

# Omer F. Yalcin

Social Data Scientist | Ph.D. in Political Science & Social Data Analytics | Statistics · Machine Learning · Network Analysis · Causal Inference · NLP | 8+ years of using Python, R, SQL | Published in ICWSM, EMNLP Findings, JQD

[ofyalcin@gmail.com](mailto:ofyalcin@gmail.com) | [omerfyalcin.com](http://omerfyalcin.com) | [linkedin.com/in/ofyalcin](https://linkedin.com/in/ofyalcin) | [github.com/omerfyalcin](https://github.com/omerfyalcin)

Authorized to work in the U.S. for any employer; no sponsorship required

## SKILLS

---

**Programming Languages:** Python, R, SQL

**Data Wrangling and Modeling:** numpy, pandas, tidyverse, scikit-learn, tidymodels, DuckDB, SQLite

**Visualization and Reporting:** ggplot2, seaborn, Quarto, LaTeX

**Natural Language Processing (NLP):** transformers, spaCy, BERT, STM

**Network Analysis:** igraph, network, sna, ergm, latentnet

**Web & APIs:** requests, BeautifulSoup, Selenium, http2

**Development Tools:** Git, Bash, (Neo)Vim, Jupyter, VS Code/Positron, RStudio

**High-Performance Computing:** Slurm, parallel processing, SSH, HPC workflows, Apptainer & Docker

## PROJECTS

---

**YouTube Platform Analytics: Political Content, Misinformation, and Engagement Inequality |**

selenium, STM (topic modeling), causal inference with hierarchical models (lme4), ggplot2

- Built a Python Selenium data collection bot to simulate user search behavior on YouTube
- Helped design a difference-in-differences analysis of YouTube's informational-intervention policy using hierarchical models in lme4; produced publication-ready visualizations in ggplot2.
- Co-authored study analyzing millions of YouTube videos, comments, and channels to measure political content and engagement inequality; ran structural topic models of comment content. Study found that 2% of commenters accounted for 50% of all comments.
- Published in *ICWSM* [\[Link\]](#) and *Journal of Quantitative Description* [\[Link\]](#)

**NLP: Topic Modeling and LLM Bias Detection |** transformers, BERT, statistical modeling with multinomial logit, AIC-based model comparison and out-of-sample predictive validation

- Co-authored paper [\[Link\]](#) on how document-aggregation strategies affect topic-model content; independently implemented the BERT-based topic-modeling component using GPU-accelerated workflows on a university HPC (High Performance Computing) cluster.
- Used NVIDIA RAPIDS and UMAP to scale dimensionality reduction on BERT embeddings for large text corpora of ~ 1 million tweets and ~ 190k Wikipedia articles
- Evaluated models using both out-of-sample prediction accuracy and in-sample fit metrics such as AIC, combining ML-style validation with statistical model assessment.
- Co-authored another paper [\[Link\]](#) on nationality bias in pretrained language models (PLMs), working on data curation, statistical uncertainty estimation, and interpretation

**Elite Political Networks and Digital Trace Data |** SPARQL, igraph, requests, latentnet

- Collected data on 270k elite political actors across the globe and their kinship ties from the Wikidata knowledge graph using the SPARQL query language. Showed kinship networks among elites to be denser in more authoritarian countries [\[Link\]](#)

- Studied the evolution of the network structure of Turkey's ruling elites—320 unique actors—over 22 years using Google Search data on pairs of actors. Modeled the network using latent-space network models to account for tie interdependencies. Showed the declining role of cabinet membership in tie formation over time. [\[Link\]](#)
- Co-authored handbook chapter on web data collection (Edward Elgar Publishing); wrote all Python implementation covering REST APIs, HTML/JSON parsing, and scraping [\[Link\]](#)

## EXPERIENCE

---

**Lecturer in Data Analytics and Computational Social Science,** Jan 2022 - present  
**University of Massachusetts Amherst, Amherst, MA**

- Ran in-class Kaggle machine learning competitions with self-collected original data on [1] 150K paintings sold at art auctions (regression, price prediction) and [2] 30K seismic events (event type classification), emphasizing feature engineering, model evaluation, and data storytelling
- Delivered graduate-level instruction in Python, R, statistics, and machine learning, to ~50 students per semester, translating complex technical concepts for mixed-experience audiences.
- Designed and launched a Python for Data Science course covering data wrangling, visualization, modeling, data collection, database interaction, and applied machine learning
- Integrated Bayesian approaches and causal inference into an existing inferential statistics course
- Supervised graduate student RA (ongoing) on a project on inauthentic editing in more than 300K edits over U.S. politicians' Wikipedia pages from 2015-2025

**Inter-university Consortium for Political and Social Research,** Summers 2019,  
**University of Michigan, Ann Arbor, MI** 2020, 2023, 2024

- Instructor (2023, 2024): taught a Python course and a LaTeX course to around 200 social science graduate students, researchers, and professionals from around the world; organized “Tidy Tuesday” R/ggplot2 visualization challenges for participants
- Teaching Assistant (2019, 2020): served as TA for introductory and advanced network analysis courses, conducted daily office hours, troubleshoot R problems, taught advanced inferential network analysis

**Graduate Assistant, Pennsylvania State University, State College, PA** Aug 2016 - Dec 2021

- Contributed to computational research on anti-vaccine content on YouTube, lobbying in Congress, and statistical topic models using R/Python, text analysis, and network methods.
- Taught research design for social data analytics as instructor-of-record; scientific study of politics as TA

## EDUCATION

---

<b>Ph.D. in Political Science and Social Data Analytics,</b> Pennsylvania State University	2021
<b>M.A. in Political Science,</b> Pennsylvania State University	2018
<b>M.A. Research Training Program,</b> Humboldt University of Berlin, Germany	2016
<b>B.A. Economics and Political Science,</b> Bogazici University, Istanbul, Turkey	2014