

ACADEMIC EXPERIENCE

- 2016.8 - **The Chinese University of Hong Kong (CUHK)** Shatin, Hong Kong
Present **Ph.D.** in Computer Science and Engineering
◦ Supervisor: Prof. James Cheng
- 2011.9 - 2015.6 **Huazhong University of Science and Technology (HUST)** Wuhan, China
B.Eng in Computer Science and Technology
Overall GPA: 3.88 / 4.00 Rank: 1/ 30/ 472 (in a 30-student Honor Class)

RESEARCH INTERESTS

My general research interests cover the broad area of distributed systems and databases, with special emphasis on distributed graph processing systems and distributed machine learning/deep learning systems. My current works focus on RDMA based OLAP/OLTP system over knowledge graphs, distributed DL training system and Graph Neural Network.

WORK EXPERIENCE

- 2019.2 - **HUAWEI, 2012 Lab, Parallel and Distributed Computing Laboratory** Shenzhen, China
Present ◦ **Research Intern** in MindSpore Team, Distributed Deep Learning Platform.
- 2017.5 - 2017.8 **University of Pennsylvania, NetDB Lab, Dept. CIS** PA, USA
◦ **Visiting Scholar**, Distributed Graph Query Optimization
Supervisor: **Prof. Boon Thau Loo**.
- 2014.6 - 2015.6 **Microsoft Research Asia, Software Analytics Group** Beijing, China
◦ **Research Intern**, Large-scale Data Analytics, Distributed System.
Supervisor: **Qingwei Lin** (Lead Researcher) and **Dr. Jianguang Lou** (Principle Researcher).
- 2013.9 - 2014.6 **HUST, IDC Lab, Dept. CSE** Wuhan, China
◦ **Research Intern**, System Optimization on Hadoop.
Supervisor: **Prof. Ruixuan Li**

PUBLICATIONS

- [1] *A Representation Learning Framework for Property Graphs* SIGKDD'19
Yifan Hou, **Hongzhi Chen**, Changji Li, James Cheng, Ming-Chang Yang.
- [2] *Large Scale Graph Mining with G-Miner* SIGMOD'19
Hongzhi Chen, Xiaoxi Wang, Chenghuan Huang, Juncheng Fang, Yifan Hou, Changji Li, James Cheng.
- [3] *Optimizing Declarative Graph Queries at Large Scale* SIGMOD'19
Qizhen Zhang, Akash Acharya, **Hongzhi Chen**, Simran Arora, Ang Chen, Vincent Liu, Boon Loo.
- [4] *Scalable De Novo Genome Assembly Using a Pregel-Like Graph-Parallel System* TCBB'19
Guimu Guo, **Hongzhi Chen**, Da Yan, James Cheng, Jake Chen, Zechen Chong.
- [5] *Lightweight Fault Tolerance in Pregel-Like Systems* ICPP'19
Da Yan, James Cheng, **Hongzhi Chen**, Cheng Long, Purushotham Bangalore.
- [6] *G-Miner: An Efficient Task-Oriented Graph Mining System.* EuroSys'18
Hongzhi Chen, Miao Liu, Yunjian Zhao, Xiao Yan, Da Yan, James Cheng.
- [7] *Norm-Ranging LSH for Maximum Inner Product Search.* NIPS'18
Xiao Yan, Jinfeng Li, Xinyan Da, **Hongzhi Chen**, and James Cheng.
- [8] *Scalable De Novo Genome Assembly Using Pregel.* ICDE'18
Da Yan, **Hongzhi Chen**, James Cheng, Zhenkun Cai, Bin Shao.
- [9] *GraphD: Distributed Vertex-Centric Graph Processing Beyond the Memory Limit.* TPDS'18
Da Yan, Yuzhen Huang, Miao Liu, **Hongzhi Chen**, James Cheng, Huanhuan Wu, Chengcui Zhang.

- [10] *Architectural Implications on the Performance and Cost of Graph Analytics Systems.* **SoCC'17**
Qizhen Zhang, **Hongzhi Chen**, Da Yan, James Cheng, Boon Thau Loo, Purushotham Bangalore.
- [11] *G-thinker: Big Graph Mining Made Easier and Faster.* **arXiv'17**
Da Yan, **Hongzhi Chen**, James Cheng, M.Tamer.Ozsu, Qizhen Zhang, John C.S. Lui.

PROJECTS

My research focuses on the design and implementation of distributed systems and the applications on the top of it. I have led or participated as the core developer of the following projects.

