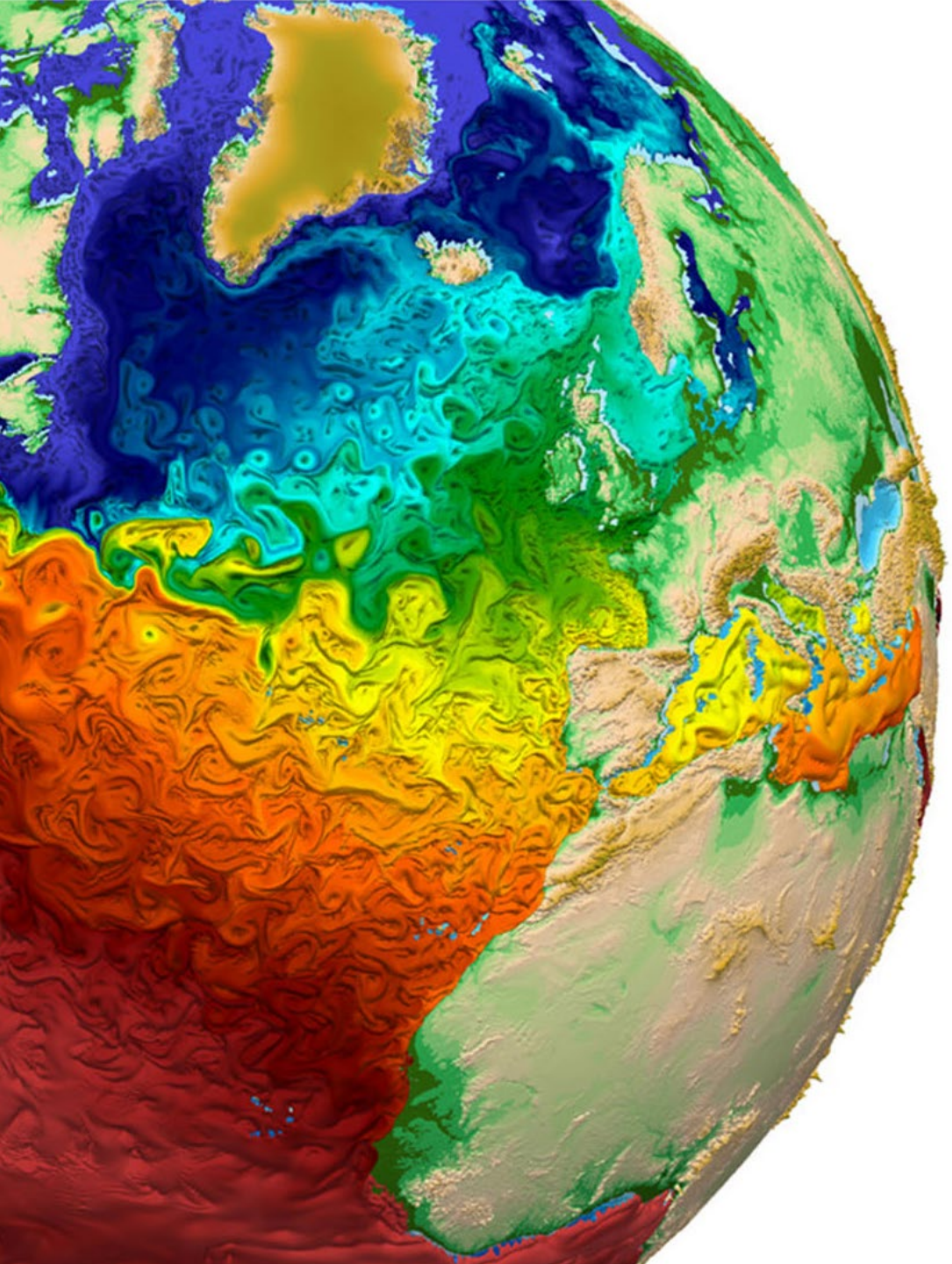


# Large Graph Visualization

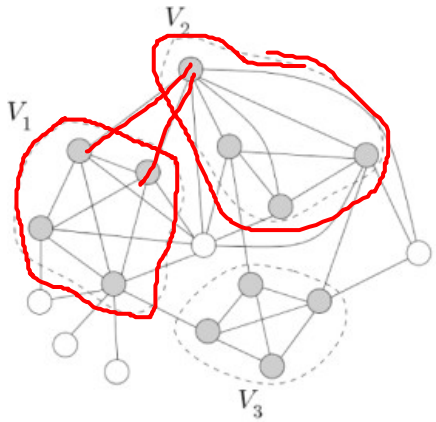
# Edge Bundling

Scientific Visualization

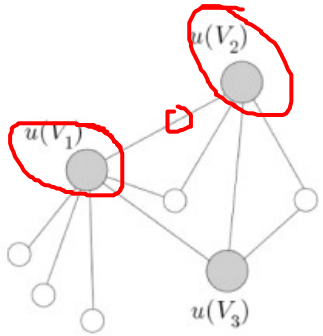
Professor Eric Shaffer

ILLINOIS

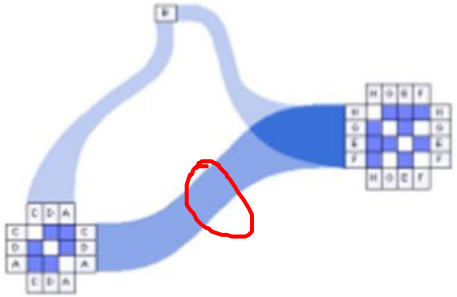
# Graph Aggregation



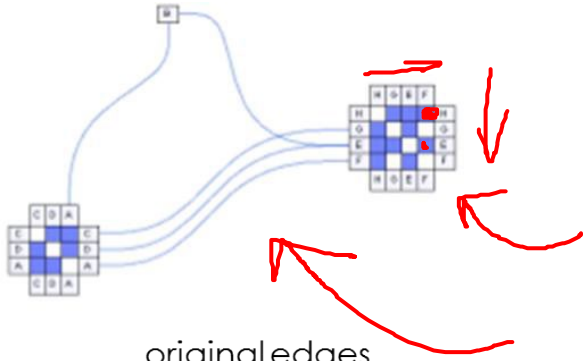

(a)




(b)



aggregated edges



original edges

Produces a simpler/smaller 'cluster graph' from a large one

