Mathematical Foundations of Multiscale Graph Representations and Interactive Learning

Mauro Maggioni, Eric Monson, R. Brady
Dept. of Mathematics and Computer Science
Duke University

FODAVA Annual Review 2011, Georgia Tech
12/8/2011

Joint work with: W.K. Allard, G. Chen.

Support: NSF/DHS, partial support from ONR

Research topics

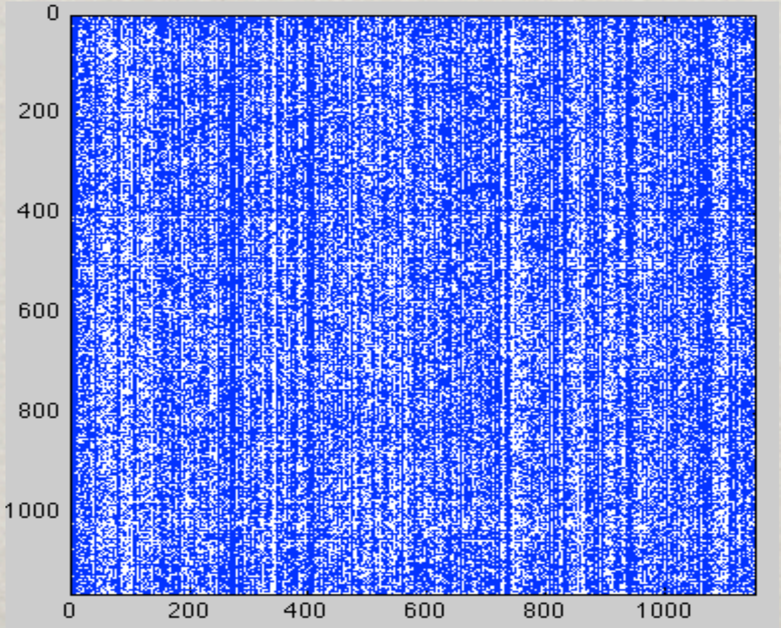
- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

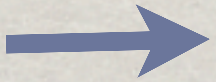
Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Random walks on data & graphs

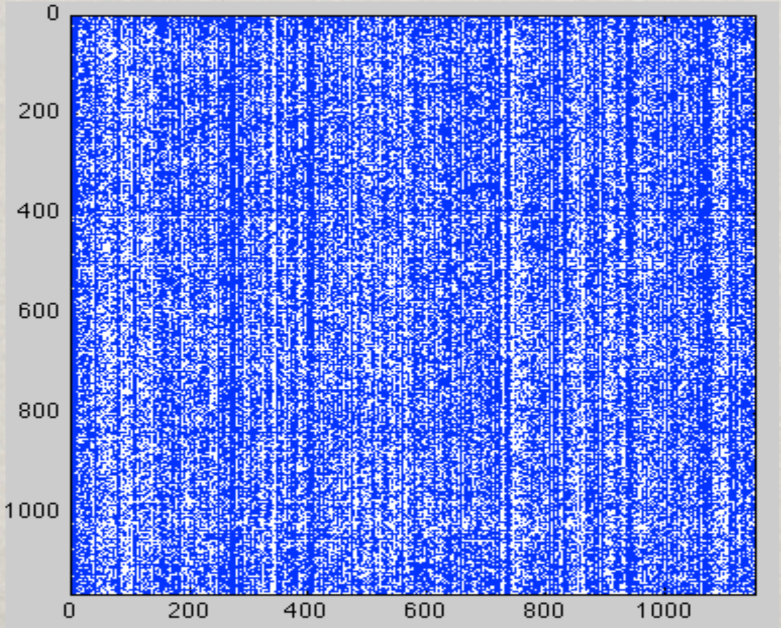
- One may connect data points to form a graph, with edges weighted by the similarity of data points.
- One can then construct a random on the data points, which may be used for a variety of tasks:
  - dimension reduction
  - clustering, classification, regression, etc..
  - diffuse information (e.g. labels) on data
  - study geometric properties of data

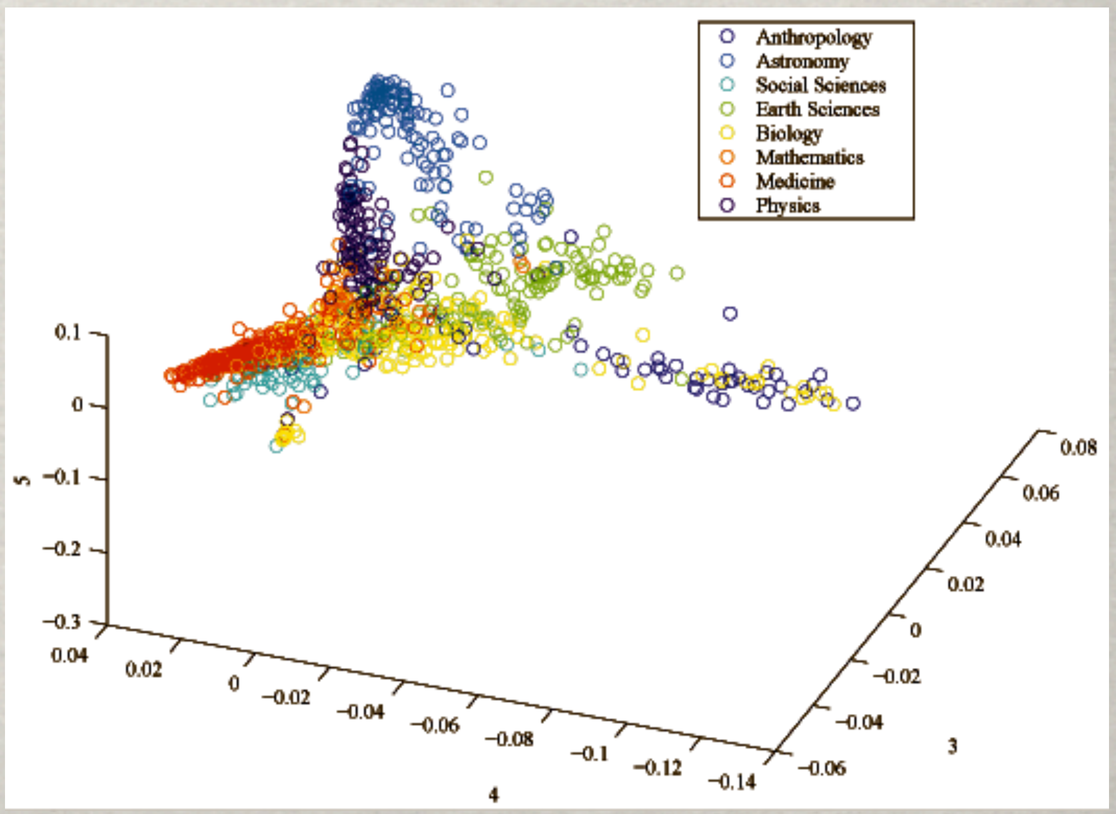




Random walks on data & graphs

- One may connect data points to form a graph, with edges weighted by the similarity of data points.
- One can then construct a random on the data points, which may be used for a variety of tasks:
  - dimension reduction
  - clustering, classification, regression, etc..
  - diffuse information (e.g. labels) on data
  - study geometric properties of data



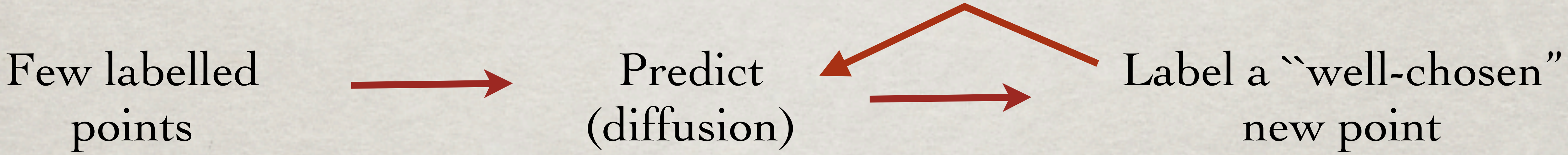


Active Learning

Given: full data set, (expensive) queries to an expert.

