## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agenda at a Glance</td>
<td>5</td>
</tr>
<tr>
<td>Message from the ICDM 2021 General Chairs</td>
<td>7</td>
</tr>
<tr>
<td>Message from the ICDM 2021 Program Chairs</td>
<td>8</td>
</tr>
<tr>
<td>Message from the Workshop Chairs</td>
<td>9</td>
</tr>
<tr>
<td>Diversity, Equity, and Inclusiveness Statement</td>
<td>10</td>
</tr>
<tr>
<td>Program Highlights</td>
<td>11</td>
</tr>
<tr>
<td>Conference Program</td>
<td>13</td>
</tr>
<tr>
<td>Tutorial Program</td>
<td>25</td>
</tr>
<tr>
<td>PhD Forum Program</td>
<td>26</td>
</tr>
<tr>
<td>Keynote Speakers</td>
<td>27</td>
</tr>
<tr>
<td>Panel</td>
<td>31</td>
</tr>
<tr>
<td>Special Sessions</td>
<td>33</td>
</tr>
<tr>
<td>Job Matching Program Highlights</td>
<td>35</td>
</tr>
<tr>
<td>Virtual Social Event Highlights</td>
<td>35</td>
</tr>
<tr>
<td>Virtual Social Event Program</td>
<td>35</td>
</tr>
<tr>
<td>Awards</td>
<td>35</td>
</tr>
<tr>
<td>Organising Committee</td>
<td>37</td>
</tr>
<tr>
<td>Conference Sponsors</td>
<td>41</td>
</tr>
</tbody>
</table>
# Agenda at a glance

## Tuesday, December 7 (Pre-conference day)
- **14:00-21:00** Workshop 01 - NeuRec: Advanced Neural Algorithms and Theories for Recommender Systems
- **14:00-21:00** Workshop 02 - SENTIRE: Sentiment Elicitation from Natural Text for Information Retrieval and Extraction
- **14:00-21:00** Workshop 03 - DMIS: Data Mining for Service
- **14:00-21:00** Workshop 04 - IncrLearn: Incremental classification and clustering, concept drift, novelty detection in big/fast data context
- **14:00-21:00** Workshop 05 - STDTM: 16th International Workshop on Spatial and Spatiotemporal Data Mining and (WAIN)/STSTD: 1st Workshop on AI for Nudging and Personalization
- **14:00-21:00** Workshop 06 - IAAA: International Workshop on Intelligence-Augmented Anomaly Analytics
- **14:00-18:30** PhD Panel
- **14:00-17:00** Workshop 07 - UDML: Utility Driven Mining and Learning
- **14:00-17:00** Workshop 08 - SFE-TSDM: Systematic Feature Engineering for Time-Series Data Mining
- **14:00-17:00** Workshop 09 - DUC: Deep Learning and Clustering
- **14:00-17:00** Workshop 10 - OEDML: Optimization Based Techniques for Emerging Data Mining Problems
- **14:00-17:00** Workshop 11 - EDMML: Evolutionary Data Mining and Machine Learning
- **14:00-17:00** Workshop 12 - LTSA: 2nd Workshop on Large-scale Industrial Time Series Analysis
- **14:00-17:00** Workshop 13 - HDM: High Dimensional Data Mining
- **14:00-17:00** Workshop 14 - DMBIH: Data Mining in Biomedical Informatics and Healthcare
- **14:00-17:00** Workshop 15 - MLLD: Mining and Learning in the Legal Domain
- **15:00-18:00** Steering Committee Meeting (by invitation only)
- **16:30-17:30** Networking (Breakout rooms and speed networking)
- **17:30-20:00** Workshop 16 - CLEATED: Continual Learning and Adaptation for Time Evolving Data
- **17:30-20:00** Workshop 17 - SDML: Social Data Mining in the Post-pandemic Era
- **17:30-20:00** Workshop 18 - DMC: Data Mining and Machine Learning in Cybersecurity

## Wednesday, December 8
- **14:00-15:00** Opening Ceremony
- **15:00-16:00** Keynote: Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond
- **16:00-18:30** Tutorial O1: Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems
- **16:00-18:30** Tutorial O2: Complement, Composite and Context: The 3C-Law to Build Multidomain Recommender Systems
- **16:00-17:30** Session 1A: Data Mining Foundations and Frontiers (Stream A)
- **16:00-17:30** Session 1B: Graph Mining and Learning (Stream B)
- **16:00-17:30** Session 1C: Sequences and Time Series 1 (Stream C)
- **16:00-17:30** Session 1D: Text 1 (Stream D)
- **16:00-17:30** Session 1E: Traffic Forecasting (Stream E)
- **17:30-18:00** Networking (Breakout rooms and speed networking)
- **18:00-19:30** Session 2A: Clustering (Stream A)
- **18:00-19:30** Session 2B: Algorithmic Methods (Stream B)
- **18:00-19:30** Session 2C: Sequences and Time Series 2 (Stream C)
- **18:00-19:30** Session 2D: Text 2 (Stream D)
- **18:00-19:30** Session 2E: Applications (Stream E)
- **18:30-21:00** Tutorial O3: Roles in Networks - Foundations, Methods and Applications
- **19:30-20:00** Networking (Breakout rooms and speed networking)
- **20:00-21:00** Job Matching Networking Session

## Thursday, December 9
- **14:00-15:00** Keynote: Sample Efficient AI with Applications in Health Care and Advanced Manufacturing
- **15:00-15:30** Networking
- **15:30-20:30** Tutorial 04: Automated Taxonomy Discovery and Exploration
- **15:30-17:00** Session 3A: Graph Learning (Stream A)
- **15:30-17:00** Session 3B: Fairness and Interpretability (Stream B)
- **15:30-17:00** Session 3C: Sequences and Time Series 3 (Stream C)
- **15:30-17:00** Session 3D: Deep Learning 1 (Stream D)
- **15:30-17:00** Session 3E: Biological and Physical Systems (Stream E)
- **15:30-17:00** Session 3F: Representation Learning (Stream F)
- **17:00-17:30** Networking (Breakout rooms and speed networking)
- **17:30-19:00** Session 4A: Networks (Stream A)
- **17:30-19:00** Session 4B: Causality and Fairness (Stream B)
- **17:30-19:00** Session 4C: Theory and Optimization (Stream C)
- **17:30-19:00** Session 4D: Deep Learning 2 (Stream D)
- **17:30-19:00** Session 4E: Bandits and Online Methods (Stream E)
- **17:30-19:00** Session 4F: Meta Learning (Stream F)
- **19:00-19:30** Networking (Breakout rooms and speed networking)
- **19:30-20:30** Panel Discussion: Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning

## Friday, December 10
- **14:00-15:00** Keynote: Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces
- **15:00-15:30** Networking
- **15:30-17:00** Session 5A: Crowdsourcing (Stream A)
- **15:30-17:00** Session 5B: Explainability (Stream B)
- **15:30-17:00** Session 5C: Graph Neural Networks 1 (Stream C)
- **15:30-17:00** Session 5D: Statistical Techniques (Stream D)
- **15:30-17:00** Session 5E: Clinical Data Analysis (Stream E)
- **15:30-17:00** Session 5F: Reinforcement Learning (Stream F)
- **15:30-17:00** Special Session in Diversity: Women in Data Science and Machine Learning
- **17:00-17:30** Networking (Breakout rooms and speed networking)
- **17:30-19:00** Session 6A: Recommender Systems (Stream A)
- **17:30-19:00** Session 6B: Adversarial Learning (Stream B)
- **17:30-19:00** Session 6C: Graph Neural Networks 2 (Stream C)
- **17:30-19:00** Session 6D: Active and Semi-Supervised Learning (Stream D)
- **17:30-19:00** Session 6E: Matrix and Tensor (Stream E)
- **19:00-19:30** Networking (Breakout rooms and speed networking)
- **19:30-20:30** Awards Ceremony
- **20:30-21:30** Closing

All times are in New Zealand time (GMT+13)
Message from the ICDM 2021 General Chairs

On behalf of the organizing committee of the IEEE ICDM 2021 conference and our virtual host Auckland, New Zealand, it is our great pleasure to welcome you to the 2021 IEEE International Conference on Data Mining. Due to the COVID-19 global pandemic, IEEE ICDM 2021 will be hosted virtually for the second time.

Our goal is to run the conference in a way that replicates an in-person conference as best as we can, while leveraging events that can only happen virtually. We have live keynotes, tutorials, and special sessions. We also have speed-networking sessions, online activities, and are running a job matching program for the first time. Additionally, this is the first time there have been Reproducibility Chairs, and a Diversity and Inclusion Chair at ICDM. We hope to ultimately encapsulate this IEEE ICDM 2021 conference with our unique New Zealand hospitality.

The organization of a successful conference would not be possible without the dedicated efforts from many individuals. In particular, we would like to express our gratitude to the Program Chairs Pauli Miettinen, University of Eastern Finland (Finland); James Bailey, The University of Melbourne (Australia); Steering Committee Chair Xindong Wu, MiningLamp Academy of Sciences (China); Workshop Chairs Bing Xue, Victoria University of Wellington (New Zealand); Mykola Pechenizkiy, Eindhoven University of Technology (The Netherlands); Tutorial Chairs Wei Liu, University of Technology Sydney (Australia); Katharina Morik, TU Dortmund (Germany); Online Experience/Virtual Chair Heitor Gomes, University of Waikato (New Zealand); Publicity Chair Diana Benvides Prado, The University of Auckland (New Zealand); Finance Chair Gillian Dobbe, The University of Auckland (New Zealand); Sponsorship Chairs Kaiqi Zhao, The University of Auckland (New Zealand); Philippe Fournier-Viger, Harbin Institute of Technology (China); Eva Garcia- Martin, Ekkono Solutions (Sweden); PhD Forum Chairs Sibylle Hess, Eindhoven University of Technology (The Netherlands); Meng-Fen Chang, The University of Auckland (New Zealand); Lisi Chen, Inception Institute of Artificial Intelligence (IIAI) (UAE); Diversity and Inclusion Chair, Richi Nayak, Queensland University of Technology (Australia); Job Matching Chairs, Albert Bifet, University of Waikato (New Zealand); Lin Liu, University of South Australia (Australia); Panel Chairs Chenliang Li, Wuhan University (China); Michael Witbrock, The University of Auckland (New Zealand); Newcomers Chairs Huan Liu, Arizona State University (USA); Katerina Taskova, The University of Auckland (New Zealand); Local Arrangement Chair Joerg Wicker, The University of Auckland (New Zealand); Reproducibility Chairs Dragi Kocić, Jožef Stefan Institute (Slovenia); Jacob Montiel, University of Waikato (New Zealand); Award Committee Chair Xia Ning, The Ohio State University (USA).

