

# Barna Saha

## CONTACT INFORMATION

School of Computer Science  
140 Governors Drive, University of Massachusetts Amherst,  
MA 01003, USA

Phone: (413) 577-2510  
barna@cs.umass.edu  
<http://barnasaha.net>

## RESEARCH INTERESTS

*Algorithm Design and Analysis, Large Scale Data Analytics, Randomization in Computation.*

My primary research interests span *design and analysis of algorithms, discrete optimization, and foundational aspects of data analytics involving large-scale data mining and learning problems*. From solving core theoretical problems, to interdisciplinary research with collaborations across computer science, and other disciplines, my niche remains in developing theory that addresses fundamental problems, and can be applied, implemented and leave broader impact.

See <https://people.cs.umass.edu/~barna/projects.html> for more details.

## POSITIONS

<b>Assistant Professor</b> , University of Massachusetts Amherst College of Information & Computer Science	Sept 2014–Present
<b>Visiting Scholar</b> , The Simons Institute, University of California Berkeley	August 2015–Dec 2015
<b>Senior Member of Technical Staff-Research</b> , AT&T Labs-Research	Aug, 2011–July 2014
<b>Visiting Faculty</b> , University of Minnesota Twin Cities Department of Computer Science	Sept, 2013–Dec 2013
<b>Industrial Research Internships</b> IBM T.J. Watson Research Center, AT&T Labs-Research, IBM Almaden Research Center,	Jun, 2010–Aug, 2010 Jun, 2009–Aug, 2009 Jun, 2008–Aug, 2008

## EDUCATION

**University of Maryland, College Park (UMD)**  
Ph.D., Computer Science [Aug, 2007–Aug, 2011]

**Indian Institute of Technology (IIT), Kanpur**  
M.Tech, Computer Science & Engineering

**Jadavpur University**  
B.E. Computer Science & Engineering

## HONORS AND AWARDS

**Google Faculty Research Award**, 2016

**Yahoo! Academic Career Enhancement Award**, 2015  
(Yearly given to top-5 young faculty doing Yahoo relevant research)

**Simons Research Fellowship**, 2015, The Simons Institute, University of California Berkeley.

**NSF CISE Research Initiation Initiative (CRII) Award**, 2015.

**Best Paper Award** at 35th International Conference on Very Large Data Bases (VLDB), 2009.

**Finalist for Best Paper Award** at 30th International Conference on Data Engineering (ICDE), 2012.

**UMD Dean's Dissertation Award**, for excellence in dissertation research, 2011.

**UMD Graduate Fellowship award**, 2007-2009.

**Travel Award**, Foundation of Computer Science (FOCS), 2010, Innovations in Theoretical Computer Science (ITCS), 2010, ACM-Siam Symposium of Discrete Algorithms (SODA), 2011.

**UMD Goldhaber Travel Award**, 2009.

**Best Presentation Award** at IRISS (Inter Research Institute Student Seminar), co-event conducted along with IJCAI 2007.

**IIT Kanpur, Academic Excellence Award**, for 10/10 GPA, 2006.

**Ministry of Human Resource and Development Scholarship**, India, 2004–06.

**Jagadish Bose National Science Talent Search (JBNSTS) Scholarship**, 2000.

**State Government Award** for rank 2nd out of 400,000 examinee at +12 examination

**State Government Award** for rank 8th out of 600,000 examinee at +10 examination.

## GRANTS

1. “Mining Structures in Massive Networks”  
Google Faculty Research Award, Single PI  
\$53,058 04/01/2016–.
2. “Scaling up Dynamic Programming for Certain Optimization Problems”,  
NSF/CISE/CCF, Primary Investigator (Single PI)  
\$174,937 (100% of total) 04/01/2015 – 03/31/2017.
3. Yahoo Academic Career Enhancement Award  
\$10,000 11/01/2015–.
4. “New Generation Scheduling and Resource Allocation Algorithms  
NSF/CISE/CCF, Primary Investigator  
\$483,301 (my share), Total ~ \$950,000 (Under Review).
5. “Algorithmic Formal Language Theory for Mining and Correcting Data”, Selected among top 5% (23 out of ~500) proposals for Bell Labs Innovation Award Competition, 2014.

