

“Search, Show Context, Expand on Demand”: Supporting Large Graph Exploration with Degree-of-Interest

Authors:

Frank van Ham and Adam Perer

Presenter: Charles Blatti

March 19, 2010

March 19, 2010

Search, Show Context, Expand on Demand

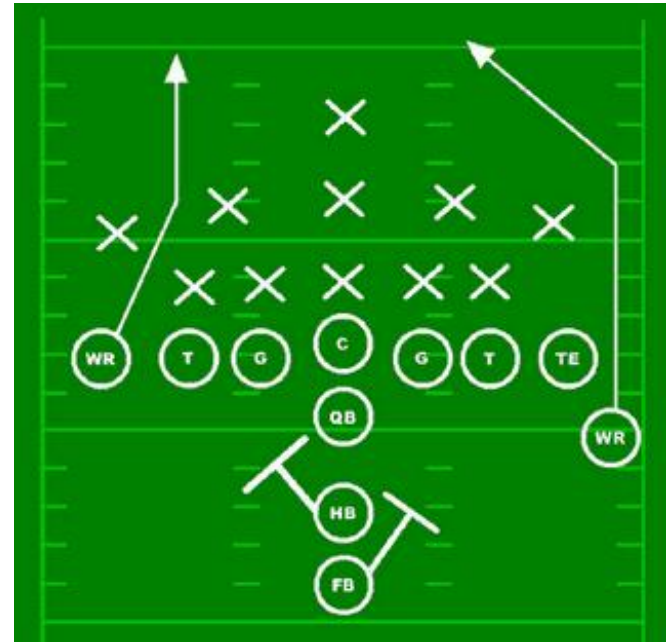
- Submitted in InfoVis 2009
- Frank van Ham

Adam Perer



The Game Plan

- Motivation
- Related Work
 - Degree-Of-Interest
- Novel Ideas and Implementation
- Anecdotal Support of Utility
- Demo
- Impressions and Future Directions



Interactive Overviews of Entire Graphs

- “Overview, Zoom, Details on Demand”
- Seven types of data representation
 - (1-, 2-, 3- dimensional data, temporal and multi-dimensional data, and tree and network data)
- Seven types of operations
 - Overview, zoom, filter, details-on-demand, relate, history, and extract
- B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations. 1996.

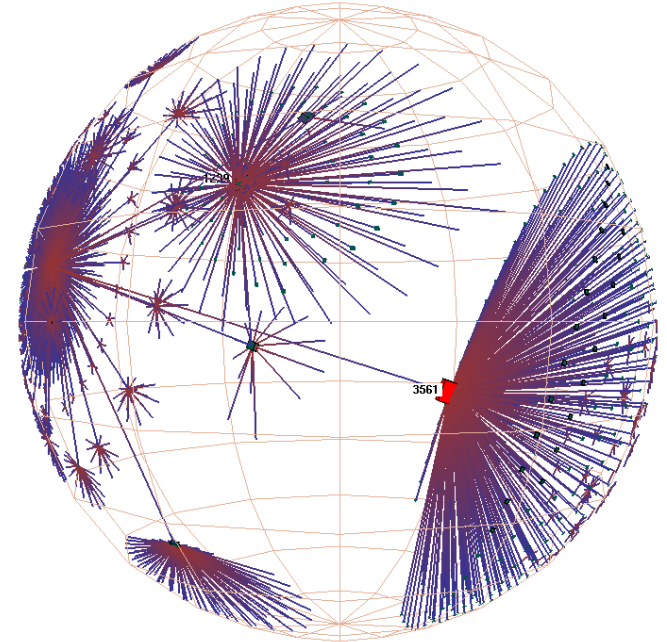
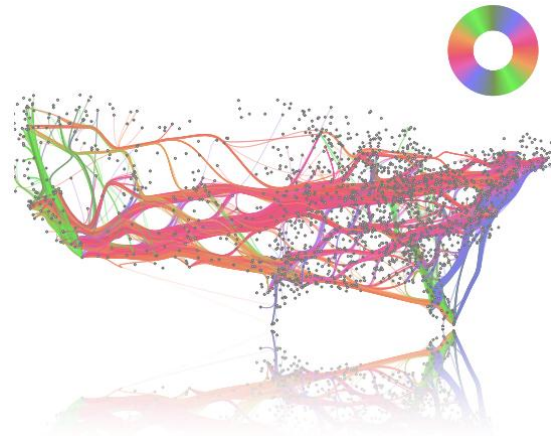
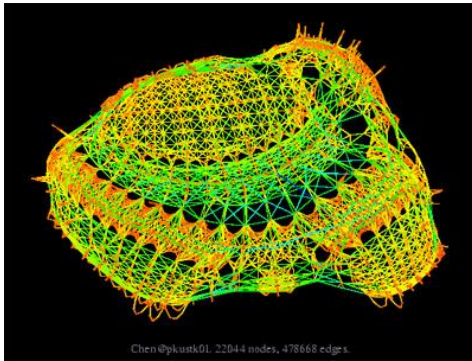
Interactive Overview of Entire Graphs

- Design novel techniques for layout
- No information missing
- Allows users to deduce central actors or clusters and outliers
- Density and size make it hard
- Difficult to follow edges and count in-degree
- Examples
 - Social scientists
 - Software architect

Semi-specific Information Needs

- “Search, Show Context, Expand on Demand”
- Do not need the entire graph
- Trying to learn more about a particular data point and determine how it relates to similar points
- Examples
 - Fraud analysts
 - Programmers

Practical Reasons



- Multiple analysts of large graph databases through online connection
- Privacy reasons and access restrictions prevent precomputed overviews
- Running on less powerful mobile devices

Main Idea of Project

- Extend degree of interest concept from graphs to trees
- Pick a focal point and display relevant context
- Allow context to be defined by graph attributes, associated content, and user interactions
- Extract maximal interest subgraph around focal point
- Allow user to select and expand context in interesting directions

Degree of Interest

- Concept introduced in classic paper
 - “Generalized fisheye views”
 - George Furnas, 1986
- Definition :
 - $DOI_{\text{fisheye}}(x|y) = API(x) - D(x,y)$
 - x is any node
 - y is current point of focus
 - $API(x)$ is the global *A Priori* Importance
 - $D(x,y)$ is the Distance between x and the current point y

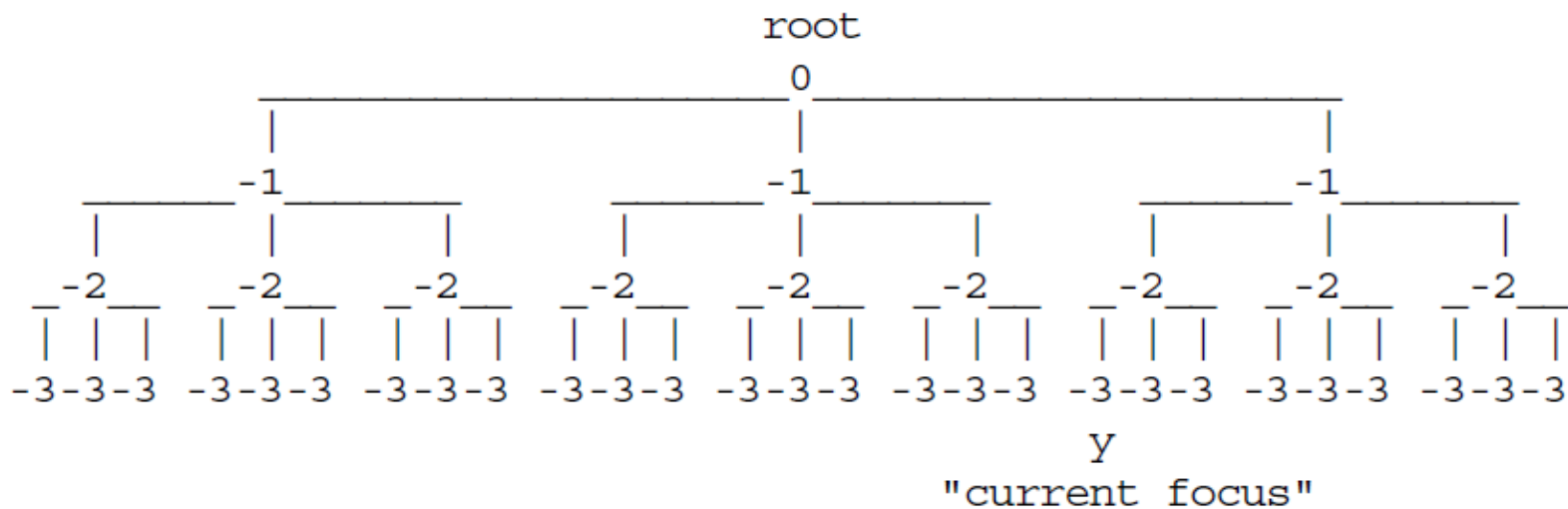


API(x)

