

Xifeng Yan

Research Staff Member
IBM T. J. Watson Research Center

Tel: (914) 784-7127
Email: xifeng@gmail.com
xifengyan@us.ibm.com
<http://www.xifengyan.net>

RESEARCH INTERESTS

Data mining, data management and machine learning, with emphasis on modeling, managing, and mining large-scale graphs and networks in bioinformatics, social networks, and computer systems. I am also working on enterprise search and text mining, integrated with social network analysis.

WORK EXPERIENCE

2006 – present **IBM T. J. Watson Research Center**
Research Staff Member, Data-Intensive Systems and Analytics

EDUCATION

2006 **University of Illinois at Urbana-Champaign**
Ph.D. in Computer Science
Dissertation: Mining, Indexing, and Similarity Search in Large Graph Datasets
Advisor: Prof. Jiawei Han.

2001 **State University of New York at Stony Brook**
M.S. in Computer Science

1999 **East-China Institute of Computer Technology, China**
M.E. in Computer Engineering

1997 **Zhejiang University, China**
B.E. in Computer Engineering, Honor Degree (Mixed Class)

HONORS & AWARDS

2007 ACM-SIGMOD Dissertation Award Honorable Mention, June 2007

2007 Best Student Paper, Proc. of 2007 IEEE International Conference on Data Engineering (ICDE), 2007

2007 Best Student Paper, Proc. of 2007 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2007

2006 One of the Two Department's Nominees for ACM Doctoral Dissertation Award, Department of Computer Science, University of Illinois at Urbana-Champaign, 2006

2006 The Computer Science Department's nominee for the Ross J. Martin Award, College of Engineering, University of Illinois at Urbana-Champaign, 2006

2005 Best Student Paper Runner-up Award, Proc. of 2005 ACM Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD), 2005

2004, 2005 Data Mining Research Gold Awards, Data Mining Research Group, University of Illinois at Urbana-Champaign, 2004, 2005

2004 IBM Invention Achievement Award, IBM T.J. Watson Research Center, May 2004

2003, 2004 Excellent Teaching Assistant Award, Department of Computer Science, University of Illinois at Urbana-Champaign, 2003, 2004

2001	Outstanding Teaching Assistant Award, Department of Computer Science, State University of New York at Stony Brook, Fall 2001
1993	First-class Scholarship, Zhejiang University, China
1992	First Prize, National Mathematical Contest, China
1989	First Prize, National Mathematical Contest (middle school), China

PATENTS

2007	“System and Method for Graph Classification with Skewed Class Distribution,” by H. Cheng, X. Yan, W. Fan and P. S. Yu, US patent filed as Docket YOR8-2007-0684-US1 by IBM (Dec., 2007).
2007	“A System for Entity Search and a Method for Entity Scoring in a (Linked) Document Database,” by T. Cheng, X. Yan, and Kevin Chen-Chuan Chang, Invention Disclosure (TF07088) by University of Illinois at Urbana-Champaign (Sep., 2007).
2005	“System and Method for Efficiently Performing Similarity Searches of Structural Data,” by X. Yan and P. S. Yu, US patent filed as Docket YOR9-2005-0047-US1 by IBM (April, 2005).
2004	“System and Method for Graph Indexing,” by X. Yan and P. S. Yu, US patent filed as Docket YOR9-2004-0013-US1 by IBM (April, 2004).

