Hongzhi Chen

RM 121(A), SHB, CUHK, Hong Kong
(+852) 6840-6304 | (+86) 150-027-34771
chzyaobaiwei@gmail.com
https://yaobaiwei.github.io/

ACADEMIC EXPERIENCE

2016.8-Present **The Chinese University of Hong Kong (CUHK),** *Shatin,* Hong Kong
Ph.D. in Computer Science and Engineering
- Advisor: 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 Talented Class**)

RESEARCH INTERESTS

My general research interests cover the broad area of distributed systems and databases, with special emphasis on large-scale graph processing systems and distributed data analytics systems. My current focus is on RDMA based OLTP/OLAP systems over big graphs, graph embedding and similarity search.

RESEARCH EXPERIENCE

2015.9-Now **Research Assistant,** *Dept of Computer Science and Engineering,* **CUHK**
Advisor: Prof. James Cheng
- G-Miner: A task-oriented distributed graph mining system that achieves orders of magnitude performance improvement over other state-of-the-art systems thanks to its novel task pipelining design.
- G-thinker: A novel subgraph-centric distributed graph processing system, which can be deployed on large graphs for graph matching, clique finding, communcity detection, etc.
- PPA-Assembly: A Pregel-based De Novo genome assembly to provide a library of adapted operations for genome sequencing with strong performance guarantee.
- Architectural Implications on the Performance and Cost of Graph Analytics Systems.
- FPM / GraphD / LWCP (please refer to details in the Projects section)

2017.5-2017.7 **Research Assistant,** *Dept of Computer and Information Science,* **University of Pennsylvania**
Advisor: Prof. Boon Thau Loo
- GraphRex: an efficient, robust, scalable, and easy-to-program framework for graph processing on datacenter infrastructure.

2014.6-2015.6 **Research Intern,** *Software Analytics Group,* **Microsoft Research Asia**
Advisor: Dr. Jianguang Lou (Senior Researcher) and **Qingwei Lin** (Researcher)
- Worked on large-scale software log analytics through Microsoft Cosmos for Microsoft Azure and Office.
- Worked on emerging pattern mining on multi-dimensional data.
- Worked on a data mining plug-in on EXCEL, which includes algorithms such as frequent pattern mining, text clustering, association rule mining, anomaly detection and mutil-dimension change detection.
- Worked on a distributed OLAP system using Akka.

PUBLICATIONS

[1] Qizhen Zhang, Akash Acharya, Hongzhi Chen, Simran Arora, Ang Chen, Vincent Liu, Boon Loo.
Optimizing Declarative Graph Queries at Large Scale
In Proceedings of the 2019 ACM International Conference on Management of Data. (**SIGMOD'19**)

[2] Xiao Yan, Jinfeng Li, Xinyan Da, Hongzhi Chen, and James Cheng.
Norm-Ranging LSH for Maximum Inner Product Search.
In Proceedings of the 31st Annual Conference on Neural Information Processing Systems, 2018. (**NIPS'18**)

[3] Hongzhi Chen, Miao Liu, Yunjian Zhao, Xiao Yan, Da Yan, James Cheng.
G-Miner: An Efficient Task-Oriented Graph Mining System.
In Proceedings of the 2018 European Conference on Computer Systems. (**EuroSys'18**)

[4] Da Yan, Hongzhi Chen, James Cheng, Zhenkun Cai, Bin Shao.
Scalable De Novo Genome Assembly Using Pregel.
In Proceedings of the 34th IEEE International Conference on Data Engineering, 2018. (**ICDE'18**)

[5] Qizhen Zhang, Hongzhi Chen, Da Yan, James Cheng, Boon Thau Loo, Purushotham Bangalore.
Architectural Implications on the Performance and Cost of Graph Analytics Systems.
In Proceedings of the 2017 ACM Symposium on Cloud Computing. (**SoCC'17**)

[6] Da Yan, Yuzhen Huang, Miao Liu, Hongzhi Chen, James Cheng, Huanhuan Wu, Chengcui Zhang.
GraphD: Distributed Vertex-Centric Graph Processing Beyond the Memory Limit.
IEEE Transactions on Parallel and Distributed Systems, 2017. (**TPDS'17**)

[7] Da Yan, Hongzhi Chen, James Cheng, M.Tamer.Ozsu, Qizhen Zhang, John C.S. Lui.
G-thinker: Big Graph Mining Made Easier and Faster.
arXiv preprint arXiv:1709.03110, 2017

PROJECTS

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

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 framework was proposed to mine frequent patterns, including frequent itemsets/sequences/graphs, using an efficient Master-Slaves model as well as dynamic work stealing strategy.

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 processing and data mining, 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

2016.8 CUHK Postgraduate Studentship.

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