DCS/CSCI 2350
Social & Economic Networks

What does a real-world network look like?

Reading: Ch 2 of EK, Ch 2 & 3 of Jackson
Graph visualization using Gephi

Mohammad T. Irfan
Email: mirfan@bowdoin.edu

1

Real-World Networks

Investigate at:

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

2

Macro-level:

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

3

1. Giant component

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

4

Examples

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

5

6

More examples

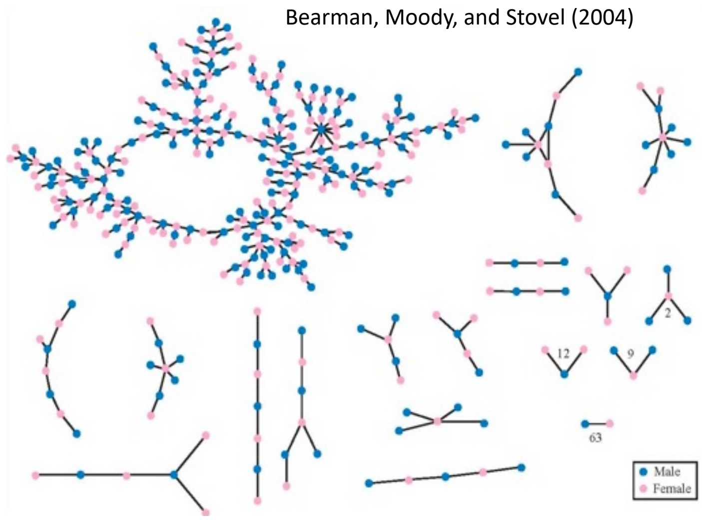
- Instant messaging
  - Microsoft IM: one giant component in a network of 240 million users (2008)
- Co-author network
- Email
- Biological networks (neural networks)
- Technology networks (power grid)
- The Internet (web of links)

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

7

Giant component: implications

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



9

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

11

Six degrees of separation

- John Guare’s play (1990) & later movie
- http://www.youtube.com/watch?v=HLIyuYwbVnA

12

Six degrees of separation

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

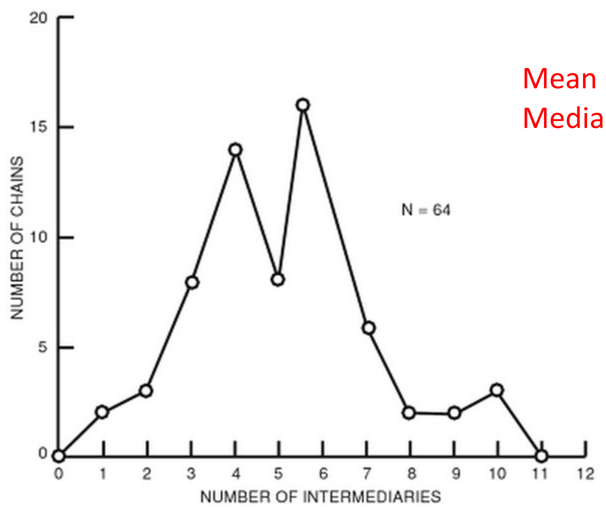
> “A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at 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.”

13

Milgram’s experiment (1963)

14

Milgram's experiment (cont…)



Mean = 5.2
Median = 5

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

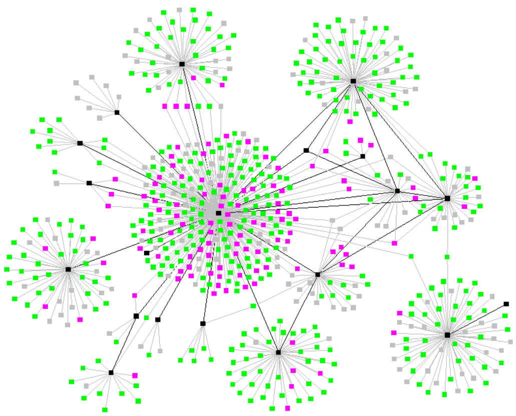
15

Critiques

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

16

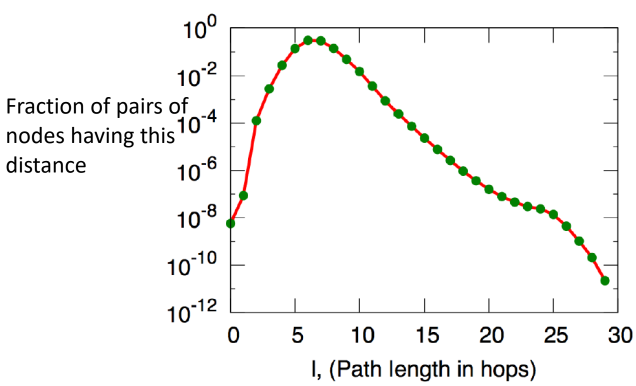
Contagion of TB
(Valdis Krebs, Oklahoma, 2002)



