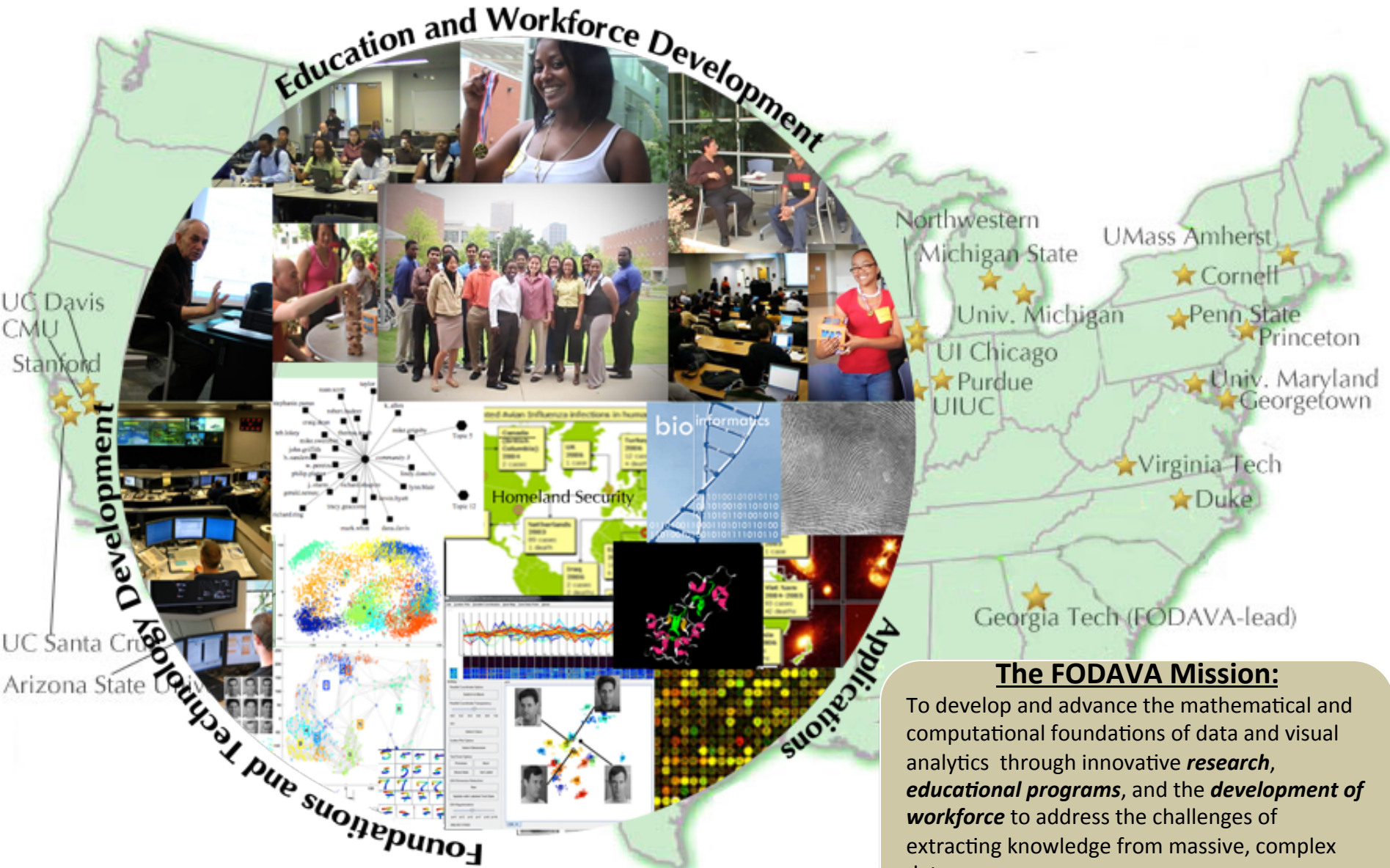


FODAVA Education and Community Building Activities

Haesun Park

School of Computational Science and Engineering
Georgia Institute of Technology

FODAVA Annual Meeting, Dec. 12-13, 2012



The FODAVA Mission:

To develop and advance the mathematical and computational foundations of data and visual analytics through innovative *research*, *educational programs*, and the *development of workforce* to address the challenges of extracting knowledge from massive, complex data.

FODAVA–Lead Mission

Research: Serve as a central facility to involve all FODAVA awardees in a common effort to develop the mathematical and computational foundations for data and visual analytics

Education: Facilitate the development of a body of knowledge, curricula, and education programs to establish and build DAVA workforce

Community Building:

- Reach out for broader participation from various communities
- Serve as a liaison between FODAVA researchers and NVAC, DHS Centers of Excellence

FODAVA Education and Curriculum Development

- **Data and Visual Analytics Education Workshops**
 - December 2008, Georgia Tech, Atlanta, GA
 - August 2010, University of Maryland, College Park, MD
 - Resources, syllabi, and discussion available (<http://www.vacommunity.org>)
 - Dissemination of homework assignments, exams, lecture notes, demonstrations, etc. via an online blog on data and visual analytics (<http://smlv.cc.gatech.edu/>)
- **New courses on Data and Visual Analytics**
 - CX 4010 Computational Problem Solving for Scientists and Engineers
 - CX 4240 Introduction to Computing for Data Analysis (proposed)
 - CX 4242/CSE 6242 Data and Visual Analytics
 - CSE 6740 Computational Data Analytics
 - CSE 8803 Advanced Machine Learning
- **Graduate specialization in visual analytics**
 - Specialization in MS, PhD in Computer Science
- **Undergraduate Minor in Computational Data Analysis (proposal pending)**
 - Targets science and engineering students

