Research Track Program Chairs' Welcome

The KDD conference has seen remarkable growth since its origins as an IJCAI workshop in Detroit in 1989, evolving into a full-fledged research conference in 1995, underscoring the important role data mining as a field has played in extracting knowledge and actionable insights from vast troves of data that is being generated in the digital world around us. This year we received a very healthy number of 726 submissions to the research program, out of which 125 papers were accepted, for an aggregate acceptance rate of 17.2% (in line with the acceptance rate in recent years). The accepted papers were split between oral presentations (66 papers) and poster presentations (59 papers).

Among the academic conferences, the KDD conference has typically more of an emphasis on research motivated by real-world applications. It is important to keep in mind that it is this synergy of research in areas like algorithms, computational geometry, database, graph theory, machine learning, natural language processing, statistics, visualization and many others when applied to problems arising in diverse fields such as web, medicine, biology and marketing that drives our field forward, makes it vibrant and fun.

The breadth of topics covered in this year's research program is truly comprehensive, including social and information networks (8.5% of papers), graph mining (4.4%), classification (4.2%), big data (4.1%), recommender systems (4.1%), probabilistic methods (3.6%), user modeling (3.4%), web mining (3.2%), supervised learning (3%), unsupervised learning (3%), social media (2.8%), time series (2.5%), healthcare (2.5%), text mining (2.5%), and more. In particular, we observe an increase in papers related to diffusion and influence in social networks, as well as works on scaling algorithms to big data. Other emerging trends are multiple papers dealing with sampling and diversity. We are very fortunate to have four world-class keynote speakers this year spanning industry and academia, providing inspirational talks on cutting-edge techniques and issues in big data, online education, web search, optimization and data mining.

The process of whittling down the initial 726 submissions to the final set of 125 accepted papers required the coordination and time of a large number of willing volunteers. The program committee (PC) consisted of 298 reviewers (PC members) and 50 senior PC members. Selection of the PC members involved rigid quantitative constraints in order to achieve a well-balanced coverage of expertise matching an expected distribution of submissions' subject areas, as well as requiring seniority and freshness in the field.

Initially, 10 papers were rejected for technical reasons. In the first reviewing phase each submitted paper was automatically assigned to 3 reviewers (after a bidding process). After the first round of reviewing was complete, the papers were assigned to the senior PC members based on their bids. The senior PC members then initiated discussions for many of the papers, e.g., if there was significant divergence in scores among reviewers, or if a paper was on the borderline of being accepted. Following the discussion phase, the senior PC members provided a recommendation score and a meta-review for each paper. In the final phase, we (the program chairs) analyzed all of this information, starting with the obvious accept and reject decisions, and then gradually focused in more detail on the papers near the borderline, seeking additional reviews and input from the PC and senior PC members where appropriate. We also initiated a shepherding phase with 22 papers having the opportunity to fix mild issues we thought would be possible to address before they can be accepted. 21 of them were accepted after thorough revisions. The top rated papers were provided to an awards committee, chaired by Charles Elkan and Aristides Gionis, who (independently of any extra input from us) chose the best paper and best student paper awards. Finally, it is quite likely that in hindsight some worthy papers may have been rejected as part of this process - these errors are an unfortunate reality of modern computer science conferences, and hard to avoid when a very large number of decisions have to be made over a short time span based on a subjective reviewing process. Nevertheless, we, the PC chairs, are responsible for those unfortunate errors and welcome suggestions on the matter.

We conclude with some well-deserved words of thanks to the large supporting cast. We thank all authors for submitting their research to KDD 2013. We are extremely grateful to the PC members and senior PC members for volunteering their time to help in the reviewing process. Diligent peer review is an essential cornerstone for research progress - we thank you all for the many detailed and conscientious reviews that were generated. Last year's program chairs, Jian Pei and Deepak Agarwal, were generous in providing us with advice and expertise based on lessons learned in 2012. We would like to extend our appreciation to two student volunteers at the University of Texas at Austin, Cho-Jui Hsieh and Hsiang-Fu Yu, who helped at various stages of the process. We would also like to convey our sincere thanks to the staff who run the Microsoft CMT system - they were more than willing to answer questions and help us at all times during the submission, reviewing, and decision process. And finally we would like to thank all of our colleagues on the KDD 2013 organization committee who were a pleasure to work with and without whom we would not have such a wonderful conference.

Let the 19th ACM SIGKDD conference begin - we hope you enjoy it!

Inderjit S. Dhillon and Yehuda Koren
ACM SIGKDD 2013 Program Chairs