

Xifeng Yan

Professor
Venkatesh Narayanamurti Endowed Chair
Department of Computer Science
Rm 2017, Henley Hall
University of California
Santa Barbara, CA 93106-5110

Email: xyan@cs.ucsb.edu
Phone: (805)893-3734

<http://www.cs.ucsb.edu/~xyan>

RESEARCH INTERESTS

The primary objective of my research is to explore foundation models in artificial intelligence, leverage these models for knowledge discovery, and develop cross-disciplinary applications in areas like finance, healthcare, and science.

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)

WORK EXPERIENCE

2016 –	University of California at Santa Barbara Professor, Department of Computer Science
2012 – 2016	University of California at Santa Barbara Associate Professor, Department of Computer Science
2008 – 2012	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

HONORS & AWARDS

2022	Test of Time Award, the International Conference on Very Large Data Bases (VLDB)
2022	Distinguished Paper Award, Programming Language Design and Implementation (PLDI)
2019	Best of IEEE International Conference on Data Mining (ICDM) 2019 Selection

2013	Best Poster Award, IEEE International Conference on Data Engineering (ICDE)
2012	Best Student Paper Award, ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)
2011	10-year Highest Impact Paper Award, IEEE International Conference on Data Mining (ICDM)
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	Dissertation Award Runner-Up, ACM Special Interest Group on Management of Data (SIGMOD)
2007	Best Student Paper, IEEE International Conference on Data Engineering (ICDE)
2007	Best Student Paper, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)
2005	Best Student Paper Runner-up Award, ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)
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

PUBLICATIONS (h-index: 75, citations: 29,000+ per Google Scholar)

Refereed Conference Papers

- [1] Financial Forecasting from Textual and Tabular Time Series, by R. Koval, N. Andrews, X. Yan, Proceedings of Findings of EMNLP 2024 (**EMNLP'24**)
- [2] Evaluating the Instruction-Following Robustness of Large Language Models to Prompt Injection, by Z. Li, B. Peng, P. He, X. Yan, Proceedings of Findings of EMNLP 2024 (**EMNLP'24**)
- [3] “Creative and Context-Aware Translation of East Asian Idioms with GPT-4,” by K. Tang, P. Song, Y. Qin, X. Yan, Proceedings of Findings of EMNLP 2024 (**EMNLP'24**)
- [4] Can Editing LLMs Inject Harm? by C. Chen, B. Huang, Z. Li, Z. Chen, S. Lai, X. Xu, J.-C. Gu, J. Gu, H. Yao, C. Xiao, X. Yan, W. Wang, P. Torr, D. Song, K. Shu, TiFA@ICML 2024 and NextGenAISafety@ICML 2024.
- [5] MMSci: A Multimodal Multi-Discipline Dataset for PhD-Level Scientific Comprehension, by Z. Li, X. Yang, K. Choi, W. Zhu, R. Hsieh, HJ Kim, JH Lim, S. Ji, B. Lee, X. Yan, L. Petzold, S. Wilson, W. Lim, W. Wang, AI for Accelerated Materials Design (**AI4MAT'24**), 2024
- [6] “Bot or Human? Detecting ChatGPT Imposters with A Single Question,” by H. Wang, X. Luo, W. Wang, M. Yu, X. Yan, Conference on Language Modeling (**COLM'24**), 2024
- [7] “Large Language Models as Zero-shot Dialogue State Tracker through Function Calling,” by Z. Li, Z. Chen, M. Ross, P. Huber, S. Moon, Z. Lin, X. Dong, A. Sagar, X. Yan, P. Crook, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'24**), 2024
- [8] “Learning to Compare Financial Reports for Financial Forecasting,” by R. Koval, N. Andrews and X. Yan, the 18th Conference of the European Chapter of the Association for Computational Linguistics (**EACL'24**), 2024

