

Mireia Triguero Roura

✉ mt3197@columbia.edu
🐙 github.com/mpnyka
🌐 [linkedin.com/in/mtriguero](https://www.linkedin.com/in/mtriguero)
Last Updated: 02.06.2024

EDUCATION

Columbia University

Ph.D in Sociology

(expected May 2024)

Columbia University

M.A. in Sociology (2019)

New York University

Dual M.A. in Int. Relations and Journalism (2016)

Pompeu Fabra University

B.A. in Humanities (2013)

SKILLS

Domain Expertise:

Online Abuse, Bad Actor Detection, Racial Inequality

Languages:

R, Python, SQL, Stan, JavaScript, HTML, CSS

Natural Languages:

Catalan, Spanish, English, French

Others:

Git, HCP, GCP, machine learning, causal inference, network analysis, survey design, multilevel model, NLP, experiments

COURSEWORK

Advanced Regression and Multilevel modeling, Causal Inference for Social Science, Modern Statistics for Data Science, Machine Learning and Causal inference, NLP and Sentiment Analysis

RESEARCH PROJECTS

User Behavior and Spam and Abuse detection

Developed a novel measure of global cultural similarity; created statistical models to predict user behavior based on metadata information as well as social network characteristics and used it to improve early spam detection

Ethnic and racial classifications across the globe

Secured external funding for (NSF grant, \$20k), designed, and implemented a survey experiment across five international markets to understand how to measure race and ethnicity across different contexts.

Immigration Discourse on Twitter

Scraped data, trained and created an issue position classifier using word embeddings, and then used a structural topic model to understand how users talk about immigration online.

WORK EXPERIENCE

Computational Science Intern May 2022 - Dec 2022; May 2023 - Aug 2023

Snap, inc

New York, NY

- Conducted internal research to model user behavior using advanced statistical methods and machine learning (using SQL, R and Python).
- Worked with Spam and Abuse team to improve early spam detection.
- Used machine learning to extract latent cultural models of communication across users in different markets and to understand factors driving users' similarity in taste across the world.
- Developed an R package to properly estimate multilevel, dyad data with large datasets in a regression framework.
- Developed causal inference techniques for network data.
- Translated research findings into strategic narratives to inform product and growth teams' strategies.

Quantitative Researcher

Feb 2020 – present

Freelance

New York, NY

- Worked with several non-profit clients to pull and create data analyses and visualization to non-technical audiences to influence decision-making and policy lobbying.
- Conducted data analysis with unstructured, payroll data from Congress.
- Used deep-learning algorithms to predict racial stratification.
- Translated statistical models to simple data visualizations and memos for non-technical audiences.

Statistics and Data Science Instructor

Sept 2022 – present

Columbia University - Sociology Dept.

New York, NY

- Created teaching materials to teach PhD students how to code, run simulations, and do data wrangling and data visualization in R.
- Taught statistical concepts and causal inference methods using data simulation, as well as real data examples.
- Mentored over 20 Ph.D. and master students on developing original research, and helped them with the statistical analyses and causal inference designs.

PAPERS AND CONFERENCE TALKS

Causal inference

"Colorblind ethnocentrism: How ancestry and race continue to define Western national identities" (under review; presented at the conferences ASA 2023 and CES 2023)

- Designed, coded and implemented (using Qualtrics and custom Javascript code) a conjoint experiment and survey across five national samples (US, UK, France, Germany, Spain).
- Analyzed data using both traditional econometrics measures and multilevel models.
- Ran post-stratification models to adjust data to population estimates.

"Unbiased Causal Estimates in Panel Data with Bayesian Additive Trees." (a version of this paper was accepted for the 2020 ASA conference)

- Proposed an alternative causal estimand to fixed effects to avoid post-treatment bias in panel data
- Implemented this causal inference identification strategy using Bayesian Additive Trees to create a posterior distribution for both potential outcomes for each treated unit.

"Causal Effect of Police Violence on 911 calls" (working paper; presented at the Columbia Stats Lab in 2020)

- Designed a causal inference framework using Gaussian Process Models to accurately model time trends in 911 calls in order to create counterfactual estimates for when a specific time intervention happened.

Computational Social Science

"Social media, language, and politics: What language do people use to make arguments about undocumented immigrants?" (Working paper, see [here] for code and draft)

- Created a topic-specific sentiment classifier using word embeddings to classify tweets about DACA.
- Used Structural Topic Modeling to extract latent themes used by each position on the issue to understand the types of semantic frameworks people used to make their arguments.

"The Ties that Bind: Online Pictorial Communication Reveals Cultural Boundaries In the World." Author order: **Triguero Roura, Mireia**, Maarten Bos, Francesco Barbieri, Julie Jiang, Yozen Liu and Ron Dotsch. (Under review)

- Created user embeddings based on their sticker usage (using IRLBA).
- Developed custom functions in R to analyze undirected dyad data and properly account for correlation in the error.

"Reciprocity, Homophily, and Social Network Effects in Pictorial Communication: A Case Study of Bitmoji Stickers." Author order: Jiang, Julie, Ron Dostch, **Mireia Triguero Roura**, Yozen Liu, Vitor Silva Souza, Maarten Bos and Francesco Barbieri. (Published at CHI '23) [link].

- Used propensity score models to estimate the effect of receiving a sticker on Snapchat engagement.
- Used social network features to understand sticker usage on the app.

"Understanding the Determinants of Message Response Behaviors." Author order: Peters, Heinrich, Ron Dotsch, **Mireia Triguero Roura**, Yozen Liu, Sandra Matz, and Maarten Bos. (Under review)

- Extracted social network features to understand messaging behavior and used Gradient Boosting algorithms for continuous and binary prediction.

Multilevel models

"Threatening for whom? National boundary-making, immigration, and support for the welfare state" (R&R at European Sociological Review)

- Used multilevel models to properly account for the three-level nested structure of respondents within regions within countries, as well as the cross-level structure of multiple years.
- Used factor analysis and principal components analysis to summarize multifaceted concepts around national belonging and support for equitable public policies.

"Inequality under the Dome: Racial disparities in Congress." co-authored with James R. Jones (Under Review)

- Analyzed nested data with multiple respondent-level observations, and respondents within offices.
- Used a pre-trained algorithm (based on LSTM) to predict racial identification of staffers.