Network Analysis

Maneesh Agrawala

CS 448B: Visualization
Fall 2021

1

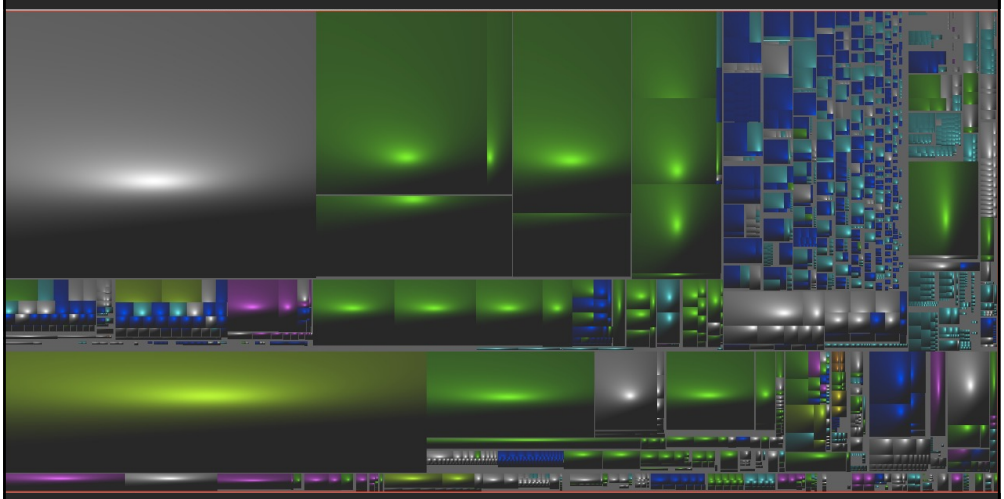
Last Time: Network Layout

2

***Squarified* layout: Try to produce square (1:1) aspect ratios**

3

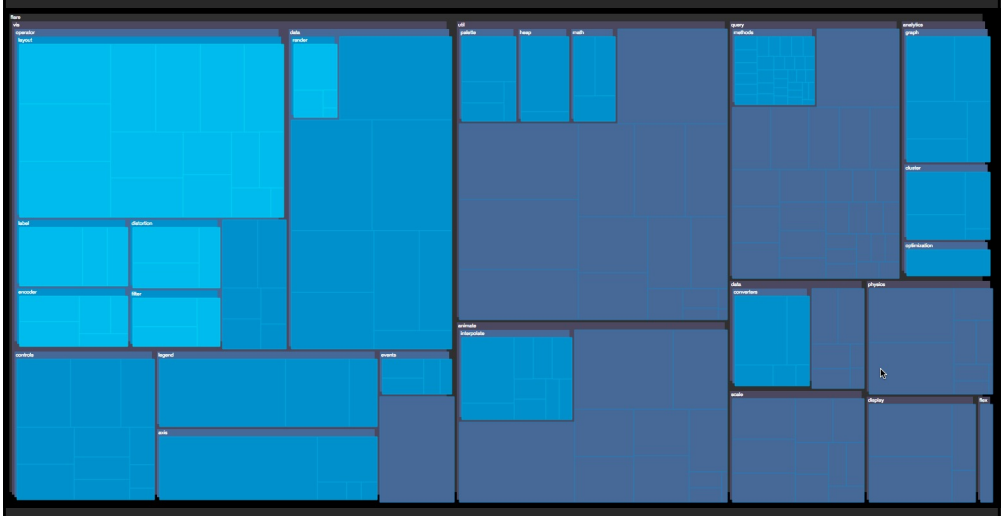
Cushion Treemaps [van Wijk 99]



Use shading to emphasize hierarchical structure

4

Cascaded Treemaps [Lü 08]



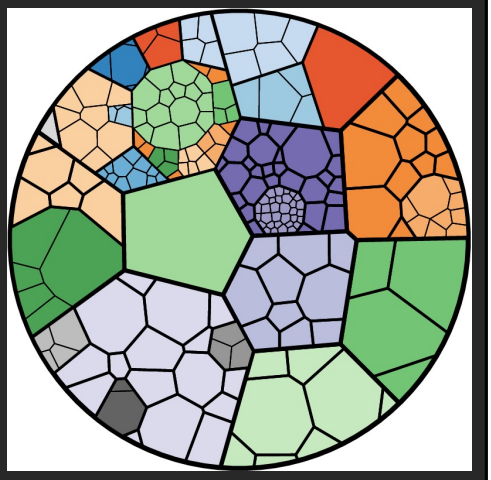
Use 2.5D effect emphasize hierarchical structure

5

Voronoi Treemaps [Balzer 05]

Treemaps with arbitrary polygonal shape and boundary

Uses iterative, weighted Voronoi tessellations to achieve cells with value-proportional areas



6

Node-Link Graph Layout

7

Spanning Tree Layout

Many graphs are tree-like or have useful spanning trees

- Websites, Social Networks

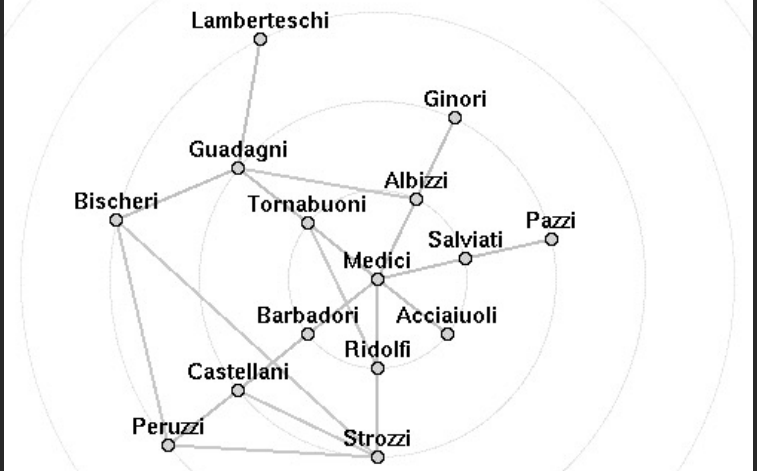
Use tree layout on spanning tree of graph

- Trees created by BFS / DFS
- Min/max spanning trees

Fast tree layouts allow graph layouts to be recalculated at interactive rates

Heuristics may further improve layout

9



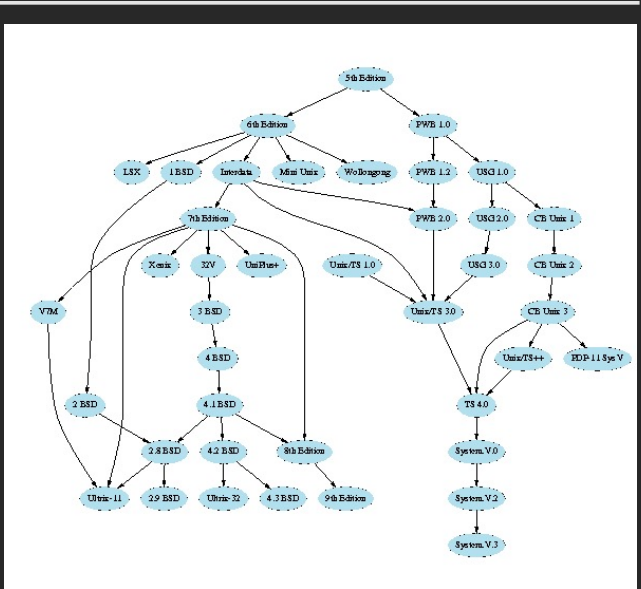
Spanning tree layout may result in arbitrary parent node

10

Sugiyama-style graph layout

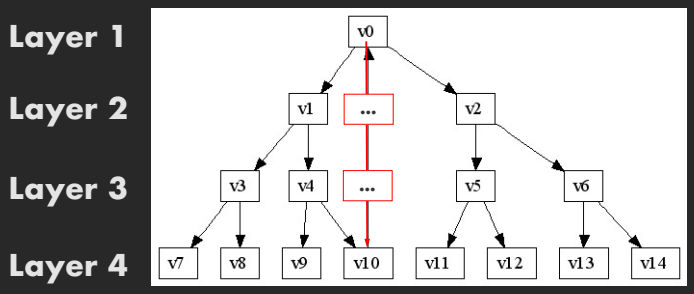
Evolution of the UNIX operating system

Hierarchical layering based on descent



11

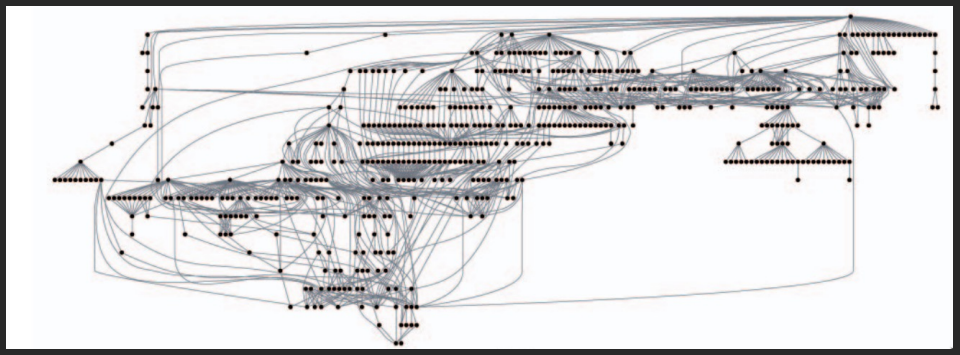
Sugiyama-style graph layout



Reverse some edges to remove cycles
Assign nodes in hierarchy layers → Longest path layering
    Create dummy nodes to "fill in" missing layers
Arrange nodes within layer, minimize edge crossings
    Route edges – layout splines if needed

12

Produces hierarchical layout



Sugiyama-style layout emphasizes hierarchy

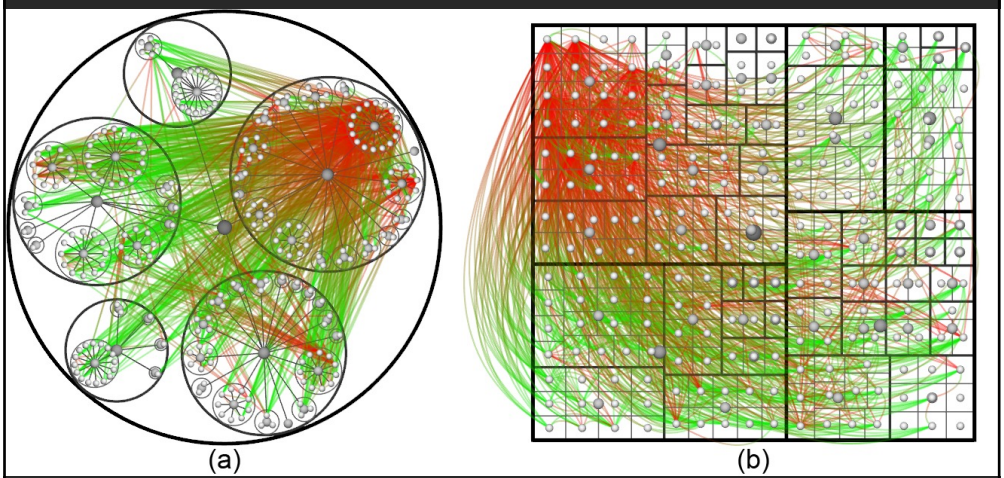
However, cycles in the graph may mislead.
Long edges can impede perception of proximity.