INVITED TALKS & TUTORIALS

Aug 2008	“Graph Mining and Graph Kernels”, by K. Borgwardt and X. Yan. Conference Tutorial, The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV, August 24 - 27, 2008
Apr 2007	“Mining and Searching Graphs and Structures”, by X. Yan. Guest Lectures in Graduate Course “Computational Biology”, Molecular and Computational Biology, University of Southern California, April 9, 2007.
Aug 2006	“Mining and Searching Graphs and Structures”, by J. Han, X. Yan, and P. S. Yu. Conference Tutorial, The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’06), Philadelphia, PA, August 20 - 23, 2006
Apr 2006	“Mining, Indexing, and Similarity Search in Graphs and Complex Structures”, by J. Han, X. Yan, and P. S. Yu. Conference Tutorial, The 22nd International Conference on Data Engineering (ICDE’06), Atlanta, Georgia, April 6 2006.
Nov 2005	“Mining and Searching of Graph-Structured Databases”, by J. Han, X. Yan, and Philip S. Yu. Invited Tutorial, The Fifth IEEE International Conference on Data Mining (ICDM’05), Houston, Texas, November 27, 2005.
Apr 2005	“Graph Pattern Mining and Searching”, by X. Yan. NEW-CEGS SEMINAR, Center of Excellence in Genomic Science, Molecular and Computational Biology, University of Southern California, April 22, 2005.

RESEARCH EXPERIENCE

2006 – present	IBM T. J. Watson Research Center Research Staff Member, Data-Intensive Systems and Analytics Research projects: Large-scale intrusion alert network modeling and machine learning; network analysis and graph mining across massive biological networks; enterprise search and text mining, integrated with social network analysis.
----------------	---

- 2002 – 2006 **University of Illinois at Urbana-Champaign**
 Research Assistant, Advisor: Prof. Jiawei Han
 Research projects: Mining sequential and graph patterns; indexing, querying, and mining of complex biological structures; statistical model-based software bug localization; graph pattern analysis including pattern compression, summarization and selection, and its applications in indexing, classification, and clustering.
- Summer 2004 **IBM T.J. Watson Research Center**
 Summer 2003 Research Intern, Advisor: Dr. Philip S. Yu
 Invented systems and methods for (1) indexing and searching large-scale graph datasets using discriminative graph patterns, and (2) performing efficient similarity search of structural data.
- Summer 2001 **FalconStor Inc.**
 Technical Staff, Storage over IP Engineering Group, Advisor: Alan Chen (VP)
 Participated in developing IP-based network storage infrastructure using Ethernet, SCSI and Fiber Channel. Responsible for the snapshot part that can backup through SCSI and Gigabit Ethernet. Developed the first SCSI interface for Compaq SmartArray disks.
- 2001 **State University of New York at Stony Brook**
 Research Assistant, Center of Visual Computing, Advisor: Prof. Arie Kaufman
 Studied volume rendering, virtual colonoscopy, and image-based rendering.
- Summer 2000 **Silicon Graphics Inc.**
 Research Intern, Data Mining and Visualization Group, Advisor: Dr. Alan Norton
 Participated in developing a browser-based thin client in MineSet, an integrated system for data mining. Simplified its data publishing/distribution infrastructure.
- 1998 – 1999 **East-China Institute of Computer Technology, China**
 Research Assistant, Advisor: Prof. Hansheng Chen
 Designed and implemented an ADA thin binding to DCE (Distributed Computing Environment).
- 1995 – 1997 **Zhejiang University, China**
 Research Assistant, State Key Lab of CAD&CG, Advisor: Prof. Jiaoying Shi
 Completed a volume rendering module in GIVE (General Interactive Visualization Environment), the first visualization environment designed in China. Designed a parallel rendering toolkit based on PVM (Parallel Virtual Machine). Developed an image database management system to facilitate the storage, retrieval and query of digitized medical films.

TEACHING AND PRESENTATION EXPERIENCE

- Spring 2003 **Data Mining: Principles and Algorithms**
 Fall 2002 Graduate level and senior level (CS498 and CS412), UIUC (Prof. Jiawei Han)
 Designed one course project, graded assignments, and lectured two-week classes.
- Spring 2002 **Operating Systems**
 Fall 2001 Undergraduate level, CS323, UIUC (Prof. Klara Nahrstedt, Prof. Roy Campbell)
 Designed 6 quizzes, graded assignments and projects, assisted in preparing and grading exams.