- [9] “Language Models Augmented with Decoupled Memory,” by W. Wang, L. Dong, H. Cheng, X. Liu, X. Yan, J. Gao, F. Wei, The Thirty-seventh Annual Conference on Neural Information Processing Systems (**NeurIPS'23**), 2023
- [10] “Guiding Large Language Models via Directional Stimulus Prompting,” by Z. Li, B. Peng, P. He, M. Galley, J. Gao, X. Yan, The Thirty-seventh Annual Conference on Neural Information Processing Systems (**NeurIPS'23**), 2023
- [11] “Time Series as Images: Vision Transformer for Irregularly Sampled Time Series,” by Z. Li, S. Li, X. Yan, The Thirty-seventh Annual Conference on Neural Information Processing Systems (**NeurIPS'23**), 2023
- [12] “Limitations of Language Models in Arithmetic and Symbolic Induction,” by J. Qian, H. Wang, Z. Li, S. Li, and X. Yan, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'23**), 2023
- [13] “Graph Reasoning for Question Answering with Triplet Retrieval,” by S. Li, Y. Gao, H. Jiang, Q. Yin, Z. Li, X. Yan, C. Zhang and B. Yin, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'23**), 2023
- [14] “Forecasting Earnings Surprises from Conference Call Transcripts,” by R. Koval, N. Andrews and X. Yan, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'23**), 2023
- [15] “Improving Medical Predictions by Irregular Multimodal Electronic Health Records Modeling,” by X. Zhang, S. Li, Z. Chen, X. Yan, L. Petzold, The Fortieth International Conference on Machine Learning (**ICML'23**), 2023
- [16] “Visually-augmented language modeling,” by W. Wang, L. Dong, H. Cheng, H. Song, X. Liu, X. Yan, J. Gao, F. Wei, Proceedings of Int. Conf. on Learning Representations (**ICLR'23**), 2023
- [17] “Language Model Detoxification in Dialogue with Contextualized Stance Control,” by J. Qian and X. Yan, Proceedings of Findings of EMNLP (**EMNLP'22**), 2022
- [18] “Controllable Dialogue Simulation with In-context Learning,” by Z. Li, W. Chen, S. Li, H. Wang, J. Qian and X. Yan, Proceedings of Findings of EMNLP (**EMNLP'22**), 2022
- [19] “Visualization Question Answering Using Introspective Program Synthesis,” by Y. Chen, X. Yan, Y. Feng, The 43rd ACM SIGPLAN Conference on Programming Language Design and Implementation (**PLDI'22** Artifacts)
- [20] “Inductive Relation Prediction by BERT,” by H. Zha, Z. Chen, X. Yan, Thirty-Sixth AAAI Conference on Artificial Intelligence (**AAAI'22**), 2022
- [21] “Composite Re-Ranking for Efficient Document Search with BERT,” by Y. Yang, Y. Qiao, J. Shao, X. Yan, T. Yang, ACM International Conference on Web Search and Data Mining (**WSDM'22**), 2022
- [22] “Task-adaptive Pre-training and Self-training are Complementary for Natural Language Understanding,” by S. Li, S. Yavuz, W. Chen and X. Yan, Proceedings of Findings of EMNLP (**EMNLP'21**), 2021
- [23] Comprehensively Computing Link-based Similarities by Building A Random Surfer Graph by M. Zhang, X. Yan, W. Wang, The 2021 ACM Int. Conf. on Information and Knowledge Management (**CIKM'21**), Nov 2021. [pdf]
- [24] “Network Intervention for Mental Disorders with Minimum Small Dense Subgroups,” by B.-Y. Hsu, C.-Y. Shen, and X. Yan, IEEE Transactions on Knowledge and Data Engineering, (**TKDE'21**), 2021
- [25] “Lifelong Learning of Hate Speech Classification on Social Media,” by J. Qian, H. Wang, M. ElSherief and X. Yan, Proc. of the 2021 North American Chapter of ACL: Human Language Technologies (**NAACL-HLT'21**), 2021
- [26] “Beyond I.I.D.: Three Levels of Generalization for Question Answering on Knowledge Bases,” by Y. Gu, S. Kase, M. Vanni, B. Sadler, P. Liang, X. Yan, Y. Su, The World Wide Web Conf. (**WWW'21**), 2021
- [27] “CoCo: Controllable Counterfactuals for Evaluating Dialogue State Trackers,” by S. Li*, S. Yavuz*, K. Hashimoto, J. Li, T. Niu, N. Rajani, X. Yan, Y. Zhou and C. Xiong (*Equal Contribution) , International Conference on Learning Representations, Leaderboard No.1, Jan 2021-present in

