BUS 243 Name:

Quiz 2.

Problem 1. What is the main advantage of using the TF-IDF representation for text data compared to other methods like bag-of-words?

Problem 2. In a corpus of text where the frequency distribution of words follows Zipf's law, why does a log-log plot of word frequency against rank produce a straight line?