



# **Networks, yeah!**

## The representation of relations

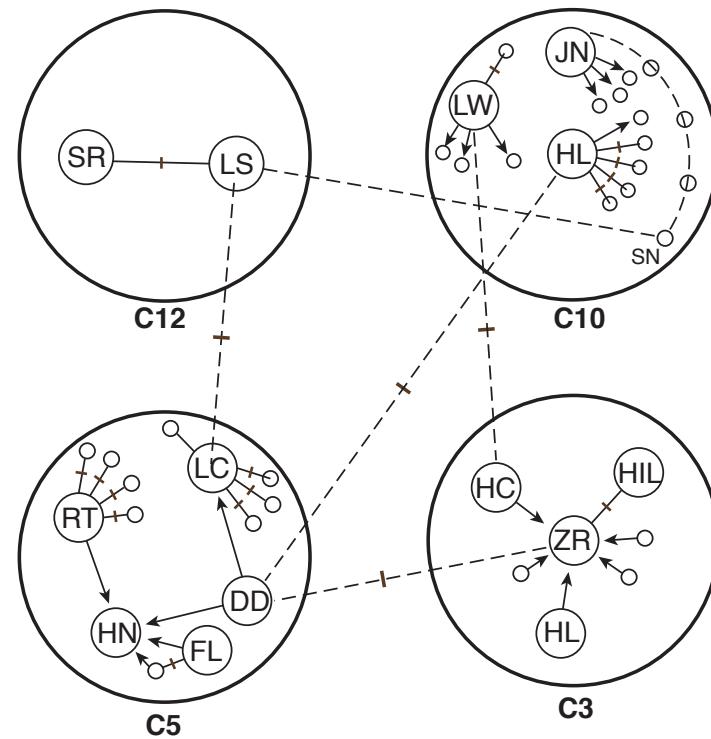
Momin M. Malik <[momin\\_malik@cyber.harvard.edu](mailto:momin_malik@cyber.harvard.edu)>  
Data Science Postdoctoral Fellow, Berkman Klein Center

