LARC BULLETIN

At the Research Frontier

Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data

One of the early research efforts that a group of researchers at LARC undertook shortly after the centre was launched, and before it had acquired much data, was to evaluate how the state-of-the-art of research would evolve in the presence of big data.

The new era has created opportunities for researchers to achieve high relevance and impact amid changes and transformations in how we study social science phenomena. With the emergence of new data collection technologies, advanced data mining and analytics support, and emerging statistical methods for social networks, there seem to be fundamental changes that are occurring with the research questions we can ask, and the research methods we can apply.

The contexts that are under study now include social networks and blogs, political discourse, corporate announcements, digital journalism, mobile telephony, home entertainment, online gaming, financial services, online shopping, social advertising, and social commerce – and much more. For example, one of the issues that is under study is the possibility of rapidly developing online experiments.

An article that addresses these issues, titled '*Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data*', and co-written by Robert KAUFFMAN (Deputy Director, LARC and Associate Dean (Research), School of Information Systems, Singapore Management University), and former LARC researchers, Rae CHANG (Research Scientist) and Young Ok KWON (Research Fellow), will appear in the journal, *Decision Support Systems*.

in this issue...

Technology advances and scientific convergence supports new knowledge discovery.

Business Issues

- Product bundling and pricing
- Effects of tech disruptions to channel management
- Geo-location-based consumer demand sensing
- Discovery of info for optimal service design
- Sensing of market sentiments in social networks
- New info on dynamics of competition

Consumer Issues

- Behavior-based consumer segmentation
- "Friends of friends" effects discovery
- Pricing and product design sensitivity
- Customer feedback and sentiments
- Churn management, service up-/cross-selling

Social Issues

- Social, political and electoral sentiments
- Human network communication patterns
- Social feedback on government policies
- Public goods, traffic and energy use management
- Support for education and human welfare

Business, Consumer and Social Issues in Computational Social Science.

Professor KAUFFMAN and his co-authors noted that the changing costs of data collection and the new capabilities that researchers have to conduct research that leverages micro-level, meso-level and macro-level data suggest the possibility of a scientific paradigm shift toward computational social science. The new thinking related to empirical regularities analysis, experimental design, and longitudinal empirical research further suggests that these approaches can be tailored for rapid acquisition of new knowledge from big data sets. This will allow business analysts and researchers to achieve frequent, controlled and meaningful observations of real-world phenomena.

A key contribution of the research is how the philosophy of science for data analytics should be changing in step with the times. The authors argued that the assertion by Chris ANDERSON of *Wired* magazine that 'theory no longer matters' is not right. The authors also contended that the scientific research process involving 'realism,' 'control' and 'generality' is being affected by much lower data acquisition costs and extraordinary advances in the methods that are being used for experimentation in social network and social media settings.
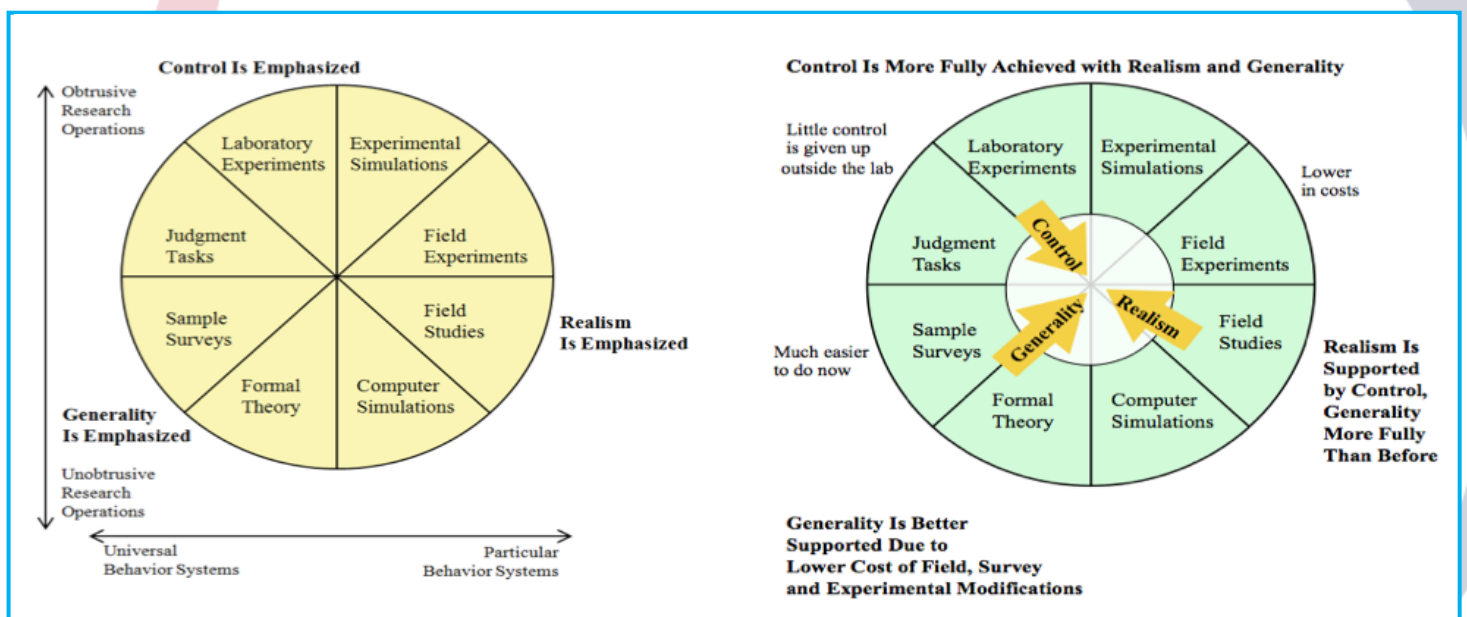
Realism, Control and Generality in Research Involving Data at Scale

Some of the new opportunities in research are being created with the emergence of new and high-tech experimental testbeds. One such example is the LiveLabs Urban Lifestyle Platform that was launched at SMU in October 2012. Researchers there are setting up new capabilities for conducting research involving mobile phones and information on the geolocations of their users. This has resulted in heretofore unseen levels of contextual awareness, and the opportunity to conduct closed-loop iterative experiments involving consumer incentives.

They also offer some new directions for research:

- **Explore the impacts of knowledge about contextual awareness**. Newly available contextual information can be collected inexpensively, and supports the analysis of new levels of contextual awareness on the part of consumers in many commercial settings.

- **Reassess the business value of personalization and extend the study to consumer and firm informed-ness, as well as hyper-differentiation**. The extraordinary amount of data that can be captured on consumers today creates dramatic changes in the basis for value maximization from the consumer's and the vendor's standpoints. Geolocation-based, and date and time-stamped data that capture information on human activities in a time-sequence (credit card transactions, mobile phone use geolocation info, etc.) can also support new valuable research.

- **Pursue big data studies that focus on patterns, explanations and predictions of events in business, consumer and social insight settings**. New means for digital sensing of events in various settings are available through technological advances that can produce behavioral and economic information at the individual, group, network, firm and societal levels. Event-focused analysis of when consumers make purchases and how they are influenced (by the setting, by an incentive, by friends in their social network, etc.) open up the possibility of the application of new research methods to discover valuable business intelligence also.

- **Conduct complementary research on information privacy issues related to big data sets in parallel with investigations of business, consumer and social insights**. The rise of big data and the new vulnerabilities make information privacy issues truly important. They affect people as individuals, in groups, in relation to private and public organizations, and with respect to government and society as a whole.

This research supports SMU's effort with thought leadership for its Area of Excellence on Analytics for Business, Consumer and Social Insights.