13

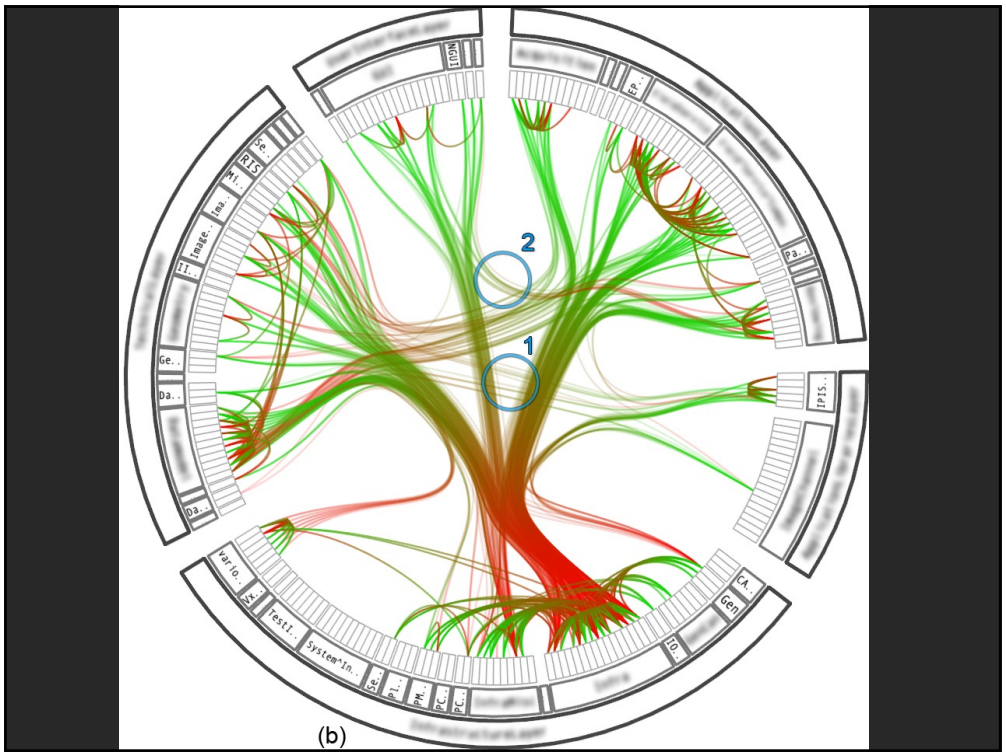
Hierarchical Edge Bundles

14

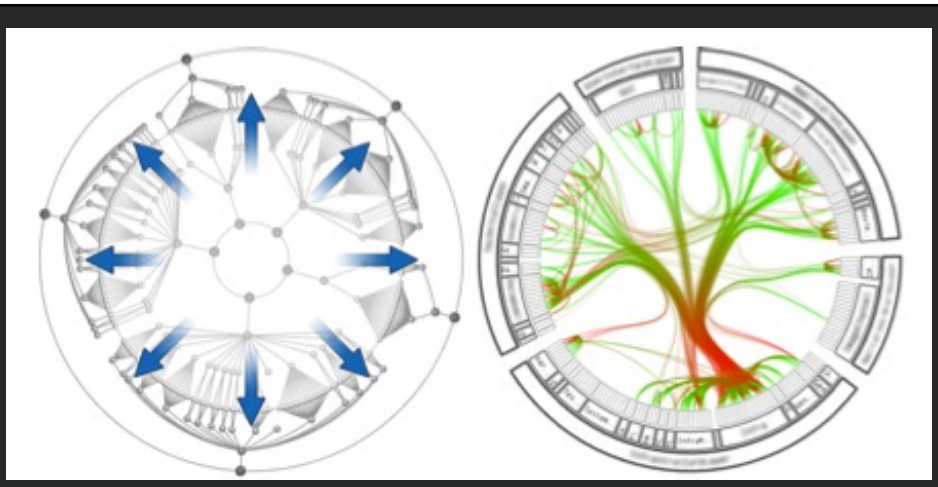
Trees with Adjacency Relations



15



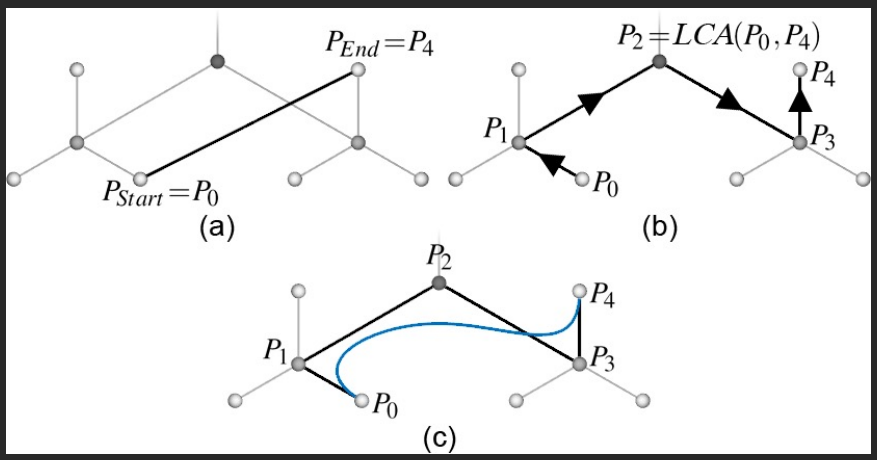
16



Use radial tree layout for inner circle
Mirror to outside
Replace inner tree with hierarchical edge bundles

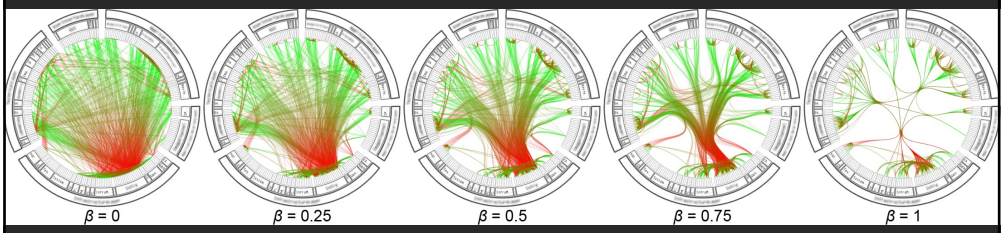
17

Bundle Edges along Hierarchy



18

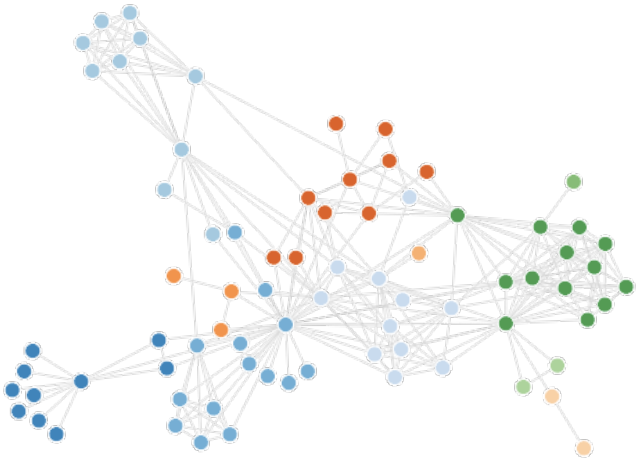
Increasing Edge Tension



20

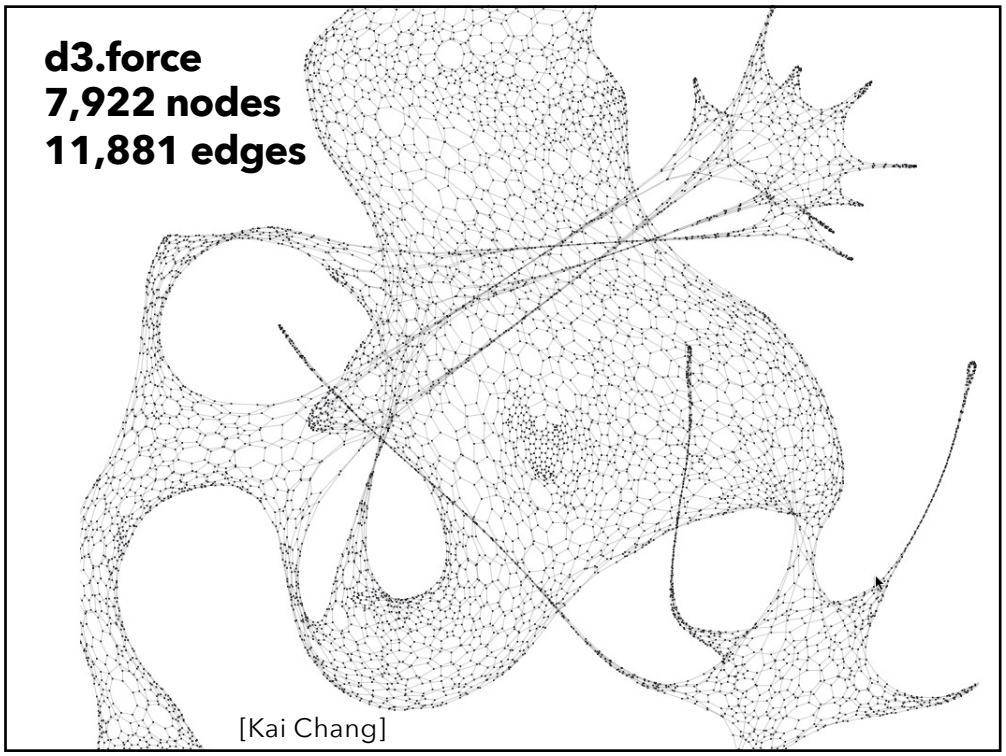
Force-Directed Layout

22



Interactive Example: Configurable Force Layout

23



24

Force-Directed Layout

Nodes = charged particles with air resistance — $F = q_i * q_j / d_{ij}^2$

Edges = springs — $F = -b * v_i$; $F = k * (L - d_{ij})$

D3's force layout uses velocity Verlet integration

Assume uniform mass *m* and timestep Δt:

$F = ma \rightarrow F = a \rightarrow F = \Delta v / \Delta t \rightarrow F = \Delta v$

Forces simplify to velocity offsets!