Goal: label all data points.

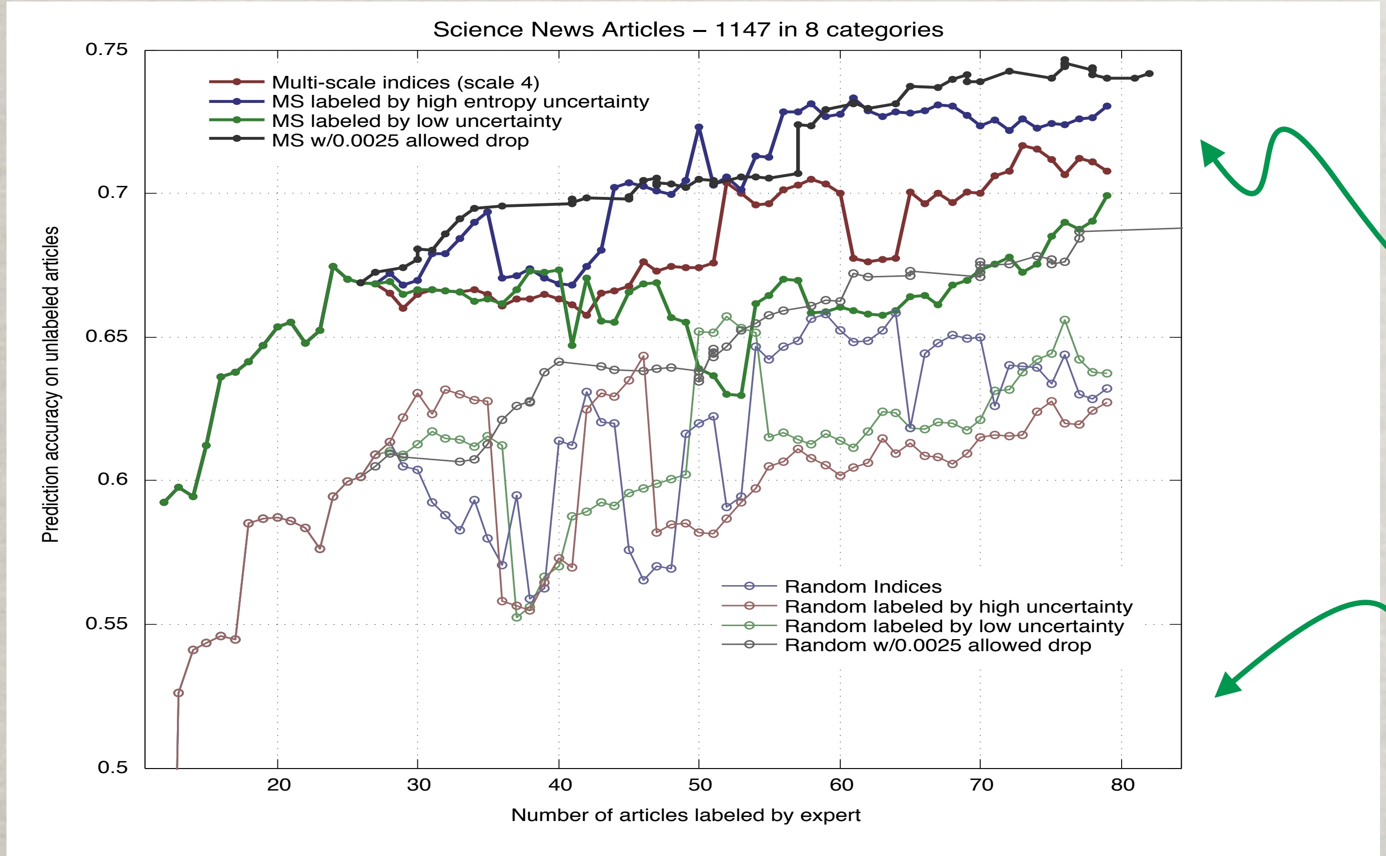
Proceed iteratively, querying labels at points with highly uncertain predictions + well-distributed on the data (multiscale)



Few labelled points

Predict (diffusion)

Label a ``well-chosen” new point



Accuracy of predictions

1147 Science News articles, 8 categories

Cost: # of labeled points

Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Example: text documents

With J. Guinney, S. Mukherjee and P. Febbo

Use dictionaries on graphs for sparse classification/regression. E.g.: N documents in $\mathbb{R}^D$, compute multiscale dictionary Φ $(D \times M)$ on the D words. If f maps documents to their topic, write $f = X\Phi\beta + \eta$ and find β by

$$\text{argmin}_\beta ||f - X\Phi\beta||_2^2 + \lambda||\{2^{-j\gamma}\beta_{j,k}\}||_1\,,$$

which is a form of sparse regression. (λ, γ) are determined by cross-validation.

Application to gene array data (prostate cancer). Not only better predictions, but more interpretable results as our multiscale genelets better related to relevant pathways than eigengenes.

Source of data: Nakagawa T, Kollmeyer T, Morlan B, Anderson, S, Bergstralh E, et al, (2008) A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy, Plos One 3:e2318.

