Joint Interface/CSNA 2005

Conference Agenda

(Invited Sessions are Fixed, Contributed Sessions are Tentative)

Wednesday, June 8, 2005	
8:00 am to 6:00 pm	Registration
9:00 am to 12:00 pm	Short Course: Introduction to Clustering, Bryant
9:00 am to 12:00 pm	Short Course: Introduction to Microarrays and Clustering, Shannon
1:00 pm to 4:00 pm	Short Course: Random Graphs for Statistical Pattern Recognition, Marchette
1:00 pm to 4:00 pm	Short Course: Latent Class Models for Clustering and Classification, Magidson
6:00 pm to 10:00 pm	Reception
7:00 pm to 11:00 pm	Interface Board of Directors Meeting
Thursday, June 9, 2005	
7:30 am to 8:30 am	Continental Breakfast
8:30 am to 8:45 am	Opening Ceremonies
8:45 am to 10:00 am	Keynote Address: **David Scott, Rice University** **Mixtures at the Interface**
10:00 am to 10:15 am	Break
10:15 am to 12:00 pm	**Invited: COLLABORATION IN DATA MINING (GOODMAN)** 10:15-10:50 03119. The role of collaborations in the Sapphire scientific data mining project, Kamath 10:50-11:25 03046. Data Mining Success = f (Definition, Collaboration, Value), Goodman 11:25-12:00 03154. Strategies for Visual Data Mining, Wegman **Invited: MIXTURE MODELS (SCOTT)** 10:15-10:55 03151. Topography of Multivariate Normal Mixtures, Lindsay, Penn State, Ray 10:55-11:35 03073. Time-dynamic mode tracking and the mean-shift algorithm, Mueller 11:35-11:50 Discussant: Szewczyk

	11:50-12:00 Open Discussion/Questions **Invited: STATISTICS AND THE LAW (BARNES)** 10:15-10:50 03137. Racial Profiling and Selection Models, Barnes 10:50-11:25 03147. To Tell the Truth: On the Probative Value of Polygraph Search Evidence, Fienberg 11:25-12:00 03021. Statistical Models for Improving Biometric Authentication, Mitra **Contributed: APPLICATIONS IN EARTH AND PHYSICAL SCIENCES** 03082. Carbon Dioxide, Global Warming, and Michael Crichton's Rust 03113. Spatially Constrained Clustering in Surveillance GeoInformation of Hotspot Detection and Early Warning, Patil 03057. Mining Massive Earth Science Data Sets for Large Scale Structure, Braverman 03094. Efficient Processing of Massive Data Obtained Using Comprehensive Two-Dimensional Gas Chromatography Coupled with Mass Spectrometry Detection (GCxGC-MS), Beagley 03105. Robust Clustering Of Positron Emission Tomography Data, Velamuru
12:00 pm to 1:00 pm	Lunch – EPNEC Lobby
1:00 pm to 2:15 pm	Keynote: **Jon Kettenring, Drew University** **The Practice of Cluster Analysis**
2:15 pm to 4:00 pm	**Invited: CLUSTERING (BANKS)** 03010. A Fast Clustering Algorithm with Application to Cosmology, Jang (2:15-2:50) 03011. Multiple Imputation for Cluster Analysis, Larsen (2:50-3:25) 03012. Selecting the Number of Components in a Finite Mixture: A Risk-Based Approach, Ray (3:25-4:00) **Invited: MASS SPECTROMETRY BASED PROTEOMICS (NAIK)** 03075. Characterization of environmental stress response in mouse lungs via mass spectrometric profiling, Wagner (2:15-2:45) 03076. Standardization and De-noising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens, Datta (2:45-3:15) 03077. Strategies for Mass Spectrometry Data Analysis in Cancer Proteomics, Naik (3:15-3:45) 3:45-4:00 Open Discussion/Questions **Invited: METHODS FOR DETECTING GENE-GENE INTERACTIONS (CULVERHOUSE)** 03159. The Restricted Partition Method used to screen for genetic interactions, Culverhouse (2:15-2:45) 03169. Dimensionality Reduction for Detecting Epistasis, Ritchie (2:45-3:15)

