CS591Han Seminar: Advanced Topics on Data Mining (Presentation Schedule: Spring 2012)

- Semester themes: (1) Information Network Analysis; and (2) data mining in cyperphysical systems
- Papers can be selected from this year or previous several year's conference proceedings or journals. We welcome students who would like to present tutorials and/or writing topic survey articles related to this theme.
- Two students per unit (20 minutes presentation and 5 minutes discussion for each research paper, i.e., two papers will be covered per class unit)
- Please book your timeslot early. Also, please upload your papers and slides to be presented at least one day before its presentation!

Week 1 (1/19/12): Class organization

Week 2 (1/26/12): One presentation

- Querying and Online Analytical Processing of Large-scale Information Networks

Abstract:

Social and technical information systems usually consist of a large number of interacting physical, conceptual, and human/societal entities. Such individual entities are interconnected to form large and sophisticated networks, which, without loss of generality, are often referred to as information networks. Noteworthy examples of information networks include communication networks, research collaboration networks, biological networks, social networks and ultimately, the Web. Information networks are ubiquitous and have formed a critical component of modern information infrastructure.

In this talk, I will discuss a series of fundamental queries of practical value arising in real world information networks, and explore the corresponding effective and scalable query processing solutions in the case of large-scale information networks. I will be mainly focused on two pieces of my recent work. SPath is a high performance graph indexing mechanism to address the general sub-graph queries in large information networks. Graph Cube is the first of its kind warehousing and online analytical processing (OLAP) tool for multidimensional information networks. In Graph Cube, we jointly considered both the multidimensional metadata together with the graph structure of information networks for query processing, thus generating a set of structure-enriched and concisely summarized aggregate networks as query results. It has been widely believed that the marriage of information network analytics and query processing technology will bring many exciting opportunities for future study, and thus I will briefly mention some prospective research topics within and beyond this direction as well in the talk.

Biography

Peixiang Zhao is the 5th year Ph.D. student in the Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC). He is affiliated with the Data Mining Research Group and the Data and Information Systems (DAIS) Laboratory at UIUC. His advisor is Prof. Jiawei Han. Peixiang's research interest lies in database systems, data mining and data-intensive computation and analytics in general. More specifically, he has been focused on the problems in modeling, querying and mining graph-structured data, such as large-scale graph databases and information networks, which have been witnessed extremely popular and flourishing in a wide spectrum of application domains, such as bioinformatics, computer systems, business processes and the Web. Peixiang got his B.S. and M.S. degree from the Department of Computer Science and Technology at Peking University in 2001 and 2004, respectively. Between 2004 and 2007, he was a Ph.D. student in the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong. For more detailed information, please visit his homepage at http://www.cs.uiuc.edu/homes/pzhao4/ or contact him via pzhao4@illinois.edu.

* Week 3 (2/2/12): *ICDM 2011 conference report

* Hyungsul Kim ICDM reports.pdf

Week 4 (2/9/12): Canceled

Week 5 (2/16/12): Two presentations

- Jingjing Wang
- Chi Wang: A survey of blockmodel and stochastic blockmodel slides
Week 6 (2/23/12): Two presentations
- Tim Weninger Slides
- Brandon Norick Slides

Week 7 (3/1/12): Two presentations
- Manish Gupta Slides Paper
- Bolin Ding Slides

Week 8 (3/8/12): INARC summary discussion
NSCTA-reporting slides
I3.1
Ming Ji: Hidden Network Discovery in Heterogeneous Information Networks (I3.1: Mining hidden networks in noisy data) (KDD12 sub)
Yizhou: Relation Strength-Aware Clustering of Heterogeneous Information Networks with Incomplete Attributes (VLDB'12) (I3.1)
Quanquan: Classification of Networks with Inconsistent Links (I3.1: Mining hidden networks in noisy data) (KDD12 sub)
Xiao Yu: *Citation Prediction in Heterogeneous Bibliographic Networks* *(I3.1: Mining hidden networks in noisy data) (SDM’12)*
I3.2
Hongbo: Exploring Opposite Opinions and Dissimilarities for Sentiment Analysis with Heterogeneous Networks (I3.2) (KDD12 sub)
Chi Wang: Hierarchical Relationships among Objects with Heterogeneous Attributes and Links (SDM12) (I3.2)
Bo: From TruthFinder to Latent Truth Model (LTM) (I3.2) (VLDB’12)
E2.2
LuAn: Mining Intruders from Sensors (E2.2 Military Mobility) (cross-center: SDM’12)
Zhenhui Li: Mining Periodicity for Sparse and Incomplete Event Data (E2.2 Military Mobility) (KDD12 sub)
E2.5
Yizhou: When Will It happen? – Relationship Prediction in Heterogeneous Information Networks (E2.5 Evolution Modeling and Prediction in Time-Varying Inter-Genre Networks) (cross-center: WSDM’12)
CS4
Manish: Integrating Community Matching and Outlier Detection for Mining Evolutionary Community Outliers (Manish Gupta, Jing Gao, Yizhou Sun, Jiawei Han) (Cyber-Security: CS4) (KDD12 sub)

Week 9 (3/15/12): Two presentations
- Yizhou Sun Relation Strength-Aware Clustering of Heterogeneous Information Networks with Incomplete Attributes
- Bo Zhao slides


Week 11 (3/29/11): Two presentations
- Xiao Yu NoSQL Slides
- Jingjing Wang: Spatial-Temporal Models in Location Prediction 591_032912.pptx

Week 12 (4/5/11): Two presentations
- Hongbo Deng Modeling and Exploiting Heterogeneous Bibliographic Networks for Expertise Ranking slides
- Quanquan Gu Locality Preserving Feature Learning slides

Week 13 (4/12/11): Two presentations
Week 14 (4/19/11): Two presentations

- Lu-An Tang: IntruMine: Mining Intruders in Untrustworthy Data of Cyber-physical Systems [Slides]
- Hyung Sul Kim

Week 15 (4/26/11): Two presentations

- Brandon Norick: Alleviating the Sparsity Problem in Collaborative Filtering by using an Adapted Distance and a Graph-based Method, SDM ’10 [Slides]

Week 16 (5/3/12): No presentation: Group semester summary