

Five Lectures on Networks

Aaron Clauset

 @aaronclauset

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

lecture 3: describing network position and understanding assortative mixing



University of Colorado **Boulder**

Network Analysis and Modeling

Instructor: Aaron Clauset

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)

<http://santafe.edu/~aarond/courses/5352/>

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](#)

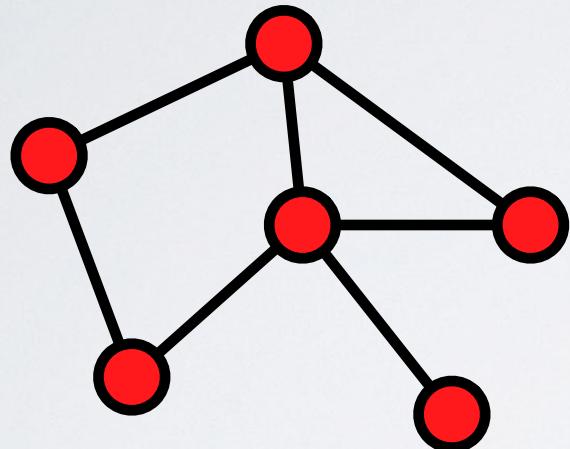
Data sets

[Mark Newman's network data sets](#)
[Stanford Network Analysis Project](#)
[Carnegie Mellon CASOS data sets](#)
[NCEAS food web data sets](#)
[UCI NET data sets](#)
[Pajek data sets](#)
[Linkgroup's list of network data sets](#)
[Barabasi lab data sets](#)
[Jake Hofman's online network data sets](#)
[Alex Arenas's data sets](#)

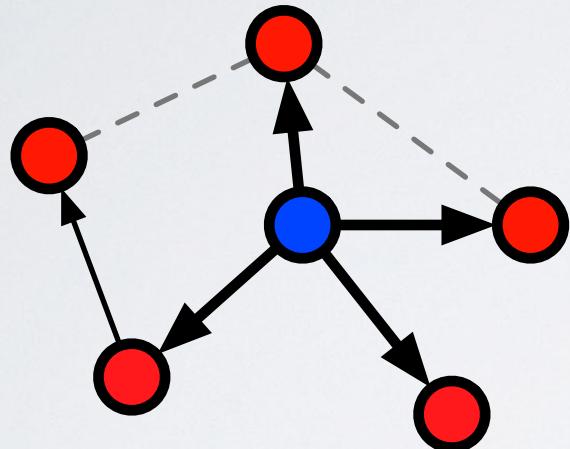
1. defining a network
2. describing a network
- 3. null models for networks**
4. statistical inference

describing networks

position



describing networks

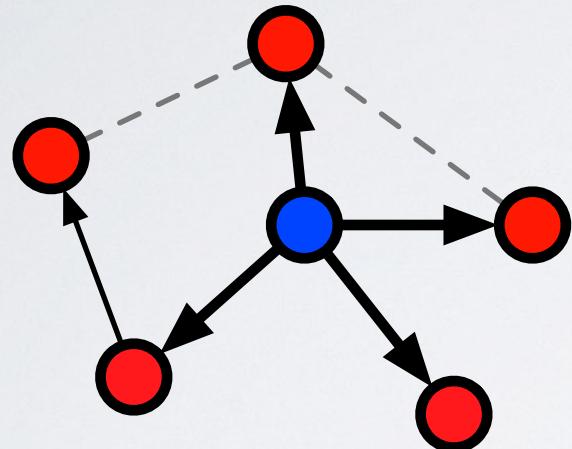


position = centrality:
measure of positional
“importance”

geometric connectivity

- harmonic centrality
- closeness centrality
- betweenness centrality
- degree centrality
- eigenvector centrality
- PageRank
- Katz centrality
- many many more...

describing networks



position = centrality:
harmonic, closeness
centrality

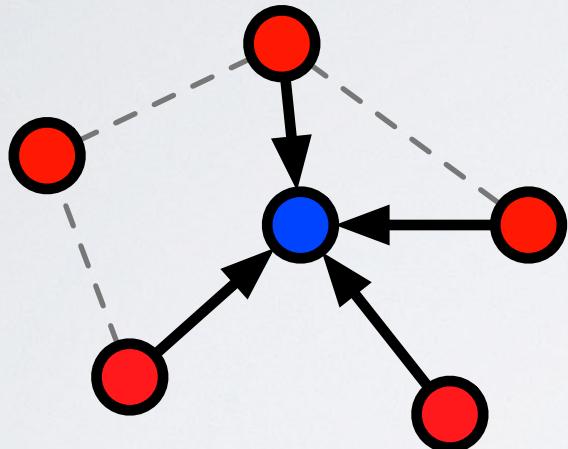
importance = being in
“center” of the network

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

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 left eigenvector of

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

*modulo several technical details

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

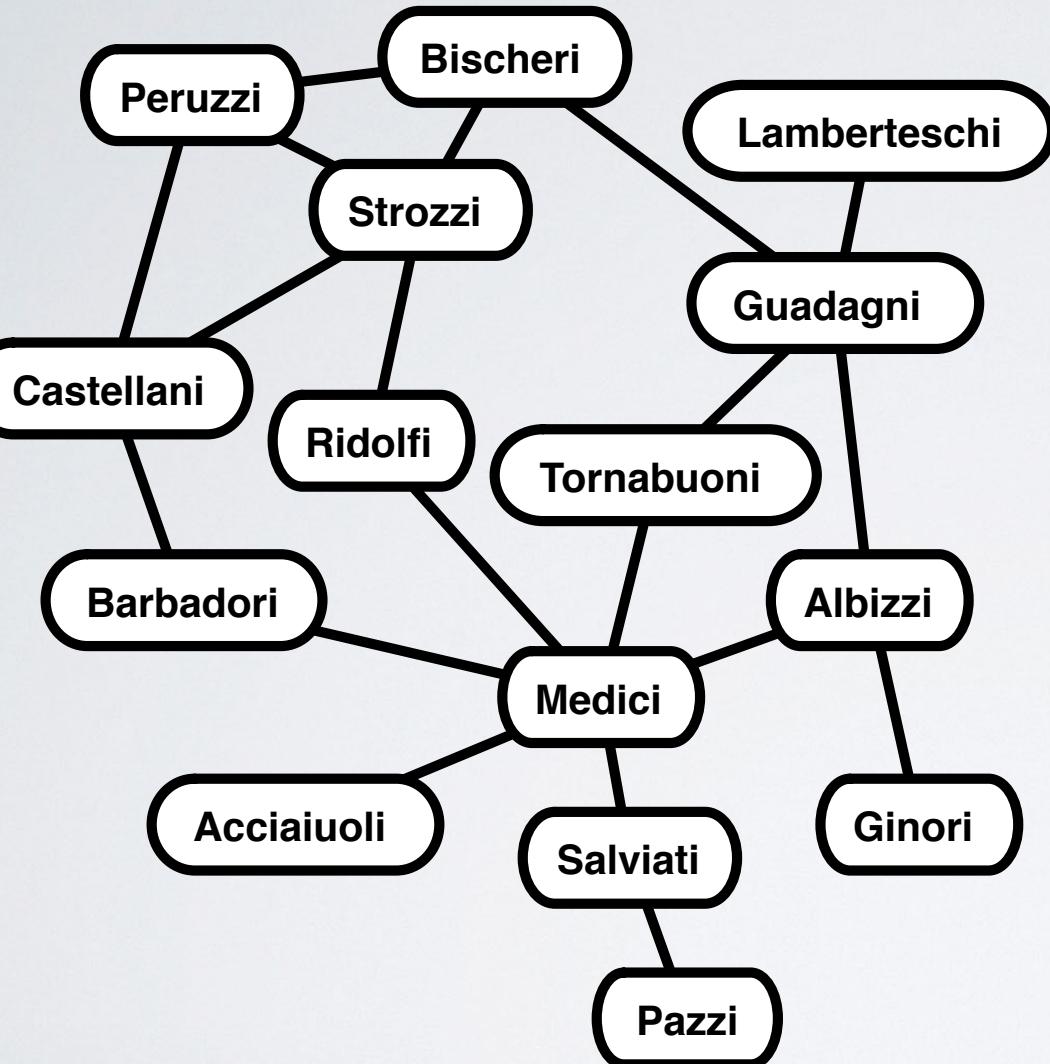


Palazzo Medici



Giovanni de Medici

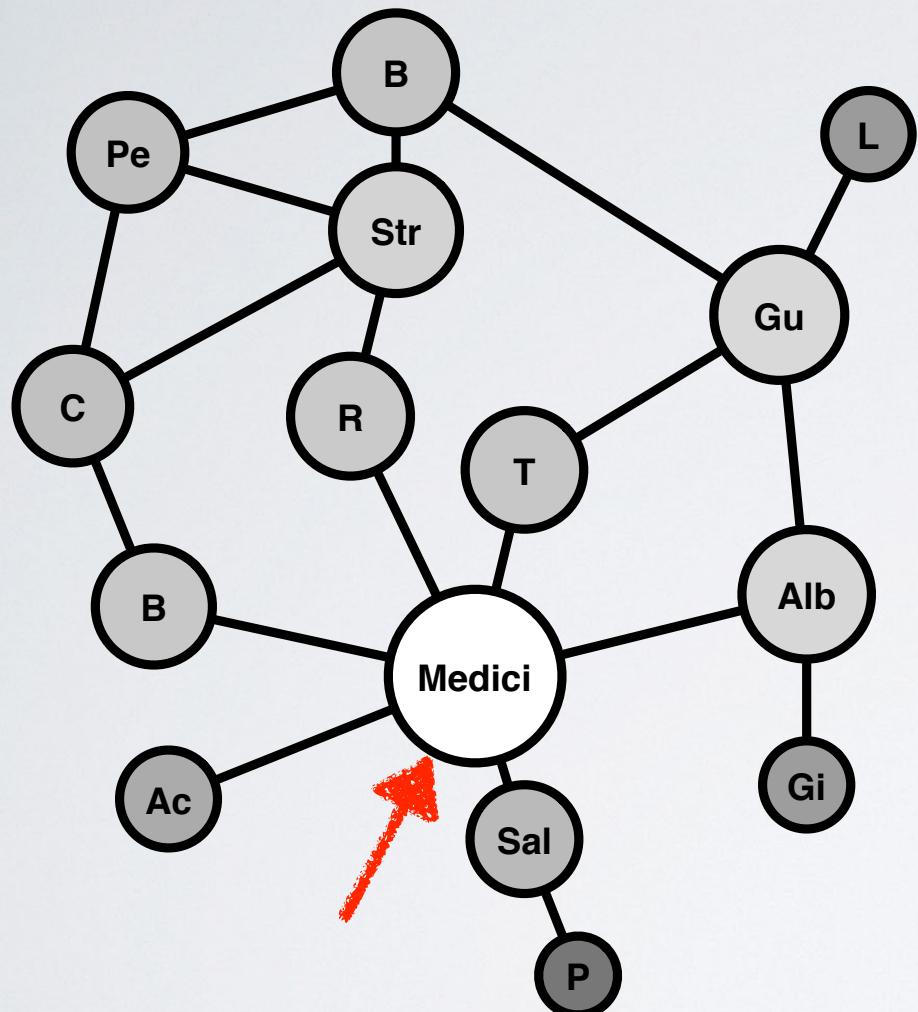
network position: closeness



nodes: Florence families
edges: inter-family marriages

which family is most central?

