



Program
for the
**Eleventh ACM SIGKDD
International Conference
on
Knowledge Discovery
and
Data Mining**

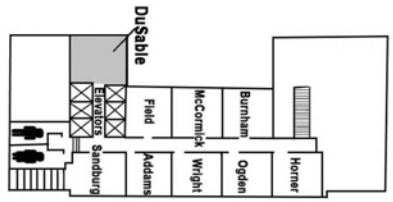
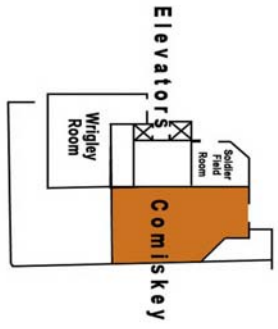
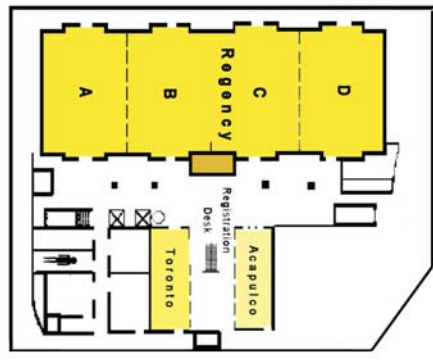
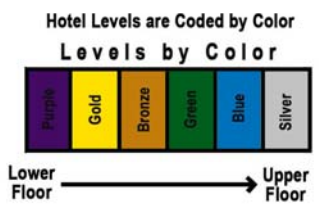
KDD-2005

**Chicago, IL, USA
August 21-24, 2005**



Hyatt Chicago

Gold Level, West Tower Bronze Level, West Tower Silver Level, West Tower



Program Highlights

Invited Talks

- Incentive Networks, Prabhakar Raghavan, *Yahoo! Research*
- Mining the Internet: The Eighth Wonder of the World, Gian Fulgoni, *comScore Networks*
- The Architecture of Complexity: The structure and the dynamics of networks, from the web to the cell, Albert-László Barabási, *Notre Dame*

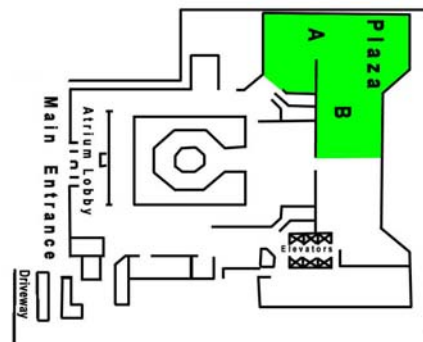
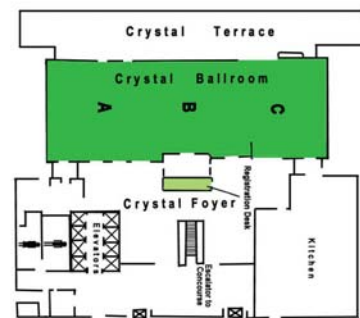
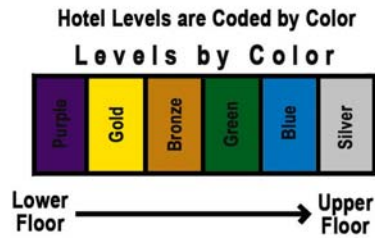
Research and Industrial/Government Tracks

- 40 research papers in 13 sessions
- 14 industrial/government papers in 3 sessions
- 36 research posters
- 11 industrial/government posters
- 2 panels
- 4 tutorials:
 - Introduction to Logistic Regression
 - Data Visualization and Mining using the GPU
 - Randomized Algorithms for Matrices and Massive Data Sets
 - Principles and Applications of Probabilistic Learning
- 9 workshops:
 - Data Mining Methods for Anomaly Detection
 - OSDM 2005: Open Source Data Mining
 - UBDM 2005: Utility-Based Data Mining
 - MRDM 2005: Multi-Relational Data Mining
 - BIODDD 2005: Data Mining in Bioinformatics
 - DM-SSP 2005: Data Mining Standards, Services, and Platforms
 - WebKDD 2005: Taming Evolving, Expanding and Multi-faceted Web Clickstreams
 - LinkKDD 2005: Link Discovery: Issues, Approaches and Applications
 - Multimedia Data Mining: "Mining Integrated Media and Complex Data"

Saturday, August 20

5:00pm-9:00pm (Regency Foyer)

Registration



Hyatt Chicago

Green Level, West Tower

Green Level, East Tower

Summarized Technical Program

Sunday

- 4 Tutorials
- 9 Workshops
- SIGKDD 2005 Opening
- Awards Ceremony
- KDD Cup 2005

Monday

- Invited Talk
- Research Track
 - Temporal Mining (3 papers)
 - Cost Sensitive Learning (2 papers)
 - Privacy (3 papers)
 - Streaming Data (2 papers)
- Industrial/Government Track
 - E-Commerce (3 papers)
- 1 Best Research Paper
- 1 Best Applications Paper
- 1 Best Student Paper, 1 Runner-up
- Poster Highlights
- Poster Session and Reception

