ADC 2020 ECR/PhD Workshop Program

6 February 2020 (ATC422/ATC423)

8:30am – 9:30am: Registration, Arrival Tea and Coffee (ATC 103 & Foyer)

9:30am – 10:30am: Plenary Session: *Machine Learning for Data Streams*, Albert Bifet (ATC 101 Lecture Theatre)

10:30am – 10:50am: Morning Tea (ATC 103 & Foyer)

11:00am – 11:10am: Workshop opening: Chengfei Liu (ATC422/ATC423)

11:10am – 12:00pm: Invited talk: *Indoor Data Management: Progress, Challenges and Opportunities*, Muhammad Aamir Cheema (ATC422/ATC423)

12:20pm – 1:10pm: Lunch & Poster Viewing (ATC 103 & Foyer)

1:10pm – 2:10pm: Plenary Session: *Towards High-Quality Big Data for Responsible Data Science*, Divesh Srivastava (ATC 101 Lecture Theatre)

2:10pm – 2:40pm: Lightning Talks (ATC 101 Lecture Theatre)

3:00pm – 4:00pm: Invited talk: *Data Privacy: From Public Policy to Formal Methods*, Benjamin I. P. Rubinstein (ATC422/ATC423)

4:10pm – 4:30pm: Afternoon Tea (ATC 103 & Foyer)

4:30pm – 5:30pm: Research Excellence Forum: Zhifeng Bao, Renata Borovica-Gajic, Lijun Chang, Muhammad Aamir Cheema, Jianxin Li, Jianzhong Qi, Benjamin I. P. Rubinstein (ATC422/ATC423)

6:30pm – 9:00pm: Workshop dinner

7 February 2020 (ATC422/ATC423)

9:30am – 10:30am: Invited talk: *Efficient Large Graph Processing via Data Reduction*, Lijun Chang (ATC422/ATC423)

10:30am – 10:50am: Morning Tea (ATC 103 & Foyer)

11:00am – 12:00pm: Invited talk: *Making Sense of Massive Urban Data*, Zhifeng Bao (ATC422/ATC423)

12:20pm – 2:00pm: Lunch (ATC 103 & Foyer)

2:00pm – 3:00pm: Invited talk: *Spatio-Visual Query Processing*, Farhana Choudhury (ATC422/ATC423)

3:00pm – 4:00pm: Invited talk: *Anchored Vertex Exploration for Community Engagement in Social Networks*, Jianxin Li (ATC422/ATC423)

4:00pm – 4:10pm: Workshop closing: Chengfei Liu (ATC422/ATC423)

4:10pm – 4:30pm: Afternoon Tea (ATC 103 & Foyer)

Speakers

Muhammad Aamir Cheema, Monash University, Australia

Title: Indoor Data Management: Progress, Challenges and Opportunities

Abstract: A large part of modern life is lived indoors such as in homes, offices, shopping malls, universities, libraries and airports. However, almost all of the existing location-based services (LBS) have been designed only for outdoor space. This is mainly because the global positioning system (GPS) and other positioning technologies cannot accurately identify the locations in indoor venues. Some recent initiatives have started to cross this technical barrier, promising huge future opportunities for research organizations, government agencies, technology giants, and enterprising start-ups - to exploit the potential of indoor LBS. Consequently, indoor data management has gained significant research attention in the past few years and the research interest is expected to surge in the upcoming years. This will result in a broad range of indoor applications including emergency services, public services, in-store advertising, shopping, tracking, guided tours, and much more. In this talk, we first highlight the importance of indoor data management and the unique challenges that need to be addressed. Subsequently, we provide an overview of some of the existing research in indoor data management. Finally, we discuss the future research directions in this important and growing research area.

