# Akhil Arora

Contact Information	Doctoral Student, DLab EPFL Lausanne, Switzerland	★ +41-779876871 (Switzerland) ⊠ akhil.arora@epfl.ch, arora.akhilcs@gmail.com WWW: http://www.cse.iitk.ac.in/users/aarora	
Research Interests	Large Scale Data Mining: Graph Mining, Social Network Analysis; Databases: Indexing & Querying Large Graphs, Text and High Dimensional Databases; Machine Learning		
Education	PhD in Computer Science EPFL, Lausanne, Switzerland Advisor: Prof. Robert West	September 2018 – Present	
	Masters in Computer Science Indian Institute of Technology (IIT), Kanpur, Ind Advisor: Prof. Arnab Bhattacharya	CGPA : 8.67/10.0 (Rank: 3/39) july 2011 – May 2013	
	Bachelor of Engineering in Computer Scient The NorthCap University, Gurgaon, India	ce July 2006 – May 2010	
Selected Publications	• WWW 2019: Jithin Vachery, Akhil Arora, Sayan Ranu, Arnab Bhattacharya: RAQ: Relationship- Aware Graph Querying in Large Networks, In: Proc. of The Web Conference, 2019		
	• EDBT 2019: Akhil Arora, Sainyam Galhotra, Sayan Ranu: Influence Maximization Revisited: The State of the Art and the Gaps that Remain, In: Proc. of Extending Database Technology Conference, 2019 (Tutorial)		
	• VLDB 2018: Akhil Arora, Sakshi Sinha <sup>*</sup> , Piyush Kumar <sup>*</sup> , Arnab Bhattacharya: <i>HD-Index:</i> <i>Pushing the Scalability-Accuracy Boundary for Approximate kNN Search in High-Dimensional</i> <i>Spaces</i> , In: Proc. of International Conference on Very Large Databases, 2018		
	• SIGMOD 2017: Akhil Arora <sup>*</sup> , Sainyam Galhotra <sup>*</sup> , Sayan Ranu: Debunking the Myths of Influence Maximization, In: Proc. of ACM International Conference on Management of Data, 2017 [ACM SIGMOD Most Reproducible Paper Award]		
	• SIGMOD 2016: Akhil Arora <sup>*</sup> , Sainyam Galhotra <sup>*</sup> , Shourya Roy: <i>Holistic Influence Maxi-</i> mization: Combining Scalability and Efficiency with Opinion-Aware Models, In: Proc. of ACM International Conference on Management of Data, 2016		
	• VLDB 2016: Satyajit Bhadange, Akhil Arora <sup>†</sup> , Arnab Bhattacharya: <i>GARUDA: A System</i> for Large-Scale Mining of Statistically Significant Connected Subgraphs, In: Proc. of Interna- tional Conference on Very Large Databases, 2016 (Demonstrations Track)		
	• WWW 2015: Sainyam Galhotra <sup>†</sup> , Akhil Arora <sup>†</sup> , Srinivas Virinchi, Shourya Roy: ASIM: A Scalable Algorithm for Influence Maximization under the Independent Cascade Model, In: Proc. of ACM International Conference on World Wide Web, 2015 (Poster: Companion Volume)		

• SIGMOD 2014: Akhil Arora, Mayank Sachan, Arnab Bhattacharya: *Mining Statistically Significant Connected Subgraphs in Vertex Labeled Graphs*, In: Proc. of ACM International Conference on Management of Data, 2014

<sup>\*</sup>Equal Contribution

 $<sup>^{\</sup>dagger} {\rm Corresponding} ~{\rm Author}$ 

Professional ACTIVITIES

- Co-Chair GRADES-NDA Workshop (2016-2019) (Co-located with SIGMOD 2016-19); Research Programming Challenge at COMAD 2014.
- Program Committee Member EDBT 2020, DASFAA (2017-19), SIGMOD 2018 (Demonstrations), ISWC 2018-19 (Posters & Demonstrations), GDAM 2015, XRCI Open 2015.
- Reviewer SIGMOD 2019, VLDB (2019, 2018, 2016), KDD (2015-19), WWW (2019, 2017), ICDE 2018, SDM (2015-16), CIKM (2018, 2015, 2014), CoDS/COMAD (2013-16), TKDD, TKDE, ToN, Journal of Data Science and Analytics, Journal of Information Science.
- Co-founder and organizer of Special Interest Group in Data (SIGDATA) at IIT Kanpur, which meets weekly to discuss relevant advances in **Databases** and **Data Mining**

### WORK EXPERIENCE Research Scientist, American Express Big Data Labs

Member of Big Data Labs

July 2017 - July 2018 Devised scalable algorithms to solve a host of complex real-world problems in the area of semisupervised and unsupervised deep learning using structured and unstructured data.

### Researcher, Xerox Research Centre India (XRCI)

Member of Text and Graph Analytics Group July 2014 - July 2017 Led multiple research projects on devising scalable algorithms to solve a gamut of complex realworld problems in the area of databases, data mining and machine learning. Work done here led to multiple publications in **SIGMOD**, **VLDB** and **WWW**.

