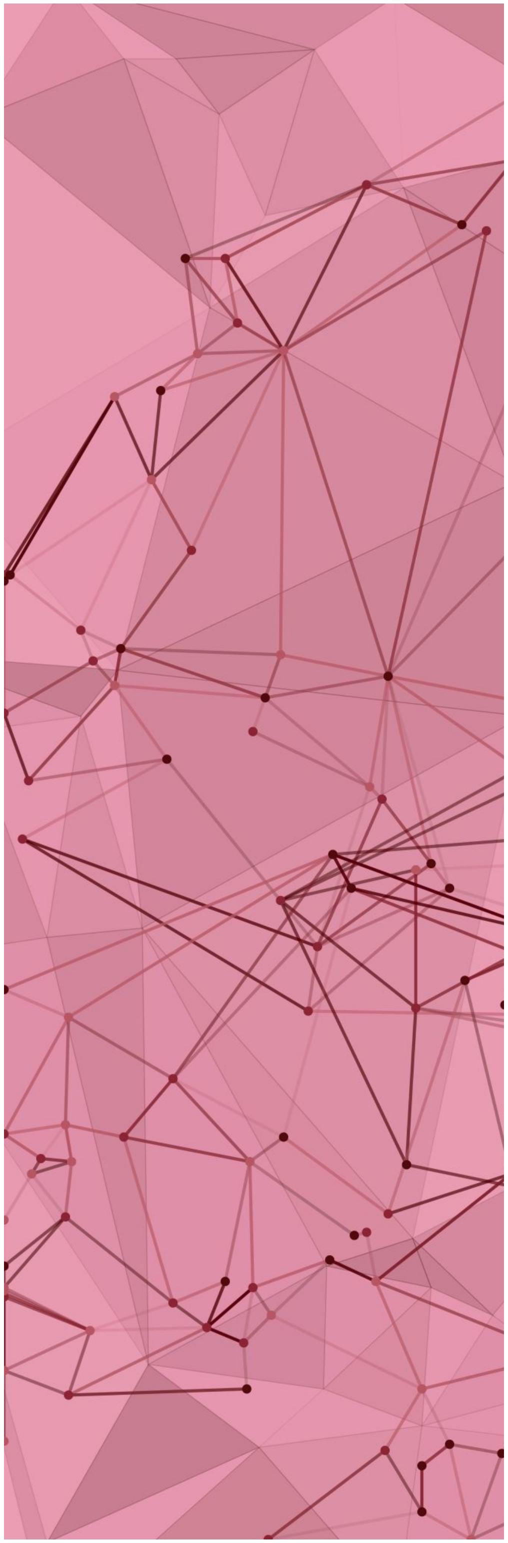
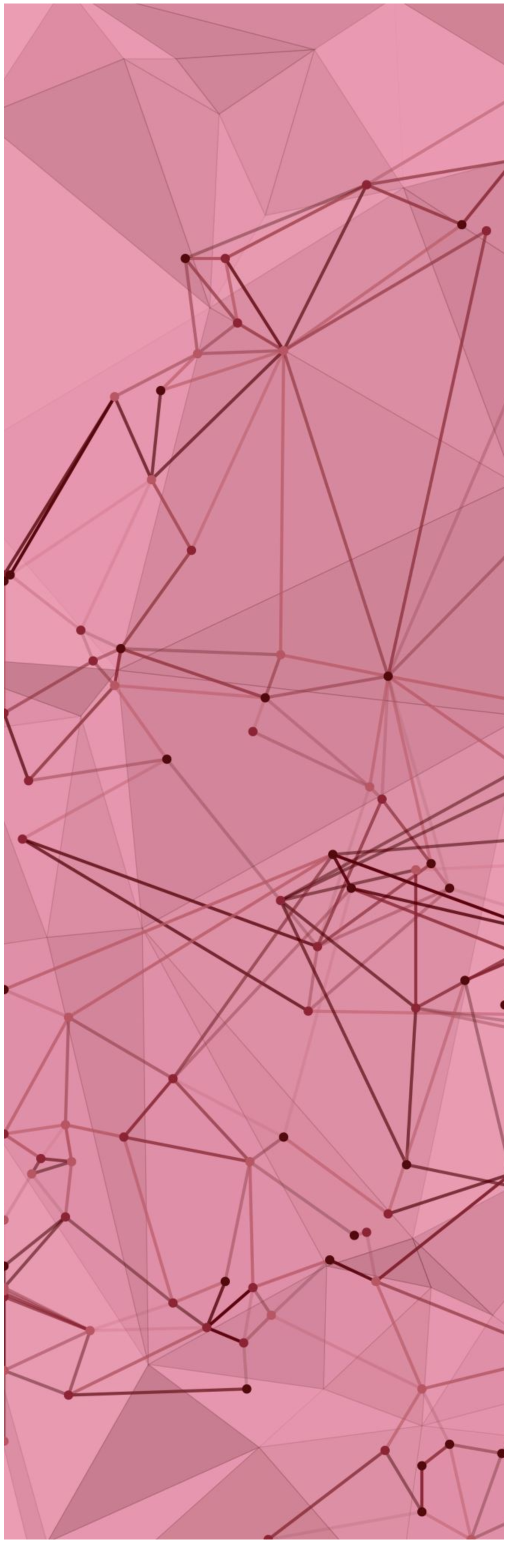


DCS/CSCI 2350
Social & Economic Networks

How can we analyze a real-world network? What are some common properties of networks?

Reading: Ch 2 of EK, Ch 2 & 3 of Jackson

Mohammad T. Irfan
https://mtirfan.com/DCS-2350



Social Network Analysis (SNA)

1. Macro-level (network-level)
2. Micro-level (node-level)

Macro-Level Analysis

1. Giant component
2. Small-world
3. Degree distribution
4. Clustering

1. Giant component

- Intuitive example– world acquaintance network
- Questions
  - Is it connected?
  - How many giant components are there?

Examples

Actor network

- Edge between two actors iff they appear together in a movie
- 98% of 449,913 actors belong to the giant component (IMDB, May 2000)


facebook
December 2010

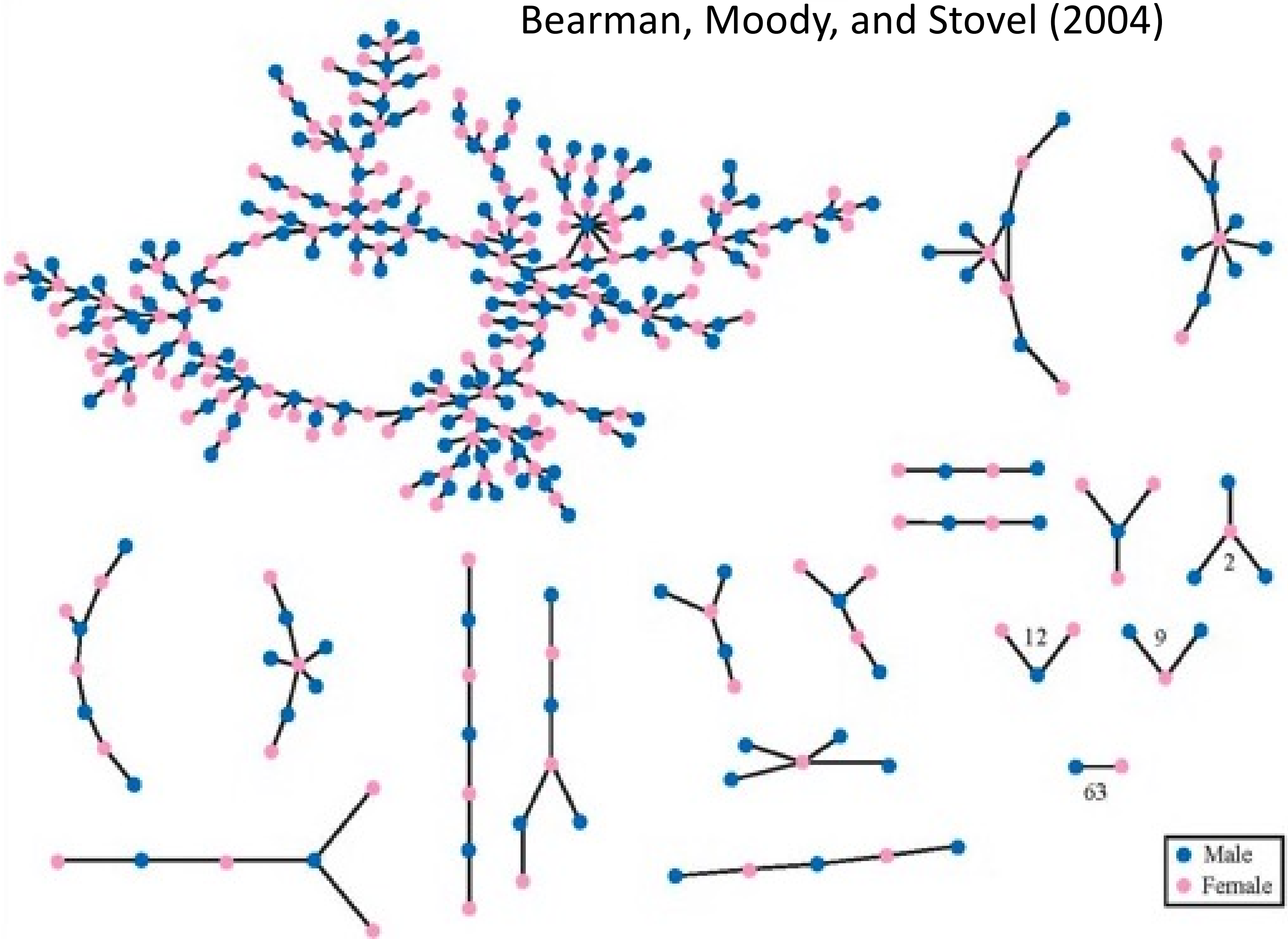
More examples

- Instant messaging
  - Microsoft IM: one giant component in a network of 240 million users (2008)
- Co-author network
- Email
- Biological networks (neural networks)
- Technology networks (power grid)
- The Internet (web of links)

Can you think of a network that doesn’t have any giant component?

