Xifeng Yan

Assistant Professor
Department of Computer Science
Rm 1111, Harold Frank Hall
University of California
Santa Barbara, CA 93106-5110

Email: xyan@cs.ucsb.edu
Fax: (805) 893-8553
http://www.cs.ucsb.edu/~xyan

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, the Web, and computer systems.

EDUCATION

2006	**University of Illinois at Urbana-Champaign** Ph.D. in Computer Science
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

2011	IEEE ICDM 10-year Highest Impact Paper Award
2010	NSF CAREER Award
2009	Microsoft New Faculty Fellow Finalist (one of 10 selected across North America)
2008-present	The Venkatesh Narayanamurti Chair, Univ. of California at Santa Barbara
2007	ACM-SIGMOD Dissertation Award Runner-Up, 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
2005	Best Student Paper Runner-up Award, Proc. of 2005 ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 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

PATENTS

2008	"Method and Apparatus for problem diagnosis using problem description information", by P. Zhou, B. Gill, W. Belluomini, X. Yan, Disclosure ARC8-2008-0126 by IBM (in processing)
2008	"System and Method for Collective Search and Recommendation Using Social Networks," by X. Yan, C.-Y. Lin, V. Griffiths-Fisher and K.-L. Wu, Disclosure YOR8-2008-0847 by IBM (in processing)
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), accepted as US Patent #7974978 (July, 2011)

TALKS & TUTORIALS

Apr 2011	Graph Pattern Mining, Search, and Distributed Processing," by X. Yan, University of Wisconsin at Madison, Database Seminar
Jan 2011	"Graph Pattern Mining and Distributed Processing," by X. Yan, INARC/NS-CTA, ARL
Nov 2010	"Graph Pattern Mining and System," by X. Yan, Microsoft Research Asia, Beijing.
Sept 2010	"Graph Pattern Mining," by X. Yan, USC Guest Lecture, in Computer Science 599 "Selected Topics in Data Analytics"
Jul 2010	"Mining Heterogeneous Information Networks," by J. Han, Y. Sun, X. Yan, P. S. Yu, Conference Tutorial, the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD'10), 2010
Jun 2010	"Mining Knowledge from Databases: An Information Network Analysis Approach," by J. Han, Y. Sun, X. Yan, P. S. Yu, Conference Tutorial, ACM Int. Conf. on Management of Data (SIGMOD'10), 2010
Oct 2009	"From Data to Knowledge, The New Frontier" by X.Yan, University of California at Santa Barbara, Venky Chair talk
Mar 2009	"Scalable OLAP and Mining of Information Networks," by Jiawei Han, X. Yan, and P. S. Yu, Conference Tutorial, 12th International Conference on Extending Database Technology (EDBT'09), Saint-Petersburg, Russia, March 23-26, 2009
Dec 2008	"Integration of Classification and Pattern Mining," by H. Cheng, J. Han, X. Yan, and P. S. Yu, Conference Tutorial, The 2008 IEEE International Conference on Data Mining (ICDM'08), Pisa, Italy, December 15-19, 2008
Nov 2008	"Robust Methods for Mining Interesting Graph Patterns," by X. Yan, Univ. of Arizona, 2008
Aug 2008	"Graph Mining and Graph Kernels," by K. Borgwardt and X. Yan. Conference Tutorial, The 14th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD'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 International Conference on Knowledge Discovery and Data Mining (SIGKDD'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

WORK EXPERIENCE

2008 –	**University of California at Santa Barbara** Assistant Professor, Department of Computer Science
2006 – 2008	**IBM T. J. Watson Research Center** Research Staff Member, Data-Intensive Systems and Analytics, Manager: Dr. Philip S. Yu/Dr. Kun-Lung Wu
2002 – 2006	**University of Illinois at Urbana-Champaign** Research Assistant, Advisor: Prof. Jiawei Han
Summer 2001	**FalconStor Inc.** Technical Staff, Storage over IP Engineering Group, Advisor: Alan Chen (VP)
Summer 2000	**Silicon Graphics Inc.** Research Intern, Data Mining and Visualization Group, Advisor: Dr. Alan Norton