17

Another example

- Microsoft instant messenger (2008)
  - 240M node network
  - Edge: Two-way conversation at some point during a month-long observation period
  - Average distance: 6.6



18

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/

19

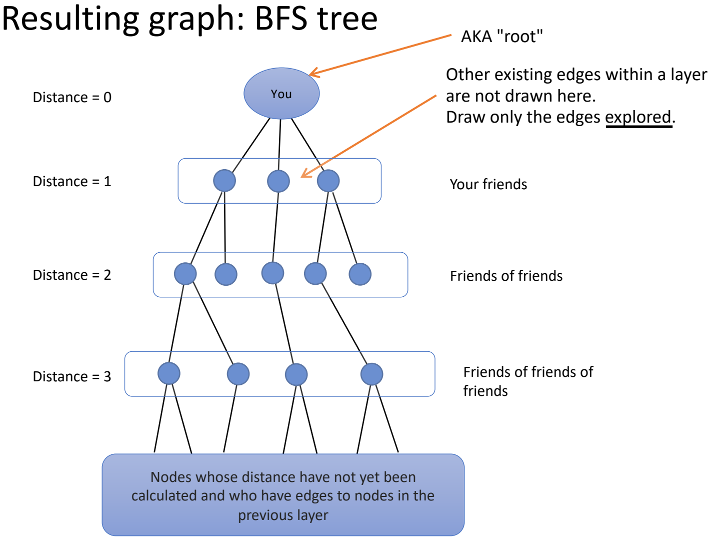
Shortest path algorithm

Breadth-First Search (BFS)