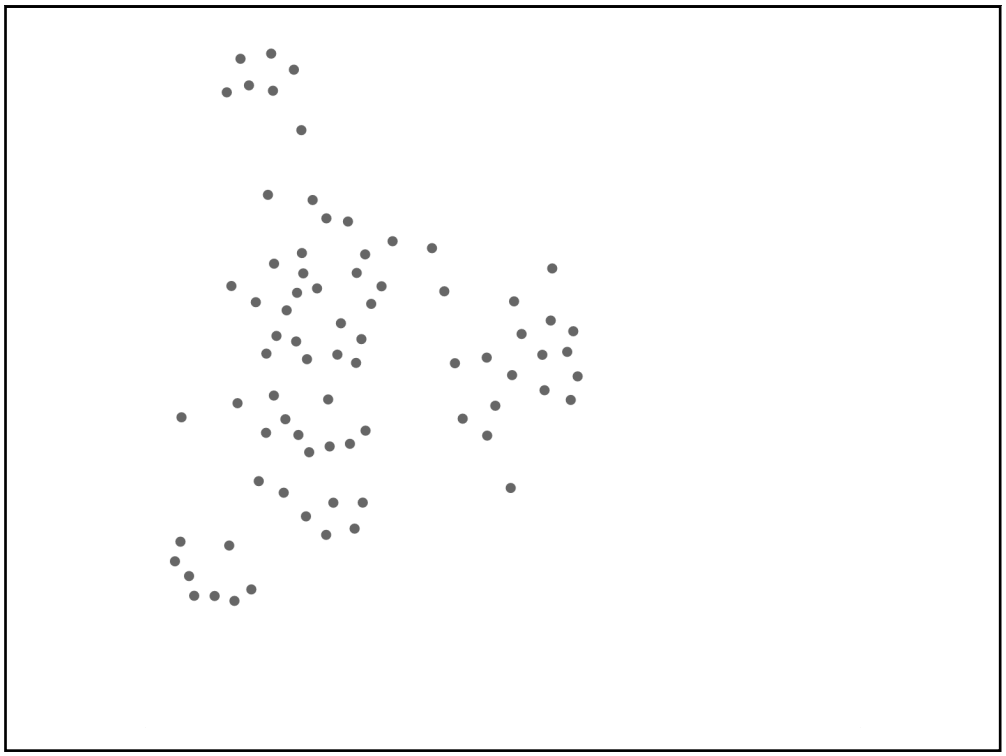
Repeatedly calculate forces, update node positions

Naïve approach $O(N^2)$

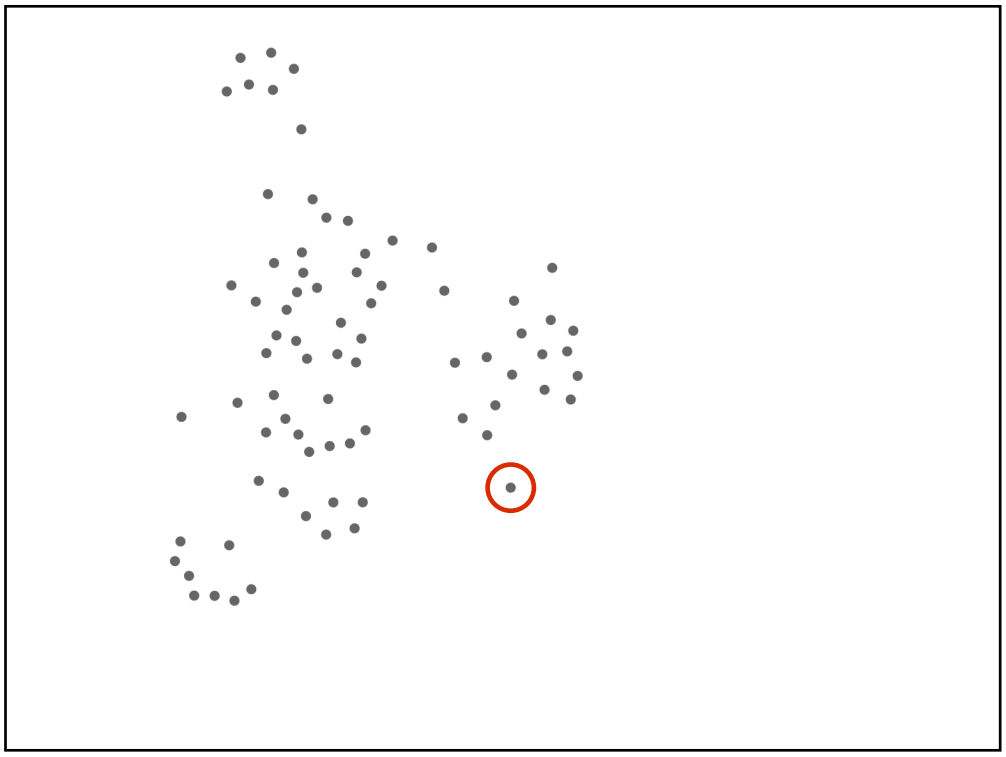
Speed up to O(N log N) using quadtree or k-d tree

Numerical integration of forces at each time step

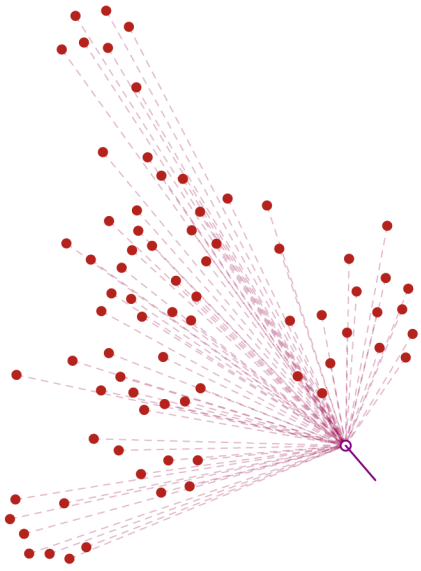
25



26

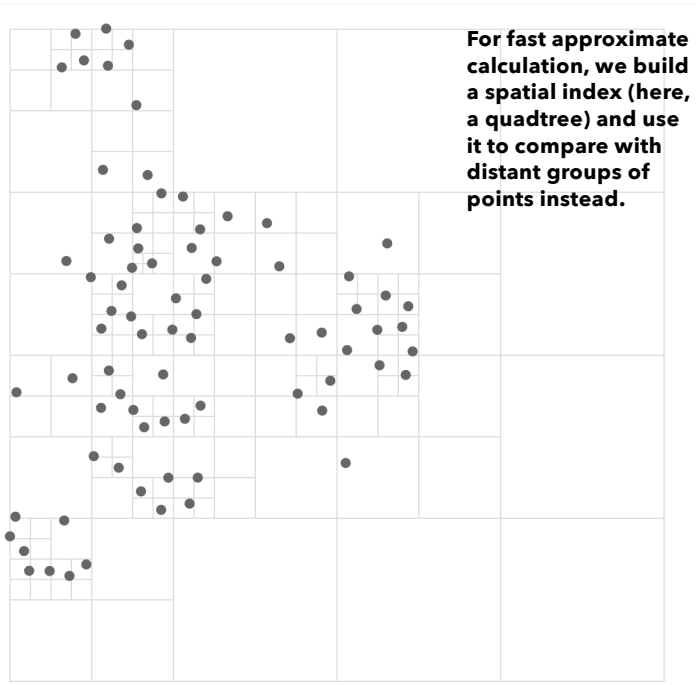


27



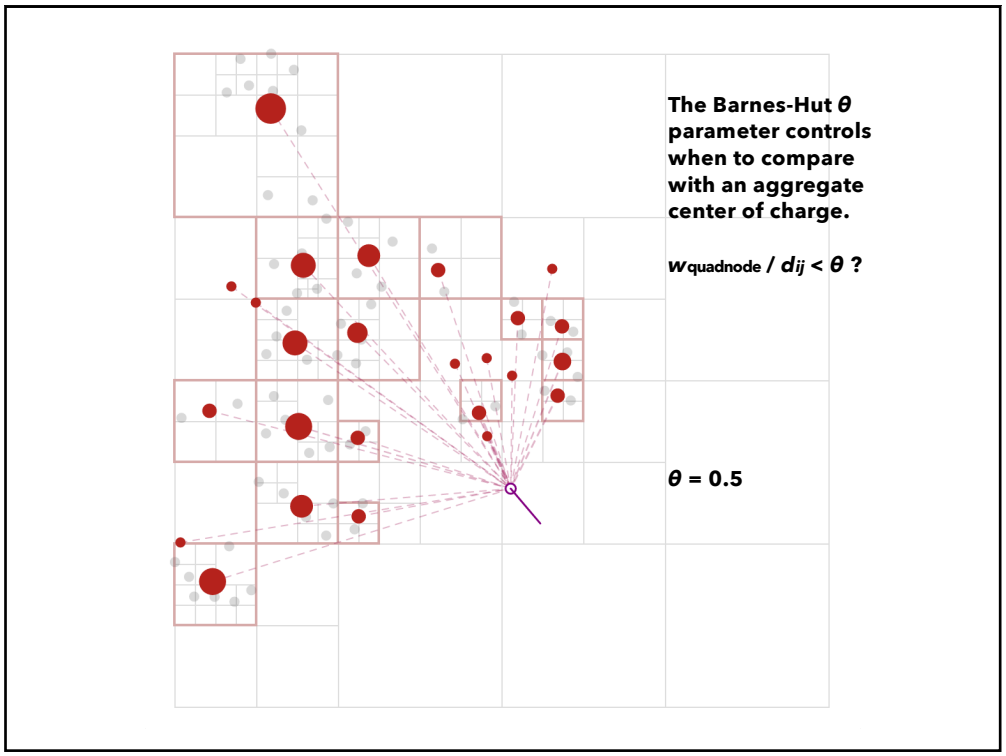
Naive calculation of forces at a point uses sum of forces from all other n-1 points.

28

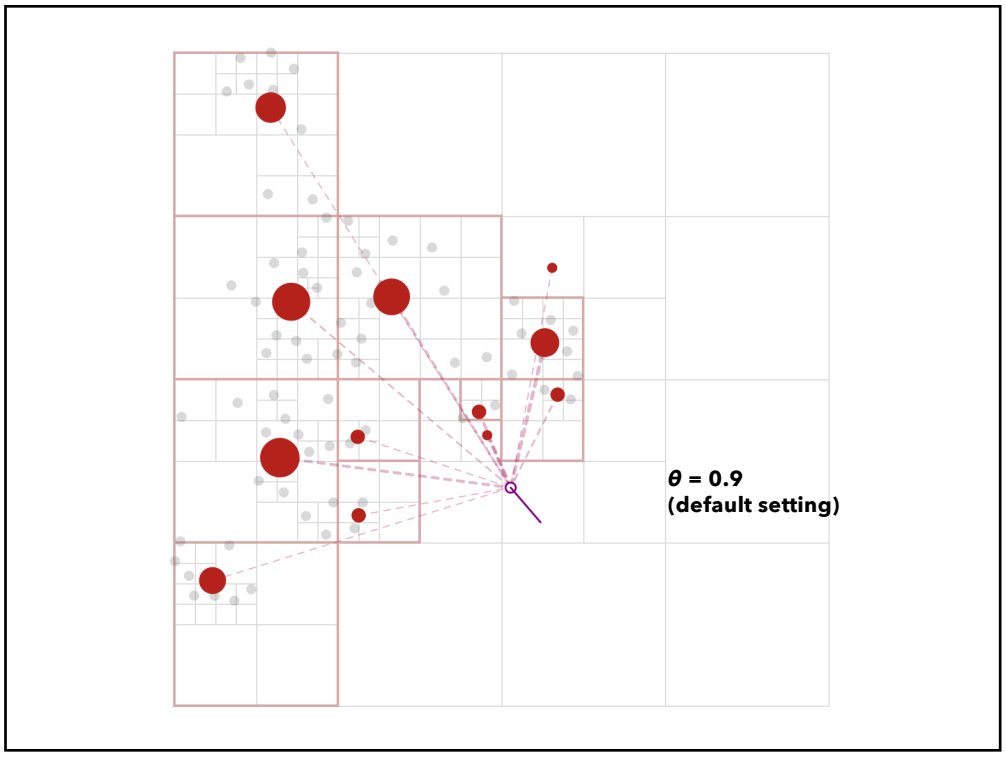


For fast approximate calculation, we build a spatial index (here, a quadtree) and use it to compare with distant groups of points instead.

