Sainyam Galhotra

✉ sainyam@cs.umass.edu
people.cs.umass.edu/~sainyam/

Research Interests

Large-scale Data Integration; Fairness and Transparency in Data Management

Education

2017-present **PhD in Computer Science**, University of Massachusetts Amherst, USA
Advisor: Prof. Barna Saha

2015-2017 **Masters in Computer Science**, University of Massachusetts Amherst, USA
CGPA: 4.0/4.0

2010-2014 **BTech in Computer Science and Engineering**, Indian Institute of Technology (IIT) Delhi, India
Advisor: Prof. Amitabha Bagchi

Selected Awards and Honors

2020 ACM SIGMOD Entity Resolution Programming Contest – Top 5 finalist
2018 PhD Candidate with Distinction
2018 Finalist for Adobe Fellowship
2017-18 Most reproducible paper award in SIGMOD 2017 and 2018
2017 Best paper award in SIGSOFT FSE 2017
2016 First recipient of Krithi Ramamritham Computer Science Scholarship
2016-18 SIGMOD Travel Award for three consecutive years
2014-20 Travel Grants Awarded: CIKM 2020, VLDB 2020, EDBT 2019, AAAI 2018, UAI 2014
2010 All India Rank 53 in IIT-JEE (of $\sim$ 450K candidates)
2010 State Rank 1 and All India Rank 20 in AIEEE (of $\sim$ 1M candidates)

Work Experience

Aug-Nov 2019 **Applied Scientist Intern**, Amazon, Seattle, USA
Mentor: Dr. Xin Luna Dong
Developed a novel deep-learning based entity linking approach that leveraged Amazon Knowledge Graph to identify product attributes from search queries.

May-Aug 2019 **Research Intern**, IBM T.J. Watson Research Center, Yorktown Heights, USA
Mentor: Dr. Udayan Khurana
Demonstrated use of external knowledge for feature enrichment with AutoML. Demo paper of this work was presented at CIKM 2020 and ICDM 2019.

May-Aug 2018 **Research Intern**, Megagon Labs, Mountain View, USA
Mentors: Dr. Wang-Chiew Tan & Dr. Behzad Golshan
Designed a scalable system to generate labelled training data with minimum human intervention. Demonstrated superior performance over a variety of NLP tasks. This work is under review at SIGMOD 2021.

May-Aug 2016 **Research Intern**, Google Research, Mountain View, USA
Mentor: Dr. Steven Whang
Developed algorithms to recommend feature engineering services for billions of internal datasets

May 2014-2015 **Budding Scientist**, Xerox Research Center India (XRCI), Bangalore India
Member of Text and Graph Analytics team

Worked on data mining algorithms to reduce customer churn and validated it on various telecommunications providers.

May-Aug 2013 **Research Intern**, Yahoo! Labs, Bangalore, India

Mentor: Dr. Vidit Jain

Designed trend prediction algorithm and validated on real-time Twitter data. This work has been published at UAI – 2014.

May-Aug 2012 **Research Intern**, École Nationale Supérieure des Télécommunications, Brest, France

Mentor: Dr. Fabien Dagnat

Extended the Continuations package of Java Virtual Machine (JVM) to support dynamic software updates.

Patents

[4] Udayan Khurana, **Sainyam Galhotra**, Oktie Hassanzadeh, Kavitha Srinivas, Horst Samulowitz. Knowledge Aided Feature Engineering. Filed: 2019.

[3] **Sainyam Galhotra** and Narayanan Unny. Method and system to predict a communication channel for communication with a customer service. USPTO Application Number: 15/077,085, Filed: 2016.

[2] 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

[1] 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.

Refereed Conference and Journal Publications

Articles Under Review

[30] **Sainyam Galhotra**, Donatella Firmani, Barna Saha and Divesh Srivastava. Hierarchical Entity Resolution using an Oracle. under review, 2021.

[29] **Sainyam Galhotra**, Anna Fariha, Raoni Lourenco, Juliana Freire, Alexandra Meliou and Divesh Srivastava. DataExposer: Exposing Disconnect between Data and Systems. under review, 2021.

[28] **Sainyam Galhotra**, Karthikeyan Shanmugam, Prasanna Sattigeri and Kush R. Varshney. Fair Data Integration. arXiv, 2020.