Fall 2000	Introduction to Scientific Visualization
Spring 2000	Undergraduate level, CSE332, SUNY at Stony Brook (Prof. Klaus Mueller)
	Computer Architecture
	Undergraduate level, CSE345, SUNY at Stony Brook (Prof. Larry Wittie)
	Introduction to Database Systems
	Undergraduate/Graduate level, CSE305/405, SUNY at Stony Brook (Prof. Michael Kifer)
	Held office hours, graded assignments, projects, and exams
Spring 2000	Introduction to C&UNIX
Fall 1999	Undergraduate Level, CSE230, SUNY at Stony Brook (Prof. Tian-Li Teng)
	Lead weekly recitations, held office hours, graded assignments and exams
2007	Conference Presentation
	A Graph-Based Approach to Systematically Reconstruct Human Transcriptional Regulatory Modules, ISMB'07, Vienna, Austria
2006	Conference Presentation
	Searching Substructures with Superimposed Distance, ICDE'06, Atlanta, GA
2005	Conference Presentation
	Summarizing Itemset Patterns: A Profile-Based Approach, SIGKDD'05, Chicago, IL
	Mining Closed Relational Graphs with Connectivity Constraints, SIGKDD'05, Chicago, IL
	Substructure Similarity Search in Graph Databases, SIGMOD'05, Baltimore, MD
2003	Conference Presentation
	CloseGraph: Mining Closed Frequent Graph Patterns, SIGKDD'03, Seattle, WA
	CloSpan: Mining Closed Sequential Patterns in Large Datasets, SDM'03, San Francisco, CA

SOFTWARE RELEASES

2007	NeMo: A graph-based approach to systematically reconstruct human transcriptional regulatory modules, http://zhoulab.usc.edu/NeMo/source.htm (USC, MIT/Whitehead, UCSC, SFU)
2006	gIndex: Index and search large graph sets, http://www.xifengyan.net/software.htm (75 citations)
2005	CODENSE Frequent Coherent Dense Subgraphs Mining Package, http://zhoulab.usc.edu/CODENSE/ (USC, NIH, Aristotle Univ.)
2005	CloSpan: Closed Sequential Pattern Mining Package, http://illimine.cs.uiuc.edu (110 citations, Rational Software)
2005	CloseGraph: Frequent Closed Graph Mining Package (165 citations)
2004	gSpan: Frequent Graph Mining Package, http://illimine.cs.uiuc.edu (338 citations)

PUBLICATIONS

Refereed Journal Papers

- [1] "Frequent Pattern Mining: Current Status and Future Directions", by J. Han, H. Cheng, D. Xin and X. Yan, Data Mining and Knowledge Discovery (**DMKD**), the 10th Anniversary Issue, invited, 2007.

- [2] “On compressing frequent patterns”, D. Xin, J. Han, X. Yan, H. Cheng: *Data Knowl. Eng. (DKE)*, 60(1): 5-29, 2007
- [3] “Integrative Array Analyzer: A Software Package for Analysis of Cross-platform and Cross-species Microarray Data”, by F. Pan, K Kamath, K. Zhang, S. Pulapura, A. Achar, J. Nunez-Iglesias, Y. Huang, X. Yan, J. Han, H. Hu, M. Xu, J. Hu, and X. Jasmine Zhou, *Bioinformatics*, Vol. 22 no. 13: 1665–1667, 2006.
- [4] “Feature-based Substructure Similarity Search”, by X. Yan, P. S. Yu, and J. Han, *ACM Transactions on Database Systems (TODS)*, 31 (4): 418 – 1453, December, 2006 (invited submission: selected from the accepted papers in SIGMOD’05).
- [5] “SOBER: Statistical Model-based Fault Localization”, by C. Liu, L. Fei, X. Yan, J. Han, and S. Midkiff, *IEEE Transactions on Software Engineering (TSE)*, 32(10): 831-848, 2006.
- [6] “Graph Indexing Based on Discriminative Frequent Structure Analysis”, by X. Yan, P. S. Yu, and J. Han, *ACM Transactions on Database Systems (TODS)*, 30(4):960 – 993, December, 2005 (invited submission: selected from the accepted papers in SIGMOD’04).
- [7] “TSP: Mining Top-K Closed Sequential Patterns”, by P. Tzvetkov, X. Yan, and J. Han, *Knowledge and Information Systems: An International Journal (KAIS)*, 7:438-457, 2005.
- [8] “From Sequential Pattern Mining to Structured Pattern Mining: A Pattern-Growth Approach”, by J. Han, J. Pei, and X. Yan, *Journal of Computer Science and Technology (JCST)*, 19(3): 257 – 279, 2004.

