

Ke Yang

PHD CANDIDATE, COMPUTER SCIENCE, NEW YORK UNIVERSITY

Visualization and Data Analytics Research Center (VIDA)

370 Jay Street

Brooklyn, NY, 11201

ky630@nyu.edu | keyang0923@gmail.com

Webpage : keyang0923.github.io

+1-212-900-5854

EDUCATION

Ph.D. in Computer Science, New York University, New York, U.S.A

Advisor: Julia Stoyanovich

GPA 3.89/4.0

01.2019 - 03.2021

Ph.D. in Computer Science, Drexel University, Philadelphia, U.S.A

Advisor: Julia Stoyanovich

GPA 3.98/4.0

09.2015 - 12.2018

M.S.E in Computer Science, Beijing Technology and Business University, Beijing, China

Advisors: Qian Mo and Zhongming Han

09.2012 - 06.2015

Graduated summa cum laude, GPA 3.51/4.0

B.Eng. in Software Engineering, Beijing Technology and Business University, Beijing, China

Advisor: Qian Mo

09.2008 - 06.2012

Graduated summa cum laude, GPA 3.45/4.0

RESEARCH INTERESTS

Responsible data management, non-discriminatory machine learning, and data ethics
My research focuses on several quality aspects of algorithmic outcome and how to quantify and ensure them in different scenarios. The quality aspects include fairness, diversity, interpretability, and stability of algorithmic outcome. I develop algorithms and systems to ensure high-quality outcome outputted from a data-driven system.

Current project focuses on quantifying fairness in rankings and interpretability of algorithmic outcome.

RESEARCH PROJECTS

Data, Responsibly

Supervisor: Julia Stoyanovich

12.2015 - Present

- Interpret causal intersectional fairness in ranking-related applications. Details in [arXiv paper](#).
- Study fairness-enhancing interventions when treat data as a first-class citizen. Details in [EDBT paper](#) and [FairPrep on Github](#).
- Instrument fairness in machine learning pipelines. Details in [HILDA paper](#) and [fairDAGs demo](#).
- Guarantee diverse and fair representation of groups of candidates in ranking-related applications and online decision making. Details in [IJCAI paper](#) and [EDBT paper](#).
- Measure fairness in ranking-related applications and design models to mitigate bias against demographic minorities. Details in [SSDBM paper](#).
- Interpret algorithmic outcomes for ranking-related applications. Details in [SIGMOD paper](#) and [RankingFacts demo](#). - More project details in dataresponsibly.github.io.

Detect spammers in online shopping website

Supervisors: Zhongming Han and Qian Mo

05.2013 - 05.2015

- Study demographic and behavioral features of the groups of spammers that are either robot accounts or accounts under manipulation in Chinese online shopping websites.
- Design graphical models to feature users' interaction within groups members and to detect the groups of spammers.
- Publish a survey paper in Chinese Journal of Software, one of the Chinese premier peer-reviewed journals in Computer Science.
- Publish a paper in Chinese Journal of Computers, one of the Chinese premier peer-reviewed journals in Computer Science.
- Publish a China invention patent for the methodology to detect the groups of spammers in online shopping websites.

Detect and predict hot topics in social network through analyzing human behavior pattern

Supervisor: Zhongming Han

12.2012 - 06.2013

- Study demographic and behavioral features of users in Sina Weibo (Chinese version of twitter).
- Design a probabilistic graphical model that outperforms the baseline SVM model to detect hot topics in Sina Weibo.
- Publish a paper in Journal of Embedded System.

