

# **2013 IEEE 13th International Conference on Data Mining (ICDM 2013)**

**Dallas, Texas, USA  
7 - 10 December 2013**

**Pages 1-666**



**IEEE Catalog Number: CFP13278-POD  
ISBN: 978-1-4799-1328-2**

# 2013 IEEE 13th International Conference on Data Mining

## ICDM 2013

### Table of Contents

|  |        |
|--|--------|
| Message from the Conference Chairs ..... | xvi    |
| Message from the Program Co-Chairs ..... | xviii  |
| Conference Organizers .....              | xx     |
| Program Committee .....                  | xxiii  |
| Keynotes .....                           | xxxi   |
| Tutorials .....                          | xxxvii |
| List of Sponsors .....                   | xl     |

---

#### Regular Papers

|   |    |
|---|----|
| Tree-Like Structure in Large Social and Information Networks .....  | 1  |
| <i>Aaron B. Adcock, Blair D. Sullivan, and Michael W. Mahoney</i>   |    |
| Subgraph Enumeration in Dynamic Graphs .....  | 11 |
| <i>Abhijin Adiga, Anil Kumar S. Vullikanti, and Dante Wiggins</i>   |    |
| A Masking Index for Quantifying Hidden Glitches .....   | 21 |
| <i>Laure Berti-Équille, Ji Meng Loh, and Tamraparni Dasu</i>  |    |
| CSI: Charged System Influence Model for Human Behavior Prediction .....   | 31 |
| <i>Yuanjun Bi, Weili Wu, and Yuqing Zhu</i>   |    |
| Context-Aware MIML Instance Annotation .....  | 41 |
| <i>Forrest Briggs, Xiaoli Z. Fern, and Raviv Raich</i>  |    |
| Maximizing Expected Model Change for Active Learning in Regression .....  | 51 |
| <i>Wenbin Cai, Ya Zhang, and Jun Zhou</i>   |    |
| The Pairwise Gaussian Random Field for High-Dimensional Data Imputation .....   | 61 |
| <i>Zhuhua Cai, Christopher Jermaine, Zografoula Vagena, Dionysios Logothetis,<br/>and Luis L. Perez</i>                   |    |
| Controlling Attribute Effect in Linear Regression .....   | 71 |
| <i>Toon Calders, Asim Karim, Faisal Kamiran, Wasif Ali, and Xiangliang Zhang</i>  |    |
| Active Matrix Completion .....  | 81 |
| <i>Shayok Chakraborty, Jiayu Zhou, Vineeth Balasubramanian,<br/>Sethuraman Panchanathan, Ian Davidson, and Jieping Ye</i> |    |

