

ARIJIT KHAN

Assistant Professor
School of Computer Science and Engineering
Nanyang Technological University
Singapore

Updated: May 20, 2019
Email: arijit.khan@ntu.edu.sg
Web: ntu.edu.sg/home/arijit.khan
Blog: big-graph-live.blogspot.com

RESEARCH INTERESTS

Data management for the emerging problems in large graphs, with a focus on user-friendly, efficient, and approximate querying and pattern mining in social and information networks, using scalable algorithms, machine learning techniques, and distributed systems.

keywords: big graphs, big data, graph systems, knowledge graphs, uncertain graphs, graph streams, databases, data mining, machine learning, algorithms, crowdsourcing.

ACADEMIC POSITIONS AND PRIOR EMPLOYMENT

- **Computer Science, Nanyang Technological University** Singapore
Assistant Professor (Tenure-track) Dec. 2015 - now
- **Systems Group, ETH Zurich** Zurich, Switzerland
Postdoctoral Researcher, Host: Donald Kossmann Sep. 2013 - Nov. 2015
- **Computer Science, UC Santa Barbara** Santa Barbara, CA, USA
Graduate Research Assistant Sep. 2008 - Sep. 2013
- **Yahoo! Research Lab** Barcelona, Spain
Research Intern, Mentor: Francesco Bonchi and Aris Gionis June 2012 - Sep. 2012
- **IBM T. J. Watson Research Center** Hawthorne, NY, USA
Research Intern, Mentor: Shu Tao, Manager: Nikos Anerousis June 2010 - Sep. 2010
- **Electrical Engineering, Indian Institute of Science** Bangalore, India
Summer Research Fellow, Mentor: Lawrence Jenkins June 2007 - Sep. 2007

EDUCATION

- **University of California, Santa Barbara** Santa Barbara, CA, USA
Ph.D., Computer Science Sep. 2008 - Sep. 2013
Thesis Towards Querying and Mining of Large-Scale Networks
Ph.D. adviser Xifeng Yan, GPA: 3.95/4
- **Jadavpur University** Kolkata, India
B.E., Computer Science and Engineering Sep. 2004 - June 2008
GPA: 9.39/10, First Class (Honors), ranked 1st among 62 students

FUNDING

- **NRF-ST Engineering-NTU Corporate Lab Grant** Singapore
PI, SG\$ 585,000 Jan. 2018 - Dec. 2019
Health Monitoring and Predictive Maintenance for UAV through Data Analytics
- **AcRF MOE Tier 1 Grant** Singapore
PI, SG\$ 100,000 Nov. 2016 - Jan. 2019
Influence Maximization in Social Networks: From Reel to Real World
- **NTU Start-Up Grant** Singapore
PI, SG\$ 500,000 and 4 PhD Students Dec. 2015 - Dec. 2019
Big-Graphs Analytics and Systems: Supporting User-friendly Querying

PUBLICATIONS

[h-index: 15, citations: 1145 per Google Scholar]

Peer-Reviewed Books

[1] Arijit Khan, Yuan Ye, and Lei Chen, *On Uncertain Graphs*, in Morgan & Claypool Publishers, Synthesis Lectures on Data Management, 2018.

Peer-Reviewed Book Chapters

[2] Yinghui Wu and Arijit Khan, *Graph Pattern Matching*, in Sherif Sakr and Albert Zomaya (eds.), *Encyclopedia of Big Data Technologies*, Springer, 2018.

[3] Arijit Khan and Sayan Ranu, *Big-Graphs: Querying, Mining, and Beyond*, in Sherif Sakr and Albert Zomaya (eds.), *Handbook of Big Data Technologies*, Springer, 2017.

Peer-Reviewed Journal and Conference Papers

Note that in many areas within computer science, including databases and data mining, *conferences* (not journals) are the primary venues for peer-reviewed publications.

[4] **SIGMOD 2019**: Francesco Bonchi, Arijit Khan, and Lorenzo Severini, *Distance-generalized Core Decomposition*, in Proc. of ACM International Conference on Management of Data 2019.