- Designed for rooted trees
- API is depth in tree

(b) *A Priori* Importance in the tree:

$$\text{Imp}(x) = -d_{\text{tree}}(x, \text{root})$$

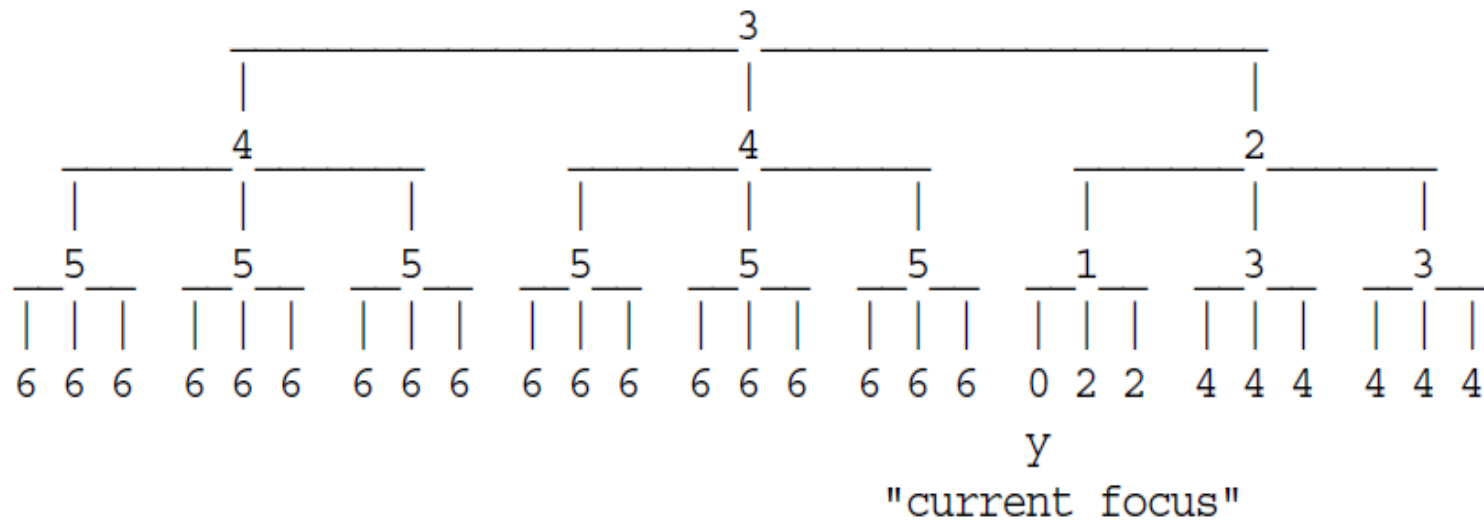


$D(x,y)$

- Distance in tree between x and y

(a) Distance from y :

$d_{\text{tree}}(x,y)$

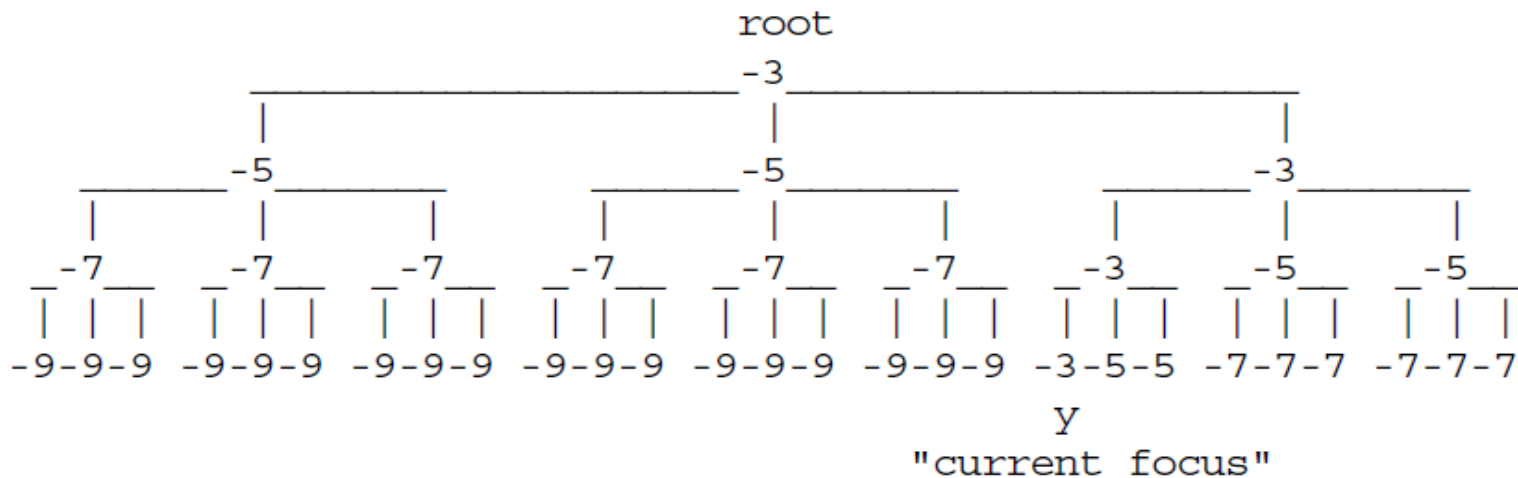


DOI_{fisheye}(x | y)

- Interestingness Measure on each node

(c) The Fisheye DOI:

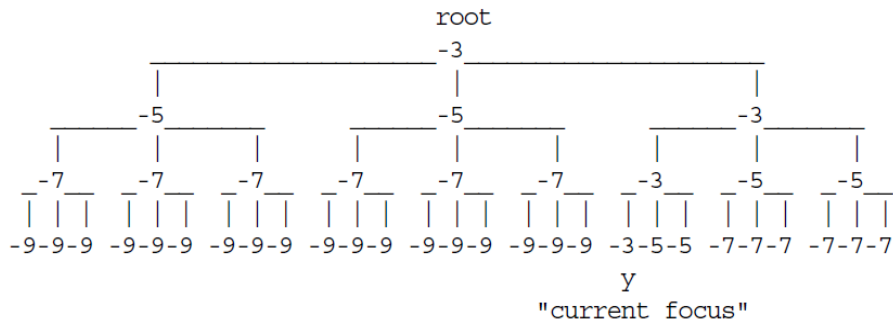
$$\begin{aligned} \text{DOI}_{\text{fisheye}(\text{tree})}(x|y) &= \text{API}(x) - D(x,y) \\ &= -(d_{\text{tree}}(x,y) + d_{\text{tree}}(x,\text{root})) \end{aligned}$$



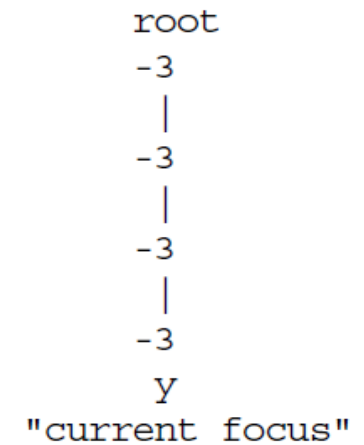
Thresholding to Create Subgraphs

(c) The Fisheye DOI:

$$\begin{aligned} \text{DOI}_{\text{fisheye}(\text{tree})}(x|,=y) &= \text{API}(x) - D(x,y) \\ &= -(d_{\text{tree}}(x,y) + d_{\text{tree}}(x,\text{root})) \end{aligned}$$



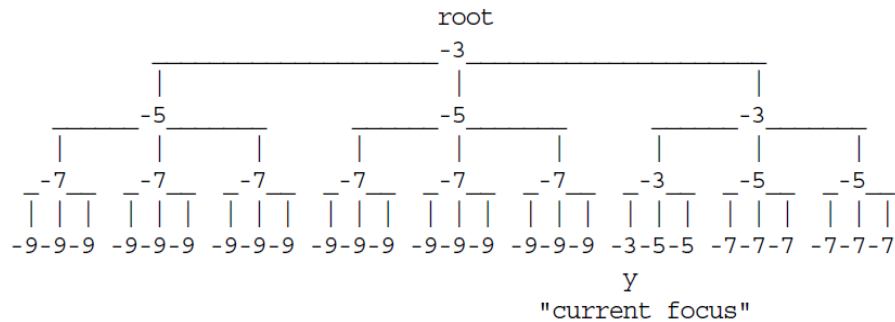
(a) Zero-order tree fisheye:



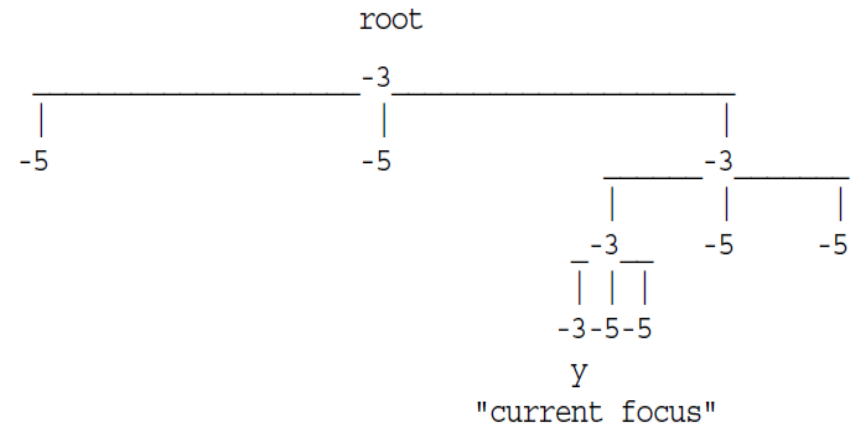
Thresholding to Create Subgraphs

(c) The Fisheye DOI:

$$\begin{aligned} \text{DOI}_{\text{fisheye}(\text{tree})}(x|y) &= \text{API}(x) - D(x,y) \\ &= -(d_{\text{tree}}(x,y) + d_{\text{tree}}(x,\text{root})) \end{aligned}$$



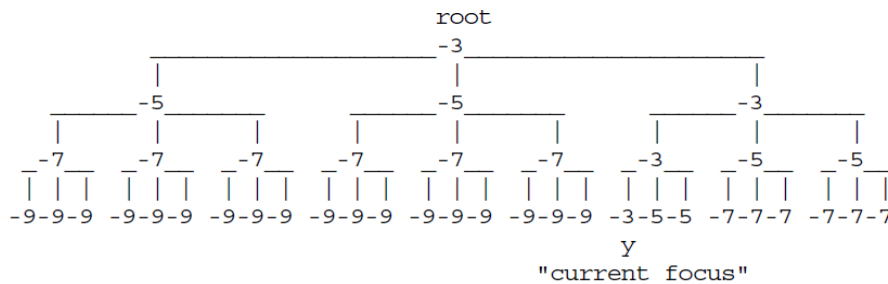
(b) First-order tree fisheye:



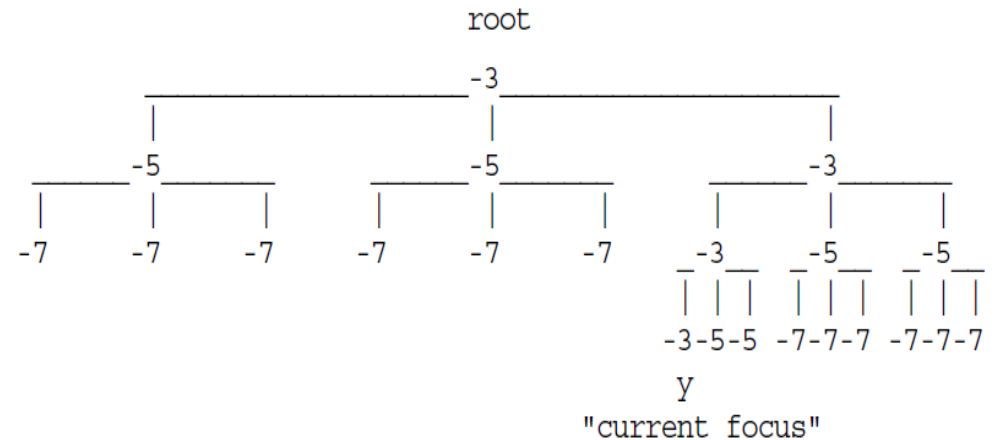
Thresholding to Create Subgraphs

(c) The Fisheye DOI:

$$\begin{aligned} \text{DOI}_{\text{fisheye}(\text{tree})}(x|y) &= \text{API}(x) - D(x,y) \\ &= -(d_{\text{tree}}(x,y) + d_{\text{tree}}(x,\text{root})) \end{aligned}$$



(c) Second-order tree fisheye:



DOITrees

March 19, 2010

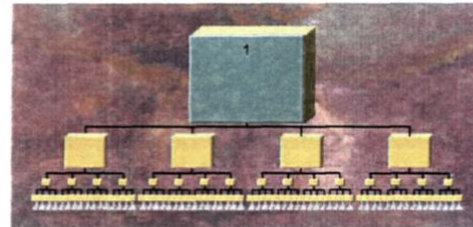


Plate 2. Display of a uniform tree of 4 levels, 4 branches each level, focus on Node 1 (number in the plate represents the node number).

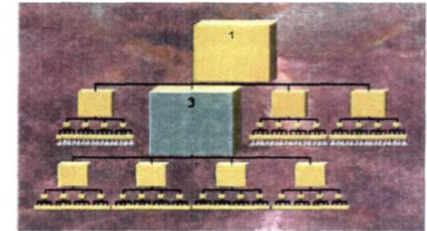


Plate 3. Same table with focus on Node 3.

- S.K. Card and D. Nation. Degree-of-Interest Trees: A Component of an Attention-Reactive User Interface. 2002.
- J. Heer and S.K. Card. DOITrees revisited: Scalable, Space-constrained Visualization of Hierarchical Data. 2004

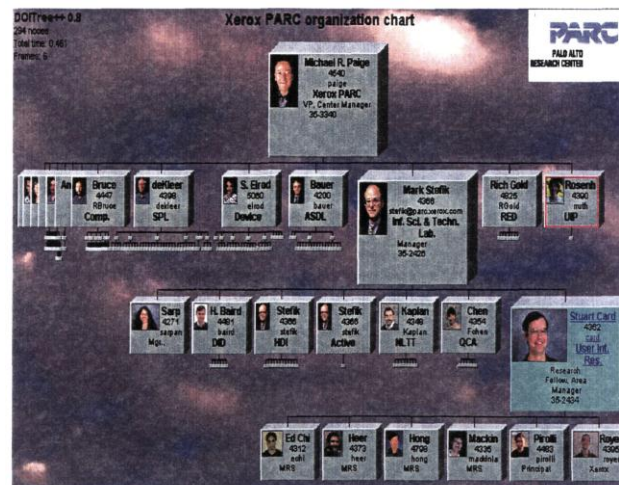


Plate 8. Organization chart with over 400 nodes accessible over WWW through Web browser.

March 19, 2010

General Graph Clustering

- DOI functions to aggressively cluster nodes further away from the focal node
- Generate abstract views where meaning of clusters are unclear
 - F. van Ham and J.J. van Wijk, *Interactive Visualization of Small World Graphs*. 2004

