Nicholas Guerrero, Jasper Hoong

Prof. Tsourakakis

CS 365: Foundations of Data Science

6 February 2024

CS365 Project Proposal: Food Pairing With GNNs

Defining the problem of focus:

The problem we will be addressing is how to utilize neural networks to create new food pairings that go well together. Furthermore, we will be investigating how accurate a neural network can be by comparing the recipes it comes up with with existing and well-known recipes. More specifically, how different types of GNN models such as GCN (Graph Convolutional Network) and GraphSAGE compare to each other and which would be better. For the project, we will be consulting Stanford's "What's Cooking? Using GNNs to redefine Culinary Boundaries" as well as "The Little Book of Deep Learning" by François Fleuret to develop a good understanding of different types of neural networks and machine learning.

Methods we plan to explore:

Our research will be centered around the use of GNNs. We will be employing the use of PyTorch Geometric to build our GNN models. PyG (PyTorch Geometric) will incorporate autoencoder decoders to help us perform link predictions which are crucial to GNNs. In regards to GNNs, we will be employing the use of GCN and GraphSAGE. For the dataset, we will be using a dataset called FlavorGraph which is a large-scale food-chemical graph for generating food representations. It shows both the food ingredients and chemical compounds as well as recommended food pairings. We will be using the FlavorGraph as a network and focus on food

ingredient data to create new food pairings. For the output of the neural networks, we will be using a multi-layer perceptron to help us output a score for the newly generated pairings of ingredients to indicate a prediction of how well the pairing would go together in a recipe. After collecting the results of the experiment, we will visualize the output of both GCN and GraphSAGE using Matplotlib.

Datasets we will use:

https://raw.githubusercontent.com/lamypark/FlavorGraph/master/input/edges_191120.csv

https://raw.githubusercontent.com/lamypark/FlavorGraph/master/input/nodes_191120.csv

Division of responsibilities:

We will split up the work by having each partner focus on a specific type of GNN. One of us will be performing research on GCN while the other on GraphSAGE.