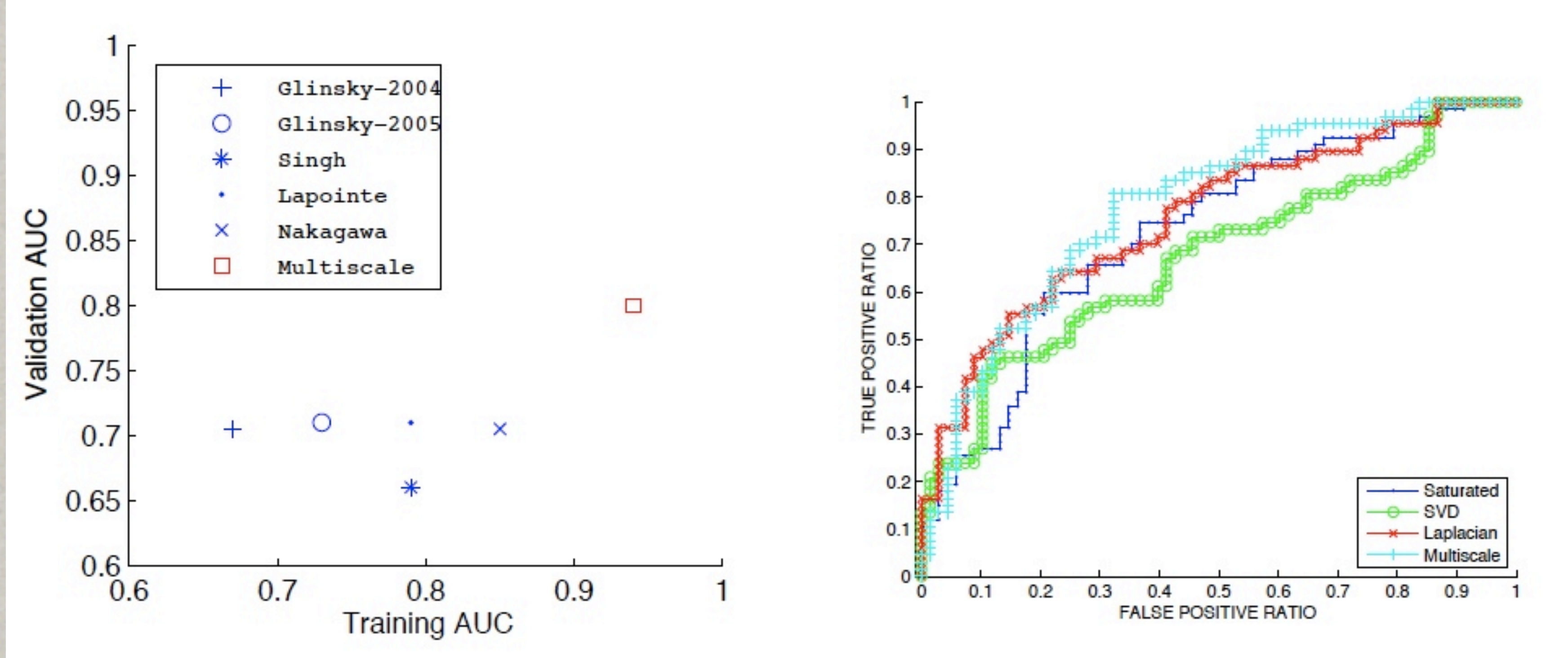
Example: text documents

With J. Guinney, S.
Mukherjee and P. Febbo

Use dictionaries on graphs for sparse classification/regression. E.g.: N documents in $\mathbb{R}^D$, compute multiscale dictionary Φ $(D \times M)$ on the D words. If f maps documents to their topic, write $f = X\Phi\beta + \eta$ and find β by

$$\text{argmin}_{\beta} ||f - X\Phi\beta||_2^2 + \lambda ||\{2^{-j\gamma}\beta_{j,k}\}||_1 \, ,$$

which is a form of sparse regression. (λ, γ) are determined by cross-validation.



Application to gene array data (prostate cancer). Not only better predictions, but more interpretable results as our multiscale genelets better related to relevant pathways than eigengenes.

Source of data: Nakagawa T, Kollmeyer T, Morlan B, Anderson, S, Bergstralh E, et al, (2008)
A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy, Plos One 3:e2318.

Research topics

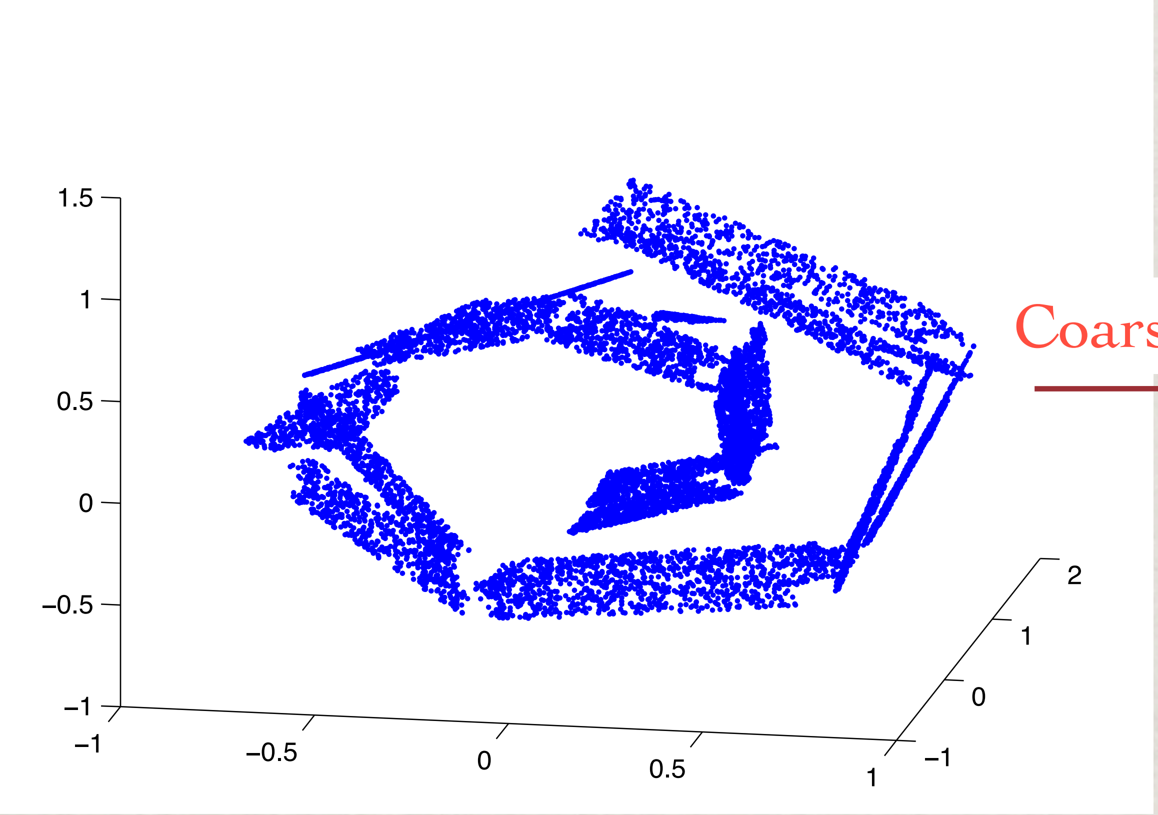
- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Research topics

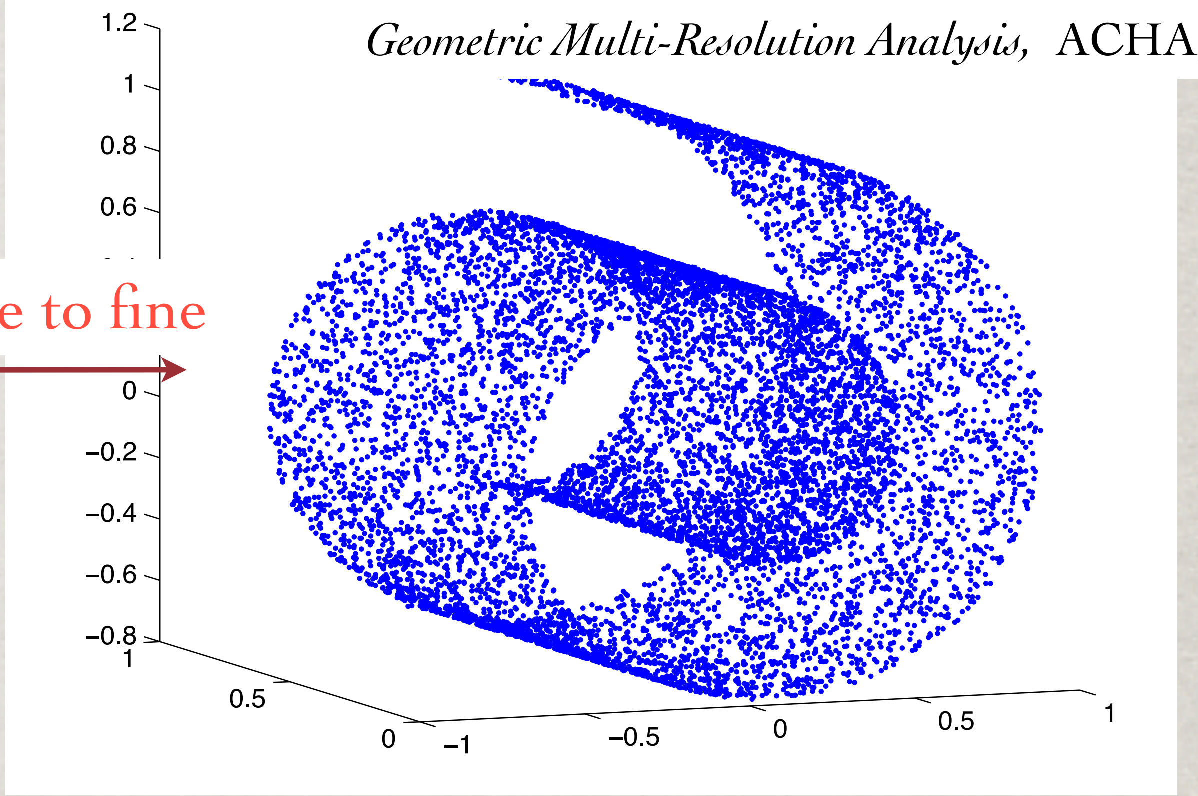
- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Geometric Wavelets, simple example