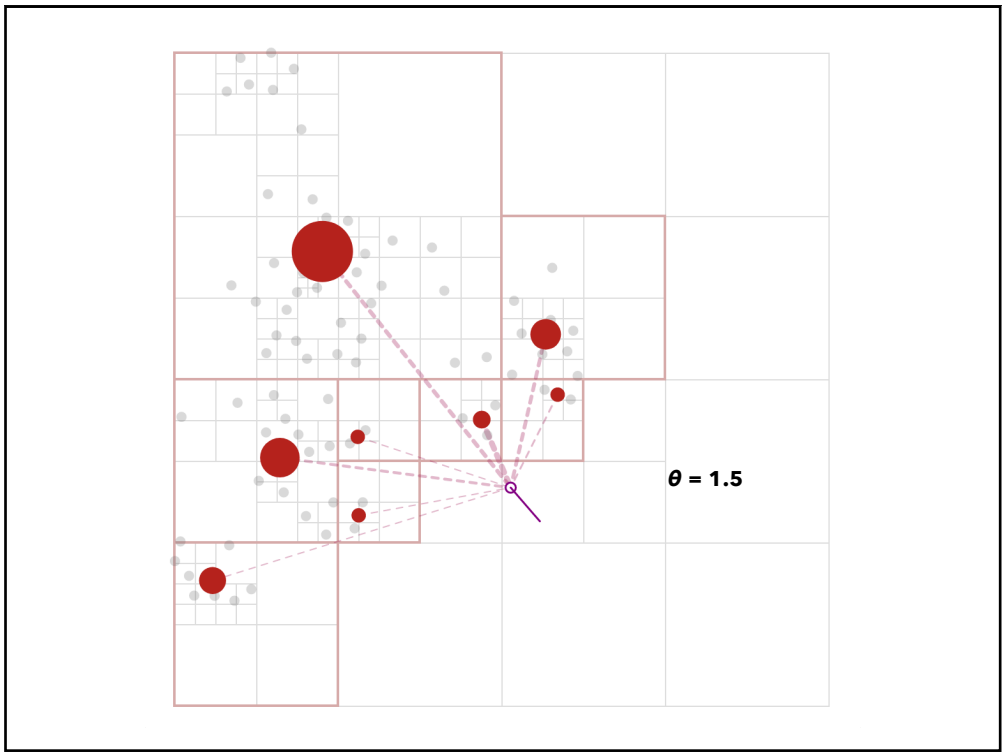
29


The Barnes-Hut θ parameter controls when to compare with an aggregate center of charge.
$w_{quadnode} / d_{ij} < \theta$?
θ = 0.5

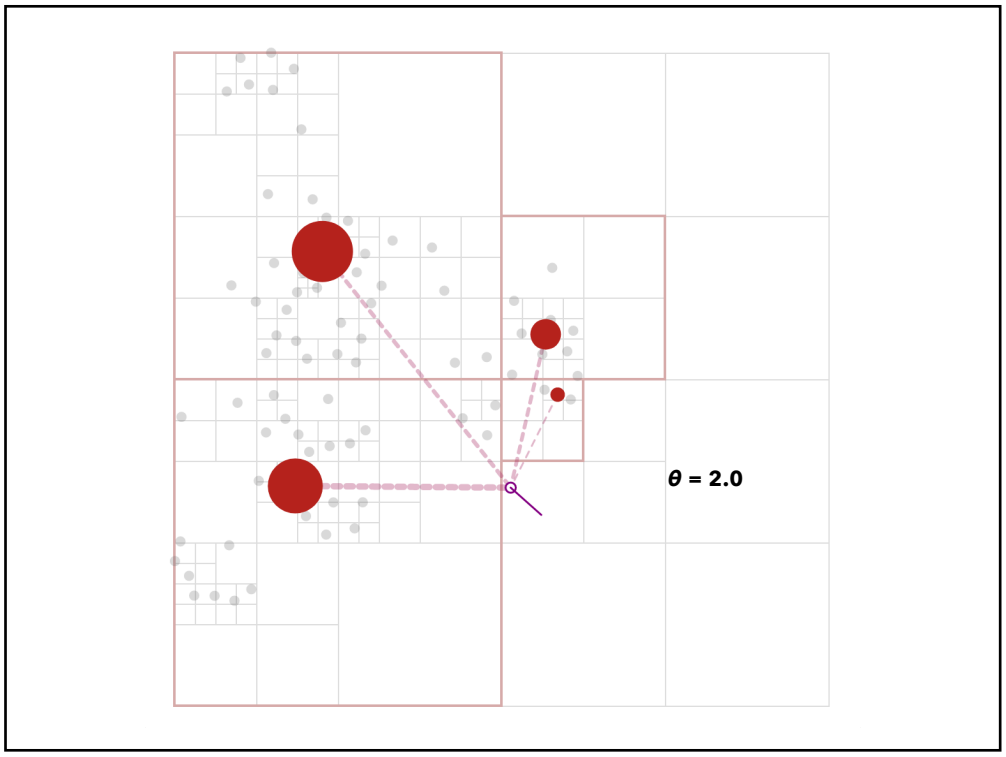
30


θ = 0.9
(default setting)

31



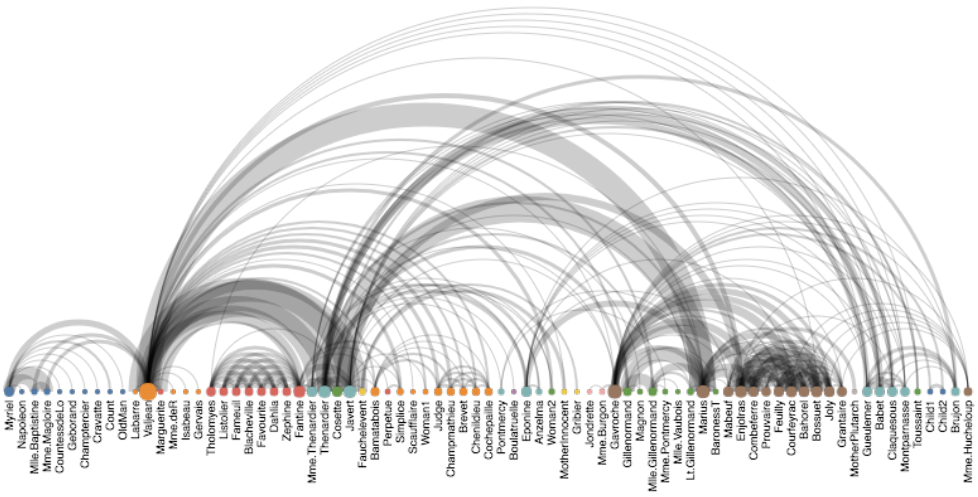
32



33

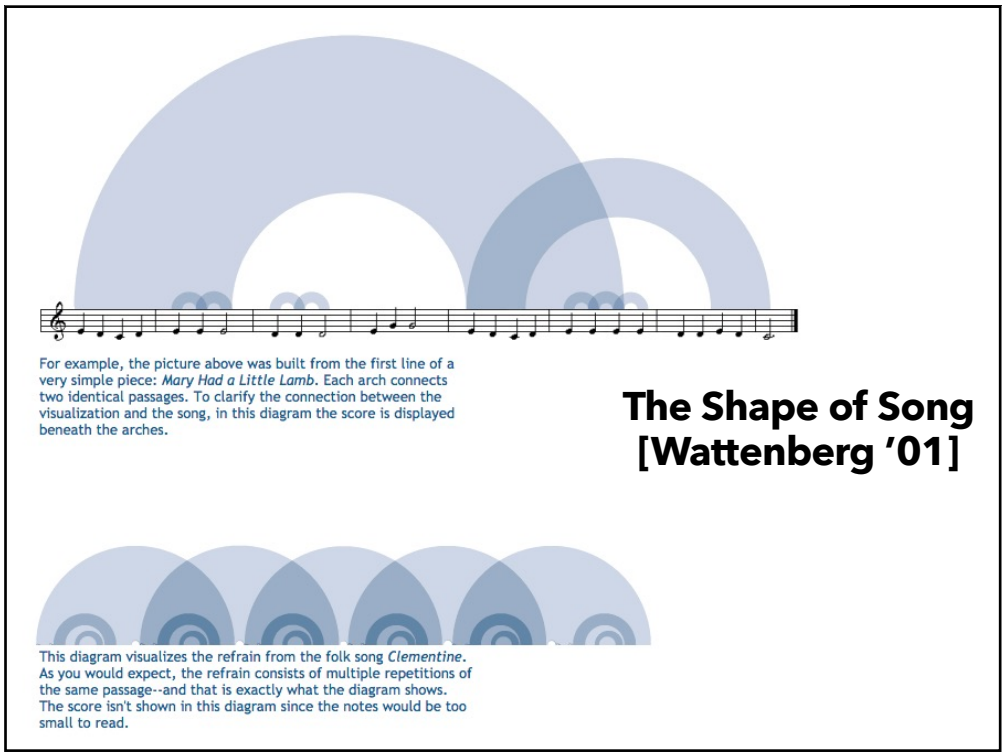
Alternative Layouts

35

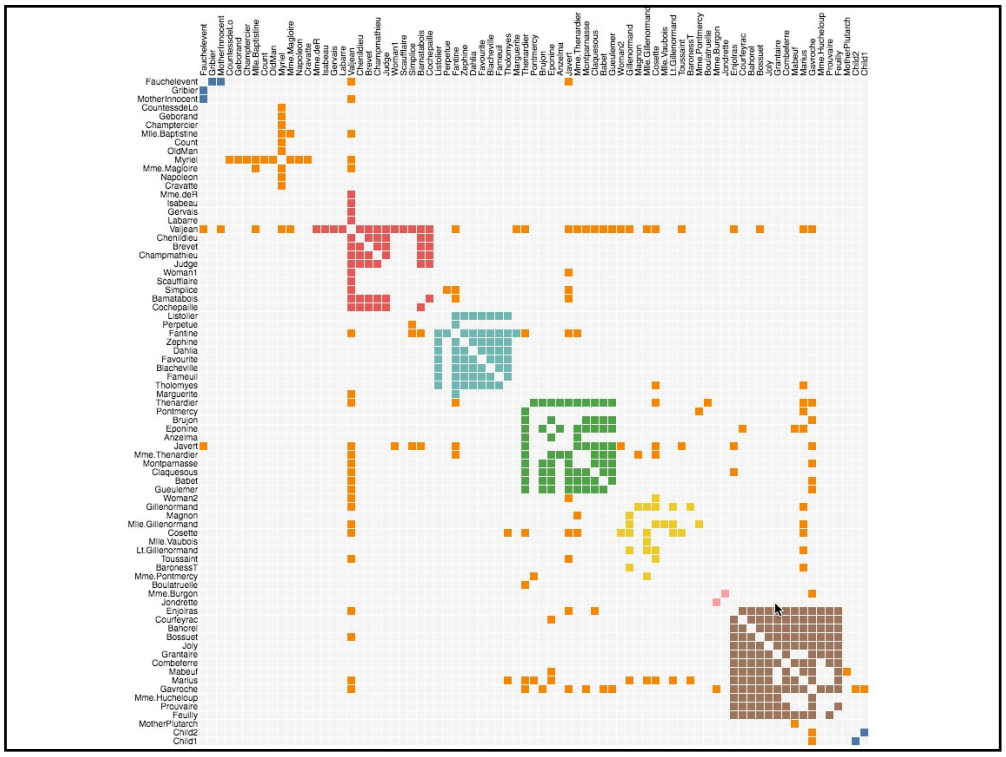


Linear node layout, circular arcs show connections. Layout quality sensitive to node ordering!