|  |     |
|--|-----|
| Modeling Temporal Adoptions Using Dynamic Matrix Factorization .....   | 91  |
| <i>Freddy Chong Tat Chua, Richard J. Oentaryo, and Ee-Peng Lim</i>   |     |
| Modeling Preferences with Availability Constraints .....   | 101 |
| <i>Bing Tian Dai and Hady W. Lauw</i>  |     |
| wRACOG: A Gibbs Sampling-Based Oversampling Technique .....  | 111 |
| <i>Barnan Das, Narayanan C. Krishnan, and Diane J. Cook</i>  |     |
| Efficient Visualization of Large-Scale Data Tables through Reordering and Entropy<br>Minimization .....        | 121 |
| <i>Nemanja Djuric and Slobodan Vucetic</i>   |     |
| Local and Global Discriminative Learning for Unsupervised Feature Selection .....                              | 131 |
| <i>Liang Du, Zhiyong Shen, Xuan Li, Peng Zhou, and Yi-Dong Shen</i>  |     |
| Generative Maximum Entropy Learning for Multiclass Classification .....  | 141 |
| <i>Ambedkar Dukkipati, Gaurav Pandey, Debarghya Ghoshdastidar, Paramita Koley,<br/>and D.M.V. Satya Sriram</i> |     |
| A Parameter-Free Spatio-Temporal Pattern Mining Model to Catalog Global Ocean<br>Dynamics .....                | 151 |
| <i>James H. Faghmous, Matthew Le, Muhammed Uluyol, Vipin Kumar,<br/>and Snigdhasu Chatterjee</i>               |     |
| Transfer Learning across Networks for Collective Classification .....  | 161 |
| <i>Meng Fang, Jie Yin, and Xingquan Zhu</i>  |     |
| Distributed Column Subset Selection on MapReduce .....   | 171 |
| <i>Ahmed K. Farahat, Ahmed Elgohary, Ali Ghodsi, and Mohamed S. Kamel</i>                                      |     |
| Compression-Based Graph Mining Exploiting Structure Primitives .....   | 181 |
| <i>Jing Feng, Xiao He, Nina Hubig, Christian Böhm, and Claudia Plant</i>                                       |     |
| Online Estimation of Discrete Densities .....  | 191 |
| <i>Michael Geilke, Eibe Frank, Andreas Karwath, and Stefan Kramer</i>  |     |
| Extraction of Interpretable Multivariate Patterns for Early Diagnostics .....                                  | 201 |
| <i>Mohamed F. Ghalwash, Vladan Radosavljevic, and Zoran Obradovic</i>  |     |
| Search Behavior Based Latent Semantic User Segmentation for Advertising Targeting .....                        | 211 |
| <i>Xueqing Gong, Xinyu Guo, Rong Zhang, Xiaofeng He, and Aoying Zhou</i>                                       |     |
| Mixed Membership Subspace Clustering .....   | 221 |
| <i>Stephan Günnemann and Christos Faloutsos</i>  |     |
| Spectral Subspace Clustering for Graphs with Feature Vectors .....   | 231 |
| <i>Stephan Günnemann, Ines Färber, Sebastian Raubach, and Thomas Seidl</i>                                     |     |
| Permutation-Based Sequential Pattern Hiding .....  | 241 |
| <i>Robert Gwadera, Aris Gkoulalas-Divanis, and Grigorios Loukides</i>  |     |
| Utilizing URLs Position to Estimate Intrinsic Query-URL Relevance .....  | 251 |
| <i>Xiaogang Han, Wenjun Zhou, Xing Jiang, Hengjie Song, Ming Zhong, and Toyooki Nishida</i>                    |     |

|   |     |
|---|-----|
| Parameter-Free Audio Motif Discovery in Large Data Archives .....   | 261 |
| <i>Yuan Hao, Mohammad Shokoohi-Yekta, George Papageorgiou, and Eamonn Keogh</i>   |     |
| Kernel Density Metric Learning .....  | 271 |
| <i>Yujie He, Wenlin Chen, Yixin Chen, and Yi Mao</i>  |     |
| Classification of Multi-dimensional Streaming Time Series by Weighting Each<br>Classifier’s Track Record .....                        | 281 |
| <i>Bing Hu, Yanping Chen, Jesin Zakaria, Liudmila Ulanova, and Eamonn Keogh</i>   |     |
| Identifying Transformative Scientific Research .....  | 291 |
| <i>Yi-Hung Huang, Chun-Nan Hsu, and Kristina Lerman</i>   |     |
| Min-Max Hash for Jaccard Similarity .....   | 301 |
| <i>Jianqiu Ji, Jianmin Li, Shuicheng Yan, Qi Tian, and Bo Zhang</i>   |     |
| GRIAS: An Entity-Relation Graph Based Framework for Discovering Entity Aliases .....  | 310 |
| <i>Lili Jiang, Ping Luo, Jianyong Wang, Yuhong Xiong, Bingduan Lin, Min Wang,<br/>and Ning An</i>                                     |     |
| Focal-Test-Based Spatial Decision Tree Learning: A Summary of Results .....   | 320 |
| <i>Zhe Jiang, Shashi Shekhar, Xun Zhou, Joseph Knight, and Jennifer Corcoran</i>  |     |
| Conformal Prediction Using Decision Trees .....   | 330 |
| <i>Ulf Johansson, Henrik Boström, and Tuve Löfström</i>   |     |
| An Unsupervised Algorithm for Learning Blocking Schemes .....   | 340 |
| <i>Mayank Kejriwal and Daniel P. Miranker</i>   |     |
| Semantic Frame-Based Document Representation for Comparable Corpora .....   | 350 |
| <i>Hyungsul Kim, Xiang Ren, Yizhou Sun, Chi Wang, and Jiawei Han</i>  |     |
| Regularization Paths for Sparse Nonnegative Least Squares Problems<br>with Applications to Life Cycle Assessment Tree Discovery ..... | 360 |
| <i>Jingu Kim, Naren Ramakrishnan, Manish Marwah, Amip Shah, and Haesun Park</i>   |     |
| A High-Dimensional Set Top Box Ad Targeting Algorithm Including Experimental<br>Comparisons to Traditional TV Algorithms .....        | 370 |
| <i>Brendan Kitts, Dyng Au, and Brian Burdick</i>  |     |
| Efficient Algorithms for Selecting Features with Arbitrary Group Constraints<br>via Group Lasso .....                                 | 379 |
| <i>Deguang Kong and Chris Ding</i>  |     |
| BIG-ALIGN: Fast Bipartite Graph Alignment .....   | 389 |
| <i>Danai Koutra, Hanghang Tong, and David Lubensky</i>  |     |
| Blocking Simple and Complex Contagion by Edge Removal .....   | 399 |
| <i>Chris J. Kuhlman, Gaurav Tuli, Samarth Swarup, Madhav V. Marathe, and S.S. Ravi</i>  |     |
| Guiding Autonomous Agents to Better Behaviors through Human Advice .....  | 409 |
| <i>Gautam Kunapuli, Phillip Odom, Jude W. Shavlik, and Sriraam Natarajan</i>  |     |