Data & Donuts, DigitalHKS, Harvard Kennedy School, November 2, 2018

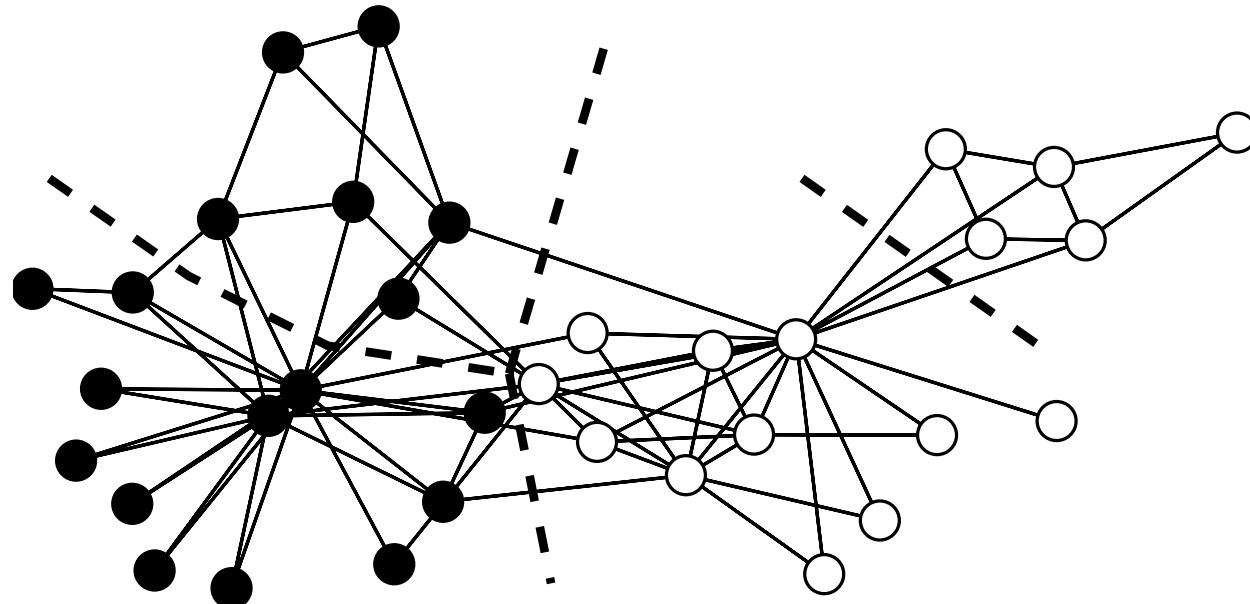
# Success stories

1. → Success stories
2. Principles
3. Problems

# Runaways from “reformatory” for delinquent girls



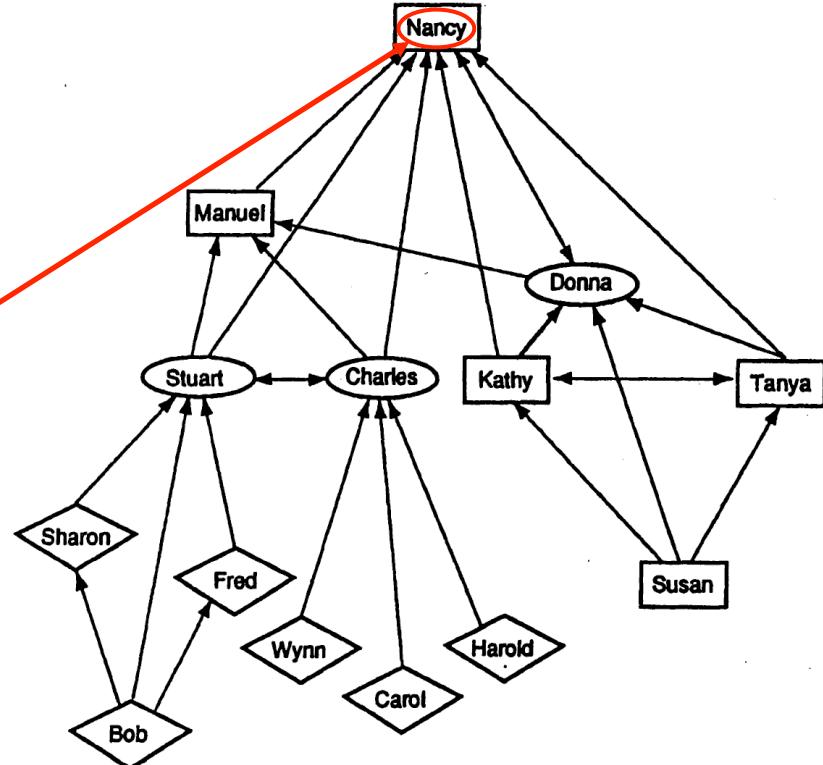
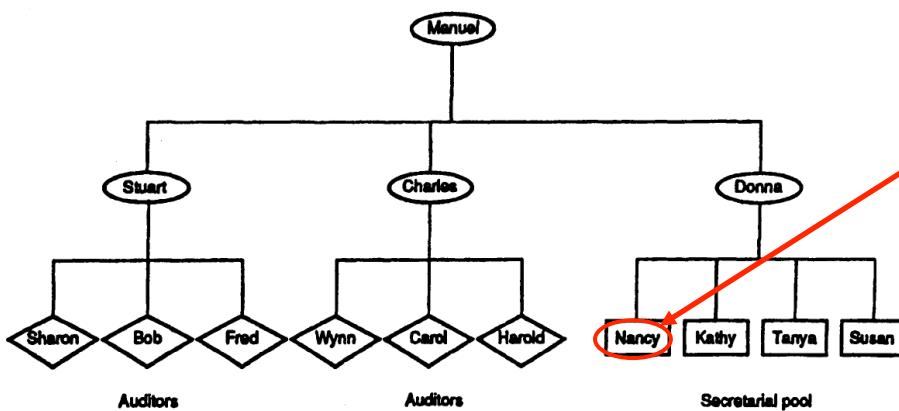
# Competing factions in a karate club



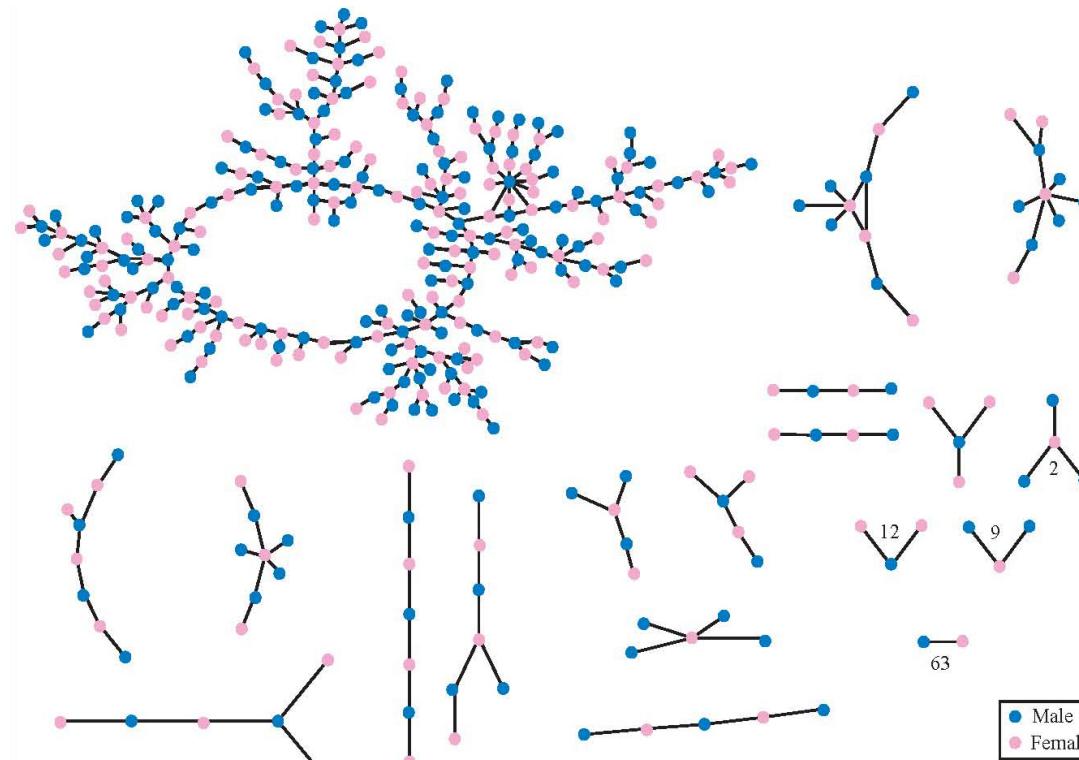
Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4), 452-473.

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. *Notices of the AMS*, 56(9), 1082-1166.