TEACHING EXPERIENCE

Spring 2011	**Advanced Data Mining** Graduate level, UC at Santa Barbara (CS290D), Lecture
Winter 2011	**Artificial Intelligence** Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Fall 2010	**Dynamic Graph Modeling** Graduate level, UC at Santa Barbara (CS595D), Seminar **Advanced Information Systems** Graduate level, UC at Santa Barbara (CS595J), Seminar
Spring 2010	**Design and Implementation Techniques of Database Systems** Undergraduate level, UC at Santa Barbara (CS174B), Lecture
Winter 2010	**Advanced Data Mining** Graduate level, UC at Santa Barbara (CS290D), Lecture **Readings in Information Networks** Graduate level, UC at Santa Barbara (CS595D), Seminar
Fall 2009	**Data and Knowledge Bases** Graduate level, UC at Santa Barbara (CS273), Lecture **BPM: Models, Process Mining, and BI** Graduate level, UC at Santa Barbara (CS595D), Seminar
Spring 2009	**Data Mining: Principles and Algorithms** Graduate level, UC at Santa Barbara (CS290D), Lecture

Winter 2009	**Graph Mining** Graduate level, UC at Santa Barbara (CS595D), Seminar

SOFTWARE RELEASES

2011	Sedge: A Self Evolving Distributed Graph Processing Environment (IBM, USC)
2007	NeMo: A graph-based approach to systematically reconstruct human transcriptional regulatory modules, http://zhoulab.usc.edu/NeMo/source.htm (USC, MIT/Whitehead, UCSC, UNC, SFU)
2006	gIndex: Index and search large graph sets, http://www.xifengyan.net/software.htm (259 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 (392 citations, Rational Software)
2005	CloseGraph: Frequent Closed Graph Mining Package (391 citations)
2004	gSpan: Frequent Graph Mining Package, http://www.cs.ucsb.edu/~xyan/software (872 citations)

PUBLICATIONS (h-index: 30, citations: 5,000+ per Google Scholar)

Refereed Journal Papers

[1] "Discriminative Frequent Subgraph Mining with Optimality," by M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. S. Yu, X. Yan, and K. Borgwardt, Statistical Analysis and Data Mining (**SAM**), 3(5), 302– 318, 2010

[2] "Graph OLAP: A Multi-Dimensional Framework for Graph Data Analysis," by C. Chen, X. Yan, F. Zhu, J. Han, P. S Yu, Knowledge and Information Systems: An International Journal **(KAIS)**, 21(1), 41– 63, 2009

[3] "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, 15(1), 55–86, 2007

[4] "On Compressing Frequent Patterns," D. Xin, J. Han, X. Yan, H. Cheng, Data Knowl. Eng. (**DKE**), 60(1): 5– 29, 2007

[5] "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

[6] "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)

[7] "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.

[8] "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)

[9] "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.

[10] "From Sequential Pattern Mining to Structured Pattern Mining: A Pattern-Growth Approach," by J. Han, J. Pei, and X. Yan, J. of Computer Science and Technology (**JCST**), 19(3): 257 – 279, 2004.

Refereed Conference Papers

[1] "Efficient multicasting for delay tolerant networks using graph indexing," by M. Mongiovi, A. Singh, X. Yan, B. Zong, K. Psounis. Proc. 2012 Int. Conf. on Computer Communications (**INFOCOM'12**), (acceptance rate, 18%).

[2] "PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks," by Y. Sun, J. Han, X. Yan, P. S. Yu, T Wu, Proc. 2011 Int. Conf. on Very Large Data Bases (**VLDB'11)**, 992 – 1003, Aug 2011 (acceptance rate, 18%).

[3] "Mining Top-K Large Structural Patterns in a Massive Network," by F. Zhu, Q. Qu, D. Lo, X. Yan, J. Han, and P. Yu, Proc. 2011 Int. Conf. on Very Large Data Bases (**VLDB'11**), 807 – 818, Aug 2011 (acceptance rate, 18%).

[4] "Neighborhood Based Fast Graph Search in Large Networks," by A. Khan, N. Li, Z. Guan, X. Yan, S. Chakraborty, and S. Tao, Proc. 2011 Int. Conf. on Management of Data (**SIGMOD'11**), 901 – 912 , June 2011 (acceptance rate, 23%).

