

Three Lectures on Networks

Aaron Clauset

 @aaronclauset

Associate Professor of Computer Science
University of Colorado Boulder
External Faculty, Santa Fe Institute

lecture 2: degrees, positions, and communities

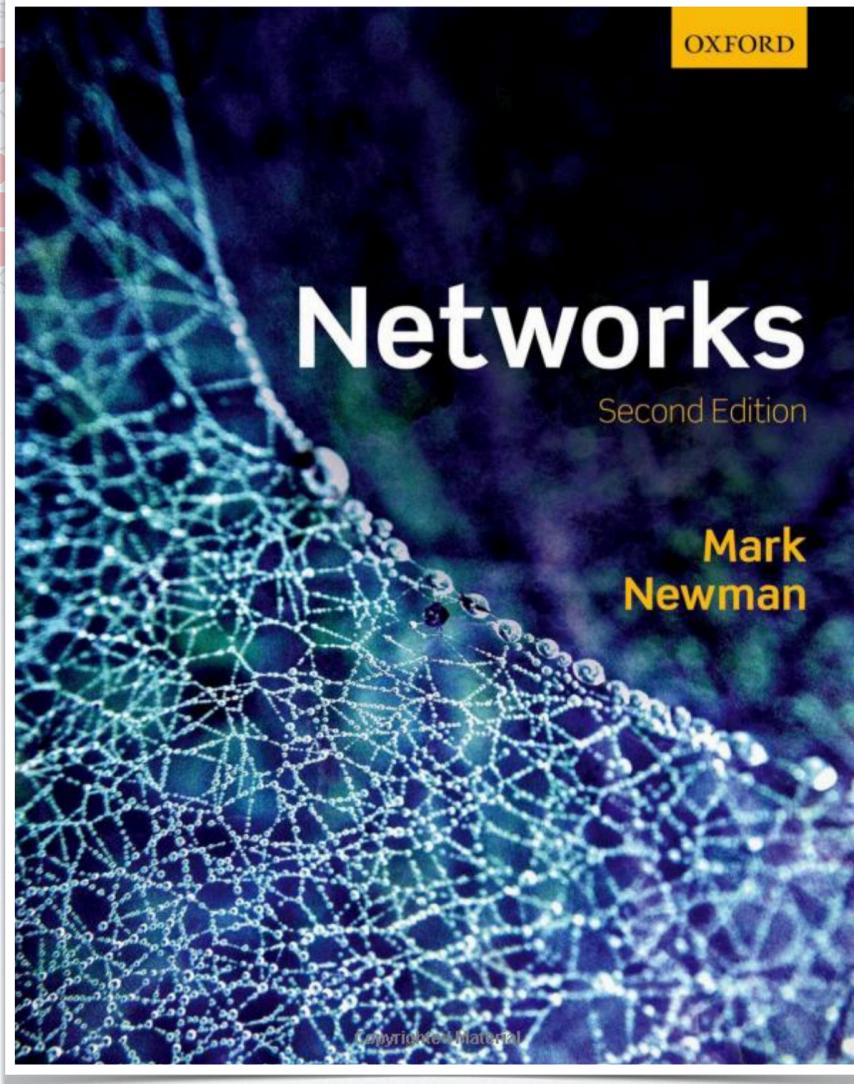


Mark Newman

Professor of Physics
University of Michigan

External Faculty
Santa Fe Institute

<http://www-personal.umich.edu/~mejn/>





University of Colorado **Boulder**

Network Analysis and Modeling

Instructor: Aaron Clauset or Daniel B. Larremore

This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Full lectures notes online (~150 pages in PDF)

<https://aaronclauset.github.io/courses/5352/>



University of Colorado **Boulder**

Biological Networks

Instructor: Aaron Clauset

This undergraduate-level course examines the computational representation and analysis of biological phenomena through the structure and dynamics of networks, from molecules to species. Attention focuses on algorithms for clustering network structures, predicting missing information, modeling flows, regulation, and spreading-process dynamics, examining the evolution of network structure, and developing intuition for how network structure and dynamics relate to biological phenomena.

Full lectures notes online (~150 pages in PDF)

<https://aaronclauset.github.io/courses/3352/>

Software

R

Python

Matlab

NetworkX [python]

graph-tool [python, c++]

GraphLab [python, c++]

Standalone editors

UCI-Net

NodeXL

Gephi

Pajek

Network Workbench

Cytoscape

yEd graph editor

Graphviz

Network data sets

Colorado Index of Complex Networks

The screenshot shows a web browser window for 'icon.colorado.edu/#!/'. The title bar says 'icon.colorado.edu/#!/'. The main content area has a dark header with 'Index of Complex Networks' on the left and 'NETWORKS', 'ABOUT', and 'SUGGEST...' tabs on the right. Below the header, there's a large white space with some text and a bar chart.

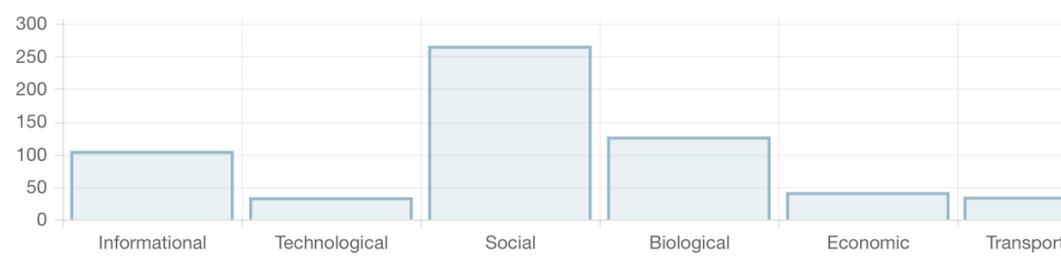
The Colorado Index of Complex Networks (ICON)

ICON is a comprehensive index of research-quality network data sets from all domains of networks, including social, web, information, biological, ecological, connectome, transportation, and technological networks.

Each network record in the index is annotated with and searchable or browsable by its graph properties, description, size, etc., and many records include links to multiple networks. The contents of ICON are curated by volunteer experts from Prof. Aaron Clauset's research group at the University of Colorado Boulder.

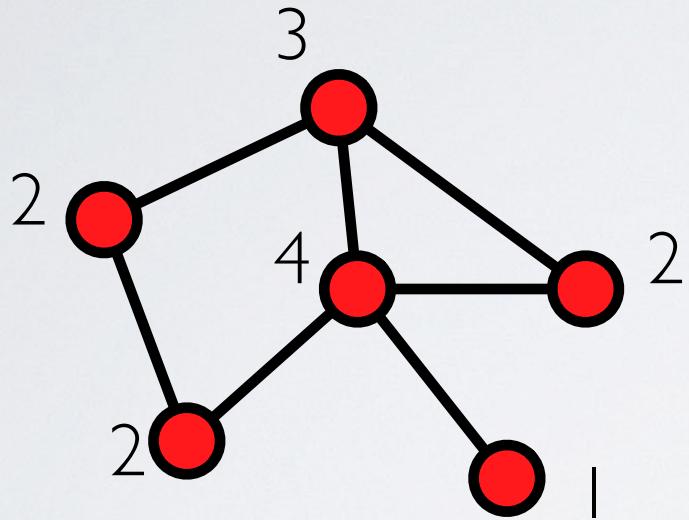
Click on the [NETWORKS tab](#) above to get started.

Entries found: 609 Networks found: 4419



1. defining a network
- 2. describing a network**
3. null models and statistical inference for networks

describing networks



degree:

number of connections k

$$k_i = \sum_j A_{ij}$$

**when does node
degree matter?**

network degrees

spreading processes on networks

network edges are the mechanism of transmission

biological (diseases)

- SIS and SIR models

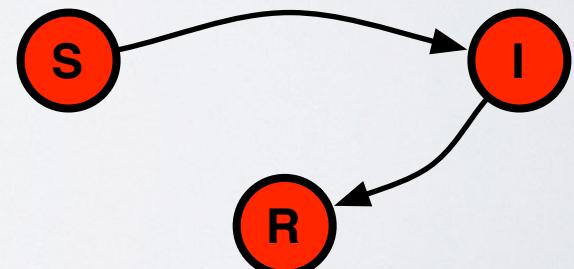
social (information)

- SIS, SIR models
- threshold models

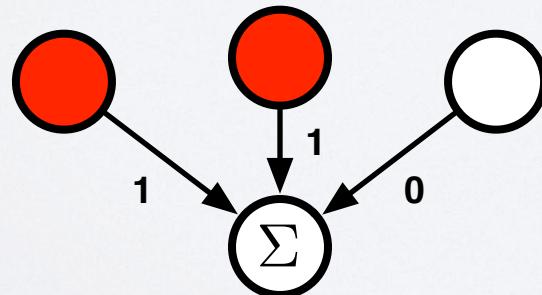
susceptible-infected-susceptible



susceptible-infected-recovered



threshold



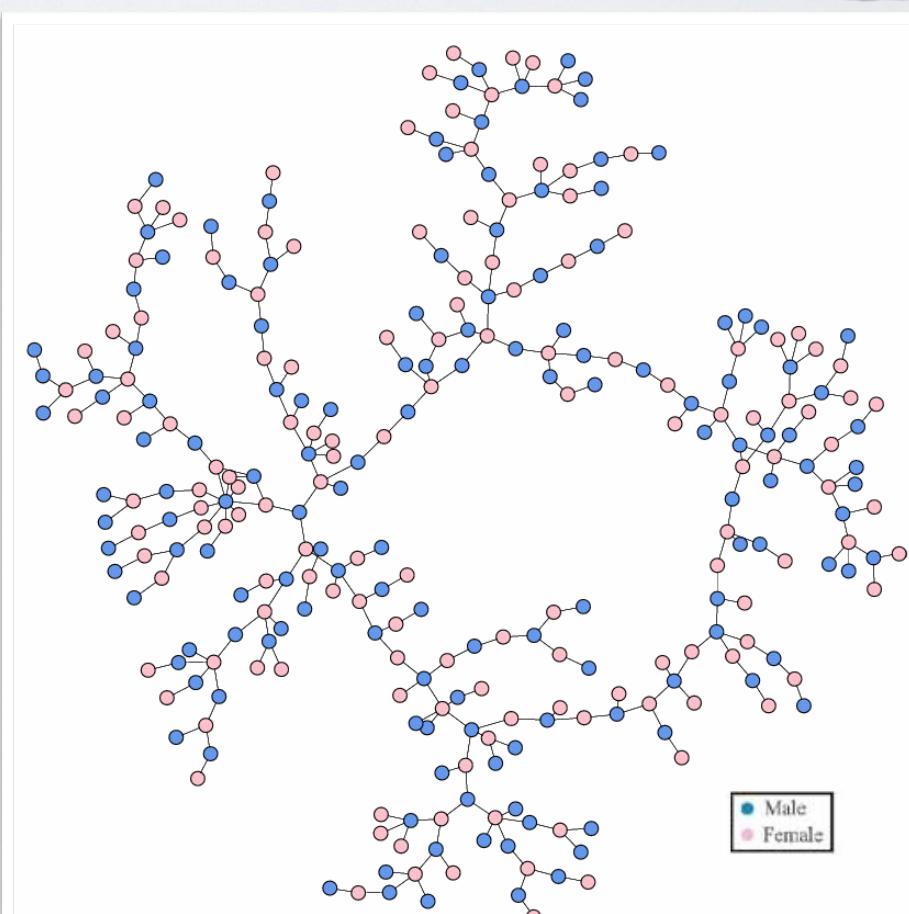
network degrees

Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks

2004

Peter S. Bearman James Moody Katherine Stovel
Columbia University *Ohio State University* *University of Washington*

- relationship network in “Jefferson High”
- this subgraph is 52% of school
- who are most important disease spreaders?



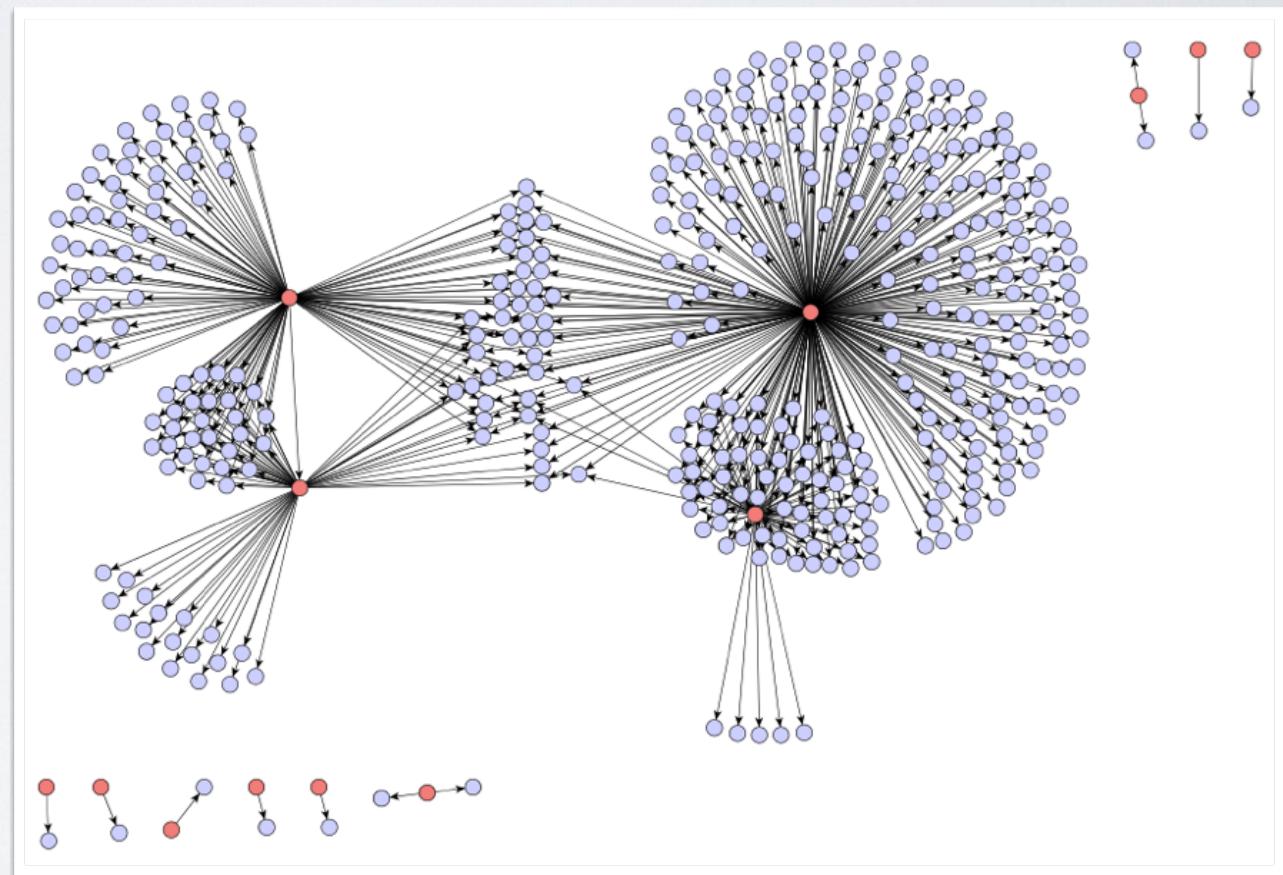
network degrees

The Dynamics of Viral Marketing

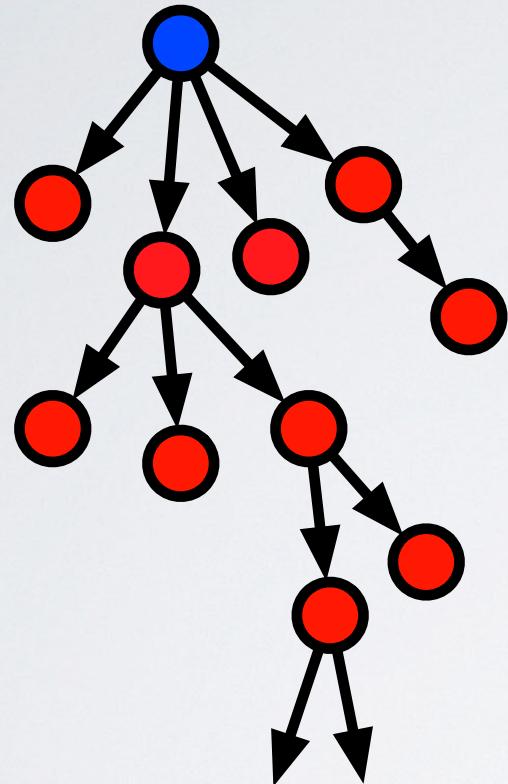
2007

JURE LESKOVEC LADA A. ADAMIC BERNARDO A. HUBERMAN

- amazon.com viral marketing
- viral trace for “Oh my Goddess!” community
- very high degrees!
- most attempts to “influence” fail



network degrees



$$R_0 = 0.923 \dots$$

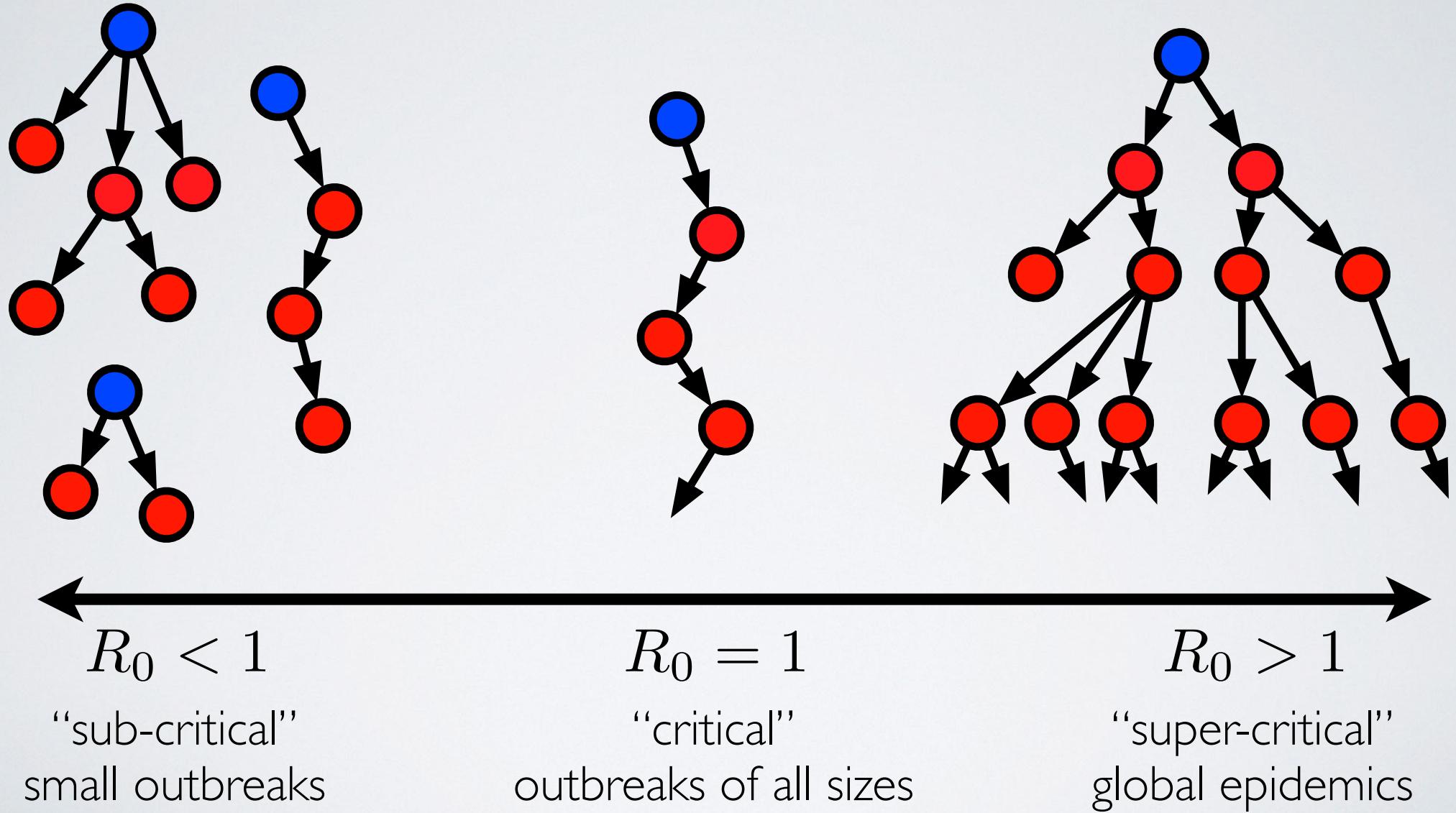
cascade
epidemic
branching process
spreading process

R_0 = net reproductive rate
= average degree $\langle k \rangle$

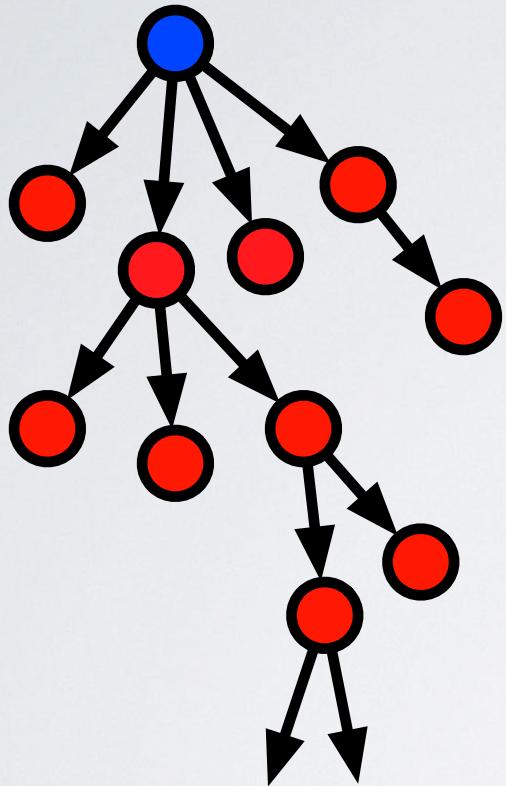
caveat:

ignores network structure,
dynamics, etc.

network degrees



network degrees

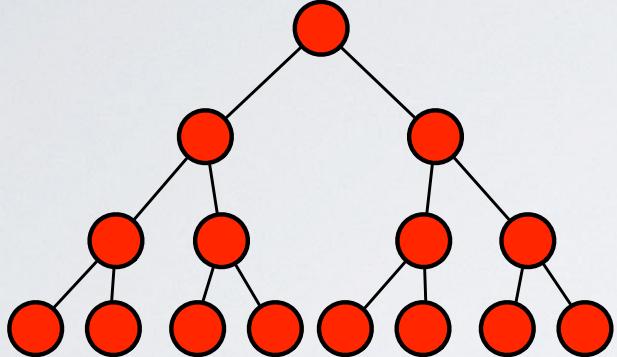


disease	R_0	transmission	vax.
measles	12 – 18	airborne	90 – 95%
chickenpox	7 – 12		85 – 90%
polio	5 – 7	fecal-oral route	82 – 87%
small pox	1.5 – 20+	airborne droplet	70 – 80%
H1N1 flu	1 – 3	airborne droplet	≈ 67%
ebola	1.5 – 2.5	bodily fluids	
zika	2		
covid-19 (wildtype)	≈ 2.4	aerosols	≈ 60%
covid-19 (alpha)	4 – 5	aerosols	75 – 80%
covid-19 (delta)	5 – 8	aerosols	80 – 88%
covid-19 (omicron)	10 – 14	aerosols	90 – 93%



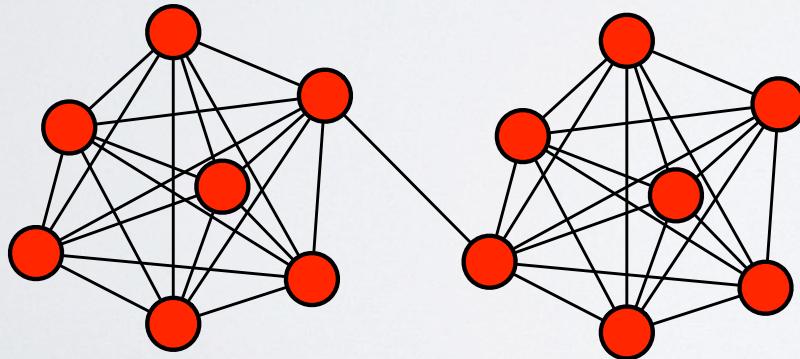
all super-critical

network degrees



bigger cascades

- smaller overlap among neighbors
- more expander-like
[more like a random graph]
- higher transmission probability
- lower activation threshold



smaller cascades

- larger overlap among neighbors
- more triangles
- smaller "communities"
- more spatial-like organization
- lower transmission probability
- higher activation threshold

[Volz, J. Math. Bio. 56, 293–310 \(2008\)](#)

[Bansal et al., J. Royal Soc. Interface 4, 879–891 \(2007\)](#)

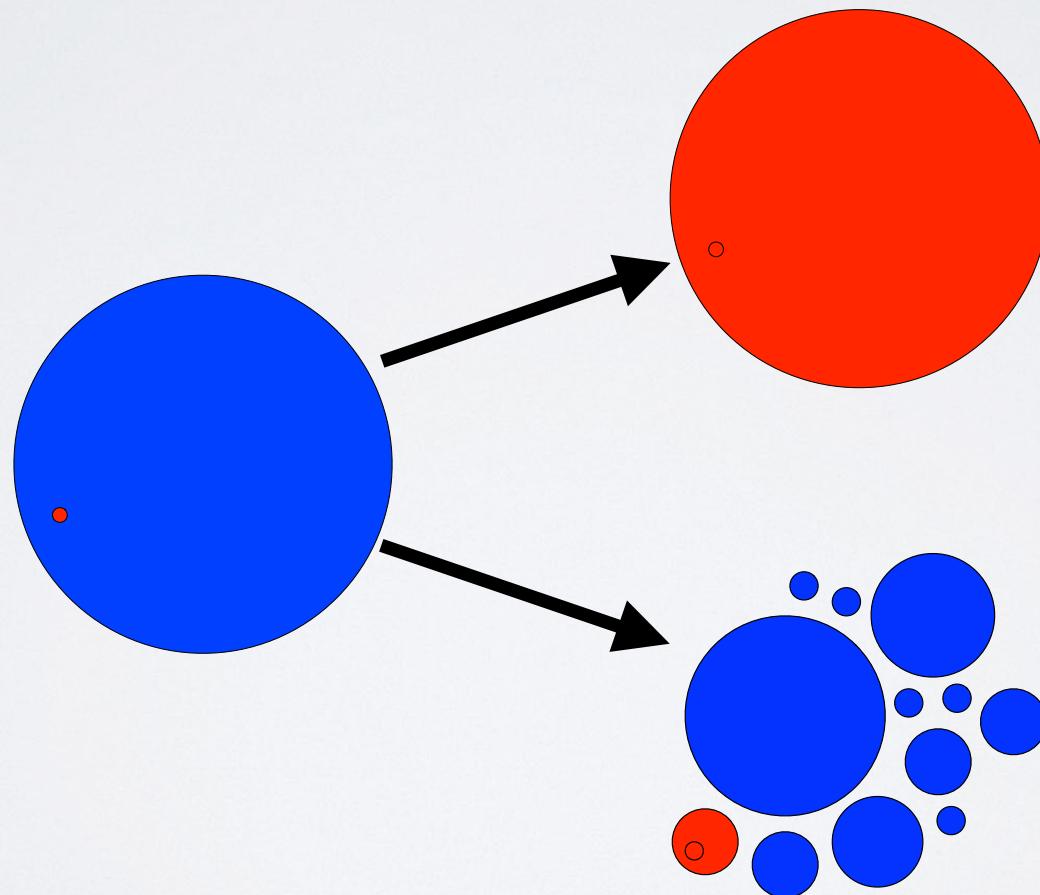
[Karrer and Newman, Phys. Rev. E 82, 016101 \(2010\)](#)

[Salathe and Jones, PLoS Comp. Bio. 6, e1000736 \(2010\)](#)

network degrees

how could we halt the spread?