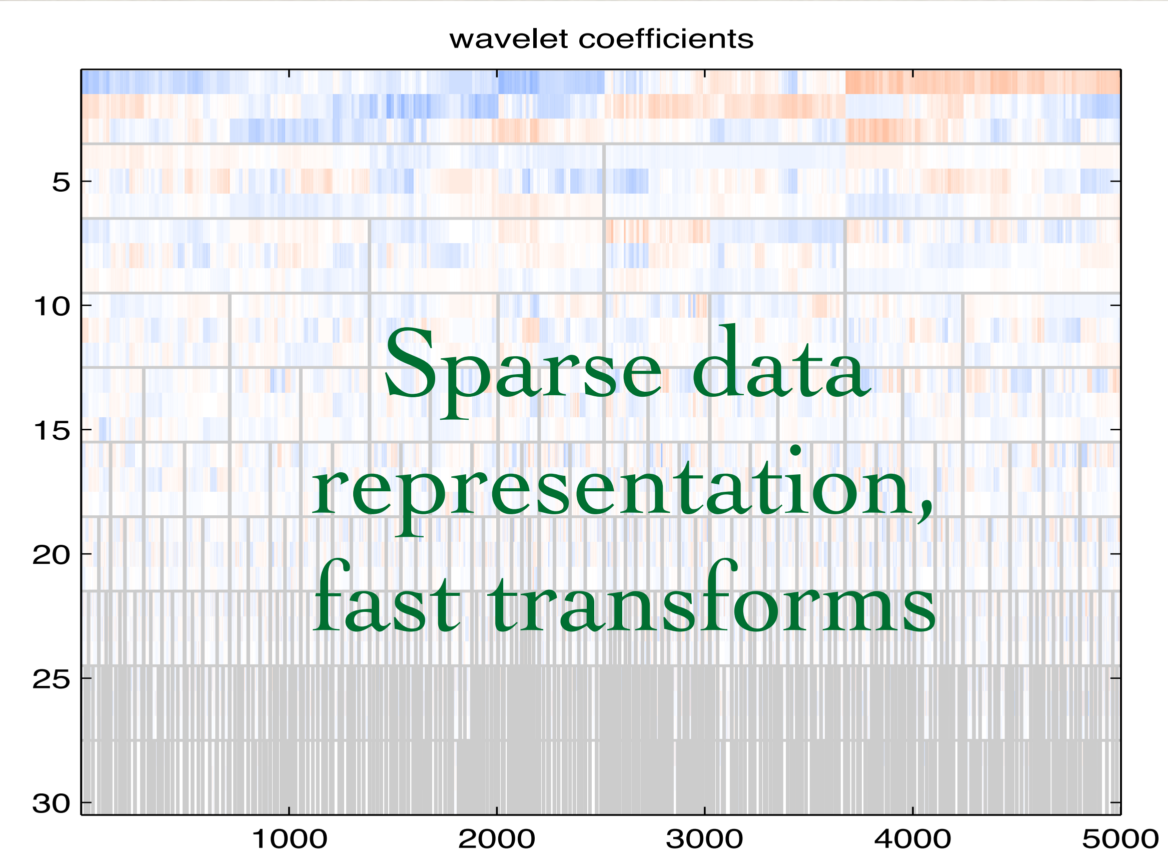
With W.K. Allard and G. Chen,
Geometric Multi-Resolution Analysis, ACHA, 2011