Refereed Conference Papers

- [1] “Efficient Ticket Routing by Resolution Sequence Mining”, by Q. Shao, Y. Chen, S. Tao, X. Yan, N. Anerousis, *Proc. of 2008 Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD’08)*, 2008. (long presentation, acceptance rate, 10%)
- [2] “Direct Mining of Discriminative and Essential Graphical and Itemset Features via Model-based Search Tree”, by W. Fan, K. Zhang, H. Cheng, J. Gao, X. Yan, J. Han, P. S. Yu, O. Verscheure, *Proc. of 2008 Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD’08)*, 2008. (long presentation, acceptance rate, 10%)
- [3] “Mining Significant Graph Patterns by Leap Search”, X. Yan, H. Cheng, J. Han, and P. S. Yu, *Proc. 2008 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’08)*, Vancouver, Canada, 2008
- [4] “Direct Discriminative Pattern Mining for Effective Classification”, H. Cheng, X. Yan, J. Han, and P. S. Yu, *Proc. 2008 Int. Conf. on Data Engineering (ICDE’08)*, 169-178, Cancun, Mexico, 2008 (acceptance rate, 12%).
- [5] “gApprox: Mining Frequent Approximate Patterns from a Massive Network”, C. Chen, X. Yan, F. Zhu, and J. Han, *Proc. 2007 Int. Conf. on Data Mining (ICDM’07)*, Omaha, NE, 2007. (short paper, acceptance rate, 19%)
- [6] “Efficient Discovery of Frequent Approximate Sequential Patterns”, F. Zhu, X. Yan, J. Han, and P. S. Yu., *Proc. 2007 Int. Conf. on Data Mining (ICDM’07)*, Omaha, NE, Oct. 2007. (short paper, acceptance rate, 19%)
- [7] “Towards Graph Containment Search and Indexing”, C. Chen, X. Yan, P. S. Yu, J. Han, D.-Q. Zhang and X. Gu, *Proc. of 2007 Int. Conf. on Very Large Data Bases (VLDB’07)*, Vienna, Austria, 2007 (acceptance rate, 17.5%)
- [8] “Entity Search: Search Directly and Holistically”, T. Cheng, X. Yan and K. Chang, *Proc. of 2007 Int. Conf. on Very Large Data Bases (VLDB’07)*, Vienna, Austria, 2007 (acceptance rate, 16.4%)
- [9] “A Graph-Based Approach to Systematically Reconstruct Human Transcriptional Regulatory Modules”, by X. Yan, M. Mehan, Y. Huang, M. S. Waterman, P. S. Yu, and X. Zhou. the 15th Annual Int. Conf. on Intelligent Systems for Molecular Biology (*ISMB’07*), Vienna, Austria, 2007 (acceptance rate, 15%)
- [10] “Systematic Discovery of Functional Modules and Context-Specific Functional Annotation of Human Genome”, by Y. Huang, H. Li, H. Hu, X. Yan, M. S. Waterman, H. Huang, and X. Zhou. the 15th Annual Int. Conf. on Intelligent Systems for Molecular Biology (*ISMB’07*), Vienna, Austria, 2007 (acceptance rate, 15%)
- [11] “Discriminative Frequent Pattern Analysis for Effective Classification”, by H. Cheng, X. Yan, J. Han,