	03157. Epistasis in a Case-Cotrol Study with 100,000 SNPs, Hoh (3:15-3:45) 3:45-4:00 Open Discussion/Questions **Contributed: INFERENCE AND FOUNDATIONS** 03079. Randomization tests for functional data based on adaptive truncation} Lee 03066. Ternary Separation and Hierarchies, Powers 03065. On the Operating Characteristics of Some Non-parametric Methodologies for the Classification of Distributions by Tail Behavior, Ott 03108. Minimum energy clustering, Rizzo 03095. Higher-Order Density Estimation and Bump Hunting, Minnotte
4:00 pm to 4:15 pm	Break
4:15 pm to 6:00 pm	**Invited: HIGH-DIMENSIONAL BIOMEDICAL DATA: VISUALIZATION, CLASSIFICATION, CLUSTERING (SOMORJAI)** 03153. Class-Preserving Mapping of High-Dimensional Biomedical Data: Visualization, Classification, Clustering, Somorjai (4:15-4:45) 03004. Generalized MDS for Data Visualization, Clustering, and Classification, Solka (4:45-5:15) 03100. Deriving Meaningful Biological Structure from Spectral Embedding and Clustering, Higgs (5:15-5:45) 5:45-6:00 Open Discussion/Questions **Invited: MIXTURE MODELS (VIELE)** 03048. On a Flexible Information Criterion for Order Selection in Finite Mixture Models, Charnigo (4:15-4:45) 03161. Applying Dirichlet Process Mixture Models to Compositional Data, Gantz (4:45-5:15) 03162. Functional Clustering of Temporal Microarray Data, Ma, Harvard University (5:15-5:45) 5:45-6:00 Open Discussion/Questions **Contributed: DATABASES AND COMPUTERS** 03043. Automated Generation of Metadata, Al-Shameri. 03109. A Database Design for Assessing Student Learning in an Online Course System, Tan 03112. Assessment Performance Of Data Interoperability In Federated Distributed Database Systems, Owunwanne , 03123. Developing Statistical COM Servers, Lemmon 03132. Clustering Heterogeneously Distributed Data, Socolovsky **Contributed: APPLICATIONS IN MEDICIN E** 03178. Improving the Sensitivity of Health Care Cost Predictions Using a Combination of Regression and Classification Procedures, Asparouhov 03124. Multilevel Classification of quantitative cytology data using Cumulative Log-

	Odds Method, Yamal 03072. Automatic classification of fMRI brain images using smooth asynchrony maps, Shinkareva 03083. Modeling exposures for DNA methylation profiles}, Kimberly Siegmund, University of Southern California
7:00 pm to 11:00 pm	Banquet, Bowling Hall of Fame
Friday, June 10, 2005	
7:30 am to 8:45 am	Continental Breakfast
8:45 am to 10:00 am	Keynote Address: **Merlise Clyde, Duke University** **Bayesian Perspectives on Combining Models**
10:00 am to 10:15 am	Break
10:15 am to 12:00 pm	**Invited: BEST OF SIAM DATA MINING CONFERENCE 2005 (KAMATH)** 03158. Dynamic Detection, Visualization and Classification of Defect Structures in Molecular Dynamics Simulations, Parthasarathy (10:15-10:40) 03052. Topic-driven Clustering for Document Datasets, Zhao (10:40-11:05) 03032. Gaussian Processes for Active Data Mining of Spatial Aggregates, Bailey-Kellogg (11:05-11:30) 03149. Clustering with Model-Level Constraints, Gondek (11:30-11:55) 11:55-12:00 Open Discussion/Questions **Invited: MODELING ALCOHOL ABUSE AND CONSEQUENCES (SAID)** 03139. Drinking and Public Health: The Role for Simulations and Models, Wieczorek (10:15-10:50) 03141. Modeling Alcohol Abuse and Consequences Said (10:50-11:25) 03140. Assessing subpopulation differences in the DSM IV constructs of alcohol use and abuse across non-Hispanic Caucasians, non-Hispanic African Americans, and Hispanics, Carle (11:25-12:00) **Contributed: MODEL AVERAGING AND ASSESSMENT** 03055. A modest improvement towards seeing the trees in a utopian forest, Izmirlian 03063. Application of Ensemble Learning for Modeling of Quantitative Structure-Activity Relations of Pharmaceutical Molecules, Tong 03111. A robust meta-classification scheme for cancer detection, Alexe 03051. Comparison of Estimators of Generalization Error, Martin
12:00 pm to 1:00 pm	Lunch – EPNEC Lobby
1:00 pm to 2:15 pm	Keynote: **Sid Chib, Washington University** **Bayesian Causal Inference from Observational Data**