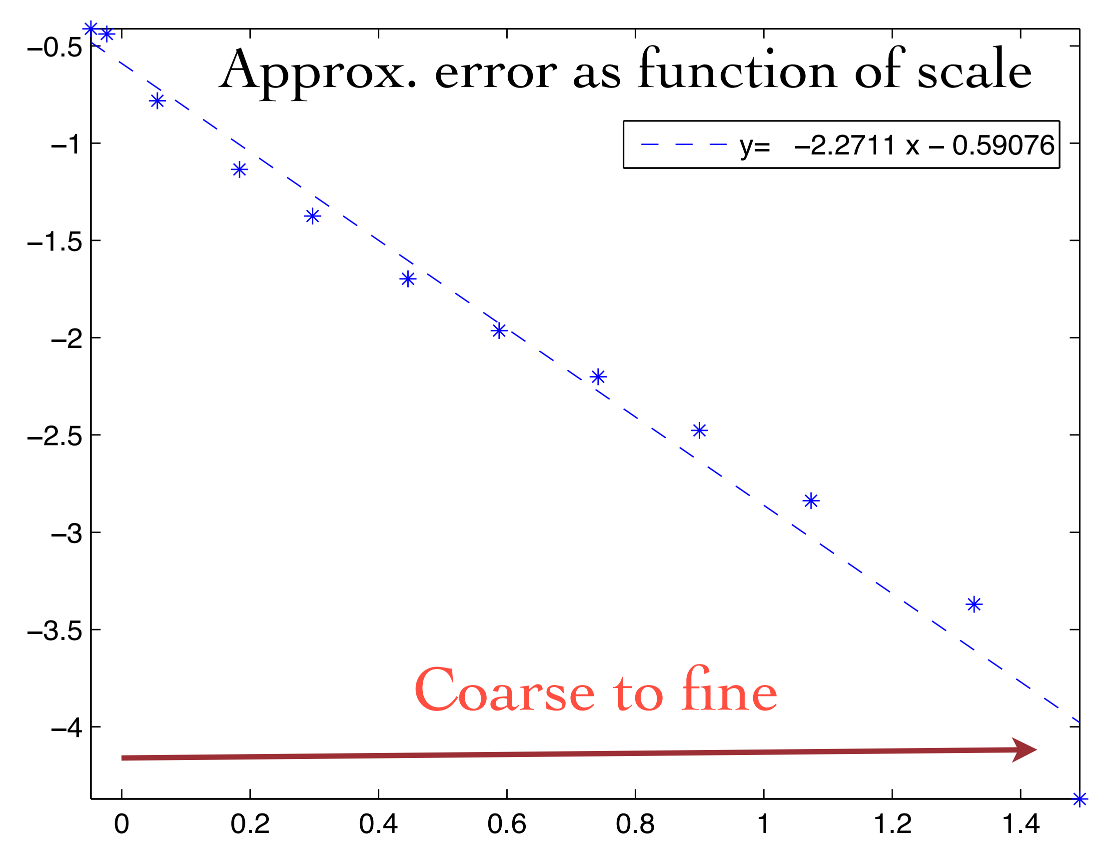
Coarse to fine



Coarse to fine



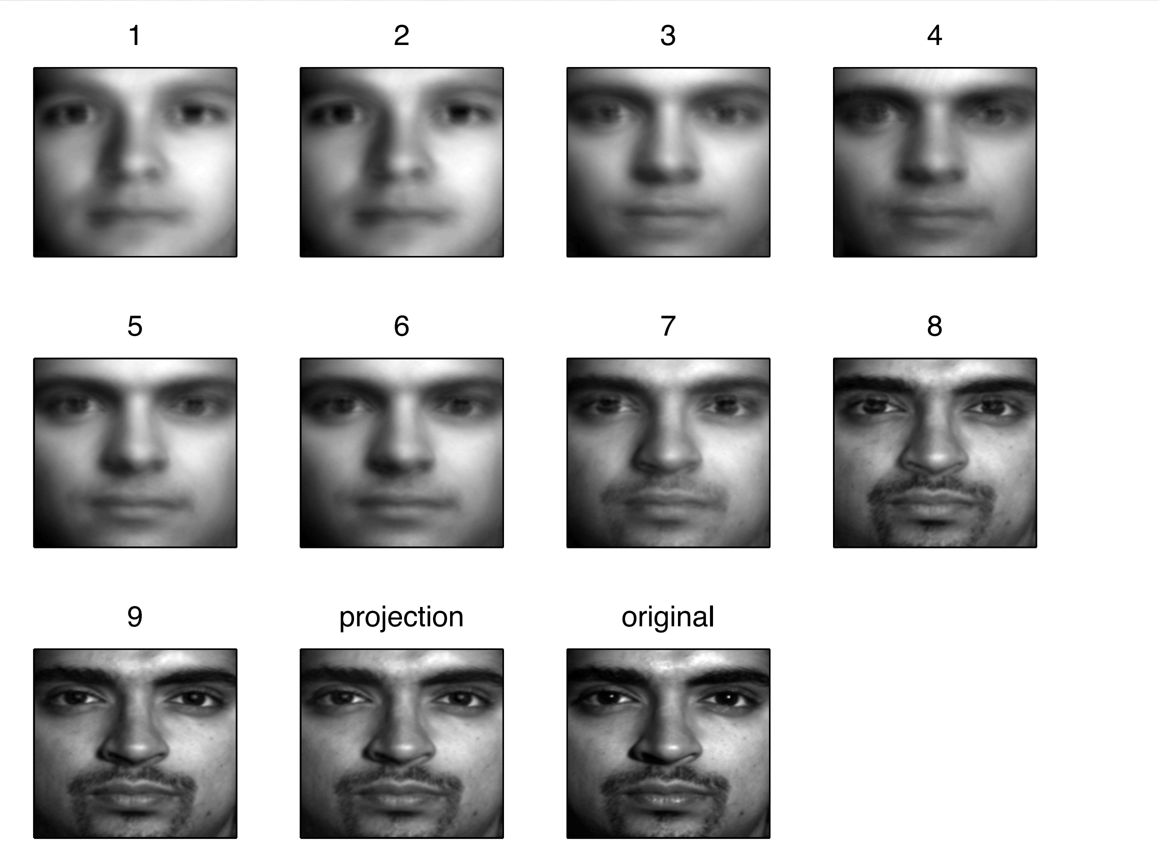
Sparse data representation, fast transforms