## NOTABLE CONTRIBUTIONS

- Improving on the complexity of *RNA folding*, considered as one of the most basic problems in Bioinformatics after more than four decades.
- Reviving the study of *Error-correcting Parsers* aka *Language Edit Distance* and unfolding its connections to many problems in Theoretical Computer Science, Data Mining, and Computational Biology.
- Developing new constructive aspects of an important probabilistic method called the the Lovász Local Lemma. This helped to resolve an outstanding open question in combinatorial optimization known as the *Santa Claus problem*.
- Providing a unified framework for probabilistic ranking that won the best paper award at the Very Large Data Bases Conference (VLDB), 2009.

## PUBLICATIONS

By TCS tradition, author names are in alphabetical order of surnames by default. In TCS, conferences are the main publication venue, and are highly competitive.

### Working Papers

- I. Karl Bringman, Fabrizio Grandoni, Barna Saha, Virginia Vassilevska Williams, “RNA Folding & Language Edit Distance: Breaking the Cubic Barrier”, to be submitted to *FOCS 2016*. (This paper improves several foundational problems in computational biology and programming language after 40 years.)
- II. Manish Purohit, Barna Saha, “A Framework for Network-Aware Scheduling Problem”, to be submitted to *FOCS 2016*. (This paper combines scheduling on parallel machine with routing constraints on networks in a novel way.)
- III. Several other papers on graphs, crowdsourcing, scheduling etc.

