

# Xifeng Yan

Professor  
Venkatesh Narayanamurti Endowed Chair  
Department of Computer Science  
Rm 1111, Harold Frank Hall  
University of California  
Santa Barbara, CA 93106-5110

Email: [xyan@cs.ucsb.edu](mailto:xyan@cs.ucsb.edu)  
Phone: (805)893-3734  
Fax: (805) 893-8553  
<http://www.cs.ucsb.edu/~xyan>

## **RESEARCH INTERESTS**

---

The primary goal of my research is to develop fundamental concepts and new principles of data mining, design intelligent algorithms and build scalable systems. My current work has concentrated on three key areas: Modeling, managing, and mining large-scale, heterogeneous graphs; Structural and statistical analysis of texts and derived applications such as information extraction, question answering, and information trustworthiness; Intelligent solutions to various cross-domain problems in bioinformatics, business intelligence, computer security, computer systems, and social science.

## **EDUCATION**

---

2006	<b>University of Illinois at Urbana-Champaign</b> Ph.D. in Computer Science
2001	<b>State University of New York at Stony Brook</b> M.S. in Computer Science
1999	<b>East-China Institute of Computer Technology, China</b> M.E. in Computer Engineering
1997	<b>Zhejiang University, China</b> B.E. in Computer Engineering, Honor Degree (Mixed Class)

## **WORK EXPERIENCE**

---

2016 –	<b>University of California at Santa Barbara</b> Professor, Department of Computer Science
2012 – 2016	<b>University of California at Santa Barbara</b> Associate Professor, Department of Computer Science
2008 – 2012	<b>University of California at Santa Barbara</b> Assistant Professor, Department of Computer Science
2006 – 2008	<b>IBM T. J. Watson Research Center</b> Research Staff Member, Data-Intensive Systems and Analytics, Manager: Dr. Philip S. Yu/Dr. Kun-Lung Wu
2002 – 2006	<b>University of Illinois at Urbana-Champaign</b> Research Assistant, Advisor: Prof. Jiawei Han
Summer 2001	<b>FalconStor Inc.</b> Technical Staff, Storage over IP Engineering Group, Advisor: Alan Chen (VP)
Summer 2000	<b>Silicon Graphics Inc.</b> Research Intern, Data Mining and Visualization Group, Advisor: Dr. Alan Norton

## **HONORS & AWARDS**

---

2013	ICDE 2013 Best Poster Award
2012	SIGKDD 2012 Best Student Paper Award
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**

---

2013	“Ranking Technique for Queries with Multiple Transformations,” by X. Yan, S. Yang, Y. Wu, 61/942,588 by University of California (Provisional Patent Filing)
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, approved as US Patent # US8112667 (Feb, 2012)
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), approved as US Patent # 8,121,967 (July, 2011)
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), approved as US Patent # 08,117,208 (Feb, 2012)
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), approved as US Patent #7,974,978 (July, 2011)

## **TALKS & TUTORIALS**

---

Dec 2015 Invited speaker, “Networks and Knowledge Bases,” by X. Yan, Army Science Planning and Strategy Meeting (ASPSM): Distributed and Collaborative Intelligent Systems, ARL at Adelphi

Oct 2015 Invited speaker, “Graph Analysis and Search in Networks,” by X. Yan, University of Southern California

Oct 2015 Invited speaker, “Network Centric Information Search and Question Answering,” by X. Yan, SnapChat

Sept 2015 “Network Centric Information Search and Question Answering,” by X. Yan, Baidu Silicon Valley AI Lab

Aug 2015 “Network Centric Information Search and Question Answering,” by X. Yan, Fudan University

Aug 2015 Keynote speaker, “Graph OLAP, Anomaly and Query-Based Outlier Detection,” by X. Yan, KDD 2015 workshop, ODDx3: Outlier Definition, Detection, and Description

July 2015 Invited speaker, “Network Centric Information Search and Question Answering,” by X. Yan, University of Electronic Science and Technology at Xi'an

July 2015 Invited speaker, “Network Centric Information Search and Question Answering,” by X. Yan, Huazhong University of Science and Technology

July 2015 Invited speaker, “Network Centric Information Search and Question Answering,” by X. Yan, Shandong University

July 2015 Invited speaker, “Network Centric Information Search and Question Answering,” by X. Yan, Big Data and Internet + Seminar, Donghua University

Apr 2015 “Schemaless Graph Querying,” by X. Yan, Computer Science Department, UC Irvine

Sept 2014 Invited Speaker, “Graph Analysis and Search in Networks,” by X. Yan, Computer Science Department, Peking University, 2014

Sept 2014 Invited Speaker, “Graph Analysis and Search in Networks,” by X. Yan, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, 2014

Aug 2014 “Network Mining and Analysis for Social Applications,” by F. Zhu, H. Sun, and X. Yan. Conference Tutorial, the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD), 2014

Jun 2013 “Towards Fast Malware Detection Using Proximity Graph Patterns,” Workshop of Science for Cyber-Security, University of Maryland

Jun 2013 “Schema-less and Structure-less Graph Search,” INARC/NS-CTA, Army Proving Ground