FODAVA Outreach Program

- GT CRUISE Program (Computing Research Undergraduate Intern Summer Experience)
 - Encourage students to consider graduate studies
 - Diverse student participation
 - Multicultural, emphasizing minorities, women
 - U.S. and international students
 - Ten week summer research projects
 - Interdisciplinary individual and group projects and CRUISE-wide events
 - Weekly seminars (technical, grad studies)
 - Symposium: conference-style presentations
 - 24 minority and women interns over 4 years
- Participation in VAST Challenge Competitions (2009-2012)
 - Six awards received



DAVA Community Development

- Birds-of-Feather Session, VAST Conference, Columbus Ohio, October 2008 (K. Cook, K. Ma, and H. Park)
- Forum on Geometric Aspects of Machine Learning and Visual Analytics: Recent Developments and Future Challenges, IEEE VisWeek, Atlantic City, October 11-12, 2009 (M. Maggioni, V. Koltchinskii, A. Varshney, H. Park)
- Large-Scale Machine Learning: Parallelism and Massive Datasets, NIPS Conference, Vancouver, B.C., Canada, December 11, 2009 (C. Guestrin, A. Gray, A. Smola, A. Gretton, J. Gonzalez)
- Statistical Machine Learning for Visual Analytics, NIPS Conference, Vancouver, B.C., Canada, December 11, 2009 (G. Lebanon and F. Sha)
- Workshop on Extreme Scale Visual Analytics, IEEE VisWeek, Salt Lake City, October 24, 2010 (D. Ebert, G. Lebanon, P. McCormick, H. Pfister, L. Wilkinson, and H. Park)
- Tutorial on Machine Learning for Information Visualization, IEEE VisWeek, Salt Lake City, October 24, 2010 (G. Lebanon, F. Sha)
- Workshop on Visual Analytics and Information Fusion, KDD, San Diego, August 21, 2011 (D. Gotz, S. Mahadevan, J. Sun, J. Ye, and H. Park)
- SAMSI/FODAVA Workshop on Interactive Visualization and Analysis of Massive Data, SAMSI, Research Triangle Park, December 10-12, 2012 (K. Ma, M. Maggioni, J. Sun, L. Wilkinson, and H. Park)

Forum on Geometric Aspects of Machine Learning and Visual Analytics: Recent Developments and Future Challenges

- IEEE VisWeek, Atlantic City, October 11-12, 2009
- Organizers: M. Maggioni, V. Koltchinskii, A. Varshney, H. Park
- Talks
 - “The Opportunities and Challenges within Visual Analytics?”, Thomas, Jim
 - “Partial Clustering and Visualization”, Carlson, Gunnar
 - “Foundations of Multi-Manifold Modeling Algorithms”, Lerman, Gilad
 - “Sparse Subspace Clustering”, Vidal, Rene
 - “Nonparametric Bandits with Covariates”, Rigollet, Philippe
 - “Algorithms and a Visual Interface to Support Query Learning Under Uncertainty”, Scott, Clayton
 - “Visualizing Conditional Dependencies in Hyper-Graph Models”, Mukherjee, Sayan
 - “Towards Understanding Mixtures of Gaussians: Spectral Methods and Polynomial Time Learning with No Separation”, Belkin, Misha
 - “Learning from Data using Matchings and Graphs”, Jebara, Tony
 - “Affine-invariant Principal Components”, Vempala, Santosh
 - “Parameterizing High-dimensional Data Sets with Kernel Map Manifolds”, Whitaker, Ross
 - “Visual Analytics Techniques for Clustering of High-dimensional Data”, Keim, Daniel
 - “Uncertainty-Aware Visualization of Networks”, Ma, Kwan-Liu

Workshop on Extreme Scale Visual Analytics

- IEEE VisWeek, Salt Lake City, October 24, 2010
- Organizers: D. Ebert, G. Lebanon, P. McCormick, H. Park, H. Pfister, L. Wilkinson
- Panel Discussion: “Challenges and Future Directions”
- Talks
 - “Science at Scale: Batch No More”, Lucy Nowell
 - “Large Scale Remote Visualization and Visual Analytics”, Kelly Gaither
 - Invited Talk by Steven Parker
 - “Divide and Recombine for the Analysis of Large, Complex Datasets”, William Cleveland
 - “Validating Visual Features”, Heike Hofmann
 - “Pass/Stream/Merge (PSM) Analytics: A Flexible Architecture for Distributed Analytics and Visualization”, Leland Wilkinson
 - “Visualization: The Dependence Maximization Point of View”, Alex Smola
 - “Beyond RAM: Fast Visual/Statistical Analysis Using Disk-Based Data Structures”, Alex Gray
 - “Novel Multiscale Representations of Data Sets for Interactive Learning”, Mauro Maggioni
 - “Large-Scale Visual Analytics in Biology”, Hanspeter Pfister
 - “Scalable Visual Analytics”, Daniel Keim

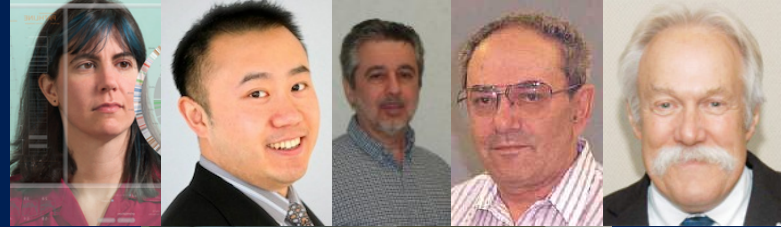
Workshop on Visual Analytics and Information Fusion

- KDD, San Diego, August 21, 2011
- Organizers: D. Gotz, S. Mahadevan, J. Sun, J. Ye, and H. Park
- Panel Discussion: “Challenges and Future Directions”
- Talks
 - “Robust Nonnegative Matrix Factorization using L1, L21 Norms”, Chris Ding
 - “Quality Metrics for Visual Analytics of High-Dimensional Data”, Daniel Keim
 - “Relation Extraction with Relation Topics”, Chang Wang
 - “Visual Analytics and Knowledge Fusion in Multidisciplinary Scientific Discovery”, Pak Chung Wong
 - “Predictive Interactive Visual Analytics”, David Ebert