Multiwoz (**ICLR'21**), 2021

- [28] “Inter-Series Attention Model for COVID-19 Forecasting,” by X. Jin, Y-X Wang, X. Yan, SIAM Int. Conf. on Data Mining (**SDM'21**), 2021
- [29] “KGPT: Knowledge-Grounded Pre-Training for Data-to-Text Generation,” by W. Chen, Y. Su, X. Yan, W. Wang, Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing, (**EMNLP'20**), 2020
- [30] “Adaptive-Step Graph Meta-Learner for Few-Shot Graph Classification,” by N. Ma, J. Bu, J. Yang, Z. Zhang, C. Yao, Z. Yu, S. Zhou, X. Yan, The 2020 ACM Int. Conf. on Information and Knowledge Management, (**CIKM'20**), 2020
- [31] “HierCon: Hierarchical Organization of Technical Documents based on Concepts,” by K. Li, S. Li, S. Yavuz, H. Zha, Y. Su, and X. Yan, Proc. 2019 IEEE Int. Conf. on Data Mining, (**ICDM'19**), 2019 (Best of ICDM 2019 selection)
- [32] “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting,” by S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, X. Yan, The Thirty-third Annual Conference on Neural Information Processing Systems, (**NeurIPS'19**), 2019
- [33] “How Large A Vocabulary Does Text Classification Need? A Variational Approach on Vocabulary Selection,” by W. Chen, Y. Su, Y. Shen, Z. Chen, X. Yan and W. Wang, Proc. of the 17th North American Chapter of ACL: Human Language Technologies, (**NAACL-HLT'19**), 2019
- [34] “Mining Algorithm Roadmap in Scientific Publications,” by H. Zha, W. Chen, K. Li and X. Yan, Proc. of the 25th Int. Conf. on Knowledge Discovery and Data Mining (**KDD'19**), 2019
- [35] “Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention,” by W. Chen, J. Chen, P. Qin, X. Yan and W. Wang, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'19**), 2019
- [36] “Global Textual Relation Embedding for Relational Understanding,” by Z. Chen, H. Zha, H. Liu, W. Chen, X. Yan and Y. Su, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'19**), 2019 (Short Paper)
- [37] “What It Takes to Achieve 100% Condition Accuracy on WikiSQL,” by S. Yavuz, I. Gur, Y. Su, X. Yan, Proc. of the Annual Conference on Empirical Methods in Natural Language Processing (**EMNLP'18**), 2018
- [38] “XL-NBT: A Cross-lingual Neural Belief Tracking Framework,” by W. Chen, J. Chen, Y. Su, X. Wang, D. Yu, X. Yan and W. Wang, Proc. of the Annual Conference on Empirical Methods in Natural Language Processing (**EMNLP'18**), 2018
- [39] “DialSQL: Dialogue Based Structured Query Generation,” by I. Gur, S. Yavuz, Y. Su, X. Yan, Proc. of the Annual Meeting of the Association for Computational Linguistics (**ACL'18**), 2018
- [40] “Variational Knowledge Graph Reasoning,” by W. Chen, W. Xiong, X. Yan and W. Wang, Proc. of the 16th North American Chapter of ACL: Human Language Technologies (**NAACL'18**) 2018
- [41] “Global Relation Embedding for Relation Extraction,” by Y. Su, H. Liu, S. Yavuz, I. Gur, H. Sun, X. Yan. Proc. of the 16th North American Chapter of ACL: Human Language Technologies (**NAACL'18**) 2018
- [42] “Concept Mining via Embedding,” by K. Li, H. Zha, Y. Su, X. Yan, Proc. 2018 Int. Conf. on Data Mining (**ICDM'18**), 2018
- [43] “Unsupervised Neural Categorization for Scientific Publications,” by K. Li, H. Zha, Y. Su, X. Yan, Proc. of the SIAM International Conference on Data Mining (**SDM'18**), 2018 (acceptance rate: 23%)
- [44] “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
- [45] “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
- [46] “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)
- [47] “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)
- [48] “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
- [49] “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
- [50] “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)
- [51] “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.
- [52] “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
- [53] “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
- [54] “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%)
- [55] “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%)
- [56] “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
- [57] “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
- [58] “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
- [59] “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%)
- [60] “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%)
- [61] “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**)
- [62] “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%)
- [63] “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%)
- [64] “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%)
- [65] “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%)
- [66] "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.
- [67] "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%)
- [68] "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.
- [69] "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**).
- [70] "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%)
- [71] "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%)
- [72] "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.
- [73] "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)
- [74] "Noise-Resistant Bicluster Recognition," by H. Sun, G. Miao, X. Yan, Proc. 2013 IEEE Int. Conf. on Data Mining (**ICDM'13**)
- [75] "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
- [76] "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%).
- [77] "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%).
- [78] "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%).
- [79] "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%).
- [80] "Characterizing Tenant Behavior for Placement and Crisis Mitigation in Multitenant DBMSs," by A. Elmore, S. Das, A. Pucher, D. Agrawal, A. El Abbad, X. Yan, Proc. 2013 Int. Conf. on Management of Data (**SIGMOD'13**).
- [81] "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%)
- [82] "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.
- [83] "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.
- [84] "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%)
- [85] "Ontology-based Subgraph Querying," Y. Wu, S. Yang, X. Yan, Proc. 2013 Int. Conf. on Data Engineering (**ICDE'13**), 2013 (acceptance rate, 19%)

- [86] “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%)
- [87] “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%)
- [88] “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%)
- [89] “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%).
- [90] “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%).
- [91] “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**)
- [92] “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%)
- [93] “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%).
- [94] “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%).
- [95] “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%).
- [96] “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%).
- [97] “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%).
- [98] “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%).
- [99] “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%).
- [100] “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.
- [101] “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%).
- [102] “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%)
- [103] “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
- [104] “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%)
- [105] “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%)
- [106] “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%)
- [107] “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%)
- [108] “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%)
- [109] “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%)
- [110] “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)
- [111] “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%)
- [112] “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)
- [113] “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%)
- [114] “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%)
- [115] “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%)
- [116] “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%)
- [117] “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%).
- [118] “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%)
- [119] “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%)