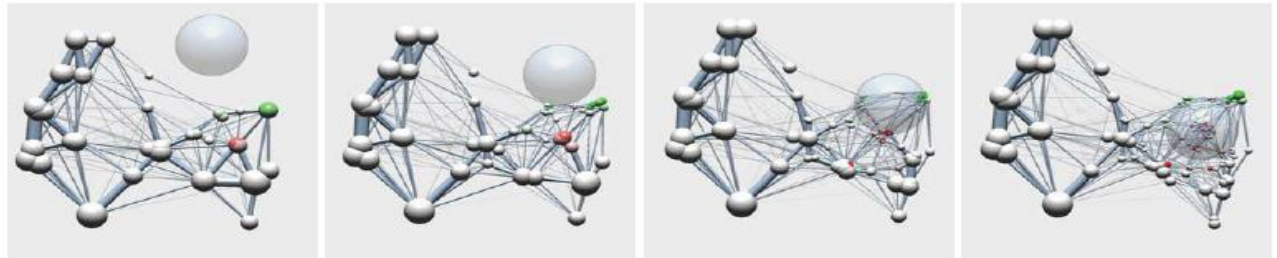
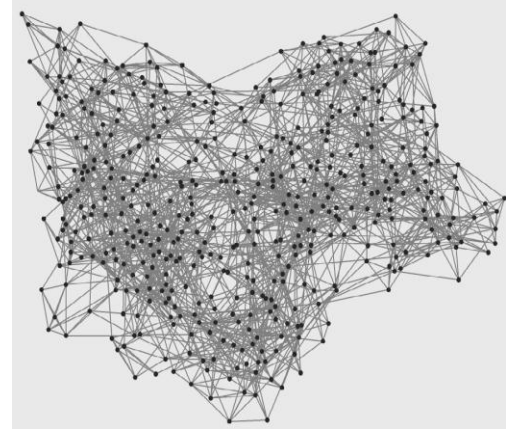
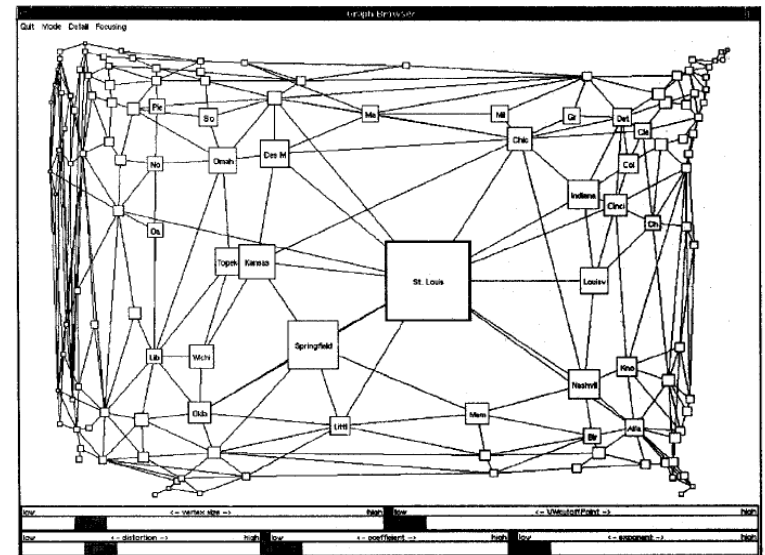


Figure 8: Series of frames showing the effect of a changing focus on the layout. Surrealist artists are indicated in red, artists belonging to the Pop-Art movement in green.

Graphical Distortion

- Use DOI to compress information
- Still render entire graph
 - M. Sarkar and M.H. Brown. Graphical fisheye views of graphs. 1992



Attribute Based Abstractions

- Use node attributes as parameters for abstraction
- Where each node is associated with several attributes
- Roll up into metanodes where they have the same value for particular dimension
 - [M. Wattenberg, Visual exploration of multivariate graphs. 2006](#)

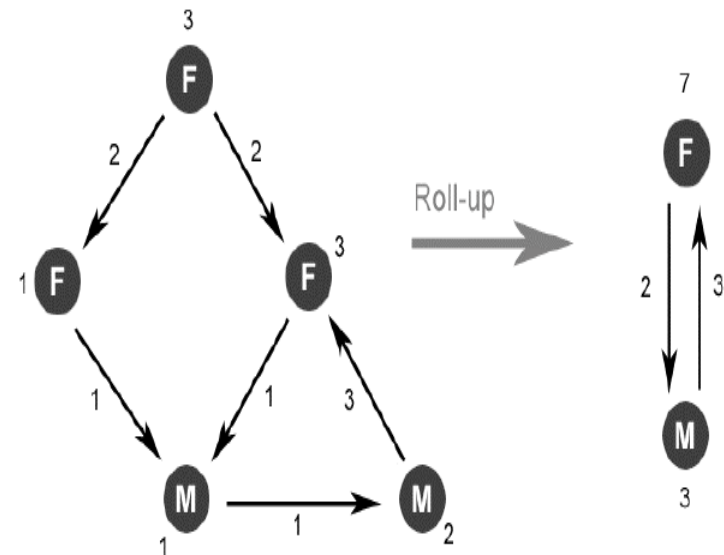
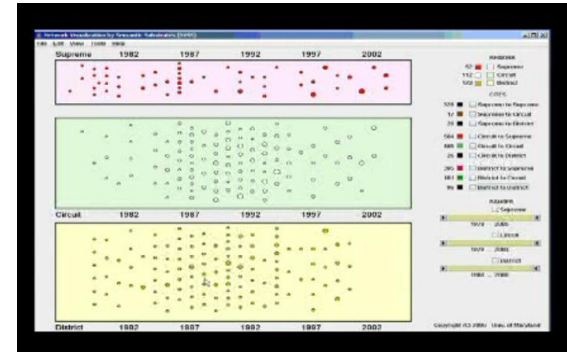


Figure 2. Roll-up of a simple social network on “gender” dimension. Numbers indicate weights of edges and sizes of nodes.

Semantic substrates



- Use attributes to assign a basic layout
- Superimpose edges on top
- <http://hcil.cs.umd.edu/video/2006/substrates.mpg>
- Can relate higher level features to application area
- Do not provide detailed information on the single node
 - B. Shneiderman and Aris, A. Network Visualization by Semantic Substrates. 2006.

Contextual Views

- Pick a point and show immediate surrounding subgraph
- Perform well when average degree is low
- Struggle with high degree

“Plant a tree and watch it grow”

- B. Lee, et al. TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts. 2006

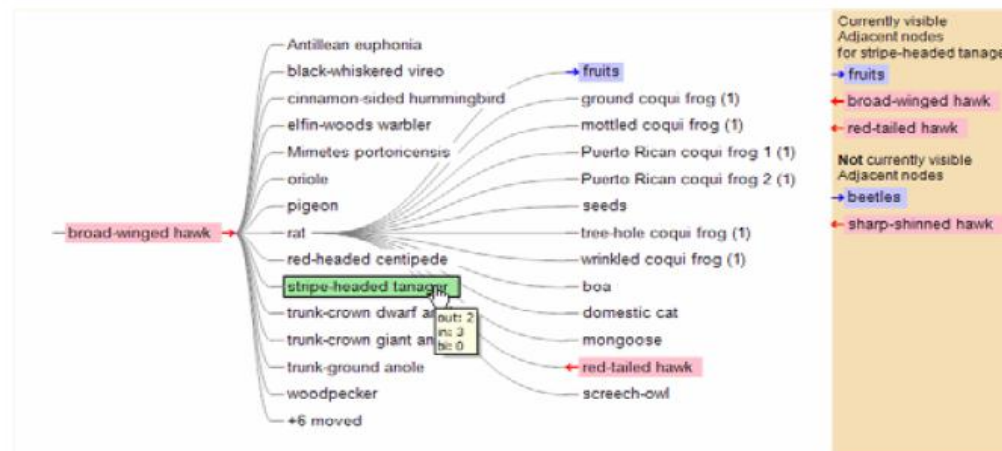


Fig. 2. “broad-winged hawk” was set as the root, and users selected “rat” which added all its adjacent nodes to the tree. A single click on “stripe-headed tanager” gives it the focus and shows a preview of its adjacent nodes in the preview panel on the right. The adjacent nodes already present in the tree are highlighted in the tree revealing that “fruits”, “red-tailed hawk” and “broad-winged hawk” are connected to both “rat” and “stripe headed tanager”. Color indicates link direction.