We owe special thanks to our sponsors of the conference, including the US National Science Foundation (NSF), School of Computer Science - The University of Auckland, IEEE Technical Committee on Intelligent Informatics and Google, and Tiwo Sigma. We thank the authors, keynote speakers, special session speakers, panelists, and tutorial speakers for agreeing to present their sessions virtually. We also thank the workshop organizers for running their workshops virtually.

Finally, we thank all researchers, practitioners and students who are working in the field of data mining for their support and promotion of ICDM over the years. We wish you a productive conference with new discoveries, new collaborations and a very enjoyable virtual experience.

Yun Sing Koh and Dacheng Tao
IEEE ICDM 2021 General Co-Chairs

Message from the ICDM 2021 Program Chairs

Since its inception in 2001, the IEEE International Conference on Data Mining (ICDM) has become a premier forum for researchers, users, practitioners, and developers to exchange and disseminate not only original research results but also new research directions in data mining. The 21st IEEE ICDM is being hosted this year in Auckland, New Zealand and run as an entirely virtual conference, due to the COVID-19 pandemic.

It is our great pleasure to welcome you to ICDM 2021 and to present its proceedings to you. The ICDM conference is truly an international forum. During its nineteen-year history, the conference has been held in ten countries around the world. This year’s conference continues this global trend: Our organizing and program committee members represent around 36 countries/regions. This year’s conference was extremely competitive. We are pleased to announce 990 paper submissions from 46 different countries/regions for review. Best efforts were made to ensure each paper was reviewed by at least three program committee members and the selection was made on the basis of discussion among the reviewers, an area chair, and the program co-chairs. Like previous years, we implemented a triple blind review process, ensuring that the reviewers do not know the identity of the authors or of the other individuals reviewing the same submission. This process is intended to remove bias during the paper discussions. This year, 98 regular papers (9.9% acceptance rate) and 100 short papers were selected for inclusion in the proceedings and program (giving a total acceptance rate of 20%). Of the papers that were submitted, 573 (57.9%) had student first authors. These authors represent the future of our field.

In keeping with the goal of advancing the state-of-the-art in data mining, paper topics span a range of areas including: novel data mining algorithms in traditional areas, models and algorithms for new, structured, data types; deep learning and its applications; mining sequences and sequential data; mining spatial and temporal datasets; mining textual and unstructured datasets; high performance implementations of data mining algorithms; stream data mining; mining and link analysis in networked settings; data mining in electronic commerce (e.g., recommendation); web search, advertising, and marketing tasks; methodological aspects and the KDD process; and healthcare, epidemic modeling, and clinical research.

In addition to the technical presentations, our program also highlights three outstanding keynotes given by internationally renowned, distinguished scientists Jian Pei (Simon Fraser University), Massaf Sugiymama (RIKEN Center for Advanced Intelligence Project/University of Tokyo) and Svetla Venkatesh (Deakin University). Four tutorials will be offered and 18 workshops will be run in conjunction with the main conference. A job matching program will also be run as part of the conference. We would like to thank all those who invested their substantial efforts into making this conference what it is, starting with all the authors of the 990 manuscripts for submitting content for the conference.

Reviewing and selecting papers from a large set of submissions required the coordinated effort of many individuals. We want to thank the 55 Area Chairs and 398 Program Committee members who provided insightful feedback to the authors and helped with this selection process.

Organizing the ICDM 2021 program required the time and expertise of numerous contributors. We are very thankful for the outstanding work of Bing Xue and Mykola Pechenizkiy who served as Workshop Co-Chairs, Wei Liu and Katharina Monk who were the Tutorial Co-Chairs, Sibylle Hess and Meng-Fen Chiang who organized the PhD Forum, Chenliang Li and Michael Witbrock who were the Panel Co-Chairs, Diana Benvides Prado who served as Publicity Chair, sponsorship chairs Kaqi Zhao, Philippe Fournier-Viger and Eva Garcia Martin, newcomers chairs Huan Liu and Katerina Taskova, job matching chairs Albert Bifet and Lin Liu, reproducibility chairs Dragi Kocić and Jacob Montiel, Xia Ning who chaired the Best Paper Award Committee, and Joerg Wicker, Gillian Dobbe, Heitor Gomes who served as Local Arrangement Chair, Finance Chair and Online experience/virtual Chair respectively.

The guidance of the Steering Committee Chair, Xindong Wu, and the General Co-Chairs, Yun Sing Koh and Dacheng Tao, were invaluable throughout each step of the conference organization and we wish to express our appreciation to them for their tireless efforts. We also would like to extend our special thanks to Juzhen Dong for the many hours she put in to maintain and enhance the Cyberchair web system to support the conference.

Finally, we thank the ICDM community for their support of this premier conference. We hope you enjoy the ICDM conference and that you are inspired by the ideas found in these papers.

James Bailey and Pauli Miettinen
ICDM 2021 Program Co-Chairs
Message from the Workshop Chairs

The 21th IEEE International Conference on Data Mining (IEEE ICDMW 2021) is a premier and truly international conference for researchers and practitioners in the broad area of data mining. The ICDM Workshops program (IEEE ICDMW) aims to provide a platform for multiple workshops with a range of more focused topics to be discussed and explored, where attendees can present their original results, exchange research ideas, identify limitations, and explore new opportunities on the theoretical development and real-world applications of data mining techniques.

Due to the global COVID-19 pandemic, IEEE ICDMW was held online on 7th December 2021, followed by the IEEE ICDM 2021 conference. This year, we received 24 proposals. After the workshop proposal review, paper review and merging of workshops, the final IEEE ICDMW 2021 program consisted of 18 workshops, including 6 full-day and 12 half-day workshops. Overall, IEEE ICDMW received 266 papers while 131 papers i.e. 50% being presented in the final program. All the papers have been through a peer-review process to ensure high-quality papers being presented in the ICDMW 2021 proceedings.

A wide range of research topics in data mining have been included in IEEE ICDMW 2021, covering both theoretical research and real-world applications. Among the more traditional areas of data mining, ICDMW 2021 includes recommendation systems, information retrieval, information extraction, natural text, high-dimensional data mining, multi-source data, incremental learning, continual learning, sentiment analysis, deep learning, clustering, concept drift, novelty detection, feature engineering, time-series data mining, and data optimization. Furthermore, emerging research areas include data mining for bioinformatics, healthcare, engineering service, anomaly analytics, utility-driven data mining and learning, data in legal domain, spatial and spatio-temporal data mining, evolutionary computation based data mining, and cybersecurity.

It is no doubt that the success of ICDMW 2021 is a collective effort from many colleagues. Our particular thanks to all the workshop organizers for their delegate time and effort in preparing, and submitting their workshop proposals, managing their technical program committees, paper review, presentations, and invited talks from experts of their domains. Last but not least, we would like to thank all the reviewers for their thorough evaluations and constructive feedback to the papers, which ensure the high quality of the papers appeared in ICDMW 2021 and the support authors receive in further improving their work.

The PhD forum is featured by IEEE ICDM and continued to this year for the 10th year. As a tradition, the PhD forum aims to provide more opportunities for research students to present their works and communicate with their peers and senior researchers in the areas related to data mining. This forum particularly beneficial for PhD students in the early stage of their doctoral study, and MSc students who are planning to PhD study. The 2021 edition of the PhD forum includes 8 high-quality papers, spanning various topics of data mining and work in related fields including deep learning, crowdsourcing, statistical learning, anomaly detection, recommendation, and reinforcement learning. Many thanks to the program committee members of the PhD forum this year, who helped with the selection of these high-quality papers and who provided high quality review comments, under the guidance of the PhD forum Chairs Lisi Chen, Meng-Fen Chang and Sibylle Hess.

The Workshop Chairs of IEEE ICDMW 2021 would like to thank the Steering Committee Chair, Xindong Wu, and the General Chairs, Yun Sing Koh and Dacheng Tao, for their previous advices and leadership. We would also like to thank the Online Experience/ Virtual Chair, Heitor Gomes, for the arrangement of the Workshops day. Furthermore, we also would like to thank the Program Chairs, Pauli Miettinen and James Bailey, for their substantial contribution to the organization of the conference, and the Finance Chair, Gillian Dobbie, for managing the finance and registration of the conference. Furthermore, special thanks to the publicity Chair, Diana Benavides Prado, for her great effort on managing the conference web and social media, and her valuable help with the call for workshop proposals and call for papers. Last but not least we would like to thank Giuseppe Di Fatta for passing their experience as Workshop Chairs from the past year edition. Thank you all for your great effort in making IEEE ICDMW 2021 a success!

ICDM 2021 Workshop Chairs:
Bing Xue, Victoria University of Wellington, New Zealand
Mykola Pechenizkiy, Eindhoven University of Technology, Netherlands

Diversity, Equity, and Inclusiveness Statement

ICDM 2021 encourages the open expression and exchange of ideas, free from all forms of discrimination, retaliation, and harassment. ICDM 2021 is committed to empowering diverse, equitable, and inclusive participation, aligned with the IEEE Diversity Statement.

“IEEE’s mission to foster technological innovation and excellence to benefit humanity requires the talents and perspectives of people with different personal, cultural, and disciplinary backgrounds. IEEE is committed to advancing diversity in the technical profession, and to promoting an inclusive and equitable culture in its activities and programs that welcomes, engages and rewards those who contribute to the field without regard to race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression.”

We require all ICDM 2021 conference participants to work towards this commitment to diversity, equity, and inclusion. ICDM 2021 should be welcoming and comfortable to all attendees and those who participate in its activities.
Program Highlights

Keynote Talks
Masashi Sugiyama (RIKEN Center for Advanced Intelligence Project, The University of Tokyo). Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond. Introduced by Dacheng Tao (The University of Sydney).

Svetha Venkatesh (Deakin University’s Applied Artificial Intelligence Institute). Sample Efficient AI with Applications in Health Care and Advanced Manufacturing. Introduced by James Bailey (The University of Melbourne).

Jian Pei (Simon Fraser University). Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces. Introduced by Pauli Miettinen (University of Eastern Finland).