SAMSI–FODAVA Workshop on Interactive Visualization and Analysis of Massive Data, Dec. 10–12, SAMSI, NC

- SAMSI (Statistical and Applied Mathematical Sciences Institute)
 - Mathematical Sciences Institute program, Division of Mathematical Sciences, National Science Foundation
 - SAMSI is a partnership of Duke University, North Carolina State University, the University of North Carolina at Chapel Hill, and the National Institute of Statistical Sciences, in collaboration with William Kenan Institute for Engineering, Technology, and Science
 - 2012-2013 Program on Statistical and Computational Methodology for Massive Datasets
- Organizers: K. Ma, M. Maggioni, J. Sun, L. Wilkinson, and H. Park
- Talks
 - “The Role of Visualization and Analytics in Solving Problems Based on Massive Data”, Daniel Keim
 - “Visual Analytics for Evidence-Based Medicine”, David Gotz
 - “On the Complexity of Statistical Algorithms”, Santosh Vempala
 - “Interactive Graphics for Data Exploration”, Heike Hofmann
 - “The Role of Perception in Visualization and Visual Analytics”, Christopher Healey
 - “Robust Subspace Modeling”, Gilad Lerman
 - “Large-Scale Visual Data Analysis”, Chris Johnson
 - “Computational Signal Processing in Smart Patient monitoring: Algorithms, Applications and Future Challenges”, Sabine Van Huffel
 - “Discovery of Mechanisms and Prognosis of Cancers from Matrix and Tensor Modeling of Large-Scale Molecular Biological Data”, Orly Alter
 - “The Combinatorial Laplacian and Dimension Reduction”, Sayan Mukherjee
 - “TB-Vis: Visualizing TB Patient-Pathogen Relationships”, Kristin Bennett
 - “VisIRR: Visual Information Retrieval and Recommendation System for Document Discovery”, Alexander Gray
 - “New Approaches for Nonlinear Dimensionality Reduction”, Fei Sha
 - “New Approaches to Storytelling from Massive Textual Datasets”, Naren Ramakrishnan
 - “Scalable Bayesian Learning for Matrix and Tensors”, Alan Qi
 - “BigData: Probabilistic Methods for Efficient Search and Statistical Learning in Extremely High-Dimensional Data”, Ping Li

FODAVA Lecture Series





- Lecture series featuring leaders in the DAVA community
- Web-cast
- Recorded lectures available on FODAVA web
- Tamara Munzner, *Dimensionality Reduction From Several Angles*. December 7, 2012.
- Jimeng Sun, *Patient Similarity Learning through Distance Metric Learning and Interactive Visualization*. November 30, 2012.
- Pat Hanrahan, *The Semiology of Graphics and Interactive Data Analysis: The Impact of Jacques Bertin on Information Visualization and Visual Analytics*. February 25, 2011.
- William Ribarsky, *Developing a Visual Analytics Approach to Analytic Problem Solving*. February 26, 2010
- Leeland Wilkensen, *The Mathematical Foundation of Analytical Visualization*. April 2, 2010
- Jim Thomas, *Three I's of Visual Analytics for FODAVA Teams: Interdisciplinary, International, Immediacy*. April 16, 2010
- Alan Turner, *Mathematical Foundations as a Key Enabler of Agile Human Performance in Visual Analytics Environments*, April 24, 2009
- William Cleveland, *The Disappearing Second Derivative of Quadratics: Perceptual, Mathematical, and Statistical Properties of Judging Dependence on Visual Displays*. March 27, 2009
- Joseph Kielman, *Visual Analytics - Past, Present, and Future*. February 27, 2009
- Alexey Chervonenkis, *Model Complexity Optimization*. January 16, 2009
- Vladimir Vapnik, *Learning with Teacher: Learning Using Hidden Information*. January 16, 2009

FODAVA Website

<http://fodava.gatech.edu>

- FODAVA Latest News and Events
- Quick Links and Information (right sidebar)
- Functionalities: Webforms to upload and communicate information back to the website
- Dissemination of FODAVA results to communities:
 - FODAVA Tech Report Series
 - Repository of over 30 Data Sets
 - FODAVA meetings/lecture/DLS materials
- Blog on DAVA Taxonomy and course material



Home About Us NSF BIGDATA Solicitation Contact Us

People of FODAVA
FODAVA-Lead
FODAVA-Partners '10
FODAVA-Partners '09
FODAVA-Partners '08

Research
Technical Reports
Projects
Data Sets

Lectures
Distinguished Lecture Series

Events
SAMS-FODAVA Workshop
FODAVA Annual Review Meeting 2012
All Events
Related Meetings

Blog
Blog on Data and Visual Analytics
Data and Visual Analytics Taxonomy

Announcements
FODAVA: Seeking a Research Scientist
PhD Fellowships Available

Education & Outreach
Short Course
Summer Intern Program

Other DAVA News
Related news


Latest News and Events

SAMS-FODAVA Workshop
The SAMS-FODAVA Workshop on Interactive Visualization and Analysis of Massive Data will be held on December 10-11, 2012.
Posted: October 02, 2012

FODAVA Annual Review Meeting 2012
The FODAVA Annual Meeting will immediately follow (Dec 12-13) the SAMS-FODAVA joint workshop at the Georgia Institute of Technology.
Posted: September 05, 2012

FODAVA Testbed Software
Many of the modern data sets such as text and image data can be represented in high-dimensional vector spaces and have associated metadata.
Posted: June 30, 2012

About FODAVA
Enormous amounts of data are being generated every day in health care, computational biology, homeland security, commerce, and many other areas. Analyzing these massive and complex data sets is essential to achieve new discoveries, but extremely difficult. An emerging research field known as data and visual analytics is concerned with synthesizing information and deriving insight from massive, dynamic, ambiguous and possibly conflicting digital data for increased understanding and effective decision making.



