

Conference Committee

General Chair

John G. Breslin (National University of Ireland, Galway, Ireland)

Program Cochairs

Nicole B. Ellison (Michigan State University, USA)

James G. Shanahan (Church and Duncan Group Inc., USA)

Zeynep Tufekci (University of North Carolina, Chapel Hill, USA)

Local Cochairs

Derek Greene (University College Dublin, Ireland)

Conor Hayes (Digital Enterprise Research Institute, Ireland)

Publicity Chair

Max L. Wilson (Swansea University, United Kingdom)

Demonstrations Chair

Alejandro Jaimes (Yahoo!, Spain)

Sponsorship Chair

Meenakshi Nagarajan (IBM Research, USA)

Workshops Chair

Sofus A. Macskassy (USC Information Sciences Institute, USA)

Tutorials Chair

Bernie Hogan (Oxford Internet Institute, United Kingdom)

Data Challenge Cochairs

Derek Ruths (McGill University, Canada)

Ian Soboroff (National Institute of Standards and Technology, USA)

Industry Chair

Sheila Kinsella (Datahug, Ireland)

Webmaster

Ritesh Agrawal (AT&T Labs Research, USA)

Regional Chair, Americas

Raquel Recuero (Universidade Católica de Pelotas, Brazil)

Regional Chair, Asia Pacific

Hideaki Takeda (National Institute for Informatics, Japan)

Regional Chair, Europe

Jan-Hinrik Schmidt (Hans-Bredow-Institute, Germany)

Regional Chair, Middle East and Africa

Sihem Amer-Yahia (Qatar Computing Research Institute, Qatar)

2012 Senior Program Committee

Lada Adamic (University of Michigan, USA)
Eugene Agichtein (Emory University, USA)
Ricardo Baeza-Yates (Yahoo!, Spain)
Cindy Chung (University of Texas, Austin, USA)
Scott Counts (Microsoft Research, USA)
Chris Diehl (Jive Software, USA)
Ying Ding (Indiana University, USA)
Tim Finin (University of Maryland, Baltimore County, USA)
Lise Getoor (University of Maryland, USA)
Eric Gilbert (Georgia Institute of Technology, USA)
Kathy Gill (University of Washington, USA)
Natalie Glance (Google, USA)
Conor Hayes (Digital Enterprise Research Institute, Ireland)
Alejandro Jaimes (Yahoo!, Spain)
Karolin Kappler (Barcelona Media, Spain)
Kristina Lerman (USC Information Sciences Institute, USA)
Winter Mason (Stevens Institute of Technology, USA)
Gilad Mishne (Twitter, USA)
Mor Naaman (Rutger University, USA)
Kate Niederhoffer (Knowable Research, USA)
Bo Pang (Yahoo! Research, USA)
Jie Tang (Tsinghua University, China)
Loren Terveen (University of Minnesota, USA)
Yi Zhang (University of California, Santa Cruz, USA)

2012 Program Committee

Muhammad Aurangzeb Ahmad (University of Minnesota, USA)
Noor Ali-Hasan (Samsung Electronics, USA)
Jisun An (University of Cambridge, United Kingdom)
Mike Ananny (Microsoft Research, USA)
Judd Antin (Yahoo! Research, USA)
Ricardo Araujo (Federal University of Pelotas, Brazil)
James Bagrow (Northwestern University, USA)
Eytan Bakshy (Facebook, USA)
Denilson Barbosa (University of Alberta, Canada)
Christian Bauchhage (Fraunhofer IAIS, Germany)
Bettina Berendt (KU Leuven, Belgium)
Michael Bernstein (Massachusetts Institute of Technology, USA)
Rushi Bhatt (Yahoo! Labs, India)