### Conference Papers.

1. Donatella Firmani, Barna Saha, Divesh Srivastava, "Online Entity Resolution with an Oracle", Accepted *42nd International Conference on Very Large Data Bases (VLDB)*, 2016.
2. Barna Saha, "Language Edit Distance & Maximum Likelihood Parsing of Stochastic Grammars: Faster Algorithms & Connection to Fundamental Graph Problems", Proc., *56th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2015.
3. Flip Korn, Barna Saha, Divesh Srivastava, Shanshan Ying, "TreeScope: Finding Structural Anomalies In Semi-Structured Data", Proc. *41st International Conference on Very Large Data Bases (VLDB)*, 2015. Demonstration paper. The demo site <http://db128gb-b.ddns.comp.nus.edu.sg/shanshan/treescope/>.
4. Lukasz Golab, Flip Korn, Feng Li, Barna Saha, Divesh Srivastava, "Size-constrained Weighted Set Cover", Proc. *31st IEEE International Conference on Data Engineering (ICDE)*, 2015.
5. Barna Saha, "The Dyck Language Edit Distance Problem in Near-linear Time", Proc. *55th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.
6. Mohammadtaghi Hajiaghayi, Wei Hu, Jian Li, Shi Li, Barna Saha, "A Constant Factor Approximation Algorithms for Fault Tolerant K-Median Problem", Proc. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.
7. Mohammadtaghi Hajiaghayi, Theodore Johnson, Reza Khani, Barna Saha, "Hierarchical Graph Partitioning", Proc. *26th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, 2014.
8. Howard Karloff, Lukasz Golab, Marios Hadjieleftheriou, Barna Saha, "Distributed Data Placement via Graph Partitioning", Proc. *26th International Conference on Scientific and Statistical Database Management (SSDBM)*, 2014.
9. Barna Saha, Divesh Srivastava, "Data Quality: The Other Face of Big Data", Tutorial, Proc. *30th IEEE International Conference on Data Engineering (ICDE)*, 2014.
10. Flip Korn, Barna Saha, Divesh Srivastava and Shanshan Ying, "On Repairing Structural Problems in Semi-structured Data", Proc. *39th International Conference on Very Large Data Bases (VLDB)*, 2013.
11. Xin Luna Dong, Barna Saha, Divesh Srivastava, "Less is More, Selecting Sources Wisely for Integration", Proc. *39th International Conference on Very Large Data Bases (VLDB)*, 2013.
12. Yury Polyanskiy, Arya Mazumdar, Barna Saha, "The Closest String Problem and the Chebyshev Radius", Proc. *63rd IEEE International Symposium on Information Theory (ISIT)*, 2013.
13. Barna Saha, "Renting a Cloud," Proc. *33rd Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, 2013.
14. Seungjoon Lee, Manish Purohit, Barna Saha "Firewall Placement on Cloud Data Centers", poster paper in Proc. *ACM Symposium on Cloud Computing (SOCC)*, 2013.
15. Barna Saha, Samir Khuller, "Set Cover Revisited: Hypergraph Cover with Hard Capacities," Proc. *39th International Colloquium on Automata, Languages, and Programming (ICALP)*, 2012.
16. Samir Khuller, Barna Saha, Kanthi Sarpatwar, "New Approximation Results for Resource Replication Problems," Proc. *The 15th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, 2012.
17. Lukasz Golab, Howard Karloff, Flip Korn, Barna Saha, Divesh Srivastava, "Discovering Conservation Rules", *28th IEEE International Conference on Data Engineering (ICDE)*, (**Among Finalists for BEST PAPER AWARD**), 2012.
18. Ravishankar Krishnaswamy, Vishwanath Nagarajan, Barna Saha "The Matroid Median Problem", Proc. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.  
Merged with a paper by Amit Kumar and Yogish Sabharwal who obtained similar results for the case of partition matroid.
19. Saeed Alaei, Mohammad Taghi Hajiaghayi, Vahid Liaghat, Dan Pei, Barna Saha, "AdCell-Ad Allocation in Cellular Networks," Proc. *19th European Symposium on Algorithms (ESA)* , 2011.
20. Nikhil Bansal, Ravishankar Krishnaswamy, Barna Saha, "On Capacitated Set Cover Problem," Proc. *The 14th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, 2011.

21. Philip Anderson, Samir Khuller, Saket Navlakha, Louiqa Raschid, Barna Saha, Andreas Thor, Xiao-Ning Zhang, "Link Prediction for Annotation Graphs," Proc. *10th International Semantic Web Conference (ISWC)*, 2011.
22. Bernhard Haeupler, Barna Saha, Aravind Srinivasan, "New Constructive Aspects of the Lovász Local Lemma," *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010.
23. Samir Khuller, Jian Li, Barna Saha, "Energy Efficient Scheduling via Partial Shutdown," Proc. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.
24. Barna Saha, Aravind Srinivasan, "A New Approximation Technique for Resource-Allocation Problems," Proc. *Innovations in Theoretical Computer Science (ITCS)*, 2010.
25. Barna Saha, Allison Hoch, Samir Khuller, Louiqa Raschid and Xiao-Ning Zhang, "Dense Subgraphs with Restrictions and Applications to Gene Annotation Graphs," Proc. *14th International Conference on Research in Computational Molecular Biology (RECOMB)*, 2010.
26. Barna Saha, Ioana Stanoi, Kenneth L. Clarkson, "Schema Covering: a Step Towards Enabling Reuse in Information Integration," Proc. *26th IEEE International Conference on Data Engineering (ICDE)*, 2010.
27. Jian Li, Barna Saha, Amol Deshpande, "A Unified Approach to Ranking in Probabilistic Databases," *35th International Conference on Very Large Data Bases (VLDB)*, 2009, **(BEST PAPER AWARD)**
28. Barna Saha, Lise Getoor, "On Maximum Coverage in the Streaming Model & Application to Multi-topic Blog-Watch," Proc. *9th SIAM International Conference on Data Mining (SDM)*, 2009.
29. Samir Khuller, Barna Saha "On Finding Dense Subgraphs," Proc. *36th International Colloquium on Automata, Languages, and Programming (ICALP)*, 2009.
30. Barna Saha, Lise Getoor, "Group Proximity Measure for Recommending Groups in Online Social Networks," Proc. *2nd ACM SIGKDD Workshop on Social Network Mining and Analysis (SNA-KDD)*, 2008.
31. Barna Saha, Pabitra Mitra, "Dynamic Graph Clustering", Proc. *7th SIAM International Conference on Data Mining (SDM)*, 2007.
32. Sumit Ganguly, Barna Saha, "On Estimating Path-Aggregates over Streaming Graphs," Proc. *17th International Symposium on Algorithms and Computations (ISAAC)*, 2006.

