Modeling Uncertainty due to Data/Visual Transformations using Sensitivity Analysis

Kwan-Liu Ma
University of California, Davis

Uncertainty Aware Visual Analysis

Let know confidence level of reasoning, and
Input sensitivity of data & visual transformations

1. Uncertainty Analysis
   - Determining the uncertainty on estimates of the output given the input data
   - Adapting data transformations to account for uncertainty
   - Process view of visual analysis
   - The focus of our first FODAVA project
2. Sensitivity Analysis
   - Determining the relationship between the uncertainty of the output and the uncertainty of the input for a number of transformations
   - Estimates of the transformation sensitivity to input data
   - Variational view of visual analysis
   - The focus of our second FODAVA project

Objectives

- Study sensitivity analysis to guide the evaluation of uncertainty of data and insights in the visual analysis process
- Create variational views of the visual analytics process
- Discover the factors that mostly contribute to output variability
- Find stable/unstable regions of the different transformations within the data space
- Understand interaction between variables, transformations, and output
- Focus on network data

Proposed Research Tasks

1. Semi-automatic extraction of sensitivity information
   - Develop sampling based methods for the types of data that differentiation cannot be used
   - Estimate importance of region in the parameter space
   - Benchmark the accuracy of the results produced by common transformations
2. Differential and sampling-based sensitivities of graph-based metrics and transformations
3. Sensitivity-guided visual representations and interaction

Centrality-based Network Analysis

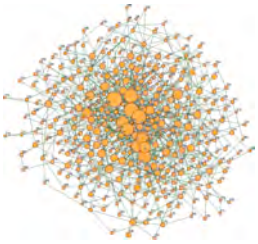
- Centralities (degree, between-ness, closeness, eigenvector, Markov, …) indicate how *important* a node is in a network.
- We can understand implicit relationships in a network by studying its ***sensitivity*** and *stability* in terms of different metrics for centrality
- The study allows us to determine **competitiveness** & **collaborativeness** between nodes or clusters, the robustness of a metric, …

A protein interaction network with lighter color nodes more central

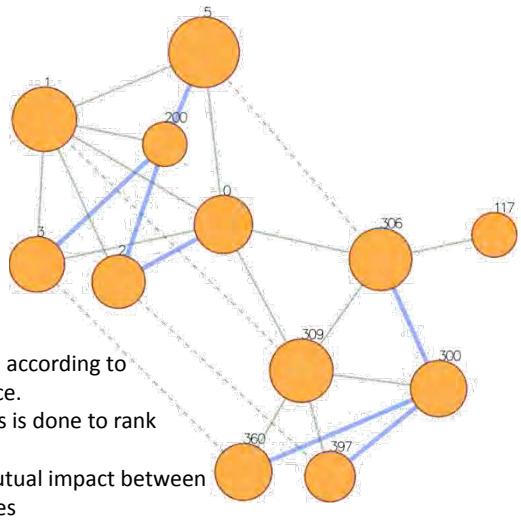
Centrality Sensitivity

- Compute sensitivity as the derivative of the centrality function
- Approximate derivatives of centrality using finite difference (before and after small changes are made stochastically)
- Validate by computing the mean square error of the linear fit between the approximated and analytical values (for the Eigenvector and Markov centralities)

Sensitivity Analysis to Discover Hidden Relations

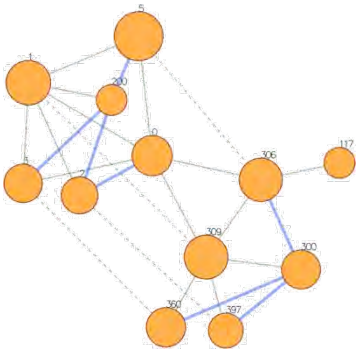


A VAST challenge dataset
400 cell phones

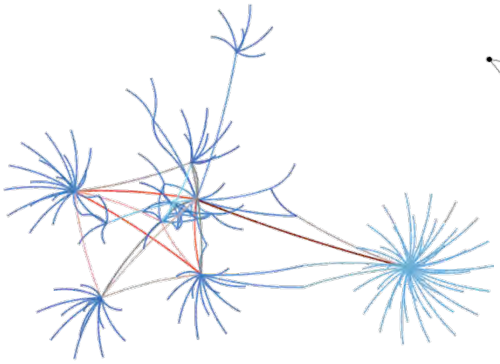


1. Nodes are filtered according to Markov importance.
2. Sensitivity analysis is done to rank the edges
3. Identify strong mutual impact between unconnected nodes