Sensing Haze in Social Media

The last haze that enveloped Singapore back in 2006 seemed a distant memory and 7 years on, the island experienced its worst spate of air pollution with the PSI (Pollution Standard Index) Index hitting a record high of 401! According to the Singapore's National Environment Agency (NEA), a PSI reading of over 300 is deemed "hazardous" and the aftermath was sufficient to force companies to scale-down their outdoor operations and ignite a frenzied rush for surgical and N95 masks! This brought to mind the smog that invaded Singapore way back in September 1997 where a PSI index of 226 was recorded. That record stood for 16 years before being surpassed by the now infamous reading of 401!

For almost 2 weeks in the month of June, there was extensive media coverage on the haze situation in Singapore, and it dominated the usual office chatter and neighborly greetings. The PSI Index took on new significance and soon, everyone was monitoring the hourly readings, and the latest news and tongue-in-cheek "hazy" images were shared relentlessly over social media. This gives rise to questions whether there is a way to measure the consciousness of Singaporeans during the haze to understand their thoughts and feelings? Is there a way to quantify the haze impact on human activity?

Combining the knowledge and expertise of researchers from SMU School of Information Systems and School of Social Science, the research team at LARC attempted to answer these questions by studying social media usage by Singapore users during the haze event so as to derive insights into people's reactions to the haze. Although the Twitter users are a select sample and the observations provided a limited perspective, the team acknowledges that these Twitter information offers a window into the psyche state of mind and could reveal stuff that ordinary observations cannot.

The study began with an overview of how the haze affected Twitter usage in general. In Figure 1, it was evident that the daily tweets from Singapore users surged on 17, 19 and 21 June when the PSI index hit peak numbers at 111, 321 and 401 respectively. Further analysis into tweets containing haze related keywords and hash tags (such as #SGHaze or #haze among other terms) confirmed our observations that the pattern of increased tweeting activity was due to the haze (Figure 2). Notably, haze-related tweets comprised 8 – 10% of all tweets generated and this is comparable to the largely popular K-pop tweets generated from the same group of Twitter users. When the haze began to clear out after 21 June, there was a noticeable decreasing trend of haze-related tweets but it continued to remain higher than the proportion before the haze.

Office workers walk through a haze blanketing the central business district in Singapore on 20 June 2013. (AP Photo/Joseph Nair) Photo courtesy of Channel NewsAsia
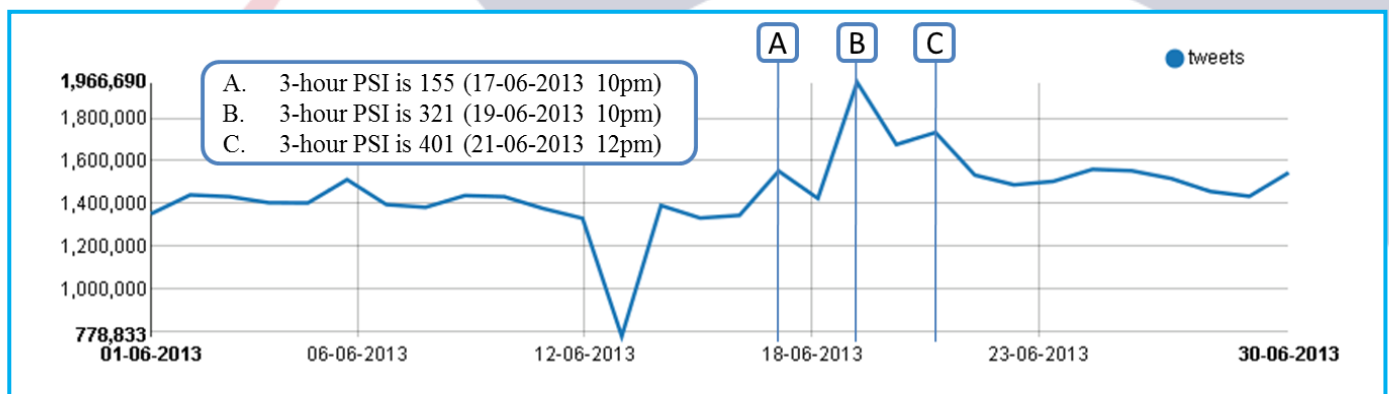

A.   3-hour PSI is 155 (17-06-2013 10pm)
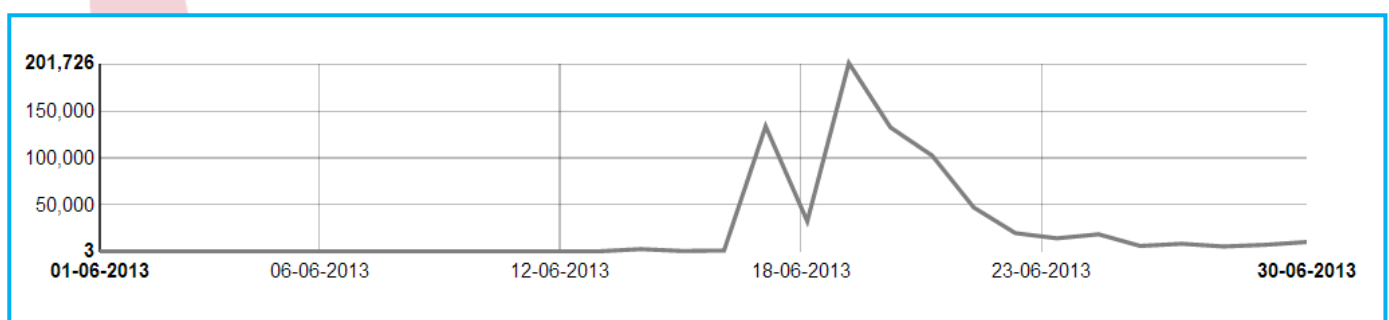B.   3-hour PSI is 321 (19-06-2013 10pm)
C.   3-hour PSI is 401 (21-06-2013 12pm)

Figure 1: Daily Tweet Count

Figure 2: Daily Count of Haze Related Tweets

The team further examined the tweets for emotion words, such as anxiety-related words (e.g. worried), anger-related words (e.g. irritated), and swear words/exclamation (e.g. damn, oh no!). Not surprisingly from Figure 3, the use of swear words and exclamations increased as the haze continued and worsened, peaking on 19 June, one of the days when the PSI hits record high!
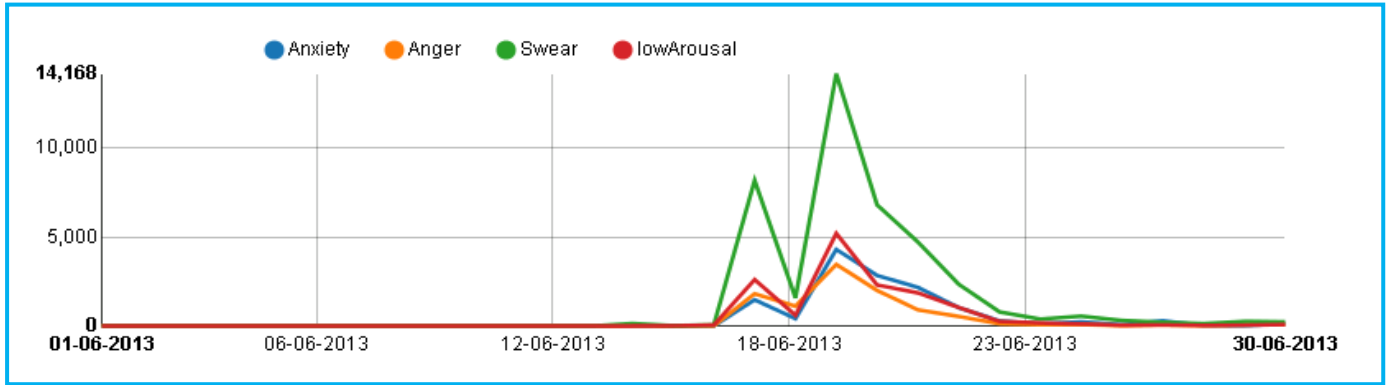

