RESEARCH STATEMENT

Siqi Wu, Australian National University, siqi.wu@anu.edu.au

Social media is increasingly transforming our perceptions and behaviors: it shapes our information feeds, consumes our scarce attention, and impacts our ideological positions. In both online and offline worlds, social media influences the society on the issues of politics, culture, and health. Despite its prevalence, the responsibilities of online social platforms remain an active matter of debate.

My long-term goal is to develop principles for responsible platforms by measuring and modeling the collective user behavior on the web and social media. My research uses a mix of computational and quantitative methods: I design the infrastructure for constructing large-scale, cross-platform datasets; I perform empirical measurement studies using the techniques of data visualization, network analysis, and text mining; I develop machine learning models to explore the prediction limits in complex social systems. My work lies in the fields of computational social science, social computing, and data mining.

My research focus has been on two prevailing online social platforms -- Twitter and YouTube. I have examined questions such as how Twitter data sampling impacts common measurements (*ICWSM '20*), how YouTube recommender system drives user attention (*CSCW '19*), and why some YouTube videos are more engaging (*ICWSM '18*). All of these demonstrate my technical skills in addressing the grand challenge of understanding the impacts of social platforms on the society:

- **Data acquisition**: I have released 2 data collection tools and 3 large datasets to the community. For example, the Twitter-intact-stream software constructs the nearly complete Twitter stream. The Vevo music graph dataset is the only public dataset containing information of video watch time. I further enriched it with YouTube recommendations, Twitter posts, and Google search interests.
- **Empirical measures:** I have presented a collection of large-scale measurements related to Twitter data sampling, YouTube recommendation network, and collective engagement. Firstly, I quantified the sampling effects of Twitter Streaming API across different timescales and different subjects (entities, networks, and cascades). Secondly, using the bow-tie structure, I characterized the video network connected by the YouTube recommender system. And I found that its core component, which occupies most of the attention, is made out of videos that are mainly recommended among themselves. Lastly, I proposed a duration-calibrated metric for the aggregate video engagement. I showed that this metric is correlated with video quality and stable over time.
- **Interpretable modeling:** I have developed machine learning models in my research. For example, I built a model that accounts for the network effects to estimate the attention flow between videos and artists. I also quantified the predictability of different metrics. And I found that engagement metrics (watch time) are more predictable than the popularity metrics (view count).

I have actively collaborated with external researchers. My collaborators spread over University of Michigan, University of Pittsburgh, and Arizona State University. The collaborations are all in remote format and over a wealth of topics, e.g., political science, computational journalism, and cybersecurity.

There are still many interesting questions to be answered around the web and social media. My future research plan includes: (1) measuring the negative effects of YouTube recommender system; and (2) measuring the public Twitter opinions of different demographic groups towards COVID-19 pandemic.