network position: closeness

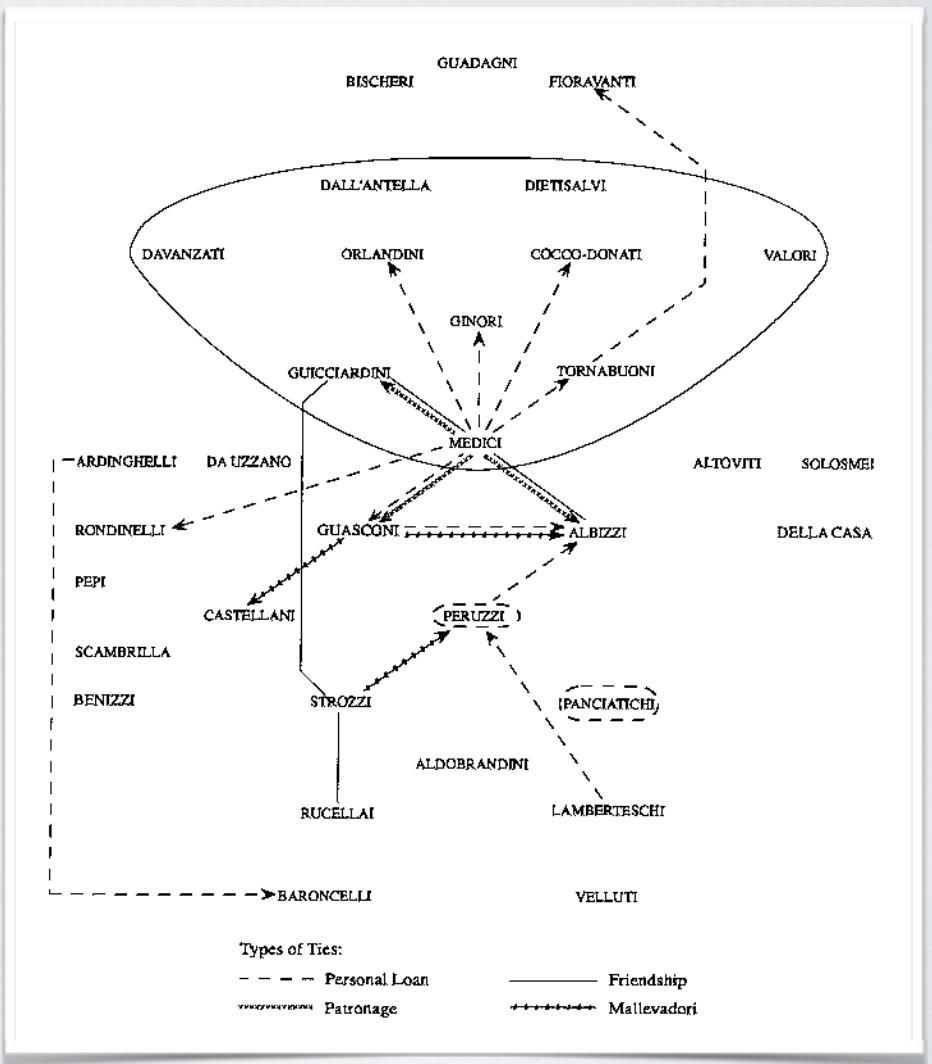
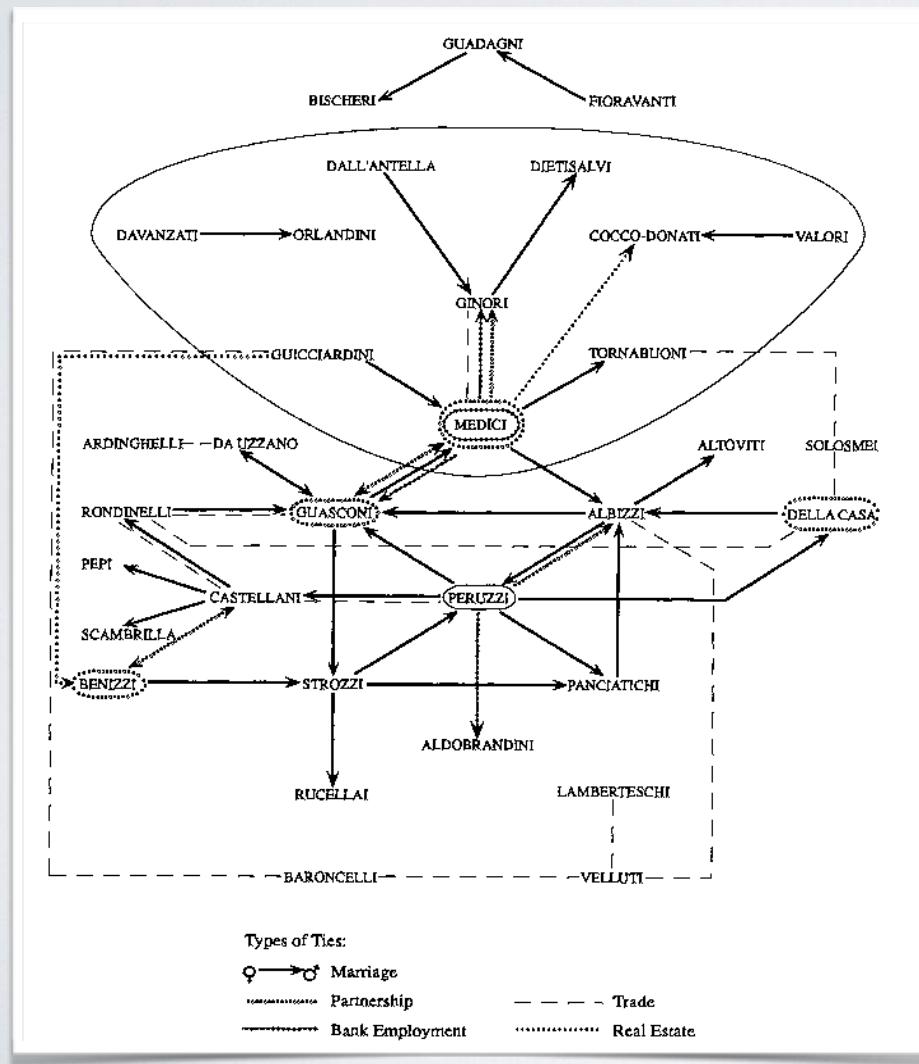


Medici	9.5
Guadagni	7.92
Albizzi	7.83
Strozzi	7.67
Ridolfi	7.25
Bischeri	7.2
Tornabuoni	7.17
Barbadori	7.08
Peruzzi	6.87
Castellani	6.87
Salviati	6.58
Acciaiuoli	5.92
Ginori	5.33
Lamberteschi	5.28
Pazzi	4.77



network position

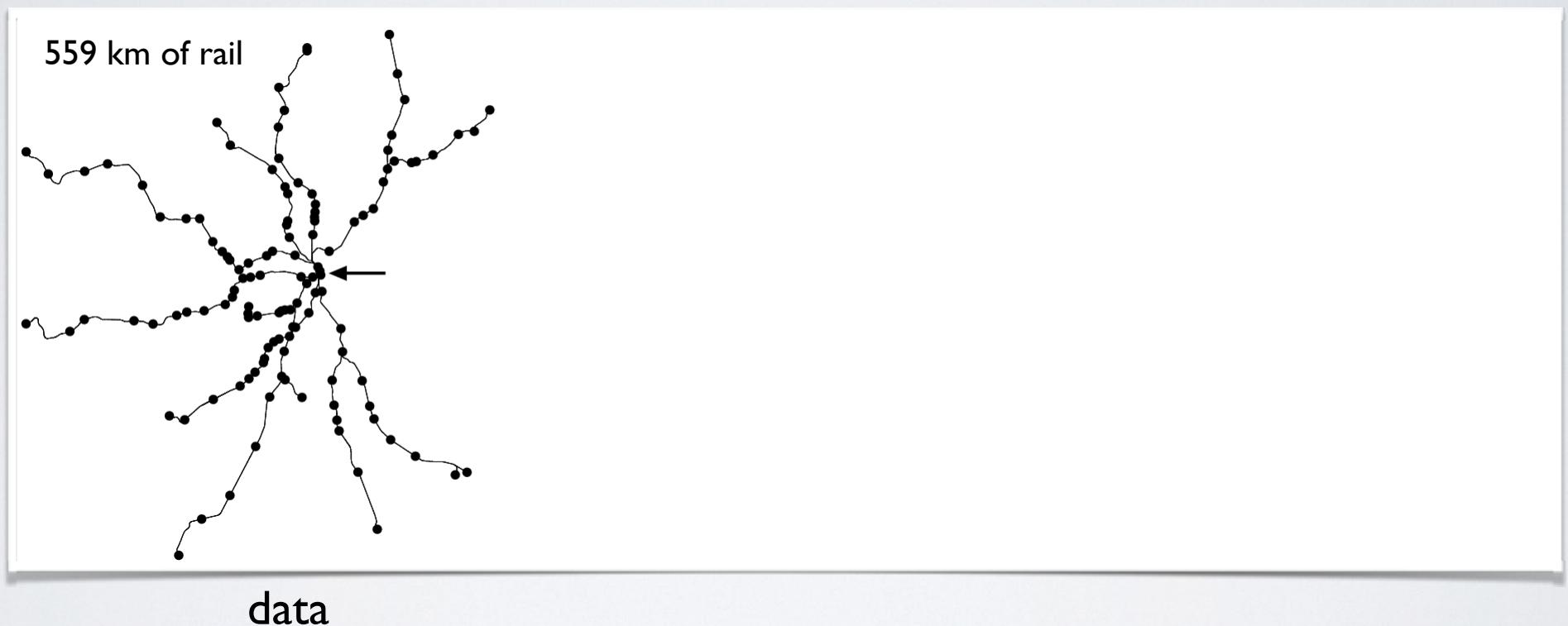
actually, it's complicated...