- and C. Hsu. Proc. of 2007 Int. Conf. on Data Engineering (**ICDE'07**), Istanbul, Turkey, 2007. (acceptance rate, 18.5%)
- [12] “Mining Colossal Frequent Patterns by Core Pattern Fusion”, by F. Zhu, X. Yan, J. Han, P. S. Yu, and H. Cheng. Proc. of 2007 IEEE Int. Conf. on Data Engineering (**ICDE'07**), Istanbul, Turkey, 2007. (**Best Student Paper**) (acceptance rate, 18.5%)
- [13] “gPrune: A Constraint Pushing Framework for Graph Pattern Mining”, by F. Zhu, X. Yan, J. Han, P. S. Yu. Proc. of 2007 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'07), Nanjing, China 2007. (**Best Student Paper**) (acceptance rate, 4.66%)
- [14] “Extracting Redundancy-aware Top-k Patterns”, by D. Xin, H. Cheng, X. Yan, J. Han, Proc. of 2006 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'06**), 2006. (acceptance rate, 11%)
- [15] “Mining Control Flow Abnormality for Logic Error Isolation”, by C. Liu, X. Yan, and J. Han, Proc. of 2006 SIAM Int. Conf. on Data Mining (**SDM'06**), 106-117, Bethesda, MD, 2006. (acceptance rate, 16%)
- [16] “Searching Substructures with Superimposed Distance”, by X. Yan, F. Zhu, J. Han, and P. S. Yu, Proc. of 2006 Int. Conf. on Data Engineering (**ICDE'06**), Atlanta, GA, 2006. (acceptance rate, 20%)
- [17] “Community Mining from Multi-Relational Networks”, by D. Cai, Z. Shao, X. He, X. Yan, J. Han, Proc. of 2005 European Conf. on Principles and Practice of Knowledge Discovery in Databases (**PKDD'05**), 445 – 452, Porto, Portugal 2005. (acceptance rate, 28%)
- [18] “SOBER: Statistical Model-based Bug Localization”, by C. Liu, X. Yan, L. Fei, J. Han, and S. Midkiff, Proc. of 2005 ACM SIGSOFT Symp. on the Foundations of Software Engineering (**FSE'05**), 286 – 295, Lisbon, Portugal, 2005. (acceptance rate, 16%)
- [19] “Mining Compressed Frequent Pattern Sets”, by D. Xin, J. Han, X. Yan and H. Cheng, Proc. of 2005 Int. Conf. on Very Large Data Bases (**VLDB'05**), 709-720, Trondheim, Norway, 2005. (acceptance rate, 16.5%)
- [20] “Mining Closed Relational Graphs with Connectivity Constraints”, by X. Yan, X. Jasmine Zhou, and J. Han, Proc. of 2005 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'05**), 324 – 333, Chicago, IL, 2005. (acceptance rate, 9%)
- [21] “Summarizing Itemset Patterns: A Profile-Based Approach”, by X. Yan, H. Cheng, J. Han, and D. Xin, Proc. of 2005 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'05**) (**Best Student Paper Runner-up**), 314 – 323, Chicago, IL, 2005. (acceptance rate, 9%)
- [22] “Mining Coherent Dense Subgraphs Across Massive Biological Networks for Functional Discovery”, by H. Hu, X. Yan, Y. Huang, J. Han, X. Jasmine Zhou, Proc. of 2005 Int. Conf. on Intelligent Systems for Molecular Biology (**ISMB'05**), 213-221, Detroit, MI, 2005 (also **Bioinformatics** Vol. 21 Suppl. 2005). (acceptance rate, 13%)
- [23] “Substructure Similarity Search in Graph Databases”, by X. Yan, P. S. Yu, and J. Han, Proc. of 2005 Int. Conf. on Management of Data (**SIGMOD'05**), 766 – 777, Baltimore, MD, 2005. (acceptance rate, 15%)
- [24] “Mining Behavior Graphs for ‘Backtrace’ of Noncrashing Bugs”, by C. Liu, X. Yan, H. Yu, J. Han, and P. S. Yu, Proc. of 2005 SIAM Int. Conf. on Data Mining (**SDM'05**), 286 – 297, Newport Beach, CA, 2005. (acceptance rate, 18%)
- [25] “SeqIndex: Indexing Sequences by Sequential Pattern Analysis”, by H. Cheng, X. Yan, and J. Han, Proc. of 2005 SIAM Int. Conf. on Data Mining (**SDM'05**), 601 – 605, Newport Beach, CA, 2005. (short paper, acceptance rate, 36%)
- [26] “Mining Closed Relational Graphs with Connectivity Constraints”, by X. Yan, X. Jasmine Zhou, J. Han, Proc. of 2005 Int. Conf. on Data Engineering (**ICDE'05**), 357 – 358, Tokyo, Japan, 2005. (short paper, acceptance rate, 19%).
- [27] “Graph Indexing: A Frequent Structure-based Approach”, by X. Yan, P. S. Yu, and J. Han, Proc. of 2004 Int. Conf. on Management of Data (**SIGMOD'04**), 335 – 346, Paris, France, 2004. (acceptance rate, 16%)
- [28] “IncSpan: Incremental Mining of Sequential Patterns in Large Database”, by H. Cheng, X. Yan, and J. Han, Proc. of 2004 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'04**), 527 – 532, Seattle, WA, 2004. (short paper, acceptance rate 25%)