Aug 2012 “On the Power of Heterogeneous Information Networks,” by Y. Sun, J. Han, X. Yan, and P. Yu, Proc. 2012 International Conference on Social Networks Analysis and Mining (ASONAM)

Aug 2012 “Mining Knowledge from Interconnected Data: A Heterogeneous Information Network Analysis Approach,” by J. Han, Y. Sun, X. Yan, and P. Yu, Conference Tutorial, Proc. 2012 Int. Conf. on Very Large Data Bases (VLDB)

Apr 2012 “Emerging Graph Queries In Linked Data,” by A. Khan, Y. Wu, and X. Yan, Conference Tutorial, The 2012 Int. Conf. on Data Engineering (ICDE)

Apr 2012 “Mining Knowledge from Data: An Information Network Analysis Approach,” by J. Han, Y. Sun, X. Yan, and P. Yu, Conference Tutorial, The 2012 Int. Conf. on Data Engineering (ICDE)

- 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), 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), 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

## **TEACHING EXPERIENCE**

---

- Spring 2016 **Advanced Data Mining (Neural Networks)**  
Graduate level, UC at Santa Barbara (CS291K), Lecture
- Winter 2016 **Artificial Intelligence**  
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Spring 2015 **Advanced Data Mining**  
Graduate level, UC at Santa Barbara (CS290D), Lecture

Winter 2015	<b>Artificial Intelligence</b> Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Fall 2014	<b>Network Science</b> Graduate level, UC at Santa Barbara (CS90A), Lecture
Fall 2014	<b>Data and Knowledge Bases</b> Graduate level, UC at Santa Barbara (CS273), Lecture
Spring 2014	<b>Advanced Data Mining</b> Graduate level, UC at Santa Barbara (CS290D), Lecture
Winter 2014	<b>Artificial Intelligence</b> Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Spring 2013	<b>Artificial Intelligence</b> Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Winter 2012	<b>Artificial Intelligence</b> Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Fall 2011	<b>Data and Knowledge Bases</b> Graduate level, UC at Santa Barbara (CS273), Lecture
Spring 2011	<b>Advanced Data Mining</b> Graduate level, UC at Santa Barbara (CS290D), Lecture
Winter 2011	<b>Artificial Intelligence</b> Undergraduate level, UC at Santa Barbara (CS165A), Lecture
Fall 2010	<b>Dynamic Graph Modeling</b> Graduate level, UC at Santa Barbara (CS595D), Seminar <b>Advanced Information Systems</b> Graduate level, UC at Santa Barbara (CS595J), Seminar
Spring 2010	<b>Design and Implementation Techniques of Database Systems</b> Undergraduate level, UC at Santa Barbara (CS174B), Lecture
Winter 2010	<b>Advanced Data Mining</b> Graduate level, UC at Santa Barbara (CS290D), Lecture <b>Readings in Information Networks</b> Graduate level, UC at Santa Barbara (CS595D), Seminar
Fall 2009	<b>Data and Knowledge Bases</b> Graduate level, UC at Santa Barbara (CS273), Lecture <b>BPM: Models, Process Mining, and BI</b> Graduate level, UC at Santa Barbara (CS595D), Seminar
Spring 2009	<b>Data Mining: Principles and Algorithms</b> Graduate level, UC at Santa Barbara (CS290D), Lecture
Winter 2009	<b>Graph Mining</b> Graduate level, UC at Santa Barbara (CS595D), Seminar

## **SOFTWARE RELEASES**

---

2014	OntQ: Ontology-based Subgraph Querying
2013	TESC: Measuring Two-Event Structural Correlations on Graphs
2013	MSP: Minimum Substring Partitioning

2013	gIceberg: Uncover graph icebergs
2011	Sedge: A Self Evolving Distributed Graph Processing Environment (IBM, USC)
2007	NeMo: A graph-based approach to systematically reconstruct human transcriptional regulatory modules, <a href="http://zhoulab.usc.edu/NeMo/source.htm">http://zhoulab.usc.edu/NeMo/source.htm</a> (USC, MIT/Whitehead, UCSC, UNC, SFU)
2006	gIndex: Index and search large graph sets, <a href="http://www.xifengyan.net/software.htm">http://www.xifengyan.net/software.htm</a> (590 citations)
2005	CODENSE: Frequent Coherent Dense Subgraphs Mining Package, <a href="http://zhoulab.usc.edu/CODENSE/">http://zhoulab.usc.edu/CODENSE/</a> (USC, NIH, Aristotle Univ.)
2005	CloSpan: Closed Sequential Pattern Mining Package, <a href="http://illimine.cs.uiuc.edu">http://illimine.cs.uiuc.edu</a> (800 citations, Rational Software)
2005	CloseGraph: Frequent Closed Graph Mining Package (660 citations)
2004	gSpan: Frequent Graph Mining Package, <a href="http://www.cs.ucsb.edu/~xyan/software">http://www.cs.ucsb.edu/~xyan/software</a> (1900 citations)