[27] Saba Ahmadi*, **Sainyam Galhotra***, Barna Saha and Roy Schwartz. Fair Correlation Clustering. arXiv, 2020.

Articles published at peer-reviewed conferences

[26] **Sainyam Galhotra**, Romila Pradhan and Babak Salimi. Explaining Black-Box Algorithms Using Probabilistic Contrastive Counterfactuals. In Proc. of ACM International Conference on Management of Data (**SIGMOD**), 2021.

[25] **Sainyam Galhotra**, Donatella Firmani, Barna Saha and Divesh Srivastava. BEER: Blocking for Effective Entity Resolution. In Proc. of ACM International Conference on Management of Data (**SIGMOD**), Demonstrations Track, 2021.

[24] **Sainyam Galhotra**, Behzad Golshan and Wang-Chiew Tan. Adaptive Rule Discovery for Labeling Text Data. In Proc. of ACM International Conference on Management of Data (**SIGMOD**), Data Science and Engineering Track, 2021.

[23] **Sainyam Galhtora** and Udayan Khurana. Semantic Search over structured data. In Proc. of Conference on Information and Knowledge Management **(CIKM)**, Demonstrations Track, 2020. **Nomination: Best Demo Paper**

[22] Sandhya Saisubramanian*, **Sainyam Galhotra*** and Shlomo Zilberstein. Balancing the Tradeoff Between Clustering Value and Interpretability. In Proc. of 3rd ACM Conference on AI, Ethics, and Society **(AIES)**, 2020.

[21] **Sainyam Galhotra**, Udayan Khurana, Oktie Hassanzadeh, Kavitha Srinivas, Horst Samulowitz, Miao Qi. Automated Feature Enhancement for Predictive Modeling using External Knowledge. In Proc. of International Conference on Data Mining **(ICDM)**, Demonstrations Track, 2019.

[20] **Sainyam Galhotra**, Soumyabrata Pal, Arya Mazumdar and Barna Saha. Connectivity in Random Annulus Graphs and the Geometric Block Model. In Proc. of International Conference on Randomization and Computation **(APPROX-RANDOM)**, 2019.

[19] **Sainyam Galhotra**, Donatella Firmani, Barna Saha and Divesh Srivastava. Robust Entity Resolution using Random Graphs. In Proc. of ACM International Conference on Management of Data **(SIGMOD)**, pages 3-18, 2018.
Most Reproducible Paper Award

[18] **Sainyam Galhotra**, Arya Mazumdar, Soumyabrata Pal and Barna Saha. The Geometric Block Model. In Proc. of the 32nd AAAI Conference on Artificial Intelligence **(AAAI)**, 2018.

[17] **Sainyam Galhotra**, Arya Mazumdar, Soumyabrata Pal and Barna Saha. The Geometric Block Model and Applications. In Proc. of Allerton Conference on Communication, Control, and Computing (Invited papers track), pages 1147-1150, 2018.

[16] **Sainyam Galhotra**, Yuriy Brun and Alexandra Meliou. Fairness Testing: Testing Software for Discrimination. In Proc. of ACM SIGSOFT Foundations of Software Engineering **(ESEC/SIGSOFT FSE)**, pages 498-510, 2017.
Best Paper Award

[15] Akhil Arora*, **Sainyam Galhotra*** and Sayan Ranu. Debunking the Myths of Influence Maximization. In Proc. of ACM International Conference on Management of Data **(SIGMOD)**, 2017.
Most Reproducible Paper Award

[14] **Sainyam Galhotra***, Akhil Arora* and Shourya Roy. Holistic IM: Combining Scalability and Efficiency with Opinion-Aware Models. In: Proc. of ACM International Conference on Management of Data **(SIGMOD)**, 2016

[13] Shourya Roy, Sandipan Dandapat, R. Mariappan, S Srivastava, **Sainyam Galhotra** and B. Peddamuthu. QA^{RT} : A System for Real-time Holistic Quality Assurance for Contact Center Dialogues. In Proc. of 30th AAAI Conference on Artificial Intelligence **(AAAI)**, 2016.