network position

an example

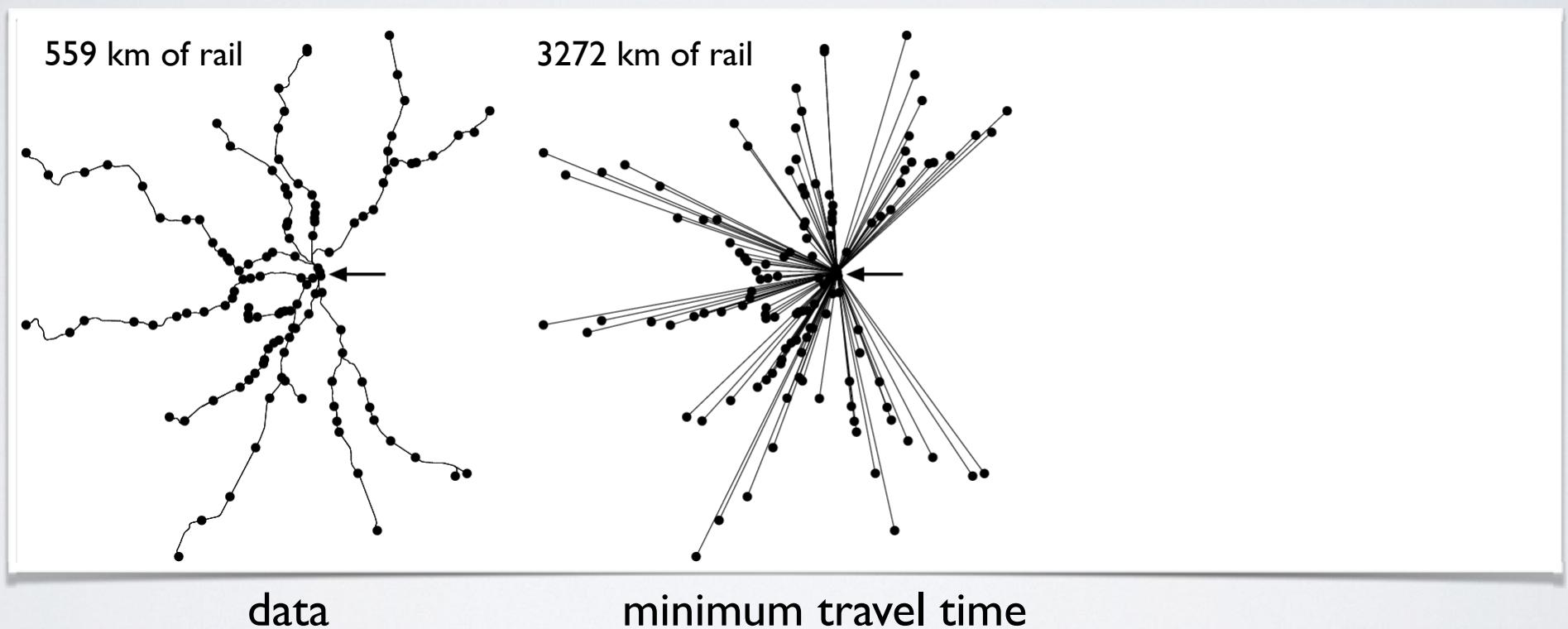
Boston commuter rail



network position

an example

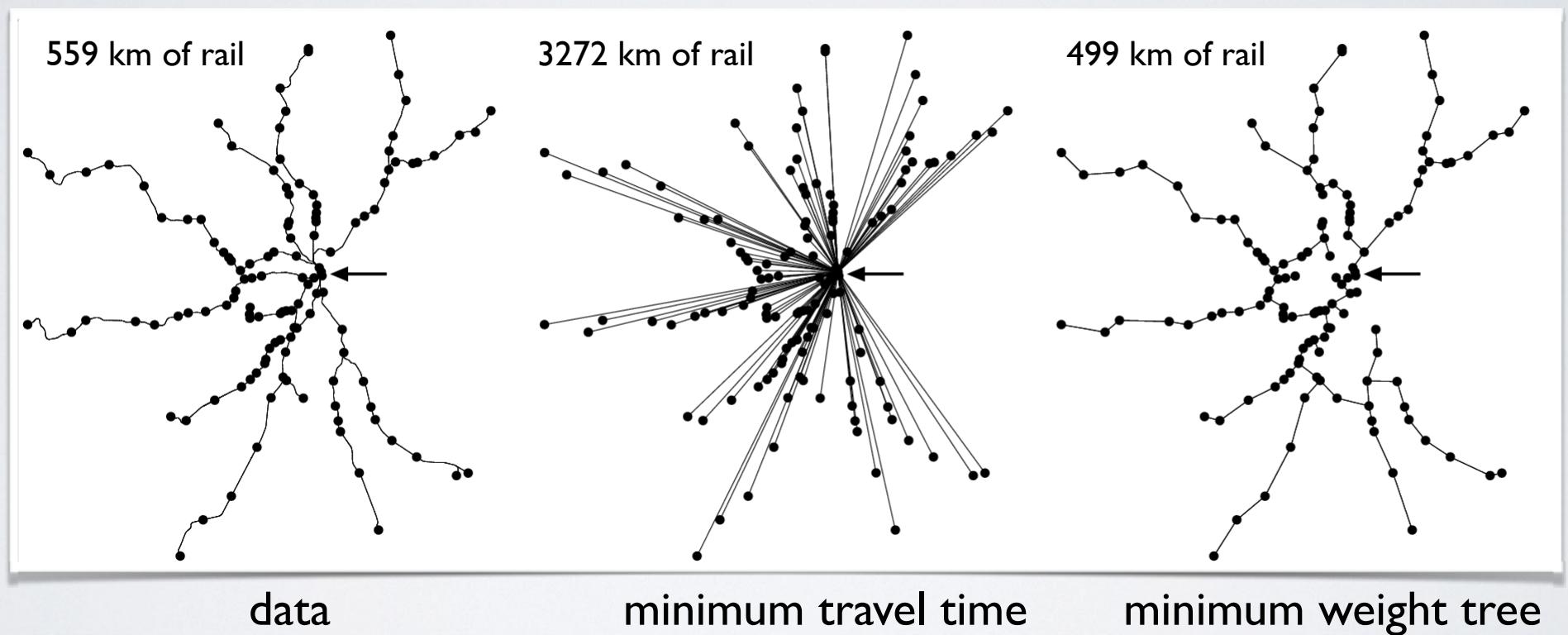
Boston commuter rail



network position

an example

Boston commuter rail

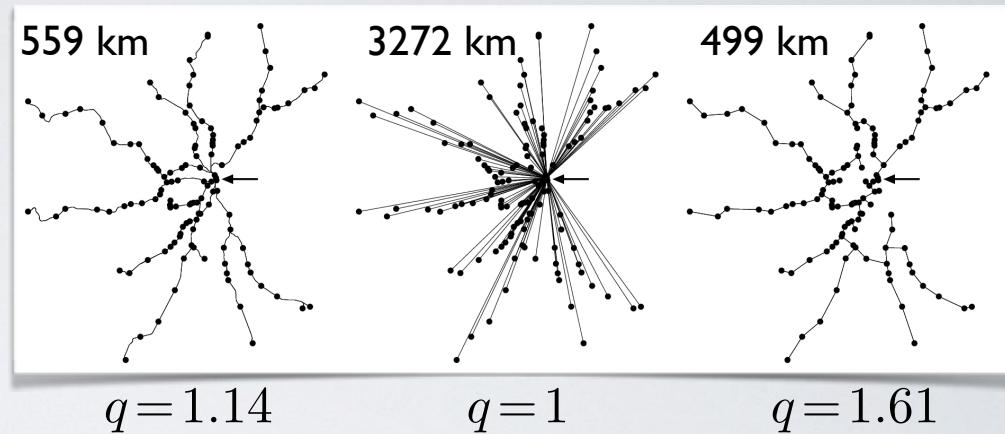


network position

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

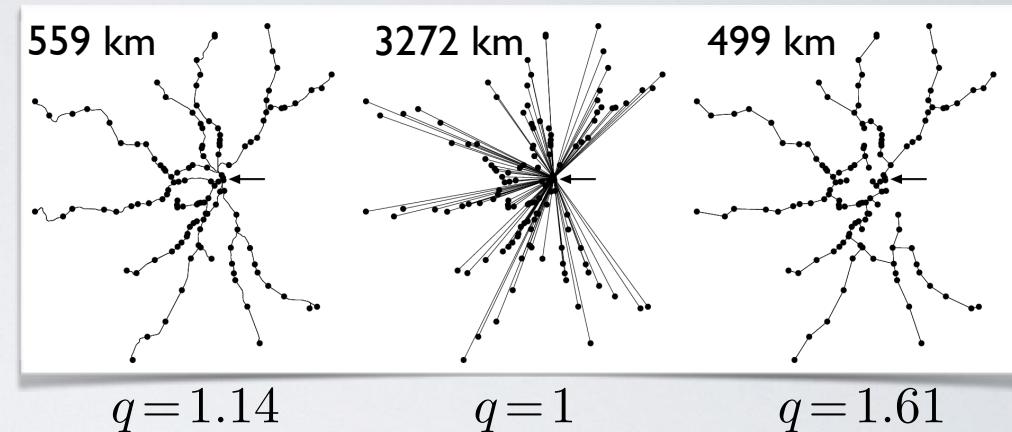
add edge (i, j) with
minimum value for

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

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

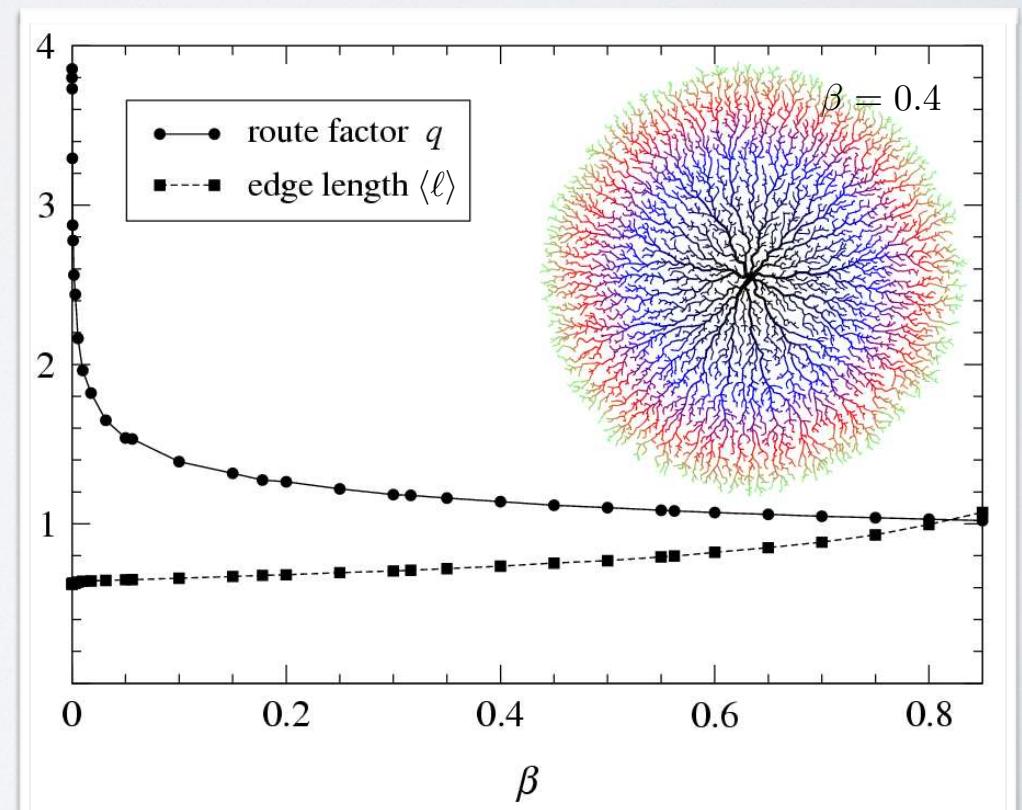
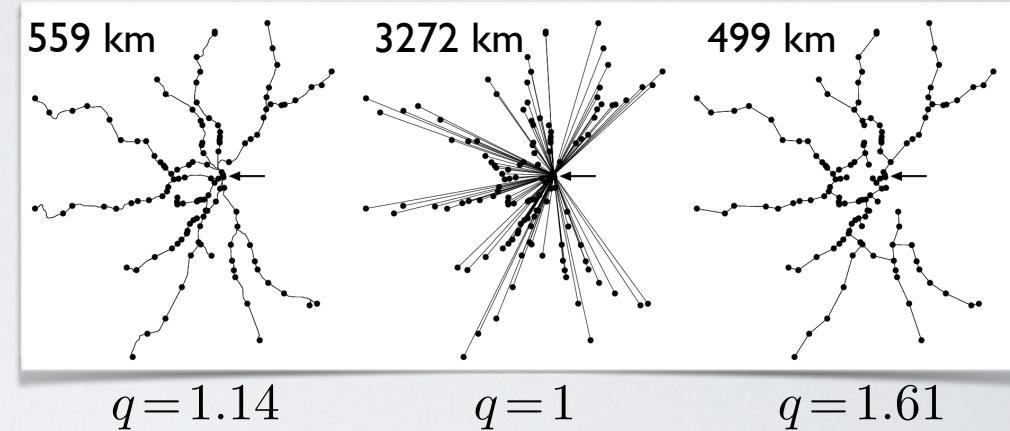
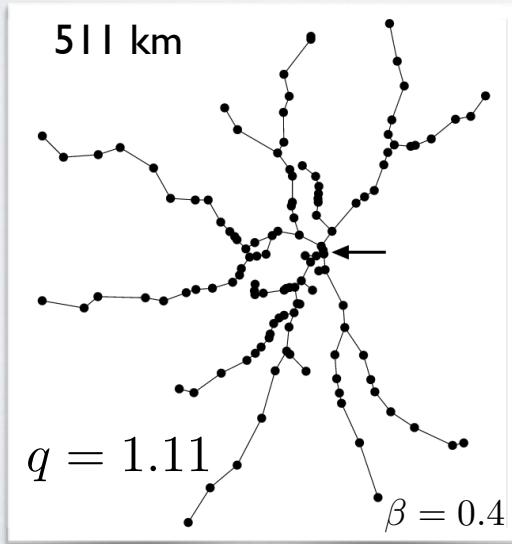
a simple model