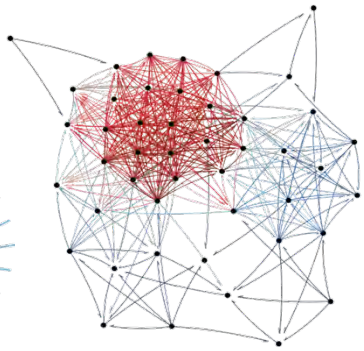
Sensitivity Analysis to Discover Hidden Relations



Overview of Sensitivity



Friendster social network
Links exhibit negative sensitivity between cluster centers

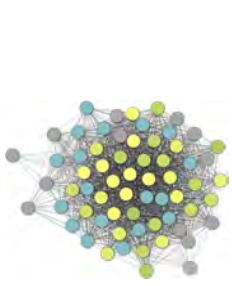


Astrophysics co-author network
One competitive network and one collaborative network

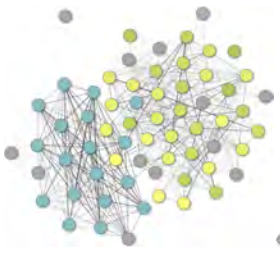
Visualizing Proximity Data

MIT Reality dataset

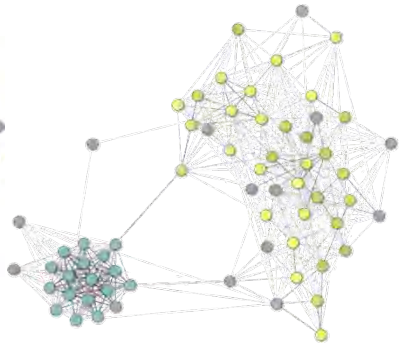
Blue: Sloan school Green: Media Lab Gray: Unidentified



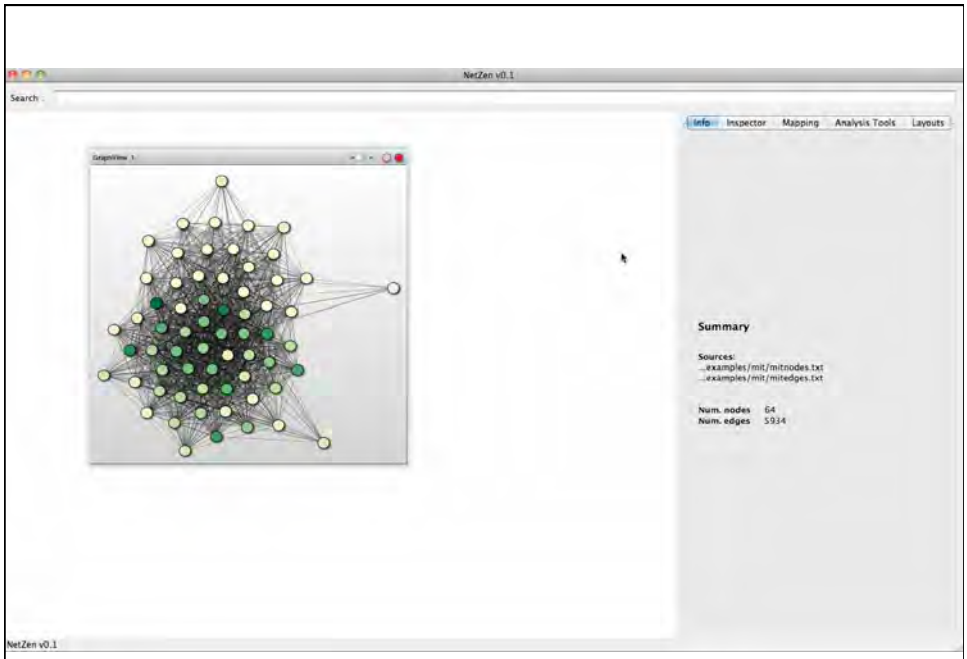
Force directed layout
Most users were in close proximity to each other!!