Danushka Bollegala (The University of Tokyo, Japan)
Helena Buhr (Tagged, USA)
Carlos Castillo (Yahoo!, Spain)
James Caverlee (Texas A&M University, USA)
Meeyoung Cha (KAIST, Korea)
Jeffrey Chan (University of Melbourne, Australia)
Jonathan Chang (Facebook, USA)
Wei Chen (Microsoft Research Asia, China)
Zhiyuan Cheng (Texas A&M University, USA)
Ed H. Chi (Google, USA)
Philipp Cimiano (University of Bielefeld, Germany)
Michael Conover (Indiana University, USA)
William Corvey (University of Colorado at Boulder, USA)
Padraig Cunningham (University College Dublin, Ireland)
Onkar Dabeer (Tata Institute of Fundamental Research, India)
Elizabeth Daly (IBM Research, USA)
Cristian Danescu-Niculescu-Mizil (Cornell University, USA)
Kushal Dave (International Institute of Information Technology, Hyderabad, India)
Brian Davison (Lehigh University, USA)
Munmun De Choudhury (Microsoft Research, USA)
Christopher DuBois (University of California, Irvine, USA)
Brian Eoff (Bitly, USA)
Donghui Feng (eBay Inc., USA)
Junlan Feng (AT&T Research, USA)
Jennifer Foster (Dublin City University, Ireland)
Daniel Gatica-Perez (Idiap/EPFL, Switzerland)
Werner Geyer (IBM Research, USA)
Rumi Ghosh (University of Southern California, USA)
Roxana Girju (University of Illinois at Urbana-Champaign, USA)
Scott Golder (Cornell University, USA)
Sam Gosling (University of Texas, Austin, USA)
Derek Greene (University College Dublin, Ireland)
Josephine Griffith (National University of Ireland, Galway, Ireland)
Maria Grineva (Yandex Labs, USA)
Ido Guy (IBM Research, Israel)
Zoltan Gyongyi (Google, USA)
Laura Hasler (University of Strathclyde, United Kingdom)
Jonathan Hurlock (Swansea University, United Kingdom)
Michal Jacovi (IBM Research, Israel)
Brian Keegan (Northwestern University, USA)
Emre Kiciman (Microsoft Research, USA)
Alek Kolcz (Twitter, USA)
Gueorgi Kossinets (Google, USA)
Vassilis Kostakos (University of Oulu, Finland)
Renaud Lambiotte (Imperial College, United Kingdom)
James Lanagan (Technicolor R&I, France)
David Laniado (Barcelona Media, Spain)
Debra Lauterbach (Google, USA)

Bu Sung Lee (HP Labs, Singapore)
Jure Leskovec (Stanford University, USA)
Elizabeth Liddy (University of Syracuse, USA)
Yu-Ru Lin (Northeastern University and Harvard University, USA)
Avishay Livne (University of Michigan, USA)
Yoelle Maarek (Yahoo! Research, Israel)
Craig Macdonald (University of Glasgow, United Kingdom)
Sofus A. Macskassy (USC Information Sciences Institute, USA)
James Martin (University of Colorado, USA)
Mikhail Masli (University of Minnesota, USA)
Bhaskar Mehta (Google, Switzerland)
Andres Monroy-Hernandez (Microsoft Research, USA)
Michael Muller (IBM Research, USA)
Sean Munson (University of Michigan, USA)
Nozomi Nori (The University of Tokyo, Japan)
Oded Nov (Polytechnic Institute of New York University, USA)
Miles Osborne (University of Edinburgh, United Kingdom)
Aditya Pal (University of Minnesota, USA)
Katrina Panovich (Massachusetts Institute of Technology, USA)
Dae Hoon Park (University of Illinois at Urbana-Champaign, USA)
Alexandre Passant (Digital Enterprise Research Institute and NUI Galway, Ireland)
Wei Peng (Xerox, USA)
Igor Perisic (LinkedIn, USA)
Ana-Maria Popescu (Yahoo! Labs, USA)
Reid Priedhorsky (Los Alamos National Laboratory, USA)
Kunal Punera (Yahoo! Research, USA)
William Rand (University of Maryland, College Park, USA)
Raquel Recuero (Universidade Católica de Pelotas, Brazil)
Paul Resnick (University of Michigan, USA)
Timo Reuter (CITEC and University of Bielefeld, Germany)
Laura Ripamonti (University of Milan, Italy)
Alan Ritter (University of Washington, USA)
Monica Rogati (LinkedIn, USA)
Jakob Rogstadius (Madeira Interactive Technologies Institute, Portugal)
Camille Roth (CNRS/EHESS, France)
Ricard Ruiz De Querol (Barcelona Media, Spain)
Derek Ruths (McGill University, Canada)
Hassan Sayyadi (University of Maryland, USA)
Salvatore Scellato (University of Cambridge, United Kingdom)
Jan-Hinrik Schmidt (Hans-Bredow-Institut, Germany)
N. Denise Senmartin (Universitat Oberta de Catalunya, Spain)
N. Sadat Shami (IBM, USA)
Hossam Sharara (University of Maryland, College Park, USA)
Amit Sheth (Wright State University, USA)
Irina Shklovski (IT University of Copenhagen, Denmark)
Matthew P. Simmons (University of Michigan, USA)
Marc Smith (Connected Action, USA)
Scott Spangler (IBM, USA)