|  |     |
|--|-----|
| TL-PLSA: Transfer Learning between Domains with Different Classes .....  | 419 |
| <i>Anastasia Krithara and Georgios Paliouras</i>   |     |
| Learning, Analyzing and Predicting Object Roles on Dynamic Networks .....  | 428 |
| <i>Kang Li, Suxin Guo, Nan Du, Jing Gao, and Aidong Zhang</i>  |     |
| Tag-Weighted Dirichlet Allocation .....  | 438 |
| <i>Shuangyin Li, Guan Huang, Ruiyang Tan, and Rong Pan</i>   |     |
| Mining Probabilistic Frequent Spatio-Temporal Sequential Patterns with Gap<br>Constraints from Uncertain Databases ..... | 448 |
| <i>Yuxuan Li, James Bailey, Lars Kulik, and Jian Pei</i>   |     |
| Mining Following Relationships in Movement Data .....  | 458 |
| <i>Zhenhui Li, Fei Wu, and Margaret C. Crofoot</i>   |     |
| Linear Computation for Independent Social Influence .....  | 468 |
| <i>Qi Liu, Biao Xiang, Lei Zhang, Enhong Chen, Chang Tan, and Ji Chen</i>  |     |
| Learning Imbalanced Multi-class Data with Optimal Dichotomy Weights .....  | 478 |
| <i>Xu-Ying Liu, Qian-Qian Li, and Zhi-Hua Zhou</i>   |     |
| Mining Statistically Significant Sequential Patterns .....   | 488 |
| <i>Cécile Low-Kam, Chedy Raïssi, Mehdi Kaytoue, and Jian Pei</i>   |     |
| Mining Summaries of Propagations .....   | 498 |
| <i>Lucrezia Macchia, Francesco Bonchi, Francesco Gullo, and Luca Chiarandini</i>   |     |
| Active Density-Based Clustering .....  | 508 |
| <i>Son T. Mai, Xiao He, Nina Hubig, Claudia Plant, and Christian Böhm</i>  |     |
| Explaining Outliers by Subspace Separability .....   | 518 |
| <i>Barbora Micenková, Raymond T. Ng, Xuan-Hong Dang, and Ira Assent</i>  |     |
| Quantification Trees .....   | 528 |
| <i>Letizia Milli, Anna Monreale, Giulio Rossetti, Fosca Giannotti, Dino Pedreschi,<br/>and Fabrizio Sebastiani</i>       |     |
| Mining Evolving Network Processes .....  | 537 |
| <i>Misael Mongiovì, Petko Bogdanov, and Ambuj K. Singh</i>   |     |
| Enumeration of Time Series Motifs of All Lengths .....   | 547 |
| <i>Abdullah Mueen</i>  |     |
| Dominance Programming for Itemset Mining .....   | 557 |
| <i>Benjamin Negrevergne, Anton Dries, Tias Guns, and Siegfried Nijssen</i>   |     |
| Least Cost Influence in Multiplex Social Networks: Model Representation<br>and Analysis .....                            | 567 |
| <i>Dung T. Nguyen, Huiyuan Zhang, Soham Das, My T. Thai, and Thang N. Dinh</i>   |     |
| Bayesian Discovery of Multiple Bayesian Networks via Transfer Learning .....   | 577 |
| <i>Diane Oyen and Terran Lane</i>  |     |

