Xiaohan Zhao

Dept. of Computer Science
University of California, Santa Barbara
Santa Barbara, CA 93106

Mobile: 805-886-1604
Email: xiaohanzhao@cs.ucsb.edu
homepage: http://cs.ucsb.edu/∼xiaohanzhao

Research Interests

I am interested in analyzing, processing, and modeling large complex networks, such as online social networks, mobile networks, data center networks, communication networks, information networks, computer networks, and collaborative networks. I am also working on privacy and security problems in large complex networks.

Education

Ph.D. in Computer Science (Advisor: Ben Y. Zhao) — Aug. 2009 - Dec. 2014
University of California, Santa Barbara

M.S. in Electronic Engineering (Advisor: Beixing Deng) — Sept. 2007 - Jul. 2009
Tsinghua University, China

B.E. in Electronic Engineering — Aug. 2003 - Jul. 2007
Tsinghua University, China

Work Experience

LinkedIn, Mountain View, CA — June 2012- Sept. 2012
Software Engineering Intern, work with **Chris Conrad**

- Analyze, design, and implement efficient algorithms to search and rank all connections between users.

Microsoft Research, Cambridge, UK — Oct. - Dec. 2012 & Jun. - Sept. 2013
Research Intern, worked with **Dr. Antony Rowstron**

- **Automatic topology generation for fabric computing systems**. Fabric computing systems are complete self-contained rack-scale clusters that support general-purpose commodity data center workloads. By connecting the switches co-located with processors in the fabric system, users can customize network topologies, such as 2D or 3D Torus. In this project, we analyze the performance of various workloads on a wide range of fabric topologies, and study the workload assignment problem on fixed fabric topologies. Through our simulation, we find that no single topology performs best, and workload optimization on a given topology provides limited improvement. To optimize the performance for a given workload, we propose a practical solution to customize the topology of a fabric computing system.

Academic Experience

UCSB SAND Lab, Santa Barbara, CA
Postdoc, work with **Prof. Ben Y. Zhao** — Dec. 2014 -Present
Research Assistant, work with **Prof. Ben Y. Zhao** — Aug. 2009- Dec. 2014

- **Massive Graph Processing Systems.** Node distance is a fundamental metric for graph analysis and network applications. Important node distances, such as shortest path and random walk distances, are computationally expensive and difficult to scale on big real graphs. To address the scalability problem, we propose an efficient and practical solution – *graph coordinate systems*. Our systems can accurately estimate node distances in microseconds. The following are the designed systems.
  - *Rigel & Orion: efficient shortest path computation on large graphs*
  - *Leo: a graph coordinate system for random walk distances*
- **Dynamic Graph Analysis and Modeling.**

- *Analyze dynamics in a large online social network at multiple scales*
  A number of interrelated processes drive dynamics in social networks, while many of the prior studies focused on capturing dynamics as a single process. In this work, we are interested in the question "how are individual user dynamics influenced by processes at different scales?" To answer this question, we conduct detailed analysis on dynamics in the largest Chinese online social network at multiple scales, including individual user level, user community level, and global network level.
- *Self-similarity analysis on dynamic graphs*
  This work aims to model both structural and temporal properties in dynamic social networks. Our work is inspired by self-similarity analysis on network traffic. Self-similarity refers to that the relative variance or volatility of traffic traces stays similar across different time scales. The existence of self-similarity in social network dynamics will fundamentally change the way we view and consider dynamic graph models. In this work, we detect the presence of self-similar properties in graph dynamics, and propose a dynamic graph model that captures both temporal and structural properties of network evolution.

- **Graph Privacy and Security.**
  - *Pygmalion: differential privacy in graphs*
    Big real graphs usually contain sensitive information. Sharing these graphs often raise the risks of leaking users' private data. Current anonymization techniques only address the privacy problems caused by specific attacks, and are vulnerable to powerful de-anonymization attacks. Our work seeks a solution to sharing meaningful graph datasets while preserving privacy. To navigate the tradeoff between strength of privacy and graph structural similarity, we design Pygmalion, a *differentially-private graph model*.
  - *Graph watermarks*
    The owners of big real graph data are facing a real challenge: how to share sensitive graphs with collaborators and authorized users. Current solutions provide limited privacy, but require modifications to graph structure. In this work, we propose a novel solution, *graph watermarks*, which are small unique graphs tailored for a given graph and a secure user key, serving as a deterrent against data leakage. We provide robust schemes to create, embed, and extract watermarks. We also develop techniques to defend against potential attacks.