Ramesh Srinivasan (University of California, Los Angeles, USA)
Fred Stutzman (Carnegie Mellon University, USA)
L. Venkata Subramaniam (IBM Research, India)
Bongwon Suh (Adobe, USA)
Eric Sun (Facebook, USA)
Tao Sun (Peking University, China)
Hari Sundaram (Arizona State University, USA)
Zoltan Szlavik (VU University Amsterdam, Netherlands)
Belle Tseng (Yahoo! Labs, USA)
Greg Ver Steeg (Information Sciences Institute, USA)
Sarah Vieweg (University of Colorado at Boulder, USA)
Yana Volkovich (Barcelona Media, Spain)
Jianshu Weng (HP Labs, Singapore)
Janyce Wiebe (University of Pittsburgh, USA)
Max L. Wilson (Swansea University, United Kingdom)
Jiang Yang (University of Michigan, USA)
Daniel Xiaodan Zhou (University of Michigan, USA)

Auxiliary Reviewers

Ofer Arazy (University of Alberta, Canada)
Jennifer Bayzick (Lehigh University, USA)
Xiaohui Bei (Tsinghua University, China)
Joan-Isaac Biel (Idiap/EPFL, Switzerland)
Bin Chen (Indiana University, USA)
Lu Chen (Wright State University, USA)
Alexander Conrad (University of Pittsburgh, USA)
Maciej Dabrowski (Digital Enterprise Research Institute, Ireland)
Na Dai (Lehigh University, USA)
Anirban Dasgupta
Trinh-Minh-Tri Do (Idiap, Switzerland)
Michael Ekstrand (University of Minnesota, USA)
Denzil Ferreira (University of Oulu, Finland)
Jessica Gould (Barcelona Media, Spain)
Brent Hecht (Northwestern University, USA)
Benjamin Mako Hill (Massachusetts Institute of Technology, USA)
Liangjie Hong (Lehigh University, USA)
Simo Hosio (University of Oulu, Finland)
Aditya Johri (Virginia Tech, USA)
Marko Jurmu (University of Oulu, Finland)
Jeon-Hyung Kang (University of Southern California, USA)
Nanda Kumar (Baruch College, City University of New York, USA)
Myriam Leggieri (NUI Galway, Ireland)
Yanhua Li (University of Minnesota, USA)
Tomas Linden (University of Oulu, Finland)
Chen Lu (Kno.e.sis, USA)
Myle Ott (Cornell University, USA)
Katherine Panciera (University of Minnesota, USA)

Mikko Perttunen (University of Oulu, Finland)
Jay Pujara (University of Maryland, College Park, USA)
Hemant Purohit (Kno.e.sis, USA)
Vahed Qazvinian (University of Michigan, USA)
Angelo Restificar (eBay Inc., USA)
Paul Russo
Owen Sacco (Digital Enterprise Research Institute, Ireland)
Mrinmaya Sachan (IBM Research, India)
Aaron Shaw (University of California, Berkeley, USA)
Xin Shuai (Indiana University, USA)
Aixin Sun (Nanyang Technological University, Singapore)
Achint Thomas (Yahoo! Labs, USA)
Jayant Venkatanathan
Wenbo Wang (Kno.e.sis, USA)
Yajun Wang (Microsoft Research Asia, China)
Morten Warncke-Wang (University of Minnesota, USA)
Yuxia Yao (HP Labs, Singapore)
Chen Ye

