Project Proposal: Network Analysis and Text Processing of Daily Kos Dataset

Team Members: Coffy Andrews-Guo and Tora Mullings

Data Source:
Konect is a database that contains an extensive collection of network datasets from different domains. It provides standardized data in a format that is easy to access and analyze.

- Konect website: http://konect.cc/networks/bag-kos/
- Dataset Name: Daily Kos, http://konect.cc/files/download.tsv.bag-kos.tar.bz2
- (Original Source) UC Irvine Machine Learning Repository: http://archive.ics.uci.edu/dataset/164/bag+of+words

Introduction:
This project aims to perform network analysis and text processing on the Daily Kos dataset, a bipartite document-word dataset.

Objectives:
To conduct network analysis to identify influential nodes and explore connections, apply topic modeling to uncover prevalent topics and their relationships with words, extract representative keywords for document summaries, analyze document similarity to identify related clusters, and perform sentiment analysis to understand the emotional tone expressed in Daily Kos articles.

Methodology:
In this project, the cleaned and processed Daily Kos dataset will be used to create a network to analyze the relationships between documents and words. Topic modeling identifies key themes, while keyword extraction summarizes document content. Document similarity analysis clusters related documents, and sentiment analysis determines the expressed sentiment.

Group Plan:
In a two-member group, the division of the roles and responsibilities are as follows:

- Coffy is responsible for preprocessing, network analysis, document similarity analysis, and presenting the results.
- Tora will focus on topic modeling, keyword extraction, sentiment analysis, and providing insights.

We will collaborate closely, validate each other's findings, and work together on the analysis and presentation.