Giant component: implications

High school relationships (1993-95)
Bearman, Moody, and Stovel (2004)



2. Small-world property

Also known as distance

- Proposition
  - The average shortest path length between any two nodes in a connected component is “small”
- Intuition

Six degrees of separation

John Guare's play (1990) & later movie

http://www.youtube.com/watch?v=HLIyuYwbVnA

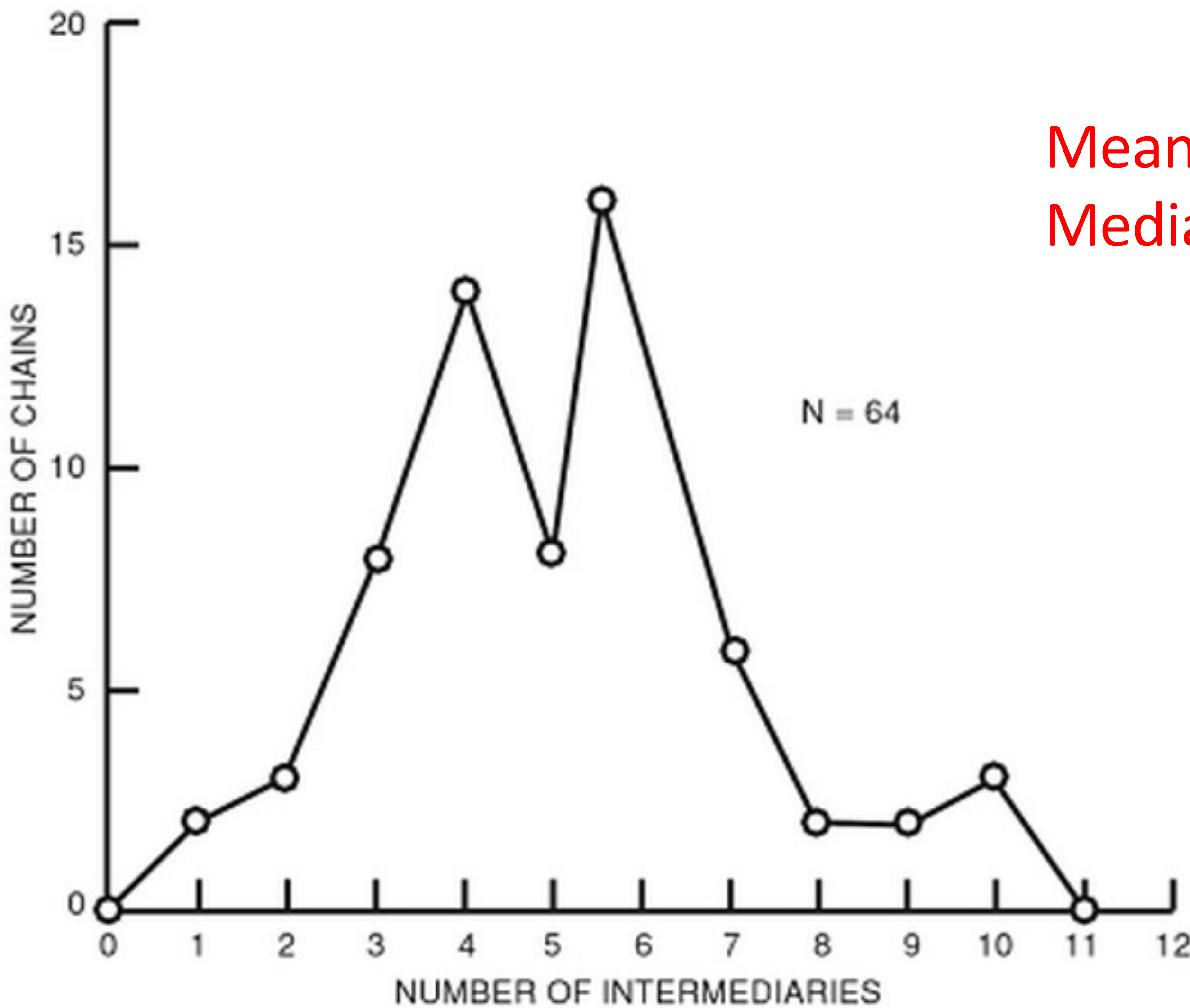
Six degrees of separation

Hungarian author Frigyes Karinthy (1929 short story “Chain-Links”)

“A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at all. He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances.”

Milgram's experiment (1963)

Milgram’s experiment (cont…)



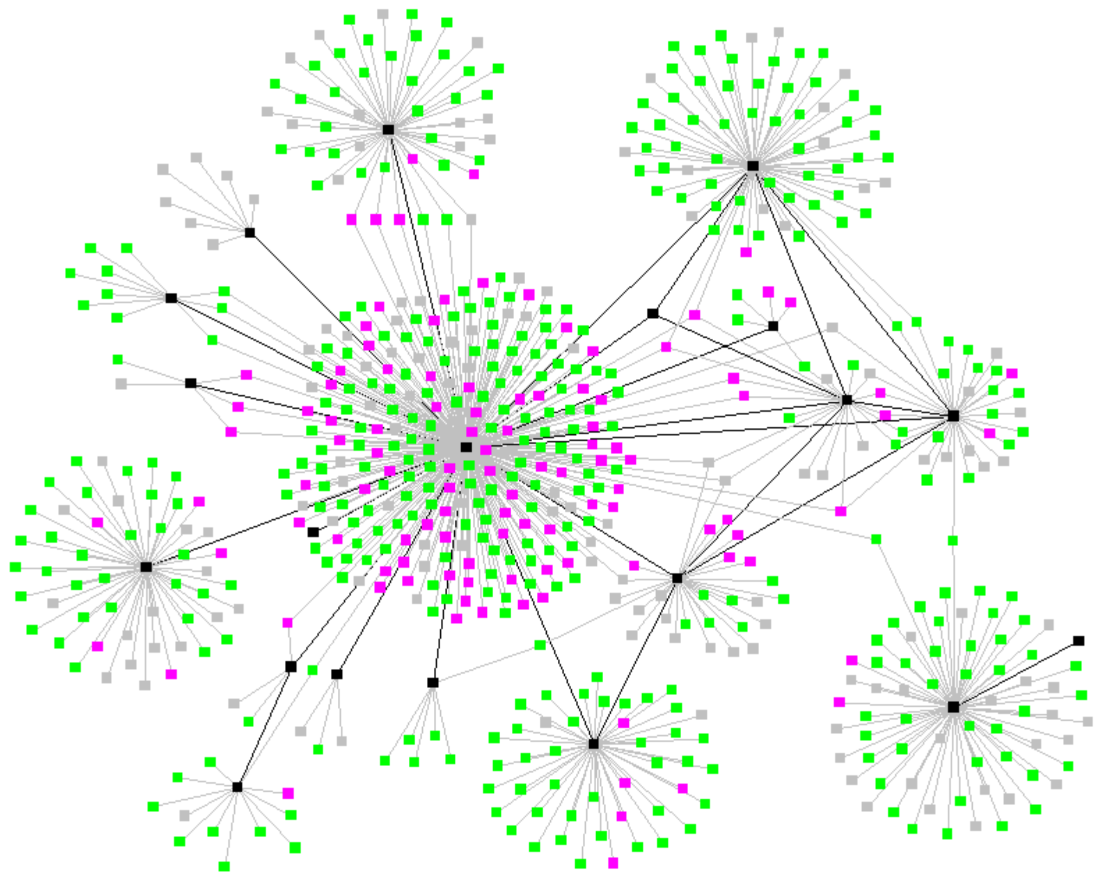
Mean = 5.2
Median = 5

Paper: https://pdodds.w3.uvm.edu/files/papers/others/1969/travers1969.pdf

Critiques

- Only 64 out of 296 cases were successful
- How useful? What is the implication?
  - Milgram: “six worlds apart”

Contagion of TB
(Valdis Krebs, Oklahoma, 2002)



Computational question

- How to find the "right 6 people?"
  - Breadth-first search (BFS) algorithm to find the shortest path
- Fun application– Bacon number
  - Bacon number of an actor = distance from Kevin Bacon
  - Average Bacon number: 2.9
  - https://oracleofbacon.org/

Shortest path algorithm

Breadth-First Search (BFS)