Vertices: partitioned between disjoint clusters
Edges: often aggregated between clusters

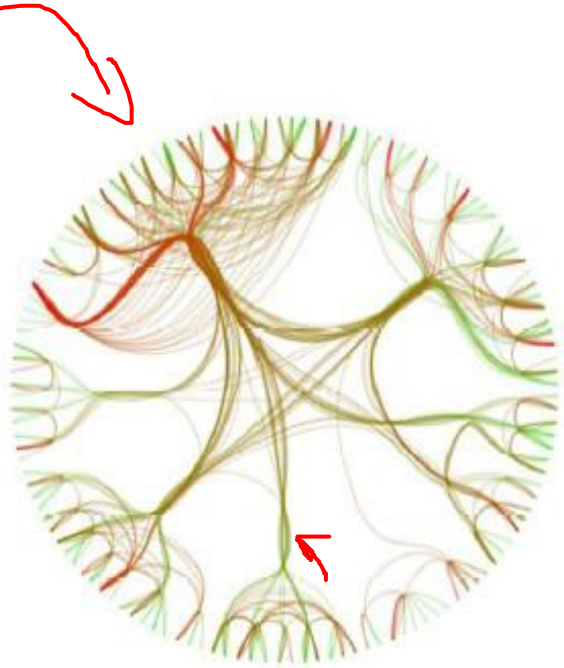
Many clustering methods (strongly-connected components, data based, …)

Visualize the cluster graph
Cluster internals shown using a cluster icon (e.g. a matrix plot)

ILLINOIS

# Graph Aggregation: Edge Bundling

- Edge bundles are clusters of similar edges
- Many approaches…usually cluster vertices
- Edges between clusters follow similar paths



