

Tuple – InfoVis Publication Browser

CS533 Project Presentation
by Alex Gukov



Project Goals

- Provide visual overview of InfoVis publication history
 - Key authors and papers
- Identify key directions
 - Major research categories
 - Influential authors and papers within a categories
 - Related categories



Project Overview

- Process article metadata to generate category subdivision
 - 10 sub-fields found
- Visualize article citation graph
 - Articles as graph nodes and citation links as edges
 - Edge instead of background color for category encoding
 - Provide interactive controls for exploration

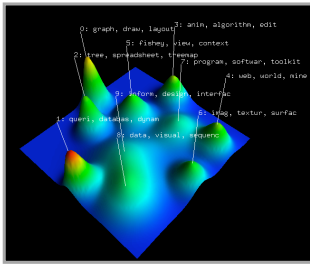


Text clustering

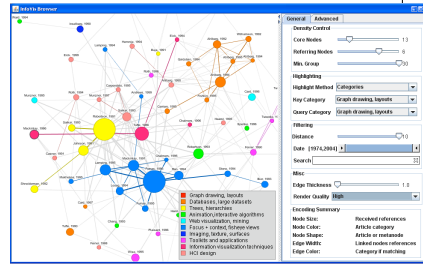
- Generate a word occurrence matrix from given metadata
 - Titles, keywords, abstracts
 - Stemming to improve search correlation
- k-means to cluster the articles
 - Best for small number of groups (10)
 - Cosine distance measure
- Use Cluto toolkit



Text clustering



Application Overview



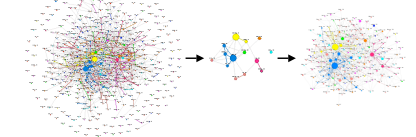
Graph Layout

- Edges as springs
 - Same category edges have lower rest length
- Node repulsion
 - Ensures clearance
- Weak centralization force
 - Handles disconnected components
- Appearing nodes positioned at the average of visible neighbours



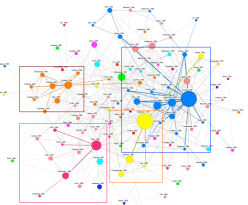
Extracting key articles

- Number of received references indicates importance
 - Use as node size
- Filter in two steps to increase coherence and connectivity
 - Want to start with the key articles and then explore details



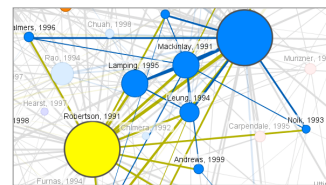
Encoding Individual Categories

- How segmented is a category ?
- How do categories compare in number or importance of nodes ?



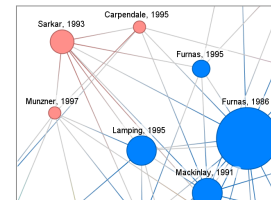
Encoding Category Pairs

- How tightly are categories connected ?
- Did one category originate from another ?



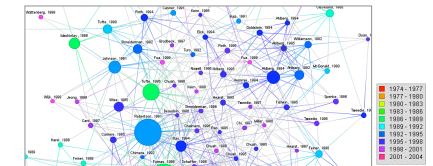
Encoding Reference Direction

- Individual paper sources
- Did one category originate from another ?



Encoding Publication Time

- Oldest / most recent papers at a glance ?
- Relationship between date and influence ?



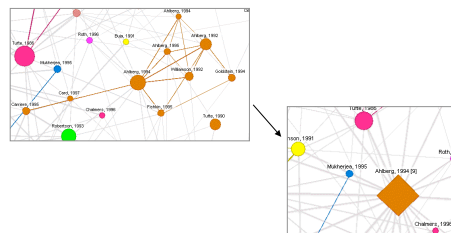
Component abstraction

- Often want to study high level features
 - Number of disconnected components
 - Relative component sizes
 - Category-level reference directions
- May want to reduce clutter



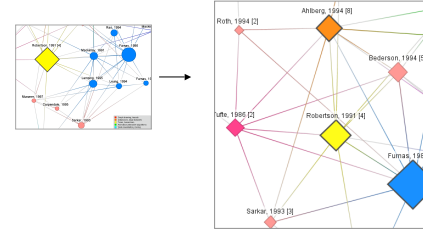
Component abstraction

- Group linked articles within the same category



Component abstraction

- Source identification made easier



Implementation Tools

- MySQL data backend
 - Initial processing and retrieval
- gCluto application for text clustering
- Java Swing and Prefuse user interface





Application Demo

- Node density control
- Additional highlighting options
 - Category connectivity
 - Date highlighting
- Filtering and search options
 - Neighbour visibility
 - Time range filtering
 - Title and author search
- Interactive features



Future Improvements

- Graph layout dynamic stability
 - Improve initial positioning when making a node visible
 - Layout calculation to minimize displacement of visible nodes
 - Perform simulation in run-once mode and smoothly interpolate
- Co-authorship graph
 - Useful for studying development of collaboration groups
 - Unclear if paper categories have any role
- Article summary table
 - Sorted table of search results, visible items, etc..
 - Immediate information lookup