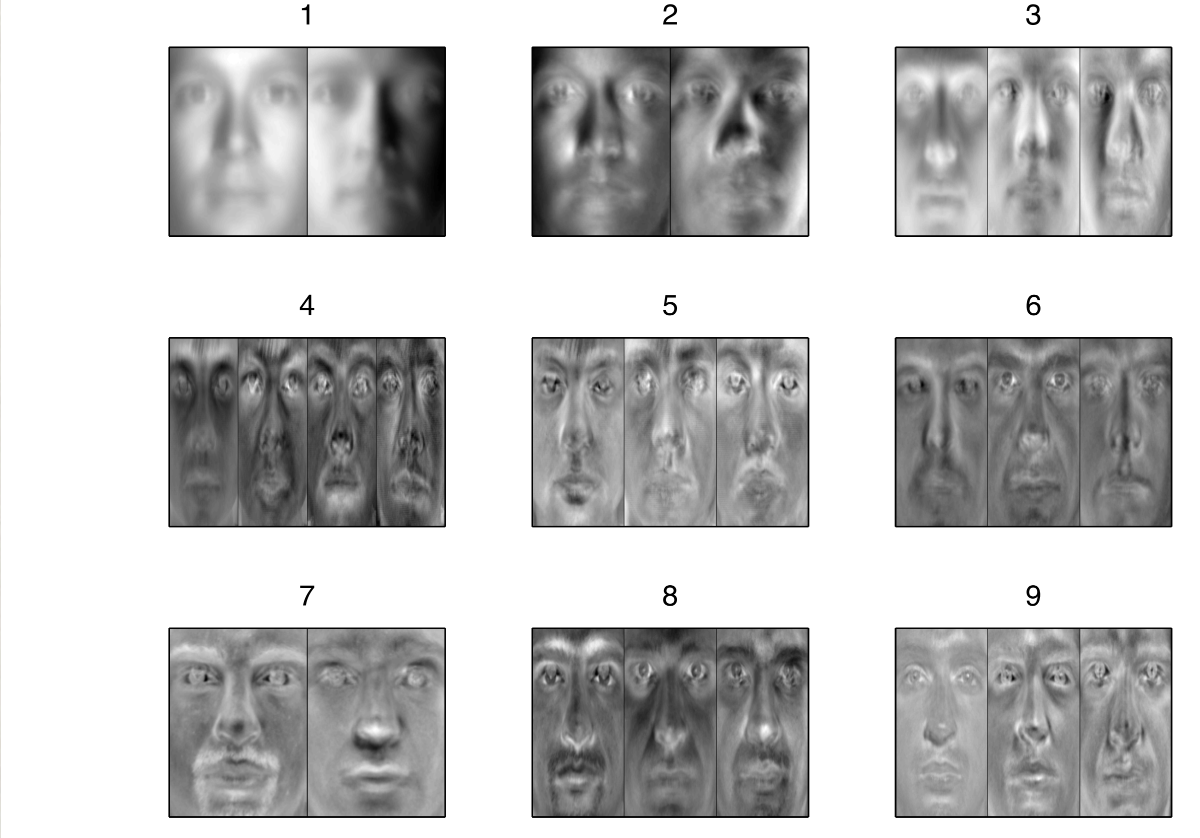


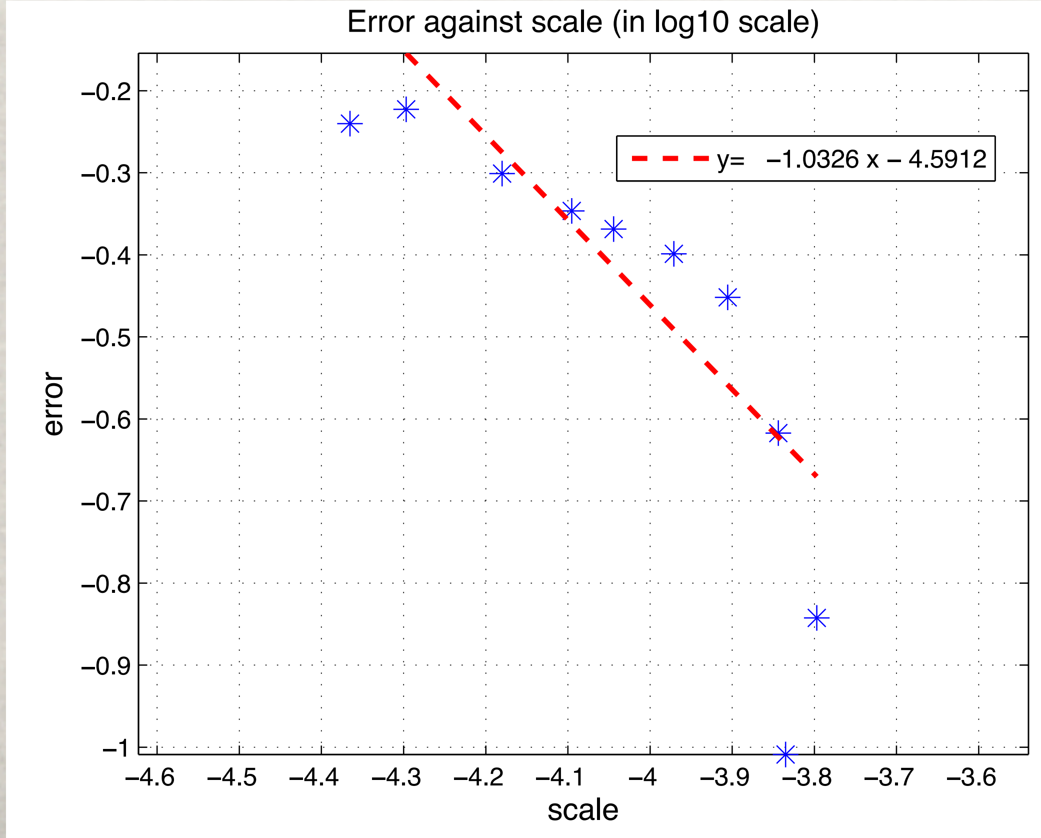
Coarse to fine

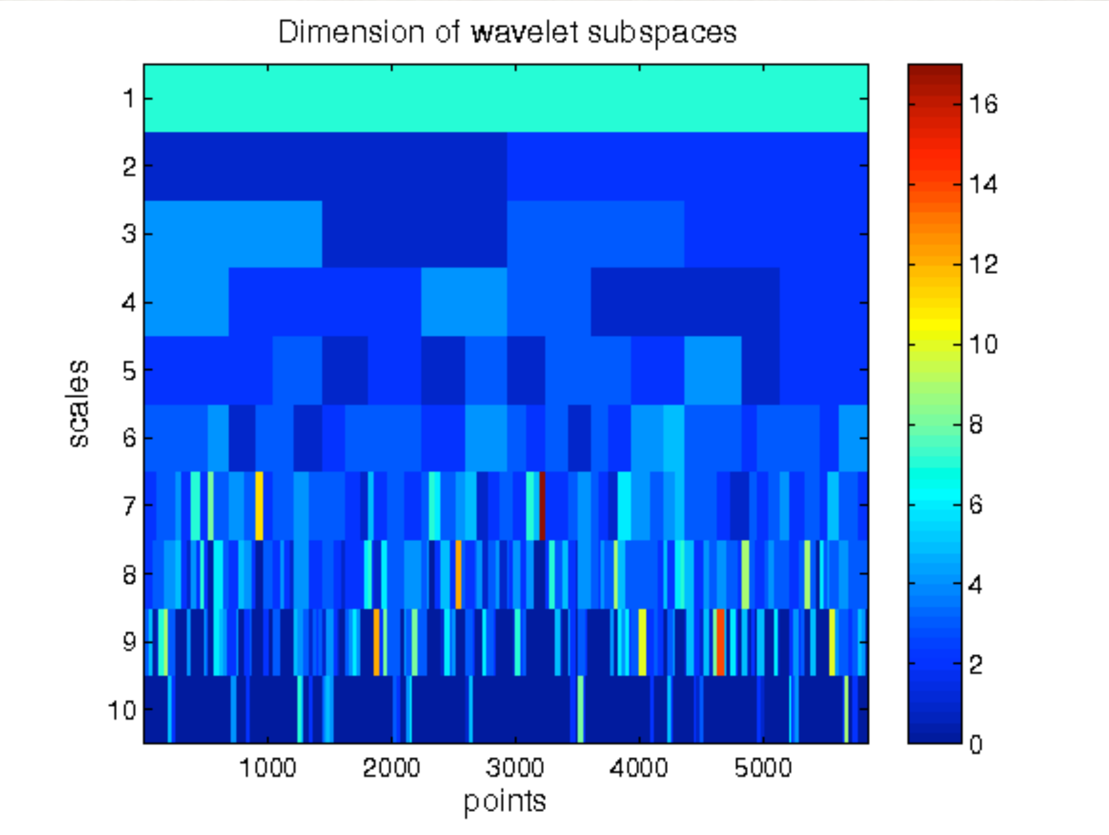
Yale Faces database

Multiscale approximation with GWT for one data point (face, 640x480)









UI for Geometric Multi-Resolution

With E. Monson, R. Brady, G. Chen, *Data Representation and Exploration with Geometric Wavelets*, VAST, 2010

Wavelet axis images control current plot coordinates

Center image of current node

Current node

Tree navigation

Icicle view of binary tree with wavelet coefficent matrix overlay

Detail view of wavelet axes for current image with coeff magnitude mapped to opacity