Figure 3: Daily Count of Emotional Tweets

URLs of external content also found haze-related tweets. The team observed that the most mentioned URL domain by haze-related tweets is straitstimes.com as its news content attracted the most tweets and re-tweets. This suggests that Twitter users were not only expressing their opinions, but also spreading and sharing the information as well. From the Twitter data, the team observed that NEA plays an important role in disseminating up-to-date haze information to the public. To determine how quickly the information gets disseminated, the team counted the number of re-tweets that mentioned the nea.gov.sg domain, and divided them into six disjoint time windows of 10 minutes each across time. Figures 4 – 6 depict the re-tweet count of each time window between 19 and 21 June respectively – the days that saw peaking PSI readings! The charts show the users' responses to NEA announcements by tweeting the URL within the first 10 to 20 minutes. This clearly illustrates how closely the users were tracking the updates from NEA.
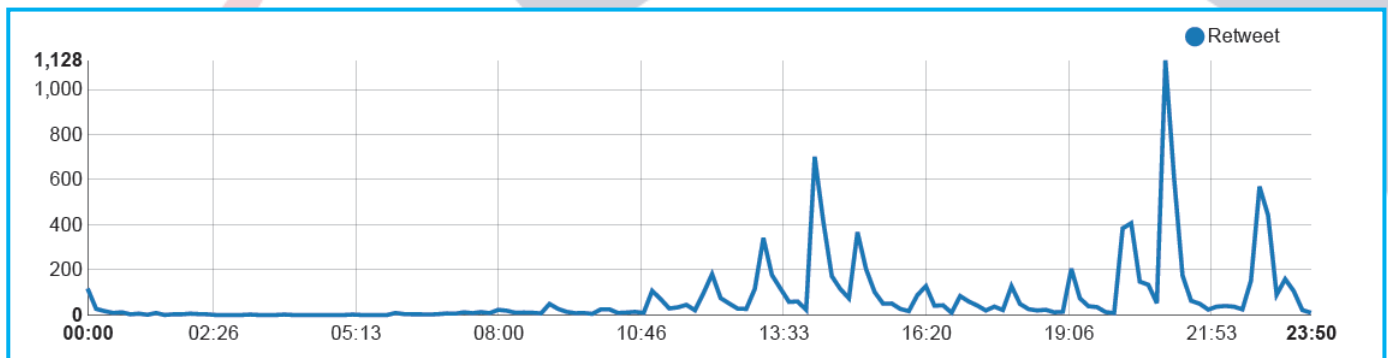

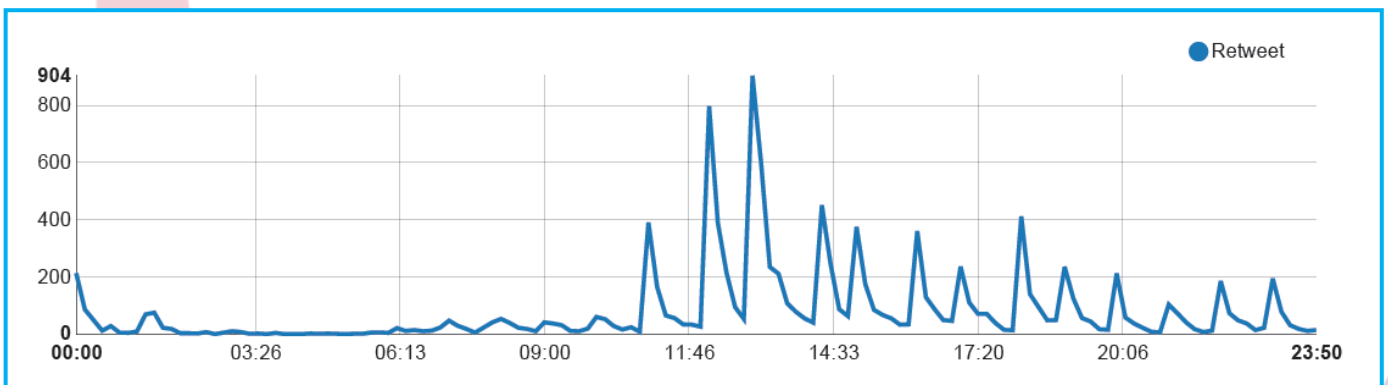*Figure 4: Retweet Count per 10 Minutes for NEA Tweets (June 19, 2013)*

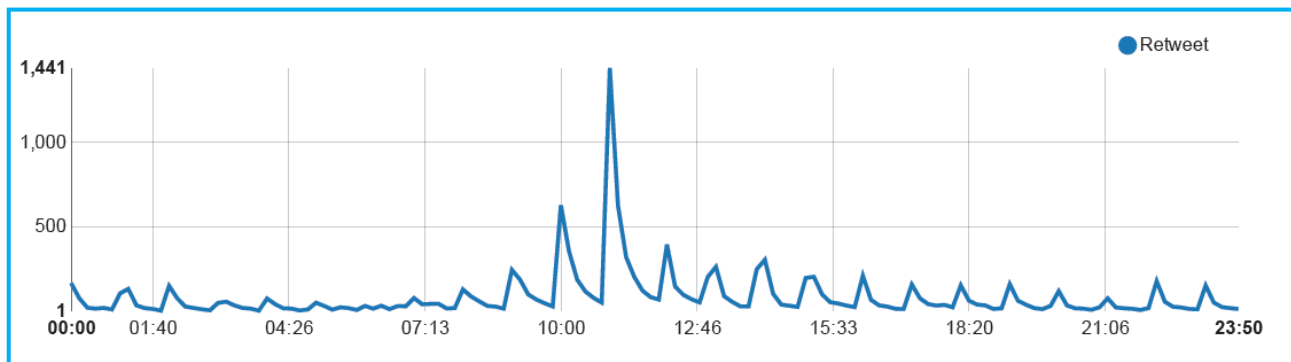*Figure 5: Retweet Count per 10 Minutes for NEA Tweets (June 20, 2013)*

Figure 6: Retweet Count per 10 Minutes for NEA Tweets (June 21, 2013)

In addition to analyzing Twitter feeds, the team examined geo-location data from FourSquare in order to quantify users' activities, in particular the haze effect on commerce. There was significantly lower number of FourSquare check-ins between 19 and 22 June 2013 and it provided clear evidence that visits to shops and eating places significantly reduced by 20 – 50% during those days that witness record-breaking PSI values. This observation corroborates with a Straits Times article published on 22 June 2013 where it reported that businesses, especially restaurants with outdoor seating suffered up to 60% drop in customer footfall.

While some may regard social media as a modern frivolity mainly as a source of entertainment, the LARC team has demonstrated how it could unearth useful and quantifiable information from analyzing Twitter feeds and FourSquare check-ins. Twitter provides a window into the stream of consciousness of Singaporeans unlike any other technology could offer at present. Geo-location data could quantify real activity to confirm or disconfirm personal observations.

Our analysis of Twitter feeds found that the impact of the haze on people's lives was undeniable and intense. People deliberately and drastically changed their daily routines by reducing their activities to food and shopping venues until the haze cleared out. People responded by turning to social media to express themselves, which primarily comprised of shock, anger, and other negative emotions. At the same time, people relied heavily on official sources of information about the haze, and they used social media to spread these information. In essence, social media is not all about entertainment and getting connected but one can definitely draw interesting insights from it.

More findings from the team's research can be found at http://research.larc.smu.edu.sg/sghaze.