Computational Approaches

- Compute network matrices to find relevant subgraphs before visualization
- Calculate different metrics
 - *bary center*
 - *betweenness centrality*
 - *closeness centrality*
 - *cut-points*
 - *degree*
 - *HITs*
 - *power-centrality*

March 19, 2010

- A. Perer and B. Shneiderman, Balancing systematic and flexible exploration of social networks. 2007

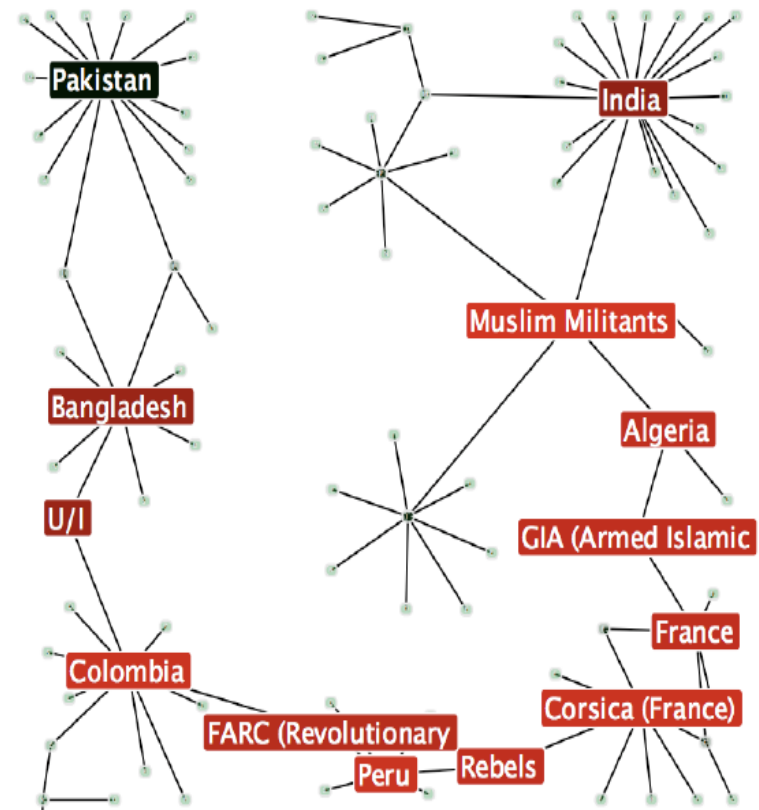


Figure 2. Users can adjust the double range slider to filter nodes that are not of interest. In this graphic, the nodes of Figure 2 that do not have a “betweenness centrality” ranking score of at least 1000 become faded and their labels are removed. The labels of nodes that meet the ranking criteria can be increased by the user. This allows users to focus on the type of nodes they are interested without ignoring the overall structure.

Extending Degree of Interest

- $\text{DOI}(x \mid y) = \alpha \cdot \text{API}(x) + \beta \cdot D(x, y)$
 - API can be calculated in terms of structural graph properties or node attributes
 - Node degree
 - Value of an attribute (Supreme Court Case)
 - $D(x, y)$
 - Minimum graph distance
 - Could depend on edge weights or edge attributes (experts)

Edge Disinterest function

- $EI(e,x,y) > 0$,
 - Disinterest function where higher values indicate less interest in following a link e between x and y
 - Can be used to define path length between to nodes in graph

Use DOI to Extract Maximal Interest Subgraph Given y

- Problems
 - Pick focal point
 - Search (ignored in many other degree of interest visualizations)
 - Traditionally use opportunistic browsing of high level overviews
 - Local maxima
 - Interest function has many local maximum
 - Many potentially interesting nodes are surrounded by uninteresting ones

“Search”

- Textual search on attribute information
- Search parameters z give important user interest hints
- User Interest function (UI), known before user picks focal node
 - Match or within a range of user defined values
- $DOI_{naive}(x \mid y, z) = \alpha \cdot API(x) + \beta \cdot UI(x, z) + \gamma \cdot D(x, y)$

Local maximum

- DOI is strictly nested in Furnas and other tree DOI implementations
 - $\text{DOI}(x)$ of any node x in a subtree of r will never be larger than $\text{DOI}(r)$.
 - Guarantees connected trees when thresholding DOI
- No guarantee on graphs
 - Simple threshold will result in disconnected graphs

Solution: Diffuse Interest in Graph

- Every node depends on its own intrinsic interest and a fraction of its highest interest neighbor
 - $API_{\text{diff}}(x) = \max(API(x), \delta \cdot \max(n \in N(x) : 1/EI(e,x,n) \cdot API_{\text{diff}}(n)))$
 - When $\delta = 0$, same as naïve with no diffusion
 - As δ approaches 1, increase diffusion
 - $EI(e,x,n)$ edge disinterest function
 - incorporated so diffusion reflective of interestingness of edge
 - $UI_{\text{diff}}(x,z) = \max(UI(x,z), \delta \cdot \max(n \in N(x,z) : 1/EI(e,x,n) \cdot UI_{\text{diff}}(n,z)))$
 - Same function can be used for UI

DOI Function Illustration

- $DOI(x | y, z) = \alpha \cdot API_{diff}(x) + \beta \cdot UI_{diff}(x, z) + \gamma \cdot D(x, y)$
 - API and EI is calculated once
 - UI is calculated every time z changes
 - D is calculated every time focal point y changes

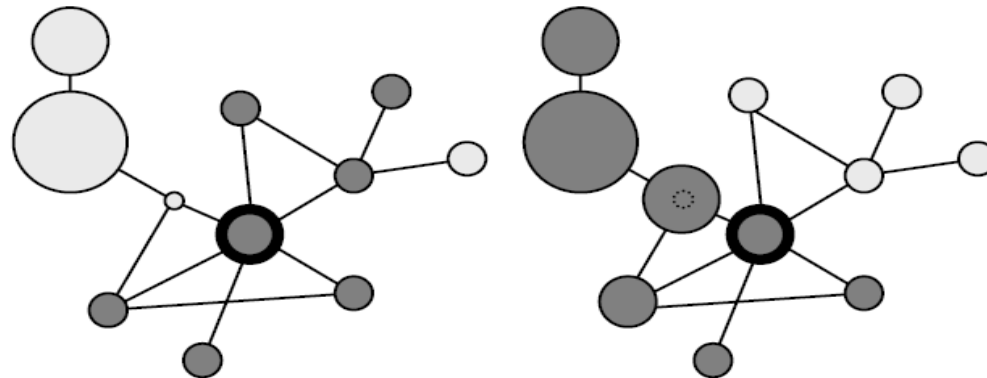


Fig. 1. By diffusing interest values over the network, we can use a greedy local search heuristic to find maximal interest subgraphs (dark grey) when starting from the focal node (circled)

“Show Context”

- Given giant graph G , calculate subgraph F that is small and captures as much relevant context around focal point as possible
- Formally, compute connected subgraph F of size at most S that contains y and has maximal total interest
- Avoid exhaustive search solution

Subgraph Calculation

- Use Greedy Approach
 - Start with subgraph $F = \{\}$
 - And candidate set $C = \{y\}$
 - Each iteration remove the candidate node with highest DOI
 - Add it to F and add its neighbors to C
 - Stop when $C = \{\}$ or $|F| = S$
- Runs in $O(S \log S)$

“Expand-on-Demand”

- User presented with local context of each node
- Decides to expand on a graph region by clicking a node, x
- Do not add all of x 's neighbors
- Implementation
 - Label each node with number of neighbors
 - Only return top N interesting neighbors with each click