[12] **Sainyam Galhotra**, Amitabha Bagchi, Srikanta Bedathur, Maya Ramanath and Vidit Jain. Tracking the Conductance of Rapidly Evolving Topic-Subgraphs. In Proc. of International Conference on Very Large Databases **(VLDB)**, 2015.

[11] **Sainyam Galhotra**, Akhil Arora, Srinivas Virinchi and Shourya Roy. ASIM: A Scalable Algorithm for Influence Maximization under the Independent Cascade Model. In Proc. of ACM International Conference on World Wide Web **(WWW)**, 2015 (Poster: Companion Volume)

[10] STAR: Real-time Spatio-Temporal Analysis and Prediction of Traffic Insights using Social Media. In Proc. of ACM Joint International Conference on Data Science and Management of Data **(CoDS-COMAD)**, 2015.

[9] Vidit Jain and **Sainyam Galhotra**. Min-d-Occur: Ensuring Future Occurrences in Streaming Sets. In Proc. of the Conference on Uncertainty in Artificial Intelligence **(UAI)**, 2014.

[8] Amitabha Bagchi, Cristina M. Pinotti, **Sainyam Galhotra** and Tarun Mangla. Optimal Radius for Connectivity in Duty-Cycled Wireless Sensor Networks. In Proc. of ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems **(MSWiM)**, 2013.

Articles in peer-reviewed journals

[7] **Sainyam Galhotra**, Donatella Firmani, Barna Saha and Divesh Srivastava. Efficient and Effective ER with Progressive Blocking. VLDB Journal, 2021.

[6] Donatella Firmani, **Sainyam Galhotra**, Barna Saha, Divesh Srivastava. Robust Entity Resolution Using a CrowdOracle. IEEE Data Engineering Bulletin, 2018.

[5] Amitabha Bagchi, Cristina M. Pinotti, **Sainyam Galhotra** and Tarun Mangla. Optimal Radius for Connectivity in Duty-Cycled Wireless Sensor Networks, In Transactions of Sensor Networks **(TOSN)**, 2015.

[4] Fuhuo Li, **Sainyam Galhotra** and Shigeru Kanemitsu. Emerging Importance of EVs in the Green Grid Era. In Pure and Applied Mathematics Journal Special issue: Mathematical aspects of engineering disciplines, 2015.

[3] **Sainyam Galhotra**, Shigeru Kanemitsu, Hiroyuki Kondo. An adaptation method for removing arsenate species from water solution. In Pure and Applied Mathematics Journal, 2015.

[2] **Sainyam Galhotra**, J. K. Bhattacharjee and Bijay K. Agarwalla. Turing - Hopf instabilities through a combination of diffusion, advection and finite size effects. In Journal of Chemical Physics, 2014.

[1] Bijay K. Agarwalla, **Sainyam Galhotra** and J. K. Bhattacharjee. Diffusion driven instability to a drift driven one: Turing patterns in the presence of an electric field. In Journal of Mathematical Chemistry, 2014.

Tutorials

[3] Akhil Arora*, **Sainyam Galhotra*** and Sayan Ranu. Navigating the Maze of Influence Maximization Algorithms. IEEE DSAA (Tutorial), 2019.

[2] Akhil Arora*, **Sainyam Galhotra*** and Sayan Ranu. Influence Maximization Revisited: The State of the Art and the Gaps that Remain. In Proc. of EDBT (Tutorial), 2019.

[1] Akhil Arora*, **Sainyam Galhotra*** and 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.

Workshops and Symposiums

[7] **Sainyam Galhotra**. Reliable Clustering with Applications to Data Integration. In PhD Workshop at VLDB, 2020.

[6] **Sainyam Galhotra**, Udayan Khurana, Oktie Hassanzadeh, Kavitha Srinivas and Horst Samulowitz. KAFE: Automated Feature Enhancement for Predictive Modeling using External Knowledge. In NeurIPS 2019 Workshop: Knowledge Representation & Reasoning Meets Machine Learning, 2019.

[5] **Sainyam Galhotra**, Donatella Firmani, Barna Saha, Divesh Srivastava. Crowd-Sourced Entity Resolution with Control Queries. In SEBD, 2019.

[4] **Sainyam Galhotra**, Arya Mazumdar, Soumyabrata Pal and Barna Saha. The Geometric Block Model. In NeurIPS 2017 Workshop on Learning on Distributions, Functions, Graphs and Groups, 2017.