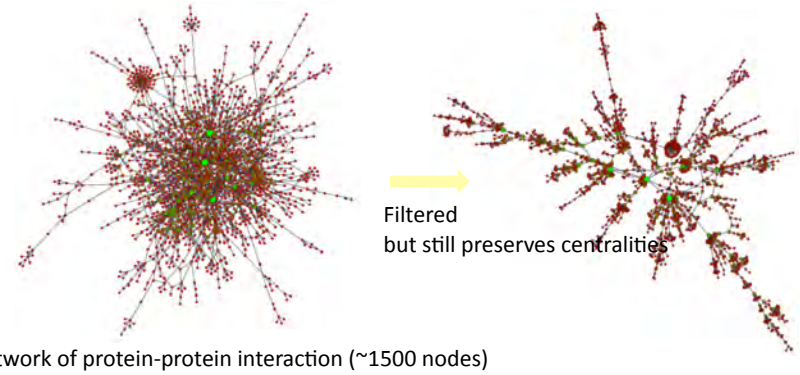
only with positive sensitivity



Filtered with sensitivity magnitude



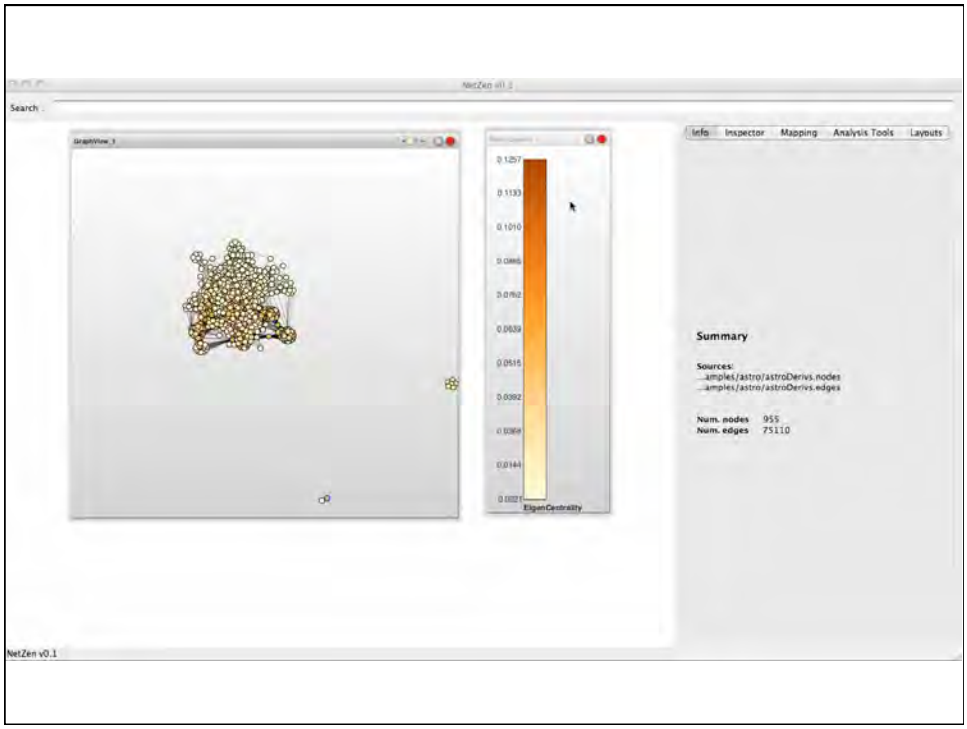
Network Simplification



Network of protein-protein interaction (~1500 nodes)

1. Find the minimum spanning tree weighted by the derivatives
2. Add back a certain number of highly weighted edges to retrain a core network

Validation: Nodes are central should remain central.

