Shan Jiang | Curriculum Vitae

660-674, Interdisciplinary Science & Engineering Complex, 805 Columbus Ave, Boston, MA 02120
(+1) 781-502-8799 | sjiang@ccs.neu.edu | shanjiang.me | printfoo | shan-jiang

Education

Northeastern University — Boston, MA
Ph.D. in Computer Science — *Sep 2016 - 2021 (Expected)*
- Advisor: Christo Wilson | GPA: 3.9/4.0

Beijing University of Posts and Telecommunications — Beijing, China
B.B.A. in Management Information Systems — *Sep 2012 - Jul 2016*
- Rank: 1/46 | GPA: 92.5/100

Experience

Google — New York, NY
Software Engineering Intern @ Fact Check Team, Google AI — *Jun 2019 - Aug 2019*
- Research on structured understanding of fact-check articles.

Dataminr — New York, NY
Research Intern @ AI and Data Science Team — *Feb 2019 - Apr 2019*
- Research on crisis sub-events detection for emergency management;
- Applied NLP methods on the Twitter firehose, e.g., BERT as a service for sentence embedding, dependency parsing and tree traversal.

Northeastern University — Boston, MA
Research Assistant @ Khoury College of Computer Sciences — *Sep 2016 - Present*
- Research on computational social science, misinformation and fact-checking, algorithmic bias and accountability;
- Utilized Spark/Hadoop frameworks to build analytical pipelines for TB-sized datasets;
- Applied statistical and causal models, mixed with machine learning flavors, for hypothesis testing on observational data;
- Paper published at top-tier web and HCI conferences, e.g., WWW, ICWSM, CSCW, FAT*.

National University of Singapore — Singapore
Research Assistant @ School of Computing — *Dec 2015 - May 2016*
- Research on economic modeling of bitcoin mining under risk aversion.

Beijing University of Posts and Telecommunications — Beijing, China
Research Assistant @ State Key Lab of Networking and Switching Technology — *Oct 2013 - Dec 2015*
- Research on game-theoretic modeling of overlay networks and traffic engineering;
- Paper published at network and system conferences, e.g., GlobeCom, LCN, ICPADS.

Publications

Bias Misperceived: The Role of Partisanship and Misinformation in YouTube Comment Moderation — ICWSM'19
Shan Jiang, Ronald E Robertson, and Christo Wilson — *Acceptance Rate: TBA*

Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation — WebSci'19
Ronald E Robertson, Shan Jiang, David Lazer, and Christo Wilson — *Acceptance Rate: TBA*

Auditing the Partisanship of Google Search Snippets — WWW'19
Desheng Hu, Shan Jiang, Ronald E Robertson, and Christo Wilson — *Acceptance Rate: 18.0%*

Who's the Guinea Pig? Investigating Online A/B/n Tests in-the-Wild — FAT*'19
Shan Jiang, John Martin, and Christo Wilson — *Acceptance Rate: 24.1%*

Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media — CSCW'18a
Shan Jiang, and Christo Wilson — *Acceptance Rate: 25.6%*

Auditing Partisan Audience Bias within Google Search — CSCW'18b
Ronald E Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson — *Honorable Mention* | *Acceptance Rate: 25.6%*

On Ridesharing Competition and Accessibility: Evidence from Uber, Lyft, and Taxi — WWW'18
Shan Jiang, Le Chen, Alan Mislove, and Christo Wilson — *Acceptance Rate:14.8%*

Conflicts in Overlay Environments: Inefficient Equilibrium and Incentive Mechanism — KSII-TIIS'16
Jianxin Liao, Jun Gong, Shan Jiang, Tonghong Li, and Jingyu Wang — *Impact Factor: 0.611*

Interactions of Overlays and Traffic Engineering: Equilibrium and Cooperation without Payment — GlobeCom'15
Shan Jiang, Jun Gong, Jingyu Wang, Jianxin Liao, and Tonghong Li — *Acceptance Rate: 35.0%*

Competitive Equilibrium and Stable Coalition in Overlay Environments — LCN'15
Shan Jiang, Jianxin Liao, Jun Gong, Jingyu Wang, and Tonghong Li — *Acceptance Rate: 30.3%*

Combination Feature for Image Retrieval in the Distributed Datacenter ICPADS'14
Di Yang, Jianxin Liao, Qi Qi, Jingyu Wang, Haifeng Sun, and Shan Jiang *Acceptance Rate: 29.8%*

Selected Projects

Crisis Sub-Events Detection for Emergency Management *Feb 2019 - Present*

Sub-events detection, e.g, building collapsed, road closed, after major events, e.g., wildfire.

