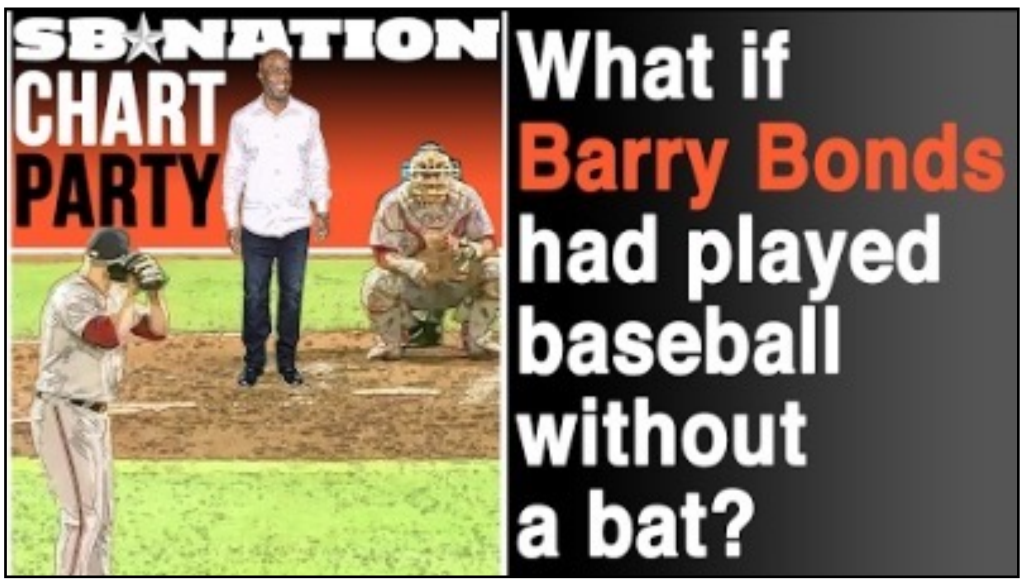
NETWORK ANALYSIS

CS 448B | Fall 2024

MANEESH AGRAWALA

1



2

ANNOUNCEMENTS

5

FINAL PROJECT

Design Reviews Dec 2 and Dec 4

Data analysis/explainer
- Analyze dataset in depth & make a visual explainer

Deliverables
- An article with multiple different interactive visualizations
- Short video (2 min) demoing and explaining the project

Schedule
- Project proposal: Today!
- Design Review and Feedback: 10th week of quarter, 12/2 and 12/4
- Final code and video: Sun 12/8 8pm

Grading
- Groups of up to 3 people, graded individually
- Clearly report responsibilities of each member

6

FINAL PROJECT GUIDELINES

Consider the audience
Your visual explainer should be of interest to a group of people beyond your immediate circle (an explainer about your own Spotify data unlikely be of interest to others you don't know)

Pick relatively less explored topics/datasets
Do some research on what has already been done for the topic/dataset(s)

Certain data like songs (e.g. Spotify) or movies (e.g. IMDB) are already well analyzed and should be avoided, unless you want to try to take a very different angle or use innovative analysis methods

Develop a narrative
In the early stages of the analysis process, try to uncover patterns to help you form and shape a narrative through-line for the explainer

7

FINAL PROJECT GUIDELINES

Design visualization interactions
Choose base visualizations that can support a high level of interactivity
- Bubble charts, tree maps, and word clouds typically aren't the most effective choices

Design interactive features that would enable viewers to interact with the data in a way that strengthens your narrative
- Tooltip is typically not enough interaction

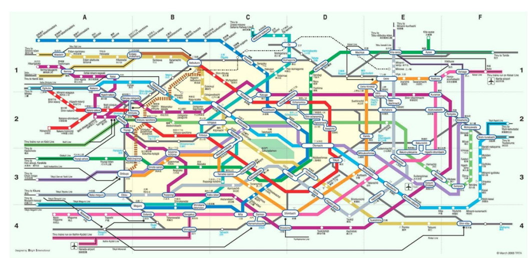
Draw inspiration from sites like the New York Times and the Pudding