-
- Grasper [Ongoing] An RDMA-enabled graph OLAP/OLTP system with good scalability, which currently has achieved order of magnitudes performance improvements over existing systems (e.g., Titan, JanusGraph, OrientDB, Neo4j).
 - PGE A representation learning framework for property graph embedding. The key idea of PGE is a three-step framework to leverage both the topology and property information into Graph Neural Networks for a better node embedding result.
 - GraphRex An efficient framework for graph processing on datacenter infrastructure. The key technical contribution of GraphRex is the identification and optimization of a set of global operators whose efficient implementation is crucial to the good performance of large, datacenter-based graph analysis.
 - G-Miner A distributed graph processing system aimed at general graph mining problems, which have intensive local computation inside a subgraph. We modeled each subgraph processing as a task and designed a task-based pipeline to improve the parallelism between computation and communication. A dynamic task stealing mechanism as well as an efficient cache strategy were also proposed to further speed up the task processing.
 - G-thinker Real applications, such as graph matching and community detection, often require computation intensive graph analytics, which cannot be represented by vertex-centric algorithms for efficient execution in systems like Pregel and GraphLab. We proposed G-thinker, a new subgraph-centric general graph processing distributed system, which is natural for subgraph finding problem.
 - PPA-Assembly A scalable toolkit for de novo genome assembly was developed based on Pregel. PPA-Assembly provides a set of key operations in genome assembly, which were implemented by practical Pregel algorithms (PPAs) with strong performance guarantees.
 - FPM A novel and general distributed framework to mine frequent patterns, including frequent item-sets/sequences/graphs.

I also joined in some other research projects during my work at CUHK, UPenn and MSRA.

-
- RANGE-LSH Using maximum inner product for similarity search, which significantly outperforms SIMPLE-LSH, and RANGE-LSH is robust to the shape of 2-norm distribution and different partitioning methods.
 - GraphD It offers out-of-core support for processing very big graphs in a small cluster of commodity PCs, with performance comparable with the state-of-the-art distributed in-memory graph systems.
 - LWCP A fault tolerance mechanism for Pregel-like systems with performance tens of times faster than the conventional checkpointing mechanisms.
 - More** If you are also interested in my previous works at **MSRA, Software Analytics Group**, please visit my homepage for more details. My works at MSRA focused more on distributed data analytics, including OLAP, pattern mining, text clustering and anomaly detection. I mainly participated in 4 projects as a core developer, i.e. *Service-Intelligence*, *Service-Insider*, *iDice* and *In4*. In particular, both **Service-Intelligence** and **iDice** have been published in **ICSE' 2016**

AWARDS & HONORS

- 2019.5 SIGMOD Travel Award
- 2018.4 EuroSys Travel Award
- 2016 - 2020 CUHK Postgraduate Studentship.
- 2015.6 The original winner of **Hong Kong PhD Fellowship**.

- 2015.6 “Stars of Tomorrow” at Microsoft Research Asia (Only 15% research interns won the Award)
- 2015.6 Outstanding Graduates (3% in HUST)
- 2014.10 CCF (China Computer Federation) National **Top 100** Outstanding Undergraduates (**Top 0.1%**)
- 2014.9 Academic Excellence Scholarship (2% in HUST)
- 2014.9 Merit Undergraduate (2% in HUST)
- 2013.9 National Undergraduate Scholarship (2% in HUST)
- 2013.9 Merit Undergraduate (2% in HUST)
- 2012.9 Most Outstanding Undergraduate (1% in HUST)
- 2012.9 Academic Excellence Scholarship (2% in HUST)

TEACHING

- Spring, 2018 CSCI1020: Hands-on Introduction to C++
- Fall, 2017 ENGG1110: Problem Solving By Programming
- Spring, 2017 ENGG1110: Problem Solving By Programming
- Fall, 2016 ENGG1110: Problem Solving By Programming

PROFESSIONAL ACTIVITIES

External Reviewer

- 2019 SIGMOD
- 2018 VLDB, ICDE
- 2017 VLDB, ICDE, CCGRID, BigData
- 2016 VLDB, KDD, SOCC, ICDM, DASFAA, BigData, APWeb

Participation in

- 2018 European Conference on Computer Systems, Porto, Portugal
- 2015 China National Computer Congress, Zhengzhou, China

SKILLS

- Programming C, C++, C#, Java, Python
- Operating Linux, Windows
- Documentation Latex, MS Office, HTML