Categorical Relationships in History of InfoVis Publications

CS533C Project Update
by Alex Gukov
Goals

- Provide visual overview of InfoVis publication history
  - Author collaboration network
  - Paper co-citation network
- Identify key influences
  - Major research categories
  - Influential authors and papers within a categories
  - Related categories
Dataset

- Filtered 2004 InfoVis Contest data
  - InfoVis publication history from 1995 to 2002
  - Original data cleaned up by Indiana University contestants

- Medium size network
  - 614 InfoVis articles with detailed metadata
  - 8502 references with limited metadata
  - 1036 authors

- Paper metadata
  - Title, year, abstract, keywords,
Previous Work

- Indiana University contest entry

- Node link diagram of highly cited papers, published authors
- Clearly identifies important papers, authors through node size
- Does not relate authors, papers to research categories
Previous Work

- IN-SPIRE by Pacific Northwest National Laboratory

- Scatter plot of publications
- Plot positioning based on themes extracted from metadata
- Clearly identifies dominant themes
- Does not make use of citation data
Criticism

Want to relate publication network data with corresponding category information
Proposed Visualization

- Reduce the data set by using highly cited papers, published authors
- Visualize collaboration and co-citation networks with node-link graphs
- Augment the plots with category information using background color
Graphing publication networks

- Node-link diagrams with papers, authors as graph nodes
- Node size proportional to the number of received citations
- Node color
  - Paper publication date
  - Number of papers written by an author
- Force-directed layout using topology and category information as cues
Visualizing categories

- Identify a small number of categories from paper metadata
  - Process titles, abstracts, keywords
  - Reduce dimensionality, cluster
- Partition space around the graph nodes after layout is complete
- Color the background of each node with corresponding category (use light colors)
  - Author is assigned the mean category of his/her publications
Mockup
Implementation

- Category identification
  - Use PCA to reduce noise
  - Use k-means on the resulting data

- Graph visualization
  - Prefuse Visualization toolkit (Java)
  - Built-in force-directed layout engine

- Background space partitioning
  - Partition using a Voronoi diagram
  - Use CGAL geometry toolkit (C++)
Current Progress

- Data graphing
  - Setup Prefuse and experimented with a sample social network graphing application

- Category identification
  - Access database converted to xml
  - Preprocessing data for use in Matlab