About fodava.gatech.edu
Our goal is to keep you informed on the progress of the FODAVA initiative while being maintained as a base for further education and outreach to the data and visual analytics community.

Read more [about FODAVA](#) and view a presentation on FODAVA's [Research, Education and Community Building!](#)

Recently updated

[About FODAVA](#)
Dec 06, 2012 - 9:12 pm

[FODAVA Annual Review Meeting 2012](#)
Dec 06, 2012 - 8:38 pm

[SAMS-FODAVA Workshop](#)
Dec 05, 2012 - 12:13 am

Data Sets

[\(ACLED\) Armed Conflict Location and Event Database](#)
World event data with time, location, casualty count, etc. Relatively low volume and only for specific countries, although there are over 70 sets of data. ACLED data are presented in two forms - the first is a simple excel sheet called "Country_X" which will give all information on the politically violent events in which actors from this country are involved in (even if abroad). The Shapefile for each country is based on the Full excel file.

[Airplane Crashes Data Set](#)
Over 5,000 Geo-temporal points with time, location, and other metadata information.

[ATnT Face Database](#)
This Database of Faces, (formerly "The ORL Database of Faces"), contains a set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department.

[Cloud Images Data Set \(MATLAB - 1.1 MB\) and documentation](#)
Contains 41 cloud images generated for weather forecasting. These cloud images are used to test clustering algorithms that segment an image into clusters of clouds. The shapes of the cloud clusters which tend to be perceived by human vision are highly non-elliptical. This poses difficulty to many widely used clustering algorithms such as k-means or mixture-model-based clustering which implicitly assume Gaussian-type clusters.
for M. Qiao and J. Li, "Two-way Gaussian mixture models for high dimensional classification"

[CNN Transcript Collection \(2000-2012\) \(April 25, 2012\)](#)
This is a collection of CNN's (Cable News Network) publicly provided transcripts of shows, events and newscasts from its broadcasts. The archive has been maintained and the text transcripts have been dependably available at transcripts.cnn.com. This is a just-in-case grab of the years of transcripts for later study and historical research. The compressed download file is approximately 1GB large.

[Enron Email Data Set](#)
This dataset was collected and prepared by the CALO Project (A Cognitive Assistant that Learns and Organizes). It contains data from

FODAVA Website Technical Reports

- Collection of FODAVA technical reports from all teams.



Home About Us NSF BIGDATA Solicitation Contact Us

People of FODAVA

- FODAVA-Lead
- FODAVA-Partners '10
- FODAVA-Partners '09
- FODAVA-Partners '08

Research

- Technical Reports
- Projects
- Data Sets

Lectures

- Distinguished Lecture Series

Events

- SAMSI-FODAVA Workshop
- FODAVA Annual Review Meeting 2012
- All Events
- Related Meetings

Blog

- Blog on Data and Visual Analytics
- Data and Visual Analytics Taxonomy

Announcements

- FODAVA: Seeking a Research Scientist
- PhD Fellowships Available

Education & Outreach

- Short Course
- Summer Intern Program

Other DAVA News

- Related news

Latest News and Events

<h4>SAMSI-FODAVA Workshop</h4> <p>The SAMSI-FODAVA Workshop on Interactive Visualization and Analysis of Massive Data will be held on December 10-</p> <p>Posted: October 02, 2012</p>	<h4>FODAVA Annual Review Meeting 2012</h4> <p>The FODAVA Annual Meeting will immediately follow (Dec 12-13) the SAMSI/FODAVA joint workshop at the</p> <p>Posted: September 05, 2012</p>	<h4>FODAVA Testbed Software</h4> <p>Many of the modern data sets such as text and image data can be represented in high-dimensional vector spaces and have</p> <p>Posted: June 30, 2012</p>
---	---	--

Technical Reports

View by year: [\[2012\]](#) [\[2011\]](#) [\[2010\]](#) [\[2009\]](#) [\[2008\]](#) [\[2007\]](#)

- [2012] [A Control Loop Structure Based on Semi-Nonnegative Matrix Factorization for Input-Coupled Systems](#)
Authors: Ryder C. Winck, Jingu Kim, Wayne J. Book, and Haesun Park
- [2012] [A First-order Block-Decomposition Method for Solving Two-easy-block Structured Semidefinite Programs](#)
Authors: Renato D.C Monteiro, Camilo Ortiz, Benar F. Svaiter
- [2012] [A Framework for Uncertainty-Aware Visual Analytics](#)
Authors: Carlos Correa, Yu-Hsuan Chan, Kwan-Liu Ma
- [2012] [A Polynomial Predictor-corrector Trust-region Algorithm for Linear Programming](#)
Authors: Guanghui Lan, Renato D.C. Monteiro, Takashi Tsuchiya
- [2012] [A Visual Analytics Approach for Protein Disorder Prediction](#)
Authors: Jaegul Choo, Fuxin Li, Keehyoung Joo, Haesun Park
- [2012] [Accelerating Bayesian Network Parameter Learning Using Hadoop and MapReduce](#)
Authors: Aniruddha Basak, Irina Brinster, Xianheng Ma, Ole J. Mengshoel
- [2012] [Adaptive Control of Bayesian Network Computation](#)
Authors: Erik Reed, Abe Ishihara, Ole Mengshoel
- [2012] [Age-Layered Expectation Maximization for Parameter Learning in Bayesian Networks](#)
Authors: Avneesh Saluja, Priya Krishnan Sundararajan, Ole J. Mengshoel

FODAVA Website Data Repository

<http://fodava.gatech.edu>

- Collection of over 30 data sets
- Description of data and link to download data

Data Sets

[\(ACLED\) Armed Conflict Location and Event Database](#)

World event data with time, location, casualty count, etc. Relatively low volume and only for specific countries of data. ACLED data are presented in two forms - the first is a simple excel sheet called "Country_X" while the politically violent events in which actors from this country are involved in (even if abroad). The Shapefile the Full excel file.