8

NETWORK ANALYSIS

9

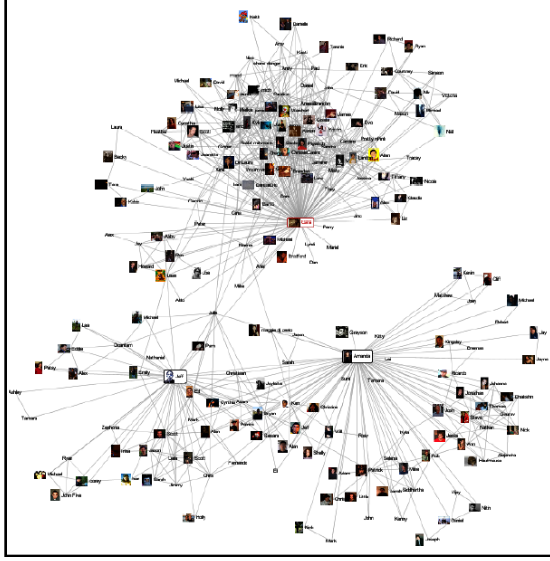
TRANSPORTATION



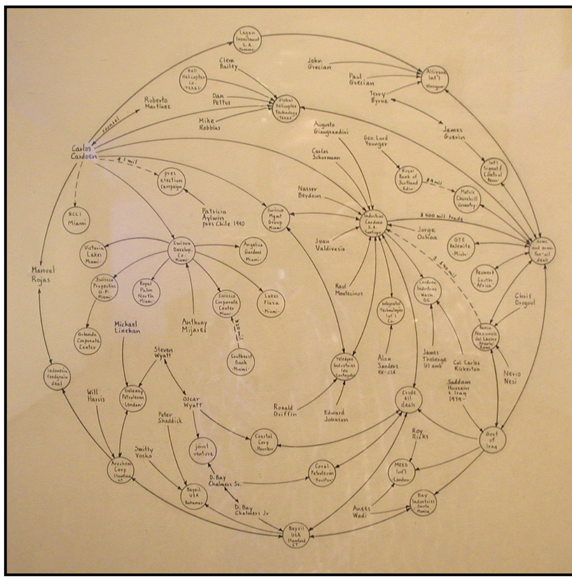
http://www.lx97.com/maps/

10

SOCIAL NETWORKS



12

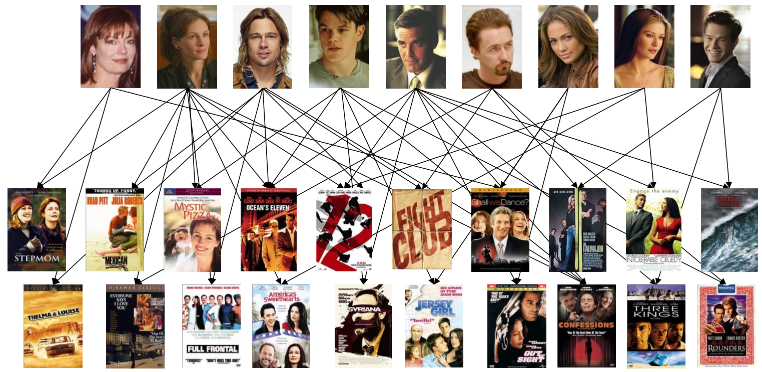


"Mark Lombardi (1951-2000) draws on the major political and financial scandals of the day to create large-scale linear diagrams that at first glance look like celestial maps; a closer reading reveals the intricate web of connections that lurk beneath current headlines.

Lombardi, M. 'George W. Bush, Harken Energy and Jackson Stephens, ca 1979–90'

13

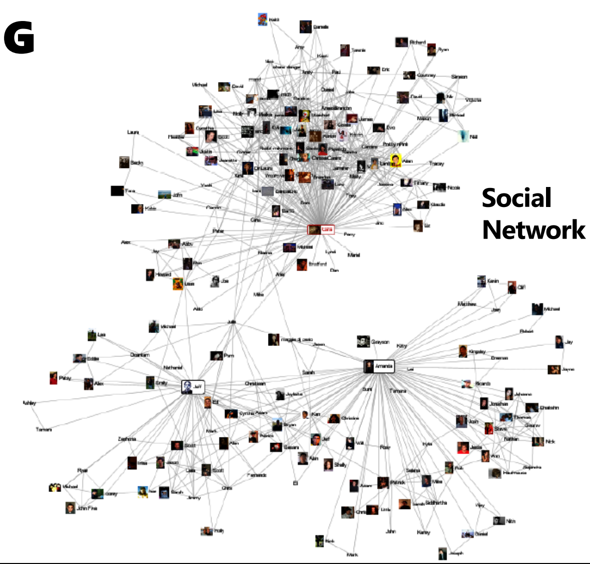
ACTORS & MOVIES (BIPARTITE)