BFS algorithm

Goal: calculate the distance from you to everyone else in a network

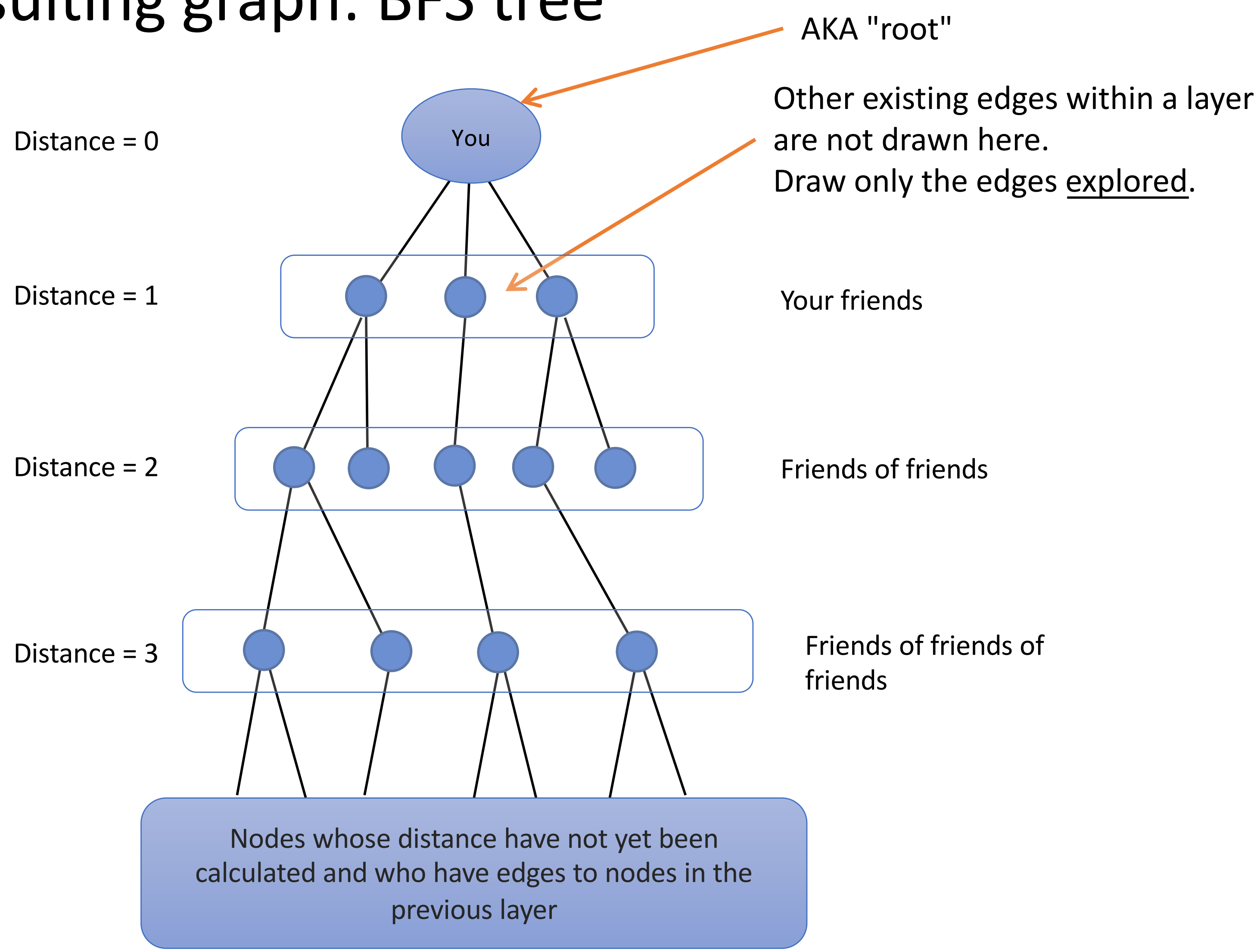
- Distance = shortest path length

BFS algorithm

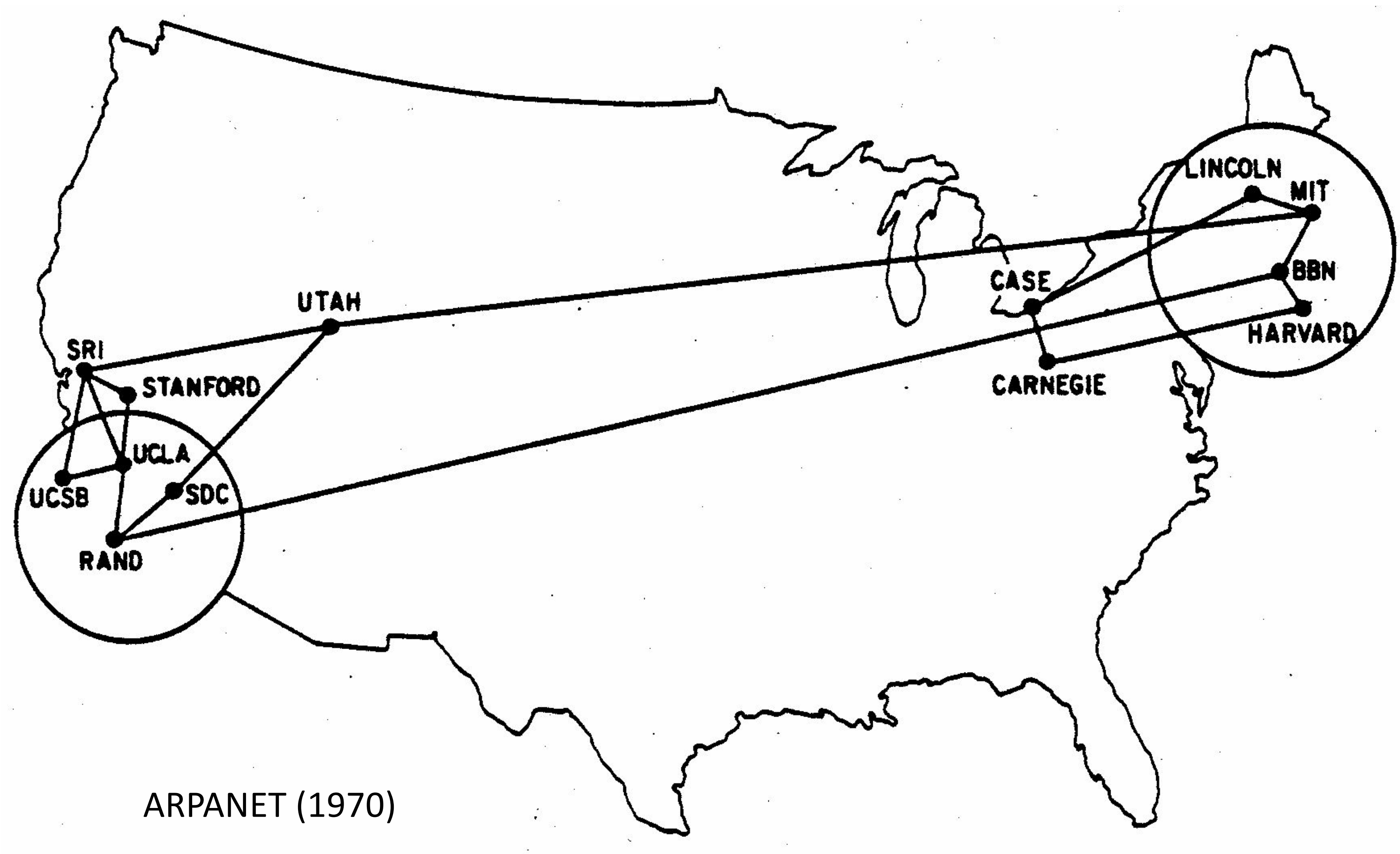
1. Define distance from you to yourself = 0
2. Define distance from you to your friends = 1
3. Define distance from you to your friends-of-friends (whose distances were not calculated) = 2
   - Do not re-define the distance of any node!
4. Continue this way until all reachable nodes' distances have been calculated.
   - Distance to an unreachable node = infinity

BFS algorithm

Resulting graph: BFS tree

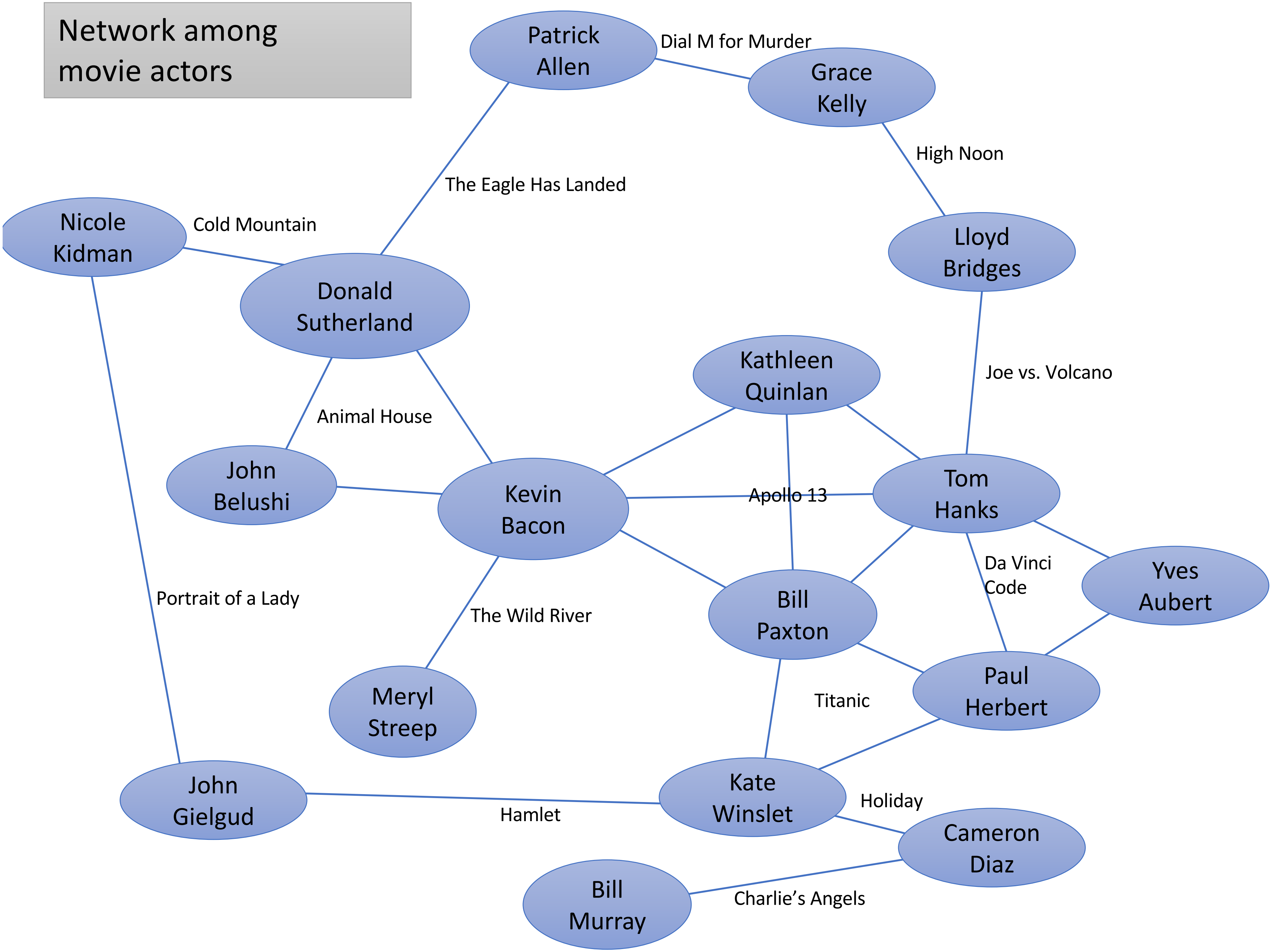


Exercise: Draw BFS tree from MIT



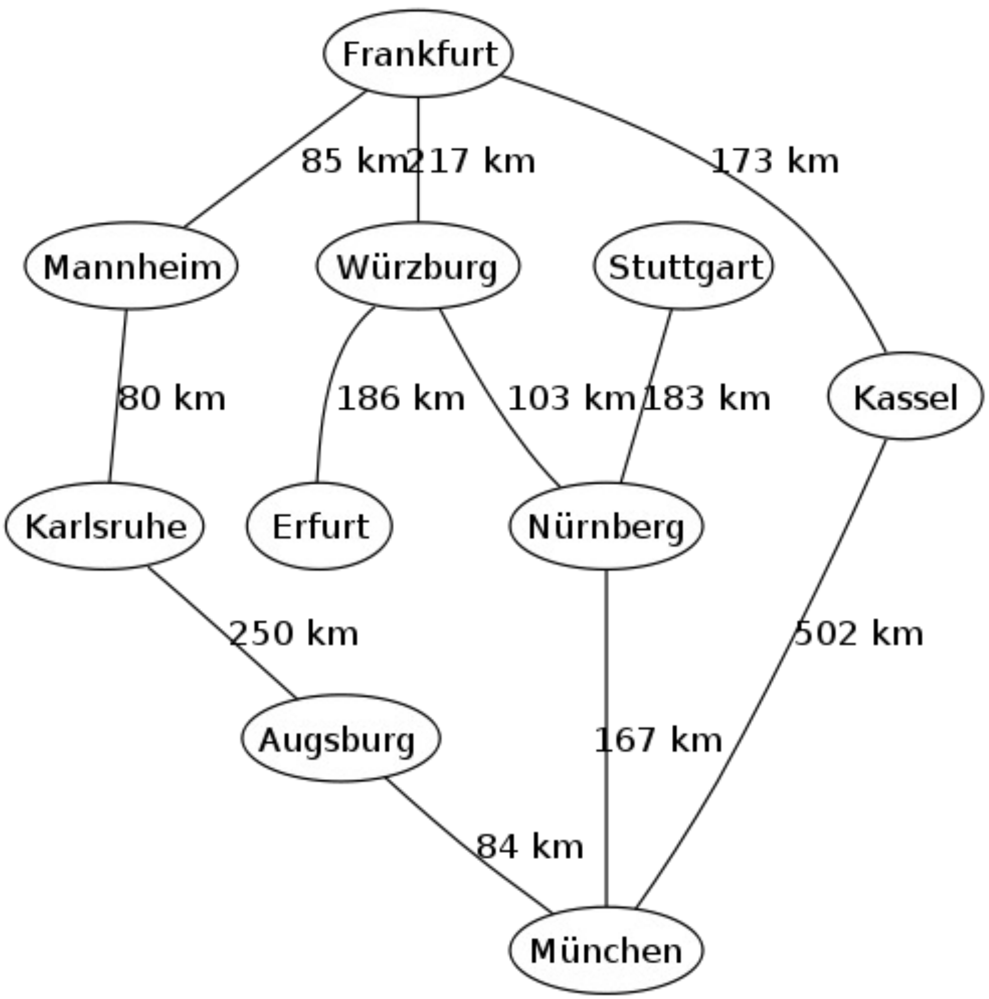
ARPANET (1970)

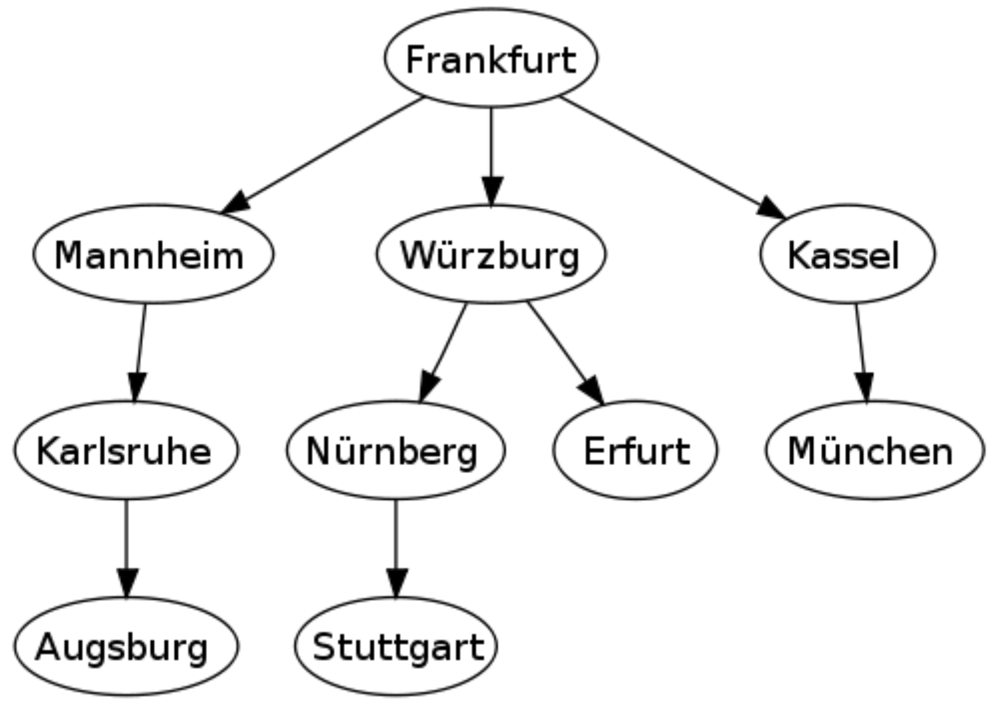
Network among movie actors



When does BFS give shortest paths?