### Software Engineer, Intel Corporation, Bangalore, India

Member of Security and Vulnerability Hacking Group July 2013 - July 2014 Worked on research problems in security while performing white hat hacking on internal Intel products, security code reviews, assessments, and code assisted penetration. Developed a framework which was published in **Black Hat** 2014.

HONORS AND AWARDS

• Travel Award to attend EDBT/ICDT 2019.

- Awarded the EDIC **Doctoral Fellowship** by EPFL for the year 2018-2019.
- Won the First Prize in the Adobe Data Mining Competition, 2013 at IIT, Madras.
- Won the Fifth Prize in Scalable String Similarity Search/Join workshop: EDBT, 2013.
- Awarded the Overall Best Hack prize in Yahoo! HackU!, 2012 at IIT, Kanpur.
- Won the Second Prize in the 10th ImageCLEF: Plant Identification Task, 2012.
- Among 7 researchers worldwide to get ELIAS Sponsorship for attending CLEF, 2012.
- Best Project Award for HiPhi: Approximate kNN Search in High-Dimensional Spaces.
- Best Project Award for FriendMiner: Inferring Relationship Based on Mobile Phone Data.
- Ranked Third in the department, among the entrant batch of M.Tech and Phd programme.
- All India Rank(AIR) 369, GATE 2011 (percentile 99.73), category: Computer Science.
- State level 'Science Award' for securing highest marks in Mathematics in Secondary Examination (ICSE), 2004.
- Uttar Pradesh State Merit Scholarship for Academic Excellence in Secondary Examination (ICSE), 2004.

INVITED TALKS

• Debunking the Myths of Influence Maximization

• University of Michigan, Ann Arbor, USA

- May, 2017
- Indian Institute of Technology (IIT), Gandhinagar, India Feb. 2017 • Indian Institute of Technology (IIT), Kanpur, India Feb, 2017
- GARUDA: A System for Large-Scale Mining of Statistically Significant Connected Subgraphs • 22<sup>nd</sup> Int. Conf. on Management of Data (Premier Papers Track) March, 2017
- Holistic Influence Maximization: Scalability and Efficiency with Opinion-Aware Models • 22<sup>nd</sup> Int. Conf. on Management of Data (Premier Papers Track) March, 2017
- Indian Institute of Technology (IIT), Delhi, India September, 2016
  - University of California, Santa Barbara, USA June, 2016
  - Facebook Inc., Menlo Park, USA June, 2016

	<ul> <li>Palo Alto Research Centre (PARC), USA</li> <li>Indian Institute of Technology (IIT), Kanpur, India</li> <li>ASIM: A Scalable Algorithm for Influence Maximization under the Independent Cascade Model</li> <li>Botathon, Forge Accelerator</li> <li>The Northcap University (NCU), Gurgaon, India</li> <li>Indian Institute of Technology (IIT), Kanpur, India</li> <li>Indian Institute of Technology (IIT), Kanpur, India</li> <li>September, 2016</li> <li>Indian Institute of Technology (IIT), Kanpur, India</li> <li>Mining Statistically Significant Connected Subgraphs in Vertex Labeled Graphs</li> <li>Palo Alto Research Centre (PARC), USA</li> <li>20<sup>th</sup> Int. Conf. on Management of Data (Premier Papers Track)</li> <li>Xerox Research Centre Europe, Grenoble, France</li> </ul>	
Masters Thesis	Mining Statistically Significant SubgraphsSpring 2012 – Spring 2013Worked with Prof. Arnab Bhattacharya to develop the first ever scalable algorithm to mine statistically significant subgraphs from a single large graph, with a wide-variety of applications ranging from Community Detection to Mining Spatial Colocations, Hotspot Detection and many more. Instead of the more commonly used frequency, we use the more involved p-value/chi-square statistic as an objective function to mine interesting patterns that deviate significantly from the expected. This work was published as a paper in the research track at SIGMOD 2014.	
Patent Applications and Disclosures	<ul> <li>Akhil Arora, Manoj Gupta, Neeta Pande, Sainyam Galhotra, Shourya Roy: System for Identifying Root Causes of Churn for Churn Prediction Refinement, USPTO Application Number: 15/132,767, Filed: 2016.</li> <li>Akhil Arora, Manoj Gupta, Shourya Roy: Transforming a Knowledge Base into a Machine Readable Format for an Automated System, USPTO Application Number: 14/887,096, Filed: 2015, Granted: 2018.</li> <li>Akhil Arora, Sainyam Galhotra, Srinivas Virinchi, Shourya Roy: Methods and Systems for Identifying Target Users of Content, USPTO Application Number: 14/628,070, Filed: 2015.</li> </ul>	
OTHER PUBLICATIONS	<ul> <li>Akhil Arora<sup>‡</sup>, Sainyam Galhotra<sup>‡</sup>, Sayan Ranu: Influence Maximization Revisited: The State of the Art and the Gaps that Remain, In: Proc. of ACM Joint International Conference on Data Science and Management of Data CoDS-COMAD (Tutorial), 2018</li> <li>Akhil Arora<sup>‡</sup>, Sainyam Galhotra<sup>‡</sup>, Sayan Ranu: Debunking the Myths of Influence Maximization, In: NEDB North East Database Day (Oral), 2017</li> <li>Sainyam Galhotra<sup>‡</sup>, Akhil Arora<sup>‡</sup>, Shourya Roy: Holistic Influence Maximization: Combining Scalability and Efficiency with Opinion-Aware Models, In: NEDB North East Database Day (Poster), 2016</li> <li>Deepali Semwal, Sonal Patil, Sainyam Galhotra, Akhil Arora, Narayanan Unny: STAR: Realtime Spatio-Temporal Analysis and Prediction of Traffic Insights using Social Media, In: ACM IKDD Conference on Data Sciences (CoDS), 2015</li> <li>Akhil Arora, Sumanth Naropanth: Android Kernel and OS Security Assessment with Iron Crow, Black Hat Europe, 2014</li> <li>Shashwat Mishra, Tejas Gandhi, Akhil Arora, Arnab Bhattacharya: Efficient Edit Distance based String Similarity Search using Deletion Neighborhoods, In: Proceedings of the Joint EDBT /ICDT 2013 Workshops</li> <li>Akhil Arora, Ankit Gupta, Nitesh Bagmar, Shashwat Mishra, Arnab Bhattacharya: A Plant Identification System using Shape and Morphological Features on Segmented Leaflets: Team IITK, CLEF 2012 In: CLEF 2012 (Online Working Notes/ Labs/Workshop)</li> </ul>	
Manuscripts under Prepara- tion/Submission	• Akhil Arora, Sakshi Sinha, Piyush Kumar, Arnab Bhattacharya: Approximate kNN Search in High-Dimensional Spaces Revisited: Myths, Mis-Claims and Beyond.	