- [29] “CloseGraph: Mining Closed Frequent Graph Patterns”, by X. Yan and J. Han, Proc. of 2003 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'03**), 286 – 295, Washington, DC, 2003. (acceptance rate, 13%)
- [30] “TSP: Mining Top-K Closed Sequential Patterns”, by P. Tzvetkov, X. Yan, and J. Han, Proc. of 2003 Int. Conf. on Data Mining (**ICDM'03**), 347 – 354, Melbourne, FL, 2003. (acceptance rate, 12%)
- [31] “CloSpan: Mining Closed Sequential Patterns in Large Datasets”, by X. Yan, J. Han, and R. Afshar, Proc. of 2003 SIAM Int. Conf. on Data Mining (**SDM'03**), 166 – 177, San Francisco, CA, 2003. (acceptance rate, 20%)
- [32] “gSpan: Graph-Based Substructure Pattern Mining”, by X. Yan and J. Han, Proc. of 2002 Int. Conf. on Data Mining (**ICDM'02**), 721 – 724, Maebashi City, Japan, 2002 (Google Scholar Rank #2 for “graph mining”) (short paper, acceptance rate, 31%).
- [33] “Accelerating Volume Rendering with L-Buffer”, by X. Yan, W. Cai and J. Shi, Proc. of 1997 Int. Conf. on CAD&GRAPHICS, Shenzhen, China, Dec, 1997.

Refereed Book Chapters

- [1] “Mining Frequent Approximate Sequential Patterns,” by F. Zhu, X. Yan, J. Han and P. S. Yu. Next Generation Data Mining, 2008.
- [2] “Discovering evolutionary classifier over high speed non-static stream”, by J. Yang, X. Yan, J. Han, and W. Wang, Advanced Methods for Knowledge Discovery from Complex Data, S. Bandyopadhyay, U. Maulik, L. Holder, D. Cook (Eds.), Springer, 2005.
- [3] “Mining Frequent Patterns in Data Streams at Multiple Time Granularities”, by C. Giannella, J. Han, J. Pei, X. Yan, and P. S. Yu, Next Generation Data Mining, H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (eds.), AAAI/MIT, 2004.
- [4] “Sequential Pattern Mining by Pattern-Growth: Principles and Extensions”, by J. Han, J. Pei, and X. Yan, Recent Advances in Data Mining and Granular Computing (Mathematical Aspects of Knowledge Discovery), W. Chu and T. Lin (eds.), Springer Verlag, 2004.