- When all the edges have the same "weight"/dist.
- Negative example:

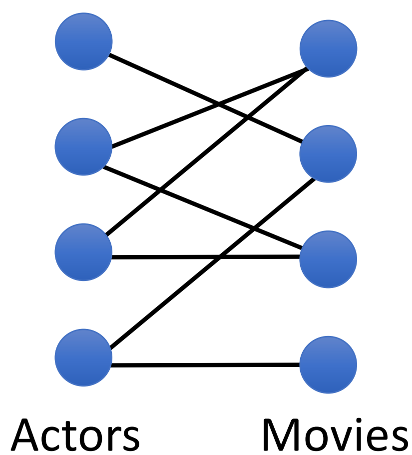




Frankfurt—Kassel—Munchen
not shortest path

Some special types of graphs

- Tree
  - Connected, acyclic graph
  - Example: BFS tree
- Bipartite graph (to be covered later)
  - Two sets of nodes with no edge within the same set of nodes
  - Example: Network between movies and actors

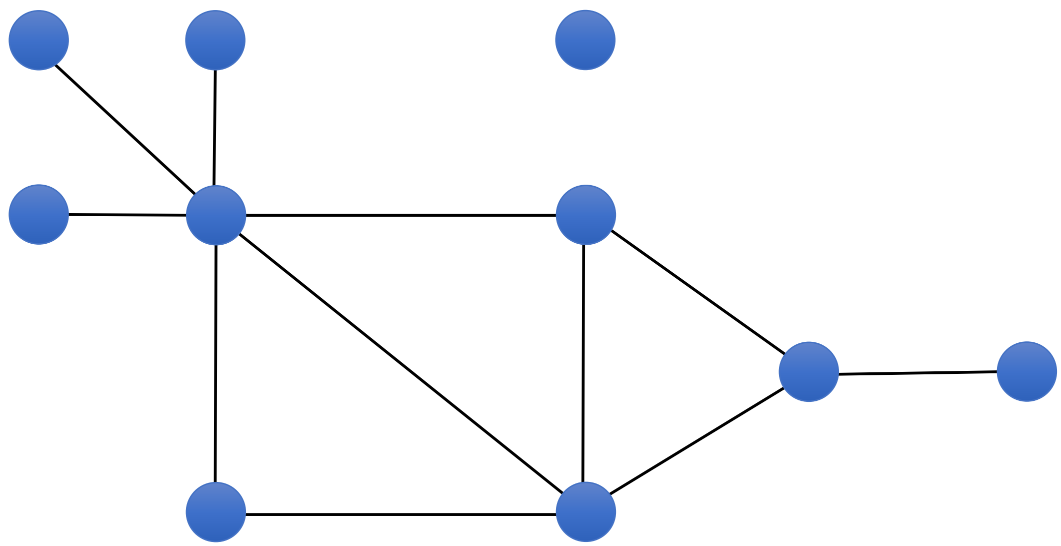


3. Degree distribution

- What's the probability of finding a node with degree k?
- What fraction of nodes have degree k?
  - Call it P_k

Example

What fraction of nodes have degree k? Call it P_k.



- P_0 = 1/10, P_1 = 4/10, P_2 = 1/10, P_3= 2/10, P_4 = 1/10, P_5 = 0/10, P_6 = 1/10
- Sum must be 1 for it to be a distribution

Real-world degree distributions

- **Power law distribution (or Pareto distrib.)**
  vs. normal distribution
- **Mathematical formulation**
- **Scale-free networks**

Extremely important
Please take note

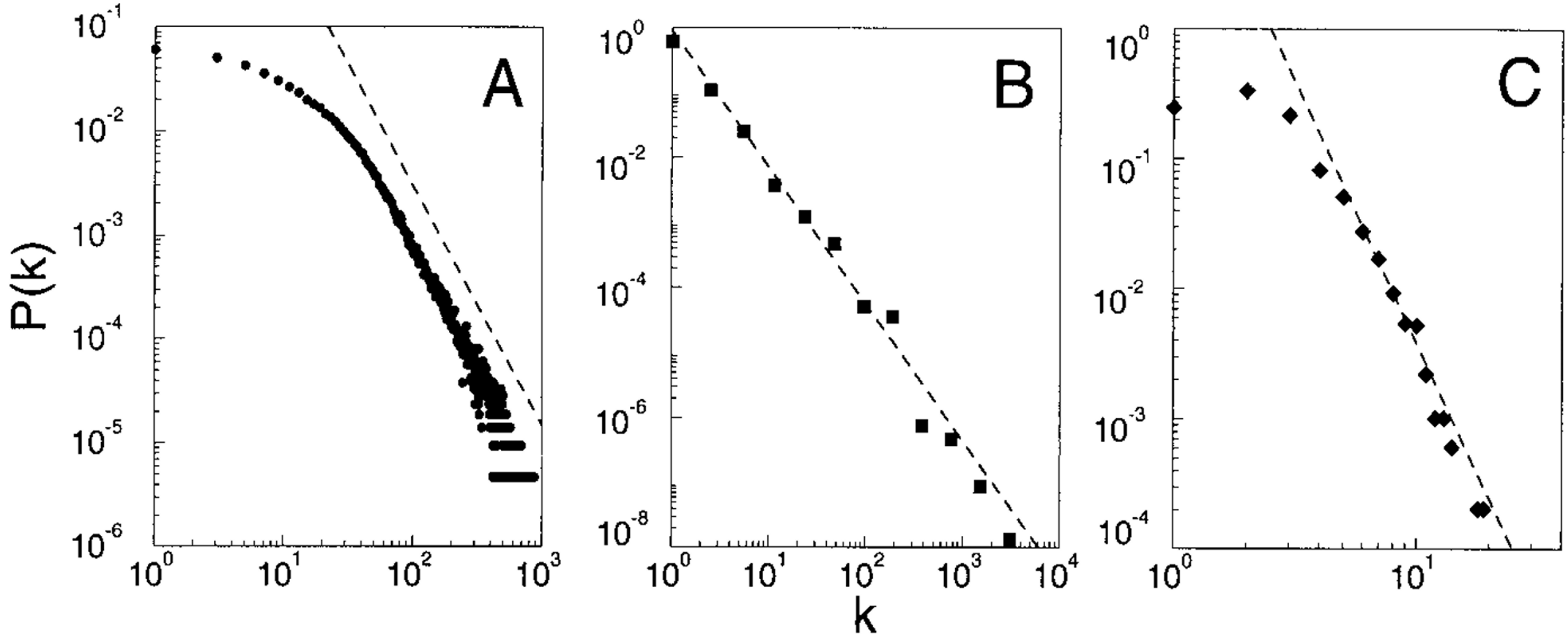


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (*6*). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Source: Emergence of Scaling in Random Networks
by Barabasi and Albert (1999)

	# nodes	# edges	avg deg			Deg variance		Slope param, α	
Network	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

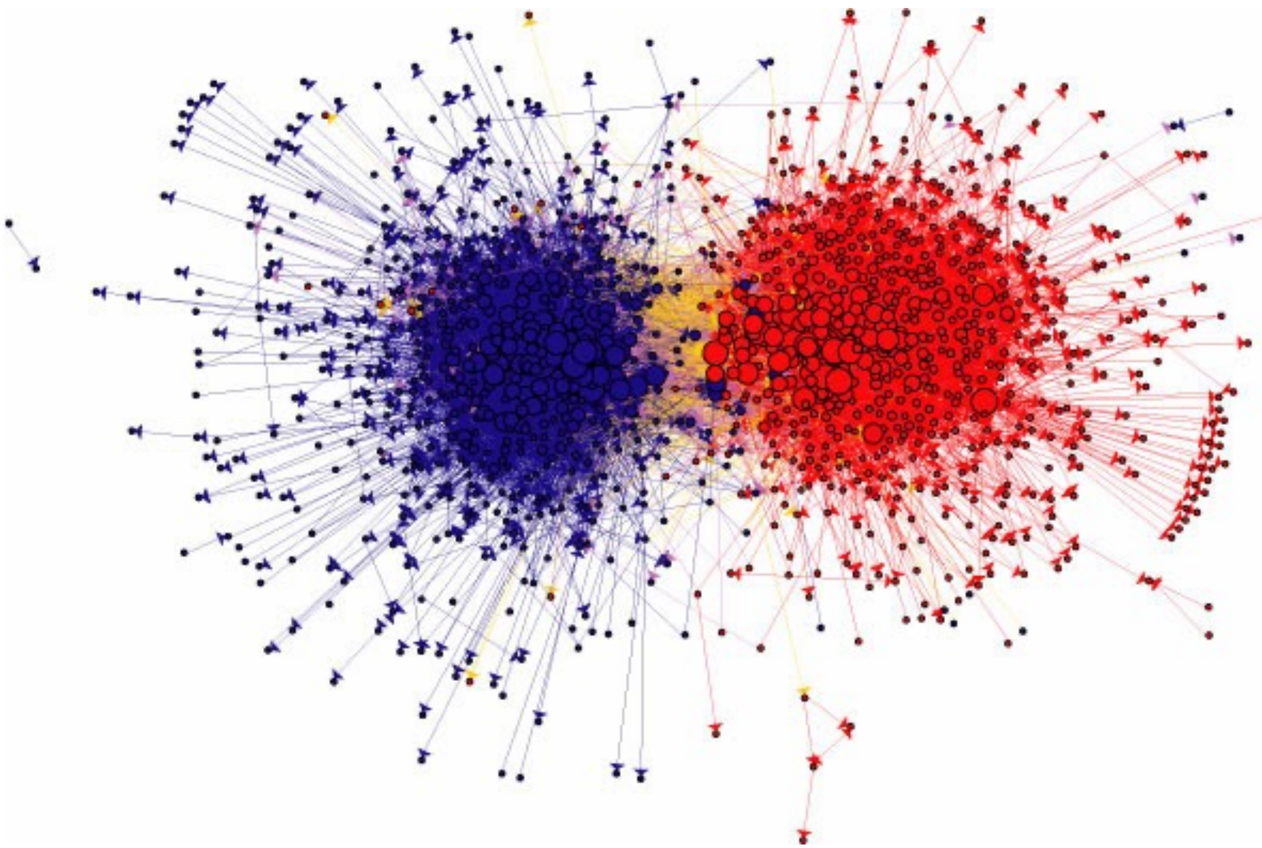
Source: Network Science Book by Barabasi (2016)

