

# YUKUN YANG

Boston, MA • [contact@yukunyang.info](mailto:contact@yukunyang.info) • [yukunyang.info](http://yukunyang.info)

## RESEARCH SUMMARY

---

My research interests lie in understanding how social identities undergird communication and interactions on social media. Specifically, I'm interested in how individuals express and construct their identities and how they perceive and react to others online. On a larger scale, I am also interested in how identity-based complex social processes influence the phenomenon of balkanization, polarization, mobilization, and user migration in online spaces.

**Keywords:** Computational Social Science, Computer-supported Communication, Online Communities, Science and Technology Studies

## EDUCATION

---

<b>University of North Carolina at Chapel Hill</b> <i>Master of Science in Information Science (13 High Passes [95+/100])</i>	<i>Chapel Hill, NC, U.S.A.</i> <i>May 2020</i>
<b>Northwest University</b> <i>Bachelor of Management in Library Science (GPA: 3.83/4.0; Rank: 1/28)</i>	<i>Xi'an, China</i> <i>July 2018</i>

## RESEARCH EXPERIENCE (\* for Independent/Self-guided Research)

---

<b>Center for Antiracist Research, Boston University</b> <i>Data Scientist</i>	<i>Boston, MA, U.S.A.</i>
---	---------------------------

**Built Environment and Racial Inequities in Health** – [J1] *May 2022-*

- Conducted primary data analysis to investigate the association between neighborhood racial composition, built environment, and health outcomes with factor analysis, clustering analysis, multi-level modeling, and mediation analysis in R
- Led paper writing and data visualization

**Self-Disclosure of Racism Experience on Reddit\*** – Paper in progress *Feb 2022-*

- Scraped, sampled, and coded self-disclosure narratives of racism experience on r/racism
- Directed research design, paper writing, and the conceptualization of thematic models

**Structural Racism in COVID-19 Vaccine Distribution** – [J2] *Nov 2021-*

- Developed and standardized the data wrangling pipeline that scraped data from public-facing websites, formatted and geocoded the addresses, and merged it with census data
- Performed secondary regression diagnostics, participated in paper writing and revision, and developed visualizations for the publication

<b>National Police Foundation (National Policing Institute)</b> <i>Research Data Scientist</i>	<i>Arlington, VA, U.S.A.</i>
---	------------------------------

**Active Bystander Training and Ethical Policing – [J3]** *May 2021-August 2021*

- Conducted exploratory data analysis on the survey data responses with Exploratory Factor Analysis and Canonical Correlation in R
- Generated static and interactive data visualizations and participated in paper writing

**Police Union Contract and Police Accountability** *Dec 2020-August 2021*

- Supervised the qualitative data collection, including MTurk worker management and double-coding review, and performed clustering on the coded documents
- Analyzed the LexisNexis news data to identify the progress of police union contract negotiations with NLP techniques in Python

**Assessment of LAPD's Response to George Floyd Protests – [R1]** *Sept 2020-March 2021*

- Identified policing hotspots and community opinion leaders with spatial-temporal clustering, and tweets analysis and developed interactive dashboards
- Scraped Twitter data and helped the team to identify key opinion leaders for interviews
- Developed an interactive E-report which was later presented to the LA City Council

**University of North Carolina-Chapel Hill**

*Chapel Hill, NC, U.S.A.*

*Master's Student*

**Computational Music Trend Analysis\* – [C1]** *Sept 2019-Jan 2020*

- Collected and wrangled, textual and acoustic data from lyrics datasets and Spotify's API
- Proposed and evaluated novel metrics for quantifying lyrical repetitiveness using closed itemset mining and performed time-series analyses with trend detection and Time-lagged Cross-Correlation

**Power Abuse and Extreme Online Moderation on Reddit\* – [C2]** *Jan 19-May 2019*

- Scraped and cleaned 110k submissions and comments from r/xkcd and r/xkcdcomic
- Conducted a quantitative event study of an infamous event of excessive moderation on Reddit with time-series analysis with intervention (ARIMA), breakpoint detection, and social network analysis in R

**Guilty Information Disclosure on Social Media\* – [C3]** *Nov 2018-Jan 2019*

- Retrieved, cleansed 225k submissions and comments from r/confession, and visualized the topic distribution of submissions with LDA in Python
- Constructed a thematic model, capturing the intents, contents, and interactions of the guilty information sharing behavior on social media with sampled data

**Communication and Interaction Patterns of Homogenous and Cross-Cutting Spaces in Online Activism\* – [P1]**

- Scraped and cleaned 60k+ historical tweets about #SilentSam online activism and classified the standpoint of tweets with word embeddings

- Conducted non-parametric significant tests to test if homophily influences their linguistic and argumentative style, and hyperlink usage using LIWC and Perspective API
- Analyzed the subgraphs pattern of the network and compared them with an ER graph