Tuesday

- Invited Talk
- Research Track
 - Ensemble Learning (3 papers)
 - Graph Mining (3 papers)
 - Clustering (4 papers)
 - Support Vector Machines (3 papers)
 - Clustering and Grouping (4 papers)
 - Text and Web Mining (4 papers)
- Industrial/Government Track
 - Sequence Mining (3 papers)
 - Anomaly Detection (4 papers)
- 1 Panel

Wednesday

- Invited Talk
- 1 Panel
- Research Track
 - Associations (3 papers)
 - Novel Learning Algorithms (3 papers)
- Industrial/Govt Track
 - Novel Learning Algorithms (3 papers)
 - Document Analysis (3 papers)

Sunday, August 21

7:30am-8:00pm (Regency Foyer)

Registration (ongoing)

9:00am-4:30pm (unless otherwise noted)

Full-Day Workshops

Data Mining Methods for Anomaly Detection
(Starts at 8:30am) (Crystal B)

OSDM 2005: Open Source Data Mining
(DuSable)

UBDM 2005: Utility-Based Data Mining
(Crystal C)

MRDM 2005: Multi-Relational Data Mining
(Toronto)

BIOKDD 2005: Data Mining in Bioinformatics
(Crystal A)

DM-SSP 2005: Data Mining Standards, Services, and Platforms
(Comiskey)

WebKDD 2005: Taming Evolving, Expanding and Multi-faceted Web Clickstreams
(Plaza A)

LinkKDD 2005: Link Discovery: Issues, Approaches and Applications
(Plaza B)

Multimedia Data Mining: "Mining Integrated Media and Complex Data"
(Acapulco)

9:00am-12:00pm (Regency A)

Tutorial: Introduction to Logistic Regression
Dave Lewis, *David D. Lewis Consulting*

9:00am-12:00pm (Regency B)

Tutorial: Randomized Algorithms for Matrices and Massive Data Sets

Petros Drineas, *Rensselaer Polytechnic Institute*
Michael W. Mahoney, *Yale University*

10:00am-10:30am (Regency Foyer)

Coffee Break

12:00pm-1:30pm

Lunch (on your own)

Sunday, August 21

Notes:

1:30pm-4:30pm (Regency A)

Tutorial: Data Visualization and Mining using the GPU

Sudipto Guha, *University of Pennsylvania*

Shankar Krishnan, *AT&T Labs*

Suresh Venkatasubramanian, *AT&T Labs*

1:30pm-4:30pm (Regency B)

Tutorial: Principles and Applications of Probabilistic Learning

Padhraic Smyth, *University of California at Irvine*

3:00pm-3:30pm (Regency Foyer)

Coffee Break

4:30pm-5:00pm

Break

5:00pm-5:45pm (Crystal Ballroom)

KDD Opening Remarks and Awards

Robert Grossman, *General Chair*

Roberto Bayardo, Kristin Bennett, *Program Chairs*

Daniela Raicu, *Student Awards Chair*

Gregory Piatetsky, *SIGKDD Chair*

5:45pm-6:15pm (Crystal Ballroom)

KDD Service Award Presentation

6:15pm-7:15pm (Crystal Ballroom)

KDD Cup Awards

Ying Li, Zijian Zheng, *KDD Cup Chairs*

Notes:

Monday, August 22

7:30am-5:00pm (Regency Foyer)

Registration (ongoing)

7:00am-8:30am (Regency Foyer)

Continental Breakfast—sponsored by SAS

8:30am-10:00am (Crystal Ballroom)

Invited Talk

Chair: Robert Bayardo

Incentive Networks

Prabhakar Raghavan, Yahoo! Research

Abstract:

We propose a notion of incentive networks, modeling online settings in which multiple participants in a network help each other find information. Within this general setting, we study query incentive networks, a natural abstraction of question-answering systems with rewards for finding answers. We analyze strategic behavior in such networks and under a simple model of networks, show that the Nash equilibrium for participants' strategies exhibits an unexpected threshold phenomenon.

(Joint work with Jon Kleinberg.)

10:00am-10:30am (Regency Foyer)

Coffee Break

10:30am-12:00pm Industrial/Government Track

Session 1 (Regency A)

E-Commerce

Chair: Ronny Kohavi

Price Prediction and Insurance for Online Auctions
Rayid Ghani

Predicting Product Purchase Patterns of Corporate Customers
Bhavani Raskutti, Alan Hershtal

Enhancing the Lift Curve Under Budget Constraints: An Application in the Mutual Fund Industry
Lian Yan, Michael Fassino, Patrick Baldasare, Robert Hull

Monday, August 22

10:30am-12:00pm Research Track Session 1
(Regency B)

Temporal Mining

Chair: Jian Pei

Finding Partial Orders from Unordered 0-1 Data
Antti Ukkonen, Mikael Fortelius, Heikki Mannila

Detection of Emerging Space-Time Clusters
Daniel Neill, Andrew Moore, Maheshkumar Sabhnani, Kenny Daniel

Probabilistic Workflow Mining
Ricardo Silva, Jiji Zhang, James G. Shanahan

10:30am-12:00pm Research Track Session 2 (Plaza A)