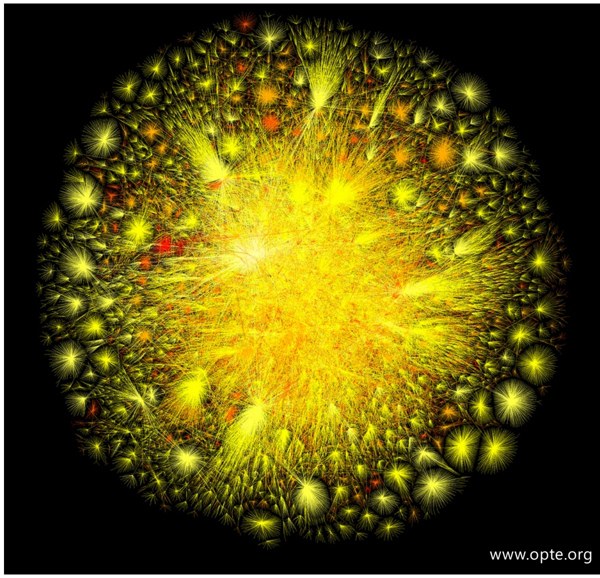
14

CHARACTERIZING NETWORKS

What does it look like?



17



Size?
Density?
Centrality?
Clustering?
Components?
Cliques?
Motifs?
Avg. path length?
...

18

TODAY

Learning Objectives

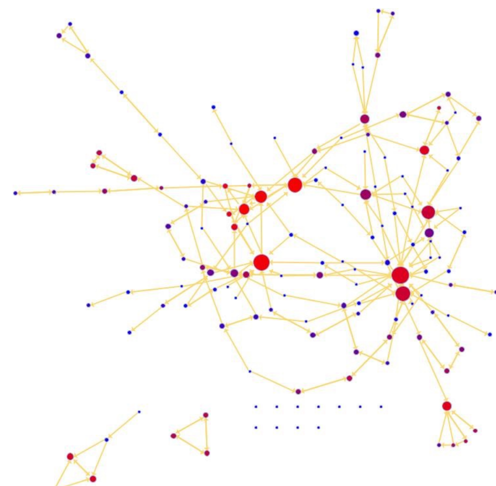
1. Measures of importance/centrality
2. Extracting community structure
3. Simulating network models

19

CENTRALITY

20

HOW FAR APART ARE THINGS?



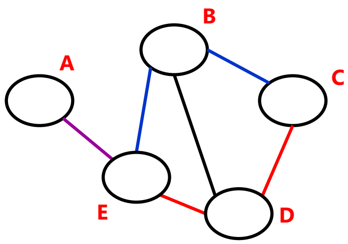
21

DISTANCE: SHORTEST PATHS

Shortest path (geodesic path)

The shortest sequence of links connecting two nodes
Not always unique

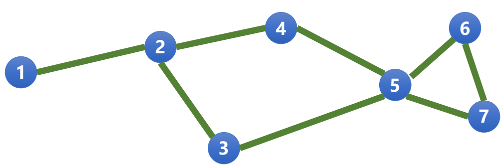
A and C are connected by 2 shortest paths
A – E – B - C
A – E – D - C



22

DISTANCE: SHORTEST PATHS

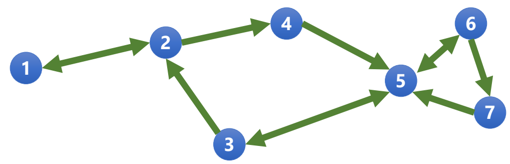
Shortest path from 2 to 3: 1



23

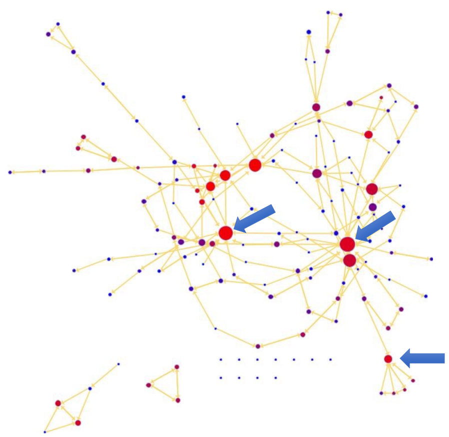
DISTANCE: SHORTEST PATHS

Shortest path from 2 to 3?