- [120] “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%)
- [121] “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%)
- [122] “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%)
- [123] “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%)
- [124] “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%)
- [125] “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%)
- [126] “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%)
- [127] “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%)
- [128] “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%)
- [129] “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%)
- [130] “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%)
- [131] “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%)
- [132] “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%)
- [133] “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%)
- [134] “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%)
- [135] “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%)
- [136] “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%)

- [137] “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%)
- [138] “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%)
- [139] “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%)
- [140] “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%).
- [141] “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%)
- [142] “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%)
- [143] “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%)
- [144] “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%)
- [145] “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%)
- [146] “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%).
- [147] “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 Journal Papers

- [1] “Improving topic disentanglement via contrastive learning,” X. Zhou, J. Bu, S. Zhou, Z. Yu, J. Zhao, X. Yan, Inf. Process. Manag. 60(2): 103164, 2023
- [2] “Heterogeneous Information Networks: the Past, the Present, and the Future,” Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi W, Proc. VLDB Endow. 15(12): 3807-3811, 2022
- [3] “Context-Guided Entropy Minimization for Semi-Supervised Domain Adaptation,” N. Ma, J. Bu, L. Lu, J. Wen, S. Zhou, Z. Zhang, J. Gu, H. Li, X. Yan, Neural Networks, 2022
- [4] “Cross-modal Image Retrieval with Deep Mutual Information Maximization,” C. Gu, J. Bu, X. Zhou, C. Yao, D. Ma, Z. Yu, X. Yan, Neurocomputing, 2022
- [5] “Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US,” Proceedings of the National Academy of Sciences (**PNAS**), Cramer et al. 2022
- [6] “Network Intervention for Mental Disorders with Minimum Small Dense Subgroups,” B.-Y. Hsu, C.-Y. Shen, and X. Yan, IEEE Transactions on Knowledge and Data Engineering (**TKDE**) Oct, 2019.
- [7] “Performance Bounds of Decentralized Search in Expert Networks for Query Answering,” by L. Ma, M. Srivatsa, D. Cansever, X. Yan, S. Kase, M. Vanni, ACM Transactions on Knowledge Discovery from Data (**TKDD**), 2019
- [8] “The genome of the jellyfish Aurelia and the evolution of animal complexity,” by D. Gold, T. Katsuki, Y. Li, X. Yan, M. Regulski, D. Ibberson, T. Holstein, R. Steele, D. Jacobs and R. Greenspan, Nature Ecology & Evolution, 2018

- [9] “Observability of Lattice Graphs,” by F. Han, S. Suri, and X. Yan, *Algorithmica*, 2015
- [10] “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
- [11] “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
- [12] “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.
- [13] “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
- [14] “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
- [15] “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
- [16] “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
- [17] “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
- [18] “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
- [19] “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
- [20] “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
- [21] “On Compressing Frequent Patterns,” D. Xin, J. Han, X. Yan, H. Cheng, *Data Knowl. Eng. (DKE)*, 60(1): 5– 29, 2007
- [22] “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
- [23] “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)
- [24] “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.
- [25] “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)
- [26] “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.
- [27] “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.

Workshop Papers, Demos, and Technical Reports

- [1] “Explanations from large language models make small reasoners better,” by S. Li, J. Chen, Y. Shen, Z.

- Chen, X. Zhang, Z. Li, H. Wang, J. Qian, B. Peng, Y. Mao, W. Chen, and X. Yan, Conference AAAI 2024 Workshop on Sustainable AI (**SAI-AAAI'24**)
- [2] “GauchoChat: Towards Proactive, Controllable, and Personalized Social Conversation,” by H. Wang, W. Wang, R. Saini, M. Zhukova, X. Yan, the 5th Proceedings of Alexa Prize SocialBot Grand Challenge (Alexa Prize 2023).
 - [3] “GauchoAI: A proactive modular embodied AI that collaborates with humans,” by J. Li, D. Nathani, T.-J. Fu, X. Wang, and X. Yan, the 1st Proceedings of Alexa Prize SimBot (Alexa Prize 2023)
 - [4] “Making Something out of Nothing: Building Robust Task-oriented Dialogue Systems from Scratch,” by Z. Li, H. Wang, A. Albalak, Y. Yang, J. Qian, S. Li, and X. Yan, the 1st Proceedings of Alexa Prize TaskBot (Alexa Prize 2021)
 - [5] “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)
 - [6] “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
 - [7] “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)
 - [8] “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)
 - [9] “Synthetic Review Spamming and Immunization,” by A. Morales, H. Sun, and X. Yan, the 22nd International World Wide Web Conference (**WWW'13**) (poster)
 - [10] “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
 - [11] “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%).
 - [12] “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.
 - [13] “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
 - [14] “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)
 - [15] “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)
 - [16] “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)
 - [17] “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.
 - [18] “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**).
 - [19] “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.