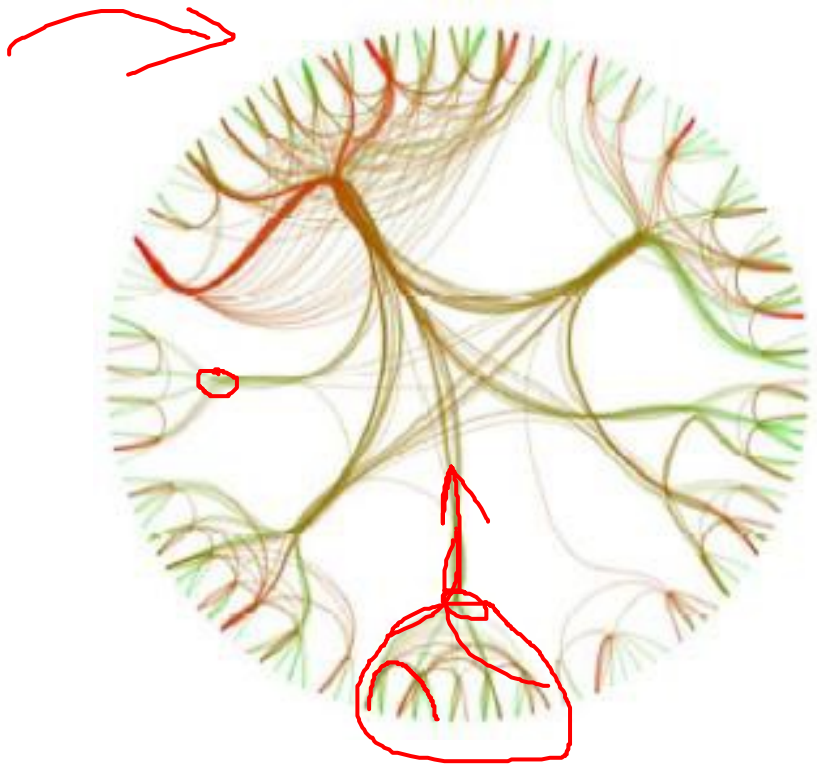
Some metrics

- Shortest path distance to a “hub-node”
- Remove high-BC edges to discover clusters
- Lots of others…..

# Hierarchical Edge Bundling: Example

Yuntao Jia, Michael Garland, John C. Hart:
*Social Network Clustering and Visualization using Hierarchical Edge Bundles.*
Computer Graphics Forum (2011)

1. Generate a hierarchical structure of vertex clusters
2. Vertices are placed radially around circle
    1. Positions from in-order traversal of hierarchy
    2. Root nodes of clusters in interior, leaves on the perimeter
3. Edges are B-Spline curves
    1. Control points are hierarchy node layout positions
    2. Positions along shortest tree path between the two nodes



ILLINOIS

# Community Discovery

- Edge betweenness centrality

$$\mathrm{BC}(u, v) = \sum_{s,t,u,v \in V} \frac{\sigma_{s,t}(u, v)}{\sigma_{s,t}}$$

- $\sigma_{st}$ is the total number of shortest paths from node s to node t
- $\sigma_{st}(u, v)$ is the number of those paths that pass through edge u,v

- Low BC edges connect nodes within a community
- High BC edges connect communities