[5] **PVLDB 2019**: Xiangyu Ke, Arijit Khan, and Leroy Lim Hong Quan, *An In-Depth Comparison of s-t Reliability Algorithms over Uncertain Graphs*, in Proc. of the VLDB Endowment, 2019.

[6] **TKDE 2019**: Tenindra Abeywickrama, Muhammad Aamir Cheema, and Arijit Khan, *K-SPIN: Efficiently Processing Spatial Keyword Queries on Road Networks*, in IEEE Transactions on Knowledge and Data Engineering Journal 2019, [Impact Factor=3.44].

[7] **SIGMOD 2018**: Xiangyu Ke, Arijit Khan, and Gao Cong, *Finding Seeds and Relevant Tags Jointly: For Targeted Influence Maximization in Social Networks*, in Proc. of ACM

International Conference on Management of Data 2018, [Acceptance Rate: 90/461 (20%)].

[8] **PVLDB 2018**: Xiangyu Ke, Michelle Teo, Arijit Khan, and Vijaya Krishna Yalavarthi, *A Demonstration of PERC: Probabilistic Entity Resolution With Crowd Errors*, (Demo), in Proc. of the VLDB Endowment, 11(12):1922–1925, 2018.

[9] **USENIX ATC 2018**: Arijit Khan, Gustavo Segovia, and Donald Kossmann, *On Smart Query Routing: For Distributed Graph Querying with Decoupled Storage*, in Proc. of USENIX Annual Technical Conference 2018, [Acceptance Rate: 76/378 (20%)].

[10] **TKDE 2018**: Arijit Khan, Francesco Bonchi, Francesco Gullo, and Andreas Nufer, *Conditional Reliability in Uncertain Graphs*, in IEEE Transactions on Knowledge and Data Engineering Journal 2018, [Impact Factor=3.44].

[11] **ICDE 2018**: Shanshan Feng, Gao Cong, Arijit Khan, Xiucheng Li, Yong Liu, and Yeow Meng Chee, *Inf2vec: Latent Representation Model for Social Influence Embedding*, in Proc. of IEEE International Conference on Data Engineering 2018.

[12] **IEEE BigData 2018**: Siyuan Liu and Arijit Khan, *An Empirical Analysis on Expressibility of Vertex Centric Graph Processing Paradigm*, in Proc. of IEEE International Conference on Big Data 2018, [Acceptance Rate: 98/518 (19%)].

[13] **IEEE BigData 2018**: Vijaya Krishna Yalavarthi and Arijit Khan, *Steering Top-k Influencers in Dynamic Graphs via Local Updates*, (Short Paper), in Proc. of IEEE International Conference on Big Data 2018, [Acceptance Rate: 103/518 (20%)].

[14] **CIKM 2017**: Vijaya Krishna Yalavarthi, Xiangyu Ke, and Arijit Khan, *Select Your Questions Wisely: For Entity Resolution With Crowd Errors*, in Proc. of ACM International Conference on Information and Knowledge Management 2017, [Acceptance Rate: 171/820 (21%)].

[15] **EDBT 2017**: Arijit Khan, *Vertex-Centric Graph Processing: Good, Bad, and the Ugly*, (Short Paper), in Proc. of International Conference on Extending Database Technology 2017, [Acceptance Rate: 22/93 (24%)].

[16] **SNAM 2017**: Arijit Khan and Charu Aggarwal, *Toward Query-Friendly Compression of Rapid Graph Streams*, in Springer Social Network Analysis and Mining Journal, 7(1):23:1-23:19, 2017, (**invited** and peer-reviewed), [Impact Factor=1.86].

[17] **SIGMOD 2016**: Pratanu Roy, Arijit Khan, and Gustavo Alonso, *Augmented Sketch: Faster and More Accurate Stream Processing*, in Proc. of ACM International Conference on Management of Data 2016, [Acceptance Rate: 20%].

[18] **ICDE 2016**: Arijit Khan, Benjamin Zehnder, and Donald Kossmann, *Revenue Maximization by Viral Marketing: A Social Network Host's Perspective*, in Proc. of IEEE International Conference on Data Engineering 2016, [Acceptance Rate: 25%].