20

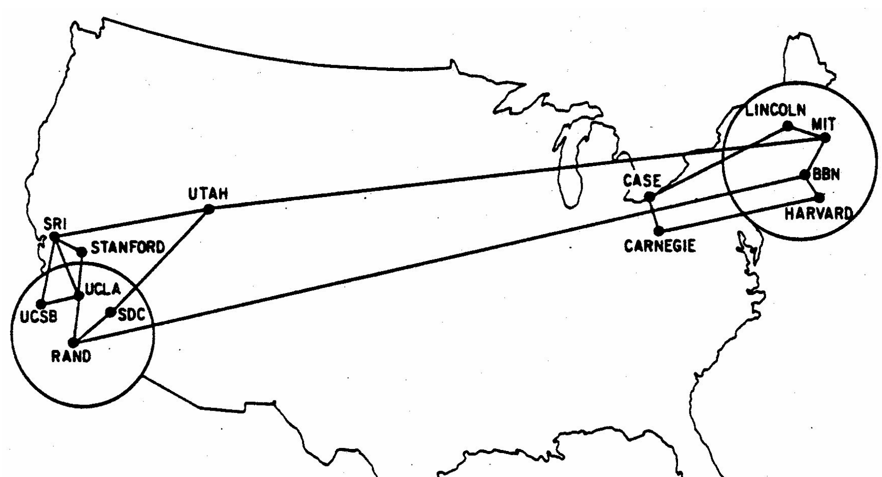
BFS algorithm

- Resulting graph: BFS tree



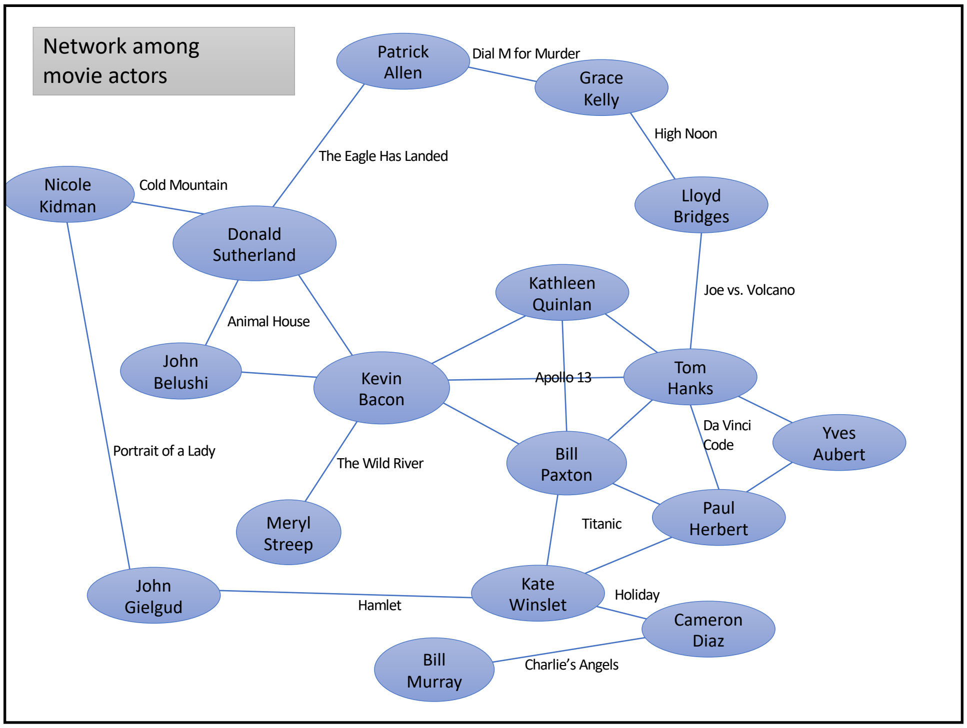
22

Exercise: Draw BFS tree from MIT



ARPANET (1970)

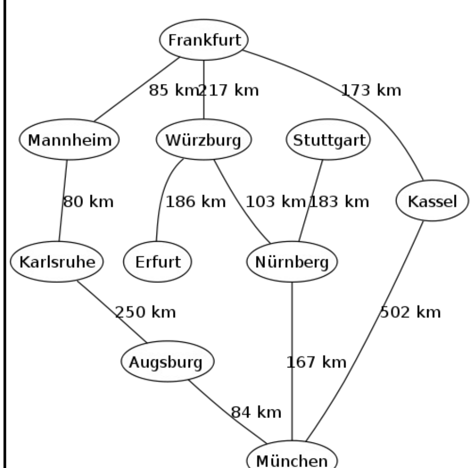
23

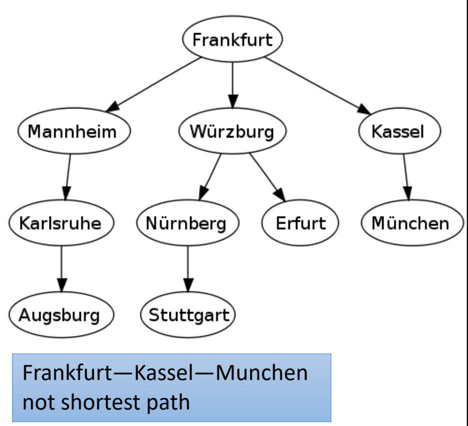


24

When does BFS give shortest paths?