Scatterplot of wavelet coeffs for current node

Mouse hover data overview

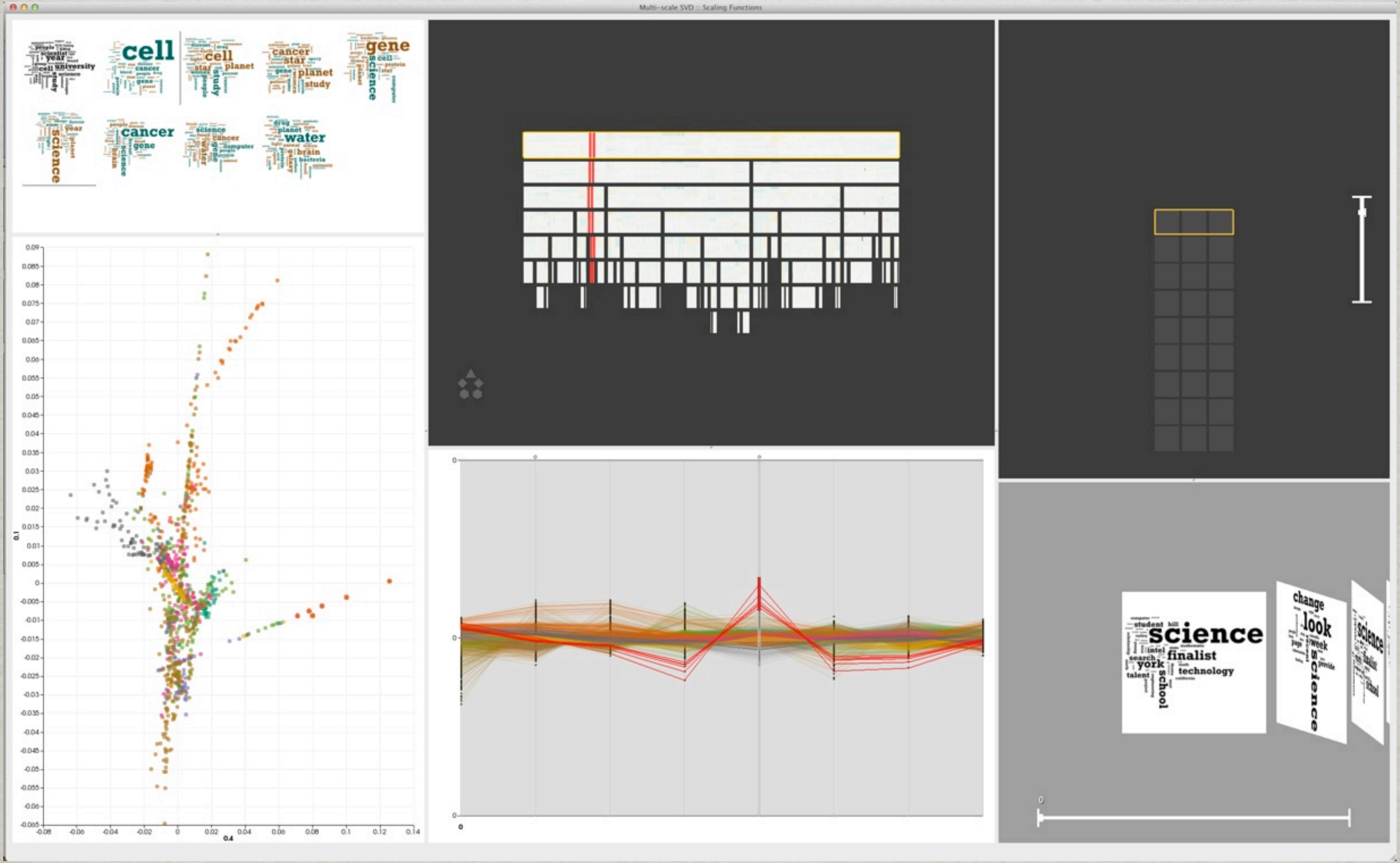
Selections linked between views

Parallel coordinates plot of wavelet coeffs at all scales for all data in current node

Scrollable view of all images selected in scatter & parallel coordinates plots (red) with individual selected (blue)

POSTER & DEMO TONIGHT

UI for Geometric Multi-Resolution



Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

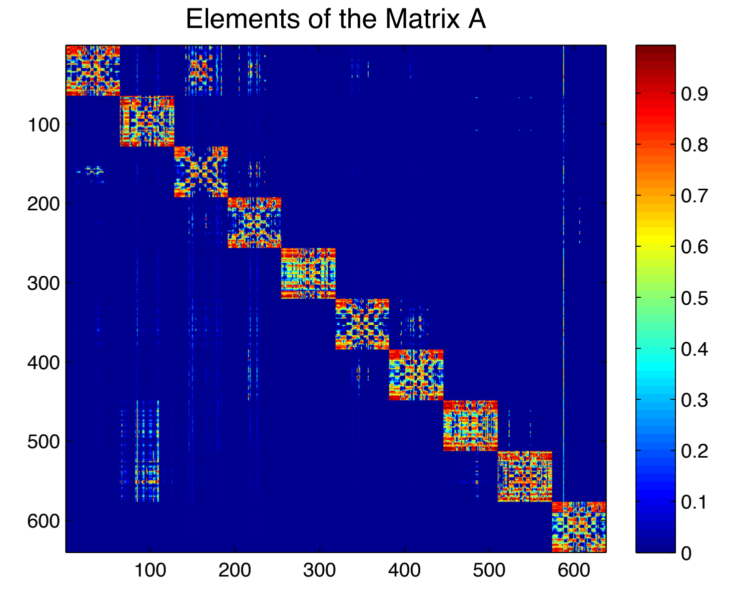
Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

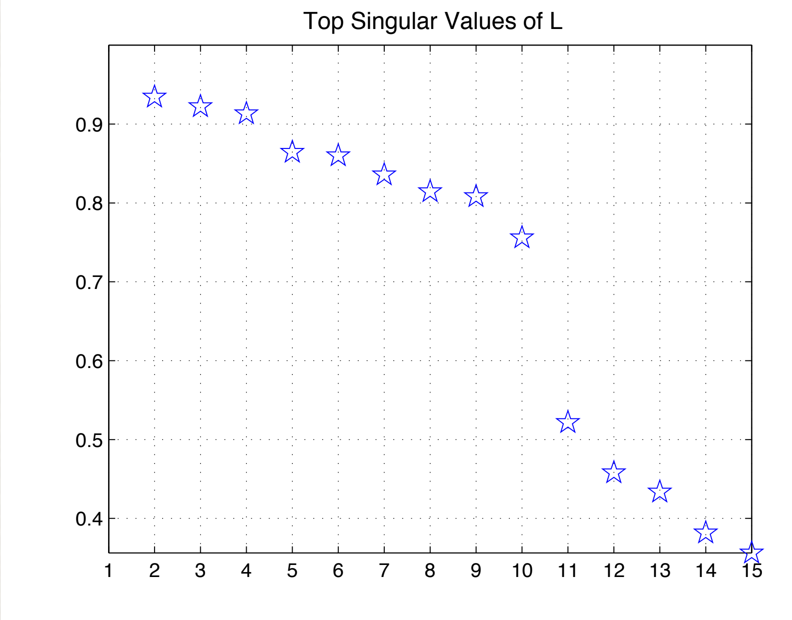
Multiscale Geometric and Spectral Analysis of Plane Arrangements

With G. Chen, *Multiscale Geometric and Spectral Analysis of Plane Arrangements*, CVPR 2011

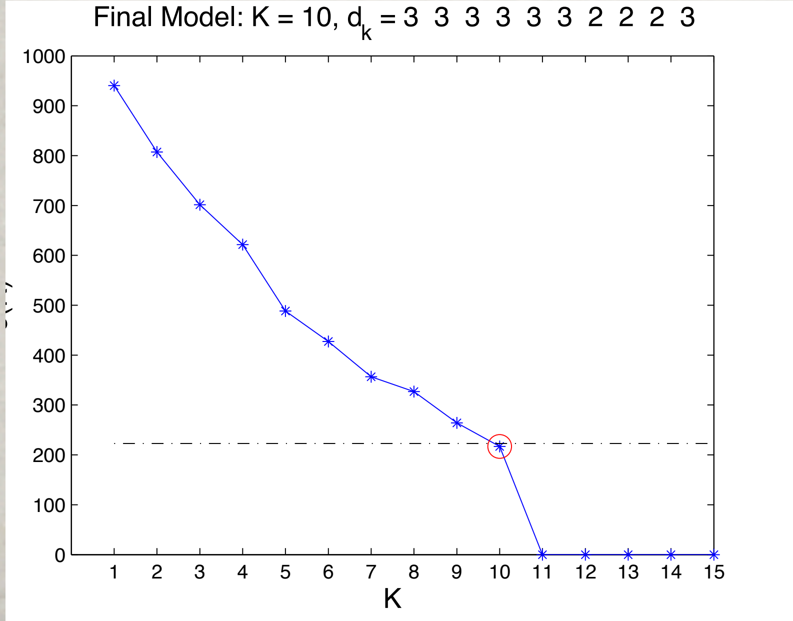
Model data by using K low-dimensional planes. Problem: estimate K and the planes, given noisy points. Classification: assign points to nearest plane.
We introduce a novel fast algorithm with strong guarantees.



