

RESEARCH INTERESTS

---

**Computer Science** Algorithms and Machine Learning for Massive Datasets

EDUCATION

---

**University of California, Berkeley**

B.A. in Mathematics, completed May 2020

B.A. in Computer Science, completed May 2021

*August 2018 - June 2021*RESEARCH EXPERIENCE

---

**Microsoft Research, India***Research Fellow, advised by Dr. Ravishankar Krishnaswamy**September 2021 - now*

- Researched theoretical and practical problems for approximate nearest neighbors search (ANNS) with Dr. Ravishankar Krishnaswamy, Dr. Harsha Vardhan Simhadri, Dr. Janardhan Kulkarni, and Dr. Sepideh Mahabadi
- Additionally worked on fairness and online variants of the Joint Replenishment Problem, a popular inventory management problem in operations research

**Stanford University***Intern, Snyder Laboratory (Stanford School of Medicine)**June 2015 - August 2018*

- Researched predicting kinase phosphorylation sites with protein sequence data with Dr. Andrew Lipchik
- Worked on data visualization techniques for complex proteomics and genomics data with Dr. Doug Phanstiel

*Intern, Goel Laboratory (Stanford School of Engineering)**June 2016 - August 2016*

- Examined municipal election voting data with clustering/statistical methods, and presented my findings (e.g. geographical bias, socioeconomic factors, etc.) to Prof. Ashish Goel

PUBLICATIONS

---

· **Filtered-DiskANN: Graph Algorithms for Approximate Nearest Neighbor Search with Filters**

**Authors:** Siddharth Gollapudi, Neel Karia, Varun Sivashankar, Ravishankar Krishnaswamy, Harsha Vardhan Simhadri, Nikit Begwani, Amit Singh, Swapnil Raz, Yiyong Lin, Yin Zhang, Neelam Mahapatro, Premkumar Srinivasan (Authors listed by contribution). *The Web Conf. (WWW) 2023*

**Summary:** Despite the fast-growing popularity of *approximate nearest neighbors* (ANNS) based information retrieval, little work has been done for efficiently handling hybrid queries. A hybrid query is a query embedding that comes with additional hard constraints (e.g. keywords, language, etc.), and an ANNS index supporting such queries must only output results matching the constraints. For this variant, we present two graph-based heuristics. Both utilize the geometric embedding data and also the hard filter constraints in adding connections.

· **Improved Approximation Algorithms for the Joint Replenishment Problem with Outliers, and with Fairness Constraints**

**Authors:** Varun Suriyanarayana, Varun Sivashankar, Siddharth Gollapudi, David B. Shmoys (Authors listed by contribution). *SODA 2023*

**Summary:** The *joint replenishment problem* is an well-studied inventory management question from operations research. We are given a series of time-ordered demands of varying types that have decision costs and holding costs: the overall cost of satisfying a demand depends on the time taken to satisfy the demand (holding cost), and the cost of deciding to satisfy the demand (general and type costs). We consider a *fairness* variant of the problem: each demand type is also given one of constantly many colors, and each color allows for a given number of demand *rejections*. For this variant, we present the first constant-factor approximation algorithm. It is based on a strong LP relaxation and a novel pipage-rounding solution.

· **Composable Coresets for Determinant Maximization: Greedy is Almost Optimal**

**Authors:** Siddharth Gollapudi, Sepideh Mahabadi, Varun Sivashankar (Authors listed alphabetically). *NeurIPS 2023*

**Summary:** Given a set of vectors, the *Determinant Maximization Problem* aims to pick a fixed-size subset of vectors that achieve the largest determinant. This problem has seen interest for modeling diversity in large datasets across many settings, namely *composable coresets*: for a series of datasets, each one can be “summarized” via a *coreset*. Operations on the coresets are considerably faster up to an approximation factor. In practice, the popular and performant greedy algorithm for determinant maximization is used to find composable coresets, but the known approximation factor for greedy is poor. We present an elegant linear algebra result to show that the approximation factor for greedy is nearly optimal, corroborating the strong performance the algorithm displays in practice.

## PROJECTS

---

### Reading Projects and Surveys

*June 2019 - December 2020*

- Discussed the textbook *Graph Theory* by Bondy and Murty with Prof. Dimitry Vaintrob from the Dept. of Mathematics
- Presented survey on the Unique Games Conjecture, a cornerstone of hardness of approximation research
- Read/Presented on optimization problems such as MAX-CUT and 3-COLOR through the Berkeley Directed Reading Program
- Presented a survey on the Restricted Burnside Problem, a key result in group theory

### Systems Projects

*January 2020 - Now*

- Developed a barebones operating system (PintOS) using elementary kernel development principles (course project)
- Utilized basic terminal APIs to implement a *vi*-style text editor in C
- Wrote a basic Intel 8080 CPU emulator in C

## SERVICE

---

### Conference Reviewing

- ESA 2022, SODA 2024

## TEACHING

---

### FIRST Robotics Team 299

*April 2016 - May 2023*

*Team Mentor*

- Introduced students to fundamentals of programming: this included teaching about languages such as Python and C++, commonly used algorithms in engineering such as sorting and searching techniques, and incorporating what was taught into developing software for robot substructures
- Worked with advanced students on a variety of topics, including systems programming, discrete optimization, and control theory

## INDUSTRY EXPERIENCE

---

### Microsoft Research, India

*September 2021 - now*

*Research Fellow*

- Implemented research algorithms for filters-based search in a production-level sponsored-search setting, giving a 20% increase in advertisement engagement and a 15% increase in gained revenue
- Contributed to open-source development for DiskANN, an ANNS library powering many product backends at Microsoft

### Amazon

*May 2020 - August 2020*

*Summer Intern, AWS Robomaker*

- Rewrote the ROS2 (Robot Operating System) cross-compilation tool, enabling all Robomaker clients to remotely build ROS2 applications and encouraging future open-source contributions
- Productionized a metrics dashboard for pinpointing future improvements and maintaining reliability of the tool

## SKILLS

---

**Programming Languages** Python, C, C++, Bash, Java

**Software & Tools** git, L<sup>A</sup>T<sub>E</sub>X, pandas, Tensorflow/Pytorch, Docker, AWS, Azure