Interesting Nodes Outside of Subgraph

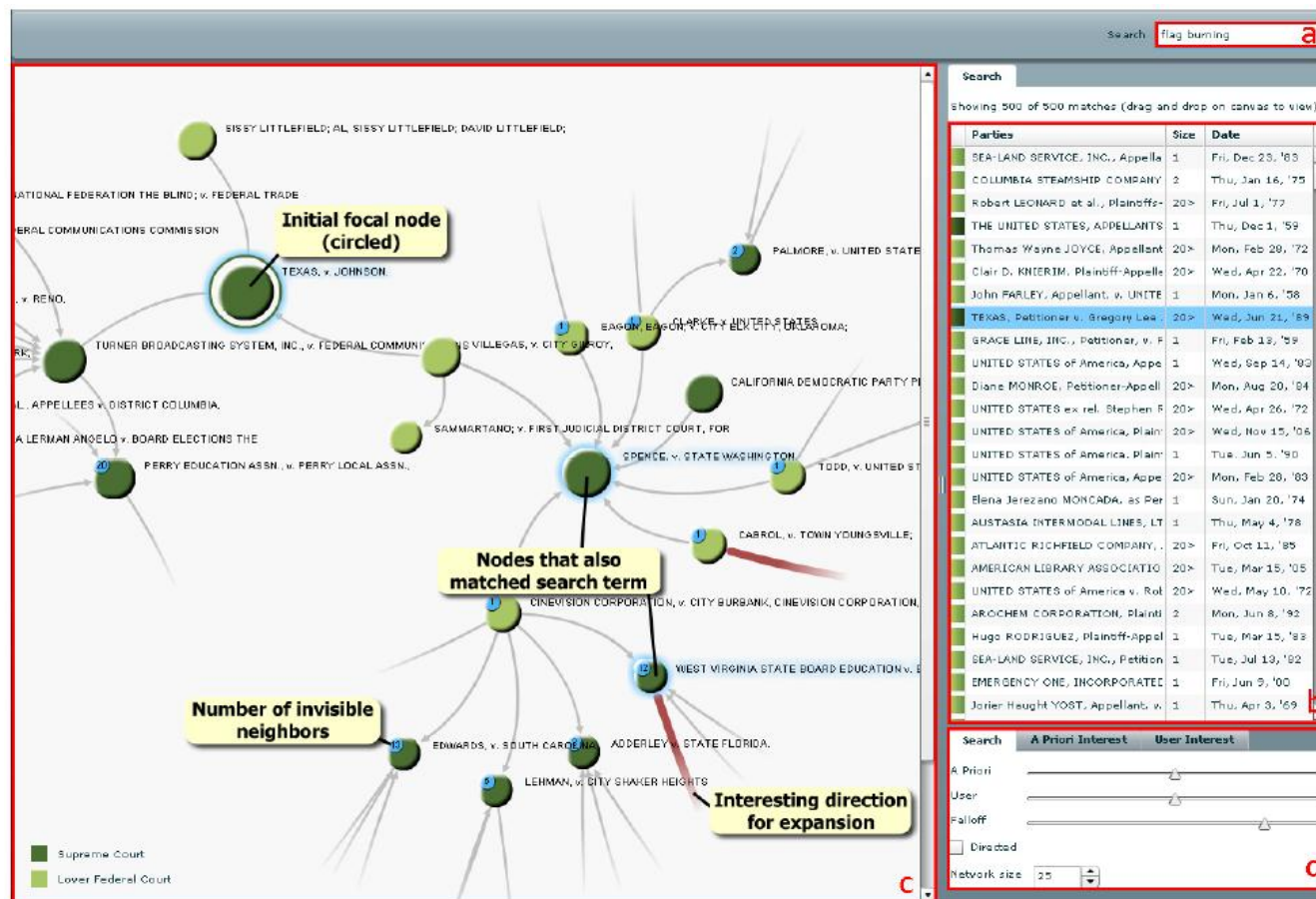
- Prevent user from having to needlessly expand dozens of items
- Use DOI function to highlight the n most interesting hidden parts of graph
 - $n < 5$ to not overburden user with too many choices
 - On expansion, top n will be relatively stable

Implementation

- **Server**
 - Maintains in memory minimum representation needed to calculate the entire graph and its interest functions
 - Precomputes all $API_{diff}(x)$
 - When given search terms and focal nodes, server maintains subgraph for each user
 - Can send layout of subgraph if user browsing from mobile phone

March 19, 2010

Client Side Implementation



Search Box

- Textual search on full dataset



March 19, 2010

List of Search Results

- Results of search
- Contains information about size of subgraph

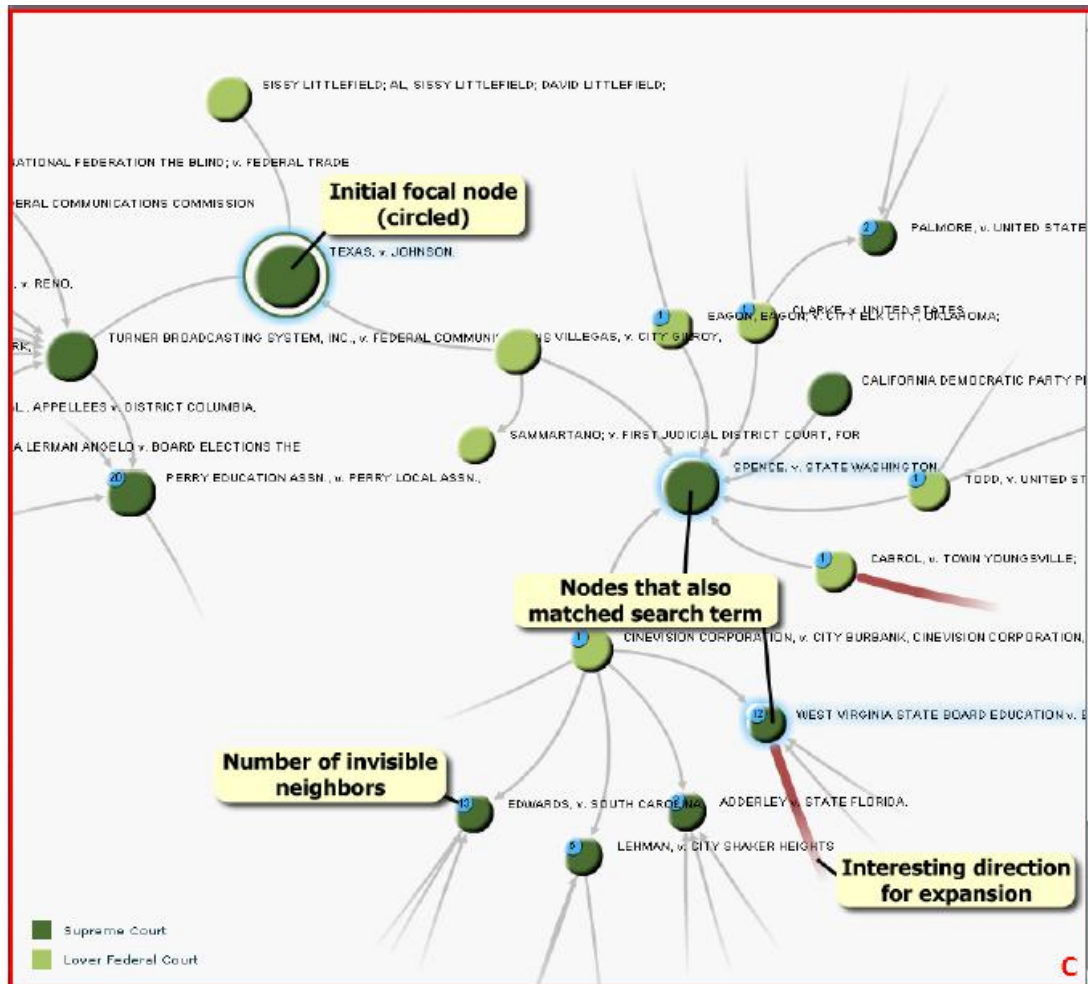
Search

Showing 500 of 500 matches (drag and drop on canvas to view)

| Parties | Size | Date |
|------------------------------------|------|------------------|
| SEA-LAND SERVICE, INC., Appella | 1 | Fri, Dec 28, '83 |
| COLUMBIA STEAMSHIP COMPANY | 2 | Thu, Jan 16, '75 |
| Robert LEONARD et al., Plaintiffs- | 20> | Fri, Jul 1, '77 |
| THE UNITED STATES, APPELLANTS | 1 | Thu, Dec 1, '59 |
| Thomas Wayne JOYCE, Appellant | 20> | Mon, Feb 28, '72 |
| Clair D. KNIERIM, Plainiff-Appelle | 20> | Wed, Apr 22, '70 |
| John FARLEY, Appellant, v. UNITE | 1 | Mon, Jan 6, '58 |
| TEXAS, Petitioner v. Gregory Lee | 20> | Wed, Jun 21, '89 |
| GRACE LINE, INC., Petitioner, v. F | 1 | Fri, Feb 18, '59 |
| UNITED STATES of America, Appe | 1 | Wed, Sep 14, '93 |
| Diane MONROE, Petitioner-Appell | 20> | Mon, Aug 20, '84 |
| UNITED STATES ex rel. Stephen F | 20> | Wed, Apr 26, '72 |
| UNITED STATES of America, Plain | 20> | Wed, Nov 15, '06 |
| UNITED STATES of America, Plain | 1 | Tue, Jun 5, '90 |
| UNITED STATES of America, Appe | 20> | Mon, Feb 28, '83 |
| Elena Jerezano MONCADA, as Per | 1 | Sun, Jan 20, '74 |
| AUSTASIA INTERMODAL LINES, LT | 1 | Thu, May 4, '78 |
| ATLANTIC RICHFIELD COMPANY, . | 20> | Fri, Oct 11, '85 |
| AMERICAN LIBRARY ASSOCIATIO | 20> | Tue, Mar 15, '05 |
| UNITED STATES of America v. Rol | 20> | Wed, May 10, '72 |
| AROCHEM CORPORATION, Plainti | 2 | Mon, Jun 8, '92 |
| Hugo RODRIGUEZ, Plainoff-Appel | 1 | Tue, Mar 15, '83 |
| SEA-LAND SERVICE, INC., Petition | 1 | Tue, Jul 13, '92 |
| EMERGENCY ONE, INCORPORATEE | 1 | Fri, Jun 9, '00 |
| Jorier Haight YOST, Appellant, v. | 1 | Thu, Apr 3, '69 |