Tutorials
4 tutorials covering a range of advanced topics in data mining research:


Tutorial 4: Automated Taxonomy Discovery and Exploration. Organisers: Jiaming Shen, Xiaozao Gu, Yu Meng, Jiawei Han.

Workshops
6 full day and 12 half-day workshops in various fields that expand the main conference technical program with presentations and discussions of new directions and applications of data mining.

Panel Session
Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning. Panelists: Min Zhang (Tsinghua University), Lina Yao (University of New South Wales), Kevin Knight (DiDi Chuxing), Bowen Zhou (3D.com), Imed Zitouni (Google). Moderated by Michael Witbrock (The University of Auckland).

Special Sessions
Special Session in Diversity: Women in Data Science and Machine Learning. Kenti Mengersen (Queensland University of Technology). Let’s talk about uncertainty. Introduced by Richi Nayak (Queensland University of Technology).

Women in Data Science and Machine Learning Networking Panel. Moderated by Richi Nayak (Queensland University of Technology).


Networking sessions in between technical sessions.

Job Matching Networking Session. Moderated by Albert Bifet (University of Waikato) and Lin Liu (University of South Australia).

Presentations by Authors of Peer-Reviewed Papers
• 33 technical sessions in various data mining fields
• 98 long papers
• 99 short papers

Awards
• 2021 IEEE ICDM Research Contributions Award
• 2021 IEEE ICDM Outstanding Service Award
• 2021 IEEE ICDM 10-Year Highest-Impact Paper Award
• 2021 IEEE ICDM Best Paper Awards
• 2021 IEEE ICDM Student Travel Awards
Conference Program

Day 1: Wednesday, December 8

Opening Ceremony
14:00-15:00

Keynote: Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond
Speaker: Prof. Masashi Sugiyama. Session Chair: Prof. Dacheng Tao
15:00-16:00

Session 1A: Data Mining Foundations and Frontiers
Stream A – 16:00-17:25
16:00 DM425
Highly Scalable and Provably Accurate Classification in Pericardial Cysts
Eli Chien, Chao Pan, Puoya Tabaghi, and Olgica Milenkovic
16:17 DM942
Deep Incremental RNN for Learning Sequential Data: A Lyapunov Stable Dynamical System
Ziming Zhang, Guojun Wu, Yufei Yan, Yansui Li, and Run Zhou
16:34 DM1055
Group-Level Cognitive Diagnosis: A Multi-Task Learning Perspective
Jie Huang, Liu Qi, Fei Wang, Zhenya Huang, Songtao Fang, Runze Wu, Chen Enhong, Yu Su, and Shixin Wang
16:51 DM632
Isolation Kernel Density Estimation
Ke Ming Ting, Takashi Washio, Jonathan Wells, and Hang Zhang
17:08 DM947
FGG-Stream: A novel joint miner for frequent generators and closed itemsets in data streams
Louis-Roman Roux, Tomas Martin, and Petrko Valkchev

Session 1B: Graph Mining and Learning
Stream B – 16:00-17:27
16:00 DM452
Graph Transfer Learning
Andrey Gritsenko, Yuan Guo, Kimia Shayestehfard, Amin Moharrer, Jennifer Dy, and Stratos Ioannidis
16:17 DM752
A Regularized Wasserstein Framework for Graph Kernels
Asif Wjeesinghe, Qing Wang, and Stephen Gould
16:34 DM943
ThyMe: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting
Geon Lee and Kijung Shin
16:51 DM818
An Effective and Robust Framework by Modeling Correlations of Multiplex Network Embedding
Pengfei Jiao, Ruilu Li, Di Jin, Zhihao Wang, and Huameng Wu
17:03 DM825
Graph Neighborhood Routing and Random Walk for Session-based Recommendation
Zihua Zhang and Baig Wang
17:15 DM390
Fair Graph Auto-Encoder for Unbiased Graph Representations with Wasserstein Distance
Wu Fan, Kunpeng Liu, Rui Xie, Hao Liu, Hui Xiong, and Yanjie Fu

Session 1C: Sequences and Time Series 1
Stream C – 16:00-17:27
16:00 DM886
Continual Learning for Multivariate Time Series Tasks with Variable Input Dimensions
Vibhor Gupta, Jyoti Narwariya, Pankaj Malhotra, Lovekesh Vig, and Gaurav Shiroff
16:17 DM758
Towards Generating Real-World Time Series Data
Hengzhi Pei, Kan Ren, Yuqing Yang, Chong Liu, Tao Qin, and Dongsheng Li
16:34 DM979
Multi-way Time Series Join on Multi-length Patterns
Md Parvez Mollah, Vinicius M. A. Souza, and Abdullah Mueen
16:51 DM399
LIFE: Learning Individual Features for Multivariate Time Series Prediction with Missing Values
Zhao-Yu Zhang, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou
17:03 DM396
MERITS: Medication Recommendation for Chronic Disease with Irregular Time-Series
Zhao Zhang, Jianxin Li, Haoyi Zhou, Qishan Zhu, Shanghang Zhang, and Danding Wang
17:15 DM1003
PIETS: Parallelised Irregularity Encoders for Forecasting with Heterogeneous Time-Series
Fulcoo M. A. Mollah, Vinicius M. A. Souza, and Abdullah Mueen

Session 1D: Text 1
Stream D – 16:00-17:27
16:00 DM813
STAN: Adversarial Network for Cross-domain Question Difficulty Prediction
Ye Huang, Wei Huang, Shihui Tong, Zhenya Huang, Qi Liu, Enhong Chen, Jianhui Ma, Liang Wang, and Shixin Wang
16:17 DM848
Expert Knowledge-Guided Length-Variant Hierarchical Label Generation for Proposal Classification
Meng Xiao, Ziyye Qiao, Yanjie Fu, Yi Du, Pengyang Wang, and Yuanchun Zhou
16:34 DM392
Topic-Noise Models: Modeling Topic and Noise Distributions in Social Media Post Collections
Robert Churchill and Lisa Singh
16:51 DM304
BioHanBERT: A Hanzi-aware Pre-trained Language Model for Chinese Biomedical Text Mining
Kai Liu, Yan Xiong, Hao Xu, Jingwen Yue, Yangyong Zhu, and Philip S. Yu
17:03 DM450
Aspect-based Sentiment Classification via Reinforcement Learning
Chen Wang, Bo Zong, Yun Yu, Can Qin, Wei Cheng, Wenchao Yu, Xuchao Zhang, Hafeng Chen, and Yun Fu
17:15 DM786
Compressibility of Distributed Document Representations
Blaz Švrlj and Matej Petković

Session 1E: Traffic Forecasting
Stream E – 16:00-17:29
16:00 DM1205
Space Meets Time: Local Spacetime Neural Network For Traffic Flow Forecasting
Song Yang, Jianou Liu, and Kaiping Zhao
16:17 DM294
Trajectory WaveNet: A Trajectory-Based Model for Traffic Forecasting
Bo Hui, Da Yan, Haqian Chen, and Wei-Shinn Ku
16:29 DM455
Temporal Multi-view Graph Convolutional Networks for Citywide Traffic Volume Inference
Shaojie Dai, Jinshuai Wang, Chao Huang, Yanwei Yu, and Junyu Dong
16:41 DM889
Exploring Reflective Limitation of Behavior Cloning in Autonomous Vehicles
Mohammad Nazemi and Mahdi Bohloul
16:53 DM1023
DanaahNet: Unstructured Vehicle Detection using Limited Computational Resources
Tarek Reza Toha, Mastipur Rahaman, Safdi Islam Salim, Manuil Hossain, Arif Mohamin Sadi, and A. B. M. Alim Al Islam
17:05 DM543
SC15: GAN: Complex-Condition-Controlled Urban Traffic Estimation through Generative Adversarial Networks
Yingrui Zhang, Yanzhu Li, Kun Zhou, Zhenming Liu, and Jun Luo
17:17 DM611
TEST-GCNN: Topologically Enhanced Spatial-Temporal Graph Convolutional Networks for Traffic Forecasting
Muhammad Afif Ali, Suriyanarayanan Venkatesan, Victor Liang, and Hannes Knoppa

★ Long  ◆ short
**Day 1: Wednesday, December 8 (continued)**

**Session 2A: Clustering**

- **Stream A** – 18:00-19:27
  - 19:03 DM247: Long: DCF: An Efficient and Robust Density-Based Clustering Method
  - 18:51 DM343: Long: K-means for Evolving Data Streams
  - 18:34 DM437: Long: Density-Based Clustering for Adaptive Density Variation
  - 18:17 DM938: Long: Towards Interpretability and Personalization: A Predictive Framework for Clinical Time-series Analysis

**Session 2B: Algorithmic Methods**

- **Stream B** – 18:00-19:27
  - 18:17 DM505: Short: Matrix Profile XXIII: Contrast Profile: A Novel Time Series Primitive that Allows Real World Applications
  - 18:00 DM883: Short: Robust BiPoly-Matching for Multi-Granular Entities

**Session 2C: Sequences and Time Series 2**

- **Stream C** – 18:00-19:27
  - 19:03 DM1167: Long: Outlier-Robust Multi-View Subspace Clustering with Prior Constraints
  - 18:51 DM343: Long: DCF: An Efficient and Robust Density-Based Clustering Method
  - 18:34 DM437: Long: Density-Based Clustering for Adaptive Density Variation

**Session 2D: Text 2**

- **Stream D** – 18:00-19:27
  - 19:17 DM869: Long: Out-of-Category Document Identification Using Target-Category Names as Weak Supervision
  - 19:05 DM964: Long: Triplet Deep Subspace Clustering via Self-Supervised Data Augmentation

---

**Session 2E: Applications**

- **Stream E** – 18:00-19:29
  - 19:17 DM869: Short: Out-of-Category Document Identification Using Target-Category Names as Weak Supervision
  - 19:05 DM964: Short: Triplet Deep Subspace Clustering via Self-Supervised Data Augmentation
  - 18:53 DM1125: Short: Towards Interpretability and Personalization: A Predictive Framework for Clinical Time-series Analysis
**Day 2: Thursday, December 9**

**Session 3A: Graph Learning**
Stream A – 15:30-16:57

**15:30 DM540**
GraphANGEL: Adaptive and Structure-Aware Sampling on Graph Neural Networks
Jingshu Peng, Yanyan Shen, and Lei Chen