- [20] “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.

TALKS & TUTORIALS

July 2024	Plenary speaker, “Guiding Large Language Models via Small Ones,” International Congress of Basic Science (ICBS), Beijing
Apr 2024	Invited speaker, “Guiding Large Language Models via Small Ones,” SRI International
Apr 2024	Invited speaker, “Guiding Large Language Models via Small Ones,” Canon Research
Sept 2023	Distinguished speaker, “From 0 to 1,” Ningbo Xiaoshi High School
Dec 2022	Invited speaker, “Limitations of Language Models in Arithmetic Induction,” ICDM 2022 workshop Foundation Models in Vision and Language
Oct 2022	Distinguished speaker, “Transformers for Time Series Forecasting,” Instacart
May 2022	Future Directions of Data Mining (Panelist), ICDE 2022 Data Mining Round Table
Apr 2021	Invited speaker, “From Retrieval Machines to Conversational Agents,” UIUC NLP Seminar
Feb 2021	Invited speaker, “From Retrieval Machines to Intelligent Chatbots,” GoogleX
Nov 2020	“History Repeats Itself: COVID-19 Forecasting,” Weekly CDC/MIDAS conference call
Nov 2020	“History Repeats Itself: COVID-19 Forecasting,” COVID-19 Research Lightning Talks: Webinar and Q&A, Columbia University
Oct 2020	“Achieving State-of-the-Art Performance in COVID-19 Hospitalization Forecasting,” 2020 Responsible Machine Learning Summit
Aug 2020	South Coast Researchers Develop New Forecast Model Using Artificial Intelligence, KCLU NPR for the California coast
July 2020	“COVID-19 Trend Forecasting Using Neural Models,” Lightning talk at the CRA Virtual Conference
Apr 2020	“New Intelligent Methods for IT Operations,” Microsoft Research
Jan 2019	Invited speaker, “Linking Intelligent Systems as Distributed Knowledge Bases,” ASPSM workshop, UT Austin
Nov 2018	Invited speaker, “Graph Query, Graph Mining, and Privacy Issues,” Privacy in Graphs (PiG) Workshop, UCSC
May 2018	“AI and Deep Learning,” Amgen Headquarter
Apr 2018	“Scalable Construction and Querying of Massive Knowledge Bases,” by Xiang Ren, Yu Su, Xifeng Yan, Conference Tutorial, the 2018 World Wide Web Conf.
Feb 2018	Invited speaker, “Some Thoughts on Research,” Doctoral Consortium, the 11th ACM International Conference on Web Search and Data Mining, (WSDM’2018)
Feb 2018	Invited speaker, “Democratize Data Science: Natural Language Interface to Data,” the First Workshop on Knowledge Base Construction, Reasoning and Mining (KBCOM’18)
Nov 2017	“Construction and Querying of Large-scale Knowledge Bases,” by Xiang Ren, Yu Su, Xifeng Yan. Conference Tutorial In Proc. of the ACM International Conference on Information and Knowledge Management (CIKM), 2017

Sept 2017 Invited speaker/panelist, "Natural Language Question Answering," Alibaba Seattle Research Center

Jul 2017 Invited speaker, "Graph Data Management and Question Answering," Alibaba Inc. Hangzhou

Jul 2017 Invited speaker, "Question Answering Benchmark and Semantic Parsing," Fudan University

Jun 2016 Invited speaker, "Knowledge Graphs and Neural Networks," by HGData, Santa Barbara

Dec 2015 Invited speaker, "Networks and Knowledge Bases," by 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," University of Southern California

Oct 2015 Invited speaker, "Network Centric Information Search and Question Answering," SnapChat

Sept 2015 "Network Centric Information Search and Question Answering," Baidu Silicon Valley AI Lab

Aug 2015 "Network Centric Information Search and Question Answering," Fudan University

Aug 2015 Keynote speaker, "Graph OLAP, Anomaly and Query-Based Outlier Detection," KDD 2015 workshop, ODDx3: Outlier Definition, Detection, and Description

July 2015 Invited speaker, "Network Centric Information Search and Question Answering," University of Electronic Science and Technology at Xi'an

July 2015 Invited speaker, "Network Centric Information Search and Question Answering," Huazhong University of Science and Technology

July 2015 Invited speaker, "Network Centric Information Search and Question Answering," Shandong University

July 2015 Invited speaker, "Network Centric Information Search and Question Answering," Big Data and Internet + Seminar, Donghua University

Apr 2015 "Schemaless Graph Querying," Computer Science Department, UC Irvine

Sept 2014 Invited Speaker, "Graph Analysis and Search in Networks," Computer Science Department, Peking University, 2014