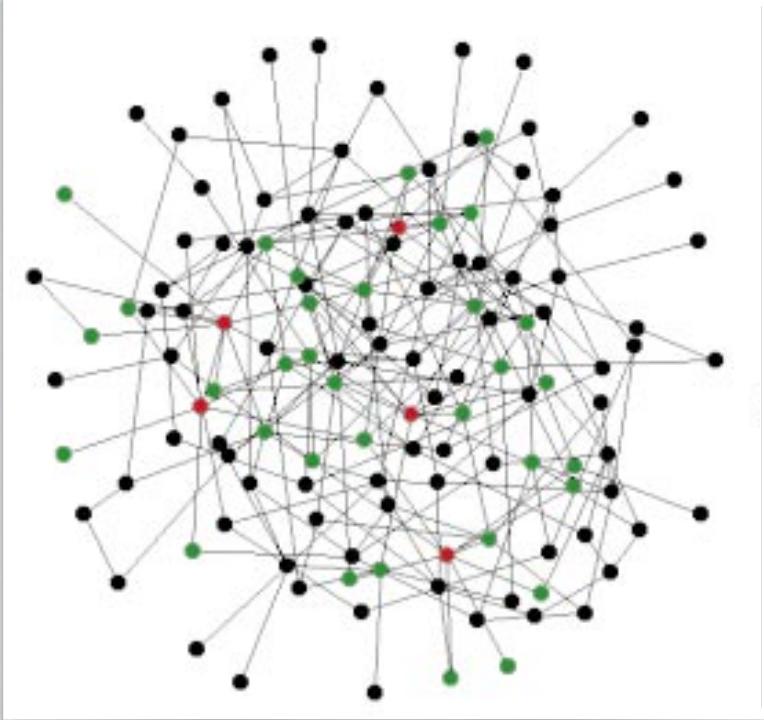
- break network into disconnected pieces



network degrees

two networks

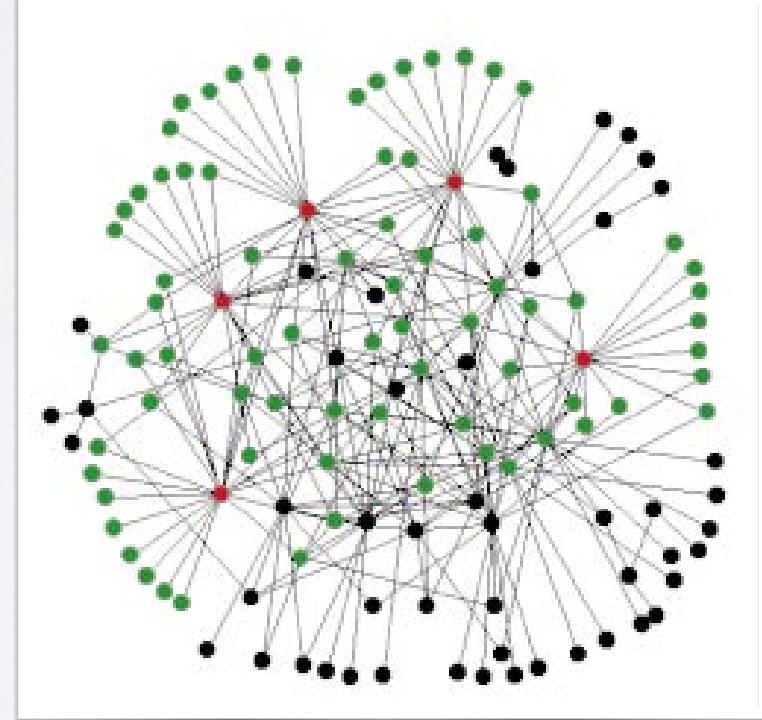
homogeneous in degree



Error and attack tolerance
of complex networks 2000

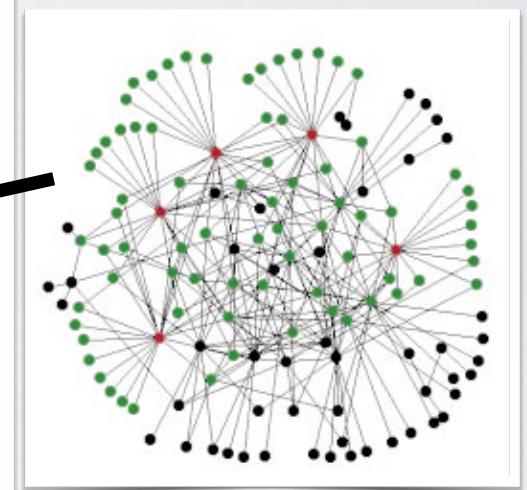
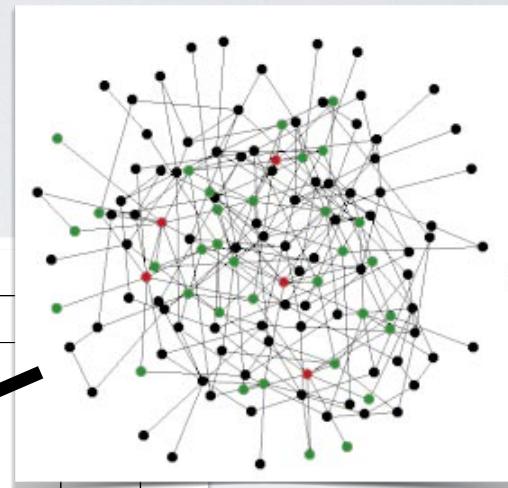
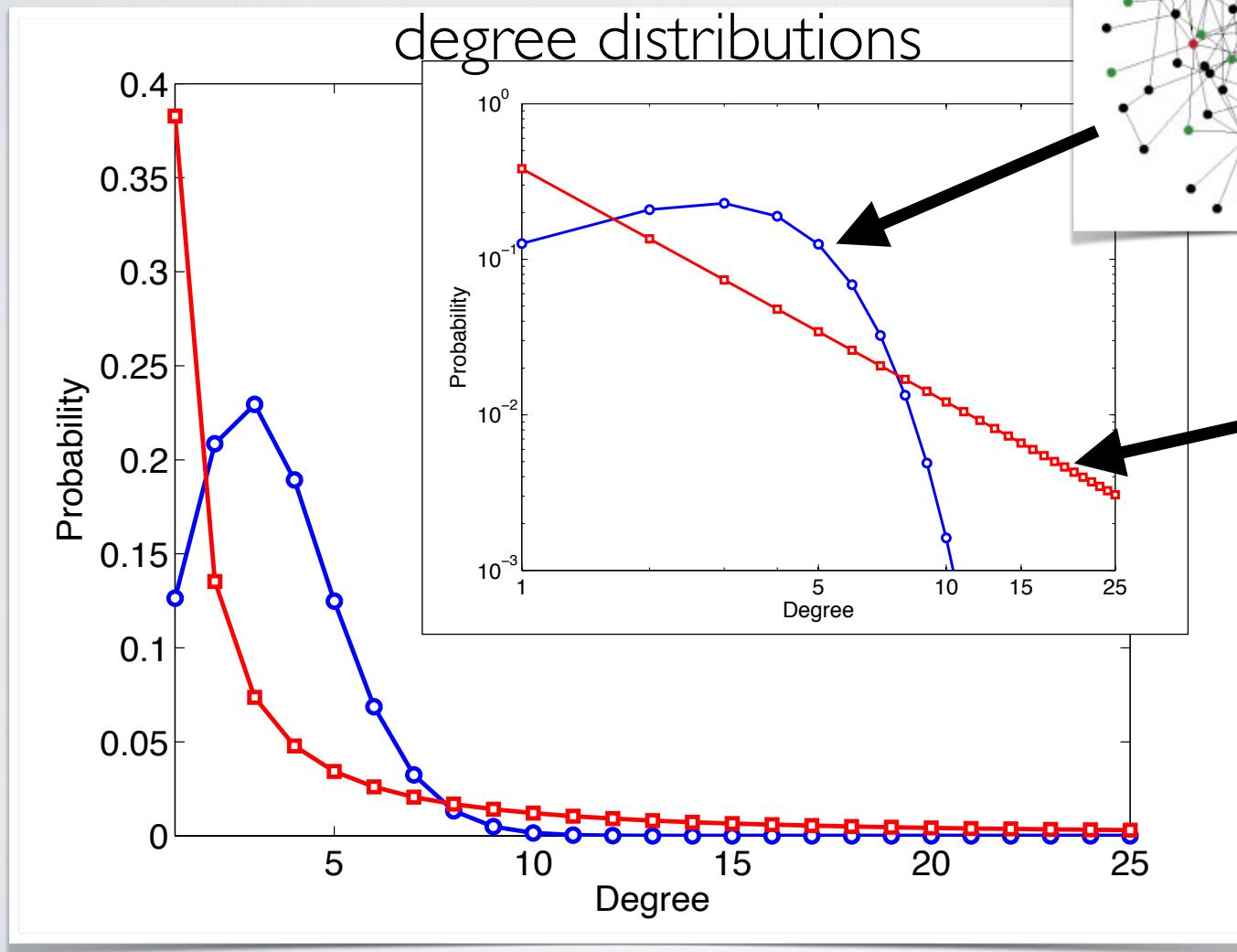
Réka Albert, Hawoong Jeong & Albert-László Barabási

heterogeneous in degree



network degrees

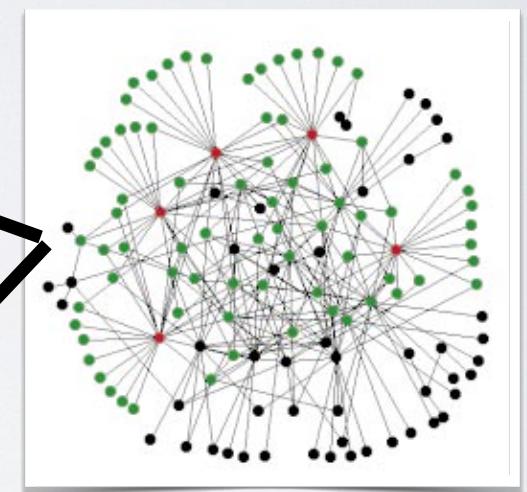
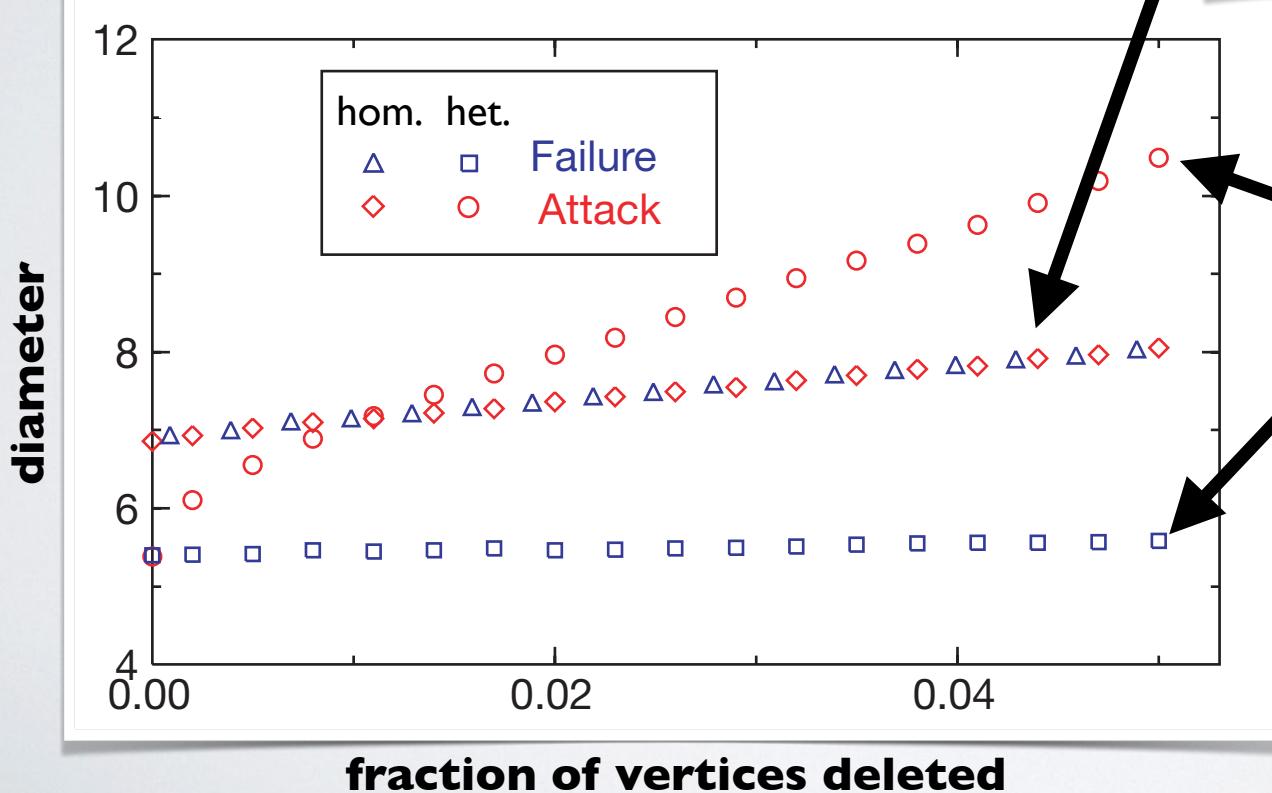
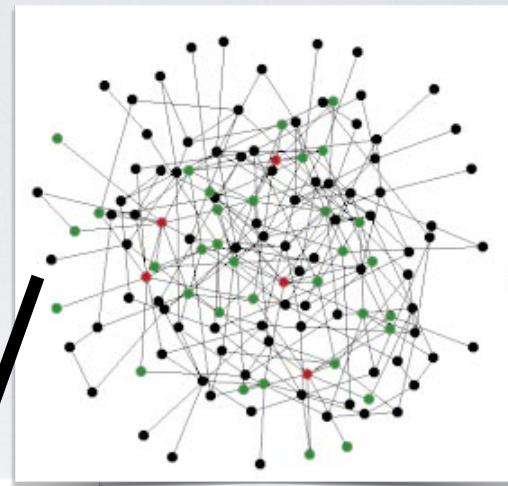
two networks



network degrees

strategy: delete vertices

1. uniformly at random ("failure")
2. in order of degree ("attack")

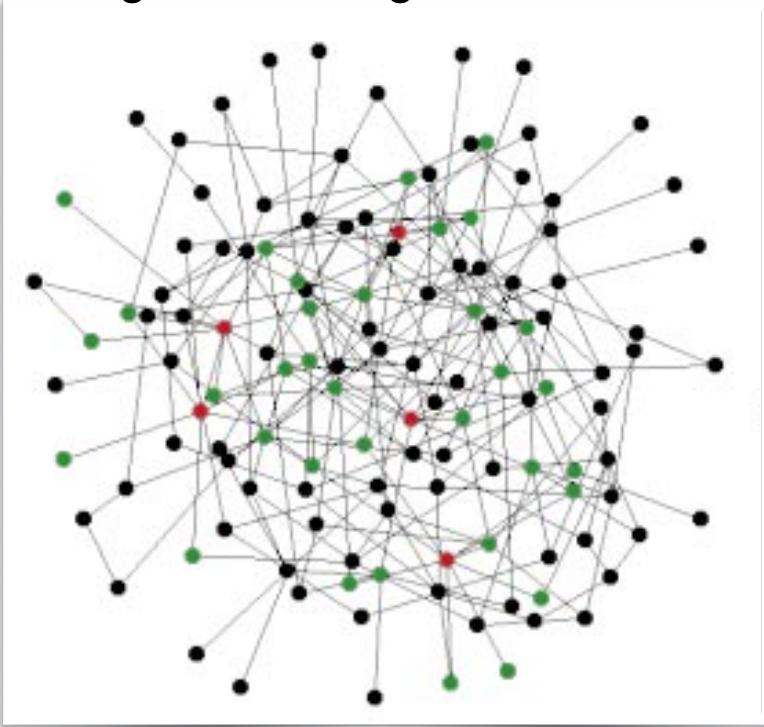


network degrees

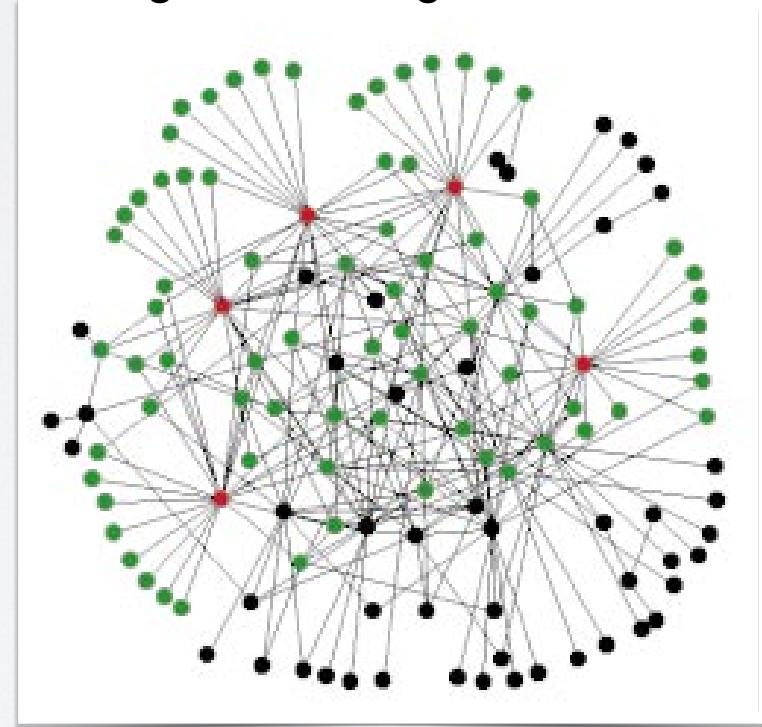
what promotes spreading?

- high-degree vertices*
- centrally-located vertices

homogeneous in degree



heterogeneous in degree

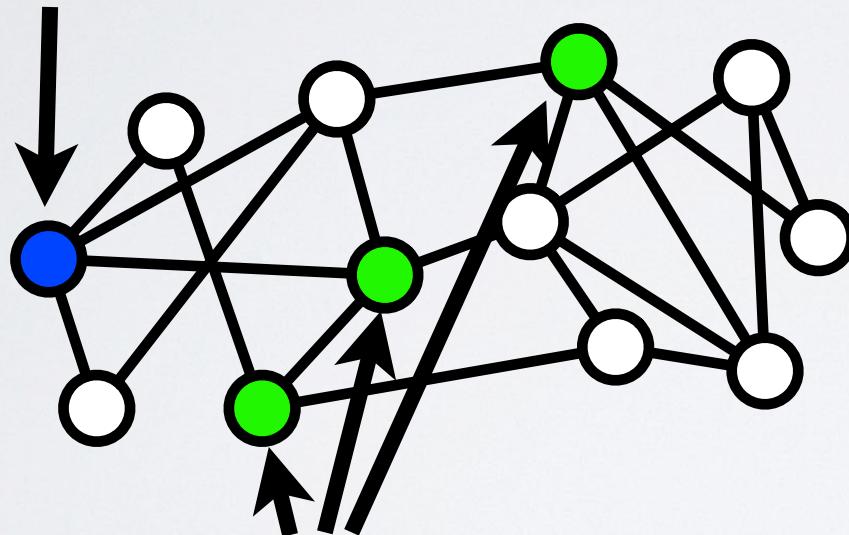


network degrees

strategy: delete vertices

3. build “fire breaks”

patient 0



vaccinated = deleted
("fire break")

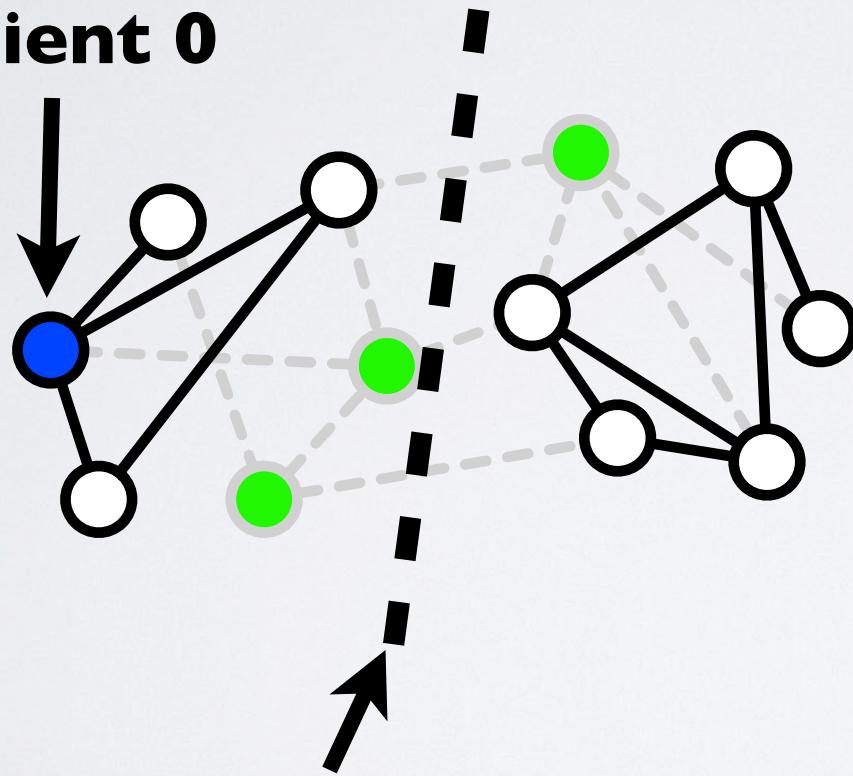
software packages for simulating epidemics on networks

1. Epidemics on Networks (EoN) <https://epidemicsonnetworks.readthedocs.io/en/latest/>

2. SEIR+ Model <https://github.com/ryansmcgee/seirsplus>

network degrees

patient 0



effective buffer

• vaccination strategies

- the “front line” (hospitals)
- high degree nodes
- the vulnerable (old/young)

software packages for simulating epidemics on networks

1. Epidemics on Networks (EoN) <https://epidemicsonnetworks.readthedocs.io/en/latest/>

2. SEIR+ Model <https://github.com/ryansmcgee/seirsplus>

network degrees

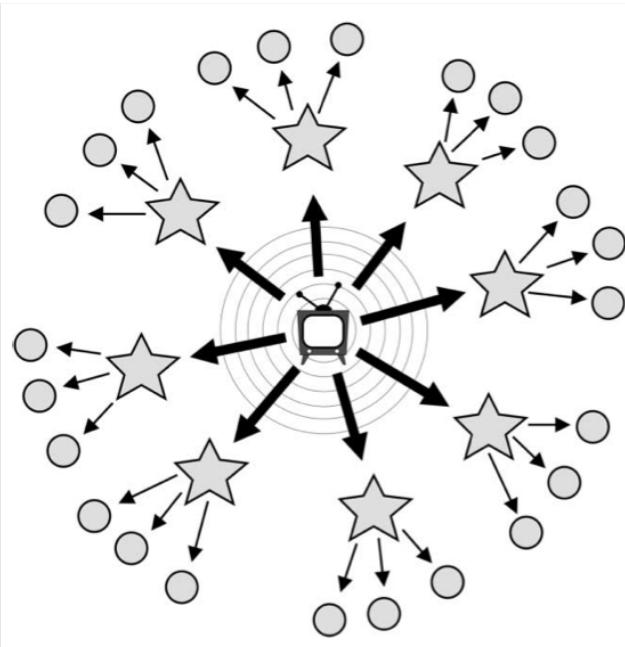
but, in social networks...

network degrees

Influentials, Networks, and Public Opinion Formation

DUNCAN J. WATTS
PETER SHERIDAN DODDS*

2007



broadcast influence

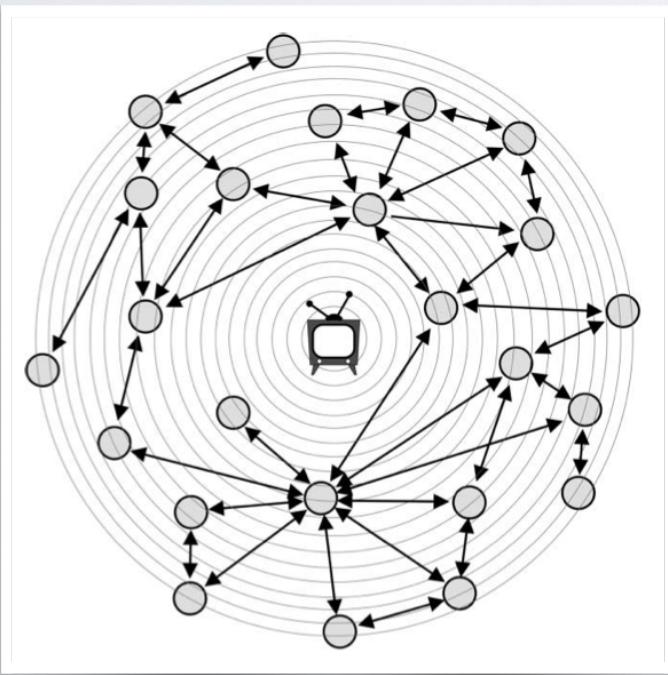
- classic information marketing
- message saturation
- **degree** is most important

network degrees

Influentials, Networks, and Public Opinion Formation

DUNCAN J. WATTS
PETER SHERIDAN DODDS*

2007



network influence

- “network” (decentralized) marketing
- high-degree = “opinion leader”
- high-degree alone = **irrelevant**
- a cascade requires a legion of *susceptibles* (a system-level property)

network degrees

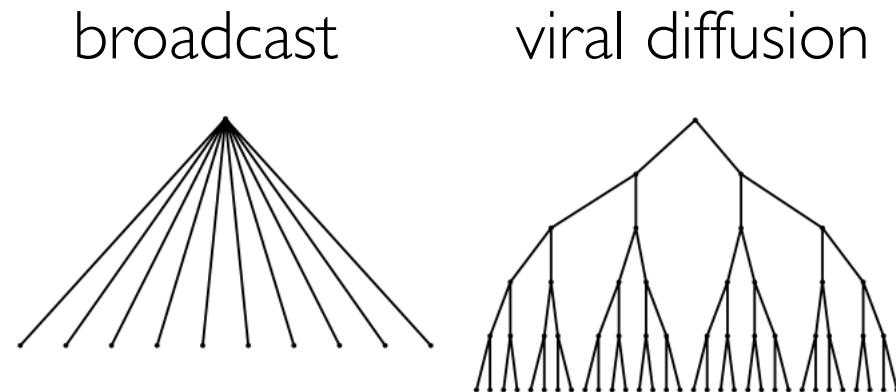
The Structural Virality of Online Diffusion

Sharad Goel, Ashton Anderson

Stanford University, Stanford, California, 94305 {scgoel@stanford.edu, ashton@cs.stanford.edu}

Jake Hofman, Duncan J. Watts

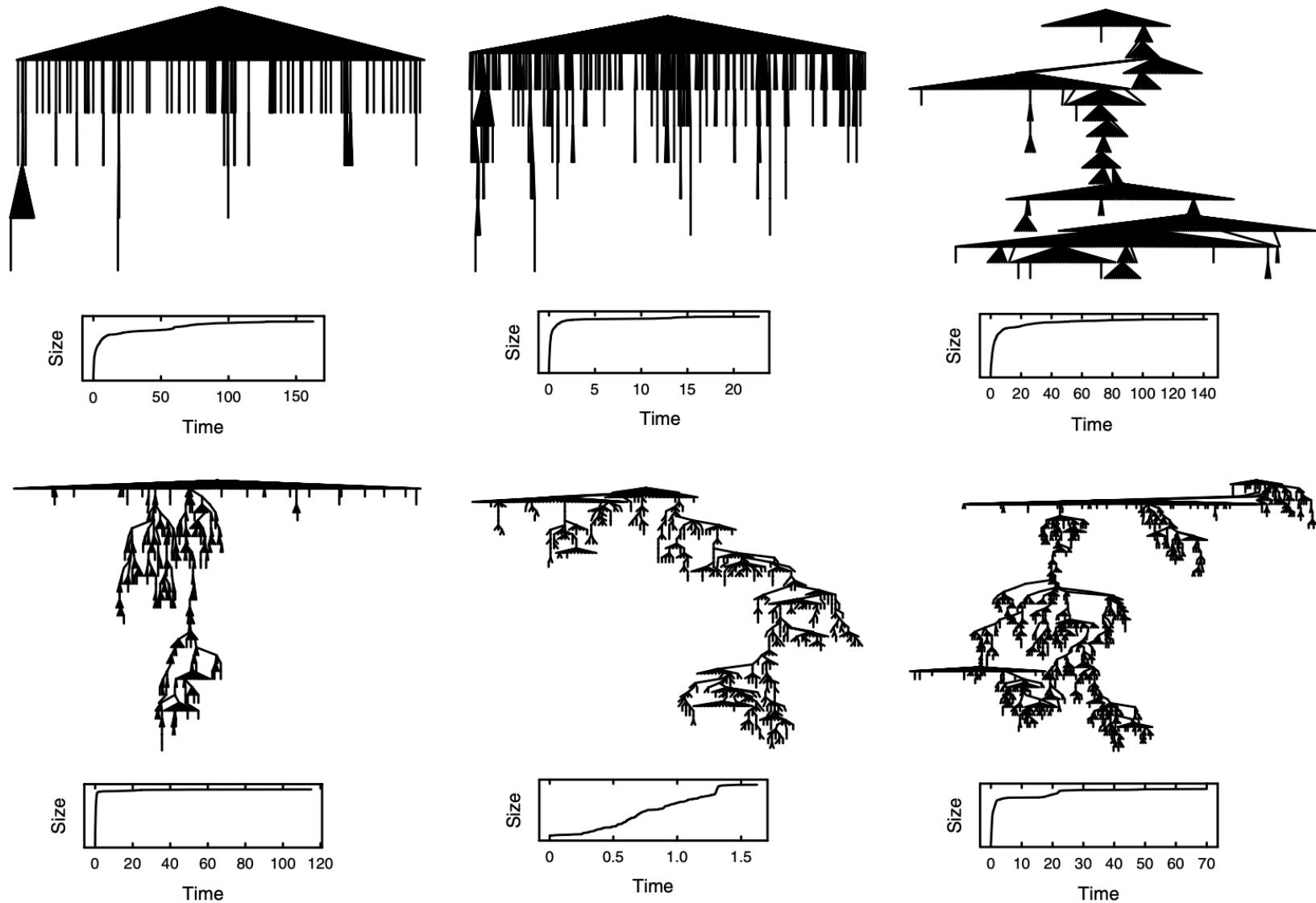
Microsoft Research, New York, New York 10016 {jmh@microsoft.com, duncan@microsoft.com} 2015