|   |     |
|---|-----|
| Forecasting Spatiotemporal Impact of Traffic Incidents on Road Networks .....   | 587 |
| <i>Bei Pan, Ugur Demiryurek, Cyrus Shahabi, and Chetan Gupta</i>  |     |
| Scaling Log-Linear Analysis to High-Dimensional Data .....  | 597 |
| <i>Francois Petitjean, Geoffrey I. Webb, and Ann E. Nicholson</i>   |     |
| Fast Pairwise Query Selection for Large-Scale Active Learning to Rank .....   | 607 |
| <i>Buyue Qian, Xiang Wang, Jun Wang, Hongfei Li, Nan Cao, Weifeng Zhi, and Ian Davidson</i>                                 |     |
| Power to the Points: Validating Data Memberships in Clusterings .....   | 617 |
| <i>Parasaran Raman and Suresh Venkatasubramanian</i>  |     |
| Weighted-Object Ensemble Clustering .....   | 627 |
| <i>Yazhou Ren, Carlotta Domeniconi, Guoji Zhang, and Guoxian Yu</i>   |     |
| Mining User Lifecycles from Online Community Platforms and their Application<br>to Churn Prediction .....                   | 637 |
| <i>Matthew Rowe</i>   |     |
| Statistical Selection of Congruent Subspaces for Mining Attributed Graphs .....   | 647 |
| <i>Patricia Iglesias Sánchez, Emmanuel Müller, Fabian Laforet, Fabian Keller,<br/>and Klemens Böhm</i>                      |     |
| Classifying Spam Emails Using Text and Readability Features .....   | 657 |
| <i>Rushdi Shams and Robert E. Mercer</i>  |     |
| Most-Surely vs. Least-Surely Uncertain .....  | 667 |
| <i>Manali Sharma and Mustafa Bilgic</i>   |     |
| On Pattern Preserving Graph Generation .....  | 677 |
| <i>Hong-Han Shuai, De-Nian Yang, Philip S. Yu, Chih-Ya Shen, and Ming-Syan Chen</i>   |     |
| Time Series Classification Using Compression Distance of Recurrence Plots .....   | 687 |
| <i>Diego F. Silva, Vinicius M.A. De Souza, and Gustavo E.A.P.A. Batista</i>   |     |
| Dynamic Pattern Detection with Temporal Consistency and Connectivity Constraints .....                                      | 697 |
| <i>Skyler Speakman, Yating Zhang, and Daniel B. Neill</i>   |     |
| Noise-Resistant Bicluster Recognition .....   | 707 |
| <i>Huan Sun, Gengxin Miao, and Xifeng Yan</i>   |     |
| Itemsets for Real-Valued Datasets .....   | 717 |
| <i>Nikolaj Tatti</i>  |     |
| Collective Response Spike Prediction for Mutually Interacting Consumers .....   | 727 |
| <i>Rikiya Takahashi, Hideyuki Mizuta, Naoki Abe, Ruby L. Kennedy, Vincent J. Jeffs,<br/>Ravi Shah, and Robert H. Crites</i> |     |
| Price Information Patterns in Web Search Advertising: An Empirical Case Study<br>on Accommodation Industry .....            | 737 |
| <i>Guanting Tang, Yupin Yang, and Jian Pei</i>  |     |
| Discovering Non-redundant Overlapping Biclusters on Gene Expression Data .....  | 747 |
| <i>Duy Tin Truong, Roberto Battiti, and Mauro Brunato</i>   |     |