# Org hierarchy versus informal network



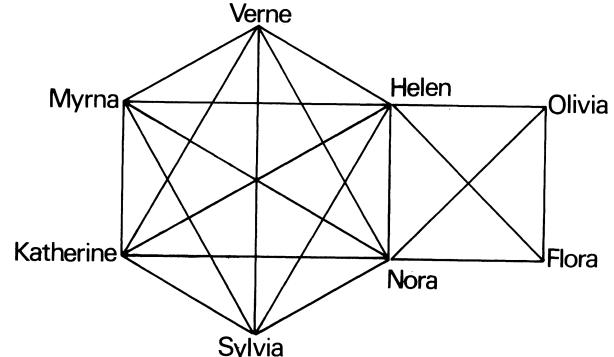
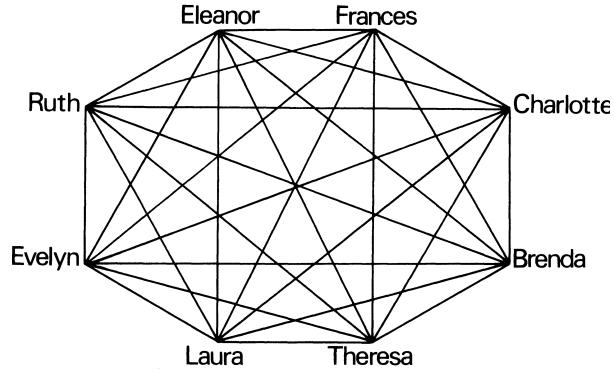
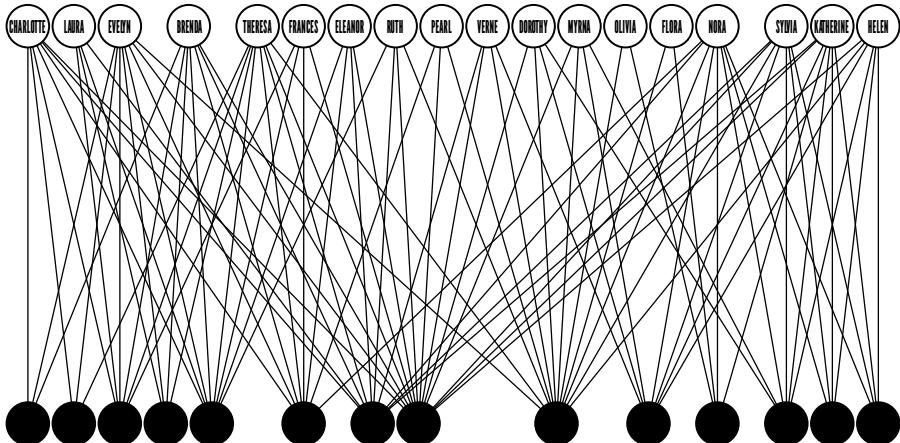
# Not dating a ex's ex in a high school



Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1), 44-91.  
Marcum, C. S., Lin, J., & Koehly, L. (2016). Growing-up and coming-out: Are 4-cycles present in adult hetero/gay hook-ups? *Network Science*, 4(3), 400-405.



# Shared events among southern debutants



Breiger, R. (1974). The duality of persons and groups. *Social Forces*, 53, 181-190.

Davis, A., & Gardner, B. B. (1941). *Deep south: A social anthropological study of caste and class*. Chicago: University of Chicago Press.

# Same principles underlies recommendations

Customers who bought this item also bought

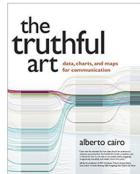
Page 1 of 13



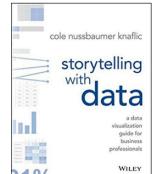
Observe, Collect, Draw!: A Visual Journal  
› Giorgia Lupi  
Diary  
\$12.76



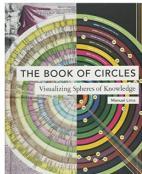
Dear Data Postcard Kit: For Two Friends to Draw and Share  
Giorgia Lupi  
Card Book  
\$16.95



The Truthful Art: Data, Charts, and Maps for Communication  
› Alberto Cairo  
 50 Paperback  
\$44.47



Storytelling with Data: A Data Visualization Guide for Business Professionals  
› Cole Nussbaumer...  
 277 #1 Best Seller in Business Mathematics Paperback  
\$22.06



The Book of Circles: Visualizing Spheres of Knowledge  
› Manuel Lima  
 12 Hardcover  
\$34.00



#MakeoverMonday: Improving How We Visualize and Analyze Data...  
› Andy Kriebel  
 1 Paperback  
\$25.31



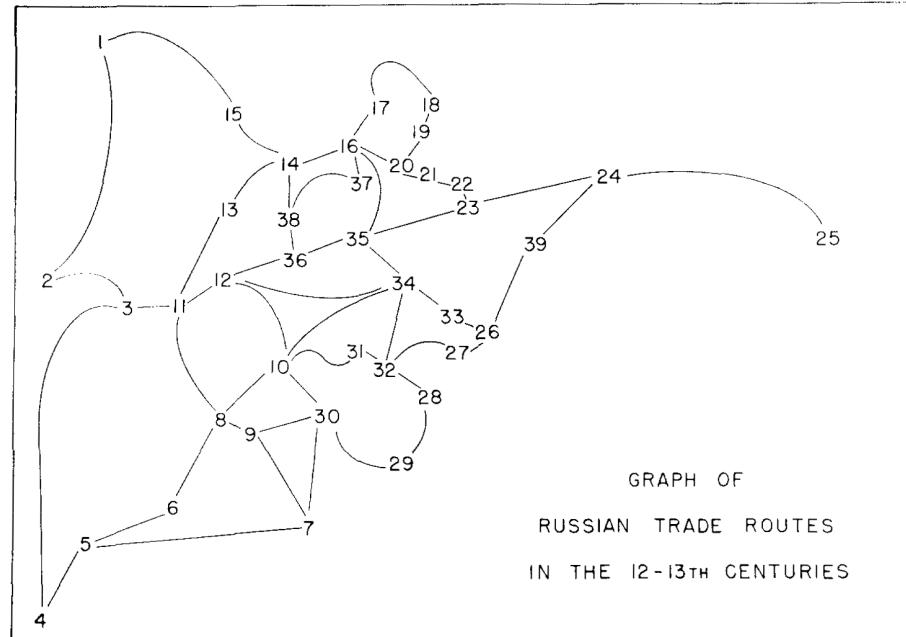
Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations...  
› Scott Berinato  
 47 Paperback  
\$20.62