- 1 billion diffusion events, on twitter
- virality measure for each cascade
- cascade sizes are extremely high variance (maybe power law...)

network degrees

Figure 3 A Random Sample of Cascades Stratified and Ordered by Increasing Structural Virality, Ranging from 2 to 50

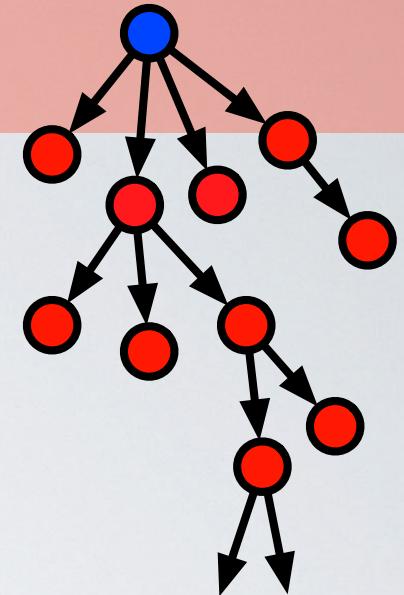


- enormous diversity of cascade shapes, depths

network degrees

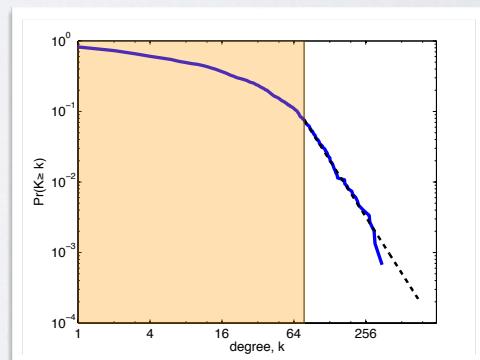
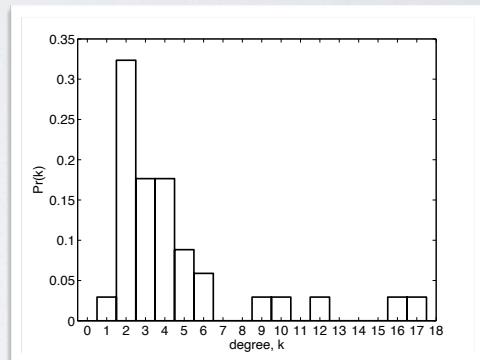
degrees:

- first-order description of network structure
- direct implications for spreading processes
- cascades require both susceptible population and spreaders



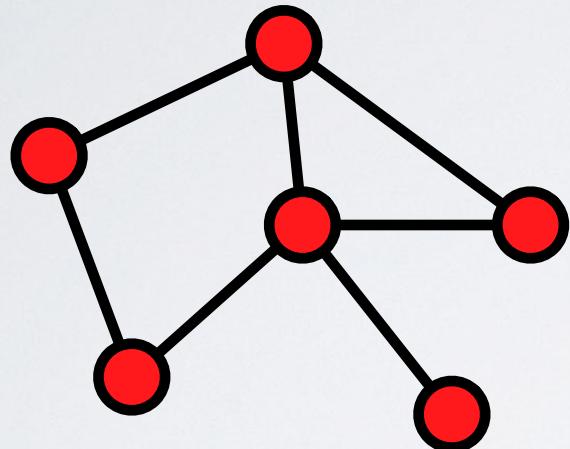
open questions:

- impact of degrees on other dynamics
- feedback from dynamics to degree [adaptive behaviors like self-quarantine, evangelism]
- when does degree not matter

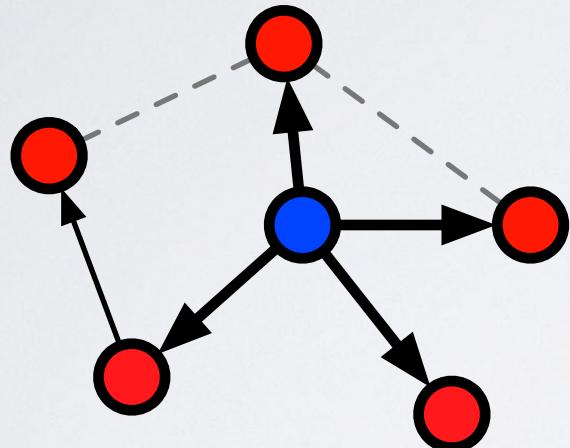


describing networks

position



describing networks

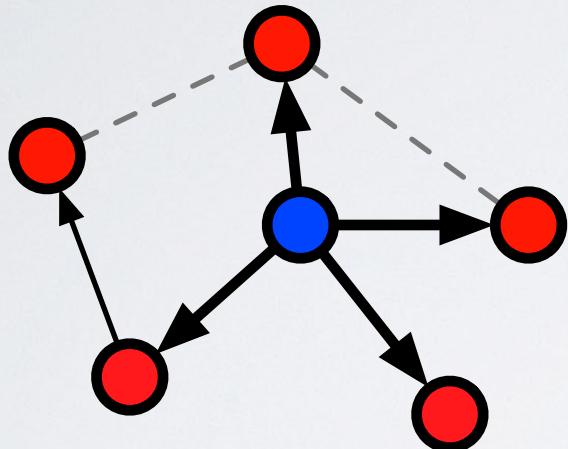


position = centrality:
structural vs. dynamical
importance

geometric connectivity	harmonic centrality
	closeness centrality
	betweenness centrality
	degree centrality
	eigenvector centrality
	PageRank
	Katz centrality
	many many more...

structural importance = cheap
estimate of dynamical importance
(aka "influence")

describing networks



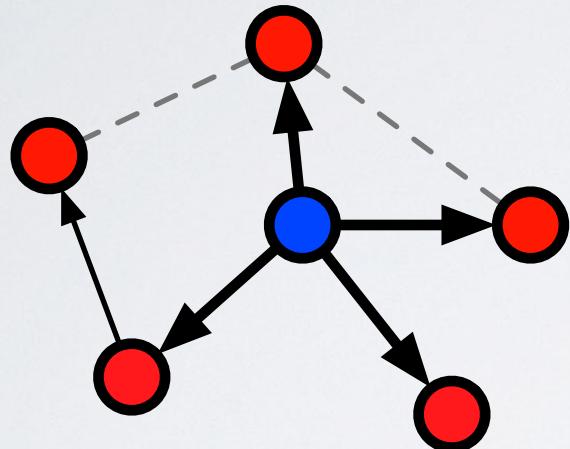
position = centrality:

structural vs. dynamical
importance

centrality = unsupervised
node ranking

$$f: G \rightarrow \vec{v}$$

describing networks



position = centrality:

harmonic, closeness
centrality

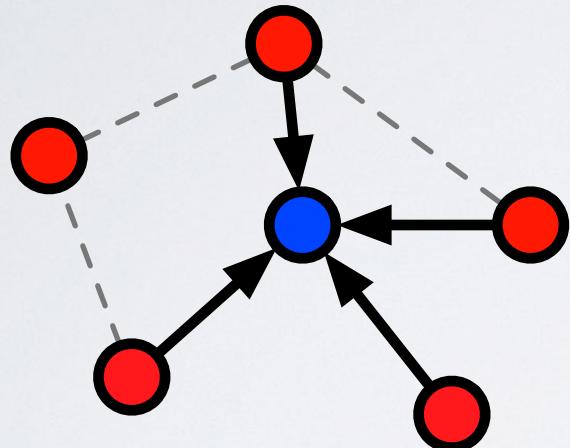
importance = being in
“center” of the network

$$\text{harmonic } c_i = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$$

length of shortest path

distance: $d_{ij} = \begin{cases} \ell_{ij} & \text{if } j \text{ reachable from } i \\ \infty & \text{otherwise} \end{cases}$

describing networks



position = centrality:

PageRank, Katz, eigenvector
centrality

importance = sum of
importances* of nodes that
point at you

$$I_i = \sum_{j \rightarrow i} \frac{I_j}{k_j}$$

or, the right eigenvector of
 $\mathbf{Ax} = \lambda \mathbf{x}$

network position

an example



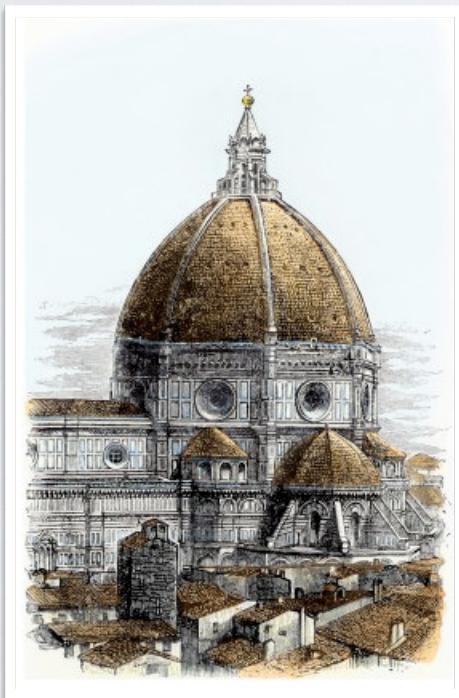
Giovanni de Medici

network position

Robust Action and the Rise of the Medici, 1400–1434¹

John F. Padgett and Christopher K. Ansell

1993



Duomo

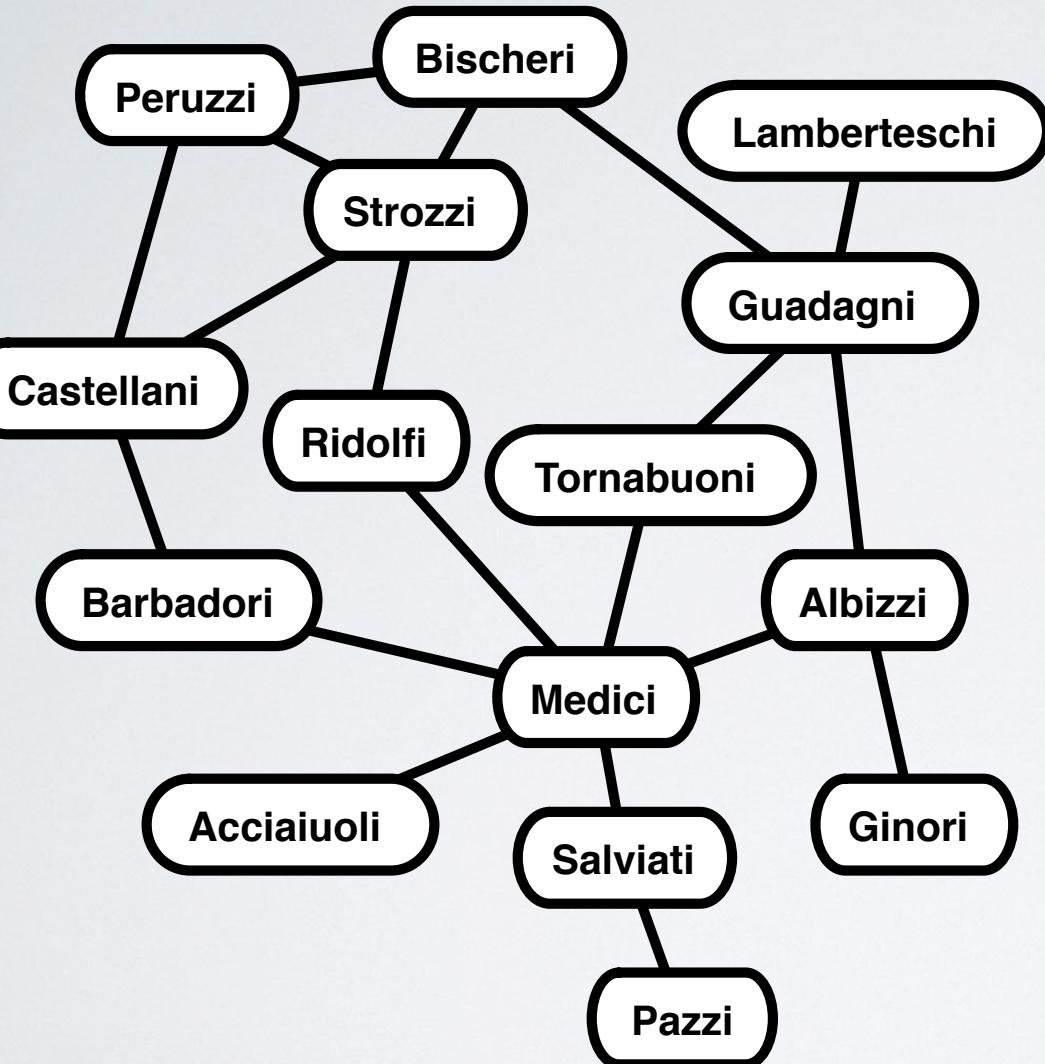


Palazzo Medici



Giovanni de Medici

network position: harmonic

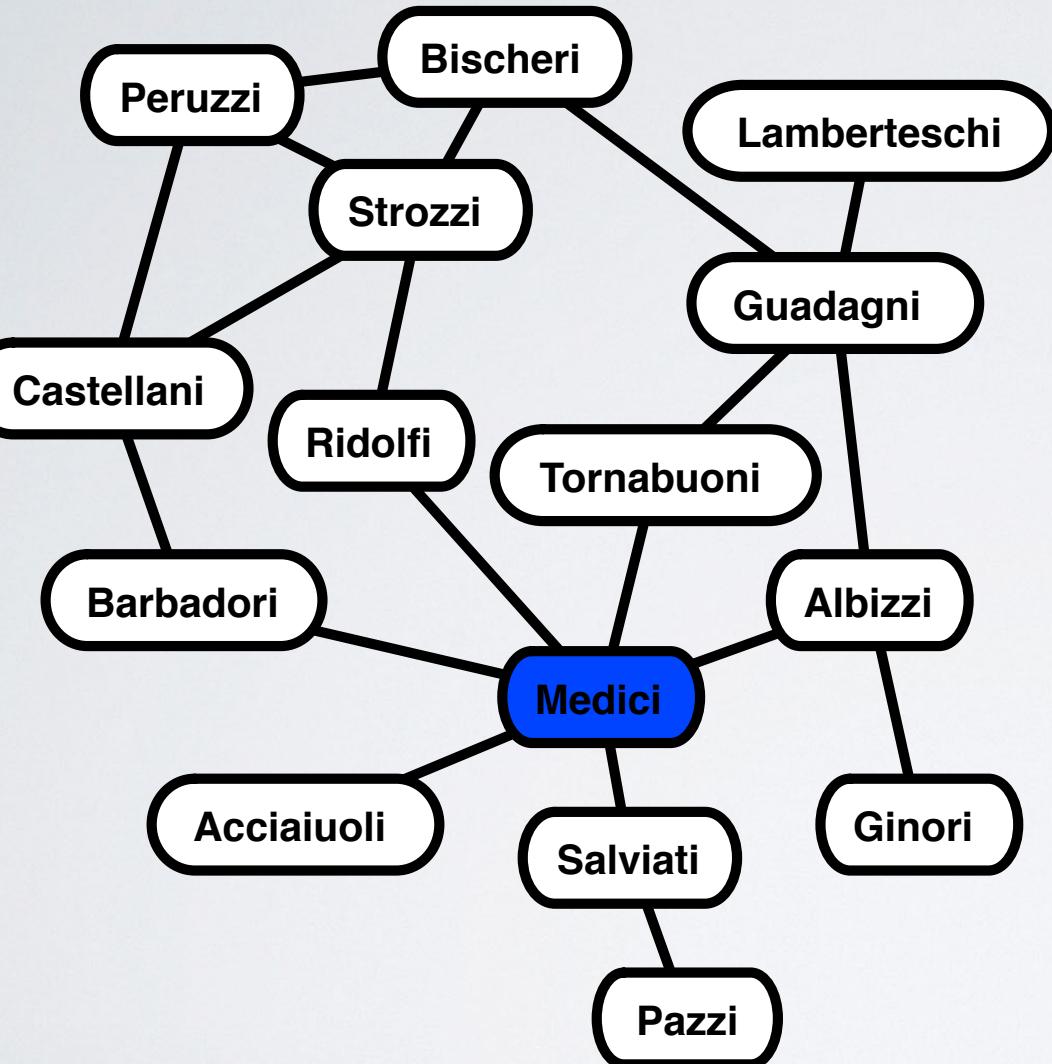


nodes: Florence families

edges: inter-family marriages

**which family is
most central?**

network position: closeness

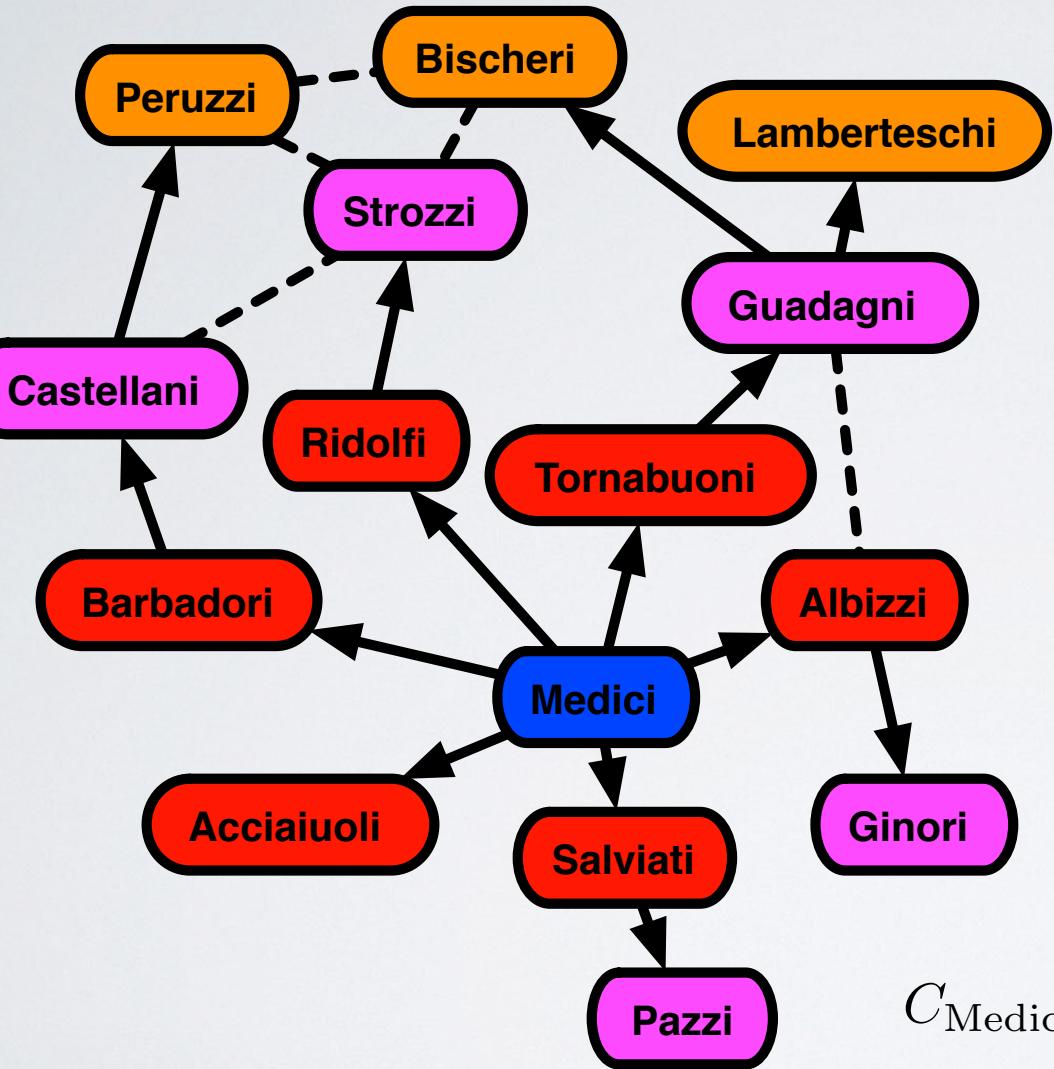


nodes: Florence families
edges: inter-family marriages

Medici?



network position: closeness



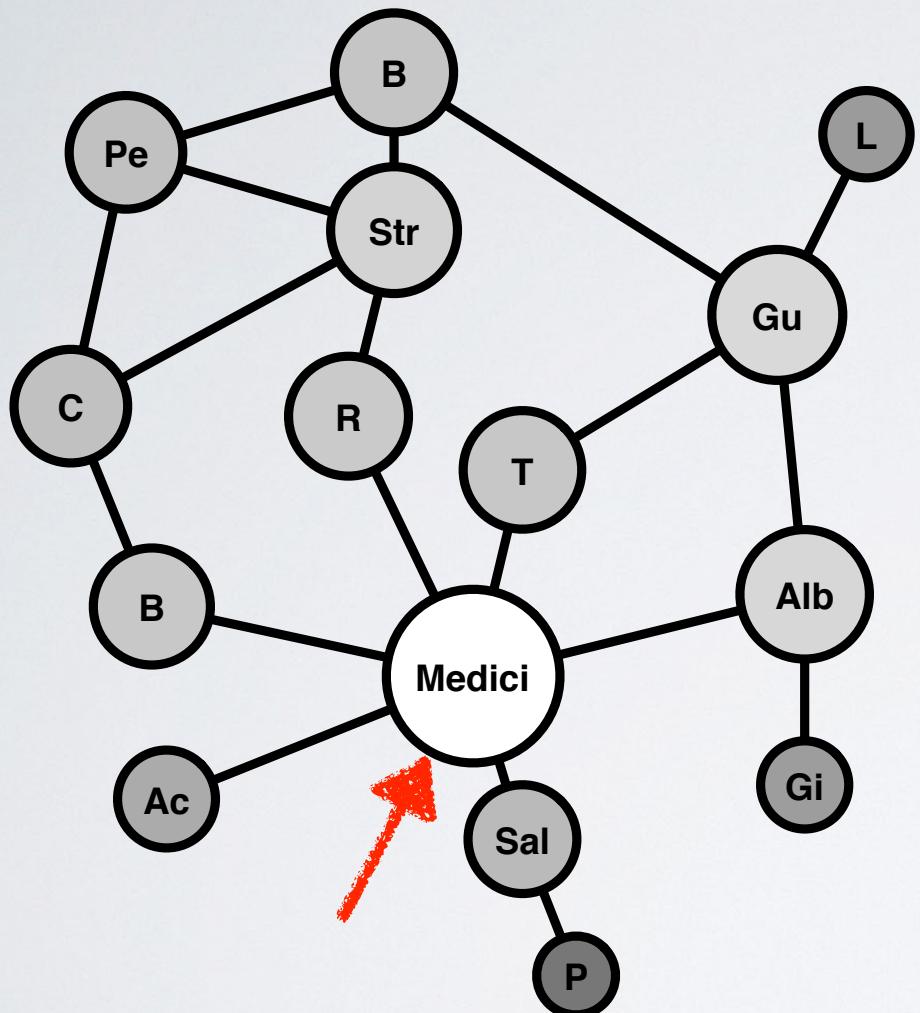
nodes: Florence families

edges: inter-family marriages

Medici.

$$C_{\text{Medici}} = 6 \left(\frac{1}{1} \right) + 5 \left(\frac{1}{2} \right) + 3 \left(\frac{1}{3} \right)$$
$$= 9.5$$

network position: harmonic

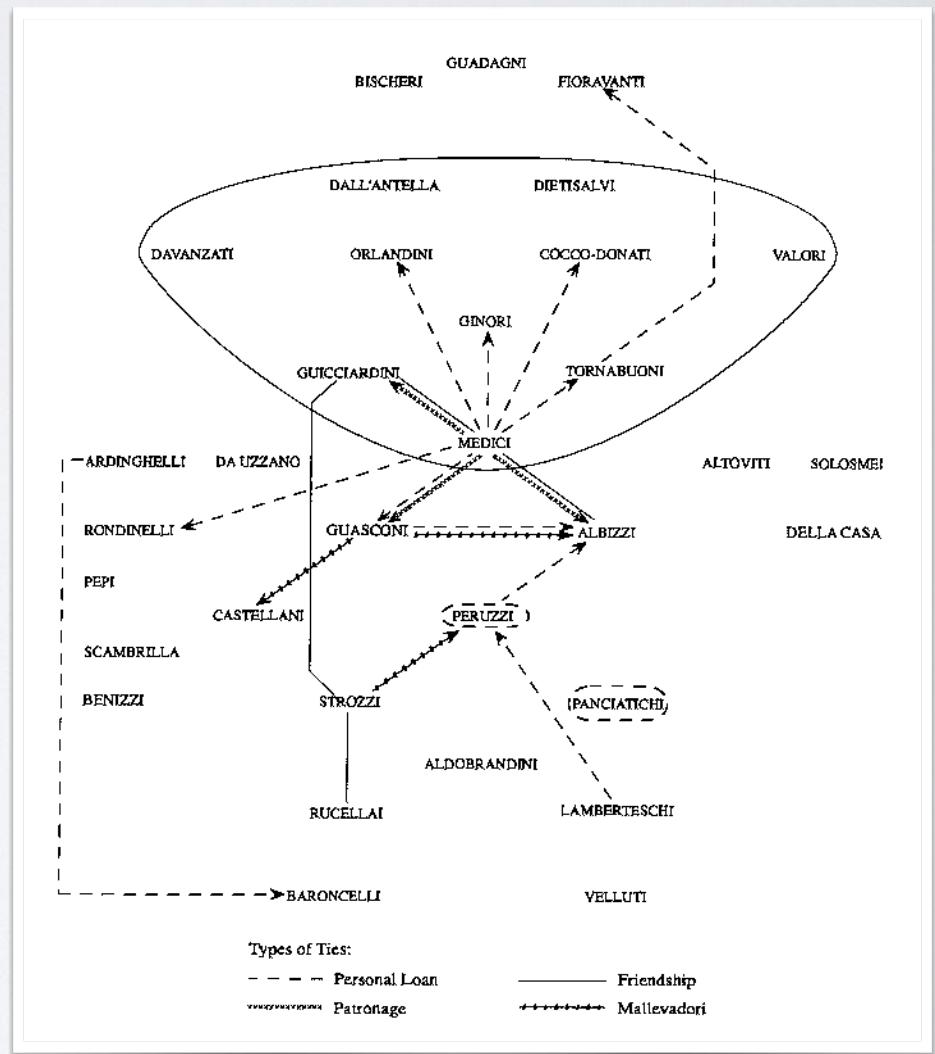
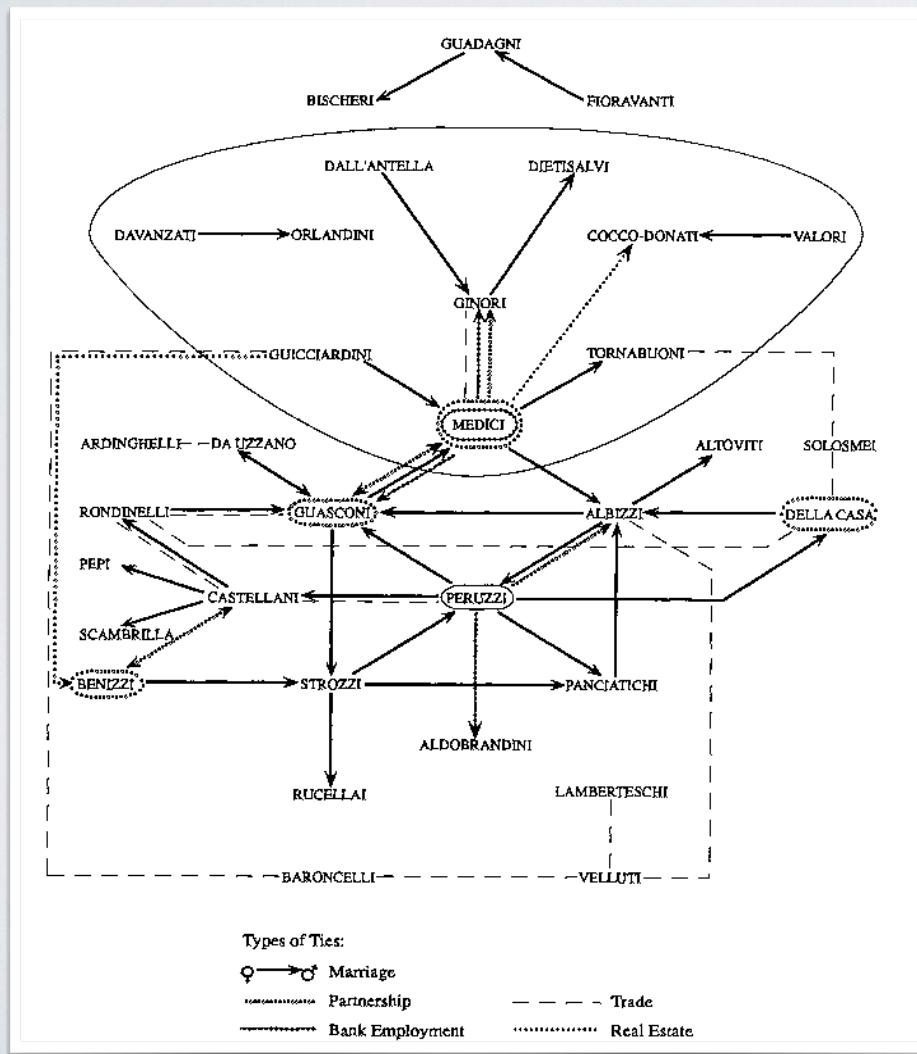