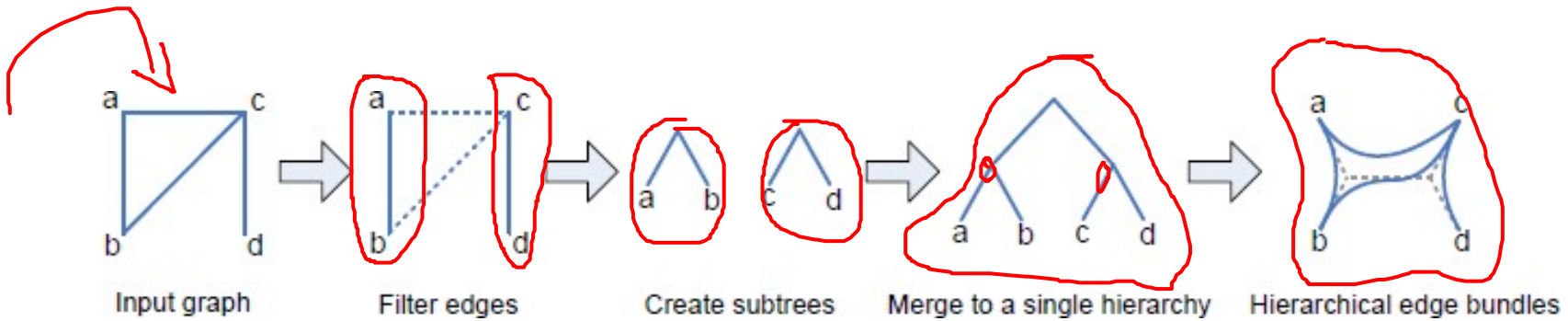
ILLINOIS

# Balanced Hierarchy Construction

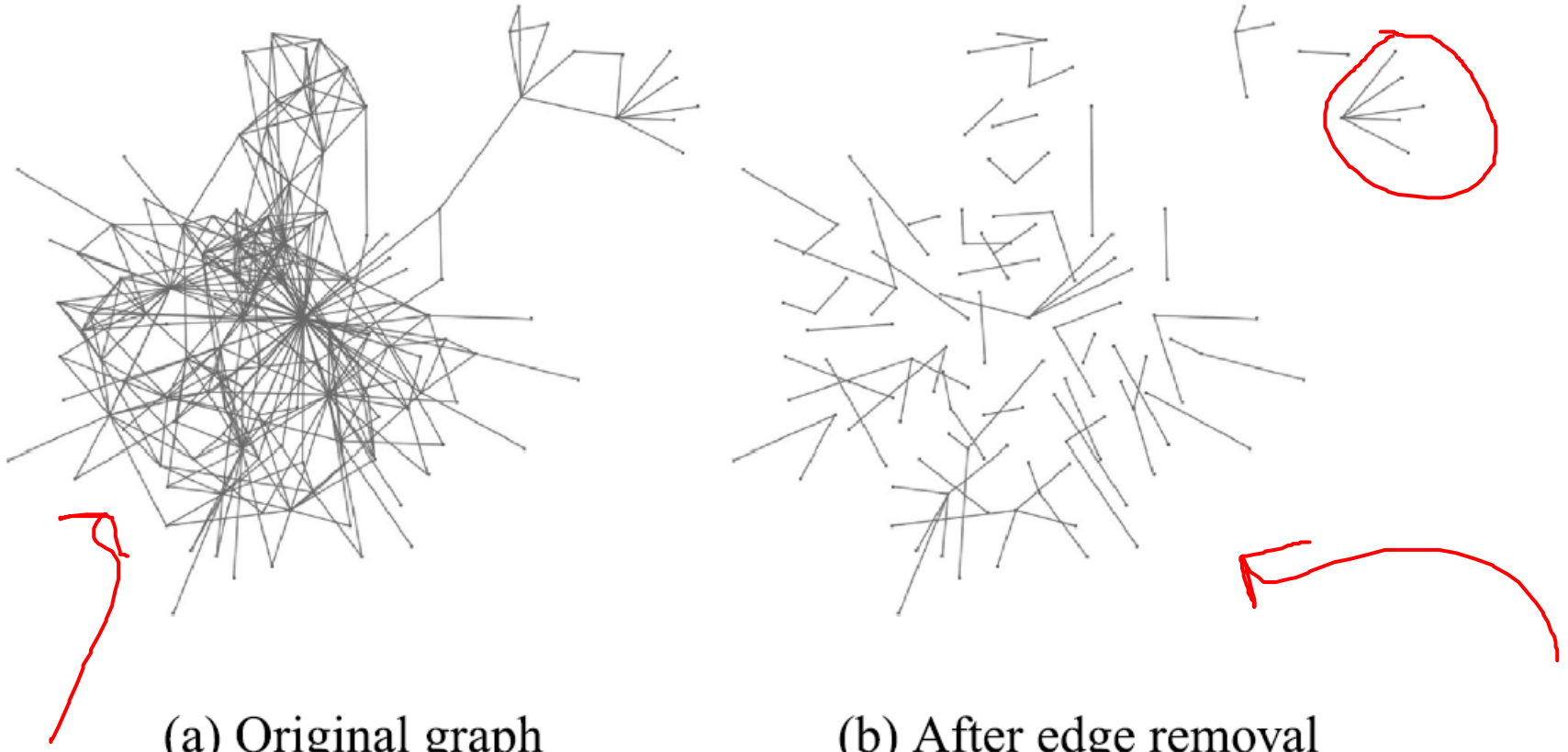
Filter edges by removing highest-bc edges first