[5] "Assessing and Ranking Structural Correlations in Graphs," by Z. Guan, J. Wu, Z. Yun, A. Singh, X. Yan, Proc. 2011 Int. Conf. on Management of Data (**SIGMOD'11**), 937 – 948, June 2011 (acceptance rate, 23%).

[6] "On Flow Authority Discovery in Social Networks," by C. Aggarwal, A. Khan, and X. Yan, Proc. 2011 SIAM International Conference on Data Mining (**SDM'11**), 522 – 533, Mesa, Arizona, April 2011 (acceptance rate, 25%).

[7] "Efficient Topological OLAP on Information Networks," by Q. Qu, F. Zhu, X. Yan, J. Han, P. Yu, and H. Li, Proc. 2011 Int. Conf. on Database Systems for Advanced Applications (**DASFAA**'**11**), 389 – 403, Hong Kong, April 2011.

[8] "Content-Aware Resolution Sequence Mining for Ticket Routing," by P. Sun, S. Tao, X. Yan, N. Anerousis, Y. Chen, the 8th International Conference on Business Process Management (**BPM' 2010**), 243 – 259, Sep. 2010 (acceptance rate, 14%).

[9] "Generative Models for Ticket Resolution in Expert Networks," by G. Miao, L. Moser, X. Yan, S. Tao, Y. Chen, N. Anerousis, Proc. 2010 Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD**'**10**), 733 – 742, Washington DC, July 2010 (acceptance rate, 17%)

[10] "Assessing Expertise Awareness in Resolution Networks," by Y. Chen, S. Tao, X. Yan, N. Anerousis, Q. Shao, Proc. 2010 International Conference on Social Networks Analysis and Mining (**ASONAM'10**), 128 – 135, Odense, Denmark, Aug. 2010

[11] "Towards Proximity Pattern Mining in Large Graphs," A. Khan, X. Yan and K.-L. Wu, Proc. 2010 Int. Conf. on Management of Data **(SIGMOD'10**), 867 – 878, Indianapolis, IN, June 2010 (acceptance rate: 21%)

[12] "Synthesizing Near-Optimal Malware Specifications from Suspicious Behaviors," M. Fredrikson, M. Christodorescu, S. Jha, R. Sailer, X. Yan, the 31st IEEE Symp. on Security & Privacy (**Oakland'10**), 45 – 60, Oakland, May 2010 (acceptance rate: 11%)

[13] "Mining Diversity on Networks," L. Liu, F. Zhu, C. Chen, X. Yan, J. Han, P. S. Yu, and S. Yang, Proc. 2010 Int. Conf. on Database Systems for Advanced Applications (**DASFAA'10**), 384 – 398, Tsukuba , Japan, March 2010 (acceptance rate: 23%)

[14] "Cross-Selling Optimization for Customized Product Promotion," by N. Li, Y. Yang, X. Yan, Proc. 2010 SIAM International Conference on Data Mining (**SDM'10**), 918 – 929, Columbus, OH, April 2010 (acceptance rate: 23%)

[15] "Top-K Aggregation Queries over Large Networks," by X. Yan, B. He, F. Zhu, and J. Han, Proc. 2010 Int. Conf. on Data Engineering (**ICDE'10**), 377 – 380, Long Beach, LA, Mar. 2010 (short

paper, acceptance rate: 20%)

[16] "Mining Graph Patterns Efficiently via Randomized Summaries," by C. Chen, C. Lin, M. Fredrikson, M. Christodorescu, X. Yan, and J. Han, Proc. 2009 Int. Conf. on Very Large Data Bases (**VLDB'09**), 742 – 753, Lyon, France, Aug. 2009 (acceptance rate: 17%)

[17] "Identifying Bug Signatures Using Discriminative Graph Mining," by H. Cheng, D. Lo, Y. Zhou, X. Wang and X. Yan, Proc. 2009 Int. Symp. On Software Testing and Analysis (**ISSTA'09**), 141-152, Chicago, IL 2009 (27% acceptance rate)

[18] "Near-Optimal Supervised Feature Selection among Frequent Subgraphs," by M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. S. Yu, X. Yan, and K. Borgwardt, Proc. 2009 SIAM Int. Conf. on Data Mining (**SDM'09**), 1075 – 1086, Sparks, NV, April 2009 (acceptance rate: 16%**)**