Privacy

Chair: Ramakrishnan Srikant

A New Scheme on Privacy-Preserving Data Classification
Nan Zhang, Shengquan Wang, Wei Zhao

Anonymity-Preserving Data Collection
Zhiqiang Yang, Sheng Zhong, Rebecca N. Wright

A Distributed Learning Framework Based on Probabilistic Models
Srujana Merugu, Joydeep Ghosh

12:00pm-1:30pm (The Riverside Center West)

Lunch—sponsored by Yahoo! Research Labs

1:30pm-2:30pm Research Track Session 3
(Regency A)

Best Student Papers

Chair: Gautam Das

Query Chains: Learning to Rank from Implicit Feedback
Filip Radlinski and Thorsten Joachims

Summarizing Itemset Patterns: A Profile-based Approach
Xifeng Yan, Hong Cheng, Dong Xin, and Jiawei Han

1:30pm-2:30pm Research Track Session 4 (Regency B)

Cost Sensitive Learning

Chair: Marko Grobelnik

Local Sparsity Control for Naïve Bayes with Extreme Misclassification Costs
Aleksander Kolcz

Combining Email Models for False Positive Reduction
Shlomo Hershkop, Salvatore Stolfo

SIGKDD-2005 Program Committee, cont.

Philip Yu, *IBM Watson, USA*
Bianca Zadrozny, *IBM Watson Research Center, USA*
Osmar Zaiane, *U of Alberta, Canada*
Carlo Zaniolo, *U of California, USA*
Aidong Zhang, *State U of New York, USA*
Tong Zhang, *IBM TJ Watson, USA*

Industrial/Government Track Program Committee

Robert Cooley. *KXEN, USA*
Mary Crissey. *SAS Institute, USA*
Tamraparni Dasu. *AT&T Labs – Research, USA*
Mayur Datar. *Google Inc, USA*
Anand Deshpande. *Persistent Systems, India*
Steve Donoho. *Donoho Analytics Inc., USA*
Mark Foresti. *Air Force Research Laboratory, USA*
Ashutosh Garg. *Google Inc, USA*
Chandrika Kamath. *LLNL, USA*
Sigal Louchheim. *Intel, USA*
Brendan Kitts. *iProspect.com, USA*
Gabor Melli. *PredictionWorks Inc., USA*
Thomas Niccum. *Lancet Software Development Inc., USA*
Ajay B. Pandey. *Government of Maharashtra, India*
Claudia Pearce. *National Security Agency, USA*
Ed Pednault. *IBM TJ Watson Research Center, USA*
Valery A. Petrushin. *Accenture Technology Labs, USA*
Bonnie Ray. *IBM TJ Watson Research Center, USA*
Greg Ridgeway. *RAND, USA*
Saharon Rosset. *IBM TJ Watson Research Center, USA*
Matthias Schonlau. *RAND Corporation, USA*
Eric V. Siegel. *Prediction Impact, USA*
Neal Rothleder. *Microsoft, USA*
Volker Tresp. *Siemens AG, Germany*
Samy Uthurusamy. *General Motors, USA*
Raju Vatsavai. *IBM, USA*
Jamshid Vayghan. *IBM TJ Watson Research Center, USA*
Chris Volinsky. *AT&T Labs – Research, USA*
Leland Wilkinson. *SPSS, USA*
Kenji Yamanishi. *NEC Corporation, Japan*
Chunsheng Yang. *National Research Council, USA*

Best Paper Awards Committee

Heikki Mannila, *Helsinki Univ. of Technology, Finland*
Gautam Das, *University of Texas, USA*

ACM SIGKDD Chair

Gregory Piatetsky, *KDNuggets, USA*

SIGKDD-2005 Program Committee, cont.