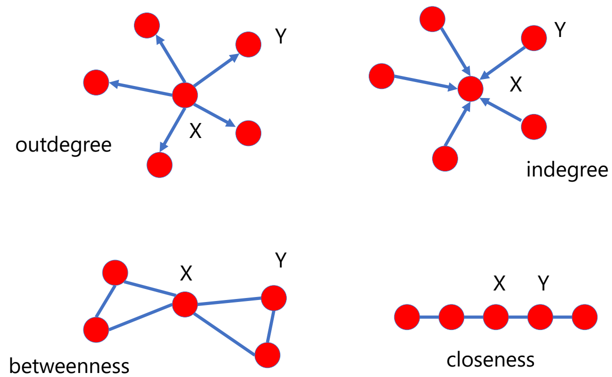
24

MOST IMPORTANT NODE



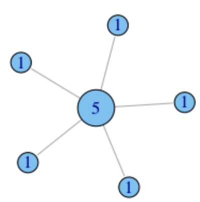
25

CENTRALITY



26

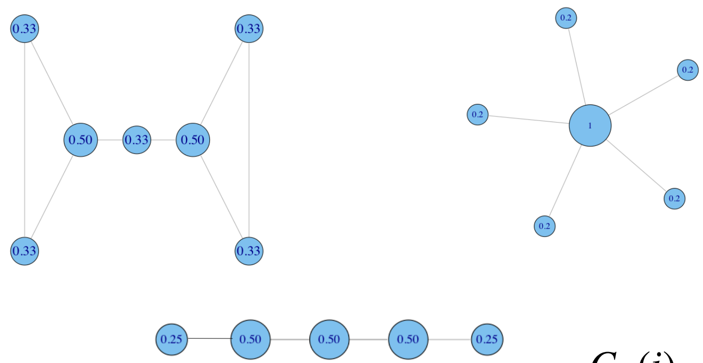
DEGREE CENTRALITY (UNDIRECTED)



$$C_D = d(i)$$

27

NORMALIZED DEGREE CENTRALITY



$$C_D(i) = \frac{d(i)}{N-1}$$

28

WHEN IS DEGREE NOT SUFFICIENT?

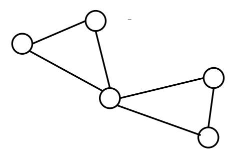
Does not capture

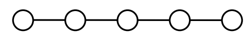
- Ability to broker between groups
- Likelihood that information originating anywhere in the network reaches you

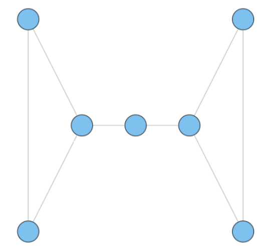
29

BETWEENESS

Assuming nodes communicate using the most direct (shortest) route, how many pairs of nodes have to pass information through target node?

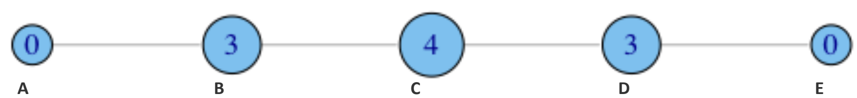


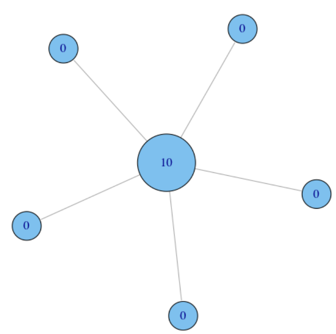




30

BETWEENESS - EXAMPLES





31

BETWEENESS - DEFINITION

$$C_B(i) = \sum_{j,k \neq i, j<k} g_{jk}(i) / g_{jk}$$

g_{jk} = the number of shortest paths connecting *jk*
$g_{jk}(i)$ = the number of shortest paths containing *i*.