---

**PUBLICATIONS (h-index: 50, citations: 14,000+ per Google Scholar)**

**Refereed Journal Papers**

- [1] “Observability of Lattice Graphs,” by F. Han, S. Suri, and X. Yan, *Algorithmica*, 2015
- [2] “Querying Knowledge Graphs by Example Entity Tuples,” by N. Jayaram, A. Khan, C. Li, X. Yan, Ramez Elmasri, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 2015
- [3] “Fine-Grained Knowledge Sharing in Collaborative Environments,” by Z. Guan, S. Yang, H. Sun, M. Srivatsa, X. Yan, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, Vol. 27, No. 8, Aug., 2015
- [4] “Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks,” by D. Agrawal, C. Budak, A. El Abbadi, T. Georgiou, X. Yan, *Databases in Networked Information Systems - Lecture Notes in Computer Science Volume 8381*, 2014, pp 1-16.
- [5] “Multi-Aspect + Transitivity + Bias: An Integral Trust Inference Model,” by Y. Yao, H. Tong, X. Yan, F. Xu, J. Lu, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 26(7) 2014
- [6] “Interpreting the Public Sentiment Variations on Twitter” by S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. Bu, C. Chen, and X. He, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 2013
- [7] “PathSelClus: Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks,” by Y. Sun, B. Norick, J. Han, X. Yan, P. S. Yu, X. Yu, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2013
- [8] “Static and Dynamic Structural Correlations in Graphs,” by J. Wu, Z. Guan, Z. Yun, A. Singh, X. Yan, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 25(9), 2013
- [9] “Co-occurrence Based Diffusion for Expert Search On the Web,” by Z. Guan, G. Miao, R. McLoughlin, X. Yan, and D. Cai, *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 25(5), 2013
- [10] “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
- [11] “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
- [12] “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

- [13] “On Compressing Frequent Patterns,” D. Xin, J. Han, X. Yan, H. Cheng, *Data Knowl. Eng. (DKE)*, 60(1): 5– 29, 2007
- [14] “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
- [15] “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)
- [16] “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.
- [17] “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)
- [18] “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.
- [19] “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] “Cross-domain Semantic Parsing via Paraphrasing,” by Y. Su, X. Yan. *Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing (EMNLP’17)*, long paper, 2017
- [2] “Recovering Question Answering Errors via Query Revision,” by S. Yavuz, I. Gur, Y. Su, X. Yan. In *Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing (EMNLP’17)*, short paper, 2017
- [3] “Privacy-Preserving Community-Aware Trending Topic Detection in Online Social Media,” by T. Georgiou, A. El Abbadi, and X. Yan, the 31st Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (**DBSec’17**), 2017 (acceptance rate: 36%, full paper)
- [4] “Extracting Topics with Focused Communities for Social Content Recommendation,” by T. Georgiou, A. El Abbadi, and X. Yan, *The 20th ACM Conf. on Computer-Supported Cooperative Work and Social Computing (CSCW’17)*, 2017 (acceptance rate: 35%, full paper)
- [5] “On Generating Characteristic-rich Question Sets for QA Evaluation,” by Y. Su, H. Sun, B. Sadler, M. Srivatsa, I. Gur, Z. Yan, and X. Yan. *Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP’16)*, short paper, 2016
- [6] “Improving Semantic Parsing via Answer Type Inference,” by S. Yavuz, I. Gur, Y. Su, M. Srivatsa, X. Yan. *Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP’16)*, long paper, 2016
- [7] “Analyzing information sharing strategies of users in online social networks,” by D. Nguyen, S. Tan, R. Ramanathan, X. Yan, *Proc. 2016 International Conference on Social Networks Analysis and Mining (ASONAM’16)*, 2016 (acceptance rate: 14%, full paper)
- [8] “Semantic SPARQL Similarity Search Over RDF Knowledge Graphs,” W. Zheng, L. Zou, W. Peng, X. Yan, S. Song, D. Zhao, *Proc. of the 42nd International Conference on Very Large Data Bases, (VLDB’16)*, 2016.
- [9] “Fast Top-K Search in Knowledge Graphs,” by S. Yang, F. Han, Y. Wu, X. Yan, *Proc. of Int. Conf. on Data Engineering (ICDE’16)*, 2016
- [10] “Fast Motif Discovery in Short Sequences,” by H. Liu, F. Han, H. Zhou, X. Yan, K. Kosik, *Proc. of Int. Conf. on Data Engineering (ICDE’16)*, 2016
- [11] “Distributed Representations of Expertise,” F. Han, S. Tan, H. Sun, M. Srivatsa, D. Cai, X. Yan, *SIAM Int. Conf. on Data Mining, 2016 (SDM’16)* (acceptance rate: 26%)
- [12] “A Fast Kernel for Attributed Graphs,” by Y. Su, F. Han, R. E. Harang, X. Yan, *SIAM Int. Conf. on*