[19] **ICDE 2016**: Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez

Elmasri, *Querying Knowledge Graphs by Example Entity Tuples*, (Extended Abstract), in Proc. of IEEE International Conference on Data Engineering 2016.

[20] **CIKM 2016**: Arijit Khan, *Towards Time-Discounted Influence Maximization*, (Short Paper), in Proc. of ACM International Conference on Information and Knowledge Management 2016, [Acceptance Rate: 300/925 (32%)].

[21] **ASONAM 2016**: Arijit Khan and Charu Aggarwal, *Query-Friendly Compression of Graph Streams*, in Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2016, [Acceptance Rate: 43/316 (14%)].

[22] **TKDE 2015**: Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez Elmasri, *Querying Knowledge Graphs by Example Entity Tuples*, in IEEE Transactions on Knowledge and Data Engineering Journal, 27(10): 2797-2811, 2015, [Impact Factor=3.44].

[23] **CIKM 2015**: Arijit Khan, Francesco Gullo, Thomas Wohler, and Francesco Bonchi, *Top-k Reliable Edge Colors in Uncertain Graphs*, (Short Paper), in Proc. of ACM International Conference on Information and Knowledge Management 2015, [Acceptance Rate: 25%].

[24] **SSDBM 2015**: Arijit Khan and Vishwakarma Singh, *Answering Top-k Representative Queries with Binary Constraints*, in Proc. of International Conference on Scientific and Statistical Database Management 2015, [Acceptance Rate: 26/71 (37%)].

[25] **SIGMOD 2014**: Arijit Khan, Pouya Yanki, Bojana Dimcheva, and Donald Kossmann, *Towards Indexing Functions: Answering Scalar Product Queries*, in Proc. of ACM International Conference on Management of Data 2014, [Acceptance Rate: 107/421 (25.41%)].

[26] **EDBT 2014**: Arijit Khan, Francesco Bonchi, Aris Gionis, and Francesco Gullo, *Fast Reliability Search in Uncertain Graphs*, in Proc. of International Conference on Extending Database Technology 2014, [Acceptance Rate: 20%].

[27] **ICDE 2014**: Nandish Jayaram, Mahesh Gupta, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez Elmasri, *GQBE: Querying Knowledge Graphs by Example Entity Tuples*, (Demo), in Proc. of IEEE International Conference on Data Engineering 2014, [Acceptance Rate: 28/65 (43.1%)].

[28] **PVLDB 2013**: Arijit Khan, Yinghui Wu, Charu Aggarwal, and Xifeng Yan, *NeMa: Fast Graph Search with Label Similarity*, in Proc. of the VLDB Endowment, 6(3): 181-192, 2013, [Acceptance Rate: 22.7%].

[29] **SIGMOD 2012**: Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan, *Towards Effective Partition Management for Large Graphs*, in Proc. of ACM International Conference on Management of Data 2012, [Acceptance Rate: 16.6%].

[30] **CIKM 2012**: Nan Li, Xifeng Yan, Zhen Wen, and Arijit Khan, *Density Index and Proximity Search in Large Graphs*, in Proc. of ACM International Conference on Infor-

mation and Knowledge Management 2012, [Acceptance Rate: 146/1088 (13.4%)].

[31] **SIGMOD 2011:** Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao, *Neighborhood Based Fast Graph Search in Large Networks*, in Proc. of ACM International Conference on Management of Data 2011, [Acceptance Rate: 23%].

[32] **SDM 2011:** Charu Aggarwal, Arijit Khan, and Xifeng Yan, *On Flow Authority Discovery in Social Networks*, in Proc. of SIAM International Conference of Data Mining 2011, [Acceptance Rate: 25.1%].

[33] **SIGMOD 2010:** Arijit Khan, Xifeng Yan, and Kun-Lung Wu, *Towards Proximity Pattern Mining in Large Graphs*, in Proc. of ACM International Conference on Management of Data 2010, [Acceptance Rate: 20.8%].

Peer-Reviewed Tutorials

[34] **PVLDB 2017:** Arijit Khan, Sourav S. Bhowmick, and Francesco Bonchi, *Summarizing Static and Dynamic Big Graphs*, in Proc. of the VLDB Endowment, 10(12): 1981-1984, 2017, [Acceptance Rate: 8/16 (50%)].