Medici	9.5
Guadagni	7.92
Albizzi	7.83
Strozzi	7.67
Ridolfi	7.25
Bischeri	7.2
Tornabuoni	7.17
Barbadori	7.08
Peruzzi	6.87
Castellani	6.87
Salviati	6.58
Acciaiuoli	5.92
Ginori	5.33
Lamberteschi	5.28
Pazzi	4.77

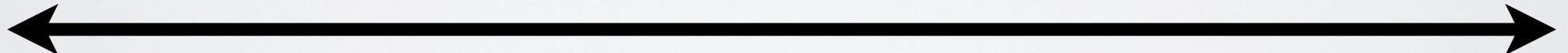


network position

actually, it's complicated...



network position

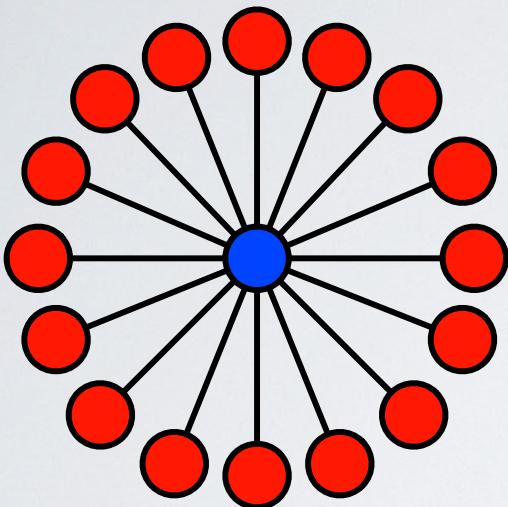


most
centralized

vast wilderness
of in-between

most
decentralized

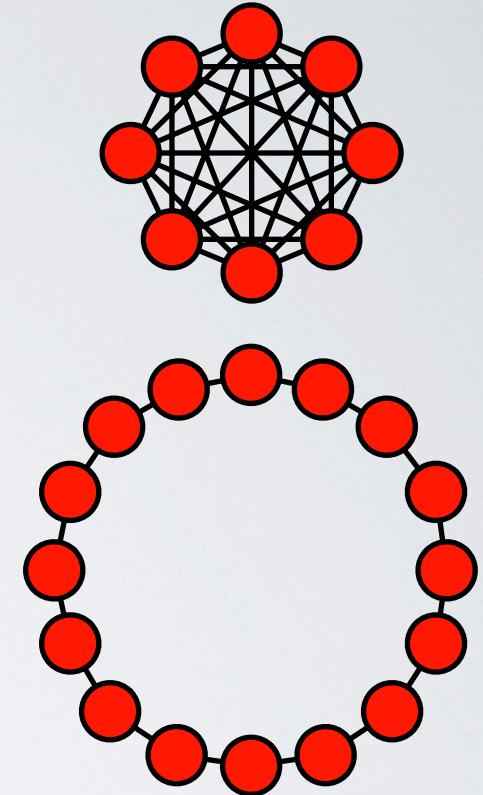
network position



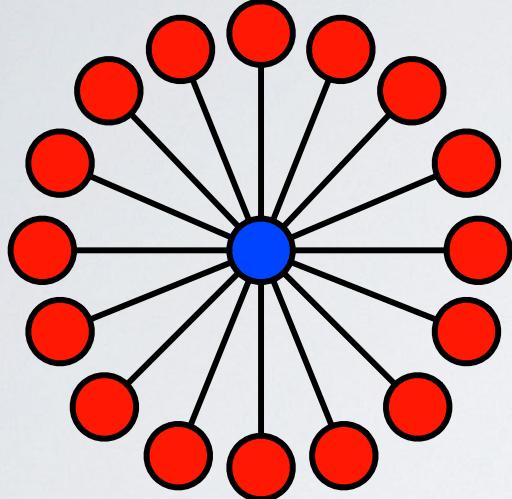
most
centralized

vast wilderness
of in-between

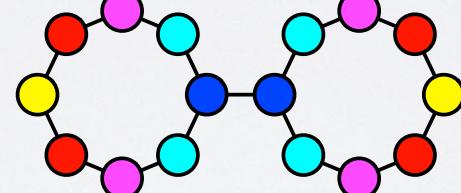
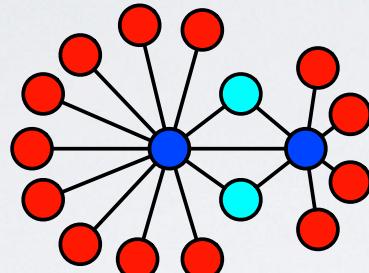
most
decentralized



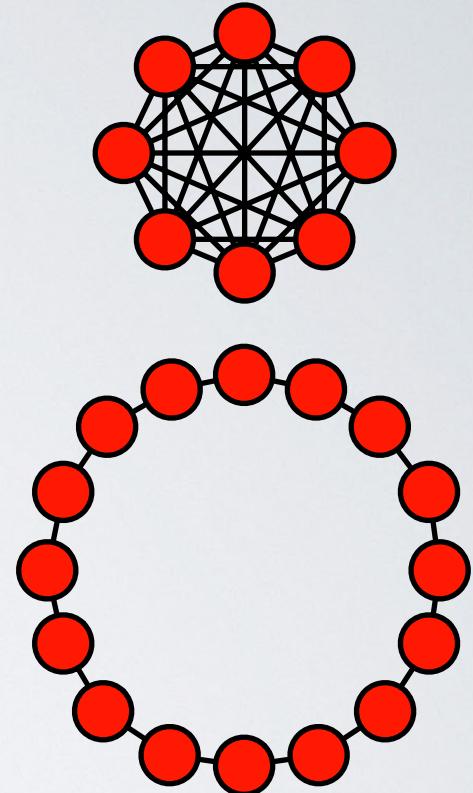
network position



most
centralized

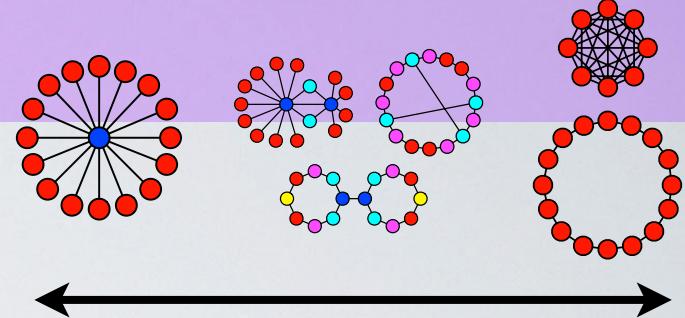


vast wilderness of in-between



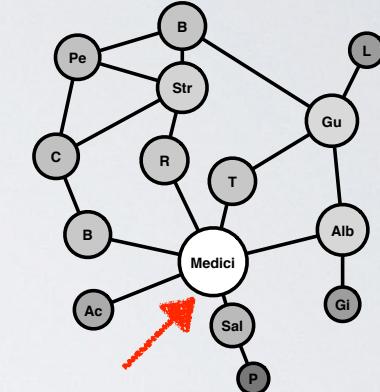
most decentralized

network position



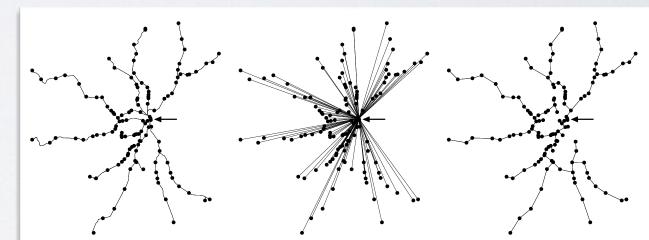
positions:

- geometric description of network structure
- core vs. periphery
- centrality = importance, influence
- nearly all centrality scores highly correlated

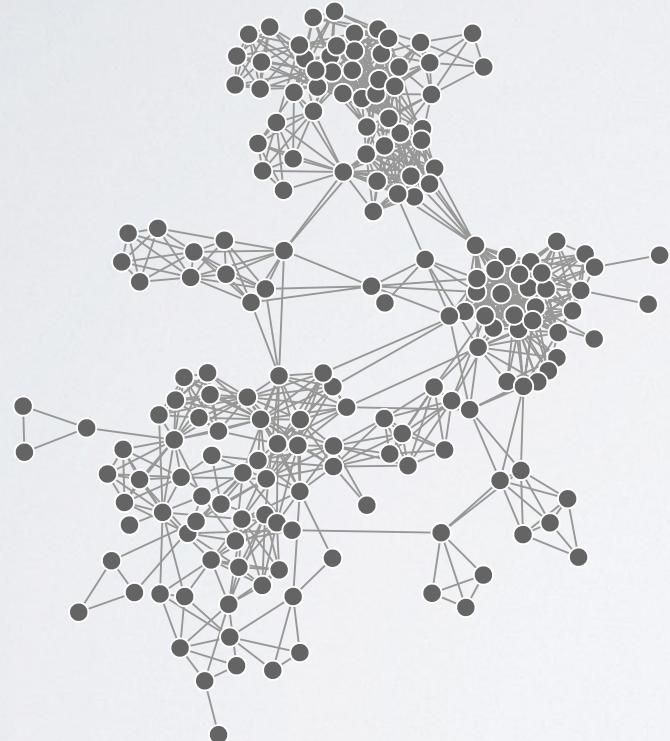


open questions:

- position and dynamics
- what does position predict?
- when does position *not* matter?

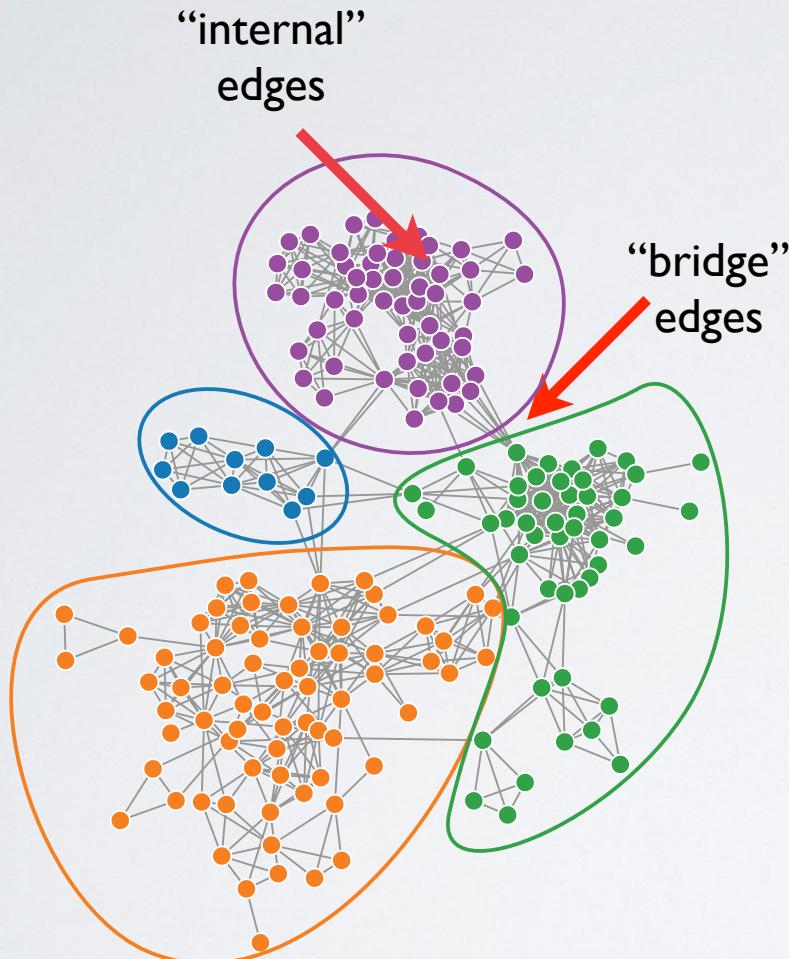


describing networks



community structure

community structure



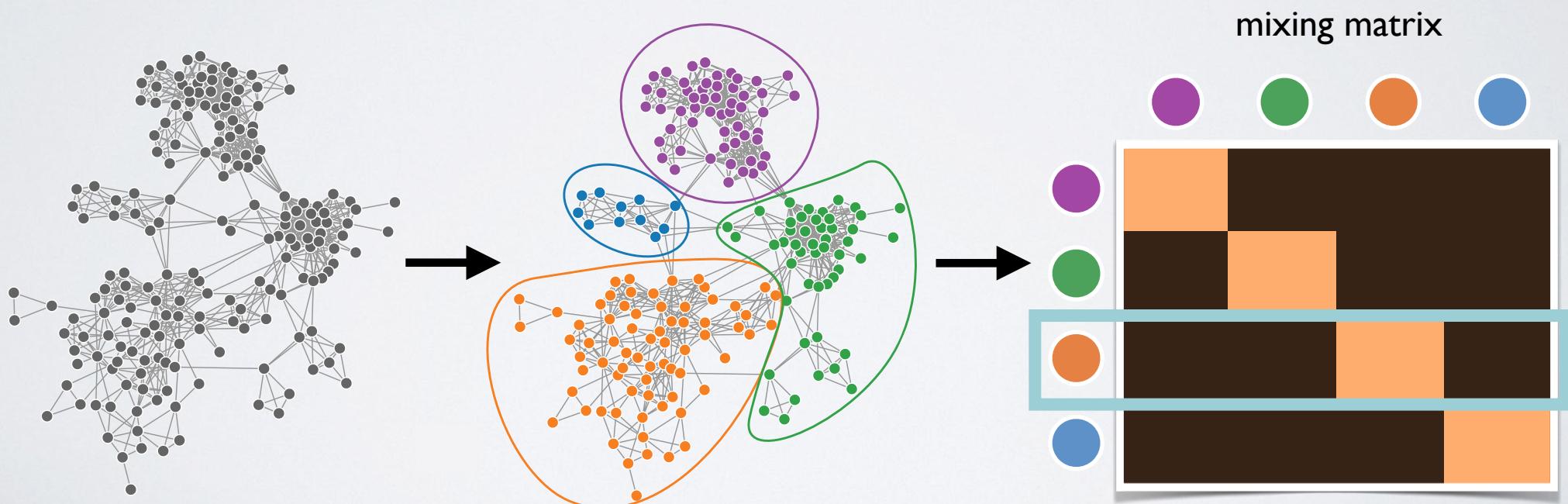
assortative community structure
(edges inside the groups)

community structure:
a group of vertices that
connect to other groups in
similar ways

community structure

community structure:

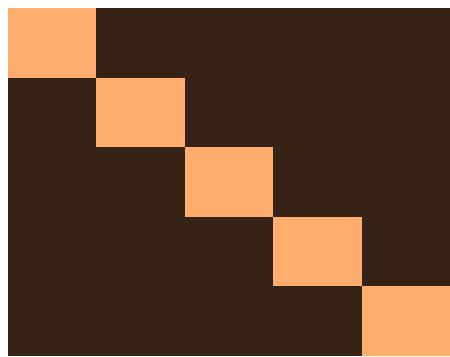
a group of vertices that connect to other groups in similar ways



community structure

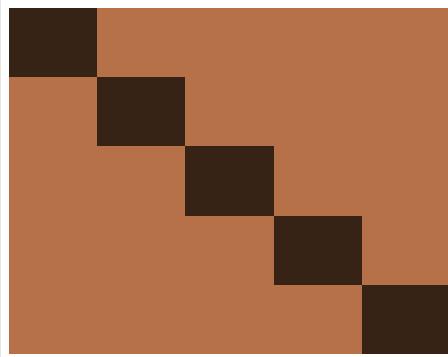
assortative

edges within groups



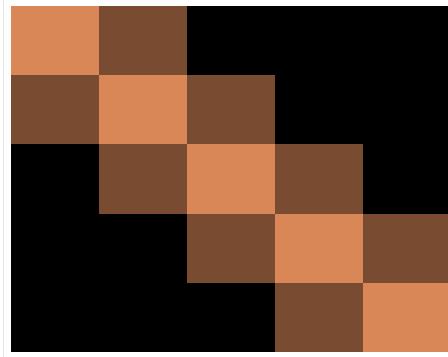
disassortative

edges between groups



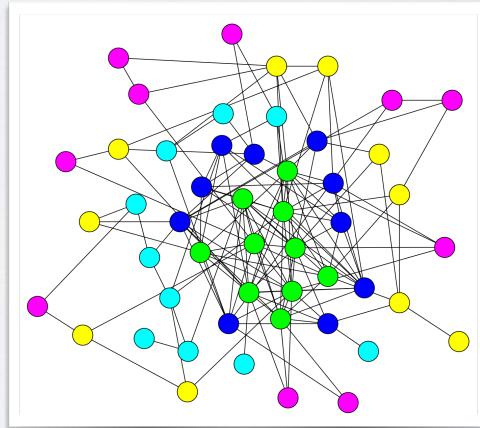
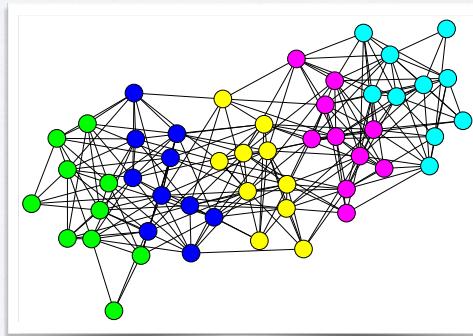
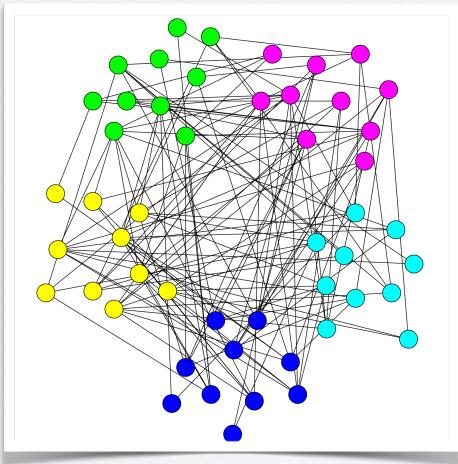
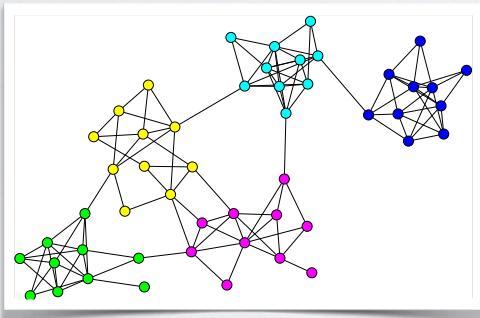
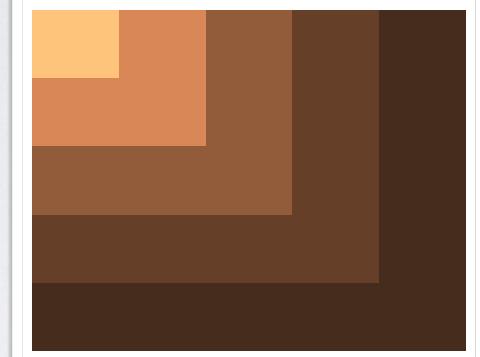
ordered

linear group hierarchy



core-periphery

dense core, sparse periphery



community structure

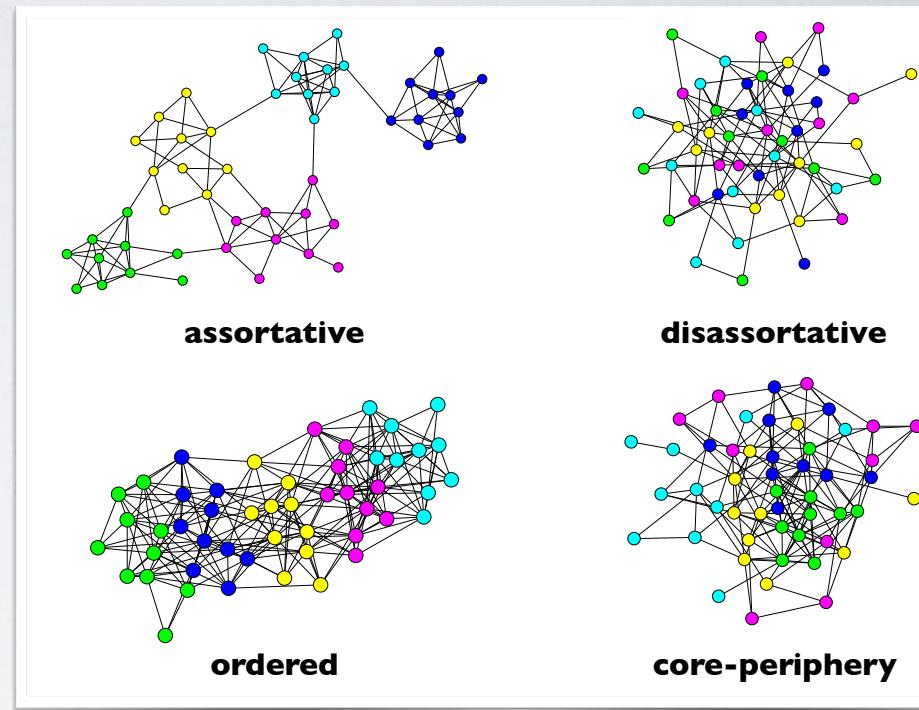
- enormous interest, especially since 2000
- dozens of algorithms for extracting various large-scale patterns
- hundreds of papers published
- spanning Physics, Computer Science, Statistics, Biology, Sociology, and more
- this was one of the first:

Community structure in social and biological networks

M. Girvan*†‡ and M. E. J. Newman*§

PNAS 2002

12,421+ citations on Google Scholar

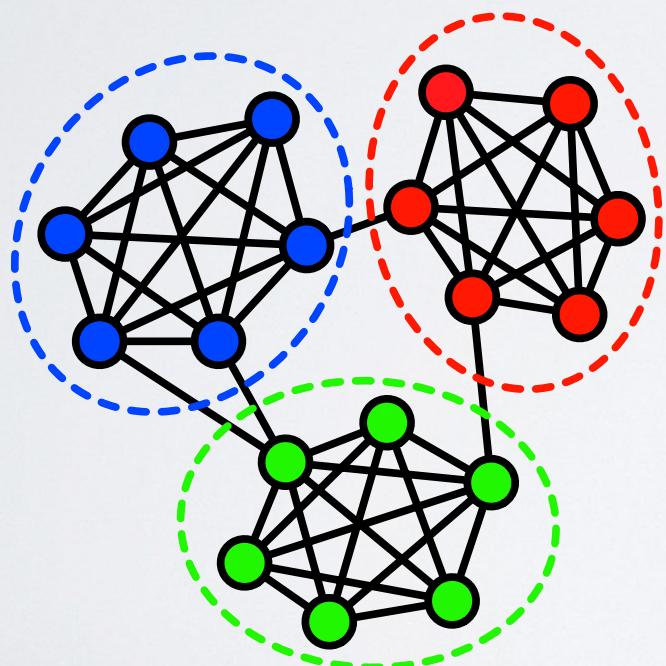


network communities

THE STRENGTH OF WEAK TIES: A NETWORK THEORY REVISITED

1983

Mark Granovetter



most new job opportunities from
“weak ties”

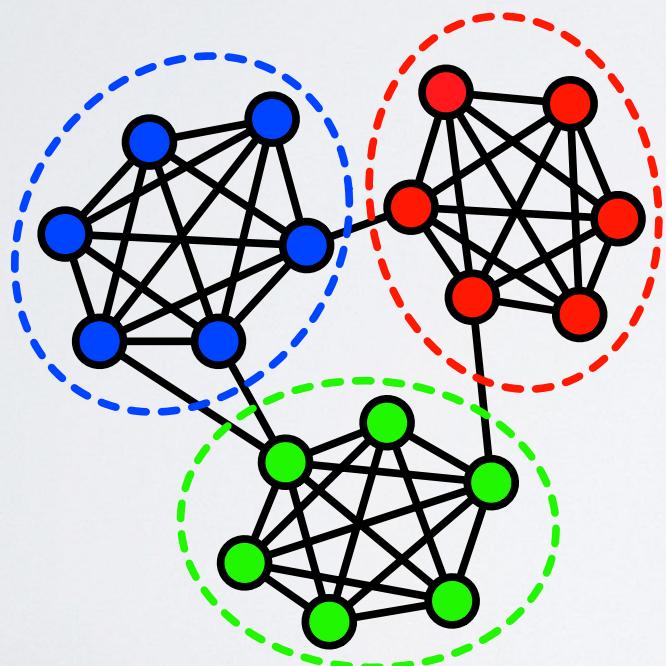
- within-community links = strong
- bridge links = weak

network communities

THE STRENGTH OF WEAK TIES: A NETWORK THEORY REVISITED

1983

Mark Granovetter



most new job opportunities from
“weak ties”

- within-community links = strong
- bridge links = weak

why?

information propagates quickly
within a community,
but slowly between communities

network communities

Finding community structure in very large networks

Aaron Clauset, M. E. J. Newman, and Cristopher Moore

2004

amazon.com

co-purchasing network

network communities

Finding community structure in very large networks

Aaron Clauset, M. E. J. Newman, and Cristopher Moore

2004

amazon.com

co-purchasing network
find partition that maximizes
modularity Q on those groups

$n = 409,687$ items

$m = 2,464,630$ edges

The screenshot shows the Amazon product page for 'Networks: An Introduction' by Mark Newman. The page includes the book cover, price (\$69.40), customer reviews (3 reviews), and shipping information. A red arrow points from the 'LOOK INSIDE!' button on the book cover to the 'Customer Who Bought This Item Also Bought' section below.

Customers Who Bought This Item Also Bought

Page 1 of 20

The screenshot shows a list of five books recommended for 'Networks: An Introduction'. Each book is shown with its title, author, price, and a small image. The first book, 'Networks, Crowds, and Markets: Reasoning About a... by David Easley', has a red circle around it, indicating it is the specific item being analyzed.