Debate on degree distribution

- Scant evidence of power law
  - https://www.quantamagazine.org/scant-evidence-of-power-laws-found-in-real-world-networks-20180215/
- Barabasi's response
  - https://tildesites.bowdoin.edu/~mirfan/files/barabasi-loveisallyouneed.pdf
- Petter Holme's take
  - https://petterhol.me/2018/01/12/me-and-power-laws/

4. Clustering coefficient (CC)

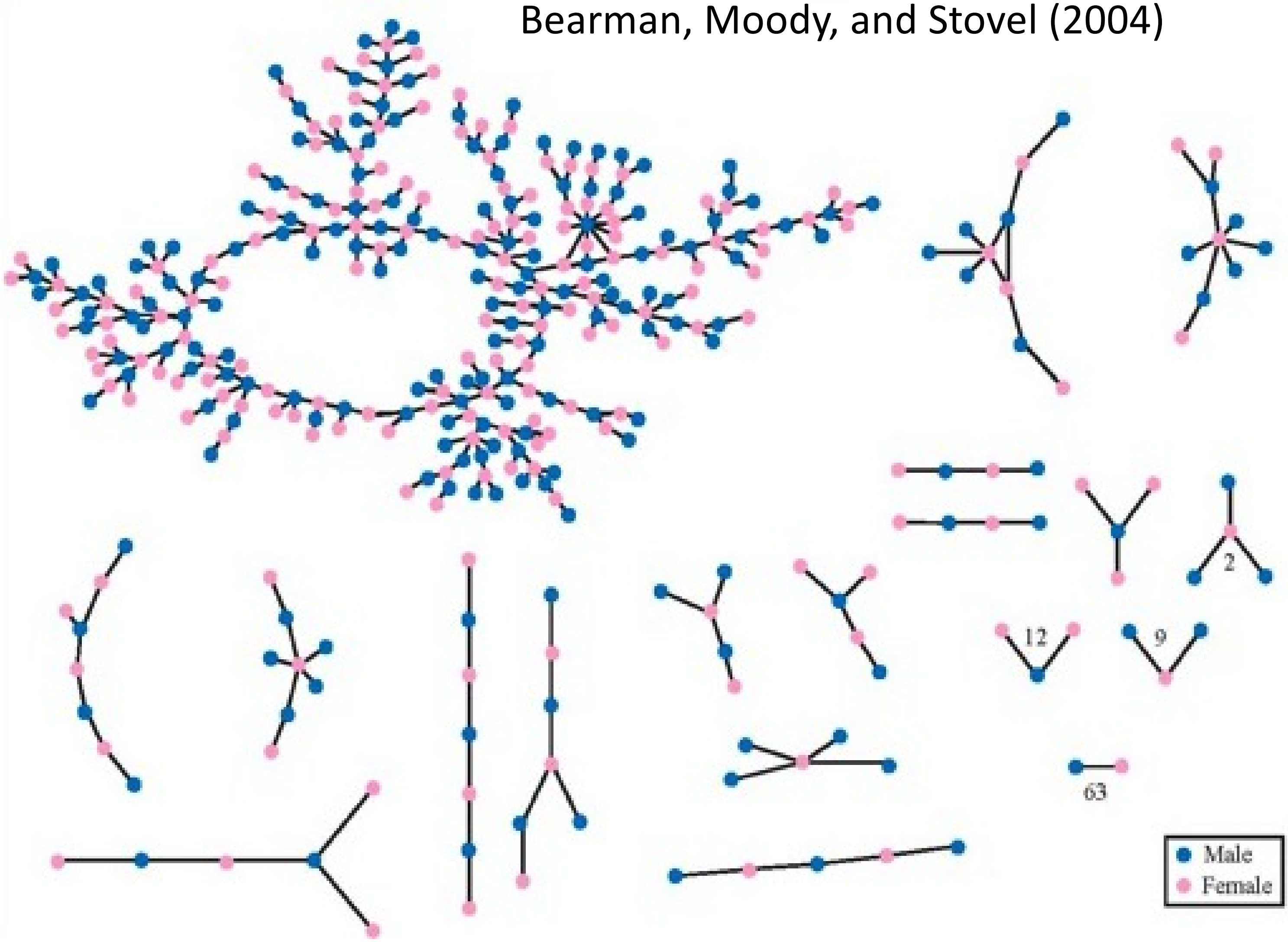
"High" clustering coefficient is observed in real-world networks



Political blogs, Adamic et al. (2005)

Example: Low CC

High school relationships (1993-95)
Bearman, Moody, and Stovel (2004)



How to compute CC?

1. Local CC of each node
2. CC of a network

4. Clustering coefficients

- Clustering coeff = Average probability that two friends of a node are also friends
- How to calculate?
  - Local clustering coeff. of node i,

Need to count

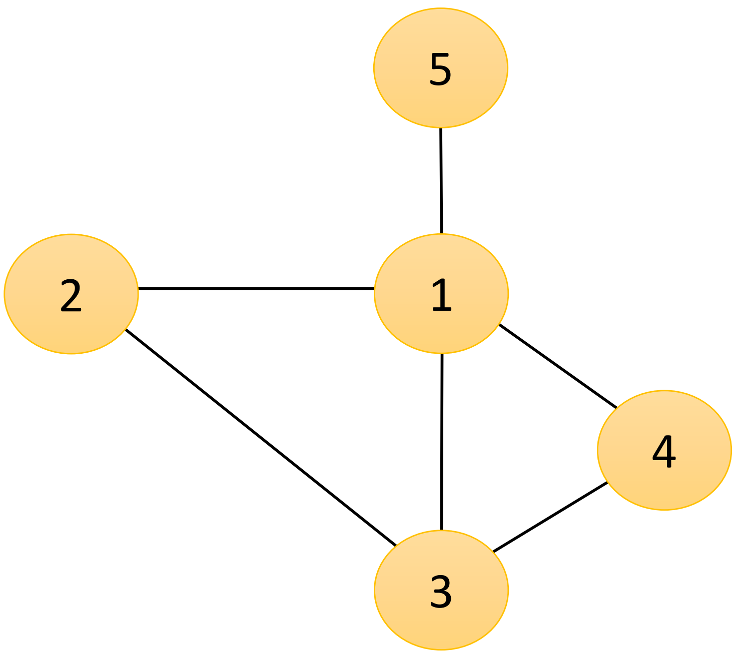
$$C_i = \frac{\text{Actual \# of edges among i's friends}}{\text{Max possible \# of edges among i's friends}}$$

$d_i (d_i - 1) / 2$
where d_i = degree of i

  - Clustering coefficient of the whole network = average C_i of all the nodes i

Example

What is the clustering coefficient of this network?



Empirical study of network properties

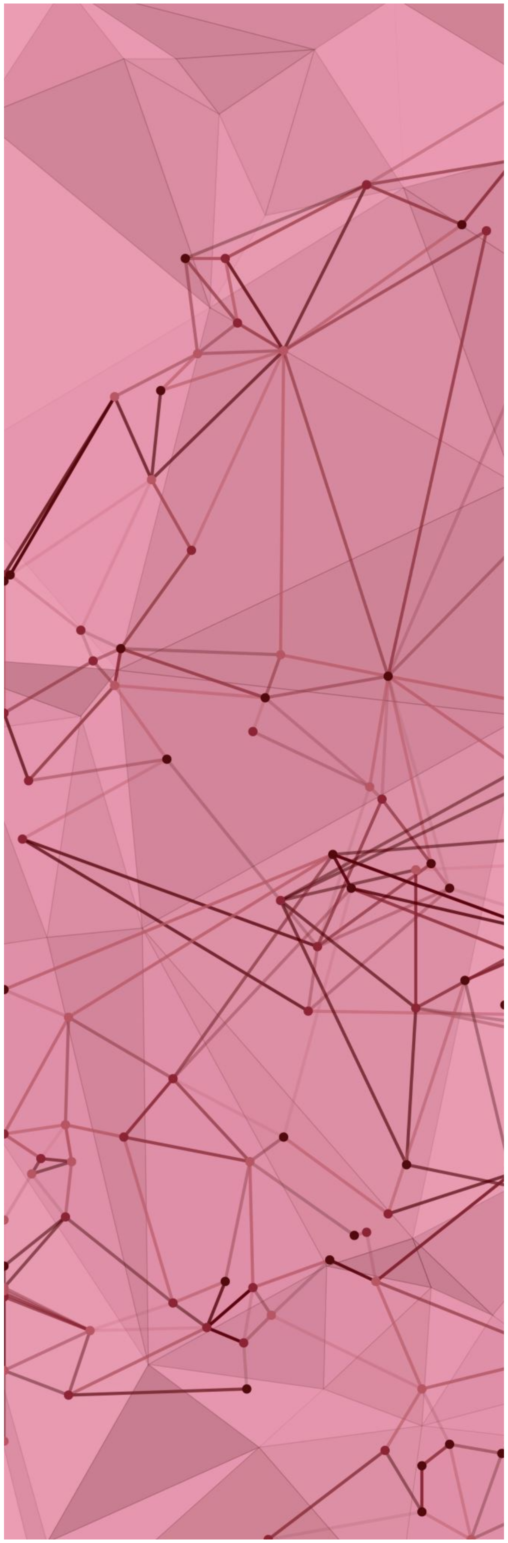
- Uzzi et al., 2007
- https://www.kellogg.northwestern.edu/faculty/uzzi/ftp/Uzzi_EuropeanManReview_2007.pdf
- N = # of nodes
  k = Avg degree
  L = Avg shortest path length
  CC = Clustering coefficient

