Yan Xie

Mail: 1 Shore Lane, Apt 1111
Jersey City, NJ 07310

Work: Room 1336 SEO
851 S. Morgan St., Chicago, IL 60607

Phone: (312)375-0606
Email: yxie8@uic.edu
URL: `http://www.cs.uic.edu/~yxie`
Nationality: P. R. China, holding F-1 Visa
Gender: Female

Highlight

Ph.D. candidate focusing on data mining and large data management, with expertise in data analytics and holding a M.S. in Statistics, as well. A responsible team player, executes fast with good hands-on capabilities, troubleshooting skills and great drive for optimization/innovation.

Skills

- Windows, Mac OS, UNIX/Linux.
- C, C++, Java, Shell Scripts, JavaScript, HTML, XML, MySQL, OPENGL, Weka, SAS Base Certification.
- Visual Studio, Eclipse, VIM.

Education

2006.8 – present **University of Illinois at Chicago (UIC)**
Ph.D. in Computer Science (expected), GPA: 4.0/4.0
– Thesis: Scalable Mining of Large Graphs and Its Applications
– Advisor: Prof. Philip S. Yu
M.S. in Statistics, GPA: 4.0/4.0
Computer Science Courses: Computer Algorithms, Computer Graphics, Database Systems, Information Retrieval, Data Mining and Text Mining, Advanced Topics in Data Mining, Theory of Artificial Intelligence
Statistics Courses: Statistical Theory and Inference, Advanced Statistical Theory, Applied Statistical Method, Linear and Non-linear Programming, Sampling Theory, Operations Research, Scientific and Statistical Software, Business Forecasting

2003.9 – 2006.4 **Zhejiang University**, P.R. China
M.E. in Computer Science and Technology, GPA: 3.52/4.0

1999.9 – 2003.7 **Hangzhou Institute of Electronics and Engineering (HZIEE)**, P.R. China
B.E. in Computer Science and Technology, GPA: 3.73/4.0

Honors & Awards

- **Best SDM 2012 paper** selected for publication in *Statistical Analysis and Data Mining* (SADM).
- **SDM Student Travel Award**, 2012.
- **SDM Student Travel Award**, 2011.
- **Excellent Graduate of Zhejiang Province** (top 1% of all seniors), Bachelor graduation, 2003.
- **Supreme Scholarship** (highest honor in university), HZIEE, 2003.
- **Outstanding Student Scholarship First Prize**, HZIEE, 2000 – 2003 consecutively.
- **Sunyard Scholarship**, HZIEE, 2002.

Research & Work Experiences

- **Knowledge and Information Management Group, University of Illinois at Chicago**
  Research Assistant with Prof. Philip S. Yu, 2009.1 – present
  - ◇ My research has been focusing on developing scalable methods to mine and manage large graphs and information networks. The framework I proposed centers around one core concept: Data reduction, which transforms big graphs into smaller ones so that they are more manageable. First, we showed the feasibility and power of this framework by efficiently mining connectivity patterns over large-scale networks [10]. Second, we extended the idea and demonstrated that it is general enough to be applied to other data domains as well [9][2]. This encourages more adoption of graph data models, as the complication to deal with complexly structured graph data, which scares many previously, can now be handled with the introduction of data reduction techniques. Third, we applied the scalable mining framework to benefit other graph data management tasks, such as indexing. To create efficient and effective indices on large graphs, we reduced the graph size by either compressing it via randomized summaries [5] or partitioning it into smaller pieces [3]. It was guaranteed that important patterns for building a high-quality index are always preserved even after data reduction.
  - ◇ I have also worked on a few other interesting large graph and information network problems, e.g., (1) prediction of new links (interactions) in non-uniform and dynamically changing networks [4][1], (2) finding the global community structure in a localized manner, which accounts for the local heterogeneity of big social networks [8], (3) data warehousing and OLAP (On-Line Analytical Processing) capabilities on large graphs [7], and (4) data processing on streaming sensor networks [6].
  - ◇ I have been in charge of many non-academic responsibilities in my research group such as on-boarding new members, accommodating visiting scholars, coordinating seminars, and other organizational facilitating activities.
- **Network Security Group, IBM T. J. Watson Research Center**
  Research Intern with Dr. Sreedhar Vugranam and Dr. Charu Aggarwal, 2009.6 – 2009.10
  - ◇ I applied sequential pattern analysis to detect fraudulent transactions in a bank database. I implemented several innovative frequent pattern mining algorithms, integrating both unsupervised learning that learns from available data and supervised learning that leverages domain expert knowledge. After conducting experiments on real data and extensive comparisons, I was able to significantly improve the fraud detection accuracy.
- **National Center of Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign**
  Summer Intern with Dr. Michael Welge, 2007.5 – 2007.8
  - ◇ State Farm Insurance was our client for this project. I analyzed text search results by developing in NCSA's D2K (Data to Knowledge) platform, created a knowledge base on collective intelligence, drafted the project's White Paper, and also helped refine it further.
- **State Key Lab of Computer Aided Design & Computer Graphics, Zhejiang University**
  Research Assistant with Dr. Xueying Qin, 2003.7 – 2006.4
  - ◇ I worked on Image Processing and Color Sciences, and more specifically, high dynamic range images and tone mapping, color correction in a panorama, and color transfer in videos.