[3] **Sainyam Galhotra***, Akhil Arora* and Sayan Ranu. Debunking the Myths of Influence Maximization, In North East Database Day (NEDB), 2017 (Oral).

[2] **Sainyam Galhotra***, Akhil Arora* and Shourya Roy. Holistic IM: Combining Scalability and Efficiency with Opinion-Aware Models. In NEDB (Poster) 2016.

[1] **Sainyam Galhotra** and Shigeru Kanemitsu. and Shigeru Kanemitsu. An adaptation method for removing arsenate species from water solution. SUDA, 2013.

Teaching Experience

2016-2018 **Teaching Assistant:** Algorithms for Data Science (CS590D), CICS, UMass Amherst

- Spring 2016
- Fall 2016
- Fall 2017
- Spring 2019

2015 **Grader:** Introduction to Computer Architecture (CS335), CICS, UMass Amherst (Fall 2015)

Professional Activities

Mentor

- Hui Wei (PhD student, UMass Amherst, 2020)
- Sidong Zhang (PhD student, UMass Amherst, 2020)
- Nicolas Van Kempen (PhD student, UMass Amherst, 2020)
- Ankita Mehta (Masters student, UMass Amherst, 2016)
- Divyesh Harit (Masters student, UMass Amherst, 2016)
- Tanvi Sahay (Masters student, UMass Amherst, 2016)
- Deepak Bhatt (Undergraduate thesis, IIT Delhi, 2016)
- Srinivas Virinchi, PhD student, UMD, College Park (2015)

Senior PC Member: IJCAI (2021)

PC Member: VLDB (2022), UAI (2021), AAAI (2020, 2021), FAcct (2021), GRADES-NDA Workshop SIGMOD (2018, 2019, 2020), SIGMOD Reproducibility 2018, ISWC Demo (2018, 2019, 2020), CoDS-COMAD (2021, 2020)

Reviewer: AISTATS (2021), ICDE (2021) NeurIPS (2020, 2019, 2018, 2016), ICML (2020,2019), NORDICHI 2020, CIKM (2020, 2019, 2018, 2017), KDD (2018, 2017), SIGMOD 2018, TKDE, SEA 2017, WWW 2017, CoDS 2015

2019-Present **Feature Editor** for the ACM XRDS Magazine.

2018-2019 **Graduate Student Representative** for all Computer Science graduate students, UMass Amherst.

2015 Organizing Committee member, XRCI Open 2015, Bangalore, India

2017-18 **Vice President**, Indian Student Association, UMass Amherst

2014-15 **Overall Coordinator, XRCI-CSR (Corporate Social Responsibility)**, Initiated the CSR group at Xerox Research Centre India and led the team for social activities like cleanliness campaign, distributing food to the needy

Invited Talks

Oct 2019 Robust Entity Resolution Using an Oracle, UC Berkeley

Sept 2018 Local Correlation Clustering, UMass Theory seminar

June 2018 Repairing Noisy Graphs, Megagon Labs

Jan 2018 Influence Maximization Revisited: The Gaps that Remain, CODS/COMAD 2018

Jan 2017 Debunking the Myths of Influence Maximization, NEDB

Sept 2016 Tracking the Conductance of Rapidly Evolving Topic-Subgraphs, IIT Delhi

July 2016 Tracking the Conductance of Rapidly Evolving Topic-Subgraphs, NDA Workshop

June 2016 Holistic IM: Combining Scalability and Efficiency with Opinion-Aware Models, UCSB, Facebook Inc, PARC

May 2015 ASIM: A Scalable Algorithm for Influence Maximization under the Independent Cascade Model, Univ. of Perugia Italy

Selected Media Coverage

1. UMass Amherst Computer Scientists Develop New Technique to Measure Social Bias in Software, ACM News and MIT Technology Download
2. Unlearning Racism and Sexism in Learning Machines, in EnterpriseTech
3. Look Who's Fighting Our Algorithmic Overlords, in Bloomberg
4. Uncovering discrimination in machine-learning software, in GCN
5. This Breakthrough Tool Detects Racism And Sexism In Software, in Co.Design