Decision Analytics – Theory Meets Practice

Two years in the making, the **10th Metaheuristics International Conference (MIC)** was staged at the Singapore Management University (SMU) from 5 till 8 August 2013. MIC is a forum for the exchange of analytics methods and tools, high-impact and novel applications, new research challenges, theoretical developments, implementation issues, and in-depth experimental studies in metaheuristics, which cuts across Management Science, Operations Research, Computer Science and Data Analytics.

Walking down memory lane, the first MIC took place in Breckenridge, Colorado in 1995 and it was subsequently showcased at various countries around the world – Sophia-Antipolis (France), Angra dos Reis (Brazil), Porto (Portugal), Kyoto (Japan), Vienna (Austria), Montréal (Canada), Hamburg (Germany) and Udine (Italy), before settling its base on the fair shores of Singapore in 2013! Of special significance, this year marks the 10th anniversary of MIC and the Living Analytics Research Centre (LARC) is proud to be an organizer and host of the biennial conference!

Adding a special buzz to the event, the conference was held to coincide with Singapore's 48th National Day celebrations on 9 August 2013! As the nation celebrates its unwavering success story over the course of history, we take pride that the allure of MIC has continued to gain traction over the years and it has grown in size, to the extent that today, MIC has a large and global following of scholars and practitioners working on metaheuristics problems and leveraging its unique toolkit of interdisciplinary methods.

In celebration of the 10th anniversary, colleagues working in related disciplines were specially invited to participate in the four-day event, with the hope that the cross-disciplinary interactions would inspire new inter-disciplinary research collaborations among the participants. Special sessions on the interaction with Metaheuristics with Data Analytics, Constraint Programming, and Multi-agent Distributed Problem Solving, as well as emerging issues relating to Multi-Objective Optimization and Dynamic Vehicle Routing were organized during the conference.

Group shot of participants of MIC 2013.

In terms of technical content, 25 accepted full papers and 34 extended abstracts were presented at the conference. Evidently, these contributions suggest that the field of metaheuristics is wide-ranging in its coverage. A good number of papers focused on the applications in transport optimization, while other application areas such as network design; data analysis; scheduling; packing; and constraint satisfaction were also aptly represented. From an algorithmic point-of-view, a broad range of techniques was considered, established metaheuristics adapted, and new variants proposed. The contents will support interesting and rich discussions among the university and industry attendees.

Notably, the success of MIC 2013 would not be made possible without the presence of four exceptional international keynote speakers who covered a board spectrum of timely topics in metaheuristics, their applications and their links with other disciplines. Specifically, they shared on the following:

− Professor Panos PARDALOS, (University of Florida) on '*Data Mining and Optimization Heuristics for Massive Networks*'

− Professor Pablo MOSCATO (University of Newcastle) on '*Personalized Information-Based Medicine: Huge Challenges, Massive Opportunities and Some Lessons Learned*'

− Professor Steven KIMBROUGH (University of Pennsylvania) on '*Solution Pluralism, Deliberation and Metaheuristics*'

− Professor Michael TRICK (Carnegie Mellon University) on '*Combining Optimization and Metaheuristics in Sports Scheduling*'.

In conjunction with MIC 2013, an '*Industry Workshop on Supply Chain Analytics*' was organized and it brought together academics, analytics executives and supply chain leaders offering unique insights into innovations that are driving supply chain management success in organizations. The event showcased speakers from Singapore's Economic Development Board, Land Transport Authority, University of Hamburg in Germany, Indian Institute of Science at Bangalore, DHL Singapore, Toll Global Logistics, Caprica International Singapore, and TNT Express ICS Singapore. The participants of this workshop were able to discuss challenges and best practices for logistics and supply chain analytics, big data analytics, methods and tools, trends and visions in the use of analytics – a broad spectrum of interesting topics.

Group shot of the industry speakers of Industry Workshop on Supply Chain Analytics on 7 August 2013.

Fun at MIC 2013! Dinner with guests.

To all the 128 local and overseas participants, your active involvement and participation at the keynote sessions and seminar talks made MIC 2013 a remarkable and memorable event to behold. We certainly hope all participants left the conference with enriching knowledge, as well as "*panda-ramic*" sights from the River Safari Singapore! Look out for the 11th edition of MIC! For the first time in the conference history, the event will be held in the Continent of Africa – Tunisia! See you there in 2015!

Fun at MIC 2013! Group shot of participants of the social visit to River Safari on 6 August 2013.

Fun at MIC 2013! Conference Dinner at Hotel Fort Canning on 7 August 2013.

Workshop on Statistical and Machine Learning Approaches to Network Experimentation

The problem of designing valid networked experiments has challenged statisticians and analytics researchers across the social and computer sciences. Networks by their very nature involve dependence among the units comprising the network and these dependencies raise questions about the validity of traditional methods for statistical experimentation. For example, a standard assumption of experimentation is that the units are independent of one another. This is demonstrably false in networked environments.

In most settings of special interest to LARC, treatments allocated to units propagate via network links and dynamics over time. Thus one can also expect substantial violations of the stable unit treatment value assumption (SUTVA) and treatment interference. While one can ask what randomization buys us in experiments on networks, we can also query how one should design a network-based experiment with randomization so that the results can be statistically analyzed.

Prof Ramayya KRISHNAN [M], Dean, Heinz College, Carnegie Mellon University addressing and welcoming the guests.

Prof Stephen FIENBERG, Maurice Falk University Professor of Statistics and Social Science, Carnegie Mellon University, gave an opening speech to kick-start the workshop.

Let us see if this works? Prof Ramayya KRISHNAN [L], Dean, Heinz College, Carnegie Mellon University and Prof Steven MILLER [R], Dean, School of Information Systems, Singapore Management University.

Brian KARRER [L] and Dean ECKLES [R] (both from Facebook) enjoying the light continental breakfast served before the start of the workshop. 1Like!

Because these issues of network experiments lie at the heart of much of the work in LARC with our partners, LARC at Carnegie Mellon University hosted a two-day workshop on Statistical & Machine Learning Approaches to Network Experimentation on 22 – 23 April 2013. Partially funded by grants from Facebook and Microsoft Research, LARC co-director Stephen FIENBERG and Heinz College Dean Ramayya KRISHNAN brought together a distinguished collection of experts on network experiments and more traditional approaches to statistical experimentation to discuss current approaches and challenges. Among the speakers were the following LARC faculty:

LARC Faculty	Talk Title
David KRACKHARDT	Network Models and Dependence
Andrew THOMAS & Michael FINEGOLD	Protocols for Randomized Experiments to Identify Network Contagion
Ee-Peng LIM & Michael FINEGOLD	Reviving Dormant ties in an Online Social Network Experiment
Pedro FERREIRA	The Impact of Likes on the Sales of Movies in Video-on-Demand: A Randomized Experiment

Several other LARC researchers from both CMU and SMU were active participants in the workshop. They were augmented by others from economics, statistics and machine learning, as well as participants from the industry. Discussions focused on alternative approaches to designing real-world network experiments and evaluating their results. Several speakers focused on the need to infer latent network structure, including from transaction data that are distributed over time and space.

Sharing insights: Prof Stephen FIENBERG [L], Prof Ramayya KRISHNAN, Dean, Heinz College [M], and Prof David KRACKHARDT, Professor of Organizations, Heinz School of Public Policy and Management, Tepper School of Business , Carnegie Mellon University.

Call for Participation

Join Us!

Inviting Applications for the PhD Programme in Information Systems at SMU

The PhD Programme in Information Systems (IS) at Singapore Management University (SMU) produces graduates with the ability to address deep technological challenges in real-world settings that impact business processes and management. Our IS PhD students are equipped with skills to develop tools and methodologies to translate business goals into technological solutions. The faculty at the School of Information Systems (SIS) and LARC think of it as a 'game-changer for your future life in technology-related research'!

Prof Steven MILLER, Dean, School of Information Systems and Vice Provost (Research), SMU [1st row, R] and SIS PhD Students.

