Shan Jiang

Social Computing · Computational Social Science · Human Computer Interaction · Online Misinformation · Algorithmic Audit
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

Ph.D. in Computer Science — Expected 2021
Northeastern University — *Boston, MA*
- Advisor: Christo Wilson

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

Publications

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'18
Shan Jiang, and Christo Wilson — *Acceptance Rate: 25.6%*

Auditing Partisan Audience Bias within Google Search — CSCW'18
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

Is YouTube's Content Moderation Biased, or Not? — *Jan. 2018 - Sep. 2018*
The claim that moderation is biased against conservatives is proven to be a misperception from correlation to causation. | *Submitted to* CHI'19.
- Collected a comprehensive dataset of the misinformation ecosystem surrounding YouTube, including veracity, bias, engagement, and 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 the above difference is not caused by political leaning but other confounders;
- Simulated model dynamics under a variety of hypotheses for robustness checks.

How do "Fake News" and Fact-Checking Affect People? — *Nov. 2017 - Aug. 2018*
People get touchy when commenting on "fake news" and misinformation, but also touchy about the truth. | *Published at* CSCW'18.
- Implemented crawlers to collect fact-check articles from Snopes and PolitiFact, and user 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 usages 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.

Do Google's Search Engine Result Pages Have Partisan Bias? — *Sep. 2016 - Aug. 2018*
Top-rank search results show less left-leaning bias than low-rank ones. | *Published at* CSCW'18.
- 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 100+ million Tweets using Apache Spark; (Visualization: polarshare.shanjiang.me)
- Performed statistical tests to show the correlation between partisan bias and rankings in Google's search engine result pages.

Is Ridesharing Services Equally Accessible? — *Sep. 2016 - Apr. 2018*
The quality of Uber and Lyft's services worsens in neighborhoods with high minority ratios or low incomes. | *Published at* WWW'18.
- Intercepted Uber and Lyft's mobile traffic using man-in-the-middle 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 spatiotemporal patterns of ridesharing services; (Visualization: tncstoday.sfcta.org)
- Used a spatial econometric model to show the inequality of ridesharing accessibility.

Miscellaneous

Skills Python, Java, Javascript, C/C++, Matlab, R, SQL, etc. | Apache Spark, Linux, Vega Lite, etc.
Reviewer CHI'19, CSCW'18, WWW'18 (external)