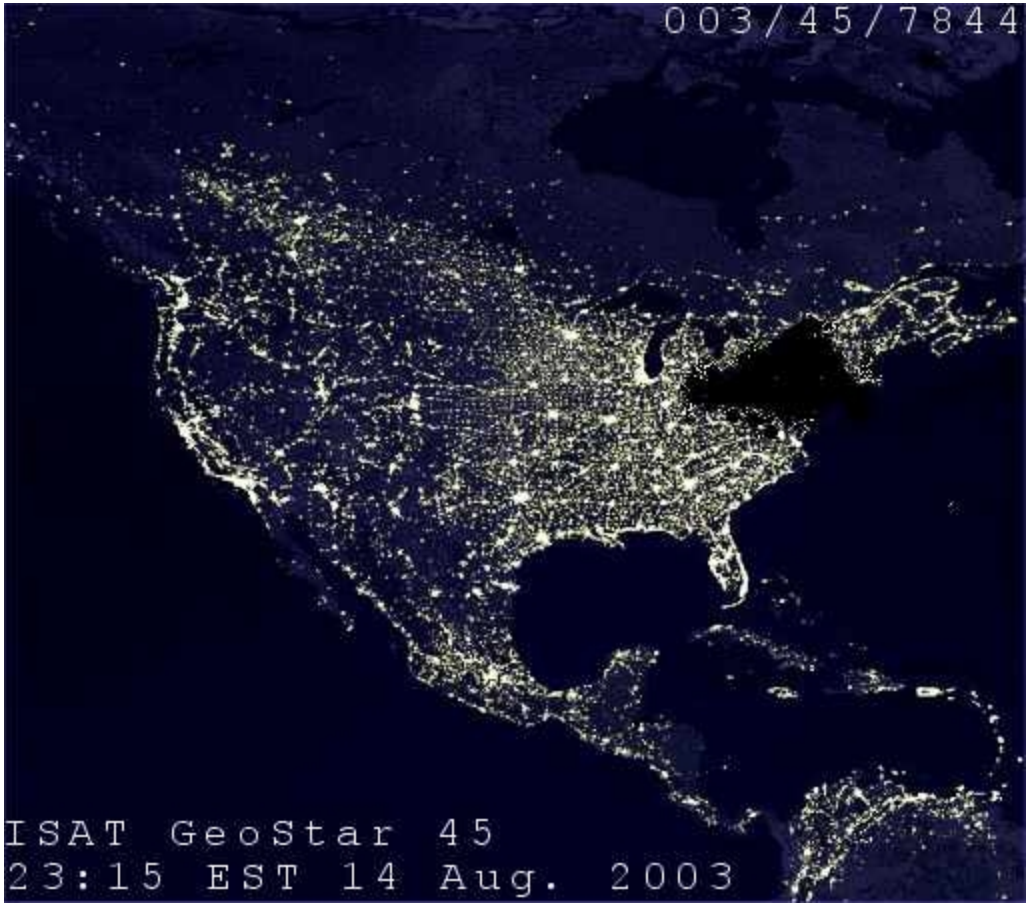
- **1:58 p.m.** The Eastlake, Ohio, First Energy generating plant shuts down (maintenance problems).
- **3:06 p.m.** A First Energy 345-kV transmission line fails south of Cleveland, Ohio.
- **3:17 p.m**. Voltage dips temporarily on the Ohio portion of the grid. Controllers take no action, but power shifted by the first failure onto another power line causes it to sag into a tree at 3:32 p.m., bringing it offline as well. While Mid West ISO and First Energy controllers try to understand the failures, they fail to inform system controllers in nearby states.
- **3:41 and 3:46 p.m.** Two breakers connecting First Energy's grid with American Electric Power are tripped.
- **4:05 p.m.** A sustained power surge on some Ohio lines signals more trouble building.
- **4:09:02 p.m.** Voltage sags deeply as Ohio draws 2 GW of power from Michigan.
- **4:10:34 p.m.** Many transmission lines trip out, first in Michigan and then in Ohio, blocking the eastward flow of power. Generators go down, creating a huge power deficit. In seconds, power surges out of the East, tripping East coast generators to protect them.