Students in our programme obtain their skills though a combination of course work involving depth and breadth research training, faculty-mentored empirical research projects, larger-scale research centre projects with private and public sector organizations, the building of innovative systems artifacts for all kinds of applied settings, teaching assistance in courses as they advance, and the publication of their research work in conferences, journals, books and other outlets.

LARC is an effective training base for students in our programme. Through LARC, IS PhD students can gain experience in the development and application of new data analytics involving social networks and social media, technology-based financial services marketing and operations, theme park operations and intelligent management decision-making, digital entertainment and mobile telephony, and information privacy in data-intensive settings. LARC supports study and research in all five of our PhD programme's niche research concentration areas: Information Security and Trust; Data Management and Analytics; Information Systems and Management; Software Systems, and Intelligent Systems and Decisions Analytics. Students have access to LARC's many research seminars, as well as academic and industry visitors, which further enriches the learning environment. You can also be as interdisciplinary as you want to, be it to cross Data Management and Analytics with Intelligent Systems and Decision Analytics; or to work on Information Security and Trust blended with Software Systems; or to mix one of the Analytics concentrations with IS and Management, and Marketing in the Lee Kong Chian School of Business. There is a wealth of opportunities awaiting your involvement.

Another unique aspect of the SMU IS Doctoral Programme is that our graduates are trained to collaborate with researchers from different disciplines to design technology solutions for real-world problems and applications, and to produce top-rated academic publications. The programme will equip you for careers with research and development laboratories, academic institutions, the business and consulting sectors, and other public and government institutions.

The SMU IS PhD programme admits students for full-time work in August and January each year. SMU awards competitive scholarships, as well as research and teaching assistantships each year to support doctoral study at the university.

To explore the qualifications, dates and deadlines, application requirements, and other information on our programme, please visit www.smu.edu.sg/programmes/doctoral/phdis. We also welcome you to make inquiries with our faculty and students!

LARC-IMS Workshop on Living Analytics: Analyzing High-Dimensional Behavioral and Other Data from Dynamic Network Environments I

Together with the Institute for Mathematical Sciences (IMS) at the National University of Singapore, LARC will be organizing a two-day workshop on **Living Analytics: Analyzing High-Dimensional Behavioral and Other Data from Dynamic Network Environments** from 27 – 28 February 2014. The organizing committee comprised of LARC Co-Directors, Professors Stephen FIENBERG and LIM Ee-Peng, and LOH Wei-Liem, from NUS Department of Statistics and Applied Probability.

"Big data" come in a variety of forms, but especially common are large high-dimensional electronic databases, e.g., involving transactions or genetic and related medical data, where the units of interest exist in a networked environment. Analytical statistical methods and machine learning tools for analyzing such data involve finding low-dimensional representations and the direct exploitation of the network links to understand dependencies among units.

The joint workshop seeks to facilitate the introduction of current living analytics research activities to other mathematical science, machine learning, and statistical researchers in Singapore. It serves to broaden statistical research underpinnings of models and computational algorithms for living analytics research and other research activities involving the analysis of large high-dimensional databases. The workshop will provide a forum for scientists to interact and develop methodology for "big data" problems using living analytics as a focal point, as well as chart new directions of research and explore possible collaborations.

The workshop will be hosted at Ngee Ann Kongsi Auditorium at SMU (on 27 February) and IMS (on 28 February). Registration is now open and more information on the workshop can be found at http://www2.ims.nus.edu.sg/Programs/014wliv/index.php.

The workshop is an initial effort of the joint collaboration between LARC and IMS. Look out for a one-week program, an extension of the workshop to be held in February 2015!

LARC/CMU Welcomes Dr. JIANG Binyan

Binyan JIANG, PhD, recently joined the research team at LARC/CMU to work with Professor Stephen FIENBERG on multiple LARC-related projects.

Dr JIANG received his Bachelor of Science degree in Statistics and Finance from the University of Science and Technology of China in July, 2007 and his PhD in July, 2012 from the National University of Singapore. His research interests include high dimensional data analysis; regularized statistical methods; multivariate analysis; model selection and survival data analysis. Prior to joining the LARC/CMU team, Dr JIANG worked as a post doctoral fellow in the Department of Statistics and Applied Probability at the National University of Singapore. He is a published author with papers submitted to professional journals such as *Biometrika*, *Biostatistics*, and *Statistics and Probability Letters*. He also has an impressive list of conference presentations and has been awarded numerous research and teaching awards during his tenure at the National University of Singapore.

The LARC/CMU team is happy to have Dr JIANG working with the CMU faculty/staff/students. Dr JIANG is located in the Heinz College (Hamburg Hall) and can be contacted at binyanj@andrew.cmu.edu.

Welcome onboard, Dr JIANG!

Dr JIANG Binyan, Visiting Research Scientist, Heinz College, Carnegie Mellon University.

LARC Accolades

## Best Demo Award (Runner-up) at DASFAA'13

The Twitalytics team from SMU won the centre's first piece of silverware – A runner-up placing for the Best Demo Award at the 18th International Conference on Database Systems for Advanced Application (DASFAA'13) in Wuhan, China. The winning team comprised of DU Juan, XIE Wei, CHENG Li, Assistant Professor Feida ZHU and Professor Ee-Peng LIM.

DASFAA is an annual international database conference, which showcases state-of-the-art R&D activities in database systems and their applications. The DASFAA 2013 demo track aimed to provide an exciting platform for researchers, developers and database users from the data management and Web communities to showcase leading-edge work and work-in-progress in every area of databases, Web, and their applications.

Chosen among 9 entries, the team's work, titled "*TwiCube: A Twitter Follow Community Analysis Tool*" described the development of an online tool that employs a novel algorithm capable of identifying a user's real-life social community, also known as the user's off-line community, purely from examining the link structure among the user's followers and followees.

The identification of a user's off-line community is important in characterizing different users and understanding their behaviour on Twitter. Notably, TwiCube not only finds a target user's off-line community in a real-time manner, it also provides further analysis of the user's off-line community by generating a summary of the user's interests, tweeting habits and neighbourhood popularity. Through manual evaluations of real Twitter users, there was clear evidence demonstrating that the team's off-line community detection approach achieves a high degree of precision and recall in most cases.

Our heartiest congratulations to the team for their achievements! Receiving this award attests to the intriguing research conducted and also the stimulating demos developed in LARC.
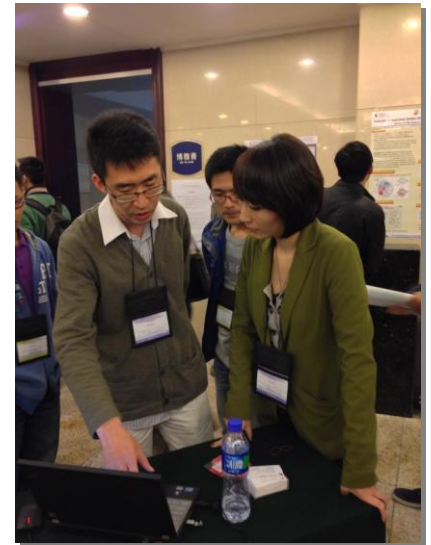
DU Juan, LARC Research Engineer showcasing her research works to participants at DASFAA 2013.

LOADING...

Recipients of 2013 SMU Presidential Doctoral Fellowship

Congratulations to Swapna GOTTIPATI and HOANG Tuan Anh who were awarded the SMU Presidential Doctoral Fellowship in July 2013. The award recognizes exceptionally qualified SMU PhD students who are outstanding in their studies. The Fellowship is a one-year award. At the end of the award term, a recipient of the Fellowship may be offered another Fellowship award or a Full Scholarship.

