Nan Li, Ph.D.

Data Scientist

Core Data Science
Facebook Inc.

☎ *(805) 886-0846*
✉ *nanli@fb.com*
✉ *lilxine0416@gmail.com*
https://whoisnan.github.io
https://alumni.cs.ucsb.edu/~nanli/

Research Interests

Data mining, machine learning, graph mining, statistical modeling, text mining

Education

2008.9-2013.9 **Ph.D. in Computer Science**, *University of California Santa Barbara (UCSB)*.
◇ Research Areas: Data Mining, Graph Mining, Statistical Modeling
◇ Thesis Title: Uncovering Interesting Attributed Anomalies in Large Graphs
◇ Advisor: Prof. Xifeng Yan, `xyan@cs.ucsb.edu`
◇ GPA: **3.99 / 4.0**

2005.9-2008.7 **M.S. in Computer Science**, *Peking University (PKU)*.
◇ Research Areas: Data Mining, Financial Forecasting, Text Mining
◇ GPA: Overall, **88.8 / 100**, Major, **90.0 / 100** (Rank: **1 out of 36**)

2001.9-2005.6 **B.S. in Computer Science**, *Wuhan University (WHU)*.
◇ GPA: Overall, **90.0 / 100**, Major, **92.5 / 100** (Rank: **1 out of 452**)

Selected Publications

Conference Publications.

•Nan Li, Huan Sun, Kyle Chipman, Jemin George, and Xifeng Yan, "A Probabilistic Approach to Uncovering Attributed Graph Anomalies", *Proc. of the 2014 SIAM International Conference on Data Mining (SDM'14)*, Philadelphia, Pennsylvania, April 2014, accepted.

• Nan Li, Ziyu Guan, Lijie Ren, Jian Wu, Jiawei Han, and Xifeng Yan. gIceberg: Towards Iceberg Analysis in Large Graphs. *Proc. of the 2013 IEEE International Conference on Data Engineering (ICDE'13)*, pp. 1021-1032, Brisbane, Australia, April 2013.

• Nan Li, Xifeng Yan, Zhen Wen, and Arijit Khan. Density index and proximity search in large graphs. *Proc. of the 2012 ACM International Conference on Information and Knowledge Management (CIKM'12)*, pp. 235-244, Maui, HI, USA, October 2012.

• Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. *Proc. of the 2011 International Conference on Management of Data (SIGMOD'11)*, pp. 901-912, Athens, Greece, June 2011.

• Nan Li and Naoki Abe. Temporal cross-sell optimization using action proxy-driven reinforcement learning. *Proc. of the ICDM 2011 Workshop on Optimization Based Methods for Emerging Data Mining Problems (ICDMW'11)*, pp. 259-266, Vancouver, Canada, December 2011.

• Charu Aggarwal and Nan Li. On node classification in dynamic content-based networks. *Proc. of the 2011 SIAM International Conference on Data Mining (SDM'11)*, pp. 355-366, Phoenix, AZ, USA, April 2011.

• Nan Li, Yinghui Yang, and Xifeng Yan. Cross-selling optimization for customized promotion. *Proc. of the 2010 SIAM International Conference on Data Mining (SDM'10)*, pp. 918-929, Columbus, Ohio, USA, April 2010.

Journal Publications.

• Charu Aggarwal and Nan Li. On supervised mining of dynamic content-based networks. *Statistical Analysis and Data Mining*, 5(1):16–34, 2012.

• Nan Li and Desheng Dash Wu. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2):354–368, 2010.

• Nan Li, Xun Liang, Xinli Li, Chao Wang, and Desheng Dash Wu. Network environment and financial risk using machine learning and sentiment analysis. *Human and Ecological Risk Assessment*, 15(2):227–252, 2009.

Programming Skills

Languages	Java, Python, Hive, SQL, Pig, C++ (LEDA), C#.	**Statistical Tools**	R, MATLAB.
IDEs	Eclipse, MS Visual Studio.	**Platforms**	Linux, Mac OS, Windows.
Databases	PostgreSQL, MS SQL Server, MySQL.		

Research Experiences

2009.1-2013.9 **Research Assistant at Department of Computer Science, UCSB**,
Advisor: *Prof. Xifeng Yan.*
Topics: data mining, machine learning, graph mining, anomaly detection, social network analysis, business analytics and optimization

• *Graph anomaly detection based on statistical modeling*
a) gAnomaly: a regularized mixture model for anomaly detection in graphs.

• *Large-scale graph indexing and query processing*
Algorithms to efficiently and effectively index large graphs and answer queries are designed. Various types of graph indices and queries are studied.
a) gDensity: label-based proximity search via density indexing;
b) gIceberg: graph iceberg search via local aggregate scoring.

• *Optimal promotion planning*
A novel formulation of product promotion value and efficient approximation algorithms are designed, using rule mining, to select the optimal set of products and customers in order to maximize promotional profitability.