Bio: Muhammad Aamir Cheema is an ARC Future Fellow and an Associate Professor at Faculty of Information Technology, Monash University, Australia. He obtained his PhD from UNSW Australia in 2011. He is the recipient of 2012 Malcolm Chaikin Prize for Research Excellence in Engineering, 2013 Discovery Early Career Researcher Award, 2014 Dean's Award for Excellence in Research by an Early Career Researcher, 2018 Future Fellowship, 2018 Monash Student Association Teaching Award and 2019 Young Tall Poppy Science Award. He has also won two CiSRA best research paper of the year awards, two invited papers in the special issue of IEEE TKDE on the best papers of ICDE, and two best paper awards at WISE 2013 and ADC 2010, respectively. He served as PC co-chair for ADC 2015, ADC 2016, 8th ACM SIGSPATIAL Workshop ISA 2016 & 2018, IWSC 2017, proceedings chair for DASFAA 2015 & ICDE 2019, tutorial co-chair for APWeb 2017 & MDM 2019 and publicity co-chair for ACM SIGSPATIAL 2017 & 2018.

Benjamin I. P. Rubinstein, The University of Melbourne, Australia

Title: Data Privacy: From Public Policy to Formal Methods

Abstract: What makes for good research directions in computer science? Papers in top ranked conferences are great, but so too is broader impact on society not reflected by citation counts. In this talk I'll summarise two research projects in data privacy that span the impact spectrum. First, I'll relate the story of how the Australian government released a detailed medical dataset, how and why we reidentified people in this dataset, and what happened next. I'll touch on how government agencies now hope to share data through an inherently flawed approach called "Five Safes". I'll then summarise the framework of differential privacy with some recent progress in the area by my group (ICML'17), and discuss why formal methods from computer science (often advanced by the databases community!) are a critical piece of any viable solution to data sharing. I hope to demonstrate how thinking about impact drives exciting technical research, and supports a raison d'être for universities in modern society.

Bio: Ben is an Associate Professor of Computing and Information Systems and Associate Dean (Research) of the Melbourne School of Engineering. After a BSc/BE(hons), MCompSci from Melbourne, he completed a PhD at Berkeley EECS at the intersection of machine learning and security & privacy, completing early work in adversarial learning and differential privacy. He spent several years in industry research labs, the most time working at Microsoft Research Silicon Valley on data integration in Bing and the Xbox, then joining Melbourne 2013. Ben has been a DECRA and Tall Poppy awardee. He was part of the team that reidentified citizens in the 2016 release of 30 years of Medicare billing records for 10% of the Australian population, and a State MP in the 2017 3-year Myki transport data release. His work on adversarial learning is funded by Defence (NGTF) and Facebook, while his group has delivered scalable end-to-end Bayesian data integration for the Australian Bureau of Statistics.

Lijun Chang, The University of Sydney, Australia

Title: Efficient Large Graph Processing via Data Reduction

Abstract: With the proliferation of graph applications and the recent advent of Big Data, research efforts have been devoted towards many fundamental problems in managing and analysing large-scale graphs. In this talk, I will illustrate the power of using data reduction techniques for processing large real-world graphs. Firstly, data reduction techniques can be used as a preprocessing step to reduce the input graph instance (e.g., for the problem of densest subgraph computation). Secondly, data reduction is used as an intrinsic part of the algorithm design (e.g., for independent set and maximum clique computations).

Bio: Dr. Lijun Chang is a Senior Lecturer and ARC Future Fellow in the School of Computer Science at the University of Sydney. He received Bachelor degree from Renmin University of China in 2007, and Ph.D. degree from The Chinese University of Hong Kong in 2011. He worked as a Postdoc and then DECRA research fellow at the University of New South Wales from 2012 to 2017. His research interests are in the fields of big graph (network) analytics, with a focus on designing practical algorithms and developing theoretical foundations for massive graph analysis. He has co-authored two monographs, and published over 50 papers in top venues such as SIGMOD, KDD, PVLDB, ICDE, VLDB Journal, TKDE, and Algorithmica. tutorial co-chair for APWeb 2017 & MDM 2019 and publicity co-chair for ACM SIGSPATIAL 2017 & 2018.

Zhifeng Bao, RMIT University, Australia

Title: Making Sense of Massive Urban Data

Abstract: With the rapid development and wide deployment of sensors, Urban data is becoming ubiquitous. In this talk we will focus on how to utilize massive urban data, such as vehicles and passenger commuting records, to enhance data-driven policy and decision making in developing low cost, reliable, resilient and efficient solutions for public and private transport planning, ridesharing, and site selection. Hopefully, the above can contribute to designing more fine-grained metrics to evaluate and monitor the impact of the policy and decision made in various urban planning tasks.

