We received 8 tutorial proposals and finally accepted 5.

The first tutorial is about “Entity Resolution for Social Network Analysis and Mining” by Lise Getoor and Ashwin Machanavajjhala. The tutorial will introduce entity resolution, a problem of extracting, matching and resolving entity mentions in structured and unstructured data, which is a long-standing challenge in artificial intelligence, database management, information retrieval and statistics. It is especially important and challenging in the context of mining and analyzing social network and social media data.

The second tutorial is entitled “Metaheuristic Algorithms for Networks Modelling and Optimization” and authored by Xin-She Yang. It introduces the fundamentals of the latest advances of metaheuristic algorithms. This is important for social network modeling and optimization, as most modeling and mining problems in social networks are NP-hard.

The third tutorial is entitled “Models and Algorithms for Social Influence Analysis” and authored by Jimeng Sun and Jie Tang. Social influence has been a widely accepted phenomenon in social networks for decades. Many applications have been built based around the implicit notion of social influence between people, such as marketing, advertisement and recommendations. This tutorial will review the research on social influence analysis with a focus on the computational aspects.

The fourth tutorial is entitled by “Advanced graph mining & community evaluation metrics for social networks and the Web” and authored by Michalis Vazirgiannis. The tutorial will first introduce the fundamental models of graph structures for social networks and their properties. It will also introduce social network analysis algorithms with an emphasis on community detection.

The last tutorial is about “On the Power of Mining Heterogeneous Information Networks”, which is authored by Yizhou Sun, Jiawei Han, Xifeng Yan, and Philip Yu. The tutorial will focus on scalable mining on heterogeneous information network, with an emphasis on basic mining algorithms on information networks such as clustering, classification, and ranking, meta-path-based similarity search and mining and relation strength-aware mining.

Ralf Klamma, RWTH Aachen University
Huan Liu, Arizona State University
Jie Tang, Tsinghua University
Entity Resolution for Social Network Analysis and Mining

Entity resolution (ER), the problem of extracting, matching and resolving entity mentions in structured and unstructured data, is a long-standing challenge in artificial intelligence, database management, information retrieval and statistics. It has been approached using a variety of techniques, including constraint-based methods, statistical methods, and methods which perform probabilistic inference. Accurate and fast entity resolution has huge practical implications in a wide variety of commercial, scientific and security domains. Entity resolutions is especially important and challenging in the context of mining and analyzing social network and social media data.

Despite the long history of work on entity resolution (which goes under a variety of names, depending on the subdiscipline, including record linkage, deduplication, co-reference resolution, reference reconciliation, object consolidation and identity uncertainty), there is still a surprising diversity of approaches, and lack of guiding theory. Meanwhile, the need for high quality entity resolution is growing, as we are inundated with more and more data, all of which needs to be integrated, aligned and matched, before further utility can be extracted.

In this tutorial, we bring together perspectives on entity resolution from a variety of fields, including databases, information retrieval, natural language processing and machine learning, to provide, in one setting, a survey of a large body of work. We discuss both the practical aspects and theoretical under-pinnings of ER. We describe existing solutions, current challenges and open research problems. We give attendees hands-on experience with entity resolution using a variety of tools.

Speakers

Lise Getoor is an associate professor in the Computer Science Department at the University of Maryland, College Park. She received her PhD from Stanford University in 2001. Her current work includes research on link mining, statistical relational learning and representing uncertainty in structured and semi-structured data. She has also done work on social network analysis and visual analytics. She has published numerous articles in machine learning, data mining, database, and artificial intelligence forums. She was awarded an NSF Career Award, is an action editor for the Machine Learning Journal, is a JAIR associate editor, has been a member of AAAI Executive council, and has served on a variety of program committees including AAAI, ICML, IJCAI, ISWC, KDD, SIGMOD, and UAI. Lise teaches several courses at UMD, including the introductory artificial intelligence course, the graduate machine learning course, a data mining course, and seminars on statistical relational learning and dynamic network analysis. She has given numerous invited talks, distinguished lectures, and keynotes, including an invited talk at AAAI 2007.