George Karypis, *U of Minnesota, USA*
Daniel Keim, *U of Constance, Germany*
David Kempe, *U of Southern California, USA*
Eamonn Keogh, *U of California Riverside, USA*
Ronny Kohavi, *Amazon.com, USA*
Aleksander Kolcz, *America Online, Inc, USA*
Vipin Kumar, *U of Minnesota, USA*
Diane Lambert, *Bell Labs, USA*
Kahn Latifur, *U of Texas at Dallas, USA*
Bing Liu, *UIC, USA*
Wei-yin Loh, *U of Wisconsin, USA*
Richard Maclin, *U of Minnesota Duluth, USA*
Brij Masand, *Data Miners, USA*
Yossi Matias, *Tel Aviv U & HyperRoll, Israel*
Rosa Meo, *U of Torino, Italy*
Nina Mishra, *HP Labs/ Stanford U, USA*
Mladenec, *J. Stefan Institute, Slovenia*
Dharmendra Modha, *IBM Almaden, USA*
Michinari Momma, *Fair Isaac Corporation, USA*
Hiroshi Motoda, *Osaka U, Japan*
Alejandro Murua, *U of Washington, USA*
Dave Musicant, *Carleton College, USA*
Ion Muslea, *SRI International, USA*
Raymond Ng, *U of British Columbia, USA*
David Page, *U of Wisconsin, USA*
D. Scott Parker Jr, *UCLA, USA*
Dmitry Pavlov, *Yahoo! Inc, USA*
Jian Pei, *Simon Fraser U, Canada*
Dan Pelleg, *Carnegie-Mellon U, USA*
Drineas Petros, *Rensselaer Polytechnic Institute, USA*
Raghu Ramakrishnan, *University of Wisconsin, USA*
Greg Ridgeway, *RAND, USA*
Cynthia Rudin, *New York U, USA*
Lorenza Saitta, *Universita del Piemonte Orientale, Italy*
Sunita Sarawagi, *IIT Bombay, India*
Thomas Seidl, *RWTH Aachen U, Germany*
Kyuseok Shim, *Seoul National U, South Korea*
Arno Siebes, *Utrecht U, The Netherlands*
Simeon Simoff, *U of Technology Sydney, Australia*
Dan Simovici, *U of Massachusetts Boston, USA*
Myra Spiliopoulou, *U of Magdeburg, Germany*
Ramakrishnan Srikant, *IBM Almaden, USA*
Salvador Stolfo, *Columbia U, USA*
Werner Stuetzle, *U of Washington, USA*
Einoshin Suzuki, *Yokohama National U, Japan*
Hannu Toivonen, *U of Helsinki, Finland*
Ke Wang, *Simon Fraser U, Canada*
Wei Wang, *U of North Carolina, USA*
Geoff Webb, *Monash U, Australia*
Ran Wolff, *U of Maryland, USA*
Rebecca Wright, *Stevens Institute of Technology, USA*
Xindong Wu, *U of Vermont, USA*

Monday, August 22

1:30pm-2:30pm Research Track Session 5 (Plaza A)

Streaming Data**Chair:** Petros Drineas

Streaming Feature Selection Using Alpha Investing
Jing Zhou, Dean Foster, Robert Stine, and Lyle Ungar

Wavelet Synopsis for Data Streams: Minimizing non-Euclidean Error
Sudipto Guha and Boulos Harb

2:30pm-3:30pm Paper Award Talks (Crystal Ballroom)

Best Paper Award Session**Chair:** Heikki Mannila**BEST RESEARCH PAPER AWARD**

Graphs Over Time: Densification Laws, Shrinking Diameters, and Possible Explanations
Jure Leskovec, Jon Kleinberg, and Christos Faloutsos

BEST APPLICATION PAPER AWARD

A Hit-Miss Model for Duplicate Detection in the WHO Drug Safety Database
Niklas Norén, Roland Orre, Andrew Bate

3:30pm-4:15pm (Crystal Ballroom)

Plenary Poster Presentations

4:15pm-4:45pm (Regency Foyer)

Coffee Break

4:45pm-5:45pm (Crystal Ballroom)

Plenary Poster Presentations

6:15pm-7:00pm

Buses begin leaving for Field Museum

7:00pm-10:00pm (Field Museum)

Poster Reception—sponsored by Fair Isaac**Poster Papers — Research Track**

CLICKS: An Effective Algorithm for Mining Subspace Clusters in Categorical Datasets
Mohammed Zaki, Markus Peters, Ira Assent, Thomas Seidl

Web Mining from Competitors Websites
Xin Chen, Yi-fang Wu

Evaluating Similarity Measures: A Large Scale Study in the Orkut Social Network
Ellen Spertus, Mehran Sahami, Orkut Buyukkokten

Monday, August 22

Poster Papers — Research Track

- Key semantics extraction by dependency tree mining
Satoshi Morinaga, Hiroki Arimura, Takahiro Ikeda, Yosuke Sakao, Susumu Akamine
- Regression Error Characteristic Surfaces
Luis Torgo
- Privacy-Preserving Distributed k-means Clustering over Arbitrarily Partitioned Data
Geetha Jagannathan, Rebecca N. Wright
- LIPED: HMM-based Life Profiles for Adaptive Event Detection
Chien Chin Chen, Meng Chang Chen, Ming-Syan Chen
- Estimating missed actual positives using independent classifiers
Sandeep Mane, Jaideep Srivastava, San-Yih Hwang
- A Hybrid Unsupervised Approach for Document Clustering
Mihai Surdeanu, Jordi Turmo, Alicia Ageno
- Mining in Anticipation: Proactive-Reactive Prediction for Data Streams
Ying Yang, Xindong Wu, Xingquan Zhu
- Optimizing time series discretization for knowledge discovery
Fabian Mörchen, Alfred Ultsch
- A Generalized Framework For Mining Spatio-temporal Patterns in Scientific Data
Hui Yang, Sameep Mehta, Srinivasan Parthasarathy
- Density-Based Clustering of Uncertain Data
Martin Pfeifle, Hans-Peter Kriegel
- Information Retrieval Based on Collaborative Filtering With Latent Interest Semantic Map
Noriaki Kawamae
- Parallel Mining of Closed Sequential Patterns
Shengnan Cong, Jiawei Han, David Padua
- Determining an Author's Native Language by Mining a Text for Errors
Moshe Koppel, Jonathan Schler, Kfir Zigdon
- Pattern Lattice Traversal by Selective Jumps
Osmar Zaiane, Mohammad El Hajj