[35] **PVLDB 2015:** Arijit Khan and Lei Chen, *On Uncertain Graphs Modeling and Queries*, in Proc. of the VLDB Endowment, 8(12): 2042-2043, 2015, [Acceptance Rate: 6/20 (30%)].

[36] **PVLDB 2014:** Arijit Khan and Sameh Elnikety, *Systems for Big-Graphs*, in Proc. of the VLDB Endowment, 7(13): 1709-1710, 2014.

[37] **ICDE 2012:** Arijit Khan, Yinghui Wu, and Xifeng Yan, *Emerging Graph Queries In Linked Data*, in Proc. of IEEE International Conference on Data Engineering 2012.

Peer-Reviewed Workshop Presentations

[38] **SIAM NS 2018:** Arijit Khan, *Conditional Reliability and Influence Maximization over Social Networks*, in SIAM Workshop on Network Science 2018.

[39] **GRADES 2014:** Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez Elmasri, *Towards a Query-by-Example System for Knowledge Graphs*, in Proc. of Graph Data-management, Experiences, and Systems 2014, co-located with International Conference on Management of Data 2014 (SIGMOD 2014).

[40] **GDM 2012:** Arijit Khan, Vishwakarma Singh, and Jian Wu, *Find Skyline Nodes in Large Networks*, in Proc. of International Workshop on Graph Data Management: Techniques and Applications 2012, co-located with International Conference in Data Engineering 2012 (ICDE 2012).

[41] **CloudMan 2012:** Arijit Khan, Xifeng Yan, Shu Tao, and Nikos Anerousis, *Workload Characterization and Prediction in the Cloud: A Multiple Time Series Approach*, in Proc. of Cloud Management 2012, co-located with IEEE/IFIP Network Operations and

Management Symposium 2012 (NOMS 2012).

[42] **WISARD 2008:** Arijit Khan and Lawrence Jenkins, *Undersea Wireless Sensor Network for Ocean Pollution Prevention*, in Proc. of Wireless Systems: Advanced Research and Development 2008, co-located with IEEE Communication Systems Software and Middleware 2008 (COMSWARE 2008).

Invited Articles

[43] Arijit Khan and Yinghui Yu, *Graph Pattern Matching Queries – Approximation and User-friendliness*, in the ACM SIGMOD Blog 2017.

Editorials and Front Matters

[44] Fusheng Wang, Gang Luo, Chunhua Weng, Arijit Khan, Prasenjit Mitra, and Cong Yu (eds.), *Biomedical Data Management and Graph Online Querying, [front matter]*, VLDB 2015 Workshops, Big-O(Q) and DMAH, in Lecture Notes in Computer Science, Springer.

HONORS & AWARDS

- **Invited** for Dagstuhl Seminar on “Big Graph Processing Systems”, December 2019, Schloss Dagstuhl - Leibniz Center for Informatics, Germany.
- **Invited** to lead a panel discussion among industry leaders on synergies between data analytics and machine learning at the 3rd Edition of GFMI Conference Optimizing Data Governance, Quality and Consistency in Financial Services, 2019, Singapore.
- **Invited** for National Institute of Informatics (NII) Shonan Meeting on “Graph Database Systems: Bridging Theory, Practice, and Engineering”, July 2018, Japan.
- **Invited** to contribute a chapter in the “Springer Encyclopedia of Big Data Technologies”.
- **Invited** to present a tutorial at the Asia Pacific Web and Web-Age Information Management Joint Conference on Web and Big Data (APWeb-WAIM 2017).
- **Invited** to contribute an article in the ACM SIGMOD Blog, 2017.
- **Invited** to present a tutorial at 21st International Conference on Management of Data (COMAD 2016).
- **Invited** to submit our IEEE/ACM ASONAM 2016 paper on graph-streams in Springer Journal of Social Network Analysis and Mining (SNAM).
- **Invited** to submit an extended version of our VLDB 2015 tutorial as a book in Morgan & Claypool’s Data Management series.
- **Invited** to contribute a chapter in the Springer handbook on Big Data Technologies.