Normalization:

$$C_B^{'}(i) = C_B(i\ ) / [(n-1)(n-2)/2]$$

number of pairs of vertices
excluding the vertex itself

32

WHEN ARE C_d, AND C_b NOT SUFFICIENT?

Does not capture
Likelihood that information originating anywhere in the network reaches you

34

CLOSENESS - DEFINITION

e.g., which node is closest to the *center* of the graph

Closeness Centrality:

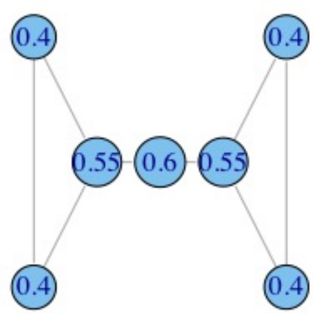
$$C_c(i) = \left[\sum_{j=1, j\neq i}^{N} d(i,j) \right]^{-1}$$

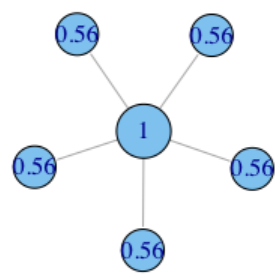
Normalized Closeness Centrality

$$C_C^{'}(i) = (C_C(i)) / (N-1) = \frac{N-1}{\sum_{j=1, j\neq i}^{N} d(i,j)}$$

35

CLOSENESS - EXAMPLES



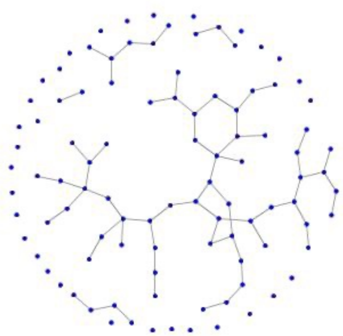


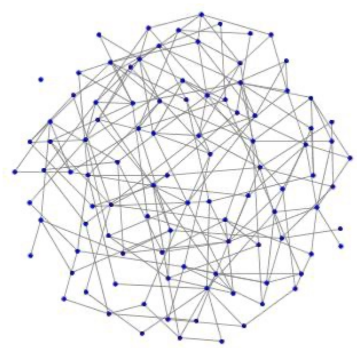
36

COMMUNITY STRUCTURE

46

HOW DENSE IS IT?





density = e / e_{max}

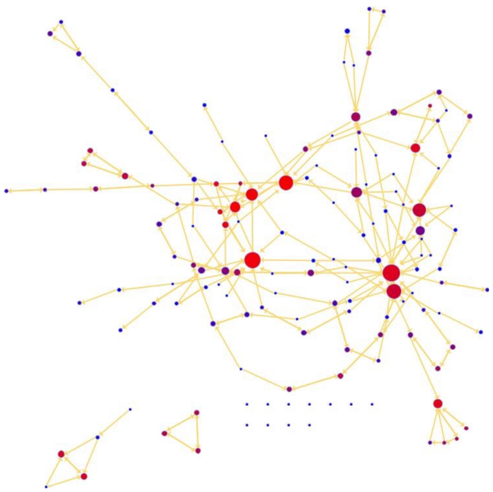
Max. possible edges:

Directed: $e_{max} = n*(n-1)$

Undirected: $e_{max} = n*(n-1)/2$

47

IS EVERYTHING CONNECTED?



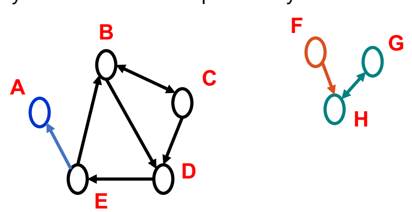
48

CONNECTED COMPONENTS - DIRECTED

Strongly connected components

Each node in component can be reached from every other node in component by following directed links