Table 1 Small world studies

Authors	Network	Period	N	k	L Actual	L Random	CC Actual	CC Random
Organizations								
Kogut and Walker (2001)	German firms	1993–1997	291	2.02	5.64	3.01	0.84	0.022
Baum *et al.* (2003)	Canadian I-banks	1952–1957	53	1.36	3.21	4.556	0.023	0.027
		1969–1974	41	2.22	2.82	3.176	0.283	0.054
		1985–1990	142	3.83	2.95	3.144	0.273	0.027
Davis *et al.* (2003)	US Co. interlocks	1982	195	6.8	3.15	2.7	0.24	0.039
		1999	195	7.2	2.98	2.64	0.2	0.039
Verspagen and Duyster (2004)	Strategic alliances*	1980–1996	5504	5.29	4.2	5.25	0.34	0.0008
Schilling and Phelps, (forthcoming)	US alliances in 11 2-digit SIC codes**	1992–2000	171 (157)	3.11 (1.42)	20.39 (18.69)	5.62 (3.01)	0.26 (0.18)	0.04 (0.039)
Persons								
Davis *et al.* (2003)	US Director interlocks	1982	2366	19.1	4.03	2.61	0.91	0.009
		1990	2078	17.4	3.98	2.65	0.89	0.009
		1999	1916	16.3	3.86	2.69	0.88	0.009
Fleming *et al.* (forthcoming)	US patenting inventors***	1986–1990	7069	4.73	2.73	1.14	0.736	0.0452
Kogut and Walker (2001)	German Co. ownership	1993–1997	429	3.56	6.09	5.16	0.83	0.008
Newman (2004)	Biology co-authorship	1995–1999	1,520,251	18.1	4.6		0.066	
	Physics co-authorship	1995–1999	52,909	9.7	5.9		0.43	
	Mathematics co-authorship	1940–2006	253,339	3.9	7.6		0.15	
Moody, 2004	Sociologists co-authorship	1963–1999	128,151		9.81	7.57	0.194	0.207
		1989–1999	87,731		11.53	8.24	0.266	0.302
Goyal *et al.*	Economists co-authorship	1980–1989	48,608	1.244			0.182	
		1990–1999	81,217	1.672			0.157	
Watts (1999)	Hollywood Film actors	1898–1997	226,000	61	3.65	2.99	0.79	0.00027
Smith (2006)	U.S. Rappers		5533		3.9		0.18	
	U.S. Jazz musicians		1275		2.79		0.33	
	Brazilian pop		5834		2.3		0.84	
Technology								
Watts (1999)	Power grids		4941	2.94	18.7	12.4	0.08	0.005
Vazquez *et al.* (2002)	Internet	1997	3112	3.5	3.8		0.18	
		1998	3834	3.6	3.8		0.21	
		1999	5287	3.8	3.7		0.24	

Micro-Level Analysis

Centrality



Centrality

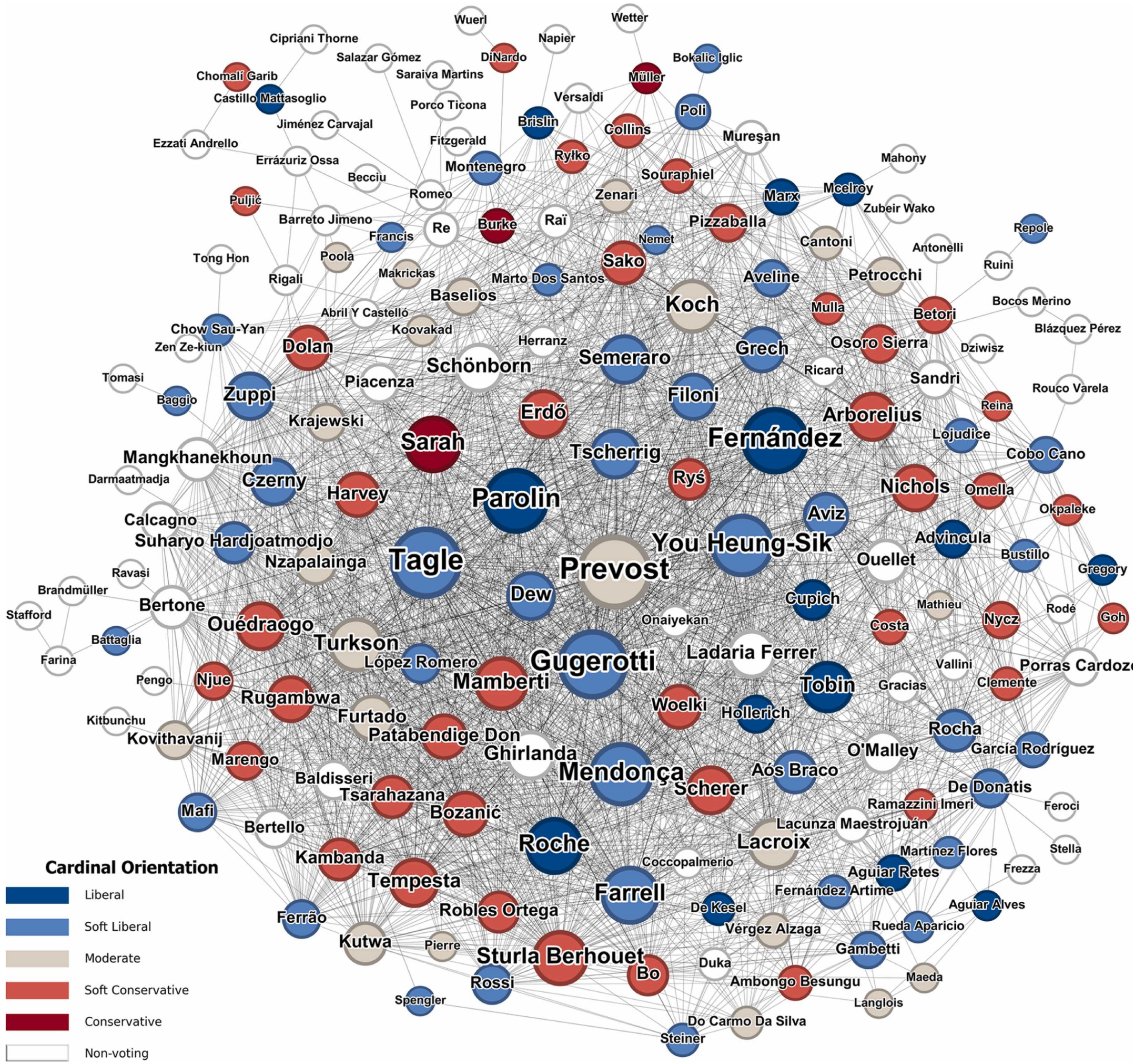
Notation: n = # of nodes

Reading: Jackson (Ch 2)

Review: Social connections and the making of a pope

"Our descriptive study—publicly released prior to the May 8, 2025 election of Pope Leo XIV—shows that Cardinal Robert F. Prevost, largely ignored by pundits, bookmakers, and AI models, held a uniquely advantageous position in the Vatican network, by virtue of being central in multiple respects."

Link to paper: https://www.sciencedirect.com/science/article/pii/S0378873325000449



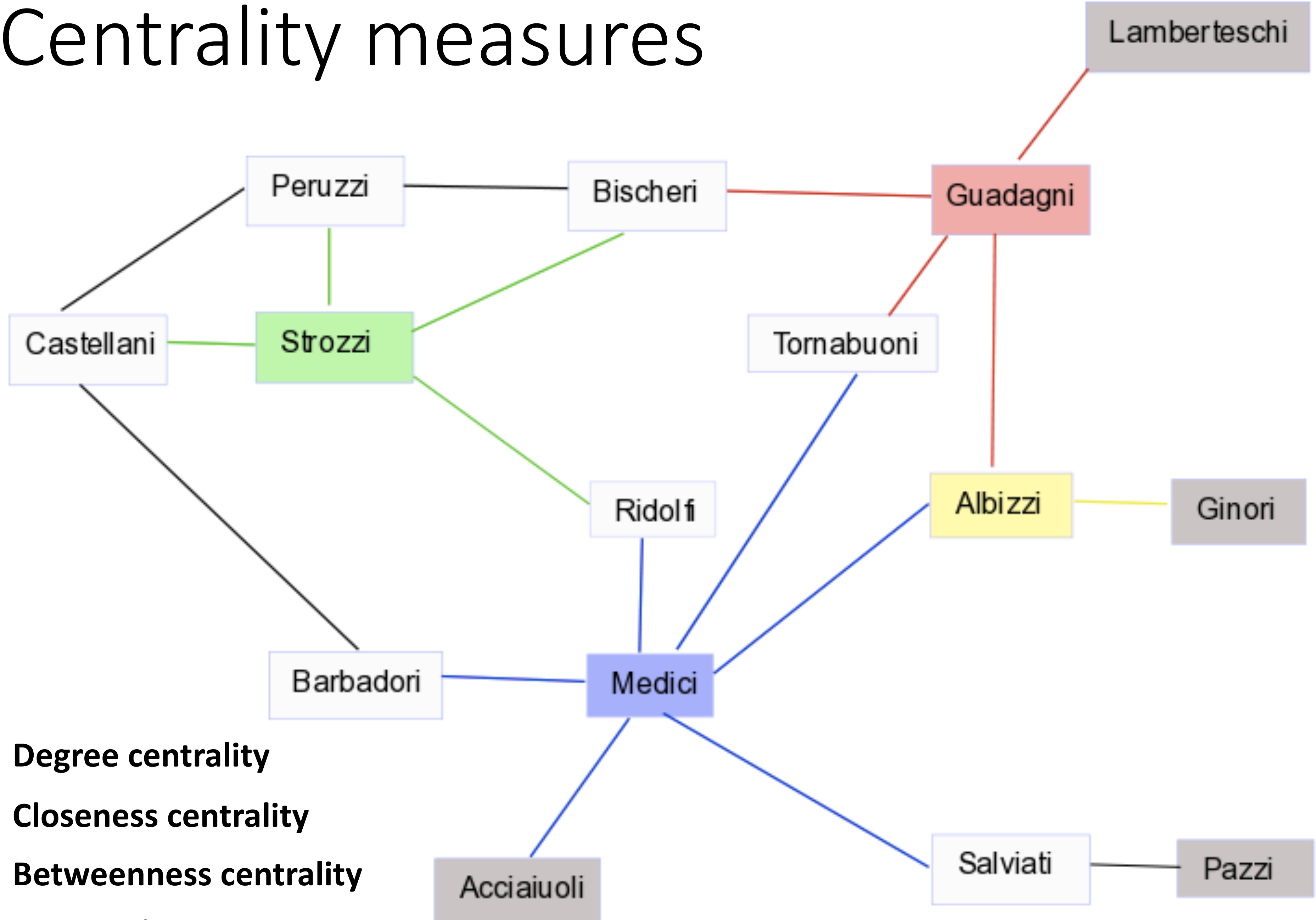
Wuerl
Wetter
Cipriani Thorne
Napier
Salazar Gómez
DiNardo
Bokalic Iglic
Chomali Garib
Saraiva Martins
Müller
Castillo Mattasoglio
Versaldi
Porco Ticona
Poli
Jiménez Carvajal
Brislin
Collins
Mureşan
Ezzati Andrello
Fitzgerald
Mahony
Errázuriz Ossa
Montenegro
Ryłko
Souraphiel
Becciu
Mcelroy
Romeo
Zenari
Marx
Puljić
Zubeir Wako
Barreto Jimeno
Re
Burke
Raï
Pizzaballa
Repole
Francis
Nemet
Cantoni
Antonelli
Tong Hon
Poola
Sako
Ruini
Rigali
Makrickas
Marto Dos Santos
Aveline
Petrocchi
Baselios
Koch
Bocos Merino
Mulla
Betori
Abril Y Castelló
Blázquez Pérez
Chow Sau-Yan
Koovakad
Herranz
Osoro Sierra
Zen Ze-kiun
Dolan
Semeraro
Grech
Dziwisz
Ricard
Tomasi
Schönborn
Sandri
Piacenza
Rouco Varela
Zuppi
Filoni
Baggio
Erdő
Reina
Arborelius
Krajewski
Fernández
Lojudice
Sarah
Mangkhanekhoun
Tscherrig
Cobo Cano
Darmaatmadja
Ryś
Czerny
Nichols
Omella
Harvey
Parolin
Okpaleke
Calcagno
Aviz
Advincula
Suharyo Hardjoatmodjo
You Heung-Sik
Tagle
Bustillo
Nzapalainga
Ouellet
Ravasi
Prevost
Gregory
Brandmüller
Dew
Cupich
Bertone
Mathieu
Stafford
Onaiyekan
Rodé
Nycz
Goh
Ouédraogo
Costa
Battaglia
Turkson
Farina
Ladaria Ferrer
Gugerotti
López Romero
Vallini
Porras Cardozo
Pengo
Njue
Mamberti
Tobin
Clemente
Gracias
Rugambwa
Woelki
Hollerich
Furtado
Kitbunchu
Patabendige Don
Rocha
Kovithavanij
Ghirlanda
O'Malley
García Rodríguez
Marengo
Mendonça
Aós Braco
Baldisseri
Scherer
De Donatis
Tsarahazana
Mafi
Ramazzini Imeri
Feroci
Bozanić
Lacunza Maestrojuán
Bertello
Roche
Lacroix
Stella
Martínez Flores
Kambanda
Coccopalmerio
Frezza
Tempesta
Aguiar Retes
Fernández Artime
Farrell
Ferrão
Robles Ortega
Aguiar Alves
De Kesel
Vérgez Alzaga
Rueda Aparicio
Kutwa
Pierre
Gambetti
Sturla Berhouet
Bo
Duka
Maeda
Rossi
Ambongo Besungu
Spengler
Langlois
Do Carmo Da Silva
Steiner
Cardinal Orientation
Liberal
Soft Liberal
Moderate
Soft Conservative
Conservative
Non-voting