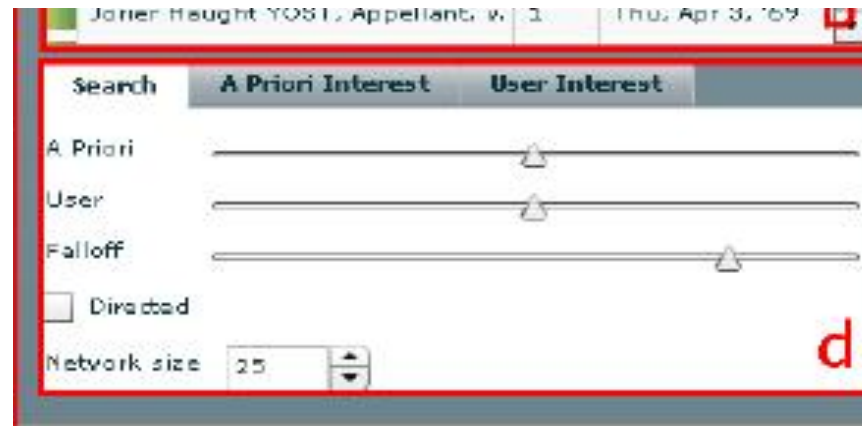
March 19, 2010

Main Canvas



Parameter Panel

- Directly manipulate the weights of DOI function and limits to display
 - $DOI(x | y, z) = \alpha \cdot API_{diff}(x) + \beta \cdot UI_{diff}(x, z) + \gamma \cdot D(x, y)$



Example: Legal Citations

- 15 gigabytes of legal documents containing all Supreme Court and federal court case decisions
- Lawyers need this information to interpret precedence and understand how relevant issues were ruled in prior cases
- Westlaw digital library
 - Text based with powerful search
 - Requires link clicking to see related documents

Citation Graph Properties

- 300,000 documents (nodes) and 3.3 million citations (edges)
 - Average degree of node is 1 to 200
 - Several cases with thousands of citations

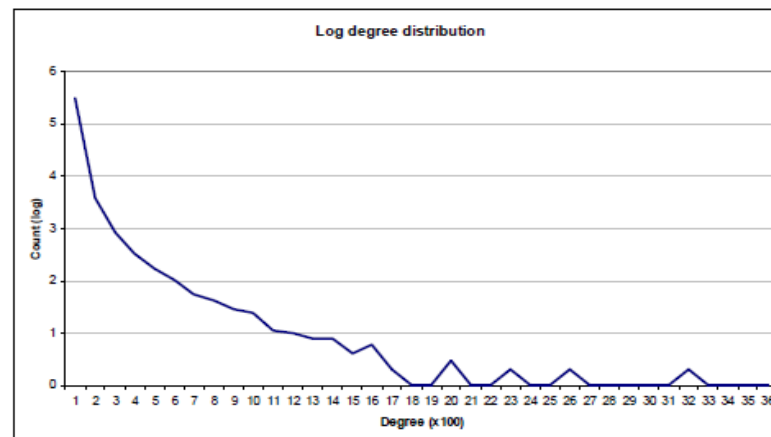


Fig. 5. Log-degree distribution of the case citation graph. Citation count includes both in and out citations.

Domain DOI Definitions

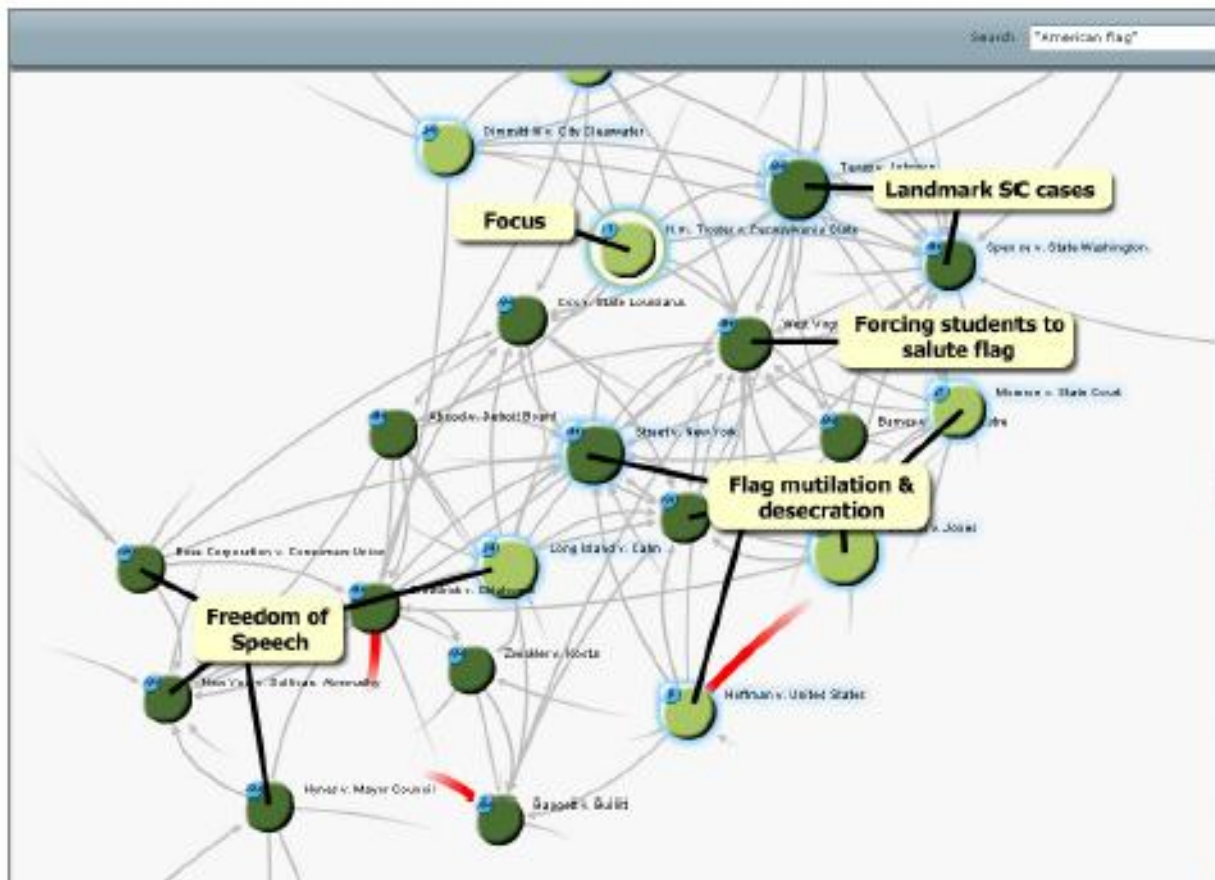
- $API(x)$ – higher interest assigned to Supreme Court cases and to high in-degree nodes
- $UI(x,z)$ – scores based on the search term
- No $EI(e,n,x)$ terms because all citations treated equally