SIGKDD-2005 Program Committee

- Brigham Anderson (*Carnegie Mellon U, USA*)
Charu Aggarwal (*IBM Watson Research Center, USA*)
Corin Anderson (*Google, USA*)
Chid Apte (*IBM T.J. Watson Research Center, USA*)
Daniel Barbara (*George Mason U, USA*)
Sugato Basu (*U of Texas at Austin, USA*)
Shai Ben-David (*University of Waterloo, Canada*)
Michael Berthold (*Konstanz Univ, Germany*)
Jinbo Bi (*Siemens Medical Solutions, Inc., USA*)
Richard Bolton (*KnowledgeBase Marketing Inc., USA*)
Jean-Francois Boulicaut (*INSA Lyon, France*)
Wray Buntine (*HIIT, Finland*)
Rich Caruana (*Cornell U, USA*)
Soumen Chakrabarti (*IIT Bombay, India*)
David Cheung (*U of Hong Kong, Hong Kong*)
Rada Chirkova (*North Carolina State U, USA*)
Diane Cook (*U of Texas at Arlington, USA*)
Gautam Das (*U of Texas at Arlington, USA*)
Inderjit Dhillon (*U of Texas at Austin, USA*)
Chabane Djeraba (*LIFL - UMR CNRS 8022, France*)
Carlotta Domeniconi (*George Mason U, USA*)
Jennifer Dy (*Northeastern U, USA*)
Tina Eliassi-Rad (*LLNL, USA*)
Charles Elkan (*U of California, USA*)
Mark Embrechts (*Rensselaer Polytechnic Institute, USA*)
Alexandre Evfimievski (*IBM Almaden, USA*)
Theos Evgeniou (*INSEAD, France*)
Chistos Faloutsos (*CMU, USA*)
Wei Fan (*IBM T.J. Watson, USA*)
Ronen Feldman (*Bar-Ilan U, Israel*)
Doug Fisher (*Vanderbilt University, USA*)
Peter Flach (*U of Bristol, UK*)
Gary Flake (*Yahoo! Research Labs, USA*)
Glenn Fung (*Siemens Medical Solutions, USA*)
Thomas Gärtner (*Fraunhofer AIS, Germany*)
Venkatesh Ganti (*Microsoft Research, USA*)
Johannes Gehrke (*Cornell U, USA*)
Rayid Ghani (*Accenture Technology Labs, USA*)
Joydeep Ghosh (*U of Texas at Austin, USA*)
Phillip Gibbons, *Intel Research Pittsburgh, USA*
C. Lee Giles, *The Pennsylvania State U, USA*
Aristides Gionis, *U of Helsinki, Finland*
Bart Goethals, *U of Antwerp, Belgium*
Mark Goldberg, *Rensselaer Polytechnic Institute, USA*
Marko Grobelnik, *Jozef Stefan Institute, Slovenia*
Sudipto Guha, *U of Pennsylvania, USA*
Dimitrios Gunopulos, *U of California Riverside, USA*
Jiawei Han, *UIUC, USA*
David Jensen, *U of Massachusetts Amherst, USA*
Chris Jeramie, *U of Florida, USA*
Thorsten Joachims, *Cornell U, USA*
Jugal Kalita, *U of Colorado at Colorado Springs, USA*
Hillol Kargupta, *U of Maryland, USA*

SIGKDD-2005 Organizing Committee

General Chair:

Robert L. Grossman, Univ. of Illinois at Chicago and
Open Data Partners, USA

Program Chairs:

Roberto Bayardo, *IBM Almaden Research, USA*
Kristin Bennett, *Rensselaer Polytechnic Institute, USA*

Industrial/Government Track Chairs:

Corinna Cortes, *Google, USA*
Jaideep Srivastava, *University of MN, USA*

Best Paper Awards Chair:

Heikki Mannila, *Helsinki Univ. of Technology, Finland*

Exhibits Chairs:

Gabor Melli, *PredictionWorks Inc., USA*

KDD Cup Chairs:

Ying Li, *Microsoft, USA*
Zijian Zheng, *Amazon.com, USA*

Local Arrangements Chairs:

Peter Caron, *SPSS, USA*
Shirley Connelly, *Univ. of Illinois at Chicago, USA*
Bamshad Mobasher, *DePaul University, USA*

Publicity Chair:

David Duling, *SAS, USA*

Local Publicity Chair:

David Turkington, *Univ. of Illinois at Chicago, USA*

Panels Chair:

Usama Fayyad, *Yahoo Inc., USA*

Proceedings Chair:

Jaideep Vaidya, *Rutgers University, USA*

Registration Chairs:

Ashfaq Khokar, *Univ. of Illinois at Chicago, USA*

Sponsorship Chairs:

Gabor Melli, *PredictionWorks Inc., USA*
Stephen G. Eick, *SSS Research and UIC, USA*

Student Awards Chair:

Daniela Raicu, *DePaul University, USA*

Treasurer:

Christopher Clifton, *Purdue University, USA*

Tutorials Chair:

Carla Brodley, *Tufts University, USA*

Webmaster:

Michal Sabala, *Univ. of Illinois at Chicago, USA*

Workshops Chair:

Mohammed Zaki, *Rensselaer Polytechnic Institute, USA*

Monday, August 22

Poster Papers — Research Track**Adversarial Learning**

Daniel Lowd, Chris Meek

Co-clustering by Block Value Decomposition

Bo Long, Zhongfei Zhang, Philip Yu

Application of kernels to link analysis

*Takahiko Ito, Masashi Shimbo, Taku Kudo, Yuji
Matsumoto*

Model-based Overlapping Clustering

*Arindam Banerjee, Chase Krumpelman, Sugato Basu,
Raymond Mooney, Joydeep Ghosh*

**Building Connected Neighborhood Graphs for Isometric Data
Embedding**

Li Yang

**Integration of Profile Hidden Markov Model Output into
Association Rule Mining**

Christopher Besemann, Anne Denton

**Towards Exploratory Test Instance Specific Algorithms for
High Dimensional Classification**

Charu Aggarwal

**Simultaneous Optimization of Complex Mining Tasks with a
Knowledgeable Cache**

Ruoming Jin, Kaushik Sinha, Gagan Agrawal

**Discovering Frequent Topological Structures from Graph
Datasets**

*Ruoming Jin, Chao Wang, Dmitrii Polshakov, Srinivasan
Parthasarathy, Gagan Agrawal*

**Efficient Computations via Scalable Sparse Kernel Partial
Least Squares and Boosted Latent Features**

Michinari Momma

Scalable Discovery of Hidden Emails from Large Folders

Giuseppe Carenini, Raymond Ng, Xiaodong Zhou

Formulating Distance Functions via the Kernel Trick

Gang Wu, Navneet Panda, Edward Chang

Fast Window Correlations Over Uncooperative Time Series

Xiaojuan Zhao, Dennis Shasha, Richard Cole

Monday, August 22

A Maximum Entropy Web Recommendation System:
Combining Collaborative and Content Features
Authors: Xin Jin, Yanzan Zhou, Bamshad Mobasher

Mining Comparable Bilingual Text Corpora for Cross-
Language Information Integration
Tao Tao, ChengXiang Zhai

Creating social networks to improve peer-to-peer networking
Andrew Fast, David Jensen, Brian Neil Levine

A Fast Kernel-based Multilevel Algorithm for Graph Clustering
Brian Kulis, Yuqiang Guan, Inderjit Dhillon

Unweaving a Web of Documents
R. Guha, Ravi Kumar, D. Sivakumar, Ravi Sundaram

Maximal Boasting
Cinda Heeren, Leonard Pitt

Poster Papers — Industrial/Government Track

Automated detection of frontal systems from numerical
model-generated data
*Xiang Li, Rahul Ramachandran, Sara Graves, Sunil
Movva, Bilahari Akkiraju, David Emmitt, Steven Greco,
Robert Atlas, Joe Terry, Juan Carlos Jusem*

Failure Detection and Localization in Component Based
Systems by Online Tracking
*Haifeng Chen, Guofei Jiang, Cristian Ungureanu, Kenji
Yoshihira*

Mining Rare and Frequent Events in Multi-camera
Surveillance Video using Self-organizing Maps
Valery Petrushin

Data Mining in the Chemical Industry
Alex Kalos, Tim Rey

Short-term performance forecasting in enterprise systems
Rob Powers, Moises Goldszmidt, Ira Cohen

Mining Risk Patterns in Medical Data
*Jiuyong Li, Ada Wai-chee Fu, Hongxing He, Jie Chen,
Huidong Jin, Damien McAullay, Graham Williams,
Ross Sparks, Chris Kelman*

Acknowledgements

The SIGKDD 2005 Conference gratefully acknowledges
the contributions of the following institutions:

Organizational Sponsor



Platinum Sponsor



Gold Supporters



Silver Supporter



Bronze Supporters



Additional Organizational Sponsors



Wednesday, August 24

11:00am-12:30pm Research Track Session 12
(Regency B)**Associations****Chair:** Bing LiuReasoning about Sets using Redescription Mining
*Mohammed Zaki, Naren Ramakrishnan*Improving Discriminative Sequential Learning with Rare-but Important Associations
*Phan Xuan-Hieu, Nguyen Le-Minh, Ho Tu-Bao, Horiguchi Susumu*A Multiple Tree Algorithm for the Efficient Association of Asteroid Observations
*Jeremy Kubica, Andrew Moore, Andrew Connolly, Robert Jedicke***11:00am-12:30pm Research Track Session 13**
(Plaza A/B)**Novel Learning Algorithms****Chair:** Dan SimoviciFast Discovery of Unexpected Patterns in Data Relative to a Bayesian Network
*Szymon Jaroszewicz, Tobias Scheffer*A Bayesian Network Classifier with Inverse Tree Structure for Voxelwise Magnetic Resonance Image Analysis
*Rong Chen, Edward Herskovits*Mining Images on Semantics via Statistical Learning
Jianping Fan Fan, Mohand-Said Hacid