Sept 2014 Invited Speaker, "Graph Analysis and Search in Networks," 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,” University of Wisconsin at Madison, Database Seminar
Jan 2011	“Graph Pattern Mining and Distributed Processing,” INARC/NS-CTA, ARL
Nov 2010	“Graph Pattern Mining and System,” Microsoft Research Asia, Beijing.
Sept 2010	“Graph Pattern Mining,” 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” 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,” 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,” 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,” NEW-CEGS SEMINAR, Center of Excellence in Genomic Science, Molecular and Computational Biology, University of Southern California, April 22, 2005

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	“Automated System Problem Diagnosing,” by P. Zhou, B. Gill, W. Belluomini, X. Yan, approved as US Patent # US8112667 (Feb, 2012)

- 2008 “Method and Apparatus for Structural Data Classification,” by H. Cheng, W. Fan, X. Yan, P. S. Yu, approved as US Patent # US8121967B2 (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 #US8117208B2 (Feb, 2012)
- 2005 “System and Method for Efficiently Performing Similarity Searches of Structural Data,” by X. Yan and P. S. Yu, approved as US Patent # US9165042B2 (Oct, 2015)
- 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)

TEACHING EXPERIENCE

- Fall 2024 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Spring 2024 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Winter 2024 **Special Topics in Conversational AI**
Graduate level, UC at Santa Barbara (CS291K), Lecture
- Spring 2023 **Machine Learning**
Undergraduate level, UC at Santa Barbara (CS165B), Lecture
- Winter 2023 **Special Topics in Deep Learning**
Graduate level, UC at Santa Barbara (CS291A), Lecture
- Fall 2022 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Spring 2022 **Special Topics in Deep Learning**
Graduate level, UC at Santa Barbara (CS291K), Lecture
- Winter 2022 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Fall 2021 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture
- Spring 2021 **Machine Learning**
Undergraduate level, UC at Santa Barbara (CS165B), Lecture
- Winter 2021 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Spring 2020 **Machine Learning**
Undergraduate level, UC at Santa Barbara (CS165B), Lecture

Spring 2020 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Spring 2019 **Data and Knowledge Bases**
Graduate level, UC at Santa Barbara (CS273), Lecture

Spring 2018 **Advanced Deep Learning**
Graduate level, UC at Santa Barbara (CS291K), Lecture

Winter 2018 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Fall 2017 **Data and Knowledge Bases**
Graduate level, UC at Santa Barbara (CS273), Lecture

Spring 2017 **Advanced Data Mining (Neural Networks)**
Graduate level, UC at Santa Barbara (CS291K), Lecture

Winter 2017 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Fall 2016 **Data and Knowledge Bases**
Graduate level, UC at Santa Barbara (CS273), Lecture

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 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Fall 2014 **Network Science**
Graduate level, UC at Santa Barbara (CS90A), Lecture

Fall 2014 **Data and Knowledge Bases**
Graduate level, UC at Santa Barbara (CS273), Lecture

Spring 2014 **Advanced Data Mining**
Graduate level, UC at Santa Barbara (CS290D), Lecture

Winter 2014 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Spring 2013 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Winter 2012 **Artificial Intelligence**
Undergraduate level, UC at Santa Barbara (CS165A), Lecture

Fall 2011 **Data and Knowledge Bases**
Graduate level, UC at Santa Barbara (CS273), Lecture

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

2024	LaViA: Fine-Tuning Multimodal LLMs as Task Assistants with Video Instructions
2023	LongMem: Augmenting Language Models with Long-Term Memory
2023	VaLM: Visually-Augmented Language Modeling
2023	DSP: Guiding Large Language Models via Directional Stimulus Prompting
2023	ViTST, Time Series as Images: Vision Transformer for Irregularly Sampled Time Series
2022	Dialogic: Controllable Dialogue Simulation with In-Context Learning
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, http://zhoulab.usc.edu/NeMo/source.htm (USC, MIT/Whitehead, UCSC, UNC, SFU)
2006	gIndex: Index and search large graph sets
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
2005	CloseGraph: Frequent Closed Graph Mining Package
2004	gSpan: Frequent Graph Mining Package, http://www.cs.ucsb.edu/~xyan/software