Ashwin Machanavajjhala is a Senior Research Scientist in the Knowledge Management group at Yahoo! Research. His primary research interests lie in data privacy and security, big-data management and statistical methods for information integration. Ashwin graduated with a Ph.D. from the Department of Computer Science, Cornell University. His thesis work on defining and enforcing privacy was awarded the 2008 ACM SIGMOD Jim Gray Dissertation Award Honorable Mention. He has also received an M.S. from Cornell University and a B.Tech in Computer Science and Engineering from the Indian Institute of Technology, Madras. Ashwin is currently leading a project on building a Web-scale distributed unsupervised information extraction system for automatically creating structured databases of entities from semi-structured Web pages, and is involved in building a large-scale distributed entity resolution system at Yahoo! as part of the Web-Of-Concepts initiative. His research specialty is in developing highly scalable statistical methods for information extraction and ER on distributed systems. He has developed solutions for extracting structured entities from semi-structured web-pages, blocking, clustering in entity resolution, and integrating temporally changing data.
Metaheuristic Algorithms for Networks Modelling and Optimization

Many problems in network modelling, data mining and optimization are NP-hard, and there is no efficient algorithm to tackle such problems. Over the last two decades, nature-inspired metaheuristic algorithms have become increasingly popular in solving large-scale, nonlinear, global optimization with many real-world applications. They also become an important part of computational intelligence as so-called “smart algorithms”. New algorithms emerge almost every year.

This 2-hour tutorial intends to introduce the fundamentals and review the latest advances of metaheuristic algorithms with the focus on the implementation and algorithm analysis. We will introduce and discuss in detail most of the new metaheuristics, including particle swarm optimization, differential evolution, bat algorithm, cuckoo search, firefly algorithm and others.

Speakers
Dr Xin-She Yang received his DPhil in Applied Mathematics from the University of Oxford, then worked at Cambridge University for a few years and now is a Senior Research Scientist at National Physical Laboratory. He has authored/edited a dozen books and published more than 140 papers. He is the Editor-in-Chief of Int. J. Mathematical Modelling and Numerical Optimisation (IJMMNO, Inderscience), serves as an editorial board member of several international journals, including Elsevier's Journal of Computational Science (JoCS). He is also a vice chair of the IEEE CIS task force.
Models and Algorithms for Social Influence Analysis

Social influence is the behavioral change of a person because of the perceived relationship with other people, organizations and society in general. Social influence has been a widely accepted phenomenon in social networks for decades. Many applications have been built based around the implicit notation of social influence between people, such as marketing, advertisement and recommendations.

With the exponential growth of online social network services such as Facebook and Twitter, social influence can for the first time be measured over a large population. In this chapter, we survey the research on social influence analysis with a focus on the computational aspects. First, we present statistical measurements related to social influence. Second, we describe the literature on social similarity and influences. Third, we present the research on social influence maximization which has many practical applications including marketing and advertisement.

Speakers

Dr. Jimeng Sun is a research staff member at Healthcare Transformation group, IBM T. J. Watson lab. He received a Bachelor and MPhil in Computer Science from Hong Kong University of Science and Technology in 2002 and 2003. After that, he obtained a MS and PhD degree in Computer Science from Carnegie Mellon University in 2006 and 2007. His research interests include data mining for streams and networks, databases and healthcare analytics. He has received the best research paper award in ICDM 2008, the best research paper award in SDM 2007. He has published over 40 refereed articles and two book chapters. He filed four patents and has given five tutorials. He has served as a program committee member of SIGKDD, ICDM, SDM and CIKM, senior committee member for SDM 2010 and SDM 2011, and a reviewer for TKDD, TKDE, VLDB, ICDE. For more details, one can refer to his personal homepage at http://www.dasfa.net/jimeng

Jie Tang is an associate professor at the Department of Computer Science and Technology, Tsinghua University. He serves as the director of Department Scientific Office. His main research interests include social network mining and fundamental learning theories. He has been visiting scholar at Cornell University, University of Illinois at Urbana-Champaign, Chinese University of Hong Kong, Hong Kong University of Science and Technology, and Leuven University. He has published over 100 research papers in major international journals and conferences including: KDD, IJCAI, SIGMOD, ACL, ISWC, Machine Learning Journal, TKDD, TKDE, JWS and JoDS. He serves as Local Chair of SIGKDD’12, Publications Co-Chairs of SIGKDD’11, Program Chair of ADMA’11, Program Chair of SocInfo’12, Poster Chair of WSDM’11, and also serves as the PC member of more than 50 international conferences. For more details, one can refer to his personal homepage at http://keg.cs.tsinghua.edu.cn/jietang
Advanced graph mining & community evaluation metrics for social networks and the Web