embed n vertices in a plane

until all vertices connected

add edge (i, j) with
minimum value for

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



network position

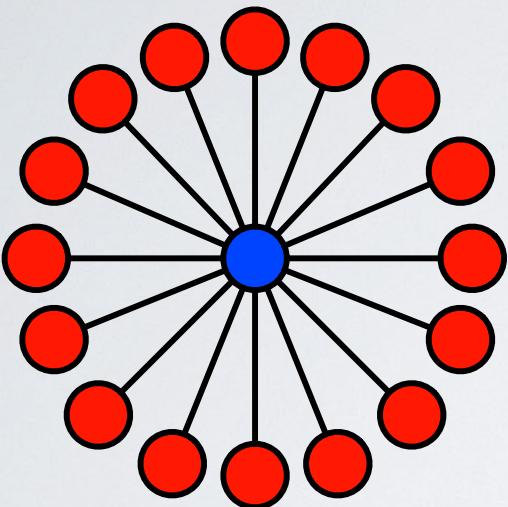


most
centralized

vast wilderness
of in-between

most
decentralized

network position

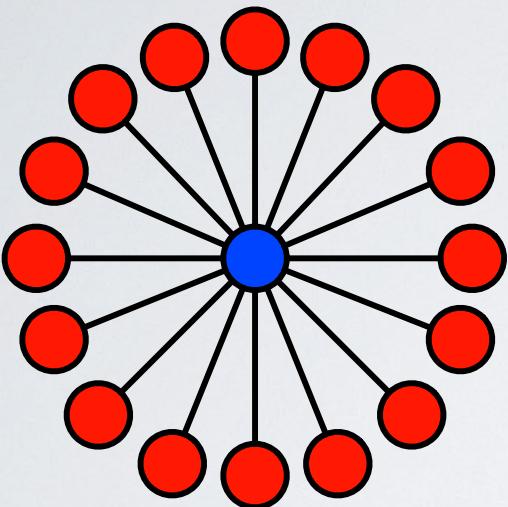


most
centralized

vast wilderness
of in-between

most
decentralized

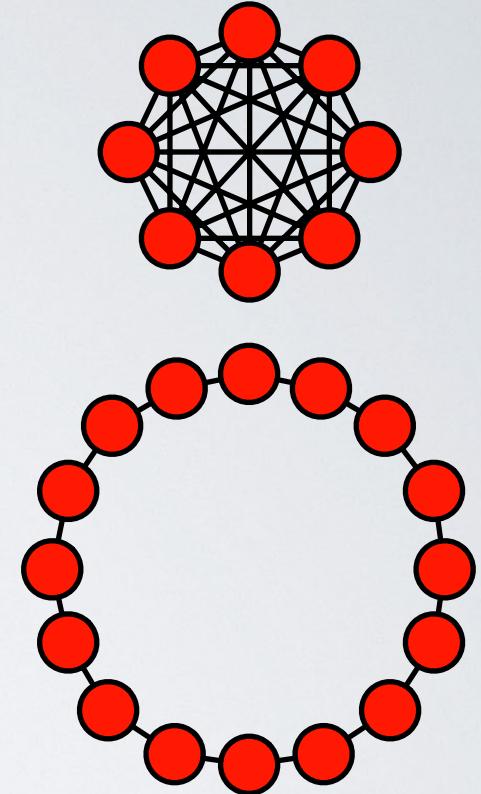
network position



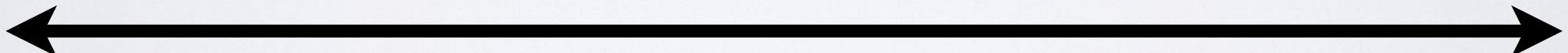
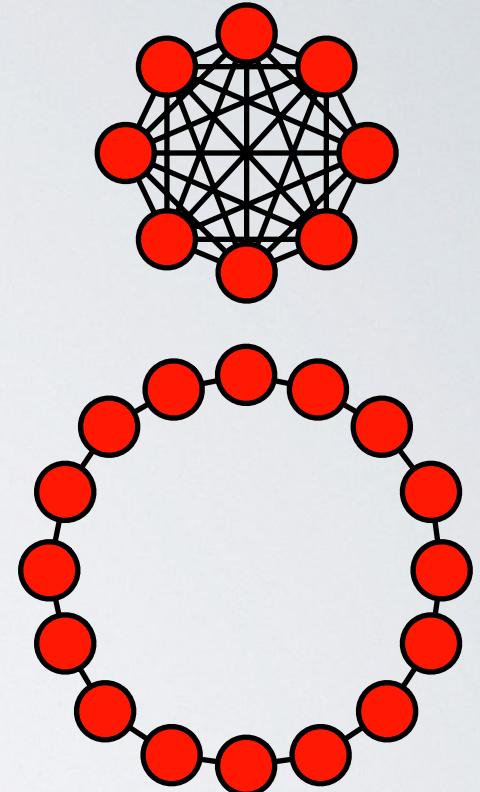
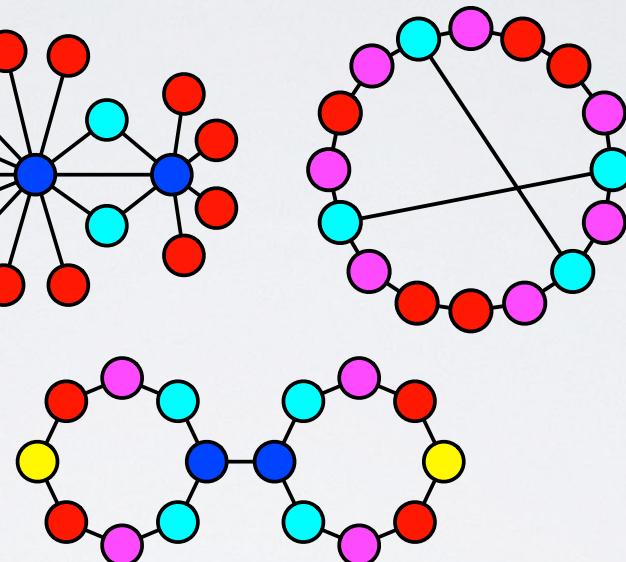
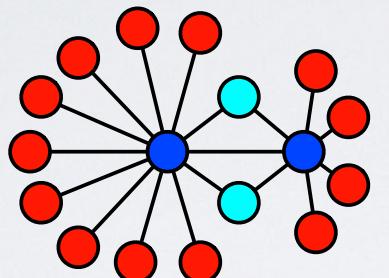
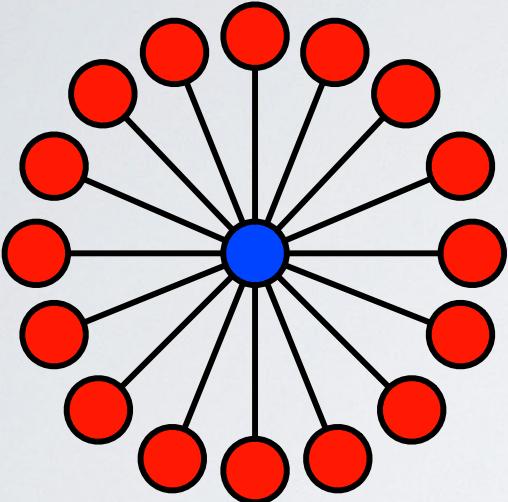
most
centralized

vast wilderness
of in-between

most
decentralized



network position

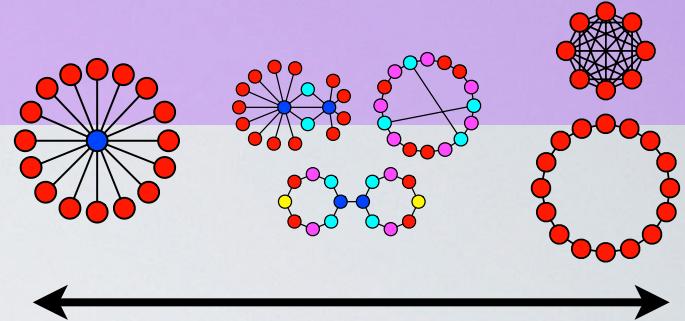


most
centralized

vast wilderness
of in-between

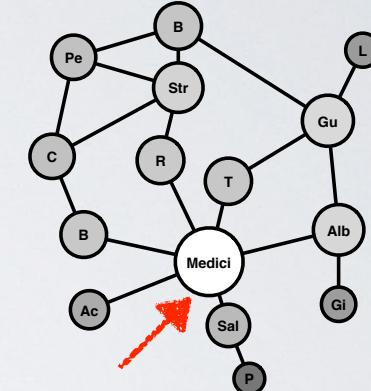
most
decentralized

network position



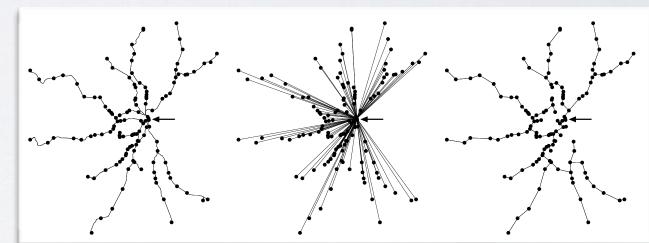
positions:

- geometric description of network structure
- core vs. periphery
- centrality = importance, influence

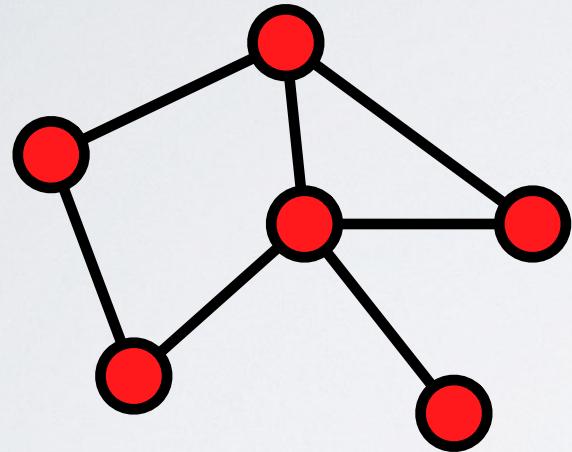


open questions:

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

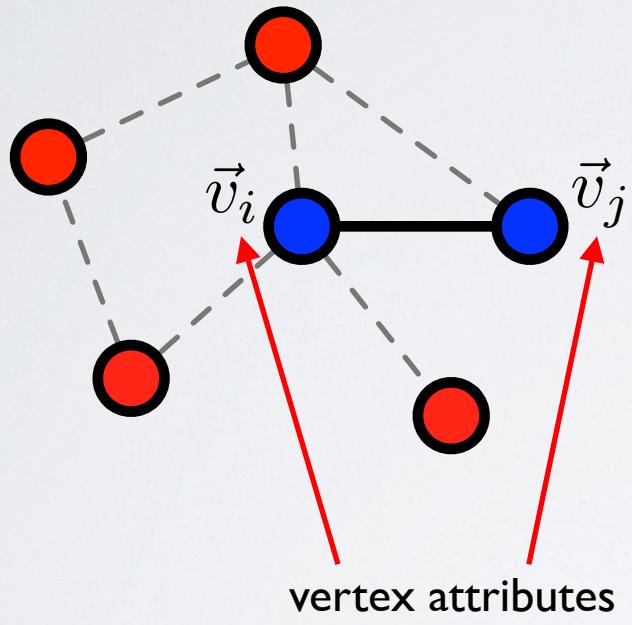


describing networks



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

assortative mixing



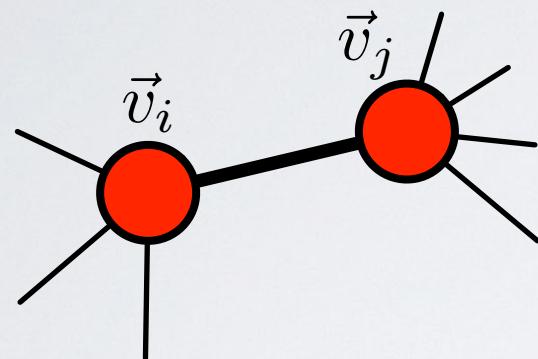
homophily and assortative mixing

like links with like

assortativity coefficient r
quantifies homophily

three types:
scalar attributes
vertex degrees
categorical variables

assortative mixing



homophily and assortative mixing

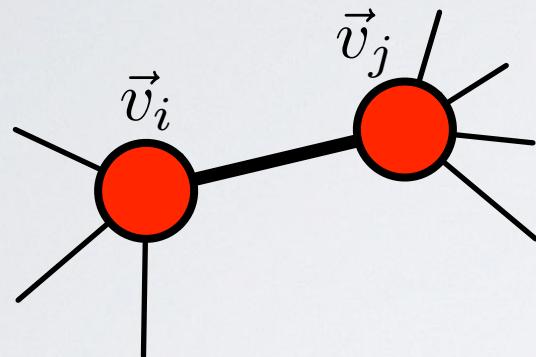
like links with like

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

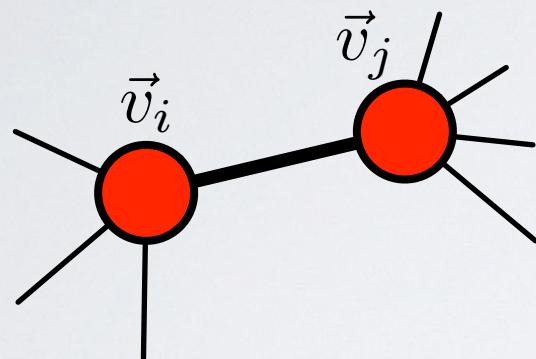
scalar attributes:
covariance across ties

$$\text{cov}(v_i, v_j) = \frac{\sum_{ij} A_{ij}(v_i - \mu)(v_j - \mu)}{\sum_{ij} A_{ij}}$$

$$= \frac{1}{2m} \sum_{ij} A_{ij} v_i v_j - \mu^2$$

$$= \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) v_i v_j$$

assortative mixing



homophily and assortative mixing

like links with like

assortativity coefficient (scalar)

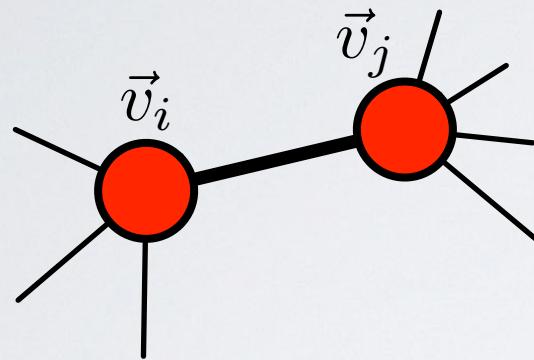
$$r = \frac{\text{cov}(v_i, v_j)}{\text{var}(v_i, v_j)}$$

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

[this is just a Pearson correlation across edges]

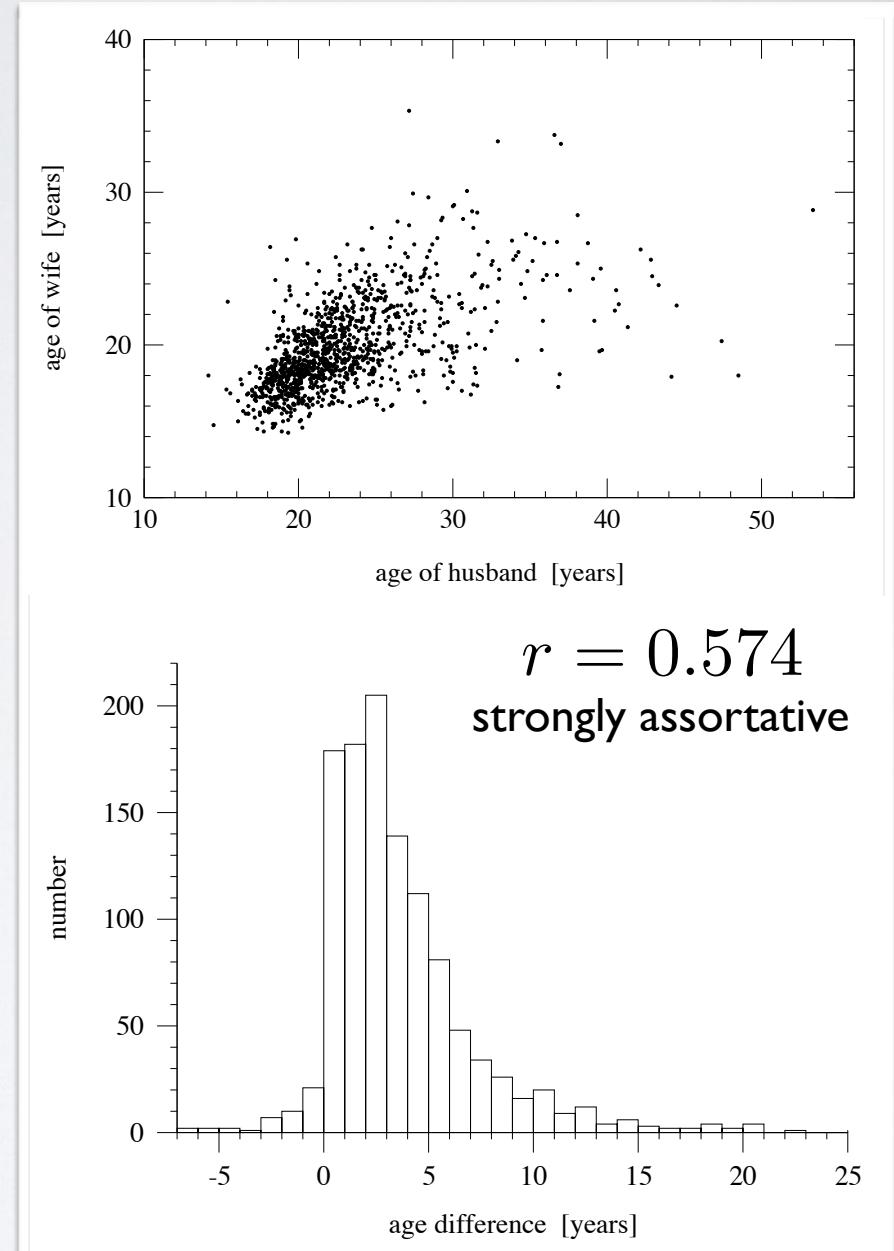
$$-1 \leq r \leq 1$$

assortative mixing

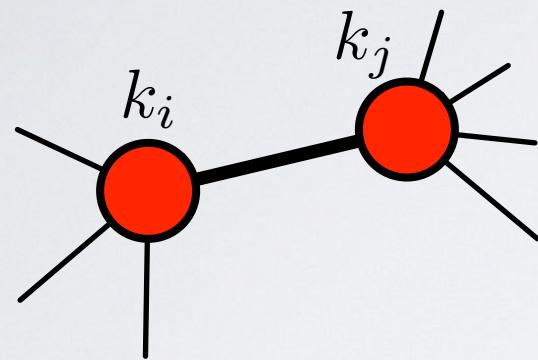


(top) scatter plot of ages of 1141 married couples at time of marriage [1995 US National Survey of Family Growth]