- An edge a,b is removed only if min(deg(a),deg(b)) >1
- And BC(a,b) > 1

Construct communities by merging in increasing BC order of removed edges



# Edge Removal

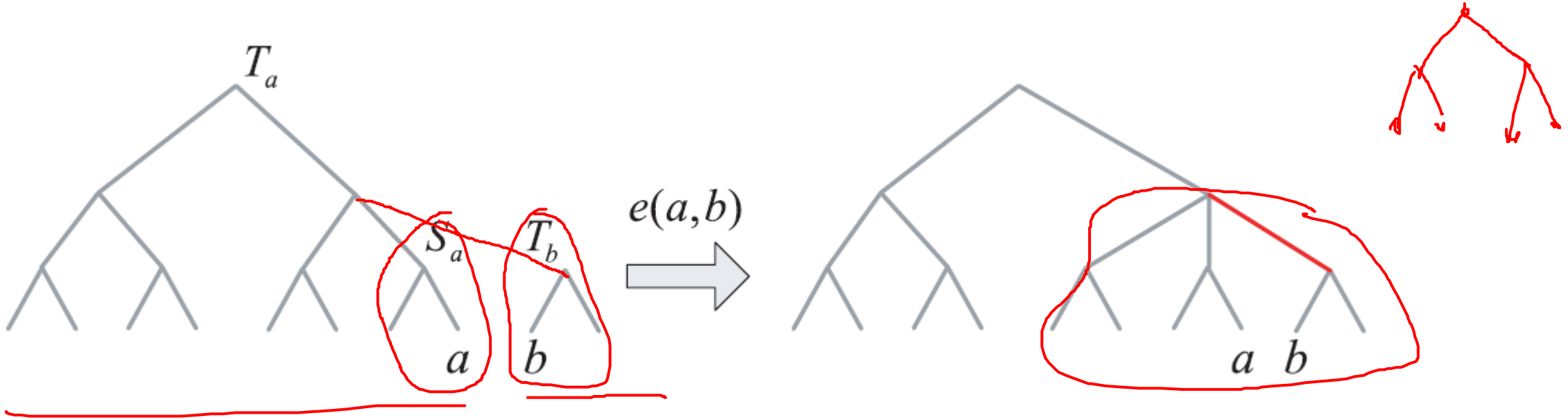


(a) Original graph

(b) After edge removal

ILLINOIS

# Merging Communities into a Hierarchy



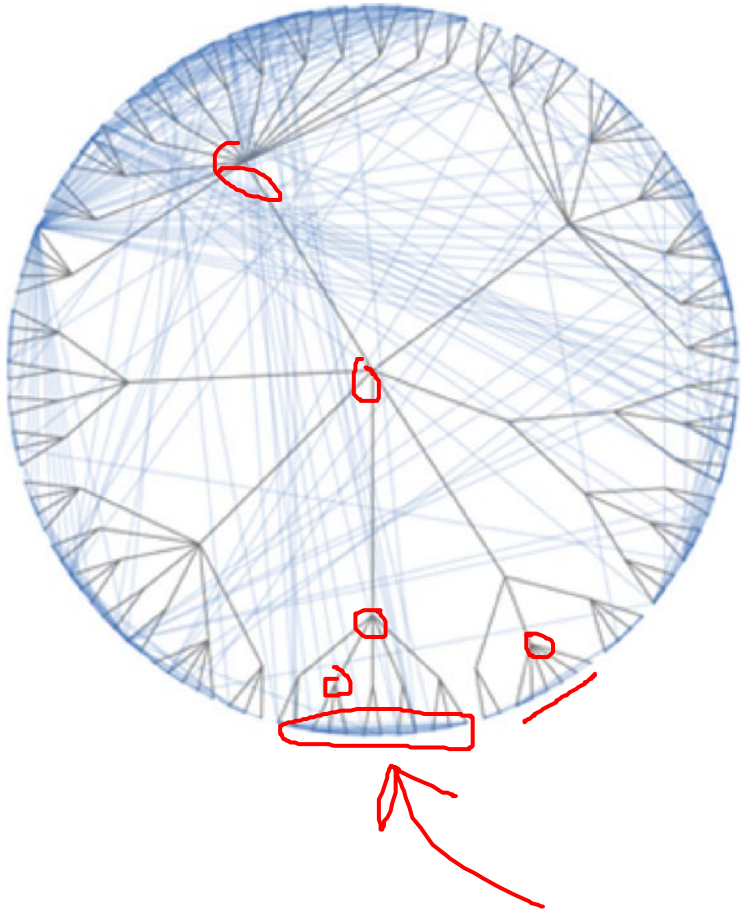

We scan the list of removed edges in order of increasing BC and merge subtrees connected by those edges.

If Ta and Tb share the same height, then they can be merged as children of the same new parent tree node.