Workshop Papers, Demos, and Technical Reports

- [1] “EasyTicket: A Ticket Routing Recommendation Engine for Enterprise Problem Resolution”, by Q. Shao, Y. Chen, S. Tao, X. Yan, N. Anerousis, Proc. of 2008 Int. Conf. on Very Large Data Bases (**VLDB'08**), Auckland, New Zealand, 2008. (system demo, acceptance rate, 28%).
- [2] “Combining near-optimal feature selection with gSpan”, by K. Borgwardt, X. Yan, M. Thoma, H. Cheng, A. Gretton, L. Song, A. Smola, J. Han, P. Yu, H.-P. Kriegel, 6th Int. Workshop on Mining and Learning with Graph (**MLG'08**), Helsinki, Finland.
- [3] “Entity Search: Search Directly and Holistically”, by T. Cheng, X. Yan, K. Chang, Proc. of 2007 Int. Conf. on Management of Data (**SIGMOD'07**), Beijing, China, 2007. (system demo)
- [4] “BioArrayMine: A software package for integrative analysis of cross-platform and cross-species microarray data”, by F. Pan, K. Kamath, H. Hu, Y. Huang, K. Zhang, M. Xu, X. Yan, J. Han, and X. Jasmine Zhou, Proc. of 2005 Int. Conf. on Intelligent Systems for Molecular Biology (**ISMB'05**), Detroit, MI, 2005. (system demo)
- [5] “GraphMiner: A Structural Pattern Mining System for Large Disk-based Graph Databases and Its Applications”, by W. Wang, C. Wang, Y. Zhu, B. Shi, J. Pei, X. Yan, and J. Han, Proc. of 2005 Int. Conf. on Management of Data (**SIGMOD'05**), 879 – 881, Baltimore, MD, 2005. (system demo)
- [6] “Mining Hidden Community in Heterogeneous Social Networks”, by D. Cai, Z. Shao, X. He, X. Yan, and J. Han. Technical Report UIUCDCS-R-2005-2538, Department of Computer Science, University of Illinois at Urbana-Champaign, 2005.
- [7] “Using Data Mining for Discovering Patterns in Autonomic Storage Systems”, by Z. Li, S. Srinivasan, Z. Chen, Y. Zhou, P. Tzvetkov, X. Yan, and J. Han, ACM Workshop on Algorithms and Architectures for Self-Managing Systems, Proc. of 2003 Federated Computing Research Conference (**FCRC'03**).
- [8] “gSpan: Graph-Based Substructure Pattern Mining”, by X. Yan and J. Han, Technical Report UIUCDCS-R-2002-2296, Department of Computer Science, University of Illinois at Urbana-

Champaign, 2002.

- [9] “A Framework for Continuous Quantile Computation over Sensor Networks”, by X. Yan, J. Yang, J. Han and W. Wang, Technical Report UIUCDCS-R-2003-2382, Department of Computer Science, University of Illinois at Urbana-Champaign, 2003.

PROFESSIONAL SERVICES

- Program Co-Chair: 2007 ICDM Workshop on Mining Graphs and Complex Structures (MGCS'2007)
- Program Committee Member: The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008 (SIGKDD'08); the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), 2006; SIAM Conference on Data Mining, 2007, 2008 (SDM'07, SDM'08); the 5th International Workshop on Mining and Learning with Graphs (MLG'07), 2007
- NSF Panelist, 2008.
- Volunteer: ACM Int. Conf. on Knowledge Discovery and Data Mining, 2005 (SIGKDD).
- Referee: VLDB journal, 2004; ACM Trans. on Database systems (TODS) 2003, 2004, 2005; ACM Transactions on Knowledge Discovery from Data (TKDD) 2007; IEEE Trans. on Knowledge and Data Engineering, TKDE 2004, 2005, 2006; Data Mining and Knowledge Discovery (DAMI) 2006, 2007; Bioinformatics 2005, 2006; Machine Learning Journal 2007
- External Referee: SIGMOD 2002, 2003, 2006, 2007; ICDE 2005, 2007, 2008; SIGKDD 2003, 2004; EDBT 2004; ICDM 2002, 2003, 2004, 2006; LinkKDD 2007; SIGKDD 2007; PAKDD 2005; PKDD 2003; SDM 2003; SAC DM 2004.
- Proposal Referee: National Aeronautics and Space Administration, Mini-Solicitation for Master NASA Research Announcement: Data Mining and Discovery Technology Demonstrations, 2004.