Graphs constitute a dominant data structure and appear essentially in all forms of information. Examples are the Web graph, numerous social networks, protein interaction networks, terms dependency graphs, network topologies etc. The main features of these graphs are their huge volume and rate of change. Presumably there is important hidden knowledge in the macroscopic topology and features of these graphs. A cornerstone issue here is the detection and evaluation of communities – bearing multiple and diverse semantics. The tutorial will report the fundamental models of graph structures for undirected/directed and signed graphs and their properties. Next we offer a review of fundamental techniques of graph/community clustering algorithms putting emphasis on those that capitalize on the concept of degeneracy as a novel means of community detection. Next we will offer a survey of community measures evaluation including both the individual node based ones as well as those that take into account aggregate properties of communities. A special mention will be made to the k-cores approach to community evaluation. We will justify the above foundational framework with applications citation graphs, trust networks, protein graphs.

Speaker
On the Power of Mining Heterogeneous Information Networks

In most of the existing network research, social and information networks are usually assumed to be homogenous, where nodes are objects from the same entity type (e.g., person) and links are relationships from the same relation type (e.g., friendship). However, in reality, objects that are of different entity types may link together via relationships of different relation types, forming heterogeneous information networks that are semi-structured. Such heterogeneous information networks are ubiquitous, and form powerful and expressive representations of very generic real-world systems in diverse domains.

In this tutorial, we will present an organized picture on scalable mining of heterogeneous information networks, with the inclusion of the following topics: (1) an introduction to information networks; (2) mining information networks: clustering, classification, and ranking; (3) meta-path-based similarity search and mining; (4) relation strength-aware mining; (5) construction of informative networks by data mining; and (6) research frontiers in information network analysis.

Speakers
Yizhou Sun is a fifth year Ph.D. candidate at the Department of Computer Science, University of Illinois at Urbana-Champaign. Her principal research interest is in mining information and social networks, and more generally in data mining, database systems, statistics, machine learning, information retrieval, and network science, with a focus on modeling novel problems and proposing scalable algorithms for large-scale, real-world applications. Yizhou has over 30 publications in book chapters, journals, and major conferences. Tutorials based on her thesis work on mining heterogeneous information networks have been given in several premier conferences, such as SIGMOD’10, SIGKDD’10 and ICDE’12.

Jiawei Han, Bliss Professor of Computer Science, University of Illinois at Urbana-Champaign. He has been researching into data mining, information network analysis, database systems, and data warehousing, with over 600 journal and conference publications. He has chaired or served on many program committees of international conferences, including PC co-chair for KDD, SDM, and ICDM conferences. He also served as the founding Editor-In-Chief of ACM Transactions on Knowledge Discovery from Data and is serving as the Director of Information Network Academic Research Center supported by U.S. Army Research Lab. He is a Fellow of ACM and IEEE, and received 2004 ACM SIGKDD Innovations Award, 2005 IEEE Computer Society Technical Achievement Award, 2009 IEEE Computer Society Wallace McDowell Award, and 2011 Daniel C. Drucker Eminent Faculty Award at UIUC.

Xifeng Yan is an associate professor at the University of California at Santa Barbara. He holds the Venkatesh Narayanamurti Chair in Computer Science. He received his Ph.D. degree in Computer Science from the University of Illinois at Urbana-Champaign in 2006. He has been working on modeling, managing, and mining graphs in bioinformatics, social networks, information networks, and computer systems. His works were extensively referenced, with over 5,000 citations per Google Scholar. He received NSF CAREER Award, IBM Invention Achievement Award, ACM-SIGMOD Dissertation Runner-Up Award, and IEEE ICDM 10-year Highest Impact Paper Award.

Philip S. Yu received his Ph.D. degree in E.E. from Stanford University. He is a Professor in Computer Science at the University of Illinois at Chicago and also holds the Wexler Chair in Information Technology. Dr. Yu spent most of his career at IBM, where he was manager of the Software Tools and Techniques group at the Watson Research Center. His research interests include data mining, database and privacy. He has published more than 680 papers in refereed journals and conferences. He holds or has applied for more than 300 US patents. Dr. Yu is a Fellow of the ACM and the IEEE. He is the Editor-in-Chief of ACM Transactions on Knowledge Discovery from Data.