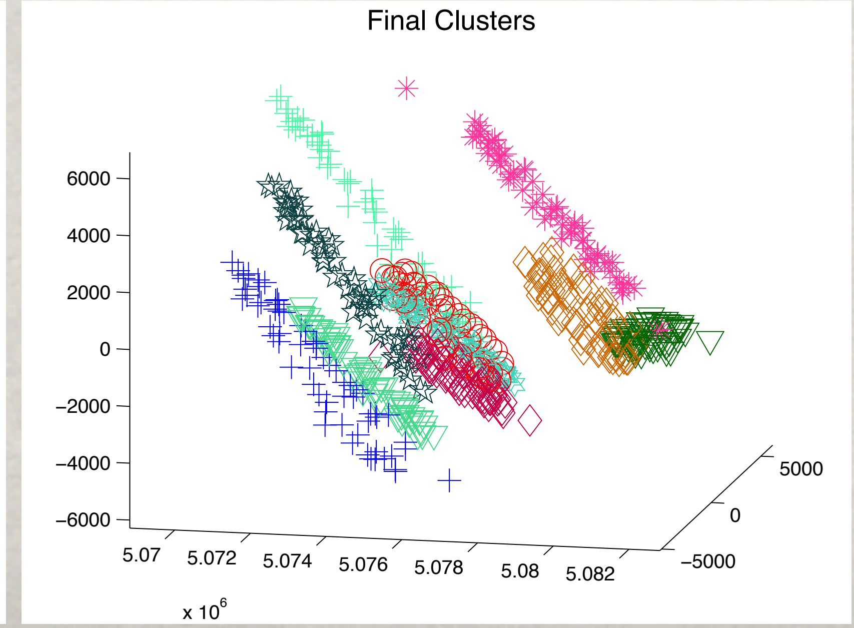
Elements of the Matrix A



Top Singular Values of L



Final Model: K = 10, d_k = 3 3 3 3 3 3 2 2 2 3



Final Clusters

Yale face database:
10 subjects, frontal,
varying illumination

POSTER TONIGHT

Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Research topics

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.
- Construct intrinsically low-dimensional models for data, in particular images and text documents.
- Exploiting the last two for clustering and classification tasks.
- Use this type of multiscale analysis to introduce new metrics between graphs, in particular for analysis of time series of graphs.

Dynamic Graphs

J. Lee, MM. See J. Lee's Thesis, 2010,
Proc. SampTA, 2011

Given: time series of graphs G_t. Objective: to analyze this time series. Desiderata:

. Sensitive to large and small significative changes in the network, and to their location.

. Should capture both topological and quantitative geometric changes.

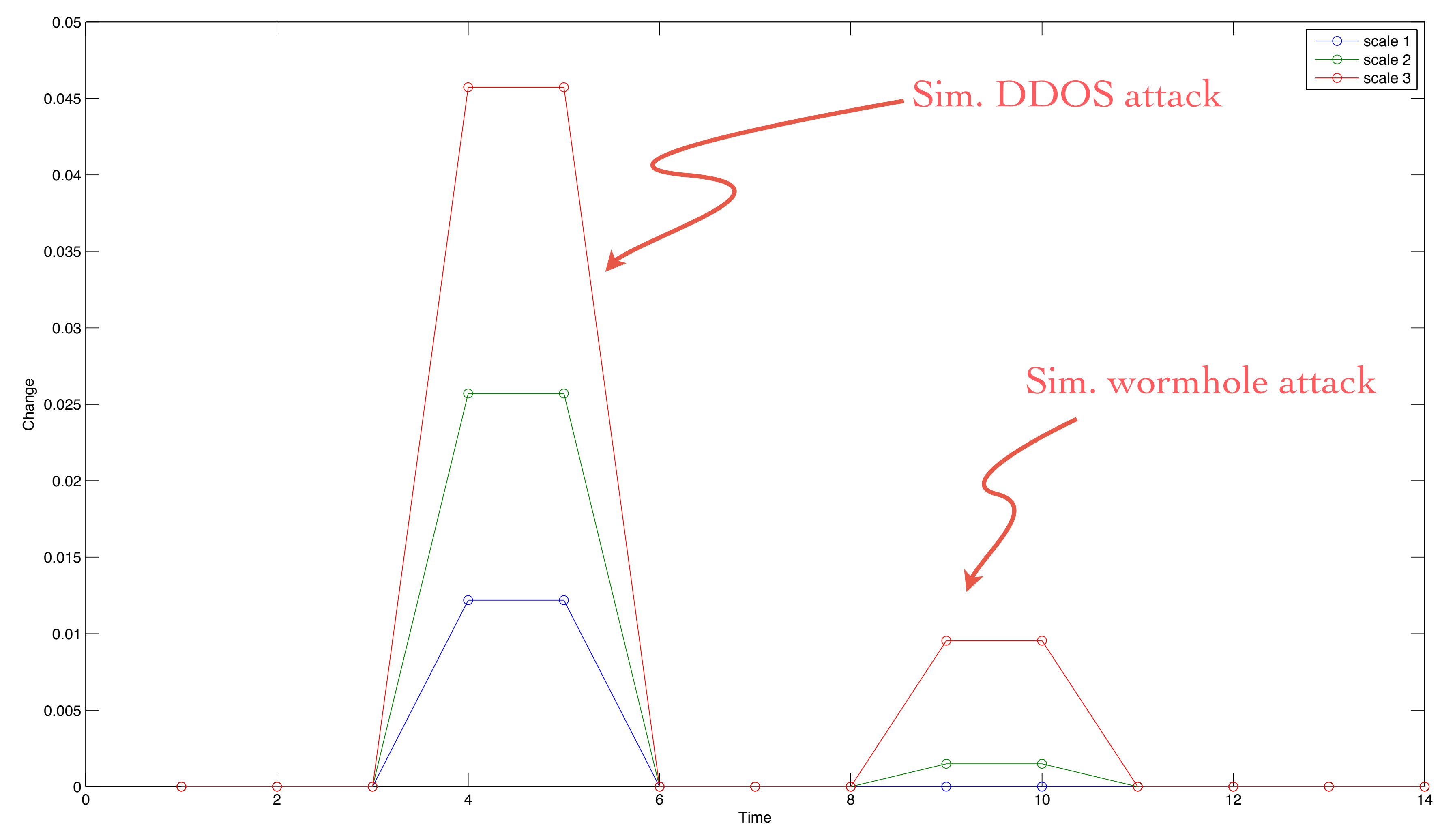
. Should yield measures of change: want to do analysis, statistics...

. Robust to "noisy" perturbations of the network.

We have introduced a framework, based on multiscale analysis on graphs, that enabled us to introduce distance measures on graphs satisfying and quantifying the above.
Basic idea: produce a multiscale decomposition of the graph G_t, match the pieces to those in G_{t-1}, and quantify change in terms of a measure of (multiscale) connectivity among the pieces. Yields a sort of "wavelet-like" analysis for time series of graphs, quantifying changes at different scales and locations.

Simulated “attacks” on blog network

Network of political blogs, 1400 nodes and 19000 edges. We simulate two attacks: a DDOS attack at time 4, when one random vertex is connected to 100 random vertices, till time 6, and then a wormhole attack at time 8, when the two farthest vertices are connected by a heavy edge.



Sim. DDOS attack

Sim. wormhole attack

L. A. Adamic and N. Glance, "The political blogosphere and the 2004 US Election", in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005)

Open problems & future dir.'s

- Bi-clustering and two-way matrix analysis with geometric methods, relationships with Bayesian methods; density estimation and anomaly detection.
- Interactivity and human-in-the-loop in the above.
- First two toolboxes (GMRA and MAPA) just released.
- Dynamic graphs and networks: scaling up, more refined approaches (and toolbox coming soon).
- Integration of our clustering and data reduction methods with J. Stasko's Jigsaw

Collaborators: Eric Monson, Rachael Brady (Duke C.S.); Guangliang Chen (Duke Math); Anna V. Little, Prakash Balachandrian (Math grad, Duke), Jason Lee (Math undergrad, Duke).

www.math.duke.edu/~mauro