Four Anecdotes of Utility

- Importance
- Overview
- Relevance
- Reasoning

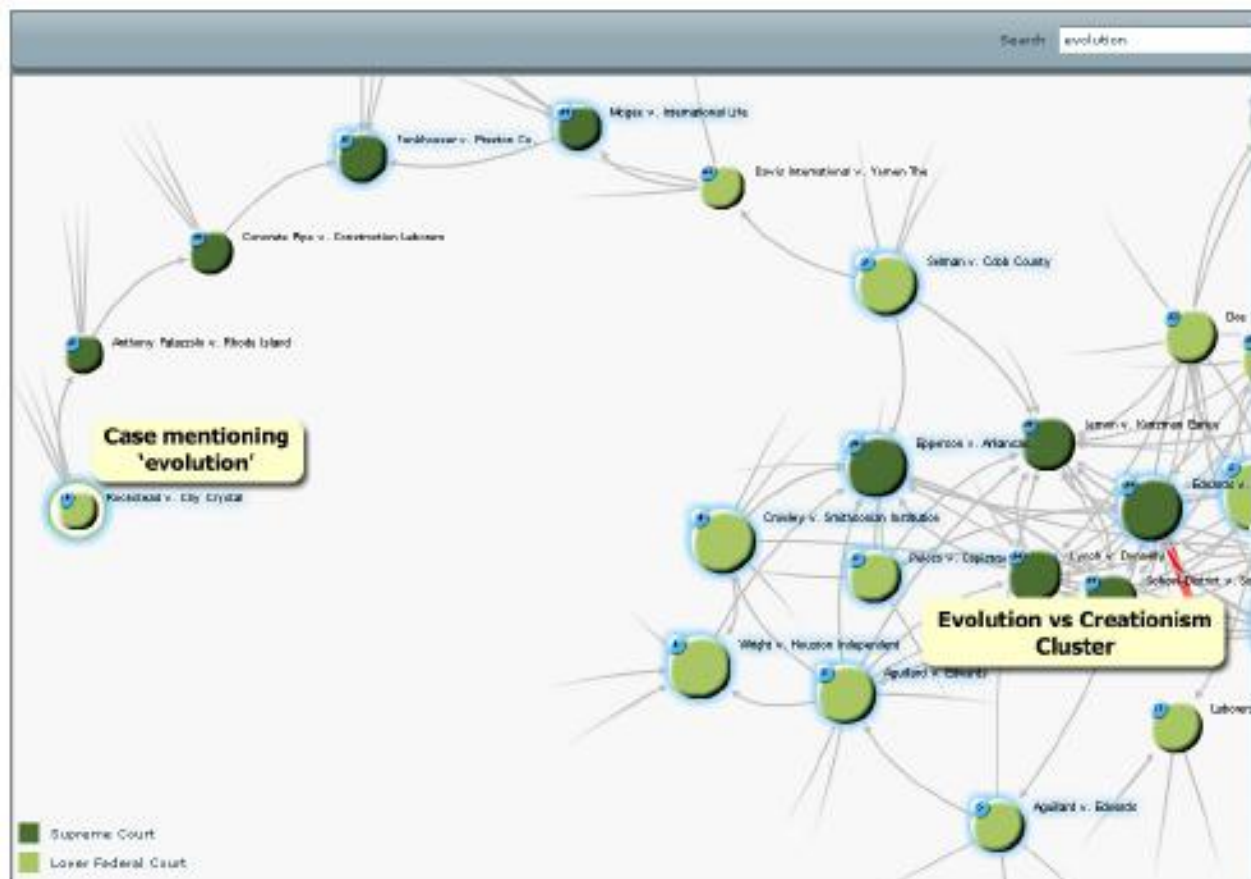
March 19, 2010

Show Overview of Subfield



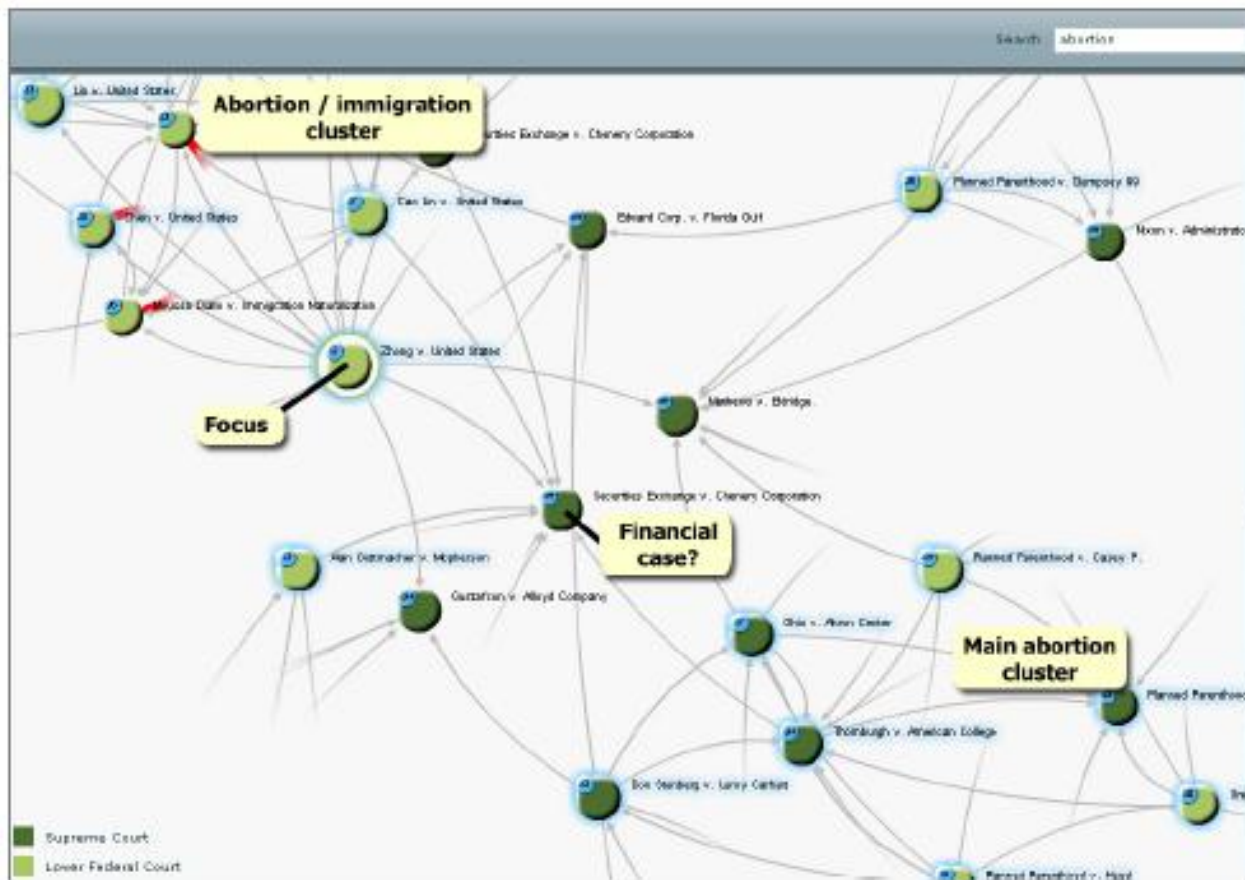
March 19, 2010

Show If Focus is As Expected



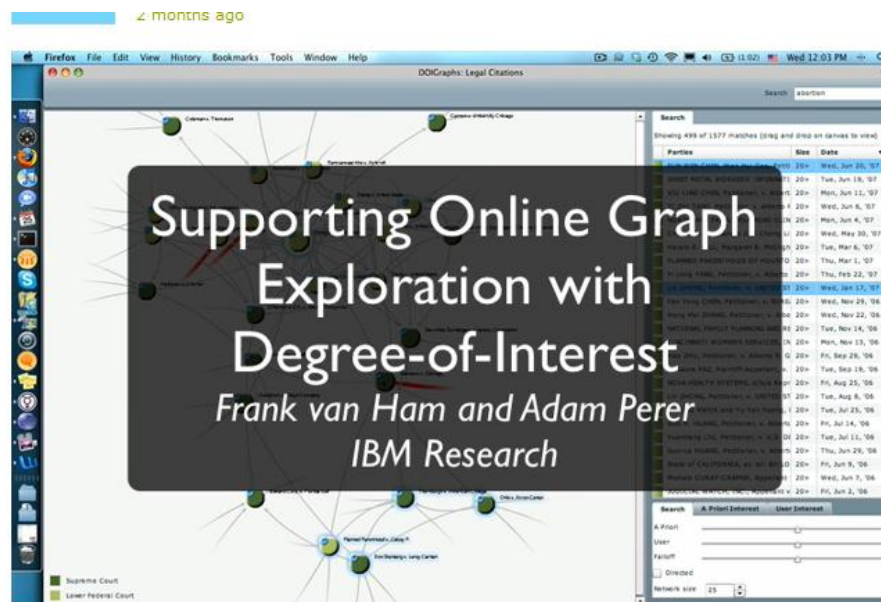
March 19, 2010

Show Why Cited Case is Relevant



March 19, 2010

Demo: <http://vimeo.com/8570431>



Discussion and Future Work

- Other domains
 - Social networking – ego centric views
 - Internet Browsing – local maps around current page using user's information needs
 - Properties and metrics for DOI should depend on the domain
 - Social networking could use betweenness centrality or PageRank like algorithms
 - These metrics are expensive

Improvements

- Improvements
 - Multiple Foci
 - UI enhancement
 - Local context problem
- Short Comings
 - Edge Finding
 - Extreme Degree Nodes
 - Following Edges

Summary

- DOI from trees to graphs
- DOI to include inferred interest from search progress
- Diffusion of interest over entire graph
- Novel Visualization to reduce complexity and guide users to hidden content