DINGYU WANG

(+1) 734-358-0383 | University of Michigan, Ann Arbor, MI | wangdy@umich.edu | ancientwang.com

EDUCATION

- Ph.D. student, Computer Science, College of Engineering, University of Michigan
  Advisor: Seth Pettie
  Sep. 2019 - Now
- B.S.E., Computer Science, College of Engineering, University of Michigan
  Sep. 2017 - May. 2019
- B.S.E., Electrical and Computer Engineering, SJTU-UM Joint Institute, Shanghai Jiao Tong University
  Sep. 2015 - Aug. 2017 and June. 2019 - Aug. 2019

RESEARCH

Sequential Cardinality Estimation in Data Streams
- Simplified the analysis of estimation algorithm that uses the martingale estimator.
- Developed new estimation algorithm that is practical and achieves a better memory-variance trade-off.

Distributed Cardinality Estimation in Data Streams
- Developed new framework to analyze the limiting performance of composable estimation algorithms.
- Constructed τ-generalized remaining area estimators which generalizes classic estimators like LogLog and HyperLogLog. It beats HyperLogLog slightly by choosing the optimal τ.

Contention Resolution for Multiple Access Channels
- Found the exact optimal contention resolution algorithms for conflict size 2.

PUBLICATIONS

- Non-Mergeable Sketching for Cardinality Estimation (ICALP21)
  Seth Pettie, Dingyu Wang and Longhui Yin
  We study sketching schemes for the *sequential cardinality estimation* problem, and advocate for measuring the efficiency of such a scheme in terms of its MVP: Memory-Variance Product, i.e., the product of its space, in bits, and the *relative* variance of its estimates. The Cohen/Ting (2014) *martingale transformation* provides a general way to obtain optimal variance for sketches. We first construct a framework that simplifies the analysis of the optimal variance. We then develop a new class of *"curtain"* sketches that are a bit more complex than the state-of-the-art Martingale LogLog (MVP ≈ 4.16) but with substantially better MVPs, e.g., Martingale Curtain has MVP ≈ 2.31. We also prove that Martingale Fishmonger has an MVP of around 1.63, and conjecture this to be an information-theoretic lower bound for the sequential cardinality estimation.

- Information Theoretic Limits of Cardinality Estimation: Fisher meets Shannon (STOC21)
  Seth Pettie and Dingyu Wang
  We study the intrinsic tradeoff between the *space complexity* of the sketch and its *estimation error* in the *distributed cardinality estimation*. We define a new measure of efficiency for cardinality estimators called the *Fisher-Shannon* (Fish) number $\mathcal{H}/\mathcal{I}$, which is the product of its space (in bits) and the relative variance *assuming perfect compression and statistically efficient estimator*. We prove that many variants of the PCSA sketch of Flajolet and Martin have Fish number H_0/I_0, where H_0, I_0 are two precisely-defined constants, and that all base-q generalizations of (Hyper)LogLog are strictly worse than H_0/I_0, but tend to H_0/I_0 in the limit as $q \to \infty$. We conjecture that the Fish-number H_0/I_0 is a universal lower bound for any such composable sketch.

- Optimal Protocols for 2-Party Contention Resolution (SSS21)
  Seth Pettie and Dingyu Wang
  Nearly all work on the Contention Resolution problem evaluated the performance of algorithms *asymptotically*, as $n \to \infty$. In this work we focus on the simplest case of $n = 2$ devices, but look for *precisely* optimal algorithms. We design provably optimal algorithms under three natural cost metrics: minimizing the expected waiting time until the first device acquires the resource (min), the expected time until the last device acquires the resource (max), and the expected average of the waiting times (avg).

MANUSCRIPT/IN SUBMISSION

- Simpler and Better Cardinality Estimators for HyperLogLog and PCSA
  Seth Pettie and Dingyu Wang
  We construct τ-generalized remaining area estimators which generalizes classic estimators like LogLog and HyperLogLog. It beats HyperLogLog slightly by choosing the optimal τ. We also provide a simple and elegant framework to analyze HyperLogLog in the asymptotic region.

TEACHING

- GSI (Graduate Student Instructor) of Introduction to Operating Systems, EECS 482, Fall2022, U of M
- GSI of Algorithms, EECS 586, Winter2022, U of M
- GSI of Foundations of Computer Science, EECS 376, Fall2021, U of M
- TA (Teaching Assistant) of Electromagnatics, 2019, SJTU
- TA of Honors Physics I, 2017, SJTU

HONORS

- Honor Competition finalist, 2021, CSE, U of M
- Jackson and Muriel Lum Scholarship, 2017-2018, U of M
- Excellent Teaching Assistant Award, 2017, Joint Institute, SJTU
- John Wu and Jane Sun Sunshine Scholarship, 2016-2017, Joint Institute, SJTU
- Excellent Student Union Member, 2015-2016, Joint Institute, SJTU