- **Received**, VLDB 2014 Tutorial Honorarium, (US\$ 350).
- Part of our VLDB 2014 Tutorial on Big-Graphs Systems included in the Large Scale Data Management (CS 848) Course 2015, University of Waterloo.
- **Invited** for Dagstuhl Seminar on “Systems and Algorithms for Large Scale Graph Analysis”, November 2014, Schloss Dagstuhl - Leibniz Center for Informatics, Germany.
- **Received**, IBM Ph.D. Fellowship 2012-13, (US\$ 30,000).
- **Received**, NSF ICDE 2012 Scholarship, (US\$ 725).
- **Received**, NSF SDM 2011 Student Travel Award, (US\$ 1,000).
- **Received**, P1 fellowship 2009-10, Computer Science, UC Santa Barbara, (US\$ 4,500).
- **Received**, CITRIX GO-TO fellowship 2008-09, Computer Science, UC Santa Barbara, (US\$ 8,000).
- **Winner**, Gold Medal from the Department of Computer Science and Engineering, Jadavpur University in 2008.
- **Winner**, Gold Medal by Tata Consultancy Services Ltd. for being the best student of the Department of Computer Science and Engineering, Jadavpur University in 2008.

INVITED TALKS

- *Synergies Between Data Analytics and Machine Learning*
 - 3rd Edition of **GFMI Conference** Optimizing Data Governance, Quality and Consistency in Financial Services, **Singapore** (March 2019).
- *Data Management for Emerging Problems in Large Networks*
 - **Singapore Management University, Singapore** (July 2018).
 - **Northeastern University, Boston, Massachusetts, USA** (July 2018).
 - **Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA** (July 2018).
 - **National Institute of Informatics (NII) Shonan Meeting, Japan** (July 2018).
- *Querying Big-graphs: Streaming and Beyond*
 - **National University of Singapore, Singapore** (September 2017).
 - **Peking University, Beijing, China** (July 2017).
 - **ISI Foundation, Torino, Italy** (February 2017).
 - **A*STAR/I2R (Institute for Infocomm Research), Singapore** (January 2017).
 - **Monash University, Melbourne, Australia** (October 2015).
 - **RMIT University, Melbourne, Australia** (October 2015).
 - **University of California, Los Angeles, USA** (September 2015).

- *On Uncertain Graphs Modeling and Queries*
 - **ISI Foundation, Torino, Italy** (June 2019).
 - **Invited Tutorial** in the Asia Pacific Web and Web-Age Information Management Joint Conference on Web and Big Data (**APWeb-WAIM 2017**).
 - **IBM Research, Delhi, India** (September 2016).
 - **Invited Tutorial** at 21st International Conference on Management of Data 2016 (**COMAD 2016**).
- *Towards Querying and Mining of Big-Data*
 - **HP Innovation Center, Singapore** (June 2017).
 - **IIT Kharagpur, India** (July 2015).
 - **Microsoft Research, Bangalore, India** (June 2015).
 - **NEC Labs, Cupertino, USA** (May 2015).
 - **University of California, Riverside, USA** (March 2015).
 - **Oregon State University, USA** (February 2015).
 - **Nanyang Technological University, Singapore** (February 2015).
 - **Indian Institute of Science, Bangalore, India** (January 2015).
 - **IIT Kanpur, India** (December 2014).
 - **Dagstuhl Seminar, Schloss Dagstuhl - Leibniz Center for Informatics, Germany** (November 2014).
 - **Microsoft Research, Redmond, USA** (June 2014).
 - **IBM Research, Almaden, USA** (June 2014).
 - **Yahoo! Research, Barcelona, Spain** (June 2014).
 - **IBM Research, Bangalore, India** (January 2014).
 - **Systems Group, ETH Zurich, Switzerland** (October 2013).
 - **IIT Bombay, India** (August 2013).
 - **Alcatel-Lucent Lab, NJ, USA** (May 2013).
 - **Bosch Research, Palo Alto, USA** (February 2013).
 - **IIT Guwahati, India** (December 2011).