2:15 pm to 4:00 pm	**Invited: BIOINFORMATICS (BANKS)** 03013. Statistical Learning Tools for Analyzing Metabolomic Datasets, Lin (2:15-2:50) 03160. Learning Variable Covariances via Gradients, Mukherjee (2:50-3:25) 03015. A Stepwise Structural Equation Modeling Algorithm to Reconstruct Genetic Networks, Shieh (3:25-4:00) **Invited: CLUSTER VALIDATION (STEINLEY)** 03142. Individual Differences Additive Tree Fitting Through (Heuristic) Iterative Projection Koehn (2:15-2:45) 03143. Multiobjective Programming Methods for Applied Cluster Analysis, Brusco (2:45-3:15) 03144. Stability analysis in K-means clustering, Steinley (3:15-3:45) 3:45-4:00 Open Discussion/Questions **Invited: SELECTED IASC PAPERS ON CLUSTERING AND CLASSIFICATION (SCHIMEK)** 03164. Data Reduction Using L_p Criteria, Schuler, Logos Technologies, Gentle (2:15-2:40) 03163. Empirical Bayes thresholding in gene expression analysis, Schimek (2:40-3:05) 03156. Fitting Cox survival model on high dimensional data, van Houwelingen (3:05-3:30) 03138. Differential co--expression. A new concept in the statistical analysis of microarrays, Kostka (3:30-3:55) 3:55-4:00 Open Discussion/Questions **Contributed: MODEL-BASED METHODS** 03053. Variable Selection for Model-Based Clustering, Dean 03172. Estimation and Selection of Normal Mixture Models Based on Spacings, Wang 03074. On Evidence Weighted Mixture Classification, Everson 03071. Model-Based Clustering Toolbox for MATLAB,. Martinez
4:00 pm to 4:15 pm	Break
4:15 pm to 6:00 pm	**Invited: CLUSTERING AND CLASSIFICATION IN TEXT DOCUMENTS (PRIEBE)** 03165. Estimating Probability Mass Functions from Very Sparse Data, Khudanpur (4:15-4:50) 03171. Knowledge Acquisition from Text, Lin (4:50-5:25) 03044. An Investigation of Text Mining Techniques for the Analysis of Abstracts, Marchette (5:25-6:00) **Invited: MODEL-BASED CLUSTERING/CLASSIFICATION IN HIGH-DIMENSIONAL DATA (RAY)**

	03031. Model-Based Clustering of High-Dimensional Data, McLachlan (4:15-4:45) 03107. Quantitating differences in two multivariate distributions, Walther (4:45-5:15) 03027. On Potts Model Clustering and Kernel K-Means, Murua (5:15-5:45) 5:45-6:00 Open Discussion/Questions **Contributed: MICROARRAYS** 03069. Clustering Methods in Microarrays, Rejto 03078. Microarray Analysis: Is an Ordered Gene List Enough?, Hearne 03176. Microarray Gene Selection Using Mantel Correlation with K-means, Shannon 03114. Genetic Algorithms for Feature Selection using Mantel Correlation Scoring, Grefenstette 03026. Gene selection using support vector machines with nonconvex penalty, Park **Contributed: APPLICATIONS** 03122. Using classification of professional to assessment vocational guidance cancellers and their clients, Mitina 03125. Visual Assessment of Simple Association Models, Hofmann 03134. Discriminant Function Analysis in Forensic Authorship Attribution, Chaski 03174. Employing Priors for Classifying High Risk Prison Inmates, Baek 03097. Measuring Relationships Between Entities in Free Text, White **Contributed: MISCELLANEOUS** 03181. SIP Load Test Automation Based on Data Mining, Doci 03180. Discovering Backbone Structure in Graphs, Lin 03173. Graph-theoretic Scagnostics, Wilkinson
7:00 pm to 11:00 pm	CSNA Board of Directors Meeting
Saturday, June 11, 2005	
7:30 am to 8:45 am	Continental Breakfast
8:45 am to 10:00 am	Keynote Address: **Jerome Friedman, Stanford University** **Predictive Learning via Rule Ensembles**
10:00 am to 10:15 am	Break
10:15 am to 12:00 pm	**Invited: COMPARING CLASSIFICATION METHODS (SHINE)** 03049. Classifying the Classifiers, Li, George Mason University, Gentle, George Mason University (10:15-10:40) 03129. Comparing neural networks and other multilayer approaches for classification, McCracken (10:40-11:05) 03130. Comparing nonlinear approaches for classification, Alzola (11:05-11:30) 03131. Comparing ensemble approaches for classification, Shine (11:30-11:55) 11:55-12:00 Open Discussion/Questions