Otherwise assume without loss of generality that Ta is taller than Tb.
Let Sa be the unique (lowest) subtree of Ta that contains a and shares the same height as Tb.
Then the communities are merged by assigning the parent of Sa as the parent of Tb.

ILLINOIS

# Using the Hierarchy to Bundle Edges



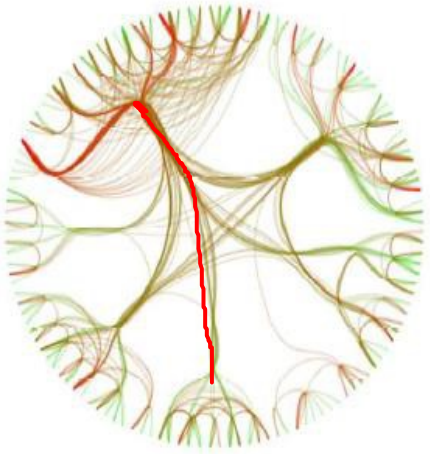
Gray lines show the computed hierarchy used to layout the graph

Blue lines show edges drawn linearly between nodes

Edge bundling is accomplished by drawing them as curves

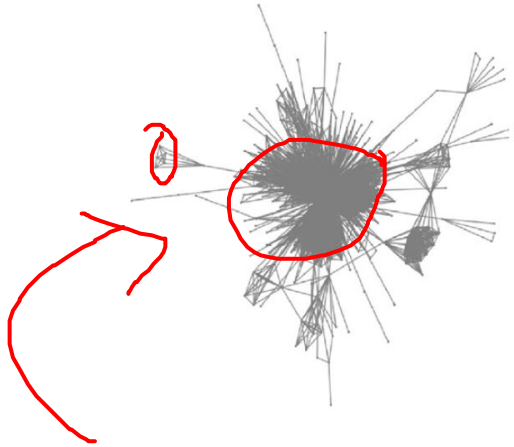
Curves with control points defined by the internal nodes