[Airplane Crashes Data Set](#)

Over 5,000 Geo-temporal points with time, location, and other metadata information.

[ATnT Face Database](#)

This Database of Faces, (formerly "The ORL Database of Faces"), contains a set of face images taken between the lab. The database was used in the context of a face recognition project carried out in collaboration with the Group of the Cambridge University Engineering Department.

[Cloud Images Data Set \(MATLAB - 1.1 MB\) and documentation](#)

Contains 41 cloud images generated for weather forecasting. These cloud images are used to test clustering algorithm image into clusters of clouds. The shapes of the cloud clusters which tend to be perceived by human vision are high poses difficulty to many widely used clustering algorithms such as k-means or mixture-model-based clustering which Gaussian-type clusters.

for M. Qiao and J. Li, "Two-way Gaussian mixture models for high dimensional classification"

[CNN Transcript Collection \(2000-2012\) \(April 25, 2012\)](#)

This is a collection of CNN's (Cable News Network) publicly provided transcripts of shows, events and newscasts from archive has been maintained and the text transcripts have been dependably available at transcripts.cnn.com. This the years of transcripts for later study and historical research. The compressed download file is approximately 1GB

[Enron Email Data Set](#)

This dataset was collected and prepared by the CALO Project (A Cognitive Assistant that Learns and Organizes). It

[Microsoft's GeoLife GPS Trajectories](#)

This GPS trajectory dataset was collected in (Microsoft Research Asia) Geolife project by 182 users in a period of over three years (from April 2007 to August 2012). A GPS trajectory of this dataset is represented by a sequence of time-stamped points, each of which contains the information of latitude, longitude and altitude. This dataset contains 17,621 trajectories with a total distance of about 1.2 million kilometers and a total duration of 48,000+ hours.

[OpenStreetMap](#)

This is a feed of individually submitted GPS traces. There are over 2 billion of these points, although they must be individually downloaded and processed into a larger data set.

[SNAP Brightkite Data Set](#)

A location-based social networking data set where users shared their locations by checking-in. The friendship network was collected using Brightkite's public API, and consists of 58,228 nodes and 214,078 edges. The network is originally directed was constructed into a network with undirected edges when there is a friendship in both ways. There is a total of 4,491,143 checkins of these users over the period of Apr. 2008 - Oct. 2010.

[SNAP Enron Data Set](#)

The graph network derived from about half a million emails. The 36,692 nodes of the network are email addresses and if an address is sent at least one email to address, the graph contains an undirected edge (total of 367,662 edges) from 1 to j. Note that non-Enron email addresses act as sinks and sources in the network as we only observe their communication with the Enron email addresses.

[SNAP Epinions Data Set](#)

Data of the friend network from Epinions.com which contains 75,789 nodes and 508,837 edges.

[SNAP Twitter Data Set](#)

467 million Twitter posts from 20 million users covering a 7 month period from June 1 2009 to December 31 2009. It is estimated to be about 20-30% of all public tweets published on Twitter during the particular time frame. It contains author, time and content information for each data point.

[SNAP Wikipedia Vote Network](#)

A collection of a complete dump of Wikipedia page edit history (from January 3 2008) with the extracted administrator elections and vote history data. It contains 2,794 elections with 103,663 total votes and 7,066 users participating in the elections (either casting a vote or being voted on).

[Synthetic Kronecker and Erdos-Renyi Data Sets](#)

Synthetic Kronecker graph generator tool with diameter 2 and synthetic Erdos-Renyi random graph generator tool.

[TDT2 \(Topic Detection and Tracking Phase 2\)](#)

The TDT2 English Corpus has been designed to include six months of material drawn on a daily basis from six English news sources. The period of time covered is from January 4 to June 30, 1998. The six sources are the New York Times News Service, the Associated Press Worldstream News Service, CNN "Headline News", ABC "World News Tonight", Public Radio International's "The World", and the Voice of America.

[VAST Benchmark Repository](#)

The repository has the VAST datasets and the materials provided by the teams who submitted entries (their answers, videos and explanations) to the Challenge, and the solutions.

[YouTube Data Set](#)

A directed graph of YouTube videos, where each video is a node in the graph. If a video b is in the related video list (first 20 only) of a video a, then there is a directed edge from a to b.

FODAVA PIs Success in Establishing New Related Projects / Funds

- Bill Cleveland
 - NSF BIGDATA
 - NSF DMS CDS&E, (Bowei, Chuanhai Liu)
- Lise Getoor
 - IARPA, “Learning and Predicting Ties in Social Networks” (Lise Getoor, Jennifer Golbeck, Aravind Srinivasan)
 - NSF, “A Theoretical Framework for Practical Entity Resolution in Network Data”
- Leonidas Guibas
 - NSF DMS, “Understanding Data Through Mappings”
- Pat Hanrahan
 - DARPA XDATA, (Pat Hanrahan, Jeff Heer, and Ryan Hafen, Bill Cleveland)
- Leanna House
 - NSF DUE, “Critical Thinking with Data Visualization” (Leanna House, Scotland Leman, Chris North)
- Mauro Maggioni
 - NSF BIGDATA (Pending), “Data Sketches for Dynamic Data Sets in High Dimensional Spaces”
 - AFOSR (Pending), “Information, Approximation, and Fast Algorithms for Data in High Dimensions”
- Chris North
 - NSF HCC, “Semantic Interaction for Visual Text Analytics”
 - DARPA, “STTR Information Salience”
- Haesun Park
 - DARPA XDATA, “Fast Algorithms on Imperfect, Heterogeneous, Distributed Data for Interactive Analysis” (Richard Fujimoto, Haesun Park, Hongyuan Zha, Alex Gray, Barry Drake, Richard Boyd)
- John Stasko
 - NSF BIGDATA (Pending), “Visualizing Semantic Themes in Large Document Collections” (Jacob Eisenstein, John Stasko)