**15:47 DM629**
Deep Generation of Heterogeneous Networks
Chen Ling, Cai Yang, and Liang Zhao

**16:04 DM1155**
Fast Attributed Graph Embedding via Density of States
Saurabh Sawlani, Lingsiao Zhao, and Leman Akoglu

**16:21 DM509**
Adapting Membership Inference Attacks to GNN for Graph Classification: Approaches and Implications
BANG Wu, Xianwen Yang, Shihui Pan, and Xingliang Yuan

**16:33 DM577**
Bi-Level Attention Graph Neural Networks
Roshni Iyer, Wei Wang, and Yichou Sun

**16:45 DM473**
A new multiple instance algorithm using structural information
Xiaojun Zha, Ting Wang, Jiaxin Wang, Ying Xu, and Yugan Liu

**Session 3B: Fairness and Interpretability**
Stream B – 15:30-16:57

**15:30 DM988**
Cardiac Complication Risk Profiling for Cancer Survivors via Multi-View Multi-Task Learning
Thi Hoang Pham, Changchang Yin, Lamzi Mehta, Xuerui Zhang, and Ping Zhang

**15:47 DM559**
Multi-objective Explanations of GNN Predictions
Yifei Liu, Chao Chen, Yizhang Liu, Xi Zhang, and Shihong Xie

**16:04 DM951**
Precise Bayes Classifier: Summary of Results
Amin Vahdadian and Iyun Zhou

**16:21 DM798**
Predictive Modeling of Clinical Events with Mutual Enhancement Between Longitudinal Patient Records and Medical Knowledge Graph
Xiao Xu, Xian Xu, Yujiao Sun, Xiaoshuang Liu, Xiang Li, Guotong Xie, and Fei Wang

**16:38 DM710**
Unified Fairness from Data to Learning Algorithm
Yandu Zhang, Lei Luo, and Meng Huang

**Session 3C: Sequences and Time Series 3**
Stream C – 15:30-16:57

**15:30 DM1002**
Attentive Neural Controlled Differential Equations for Time-series Classification and Forecasting
Shouyou Jin, Heepoo Shin, Seoyoung Hong, Solhee Park, and Noseong Park

**15:47 DM1006**
SDNet: State Space Decomposition Neural Network for Time Series Forecasting
Yang Lin, Irena Koprinska, and Mashud Rana

**16:04 DM1104**
Ultra fast warping window optimization for Dynamic Time Warping
Chang Wei Tan, Matthieu Herrmann, and Geoffrey J. Webb

**16:21 DM475**
STING: Self-attention based Time-series Imputation Networks using GAN
Eunlysh Oh, Taehun Kim, Yunhu Ji, and Sushil Khyanla

**16:33 DM408**
Makoto Imamura and Takaaki Nakamura

**16:45 DM656**
Streaming Dynamic Graph Neural Networks for Continuous-Time Temporal Graph Modeling
Sheng Tian, Tao Xiong, and Leilei Shi

**Session 3D: Deep Learning 1**
Stream D – 15:30-16:57

**15:30 DM286**
Physics Interpretable Shallow-Deep Neural Networks for Physical System Identification with Unobservability
Jingyi Yuan and Yang Weng

**15:47 DM563**
Ball: a Beat-aligned Transformer for Electrocardiogram Classification
Kaiyu Li, Chen Lin, Yuhua Wei, Yuyao Sun, Jiyang Wei, Xiang Li, and Buiyue Ghan

**16:04 DM424**
Robustlydarts by Eliminating Information Bypass Leakage via Explicit Sparse Regularization
Jiuling Zhang and Zhiming Ding

**16:21 DM585**
Operation-level Progressive Differentiable Architecture Search
Arun Yu, Jian Li, Yong Li, and Weiping Wang

**16:33 DM487**
Improving Deep Forest by Exploiting High-order Interactions
Yi-He Chen, Shen-Huan Lyu, and Yuan Jang

**16:45 DM976**
Recurrent Neural Networks Meet Context-Free Grammar: Two Birds with One Stone
Hei Guan, Umang Chaudhary, Yuanyu Li, Lin Xing, Lijun Zhang, and Xiping Shen

**Session 3E: Biological and Physical Systems**
Stream E – 15:30-16:57

**15:30 DM628**
Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems
Xiaowei Jia, Yiqun Xie, Sheng Li, Shengyu Chen, Jacob Zwart, Jeffrey Sadler, Alison Appling, Samantha Oliver, and Jordan Read

**15:47 DM603**
Partial Differential Equation Driven Dynamic Graph Networks for Predicting Stream Water Temperature
Tianshu Bao, Xiaowei Jia, Jacob Zwart, Jeffrey Sadler, Alison Appling, Samantha Oliver, and Taylor Johnson

**16:04 DM986**
Climate Modeling with Neural Diffusion Equations
Heawyoung Choi, Jeongwhan Choi, Jeeyun Hyun, and Noseong Park

**16:21 DM521**
Pest-TOYO: Deep Image Mining and Multi-Feature Fusion for Real-Time Agriculture Pest Detection
Zhe Tang, Zhengyu Chen, Fan Qi, Lingsan Zhang, and Shuhong Chen

**16:33 DM743**
DIVINA: Rare Object Localization and Search in Overhead Imagery
Jonathan Amazon, Khurram Shafique, Zeeshan Rashdee, and Aaron Reite

**16:45 DM261**
Exploring the Long Short-Term Dependencies to Infer Shot Influence in Badminton Matches
Wen-Yao Wang, Tong-Fong Chan, Hui-Kuo Yang, Chih-Chuan Wang, Yao-Chun Fang, and Wen-Chih Peng

**Session 3F: Representation Learning**
Stream F – 15:30-16:57

**15:30 DM828**
Efficient Reinforced Feature Selection via Early Stopping Traversing Strategy
Kunpeng Liu, Dongjie Wang, Pengfei Wang, Man Mu, Dapeng Oliver Wu, and Yanyue Fu

**15:47 DM760**
PRGAN: Personalized Recommendation with Conditional Generative Adversarial Networks
Jing Wen, Bi-Yi Chen, Chang-Dong Wang, and Zhihong Tian

**16:04 DM616**
Robust Low-rank Deep Feature Recovery in CMA: Toward Low Information Loss and Fast Convergence
Jiahuan Ren, Zhao Zhang, Jicong Fan, Haijun Zhang, Mingliang Xu, and Meng Wang

**16:21 DM464**
A Lookahead Algorithm for Robust Subspace Recovery
Guifeng Wan and Haim Schweitzer

**16:33 DM556**
Constrained Non-Affine Alignment of Embeddings
Yuwei Wang, Yan Zheng, Yanying Peng, Michael Ye, Zhongfang Zeng, Das Mahastweta, Benrejas Mangee, Feishe Li, Wei Zhang, and Jeff Phillips