ICWSM 2012 Sponsors

Platinum Sponsor

Science Foundation Ireland (SFI)

Silver Sponsors

Google

Bing / Microsoft Research

Fáilte Ireland

Bronze Sponsors

Yahoo! Research

IBM

Kno.e.sis

General Sponsor

Church and Duncan Group Inc.

With support from the National University of Ireland, Galway (NUI Galway) and Digital Enterprise Research Institute, Galway, Ireland.

We are grateful for the sponsorship and support of the Association for the Advancement of Artificial Intelligence (AAAI).

Preface

C*eád míle fáilte* — one hundred thousand welcomes — to the sixth meeting of the International AAAI Conference on Weblogs and Social Media in Dublin.

Research into social media sits at the intersection of several fields in both computing and social science, so since its inception, ICWSM has always had an interdisciplinary flavor. As you read through the program, you will see that this is reflected not just in a balance in papers from several fields, but also within the contributions themselves, since many of the papers and posters demonstrate excellence in multiple areas. We believe this increasing integration reflects a new level of maturity in the field: investigators have moved beyond simply applying their trusted methods to social media topics, and have started to implement the ideas and approaches that are best suited to their research questions.

This integrative approach to understanding social networks and social media is exemplified in this year's high quality roster of keynotes and invited talks. Robin Dunbar, Head of the Institute of Cognitive and Evolutionary Anthropology at the University of Oxford, will present one of our keynote addresses. Dunbar is the originator of Dunbar's number, a theoretical cognitive limit to the number of people with whom one can maintain stable social relationships. Lada Adamic, University of Michigan and serial ICWSM best paper winner, will present on information dynamics in networks. Andrew Tomkins, Engineering Director at Google, will speak about the first year of Google's social networking and identity service Google+, launched in 2011. Interestingly, Tomkins gave an invited talk at the first ICWSM conference in 2007 when he was at Yahoo!. We will also have an invited industry talk and discussion panel, bringing a strong enterprise element to ICWSM this year. Ireland is

home to both the European headquarters of multinationals (Google, Microsoft, Facebook, LinkedIn, Populis, and others) and emerging indigenous tech companies (Storyful, Datahug, Distilled Media, Newswhip, and others), all operating in the social media space. We thank our local chairs Conor Hayes and Derek Greene for all their work in helping realize this conference and for communicating with the social media research community in Ireland.

Organizing a conference is a year-round commitment — from the conference committee, our AAAI colleagues, and the program committee and reviewers. We also recognize, however, that it is only with the passion and commitment of those carrying out the research work and submitting papers, posters and demos that our meeting would take place at all, and so we reserve the greatest portion of our gratitude to everyone who made a submission. As our conference grows, so does the quality of submissions and the quality of the final accepted papers. This year's full paper submission level again increased over last year's. We accepted 47 papers out of the 232 submitted (corresponding to an acceptance rate of 20 percent). Of those, 20 papers were selected for oral presentation, while the remaining 27 papers were selected as full paper posters. The latter is a track that we introduced in 2011 whilst maintaining the integrative spirit and single-track nature of ICWSM. The full paper poster sessions will be intimate (nine papers each) and will begin with a poster-booster by the authors, followed by 45 minutes where conference participants get to visit the posters. An additional 67 papers were accepted for poster presentation, and three demos were accepted.