Edges between 2 communities will be drawn with similar curves

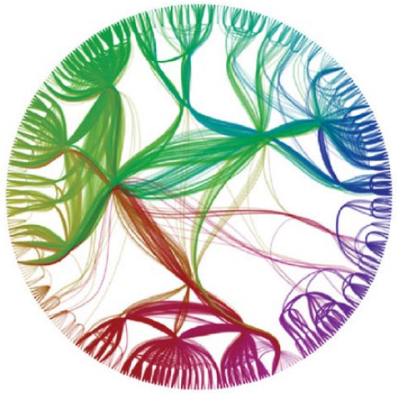


ILLINOIS

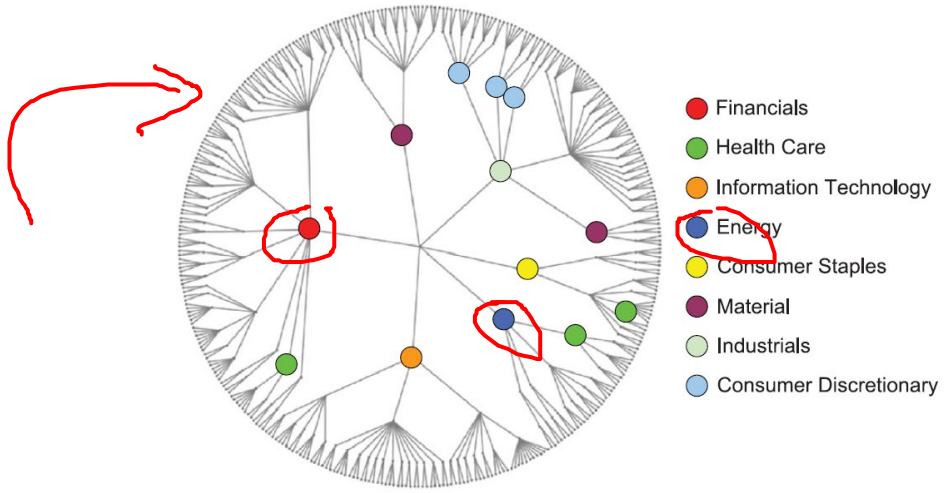
# Example: E-mail in the Enron Scandal



(a) Original graph



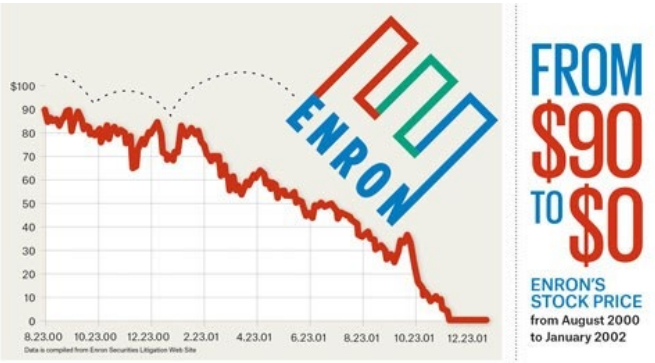
(b) Edge Bundles




(c) Generated Hierarchy (with user supplied labels)

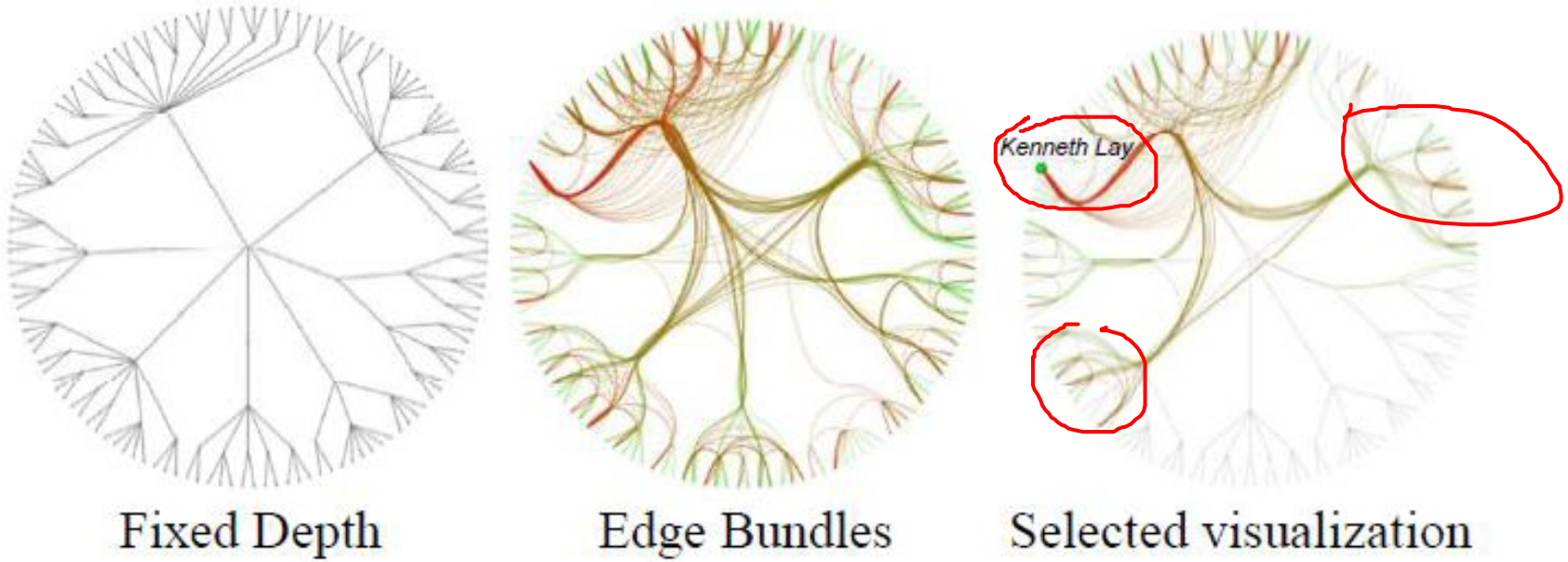
The Enron scandal was an accounting scandal of Enron Corporation, an American energy company based in Houston, Texas. It was publicized in October 2001, and led to the bankruptcy of the company, and the de facto dissolution of Arthur Andersen, which was one of the five largest audit and accountancy partnerships in the world. In addition to being the largest bankruptcy reorganization in American history at that time, Enron was cited as the biggest audit failure.
- Wikipedia

