Big Graph Analytics and Query Processing

Lijun Chang

University of Sydney

Short Bio

- Lecturer, The University of Sydney
  - August 2017—Present
- DECRA(Discovery Early Career Researcher Award) Fellow at UNSW
  - January 2015—July 2017
- Research Fellow, first at CUHK, then at UNSW
  - September 2011—December 2014
- PhD degree from The Chinese University of Hong Kong in August 2011, and B.Eng degree from Renmin University of China in July 2007

Short Bio

- Research Grants
  - **Australian Research Council Discovery Projects**: Efficient and Scalable Subgraph Search from Big Graphs in Cloud. **AUD$335,000**
  - **Australian Research Council Discovery Early Career Researcher Award**: Efficient Cohesive-Subgraph Search over Large Graphs. **AUD$372,000**
- Algorithmic Programming Competition Awards
  - Google Code Jam World Finals 2008, Mountain View, California, USA
  - ACM/ICPC World Finals 2006, San Antonio, Texas, USA

Requirement & Outcome

- Requirement
  - Good C/C++ programming skills and/or mathematical background
  - Optional: award in Informatics Competitions is a plus
- Anticipated Outcome after One Year
  - Submit one research paper
  - Improved programming and problem solving skills

Graphs

- Graphs are everywhere

Graphs

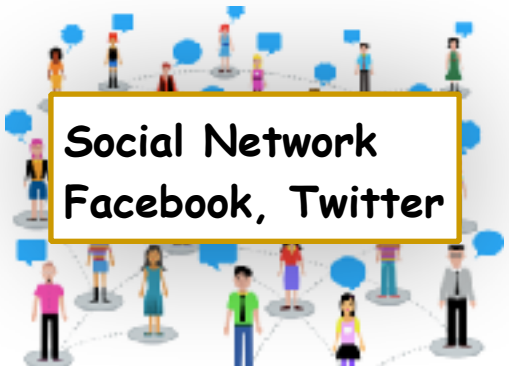
- Graphs are everywhere

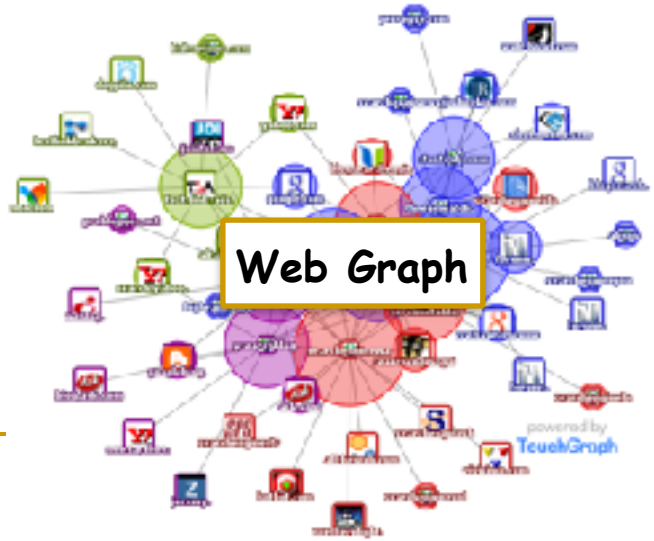
Graphs

- Graphs are everywhere

Graphs

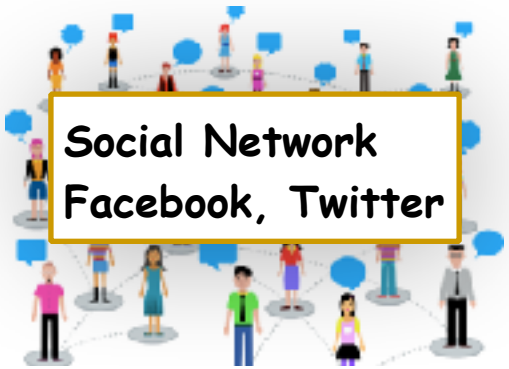
- Graphs are everywhere

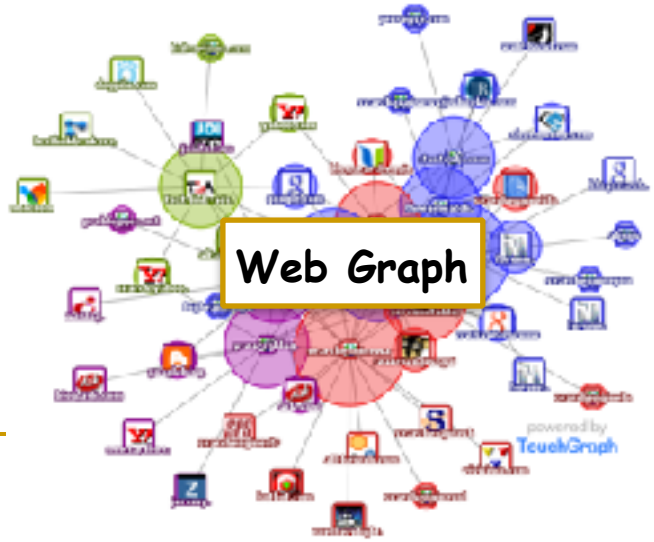


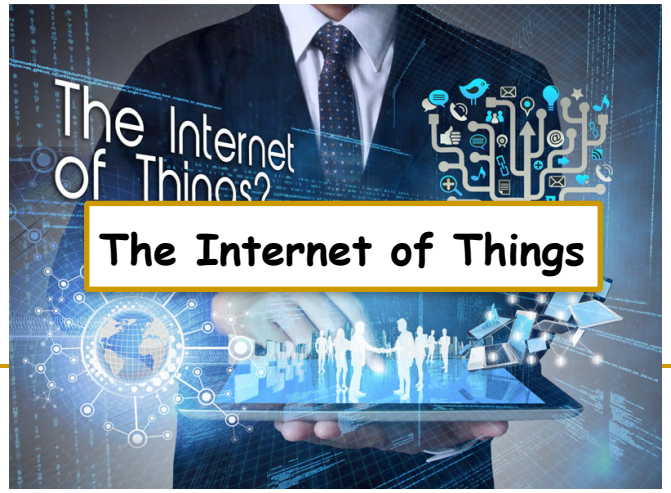


Graphs

- Graphs are everywhere







Big Graphs

- Graphs are Big (Volume)

- 302 million monthly active users
- 208 followers on average

- 1.4 billion users in 2014
- 0.4 **trillion** relationships in 2014

- 2.1 billion webpages in 2000
- 15 billion edges in 2000
- 20 PB data/day in 2008

Graph Research Problems

- New Graph Query Semantics
  - Find meaningful results
  - Keyword search, community search
- New Indexing Techniques
  - Efficiently indexing paths, trees, and subgraphs
- New Graph Processing Algorithms
  - Speed, speed, speed
  - Many graph problems are NP-hard
- New Computational Model