This year's papers cover both old and new ground, from the advancement of techniques to extract, summarize, and detect information from social streams, to understanding phe-

nomena in society by how they are reflected in and shaped by social media. For example, an analysis of the YouTube social network will show how it differs from traditional social networking sites but bears some resemblance to Twitter in terms of its network properties. In a very large study of browsing behaviors, web histories for a quarter of a million anonymized individuals are paired with user-level demographics to examine behavior change over time online, to revisit notions of the digital divide, and to show that user attributes may be inferred from browsing histories and potentially used for ad targeting. A method that can identify experts in social question-and-answer sites through evolutionary pattern detection will be presented that improves on previous non-evolutionary methods. A framework that accurately identifies sick individuals from the content of online communication will be presented and evaluated based on a large Twitter dataset. Research on the World of Warcraft multiplayer system will describe how destructive group dynamics can be predicted with medium-to-high accuracy. A clustering model and methodology for studying the structure and composition of a city, based on the social media its residents generate, will be described that has been tested using 18 million check-ins collected from a location-based online social network. These are just some samples of the high-quality work embedded in this year's topically and methodically diverse research papers contributing to the future of online social media.

We are optimistic that the marriage of techniques from the computer and social sciences that is represented at this year's conference will contribute to the continued synergistic collaborations among the increasing number of disciplines associated with research on social media.

The conference starts with a day of workshops and tutorials. In keeping with the interdisciplinary balance of the conference, we selected five tutorials this year (out of ten tutorial proposals) covering relevant foundational material from the areas of social design and social network analysis. Derek Ruths will give an

overview of how to analyze a massive social network without the need for a computing cluster. Marc Smith will demonstrate how to map networks found on social media sites and in email collections. Paul Resnick and Robert Kraut will give a tutorial on how to design an online community, using evidence-based social design principles. Denilson Barbosa will present on information extraction for social media analysis. Our final tutorial is by Lyle Ungar and Ronen Feldman, describing how to carry out sentiment mining on user-generated content, and building on their popular tutorial from last year. We are excited about this year's lineup of tutorial presenters and look forward to the breadth of ideas they will share. Sincere thanks goes out to Bernie Hogan, Tutorial Chair, and his great team of reviewers.

This year, we have accepted four workshops from ten submissions. They are Social Media Visualization, featuring an invited talk by University of Maryland's Ben Shneiderman; Real-Time Analysis and Mining of Social Streams; When the City Meets the Citizen Workshop, with guest speaker George MacKerron who created Mappiness; and The Potential of Social Media Tools and Data for Journalists in the News Media Industry. Sofus Macskassy led the charge to solicit and select workshops. Thanks to Sofus Macskassy and his team of reviewers.

The organizers would like to thank the program committee for their hard work in selecting this set of papers. Even more importantly, we would like to thank the broader ICWSM community for the excellent technical work that was submitted. Finally, we thank our sponsors — Science Foundation Ireland, Google, Bing / Microsoft Research, Fáilte Ireland / Meet in Ireland, Knoesis at Wright State University, Yahoo! Research, IBM, and Church & Duncan Group — that have made this event a reality.

We look forward to hosting you in Dublin!

– John Breslin, James G. Shanahan,
Nicole Ellison, Zeynep Tufekci

Invited Talks

Keynote

Invited Talk

Andrew Tomkins

Andrew Tomkins is an engineering director at Google working on measurement, modelling, and analysis of content, communities, and users on the world wide web. Prior to joining Google, he spent four years at Yahoo! as chief scientist of search, and eight years at IBM's Almaden Research Center, where he cofounded the WebFountain project. Andrew holds Bachelors degrees in mathematics and computer science from the Massachusetts Institute of Technology, and a PhD in computer science from Carnegie Mellon University; he has published over a hundred technical papers.

Keynote

Invited Talk: The Information Life of Social

Lada Adamic

Lada A. Adamic is an associate professor in the School of Information and the Center for the Study of Complex Systems at the University of Michigan. She is also affiliated with EECS. Her research interests center on information dynamics in networks: how information diffuses, how it can be found, and how it influences the evolution of a network's structure. Her projects have included identifying expertise in online question and answer forums, studying the dynamics of viral marketing, and characterizing the structure in blogs and other online communities.