**16:45 DM695**
Jointly Multi-Similarity Loss for Deep Metric Learning
Li Zhang, Shitian Shen, Lingxiao Li, and Han Wang
<table>
<thead>
<tr>
<th>Stream A</th>
<th>Session 4A: Networks</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:30</td>
<td>DM987</td>
<td>18:45</td>
<td>Hypergraph Ego-networks and Their Temporal Evolution</td>
<td>Cazamere Coville and Jon Kleinberg</td>
</tr>
<tr>
<td>18:04</td>
<td>DM619</td>
<td>17:30</td>
<td>Better Prevent Than React: Deep Stratified Learning to Predict Hate Intensity of Twitter Reply Chains</td>
<td>Dhruv Sahni, Snehi Dahiya, Vasu Goel, Anil Bandhakavi, and Tanmay Chakraborty</td>
</tr>
<tr>
<td>18:45</td>
<td>DM625</td>
<td>18:04</td>
<td>Heterogeneous Stream-reservoir Graph Networks with Data Assimilation</td>
<td>Shengyu Chen, Alison Aplling, Samantha Oliver, Hayley Corson-Dossch, Jordan Read, Jeffrey Saddler, Jacob Zwart, and Kaiwen Ji</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream B</th>
<th>Session 4B: Causality and Fairness</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:30</td>
<td>DM663</td>
<td>18:21</td>
<td>Nonlinear Causal Structure Learning for Mixed Data</td>
<td>Wenyuan Wei and Lu Feng</td>
</tr>
<tr>
<td>17:47</td>
<td>DM762</td>
<td>18:04</td>
<td>A Robust Algorithm to Unify Online Causal Inference and Online Multi-armed Bandit Learning</td>
<td>Giao Tang and Hong Xie</td>
</tr>
<tr>
<td>18:04</td>
<td>DM1067</td>
<td>18:21</td>
<td>Fair Decision-making Under Uncertainty</td>
<td>Wenbin Zhang and Jeremy Weiss</td>
</tr>
<tr>
<td>18:21</td>
<td>DM520</td>
<td>18:33</td>
<td>Joint Scence Network and Attention-Guided for Image Captioning</td>
<td>Dongming Zhou and Jing Yang</td>
</tr>
<tr>
<td>18:33</td>
<td>DM792</td>
<td>18:45</td>
<td>Promoting Fairness through Hyperparameter Optimization</td>
<td>André Cruz, Pedro Sales, Catarina Belém, Carlos Soares, and Pedro Bizarro</td>
</tr>
<tr>
<td>18:45</td>
<td>DM934</td>
<td>18:45</td>
<td>Causal Discovery with Flow-based Conditional Density Estimation</td>
<td>Shaoqiang Ren, Haiyan Yin, Mingming Sun, and Ping Li</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream C</th>
<th>Session 4C: Theory and Optimization</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:30</td>
<td>DM452</td>
<td>18:04</td>
<td>LAGA: Lagged AllReduce with Gradient Accumulation for Minimal Idle Time</td>
<td>Ilde Haikari, Rishen Zanjir Avis, Kif Yehuda Levy, and Assaf Schuster</td>
</tr>
<tr>
<td>17:47</td>
<td>DM1168</td>
<td>18:04</td>
<td>A Linear Primal-Dual Multi-Instance SVM for Big Data Classifications</td>
<td>Lodewijk Brand, Lauren Baker, Carla Ellefsen, Jackson Sargent, and Hua Wang</td>
</tr>
<tr>
<td>18:04</td>
<td>DM881</td>
<td>18:21</td>
<td>Fast computation of distance-generalized cores using sampling</td>
<td>Nikolay Tatti</td>
</tr>
<tr>
<td>18:21</td>
<td>DM447</td>
<td>18:33</td>
<td>Limited-memory Common-directions Method With Subsampled Newton Directions for Large-scale Linear Classification</td>
<td>Jun-Nan Yen and Chih-Jen Lin</td>
</tr>
<tr>
<td>18:33</td>
<td>DM465</td>
<td>18:45</td>
<td>Online Testing of Subgroup Treatment Effects Based on Value Difference</td>
<td>Miao Yu, Wendri Lu, and Rui Song</td>
</tr>
<tr>
<td>18:45</td>
<td>DM815</td>
<td>18:45</td>
<td>Scalable Pareto Front Approximation for Deep Multi-Objective Learning</td>
<td>Michael Ruchte and Joost Grabocka</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream D</th>
<th>Session 4D: Deep Learning 2</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:47</td>
<td>DM3148</td>
<td>18:04</td>
<td>GAINDR: A Tabular Data Generation Model</td>
<td>Yishuo Zhang, Nayar Zaidi, Jiahui Zhou, and Gang Li</td>
</tr>
<tr>
<td>18:04</td>
<td>DM1031</td>
<td>18:21</td>
<td>Temporal Clustering with External Memory Network for Disease Progression Modeling</td>
<td>Zhong Zhang, Changchong Yin, and Ping Zhang</td>
</tr>
<tr>
<td>18:33</td>
<td>DM971</td>
<td>18:45</td>
<td>Fair Decision-making Under Uncertainty</td>
<td>Meet Taraiya, Anurag Beniwal, Yen-Liang Lin, and Larry Davis</td>
</tr>
<tr>
<td>18:45</td>
<td>DM878</td>
<td>18:45</td>
<td>Adversarial Regularized Reconstruction for Anomaly Detection and Generation</td>
<td>Angelica Liquori, Giuseppe Manco, Francesco Sergio Pisani, and Ettore Rittaco</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream E</th>
<th>Session 4E: Bandits and Online Methods</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:30</td>
<td>DM673</td>
<td>18:04</td>
<td>Cutting to the Chase with Warm-Start Contextual Bandits</td>
<td>Bastian Oelotto, R. Malinga Pereira, Renata Borevicha-Gajic, and Benjamin I. P. Rubinstein</td>
</tr>
<tr>
<td>17:47</td>
<td>DM911</td>
<td>18:04</td>
<td>Online Learning in Variable Feature Spaces with Mixed Data</td>
<td>Yi He, Xiaoxian Dong, Bo-Jian Hou, Yu Wang, and Fei Wang</td>
</tr>
<tr>
<td>18:21</td>
<td>DM486</td>
<td>18:38</td>
<td>Attention-based Feature Interaction for Efficient Online Knowledge Distillation</td>
<td>Tongtong Su, Qiyi Jiang, Linsong Zhang, Zhaoxiang Yu, Gang Wang, and Xiaoguang Liu</td>
</tr>
<tr>
<td>18:38</td>
<td>DM560</td>
<td>18:38</td>
<td>Dictionary Pair-based Data-Free Fast Deep Neural Network Compression</td>
<td>Yangcheng Gao, Zhao Zhang, Huijun Zhang, Minibo Zhao, Yang Yi, and Meng Wang</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream F</th>
<th>Session 4F: Meta Learning</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:30</td>
<td>DM1208</td>
<td>18:30</td>
<td>Learning to Reweight Samples with Offline Loss Sequence</td>
<td>Yuhua Wei, Chen Li, Xiaoyu Li, Jishang Wei, and Bayue Qian</td>
</tr>
<tr>
<td>17:47</td>
<td>DM661</td>
<td>18:04</td>
<td>Few-Shot Partial Multi-Label Learning</td>
<td>Yunfeng Zhao, Guoxian Yu, Lei Liu, Zhongming Yan, Carlotta Domeniconi, and Lihen Cui</td>
</tr>
<tr>
<td>19:30</td>
<td>Panel Discussion</td>
<td>17:30</td>
<td>Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning. Moderator: Michael Witbrock</td>
<td>Panel: Min Zhang, Lina Yao, Kevin Knight, Bowen Zhou, and Emre Zitouni</td>
</tr>
</tbody>
</table>
Day 3: Friday, December 10

Keynote: Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces
Speaker: Prof. Jian Pei
14:00-15:00

Session SA: Crowdsourcing
Stream A – 15:30-16:57
15:30 DM535 Truth Discovery in Sequence Labels from Crowds
Nasim SaberiPour, Aditya Kulkarni, Shong Xie, and Qi Li
15:47 DM1069 Crowdsourcing with Self-paced Workers
Xiangqiao Kang, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Wei Guo, Yanzhou Ren, and Lotan Cui
16:04 DM1197 Preference-aware Group Task Assignment in Spatial Crowdsourcing: A Mutual Information-based Approach
Yunchuan Li, Yan Zhao, and Kai Zheng
16:21 DM994 Practitioner-Centric Approach for Early Incident Detection Using Crowdsourced Data for Emergency Services
Yasas Senaratna, Ayan Mukhopadhyay, Sayyed Mohsen Vazirizade, Hemant Purohit, Saideep Nannapaneni, and Ashishk Dubey
16:33 DM1007 Detecting Adversaries in Crowdsourcing
Panagiotis Traganitis and Georgios B. Giannakis
16:45 DM1179 Zero-shot Key Information Extraction from Mixed-Style Tables: Pre-training on Wikipedia
Ge Zhang, Jia Wu, Jian Yang, Amin Beheshti, Shan Xue, Chuan Zhou, and Michael Sheng

Session SB: Explainability
Stream B – 15:30-16:49
15:30 DM847 GNES: Learning to Explain Graph Neural Networks
Yuyang Gao, Tong Sun, Rohith Bhatt, Daizhou Yu, Sungsoo Hong, and Liang Zhao
15:47 DM291 Generating Explanations for Recommendation Systems via Injective VAE
Zefu Cai and Zefeng Cai
15:49 DM457 An Ensemble of Nave Bayes Classifiers for Uncertain Categorical Data
Marcelo Malla, Alexandre Plastino, and Alex Freitas
16:01 DM459 Self-learn to Explain Siamese Networks Robustly
Chao Chen, Yifan Shen, Guixiang Ma, Xiangnan Kong, Sinhkias Rangarajan, Xi Zhang, and Shihong Xie
16:13 DM624 Alternative RuleSet Discovery to Support Black-box Model Predictions
Yochi Sasaki and Yuzuru Okajima
16:25 DM950 A Multi-view Confidence-calibrated Framework for Fair and Stable Graph Representation Learning
Ku Zhang, Liang Zhang, Bo Jin, and Xinyang Lu
16:39 DM1099 GCNN-SE: Attention as Explainability for Node Classification in Dynamic Graphs
Yucai Fan, Yuhang Yao, and Carlee Joe-Wong

Session SC: Graph Neural Networks I
Stream C – 15:30-16:57
15:30 DM1000 FRAUDRE: Fraud Detection Dual-Resistant to Graph Inconsistency and Imbalance
Ge Zhang, Ji Wu, Jian Yang, Amin Beheštizadeh, Shan Xue, Chuan Zhou, and Michael Sheng
15:47 DM435 Accurate Graph-Based PU Learning without Class Prior
Jaemin Yoo, Junghun Kim, Hoyoung Yoon, Geonsoo Kim, Changwon Jang, and Ji Kang
16:04 DM484 AS-AGCN: Adaptive Semantic Architecture of Graph Convolutional Networks for Text-Rich Networks
Zhuhao Xu, Qi Jin, Ziyang Liu, Dongxiao He, Xiao Wang, Henghang Tong, and Jiawei Han
16:21 DM214 Dynamic Attribution Graph Prediction with Conditional Normalizing Flows
Daheng Wang, Tong Zhao, Nitish Chawla, and Meng Jiang

16:33 DM1113 Adaboosting Clusters on Graph Neural Networks
Li Zheng, Jun Gao, Zhao Li, and Ji Zhang
16:45 DM1150 Heterogeneous Graph Neural Architecture Search
Yang Gao, Peng Zhang, Zhao Li, Chuan Zhou, Hong Yang, Yongzhuo Liu, and Yue Hu

Session SA: Stream C – 15:30-16:57
15:30 DM904 Combining Rating and Point-wise Losses for Training Deep Survival Analysis Models
Lu Wang, Mark Chipperfield, and Yan Li
15:47 DM1200 Spatially and Robustly Hybrid Mixture Regression Model for Inference of Spatio-temporal Dependence
Wenhan Chang, Pengtao Dang, Changlin Wan, Xiaoyu Lu, Yue Fang, Tong Zhao, Yong Zang, Bo Li, Chi Zhang, and Sha Cao
16:04 DM461 A Statistically-Guided Deep Network Transformation and Modification Framework for Data with Spatial Heterogeneity
Yuan Xie, Erhu He, Xiaowei Jia, Han Bao, Xun Zhou, Rahul Ghosh, and Praveen Raviratnam
16:21 DM802 Accurately Quantifying under Score Variability
André Maltez, Denis dos Reis, Waiqar Hassan, and Gustavo Batista
16:33 DM423 Gain-Some-Lose-Some: Reliable Quantification Under General Dataset Shift
Benjamin Denham, Edmund Lai, Roopak Sinha, and M. Asif Naem
16:45 DM1380 Boosting Deep Ensemble Performance with Hierarchical Pruning
Yanzhuo Wu and Ling Liu

Session SD: Statistical Techniques
Stream D – 15:30-16:57
15:30 DM1032 RIDE-HGNN: Adaptive Curvature Exploration Hyperbolic Graph Neural Network
Qingping Peng, Guodong Long, Tao Shen, Sen Wang, and Jing Jiang
15:47 DM757 Lab-adaptive Attention Model for Disease Risk Prediction Using Multimodal Electronic Health Records
Shuai Niu, Qing Yin, Yuyang Song, Yike Guo, and Xian Yang
16:04 DM817 SCEHR: Supervised Contrastive Learning for Clinical Risk Prediction using Electronic Health Records
Chengyi Zang and Fei Wang
16:21 DM588 SCALP - Supervised Contrastive Learning for Cardiopulmonary Disease Classification and Localization in Chest X-rays using Patient Metadata
Aly Jaiswal, Tianhao Li, Cyprian Zander, Yan Han, Justin Rousseau, Yifan Peng, and Ying Ding
16:33 DM956 ENGINE: Enhancing Neuroimaging and Genetic Information by Neural Embedding
Wanjun Ao, Wonsik Jung, Eunjin Jeon, Ahmad Wusni Muljadi, and Heung-Ill Suk
16:45 DM1103 Multi-Classification Prediction of Alzheimer’s Disease based on Fusing Multi-modal Features
Qiao Pan, Ke Ding, and Dehua Chen

