

Zifei SHAN

28-306, Yiheyuan Road No.5
Beijing, China, 100871

zifeishan@gmail.com

Tel: (+86) 152-1062-6373
<http://www.zifeishan.org>

Research Objective

- My research interests are data mining and large social networks, in particular network dynamics and robustness, and human behaviors in real-world settings.

Academics

- **Peking University** Beijing, China
B.S. Candidate in Computer Science Sept. 2009 – Jun. 2013
 - GPA: 3.82/4 (CS major); 3.63/4 (overall), ranked 25/150 (department)
- **Technion—Israel Institute of Technology** Haifa, Israel
Research Exchange Student in Faculty of Electrical Engineering Sept. 2012 – Feb. 2013

Research

- **Institute of Network Computing and Information Systems, Peking University** Beijing, China
Advised by Prof. Yafei Dai Jun. 2011 – Present
 - Collaborated a research project on *Sybil Attacks in Online Social Networks (OSNs)*, supervised by Jing Jiang, in cooperation with Renren.com, the largest OSN in China. Detected ~1M sybil accounts and found sybil groups in Renren, based on the attribute of social degree and low popularity. Evaluated the detector by action time similarity. Done measurements on user statistics, relationship between communities, and network evolution including increasing and merging pattern.
 - Involved in *Assessing the Impact of User-interaction Transparency on Renren's User Behavior*. Constructed latent interaction graphs of user browsing behavior, and compared their structural properties against those of both visible interaction graphs and social graphs. Analyzed their conductance, modularity and mixing time.
 - Led independent research on *Profile-Cloning Attacks in OSNs*. Improved the attack pattern by snowball sampling and iteration attack. Conducted experiments in Renren to test its threats. Raised a back-end defending architecture called *CloneSpotter*, based on IP prefix. Discussed further defenses.
 - Led independent research on *Ranking and Analyzing Baseball Networks*. Raised a ranking algorithm named *GameRank*, for networks with multiple interplaying indicators. Evaluated the algorithm by comparing our rankings with ESPN Ratings, on Major League Baseball (MLB) data. Mined the MLB network on its evolution and pitchers' batting ability. Conducted visual analysis on the communities and bridges in the network.
- **Technion-Peking Research-Exchange Program** Haifa, Israel
Advised by Prof. Daniel A. Freedman Feb. 2012 – Present
 - Co-leading a long-term research project on the intersection of Programming Language, Systems, and Human Computer Interaction. Aiming at designing description languages and their compilers, to automate the creation of back-end services in various network architectures, and front-end user interfaces in heterogeneous platforms.

Peer-Reviewed Publications

AVAILABLE AT: <http://www.zifeishan.org>

- Jing Jiang, **Zifei Shan**, Wenpeng Sha, Xiao Wang, and Yafei Dai. Detecting and Validating Sybil Groups in the Wild. In *Proc. 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW '12)*.
- **Zifei Shan**, Shiyongxue Li, and Yafei Dai. GameRank: Ranking and Analyzing Baseball Network. To appear in *2012 ASE International Conference on Social Informatics*. (Acceptance rate: 11.5%)
- **Zifei Shan**, Haowen Cao, Jason Lv, Cong Yan, and Annie Liu. Enhancing and Identifying Cloning Attacks in Online Social Networks. To appear in *ACM International Conference on Ubiquitous Information Management and Communication 2013 (ICUIMC '13)*. (Acceptance rate: 29%)

Manuscripts in Submission

- Jing Jiang, **Zifei Shan**, Xiao Wang, and Yafei Dai. Detecting, Validating and Measuring Sybil Groups in the Wild, submitted to *ACM Transactions on Knowledge Discovery from Data (TKDD)*.

Course / School Projects

DATA MINING AND VISUALIZATION:

- **MLB illustrator: visualizing and ranking baseball networks** <http://mlbillustrator.com/>
Course project of SI 508, advised by Prof. Qiaozhu Mei (University of Michigan) Aug. 2012 – Sept. 2012
 - Visualized the MLB game data as a heterogeneous network, provided statistics, and ranked the batting and pitching ability of players, for all historical data over 50 years. Built an online visualization system. Used this system for visual analysis in baseball networks, in my independent research (see *GameRank* above).
- **Question answering system on Chinese Wikipedia** Oct. 2011 – Dec. 2011
Team leader; designed QA algorithm using Chinese NLP techniques.
- **WordNet viewer featuring force-driven graph of words** May. 2011 – Jun. 2011
Designed the dynamic relationship graph with an originated spring-model layout algorithm.

SYSTEMS:

- **Beijing 3-Day Startup Website** <http://www.beijing3ds.org/>
Global startup event & school project in PKU and THU. Website back-end designer. Aug. 2012 – Oct. 2012
 - Set up the web server, email server, and application form handler.
 - Handled verification, text extraction, PDF generation, and email notification using local email service; successfully processed all the applications in Beijing3DS event.
- **Kademlia network distributed simulation** Oct. 2010 – Jan. 2011
Team leader; implemented a UDP-based P2P network using Kademlia DHT.
- **Minijava compiler for Android** Sept. 2011 – Dec. 2011
Worked on optimization; realized general optimizations based on dataflow analysis.

GAMES:

- **AI for game “Blokus” generated by Genetic Algorithm** May. 2010 – Jun. 2010
Designed game AI; used GA to refine arguments for AI; ranked top 20% in department.
- **2D Shooting Game programmed with Haaf’s Game Engine** Jan. 2010 – Mar. 2010
Sole developer; used C++ Object Oriented programming to design stages and barrages.

Honors

- **National Scholarship (Rank: 1/35)**, awarded twice, ranking first in both 2011 and 2012
- **Kwang-Hua Scholarship (Rank: 3/36)**, Kwang-Hua Education Foundation Dec. 2010
- **Meritorious Winner of Interdisciplinary Contest in Modeling (ICM)** Feb. 2012
- **The CCF Outstanding Undergraduate Award**, China Computer Federation Oct. 2012
- **Outstanding Student Award**, Peking University 2010, 2011 and 2012
- **Honor Track of Computer Science Completion**, Peking University Apr. 2012
 - Practice for Compiler Design, ranked as Top 5%; Data Structure and Algorithms(A), ranked as Top 10%; Practice of Programming in C&C++, ranked as Top 10%; Introduction to Computing(A).

Technical and Language Skills

- **Programming languages:** C/C++, Java, JavaScript, Python, PHP, HTML, SQL
- **Unix tools:** Bash scripting, MakeFile, Git, NGINX
- **Manuscript preparation:** L^AT_EX, Matplotlib, gnuplot, Inkscape
- **Miscellaneous:** Hadoop, OpenMP, MPICH2, D3.js, Prefuse, Gephi, ANTLR, StringTemplate, C++ Boost
- **English proficiency:** TOEFL 104 (R26, L26, S23, W29); GRE V400, Q800, AW3.5.

Personal

- **Captain of EECS Baseball Team**, Peking University 2010 – 2012
- **Student class leader**, Peking University 2009 – Present
- **Other hobbies:** the Game of Go, music, photography, cooking, studying Chinese tea.