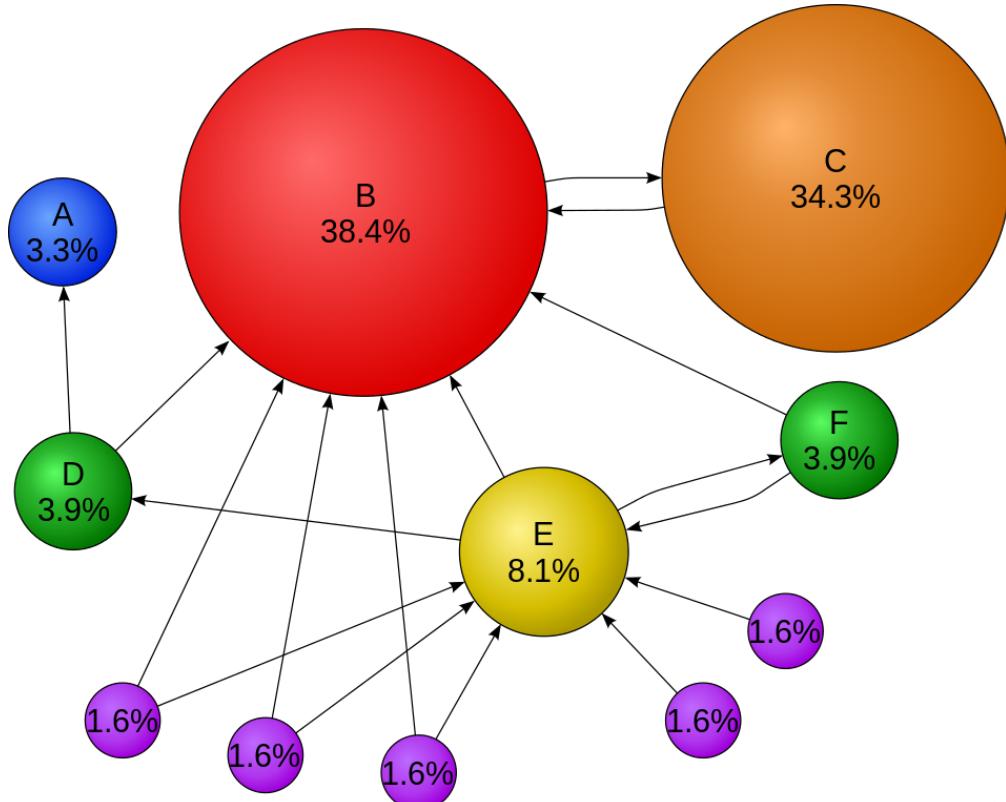
Knowledge Is Beautiful: Impossible Ideas, Invisible Patterns, Hidden Relationships...  
› David McCandless  
 38 Flexibound  
\$25.50



# Medieval Moscow at center of river trade



# Similar idea (centrality) behind Pagerank



# Principles

1. Success stories
2. → Principles
3. Problems

# A “network” is any sort of *relation*

Similarities			Social Relations					Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,	
e.g., Same spatial and temporal space	e.g., Same clubs Same events etc.	e.g., Same gender Same attitude etc.	e.g., Mother of Sibling of	e.g., Friend of Boss of Student of Competitor of	e.g., Likes Hates etc.	e.g., Knows Knows about Sees as happy etc.	e.g., Sex with Talked to Advice to Helped Harmed etc.	e.g., Information Beliefs Personnel Resources etc.	

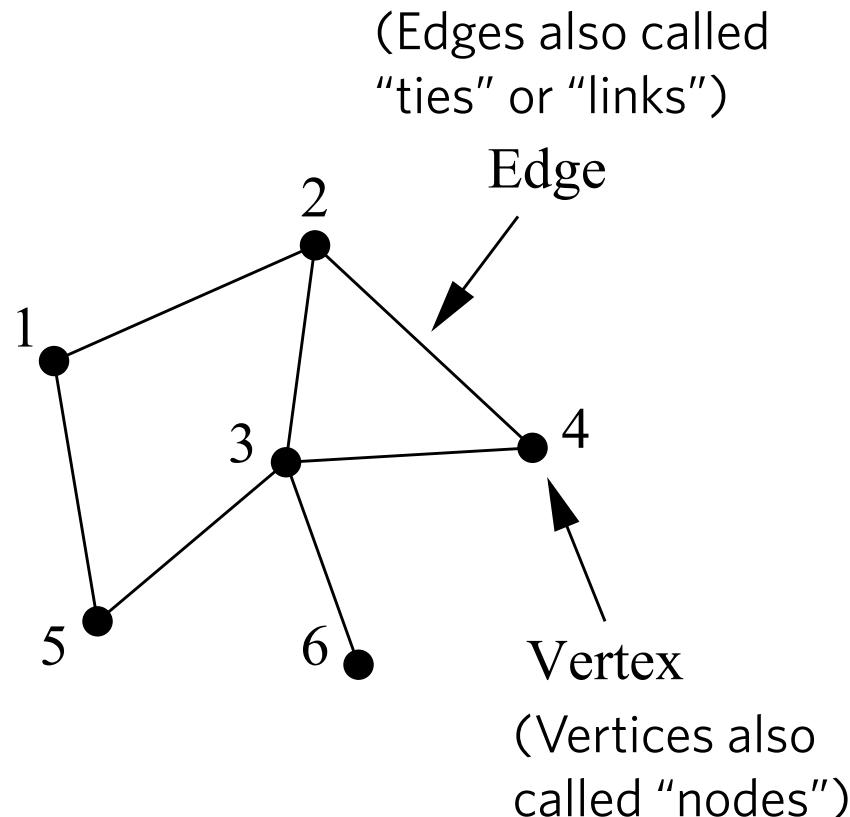
# From $n$ to $n^2$

	$Y$	$X_1$	$X_2$	$\dots$	$X_k$
$v_1$	$y_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1k}$
$v_2$	$y_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$v_n$	$y_n$	$x_{n1}$	$x_{n2}$	$\dots$	$x_{nk}$