Monday, August 22

Poster Papers — Industrial/Government TrackDisease Progression Modelling from Historical Clinical Databases
*Ronald Pearson, Robert Kingan, Alan Hochberg*An Integrated Framework on Mining Logs Files for Computing System Management
*Tao Li, F. Liang, Sheng ma, W. Peng*Generation of Synthetic Data Sets for Evaluating the Accuracy of Knowledge Discovery Systems
*Daniel Jeske, Behrokh Samadi, James Lin, Lan Ye, Sean Cox, Rui Xiao, Ted Younglove, Minh Ly, Doug Holt, Ryan Rich*Pattern-based Similarity Search for Microarray Data
*Haixun Wang, Jian Pei, Philip Yu*A Multinomial Clustering Model for fast simulation of computer architecture designs
Kaushal Sanghai, Ting Su, Jennifer Dy, David Kaeli

Tuesday, August 23

7:30am-5:00pm (Regency Foyer)**Registration** (ongoing)**7:00am-8:30am** (Regency Foyer)**Continental Breakfast** – sponsored by SPSS**8:30am-10:00am** (Crystal Ballroom)**Invited Talk****Chair: Robert Grossman****Mining the Internet: The Eighth Wonder of the World**

Gian Fulgoni, comScore Networks

Abstract:

The Internet takes behavioral consumer research to a new level by providing the ability to passively and continuously monitor the complete online behavior of millions of consumers in an opt-in, privacy protected manner. Imagine the analytical possibilities if every site visited, every page viewed, content seen, transaction conducted..... all of this granularity in behavior --- was continuously captured with explicit consumer permission for millions of consumers and privacy was protected. What unique insights could one gain into consumers' behavior, their interests, passions and lifestyles? What behavior could be predicted? What commercial applications would be possible?

Tuesday, August 23

10:00am-10:30am (Regency Foyer)

Coffee Break

10:30am-12:00pm Industrial/Government Track

Session 2 (Regency B)

Sequence Mining

Chair: Myra Spiliopoulou

Exploiting Retrieval Measures in the Early Stages of Mining Evolving Web Clickstreams

Ofa Nasraoui, Cesar Cardona, Carlos Rojas

Email Data Cleaning

Jie Tang, Hang Li, Yunbo Cao, ZhaoHui Tang

Modeling and Predicting Personal Information Dissemination Behavior

Xiaodan Song, Ching-Yung Lin, Belle L. Tseng, Ming-Ting Sun

10:30am-12:00pm Research Track Session 6

(Regency A)

Ensemble Learning

Chair: Jennifer Dy

Robust Boosting and its Relation to Bagging

Saharon Rosset

Feature Bagging for Outlier Detection

Aleksandar Lazarevic, Vipin Kumar

Combining Partitions by Probabilistic Label Aggregation

Tilman Lange, Joachim Buhmann

10:30am-12:00pm Research Track Session 7

(Plaza A/B)

Graph Mining

Chair: Tina Eliassi-Rad

Mining Tree Queries in a Graph

Bart Goethals, Eveline Hoekx, Jan Van den Bussche

On Mining Cross-Graph Quasi-Cliques

Jian Pei, Daxin Jiang, Aidong Zhang

Mining Closed Relational Graphs with Connectivity Constraints

Xifeng Yan, X. Jasmine Zhou, Jiawei Han

12:00pm-2:00pm (The Riverside Center West)

SIGKDD Business Lunch—sponsored by Microsoft SQL Server 2005

Wednesday, August 24

9:30am-10:30am (Regency C&D)

Invited Talk

Chair: Christos Faloutsos

The architecture of complexity: The structure and the dynamics of networks from the web to the cell
Albert-László Barabási, Notre Dame

Networks with complex topology describe systems as diverse as the cell, the World Wide Web or the society. The emergence of most networks is driven by self-organizing processes that are governed by simple but generic laws. The analysis of the cellular network of various organisms shows that cells and complex man-made networks, such as the Internet or the world wide web, and many social and collaboration networks share the same large-scale topology. I will show that the scale-free topology of these complex webs have important consequences on their robustness against failures and attacks, with implications on drug design, the Internet's ability to survive attacks and failures, and the ability of ideas and innovations to spread on the network.

10:30am-11:00am (Regency Foyer)

Coffee Break

11:00am-12:30pm Industrial/Government Track

Session 4 (Regency A)

Document Analysis

Chair: Gabor Melli

Finding Similar Files in Large Document Repositories

George Forman, Kave Eshghi, Stephane Chiochetti

Making Holistic Schema Matching Robust: An Ensemble Approach

Bin He, Kevin Chen-Chuan Chang

Deriving Marketing Intelligence from Online Discussion

Natalie Gance, Matthew Hurst, Kamal Nigam, Matthew Siegler, Robert Stockton, Takashi Tomokiyo

Wednesday, August 24

7:30am-8:30am (Regency Foyer)

Continental Breakfast – sponsored by Teradata

8:30am-9:30am (Regency C&D)

Plenary Panel: Selling Vitamins Instead of Aspirin - the Data Mining Adoption Challenge

Chair: George John, *Yahoo!*

Ultimately computer scientists trade in bits. Some build boxes that take in bits, remember them, then give them back when asked, maybe adding up a few numbers in the process. This is the database and enterprise applications business, \$100B/year in revenues. Other computer scientists build boxes that take in bits on one end and shoot them out the other end. This is the networking business, also \$100B/year.