|   |     |
|---|-----|
| Cox Regression with Correlation Based Regularization for Electronic Health Records .....                      | 757 |
| <i>Bhanukiran Vinzamuri and Chandan K. Reddy</i>  |     |
| Constructing Topical Hierarchies in Heterogeneous Information Networks .....                                  | 767 |
| <i>Chi Wang, Marina Danilevsky, Jialu Liu, Nihit Desai, Heng Ji, and Jiawei Han</i>                           |     |
| Binary Time-Series Query Framework for Efficient Quantitative Trait Association<br>Study .....                | 777 |
| <i>Hongfei Wang and Xiang Zhang</i>   |     |
| Communication-Efficient Distributed Multiple Reference Pattern Matching for M2M<br>Systems .....              | 787 |
| <i>Jui-Pin Wang, Yu-Chen Lu, Mi-Yen Yeh, Shou-De Lin, and Phillip B. Gibbons</i>                              |     |
| Sparse K-Means with the $L_{1-q}$ ( $0 \leq q < 1$ ) Constraint for High-Dimensional Data<br>Clustering ..... | 797 |
| <i>Yu Wang, Xiangyu Chang, Rongjian Li, and Zongben Xu</i>  |     |
| A Comparative Analysis of Ensemble Classifiers: Case Studies in Genomics .....                                | 807 |
| <i>Sean Whalen and Gaurav Pandey</i>  |     |
| Stochastic Blockmodel with Cluster Overlap, Relevance Selection,<br>and Similarity-Based Smoothing .....      | 817 |
| <i>Joyce Jiyoungh Whang, Piyush Rai, and Inderjit S. Dhillon</i>  |     |
| Multi-instance Multi-graph Dual Embedding Learning .....  | 827 |
| <i>Jia Wu, Xingquan Zhu, Chengqi Zhang, and Zhihua Cai</i>  |     |
| TopicSketch: Real-Time Bursty Topic Detection from Twitter .....  | 837 |
| <i>Wei Xie, Feida Zhu, Jing Jiang, Ee-Peng Lim, and Ke Wang</i>   |     |
| Efficient Learning on Point Sets .....  | 847 |
| <i>Liang Xiong, Barnabás Póczos, and Jeff Schneider</i>   |     |
| Markov Blanket Feature Selection with Non-faithful Data Distributions .....                                   | 857 |
| <i>Kui Yu, Xindong Wu, Zan Zhang, Yang Mu, Hao Wang, and Wei Ding</i>   |     |
| An Efficient Approach to Updating Closeness Centrality and Average Path Length<br>in Dynamic Networks .....   | 867 |
| <i>Chia-Chen Yen, Mi-Yen Yeh, and Ming-Syan Chen</i>  |     |
| Reconstructing Individual Mobility from Smart Card Transactions: A Space<br>Alignment Approach .....          | 877 |
| <i>Nicholas Jing Yuan, Yingzi Wang, Fuzheng Zhang, Xing Xie, and Guangzhong Sun</i>                           |     |
| Feature Transformation with Class Conditional Decorrelation .....   | 887 |
| <i>Xu-Yao Zhang, Kaizhu Huang, and Cheng-Lin Liu</i>  |     |
| Efficient Learning for Models with DAG-Structured Parameter Constraints .....                                 | 897 |
| <i>Leon Wenliang Zhong and James T. Kwok</i>  |     |
| UBLF: An Upper Bound Based Approach to Discover Influential Nodes in Social<br>Networks .....                 | 907 |
| <i>Chuan Zhou, Peng Zhang, Jing Guo, Xingquan Zhu, and Li Guo</i>   |     |