Over the next one year, Swapna and Tuan Anh stand to receive a higher monthly stipend and they could be sponsored for a research trip during the award period. Incidentally, both students participated in the LARC overseas training residency programme at Carnegie Mellon University and they completed their stints in December 2012 and May 2013 respectively.

LOADING...

Best Research Paper Award at CSWIM'13

A+ Excellent!

S CSWIM 2013

Good news on the research front from Tianjin, China!

The LARC Consumer Insights Research Group won the '**Best Research Paper Award**' at the 7th China Summer Workshop on Information Management (CSWIM'13). The workshop was held from 29 till 30 June 2013 in Tianjin, China. Dean ZHANG Wei from the College of Management and Economics of Tianjin University, together with other technology research leaders from the Chinese Academy of Sciences' Institute of Automation, the University of Arizona and the University of Delaware hosted the event. Over 140 participants attended the workshop, with about 100 from China and 40 from North America.

The research theme for CSWIM'13 focused on 'Big Data Analytics and Emerging Applications'. The Call for Papers encouraged the submission of 'innovative and thought-provoking' work on 'issues concerning the theory, design, development, evaluation and application of information systems and management'. The workshop also sought to include research that solved a 'real-world business problem' with validation by 'proper research methodologies.' Our research in LARC seemed to fit their specifications this year very well!

The award winning paper is titled "*Do Household Cable TV Viewing Patters Demonstrate Efficiency and Concentration?*" It was co-authored by Peiran ZHANG (Research Assistant, LARC, and visiting doctoral student from the business school at Fudan University), JUNG Gwangjae (Research Fellow, LARC), Rob KAUFFMAN (Deputy Director, LARC and Associate Dean of Research, School of Information Systems at Singapore Management University), Rae CHANG (Research Scientist, LARC) and Pulak GHOSH (Visiting Professor, School of Information Systems at Singapore Management University and Professor of Statistics, Indian Institute of Management Bangalore).

The research was based on joint work with StarHub's Customer Lifecycle Management Group, including StarHub colleagues Timothy GOH (Vice President), Darren CHOO (Assistant Vice President), Adrian KOH (Senior Manager) and TAN Yoke Kum (Data Mining Manager). They supported LARC's research every step along the way.

The research sought to understand household-level patterns for cable TV services consumption. It creates the basis for greater value for the stakeholders involved: consumers through higher viewership satisfaction, and providers through higher average revenues per user. With detailed data on many households in Singapore, the team was able to harvest the 'digital traces of in-home TV viewing,' and leverage it to create business and consumer insights.

*StarHub

The LARC research team explored a large data set on cable TV services subscriptions and viewing at the household level of analysis. They constructed household viewership preference clusters, and then used econometric methods to assess the relative efficiency and concentration of channel viewership patterns.

Methods-wise, the team estimated a system of 'limited dependent variable models' for efficiency and concentration, which vary in their values between 0 and 1 – so-called 'limited dependent variables'. For the modeling estimation, the team used a beta distribution for efficiency and concentration. They also used another approach called 'quasi-likelihood function-based estimation'. It exploits the asymptotic requirements of the model for the large data sets that were used to produce meaningful estimates of the main effects variables. The second approach is also notable because it does not employ a fully-specified distribution for the dependent variables. This gave the researchers some modeling flexibility and allowed them to implement their key perspective to 'let the data speak for themselves.' They also applied quantile regression and piecewise spline regression to further explore the data, which provided more fine-grained insights.

The main findings from the research suggest that household-level cable TV viewing patterns are affected by the household's subscribed bundle of programs, program genre preferences, available time for viewing, and demographics. The research team used the findings as a basis for drawing conclusions about how service providers can redesign their program offerings to enhance the consumer's experience with cable TV services.

LARC would like to thank the leadership of CSWIM'13 for selecting this research for its 'Best Research Paper Award.' You are welcome to contact Dr JUNG Gwangjae (gwangjaejung@smu.edu.sg) if you would like to get in touch with the research team to further discuss this piece of work.

Photos courtesy of CSWIM 2013.

LOADING...

Travel Fellowship to 'Force 11: The Future of Research Communications and e-Scholarship' Event in the Netherlands

Visiting doctoral student Peiran ZHANG from Fudan University in Shanghai attended the 'Beyond the PDF2 Conference,' as a student representative of the School of Information Systems and the Living Analytics Research Centre. The event was held at Pakhuis de Zwijger in Amsterdam, Netherlands on 18 – 19 March 2013.

The conference call for participation indicated that the event was intended to explore issues related to scholarly communication across multiple disciplines that are being rapidly changed under the influence of new technologies. New models, tools and standards are being developed that aim to enhance, enable or entirely replace prior approaches to communication, including publications, courses, conferences and policies. The conference brought together scholars, librarians, archivists, publishers, graduate students, and research funders in a discussion forum to encourage broader awareness of current efforts across disciplines, and define the future of modern scholarly communications through the effective use of information technology.

Professor Robert KAUFFMAN, Associate Dean (Research) at the School of Information Systems, Singapore Management University, acting in the role of Editor-in-Chief for *Electronic Commerce Research and Applications* (an Elsevier journal), nominated Peiran for a travel fellowship worth about S$2,000. The publisher was interested to reach out to universities in Asia, including China and Singapore, and to especially encourage doctoral students to participate. Since one of the key areas of change is related to the publication of research that involves the creation of large data sets, it was natural that Elsevier would be interested to invite a junior colleague from LARC who had some experience with big data.

Participants at the event included some of the following, in addition to Singapore Management University: Stanford University, UCLA, Harvard Medical School, the Public Library of Science (PLOS), University of Oxford, Wolfram Research, SSRN, UC Berkeley, Columbia University Libraries, Lawrence Berkeley National Laboratory, University of Cambridge, Carnegie Mellon University, Mendeley, Springer, Elsevier, Erasmus University Medical Center, and other digital media, publishing, labs, societies and research organizations.

Prior to Peiran's trip, he had a chance to be briefed by SMU's University Librarian, Gulcin CRIBB, on a variety of related issues that would be discussed in Amsterdam. They included current changes that are underway in scholarly communications, the structure of the market for academic publishing, the new technologies that are resulting in the current pressures, and other issues, including research article and data set peer reviewing and various citation tracking mechanisms.

A number of interesting initiatives were discussed at the conference, including why an alternative to Google Scholar is needed, principles and practices for data set citation and curation, audio visual resources that will transform academic publishing, and other alternatives to publishing and scholarly document sharing that go beyond the PDF. Program information for the conference is available at www.force11.org/abstracts, and the homepage, which is worthwhile to explore, is at www.force11.org/beyondthepdf2.

In addition to the conference, Peiran was able to take in some sights in Amsterdam. He visited places such as the Van Gogh Museum, the Heineken Brouwerijen Amsterdam (brewery and museum), the Amsterdam Museum. He also saw all the bicycles, the canals and the lovely Dutch architecture from centuries gone by. This was Peiran's first trip to Europe, and surely a memorable one.

Many thanks to Elsevier in Amsterdam for sponsoring Peiran's travel to the conference! And to the Publisher of Electronic Commerce Research and Applications, Gaia LUPO, for bringing this opportunity to our attention.

Looks like Peiran had fun!
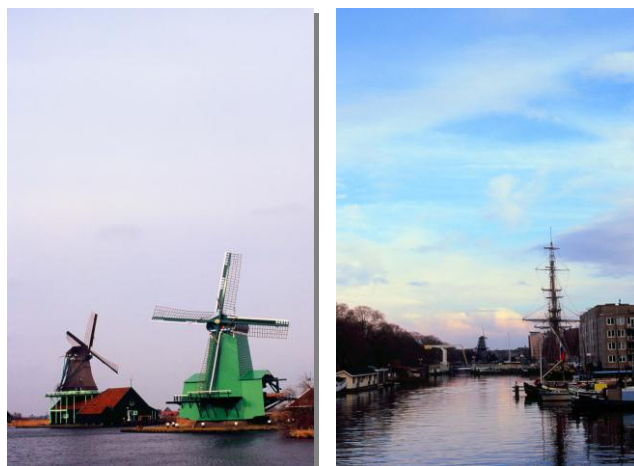
Photos courtesy of Peiran ZHANG, LARC Research Assistant

LOADING...