As data miners, we build boxes that take in bits, perform magical computations, and create models that can actually predict future behavior and events in a way that allows a business to significantly grow revenues or reduce costs, or we discover structure or patterns that allow knowledge workers or scientists to make more rapid progress towards significant discoveries.

So why isn't KDD also a \$100B business? Where is our Bill Gates, our Larry Ellison, our Cisco, our SAP? Does Usama Fayyad have a house with a trampoline in his 13th bedroom? Does Jim Goodnight race in the America's cup? Did you take a vitamin today? The last time you had a bad headache, did you take an aspirin?

At the Vitamins vs Aspirin panel, representatives from Fortune 500 companies will give their views on prioritizing investments in data mining, representatives from data mining companies will describe the ups and downs of corporate adoption, and we will get to the bottom of how to make sure everyone takes their vitamins.

Panelists:

Usama Fayyad, *Yahoo!*

Robert Grossman, *Open Data Partners and University of Illinois at Chicago*

Ronny Kohavi, *Microsoft*

Tuesday, August 23

2:00pm-4:00pm Research Track Session 8 (Regency A)

Clustering

Chair: Sugato Basu

Dimension Induced Clustering

Aris Gionis, Alexander Hinneburg, Spiros Papadimitriou, Panayiotis Tsaparas

On the Use of Linear Programming for Unsupervised Text Classification

Mark Sandler

A General Model for Clustering Binary Data

Tao Li

Consistent Bipartite Graph Co-Partitioning for Star-Structured High-Order Heterogeneous Data Co-Clustering

Bin Gao, Tie-Yan Liu, Xin Zheng, Qian-sheng Chen, Wei-Ying Ma

2:00pm-3:30pm Research Track Session 9 (Regency B)

Support Vector Machines

Chair: Dave Musicant

SVM Selective Sampling for Ranking with Application to Data Retrieval

Hwanjo Yu

Rule Extraction from Hyperplane-based Classifiers

Glenn Fung, Sathyakama Sandilya, Bharat Rao

Nomograms for Visualizing Support Vector Machines

Aleks Jakulin, Martin Mozina, Janez Demsar, Ivan Bratko, Blaz Zupan

2:00pm-3:30pm (Crystal B)

Panel: Text mining—the discipline that never was

Chair: Prabhakar Raghavan, *Yahoo! Research*

Hundreds of papers later, we are still unable to define just what text mining is. Is there a definitive, valuable discipline here with firm scientific foundations? Or is it too nascent to tell? Or is it just a special case of structured data mining? Is it just IR re-invented or is there something new here?

Join our panelists in debating this audience-interactive panel.

Panelists:

Andrei Broder, *IBM*

Natalie Glance, *Intelliseek*

Jon Kleinberg, *Cornell*

Tuesday, August 23

3:30pm-4:00pm (Regency Foyer)

Coffee Break

4:00pm-6:00pm Research Track Session 10

(Regency A)

Clustering and Grouping

Chair: Wei Wang

Non-Redundant Clustering with Conditional Ensembles
David Gondek, Thomas Hofmann

Cross-Relational Clustering with User's Guidance
Xiaoxin Yin, Jiawei Han, Philip Yu

Sampling-Based Sequential Subgroup Mining
Martin Scholz

Simple and Effective Visual Models for Gene Expression
Cancer Diagnostics
Gregor Leban, Minca Mramor, Ivan Bratko, Blaz Zupan

4:00pm-6:00pm Industrial/Government Track Session 3
(Regency B)

Anomaly Detection

Chair: Valery A. Petrushin

Dynamic Syslog Mining for Network Failure Monitoring
Kenji Yamanishi, Yuko Maruyama

Learning to Predict Train Wheel Failures
Chunsheng Yang, Sylvain Letourneau

Using Relational Knowledge Discovery to Prevent
Securities Fraud
Jennifer Neville, Ozgur Simsek, David Jensen, John Komoroske, Kelly Palmer, Henry Goldberg

An Approach to Spacecraft Anomaly Detection Problem
Using Kernel Feature Space
Ryohei Fujimaki, Takehisa Yairi, Kazuo Machida

Tuesday, August 23

4:00pm-6:00pm Research Track Session 11 (Plaza A/B)

Text and Web Mining

Chair: Eric V. Siegel

The Predictive Power of Online Chatter
Daniel Gruhl, R. Guha, Ravi Kumar, Jasmine Novak, Andrew Tomkins

Discovering Evolutionary Theme Patterns from Text - An
Exploration of Temporal Text Mining.
Qiaozhu Mei, ChengXiang Zhai

Variable Latent Semantic Indexing
Anirban Dasgupta, Ravi Kumar, Prabhakar Raghavan, Andrew Tomkins

Web Object Indexing Using Domain Knowledge
Muyuan Wang, Zhiwei Li, Lie Lu, Wei-Ying Ma, Naiyao Zhang

6:00pm-6:45pm (Regency A)

KDD Transfer meeting (organizing committee only)

7:15pm-10:30pm (Off-site)

Program Committee and Organizing Committee Dinner
(by invitation only). See Peter Caron for details.