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



assortative mixing

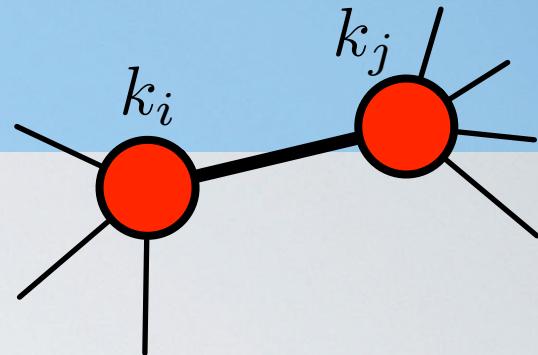


homophily and assortative mixing

like links with like

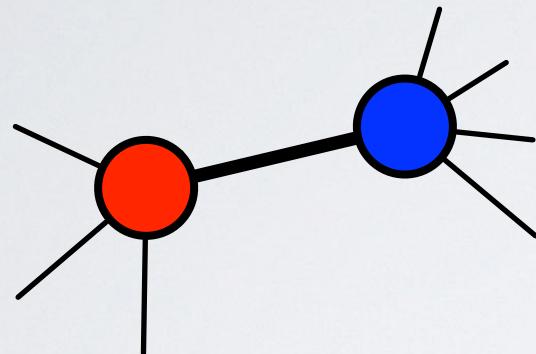
degree:
just another scalar^{*}

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

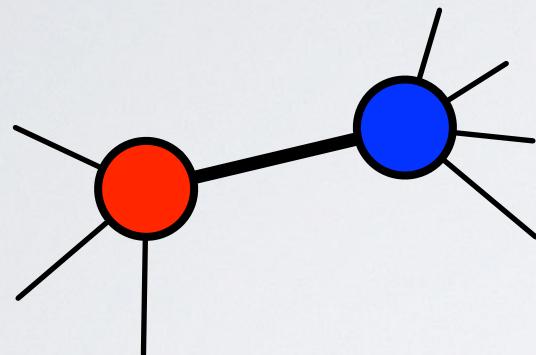
$$\sum_{ij} e_{ij} = 1$$

marginals

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

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

assortative mixing



homophily and assortative mixing

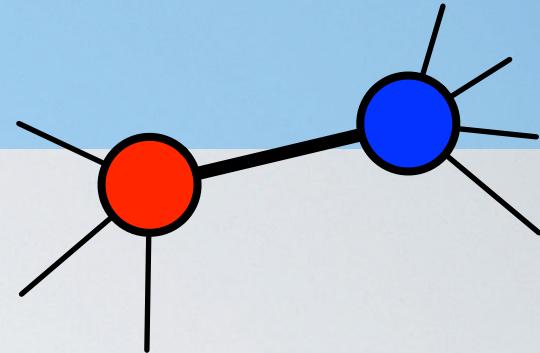
like links with like

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



1992 study of heterosexual partnerships in San Francisco*
(bipartite network)

		women				a_i
		black	hispanic	white	other	
men	black	0.258	0.016	0.035	0.013	0.323
	hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	other	0.005	0.007	0.024	0.016	0.053
b_i		0.289	0.204	0.423	0.084	

$$r = 0.621$$

strongly assortative

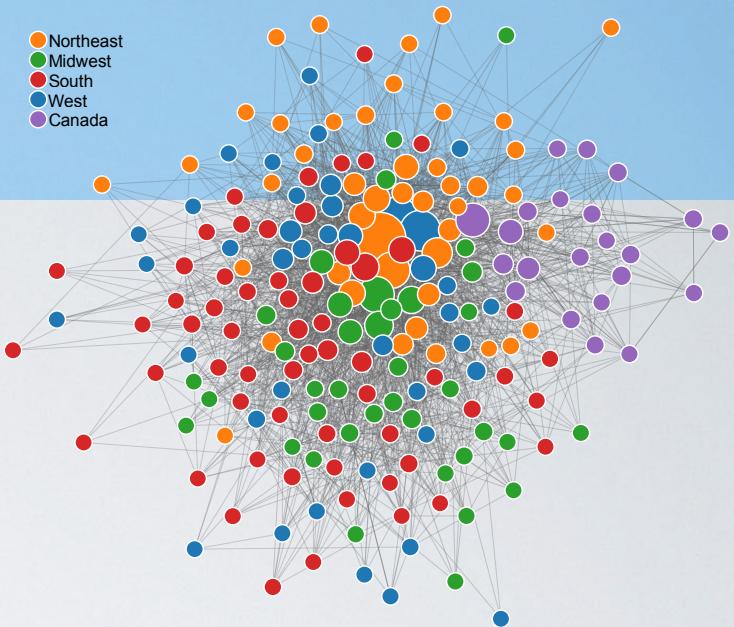
assortative mixing

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

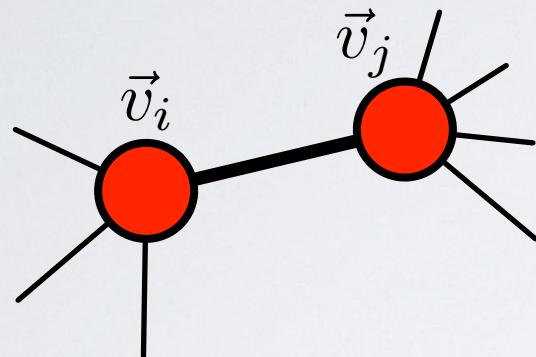


	Northeast	Midwest	South	West	Canada	a_i
Northeast	0.119	0.053	0.074	0.055	0.022	0.322
Midwest	0.031	0.067	0.061	0.026	0.011	0.196
South	0.025	0.027	0.083	0.024	0.006	0.166
West	0.049	0.033	0.043	0.073	0.011	0.209
Canada	0.006	0.005	0.005	0.005	0.085	0.107
b_i	0.229	0.185	0.267	0.184	0.135	

$$r = 0.264$$

moderately assortative

assortative mixing



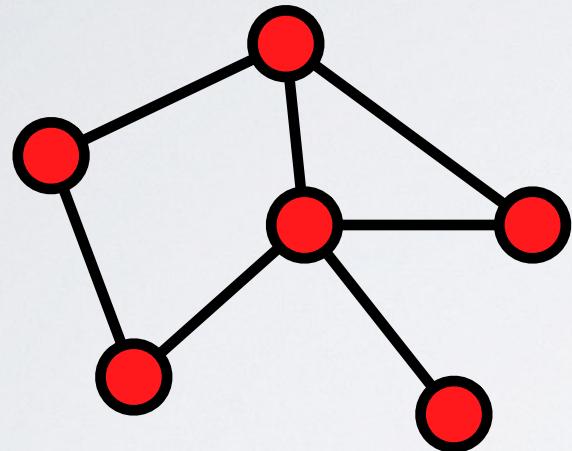
homophily and assortative mixing

like links with like

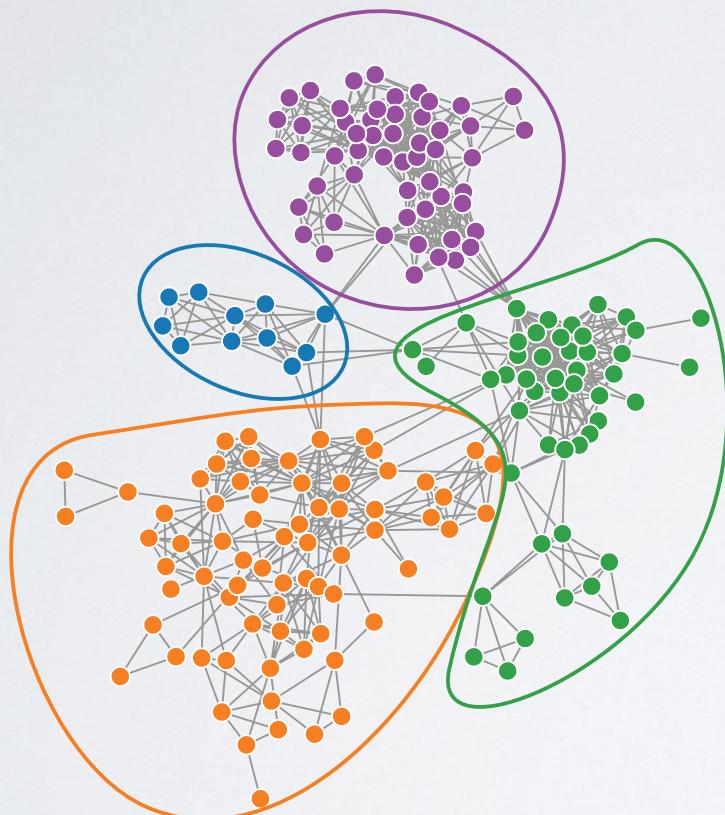
- random graphs tend to be disassortative $r \leq 0$ because the mixing is uniform
- social networks (apparently) highly assortative, in every way (attribute, degree, category)
- extremal values $r \approx \{-1, 1\}$ suggest underlying mechanism on that variable

describing networks

community structure



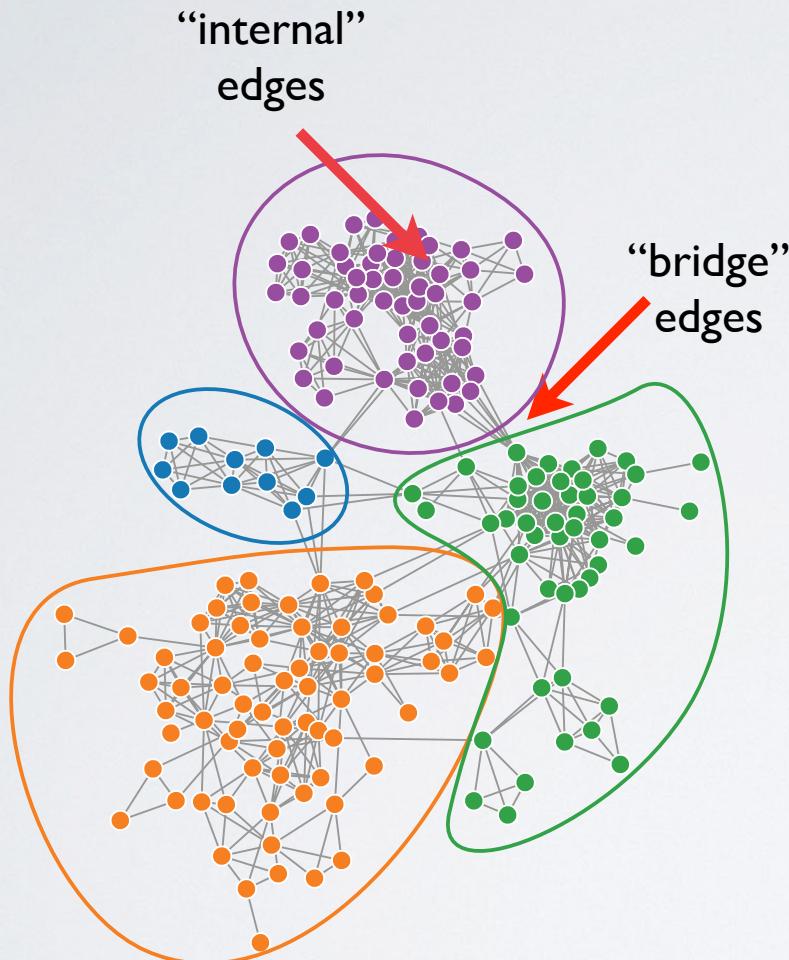
describing networks



assortative community structure
(edges inside the groups)

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

community structure

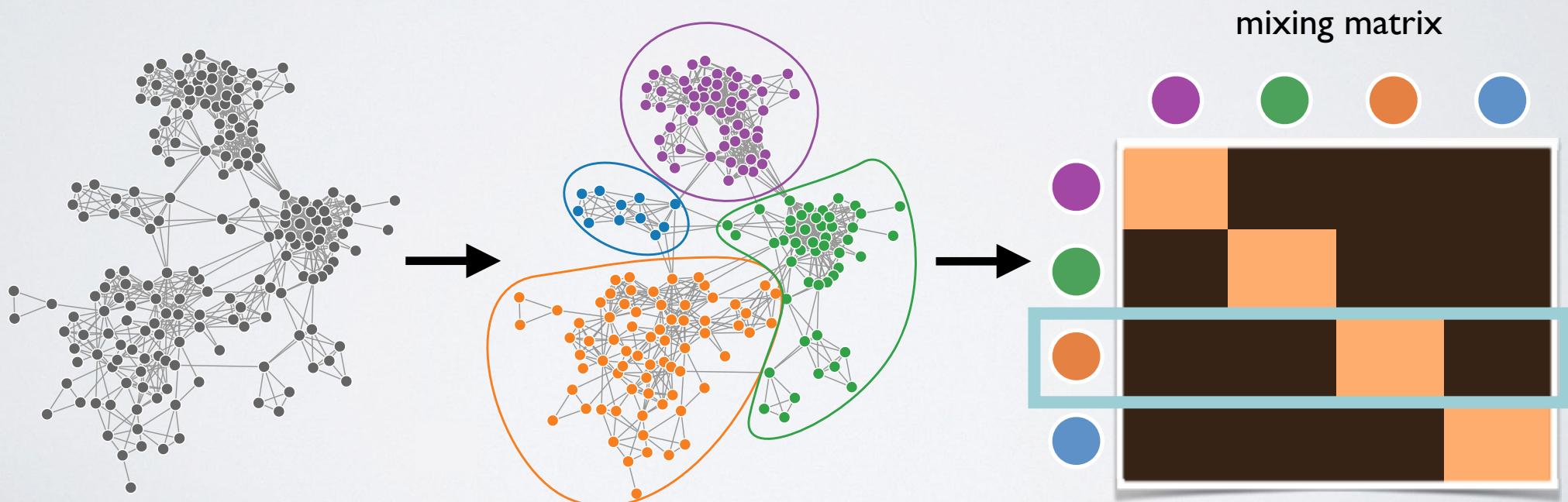


assortative community structure
(edges inside the groups)