- **Analysis of Popular Complex Networks.**
  - *Crowdturfing networks*
    Crowdsourcing networks provide a new power to solve complex problems. However, the success of the crowdsourcing networks also poses a real challenge to existing security mechanisms: malicious activities are generated by human beings instead of automated programs. In this work, we study malicious crowdsourcing networks, referred to as *crowdturfing* networks. Through measurement results from our detailed data collected from these crowdturfing networks, we find surprising evidence showing that not only do these systems exist in a number of countries, but also they are rapidly growing in both user base and total revenue.
  - *Geosocial networks* (Short paper in HotNets-XII & Full paper in submission)
    Mobile networking researchers have long searched for large-scale realistic mobility traces, which have remained elusive for both privacy and logistical reasons. Recently, researchers have begun to use geosocial mobility traces, such as Foursquare checkin traces, because of their availability and scale. But are these traces reflective of users' true mobility patterns? In this work, we answer this question by performing a large user mobility study, and comparing a ground-truth of user mobility (via GPS data) to a Foursquare dataset for the same users. Our measurement results

show that 75% of Foursquare checkin events are extraneous checkins generated by users to achieve in-system rewards, and checkin events only capture 10% of user actual visited locations. Based on this observation, we develop a method to detect extraneous checkins, and design an algorithm to recover missing locations.

Press Coverage

A dark force, unleashed online
Boston Globe, Jan 8, 2012.

Million Dollar Crowdturfing Industry Dupes Social Networks
Slashdot, Dec. 13, 2011.

Growing Number Of Cyber Shills Invade Online Reviews
The Consumerist, Dec. 13, 2011.

Cyber shill business is booming
InfoWorld, Dec. 13, 2011.

Hidden Industry Dupes Social Media Users
Technology Review, Dec. 12, 2011 .

Honors

Regents Special International Fellow, UCSB 2009 - 2010

Publications

Matteo Zignani, Sabrina Gaito, Gian Paolo Rossi, **Xiaohan Zhao**, Haitao Zheng, and Ben Y. Zhao. Link and Triadic Closure Delay: Temporal Metrics for Social Network Dynamics. *International Conference on Weblogs and Social Media (ICWSM), 2014*.

Zengbin Zhang, Lin Zhou, **Xiaohan Zhao**, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao. On the Validity of Geosocial Mobility Traces. *ACM Workshop on Hot Topics in Networks (HotNets), 2013.*

Xiaohan Zhao, Adelbert Chang,Atish Das Sarma, Haitao Zheng, and Ben Y. Zhao. On the Embeddability of Random Walk Distances. *Very Large Data Bases (VLDB), 2013*.

Xiaohan Zhao, Alessandra Sala, Christo Wilson, Xiao Wang, Sabrina Gaito, Haitao Zheng, and Ben Y. Zhao. Multi-scale Dynamics in a Massive Online Social Network. *Internet Measurement Conference (IMC), 2012*

Sabrina Gaito, Matteo Zignani, Gian Paolo Rossi, Alessandra Sala, **Xiaohan Zhao**, Xiao Wang, Haitao Zheng, and Ben Y. Zhao. On the Bursty Evolution of Online Social Networks. *ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research (HotSocial), 2012*.

Gang Wang, Christo Wilson, **Xiaohan Zhao**, Yibo Zhu, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. Serf and Turf: Crowdturfing for Fun and Profit. *International World Wide Web Conference (WWW), 2012*.

Xiaohan Zhao, Alessandra Sala, Haito Zheng, and Ben Y. Zhao. Efficient Shortest Paths on Massive Social Graphs. *IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2011*. (Invited Paper)

Alessandra Sala, **Xiaohan Zhao**, Christo Wilson, Haitao Zheng, and Ben Y. Zhao. Sharing Graphs using Differentially Private Graph Models. *Internet Measurement Conference (IMC), 2011*.

Xiaohan Zhao, Alessandra Sala, Christo Wilson, Haitao Zheng, and Ben Y. Zhao. Orion: Shortest Path Estimation for Large Social Graphs. *Workshop on Online Social Networks (WOSN), 2010*.

Yang Chen, Xiao Wang, Xiaoxiao Song, Eng Keong Lua, Cong Shi, **Xiaohan Zhao**, Beixing Deng, and Xing Li. Phoenix: Towards an Accurate, Practical and Decentralized Network Coordinate System. *International IFIP-TC6 Networking Conference (NETWORKING), 2009.*

Xiaohan Zhao, Xiaoxiao Song, Xiao Wang, Yang Chen, Beixing Deng, and Xing Li. Analysis of Security Policy in Practical Internet Coordinates. *International Journal of Security and Its Applications, Vol.3, No.1, 2009.*

Xiaoxiao Song, **Xiaohan Zhao**, Eng Keong Lua, Zengbin Zhang, Beixing Deng, and Xing Li. SLINCS: A Social Link based Evaluation System for Network Coordinate Systems. *IEEE Consumer Communications & Networking Conference (CCNC), 2009*. (Short Paper)

Xiaohan Zhao, Xiaoxiao Song, Xiao Wang, Yang Chen, Beixing Deng, and Xing Li. Attacks against Network Coordinate System: Vulnerable PIC. *International Symposium on Computer Science and its Applications (CSA), 2008*.