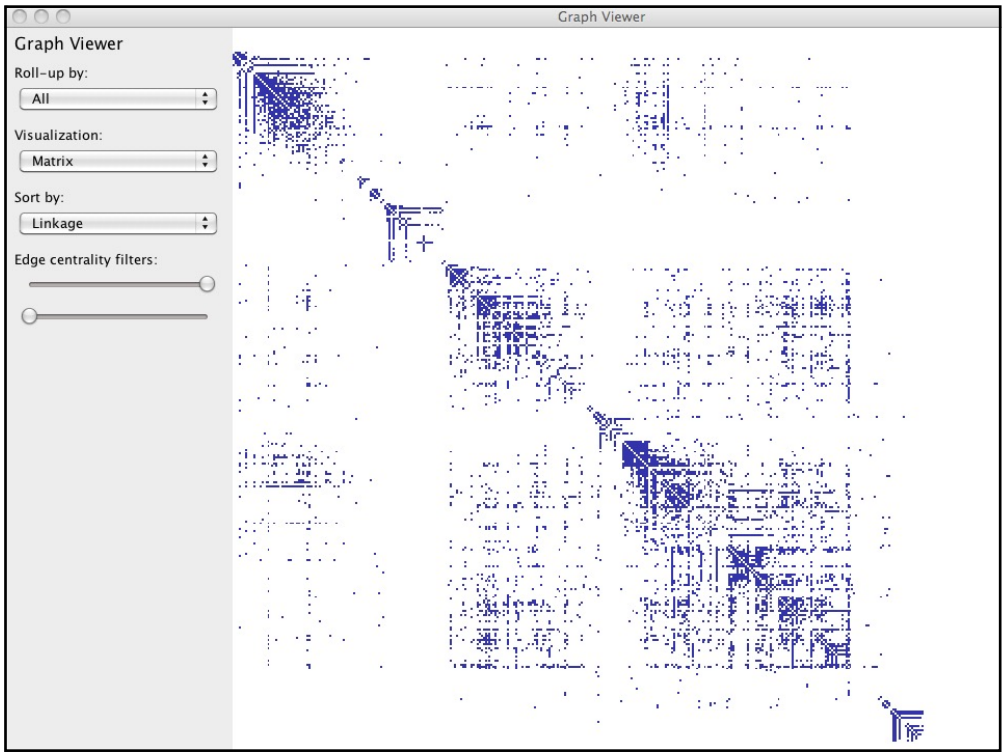
36


For example, the picture above was built from the first line of a very simple piece: *Mary Had a Little Lamb*. Each arch connects two identical passages. To clarify the connection between the visualization and the song, in this diagram the score is displayed beneath the arches.
The Shape of Song [Wattenberg '01]
This diagram visualizes the refrain from the folk song *Clementine*. As you would expect, the refrain consists of multiple repetitions of the same passage--and that is exactly what the diagram shows. The score isn't shown in this diagram since the notes would be too small to read.

37


Fauchelevent
Gribier
MotherInnocent
CountessdeLo
Geborand
Champtercier
Mlle.Baptistine
Count
OldMan
Myriel
Mme.Magloire
Napoleon
Cravatte
Mme.deR
Isabeau
Gervais
Labarre
Valjean
Chenildieu
Brevet
Champmathieu
Judge
Woman1
Scaufflaire
Simplice
Bamatabois
Cochepaille
Listolier
Perpetue
Fantine
Zephine
Dahlia
Favourite
Blacheville
Fameuil
Tholomyes
Marguerite
Thenardier
Pontmercy
Brujon
Eponine
Anzelma
Javert
Mme.Thenardier
Montparnasse
Claquesous
Babet
Gueulemer
Woman2
Gillenormand
Magnon
Mlle.Gillenormand
Cosette
Mlle.Vaubois
Lt.Gillenormand
Toussaint
BaronessT
Mme.Pontmercy
Boulatruelle
Mme.Burgon
Jondrette
Emporas
Courfeyrac
Bahorel
Bossuet
Joly
Grantaire
Combeferre
Mabeuf
Marius
Gavroche
Mme.Hucheloup
Prouvaire
Feuilly
MotherPlutarch
Child2
Child1

38



39

Attribute-Driven Layout

Large node-link diagrams get messy!
Is there additional structure we can exploit?

Idea: Use data attributes to perform layout

- e.g., scatter plot based on node values

Dynamic queries and/or brushing can be used to explore connectivity

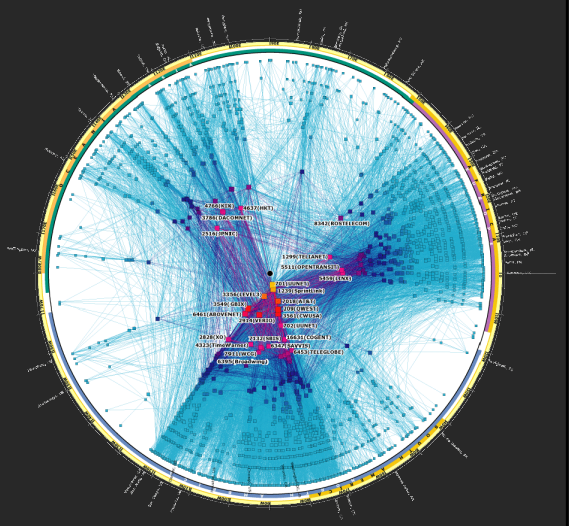
40

Attribute-Driven Layout

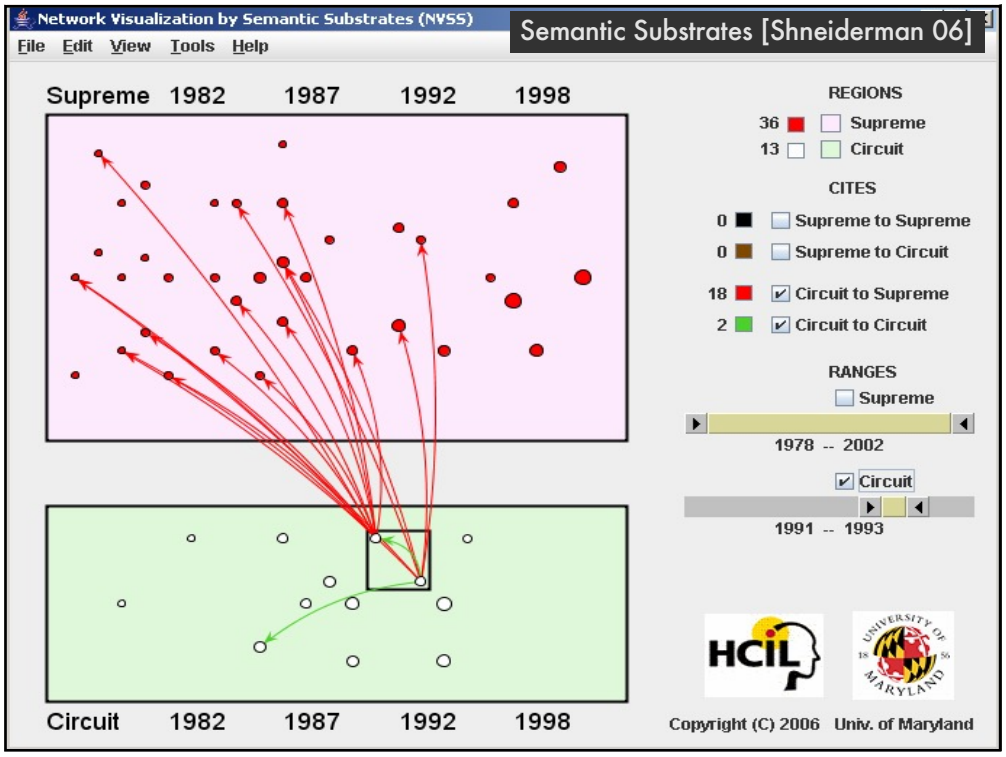
The "Skitter" Layout
- Internet Connectivity
- Radial Scatterplot

Angle = Longitude
- Geography

Radius = Degree
- # of connections
- (a statistic of the nodes)



41



42

Summary

Tree Layout

Indented / Node-Link / Enclosure / Layers

How to address issues of scale?

- Filtering and Focus + Context techniques

Graph Layout

Tree layout over spanning tree

Hierarchical “Sugiyama” Layout

Optimization (Force-Directed Layout)

Attribute-Driven Layout

43

Announcements

44

Final project

Data analysis/explainer or conduct research

- **Data analysis**: Analyze dataset in depth & make a visual explainer
- **Research**: Pose problem, Implement creative solution

Deliverables

- **Data analysis/explainer**: Article with multiple different interactive visualizations
- **Research**: Implementation of solution and web-based demo if possible
- **Short video (2 min)** demoing and explaining the project

Schedule

- Project proposal: Wed 11/3
- Design Review and Feedback: 10th week of quarter
- Final code and video: Fri 12/10 11:59pm

Grading

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

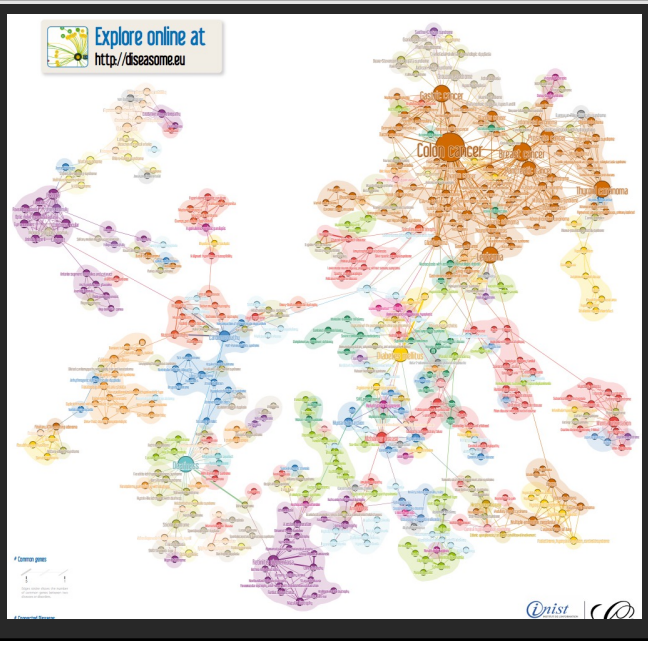
45

Network Analysis

*Slides adapted from E. Adar's / L. Adamic's Network Theory and Applications course slides.