New Computational Models

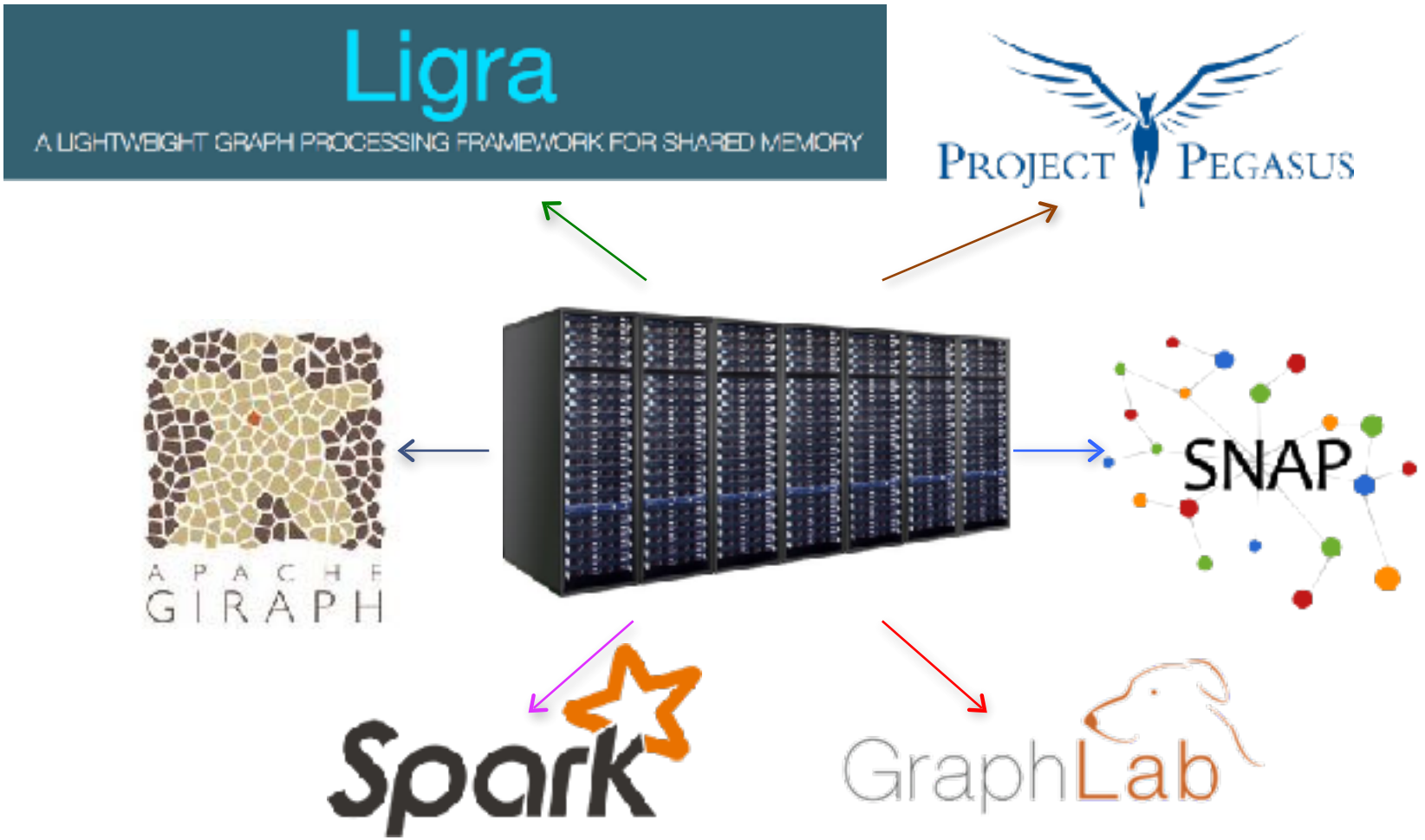
New Computational Models

Single Machine vs Multiple Machines

In-memory Algorithms vs External-memory Algorithms

Single Core vs Multiple Cores

New Computational Models



Our Work—Query Semantics

- Structural Keyword Search On Graphs
  - Given a set of keywords, find minimal connected trees and subgraphs
  - One book **(Morgan & Claypool, 2009**), **SIGMOD'09**, **ICDE'11, ICDE'09a, ICDE'09b**, **VLDB Journal'11**, **TKDE'12**
- Diversified Top-k Search
  - Find a set of results with minimal overlaps
  - **ICDE'15, VLDB'12, VLDB Journal'15**

Our Work—Indexing Techniques

- Graph/Subgraph Index
  - Given a query graph, find all the graphs in a database that are subgraphs of the query graph
  - **ICDE'16**
- Shortest Path & Reachability Index
  - Build indexes for fast answering shortest path and reachability queries
  - **ICDE'12, VLDB Journal'12, VLDB Journal'13, TKDE'14**

Our Work—Efficient Algorithms

- Dense Subgraph Discovery
  - Given a large input graph, compute subgraphs that are dense
  - **SIGMOD'15, SIGMOD'13, KDD'15, Algorithmica'13**
  - One book in preparation: have signed a contract with springer publisher
- Optimal Algorithms
  - We propose worst-case optimal algorithms for specific problems, e.g., linear to the query size and result size
  - **SIGMOD'15, VLDB'15a, ICDE'16a, TKDE'17**

Our Work—Efficient Algorithms

- Efficient Algorithms for graph primitive operations
  - Subgraph Matching (**SIGMOD'16**)
  - Maximum Independent Set (**SIGMOD'17**)
  - Graph Edit Distance (under review)
  - Maximum Clique (in preparation)
  - …

Our Work—Computation Model

- I/O Efficient Algorithms
  - Study I/O efficient algorithms for computing SCCs and ECCs
  - **SIGMOD'13, VLDB'16, VLDB Journal'15, VLDB Journal'17**
- Multi-Core Algorithms
  - Multi-core algorithm for keyword query (**VLDB'10)**
- Distributed Algorithms
  - Scalable computation class, subgraph enumeration, (bi)connected components
  - **SIGMOD'14, VLDB'15b, VLDB'17, ICDE'16c, VLDB Journal'17**

Thank you!

Welcome to join me on the journey exploring big graphs!

Email: Lijun.Chang@sydney.edu.au
Homepage: lijunchang.github.io
Github: www.github.io/lijunchang