 $^{\ddagger}\mathrm{Equal}$  Contribution

Teaching and Mentoring Experience	<ul> <li>Teaching Assistant, Introduction to Databases, EPFL</li> <li>Teaching Assistant, Database Management Systems, IIT Kanpur</li> <li>Teaching Assistant, Data Mining, IIT Kanpur</li> <li>Teaching Assistant, Introduction to Computing, IIT Kanpur</li> </ul>	Spring 2019 Spring 2013 Fall 2012 Fall 2011, Spring 2012
	<ul> <li>Masters Thesis Co-Supervisor (with Prof. Arnab Bhattach co-supervised students and masters thesis as follows         <ul> <li>Piyush Kumar, Indexing High-Dimensional Databases</li> <li>Sakshi Sinha, Indexing High-Dimensional Databases</li> <li>Satyajit Bhadange, Distributed Significant Subgraph Mining</li> </ul> </li> <li>Mentor (with Prof. Sayan Ranu), IIT Madras mentored PhD thesis as follows         <ul> <li>Jithin Vachery, PhD student (IIT Madras) working on novel research questions in (sub-)graph mining and set</li> </ul> </li> </ul>	arya), IIT Kanpur Spring 2015 – Spring 2016 Spring 2015 – Spring 2016 Spring 2015 – Fall 2015 Aug. 2015 – Present
	<ul> <li>Mentor, Xerox Research Centre India (XRCI) mentored intern projects as follows         <ul> <li>Prajna Upadhyay, PhD student (IIT Delhi) designed a system for automatic knowledge extraction and organi</li> <li>Srinivas Virinchi, PhD student (UMD, College Park)</li> </ul> </li> </ul>	Aug. 2015 – Dec. 2015 zation from enterprise data. Aug. 2014 – Dec. 2014

developed a scalable algorithm for influence maximization in large networks.

#### Prof. Robert West References

		I ION MILLION DIRUUG
	Assistant Professor	Associate Professor
	School of Computer and Communication	Department of Compu
	Sciences	IIT Kanpur, India
	EPFL, Switzerland	$\boxtimes$ arnabb@cse.iitk.ac.
	$\boxtimes$ robert.west@epfl.ch	$\boxtimes$ arnabb@iitk.ac.in

# Prof. Sayan Ranu

Assistant Professor Department of Computer Science IIT Delhi, India  $\boxtimes$  sayanranu@cse.iitd.ac.in  $\boxtimes$ sayan<br/>ranu@gmail.com

# Dr. Manish Gupta

CEO & Co-Founder / Professor VideoKen / IIIT Bangalore Bangalore, India  $\boxtimes$  manish.gupta01@gmail.com

# Prof. Arnab Bhattacharya

uter Science .in

## Dr. Shourya Roy

Vice President American Express Big Data Labs Bangalore, India ⊠ shourya.roy@gmail.com