46

Diseases

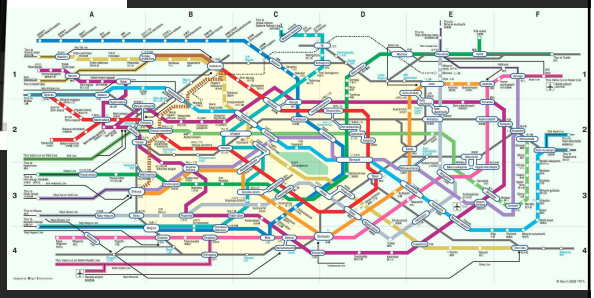


http://diseasome.eu/

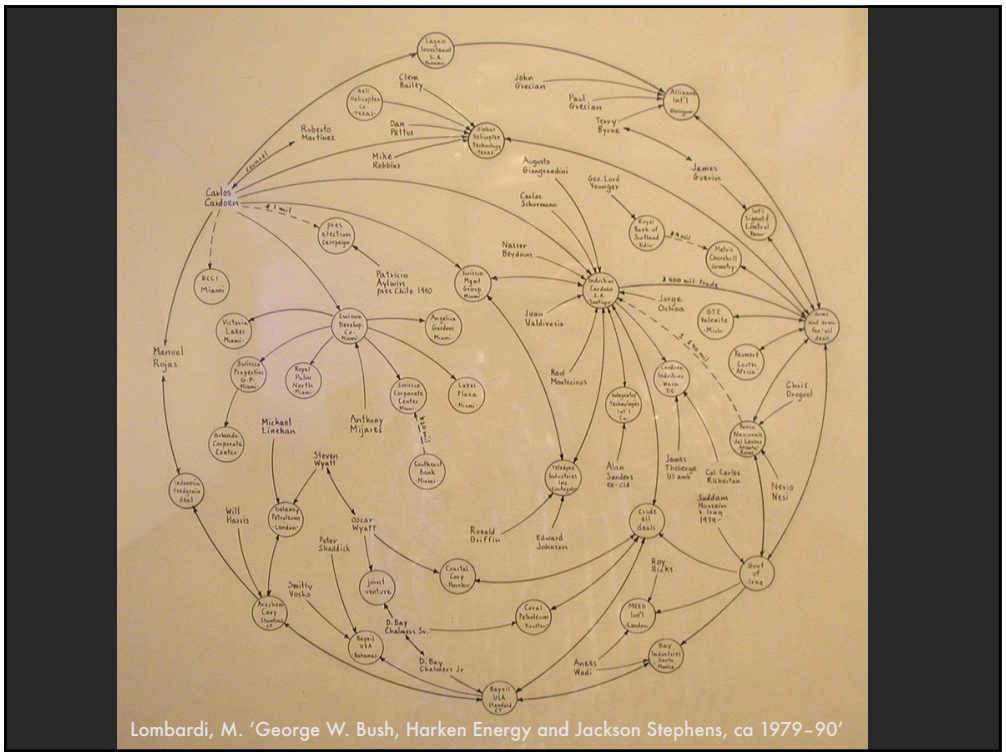
48

Transportation

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



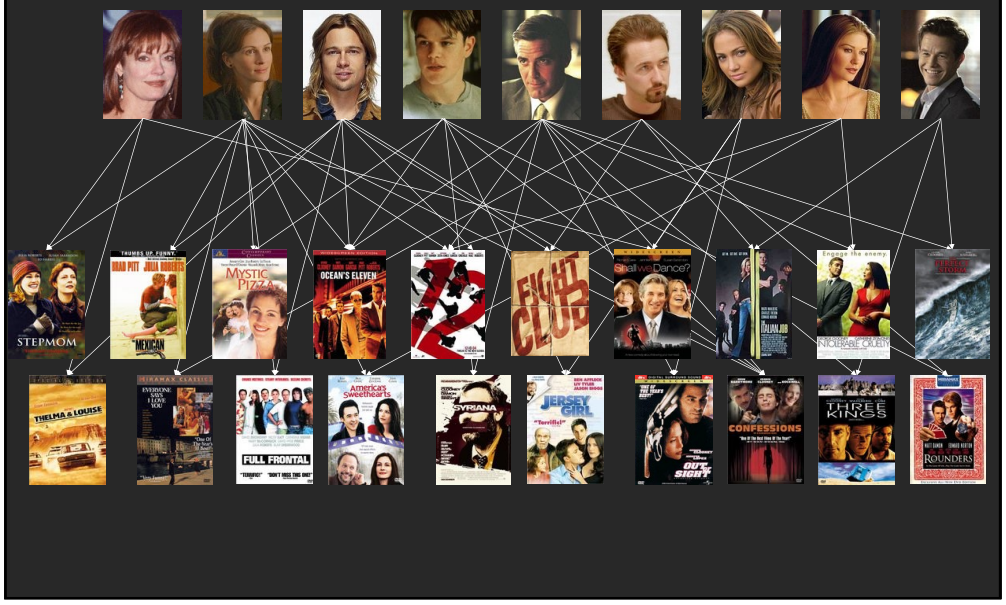
49



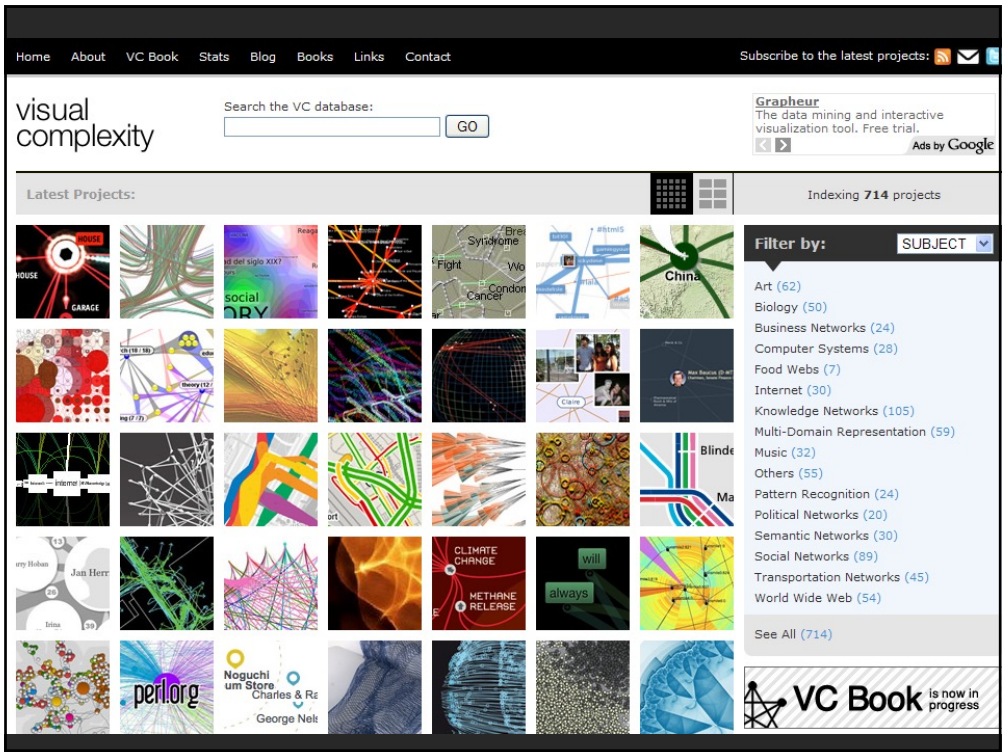
Lombardi, M. 'George W. Bush, Harken Energy and Jackson Stephens, ca 1979-90'

51

Actors and movies (bipartite)



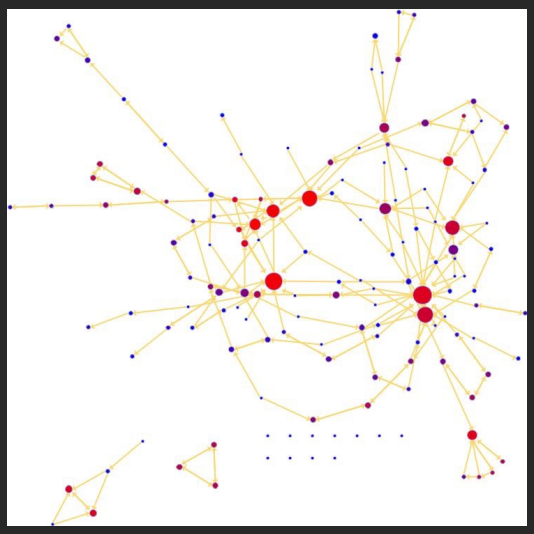
52



53

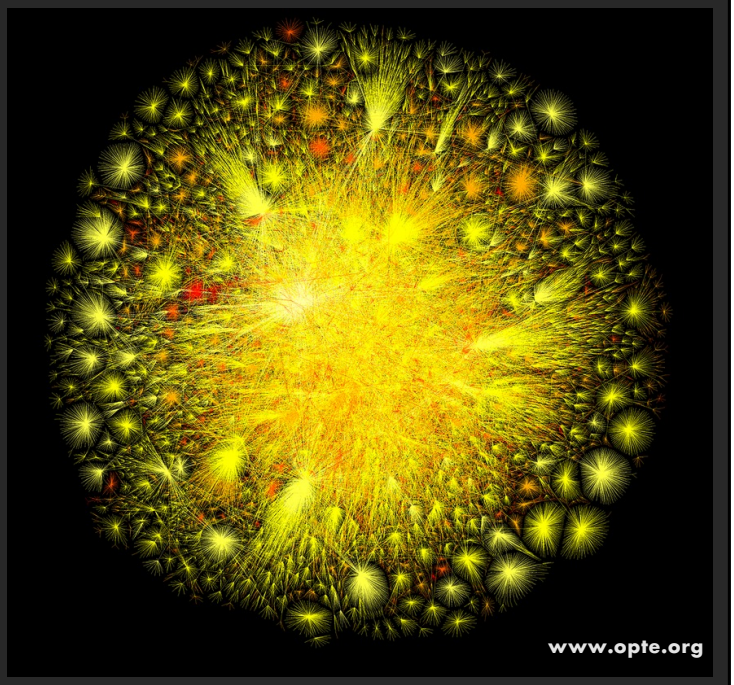
Characterizing networks

What does it look like?