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

community structure

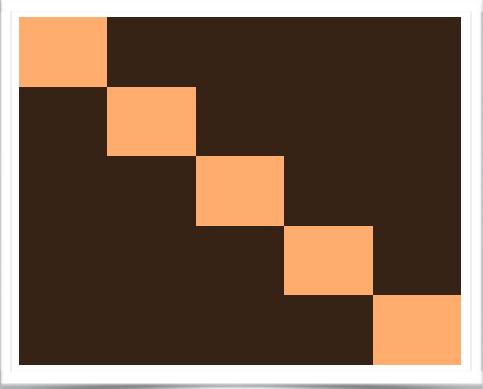
community structure:
a group of vertices that
connect to other groups in
similar ways



community structure

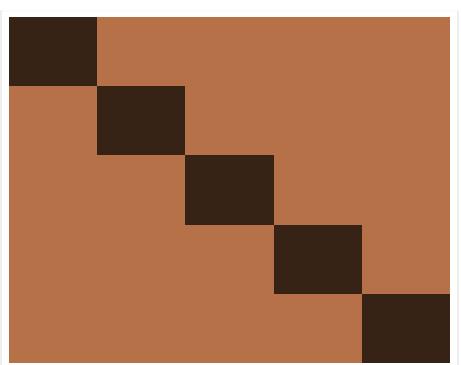
assortative

edges within groups



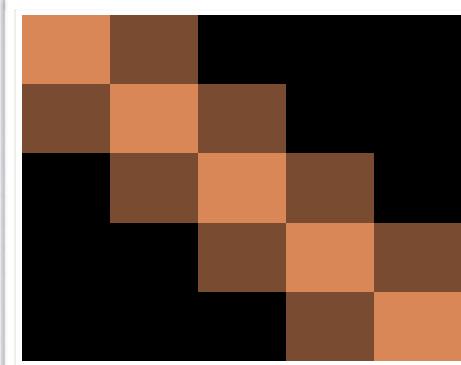
disassortative

edges between groups



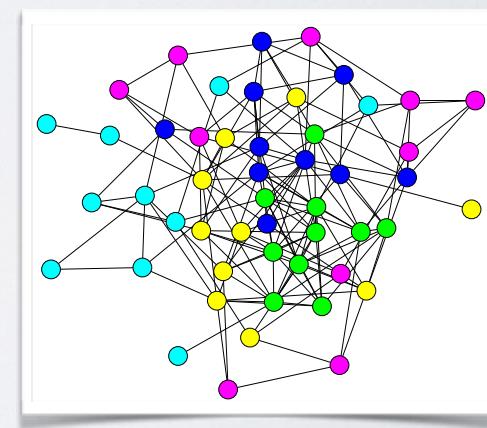
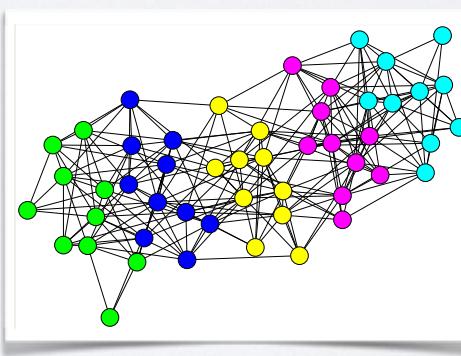
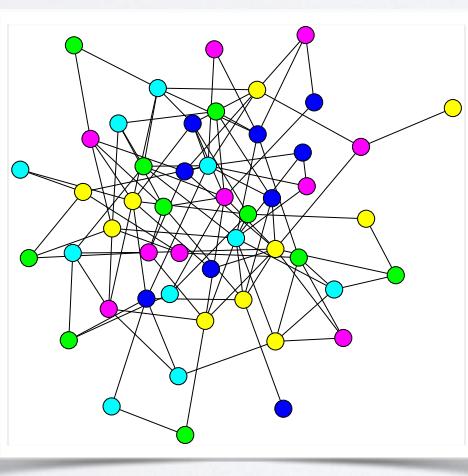
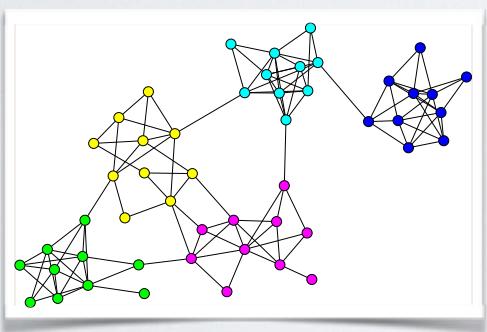
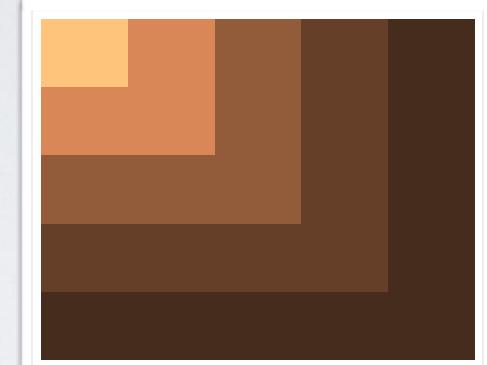
ordered

linear group hierarchy



core-periphery

dense core, sparse periphery



community structure

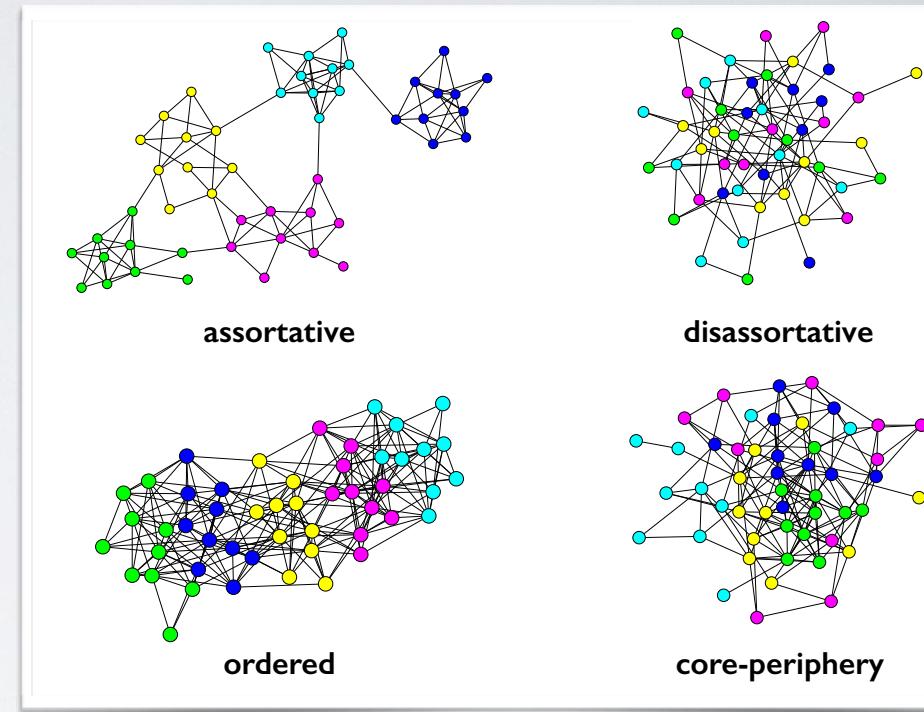
- enormous interest, especially since 2000
- dozens of algorithms for extracting various large-scale patterns
- hundreds of papers published
- spanning Physics, Computer Science, Statistics, Biology, Sociology, and more
- this was one of the first:

Community structure in social and biological networks

M. Girvan*†‡ and M. E. J. Newman*§

PNAS 2002

5700+ citations on Google Scholar

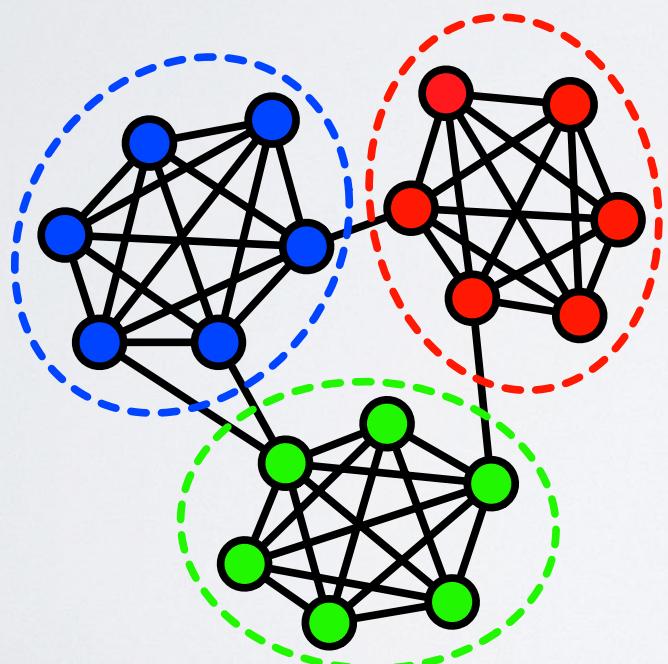


network communities

THE STRENGTH OF WEAK TIES: A NETWORK THEORY REVISITED

1983

Mark Granovetter



most new job opportunities from
“weak ties”

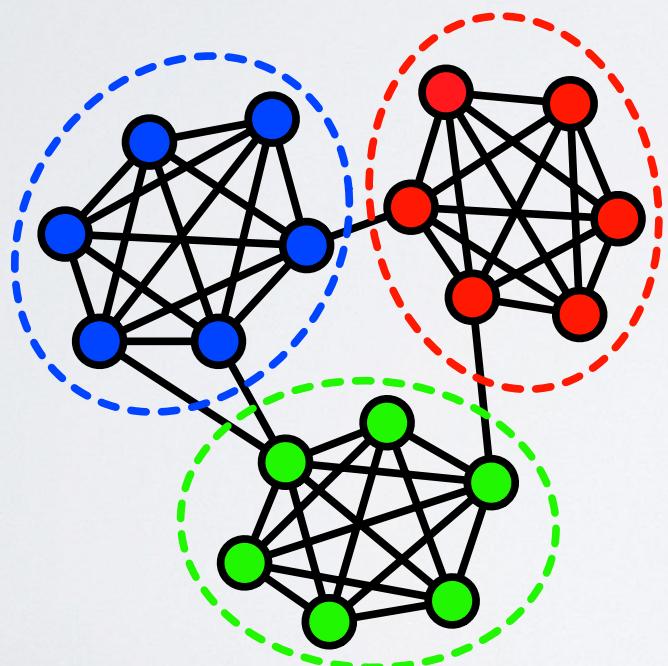
- within-community links = strong
- bridge links = weak