ENRON®
Endless possibilities.™




ILLINOIS

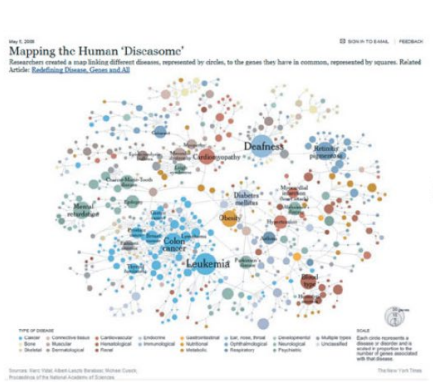
# Enron E-Mail Graph



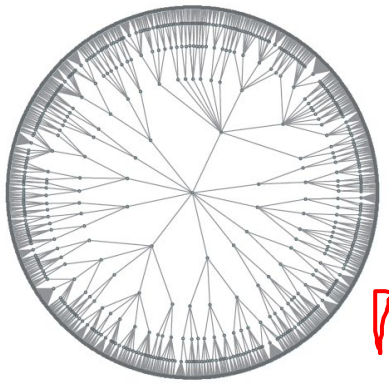

Enron scandal 2001
389 e-mails, 132 employees
Red = sender, Green = recipient
Can select node to see which communities that person contacted

ILLINOIS

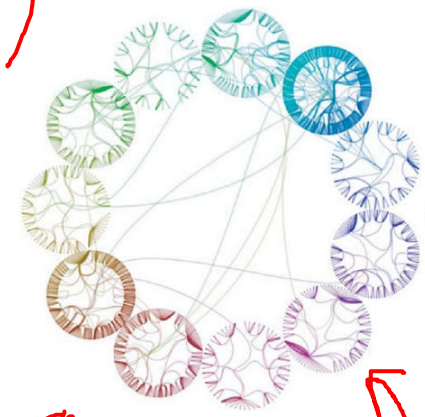
# Example: Diseasome



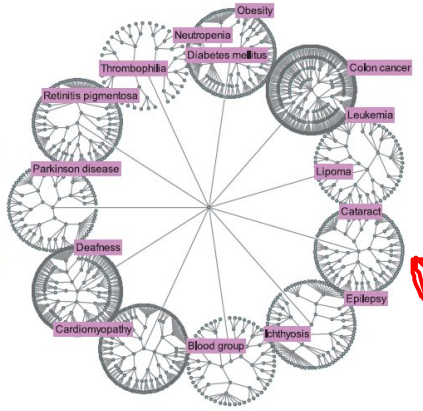

(a) "diseasome"



(b) Generated hierarchy



(c) Edge bundles




(d) Adaptive hierarchy drawing

A bipartite human disorder-gene network
1550 associations between 1419 disorders and genes
A disorder and gene connect if disorder arises from mutation of the gene

(a) shows a visualization of the network published in New York Times
- squares correspond to genes and circles correspond to disorders
- coloured by disorder types
- sized by the number of gene links

(b) HEB clusters form a community hierarchy

(d) Hierarchy laid out using adaptive tree drawing
- all first level internal nodes are drawn on different layout circles
- disorders are also labelled
- same type of disorders is likely clustered into the same community

(c) disorder–gene edge bundles are drawn on top of hierarchy
- seldom connections between communities

ILLINOIS