New Center for Data Analytics (CDA) at Georgia Tech

- Established in Nov. 2012 as one of the two centers (in addition to the HPC center) in the Institute for Data and HPC (IDH)
<http://idh.gatech.edu>
- Director: H. Park, pre-Steering committee formed Dec. 2012
(Koltchinskii, Gray, Stasko, ...)
- To bring together researchers at Georgia Tech across the colleges, GTRI (Georgia Tech Research Institute), and Emory University
- To develop a research community and collaborative research projects in large-scale data analytics and its application to important domains
- Current collaboration with Emory Center for Comprehensive Informatics
 - Data and Visual Analytics of pathology data for cancer
 - Cancer subtype discovery for glioblastoma and lung adenocarcinoma for better treatment (note: plain words are brain and lung cancers)
 - Simultaneous analysis of multiple data sets with different geometric properties (e.g. mRNASeq, Methylation)
- Current collaboration with GTRI
 - Efficient distributed algorithms for large scale machine learning
 - User interface improvement on FODAVA Testbed
 - Software development and user study for VisIRR and Testbed
 - Framework design of large-scale visual analytics with GPU
 - Integration of FODAVA algorithm library with GTRI Test Matrix Tool vis package

IEEE VisWeek 2010 Activities (FODAVA–Lead)

- Papers

- Z. Liu, J. Stasko, “Mental Models, Visual Reasoning and Interaction in Information Visualization: A Top–down Perspect,” InfoVis, 2010.
- Z. Liu, J. Stasko, “The Role of Theory in Information Visualization” Theories in Information Visualization: What, Why and How, InfoVis Workshop, 2010.
- J. Choo, H. Lee, J. Kihm, H. Park, “iVisClassifier: An Interactive Visual Analytics System for Classification Based on Supervised Dimension Reduction,” IEEE VAST 2010.

- Posters

- R. Basole, M. Hu, P. Patel, J. Stasko, Visualizing Converging Business Ecosystems for Competitive Intelligence, InfoVis 2010.

- Panel

- Challenges in Visualizing Biological Data: N. Gehlenborg, C. Gorg, M. Meyer, C. Nielsen, InfoVis 2010.

- VAST Challenge Competition: Two Awards

- Good Support for Data Ingest: Data Ingestion and Evidence Marshalling in Jigsaw (Z. Liu, C. Gorg, J. Kihm, H. Lee, J. Cho, H. Park, J. Stasko)
- Excellent Process Explanation: GeneTracer: Gene Sequence Analysis of Disease Mutations (H. Lee, J. Choo, C. Gorg, J. Shim, J. Kihm, Z. Liu, H. Park, J. Stasko)

- Doctoral Colloquium: 2 students participated

- Conference Organization

- Infoviz Program Committee: Edward Clarkson, Carsten Gorg, John Stasko
- InfoViz Steering Committee, VizWeek Executive Committ: John Stasko
- IEEE VAST Program Committee: Haesun Park

IEEE VisWeek 2011–2012 Activities (FODAVA–Lead)

- Papers
 - Y. Kang and J. Stasko, Characterizing the Intelligence Analysis Process: Informing Visual Analytics Design through a Longitudinal Field Study, VAST 2011
 - Z. Liu, S. Navathe, and J. Stasko, Network-based Visual Analysis of Tabular Data, [Honorable Mention Best Paper], VAST 2011.
 - Y. Kang and J. Stasko, Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts, VAST 2012
 - H. Pileggi, C. Stolper, J. Boyle, and J. Stasko, SnapShot: Visualization to Propel Ice Hockey Analytics, InfoVis 2012
- VAST Challenge Competition
 - Mini Challenge 3 Award - Good Use of the Analytic Process: Jigsaw to Save Vastopolis, E. Braunstein, C. Görg, Z. Liu, J. Stasko
- Meeting Organization
 - 2011 InfoViz Steering Committee, VizWeek Executive Committ, InfoViz Best Papers Committee Chair: John Stasko
 - 2011 IEEE VAST Program Committee: John Stasko and Haesun Park
 - InfoVis Steering Committee, VisWeek Executive Committee: John Stasko
 - InfoVis Program Committee: John Stasko

Additional Activities by FODAVA Members (Meeting Organization/Committee Work, etc.)

- Co-chair, SIAM International Conference on Data Mining, 2009. (Haesun Park)
- IEEE Pacific Visualization 2010 Symposium, 2010. (Kwan-Liu Ma)
- Joint Statistical Meetings (JSM) on Statistical Modeling and Learning for Information Visualization and Dimension Reduction, 2010. (Jia Li)
- KDD 2010 Workshop on Mining and Learning from Graphs, 2010. (Lise Getoor)
- NSF III PI Meeting, 2010. (Lisa Singh and Lise Getoor)
- VisWeek 2010 Birds-of-a-Feather Session on Scalable Interactive Visualizations for Visual Analytics, 2010. (Ted Selker and Ole J. Mengshoel)
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011. (Haesun Park)
- AAAI-11 Workshop on Scalable Integration of Analytics and Visualization, 2011. (Ole Mengshoel, Ted Selker, Henry Lieberman)
- Co-Chair, ACM SIGKDD Workshop on Data Mining and Knowledge Discovery Applications in Sustainability, 2011. (Naren Ramakrishnan)
- CHI 2011 Workshop on Analytic Provenance: Process + Interaction + Insight, 2011. (Chris North)
- Co-founder and Chair, IEEE Symposium on Large Data Analysis and Visualization, 2011. (Kwan-Liu Ma)
- IEEE Visualization 2011 Workshop on Working with Uncertainty, 2011. (Alex Pang)
- SMF Workshop on Learning Theory, 2001. (Vladimir Koltchinskii)
- AVI Conference Workshop on Supporting Asynchronous Collaboration in Visual Analytics Systems, 2012. (John Stasko)
- Pacific Graphics, LDAV, InfoVis, Vis, VAST, EGPGV, IVAPP, CDVE, 2012. (Kwan-Liu Ma)
- SC Ultrascale Visualization Workshop, 2010, 2011, 2012. (Kwan-Liu Ma)
- VisWeek 2012 Workshop on Interactive Visual Text Analytics, 2012 (John Stasko)
- Conference on Learning Theory, 2010, 2012. (Vladimir Koltchinskii)