PUBLICATIONS

Ke Yang, Joshua R. Loftus, and Julia Stoyanovich, “Causal intersectionality for fair ranking”, under submission

Ke Yang, Biao Huang, Julia Stoyanovich, and Sebastian Schelter, “Fairness-Aware Instrumentation of Preprocessing Pipelines for Machine Learning”, in *Proceedings of the 4th Workshop on Human-In-the-Loop Data Analytics, in conjunction with the 2020 ACM SIGMOD/PODS Conference (HILDA@SIGMOD 2020)*, Portland, OR, USA

Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich, “Balanced ranking with diversity constraints”, in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, Macao, China, DOI:10.24963/ijcai.2019/836

Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Jerome Miklau, “A Nutritional Label for Rankings” (demo) to appear in *Proceedings of the 2018 ACM International Conference on Management of Data (SIGMOD 2018)*, Houston, USA

Julia Stoyanovich, **Ke Yang**, HV Jagadish, “Online Set Selection with Fairness and Diversity Constraints” in *Proceedings of the 21th International Conference on Extending Database Technology (EDBT 2018)*, Vienna, Austria, DOI:10.5441/002/edbt.2018.22

Ke Yang, Julia Stoyanovich, “Measuring Fairness in Ranked Outputs” in *Proceedings of the 29th International Conference on Scientific and Statistical Database Management (SSDBM 2017)*, Chicago, USA, DOI:10.1145/3085504.3085526

Zhongming Han, **Ke Yang**, Xusheng Tan, “Analyzing Spectrum Features of Weight User Relation Graph to Identify Large Spammer Groups in Online Shopping Websites” in *Chinese Journal of Computers*, 40(4): 939-954 (2017) (in Chinese), DOI:10.11897/SP.J.1016.2017.00939

Zhongming Han, **Ke Yang**, Fengmin Xu, Dagao Duan, “Probabilistic Graphical Model for Detecting Spammers in Microblog Websites” in *Journal of Embedded System*, 8(1): 12-23 (2016), DOI:10.1504/IJES.2016.073747

Qian Mo, **Ke Yang**, “Overview of Web spammer detection” in *Journal of Software*, 25(7): 1505-1526 (2014) (in Chinese), DOI:10.13328/j.cnki.jos.004617

Kai Wang, **Ke Yang**, “A method and application system of detecting web spammers” in *China Invention Patent*, (2015) (in Chinese), No. CN201510012860

TEACHING EXPERIENCE	<p>Research advisor at New York University</p> <ul style="list-style-type: none"> - Independent study: Biao Huang <i>01.2020 - 05.2020</i>, Jensine Raihan <i>09.2020 - current</i> - Research assistant: Samasth Norway Ananda <i>03.2020 - 08.2020</i> <p>Teaching Assistant at Drexel University</p> <ul style="list-style-type: none"> - CS500 Fundamentals of Databases <i>01.2017 - 03.2017</i> - CS461 Database Systems <i>03.2017 - 05.2017</i>
WORK EXPERIENCE	<p>Research intern at AT&T Lab Research (New York Office) <i>06.2019 - 08.2019</i></p> <p><i>Supervisors: Emily Dodwell, Ritwik Mitra, and Balachander Krishnamurthy</i></p> <p>Worked on a project to insure fairness and transparency for internal AT&T machine learning projects. This project aims to ensure that the outcome of a machine learning project meets reasonable guidelines for fairness and does not include illegal bias and discrimination. During the internship, we developed a tool for project managers and data scientists to check whether there are bias-related issues through the whole pipeline of machine learning.</p>
	<p>Research engineer at Elite & Resource (start-up company) <i>11.2014 - 08.2015</i></p> <p><i>Supervisor: Peng Sun</i></p> <p>Worked on a project to analyze historical flood disaster data and to predict a specific characteristic that can help with the prevention of future flood. In this project, we proposed a model to predict the water flow rate of small watershed torrents that can be used in the prevention of future flood. Our model is integrated as a core component of national watershed data management system.</p>
SOFTWARE PROJECTS	<p>Ranking Facts: full-stack development using Python, Django, JavaScript, and JQuery.</p> <p>Data Synthesizer: front-end development using JavaScript and JQuery.</p> <p>Researcher Questionnaire: database design and front-end development using PHP, Drupal, and MySql.</p> <p>DataResponsibly: website development using Jekyll.</p>
EXTRA INTERESTS	<p>Start-up company co-founder (<i>since September 2018</i>): work as technical consultant for a start-up company that aims to preserve cultural heritage through collecting and presenting information about heritage-preserving services and products.</p> <p>Hobbies: an amateur marathon runner and an enthusiastic language learner.</p>