B C D E
A
G H
F



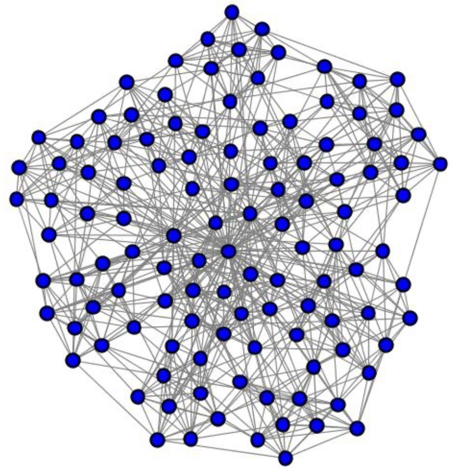
Weakly connected components

Each node can be reached from every other node by following links in either direction

A B C D E
G H F

49

COMMUNITY FINDING - CLUSTERING



52

HIERARCHICAL CLUSTERING

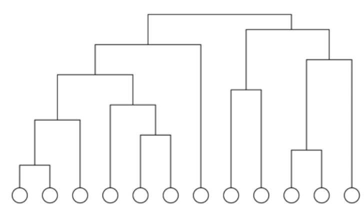
Process

Calculate affinity weights *W* for all pairs of vertices

Start: *N* disconnected vertices

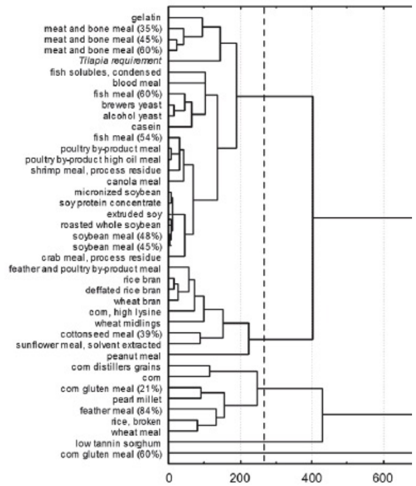
Add edges (one by one) between pairs of vertices/clusters in order of decreasing weight (use closest distance to compare clusters)