54

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

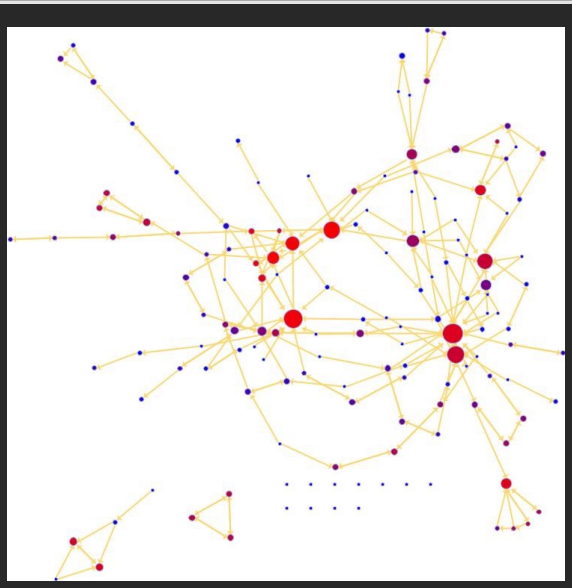


55

Centrality

57

How far apart are things?



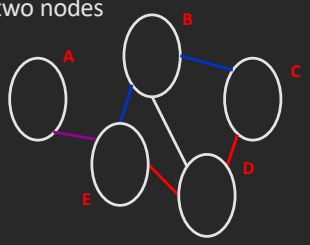
58

Distance: shortest paths

Shortest path (geodesic path)

- The shortest sequence of links connecting two nodes
- Not always unique

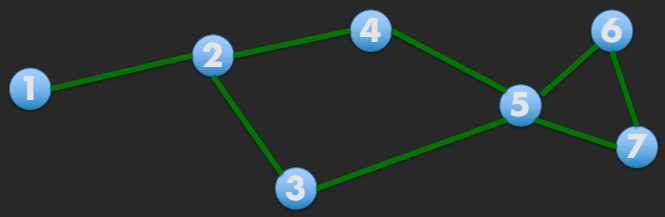
- A and C are connected by 2 shortest paths
  - A – E – B - C
  - A – E – D - C



59

Distance: shortest paths

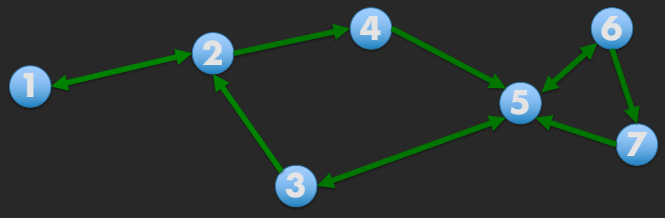
Shortest path from 2 to 3: 1



60

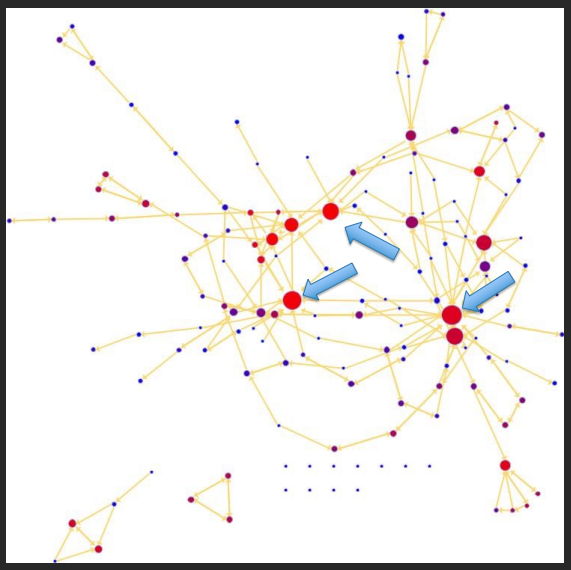
Distance: shortest paths

Shortest path from 2 to 3?



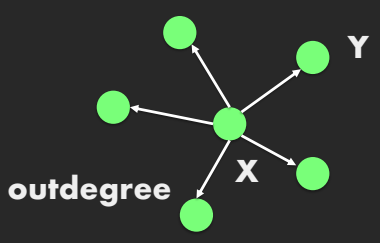
61

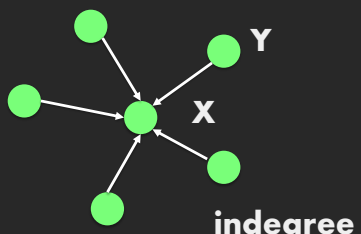
Most important node?

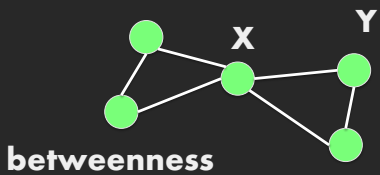


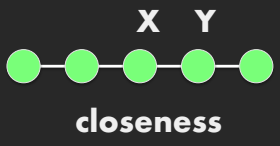
62

Centrality



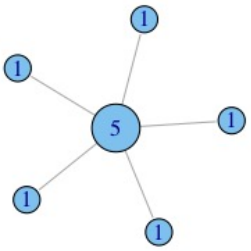






63

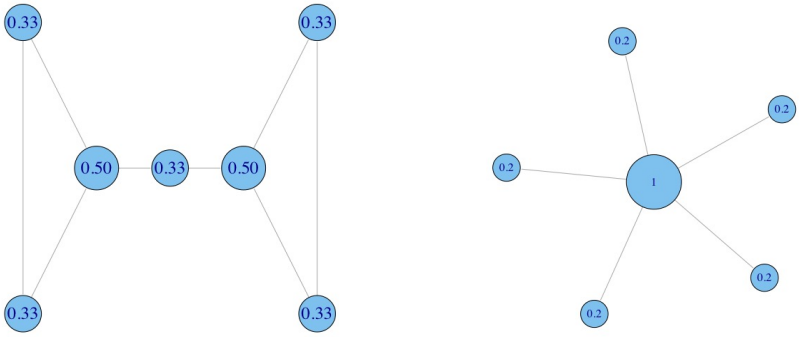
Degree centrality (undirected)



$$C_D = d(n_i) = A_{i+} = \sum_j A_{ij}$$

64

Normalized degree centrality



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

65

When is degree not sufficient?

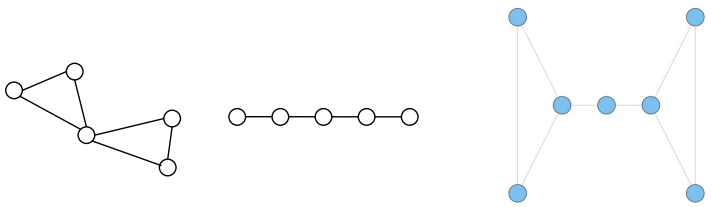
Does not capture

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

66

Betweenness

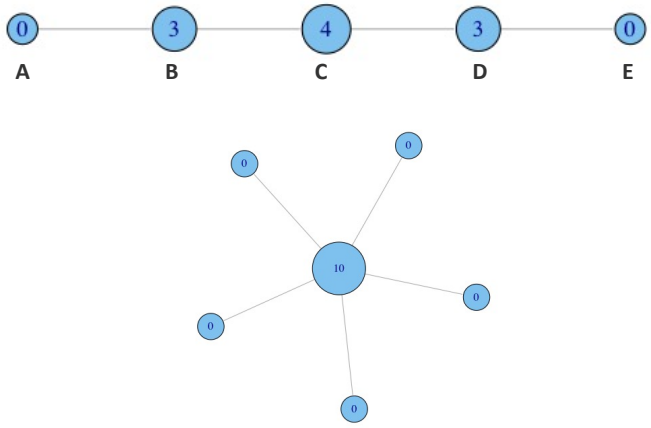
Assuming nodes communicate using the most direct (shortest) route, how many pairs of nodes have to pass information through target node?



67

Betweenness - examples

non-normalized:



68

Betweenness: 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 that node *i* is on.

Normalization:

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

number of pairs of vertices excluding the vertex itself

69

When are C_d, C_b not sufficient?

Do not capture

Likelihood that information originating anywhere in the network reaches you

71

Closeness: definition

Being close to the center of the graph

Closeness Centrality:

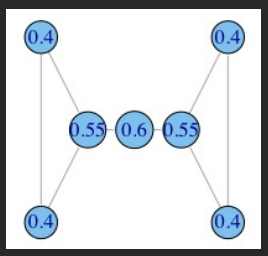
$$C_c(i) = \left[\sum_{j=1, j \neq i}^{N} d(i,j) \right]^{-1}$$

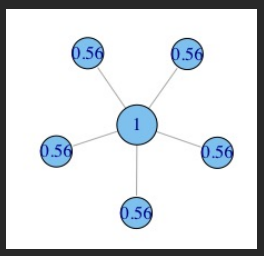
Normalized Closeness Centrality

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

72

Examples - closeness



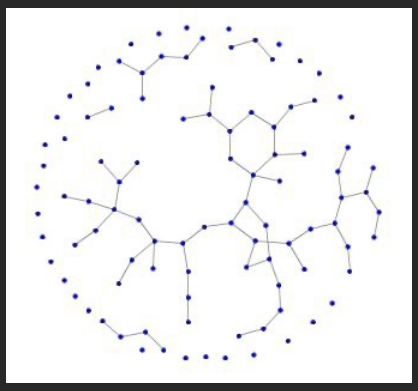


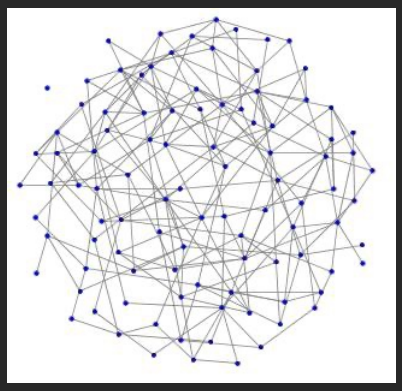
73

Community Structure

83

How dense is it?





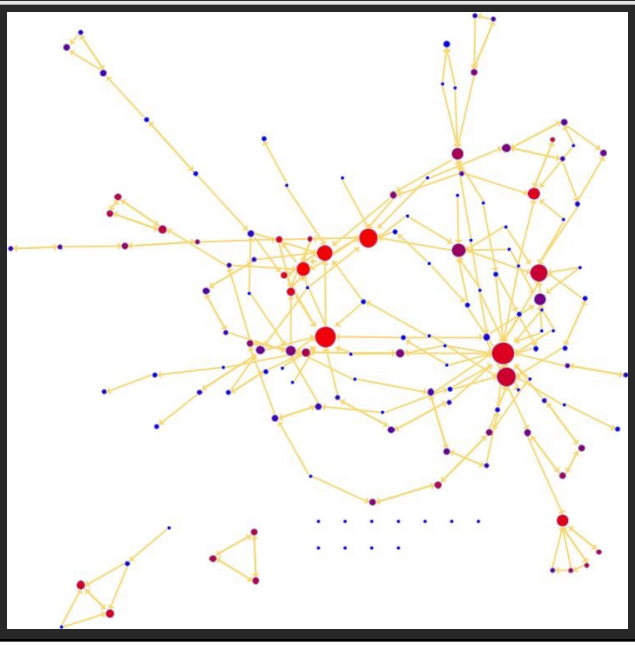
density = e/ e_{max}

Max. possible edges:

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

84

Is everything connected?