	<i>from</i>	<i>to</i>	$Y$	$W_1$	$W_2$	$W_3$	$\dots$
$e_1$	$v_1$	$v_2$	$y_{12}$	$\mathbb{1}(x_{11} = x_{21})$	$x_{12} - x_{22}$	$x_{13}$	$\dots$
$e_2$	$v_2$	$y_3$	$y_{23}$	$\mathbb{1}(x_{11} = x_{31})$	$x_{12} - x_{32}$	$x_{13}$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$e_{n+1}$	$v_2$	$v_1$	$y_{21}$	$\mathbb{1}(x_{21} = x_{11})$	$x_{22} - x_{12}$	$x_{23}$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$e_{2 \binom{n}{2}}$	$v_{n-1}$	$v_n$	$y_{(n-1)n}$	$\mathbb{1}(x_{(n-1)1} = x_{n1})$	$x_{(n-1)2} - x_{n2}$	$x_{(n-1)3}$	$\dots$

# Key representations: Graph, adjacency matrix



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

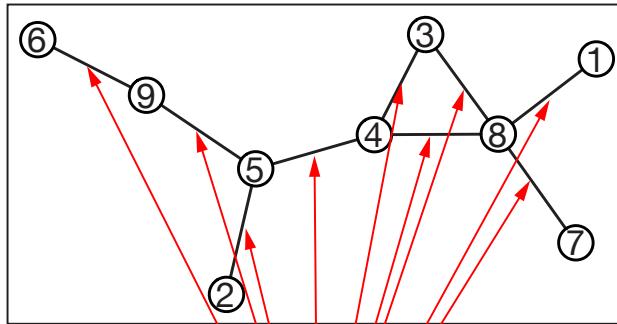
# Data collection, software

- Collection:
  - Surveys
  - Social media data
  - Communications/log data
  - Get into an *edgelist* (only with existing ties)
  - Directed/undirected, weighted/unweighted, bipartite\* (two node types, e.g., affiliation networks of people and organizations, where ties represent membership)
  - Other relevant terms: self-loops, multi-edges, “multiplex” networks (multiple edge types), “multi-modal” networks (multiple node types, generalization of bipartite)
- Recommended software
  - Gephi, overall (GUI-based)
  - (Cytoscape?)
  - Most powerful GUI-based: Pajek
  - R: igraph
  - Python: networkx
  - C++: SNAP

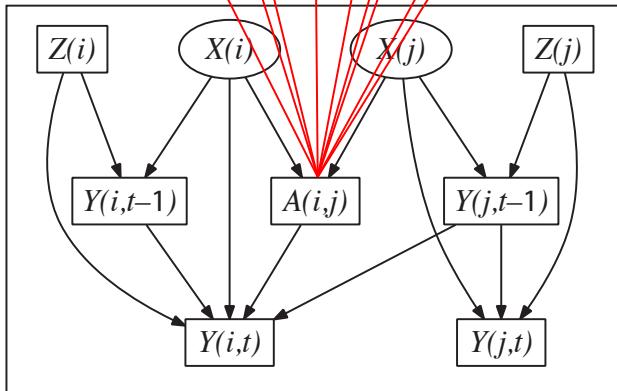
\*“Bipartite” means you can partition the network into two sets of nodes, where nodes in set 1 only link to nodes in set 2 and vice versa, regardless of what those nodes are. “Two-mode” means there are two node types, regardless of how they connect (maybe one organization can be a member of other organizations). Usually an affiliation network will be both two-mode and bipartite, so they get used interchangeably.

# Clarification 1: Graphs are overloaded

Network model  
(observations)

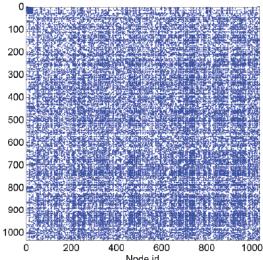


Graphical model  
(variables), e.g.  
representing  
causality

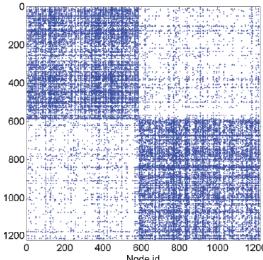


Note: This is also a Markov chain, as it has a tie from  $Y(t-1)$  to  $Y(t)$ . Markov chains are themselves graphs, but also have *transition* graphs, which is yet another type of graph.

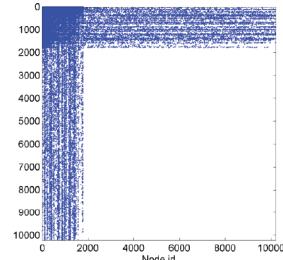
# Clarification 2: Networks are similarities (e.g., can use heatmaps), but similarities are not networks