Teaching Experiences

- *Instructor*: IDS371 - Business Statistics II, Fall 2008, UIC.
  – Taught this full-credit course to 50 people in the Department of Information and Decision Sciences.
- *Teaching Assistant*: CS202 - Data Structure II, Spring 2008, UIC.
- *Teaching Assistant*: CS107 - Java Programming, Fall 2007 / Spring 2008, UIC.
- *Teaching Assistant*: CS109 - Introduction to C++ Programming, Spring 2007 / Spring 2009, UIC.
- *Teaching Assistant*: CS108 - Fortran Programming, Fall 2006, UIC.

Publications

1. Charu Aggarwal, **Yan Xie**, and Philip S. Yu, "A Framework for Dynamic Link Prediction in Heterogeneous Networks", invited submission to *Statistical Analysis and Data Mining* (**SADM**).
2. **Yan Xie**, Philip S. Yu, and Feida Zhu, "Max-Clique: A Top-Down Graph-based Approach to Frequent Pattern Mining", under review, *IEEE Transactions on Knowledge and Data Engineering* (**TKDE**).
3. **Yan Xie** and Philip S. Yu, "Storage Efficient Graph Search by Composite Dynamic-and-Static Indexing of a Large Network", submitted to *16th International Conference on Extending Database Technology* (**EDBT 2013**).
4. Charu Aggarwal, **Yan Xie**, and Philip S. Yu, "On Dynamic Link Inference in Heterogeneous Networks", *Proceedings of the 12th SIAM International Conference on Data Mining* (**SDM 2012**), Anaheim, CA, Apr. 2012.
   – Selected as a **best SDM 2012 paper** for publication in *Statistical Analysis and Data Mining* (**SADM**).
5. **Yan Xie** and Philip S. Yu, "CP-Index: On the Efficient Indexing of Large Graphs", *Proceedings of the 20th ACM Conference on Information and Knowledge Management* (**CIKM 2011**), Glasgow, United Kingdom, Oct. 2011.
6. Charu Aggarwal, **Yan Xie**, and Philip S. Yu, "On Dynamic Data-driven Selection of Sensor Streams", *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (**KDD 2011**), San Diego, CA, Aug. 2011.
7. Chuan Li, Philip S. Yu, Lei Zhao, **Yan Xie**, and Wangqun Lin, "InfoNetOLAPer: Integrating InfoNetWarehouse and InfoNetCube with InfoNetOLAP", *Proceedings of the 37th International Conference on Very Large Data Bases* (**VLDB 2011**), Seattle, WA, Aug. 2011.
8. Charu Aggarwal, **Yan Xie**, and Philip S. Yu, "Towards Community Detection in Locally Heterogeneous Networks", *Proceedings of the 11th SIAM International Conference on Data Mining* (**SDM 2011**), Mesa, AZ, Apr. 2011.
9. **Yan Xie** and Philip S. Yu, "Max-Clique: A Top-Down Graph-based Approach to Frequent Pattern Mining", *Proceedings of the 10th IEEE International Conference on Data Mining* (**ICDM 2010**), Sydney, Australia, Dec. 2010.
10. Charu Aggarwal, **Yan Xie**, and Philip S. Yu, "GConnect: A Connectivity Index for Massive Disk-resident Graphs", *Proceedings of the 35th International Conference on Very Large Data Bases* (**VLDB 2009**), Lyon, France, Aug. 2009.
11. **Yan Xie**, Mingfu Wang, and Xueying Qin, "Color Correction Algorithm for Panorama", *Proceedings of the 5th Chinese Conference on Virtual Reality and Visualization* (**CCVRV 2005**), Beijing, China, Sep. 2005.

Professional Services

- *Journal Reviewer*: IEEE Transactions on Knowledge and Data Engineering (**TKDE**), ACM Transactions on Knowledge Discovery from Data (**TKDD**), Knowledge and Information Systems (**KAIS**).
- *External Conference Referee*: ICDM 2012, KDD 2012, KDD 2011, SDM 2011, WWW 2011, WSDM 2011, SocialCom 2011, KDD 2010, SocialCom 2010.
- *External Journal Reviewer*: IEEE Internet Computing.

References

Available upon request.