	Invited: QUANTILE REGRESSION (SHANNON) 03170. Quantile Regression: Beyond the Average Man, Koenker (10:15-10:50) 03001. Quantile Volcano Plots for Identifying Significant Genes in Microarray Data, Li (10:50-11:25) 03050. Quantile Regression for Gene Expression Analysis in GeneChip Arrays, Wang (11:25-12:00) **Invited: SPECTRAL METHODS IN DATA ANALYSIS (SOLKA)** 03098. From Text Data Mining to Gene Expression Mining and Back Again, Solka (10:15-10:50) 03054. Dissimilarity Matrices and Spectral Projections, Leeds (10:50-11:25) 03056. Co-clustering of Social Network Data, Rigsby (11:25-12:00) **Contributed: MULTIDIMESNIONAL SCALING AND OTHER METHODS** 03037. Multidimensional Scaling Algorithms for Large Data Sets, Trosset 03060. Local Multidimensional Scaling for Nonlinear Dimension Reduction, Chen 03002. Maximal Data Piling in Discrimination, Ahn 03110. Estimating the Sparse Directions in the Effective Dimension Reduction Space, Qiao 03115. Uncovering Curvature in Data, Spencer-Smith
12:00 pm to 1:00 pm	Lunch – EPNEC Lobby
1:00 pm to 2:15 pm	Keynote: **John Quackenbush, Harvard University** **TBD**
2:15 pm to 4:00 pm	**Invited: BEST OF JOURNAL OF CLASSIFICATION (HEISER)** 03150. Looking at different options within the COSA clustering algorithm, Meulman (2:15-2:50) 03151. A Hierarchical Methodology for Class Detection Problems with Skewed Priors, Eveland (2:50-3:25) 03152. Estimating the Cluster Tree of a Density, Stuetzle (3:25-4:00) **Invited: NON-NUMERIC DATA ANALYSIS (WHITNEY)** 03145. Conditional Independence Modeling for Categorical Anomaly Detection, Scherrer (2:15-2:45) 03061. Usage-Based Evolution of Visual Analysis Tools, Hetzler (2:45-3:15) 03146. Scenario Analysis for Homeland Security, Kuchar (3:15-3:45) 3:45-4:00 Open Discussion/Questions **Contributed: DISCRIMINATION METHODS AND APPLICATIONS** 03179. PLS-logistic regression for supervised classification, Acuna 03106. Choosing Weights for Nearest-Neighbor Classification with Linear Programming, Buttrey

	03120. Classification via interpoint distance profiles, Wilbur 03091. Polya tree priors for classification error distributions, Neath 03067. Some Classification Rules in Reliability and Survival Analysis, Basu **Contributed: CLUSTERING METHODS AND APPLICATIONS** 03121. Imputation-free robust clustering using soft constraints, Lin 03033. Some Applications of Parametric and Nonparametric Bayesian Predictive Clustering, Quintana 03085. Propensity scoring and the LATE distribution from unsupervised clustering, Lewin-Koh 03041. Profiling Price Dynamics in Online Auctions Using Curve Clustering, Jank 03081. Clustering for measurement error in expenditure survey data, Dixon
4:00 pm to 4:15 pm	Break
4:15 pm to 6:00 pm	**COMPUTATIONAL BIOLOGY (MELLER)** 03040. Spike and Slab Gene Selection for Multigroup Microarray Data, Rao (4:15-4:50) 03038. Bayesian infinite mixture model-based clustering of functional genomics data, Medevedovic (4:50-5:25) 03039. Classification and regression-based approaches to protein structure prediction, Meller (5:25-6:00) **Invited: GRAPH THEORETIC METHODS IN PATTERN RECOGNITION (MARCHETTE)** 03093. Juggling: Ensembles of Class Cover Classifiers, DeVinney, Center For Computing Sciences, Marchette (4:15-4:50) 03007. Geometry of Learning: from Graphs to Continuous Spaces, Belkin (4:50-5:25) 03058. Local intrinsic dimension estimation with kNNGs, Hero (5:25-6:00) **Contributed: ALGORITHMS** 03102. Parallel Computation of the kth Nearest Neighbor Estimate of Entropy of Molecules Using Circular Distances, Harner 03117. Characterizing the Solution Path of Multicategory Support Vector Machines, Cui 03135. Accelerating Linearly-convergent Algorithms, Hesterberg 03070. Alternative Visualization of Andrews' Curves, Martinez 03064. Optimal Linear Combination of Longitudinal Markers for Disease Classification, Ji **Contributed: TIME SERIES, NEURAL NETWORKS, AND HIERARCHICAL MODELS** 03068. The Time Series Prediction Based On Arma And Grnn Technology, Weimin 03084. Discrimination of locally stationary time series, Polonik 03116. Bayesian Neural Networks and Variable Selection, Molnar

