Hirak Sarkar

5812 Quebec St.
Berwyn Heights, MD-20740
Mob No.+1 6315208131

✉ hsarkar@cs.umd.edu
www.hiraksarkar.com

Education

Ph.D in Computer Science 2020 (expected)
Statistical Inference in Biological Data, *advisor: Prof. Rob Patro*
University of Maryland, MD

MS in Computer Science 2014 - transferred
Improving storage and alignment methodologies for RNA-seq data, *advisor: Prof. Rob Patro*
Stony Brook University, NY
GPA: 3.9/4.00

Masters of Technology (Computer Science) 2011-2013
Indian Statistical Institute, Calcutta
1^{st} Class (Hons.)

Bachelor of Technology (Computer Science and Engineering) 2007-2011
West Bengal University of Technology
GPA: 8.88/10

Publications

1. *Terminus enables the discovery of data-driven, robust transcript groups from RNA-seq data.*, by Hirak Sarkar, Avi Srivastava, Hector Corrada Bravo, Michael I. Love and Rob Patro. [***ISMB' 20***]
2. *A Bayesian framework for inter-cellular information sharing improves dscRNA-seq quantification*, by Avi Srivastava, Laraib Malik, Hirak Sarkar, Rob Patro. [***ISMB' 20***]
3. *Alignment and mapping methodology influence transcript abundance estimation*, by Avi Srivastava, Laraib Malik, Hirak Sarkar, Mohsen Zakeri, Charlotte Soneson, Michael I. Love, Carl Kingsford, Rob Patro. [Submitted ***Genome Bio.***]
4. *Minnow: A principled framework for rapid simulation of dscRNA-seq data at the read level*, by Hirak Sarkar, Avi Srivastava and Rob Patro [***ISMB'19***].
5. *Towards Selective-Alignment: Producing Accurate And Sensitive Alignments Using Quasi-Mapping*, by Hirak Sarkar*, Mohsen Zakeri*, Laraib Malik and Rob Patro [***ACM BCB'18***].
6. *An Efficient, Scalable and Exact Representation of High-Dimensional Color Information Enabled via de Bruijn Graph Search*, by Fatemeh Almodaresi*, Hirak Sarkar*, Avi Srivastava and Rob Patro [***ISMB'18***].
7. *Quark enables semi-reference-based compression of RNA-seq data* by Hirak Sarkar and Rob Patro [*accepted* ***Bioinformatics'17***].
8. *Fast, Lightweight Clustering of de novo Transcriptomes using Fragment Equivalence Classes* by A Srivastava*, Hirak Sarkar*, Laraib Malik and Rob Patro (* *Joint first authors*) [***RECOMB-seq'16***].
9. *RapMap: A Rapid, Sensitive and Accurate Tool for Mapping RNA-seq Reads to Transcriptomes* by A Srivastava, Hirak Sarkar, Nitish Gupta and Rob Patro [***ISMB'16***].
10. *Voronoi Game on Graphs* (Extended version) by S. Bandyapadhyay, A. Banik, S. Das and H. Sarkar (in alphabetical order of surnames) *Journal of Theoretical Computer Science* [***Journal of Theoretical Computer Science'15***].
11. *Voronoi Game on Graphs* by S. Bandyapadhyay, A. Banik, S. Das and H. Sarkar (in alphabetical order of surnames) Seventh International Workshop on Algorithms and Computation. ***WALCOM'13***.

Posters

1. *Pufferfish: A fast graph-based indexing and query strategy for large genomic sequences* by Fatemeh Almodaresi*, Hirak Sarkar*, and Rob Patro, Poster presented in [***WABI'17***].
2. *Joint probabilistic model for multiple steps of gene regulation* by Hirak Sarkar, Yi-Fei Huang and Adam Siepel, Poster presented in ***BioData'16***

Professional Experience

- **Facebook Inc.** Worked as Ph.D data scientist intern. I worked on designing scalable pipelines to analyze the data from millions of users. Applied transfer learning based methods in order to measure the effect of natural calamity such as cyclone or wild fire from open source satellite images.
- **Simons Center for Quantitative Biology, Cold Spring Harbor Lab:** Worked under the supervision of Prof. Adam Seipel from May, 2016 to July, 2016. We designed probabilistic graphical model to infer transcription and degradation rates from different assays such as GRO-seq and RNA-seq.
- *Summer Assistantship '15,'17* with Prof. Rob Patro. We worked on various problems ranging from
- Teaching Assistant for CSE549 (Computational Biology), CSE219 (Game Programming)
- *Visiting Researcher* at Advanced Computing & Microelectronics Unit, Indian Statistical Institute from October, 2013 to December, 2013. I worked on Computational Geometry and Graph Theory
- *Junior Research Fellow* in Department of Computer Science & Engineering at Indian Institute of Technology, Kharagpur (IIT) from July, 2013 to Sept, 2013. I was a member of Complex Network Engineering Group. I did TA-ship for Introductory Programming Course in that brief stint.

Relevant Course Projects

- *IPID Header Survey:* We used IPID headers to estimate the load over different servers, sampled from alexa list. The main challenge of the project is to detect the wrapping pattern and navigate through the global vs local IPID counter. We also looked at the temporal pattern of network traffic for the different regional websites which shows interesting correlation with possible working load at the server end.
  Instructor: Prof. Phillipa Gill
- *Some Geometric and Combinatorial Properties of Binary Matrices Related to Discrete Tomography:* Here we are trying to decompose an image matrix into matrices each having orthogonal convex polygon also known as Ferrer?s digraph. An operation could regenerate the original image from these matrices. The methods can be applied to image and data compression. (*Masters dissertation*) *Advisor: Prof. Bhargab B Bhattacharya & Prof. Sandip Das*
- *GameSAT- A Structured Approach to Combine SLS SAT Solvers:* Here we used several existing heuristic algorithms to mix up with each other in a customized probabilistic fashion in order to solve combinatorial hard problems encoded as SAT instances.
  (*B.Tech dissertation*) *Advisor: Dr. Ashiqur KhudaBukhsh, CMU*

Awards and Honors

- Awarded *Research Assistantship*, SBU (*2016-present*)
- Awarded *Special CS Chair Fellowship* (*of $10000*), SBU (*2014-2015*)
- Awarded *NUS Research Scholarship*, NUS (*Jan'14-June'14*)
- Awarded *Post-graduate Scholarship* by, Govt. of India. (*2011-2013*)
- Received First Prize for Software Competition (IEM), Calcutta.

Programming Skills

Python, C++,C

Open Source Tools Used

Dendropy, BioNet (Comp Bio)
NLTK, Scrapy, Scikit-learn, Stanford Parser, Pandas (Data Science)

Relevant Coursework

- Artificial Intelligence, Computational Biology, Analysis of Algorithms, Fundamental of Networks. (at *SBU*)
- Machine Learning & Pattern Recognition, Image Processing, Stochastic Process, Optimization Algorithms, Computer Graphics. (at *Indian Statistical Institute*)