Caution: centrality

Six Degrees, pg. 51

An important example of how a purely structural approach to networks has led many analysts into a reassuring but ultimately misleading view of the world is the case of *centrality*. One of the great mysteries of large distributed systems—from communities and organizations to brains and ecosystems—is how globally coherent activity can emerge in the absence of centralized authority or control. In systems like dicta-

Centrality measures



1. **Degree centrality**
2. **Closeness centrality**
3. **Betweenness centrality**
4. **Prestige/eigenvector centrality**

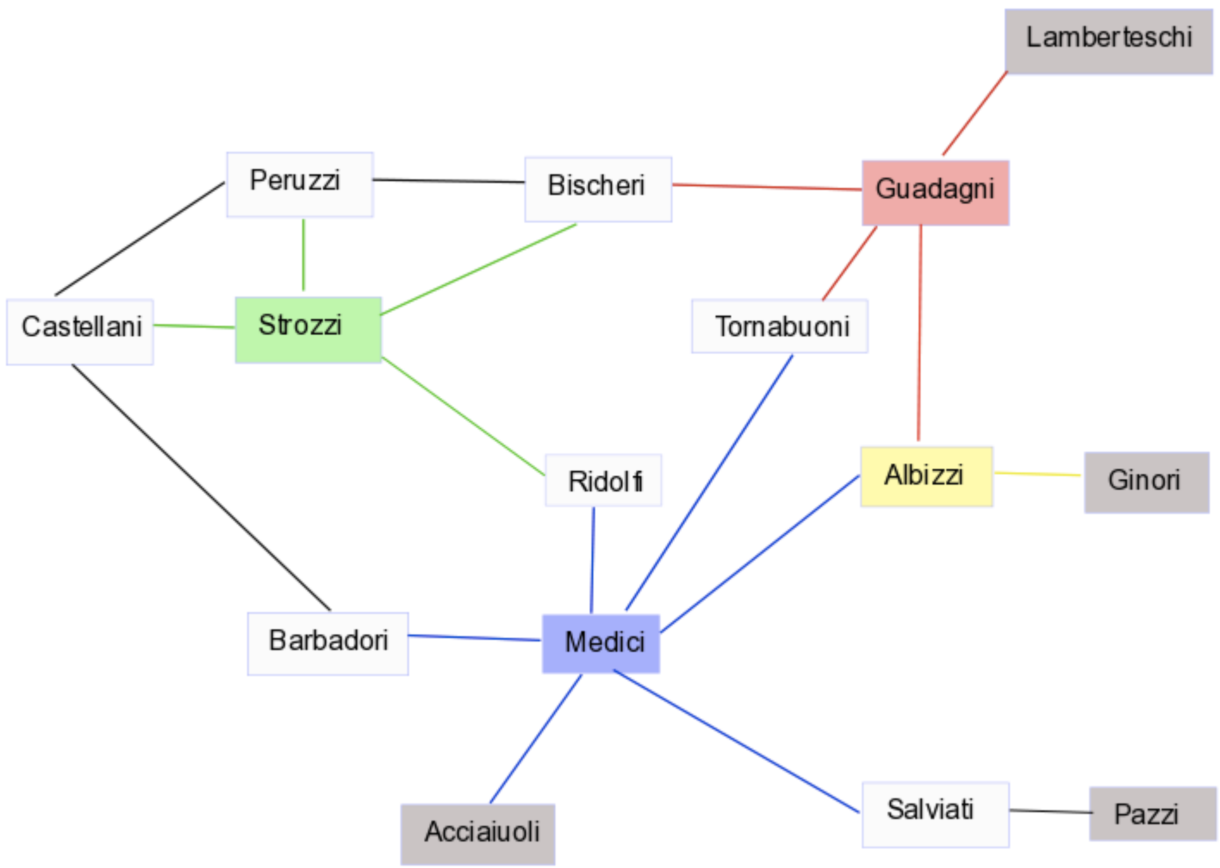
Idea | Math | Example

1. Degree centrality

Idea: Higher centrality nodes have higher degree

1. Degree centrality

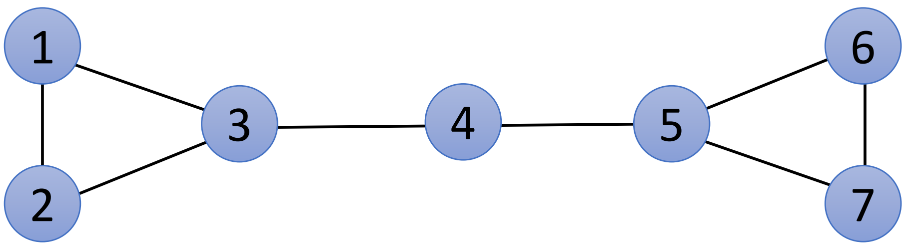
Who is the most central here?



1. Degree centrality

Downside:

How about node 4 in this network?

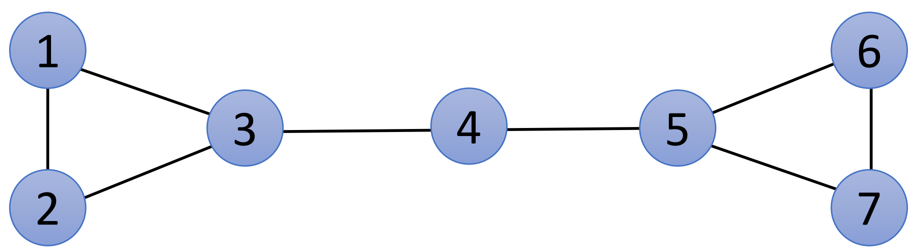


2. Closeness centrality

Idea: A node is very central if it’s very close to the other nodes

Example

Compute the closeness centralities of nodes 1 and 4



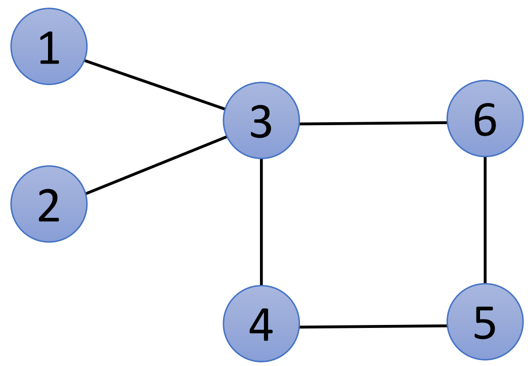
3. Betweenness centrality

Idea: a node *i* is very central if a lot of shortest paths go through *i*

Example

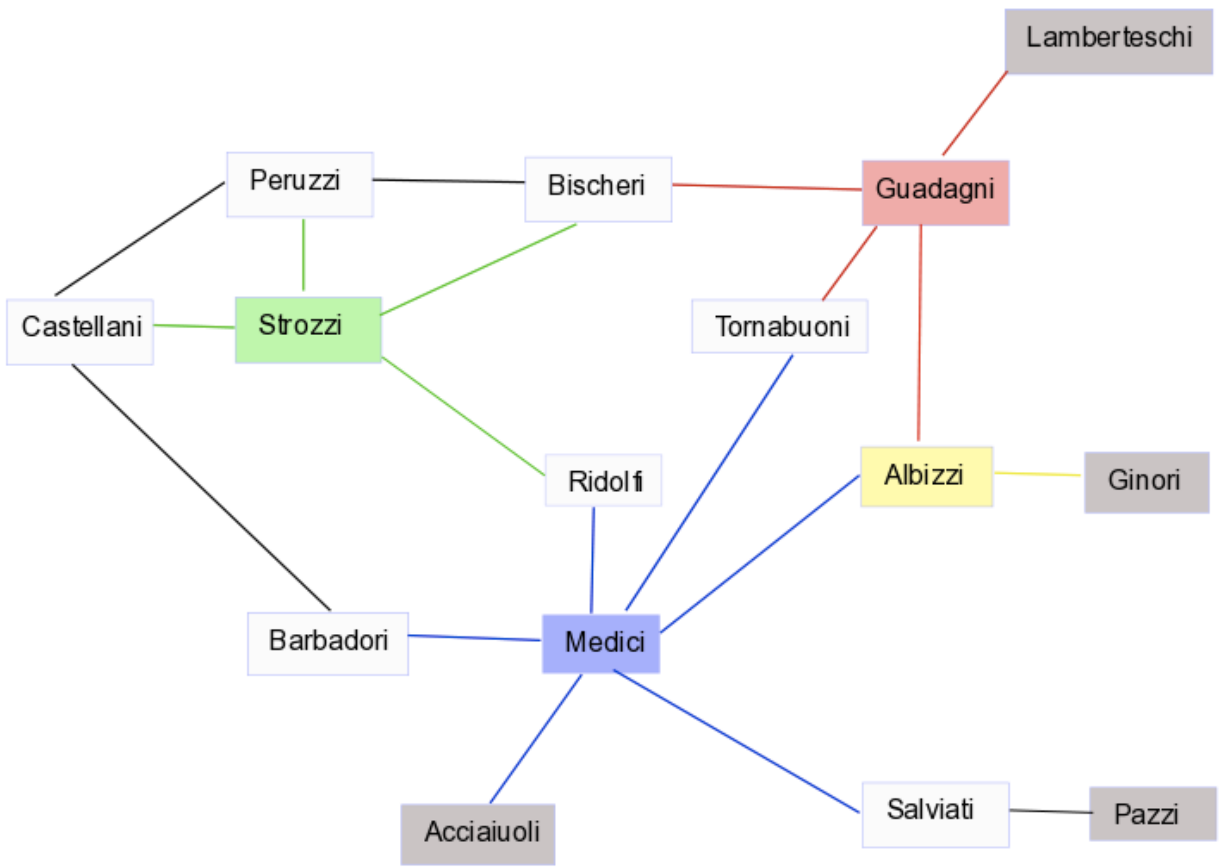
Compute the between centralities of nodes 1, 2, and 3

- $\beta_1 = 0$
- $\beta_2 = 0$
- $\beta_3 = ?$



Example

Highest betweenness centrality: Medici



Matrix algebra

Tutorial (Sec. 3 & 4) due before next class

http://www.intmath.com/matrices-determinants/3-matrices.php

Matrix algebra

- 4x1 matrix (AKA vector)

$$\begin{bmatrix} 4 \\ 2.6 \\ -8.1 \\ 7 \end{bmatrix}$$

- 3x3 matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 8 & 4 & 5 \\ 4 & -2 & 6 \end{pmatrix}$$

Graph example

Adjacency matrix

Matrix multiplication

- 2x3 matrix multiplied by 3x2 matrix

must match

- Result is a 2x2 matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix} = \begin{bmatrix} au + bw + cy & av + bx + cz \\ du + ew + fy & dv + ex + fz \end{bmatrix}$$

4. Eigenvector/prestige/power centrality

Applications

- PageRank algorithm for ranking websites
- Brain activity pattern: use fMRI data to understand how brain regions interact
- Finding critical nodes in infrastructure networks, financial networks, etc.