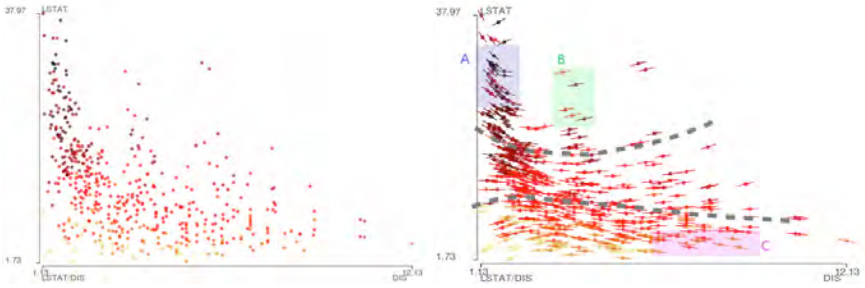


Sensitivity Guided Visualization of Multidimensional Data

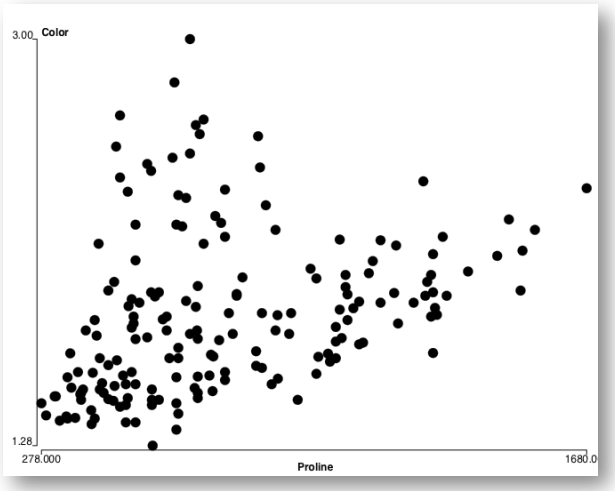
- Multi-dimensional sensitivity visualization
- Encoding sensitivity in summary visual representations
- Sensitivity-guided navigation of complex multidimensional data

Flow-Based Scatterplots

- Augment scatterplots for visualizing multidimensional data by using sensitivity coefficients to highlight local variation of one variable with respective to another.
- The resulting visualization resembles a flow field when treating sensitivity as velocity.



Flow-based Scatterplots

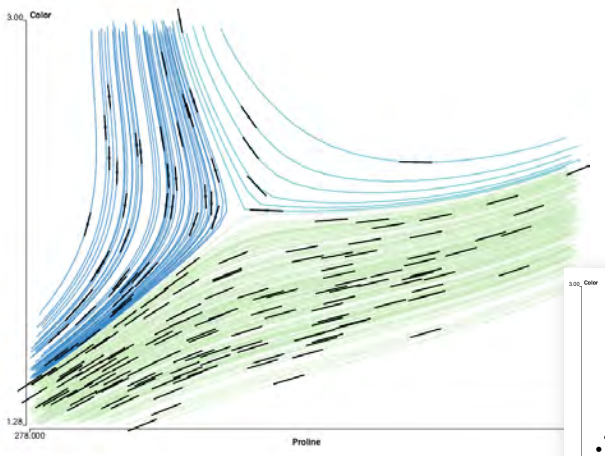


A wine data set

X: Proline

Y: Color

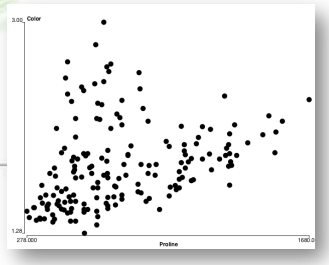
Flow-based Scatterplots



A wine data set

X: Proline

Y: Color



Other Considerations

- Scalability
  - Scalable sensitivity analysis of large graphs
- Discontinuities
  - We have assumed that variables have an underlying continuous distribution, but some data not.
- Generalization
  - We want to generalize our process towards a more theoretical view of visual analytics
- NetZen