Bio: Zhifeng Bao is an Associate Professor in Computer Science, RMIT University and an Honorary Senior Fellow at the University of Melbourne, Australia. His research interests include spatial data management and mining, data quality, and big data algorithm. He received his PhD from the CS Dept at NUS in 2011 and was the only recipient of the Best PhD Thesis Award in School of Computing. He regularly published at top venues across DB, IR and DM, such as SIGMOD, VLDB, KDD, SIGIR and WSDM. He serves the Associate Editor of PVLDB Vol 14, and was the PC Co-chair of WSDM19 Cup. Please refer to https://baozhifeng.net for more about him and http://civilcomputing.com/ for more about urban data analytics projects conducted by his team.

Farhana Choudhury, The University of Melbourne, Australia

Title: Spatio-Visual Query Processing

Abstract: The increasing availability of large-scale data from map services, such as Google Maps, OpenStreetMap inspire the applications involving visibility in spatial databases, as people explore and describe locations and experiences with a strong emphasis on visual senses. A key challenge is, the visibility of an object with respect to a viewer depends not only on their own locations, but also the on the other objects (visual obstacles) between them. This presentation will discuss the efficient indexing and query processing techniques based on the visibility of geo-spatial objects in the presence of obstacles.

Bio: Dr. Farhana Choudhury is a lecturer in the School of Computing and Information Systems at The University of Melbourne. She has received her PhD in 2018 from RMIT University, Melbourne. Her research focuses on spatial databases, data visualization, trajectory queries, and applying machine learning techniques to solve spatial problems.

Jianxin Li, Deakin University, Australia

Title: Anchored Vertex Exploration for Community Engagement in Social Networks

Abstract: User engagement has recently received significant attention in understanding decay and expansion of communities in social networks. However, the problem of user engagement hasn't been fully explored in terms of users' specific interests and structural cohesiveness altogether. Therefore, in this talk, the presenter will introduce one of his recent works published in IEEE ICDE 2020. It filled the gap by investigating the problem of community engagement from the perspective of attributed communities. Given a set of keywords W, a structure cohesive parameter k, and a budget parameter l, the objective is to find l number of users who can induce a maximal expanded community. Meanwhile, every community member must contain the given keywords in W and the community should meet the specified structure cohesiveness constraint k. This problem was formalised as best-Anchored Vertex set Exploration (AVE).

To solve the AVE problem, a Filter-Verify framework was developed by maintaining the intermediate results using multiway tree, and probe the best anchored users in a best search way. To accelerate the efficiency, they further design a keyword-aware anchored and follower index, and also develop an index-based efficient algorithm. The proposed algorithm can greatly reduce the cost of computing anchored users and their followers. Additionally, they present two bound properties that can guarantee the correctness of the solution. Finally, they demonstrate the efficiency of the proposed algorithms and index, as well as the effectiveness of attributed community-based community engagement model using five real-world datasets.

Bio: Dr Jianxin Li is an A/Professor in the School of IT, Deakin University. His research interests include social computing, query processing and optimization, and big data analytics. He has published 90 high quality research papers in top international conferences and journals, including PVLDB, IEEE ICDE, ACM WWW, AAAI, IEEE ICDM, EDBT, ACM CIKM, IEEE TKDE, The VLDB Journal, IEEE TII, and WWW Journal. His professional service can be identified by different roles in academic committees, e.g., the technical program committee members in ACM SIGMOD, PVLDB, AAAI, PAKDD, IEEE ICDM, and ACM CIKM; the journal reviewer in IEEE TKDE, ACM TKDD, WWW Journal and VLDB Journal; the proceeding chairs in DASFAA 2018, ADMA 2016 and ADC 2015; and the program committee chair in ADMA 2019, and the International Workshop on Social Computing 2017 and 2018; the tutorial chair in the 26th International Conference on WWW 2017; and the guest editors in international journals, such as IEEE Transactions on Industrial Informatics, Computational Intelligence, IET Intelligent Transport Systems, Complexity, Data Science and Engineering.