IDS 703 Final Project Proposal

Team Orange

Michael Tang, Yi Feng, Shangwen Yan, Jennie Sun

- **Title**
  - Citation Network Classification with Graph Neural Network
- **Project Aims**
  - To understand the concepts of graph theory and the implementation of graph neural networks (in terms of how it differs from other deep learning techniques).
  - To use Graph Convolutional Neural Networks (GCNs) to extract information from structured scientific publications.
  - To create a graph neural network to classify publication topics based on publication context and relationship with other publications.
- **Methodologies**
  - Semi-supervised classification with Graph Convolutional Networks
  - Word2Vec
  - Word embedding
  - Node classification on citation networks using Spektral (Tensorflow API)
  - PyTorch Geometric for data handling of graphs
- **Data**
  - The Cora dataset consists of 2708 scientific publications classified into one of seven classes. The citation network consists of 5429 links. Each publication in the dataset is described by a 0/1-valued word vector indicating the absence/presence of the corresponding word from the dictionary. The dictionary consists of 1433 unique words.
  - Link: https://graphsandnetworks.com/the-cora-dataset/