Past Activities in LARC

Seminars

- **4 Apr 2013:** Ontologies and Information Extraction by Fabian M. SUCHANEK (Max-Planck Institute for Informatics) and Nicoleta PREDA (University of Versailles Saint-Quentin-en-Yvelines)

- **20 May 2013:** Retrieving and Analyzing Location-Aware Information by Sensing Data by Xiaonan GUO (Hong Kong University of Science and Technology)

- **3 Jun 2013:** Multivariate Density Forecast Evaluation: Smooth Test Approach (Job Market Paper) by Stanley KO (Chinese University of Hong Kong)

- **14 Jun 2013:** Quantifying Privacy Loss in (LB)OSNs: from Personal Attributes to Geolocation Trajectories by Roksana BORELI (National ICT Australia)

- **25 Jun 2013:** Flexible QoS Management in Web Services Orchestrations by Ajay KATTEPUR (French National Institute for Research in Computer Science and Control)

- **15 Aug 2013:** High-level Search: From Hyper-heuristics to Algorithm Selection by Mustafa MISIR (French National Institute for Research in Computer Science and Control)

- **30 Aug 2013:** Crowdsourcing Data Analytics by Meihui ZHANG (National University of Singapore)

- **4 Sep 2013:** Media-aware Quantitative Trading Based on Public Web Information by Zhangxi LIN (Texas Tech University)

- **20 Sep 2013:** Probabilistic Latent Space Modeling of Large-Scale Networks via Triangle Motifs by Qirong HO (Carnegie Mellon University)

Workshops

- **28 May 2013:** Analytics for Business and Consumers Insights for the Use of IT in Telecommunications and Financial Services

- **1 Aug 2013:** Social Network Analysis and Mining

- **2 Aug 2013:** Mobile and Networking Systems

Recent and Upcoming Events

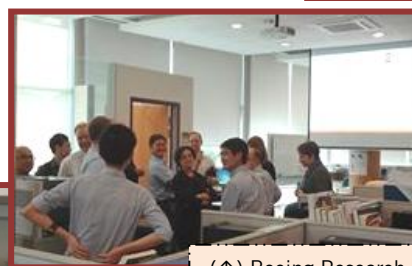|	
BCSI 2013	2nd Annual Workshop on Analytics for Business, Consumer and Social Insights (BCSI 2013) 3 - 5 August 2013 Singapore
MIC 2013	10th Metaheuristics International Conference (MIC 2013) 5 – 8 August 2013 Singapore
SNA KDD 2013	7th International Workshop on Social Network Mining and Analysis (SNA-KDD 2013) 11 August 2013 Chicago, United States *A LARC-sponsored international event*
SocInfo 2013	5th International Conference on Social Informatics (SocInfo 2013) 25 – 27 November 2013 Kyoto, Japan *A LARC-sponsored international event*
LARC / NUS / IMS Institute for Mathematical Sciences	Workshop on Living Analytics: Analyzing High-Dimensional Behavioral and Other Data from Dynamic Network Environments 26 – 28 February 2013 Singapore *Jointly organized by LARC and the Institute for Mathematical Sciences (IMS)*

Scene@LARC

LARC Seminars & Workshops

LARC Posters/Demos Presentations

Visitors to LARC

(←)(↓) Nielsen Measurement Institute

(↓) Finalists of Lee Kuan Yew Global Business Plan Competition

(↑) Boeing Research and Technology and The National Academies
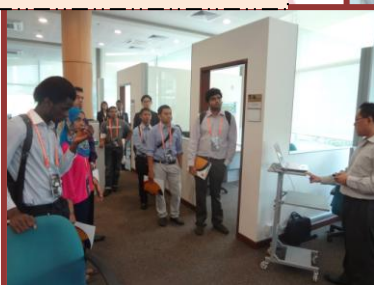
(↑) Carnegie Mellon University

(↑) Interactive Digital Media Programme Office

Flying Through LARC

Since April of 2013, we have had the privilege of hosting the following academics at LARC...

LARC Visitors Badge

Carnegie Mellon University

Michael FINEGOLD
Visiting Assistant Professor, Department of Statistics
Period of Visit:
2 Jul 2012 –
25 Jun 2013

LIU Siyuan
Research Scientist
Heinz College
Period of Visit:
1 Feb 2013 –
31 Jan 2014

Ramayya KRISHNAN
Dean, H. John Heinz III College
William W. And Ruth F. Cooper Professor of Management Science and Information Systems
Period of Visit:
14 May 2013

Michael TRICK
Senior Associate Dean Education and Professor of Operations Research
Tepper School of Business
Period of Visit:
8 Jul – 9 Aug 2013

Srinivasan SESHAN
Professor
School of Computer Science
Period of Visit:
13 Jul – 19 Aug 2013

Alejandro ZENTNER
Visiting Assistant Professor
Heinz College
Period of Visit:
17 Jul – 13 Aug 2013

Pedro FERREIRA
Research Assistant Professor of Economics of Information Networks and Public Policy
Period of Visit:
21 Jul – 4 Sep 2013

Alan MONTGOMERY
Associate Professor of Marketing
Tepper School of Business
Period of Visit:
22 Jul 2013

David KRACKHARDT
Professor of Organizations
Heinz School of Public Policy and Management
Tepper School of Business
Period of Visit:
26 Jul – 6 Aug 2013

Stephen FIENBERG
Director, LARC
Maurice Falk University Professor of Statistics and Social Science, Department of Statistics, Machine Learning Department, Heinz College, and Cylab
Period of Visit:
29 Aug – 3 Sep 2013

HO Qirong
PhD Student
Machine Learning Department
School of Computer Science
Period of Visit:
20 Sep 2013

Other Universities

1. **Latifur KHAN,** Professor, The University of Texas at Dallas
(Period of Visit: 19 Apr 2013)

2. **WANG Ke,** Professor, Simon Fraser University
(Period of Visit: 1 May 2013 – 30 Apr 2014)

3. **Alan S. ABRAHAMS,** Assistant Professor, Virginia Tech
(Period of Visit: 27 May 2013)

4. **David FUHRY,** PhD Candidate, The Ohio State University
(Period of Visit: 10 Jun – 2 Aug 2013)

5. **ZHANG Meihui,** Research Assistant, National University of Singapore *(Period of Visit: 30 Aug 2013)*

6. **LIN Zhangxi,** Associate Professor, Texas Tech University
(Period of Visit: 4 Sep 2013)

LARC Publications (Apr – Sep 2013)

Conference Papers

1. Kyriakos MOURATIDIS and Hwee Hwa PANG. Computing Immutable Regions for Subspace Top-k Queries. 39th International Conference on Very Large Data Bases (VLDB'13), Trento, Italy, August 2013.

2. Tuan-Anh HOANG, William W. COHEN, Ee-Peng LIM, Doug PIERCE and David P. REDLAWSK. Politics, Sharing and Emotion in Microblogs. 5th International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13), Niagara Falls, Canada, August 2013.

3. Truc Viet LE and Minh Thap NGUYEN. An Empirical Analysis of a Network of Expertise. 5th International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13) in Conjunction with International Workshop on Web Behavior Analytics (WBA'13), Niagara Falls, Canada, August 2013.

4. Youngsoo KIM. The User's Communication Patterns on A Mobile Social Network Site. 7th International Workshop on Social Network Mining and Analysis (SNA-KDD'13), Chicago, United States, August 2013.

