

# Recent Advances in Text Analysis

**Tracy Ke, PhD**

Associate Professor, Department of Statistics,  
Harvard University



Text analysis is an interesting research area in data science and has various applications, such as in artificial intelligence, biomedical research, and engineering. In this talk, Dr. Ke will review popular methods for text analysis, ranging from topic modeling to the recent neural language models. In particular, she will introduce Topic-SCORE (Ke and Wang, 2022), a statistical approach to topic modeling, and discuss how to use it to analyze MADStat - a dataset on statistical publications that we collected and cleaned on our own. The application of Topic-SCORE and other methods on MADStat leads to interesting findings. For example, 11 representative topics in statistics are identified. For each journal, the evolution of topic weights over time can be visualized, and these results are used to analyze the trends in statistical research. In particular, we propose a new statistical model for ranking the citation impacts of 11 topics, and we also build a cross-topic citation graph to illustrate how research results on different topics spread to one another.

**Thursday, October 5, 2023, 3:30-4:30PM Eastern**

**133 Rosenau Hall**

Zoom Link: <https://unc.zoom.us/j/95131277245>

**Meeting ID: 951 3127 7245**

**Passcode: 203117**