

# **Analysis of word co-occurrence networks from paper abstracts in Semantic Scholar database**

*Frederick Kin Hing Phoa*

*Institute of Statistical Science, Academia Sinica, Taiwan*

## Abstract

The abstract is a crucial frontmatter element that provides readers with key insights into a manuscript's core ideas and subject categories. Identifying the most important words in abstracts can offer valuable clues about the central themes and evolving trends within a particular subject area. This work introduces a novel analysis method to determine the importance of words within a subject category over time, based on various centrality measures in a word co-occurrence network. The network is constructed from words extracted from the abstracts of manuscripts within a specific scientific subject. We demonstrate the effectiveness of this method using a subset of the Semantic Scholar database, focusing on the field of Statistics from 2019 to 2023. This is a joint work with Tiffany Yoonjin Lee, Judy Nakyoung Shin, and Dr. Hohyun Jung of Sungshin Women's University, South Korea.