Keynote

Invited Talk: Why Facebook Won't Get You Anymore Friends

Robin Dunbar

Robin Dunbar graduated with a BSc in psychology and philosophy from the University of Oxford and a PhD in psychology from the University of Bristol. He has held research fellowships at Cambridge and Liverpool Universities, and teaching posts at the University of Stockholm, University College London, and the University of Liverpool. He is currently a professor of evolutionary anthropology and director of the Institute of Cognitive and Evolutionary Anthropology in the School of Anthropology, and a Fellow of Magdalen College. He was elected a Fellow of the British Academy in 1998. He is codirector of the British Academy's Centenary Research Project "Lucy to Language: The Archaeology of the Social Brain," a multidisciplinary project involving, in addition to the University of Oxford, research groups at Liverpool University, Royal Holloway (University of London), Southampton University, and the University of Kent. His principal research interest is the evolution of sociality, with specific focus on humans, nonhuman primates and ungulates.

Tutorials

MA1

How to Analyze Massive Social Network Datasets without a Cluster

Presented by Derek Ruths

For all the wonderful things we hear about how compute clusters enable the analysis of massive datasets, the sad truth is that few researchers can use them. This is due to practical issues: clusters are expensive, their administration requires nontrivial time and technical knowledge, and the tools for using them aren't user friendly. Nonetheless, the average size of network-based datasets is growing. This means that more and more researchers need to do large network analysis but don't have a cluster at their disposal.

In this tutorial, we will teach attendees several easy-to-use strategies for analyzing large social network datasets (upwards of millions of nodes and edges with metadata) on a desktop or high-power laptop. These strategies involve using a combination of tools and techniques that capitalize on their individual strengths.

Over the course of the tutorial, attendees will first learn how to use several python-based libraries for large-scale data and network analysis as well as easy ways to store data to flat-files to minimize retrieval time. Then we will discuss how to combine these tools together. Finally, we will use the newly learned approaches to analyze two large network datasets.

Derek Ruths is a computer science professor at McGill University. Much of his current work involves analysis and modeling of large networked datasets from online social platforms and general ontologies. He likes clusters, but believes that laptops and desktops (but not iPads) can do large-scale network analysis too.

MA2

Charting Collections of Connections in Social Media: Creating Maps and Measures with NodeXL

Presented by Marc Smith

Networks are a data structure commonly found across all social media services that allow populations to author collections of connections. The Social Media Research Foundation's NodeXL project makes analysis of social media networks accessible to most users of the Excel spreadsheet application. With NodeXL, Networks become as easy to create as pie charts. Applying the tool to a range of social media networks has already revealed the variations present in online social spaces. A review of the tool and images of Twitter, flickr, YouTube, and email networks will be presented.

Marc Smith is a sociologist specializing in the social organization of online communities and computer mediated interaction. Smith leads the Connected Action consulting group and lives and works in Silicon Valley, California. Smith cofounded the Social Media Research Foundation, a nonprofit devoted to open tools, data, and scholarship related to social media research. Smith received a B.S. in International Area Studies from Drexel University in Philadelphia in 1988, an M.Phil. in social theory from Cambridge University in 1990, and a Ph.D. in sociology from UCLA in 2001. He is an adjunct lecturer at the College of Information Studies at the University of Maryland. Smith is also a distinguished visiting scholar at the Media-X Program at Stanford University.

MP1

Evidenced-Based Social Design of Online Communities: Getting to Critical Mass and Encouraging Contributions

Presented by Paul Resnick and Robert Kraut

To become or remain successful, online communities and networks must meet a number of challenges that are common to many groups and organizations, offline as well as online. For example, online communities must handle the start-up paradox, when early in their lifecycle they have few members to generate content and little content to attract members. Throughout their lifecycle, they must recruit and socialize newcomers, encourage commitment and contribution from members, solve problems of coordination and encourage appropriate behavior among members and interlopers alike.