Session SF: Reinforcement Learning
Stream F – 15:30-16:50
15:30 DM1032 RIDE-HGNN: Adaptive Curvature Exploration Hyperbolic Graph Neural Network
Qingping Peng, Guodong Long, Tao Shen, Sen Wang, and Jing Jiang
15:47 DM851 Deep Reinforced Attention Regression for Partial Sketch Based Image Retrieval
Dingrong Wang, Hileth Sapkota, Xumin Liu, and Qi Yu
16:04 DM999 Impression Allocation and Policy Search in Display Advertising
Yi Wang, chenming chen, xijun chen, jinli wu, xun yang, qing tan, jian xu, and Xuan-Chih lee
16:21 DM566 Mcore: Multi-Agent Collaborative Learning for Knowledge-Graph Enhanced Recommendation
Xia Li, Yanyan Shen, and Lei Chen
16:38 DM638 Multi-Objective Distributional Reinforcement Learning for Large-Scale Order Dispatching
Fan Zhou, Xiaocheng Tang, Chenfan Lu, Fan Zhang, Zhivei Qian, Jeping Ye, and Hongtu Zhu
**Day 3: Friday, December 10 (continued)**

<table>
<thead>
<tr>
<th>Session 6A: Recommender Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream A – 17:30-18:57</td>
</tr>
<tr>
<td>17:30 DM468 Hyper Meta-Path Contrastive Learning for Multi-Behavior Recommendation</td>
</tr>
<tr>
<td>17:47 DM980 Memory Augmented Multi-Instance Contrastive Predictive Coding for Sequential Recommendation</td>
</tr>
<tr>
<td>18:04 DM1012 Learning Transferable User Representations with Sequential Behaviors via Contrastive Pre-training</td>
</tr>
<tr>
<td>18:21 DM217 Composition-Enhanced Graph Collaborative Filtering for Multi-behavior Recommendation</td>
</tr>
<tr>
<td>18:31 DM630 Stochastic Tunable Cool item Integration Mechanisms for Recommender Systems</td>
</tr>
<tr>
<td>18:45 DM722 MetaEDL: Meta Evidential Learning For Uncertainty-Aware Cool-Start Recommendations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6A: Recommender Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream B – 17:30-18:52</td>
</tr>
<tr>
<td>17:30 DM277 Structure-Aware Stabilization of Adversarial Robustness with Massive Contrastive Adversaries</td>
</tr>
<tr>
<td>17:47 DM474 Graph-based Adversarial Online Kernel Learning with Adaptive Embedding</td>
</tr>
<tr>
<td>18:04 DM626 Towards Stochastic Neural Network via Feature Distribution Calibration</td>
</tr>
<tr>
<td>18:16 DM637 Adversarial Learning of Balanced Triangles for Accurate Community Detection on Signed Networks</td>
</tr>
<tr>
<td>18:28 DM831 Detecting and Mitigating Test-time Failure Risks via Model-agnostic Uncertainty Learning</td>
</tr>
<tr>
<td>18:40 DM842 Attacking Similarity-Based Sign Prediction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6A: Recommender Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream C – 17:30-18:59</td>
</tr>
<tr>
<td>17:30 DM124 Hypergraph Convolutional Network for Group Recommendation</td>
</tr>
<tr>
<td>17:47 DM803 Heterogeneous Graph Neural Network with Distance Encoding</td>
</tr>
<tr>
<td>17:59 DM733 MC-RGCN: A Multi-Channel Recurrent Graph Convolutional Network to Learn High-Order Social Relations for Diffusion Prediction</td>
</tr>
<tr>
<td>18:11 DM854 HyperTeNet: Hypergraph and Transformer-based Neural Network for Personalized List Continuation</td>
</tr>
<tr>
<td>18:23 DM957 Learnable Structural Semantic Readout for Graph Classification</td>
</tr>
<tr>
<td>18:35 DM1123 GQNAS: Graph Q Network for Neural Architecture Search</td>
</tr>
<tr>
<td>18:47 DM418 StarGAT: Star-Shaped Hierarchical Graph Attentional Network for Heterogeneous Network Representation Learning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6B: Adversarial Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream B – 17:30-18:52</td>
</tr>
<tr>
<td>17:30 DM277 Structure-Aware Stabilization of Adversarial Robustness with Massive Contrastive Adversaries</td>
</tr>
<tr>
<td>17:47 DM474 Graph-based Adversarial Online Kernel Learning with Adaptive Embedding</td>
</tr>
<tr>
<td>18:04 DM626 Towards Stochastic Neural Network via Feature Distribution Calibration</td>
</tr>
<tr>
<td>18:16 DM637 Adversarial Learning of Balanced Triangles for Accurate Community Detection on Signed Networks</td>
</tr>
<tr>
<td>18:28 DM831 Detecting and Mitigating Test-time Failure Risks via Model-agnostic Uncertainty Learning</td>
</tr>
<tr>
<td>18:40 DM842 Attacking Similarity-Based Sign Prediction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6B: Adversarial Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream C – 17:30-18:59</td>
</tr>
<tr>
<td>17:30 DM124 Hypergraph Convolutional Network for Group Recommendation</td>
</tr>
<tr>
<td>17:47 DM803 Heterogeneous Graph Neural Network with Distance Encoding</td>
</tr>
<tr>
<td>17:59 DM733 MC-RGCN: A Multi-Channel Recurrent Graph Convolutional Network to Learn High-Order Social Relations for Diffusion Prediction</td>
</tr>
<tr>
<td>18:11 DM854 HyperTeNet: Hypergraph and Transformer-based Neural Network for Personalized List Continuation</td>
</tr>
<tr>
<td>18:23 DM957 Learnable Structural Semantic Readout for Graph Classification</td>
</tr>
<tr>
<td>18:35 DM1123 GQNAS: Graph Q Network for Neural Architecture Search</td>
</tr>
<tr>
<td>18:47 DM418 StarGAT: Star-Shaped Hierarchical Graph Attentional Network for Heterogeneous Network Representation Learning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6D: Active and Semi-Supervised Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream D – 17:30-18:57</td>
</tr>
<tr>
<td>17:30 DM389 Flexible, Robust, Scalable Semi-supervised Learning via Reliability Propagation</td>
</tr>
<tr>
<td>17:47 DM843 PARWS: Winner determination from Active Pairwise Comparisons under a Shoestring Budget</td>
</tr>
<tr>
<td>18:04 DM1008 Finding Age Path of Self-Paced Learning</td>
</tr>
<tr>
<td>18:21 DM1154 Incomplete Multi-view Multi-label Active Learning</td>
</tr>
<tr>
<td>18:33 DM298 Self-supervised Universal Domain Adaptation with Adaptive Memory Separation</td>
</tr>
<tr>
<td>18:45 DM608 Multimodal N-best List Rescoring with Weakly Supervised Pre-training in Hybrid Speech Recognition</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 6E: Matrix and Tensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream E – 17:30-18:57</td>
</tr>
<tr>
<td>17:30 DM868 MASCO: A Quanzation Framework for Efficient Matrix Factorization in Recommender Systems</td>
</tr>
<tr>
<td>17:47 DM544 Anomaly Detection with Prototype-Guided Discriminative Latent Embeddings</td>
</tr>
<tr>
<td>18:04 DM724 Discriminative Additive Scale Loss for Deep Imbalanced Classification and Embedding</td>
</tr>
<tr>
<td>18:21 DM1044 Overfitting Avoidance in Tensor Tram Factorization and Completion: Prior Analysis and Inference</td>
</tr>
<tr>
<td>18:33 DM589 Communication Efficient Tensor Factorization for Decentralized Healthcare Networks</td>
</tr>
<tr>
<td>18:45 DM641 Summarizing User-item Matrix By Group Utility Maximization</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Special Session in Diversity: Women in Data Science and Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:30 Let’s talk about uncertainty</td>
</tr>
<tr>
<td>Speaker: Prof Kerri Mengersen</td>
</tr>
<tr>
<td>Session Chair: Prof Richi Nayak</td>
</tr>
<tr>
<td>Panel Chair: Prof Richi Nayak</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Women in Data Science and Machine Learning Learning Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker: Prof Kerri Mengersen</td>
</tr>
<tr>
<td>Session Chair: Prof Richi Nayak</td>
</tr>
<tr>
<td>Panel Chair: Prof Richi Nayak</td>
</tr>
</tbody>
</table>
### Tutorials Program

<table>
<thead>
<tr>
<th>Date (NZT)</th>
<th>Title</th>
<th>Organisers</th>
</tr>
</thead>
<tbody>
<tr>
<td>16:00-18:30, Wednesday 8 December 2021</td>
<td>Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems</td>
<td>Denis Parra (PUC Chile), Antonio Ossa-Guerra (PUC Chile), Manuel Cartagena (PUC Chile), Patricio Cerdá-Martíni (PUC Chile and MinatsuDB), Felipe del Río (PUC Chile), Isidora Palma (PUC Chile), Diego Saiz-Trumper (Wikimedia Foundation), and Miriam Redi (Wikimedia Foundation)</td>
</tr>
<tr>
<td>16:00-18:30, Wednesday 8 December 2021</td>
<td>Complement, Composite and Context: The 3C-Law to Build Multidomain Recommender Systems</td>
<td>Liang Hu, Shoujun Wang (Macquarie University), Qi Zhang, Dora D. Liu, and Longbing Cao (University of Technology Sydney)</td>
</tr>
<tr>
<td>18:30-21:00, Wednesday 8 December 2021</td>
<td>Roles in Networks - Foundations, Methods and Applications</td>
<td>Yulong Pei (Eindhoven University of Technology), Pengfei Jiao (Tianjin University), Xuan Guo (Tianjin University), George Fletcher (Eindhoven University of Technology), and Mykola Pechenizkiy (Eindhoven University of Technology)</td>
</tr>
<tr>
<td>15:30-20:30, Thursday 9 December 2021</td>
<td>Automated Taxonomy Discovery and Exploration</td>
<td>Jiaming Shen (Univ. of Illinois at Urbana-Champaign), Xiaotao Gu (Univ. of Illinois at Urbana-Champaign), Yu Meng (Univ. of Illinois at Urbana-Champaign), and Jiawei Han (Univ. of Illinois at Urbana-Champaign)</td>
</tr>
</tbody>
</table>

Tutorials Chairs: Wei Liu, University of Sydney, Australia; Katharina Morik, TU Dortmund, Germany

### PhD Forum Program

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00-14:05</td>
<td>Opening Remarks</td>
<td>Speaker: Sibo Wang, The Chinese University of Hong Kong</td>
</tr>
<tr>
<td>14:05-15:00</td>
<td>Keynote</td>
<td>Title: Managing and Mining with Massive Graphs</td>
</tr>
<tr>
<td>15:00-16:30</td>
<td>Students' Presentations</td>
<td>Long: 12-15min, short: 7-10min, Compress 5 long + 3 short into 81~105 mins in total before coffee 3hr break at 4:30pm-5:30pm</td>
</tr>
<tr>
<td>16:30-18:30</td>
<td>Coffee Break</td>
<td>Gather Town 2hr Coffee Break</td>
</tr>
</tbody>
</table>
Keynote speakers

Masashi Sugiyama
RIKEN Center for Advanced Intelligence Project/
The University of Tokyo

Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond

Abstract
When machine learning systems are trained and deployed in the real world, we face various types of uncertainty. For example, training data at hand may contain insufficient information, label noise, and bias. In this talk, I will give an overview of our recent advances in robust machine learning.