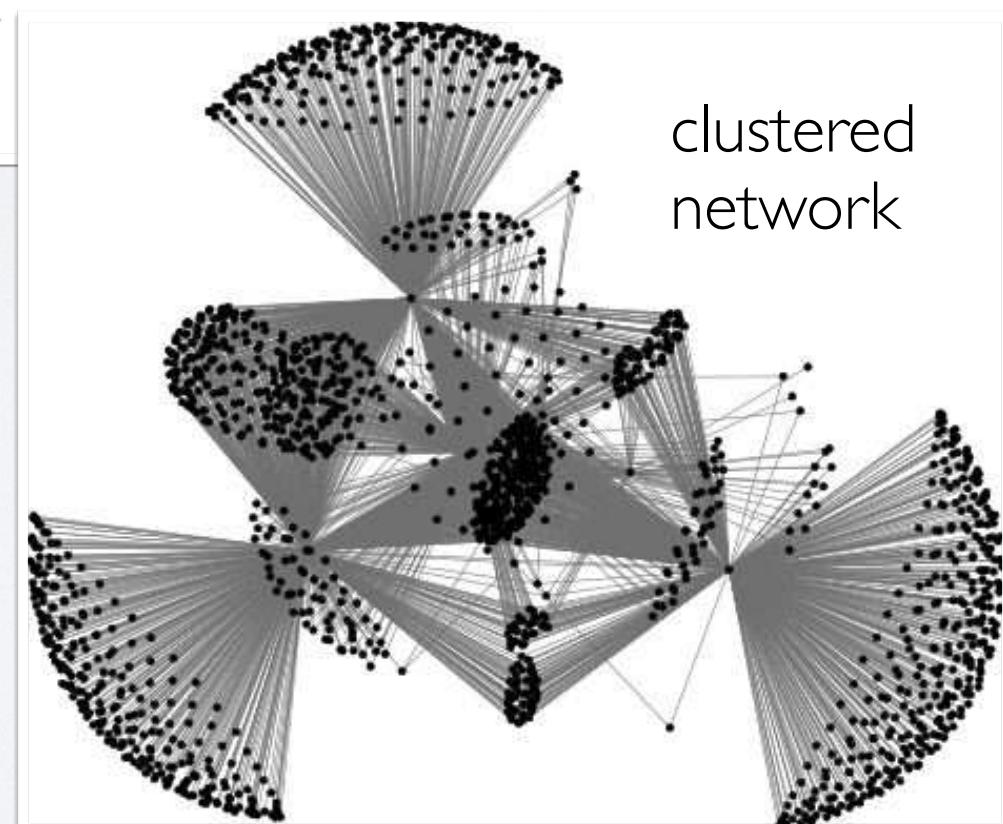
Book Title	Author	Price
Networks, Crowds, and Markets: Reasoning About a... by David Easley	David Easley	\$41.47
Dynamical Processes on Complex Networks	Alain Barrat	\$71.51
Social Network Analysis: Methods and Applications by Stanley Wasserman	Stanley Wasserman	\$44.98
Simply Complexity: A Clear Guide to Complexity Theory by Neil Johnson	Neil Johnson	\$9.81
Social and Economic Networks by Matthew O. Jackson	Matthew O. Jackson	\$33.64

network communities

Rank	Size	Description
1	114538	General interest: politics; art/literature; general fiction; human nature; technical books; how things, people, computers, societies work, etc.
2	92276	The arts: videos, books, DVDs about the creative and performing arts
3	78661	Hobbies and interests I: self-help; self-education; popular science fiction, popular fantasy; leisure; etc.
4	54582	Hobbies and interests II: adventure books; video games/comics; some sports; some humor; some classic fiction; some western religious material; etc.
5	9872	classical music and related items
6	1904	children's videos, movies, music and books
7	1493	church/religious music; African-descent cultural books; homoerotic imagery
8	1101	pop horror; mystery/adventure fiction
9	1083	jazz; orchestral music; easy listening
10	947	engineering; practical fashion

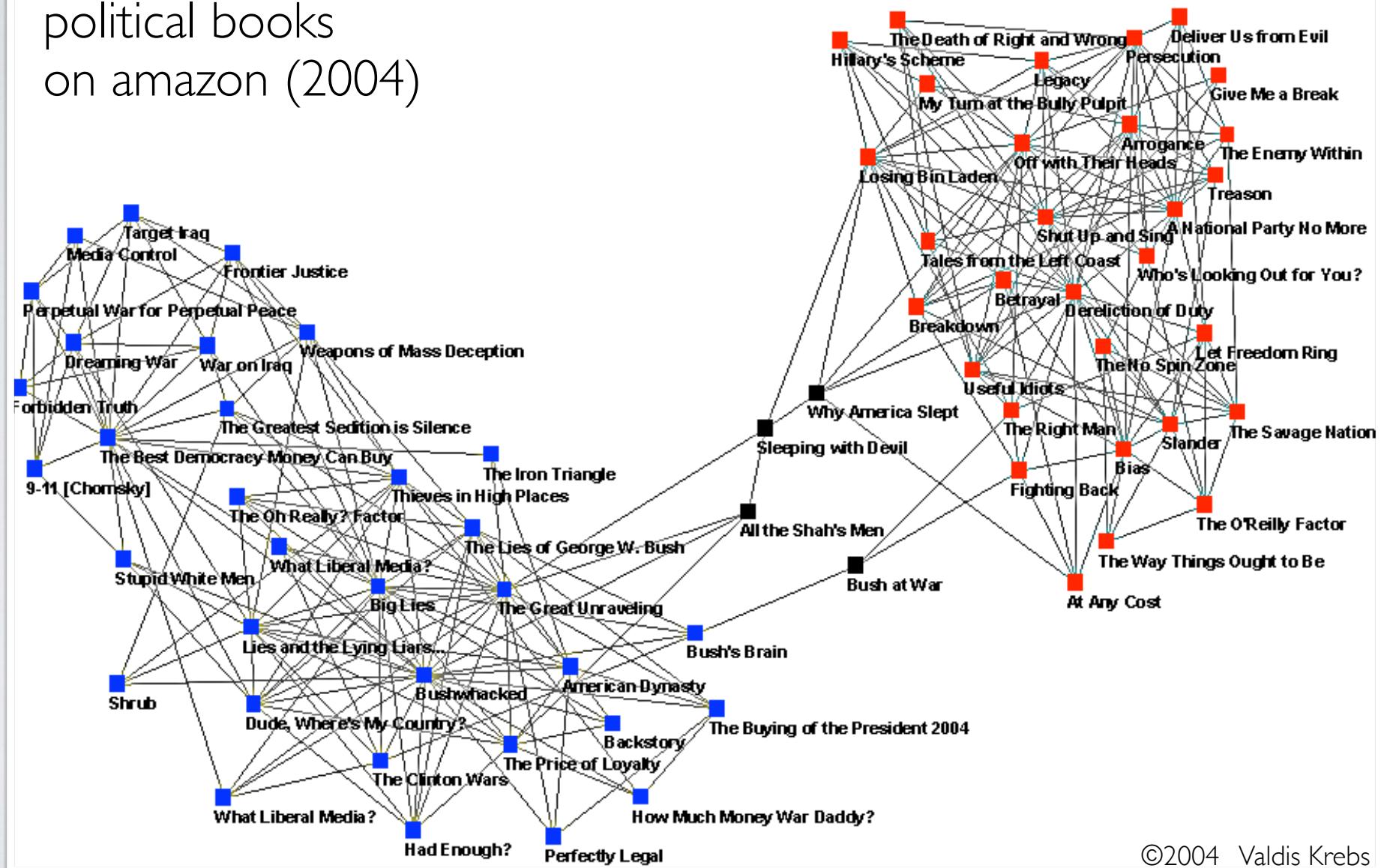
purchases = interests

interests = clustered



network communities

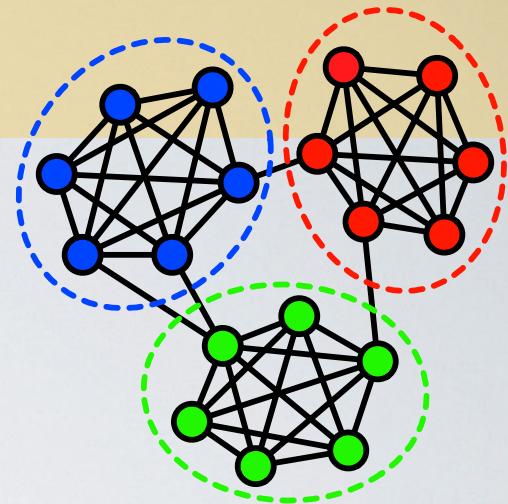
political books
on amazon (2004)



©2004 Valdis Krebs

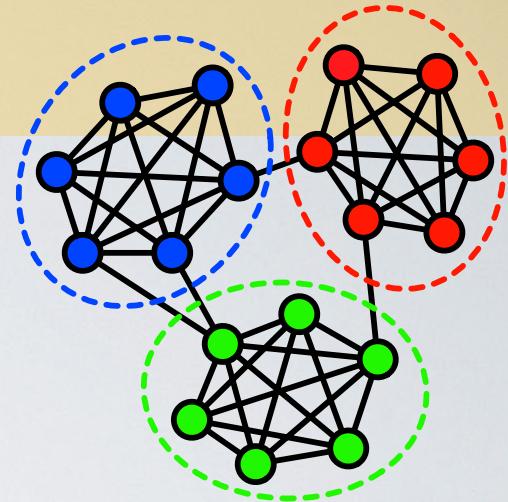
network communities

- community = vertices with same pattern of inter-community connections
- network macro-structure
- finding them like “network clustering”
[there is *no best algorithm*, and there is No Free Lunch]
- allow us to coarse grain system structure
[decompose heterogeneous structure into homogeneous blocks]
- constrains network synchronization,
information flows, diffusion, influence



network communities

- community = vertices with same pattern of inter-community connections
- network macro-structure
- finding them like “network clustering”
[there is *no best algorithm*, and there is No Free Lunch]
- allow us to coarse grain system structure
[decompose heterogeneous structure into homogeneous blocks]
- constrains network synchronization,
information flows, diffusion, influence



open questions:

- what processes generate communities?
- what impact on dynamics? network function?

describing networks

aka, summarizing a network's structure

$$f : G \rightarrow \underbrace{\{x_1, \dots, x_k\}}_{\text{summary statistics}}$$

describing networks

aka, summarizing a network's structure

at the level of

nodes	meso	whole network
degree	group degree	size (num. nodes)
centrality (various)	group size	mean degree
reciprocity (local)	modularity	mean geodesic dist.
clustering coeff. (local)	mixing matrix	diameter
eccentricity	hierarchy	assortativity (degree)
...	motif counts	modularity
	...	reciprocity (global)
		clustering coeff. (global)
		...

describing networks

aka, summarizing a network's structure

- just counting things : $f : G \rightarrow \{x_1, \dots, x_k\}$
- an infinite number of things you could count — which ones are meaningful to count?
- **warning** : *nearly all summary statistics correlate with degree*
- **things to ponder** : what is a node? what is an edge?
 - how do nodes **interact**?
 - what **causes** connections to change over time?
 - where is the **structure** : nodes? communities? network?
 - what is the role of node **degree** on dynamics?
 - what is role of node **position** on dynamics?



end of lecture 2

lecture 3 : null models & inference for networks

network position

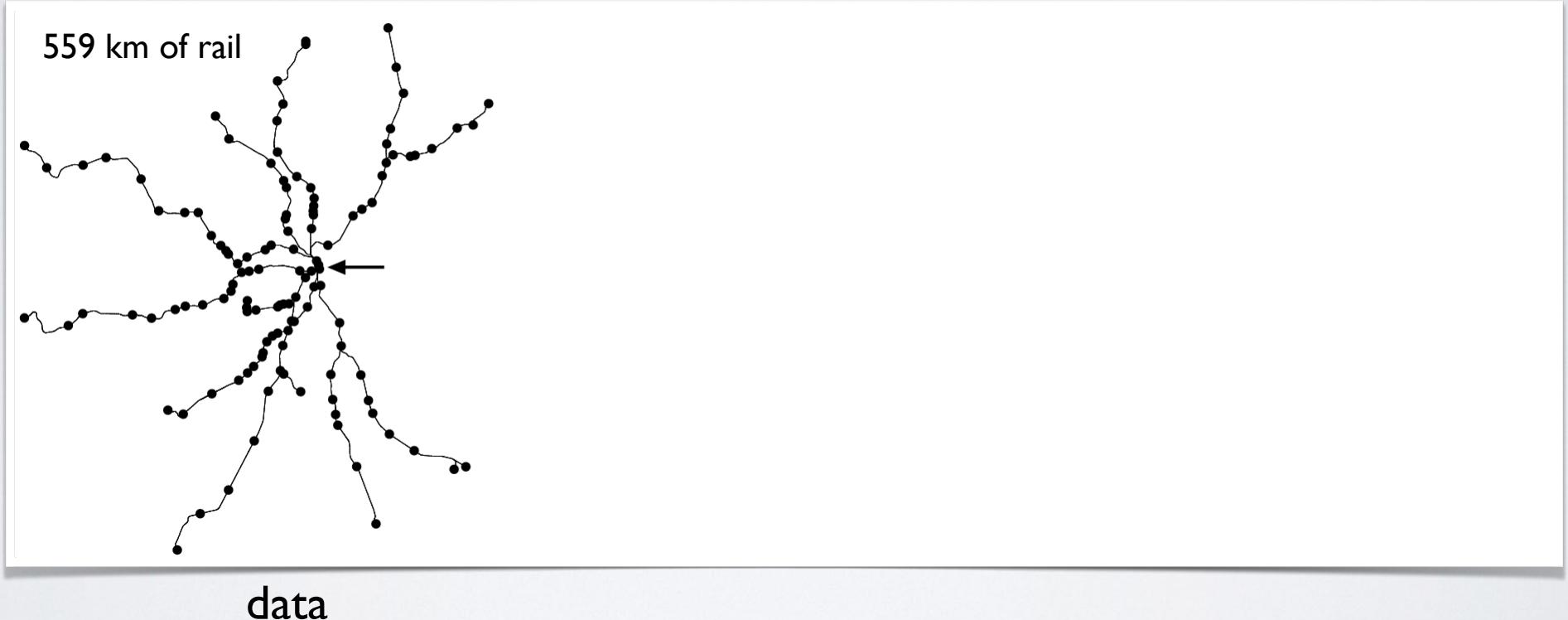
an example

how does a network become centralized?

network position

an example

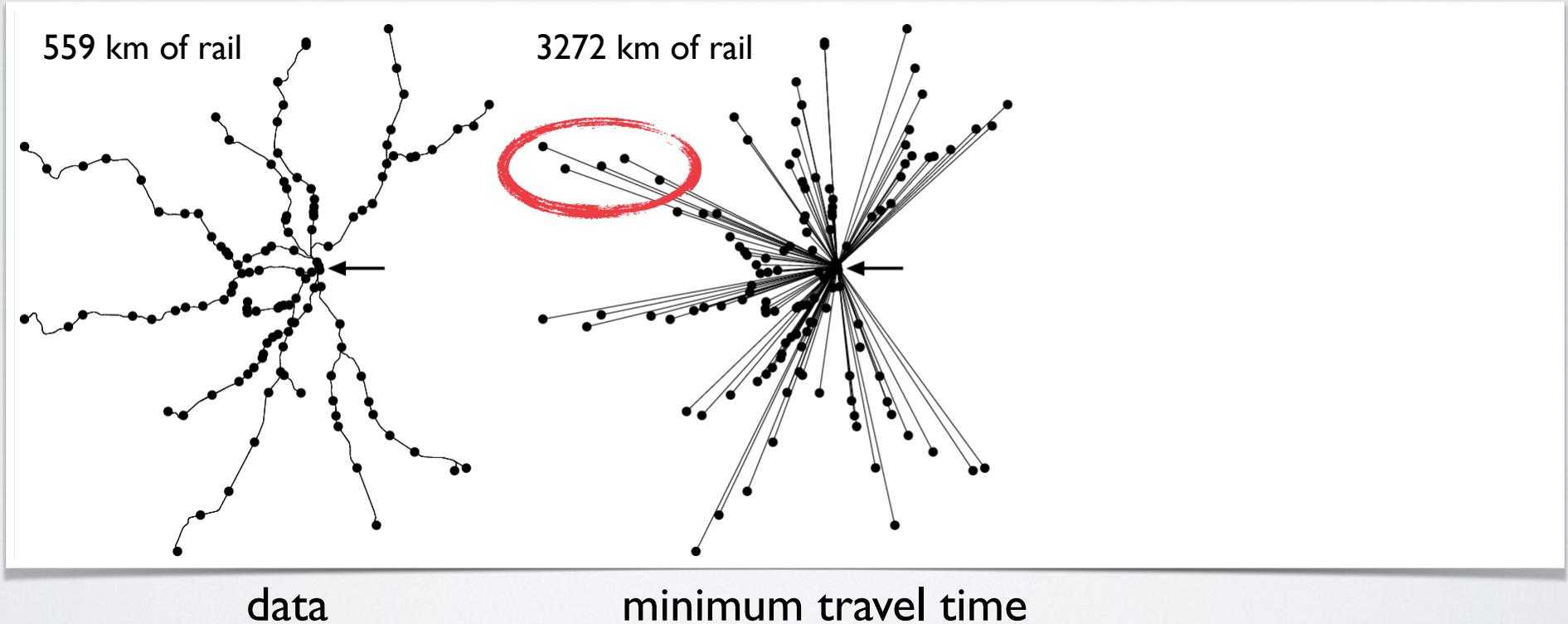
optimizing paths for Boston commuter rail



network position

an example

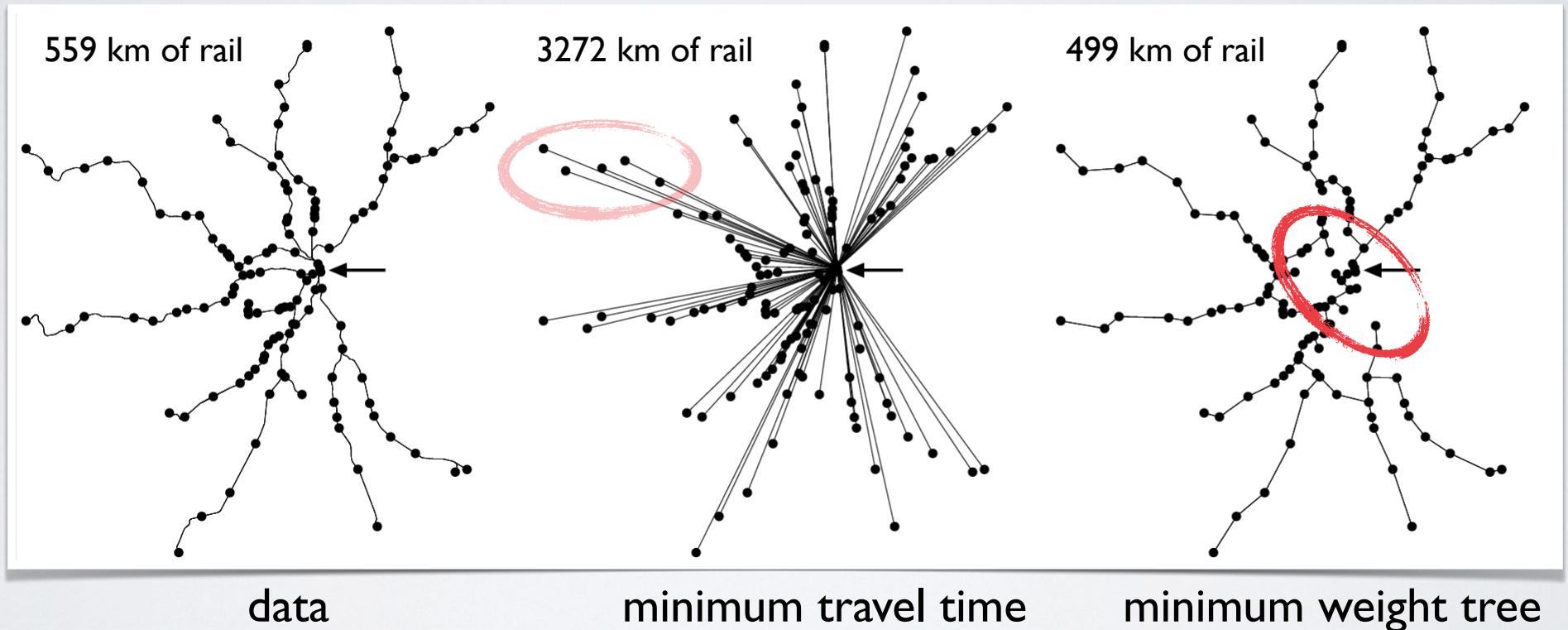
optimizing paths for Boston commuter rail



network position

an example

optimizing paths for Boston commuter rail

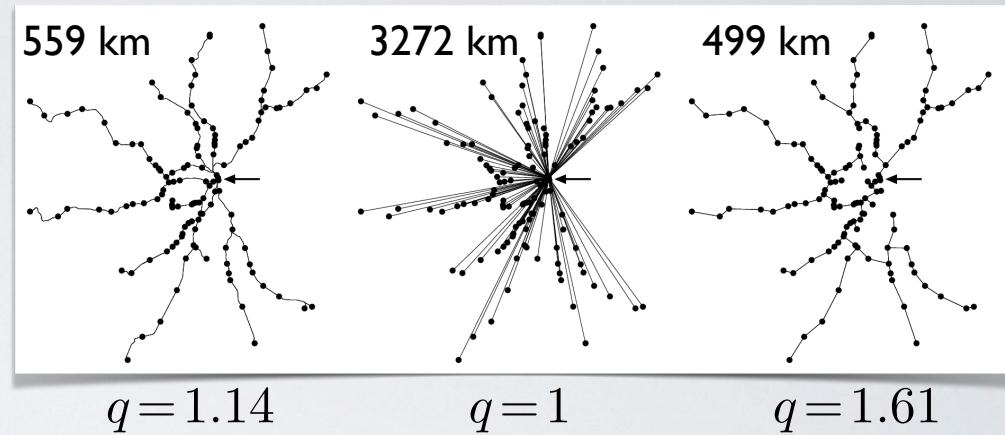


network position

route factor

$$q = \frac{1}{n} \sum_{i=1}^n \frac{\ell_{i0}}{d_{i0}}$$

mean ratio of distance along edges ℓ_{i0} to direct Euclidean distance d_{i0} to root 0



network position

a simple model

embed n vertices in a plane

until all vertices connected

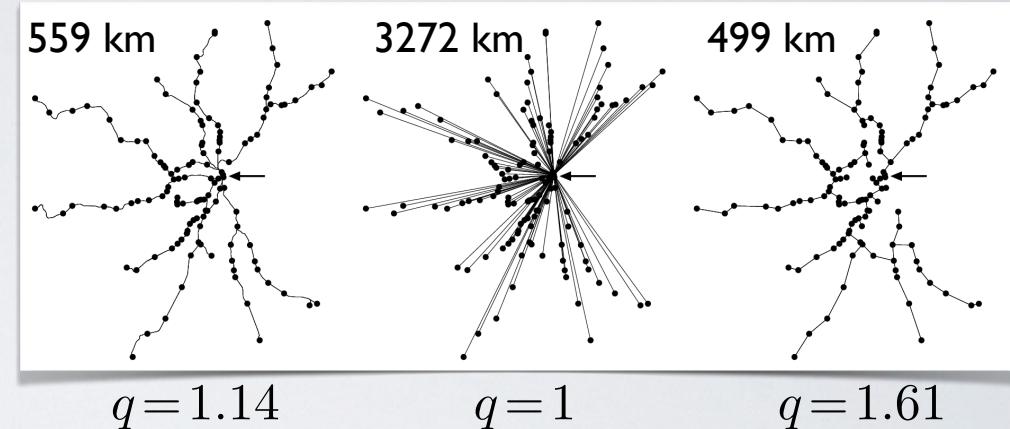
add edge (i, j) with
minimum value for

$$w_{ij} = d_{ij} + \beta \ell_{j0}$$

distance from i to j

parameter

route length to root



$\beta = 0 \rightarrow$ minimum spanning tree*

$\beta > 0 \rightarrow$ prefer shorter paths to root

*this is exactly Prim's algorithm for MSTs

network position

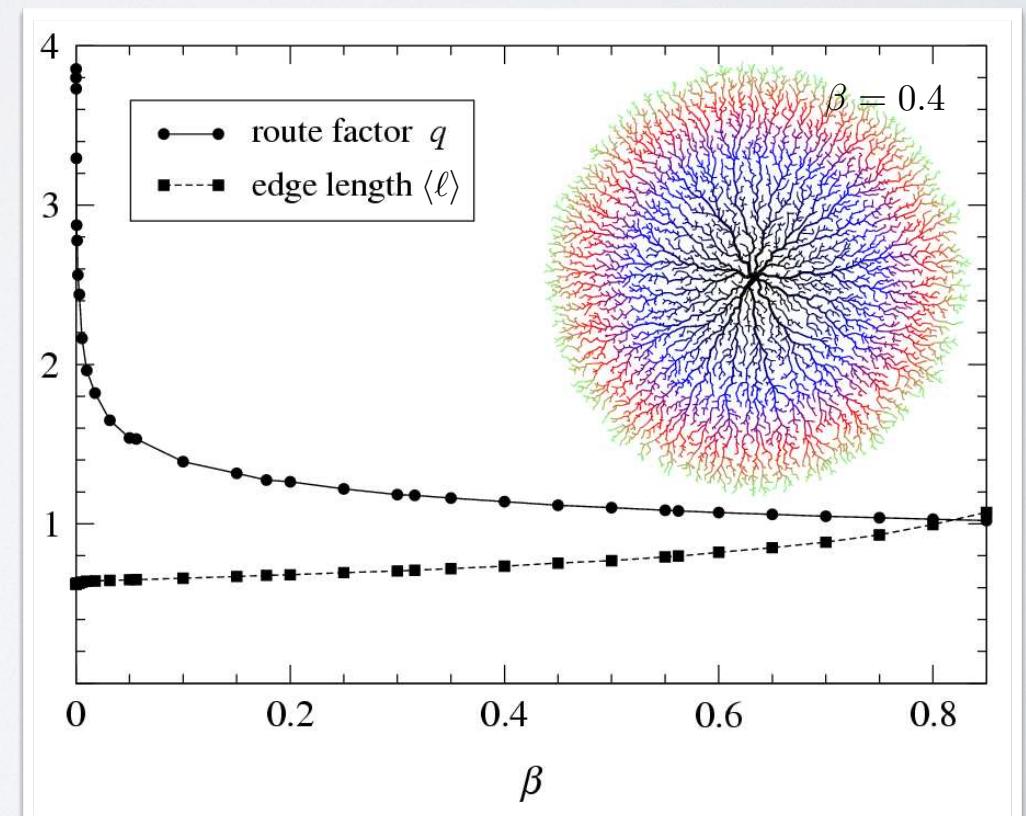
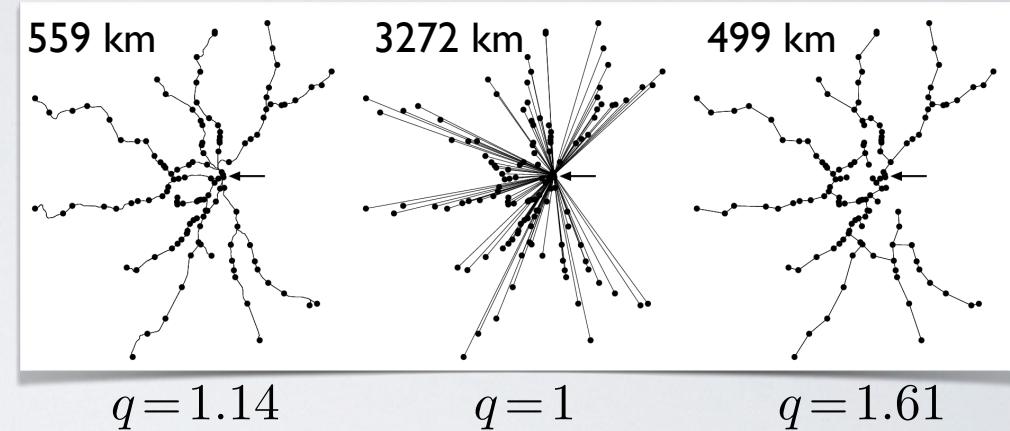
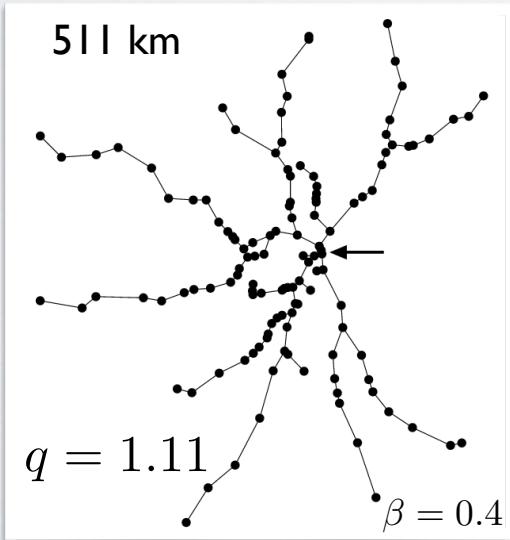
a simple model

