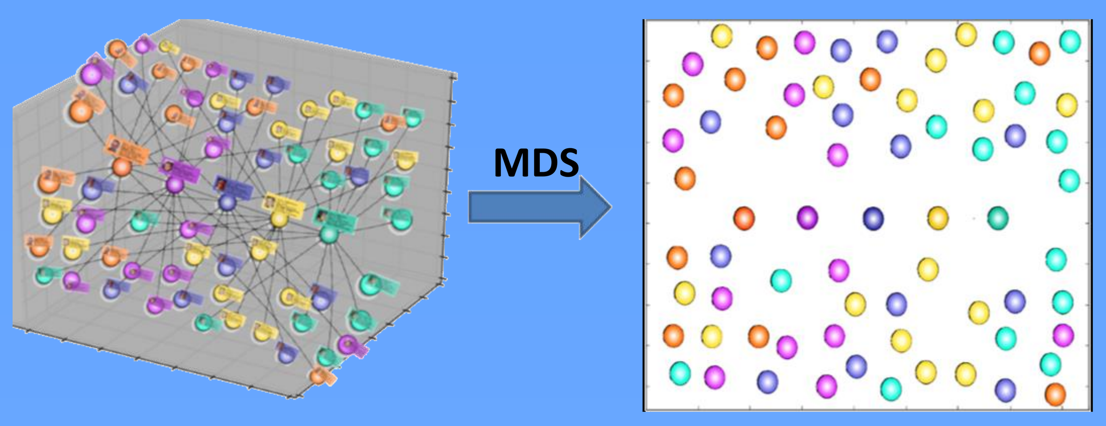
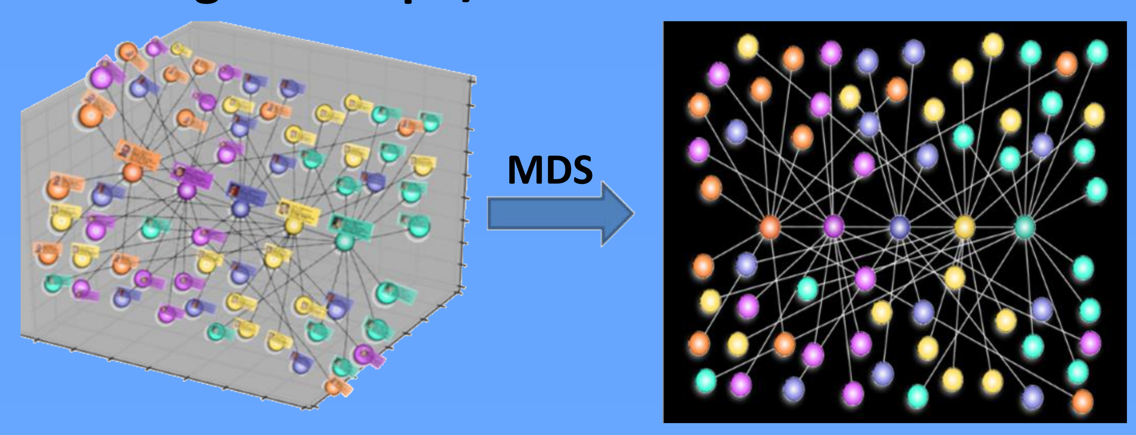
Visualization of Multivariate Data with Network Constraints using Multi-Objective Optimization

Bhavya Ghai[¥], Alok Mishra[‡], Klaus Mueller[*]
Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University, SUNY, USA

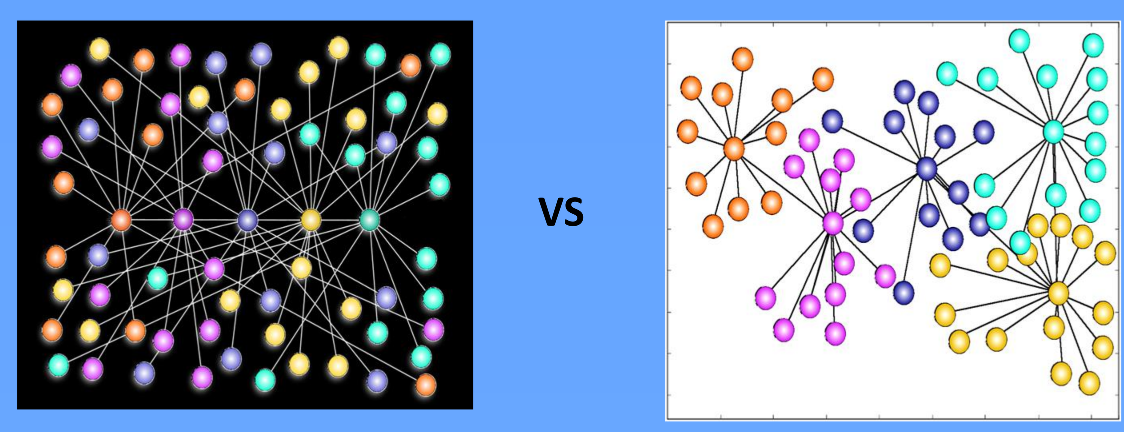
High Dimensional Data Visualization



High-D Graph/Network Visualization



High-D Graph/Network Visualization



MDS

Our Approach

Motivation

- Dimensionality reduction techniques play a key role in data visualization.
- Common dimensionality reduction techniques such as MDS, PCA, etc. do a decent job in projecting high-dimensional data into lower dimensions.
- But in the case of high-dimensional data with network constraints, they simply ignore the relationship between the nodes which might result in non-planar graphs with many intersecting edges.

