Yuanmo He

+44 7928549255 | y.he54@lse.ac.uk | yuanmohe.com

EDUCATION

PhD Social Research Methods (specialising in computational social science) 09/2020 - 09/2024
The London School of Economics and Political Science (LSE)

- Applying advanced data analysis and computational methods (e.g., machine learning, natural language processing, social network analysis) on digital trace data to study social networks, culture, and inequality.
- Supervisors: Dr Milena Tsvetkova and Professor Kenneth Benoit.
- Affiliations: Data Science Institute, International Inequality Institute.
- Awarded the **LSE PhD Studentship** for four years.

MSc Applied Social Data Science (Distinction) 09/2019 - 08/2020
The London School of Economics and Political Science

- Relevant modules: Computer Programming, Data for Data Scientist, Applied Machine learning, Quantitative Text Analysis, Multivariate Analysis and Measurement, Fundamentals of Social Science Research Design.
- **Distinction in all modules**.

BSc Social Sciences (First Class Honours) 09/2016 - 06/2019
University College London (UCL)

- Relevant modules: Social Network Analysis, Causal Analysis in Data Science, Quantitative Research Methods.
- Awarded the UCL Institute of Education **Faculty Medal** (the best final year undergraduate student).
- Achieved **the highest final mark** in the Department of Social Science.

Coursera:

- Statistics with R Specialization: probability, inferential statistics, Bayesian statistics 06 - 08/2018
- Mathematics for Machine Learning Specialization: linear algebra, multivariate calculus, PCA 06 - 08/2018

WORKING PAPER

He, Y and Tsvetkova, M. *Estimating Individual Socioeconomic Status of Twitter Users.* (Manuscript available upon request.)

- Developed a method that uses correspondence analysis to estimate Twitter users' socioeconomic status (SES) based on the brands they follow, which is firmly embedded in classical social theory.
- Worked on a complete data science workflow: from data collection, data cleaning, pre-processing, building model, evaluating results, to oral and written communication.
- Utilised and experienced a range of data science tools, including R, Python, Twitter API, Facebook Marketing API, SQL, web scraping, Google Geocoding API, Azure Cloud Computing, Stan, and parallel computing.
- Estimated the socioeconomic status of 3,482,657 Twitter users and 339 brands.
- Validated the estimation of brands with data on audience composition from the Facebook Marketing API. Brands with higher estimated SES have significantly more audience with master's or higher degree (Spearman's $\rho = 0.493$, $p < 0.001$), and significantly less audience who are high school graduates ($\rho = 0.499$, $p < 0.001$).
- Validated the estimated SES of a group of users who stated job titles on their Twitter profiles ($n = 58{,}466$). Aggregated for job titles, the median estimated SES is significantly associated with higher income ($\rho = 0.675$, $p < 0.001$) and occupational social class ($\rho = 0.666$, $p < 0.001$). Individual level validation is limited by data.

CONFERENCE PRESENTATIONS

Estimating Individual Socioeconomic Status of Twitter Users

- General Online Research, online 09/2021
- The Annual Meeting of the American Sociological Association (Section on Inequality, Poverty and Mobility: New Approaches to Understanding and Addressing Inequality), online 08/2021
- International Conference on Computational Social Science, online 07/2021

TEACHING EXPERIENCE

MY474 Applied Machine Learning, Teaching Assistant, LSE 01 - 05/2022
MY470 Computer Programming, Teaching Assistant, LSE 09 - 12/2021
Introduction to Python Programming, Teaching Assistant, Data Science Summer School 08/2021

SKILLS

Programming language & statistical software: Python, R, SQL, Stata*, Stan*, SPSS*
Python packages: NumPy, pandas, scikit-learn (non-exhaustive)
R packages: tidyverse, tm, quanteda, glmnet, randomForest, e1071 (non-exhaustive)
Advanced Data Analysis: machine learning, natural language processing, social network analysis, multivariate analysis, causal analysis*, multilevel modelling*, parallel computing*, cloud computing* (**indicates basic understanding*)
Languages: Chinese (native), English (full professional proficiency)