[19] "SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence," by C. Lin, N. Cao, S. Liu, S. Papadimitriou, J. Sun, X. Yan, Proc. of 2009 Int. Conf. on Data Engineering (**ICDE'09**), 1483 – 1486, Shanghai, China, March 2009 (Industry Demo)

[20] "Graph OLAP: Towards Online Analytical Processing on Graphs," by C. Chen, X. Yan, F. Zhu, J. Han, and P. S. Yu, Proc. 2008 Int. Conf. on Data Mining (**ICDM'08**), Pisa, Italy, 2008. (acceptance rate: 9.7%)

[21] "On Effective Presentation of Graph Patterns: A Structural Representative Approach," by C. Chen, X. Lin, X. Yan, and J. Han, Proc. 2008 ACM Conf. on Information and Knowledge Management (**CIKM'08**), 299 – 308, Napa Valley, CA 2008. (acceptance rate, 17%)

[22] "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**), 605 – 613, Las Vegas, NV, 2008 (long presentation, acceptance rate, 10%)

[23] "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**), 230 – 238, Las Vegas, NV, 2008 (long presentation, acceptance rate, 10%)

[24] "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**), 433 – 444, Vancouver, Canada, 2008 (acceptance rate, 18%).

[25] "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%)

[26] "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**), 445 – 450, Omaha, NE, 2007 (short paper, acceptance rate, 19%)

[27] "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**), 751 – 756, Omaha, NE, 2007 (short paper, acceptance rate, 19%)

[28] "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**), 926 – 937, Vienna, Austria, 2007 (acceptance rate, 17.5%)

[29] "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**), 387 – 398, Vienna, Austria, 2007 (acceptance rate, 16.4%)

[30] "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**), i577 – i586, Vienna, Austria, 2007 (acceptance rate, 15%)

[31] "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**), i222 – i229, Vienna, Austria, 2007 (acceptance rate, 15%)

[32] "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**), 716 – 725, Istanbul, Turkey, 2007. (acceptance rate, 18.5%)

[33] "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**), 706 – 715, Istanbul, Turkey, 2007. (**Best Student Paper**) (acceptance rate, 18.5%)

[34] "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**), 388 – 400, Nanjing, China 2007. (**Best Student Paper**) (acceptance rate, 4.66%)

[35] "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**), 444 – 453, 2006. (acceptance rate, 11%)

[36] "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%)

[37] "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**), 88 – 98, Atlanta, GA, 2006. (acceptance rate, 20%)

[38] "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%)

[39] "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%)

[40] "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%)

[41] "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%)

[42] "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%)

[43] "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%)

[44] "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%)

[45] "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%)

[46] "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%)

[47] "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%).

[48] "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%)

[49] "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%)

[50] "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%)

[51] "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%)

[52] "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%)

[53] "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 #1 for "graph pattern mining") (short paper, acceptance rate, 31%).

[54] "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] "Contrast Data Mining: Concepts, Algorithms and Applications," by Hong Cheng, Xifeng Yan, Jiawei Han, and Philip S. Yu, Contrast Data Mining: Concepts, Algorithms and Applications. G. Dong and J. Bailey (Eds.), Chapman & Hall/CRC 2012 (to appear)

[2] "Pattern Mining Across Many Massive Networks," by W. Li, H. Hu, Y. Huang, H. Li, M. R. Mehan, J. Nunez-Iglesias, M. Xu, X. Yan, and X. Zhou, Functional Coherence of Biological Networks. M. Koyuturk, S. Subramaniam, and A. Grama (Eds.), Springer, in press.