Additional Activities by FODAVA Members (Research Presentations)

- Panel Presentation, 'Interactive Visualization of Large Data Sets: Challenges and Some Preliminary Answers'. (Jia Li)
- Invited Talk, 'Mode Based Clustering with Applications to Information Visualization'. (Jia Li and Xiaolong Zhang)
- Poster/Demo Session, VAC Consortium Meeting. (Eric E Monson, Rachael Brady, Guangliang Chen & Mauro Maggioni)
- Invited Talk, VAC Consortium, 2009. (Mark Hasegawa-Johnson)
- Invited Talk, 'Structure Discovery in 3-D Point-Cloud Data', Industrial Light and Magic, 2009. (Leonidas Guibas)
- Invited Talk, 'Sensing Mobile Objects and Applications', NTT Laboratories, 2009. (Leonidas Guibas)
- 'The Information is in the Maps', Google, Mountain View, December 2009, Microsoft Asia Research Center, Beijing, China, May 2010, Qualcomm Research, Santa Clara, July 2010, INRIA, Sophia-Antipolis, France, August 2010. (Leonidas Guibas)
- Invited Talk, 'Structure Discovery in 3-D Geometry', SIAM/ACM Joint Conference Geometric and Physical Modeling, 2009; KAUST Jeddah, 2010. (Leonidas Guibas)
- Invited Talk, 'Image Webs', A9 Amazon Research, 2009. (Leonidas Guibas)
- Invited Talk, 'Challenges of Data Visualization', NIPS, 2010. (Lise Getoor)
- Invited Talk, 'Networks Across Disciplines: Theory and Applications', NIPS, 2010. (Lise Getoor)
- Panel of Visualization and Rich Data Sets, HCIC, 2010. (Jia Li)
- Invited Talks, International Conference on Data Analysis, 2010. (Lise Getoor)
- Poster Session, 'Exploration & Representation of Data with Geometric Wavelets', IEEE VisWeek 2010 Poster Session. (Eric E Monson, Rachael Brady, Guangliang Chen & Mauro Maggioni)

Additional Activities by FODAVA Members (Research Presentations)

- Invited Talk, IEEE VisWeek, 2010. (Mauro Maggioni)
- Workshop, Modern Massive Data Set 2010 Workshop, 2010. (Ping Li)
- Invited Talk, 'Information Dissemination and Cross-Correlation under Mobility', Army Research Labs, 2010. (Guibas)
- Invited Talk, 'Voronoi Diagrams in Geometry Processing and Network Routing', International Symposium on Voronoi Diagrams in Science and Engineering, 2010. (Guibas)
- Invited Talk, 'The Structure of Isometric Maps and Symmetries', EPFL Bernoulli Symposium, 2010. (Guibas)
- Invited Talks, NEH/IPAM Institute on Networks for the Humanities, 2010. (Lise Getoor)
- Invited Talk, 'Visual Analytics for Relational Data', NIPS, 2010. (Lise Getoor)
- Invited Talk, 'Statistical Issues in Analyzing Information from Diverse Sources', DIMACS/CCICADA Workshop, 2010. (Lise Getoor)
- Invited Talks, National Endowment for the Humanities (NEH) Summer Institute on Networks and Network Analysis for the Humanities, IPAM Institute for Pure and Applied Mathematics, 2010. (Lise Getoor)
- Invited Talk, 'Nonnegative Matrix and Tensor Factorizations: Fast Algorithms and Applications', Conference on Numerical Linear Algebra: Perturbation, Performance, and Probability, 2010. (Haesun Park)
- Keynote Lecture, 'Data and Visual Analytics and Beyond Fusion', National Research Council Workshop on New Research Directions for National Geospatial-Intelligence Agency, 2010. (Haesun Park)

Additional Activities by FODAVA Members (Research Presentations)

- Keynote Lecture, 'Advanced Concepts and Techniques for Visualizing Large Data', IEEE Pacific Visualization Symposium Keynote Speech, 2011. (Kwan-Liu Ma)
- Open House Exhibit, 'Audio Easter Egg Hunt', Beckman Institute Open House, 2011. (Mark Hasegawa-Johnson)
- Invited Talk, 'Bayesian Network Computation with System Health Management Applications', Norwegian University of Science and Technology, 2011. (Ole J. Mengshoel)
- Invited Talk, 'Collective Graph Identification', Rutgers Yahoo! Machine Learning Seminar, 2011. (Lise Getoor)
- Invited Talk, 'Complexity Penalization in Low Rank Matrix Recovery', High Dimensional Problems in Statistics Workshop, 2011. (Vladimir Koltchinskii)
- Invited Presentation, 'Data Analysis and Visualization Computing Requirements: A Case Study', Workshop on Large Scale Computing and Storage Requirements for ASCR, 2011. (Kwan-Liu Ma)
- Keynote Lecture, 'Data and Visual Analytics for Massive High-dimensional Data', G20 International Symposium on Convergence Technologies, 2011. (Haesun Park)
- Invited Talk, 'Data Quality Metrics', DIMACS/CCICADA Workshop, 2011. (Lise Getoor)
- Invited Talk, 'Discovering Network Structure beyond Communities', Network Frontier Workshop, 2011. (Takashi Nishikawa)
- Invited Talk, 'Low Rank Matrix Estimation', International Conference on High Dimensional Probability, 2011. (Vladimir Koltchinskii)
- Invited Talk, 'Network Analysis and Visualization', National Chiao Tung University, 2011. (Kwan-Liu Ma)