embed n vertices in a plane

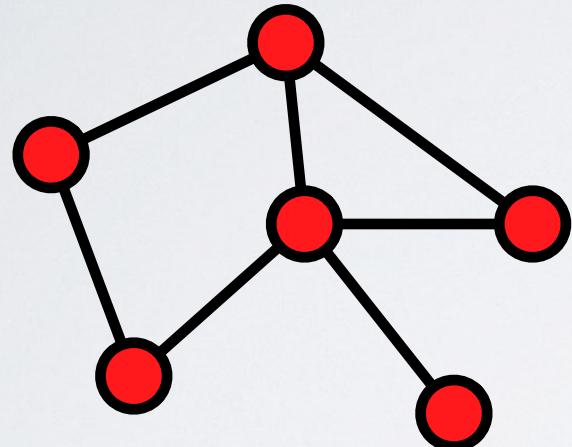
until all vertices connected

add edge (i, j) with
minimum value for

$$w_{ij} = d_{ij} + \beta \ell_{j0}$$

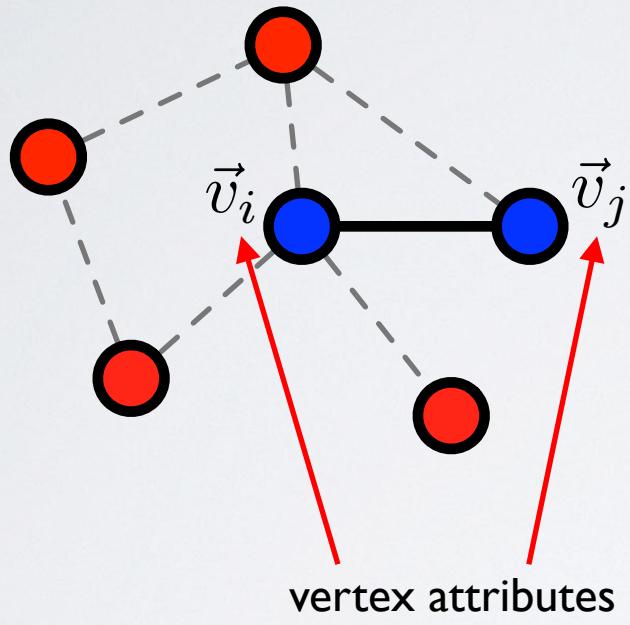


describing networks



**homophily and
assortative mixing**
like links with like

assortative mixing



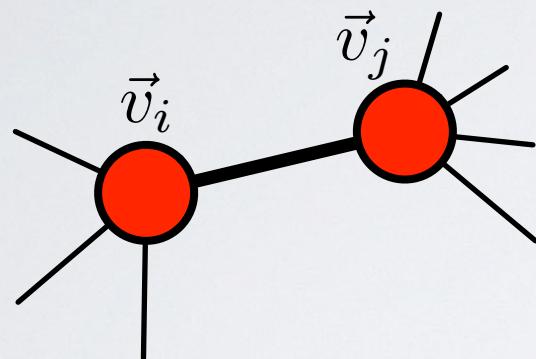
homophily and assortative mixing

like links with like

assortativity coefficient r
quantifies homophily

three types:
scalar attributes
vertex degrees
categorical variables

assortative mixing



homophily and assortative mixing

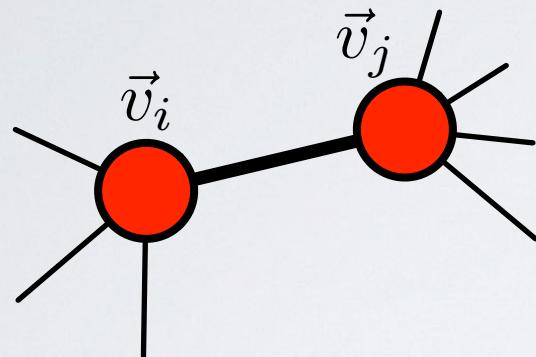
like links with like

scalar attributes:
mean value across ties

$$\mu = \frac{1}{2m} \sum_i \sum_j A_{ij} v_i$$

$$= \frac{1}{2m} \sum_i k_i v_i$$

assortative mixing



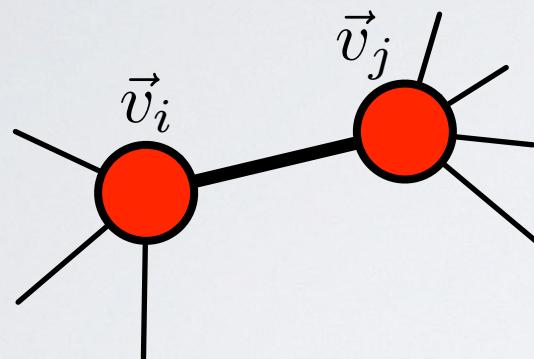
homophily and assortative mixing

like links with like

scalar attributes:
covariance across ties

$$\begin{aligned}\text{cov}(v_i, v_j) &= \frac{\sum_{ij} A_{ij}(v_i - \mu)(v_j - \mu)}{\sum_{ij} A_{ij}} \\ &= \frac{1}{2m} \sum_{ij} A_{ij} v_i v_j - \mu^2 \\ &= \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) v_i v_j\end{aligned}$$

assortative mixing



homophily and assortative mixing

like links with like

assortativity coefficient (scalar)

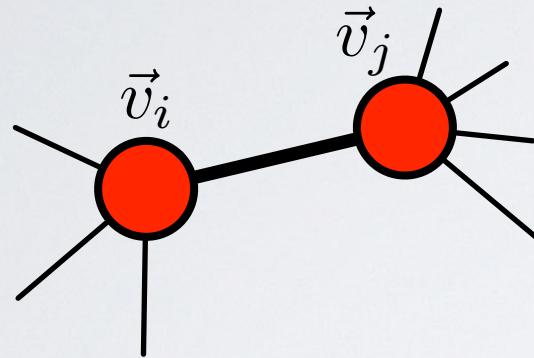
$$r = \frac{\text{cov}(v_i, v_j)}{\text{var}(v_i, v_j)}$$

$$= \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) v_i v_j}{\sum_{ij} k_i \delta_{ij} - k_i k_j / 2m}$$

[this is just a Pearson correlation across edges]

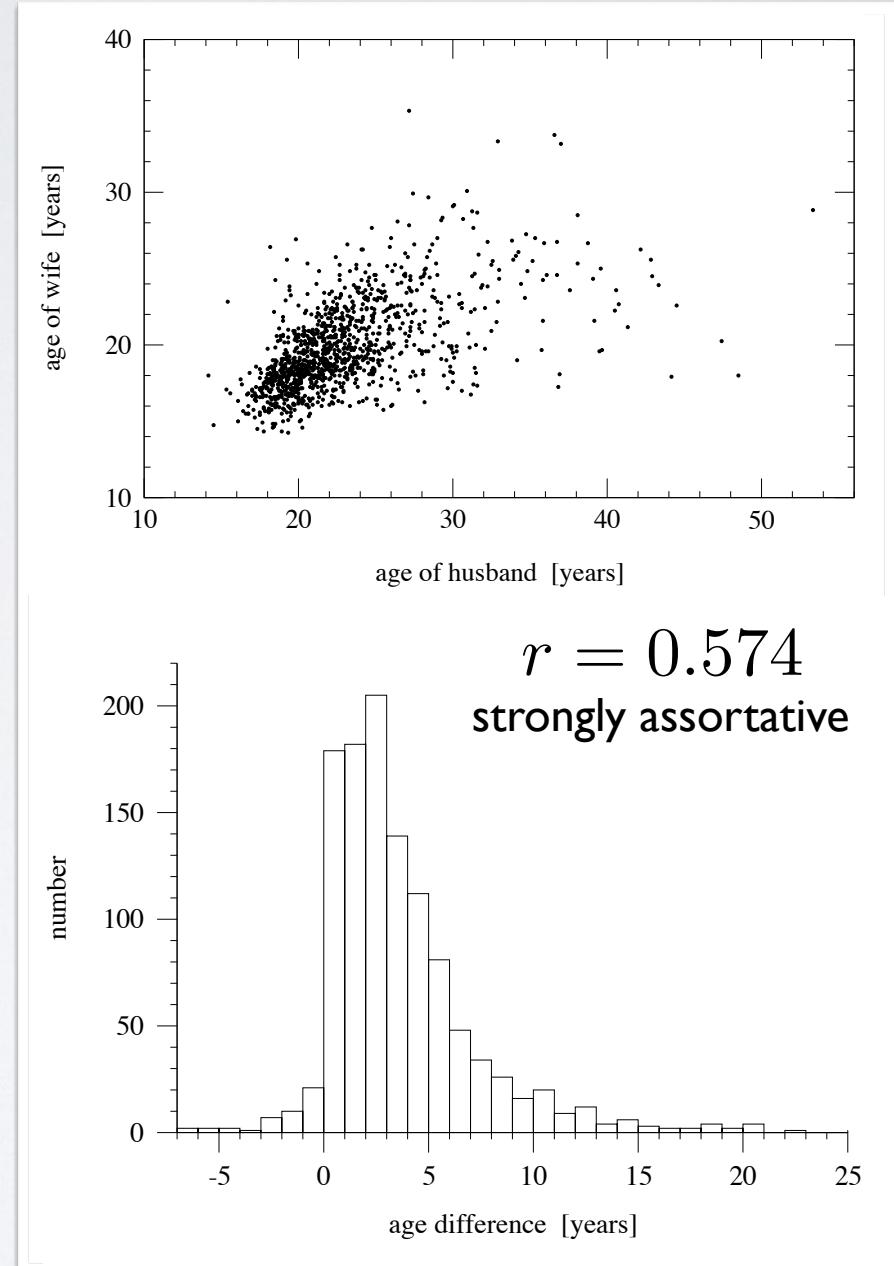
$$-1 \leq r \leq 1$$

assortative mixing

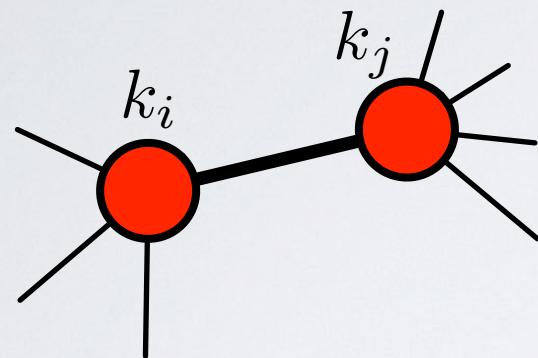


(top) scatter plot of ages of 1141 married couples at time of marriage [1995 US National Survey of Family Growth]

(bottom) histogram of age differences (M-F) for same data



assortative mixing

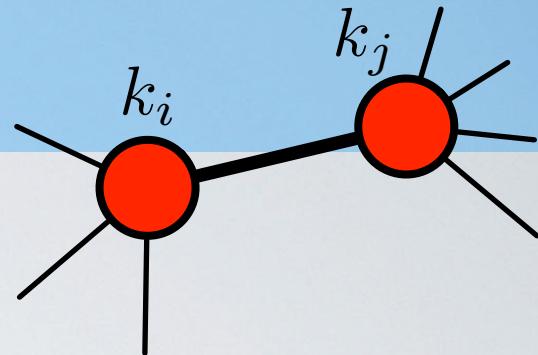


homophily and assortative mixing

like links with like

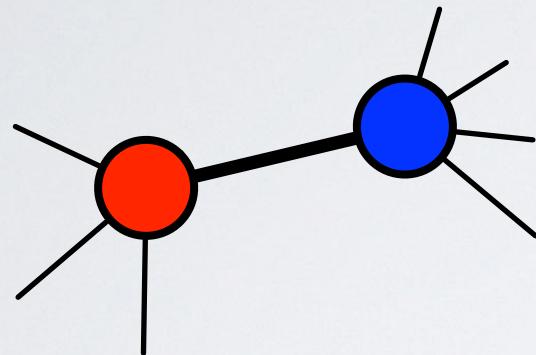
degree:
just another scalar*

assortative mixing



	network	type	size n	degree assortativity r	error σ_r
social	physics coauthorship	undirected	52 909	0.363	0.002
	biology coauthorship	undirected	1 520 251	0.127	0.0004
	mathematics coauthorship	undirected	253 339	0.120	0.002
	film actor collaborations	undirected	449 913	0.208	0.0002
	company directors	undirected	7 673	0.276	0.004
	student relationships	undirected	573	-0.029	0.037
	email address books	directed	16 881	0.092	0.004
technological	power grid	undirected	4 941	-0.003	0.013
	Internet	undirected	10 697	-0.189	0.002
	World-Wide Web	directed	269 504	-0.067	0.0002
	software dependencies	directed	3 162	-0.016	0.020
biological	protein interactions	undirected	2 115	-0.156	0.010
	metabolic network	undirected	765	-0.240	0.007
	neural network	directed	307	-0.226	0.016
	marine food web	directed	134	-0.263	0.037
	freshwater food web	directed	92	-0.326	0.031

assortative mixing



homophily and assortative mixing

like links with like

categorical variables:

let e_{ij} be fraction of edges connecting vertices of type i to vertices of type j

matrix sum

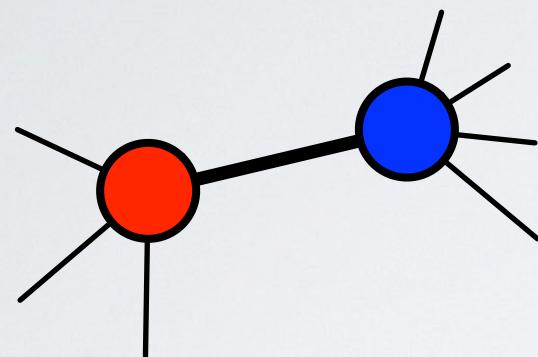
$$\sum_{ij} e_{ij} = 1$$

marginals

$$\sum_j e_{ij} = a_i$$

$$\sum_i e_{ij} = b_j$$

assortative mixing



homophily and assortative mixing

like links with like

categorical variables:
assortativity coefficient*

$$\begin{aligned} r &= \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \\ &= \frac{\text{Tr } \mathbf{e} - \|\mathbf{e}^2\|}{1 - \|\mathbf{e}^2\|} \end{aligned}$$

* this equation is equivalent to the popular modularity measure Q used to score the strength of community structure

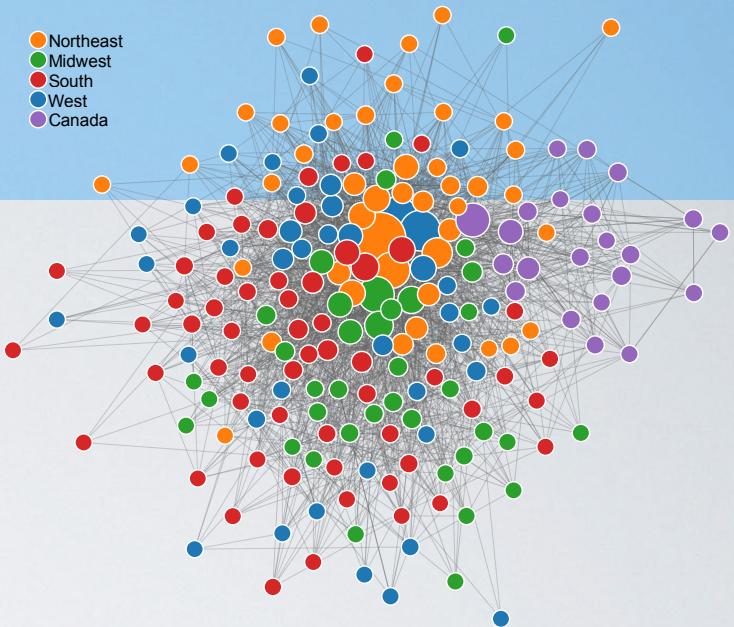
assortative mixing

4388 Computer Science faculty

vertices are PhD granting institutions in North America

edge (u, v) means PhD at u and now faculty at v

labels are US census regions + Canada

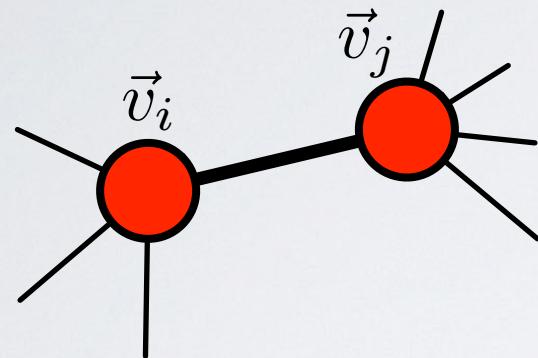


	Northeast	Midwest	South	West	Canada	a_u
Northeast	0.237	0.084	0.098	0.104	0.028	0.552
Midwest	0.084	0.134	0.088	0.059	0.016	0.381
South	0.098	0.088	0.166	0.068	0.012	0.432
West	0.104	0.059	0.068	0.145	0.017	0.393
Canada	0.028	0.016	0.012	0.017	0.170	0.242
a_u	0.552	0.381	0.432	0.393	0.242	

$$r = 0.215$$

moderately assortative

assortative mixing



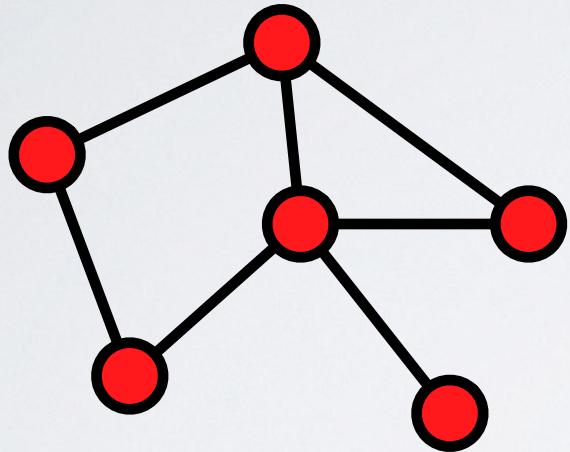
homophily and assortative mixing

like links with like

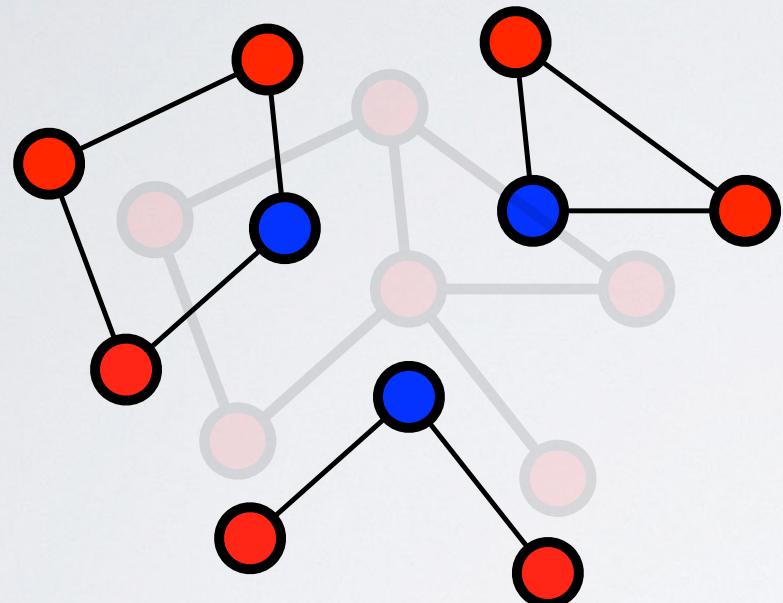
- random graphs tend to be disassortative $r \leq 0$ because the mixing is uniform
- social networks (apparently) highly assortative, in every way (attribute, degree, category)
- extremal values $r \approx \{-1, 1\}$ suggest underlying mechanism on that variable

describing networks

motifs

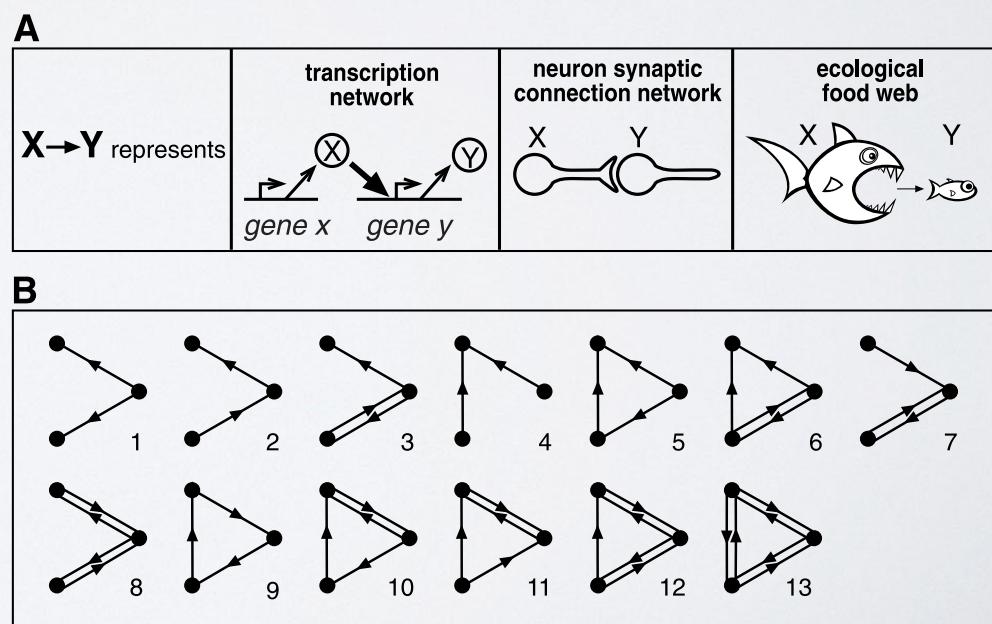


describing networks

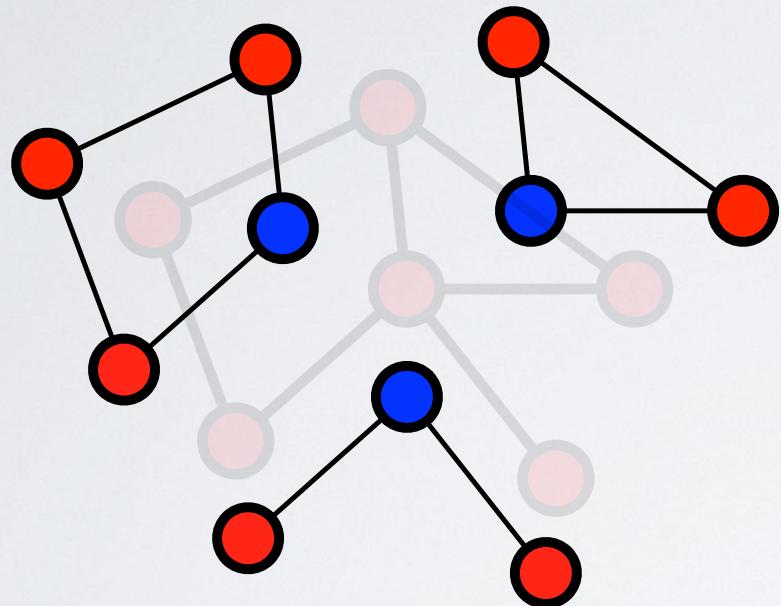


motifs:

small subgraphs (of interest),
which we then count
compare counts against null
model (random graph model)



describing networks



motifs:

small subgraphs (of interest),
which we then count

compare counts against null
model (random graph model)

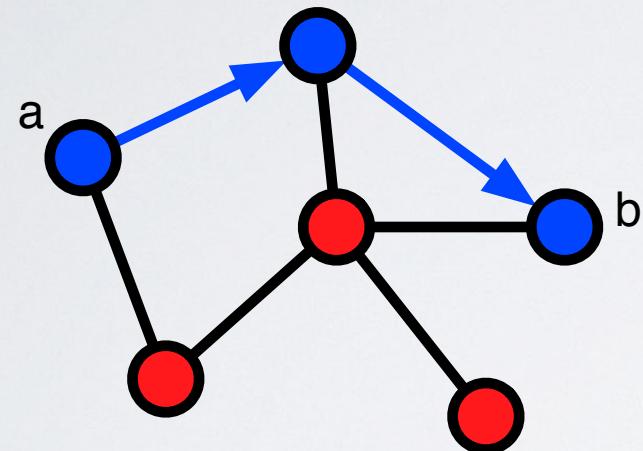
- efficient counting is tricky
(combinatorics + graph isomorphism)
- choice of null model key
- lots of work in this area, mainly in
molecular biology and neuroscience
- see

Sporns and Kotter, *PLoS Biol.* **2**, e369 (2004)

Matias et al., *REVSTAT* **4**, 31-51 (2006)

Wong et al., *Brief. in Bioinfo.* **13**, 202-215 (2011)

describing networks



path:

number of “hops”
between two nodes

$$\ell_{a \rightarrow b} = 2$$

network paths

THE ORACLE OF BACON



Tina Fey has a Bacon number of 2.