The social sciences tell us a lot about how to make thriving online communities. Economics and various branches of psychology offer theories of individual motivation and of human behavior in social situations. The theories generalize from observations of naturally occurring behavior, from controlled experiments, and from abstract mathematical models. Properly interpreted, they can inform choices about how to meet the challenges described. This tutorial will focus in particular on problems of encouraging contributions and getting new communities to critical mass.

It is based on selected sections of *Building Successful Online Communities: Evidence-based Social Design*, coauthored with Robert Kraut, on using the social sciences as a guide to designing online communities.

Paul Resnick is a professor at the University of Michigan School of Information. He received the master's and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology. His researches focus on online communities, recommender systems, and applications in politics and health.

MP2

Sentiment Mining from User Generated Content

Presented by Lyle Ungar and Ronen Feldman

The proliferation of user generated content on the web is driving a new wave of work on the determination of user sentiment from web texts such as message boards, blogs, tweets, and Facebook status updates. Both researchers and practitioners are developing and applying new methods to determine how users feel about everything: products and politicians, friends and family, scientific articles and celebrities. This tutorial will cover the state of the art in this rapidly growing area, including recent advances that combine information extraction with sentiment analysis to improve accuracy in assessing sentiment about specific entities. We will present several real world applications of sentiment analysis. Special emphasis will be given to lessons learned from years of experience in developing real world sentiment analysis systems.

Lyle H. Ungar is an associate professor of computer and information science at the University of Pennsylvania. Ungar received a B.S. from Stanford University and a Ph.D. from the Massachusetts Institute of Technology. He directed Penn's Executive Masters of Technology Management Program for a decade, and is currently associate director of the Penn Center for BioInformatics. He has published over 100 articles and holds eight patents. His current research focuses on developing scalable machine learning methods for data mining and text mining.

Ronen Feldman is one of the world's most recognized experts in the field of text mining, link analysis and the semantic analysis of data. In 1997, he founded ClearForest, a Boston-based business intelligence company later acquired by Reuters. He coined the term "text mining" in 1995, and his textbook, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* is considered the world's premier authority on this complex topic. He currently serves as the head of the Information Systems Department at the Business

School of the Hebrew University of Jerusalem and was an adjunct professor at New York University's Stern Business School. He has given over 30 tutorials on text mining and information extraction and has written numerous scholarly papers on these topics. He received his Ph.D. in computer science from Cornell University and his B.Sc. in mathematics, physics and computer science from the Hebrew University of Jerusalem. He began his career in the military assigned to the elite Talpiot Group, and served for eight years as an officer in the Israel Air Force.

MP3

Information Extraction for Social Media Analysis

Presented by Denilson Barbosa

More and more regular users use the blogosphere to express and discuss their opinions, the facts, events, and ideas pertaining to their own lives, their community, their profession, or society at large. It goes without saying that being

able to extract reliable data from this medium opens the door to the most varied kinds of analysis and using datasets of massive proportions. As a result, a great deal of attention has been devoted lately to applying information extraction to the blogosphere. In this tutorial, I focus on a specific subproblem: extracting information networks that act as summaries of the blogosphere as a whole. These networks consist of nodes representing entities and edges representing the relationship between such entities. I will cover fundamental tools from NLP and network science that allow the unsupervised extraction information networks from social media content.

Denilson Barbosa is an associate professor at the University of Alberta, working on databases, information retrieval, and the management of linked data. He received a PhD from the University of Toronto (2005), and is a member of the NSERC Strategic Network on Business Intelligence and the Canadian Writing Research Collaboratory.

Workshops

Social Media Visualization

socmedvis.ucd.ie

Social media study and analysis brings researchers from many fields into a single setting. Even though the tasks of these researchers are varied, data visualization and analytics plays an important role. For industry and academics alike, visualization of social media data helps with hypothesis formation and supports the explanation of phenomena. The Social Media Visualization workshop provides a venue to discuss how we can apply data visualization and analytic techniques to social media data.

The workshop caters to designers and consumers of interactive visualization methods to better understand the applications of these techniques to social media data for the purposes of analysis and entertainment. The workshop is open to both novel visualization techniques and applications of visualization techniques to social media data sources. We strongly encourage interdisciplinary contributions that discuss the application of visualization to social media research and how these techniques can better support user tasks in other domains.