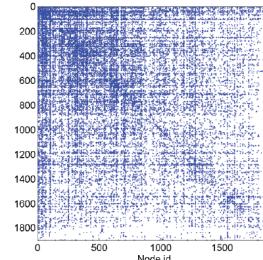
(a) facebook107



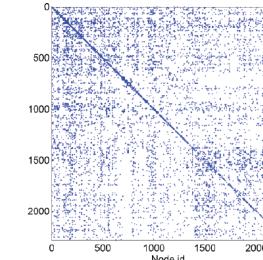
(b) polblogs



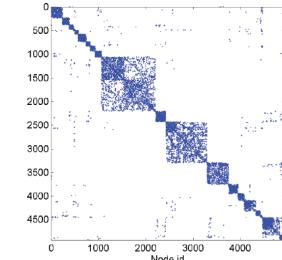
(c) USairport



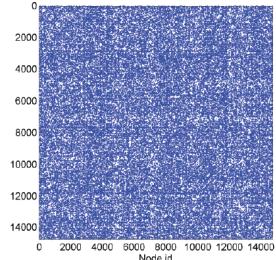
(d) UC Irvine



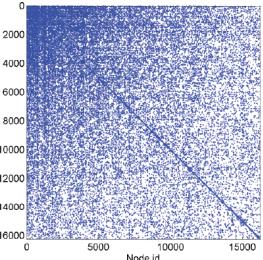
(e) yeast



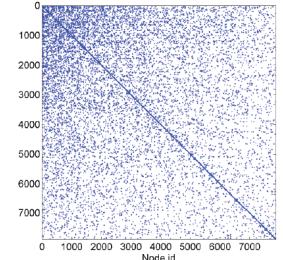
(f) USpower



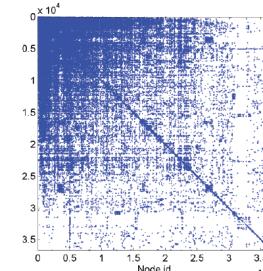
(g) IMDB



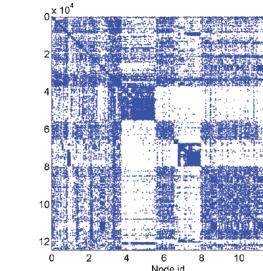
(h) cond-mat1



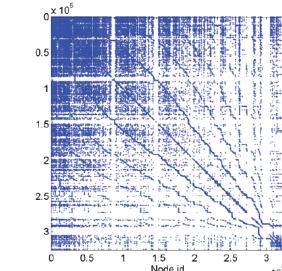
(i) cond-mat2



(j) enron



(k) internet



(l) www

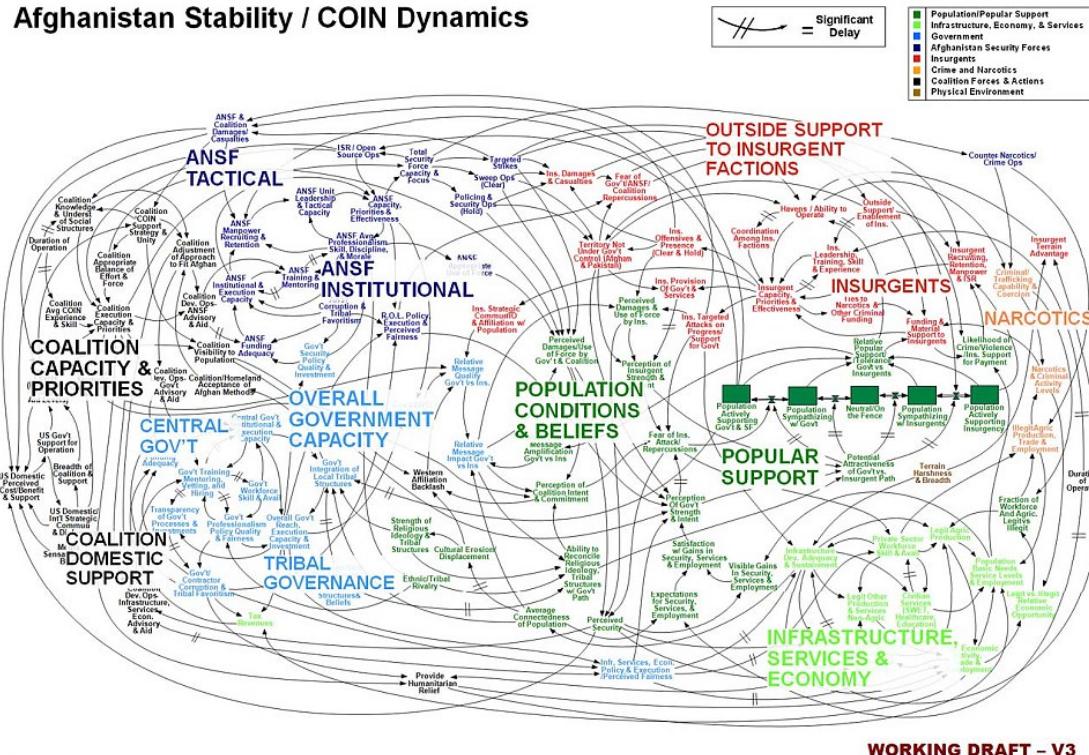
# Theoretical benefits?

- Networks, in representing relations, have been said to escape reductionism
- Networks a hallmark of “complex systems”
- But long tradition of simply displacing the endpoint of reduction from individuals to network processes (seeing people as atoms, and find regular rules determining how they configure)
- Networks lead to lots of exciting math, and are a powerful *formal* representation, but can also be used qualitatively, and theoretically (e.g., nodes emerge out of the network, rather than vice versa)
- *Statistics* on networks: Extremely difficult.