### **Journal Papers.**

33. Barna Saha, "The Dyck Language Edit Distance Problem in Near-linear Time", Under review *Journal of ACM (JACM)*, November 2014.  
Conference version, Proc. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.
34. Mohammadtaghi Hajiaghayi, Wei Hu, Jian Li, Shi Li, Barna Saha, "A Constant Factor Approximation Algorithms for Fault Tolerant K-Median Problem", To Appear *Transaction on Algorithms (TALG)*, 2015.  
Conference version, Proc. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.
35. Ravishankar Krishnaswamy, Amit Kumar, Vishwanath Nagarajan, Yogish Sabharwal, Barna Saha "Facility location with matroid or knapsack constraints" *Mathematics of Operations Research (MOR)* , 2015.  
Conference version, Proc. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.
36. Lukasz Golab, Howard Karloff, Flip Korn, Barna Saha, Divesh Srivastava, "Discovering Conservation Rules", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 26(6), 1332-1348, 2014, **Special issue on the Best Papers of ICDE 2012**.  
Conference version *28th IEEE International Conference on Data Engineering (ICDE)*, **(Among Finalists for BEST PAPER AWARD)**, 2012.
37. Samir Khuller, Barna Saha, Kanthi Sarpatwar, "New Approximation Results for Resource Replication Problems", *Algorithmica*, 2015.  
Conference version *15th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, 2012.

38. Saeed Alaei, Mohammadtaghi Hajiaghayi, Vahid Liaghat, Dan Pei, Barna Saha, “Online Competitive Algorithms for Ad Allocation in Cellular Networks (AdCell)”, Revision under review, *Transaction on Algorithms (TALG)*, 2014.  
Conference version *19th. European Symposia on Algorithms (ESA)*, 2011.
39. Bernhard Haeupler, Barna Saha, Aravind Srinivasan, “New Constructive Aspects of the Lovász Local Lemma,” *Journal of the ACM (JACM)*, 58(6), 28, 2011.  
Conference version, *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010.
40. Jian Li, Barna Saha, Amol Deshpande, “A Unified Approach to Ranking in Probabilistic Databases,” *The VLDB Journal*, 2011, **Special Issue on the Best Papers from VLDB**.  
Conference version, *35th International Conference on Very Large Data Bases (VLDB), 2009, (BEST PAPER AWARD)*
41. Bogdan Alexe, Michael Gubanov, Mauricio A. Hernandez, Howard Ho, Jen-Wei Huang, Yannis Katsis, Lucian Popa, Barna Saha, Ioana Stanoi, “Simplifying Information Integration: Object-Based Flow-of-Mappings Framework for Integration,” *Lecture Notes in Business Information Processing, Revised Selected Papers from Second International VLDB Workshop, BIRTE*, 2009, Volume 27.

