

Progress Report on Paper Writing

WU Zihan

January 4, 2024

Progress on paper writing

- ▶ Introduction, abstract, conclusion and own experiment parts are finished.
- ▶ Contributions: Speed, memory usage and probability guarantees.
- ▶ Complexity analysis based on auto-optimal argument picker.

Comparative Analysis and Data Cleaning

- ▶ Comparison of 3 approaches: SB (Baseline), GLBCO (learning-based), and PNMTF (parallel).
- ▶ Data cleaning using stop words, which needs more tuning.
- ▶ PNMTF is the fastest, but mainly focuses on sparse matrices.

To do list next week

- ▶ Comparing speed and biggest memory usage on three datasets (small, medium, large).
- ▶ NMI, ARI (gives the accuracy of clustering)
- ▶ Change of above metrics within progressing
- ▶ Analysis on
 - ▶ document number (scalability)
 - ▶ cross-dataset performance (adaptability)