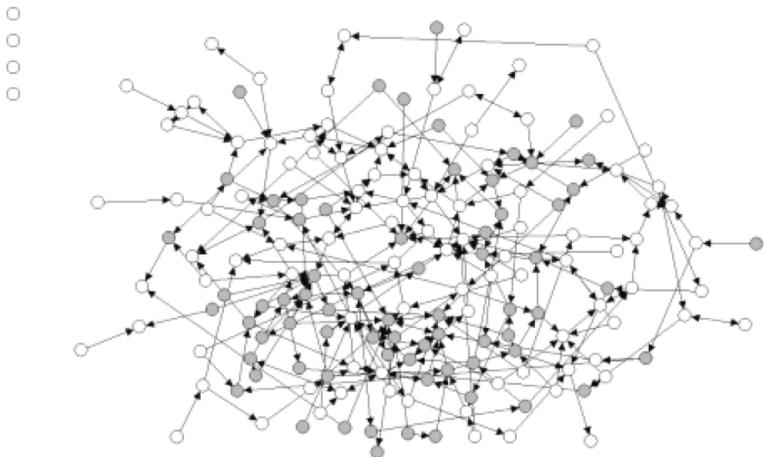
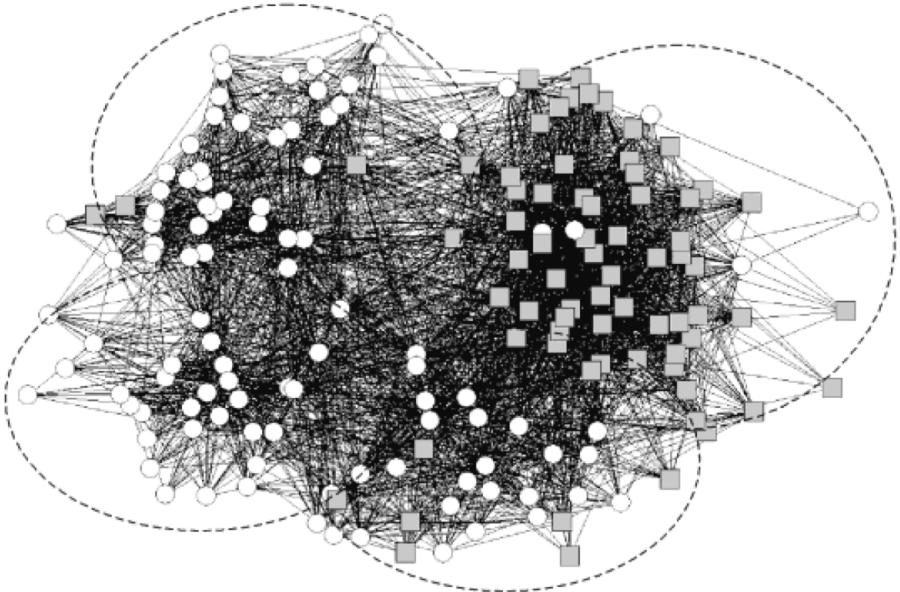
# Problems