Source: Eric J. Lerner, "What's wrong with the electric grid?" http://www.aip.org/tip/INPHFA/vol-9/iss-5/p8.html

(dis) information cascades

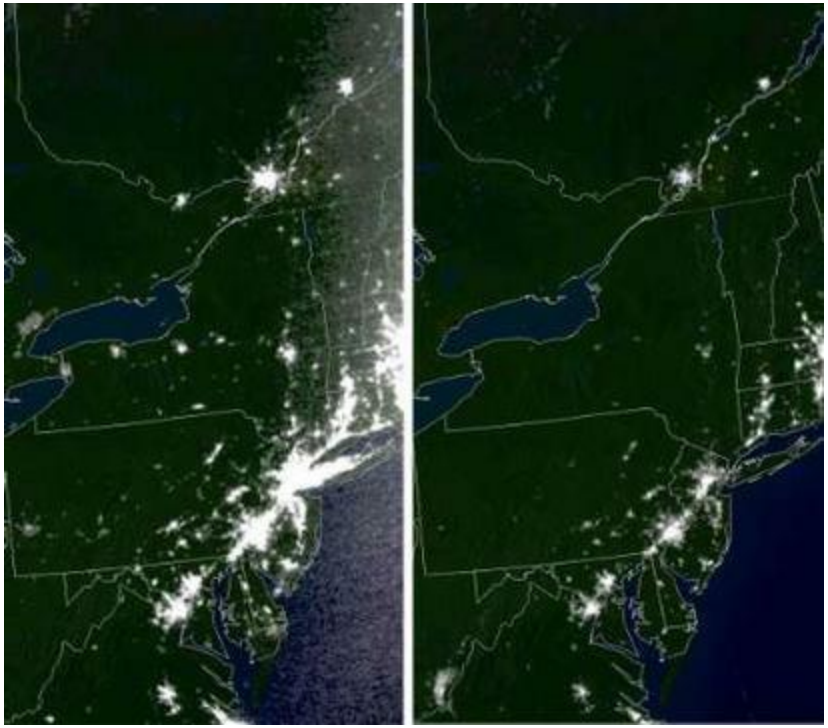
- Rumor spreading
- Urban legends
- Word of mouth (movies, products)

- Web is self-correcting:
  - Satellite image hoax is first passed around, then exposed, hoax fact is blogged about, then written up on urbanlegends.about.com



