# Xiaofeng Yang

335 Huntington Ave, Boston, MA 02115
617-275-9020
xiaofeng@ccs.neu.edu
http://www.ccs.neu.edu/home/xiaofeng/

## Education

**Northeastern University, Boston**                    **2013-Present (GPA: 3.8/4)**
Ph.D in Computer Science

**Tsinghua University, Beijing, China**                    **2009-2013 (GPA: 3.7/4)**
B.S. in Applied Mathematics
B.Eng. in Electronic Engineering

## Research Projects

• **Any-k Tree Pattern Retrieval in Heterogeneous Networks**          **2016 —2017**
We proposed a novel algorithm for finding the top-k acyclic query patterns in large heterogeneous networks. In contrast to previous work, k is not given in advance. Instead, the user keeps requesting the top-ranked answers until s/he is satisfied with the result. I optimized and analyzed the algorithm (implemented in C++) and applied it to a variety of very large real datasets. This work was recently accepted for publication in The Web Conference 2018 (WWW 2018), the premier peer-reviewed research venue on future directions of the World Wide Web (acceptance rate below 20%).

• **Graph Sparsification for Social Network Analysis**          **2014 — 2015**
We explored graph sparsification algorithms that reduce query cost while (approximately) preserving multi-hop neighborhood structures in massive graphs. These algorithms can be used to compress the size of large social networks, improving the performance of popular queries that explore neighborhoods, e.g., properties of friends of users.

• **Detecting Hired Politically Active Groups on Sina Weibo**          **2013 — 2014**
We studied organized political propaganda on Sina Weibo. I applied NLP tools (LDA, hashing, Louvin clustering, sentiment analysis) to the crawled tweets and comments, clustered users based on Jacquard similarity, performed spammer detection and used Spark to compute results in parallel. This work was published at the 2015 AAAI Conference on Web and Social Media (ICWSM-15) under the title "Penny for Your Thoughts: Searching for the 50 Cent Party on SinaWeibo."

- **Twitter User Behavior Modeling** 2013

In this project, we had access to data collected from an Android app. I wrote a distributed crawler to send API requests through 60 IP addresses and collected the timeline data from Twitter. By studying user behavior, we could potentially measure information consumption and build statistical models to predict which messages a user would read.

- **A Locality Preserving Approach for Kernel PCA** 2013

We integrated a locality preservation constraint into the kernel PCA dimensionality reduction method. I implemented an algorithm proposed by our collaborators in Matlab, and applied it to recognize faces in a variety of human face datasets. This work was published at the 2015 International Conference on Image and Graphics (ICIG) in a paper titled "A Locality Preserving Approach for Kernel PCA."

# Work Experience

**Northeastern University, Boston** 2013-Present
Research Assistant, The Data Lab
Advisors: Christo Wilson, Mirek Riedewald

**HP Enterprise Vertica, Cambridge** 2017.6-2017.8
Software Engineering Intern, Executive Engine Team
Advisor: Nga Tran

**Nokia Bell Labs, Dublin** 2016.6-2016.9
Research Intern, Data Analytics group
Advisor: Alessandra Sala

**Chinese Academy of Sciences** 2012.6-2012.9
Intern, Institute of Automation and Pattern Recognition
Advisor: Xiangsheng Huang

# Programming Languages

**proficient**: Python, C++
**experienced**: Java, Matlab, Bash
**knowledge of**: Javascript, Scheme, Scala

# Awards

Northeastern Academic Excellency Scholarship 2013, 2014, 2015
1st prize, China Physics Competition of College Students 2013
1st prize, CPHO Chinese Physics Olympiad 2008

## Teaching

Teaching Assistant, Discrete Math and Data Structures, Northeastern University, Boston 2016, 2017—An introductory graduate course, taught using C to introduce data structure concepts.

Teaching Assistant, Parallel Data Processing in MapReduce, Northeastern University, Boston 2017—An advanced graduate course, taught using MapReduce, Spark and the Amazon cloud (AWS, EMR).