1. Success stories
2. Principles
3. → Problems

# Problems: Bad networks

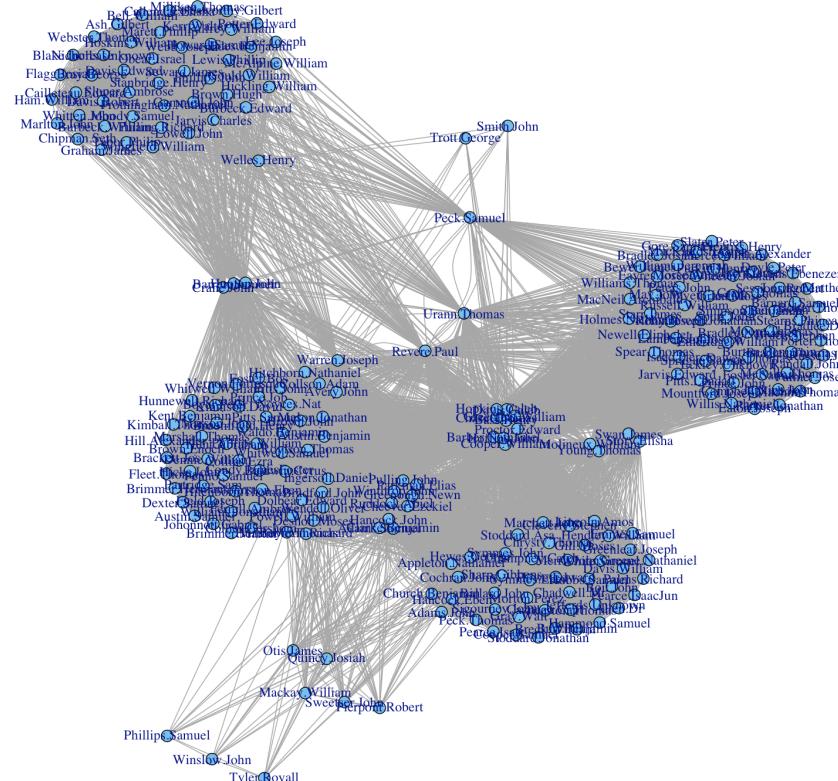


# Problems: “Spaghetti”

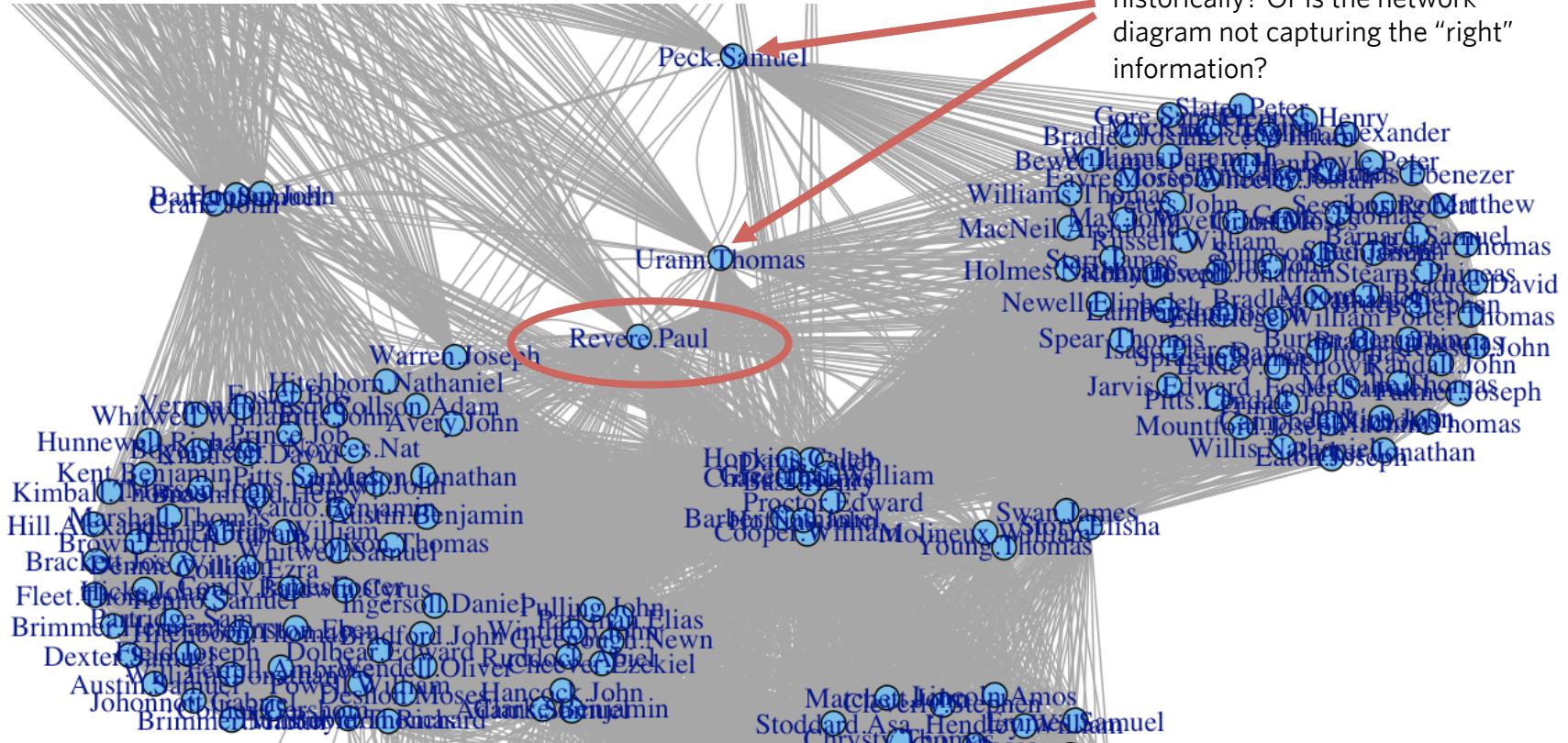


Durland, M. M. (2005). A formative evaluation of the integration of two departments. *New Directions for Evaluation*, 107, 81-94.

# Problems: Retrospective, just-so storytelling



# Problems: Retrospective, just-so storytelling

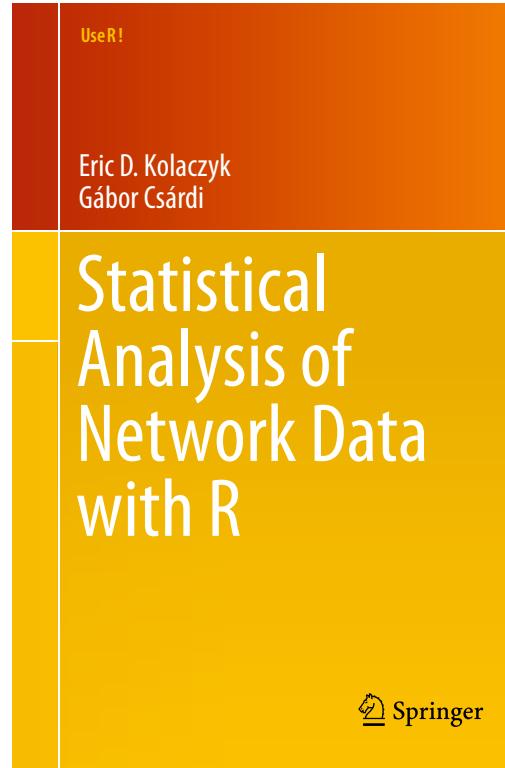


# The top three books I'd recommend for more

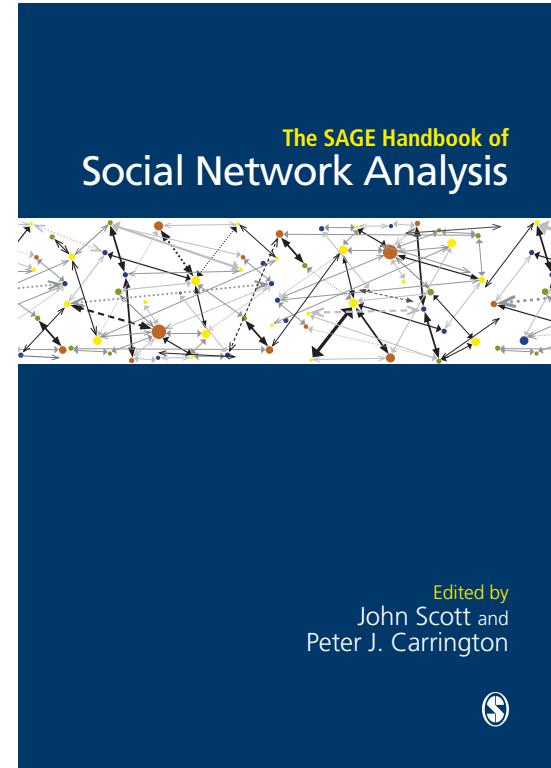
Robins, G. (2015). *Doing social network research: Network based research design for social scientists*. SAGE. Best overview of network concepts, terms, theory.



Csárdi, G., & Kolaczyk, E. D. (2014). *Statistical analysis of network data with R*. Springer. Best guide for analyzing network data.



Scott, J., & Carrington, P. J. The SAGE handbook of social network analysis. SAGE. Best overviews of dozens of areas of application.



# Thank you!

momin\_malik@cyber.harvard.edu