# CSCI446/946 Big Data Analytics

## Lab Week 8
## Text Analysis

School of Computing and Information Technology

University of Wollongong Australia

# Task One (beginner)

- R comes with an `lda` package that has a built-in sample dataset `Cora`.
  - Following the code given in Section 9.6 of the reference book to generate Figure 9-5.
  - Understand the given code and be able to explain the main procedure and Figure 9-5.
  - Know the key function of `lda` and understand the meaning for the parameters used in this function in the code.

# Task Two (intermediate)

- R comes with an `textmineR` package which provides tools that can help identify the number k of topics in a corpus.

  – Use the Sentiment.csv dataset and following the tutorial of [https://towardsdatascience.com/beginners-guide-to-lda-topic-modelling-with-r-e57a5a8e7a25](https://towardsdatascience.com/beginners-guide-to-lda-topic-modelling-with-r-e57a5a8e7a25)

  to obtain the value k and to visualize the results as a word cloud.

# Task Three (advanced)

- Verify your understanding by repeating task 2 using the cora dataset.