TEACHING

- **Instructor**, Intro. Databases, NTU Singapore
- **Instructor**, Network Science, NTU Singapore
- **Instructor**, Algorithms for Databases Seminar, ETH Zurich
- **Instructor**, Big-Graph Systems Seminar, ETH Zurich
- **Teaching Assistant**, Data Mining, UCSB
- **Teaching Assistant**, NP-Completeness, UCSB
- **Teaching Assistant**, Theory of Computation, UCSB
- **Teaching Assistant**, Introduction to Programming, UCSB

CURRENT AND FORMER STUDENTS SUPERVISED

• PhD Students

- Xiangyu Ke (NTU, paper in SIGMOD 2018, CIKM 2017, VLDB 2018, 2019) (SIGMOD 2018 Travel Award)
- Arkaprava Saha (NTU)
- Appan Rakaraddi (NTU)
- Tenindra Abeywickrama (Visiting PhD student from Monash University) (paper in TKDE 2019)

• Research Staffs

- Tianxing Wu (NTU, Postdoctoral Researcher)
- Aiqin Zhao (NTU, Postdoctoral Researcher)
- Xuan Liu (NTU, Research Associate)
- Vijaya Krishna Yalavarthi (NTU, Research Associate) (paper in CIKM 2017, VLDB 2018, IEEE BigData 2018)

• Master Thesis Supervisor

- Andreas Nufer (ETH, paper in TKDE 2018)
- Gustavo Segovia (ETH, paper in USENIX ATC 2018)
- Benjamin Zehnder (ETH, paper in ICDE 2016)
- Bojana Dimcheva (ETH, paper in SIGMOD 2014)

• Undergraduate Thesis Supervisor

- Michelle Teo Wan Teng (NTU, paper in VLDB 2018)
- Siyuan Liu (NTU, paper in IEEE BigData 2018)
- Leroy Lim Hong Quan (NTU, paper in VLDB 2019)
- Rojin Rezvan (Intern from Sharif University of Technology)
- Thomas Wohler (ETH, paper in CIKM 2015)

SERVICE

- Associate Editor for IEEE Transactions on Knowledge and Data Engineering (TKDE), 2019 - 2021 [Impact Factor=3.44].
- Proceedings Chair for EDBT 2020.
- Co-Chair for Big-O(Q) 2015 Workshop (Co-located with VLDB 2015).
- Senior Program Committee Member for TheWebConf (2018).
- Program Committee Member for SIGMOD (2020, 2018 Demo Track, 2016 Demo Track, 2014 Best Demo Selection Committee), VLDB (2020, 2019, 2016 PhD Workshop Track), ICDE (2018 Demo Track, 2017 Demo Track), KDD (2016, 2015), TheWebConf (2019), WWW (2017, 2016), ICDM (2018, 2016, 2015), CIKM (2019, 2018, 2017, 2016, 2015), EDBT (2018, 2017, 2015), DSAA (2019), DASFAA (2017), Australasian Database Conference (2017, 2016).
- Reviewer for ACM Transactions on Database Systems (TODS), IEEE Transactions on

Knowledge and Data Engineering (TKDE), VLDB Journal, ACM Transactions on Knowledge Discovery from Data (TKDD), Distributed and Parallel Databases (DAPD), ACM Transaction on Information Systems (TOIS), ACM Computing Surveys, Elsevier Journal of Parallel and Distributed Computing (JPDC), Knowledge and Information Systems (KAIS), Elsevier Journal of Knowledge Based Systems (KBS), Elsevier Journal of Information Systems (IS), Fundamenta Informaticae, International Conference on Information Systems (ICIS).

- External Reviewer for ICDT 2017, SIGMOD 2015 (Industrial Track), ICDCS 2014, IEEE Transactions on Neural Networks, VLDB 2014, MLG 2013, SIGMOD 2013, ICDE 2013, WWW 2013, VLDB 2013, ICDM 2012, SIGMOD 2011, SDM 2011, SIGKDD 2010, SDM 2010, SIGMOD 2010, ICDM 2010, ICDM 2009, ICDE 2009.
- Conference Session Chair for CIKM 2017, COMAD 2016, VLDB 2015.
- Industry Panel Chair for GFMI Conference Optimizing Data Governance, Quality and Consistency in Financial Services, Singapore 2019.