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

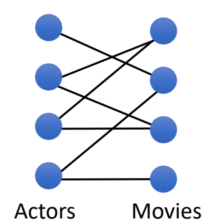




25

Some special types of graphs

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



26

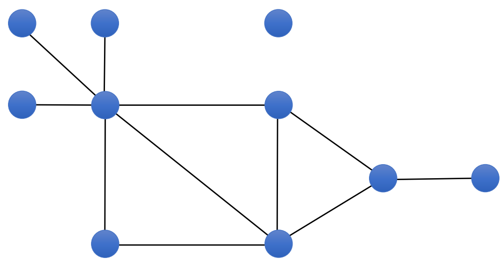
3. Degree distribution

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

27

Review: 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

28

Real-world degree distributions

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

Extremely important
Please take note

29

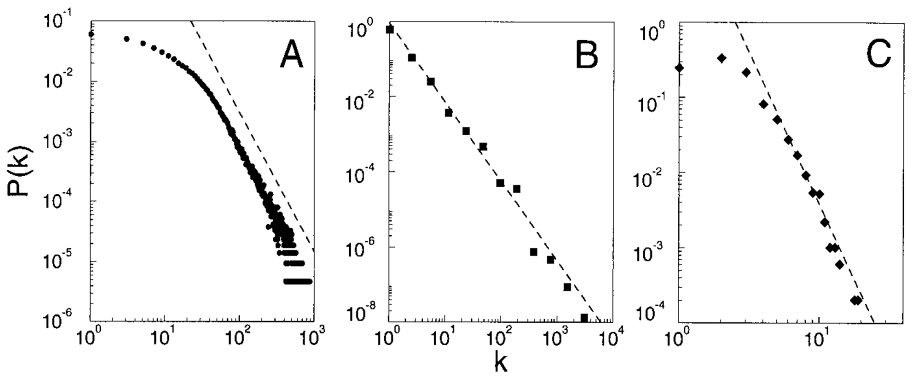


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (*6*). (**C**) Power grid data, $N = 4941$, $\langle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

30

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)

32

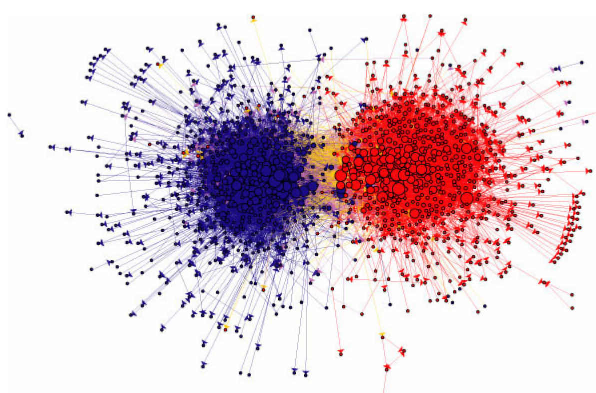
Debate on degree distribution

- 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/

33

4. Clustering coefficient (CC)

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

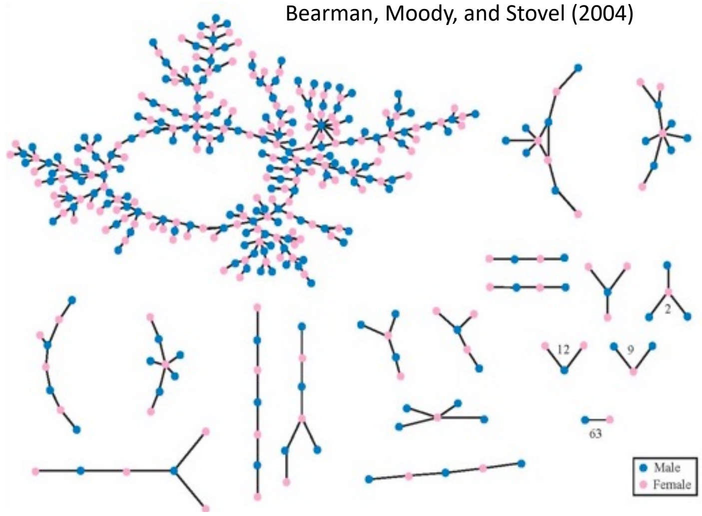


Political blogs, Adamic et al. (2005)

35

Example: Low CC

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



36

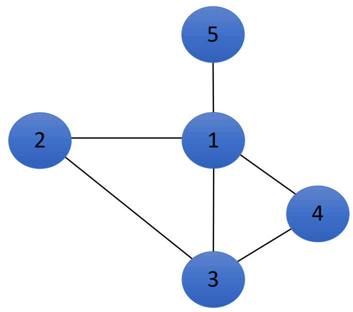
How to compute CC?

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

37

Example

- What is the clustering coefficient of this network?