Bio
Masashi Sugiyama received Doctor of Engineering in Computer Science from Tokyo Institute of Technology, Japan in 2001. Experiencing Assistant Professor and Associate Professor at Tokyo Institute of Technology, he became Professor at the University of Tokyo in 2014. Since 2016, he has been concurrently serving as Director of RIKEN Center for Advanced Intelligence Project. His research interests include theories and algorithms of machine learning. He received the Japan Academy Medal in 2017 for his series of machine learning research.

Svetha Venkatesh
Deakin University’s Applied Artificial Intelligence Institute (A²I²)

Sample Efficient AI with Applications in Health Care and Advanced Manufacturing

Abstract
From Guglielmo Marconi who developed the radio telegraph to the Wright brothers who invented flying machines, curiosity driven experimentation has powered human innovation. Such experimental optimisation remains an integral part of the Scientific Method. This time-honoured method needs a step change to accelerate scientific innovation because this iterative method quickly hits limits.

To speed-up innovation, it is imperative to expand the capability of experimental optimisation and improve its efficiency. This talk will demonstrate how sample efficient AI can be used to deliver this acceleration in experimental design. I will discuss how the methods can be applied widely, focusing on health and advanced manufacturing particularly in settings where data is scarce and experimentation is expensive. In healthcare, I show how these methods can accelerate the design of clinical trials to efficiently determine the optimal strategy. In advanced manufacturing, I will show how it can be applied broadly from inventing new materials and alloys to accelerating industrial processes.

The second part of the talk will focus on the new machine learning innovations that have been formulated and solved to advance experimental design. These include incorporating experimental design constraints such as process constraints, transferring knowledge from previous experiments or experimenter “hunches”, and high dimensional Bayesian optimisation so that the number of experimental control variables can be increased.

Bio
Svetha Venkatesh is an ARC Australian Laureate Fellow, Alfred Deakin Professor and a co-Director of Applied Artificial Intelligence Institute (A²I²) at Deakin University. She was elected a Fellow of the International Association of Pattern Recognition in 2004 for contributions to formulation and extraction of semantics in multimedia data, a Fellow of the Australian Academy of Technological Sciences and Engineering in 2006, and a Fellow of the Australian Academy of Science in 2021 for groundbreaking research and contributions that have had clear impact. In 2017, Professor Venkatesh was appointed an Australian Laureate Fellow, the highest individual award the Australian Research Council can bestow.

Professor Venkatesh and her team have tackled a wide range of problems of societal significance, including the critical areas of autism, security and aged care. The outcomes have impacted the community and evolved into publications, patents, tools and spin-off companies. This includes 650+ publications, three full patents, one start-up company (iCetana) and two significant products (TOBY Playpad, Virtual Observer).

Professor Venkatesh has tackled complex pattern recognition tasks by drawing inspiration and models from widely diverse disciplines, integrating them into rigorous computational models and innovative algorithms. Her main contributions have been in the development of theoretical frameworks and novel applications for analysing large scale, multimedia data. This includes development of several Bayesian parametric and non-parametric models, solving fundamental problems in processing multiple channel, multi-modal temporal and spatial data.
Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces

Abstract
We believe data science and AI will change the world. No matter how smart and powerful an AI model we can build, the ultimate testimony of the success of data science and AI is users' trust. How can we build trustworthy data science? At the level of user-model interaction, how can we convince users that a data analytic result is trustworthy? At the level of group-wise collaboration for data science and AI, how can we ensure that the parties and their contributions are recognized fairly, and establish trust between the outcome (e.g., a model built) of the group collaboration and the external users? At the level of data science participant eco-systems, how can we effectively and efficiently connect many participants of various roles and facilitate the connection among supplies and demands of data and models?

In this talk, I will brainstorm possible directions to the above questions in the context of an end-to-end data science pipeline. To strengthen trustworthy interactions between models and users, I will advocate exact and consistent interpretation of machine learning models. Our recent results show that exact and consistent interpretations are not just theoretically feasible, but also practical even for API-based AI services. To build trust in collaboration among multiple participants in coalition, I will review some progress in ensuring fairness in federated learning, including fair assessment of contributions and fairness enforcement in collaboration outcome. Last, to address the need of trustworthy data science eco-systems, I will review some latest efforts in building data and model marketplaces and preserving fairness and privacy. Through reflection I will discuss some challenges and opportunities in building trustworthy data science for possible future work.

Bio
Jian Pei is a Professor in the School of Computing Science at Simon Fraser University. He is a well-known leading researcher in the general areas of data science, big data, data mining, and database systems. His expertise is on developing effective and efficient data analysis techniques for novel data-intensive applications, and transferring his research results to products and business practice. He is recognized as a Fellow of the Royal Society of Canada (Canada’s national academy), the Canadian Academy of Engineering, the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE). He is one of the most cited authors in data mining, database systems, and information retrieval. Since 2000, he has published one textbook, two monographs and over 300 research papers in refereed journals and conferences, which have been cited extensively by others. His research has generated remarkable impact substantially beyond academia. For example, his algorithms have been adopted by industry in production and popular open source software suites. Jian Pei also demonstrated outstanding professional leadership in many academic organizations and activities. He was the editor-in-chief of the IEEE Transactions of Knowledge and Data Engineering (TKDE) in 2013-16, the chair of the Special Interest Group on Knowledge Discovery in Data (SIGKDD) of the Association for Computing Machinery (ACM) in 2017-2021, and a general co-chair or program committee co-chair of many premier conferences. He maintains a wide spectrum of industry relations with both global and local industry partners. He is an active consultant and coach for industry on enterprise data strategies, healthcare informatics, network security intelligence, computational finance, and smart retail. He received many prestigious awards, including the 2017 ACM SIGKDD Innovation Award, the 2015 ACM SIGKDD Service Award, the 2014 IEEE ICDM Research Contributions Award, the British Columbia Innovation Council 2005 Young Innovator Award, an NSERC 2008 Discovery Accelerator Supplements Award (100 awards cross the whole country), an IBM Faculty Award (2006), a KDD Best Application Paper Award (2008), an ICDM Influential Paper Award (2018), a PAKDD Best Paper Award (2014), a PAKDD Most Influential Paper Award (2009), and an IEEE Outstanding Paper Award (2007).
Panel

Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning

Min Zhang
Tsinghua University

Dr. Min Zhang is a tenured associate professor in the Department of Computer Science & Technology (DCST), Tsinghua University, and is the vice director of the AI lab, DCST. She specializes in Web search, personalized recommendation, and user modeling. Currently she serves as the Editor-in-Chief of ACM Transaction on Information Systems (TOIS), and ACM SIGIR Executive Committee member, and PC chair or Area Chairs of top conferences such as SIGIR, WSDM, theWebConf, etc. She was awarded IBM Global Faculty Award in 2020. She also owns 12 patents and cooperates with many international and domestic enterprises.

Kevin Knight
Didi Chuxing

Kevin Knight is Chief Scientist for Natural Language Processing at Didi Chuxing. He received a PhD in computer science from Carnegie Mellon University and a bachelor’s degree from Harvard University. Prof. Knight’s research interests include machine translation, dialog processing, natural language generation, automata theory, and decipherment of historical manuscripts. He taught computer science at USC for 25 years and has authored over 150 NLP research papers. Prof. Knight also co-wrote the widely-adopted textbook “Artificial intelligence” (McGraw-Hill).

In 2001, he co-founded the machine translation company Language Weaver, Inc. Dr. Knight served as President of the Association for Computational Linguistics (ACL) in 2011, as General Chair for ACL conferences in 2005 and 2016, and as Program Chair in 2001 and 2020. He is a Fellow of the ACL, and the Association for the Advancement of Artificial Intelligence (AAAI).

Bowen Zhou
JD.com

Dr. Bowen Zhou is a Senior Vice President of JD Group (JD.com), President of Foundational Technologies of JD Tech & Director of JD AI Research. He has also held a number of top technology executive positions at JD Group, including the Chair of JD Technology Committee and President of JD Cloud & AI. Dr. Zhou has decades of experience in natural language technology, machine learning, and artificial intelligence. Prior to joining JD.com in 2017, Dr. Zhou held several key leadership positions during his 15-year tenure at IBM Research’s headquarters. He most recently served as Director of the AI Foundations Lab at IBM Research in New York, Chief Scientist of IBM Watson Group, and a Distinguished Engineer of IBM. As an IEEE Fellow, Dr. Zhou has published over 100 peer-reviewed papers at top international conferences and journals, with over 11,200 citations, including a few of his pioneering works were cited more than 1500 times each. He previously served as a member of the IEEE Speech and Language Technical Committee, Associate Editor of IEEE Transactions, ICASSP Area Chair (2011-2015), ACL, and NAACL Area Chair.