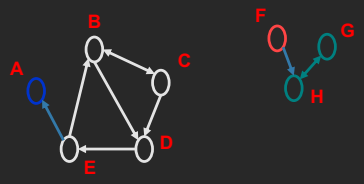


85

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

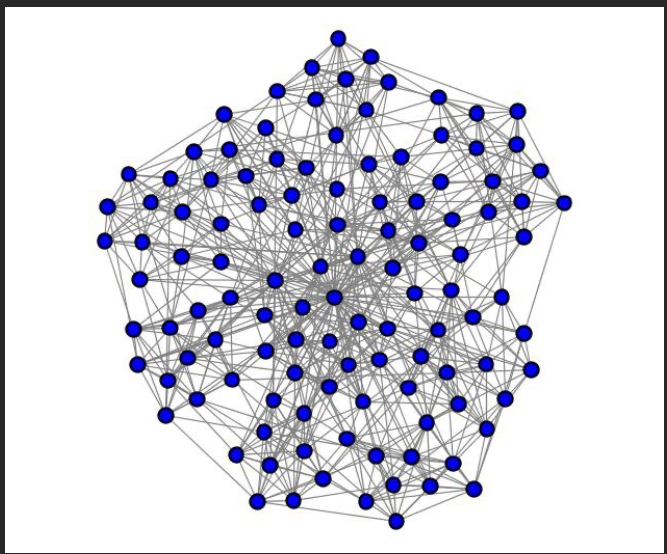


Weakly connected components

- Each node can be reached from every other node by following links in either direction
  - A B C D E
  - G H F

86

Community finding (clustering)

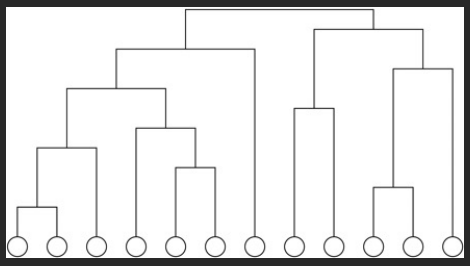


89

Hierarchical clustering

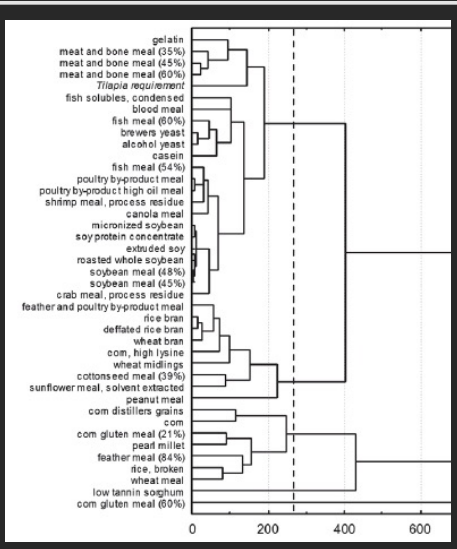
Process:

- Calculate affinity weights W for all pairs of vertices
- Start: N disconnected vertices
- Adding edges (one by one) between pairs of clusters in order of decreasing weight (use closest distance to compare clusters)
- Result: nested components

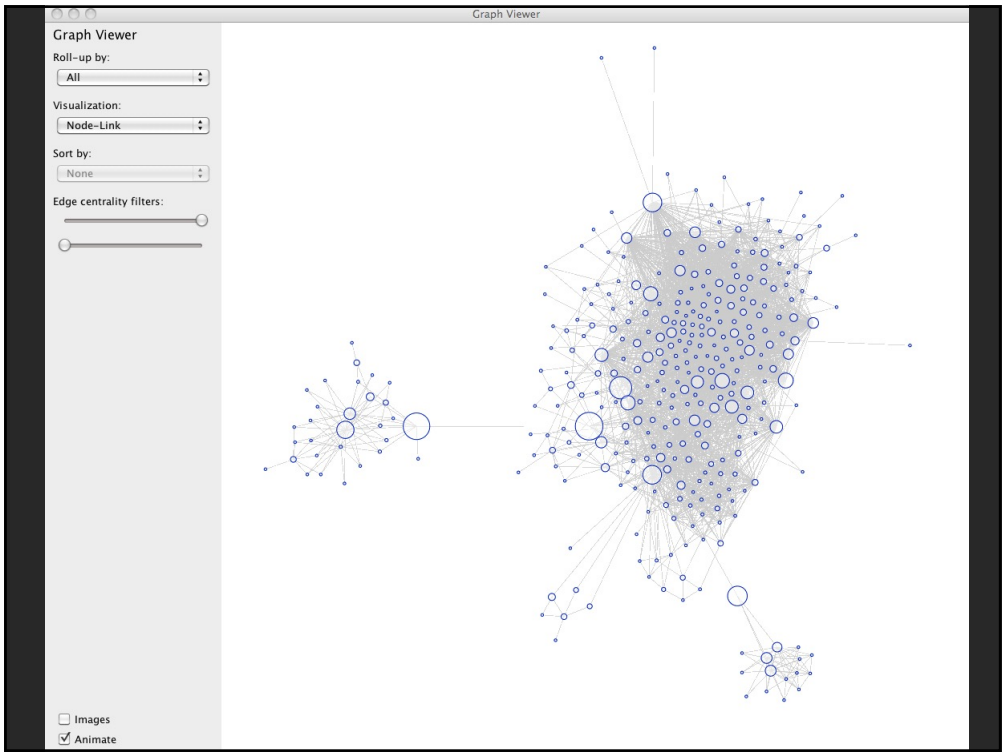


90

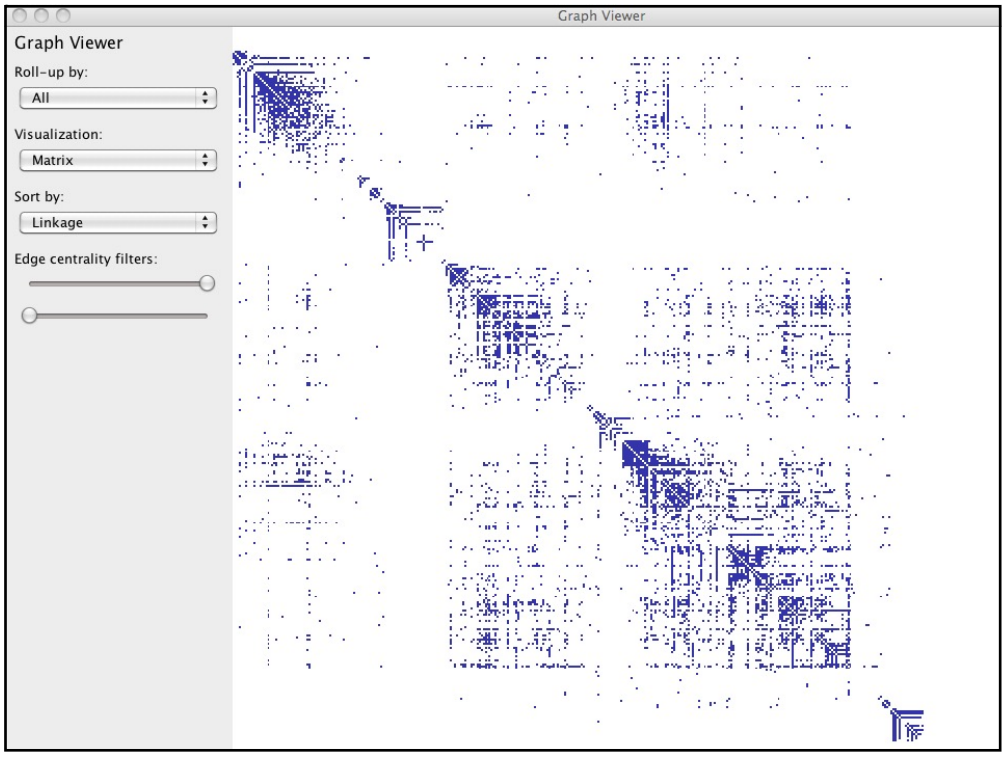
Cluster Dendrograms



91


Graph Viewer
Graph Viewer
Roll-up by:
All
Visualization:
Node-Link
Sort by:
None
Edge centrality filters:
Images
Animate

92

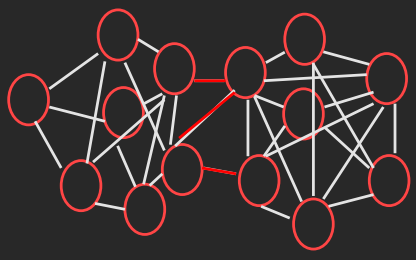

Graph Viewer
Graph Viewer
Roll-up by:
All
Visualization:
Matrix
Sort by:
Linkage
Edge centrality filters:

93

Betweenness 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



96

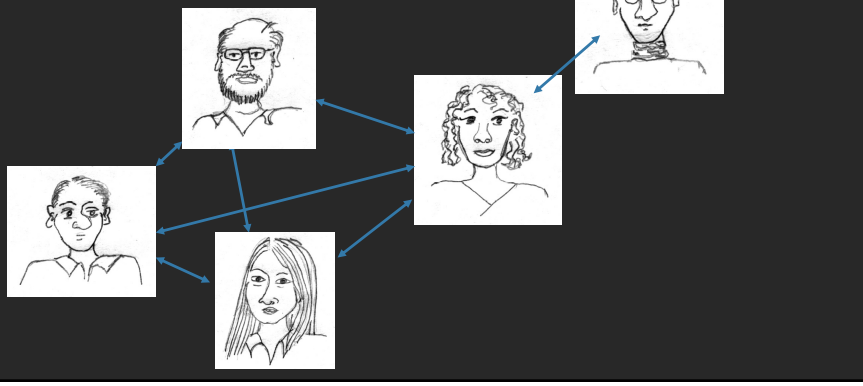
Simulating network models

114

Small world network

Milgram (1967)

- Mean path length in US social networks
- ~ 6 hops separate any two people

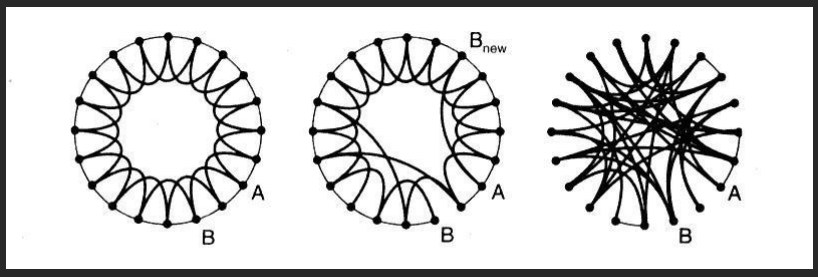


115

Small world networks

Watts and Strogatz 1998

- a few random links in an otherwise structured graph make the 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

116

Defining small world phenomenon

Pattern:

- 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

117

Summary

Structural analysis

- Centrality
- Community structure
- Pattern finding

Widely applicable across domains

122