- Data Mining, 2016 (**SDM'16**) (acceptance rate: 26%)
- [13] “Table Cell Search for Question Answering,” by H. Sun, H. Ma, X. He, W.-T. Yih, Y. Su, X. Yan, Proc. of the 25th Int. World Wide Web Conf. (**WWW'16**), 2016
  - [14] “Entity Disambiguation with Linkless Knowledge Bases,” by Y. Li, S. Tan, H. Sun, J. Han, D. Roth and X. Yan, Proc. of the 25th Int. World Wide Web Conf. (**WWW'16**), 2016
  - [15] “Behavior Query Discovery in System-Generated Temporal Graphs,” by B. Zong, X. Xiao, Z. Li, Z. Wu, Z. Qian, X. Yan, A. Singh, G. Jiang, Proc. of the 42th Int. Conf. on Very Large Databases (**VLDB'16**), Vol. 9 No. 4, 2016
  - [16] “Mining Complaints for Traffic-Jam Estimation: A Social Sensor Application,” by T. Georgiou, A. Abbadi, X. Yan, J. George, Proc. 2015 International Conference on Social Networks Analysis and Mining (**ASONAM'15**), 2015 (acceptance rate: 18%)
  - [17] “Exploiting Relevance Feedback in Knowledge Graph Search,” by Y. Su, S. Yang, H. Sun, M. Srivatsa, S. Kase, M. Vanni and X. Yan, Proc. of Int. Conf. on Knowledge Discovery and Data Mining (**SIGKDD'15**), 2015 (acceptance rate: 19%)
  - [18] “Query-Based Outlier Detection in Heterogeneous Information Networks,” by H. Zhuang, J. Zhang, G. Brova, J. Tang, H. Cam, X. Yan, and J. Han, 18th International Conference on Extending Database Technology (**EDBT'15**), 2015 (**Best of EDBT 2015, invited to TODS**)
  - [19] “Expertise-Based Data Access in Content-Centric Mobile Opportunistic Networks,” by J. Zhao, X. Zhang, G. Cao, M. Srivatsa, and X. Yan, the 11th IEEE Int. Conf. on Mobile Ad hoc and Sensor Systems (**MASS'14**), 2014 (acceptance rate: 27%)
  - [20] “Mining Query-Based Subnetwork Outliers in Heterogeneous Information Networks,” by H. Zhuang, J. Zhang, G. Brova, J. Tang, H. Cam, X. Yan, and J. Han, Proc. 2014 Int. Conf. on Data Mining (**ICDM'14**), 2014. (short paper, acceptance rate: 19%)
  - [21] “Analyzing Expert Behaviors in Collaborative Networks,” by H. Sun, M. Srivatsa, S. Tan, Y. Li, L. Kaplan, S. Tao and X. Yan, Proc. of the 20th Int. Conf. on Knowledge Discovery and Data Mining (**KDD'14**), 2014 (acceptance rate: 15%)
  - [22] “Towards Scalable Critical Alert Mining,” by B. Zong, Y. Wu, J. Song, A. Singh, H. Cam, J. Han and X. Yan, Proc. of the 20th Int. Conf. on Knowledge Discovery and Data Mining (**KDD'14**), 2014 (acceptance rate: 15%)
  - [23] “Schemaless and Structureless Graph Querying,” by S. Yang, Y. Wu, H. Sun, X. Yan, Proc. of the 40th Int. Conf. on Very Large Databases (**VLDB'14**), Vol 7, No. 7, pp. 565 – 576, 2014.
  - [24] “SLQ: A User-friendly Graph Querying System,” by S. Yang, Y. Xie, Y. Wu, T. Wu, H. Sun, J. Wu, X. Yan, Proc. 2014 Int. Conf. on Management of Data (**SIGMOD'14**) (demo paper), 2014 (acceptance rate, 38%)
  - [25] “A Probabilistic Approach to Uncovering Attributed Graph Anomalies,” by N. Li, H. Sun, K. Chipman, J. George, X. Yan, Proc. 2014 SIAM Int. Conf. on Data Mining (**SDM'14**), 2014.
  - [26] “Extracting Probable Command and Control Signatures for Detecting Botnets,” by A. Zand, G. Vigna, X. Yan and C. Kruegel, The Security Track of the 2014 ACM Symposium on Applied Computing (**SAC'14**).
  - [27] “Cloud Service Placement via Subgraph Matching,” by B. Zong, R. Raghavendra, M. Srivatsa, X. Yan, A. Singh, and K.-W. Lee, Proc. 2014 Int. Conf. on Data Engineering (**ICDE'14**), 2014 (acceptance rate, 20%)
  - [28] “Top-K Interesting Subgraph Discovery in Information Networks,” by M. Gupta, J. Gao, X. Yan, H. Cam, and J. Han, Proc. 2014 Int. Conf. on Data Engineering (**ICDE'14**), 2014 (acceptance rate, 20%)
  - [29] “Summarizing Answer Graphs Induced by Keyword Queries,” by Y. Wu, S. Yang, M. Srivatsa, A. Iyengar, X. Yan, Proc. of the 40th Int. Conf. on Very Large Databases (**VLDB'14**), 2014.
  - [30] “Synthesizing near-optimal malware specifications from suspicious behaviors,” by S. Jha, M. Fredrikson, M. Christodorescu, R. Sailer, and X. Yan, **MALWARE** 2013: 41-50 (invited paper, an extension of **Oakland'10** work)
  - [31] “Noise-Resistant Bicluster Recognition,” by H. Sun, G. Miao, X. Yan, Proc. 2013 IEEE Int. Conf. on