|  |      |
|--|------|
| Divide-and-Conquer Anchoring for Near-Separable Nonnegative Matrix Factorization and Completion in High Dimensions ..... | 917  |
| <i>Tianyi Zhou, Wei Bian, and Dacheng Tao</i>  |      |
| A Novel Relational Learning-to-Rank Approach for Topic-Focused Multi-document Summarization .....                        | 927  |
| <i>Yadong Zhu, Yanyan Lan, Jiafeng Guo, Pan Du, and Xueqi Cheng</i>  |      |
| <b>Short Papers</b>  |      |
| Cartification: A Neighborhood Preserving Transformation for Mining High Dimensional Data .....                           | 937  |
| <i>Emin Aksehirli, Bart Goethals, Emmanuel Müller, and Jilles Vreeken</i>  |      |
| Improved Electricity Load Forecasting via Kernel Spectral Clustering of Smart Meters .....                               | 943  |
| <i>Carlos Alzate and Mathieu Sinn</i>  |      |
| Large-Scale Elastic Net Regularized Linear Classification SVMs and Logistic Regression .....                             | 949  |
| <i>P. Balamurugan</i>  |      |
| Influence-Based Network-Oblivious Community Detection .....  | 955  |
| <i>Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco</i>   |      |
| Beyond Boolean Matrix Decompositions: Toward Factor Analysis and Dimensionality Reduction of Ordinal Data .....          | 961  |
| <i>Radim Belohlavek and Marketa Krmelova</i>   |      |
| Validating Network Value of Influencers by Means of Explanations .....   | 967  |
| <i>Glenn S. Bevilacqua, Shealen Clare, Amit Goyal, and Laks V.S. Lakshmanan</i>  |      |
| External Evaluation of Topic Models: A Graph Mining Approach .....   | 973  |
| <i>Hau Chan and Leman Akoglu</i>   |      |
| Multimedia LEGO: Learning Structured Model by Probabilistic Logic Ontology Tree .....                                    | 979  |
| <i>Shiyu Chang, Guo-Jun Qi, Jinhui Tang, Qi Tian, Yong Rui, and Thomas S. Huang</i>                                      |      |
| Discriminatively Enhanced Topic Models .....   | 985  |
| <i>Snigdha Chaturvedi, Hal Daumé III, and Taesun Moon</i>  |      |
| PerturBoost: Practical Confidential Classifier Learning in the Cloud .....   | 991  |
| <i>Keke Chen and Shumin Guo</i>  |      |
| A Feature-Enhanced Ranking-Based Classifier for Multimodal Data and Heterogeneous Information Networks .....             | 997  |
| <i>Scott Deeann Chen, Ying-Yu Chen, Jiawei Han, and Pierre Moulin</i>  |      |
| Nonlinear Causal Discovery for High Dimensional Data: A Kernelized Trace Method .....                                    | 1003 |
| <i>Zhitang Chen, Kun Zhang, and Laiwan Chan</i>  |      |
| Discriminative Link Prediction Using Local Links, Node Features and Community Structure .....                            | 1009 |
| <i>Abir De, Niloy Ganguly, and Soumen Chakrabarti</i>  |      |



|   |      |
|---|------|
| Most Clusters Can Be Retrieved with Short Disjunctive Queries .....   | 1019 |
| <i>Vinay Deolalikar</i>   |      |
| Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic<br>Guarantee .....  | 1025 |
| <i>Boxiang Dong, Ruilin Liu, and Wendy Hui Wang</i>   |      |
| Progression Analysis of Community Strengths in Dynamic Networks .....   | 1031 |
| <i>Nan Du, Jing Gao, and Aidong Zhang</i>   |      |
| Walk 'n' Merge: A Scalable Algorithm for Boolean Tensor Factorization .....   | 1037 |
| <i>Dóra Erdős and Pauli Miettinen</i>   |      |
| Online Active Learning with Imbalanced Classes .....  | 1043 |
| <i>Zahra Ferdowsi, Rayid Ghani, and Raffaella Settini</i>   |      |
| Efficient Invariant Search for Distributed Information Systems .....  | 1049 |
| <i>Yong Ge, Guofei Jiang, and Yuan Ge</i>   |      |
| Coupled Heterogeneous Association Rule Mining (CHARM): Application Toward<br>Inference of Modulatory Climate Relationships .....  | 1055 |
| <i>Doel L. Gonzalez II, Saurabh V. Pendse, Kanchana Padmanabhan, Michael P. Angus,<br/>Isaac K. Tetteh, Shashank Srinivas, Andrea Villanes, Fredrick Semazzi, Vipin Kumar,<br/>and Nagiza F. Samatova</i> |      |
| Leveraging Supervised Label Dependency Propagation for Multi-label Learning .....   | 1061 |
| <i>Bin Fu, Guandong Xu, Zhihai Wang, and Longbing Cao</i>   |      |
| A Mobility Simulation Framework Of Humans With Group Behavior Modeling .....  | 1067 |
| <i>Anshul Gupta, Aurosish Mishra, S.G. Vadlamudi, P.P. Chakrabarti, Sudeshna Sarkar,<br/>Tridib Mukherjee, and Nathan Gnanasambandam</i>  |      |
| Spatio-Temporal Topic Modeling in Mobile Social Media for Location<br>Recommendation .....  | 1073 |
| <i>Bo Hu, Mohsen Jamali, and Martin Ester</i>   |      |
| Active Query Driven by Uncertainty and Diversity for Incremental Multi-label<br>Learning .....  | 1079 |
| <i>Sheng-Jun Huang and Zhi-Hua Zhou</i>   |      |
| Accelerating Active Learning with Transfer Learning .....   | 1085 |
| <i>David Kale and Yan Liu</i>   |      |
| Statistical Inference of Protein "LEGO Bricks" .....  | 1091 |
| <i>Arun S. Konagurthu, Lloyd Allison, David Abramson, Peter J. Stuckey, and Arthur M. Lesk</i>  |      |
| A Probabilistic Behavior Model for Discovering Unrecognized Knowledge .....   | 1097 |
| <i>Takeshi Kurashima, Tomoharu Iwata, Noriko Takaya, and Hiroshi Sawada</i>   |      |
| Prominent Features of Rumor Propagation in Online Social Media .....  | 1103 |
| <i>Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang</i>  |      |
| Group Feature Selection with Streaming Features .....   | 1109 |
| <i>Haiguang Li, Xindong Wu, Zhao Li, and Wei Ding</i>   |      |