39

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

40

Table 1 Small world studies

Authors	*Network*	*Period*	*N*	*k*	*L Actual*	*L Random*	*CC Actual*	*CC Random*	*Lr*	*CCr*	*Q*
Organizations											
Kogut and Walker (2001)	German firms	1993–1997	291	2.02	5.64	3.01	0.84	0.022	1.87	38.18	20.38
Baum *et al.* (2003)	Canadian I-banks	1952–1957	53	1.36	3.21	4.556	0.023	0.027	0.70	0.85	1.21
		1969–1974	41	2.22	2.82	3.176	0.283	0.054	0.89	5.24	5.90
		1985–1990	142	3.83	2.95	3.144	0.273	0.027	0.94	10.11	10.78
Davis *et al.* (2003)	US Co. interlocks	1982	195	6.8	3.15	2.7	0.24	0.039	1.17	6.15	5.27
		1999	195	7.2	2.98	2.64	0.2	0.039	1.13	5.13	4.54
Verspagen and Duyster (2004)	Strategic alliances*	1980–1996	5504	5.29	4.2	5.25	0.34	0.0008	0.80	425.00	531.25
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)	3.85 (2.84)	10.44 (7.53)	2.71 (2.65)
Persons											
Davis *et al.* (2003)	US Director interlocks	1982	2366	19.1	4.03	2.61	0.91	0.009	1.54	101.11	65.48
		1990	2078	17.4	3.98	2.65	0.89	0.009	1.50	98.89	65.84
		1999	1916	16.3	3.86	2.69	0.88	0.009	1.43	97.78	68.14
Fleming *et al.* (forthcoming)	US patenting inventors***	1986–1990	7069	4.73	2.73	1.14	0.736	0.0452	2.394737	16.28	6.80
Kogut and Walker (2001)	German Co. ownership	1993–1997	429	3.56	6.09	5.16	0.83	0.008	1.18	103.75	87.91
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	1.30	0.94	0.72
		1989–1999	87,731		11.53	8.24	0.266	0.302	1.40	0.88	0.63
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	1.22	2925.93	2396.85
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	1.51	16.00	10.61
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				

41

Micro-level:

Centrality

42

Graph Visualization

Gephi

43

Links

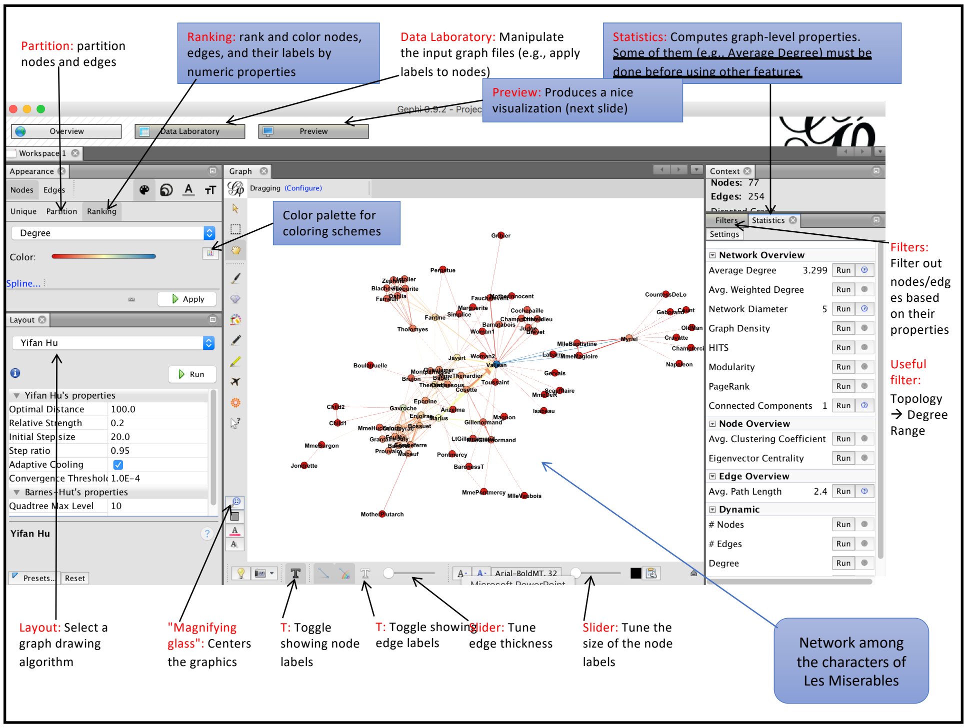
- Download
  - https://gephi.org/
- Video Tutorial: https://bit.ly/gephi-panopto
- Text tutorial: http://bit.ly/gephi_tutorial
- Dataset: http://bit.ly/gephi_dataset

Windows: Gephi requires Java. Most modern Windows PCs already have it.