2005.11-2008.7 **Research Assistant at Department of Computer Science, PKU**,
Advisor: *Prof. Xun Liang.*
Topics: data mining, machine learning, text mining

• *Impacts of Web data on stock markets*
Correlations between online financial news and volatility exhibited by both stock price and trading volume time series are modeled.

• *Web sentiment analysis*
Correlations between text sentiment and online social network patterns are investigated in order to efficiently detect ongoing and forecast incoming events.

Work Experiences

2015.02-Present **Core Data Science, Facebook**, *Menlo Park, CA*,
Position: Data Scientist.
Areas: predictive modeling, time series modeling, natural language processing (NLP), multi-armed bandit algorithms

2014.05-2015.02 **Applied Machine Learning, Apple**, *Cupertino, CA*,
Position: Data Scientist.
Areas: recommender systems, anomaly detection, NLP, graph mining

2013.08-2014.05 **Data Products & Research, oDesk**, *Redwood City, CA*,
Position: Data Scientist.
Areas: classification/regression, clustering, NLP, Bayesian online learning
Various statistical models and machine learning tools are built to solve interesting data problems emerging in one of the largest online work marketplaces.

2012.12-2013.3 **Microsoft Research**, *Cambridge, UK*,
Position: Research Intern, Advisors: Milan Vojnovic, Bozidar Radunovic.
Areas: Bayesian online learning, graphical models, inference
• *User skill ranking and competition outcome prediction*
A probabilistic model is proposed to characterize and predict user behaviors for online crowd-sourcing services.

2012.6-2012.9 **Bing Indexing and Knowledge Team, Microsoft**, *Bellevue, WA*,
Position: Research and Software Development Intern, Advisor: Kang Li.
Areas: classification, NLP
• *Full-document entity extraction and disambiguation*
Given a knowledge base, the developed entity recognition system applies surface form spotting and entity disambiguation on the entire document.

2010.6-2010.9 **Customer Insight & Data Analytics Team, IBM Research**, *Yorktown Heights, NY*,
Position: Research Intern, Advisor: Naoki Abe.
Areas: Markov decision process, reinforcement learning
• *Lifetime value maximization using action proxy-driven reinforcement learning*
Customer lifetime value maximization is done by applying reinforcement learning to solve an MDP model. Action proxies are designed to cope with scenarios without the presence of historical action data.

2007.9-2007.12 **Business Intelligence Team, IBM Research**, *Beijing, China*,
Position: Research Intern, Advisor: Bo Li.
• *Connection network intelligence*
Inter-company relationships, transactions and other financial information are conglomerated into a network, on which various queries can be studied.

2006.10-2007.4 **Autonomic Middleware & Service Delivery Team, IBM Research**, *Beijing, China*,
Position: Research Intern, Advisor: Xinhui Li.
• *CUDA resource management project for Java platform*
A review of Java Virtual Machine (JVM), including dynamic class loading, link-time verification, method dispatching, etc.

Teaching Experience

2011.6-2011.8 **Research mentor for Internships in Nanosystems Science, Engineering and Technology (INSET)**.

2008.9-2009.12 **Teaching Assistant at UCSB**.
• CS263 (graduate course): Modern Programming Languages and Their Implementations, Winter 2009, Fall 2009, Prof. Chandra Krintz.
• CS30: Introduction to Computer Systems, Fall 2008, Prof. Heather Zheng.
• CS20: Programming Methods, Spring 2009, for Professor Jianwen Su.

Professional Activities

Speaker/Attendee **Workshops**.
• 2012 Grace Hopper Celebration of Women in Computing, Baltimore, MD, Oct 3-6, 2012.
• Invited speaker at 2009 Google Workshop for Women Engineers, Mountain View, CA, Jan 22-25, 2009.
• 2009 Grad Cohort Program, San Mateo, CA, Mar 27-28, 2009.

Reviewer **Journals**.
IEEE Transactions on Neural Networks, Journal of Neurocomputing.

Conferences.
VLDB'14, KDD'13, WWW'13, ICDM'12, SIGMOD'11, SDM'11, KDD'10, SDM'10, SIGMOD'10, ICDM'10, ICDM'09, ICDE'09, ISNN'07.

Honors and Awards

2012 2012 CIKM Student Travel Grant

2012 2012 Grace Hopper Scholarship

2011 2011 SDM Conference Travel Award

2010 2010 SDM Conference Travel Award

2008-2009 UCSB Department of Computer Science Merit Fellowship

UCSB Department of Computer Science Teaching Assistantship

2006 PKU "DongShi DongFang" Scholarship for Outstanding Students

2004 WHU "Huawei" Scholarship for Outstanding Students

2001-2004 WHU Scholarships of WHU for Outstanding Students

WHU Merit Students of Excellence of WHU