## PATENTS

1. Lukasz Golab, Howard Karloff, Flip Korn, Barna Saha, Divesh Srivastava, “Conservation Dependency,” US patent filed US20120130935, AT&T Research Laboratory, New Jersey, 2012.
2. Seungjoon Lee, Manish Purohit, Barna Saha, “Firewall Placement for Secure Communication Among Cloud Virtual Machines” US patent pending, AT&T Research Laboratory, New Jersey, 2013.
3. Kook Jin Ahn, Seungjoon Lee, Barna Saha, “Resource Management for Batch Processing for Infrastructure/ Platform as a Service,” US patent pending, AT&T Research Laboratory, New Jersey, 2013.

## TEACHING EXPERIENCE

- **Spring 2016** ALGORITHMS FOR DATA SCIENCE, University of Massachusetts Amherst
- **Spring 2015** REASONING WITH UNCERTAINTY, University of Massachusetts Amherst
- **Spring 2015** BIG DATA ALGORITHMS & APPLICATIONS, University of Massachusetts Amherst, <https://people.cs.umass.edu/~barna/CMPSCI691DA.html>
- **Spring 2015** APPROXIMATION ALGORITHMS & COMBINATORIAL OPTIMIZATION, University of Massachusetts Amherst, <https://people.cs.umass.edu/~barna/CMPSCI690AA.html>
- **Fall 2013** ALGORITHMIC TECHNIQUES FOR BIG DATA ANALYSIS as a Visiting Adjunct Faculty at the Computer Science Department of University of Minnesota Twin Cities <http://csci8980bigdataalgo.wordpress.com/>

## PROFESSIONAL ACTIVITIES

1. **Journal Editor:** Guest Editor, Transactions on Algorithms, SODA 2016 Special Issue.
2. **Committee Member:** SODA 2016, WSDM 2016, CIKM 2015, ECML/PKDE 2015 (Industrial Track), VLDB 2015, CIKM 2014, APPROX 2014, Grace Hopper Celebration of Women in Computing (GHC), 2012.
3. **Grant Review Panel** NSF Big Data Panel (2012, 2014), NSF Panel (2015).
4. **Organizer** New England Celebration of Women in Computing, 2016, ACM-CRA-W.
5. **Internal Committee** Faculty hiring committee (2014-15, 2015-16), Award Committee–Distinguished Lecture Series (2015-16), UMass Amherst.
6. **Ph.D. Committee**  
Manish Purohit, University of Maryland College Park.  
Ryan Cybulski, UMass Amherst, Astronomy

7. **Reviewer (Journals) (partial list):** IEEE Transactions on Knowledge Discovery; Internet Mathematics; Operations Research Letters, Maths of Operations Research, Algorithmica, Transactions on Algorithms, Transaction on Database Systems, Maths of Programming.
8. **Reviewer (Conferences) (partial list):** FOCS, SODA, STOC, ESA, VLDB, PODS, SIGMOD, ICDE, ISAAC, KDD, SDM, APPROX, RANDOM.

## STUDENTS

My Phan, (University of Massachusetts Amherst, Ph.D. Student, 2015-).  
 Sainyam Galhotra, (University of Massachusetts Amherst, Masters Thesis, 2015-).  
 Vivek Krishnamurthy, (University of Massachusetts Amherst, Undergraduate Honors Thesis, 2015-).  
 Brendan Tech, (University of Massachusetts Amherst, Undergraduate Honors Thesis, Completed, 2014-15).  
 Ari Kobren, (University of Massachusetts Amherst, Masters Thesis, Completed 2014-15).

## Student Interns

Manish Purohit, University of Maryland College Park, (June 2013-August 2013), Project: *Network Design Problem on Cloud*.  
 Donatella Firmani, Sapienza University of Rome, (April 2012-July 2012), Project: *Map Reduce Algorithms for Graph Clustering*.  
 Shanshan Ying, National University of Singapore, (June 2012-August 2012), Project: *Repairing Semi-structured XML Documents*.  
 Kook Jin Ahn, University of Pennsylvania, (June 2012-August 2012), Project: *Resource Allocation Problems on Cloud*.  
 Harmeet Jandu, Rutgers University, (September 2012-December 2012), Project: Supervised Master's thesis on *Network Optimization on Hadoop*.