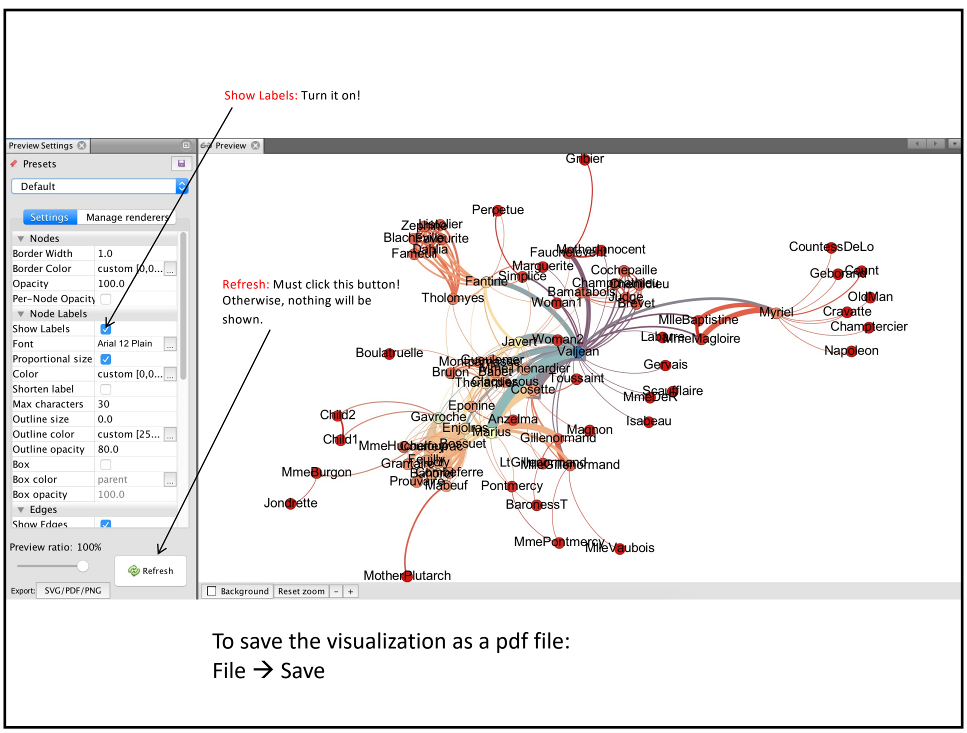
How to find Java version in Windows?
https://www.java.com/en/download/help/version_manual.xml
Where to get Java for Windows?
https://www.java.com/en/download/

Mac OS X: Java is bundled with the application so it doesn't have to be installed separately.

44


Partition: partition nodes and edges
Ranking: rank and color nodes, edges, and their labels by numeric properties
Data Laboratory: Manipulate the input graph files (e.g., apply labels to nodes)
Statistics: Computes graph-level properties. Some of them (e.g., Average Degree) must be done before using other features
Preview: Produces a nice visualization (next slide)
Color palette for coloring schemes
Filters: Filter out nodes/edges based on their properties
Useful filter: Topology → Degree Range
Layout: Select a graph drawing algorithm
"Magnifying glass": Centers the graphics
T: Toggle showing node labels
T: Toggle showing edge labels
Slider: Tune edge thickness
Slider: Tune the size of the node labels
Network among the characters of Les Miserables

45


Show Labels: Turn it on!
Refresh: Must click this button! Otherwise, nothing will be shown.
To save the visualization as a pdf file:
File → Save

46

Red: Graph level
Black: Node/edge level

Gephi Vocabulary

Term	Meaning
betweenness centrality of a node	how often the node appears on the shortest path between nodes in the network
closeness centrality of a node	average distance from that node to all other nodes in the network
degree of a node	the number of edges connected to the node (also connectedness); in a directed graph a node can have in-degree and out-degree measures
diameter of a graph	the longest shortest path between any two nodes in the graph
directed graph	this means relationships occur one way only (I follow you, but you do not follow me on Twitter); opposite of undirected (we are friends with each other on Facebook)
eccentricity of a node	the distance (shortest-path length) from the node to the farthest node from it in the network
edge	a representation of the connection between two nodes, expresses a relationship (a line)
eigenvector centrality of a node	in social network analysis, a measure of influence (a node is very influential if it is connected to other influential nodes)
layout algorithms	also known as graph drawing algorithm; e.g., force-directed drawing where linked nodes attract and non-linked nodes repel
leaf node	node with a single edge in a "tree-structured" graph
modularity	a measure of connectedness among groups of nodes (greater than 0.4 is usually considered meaningful)
node	also called a vertex by mathematicians; a person in a social network graph (a dot or bubble)
distance from one node to another	the length of the shortest path (counted in the number of edges) from one node to another
path length	the number of edges in a path
singleton node or isolated node	node with no edge/connection