POSTDOC and PHD. ALUMNI

2024	Alon Albalak, “Improving the Understanding Through a Data-Centric Lens,” SynthLabs
2023	Hong Wang, “Efficient Natural Language Processing with Limited Data and Resources,” Applied Scientist, Amazon AWS
2023	Shiyang Li, “Label-efficient Learning in Natural Language Processing,” Applied Scientist, Amazon Research
2022	Jing Qian, “Text Detoxification in Natural Language Processing,” Research Scientist, Microsoft
2022	Zhiyu Chen, “Knowledge-Grounded Natural Language Processing,” Assistant Professor, Univ. of Texas at Dallas
2022	Xiaoyong Jin, “Modern Machine Learning in Time Series Forecasting,” Applied Scientist, Amazon AWS
2021	Hanwen Zha, “Towards Effort-Saving Knowledge Mining and Reasoning Over Web,” Research Scientist, Facebook
2021	Wenhu Chen, “Accessing Diverse Web Knowledge with Natural Language Interface,” Assistant Professor, University of Waterloo
2019	Bay-Yuan Hsu, “Therapy Group Formation and Social Network Intervention with Effectiveness Optimization,” Assistant Professor, National Taipei University
2019	Keqian Li, “Mining and Analyzing Technical Knowledge Based on Concepts,” Research Scientist, Yahoo Research
2019	Semih Yavuz, “DeepAssist: Deep Knowledge Grounding for Factual and Conversational Natural Language Interfaces, Research Scientist,” Salesforce Research
2019	Izzeddin Gur, “Learning Natural Language Interfaces using Deep Neural Networks,” Research Scientist, Google Research
2018	Yu Su, “Bridging the Gap between Human and Data with AI,” Assistant Professor, Ohio State University
2017	Honglei Liu, “Multi-level Knowledge Extraction from Sequence Data,” Research Scientist, Facebook
2017	Theodore Georgiou, “Understanding the Real World through the Analysis of User Behavior and Topics in Online Social Media,” Software Engineer, Google
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

2024	Krushna Chirag Shah, "NutriFacts: A Pipeline for Nutrition Label Information Extraction and Retrieval"
2024	Sahil Naik, "Fair Gain," Toyon Research Corporation
2024	Vihaan Akshaay, "AI for Landslide Detection: Current Models and Challenges in Dataset Development"
2024	Connor Rapelje, "Robotic Process Automation with Large Language Models: Generating and Modifying Scripts from Video Demonstrations"
2024	Luke Yoffe, "Mitigating Hallucinations in Large Language Model Agent Communication with Uncertainty Metrics," Meta
2023	Rajan Saini, "Building an Engaging Socialbot with Open-source LLMs," BS/MS, Google
2023	Aaron Jimenez, M.S., "From Videos to Actions: Exploiting Video Captioning and a Large Language Model for Sequential Action Identification and Prediction," Yardi Systems
2023	Reeva Mishra, M.S., "Online Text Summarization," Microsoft
2022	Apoorva Jakalannanavar Halappa Manjula, M.S., "Abstractive Dialogue Summarization for Customer Service," Roku
2021	Krishna Chaitanya Gudipati, M.S., "Tools to Automate RPA Script Generation," TruckX
2021	Benu Changmai, M.S. "Robotic Process Automation (RPA) as a Tool for the Visually

	Impaired,” Microsoft
2020	Yuxiao Zhang, M.S., “Paraphrasing via Transfer Learning”
2020	Xiyou Zhou, M.S. “Benchmarking Energy Efficiency for Responsible Natural Language Processing,” OctoML
2020	Hongmin Wang, M.S. “Revisiting Challenges in Data-to-Text Generation with Fact Grounding,” Microsoft
2019	Mu Xu, M.S., “Blog Mining and Retrieval System for Easier Paper Reading,” Facebook
2017	Anh Nguyen, M.S. “A stacked attention model for finding similar questions in community question answering systems,” Google
2016	Anastasiya Lazareva, M.S. “ECNN: A Convolutional Neural Network Generator based on Adaptive Evolutionary Algorithms,” Elemental Analytics
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

2024	Andre Braga, “Adaptive Large Language Models in Financial Contexts,” UCSB McNair Scholars
2024	Ivy Wang, “Transformer based Time Series Forecasting,” UCSB
2024	James Michael Pflaging, “Robotic Process Automation Using LLMs,” BS/MS, UCSB
2023	Xiyue Wang, Univ. of California at Santa Barbara (Columbia University)
2022	Xuan Luo, Univ. of California at Santa Barbara

2022	Ari Polakof, Univ. of California at Santa Barbara (Appfolio, GaTech)
Summer 2021	Qiru Hu, "Amazon Alexa Taskbot Prize," Univ. of California at Santa Barbara
Summer 2021	Carina Quan, "Amazon Alexa Taskbot Prize," Univ. of California at Santa Barbara
Summer 2021	Tom Zu, "Amazon Alexa Taskbot Prize," Univ. of California at Santa Barbara
Spring 2021	Tanay Komarlu, "Video Subtitle Extraction and Speaker Diarization," Univ. of California at Santa Barbara (University of Illinois at Urbana Champaign)
Summer 2020	Qiusi Zhan, Peking University (University of Illinois at Urbana Champaign)
Summer 2018	Chen Song, Beijing University of Posts and Telecommunications
Summer 2018	Nuan Wen, Univ. of California at Santa Barbara
Summer 2018	Yunkai Zhang, "Time Series Forecasting," Univ. of California at Santa Barbara (UC Berkeley)
Summer 2018	Andy Jin, Stanford
Summer 2017	Fernando Diaz Piaz, California Polytechnic State University, supervised by Yi Ding, "Query expansion with multiple word embeddings." (Network Science IGERT)
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."
Winter 2012	Lijia Hong, "Nokia Challenge," HKUST
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