|  |      |
|--|------|
| Bayesian Multi-Task Relationship Learning with Link Structure .....  | 1115 |
| <i>Yingming Li, Ming Yang, Zhongang Qi, and Zhongfei (Mark) Zhang</i>  |      |
| Quantitative Prediction of Glaucomatous Visual Field Loss from Few Measurements .....  | 1121 |
| <i>Zenghan Liang, Ryota Tomioka, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi</i>   |      |
| On the Feature Discovery for App Usage Prediction in Smartphones .....   | 1127 |
| <i>Zhung-Xun Liao, Shou-Chung Li, Wen-Chih Peng, Philip S. Yu, and Te-Chuan Liu</i>  |      |
| How Many Zombies Around You? .....   | 1133 |
| <i>Hongfu Liu, Yuchao Zhang, Hao Lin, Junjie Wu, Zhiang Wu, and Xu Zhang</i>   |      |
| Hibernating Process: Modelling Mobile Calls at Multiple Scales .....   | 1139 |
| <i>Siyuan Liu, Lei Li, and Rammaya Krishnan</i>  |      |
| On Good and Fair Paper-Reviewer Assignment .....   | 1145 |
| <i>Cheng Long, Raymond Chi-Wing Wong, Yu Peng, and Liangliang Ye</i>   |      |
| Community Detection in Networks with Node Attributes .....   | 1151 |
| <i>Jaewon Yang, Julian McAuley, and Jure Leskovec</i>  |      |
| Multitask Learning with Feature Selection for Groups of Related Tasks .....  | 1157 |
| <i>Meenakshi Mishra and Jun Huan</i>   |      |
| Network Hypothesis Testing Using Mixed Kronecker Product Graph Models .....  | 1163 |
| <i>Sebastian Moreno and Jennifer Neville</i>   |      |
| Graph Partitioning Change Detection Using Tree-Based Clustering .....  | 1169 |
| <i>Sho-Ichi Sato and Kenji Yamanishi</i>   |      |
| SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space<br>Model .....  | 1175 |
| <i>Pavel Senin and Sergey Malinchik</i>  |      |
| Clustering on Multiple Incomplete Datasets via Collective Kernel Learning .....  | 1181 |
| <i>Weixiang Shao, Xiaoxiao Shi, and Philip S. Yu</i>   |      |
| MLI: An API for Distributed Machine Learning .....   | 1187 |
| <i>Evan R. Sparks, Ameet Talwalkar, Virginia Smith, Jey Kottalam, Xinghao Pan,<br/>Joseph Gonzalez, Michael J. Franklin, Michael I. Jordan, and Tim Kraska</i> |      |
| Co-ClusterD: A Distributed Framework for Data Co-Clustering with Sequential<br>Updates .....   | 1193 |
| <i>Sen Su, Xiang Cheng, Lixin Gao, and Jiangtao Yin</i>  |      |
| Non-negative Multiple Tensor Factorization .....   | 1199 |
| <i>Koh Takeuchi, Ryota Tomioka, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada</i>   |      |
| Multiclass Semi-Supervised Boosting Using Similarity Learning .....  | 1205 |
| <i>Jafar Tanha, Mohammad Javad Saberian, and Maarten van Someren</i>   |      |
| Mining Dependent Frequent Serial Episodes from Uncertain Sequence Data .....   | 1211 |
| <i>Li Wan, Ling Chen, and Chengqi Zhang</i>  |      |