[Find a different link](#)

```
graph TD; TinaFey["Tina Fey"] -- "was in" --> Movie1["Man of the Year (2006)"]; TinaFey -- "with" --> AudreyDwyer["Audrey Dwyer"]; AudreyDwyer -- "was in" --> Movie2["Where the Truth Lies (2005)"]; AudreyDwyer -- "with" --> KevinBacon["Kevin Bacon"];
```

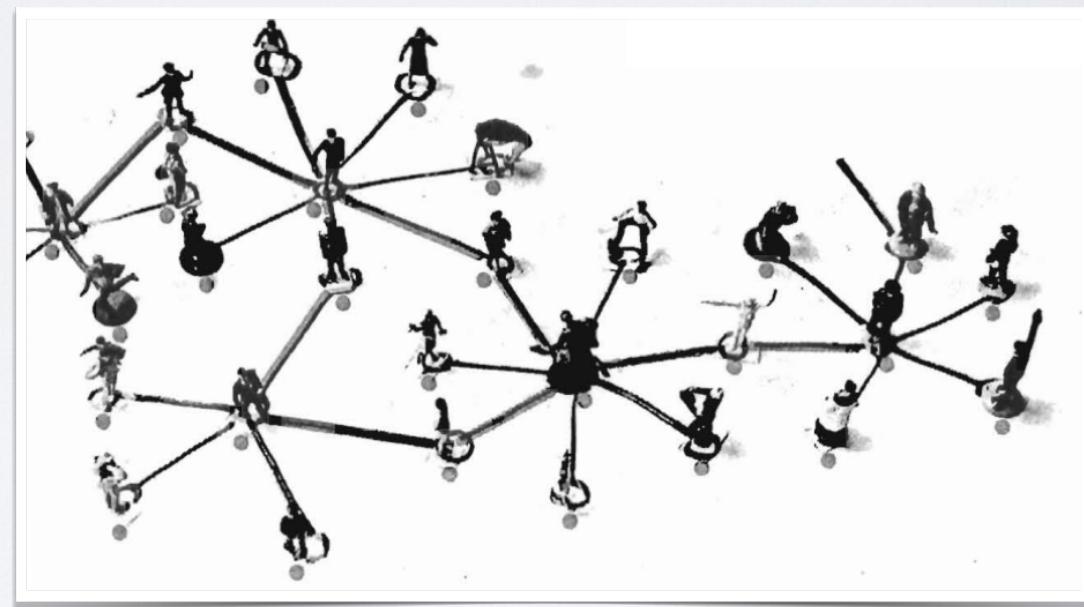
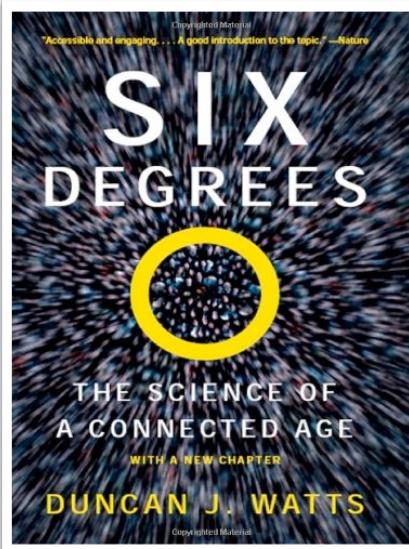
Kevin Bacon to Tina Fey [Find link](#) [More options >>](#)

network paths

The Small-World Problem

By Stanley Milgram

1967

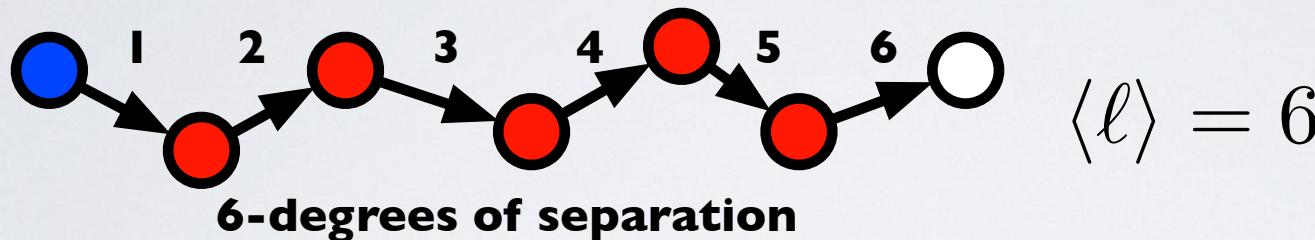


network paths

The Small-World Problem

By Stanley Milgram

1967



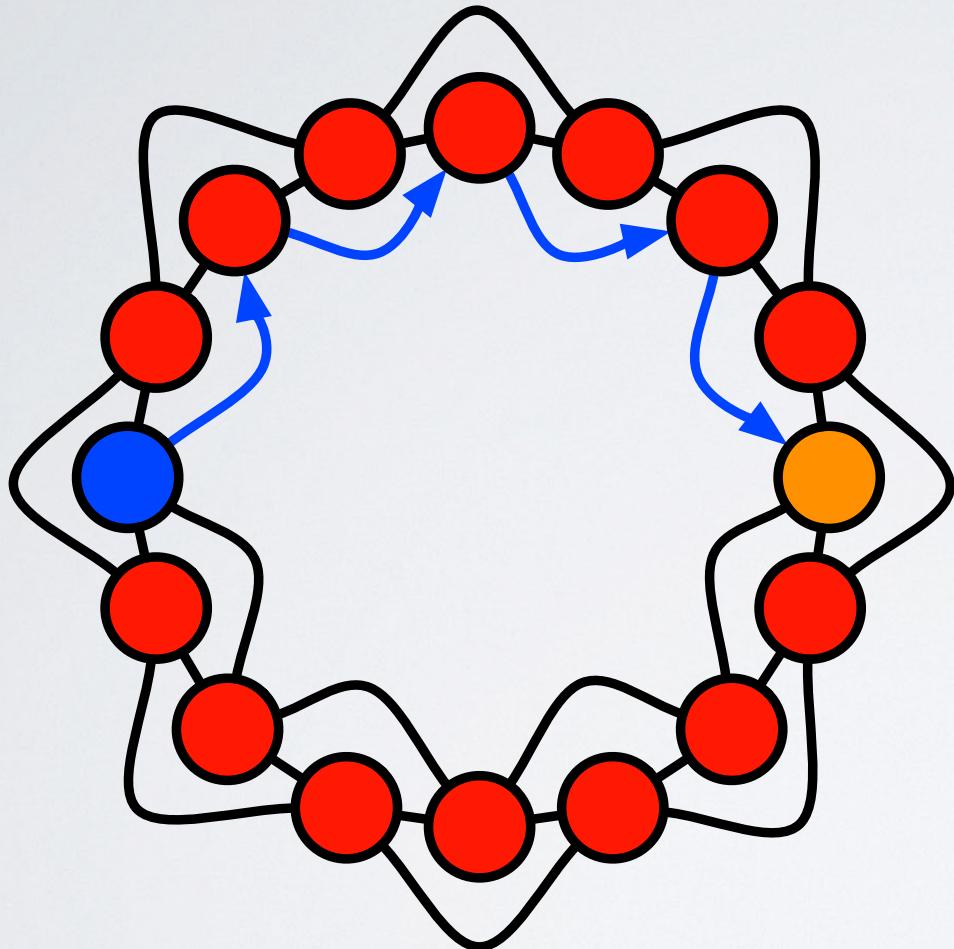
network paths

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

1998

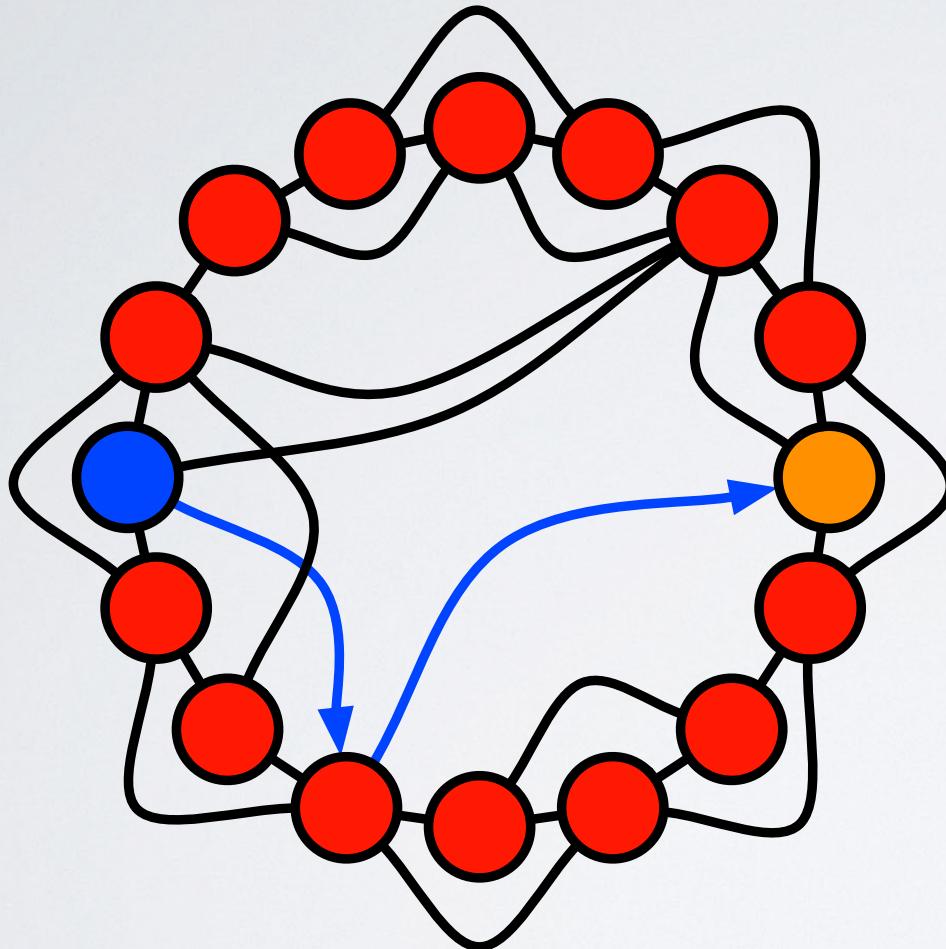
network paths



all links “local”

- most nodes far away
- high “clustering”

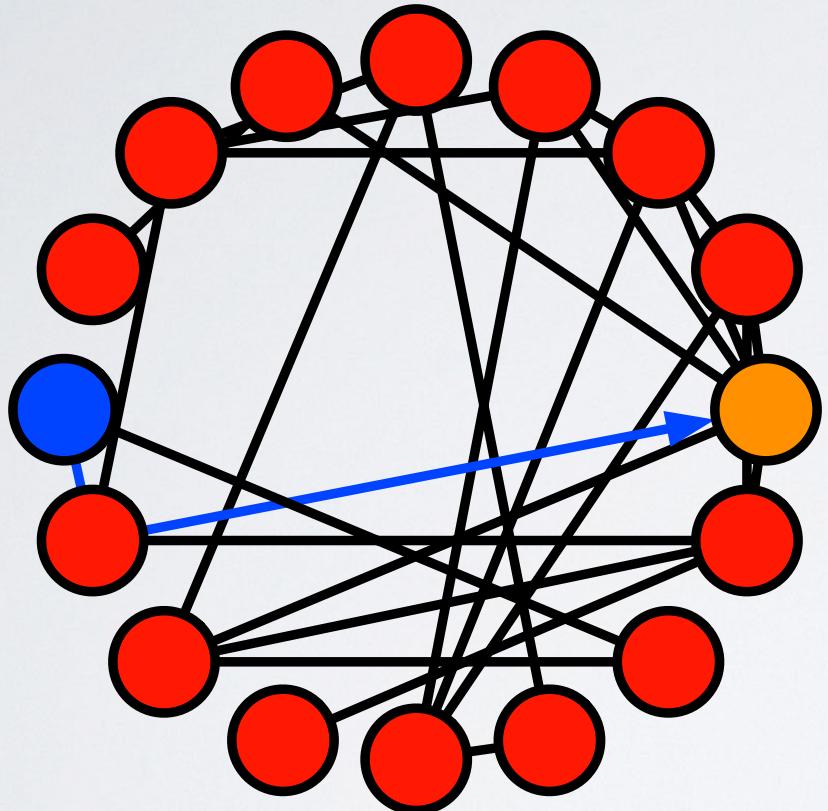
network paths



**most links “local”
some links random**

- most nodes near
- high “clustering”
- short paths can be found

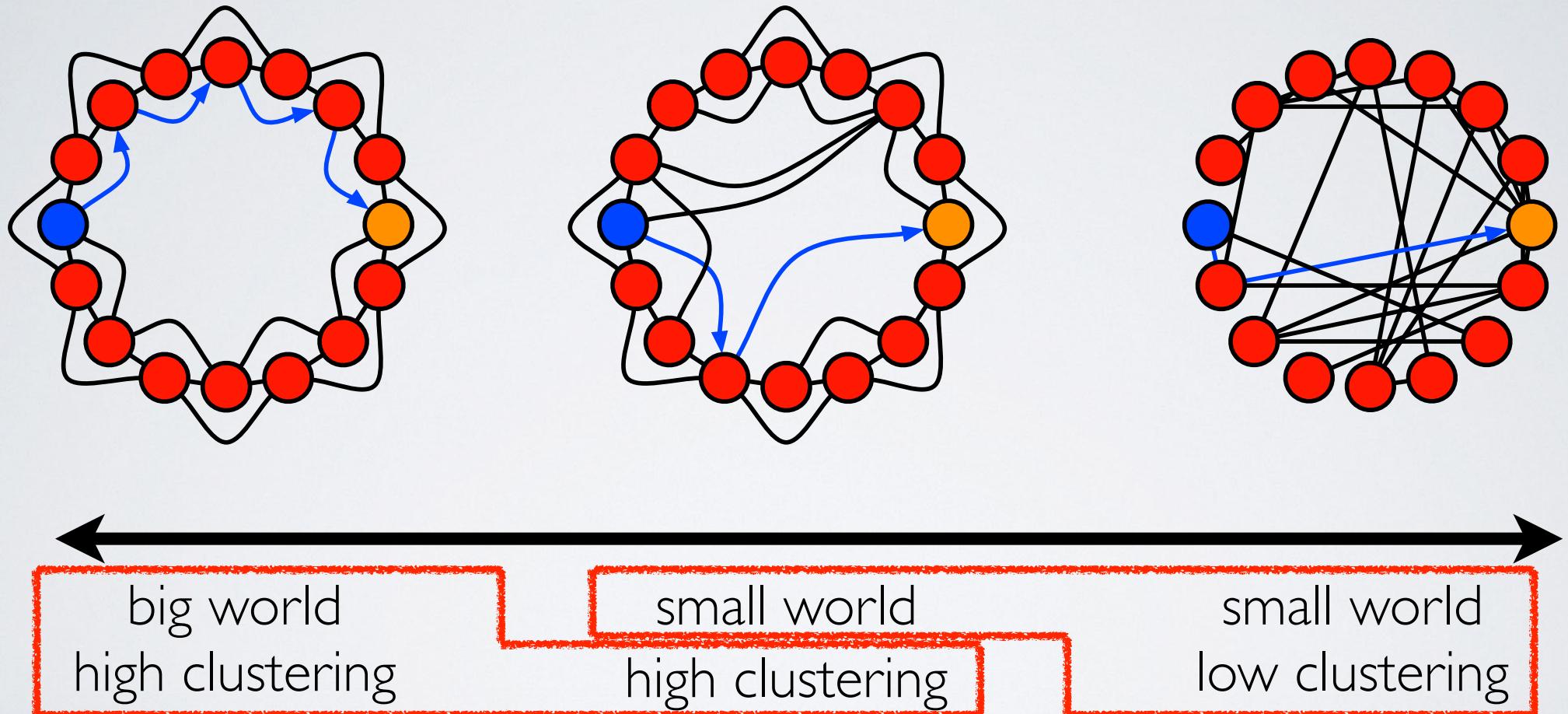
network paths



all links random

- Erdos-Renyi graph
- most nodes near
- short paths hard to find
- no “clustering”

it's a small world after all



it's a small world after all

Geographic routing in social networks

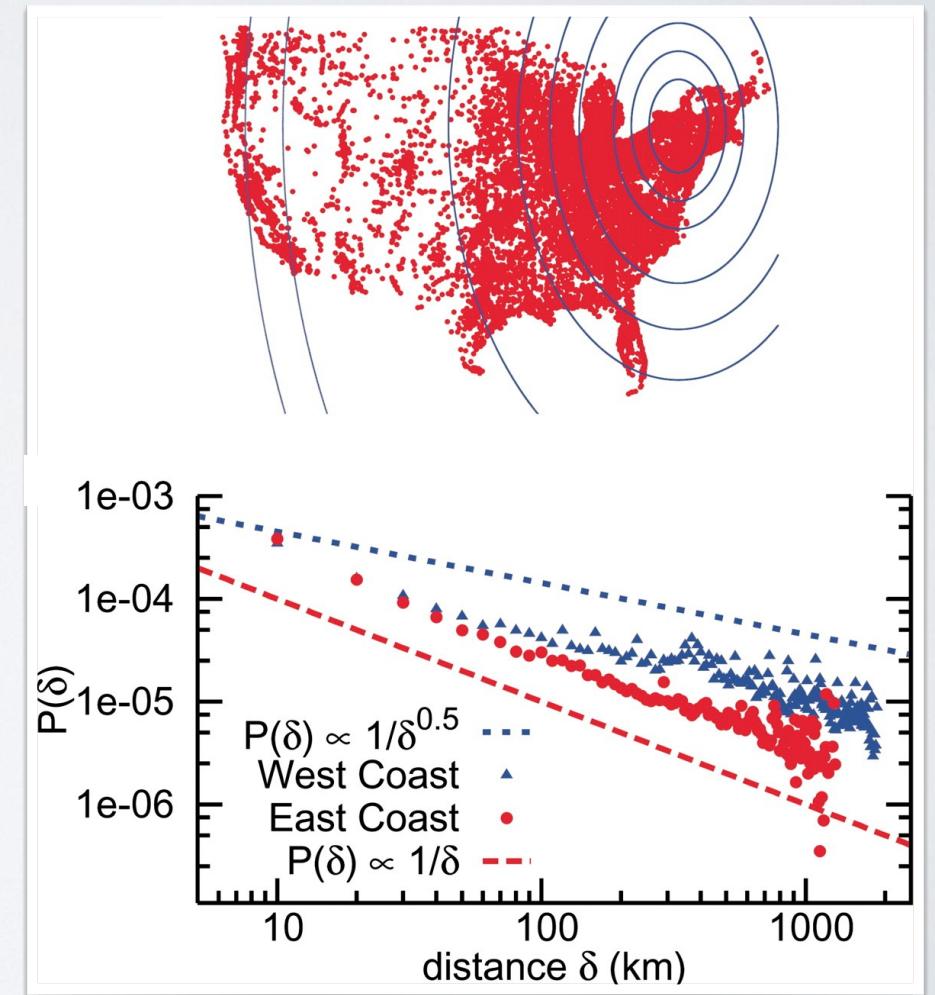
2005

David Liben-Nowell^{*†‡§}, Jasmine Novak[†], Ravi Kumar^{†¶}, Prabhakar Raghavan^{†||}, and Andrew Tomkins^{†¶}



495,836 geo-located users

- most links “local”
- remaining links span all scales
- high clustering
- small “diameter”



network paths



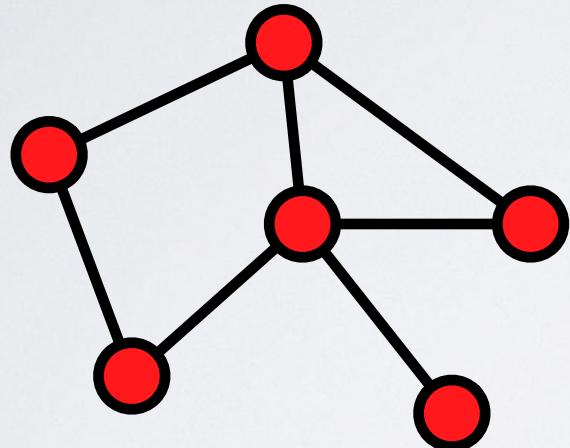
- path = sequence of edges $a \rightarrow \dots \rightarrow b$
- many short paths = “small world”
- social world is surprisingly small, yet highly “clustered”
(many locally dense groups)

open questions:

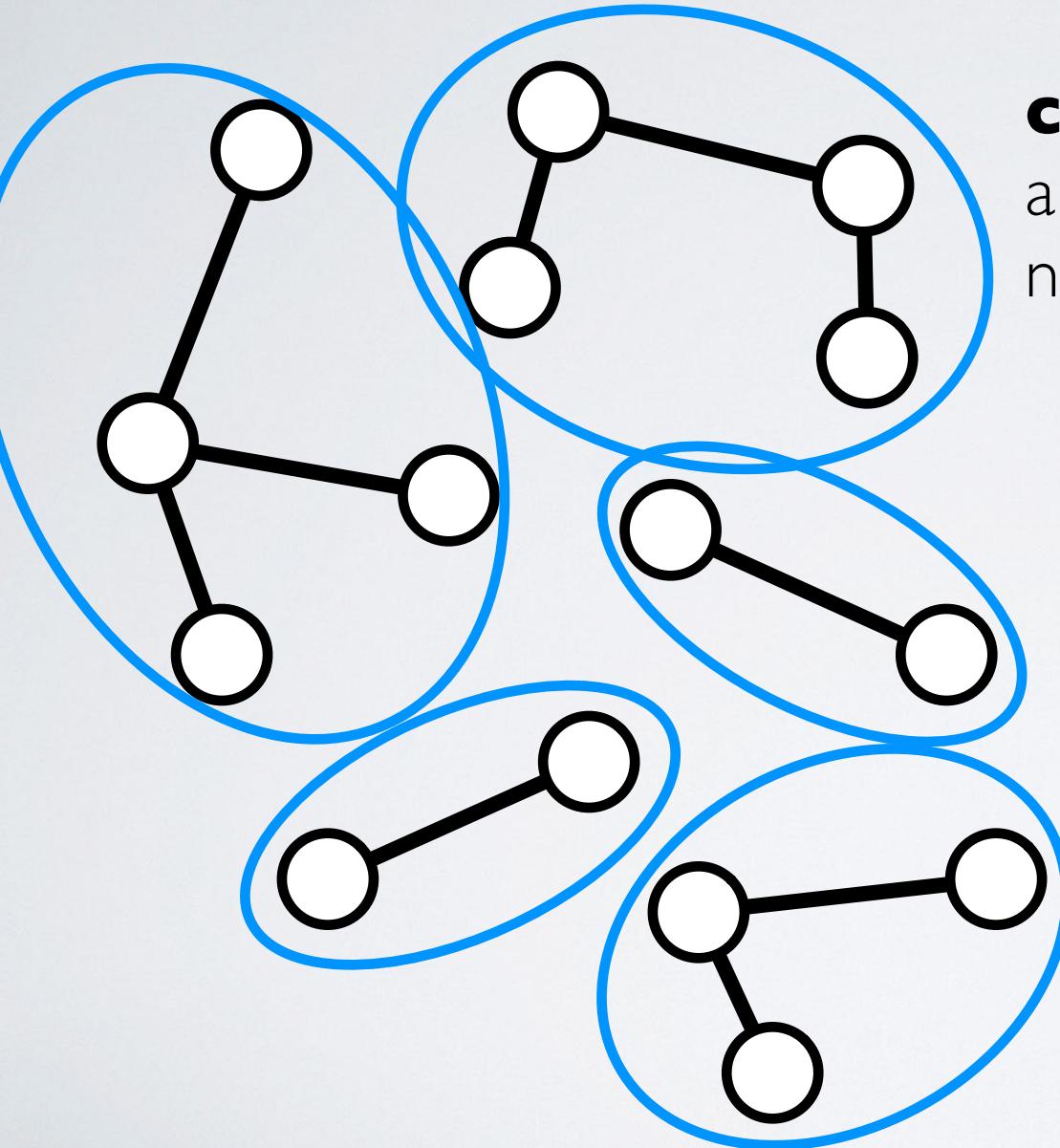
- how do big social networks self-organize?
- what processes shrink big worlds?
- social information filtering

describing networks

components



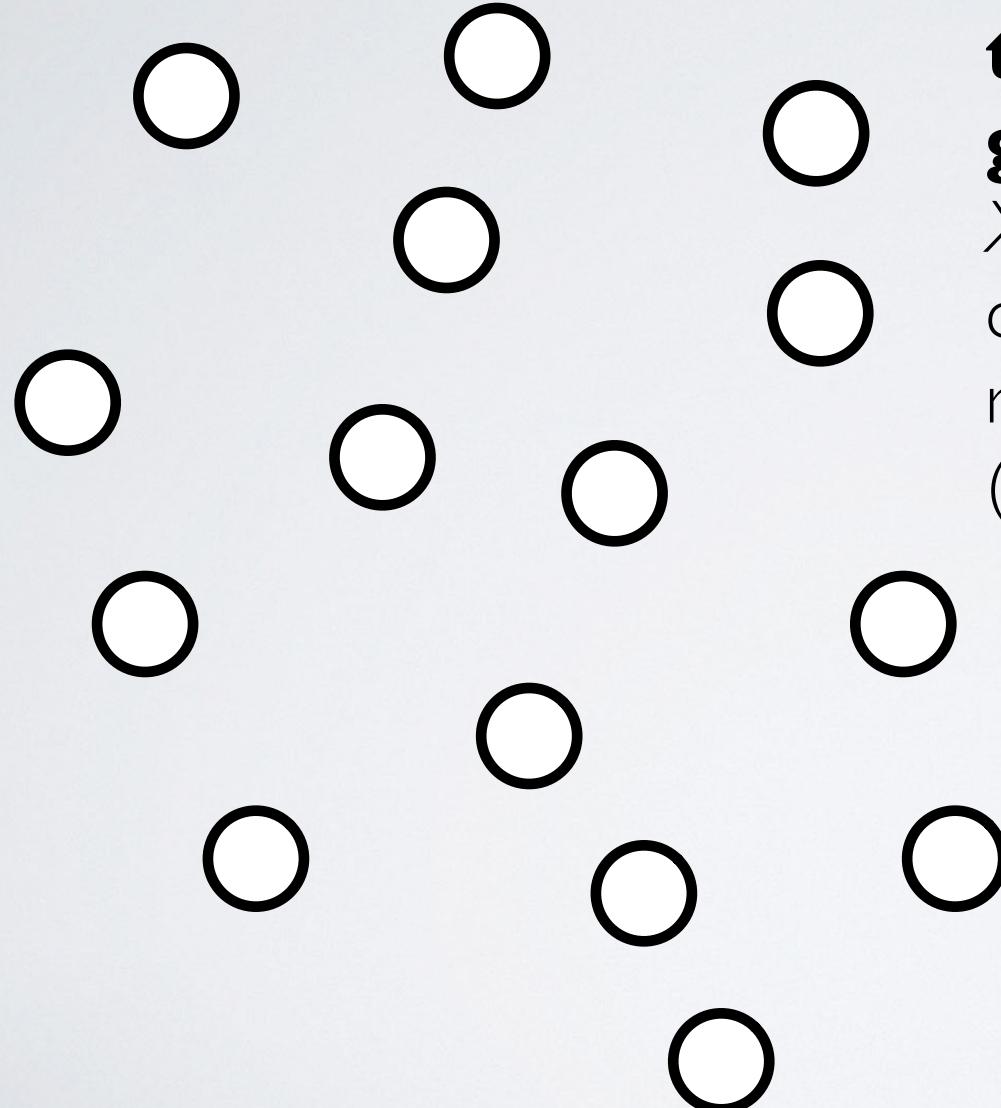
network terminology



component:

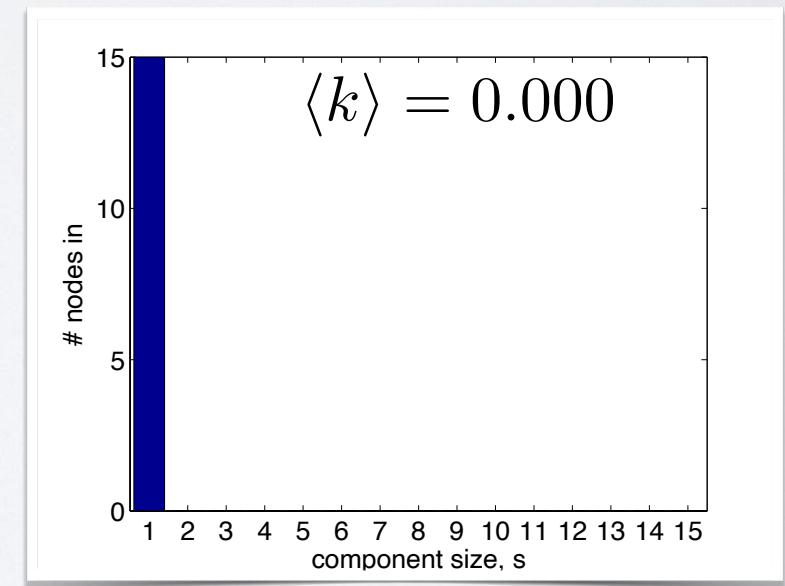
a group of connected
nodes

network components

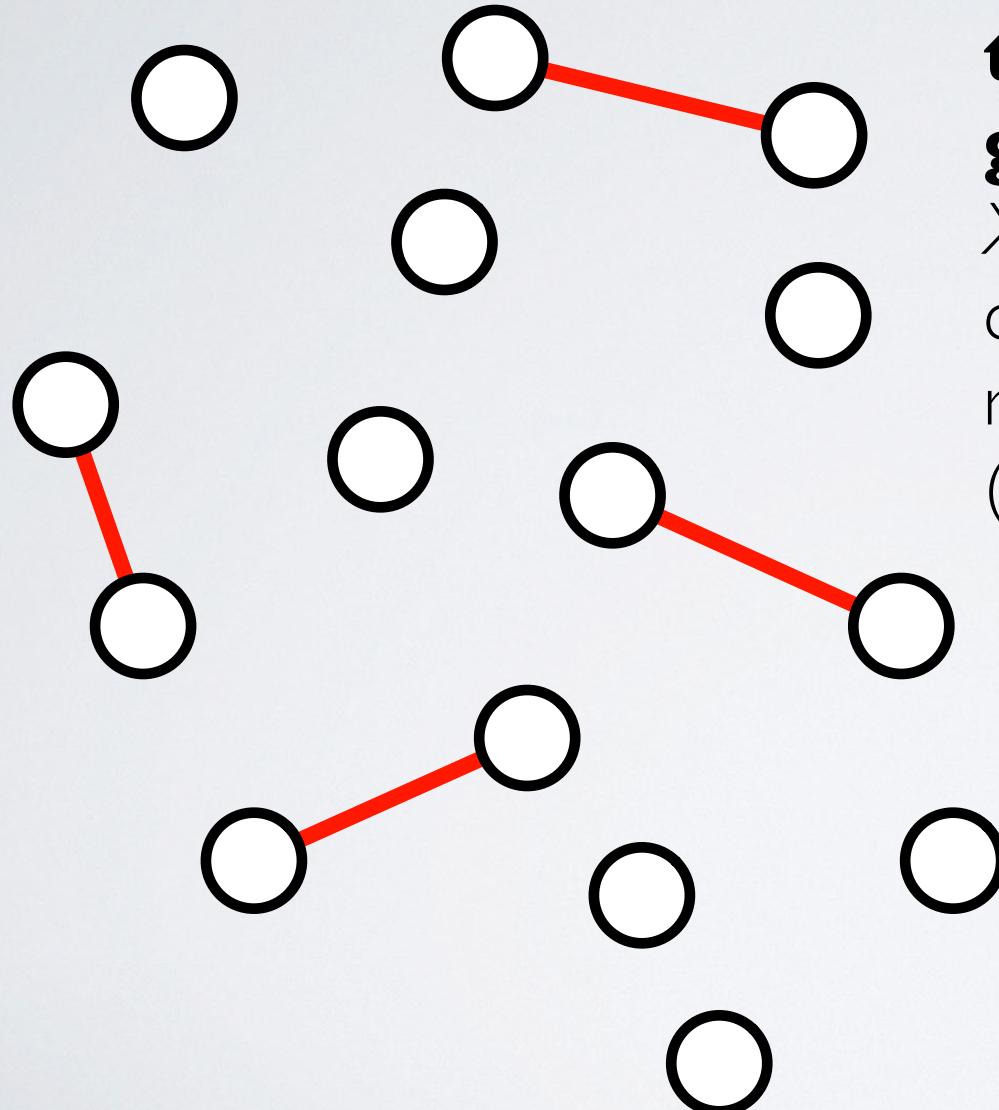


the percolation game:

choose
X random pairs
connect them
repeat
(count components)