## REU Students

James Kersh (Summer 2015, Rutgers University)  
 I am an active mentor of women students through the NSF Funded Organization: [Computing Behind Double Bind](#).

## SELECTED INVITED TALKS

1. "Playing with Grammars: What have we known so far?"
  - The Simons Institute Workshop on Computational Complexity of Low-Polynomial Time Problems.
2. "Randomization in Data & Design"
  - University of California Santa Barbara, Nov 2015
  - University of Southern California, Nov 2015
3. "The Language Edit Distance Problem"
  - MIT, Algorithms & Complexity Seminar, Oct 2015
  - Association for Women in Mathematics, April 2015
  - Google Mountain View, Feb 2015
  - Bellairs Workshop on Combinatorial Optimization, April 2015
4. "Perfecting the Imperfect: Reasoning with "real" Data via a Probabilistic Lens"
  - University of Illinois at Urbana-Champaign, March, 2014
  - Columbia University, March, 2014
  - University of Arizona, February 2014
  - Arizona State University, February 2014
5. "Clean Data and Unlimited Resources: A Probabilistic Journey to a Fantasy Land"

- University of California Davis, April 2014
  - Washington University in St. Louis, April 2014
  - Purdue University, April 2014
  - University of Massachusetts Amherst, March 2014
  - Dartmouth College, March 2014
  - Boston University, March 2014
  - Virginia Tech, February 2014
6. “Making your Clean Data Big: Scalable Algorithms for Data Quality Problems,” Massachusetts Institute of Technology, January, 2014.
  7. “Energy Efficient Scheduling & Covering with Hard Capacities,” University of Illinois at Urbana-Champaign, October, 2013.
  8. “Computing Dense Subgraphs from Theory to Practice,” University of Minnesota Twin Cities, October, 2013.
  9. “Finding Nemo: The Power of Probabilistic Methods,” Johns Hopkins University, May, 2013.
  10. “Resource Replication Problems,” Information Theory Workshop and Applications (ITA), Invited Talk, February, 2013.
  11. “Allocating Resources & Processing Top-k Queries: The Power of Probabilistic Methods,” Purdue University, March, 2011.
  12. “How to Approximate Better: Computing Dense Subgraphs and Allocating Resources,” Yahoo! Research, Santa Clara, February, 2011.
  13. “New Constructive Aspects of The Lovász Local Lemma,” IBM T.J. Watson Research Center, IP (Integer Programming) Talk, June, 2010.
  14. “Dense Subgraphs with Restrictions and Applications to Gene Annotation Graphs,” Center for Bioinformatics & Computational Biology, University of Maryland, April, 2010.

## REFERENCES

Prof. Samir Khuller  
 Chair of Computer Science Department,  
 Department of Computer Science,  
 University of Maryland College Park,  
 College Park, MD 20742  
 +1-(301) 405-6765  
 Email: samir@cs.umd.edu

Dr. Divesh Srivastava  
 Head of Databases Department  
 AT&T Shannon Research Laboratory,  
 180 Park Avenue, Building 103  
 Florham Park, NJ 07932-0971  
 +1-(917) 360-8776  
 divesh@research.att.com

Prof. Piotr Indyk  
 Department of Electrical Engineering & Computer Science  
 MIT Computer Science and Artificial Intelligence Laboratory (CSAIL),  
 Massachusetts Institute of Technology (MIT),  
 Cambridge, MA 02319  
 Phone: +1 (617) 452-3402  
 Email: indyk@csail.mit.edu

Prof. Robert Krauthgamer  
 Faculty of Mathematics and Computer Science  
 The Weizmann Institute of Science  
 234 Herzl St., P.O.B. 26, Rehovot 76100  
 Phone: 08-9344281  
 Email: robert.krauthgamer@weizmann.ac.il