|  |      |
|--|------|
| Efficient and Scalable Information Geometry Metric Learning .....  | 1217 |
| <i>Wei Wang, Bao-Gang Hu, and Zengfu Wang</i>  |      |
| Exploring Patient Risk Groups with Incomplete Knowledge .....  | 1223 |
| <i>Xiang Wang, Fei Wang, Jun Wang, Buyue Qian, and Jianying Hu</i>   |      |
| Classification-Based Clustering Evaluation .....   | 1229 |
| <i>John S. Whissell and Charles L.A. Clarke</i>  |      |
| Efficient Online Sequence Prediction with Side Information .....   | 1235 |
| <i>Han Xiao and Claudia Eckert</i>   |      |
| Multilabel Consensus Classification .....  | 1241 |
| <i>Sihong Xie, Xiangnan Kong, Jing Gao, Wei Fan, and Philip S. Yu</i>  |      |
| Sampling Heterogeneous Networks .....  | 1247 |
| <i>Cheng-Lun Yang, Perng-Hwa Kung, Cheng-Te Li, Chun-An Chen, and Shou-De Lin</i>                                      |      |
| A Model for Discovering Correlations of Ubiquitous Things .....  | 1253 |
| <i>Lina Yao, Quan Z. Sheng, Byron J. Gao, Anne H.H. Ngu, and Xue Li</i>  |      |
| Efficiently Mining Top-K High Utility Sequential Patterns .....  | 1259 |
| <i>Junfu Yin, Zhigang Zheng, Longbing Cao, Yin Song, and Wei Wei</i>   |      |
| Efficient Proper Length Time Series Motif Discovery .....  | 1265 |
| <i>Sorrachai Yingchareonthawornchai, Haemwaan Sivaraks, Thanawin Rakthanmanon,<br/>and Chotirat Ann Ratanamahatana</i> |      |
| On Anomalous Hotspot Discovery in Graph Streams .....  | 1271 |
| <i>Weiren Yu, Charu C. Aggarwal, Shuai Ma, and Haixun Wang</i>   |      |
| From Social User Activities to People Affiliation .....  | 1277 |
| <i>Guangxiang Zeng, Ping Luo, Enhong Chen, and Min Wang</i>  |      |
| Transfer Learning across Cancers on DNA Copy Number Variation Analysis .....   | 1283 |
| <i>Huanan Zhang, Ze Tian, and Rui Kuang</i>  |      |
| Predicting Social Links for New Users across Aligned Heterogeneous Social Networks .....                               | 1289 |
| <i>Jiawei Zhang, Xiangnan Kong, and Philip S. Yu</i>   |      |
| Combating Sub-Clusters Effect in Imbalanced Classification .....   | 1295 |
| <i>Yang Zhao and Abhishek K. Shrivastava</i>   |      |
| Influence and Profit: Two Sides of the Coin .....  | 1301 |
| <i>Yuqing Zhu, Zaixin Lu, Yuanjun Bi, Weili Wu, Yiwei Jiang, and Deying Li</i>   |      |
| Constrained Clustering: Effective Constraint Propagation with Imperfect Oracles .....                                  | 1307 |
| <i>Xiatian Zhu, Chen Change Loy, and Shaogang Gong</i>   |      |
| Influence Maximization in Dynamic Social Networks .....  | 1313 |
| <i>Honglei Zhuang, Yihan Sun, Jie Tang, Jialin Zhang, and Xiaoming Sun</i>   |      |
| Adaptive Model Tree for Streaming Data .....   | 1319 |
| <i>Anca M. Zimmer, Michael Kurze, and Thomas Seidl</i>   |      |

Structural-Context Similarities for Uncertain Graphs .....1325  
*Zhaonian Zou and Jianzhong Li*

**Author Index**