Imed Zitouni
Google

Imed Zitouni is a director of engineering at Google leading R&D efforts on NLU to enhance and enable capabilities for Semantic Search and Answers leveraging the power of the Knowledge Graph. Before joining Google in 2019, Imed was at Microsoft leading the NLU and the Conversation Engine effort for the digital assistant Cortana. Prior to joining Microsoft in 2012, Imed was a Senior Researcher at IBM Watson for almost a decade, working on several NLP initiatives including the Watson initiative around informatics extraction, language modeling and automatic machine translation. Prior to IBM, Imed was a research member at Bell Labs, Lucent Technologies, for half a dozen years working on speech recognition, language modeling and spoken dialog systems. Imed received his M.Sc. and Ph.D. from the University-of-Nancy1 and INRIA in France. He also obtained a MEng degree in computer science from ENSI in Tunisia.

Imed is the Editor-in-Chief of ACM Transactions on Asian and Low-Resources Language Processing. He is a senior member of IEEE, served as a board member of the IEEE Speech and Language Processing Technical Committee, and was the associate editor of IEEE Trans. on Audio, Speech and Language Processing. He also served as the Information Officer of the ACL SIG on Semantic-Languages and served as chair as well as reviewing-committee-member of several conferences and journals in the area of machine learning, information retrieval and natural language technologies. Imed is the author/co-author of two books, half-dozen book-chapters as well as more than 120 patents and scientific papers.

Moderator:

Michael Witbrock
University of Auckland

Michael Witbrock is a Professor at the School of Computer Science, The University of Auckland. He is leading the research group, the Strong AI Lab, at the intersection of machine learning, reasoning and natural language understanding, with an additional focus on achieving the best social and civilisational impacts of increasingly powerful AI.
Special Sessions

**Special Session in Diversity: Women in Data Science and Machine Learning**

**Kerri Mengersen**
Queueing University of Technology

*Let’s talk about uncertainty*

**Abstract**

Statistical and machine learning analyses provide a wide array of insights to expand knowledge and enhance data-focused decision making. This has led to a surge in demand for, and indeed understanding of, estimates and predictions based on these analyses, not only by scientists but also policy makers and the public. A concomitant issue that is perhaps less well addressed or understood is the uncertainty of these values. In this presentation, I will touch on the many forms of uncertainty in statistical and machine learning analyses, and discuss by example some of our approaches to quantification and communication of uncertainty. The argument will be made that while we have come a long way, the expression and use of uncertainty in science and decision-making is still in its infancy. Indeed, I will posit that SML uncertainty is so important for all members of the community that it could be considered the ‘4th R’ after reading, writing and arithmetic. I will preface the talk with a brief discussion about women in data science, and I will mention our ambition at QUT to create a Centre of Excellence to amplify our research.

**Bio**

Kerrie Mengersen graduated in 1985 with a Bachelor of Arts (Honours Class 1), majoring in Mathematics (Statistics) and Computing, and received her PhD in Mathematical Statistics in 1989 from the University of New England, New South Wales. Her PhD thesis was on the topic of ranking and selection under the supervision of Professor Eve Bofinger, one of the pioneer female university researchers in regional Australia.

Following graduation, she was recruited to a commercial statistical consulting company, which provided her with strong experience in a wide range of statistical methods in the context of diverse applied problems. Her career since then has been characterised by a dual focus of engaging with and developing new statistical methodology motivated by, and motivating, challenging statistical applications.

In 2016, QUT awarded the title of Distinguished Professor to Professor Kerrie Mengersen in recognition of her outstanding achievements, both nationally and internationally, in mathematics and statistical research. Distinguished Professor Mengersen is acknowledged to be one of the leading researchers in her discipline.

In 2016 Professor Mengersen also received two more prestigious awards: the Statistical Society of Australia’s Pitman Medal, the highest award presented by the Society and the first time it has been presented to a woman, and the Research Excellence award by the Cooperative Research Centre for Spatial Analysis (CRCSA).

In 2018 Professor Mengersen was elected a Fellow of the Australian Academy of Science (AAS), a Fellow of the Academy of Social Sciences in Australia (ASSA), and an Invited Fellow of the Queensland Academy of Arts and Sciences (QQAS).

---

**Special Session: PhD Forum**

**Sibo Wang**
The Chinese University of Hong Kong

*Managing and Mining with Massive Graphs*

**Abstract**

In this talk, I will share some of my research on massive graph data analysis and processing that can be applied to many real scenarios in industries. The topics include personalized PageRank query processing and graph embedding.

**Bio**

Sibo Wang is an Assistant Professor in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong since Dec 2018. He received his B.E. in Software Engineering in 2011 from Fudan University and his Ph.D. in Computer Science in 2016 from Nanyang Technological University. His main research area is database and data mining, especially big data analytics and processing, graph mining and graph representation learning. Most of his research works have been published in top conferences like SIGMOD, VLDB, and SIGKDD. His professional services include Workshop Chair at ICDE 2022 Conference and Local Organization Chair of ADC 2018, Program Committee for VLDB 2020-2021, SIGMOD 2019-2021, WWW2020-2022, ICDE 2021-2022, UCAI 2020, AAAI 2021-2022, PAKDD 2018-2022, DASFAA 2019-2022. He also served as a peer reviewer in journals like TODS, TKDE, VLDBJ, TOS, and TKDD.
Job matching program highlights

The IEEE ICDM 2021 Job Matching Program provides an opportunity for ICDM attendees to apply for data science jobs posted by the sponsors and other attendees of the conference. A job matching session will be held on Day 1 of the conference, to provide a platform for registered job seekers and recruiters to meet, and for general conference attendees to network. Please refer to our Job Matching page (https://icdm2021.auckland.ac.nz/job-matching/) to register your interest prior to the conference, or contact icdm2021@auckland.ac.nz for more details or queries.

Virtual social event highlights

We look forward to welcoming you to the IEEE ICDM virtual conference. While we regret that the COVID pandemic prevented us from holding the conference in Auckland, we are excited about the opportunities of holding an engaging virtual conference.

The University of Auckland developed this year’s virtual platform. It features live polls, video chat and Q&A, group networking capabilities, and much more.

Based on our experience with previous virtual events, we have designed simple and effective interactions to make the event as engaging as possible.

• Dedicated coffee break rooms after every stream. Attendees can immediately join the room after the stream has finished and continue the discussion.
• Speed networking during the coffee break. Attendees can meet with other researchers without the need to choose a room or approach people directly actively.
• Meeting hub. If you want to catch up with other attendees at any time, you can always join the meeting hub!
• PhD Forum. The attendees of the PhD forum will have access to gather town during the event.

Virtual social event program

On the first day of the conference, we will have a one-hour coffee break. It will include dedicated coffee break rooms and speed networking events.

In the following days of the conference, we will have three coffee breaks of 30 minutes each. Each of these breaks will feature the dedicated coffee break rooms and speed networking events.

Awards

Awards:
• 2021 IEEE ICDM Research Contributions Award
• 2021 IEEE ICDM Outstanding Service Award
• 2021 IEEE ICDM 10-Year Highest-Impact Paper Award
• 2021 IEEE ICDM Best Paper Awards
• 2021 IEEE ICDM Student Travel Awards

Awards Committee:
• Xia Ning, The Ohio State University (Awards Chair)
• Pauli Pietiläinen, University of Eastern Finland, and James Bailey, The University of Melbourne (Program Chairs)
• Xindong Wu, Miningamp Academy of Sciences (Steering Committee Chair)
• Three established data mining researchers:
  – Jian Pei, Simon Fraser University
  – Diane Joyce Cook, Washington State University
  – Eamonn Keogh, University of California Riverside

Life at Two Sigma

At Two Sigma, we’re proud of our company culture: a collaborative and diverse environment that challenges and empowers our employees to find smart solutions to the world’s hardest problems.

We want to help everyone here grow and get better, so we offer a myriad of internal programs and initiatives that encourage teaching, growth, and mobility.

Support

We@TS
Women at Two Sigma, or W@TS, seeks to build a strong community of women and allies at Two Sigma and beyond by providing career development opportunities, creating the space for genuine connection and mentorship, using intersectional advocacy, and promoting ally education.

Be@TS
Be@TS aims to foster racial and ethnic diversity at Two Sigma by supporting a culture that is inclusive to people of color.

Grow

NewTS
Two Sigma’s Early Careers Program integrates Engineering and Modeling campus hires into life at Two Sigma.

Interns
Spend the summer working in a creative and hands-on environment that gives you the freedom to innovate, explore, and dig into unique and complex challenges.

Disability Alliance
Disability Alliance advocates a welcoming work environment for people with disabilities by building awareness, educating employees, cultivating community, and serving as a resource and support system.

OutTSigma
OutSigma brings focus and resources to the recruitment, development, and retention of LGBTQ+ talent.

Learn more at twosigma.com/careers/culture

Data Clinic
Two Sigma’s Data Clinic connects our experts in data and technology with nonprofits, government agencies, and academic institutions to have greater impact on the communities they serve.

Open Source
Our teams are encouraged to contribute to and originate open source projects. As a company, we actively support core maintainers of OSS projects, host hackathons, and sponsor open source nonprofits.

ImpacTS
ImpacTS surfaces opportunities for employees to volunteer in activities that strengthen our neighbors and communities around the world.
Organising Committee

General Chairs

Yun Sing Koh
The University of Auckland, New Zealand

Dacheng Tao
The University of Sydney, Australia

Program Chairs

Pauli Miettinen
University of Eastern Finland, Finland

James Bailey
The University of Melbourne, Australia

Finance Chair

Gillian Dobbie
The University of Auckland, New Zealand

Sponsorship Chairs

Kaiqi Zhao
The University of Auckland, New Zealand

Local Arrangement Chair

Joerg Wicker
The University of Auckland, New Zealand

Eva García-Martín
Ekkono Solutions, Sweden

Tutorial Chairs

Wei Liu
University of Technology Sydney, Australia

Katharina Morik
TU Dortmund, Germany

Worship Chairs

Bing Xue
Victoria University of Wellington, New Zealand

Mykola Pechenizkiy
Eindhoven University of Technology, Netherlands

Steering Committee Chair

Xindong Wu
Mininglamp Academy of Sciences, China

Philippe Fournier-Viger
Harbin Institute of Technology, China
Conference Sponsors

IEEE ICDM 2021 is sponsored by

Main Conference Sponsor
Google

Gold Sponsor
TWO SIGMA

Student Attendance Award Sponsor
NSF

Keynote Sponsor

Best Student Paper, Best Paper, 10-Year Highest Impact Paper Awards Sponsor
TCII