**Digital Innovation Lab, University of North Carolina-Chapel Hill**

*Chapel Hill, NC, U.S.A.*

*Social Media Data Analyst*

**#SilentSam Tweets Analysis**

*Feb 2019-March 2020*

- Analyzed the tweets of a campus activism event with visual and spatial analytics, topic modeling, and network analysis and delivered multiple interactive dashboards in Python and R

**Pinterest for Digital Humanities**

*Sept 2019-May 2020*

- Designed, developed, and deployed an online application for a graduate-level digital humanities class that allows scraping images from search results on Pinterest and provides a descriptive analysis of the metadata and the images

## **OTHER WORK EXPERIENCE**

---

**ArcLight Analytics**

*Chapel Hill, NC, U.S.A.*

*Data Scientist*

*July 2020 -*

- Scraped and wrangled multiple large-scale datasets from public/private sectors, including Crunchbase, NSF, PPP loans FIOA, Federal Workforce, National Conference of State Legislatures, BLS, etc.
- Performed data analysis with survey data, administrative data, and text data in Python
- Developed and deployed a static project website and two interactive data dashboards

## **PUBLICATIONS**

---

**Peer-Reviewed Conference Papers:**

- [C1] **Yang Y.** (2020). "Musicalization of the Culture": Is Music Becoming Louder, More Repetitive, Monotonous, and Simpler?. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 14(1), 750-761. Retrieved from <https://aaai.org/ojs/index.php/ICWSM/article/view/7340>. (Acceptance Rate: 24%)
- [C2] **Yang Y.** (2019). When Power Goes Wild Online: How Did a Voluntary Moderator's Abuse of Power Affect an Online Community. In *Proceedings of the Association for Information Science and Technology (ASIS&T)*, 56: 504-508. Retrieved from doi:10.1002/pr2.55.
- [C3] **Yang Y., & Huang, Y.** (2019). Dumping the Closet Skeletons Online: Exploring the Guilty Information Disclosure Behavior on Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 13(01), 663-666. Retrieved from <https://www.aaai.org/ojs/index.php/ICWSM/article/view/3267>. (Acceptance Rate: 21%)

## Peer-Reviewed Journal Papers:

- [J1] Yang, Y., Cho, A., Nguyen, Q., Nsoesie, E. O. Neighborhood Racial Composition, Built Environment and Health Outcomes. *JAMA Network Open*. Submitted.
- [J2] Holtzman, G., Yang, Y., Louis, P., Kandaswamy P. Quantifying Structural Racism in COVID-19 Vaccine Distribution. *Nature Human Behaviour*. Revise and Resubmit.
- [J3] Taniguchi, T., Vovak, H., Cordner, G., Amendola, K., Yang, Y., Hoogesteyn, K., & Bartness, M. (2022). The Impact of Active Bystander Training on Officer Confidence and Ability to Address Ethical Challenges. *Policing: A Journal of Policy and Practice*.

## Reports:

- [R1] National Police Foundation. (2021). A crisis of trust: A National Police Foundation report to the Los Angeles Board of Police Commissioners on the Los Angeles Police Department's response to First Amendment assemblies and protests occurring May 27 – June 7, 2020.

## Master's Paper:

- [P1] Yang, Y. (2020). Friends or Foes: Understanding Communication and Interaction Patterns of Homogeneous and Cross-Cutting Spaces in Online Activism. <https://doi.org/10.17615/efhv-bf11>.

## SKILLS

---

### Quantitative Research Methods

Exploratory Data Analysis, Supervised & Unsupervised Machine Learning, Data Mining (Pattern, Time-series, Network), Statistical Modelling (Regressions, CFA/EFA, SEM), Data Visualization

### Qualitative Research Methods

Content Analysis, Thematic Analysis, Survey Testing & Administration

### Tools/Languages

Python, R, Tableau, MS Office, Git, Arc/QGIS, AWS, Markdown, SQL, JavaScript, HTML/CSS, Gephi, Java, Qualtrics, Figma

## HONORS & AWARDS

---

Burmeister Fellowship, UNC School of Information and Library Science (SILS)	<i>August 2019</i>
Best Project Award, 5th Annual UNC SILS Project Fair	<i>April 2019</i>
China National Scholarship (3 times), Chinese Ministries of Education	<i>2014-2017</i>

## SERVICE & LEADERSHIP

---

Master Student Representative, Diversity Committee, UNC SILS	<i>2019-2020</i>
Master Student Mentor, UNC SILS MENTOR Program	<i>2019-2020</i>
Peer Reviewer (Full papers/Posters/Datasets): ICWSM & ASIS&T	<i>2020-2022</i>