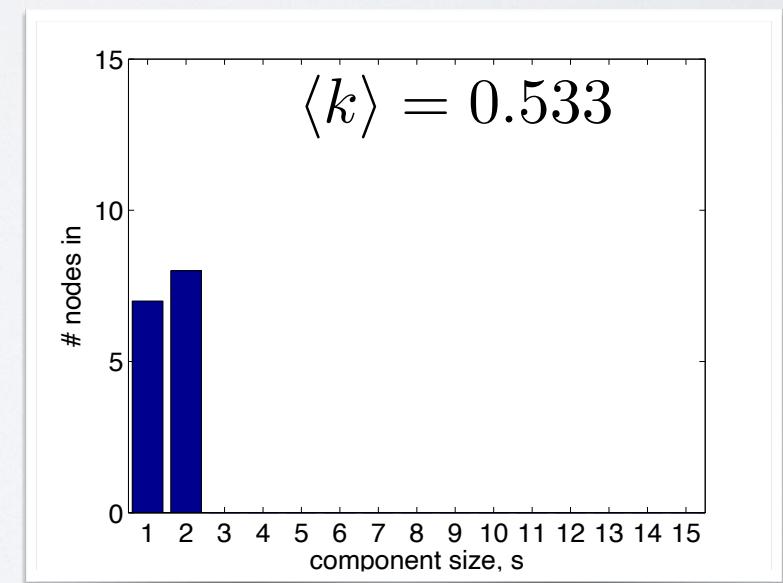


network components

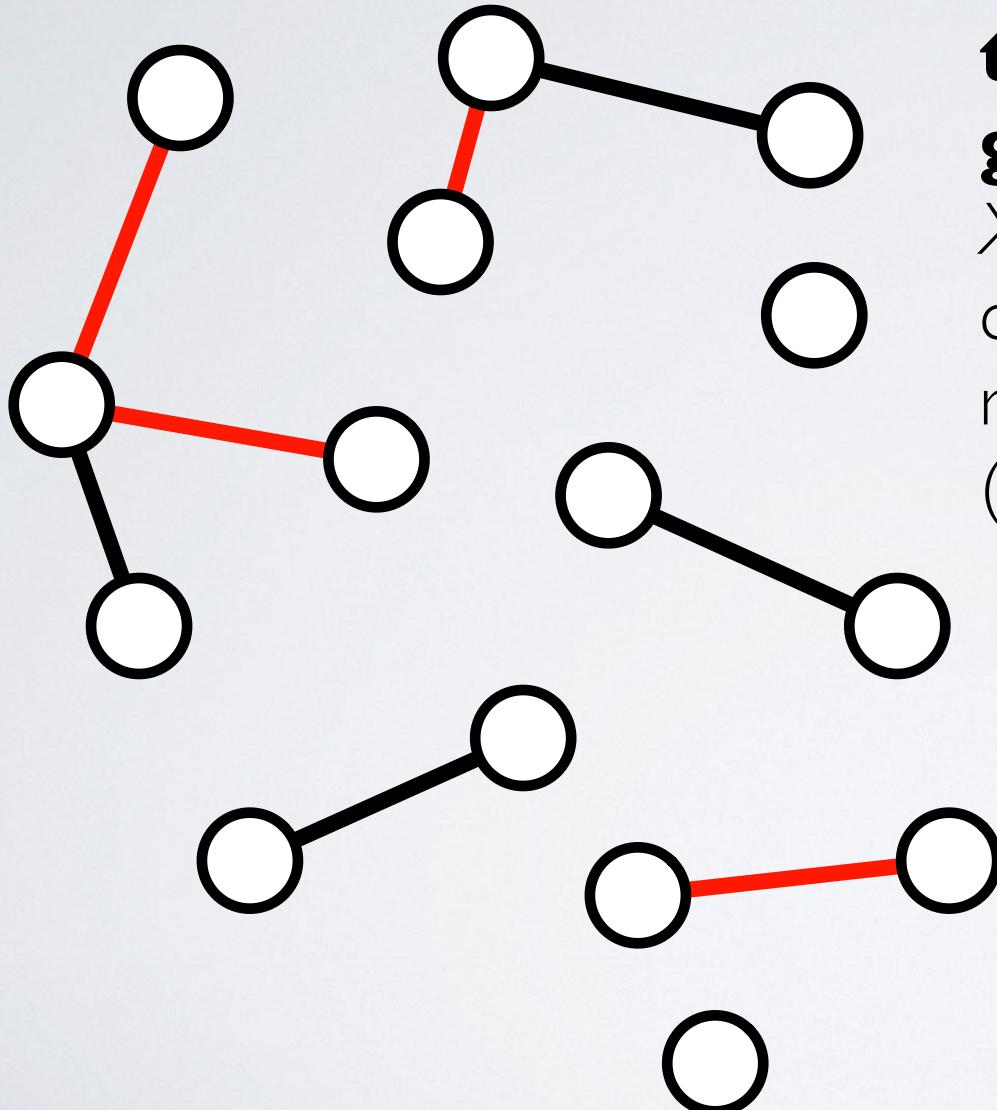


the percolation game:

choose
X random pairs
connect them
repeat
(count components)

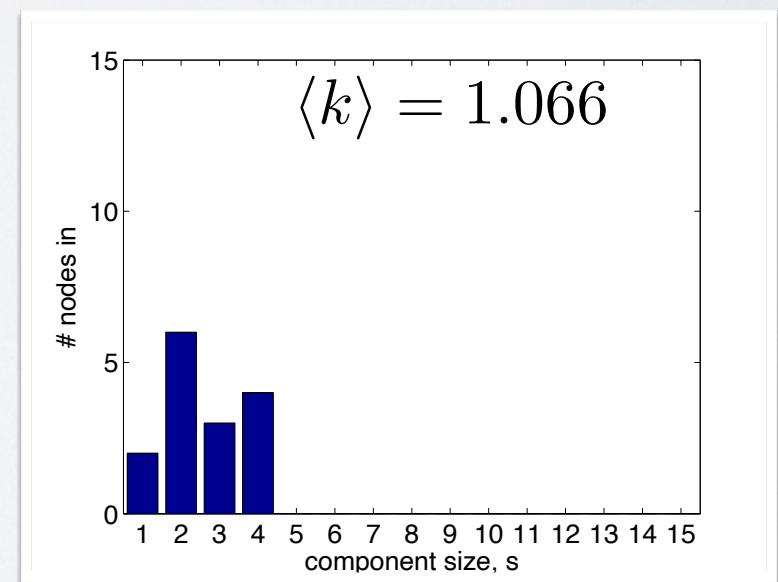


network components

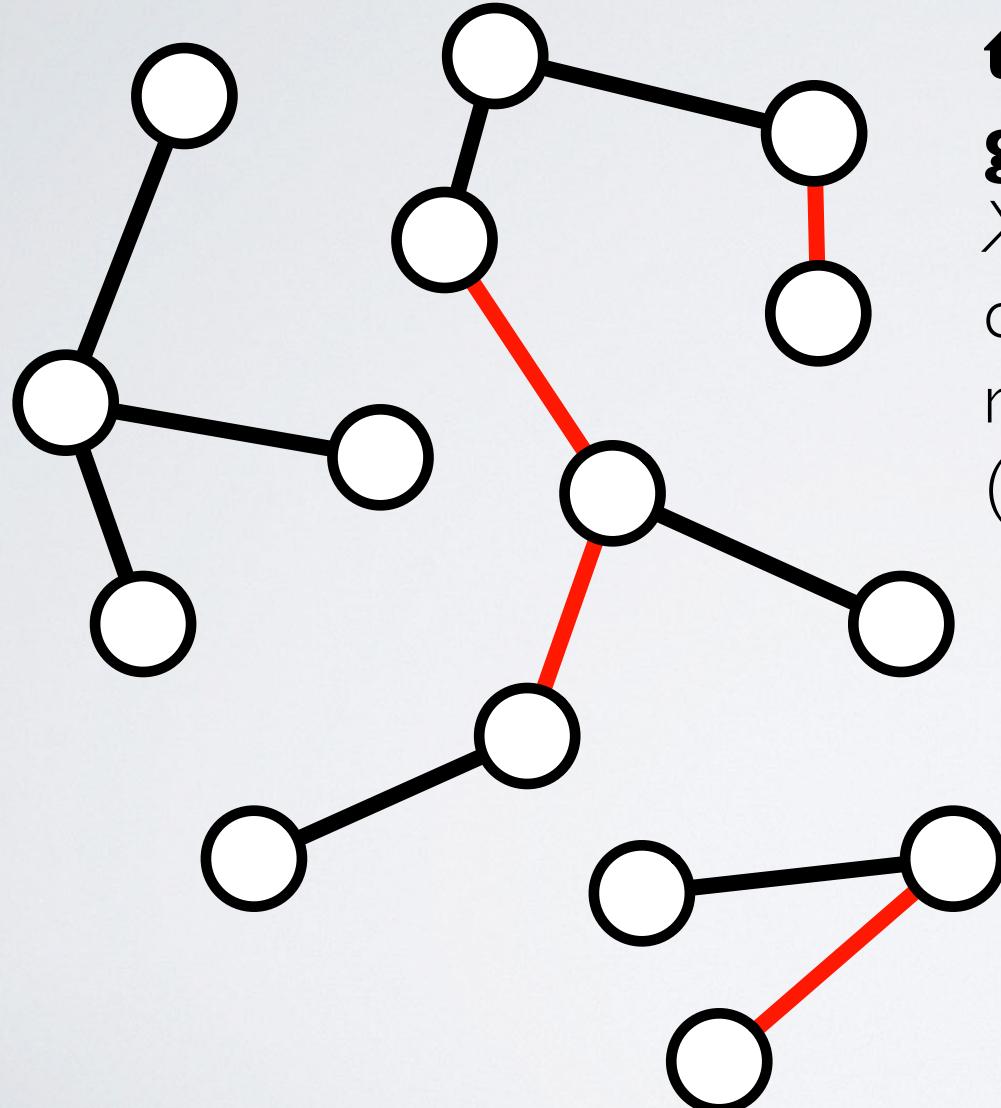


the percolation game:

choose
X random pairs
connect them
repeat
(count components)

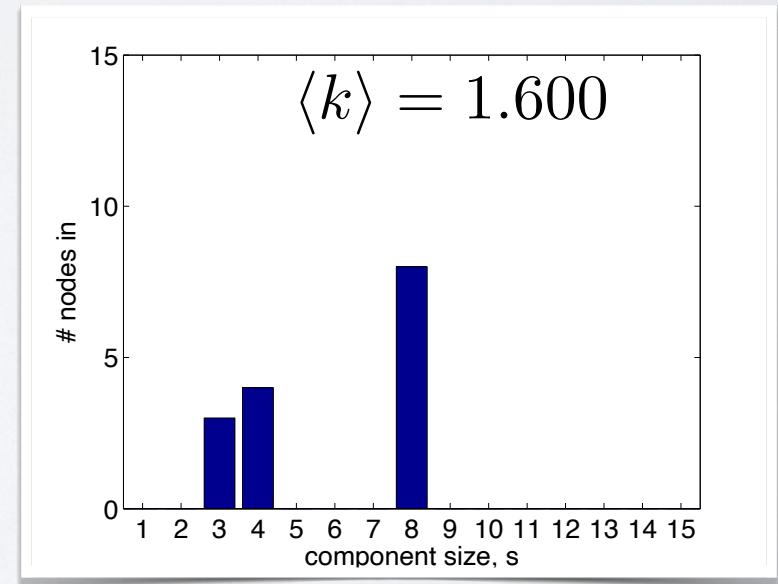


network components

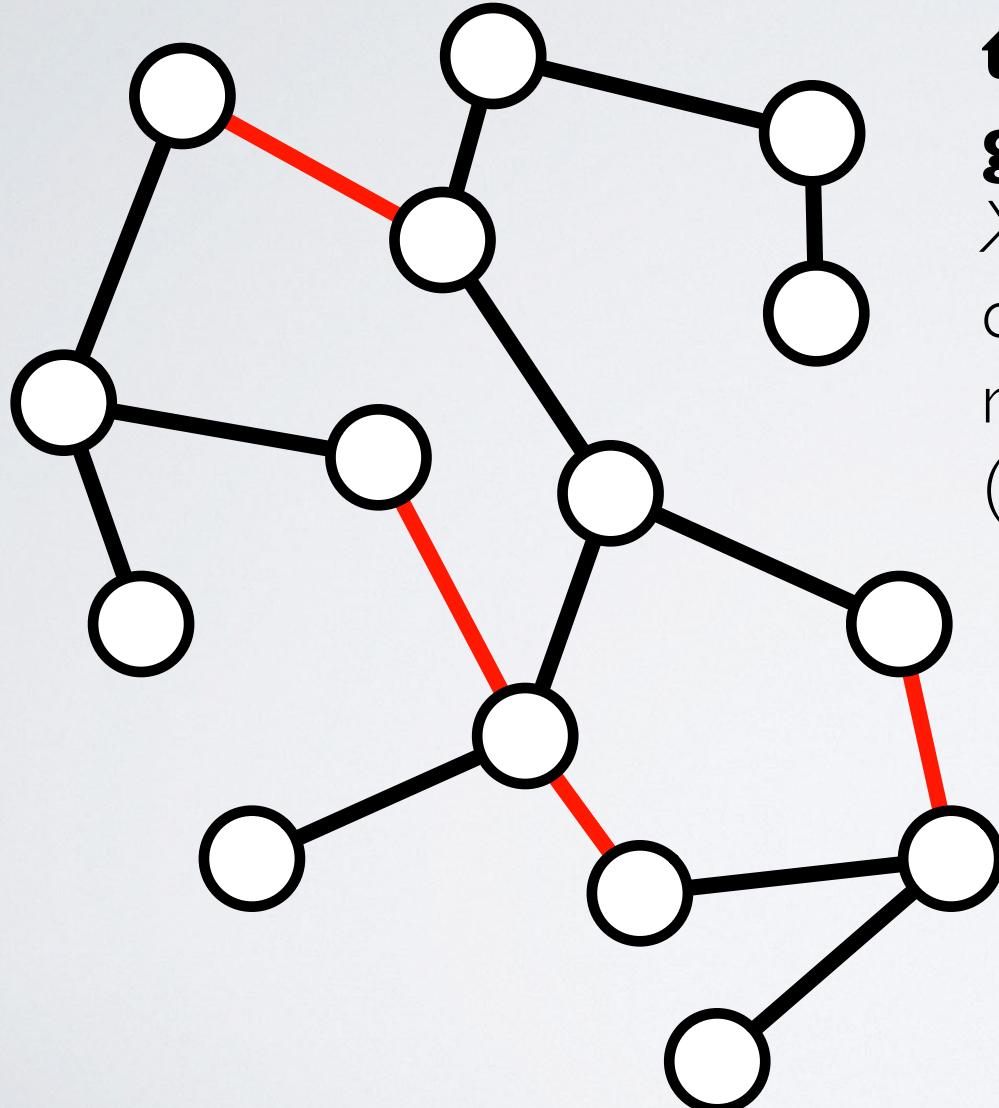


the percolation game:

choose
X random pairs
connect them
repeat
(count components)

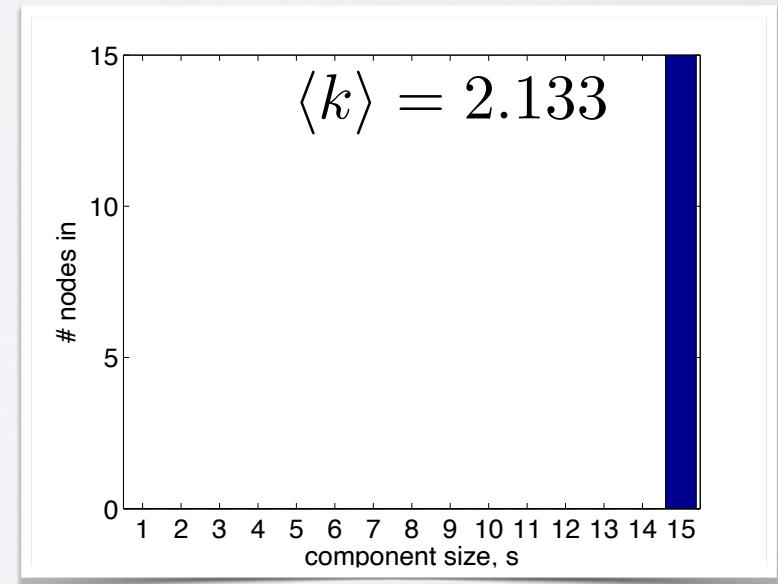


network components

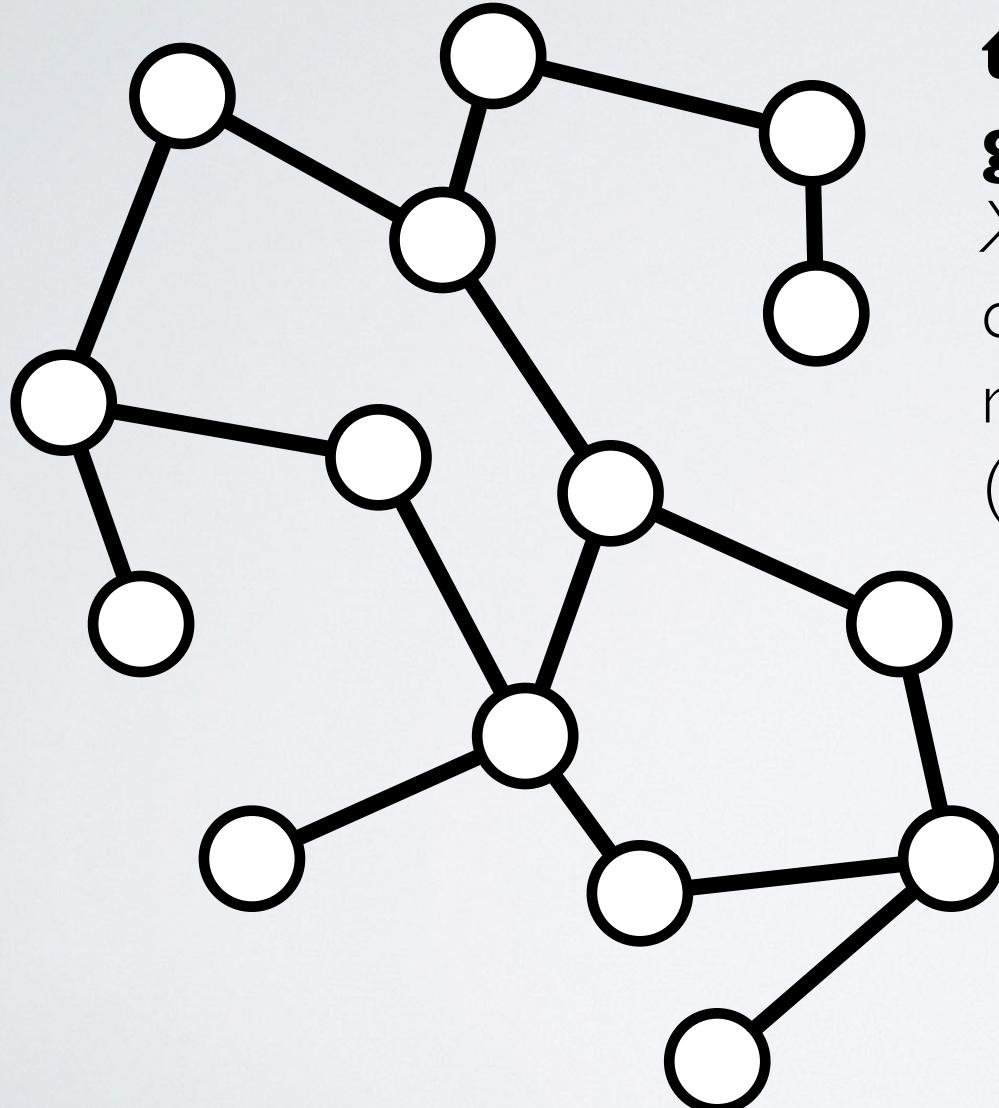


the percolation game:

choose
X random pairs
connect them
repeat
(count components)

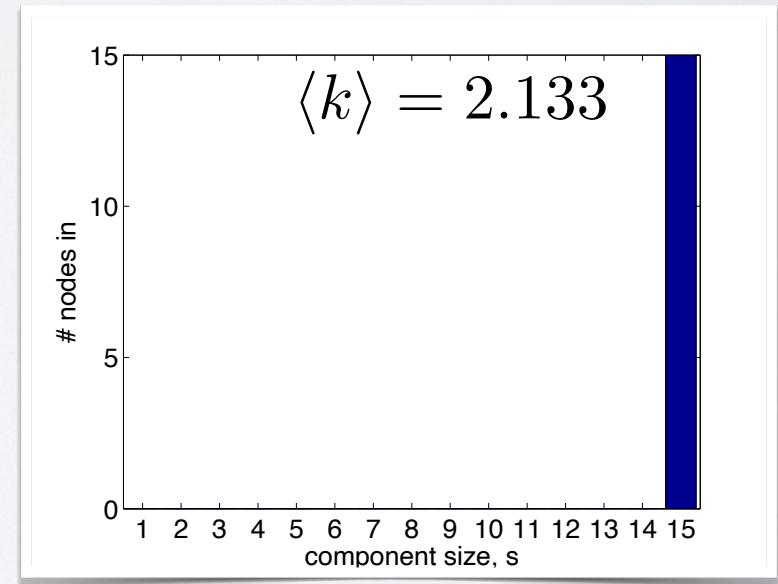


network components



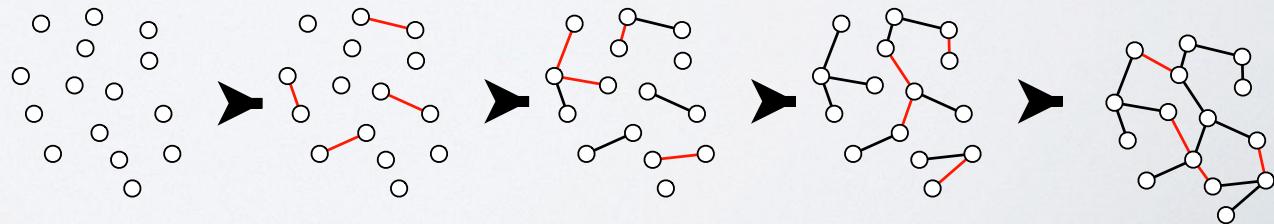
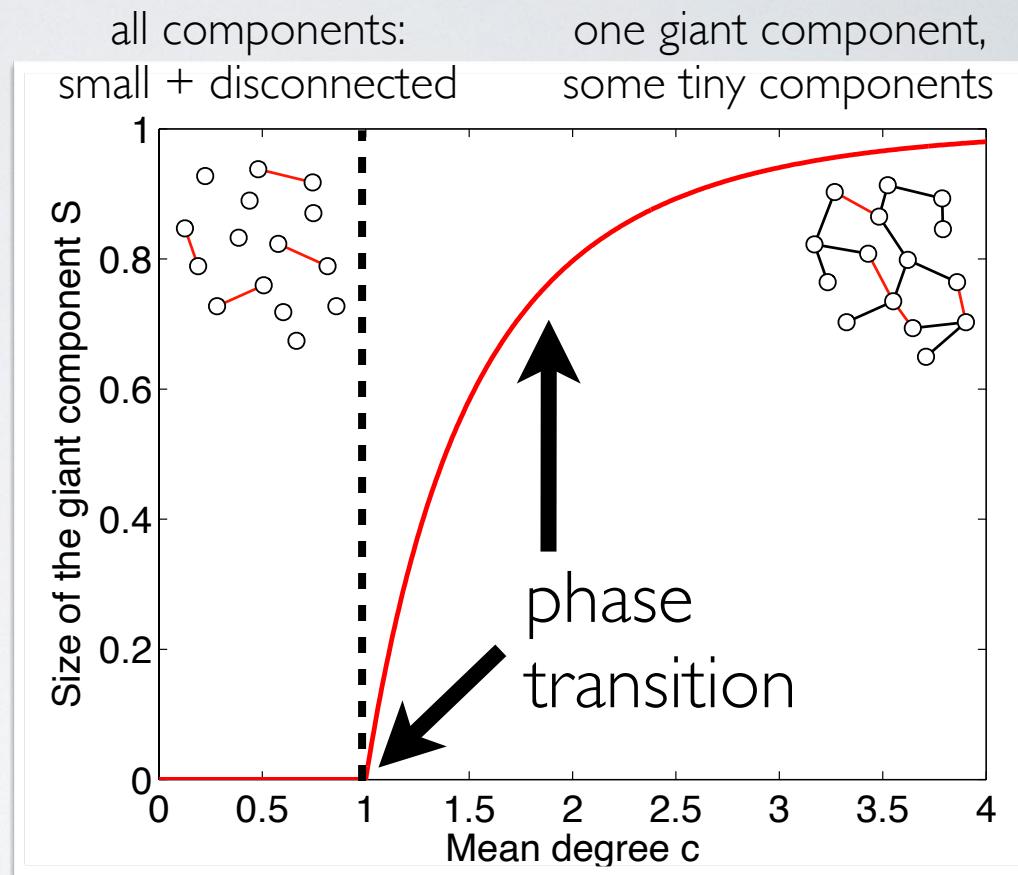
the percolation game:

choose
X random pairs
connect them
repeat
(count components)



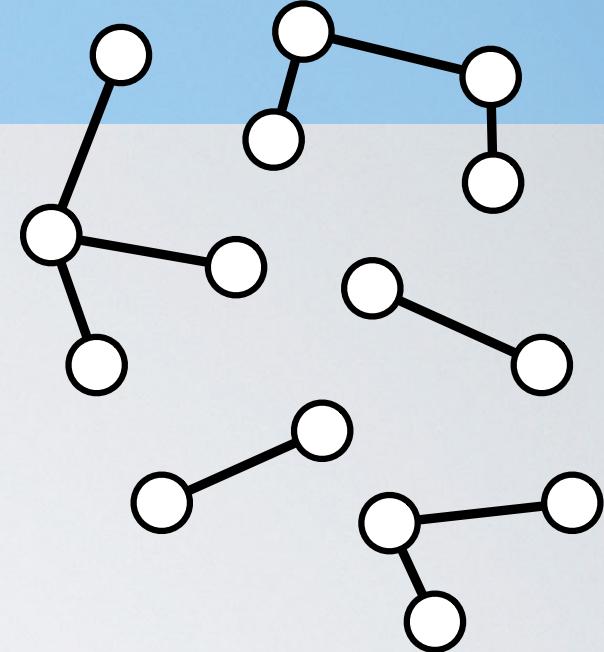
the “giant” component

- add edges randomly
- at first, components are small and disconnected
- at critical value, these components begin linking
- beyond, all nodes in single “giant” component



network components

- component = connected group
- component dynamics are independent (no information flow)
- *phase transition*: sudden emergence of new behavior (giant component)



open questions:

- other network properties + phase transitions
- adaptive wiring
- local vs. global connectivity rules