### Data Mining (ICDM'13)

- [32] “Automated Trauma Incident Cubes,” by A. Srivastava, L. Ferrigno, S. Kaminski, X. Yan and J. Su, IEEE Int. Conf. on Healthcare Informatics (ICHI'13), 2013
- [33] “On Detecting Association-Based Clique Outliers in Heterogeneous Information Networks,” by M. Gupta, J. Gao, X. Yan and J. Han, Proc. 2013 International Conference on Social Networks Analysis and Mining (ASONAM'13), 2013 (acceptance rate: 13%).
- [34] “I act, therefore I judge: Network sentiment dynamics modeling based on user activity,” K. Macropol, P. Bogdanov, A. Singh, L. Petzold and X. Yan, Proc. 2013 International Conference on Social Networks Analysis and Mining (ASONAM'13), 2013 (acceptance rate: 15%).
- [35] “Synthetic Review Spamming and Defense,” by H. Sun, A. Morales, and X. Yan, Proc. of the 19th Int. Conf. on Knowledge Discovery and Data Mining (KDD'13), 2013 (acceptance rate: 17%).
- [36] “Mining Evidences for Named Entity Disambiguation,” by Y. Li, C. Wang, F. Han, J. Han, D. Roth, and X. Yan, Proc. of the 19th Int. Conf. on Knowledge Discovery and Data Mining (KDD'13), 2013 (acceptance rate: 17%).
- [37] “Characterizing Tenant Behavior for Placement and Crisis Mitigation in Multitenant DBMSs,” by A. Elmore, S. Das, A. Pucher, D. Agrawal, A. El Abbadi, X. Yan, Proc. 2013 Int. Conf. on Management of Data (SIGMOD'13).
- [38] “MATRI: a multi-aspect and transitive trust inference model,” by Y. Yao, H. Tong, X. Yan, F. Xu, J. Lu, the 22nd International World Wide Web Conference (WWW'13), 2013 (acceptance rate, 15%)
- [39] “Memory Efficient Minimum Substring Partitioning,” by Y. Li, P. Kamousi, F. Han, S. Yang, X. Yan, S. Suri, Proc. of the 39th Int. Conf. on Very Large Databases (VLDB'13), 2013.
- [40] “NeMa: Fast Graph Search with Label Similarity,” by A. Khan, Y. Wu, C. Aggarwal, X. Yan, Proc. of the 39th Int. Conf. on Very Large Databases (VLDB'13), 2013.
- [41] “gIceberg: Towards Iceberg Analysis in Large Graphs,” by N. Li, Z. Guan, L. Ren, J. Wu, J. Han, X. Yan, Proc. 2013 Int. Conf. on Data Engineering (ICDE'13), 2013 (acceptance rate, 19%)
- [42] “Ontology-based Subgraph Querying,” Y. Wu, S. Yang, X. Yan, Proc. 2013 Int. Conf. on Data Engineering (ICDE'13), 2013 (acceptance rate, 19%)
- [43] “Inferring the Underlying Structure of Information Cascades,” by B. Zong, Y. Wu, A. Singh, and X. Yan, Proc. 2012 Int. Conf. on Data Mining (ICDM'12), Brussels, Belgium, 2012. (short paper, acceptance rate: 20%)
- [44] “A General Framework to Encode Heterogeneous Information Sources for Contextual Pattern Mining,” by W. Dong, W. Fan, L. Shi, C. Zhou, and X. Yan, The 21st ACM Int. Conf. on Information and Knowledge Management (CIKM'12), Oct 2012. (acceptance rate, 13%)
- [45] “Density Index and Proximity Search in Large Graphs,” by N. Li, X. Yan, Z. Wen, and A. Khan, The 21st ACM Int. Conf. on Information and Knowledge Management (CIKM'12), Oct 2012. (acceptance rate, 13%)
- [46] “Measuring Two-Event Structural Correlations on Graphs,” by Z. Guan, X. Yan, L. M. Kaplan, Proc. of the 38th Int. Conf. on Very Large Databases (VLDB'12), Aug 2012 (acceptance rate, 20%).
- [47] “Latent Association Analysis of Document Pairs,” by G. Miao, Z. Guan, L. Moser, X. Yan, S. Tao, N. Anerousis, and J. Sun, Proc. of the 18th Int. Conf. on Knowledge Discovery and Data Mining (KDD'12), Aug 2012 (acceptance rate, 13%).
- [48] “Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks,” by Y. Sun, B. Norick, J. Han, X. Yan, P. S. Yu, X. Yu, Proc. of the 18th Int. Conf. on Knowledge Discovery and Data Mining (KDD'12), Aug 2012. (acceptance rate, 13%) (**Best Student Research Paper**)
- [49] “Towards Effective Partition Management for Large Graphs,” by S. Yang, X. Yan, B. Zong, A. Khan Proc. 2012 ACM Int. Conf. on Management of Data (SIGMOD'12), Jun 2012 (acceptance rate, 17%)
- [50] “Understanding Task-driven Information Flow in Collaborative Networks,” by G. Miao, S. Tao, W. Cheng, J. Moulic, L. Moser and X. Yan. Proc. the 21st Int. Conf. on World Wide Web 2012 (WWW'12), (acceptance rate, 12%).