4. Eigenvector/prestige/power centrality

- Idea (Phillip Bonacich, 1987): A node's importance is determined by its friends' importance
- Mathematical formulation
- Example

Perron-Frobenius Theorem

For the largest eigenvalue, the corresponding eigenvector is nonnegative (for any nonnegative matrix)

- Reassuring!

Eigenvector calculator

WolframAlpha computational intelligence.

eigenvector calculator

Extended Keyboard Upload Examples Random

Computational Inputs:

» matrix:

{{0, 1, 0, 0}, {1, 0, 1, 1}, {0, 1, 0, 1},

Compute

Input:

$$\text{eigenvectors}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Results: Exact forms Step-by-step solution

$v_1 \approx (0.539189, 1.17009, 1, 1)$

$v_2 \approx (1.67513, -2.48119, 1, 1)$

$v_3 = (0, 0, -1, 1)$

$v_4 \approx (-2.21432, -0.688892, 1, 1)$

Corresponding eigenvalues: Exact forms Step-by-step solution

$\lambda_1 \approx 2.17009$

$\lambda_2 \approx -1.48119$

$\lambda_3 = -1$

$\lambda_4 \approx 0.311108$

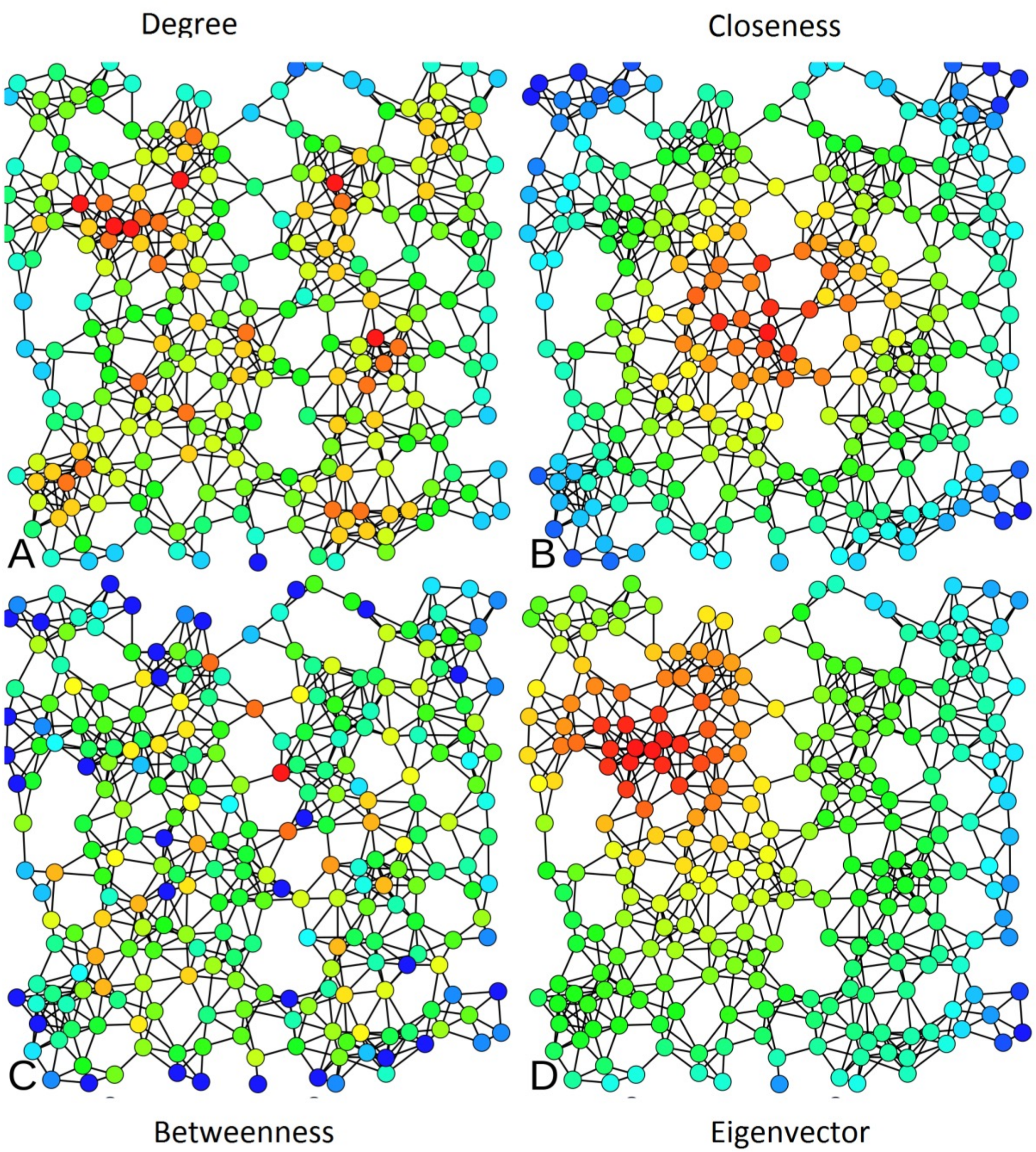
More on eigenvector centrality

- Tutorial on eigenvector
  - Jackson's Section 2.4 (Appendix)

Comparison

- What are the differences among:
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvector centrality

Comparison of centrality measures



Gephi demo of centrality

- Dataset: http://bit.ly/gephi_dolphin (dolphin network)
- Statistics tab
  - Average degree
  - Network diameter
  - Eigenvector centrality
- Appearance tab
  - Rank and color nodes according to centrality measures

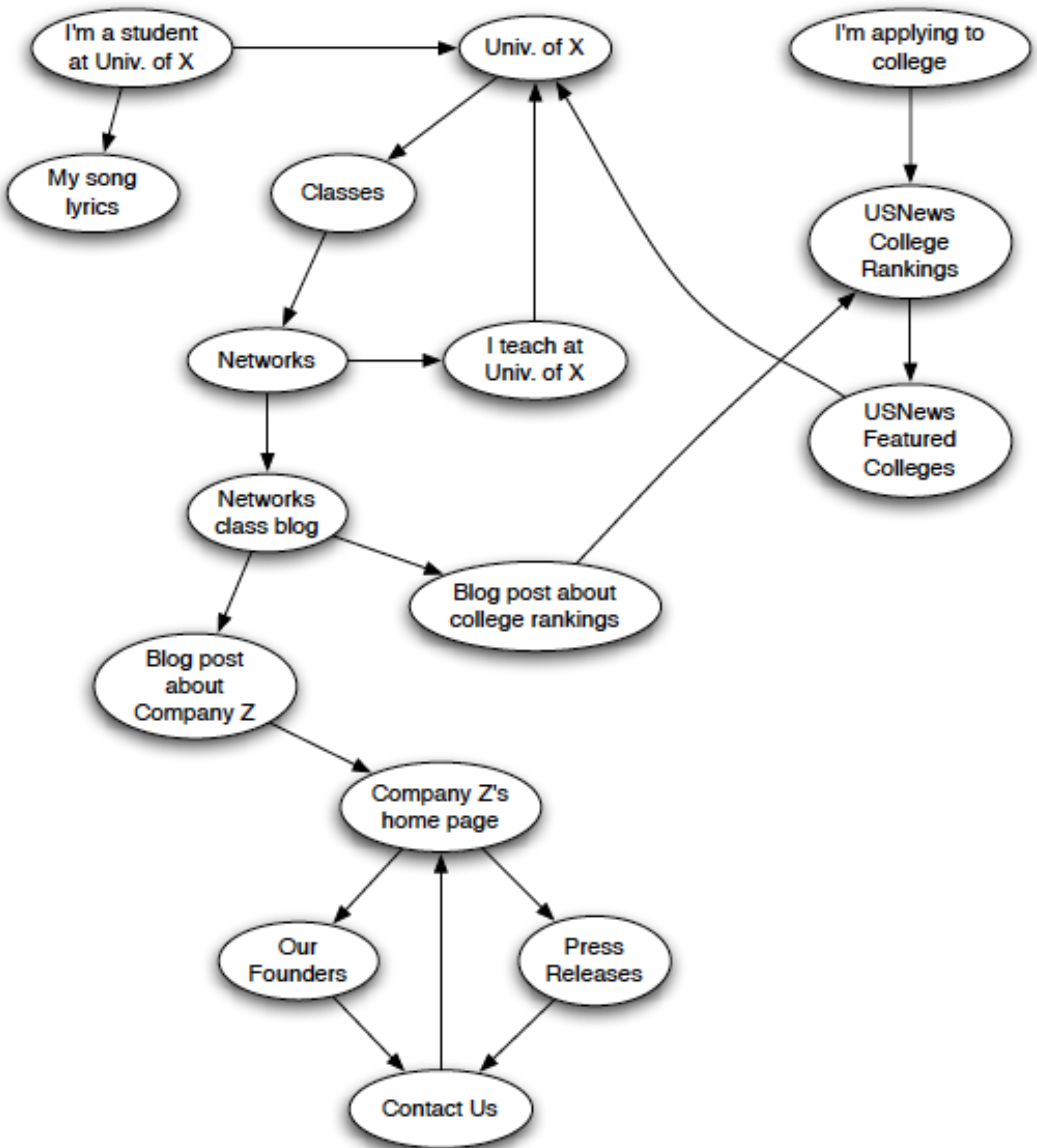
PageRank:
An application of eigenvector centrality

Chapter 13 [EK]

Note: the details of PageRank are out of scope for this topic. We'll cover it later in the semester.

Web as a directed graph

- Nodes: Web pages
- Directed edges: Links
- bowdoin.edu → Arts → Museum of Art → Exhibitions → … → bowdoin.edu
  - A directed cycle


I'm a student at Univ. of X
Univ. of X
I'm applying to college
My song lyrics
Classes
USNews College Rankings
Networks
I teach at Univ. of X
USNews Featured Colleges
Networks class blog
Blog post about college rankings
Blog post about Company Z
Company Z's home page
Our Founders
Press Releases
Contact Us

PageRank: idea

- A webpage is important if it is cited by other important webpages
  - Bonacich’s idea on centrality (1987)
- Iteratively refine the PR of each webpage

PageRank (PR) algorithm

- Input: directed network with n nodes and desired number of rounds k
- Step 1: Assign each node initial PR = 1/n
- Step 2: Repeat for k rounds:
    - ❑ **Out-Phase:** Each node divides its current PR equally across its outgoing links and passes these equal shares to the nodes it points to.
    - ❑ **In-Phase:** Each node replaces its PR with the sum of the shares it receives.

Q: What if PR values do not change between two consecutive rounds?