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

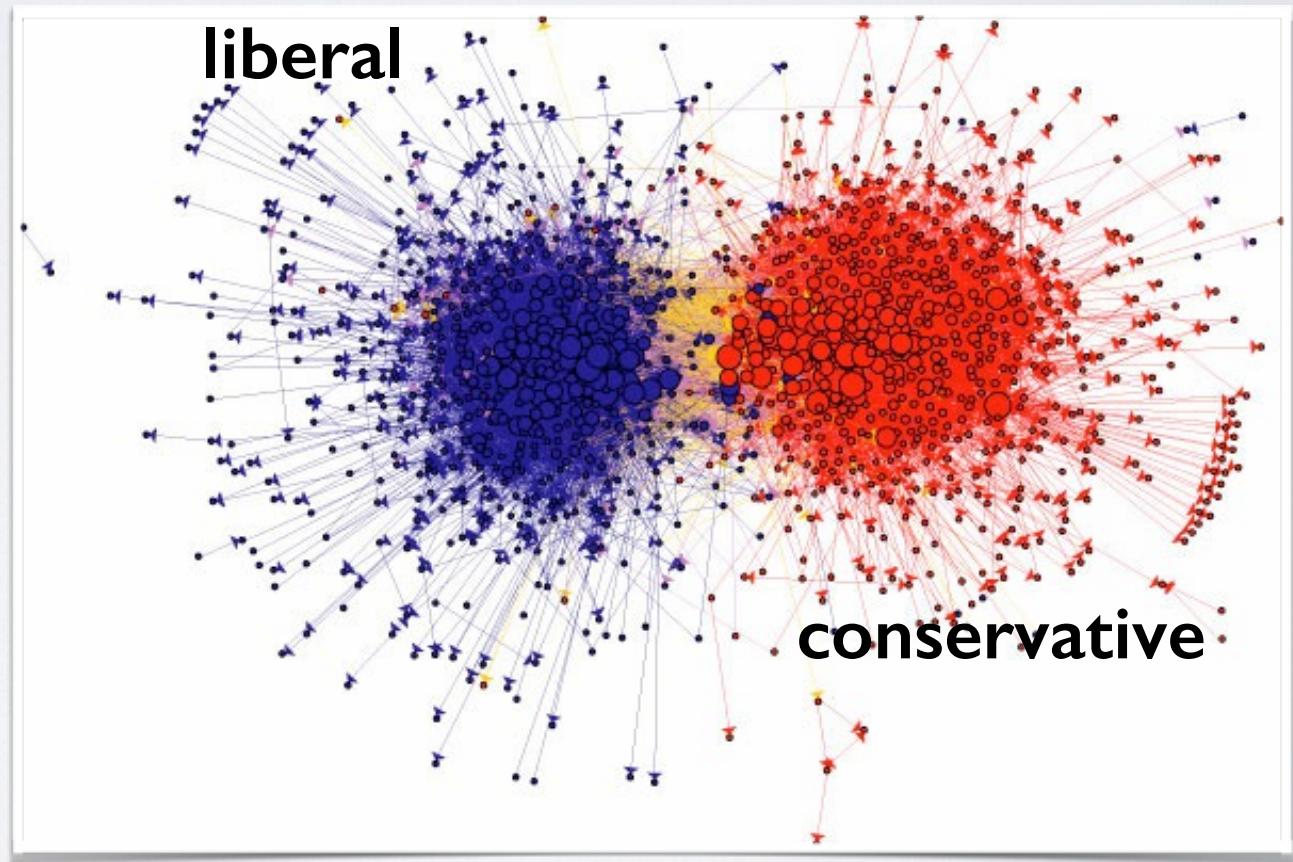
The Political Blogosphere and the 2004 U.S. Election: Divided They Blog

Lada Adamic

Natalie Glance

2004

1494 blogs
759 liberal
735 conservative



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 assortativity r on those groups

$n = 409,687$ items

$m = 2,464,630$ edges

Hello, Aaron J Clauset. We have recommendations for you. (Not Aaron?)
Aaron's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Shop All Departments | Search Books

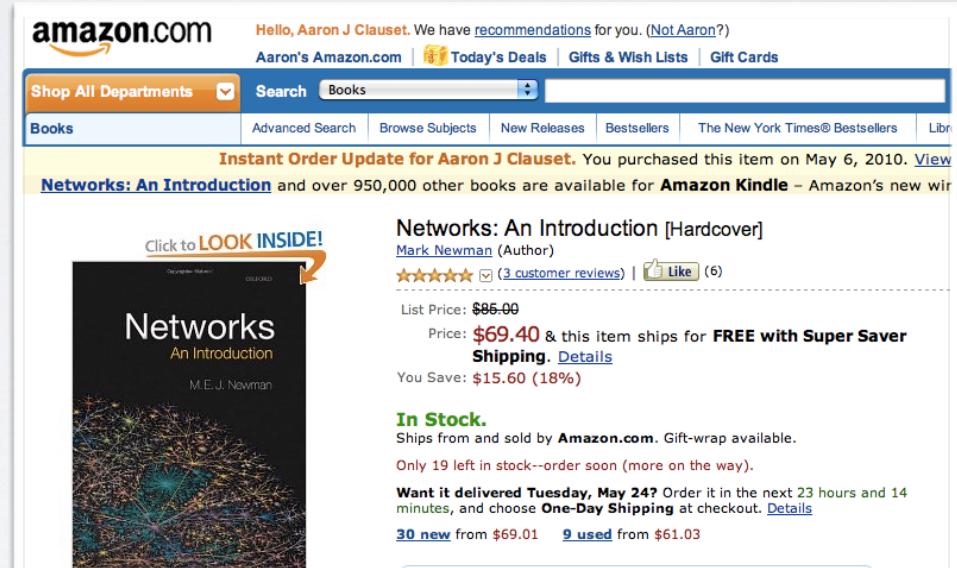
Books Advanced Search Browse Subjects New Releases Bestsellers The New York Times® Bestsellers Library

Instant Order Update for Aaron J Clauset. You purchased this item on May 6, 2010. View Networks: An Introduction and over 950,000 other books are available for Amazon Kindle – Amazon's new wireless reading device.

Networks: An Introduction [Hardcover]
Mark Newman (Author)
★★★★★ (3 customer reviews) | Like (6)

List Price: \$85.00
Price: \$69.40 & this item ships for FREE with Super Saver Shipping. Details
You Save: \$15.60 (18%)

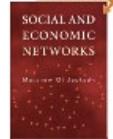
In Stock.
Ships from and sold by Amazon.com. Gift-wrap available.
Only 19 left in stock--order soon (more on the way).
Want it delivered Tuesday, May 24? Order it in the next 23 hours and 14 minutes, and choose One-Day Shipping at checkout. Details
30 new from \$69.01 9 used from \$61.03



Customers Who Bought This Item Also Bought

Page 1 of 20



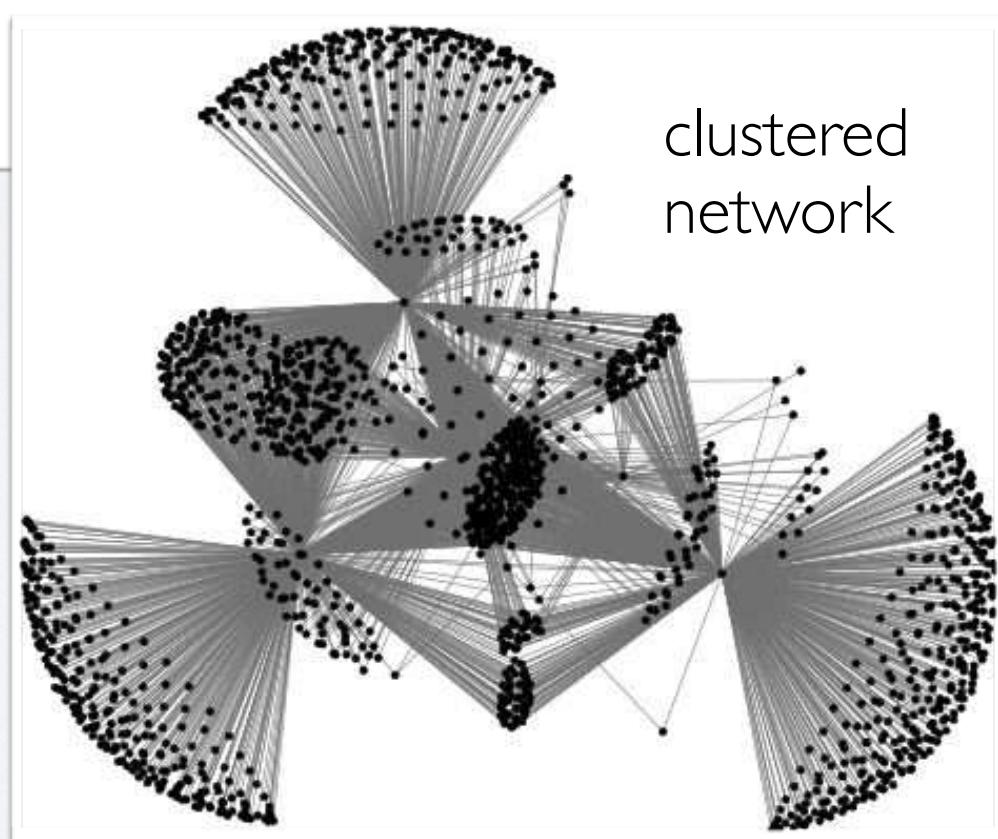
 Networks, Crowds, and Markets: Reasoning About a... by David Easley ★★★★★ (3) \$41.47	 Dynamical Processes on Complex Networks by Alain Barrat ★★★★★ (3) \$71.51	 Social Network Analysis: Methods and Applica... by Stanley Wasserman ★★★★★ (9) \$44.98	 Simply Complexity: A Clear Guide to Complexity Th... by Neil Johnson ★★★★★ (8) \$9.81	 Social and Economic Networks by Matthew O. Jackson ★★★★★ (2) \$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

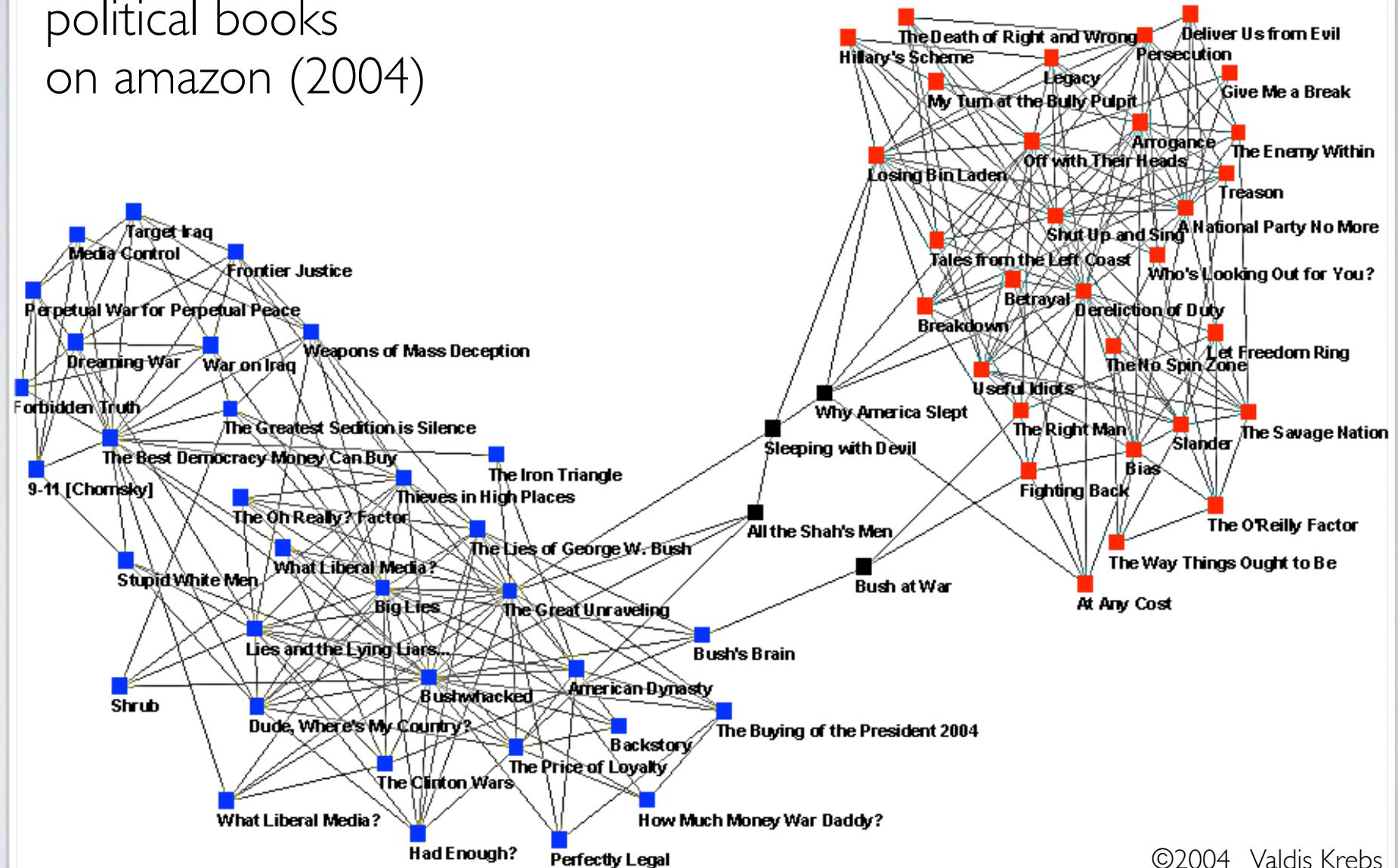
purchases = interests

interests = clustered



network communities

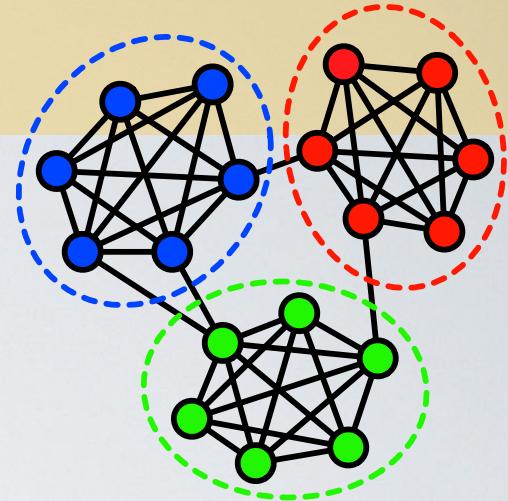
political books
on amazon (2004)



©2004 Valdis Krebs

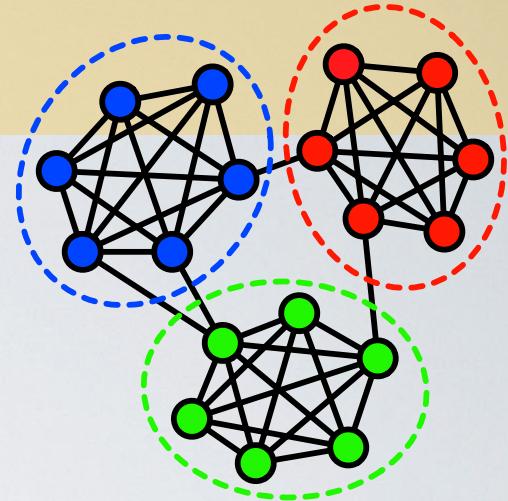
network communities

- community = vertices with same pattern of inter-community connections
- network macro-structure
- finding them like “network clustering”
- 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”
- allow us to coarse grain system structure
[decompose heterogeneous structure into homogeneous blocks]
- constrains network synchronization, information flows, diffusion, influence