- [51] “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%).
- [52] “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%).
- [53] “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%).
- [54] “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%).
- [55] “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%).
- [56] “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%).
- [57] “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.
- [58] “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%).
- [59] “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%).
- [60] “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
- [61] “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%)
- [62] “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%)
- [63] “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%)
- [64] “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%)
- [65] “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%)
- [66] “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%)
- [67] “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)

- [68] “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%)
- [69] “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)
- [70] “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%)
- [71] “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%)
- [72] “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%)
- [73] “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%)
- [74] “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%).
- [75] “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%)
- [76] “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%)
- [77] “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%)
- [78] “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%)
- [79] “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%)
- [80] “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%)
- [81] “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%)
- [82] “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%)
- [83] “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%)
- [84] “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%)
- [85] “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%)
- [86] “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%)
- [87] “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%)
- [88] “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%)
- [89] “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%)
- [90] “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%)
- [91] “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%)
- [92] “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%)
- [93] “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%)
- [94] “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%)
- [95] “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%)
- [96] “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%)
- [97] “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%).
- [98] “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%)
- [99] “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%)
- [100] “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%)
- [101] “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%)
- [102] “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%)
- [103] “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%).
- [104] “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] “Efficient Direct Mining of Selective Discriminative Patterns for Classification,” 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
- [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, 2012.
- [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] “On the Efficiency of Decentralized Search in Expert Networks,” by L. Ma and M. Srivatsa, D. Cansever, X. Yan, and S. Kase and M. Vanni, the 36th IEEE International Conference on Distributed Computing Systems (**ICDCS 2016**) (poster)
- [2] “Towards a Query-by-Example System for Knowledge Graphs,” by N. Jayaram, A. Khan, C. Li, X. Yan, and R. Elmasri, Graph Data-management Experiences & Systems (**GRADES'14**), 2014
- [3] “GQBE: Querying Knowledge Graphs by Example Entity Tuples,” by N. Jayaram, M. Gupta, A. Khan, C. Li, X. Yan, R. Elmasri, Proc. 2014 Int. Conf. on Data Engineering (**ICDE'14**), 2014 (system demo)

- [4] “Noise-Resistant Bicluster Recognition,” by H. Sun, G. Miao, Y. S. Huang, and X. Yan, the 17th Annual International Conference on Research in Computational Molecular Biology (**RECOMB’13**) (poster)
- [5] “Synthetic Review Spamming and Immunization,” by A. Morales, H. Sun, and X. Yan, the 22nd International World Wide Web Conference (**WWW’13**) (poster)
- [6] “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
- [7] “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%).
- [8] “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, 2008.
- [9] “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
- [10] “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)
- [11] “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)
- [12] “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)
- [13] “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.
- [14] “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**).
- [15] “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.
- [16] “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.

## **POSTDOC and PHD. ALUMNI**

---

2016	Fangqiu Han, “Representation Learning on Unstructured Data,” Research Scientist, Snapchat
2015	Huan Sun, “Intelligent and Collaborative Query Resolution,” Assistant Professor, Ohio State University
2015	Bo Zong, “Towards Mining and Managing Large-Scale Temporal Graphs,” Research Staff, NEC Labs America
2015	Shulong Tan, Project Scientist, Research Scientist, Baidu Silicon Valley AI Lab
2015	Yang Li, “Connecting Text with Knowledge,” Google

2015	Shengqi Yang, “Fast Search in Large Scale Knowledge Graphs,” Research Scientist, Facebook
2014	Yinghui Wu, Project Scientist, Assistant Professor, Washington State University
2014	Ying Tang, Project Scientist, Associate Professor, Chengdu University of Technology, China
2013	Nan Li, Ph.D., “Uncovering Anomalous Patterns in Large Attributed Graphs,” ODESK
2013	Arijit Khan, Ph.D., “Towards Querying and Mining of Large-Scale Networks,” ETH Postdoc
2012	Ziyu Guan, Project Scientist, Professor, Northwest University, China
2012	Gengxin Miao, Ph.D., “Understanding the Semantics of Networked Text,” Google

### **M.S. ALUMNI**

---

2016	Anastasiya Lazareva, M.S. “ECNN: A Convolutional Neural Network Generator based on Adaptive Evolutionary Algorithms”
2016	Tianyu Wu, M.S., “Top-k graph querying with Spark GraphX,” Google
2016	Arvind Chitra Rajasekaran, M.S., “Information extraction from bio-medical documents”
2015	Yanbo Ma, M.S. “Open Source Recommendation System - Survey and Case Study,” ORACLE
2014	Alex Morales, M.S. “Detecting Synthetic Review Spam,” Univ. of Illinois at Urbana-Champaign, Ph.D. program (5-year fellowship)
2014	Nadav Caspi, M.S. “Solar Rating - Better Insight into Solar Panels,” Microsoft
2014	Bo Yang, M.S., “Comparison of Collaborative Filtering Algorithms,” CISCO
2013	Liu Chen, M.S., “What will affect the success of open source projects on Github?” ORACLE
2012	Matt Weiden, M.S., “An Adaptable Legislative Advocacy and Accountability Framework,” HRL Laboratories
2011	Lei Zhao, M.S., “Frequent Subgraph Pattern Mining With Substructural Constraints,” LinkedIn
2010	Russell McLoughlin, M.S., “Pruning Strategies for Tractable Sequence Assembly with Moderate Memory Requirements,” Lawrence Livermore National Lab