This workshop session will also feature an invited talk by Ben Shneiderman, a professor of computer science at University of Maryland Institute for Advanced Computer Studies and founding director of the Human-Computer Interaction Laboratory. Shneiderman was elected as a Fellow of the Association for Computing (ACM) in 1997 and a Fellow of the American Association for the Advancement of Science (AAAS) in 2001. He received the ACM SIGCHI Lifetime Achievement Award in 2001. He is a member of the National Academy of Engineering.

Organizers: Daniel Archambault (Cliques Strategic Research Cluster, University College Dublin, Ireland), Eser Kandogan (IBM Re-

search, Almaden, USA), Martin Harrigan (Cliques Strategic Research Cluster, University College Dublin, Ireland)

Real-Time Analysis and Mining of Social Streams

www.ramss.ws

The recent increase of real-time data provided by users on social networking services has leveraged an importance gain of the real-time processing of social streams. Processing the streams in real-time can help enhance search engines, news media, and many other systems by feeding them with fresh knowledge about current affairs. Performing such analysis in real-time is of utmost importance for early reporting of breaking news, events, trends, and knowledge related to current affairs. However, analyzing social streams in real-time makes the task more challenging as it requires making decisions without clue of what will be next in the stream. The Real-Time Analysis and Mining of Social Streams workshop aims to bring together experts in the real-time analysis and mining of social streams, as well as to further develop and exchange knowledge around these tasks.

Organizers: Arkaitz Zubiaga (City University of New York, USA), Damiano Spina (Universidad Nacional de Educación a Distancia, Spain), Maarten De Rijke (University of Amsterdam, The Netherlands), Markus Strohmaier (Graz University of Technology, Austria), Mor Naaman (Rutgers University, USA)

When the City Meets the Citizen Workshop

bit.ly/wcmcw

Social media has played a key role in providing insights into people's activities, opinions and day to day lives. These detailed user-generated information-streams offer a unique opportunity for cities to understand and engage their citizens. The research domain of smarter cities aims to monitor disruptive events (such as emergencies, Olympics), analyze social behaviour, identify citizens' sentiment and understand their interactions with services. On the other side, cities can use their understanding of the citizen to foster stronger relationships with the diverse communities in their constituencies. This understanding could be applied to mobilize people on important issues such as education, health care, political engagement and community awareness. This workshop is interested in research that aims to progress both sides of this relationship, but particularly research that works towards closing the loop between the city and the citizens.

Organizers: Elizabeth Daly (IBM Research, Dublin, Ireland), Giusy Di Lorenzo (IBM Research, Dublin, Ireland), Daniele Quercia (University of Cambridge, UK), Michael Muller (IBM Research, Cambridge, USA)

The Potential of Social Media Tools and Data for Journalists in the News Media Industry

www.arcomem.eu/icwsm-2012-workshop

The news media industry has recognized that there is an undeniable shift in the way content is produced, consumed and distributed, including the sharing thereof. Nowadays, there are numerous cases in which established news agencies and news outlets are not the first point of call for users because everybody equipped with a smartphone (for example) can capture and publish events as they unfold on a platform of their choice (such as YouTube, Facebook, Twitter, and others). For this reason, news media organizations are increasingly depending on content from and information residing in social networks. However, news organizations are facing one crucial question when it comes to using content from the social web: "Can I trust this source? Is the source reliable?" The aim of this workshop is to encourage participants to discuss, share, and exchange ideas and results on social media research, technologies, and applications in two essential areas: newsgathering in the social web and filtering and analyzing the available data (for example, by relevance, reliability, accuracy, and other criteria).

Organizers: Cosmin Cabulea (Deutsche Welle, Germany), Alejandro Jaimes (Yahoo! Research, Spain), Jonathon Hare (University of Southampton, UK), Jochen Spangenberg (Deutsche Welle, Germany), Dominik Frey (Südwestrundfunk, Germany)