[3] "InfoNetOLAP: OLAP and Mining of Information Networks," by C. Chen, F. Zhu, X. Yan, J. Han, P. Yu, and R. Ramakrishnan, Link Mining: Models, Algorithms, and Applications, P. S. Yu, J. Han, C. Faloutsos (Eds.), Springer, 2010

[4] "Mining Large Information Networks by Graph Summarization," by C. Chen, X. Lin, M. Fredrikson, M. Christodorescu, X. Yan, and J. Han, Link Mining: Models, Algorithms, and Applications, P. S. Yu, J. Han, C. Faloutsos (Eds.), Springer, 2010

[5] "Discriminative Frequent Pattern-Based Graph Classification," by H. Cheng, X. Yan and J. Han, Link Mining: Models, Algorithms, and Applications, P. S. Yu, J. Han, C. Faloutsos (Eds.), Springer, 2010

[6] "Graph Indexing", by X. Yan and J. Han, Managing and Mining Graph Data, C. Aggarwal, H. Wang (Eds.), Springer, 2010

[7] "Mining Graph Patterns," by H. Cheng, X. Yan and J. Han, Managing and Mining Graph Data, C. Aggarwal, H. Wang (Eds.), Springer, 2010

[8] "Mining Frequent Approximate Sequential Patterns," by F. Zhu, X. Yan, J. Han and P. S. Yu. Next Generation Data Mining, 2008.

[9] "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.

[10] "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.

[11] "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] "Frequent Trajectory Mining on GPS Data," by N. Savage, S. Nishimura, N. Chavez, X. Yan, the Third

International Workshop on Location and the Web (**LocWeb 2010**) , Tokyo, Japan, 2010

[2] "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%).

[3] "Combining near-optimal feature selection with gSpan," by K. Borgwardt1, 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, 2008.

[4] "Report on the First International Workshop on Mining Graphs and Complex Structures," by L. Holder and X. Yan, **SIGMOD Record** 37(1): 53-55, 2008

[5] "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)

[6] "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)

[7] "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)

[8] "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.

[9] "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**).

[10] "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.

[11] "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

- Area Editor: Information Systems, an Elsevier journal (Starting in Oct, 2011)
- Program Vice Chair: 2011 International Conference on Data Mining (ICDM'11)
- Program Co-Chair: 2010 SIAM Data Mining Workshop on High Performance Analytics Algorithms, Implementations, and Applications
- Program Co-Chair: the 7th International Workshop on Mining and Learning with Graphs (MLG'09)
- Program Co-Chair: 2008 NIPS Workshop on Structured Input - Structured Output (SISO'08)
- Program Co-Chair: 2007 ICDM Workshop on Mining Graphs and Complex Structures (MGCS'2007)
- Guest Editor: The Machine Learning Journal (2010)
- Program Committee Member: The ACM Int. Conf. on Management of Data (SIGMOD'10, 11); The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'08-11); WWW 2012; The IEEE Int. Conf. Data Engineering (ICDE'10, 11, 12); The IEEE International Conference on Data Mining series (ICDM'09-11); SIAM Conference on Data Mining (SDM'07-11); the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06); the 5th International Workshop on Mining and Learning with Graphs (MLG'07)
- NSF Panelist, 2008, 2009, 2010, 2011
- Czech Science Foundation, 2011

- DOE panel 2009
- UC Discovery Panelist, 2009
- Volunteer: ACM Int. Conf. on Knowledge Discovery and Data Mining, 2005 (SIGKDD).
- Referee: JACM, 2010; VLDB journal, 2004; ACM Trans. on Database systems (TODS) 2003, 2004, 2005, 2009-2011; ACM Transactions on Knowledge Discovery from Data (TKDD) 2007-2011; IEEE Trans. on Knowledge and Data Engineering, TKDE 2004-2006, 2008-2011; Data Mining and Knowledge Discovery (DAMI) 2006, 2007; Data & Knowledge Engineering (DKE) 2008-2010; Knowledge and Information Systems (KAIS) 2009, 2010; Statistical Analysis and Data Mining (SAM), 2009-11; Bioinformatics 2005, 2006, 2008; Machine Learning Journal 2007, 2009, 2010; Information Processing & Management 2011.
- External Referee: SIGMOD 2002, 2003, 2006, 2007; ICDE 2005, 2007, 2008, 2009; SIGKDD 2003, 2004; EDBT 2004; ICDM 2002, 2003, 2004, 2006; LinkKDD 2007; SIGKDD 2007, 2008; PAKDD 2005; PKDD 2003; SDM 2003; SAC DM 2004; WWW 2004
- External Dissertation Examiner: NUS School of Computing
- Proposal Referee: National Aeronautics and Space Administration (NASA), 2004.