Spring 2024	Melody Yu, Sage Hill School, “Bot or human? detecting ChatGPT imposters with a single question”
Summer 2017	Albert Miao, Dos Pueblos High School, supervised by Xiaoyong Jin, “Gender classification in images.”
Summer 2017	Andy Jin, Westlake High School, supervised by Xiaoyong Jin, “Fake news detection,” (UCSB Research Mentorship Program), admitted by Harvard, Princeton, Stanford, and Yale
Summer 2016	Aidan Barbieux, Santa Barbara High School, supervised by Yi Ding, “Graph analysis and visualization of flight safety.” (Network Science IGERT)
Summer 2016	Jared Fitton, Alberto Gil, Aaron Peyton, Santa Barbara High School, supervised by Yi Ding, “Modeling election patterns using graph techniques.” (Network Science IGERT)
Summer 2016	Byron Osborne, Santa Barbara High School, supervised by Yi Ding, “Network security and graphs.” (Network Science IGERT)
Summer 2016	Nick Ding, Princeton Plainsboro High School North, supervised by Yi Ding, “Graph analysis and visualization of flight safety.” (Network Science IGERT)
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.

VISITING SCHOLAR

2018	Shuguang Zhu, Beijing University of Posts and Telecommunications
------	--

STUDENT HONORS AND AWARDS

2023	Team GauchoChat won the overall first-place prize for the Alexa Prize SocialBot Grand Challenge 5
2023	Team GauchoAI won the second-place prize for the Alexa Simbot Challenge
2022	Team GauchoBot entered the finalist of Amazon Alexa Taskbot Challenge and constantly received the highest rate by Amazon Alexa users
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

2022-2023 “Alexa Prize SocialBot Grand Challenge 5,” X. Yan (PI). Grant: \$250K

2022-2023 “Amazon SimBot Prize Challenge,” X. Yan (coPI). Grant: \$250K

2021-2025 PARC Digital Workforce in Perceptually-enabled Task Guidance (PTG), X. Yan (Subcontractor). Grant: \$600K

2021-2022 “Amazon TaskBot Prize Challenge,” X. Yan (PI). Grant: \$250K

2020-2021 NSF RAPID: “Interventional COVID-19 Response Forecasting in Local Communities Using Neural Domain Adaptation Models,” X. Yan (PI). Grant: \$200K

2019-2020 VISA Faculty Research Gift Award. Grant: \$50K

2017-2018 VISA Faculty Research Gift Award. Grant: \$35K

2017-2018 Huawei Technologies Co., Ltd. Machine Learning Algorithm Selection for Storage Resource Management, X. Yan (PI). Grant: \$180K

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

2015-2017 NSF CCF: “EAGER: Collaborative Research: Leveraging Graph Databases for Incremental and Scalable Symbolic Analysis and Verification of Web Applications,” X. Yan (coPI). Grant: \$150K

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

- Program Co-chair: 2023 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
- Program Area Chair: NeurIPS 2023 Datasets and Benchmarks Track
- Research Award Committee: SIAM International Conference on Data Mining, 2022, 2024
- Senior Program Committee Member: ACM International Conference on Information and Knowledge Management 2013, 2022
- Program Co-Chair: 2019 SIAM Conference on Data Mining (SDM’19)

- 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)
- 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: AAAI Conference on Artificial Intelligence (AAAI' 21); ACM Int. Conf. on Management of Data (SIGMOD'10-13, 22); ACM Int. Con. on Knowledge Discovery and Data Mining (SIGKDD'08-15); Int. Conf. on Very Large Databases (VLDB 2013-15); WWW 2012-14, 2022; IEEE Int. Conf. Data Engineering (ICDE'10-12, 20); 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, 2020-2022
- Ireland Science Foundation, 2016, 2019
- 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, 2022; 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, 2019-20; Data Mining and Knowledge Discovery (DAMI) 2006, 2007, 2022; Data & Knowledge Engineering (DKE) 2008-2010; Knowledge and Information Systems (KAIS) 2009-2012, 2020; Statistical Analysis and Data Mining (SAM), 2009-11; Bioinformatics 2005, 2006, 2008; Machine Learning Journal 2007, 2009-10, 2020; Journal of Machine Learning 2020, 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: Univ. of British Columbia, NUS School of Computing, NTU
- Proposal Referee: National Aeronautics and Space Administration (NASA), 2004.