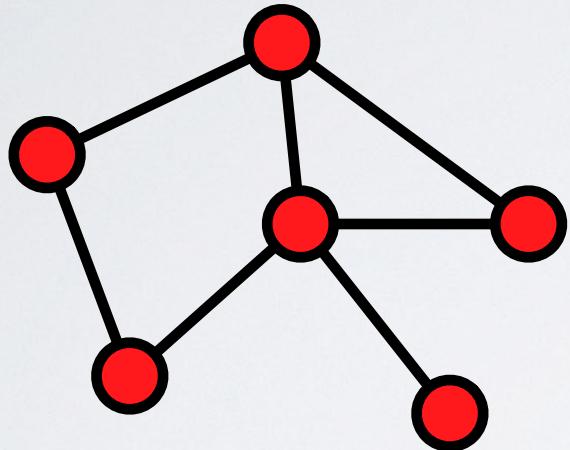


open questions:

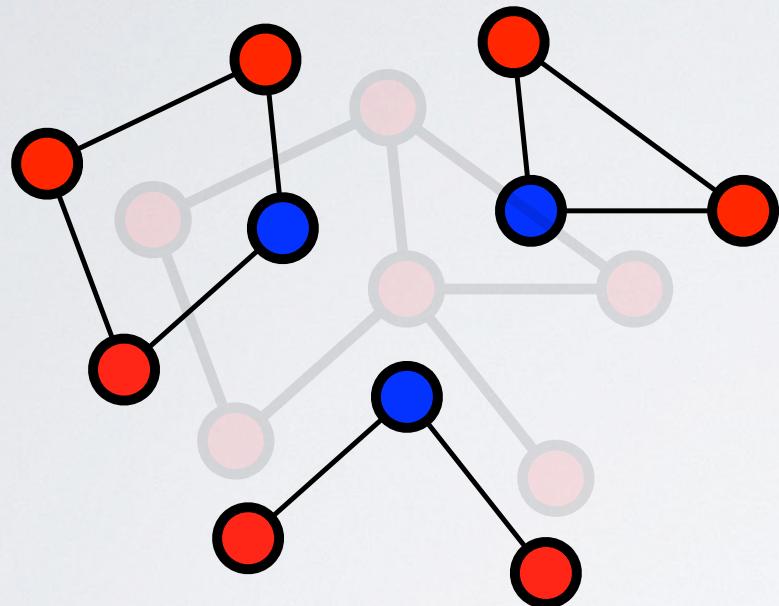
- what processes generate communities?
- what impact on dynamics? network function?

describing networks

motifs



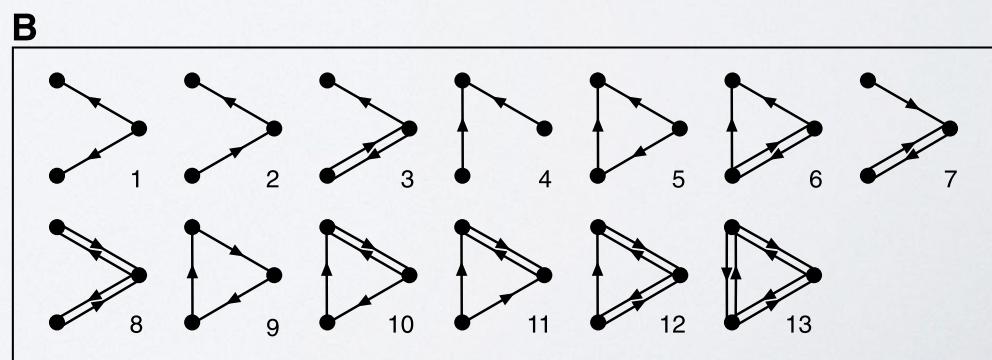
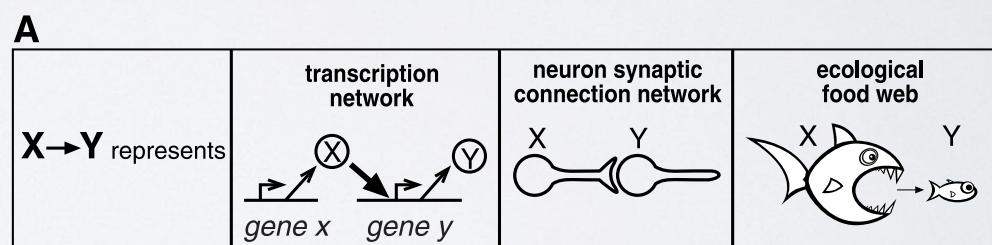
describing networks



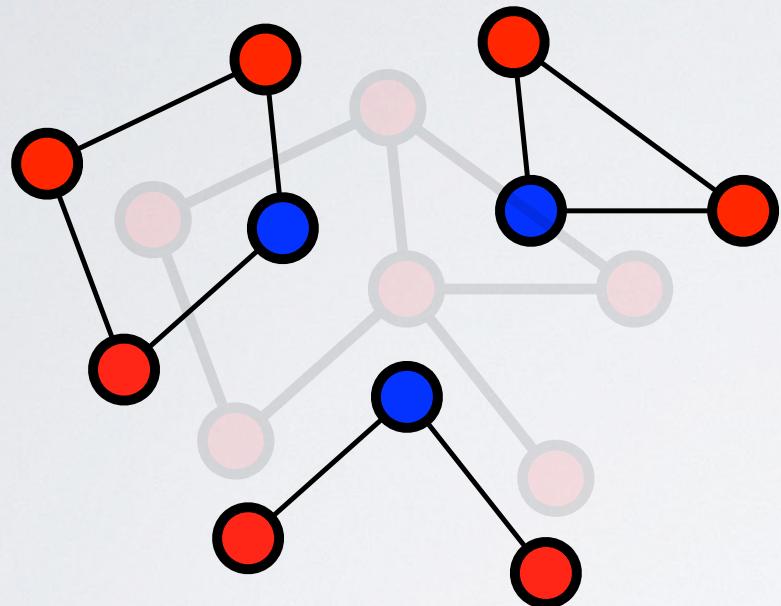
motifs:

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

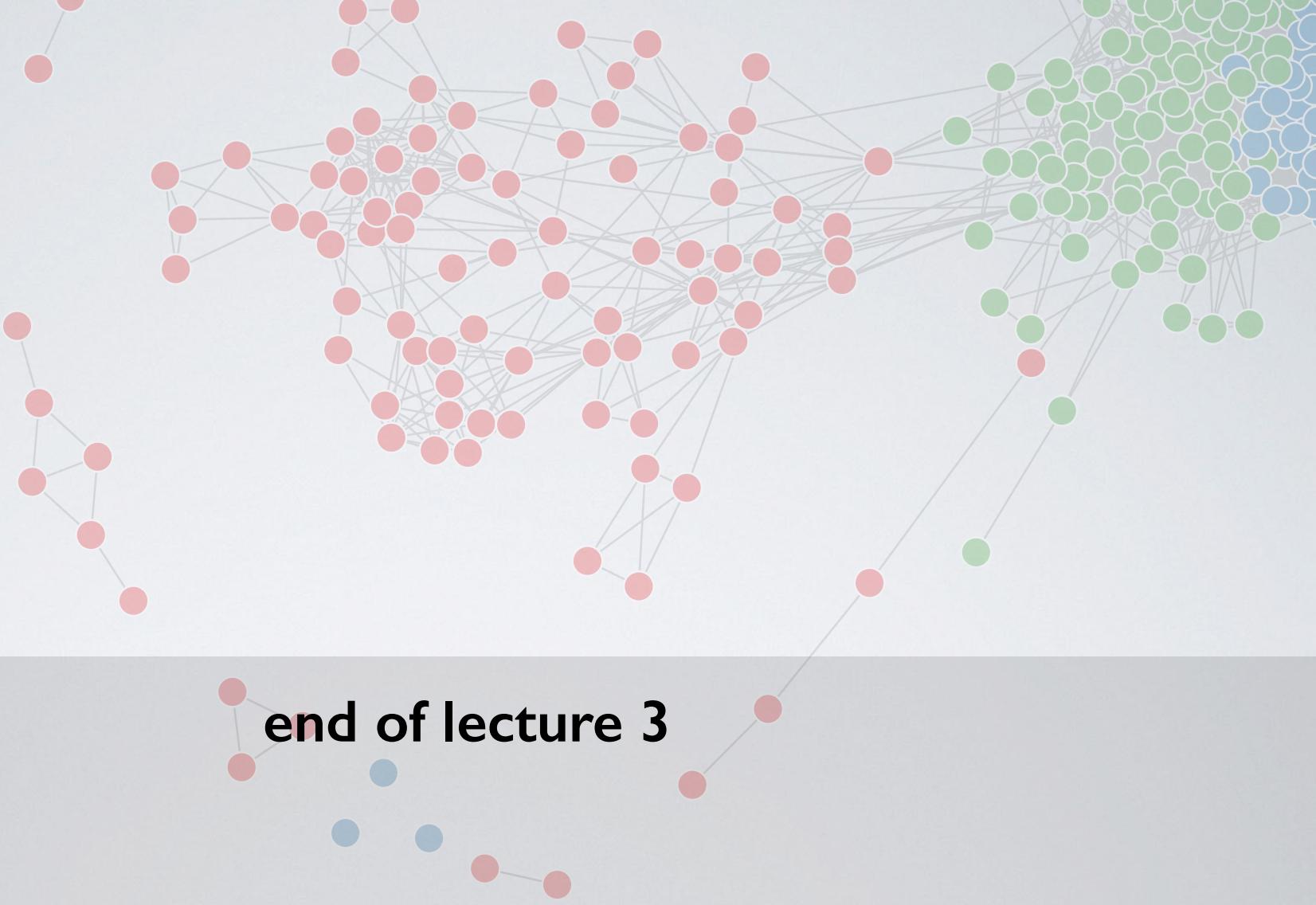
compare counts against null
model (random graph model)

- efficient counting is tricky
(combinatorics + graph isomorphism)
- choice of null model key
(see Lecture 2)
- 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)



selected references

- The structure and function of complex networks. M. E. J. Newman, *SIAM Review* **45**, 167–256 (2003).
- *The Structure and Dynamics of Networks*. M. E. J. Newman, A.-L. Barabási, and D. J. Watts, Princeton University Press (2006).
- Hierarchical structure and the prediction of missing links in networks. A. Clauset, C. Moore, and M. E. J. Newman, *Nature* **453**, 98–101 (2008).
- Modularity and community structure in networks. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
- Why social networks are different from other types of networks. M. E. J. Newman and J. Park, *Phys. Rev. E* **68**, 036122 (2003)
- Random graphs with arbitrary degree distributions and their applications. M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- Comparing community structure identification. L. Danon, A. Diaz-Guilera, J. Duch and A. Arenas. *J. Stat. Mech.* P09008 (2005).
- Characterization of Complex Networks: A Survey of measurements. L. daF. Costa, F. A. Rodrigues, G. Travieso and P. R. VillasBoas. arxiv:cond-mat/050585 (2005).
- Evolution in Networks. S.N. Dorogovtsev and J. F. F. Mendes. *Adv. Phys.* **51**, 1079 (2002).
- Revisiting “scale-free” networks. E. F. Keller. *BioEssays* **27**, 1060-1068 (2005).
- Currency metabolites and network representations of metabolism. P. Holme and M. Huss. arxiv:0806.2763 (2008).
- Functional cartography of complex metabolic networks. R. Guimera and L. A. N. Amaral. *Nature* **433**, 895 (2005).
- Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. J. Leskovec, J. Kleinberg and C. Faloutsos. *Proc. 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* 2005.
- The Structure of the Web. J. Kleinberg and S. Lawrence. *Science* **294**, 1849 (2001).
- Navigation in a Small World. J. Kleinberg. *Nature* **406** (2000), 845.
- Towards a Theory of Scale-Free Graphs: Definitions, Properties and Implications. L. Li, D. Alderson, J. Doyle, and W. Willinger. *Internet Mathematics* **2**(4), 2006.
- A First-Principles Approach to Understanding the Internet’s Router-Level Topology. L. Li, D. Alderson, W. Willinger, and J. Doyle. *ACM SIGCOMM* 2004.
- Inferring network mechanisms: The *Drosophila melanogaster* protein interaction network. M. Middendorf, E. Ziv and C. H. Wiggins. *Proc. Natl. Acad. Sci. USA* **102**, 3192 (2005).
- Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. S. Ciliberti, O. C. Martin and A. Wagner. *PLoS Comp. Bio.* **3**, e15 (2007).
- Simple rules yield complex food webs. R. J. Williams and N. D. Martinez. *Nature* **404**, 180 (2000).
- A network analysis of committees in the U.S. House of Representatives. M. A. Porter, P. J. Mucha, M. E. J. Newman and C. M. Warmbrand. *Proc. Natl. Acad. Sci. USA* **102**, 7057 (2005).
- On the Robustness of Centrality Measures under Conditions of Imperfect Data. S. P. Borgatti, K. M. Carley and D. Krackhardt. *Social Networks* **28**, 124 (2006).