### **UNDERGRADUATE STUDENTS SUPERVISED**

---

Summer 2014	Lawrence McClendon, Jackson State University, “Detecting Spam Email Using Machine Learning Algorithms,” supervised by Xifeng Yan, Research Internships in Science and Engineering (RISE), UCSB.
-------------	---

Summer 2014	Lennon Ganz, Santa Barbara City College, “Finding Patterns in Complex Social Networks,” supervised by Yinghui Wu/Xifeng Yan, Internships in Nanosystems Science, Engineering and Technology (INSET). He will attend the graduate school at CS@UC Berkeley.
Summer 2013	Miranda Z. Aperghis, University of Bristol, supervised by Xifeng Yan.
Summer 2013	Bruce Liu, Pasadena Community College/UCI, supervised by Fangqiu Han and Xifeng Yan, “Image Classification with Neural Networks.”
Summer 2012	Amber Johnson, Jackson State University, supervised by Xifeng Yan, “Investigating Techniques to Detect Fake Reviews in Social Media.” Research Internships in Science and Engineering (RISE), UCSB. She became a Ph.D student at CS@purdue.
Summer 2012	Juan Zepeda, Santa Barbara City College, supervised by Yinghui Wu and Xifeng Yan, “Building a Server-Client Framework: Visualizing Social Graphs and Querying Human Relations.”
Summer 2012	Sang Nguyen, Oxnard College, supervised by Huan Sun, “Perceptron: A Classification Method.”
Summer 2011	Peiyang Shi, Moorpark College, supervised by Arijit Khan, “Graph Algorithms - Efficient Shortest Path Estimation.”
Summer 2011	Daniel Vicory, Allan Hancock College, supervised by Nan Li, “Novel Optimization of Task Scheduling Within MapReduce/Hadoop.”

### **K-12 STUDENTS SUPERVISED**

Summer 2017	Albert Miao, Dos Pueblos High School, supervised by Xiaoyong Jin
Summer 2017	Andy Jin, Westlake High School, supervised by Xiaoyong Jin
Summer 2015	Allen Chen, Lynbrook High School, supervised by Izzeddin Gür, “Knowledge Driven Paraphrase Identification.”
Summer 2015	Rohan Divate, Moreau Catholic High School, supervised by Bo Zong, “Graph Feature Mining via Machine Learning.”
Summer 2014	Jonah Ezra Rubin, Ridley College, Canada, “Review Rating Adjustment to Incorporate User Preferences.” supervised by Honglei Liu/Xifeng Yan. He later studied in Harvey Mudd College.

### **STUDENT HONORS AND AWARDS**

2016	Huan Sun, “Mining Disparate Sources for Question Answering,” SIGKDD 2016 Ph.D. Dissertation Runner Up Award. She is also elected as one of KDD 2016 Rising Stars.
2015	Huan Sun, Department Best Dissertation Award
2012-2013	Arijit Khan, IBM Ph.D. Fellowship
2011-2012	Gengxin Miao, IBM Ph.D. Fellowship

### **GRANTS**



2016-2019 Army Research Lab, Network Science CTA, Information Network Academic Research Center, J. Han (PI), extended. (Years 8-10, PI \$1.1M)

2015-2018 NSF III: "Small: Knowledge Graph Query Processing and Benchmarking," X. Yan (PI). Grant: \$500K

2014-2016 Army Research Lab, Network Science CTA, Information Network Academic Research Center, J. Han (PI), extended. (Years 6/7, PI \$490K)

2014-2016 DARPA SBIR PHASE II, "CERTAIN: Certainty Enrichment via Relational and Temporal Analytical Indexing of Networks," X. Yan (PI) subcontract of Aptima, Grant: 160K+80K(option) (declined by UCSB, due to a disclosure policy of information)

2013-2017 DARPA, Air Force Research Laboratory, "PLAN-X: Revolutionary Cyber Battlespace Analytics," X. Yan (PI) subcontract of Aptima, Grant: \$460K (declined by UCSB, due to a disclosure policy of information)

2013-2018 NSF IGERT-CIF21 "Interdisciplinary Graduate Education Research and Training in Network Science," A. Singh (PI). Grant: \$2.8M

2012-2013 ICB - Institute for Collaborative Biotechnologies, "Multimodal Sensor Network: A Systems Approach to Analyzing Patient Well-Being and Disease States," Scott Hammond (PI), X. Yan. Grant: \$100K

2012-2015 Rainwater Foundation, "Tauopathy Research Consortium – Bioinformatics," K. Kosik (PI), X. Yan. Gift: \$1.7M (830+890K+595K)