Source: undetermined

Actual satellite images of the effect of the blackout



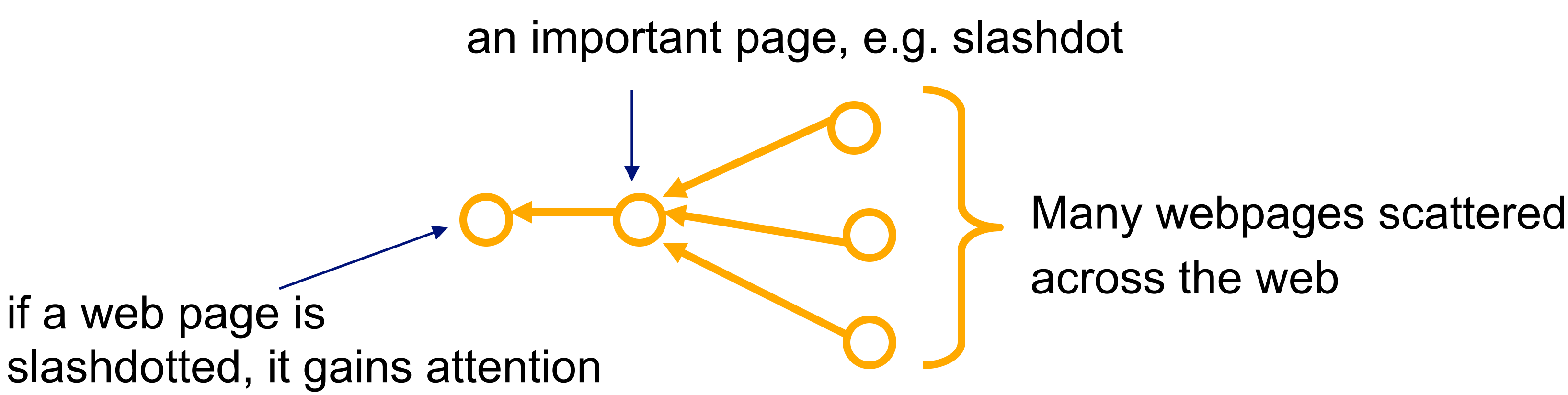
20 hours prior to blackout

7 hours after blackout

Source: NOAA, U.S. Government

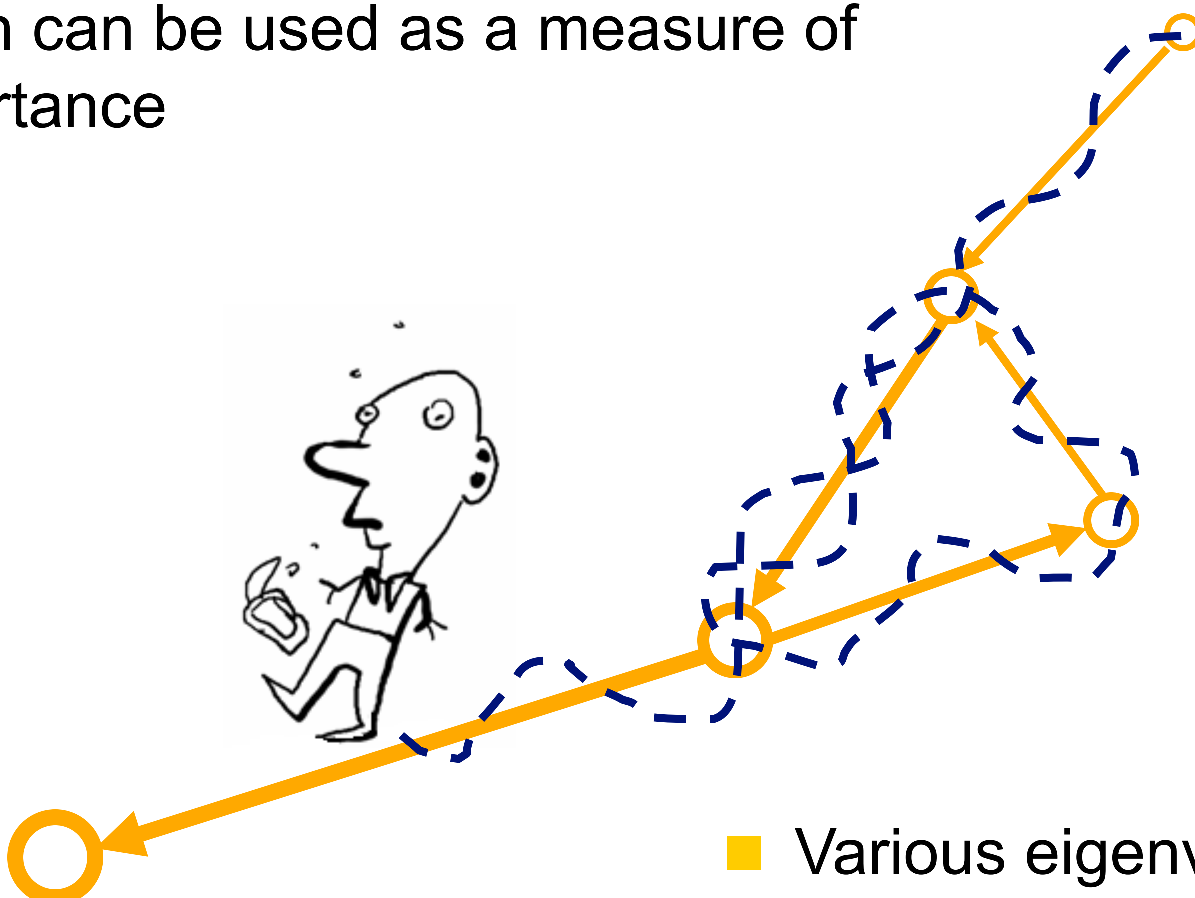
IR applications: online info retrieval

- It's in the links:
  - links to URLs can be interpreted as endorsements or recommendations
  - the more links a URL receives, the more likely it is to be a good/ entertaining/provocative/authoritative/interesting information source
  - but not all link sources are created equal
    - a link from a respected information source
    - a link from a page created by a spammer



Ranking pages by tracking a drunk

- A random walker following edges in a network for a very long time will spend a proportion of time at each node which can be used as a measure of importance



- Various eigenvalue metrics yield variations of importance measures

Wrap up

- networks are everywhere and can be used to describe many, many systems
- by modeling networks we can start to understand their properties and the implications those properties have for processes occurring on the network