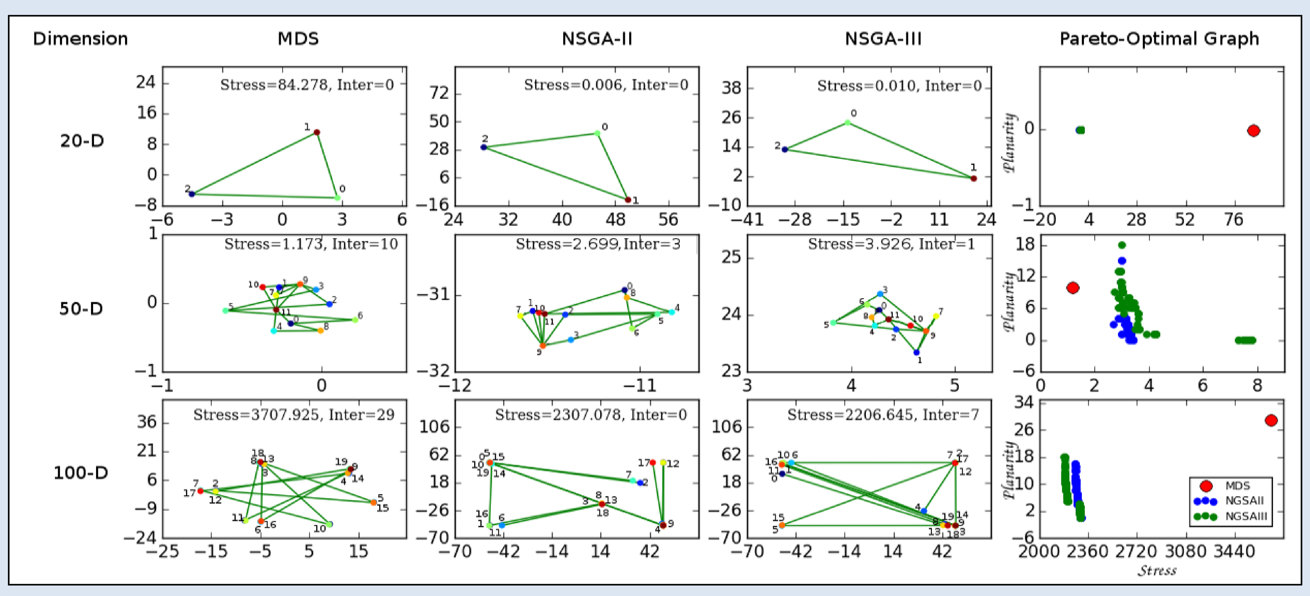
Objective

This problem of dimensionality reduction can be visualized as searching for an Optimal configuration of points in lower dimensional space. The term 'Optimal' here means the configuration which satisfies two objectives i.e. minimize stress and maximize planarity.

Method

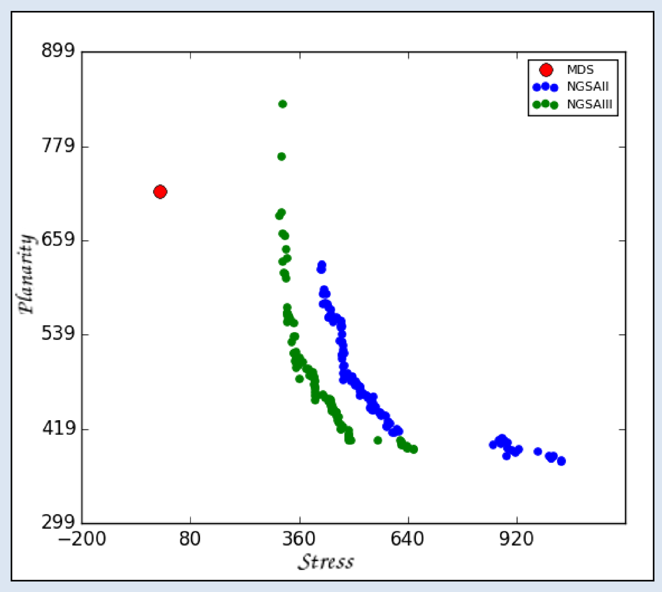
- Models dimensionality reduction for network data as a multi-objective optimization problem.
- Draws graph/network in lower dimensions such that **Planarity is Maximized** and **Stress is Minimized** simultaneously.
- Stress — Stress is a loss function which quantifies the preservation of relative distances in higher dimensional space to lower dimension space.
- Planarity — In graph theory, a graph is said to be planar if no two edges intersect. We have tried to quantify planarity in terms of number of intersections.
- Both objectives are equally important and are optimized together.
- In our approach we are using two genetic algorithms namely, NSGA-II and NSGA-III.

Observation



Comparison of Metric MDS, NSGA-II and NSGA-III using 3 datasets with 20, 50 and 100 dimensions respectively.

- Our Algorithms return a set of non-dominated solutions represented by a Pareto-Optimal front.
- Genetic algorithms outperformed MDS for some cases.
- In other cases, genetic algorithms gave solutions with significantly lower number of intersections for a slight increase in stress value.
- Apart from our own datasets, we also tested our algorithm on the World Football Dataset. As expected, Genetic algorithms performed fairly well on planarity for a slight increase in stress.



Pareto-Optimal graph comparing MDS, NSGA-II and NSGA-III for the World Soccer Dataset

Future Work

- Genetic Algorithms are computationally expensive. We intend to develop a GPU version of the genetic algorithms for better performance.
- Explore other optimization models which might achieve comparable accuracy in less time
- For dense graphs we intend to integrate edge bundling into our approach and investigate the effectiveness for different type of graphs

Emails: [¥]bghai@cs.stonybrook.edu,
[‡]almishra@cs.stonybrook.edu,
[*]mueller@cs.stonybrook.edu

This research was partially supported by NSF grant IIS 1527200
.

VISUAL ANALYTICS AND IMAGING LAB

Stony Brook University