47

Centrality

Notation: n = # of nodes

Reading: Jackson (Ch 2)

49

Gephi demo: centrality

- Data: .gml

http://bit.ly/gephi_dataset (Les Miserables data)
http://bit.ly/gephi_dolphin (Dolphin network data)

Dolphin network data: Social network (by association) among 62 dolphins in Doubtful Sound, New Zealand

50

Gephi demo: centrality

- Statistics tab
  - Average degree
  - Network diameter
  - Eigenvector centrality
- Appearance tab
  - Rank and color nodes according to centrality measures

51

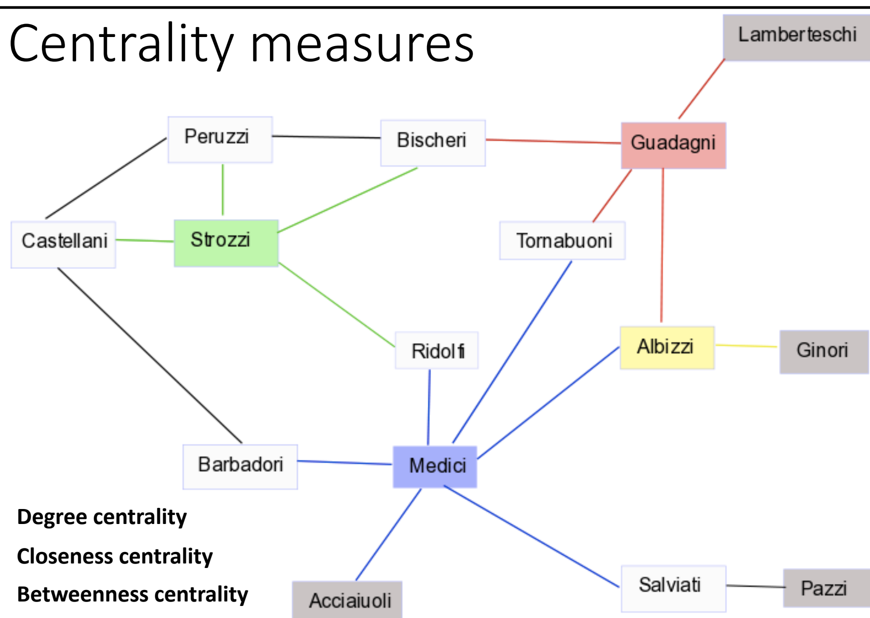
Caution: centrality

- 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-

52

Centrality measures



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

Idea | Math | Example

53

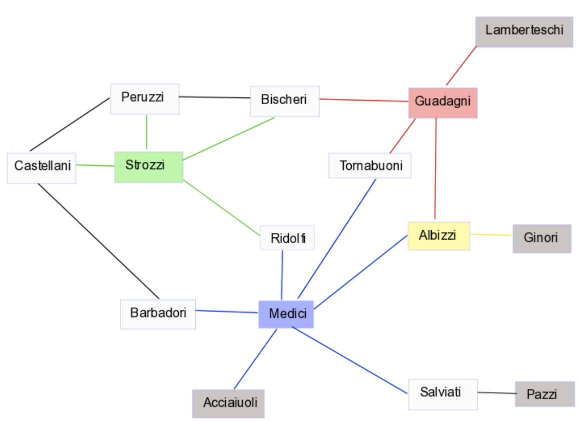
1. Degree centrality

Idea: Higher centrality nodes have higher degree

54

1. Degree centrality

Who is the most central here?

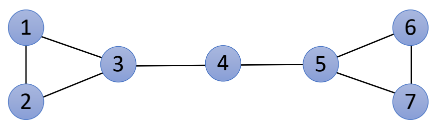


55

1. Degree centrality

Downside:

How about node 4 in this network?