Result: nested components



53

CLUSTER DENDROGRAMS

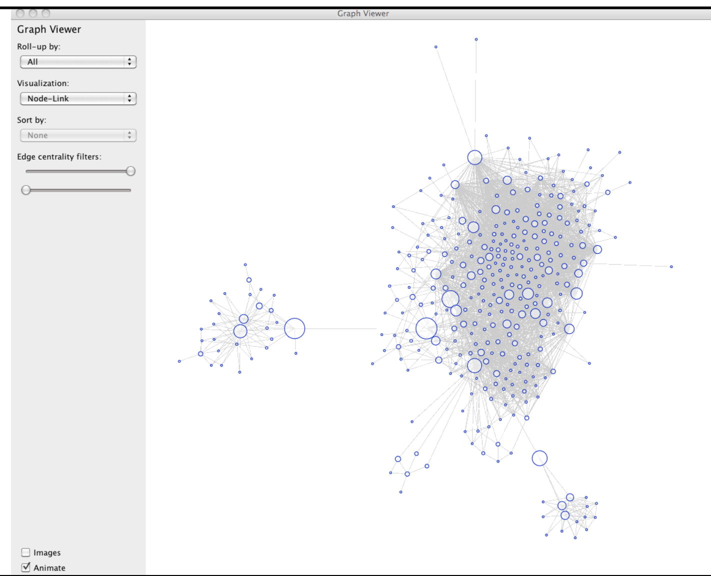


Depicts cluster trees produced by hierarchical clustering algorithms

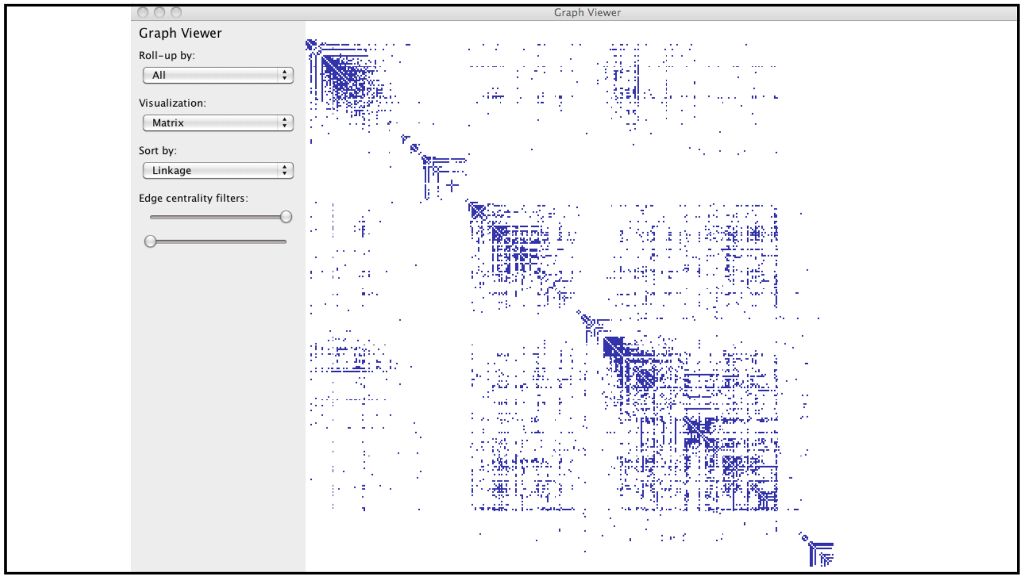
Leaf nodes arranged in line, internal node depth indicates order/value at which clusters merge

Basic recursive layout with orthogonal two-segment edges

54



55



56

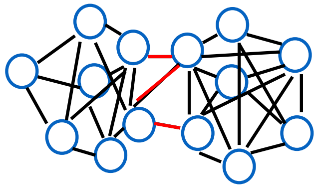
BETWEENESS CLUSTERING

Girvan and Newman 2002 iterative algorithm:

Compute C_b of all *edges*

Remove edge i where $C_b(i) == max(C_b)$

Recalculate betweenness



59

SIMULATING NETWORK MODELS

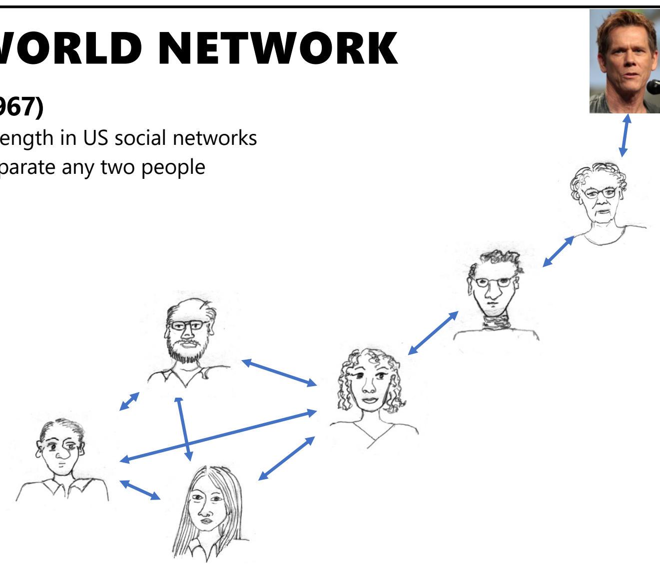
77

SMALL WORLD NETWORK

Milgram (1967)

Mean path length in US social networks

~ 6 hops separate any two people

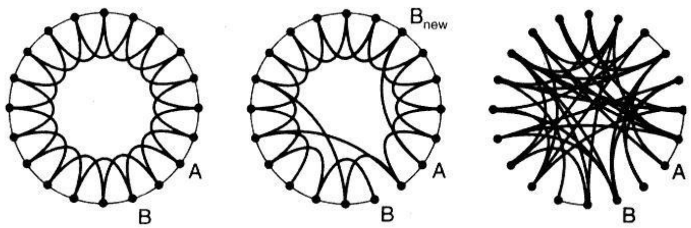


78

SMALL WORLD NETWORK

Watts and Strogatz 1998

a few random links in otherwise structured graph make network a small world



regular lattice: my friend's friend is always my friend

small world: mostly structured with a few random connections

random graph: all connections random

79

DEFINING SMALL WORLD PHENOMENA

Properties

high clustering

low mean shortest path

$$C_{\text{network}} >> C_{\text{random graph}}$$

$$l_{\text{network}} \approx \ln(N)$$

Examples

neural network of C. elegans

semantic networks of languages

actor collaboration graph

food webs

80

SUMMARY

Structural analysis
- Centrality
- Community structure

Simulation models enable further analysis

Network analysis applicable in many domains

85