- Filtered Twitter firehose using SQL queries and collected Tweets on major crisis events, e.g., wildfires, hurricanes;
- Parsed Tweets to dependency trees and traversed trees to extract connected noun-verb pairs as sub-events, e.g, building collapsed;
- Used BERT as a service to generate sentence embedding and clustered sub-events of similar semantic meanings;
- Mapped Tweets with sub-events and studied the temporal cascading of sub-events.

Is YouTube's Content Moderation Biased, or Not? *Jan 2018 - Mar 2019*

The claim of content moderation being biased against conservatives is but a misperception from correlation to causation.

- Built a dataset of the ecosystem surrounding YouTube, including video veracity, political leaning, user engagement for 80K+ comments;
- Performed statistical tests to show the difference in moderation likelihood for user comments under left- and right- leaning videos;
- Used a causal model (propensity score matching) to show that above difference is not caused by political leaning but other confounders;
- Simulated model dynamics under a variety of hypotheses for robustness checks;
- A paper published at ICWSM'19.

How do "Fake News" and Fact-Checking Affect People? *Nov 2017 - Nov 2018*

Social media users use more emojis and swear words under misinformation. Fact-checking has both corrective and "backfire" effects.

- Collected 5K+ fact-check articles from Snopes and PolitiFact, and 2M+ comments from Facebook, Twitter and YouTube;
- Built a topical lexicon ComLex using a hybrid method of unsupervised learning (word2vec, spectral clustering) and human evaluation;
- Performed statistical tests to show different word usage in user comments for truthful/fake news and before/after fact-check;
- Built predictive models to show that such difference in user comments can help with fake news detection;
- A paper published at CSCW'18a.

Do Google's Search Engine Result Pages Have Partisan Bias? *Sep 2016 - Nov 2018*

Search results show consistent bias with input queries, and no significant evidence for "filter bubbles" on political ideology.

- Recruited 200+ participants to install browser extensions that enabled us to collect search data from their computers;
- Calculated partisan bias score based on a dataset of 100M+ Tweets using Apache Spark;
- Performed statistical tests to show the correlation between partisan bias and rankings in Google's search engine result pages;
- A paper published at CSCW'18b, a visualization system available at polarshare.shanjiang.me.

Are Ridesharing Services Equally Accessible? *Sep 2016 - Apr 2018*

The quality of Uber and Lyft's services worsen in high-diversity areas in San Fransisco and low-income areas in New York City.

- Intercepted Uber and Lyft's mobile traffic using man-in-the-middle (MITM) proxy and built structured requests for data collection;
- Implemented crawlers to collect driver's trajectory data from Uber and Lyft in San Fransisco and New York City for 2 months;
- Analyzed 10TB+ data using Apache Spark to discover spatio-temporal patterns of ridesharing services;
- Used a spatial econometric model to show the inequality of ridesharing accessibility;
- A paper published at WWW'18, a report published by SFCTA, a visualization system available at tncstoday.sfcta.org.

Honors and Awards

Honorable Mention	for top 2.7% (30/1,106) papers at CSCW, awarded for CSCW'18b	*2018*
Graduate Fellowship	for first-year Ph.D. students at Northeastern University	*2016-2017*
Outstanding Undergraduates	for top undergraduate students in the city of Beijing	*2016*
National Scholarship	for top 1% students at Beijing University of Posts and Telecommunications	*2014-2015*
First-Class Scholarship $\times$**2**	for top 1% students at Beijing University of Posts and Telecommunications	*2013-2014, 2015-2016*

Skills

Languages	Python, Java, C/C++, R, Matlab, SQL, HTML/CSS, JavaScript
Platforms	Apache Spark, Apache Hadoop, TensorFlow, PyTorch, Vega/Vega Lite

Service

Reviewer 2020: CSCW, ICWSM | 2019: CSCW, ICWSM, CHI | 2018: CSCW, WWW