	03177. Hierarchical bayesian models for frequent terms in text, Airoldi **Contributed: APPLICATIONS IN BIOLOGY** 03096. Multivariate regression tree: classification of bird assemblages based on their habitat characteristics, Ouellette 03080. Global ClClassification of (Plant) Proteins across Multiple Species, Altman 03099. Phyloinformatics of Genes in Complete Chloroplast Genomes, Kilel 03103. Simulation-based detection of water-borne bacterial contamination, West 03133. Visualizing primate evolution – reification of a statistical model, Rohlf
Sunday, June 12, 2005	
7:30 am to 8:45 am	Continental Breakfast
8:45 am to 10:30 am	**Invited: CLASSIFICATION AND CLUSTERING IN ENVIRONMENTAL APPLICATIONS (WILBUR)** 03029. Statistical modeling and evaluation of microbial source tracking data from repPCR DNA fingerprints, Tenorio (8:45-9:20) 03104. Simulation-based detection of water-borne bacterial contamination, Brian (9:20-9:55) 03028. Scaling by Reference Conditions for Ecological Assessment, Prins (9:55-10:30) **Invited: MODEL BUILDING WITH APPLICATIONS TO MIXTURES AND BIOINFORMATICS (PILLA AND LOADER)** 03148. Analysis of Semiparametric Mixture Models, with Application to QTL Analysis, Fine (8:45-9:20) 03030. A Mixture Model Approach to Multiple Hypothesis Testing, McLachlan (9:20-9:55) 03092. Searching high-dimensional parameter spaces for a parsimonious mixture model, Pilla (9:55-10:30) **Invited: TREES AND NEURAL NETWORKS (LEE)** 03019. Gaussian process trees, Gramacy, Lee (8:45-9:20) 03036. Making use of small samples for classification (for beginners), Steele (9:20-9:55) 03008. Default Bayesian Neural Network Classification, Lee (9:55-10:30) (Session: 8:45 – 10:30)
10:30 am to 10:45 am	Break
10:45 am to 12:30 pm	**Invited: DATA MINING (BANKS)** 03016. Learning classifiers under a limited budget for acquiring training data, Krishnapuram (10:15-10:50) 03017. Nonparametric Statistics and Data Mining, Dasu (10:50-11:25) 03018. Mining Temporal Patterns, Liang, ISDS (11:25-12:00)

	Invited: RECENT DEVELOPMENT OF BIOINFORMATICS (SHIEH) 03127. Selecting an Appropriate Clustering Algorithm for Analyzing Microarray Data, Datta (10:45-11:20) 03128. Estimating Network Topology and Latent Factors in A Gene Network Based on Independent Component Analysis, Wang (11:20-11:55) 03126. Statistical methods for constructing transcriptional regulatory networks using gene expression and DNA sequence data, Xing (11:55-12:30) **Invited: STATISTICAL AUTHOR IDENTIFICATION: FROM CLASSICS TO LIST SERVERS (KANTOR)** 03166. Lexical Predictors of Personality Type, Argamon (10:45-11:20) 03167. Multiple Methods and the Federalist Papers, Sowell, University of the South, Michalek (11:20-11:55) 03168. Author Identification on the Large Scale, Genkin (11:55-12:30)
12:30 pm	Conference Adjourns – Go Home