56

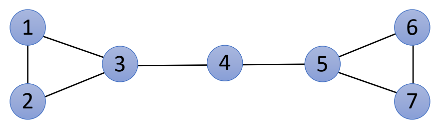
2. Closeness centrality

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

57

Example

Compute the closeness centralities of nodes 1 and 4



59

3. Betweenness centrality

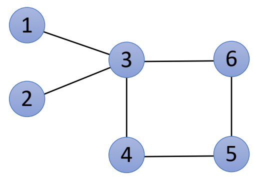
Idea: a node *i* is very central if a lot of shortest paths go through *i*

60

Example

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

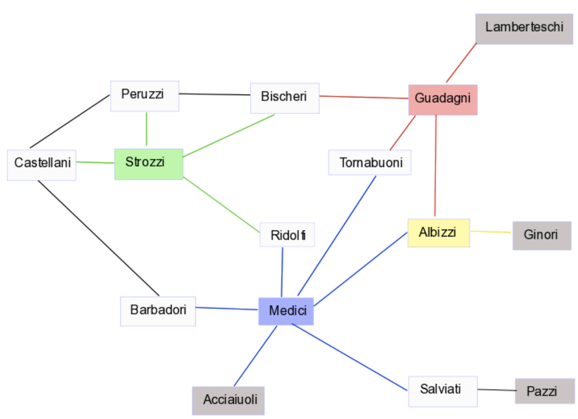
- $\beta_1 = 0$
- $\beta_2 = 0$
- $\beta_3 = ?$



62

Example

Highest betweenness centrality: Medici



64

Matrix algebra

- Images from this tutorial: http://www.intmath.com/matrices-determinants/3-matrices.php
- 4x1 matrix (AKA vector)

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

- 3x3 matrix

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

65

Graph example

Adjacency matrix

66

Matrix multiplication

- 2x3 matrix multiplied by 3x2 matrix

  must match

- Result is a 2x2 matrix

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

67

4. Eigenvector/prestige/power centrality

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

69

Perron-Frobenius Theorem

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

- Reassuring!

70

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:

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

$\lambda_1 \approx 2.17009$

$\lambda_2 \approx -1.48119$

$\lambda_3 = -1$

$\lambda_4 \approx 0.311108$

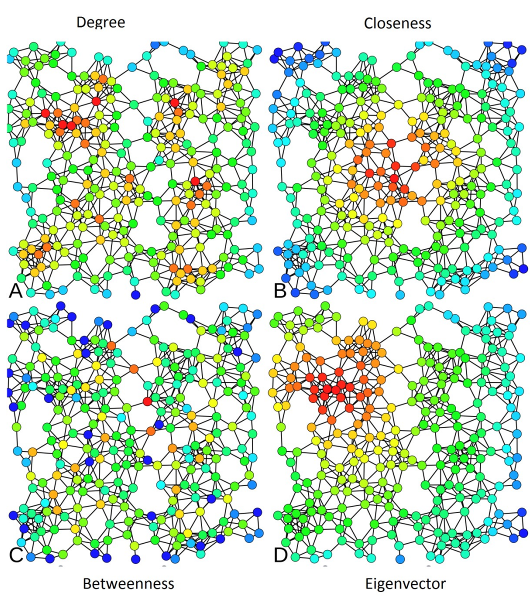
71

More on eigenvector centrality

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

72

Comparison of centrality measures



73

Comparison

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

76

Red: Graph level
Black: Node/edge level

Review: Gephi Vocabulary

Term	Meaning
betweenness centrality of a node	how often the node appears on the shortest path between nodes in the network
closeness centrality of a node	average distance from that node to all other nodes in the network
degree of a node	the number of edges connected to the node (also connectedness); in a directed graph a node can have in-degree and out-degree measures
diameter of a graph	the longest shortest path between any two nodes in the graph
directed graph	this means relationships occur one way only (I follow you, but you do not follow me on Twitter); opposite of undirected (we are friends with each other on Facebook)
eccentricity of a node	the distance (shortest-path length) from the node to the farthest node from it in the network
edge	a representation of the connection between two nodes, expresses a relationship (a line)
eigenvector centrality of a node	in social network analysis, a measure of influence (a node is very influential if it is connected to other influential nodes)
layout algorithms	also known as graph drawing algorithm; e.g., force-directed drawing where linked nodes attract and non-linked nodes repel
leaf node	node with a single edge in a “tree-structured” graph
modularity	a measure of connectedness among groups of nodes (greater than 0.4 is usually considered meaningful)
node	also called a vertex by mathematicians; a person in a social network graph (a dot or bubble)
distance from one node to another	the length of the shortest path (counted in the number of edges) from one node to another
path length	the number of edges in a path
singleton node or isolated node	node with no edge/connection

77