2012-2013 Army Research Lab, "Developing a Science of Cybersecurity: Graph Search and Pattern Discovery for Malware Analysis," X. Yan (PI). Grant: \$85K

2012-2013 ICB - Institute for Collaborative Biotechnologies, "Automated Analysis and Modeling of Motivationally Relevant Narratives from Online Communication Sources," René Weber (PI), X. Yan. Grant: \$100K

2011-2013 Army Research Lab, "Developing a Science of Cybersecurity: Graph-Centric Metrics, Monitoring, and Composite Analysis of Adversarial Activities on Networks," X. Yan (PI), J. Han, Grant: \$400K

2011-2014 ICB - Institute for Collaborative Biotechnologies, "Metrics and Models of Persuasion over Networks," L. Petzold (PI), X. Yan, A. Singh, Grant: \$250K+150K

2010-2015 NSF Career Proposal, "CAREER: Graph Information System: Deciphering Complex Networks," X. Yan (PI). Grant: \$495K

2009-2014 Army Research Lab, "Information Network Academic Research Center: An Integrated Approach towards Information Integration, Modeling, Retrieval, and Discovery," Network Science CTA (Core Members: UIUC, UCSB, IBM, CUNY, Subcontractors: CMU, PARC, UMichigan, Northwestern Univ.), J. Han (PI), X. Yan (one of the five core members). Grant: \$16.75M

2010-2013 Alzheimer's Association, "Lifestyle Improvement Game to Delay Alzheimer's Onset and Support Treatment," D. Lieberman (PI), K. Kosik, X. Yan. Grant: \$160K

2010-2012 Zhejiang University CAD&CG Lab, "Large-Scale Graph Information System," X. Yan (PI). Grant: \$60K Yuan

2010-2011 ICB - Institute for Collaborative Biotechnologies, "Network Inference and Experimental Design from Time-Series Microarray Data," L. Petzold (PI), X. Yan. Grant: \$100K

2009-2010	ICB - Institute for Collaborative Biotechnologies, "Multiscale Computational Modeling, Network Interference, and the Effects of Exercise," L. Petzold (PI), X. Yan. Grant: \$100K
2009-2012	NSF IIS Award #0905084, "Medium: Collaborative Research: Towards On-Line Analytical Mining of Heterogeneous Information Networks," J. Han (PI), X. Yan, P. S. Yu. Grant: \$368K (\$1.2M)
2009-2012	NSF IIS Award #0917228, "III: Small: Collaborative Research: Mining and Optimizing Ad Hoc Workflows", X. Yan (PI), Y. Chen. Grant: \$248K (\$498K)
2009-2011	NSF IIS Award #0847925, "CluE: Towards Scalable Primitives for Graph Operations," B. Zhao(PI) , D. Agrawal, A. Abbadi, R. Wolski, and X. Yan. Grant: \$482K
2008 -	The Venkatesh Narayanamurti Chair, Univ. of California at Santa Barbara. Grant: \$200K

## **PROFESSIONAL SERVICES**

---

- Area Editor: Information Systems, an Elsevier journal (2011-2016)
- Senior Program Committee Member: 2014, 2015, 2016 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)
- Senior Program Committee Member: ACM 2012, 2014, 2015 Int. Con. on Knowledge Discovery and Data Mining (KDD)
- Workshop Co-Chair: 2015 SIAM Conference on Data Mining (SDM'15)
- Program Area Chair: IEEE 2014 International Conference on Data Mining (ICDM'14)
- Senior Program Committee Member: ACM 2013 International Conference on Information and Knowledge Management (CIKM'13)
- Area Chair: "Social Networks, Web, and Personal Information Management," IEEE 2013 Int. Conf. Data Engineering (ICDE'13)
- 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: ACM Int. Conf. on Management of Data (SIGMOD'10-13); ACM Int. Con. on Knowledge Discovery and Data Mining (SIGKDD'08-15); Int. Conf. on Very Large Databases (VLDB 2013-15); WWW 2012-14; IEEE Int. Conf. Data Engineering (ICDE'10-12); IEEE International Conference on Data Mining series (ICDM'09-13); SIAM Conference on Data Mining (SDM'07-13); IEEE/ACM Int. Con. on Advances in Social Networks Analysis and Mining (ASONAM'10-12); the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06); the 5th International Workshop on Mining and Learning with Graphs (MLG'07,09-12), The 24th ACM International Conference on Information and Knowledge Management (CIKM'2013, 2014)
- NSF Panelist, 2008-2017
- Ireland Science Foundation, 2016
- Chilean Government, 2016

- Czech Science Foundation, 2011
- Portuguese Foundation for Science and Technology (FCT), 2012
- 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) 2003-2005, 2009-2012; ACM Transactions on Knowledge Discovery from Data (TKDD) 2007-2011; IEEE Trans. on Knowledge and Data Engineering, TKDE 2004-2006, 2008-2012; Data Mining and Knowledge Discovery (DAMI) 2006, 2007; Data & Knowledge Engineering (DKE) 2008-2010; Knowledge and Information Systems (KAIS) 2009-2012; 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, NTU
- Proposal Referee: National Aeronautics and Space Administration (NASA), 2004.