Additional Activities by FODAVA Members (Research Presentations)

- Invited Talks, 'Network Visualization', 'Visualizing Large, Complex Data: An Overview', 'Advanced Concepts for Large Data Visualization', CEA/EDF/INRIA Summer Schools, June 2011. (Kwan-Liu Ma)
- Keynote Lecture, 'New Approaches to Large Data Visualization', IEEE Pacific Visualization Symposium, 2011. (Kwan-Liu Ma)
- Keynote Lecture, 'Nonnegative Matrix Factorization: Algorithms and Applications', SIAM International Conference on Data Mining (SDM11), 2011. (Haesun Park)
- Keynote Lecture, 'Research Directions in Data Visualization', Computer Graphics 2011 Workshop, 2011. (Kwan-Liu Ma)
- Invited Talk, 'Recent Advances in Modeling and Computation Using Bayesian Networks', Ericsson Research, 2011. (Ole J. Mengshoel)
- Invited Talks, 'Sparse Recovery in Infinite Dictionaries' Lecture Series, ENSAE-CREST, 2011. (Vladimir Koltchinskii)
- Invited Talk, 'Ultrascale Visualization', ITER, 2011. (Kwan-Liu Ma)
- Invited Talk, 'Visual Analytics for Discovering Node Groups in Complex Networks', APS March Meeting, 2011. (Takashi Nishikawa)
- Invited Talk, 'Visual Analytics in Cyber Security', Sandia CyberSecurity Open House and Workshop, 2011. (Chris North)
- Invited Talk, 'Visualizing Contact Diaries', Academia Sinica, 2011. (Kwan-Liu Ma)
- Invited Seminar, 'Visualizing Large, Complex Data', 5th XLDB, 2011; National Tsing Hua University, 2011. (Kwan-Liu Ma)
- Invited Talk, 'Visualizing Large Data', AMP UC Berkeley, 2011. (Kwan-Liu Ma)

Additional Activities by FODAVA Members (Research Presentations)

- Invited Talk, CMU Machine Learning – Google Distinguished Lecture Series, 2012. (Lise Getoor)
- Invited Talk, Michigan State University, Columbia University, CREST, Paris, Florida State University, 2012. (Vladimir Koltchinskii)
- Invited Talk, ‘Advanced Concepts and Techniques for Extreme-Scale Data Analysis and Visualization’, DOE Exascale Computing Conference, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Advanced Concepts for Large Data Visualization’, Seoul National University, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Big Data Visualization’, UI Irvine, 2012, HP Labs, 2012; Chunghwa Telecom Laboratories, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Big Data Visualization: Advanced Concepts and Techniques’, SIBGRAPI, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Information Visualization for Studying Behaviors, Connections and Evolution’, Taiwan Tech, 2012; Zhejiang University, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Large Data Visualization’, National Chao Tung University, Zhejiang University of Technology, 2012. (Kwan-Liu Ma)
- Invited Talk, ‘Nonnegative Matrix Factorization for Clustering’, Workshop on Algorithms for Modern Massive Data Sets (MMDS2012), 2012. (Haesun Park)
- Invited Talk, ‘Recent Advances in Social Network Visualization’, Academia Sinica, 2012. (Kwan-Liu Ma)

Additional Activities by FODAVA Members (Research Presentations)

- Invited Talk, 'Visual Analytics for Knowledge Discovery in High Dimensional Data', Massive Datasets Program, SAMSI, 2012. (Haesun Park)
- Invited Talk, 'Visualization: An Enabling Tool for Discovery and Communication', St. John's University of Technology, 2012. (Kwan-Liu Ma)
- Invited Talk, 'Visualization Research Challenges and Opportunities', National Taipei University of Technology, 2012. (Kwan-Liu Ma)
- Invited Talk, 'Visualizing Large and Complex Data', IBM Almaden Research, 2012. (Kwan-Liu Ma)

Additional Activities by FODAVA Members

(Educational Activities and Tutorial Presentations)

- Birds-of-a-Feather Session on Scalable Interactive Visualizations for Visual Analytics at VisWeek 2010. (Ted Selker and Ole J. Mengshoel)
- Tutorial Series on 'Exploiting Statistical and Relational Information on the Web and in Social Media', AAAI10, July 2010 and WSM 2011, February 2011, and SDM 2011, April 2011. (Lise Getoor)
- Tutorial on 'Learning Statistical Models from Relational Data', SIGMOD, June 2011. (Lise Getoor and Lily Mihalkova)
- Georgetown Conference on Computer Science Education for K – 12, April 2012. (Lisa Singh)
- New Class: 'Geometry Processing Algorithms', Stanford, 2011-2012. (Leonidas Guibas)
- New Class: 'High-Dimensional Data', Duke, 2012. (Mauro Maggioni)
- New Class: 'Visual Analytics', CMU Silicon Valley, 2011-2012. (Ole J. Mengshoel)
- New Class Topic: "Visual Analytics", Georgetown University, Spring 2012. (Lisa Singh)
- New Class: 'Visualization', UCSC, 2012. (Alex Pang)
- Tutorial on 'Representation, Inference, and Learning in Structured Statistical Models', NIPS 2012. (Lise Getoor)
- Tutorial on 'Sparse and Low Rank Recovery', Mathematical and Computational Foundations of Learning Theory Workshop, Dagstuhl, Germany, July 2011. (Vlad Koltchinskii)
- Tutorial on 'Uncertainty and Parameter Space Analysis in Visualization', IEEE VisWeek 2012. (Alex Pang)
- Tutorials on 'Entity Resolution', VLDB 2012, AAAI 2012, ANOSAM 2012. (Lise Getoor)