Hirak Sarkar

8 Acorn Ln
Stony Brook, NY-11790
Mob No.+1 6315208131

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

Research Interest

I am interested in applying machine learning techniques such as statistical inference and deep learning to analyse and extract information from big data in the field of genomics, social and computer networks.

Education

Ph.D in Computer Science 2014-2019
Statistical Inference in Biological Data, *advisor: Prof. Rob Patro*
Stony Brook University, NY
GPA: 3.99/4.00

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

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

Publications

1. *Towards selective-alignment: Bridging the accuracy gap between alignment-based and alignment-free transcript quantification*, by Hirak Sarkar*, Mohsen Zakeri*, Laraib Malik, Rob Patro. [*Submitted to Bioinformatics, 2017*]
2. *Quark enables semi-reference-based compression of RNA-seq data* by Hirak Sarkar and Rob Patro [*accepted* ***Bioinformatics'17****, impact factor: 7.307*].
3. *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***].
4. *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****, acceptance rate: 17%*].
5. *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* [***TCS'15***].
6. *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***.
7. *Pufferfish: A fast graph-based indexing and query strategy for large genomic sequences* by Fatemeh Almodaresi*, Hirak Sarkar*, Yi-Fei and Rob Patro, Poster presented in [***WABI'17***].
8. *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

- **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 from May, 2015 to July, 2015. 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
- *Classification & Clustering of satellite images:* We benchmarked different classification algorithms to cluster different kind of landmasses from infrared images taken by satellite.
  Instructor: Prof. C A Murthy
- *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 probability to solve combinatorial hard problems encoded as SAT problems. We used UBCSAT framework for experimentation.
  (*B.Tech dissertation*) *Advisor: 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*)