5. Ajay ARAVAMUDHAN, Archan MISRA and Hoong Chuin LAU. "Network-Theoretic" Queuing Delay Estimation in Theme Park Attractions. 9th IEEE International Conference on Automation Science and Engineering (CASE'13), Madison, Wisconsin, USA, August 2013.

6. Kar Way TAN, Wei Hao TAN and Hoong Chuin LAU. Improving Patient Length-of-Stay in Emergency Department through Dynamic Resource Allocation Policies. 9th International Conference on Automation Science and Engineering (CASE'13), Madison Wisconsin, USA, August 2013.

7. Siyuan LIU, Yisong YUE and Ramayya KRISHNAN. Adaptive Collective Routing Using Gaussian Process Dynamic Congestion Models. 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM SIGKDD'13), Chicago, USA, August 2013.

8. Cen CHEN, Shih-Fen CHENG and Hoong Chuin LAU. Multi-Agent Orienteering Problem with Time-Dependent Capacity Constraints. 10th Metaheuristics International Conference (MIC'13), Singapore, August 2013.

9. Truong-Huy D. NGUYEN, Pradeep VARAKANTHAM, Hoong Chuin LAU and Shih-Fen CHENG. Interacting Knapsack Problem in Designing Resource Bundles. 10th Metaheuristics International Conference (MIC'13), Singapore, August 2013.

10. LINDAWATI, Feida ZHU and Hoong Chuin LAU. FloTra: Flower-shape Trajectory Mining for Instance-specific Parameter Tuning. 10th Metaheuristics International Conference (MIC'13), Singapore, August 2013.

11. Zhi YUAN, Thomas STUTZLE, Marco A. MONTES DE OCA, Hoong Chuin LAU and Mauro BIRATTARI. An Analysis of Post-Selection in Automatic Configuration. 22nd Genetic and Evolutionary Computation Conference (GECCO'13), Amsterdam, Netherlands, July 2013.

12. Ee-Peng LIM, Denzil CORREA, David LO, Michael FINEGOLD and Feida ZHU. Reviving Dormant Ties in an Online Social Network Experiment. 7th International AAAI Conference on Weblogs and Social Media (ICWSM'13), Boston, USA, July 2013.

13. Rae CHANG, Pulak GHOSH, Gwangjae JUNG, Robert J. KAUFFMAN and Peiran ZHANG. Do Household Cable TV Viewing Patterns Demonstrate Efficiency and Concentration? 7th China Summer Workshop on Information Management (CSWIM'13), Tianjin, China, June 2013.

14. Azeem J. KHAN, Kasthuri JAYARAJAH, Dongsu HAN, Archan MISRA, Rajesh BALAN and Srinivasan SESHAN. CAMEO: A Middleware for Mobile Advertisement Delivery. 11th International Conference on Mobile Systems, Applications and Services (MobiSys'13), Taipei, Taiwan, June 2013.

15. Feida ZHU, Zequn ZHANG and Qiang QU. A Direct Mining Approach To Efficient Constrained Graph Pattern Discovery. 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD'13), New York, USA, June 2013.

16. Yueqiang CHENG and Xuhua DING. Guardian: Hypervisor as Security Foothold for Personal Computers. 6th International Conference on Trust & Trustworthy Computing (TRUST'13), London, United Kingdom, June 2013.

17. Xun WANG, Feida ZHU, Jing JIANG and Sujian LI. Real Time Event Detection in Twitter. 14th International Conference on Web-Age Information Management (WAIM'13), Beidaihe, China, June 2013.

18. Siyuan LIU, Miguel ARAÚJO, Emma BRUNSKILL, Rosaldo ROSSETTI, João BARROS and Ramayya KRISHNAN. Understanding Sequential Decisions via Inverse Reinforcement Learning. 14th IEEE International Conference on Mobile Data Management (IEEE MDM'13), Milan, Italy, June 2013.

19. Jiansu PU, Siyuan LIU, Ye DING, Huamin QU and Lionel NI. T-Watcher: A New Visual Analytic System for Effective Traffic Surveillance. 14th IEEE International Conference on Mobile Data Management (IEEE MDM'13), Milan, Italy, June 2013.

20. Ye DING, Siyuan LIU, Jiansu PU and Lionel M. NI. HUNTS: A Trajectory Recommendation System for Effective and Efficient Hunting of Taxi Passengers. 14th IEEE International Conference on Mobile Data Management (IEEE MDM'13), Milan, Italy, June 2013.

21. Yan LI, Yingjiu LI, Qiang YAN and Robert H. DENG. Think Twice before You Share: Analyzing Privacy Leakage under Privacy Control in Online Social Networks. 7th International Conference on Network and System Security (NSS'13), Madrid, Spain, June 2013.

22. Divyan Munirathnam KONIDALA, Robert H. DENG, Yingjiu LI, Hoong Chuin LAU and Stephen E. FIENBERG. Anonymous Authentication of Visitors for Mobile Crowd Sensing at Amusement Parks. 9th International Conference on Information Security Practice and Experience (ISPEC'13), Lanzhou, China, May 2013.

23. Shih-Fen CHENG, Larry LIN, Jiali DU, Hoong Chuin LAU and Pradeep VARAKANTHAM. An Agent-Based Approach to Dynamic Experience Management in Theme Parks. 12th International Conference on Autonomous Agents and Multiagents Systems (AAMAS'13), Minnesota, USA, May 2013.

24. Duc Thien NGUYEN, William YEOH and Hoong Chuin LAU. Distributed Gibbs: A Memory-Bounded Sampling-Based DCOP Algorithm. 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'13), Minnesota, USA, May 2013.'

25. Ming GAO, Ee-Peng LIM and David LO. R-Energy for Evaluating Robustness of Dynamic Networks. 5th ACM International Conference on Web Science (WebSci'13), Paris, France, May 2013.

26. Minghui QIU, Feida ZHU and Jing JIANG. It Is Not Just What We Say, But How We Say Them: LDA-based Behavior-Topic Model. 2013 SIAM International Conference on Data Mining (SDM'13), Texas, USA, May 2013.

27. Tuan-Anh HOANG and Ee-Peng LIM. Retweeting: An Act of Viral Users, Susceptible Users, or Viral Topics? 2013 SIAM International Conference on Data Mining (SDM'13), Texas, USA, May 2013.

28. Juan DU, Wei XIE, Cheng LI, Feida ZHU and Ee-Peng LIM. TwiCube: A Real-time Twitter Offline Community Analysis Tool. 18th International Conference on Database Systems for Advanced Applications (DASFAA'13), Wuhan, China, April 2013.

29. Juan DU, Feida ZHU and Ee-Peng LIM. Dynamic Label Propagation in Social Networks. 18th International Conference on Database Systems for Advanced Applications (DASFAA'13), Wuhan, China, April 2013.

30. Siyuan LIU, Ramayya KRISHNAN, Emma BRUNSKILL and Lionel M. NI. Modeling Social Information Learning among Taxi Drivers. 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'13), Gold Coast, Australia, April 2013.

Journal Papers

1. Siyuan LIU, Jiansu PU, Qiong LUO, Huamin QU, Lionel M. NI and Ramayya KRISHNAN. VAIT A Visual Analytics System for Metropolitan Transportation, PP(99), IEEE Transactions on Intelligent Transportation Systems, (2013).

2. Siyuan LIU, Yunhuai LIU, Lionel M. NI, Minglu LI and Jianping FAN. Detecting Crowdedness Spot in City Transportation, 62(3), IEEE Transactions on Vehicular Technology (IEEE TVT), (2013).

3. Alessandro RINALDO, Sonja PETOVIC* and Stephen E. FIENBERG. Maximum Likelihood Estimation in the β-Model, 41(3), Annals of Statistics, (2013).