PSG COLLEGE OF TECHNOLOGY, COIMBATORE 641004

Department of Computer Science & Engineering

19ZO02 – SOCIAL AND ECONIMIC NETWORK ANALYSIS

FOOD RECIPE RECOMMENDATION SYSTEM

By

19Z304 – Aditya Sriram

19Z307 – Barath Kumar G

19Z324 - Kamalraj D

19Z357 – T S Swaminathan

19Z358 - Tarun Visva R

BE CSE G2 (2019 – 2023)

PROBLEM STATEMENT

The challenge of preparing enticing and nourishing food has plagued humanity for ages. People learned about good dishes either by word of mouth or trial and error and neither method placed focus on the dishes nutritional value. With the advent of techniques in machine learning, it is now possible to leverage other people's food preferences to not only suggest dishes one might like but also personalize them in a way to help satisfy the nutritional requirements people need on a daily basis. The existing food recipe recommendation system uses techniques such as collaborative filtering and content based algorithms. But the major drawback of this existing system is that it recommends only the recipes that already exist. Our proposed system tackles this problem by suggesting nutritional recipes that are tailored made for the user of the system.

DATASET DESCRIPTION

The dataset used for the application involves multiple CSVs that contain various required features. The important dataset used for the recommendation is the 'ratings' dataset which comprises of the ratings provided by all the users for the various recipes they have tried out. The dataset used consists of approximately 677,000 such ratings. The ratings provided by the user would be in the scale to 0 to 5. The users are represented by their User ID and the recipes and represented by their Recipe ID. The information about the recipes is fetched from the 'recipe' dataset that contains the Recipe ID mapped to the recipe name, image URL of the recipe, cooking directions, ingredients and nutrition of the recipe. The dataset used contain the aforementioned details about 45,600 recipes, approximately.

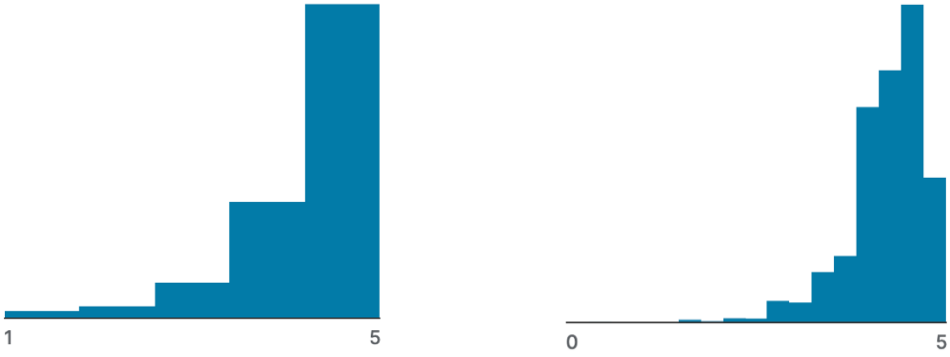

Figure.1. Distribution of the ratings provided by the users(L) and the distribution average ratings of the recipes(R).

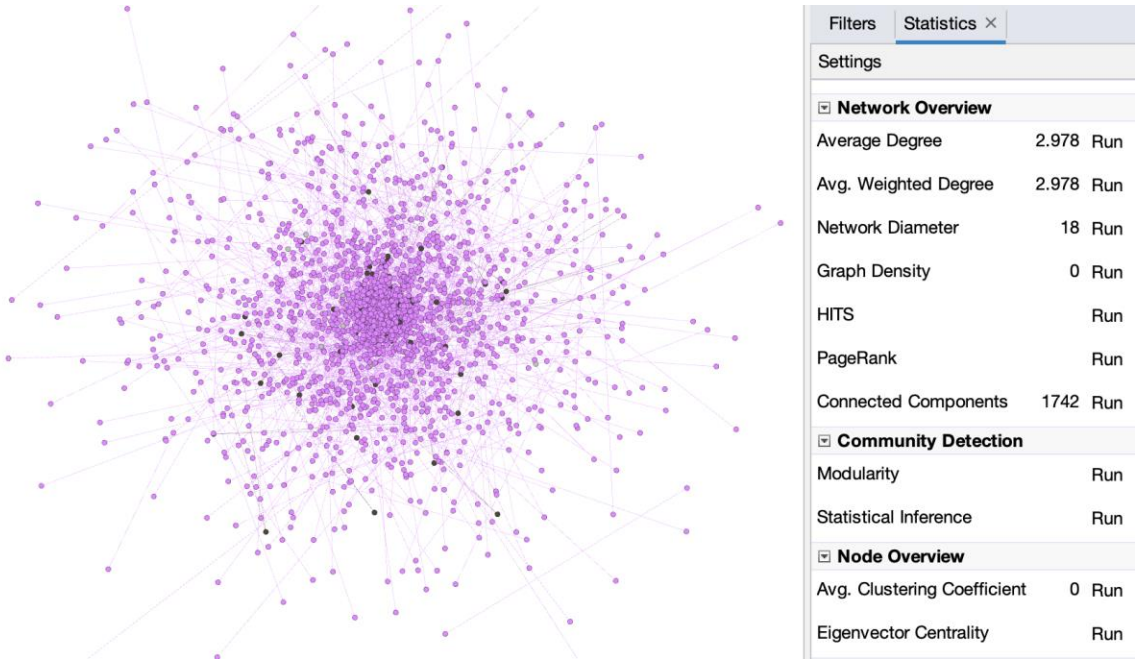

Figure.2. Network Statistics of the user-recipe ratings dataset.

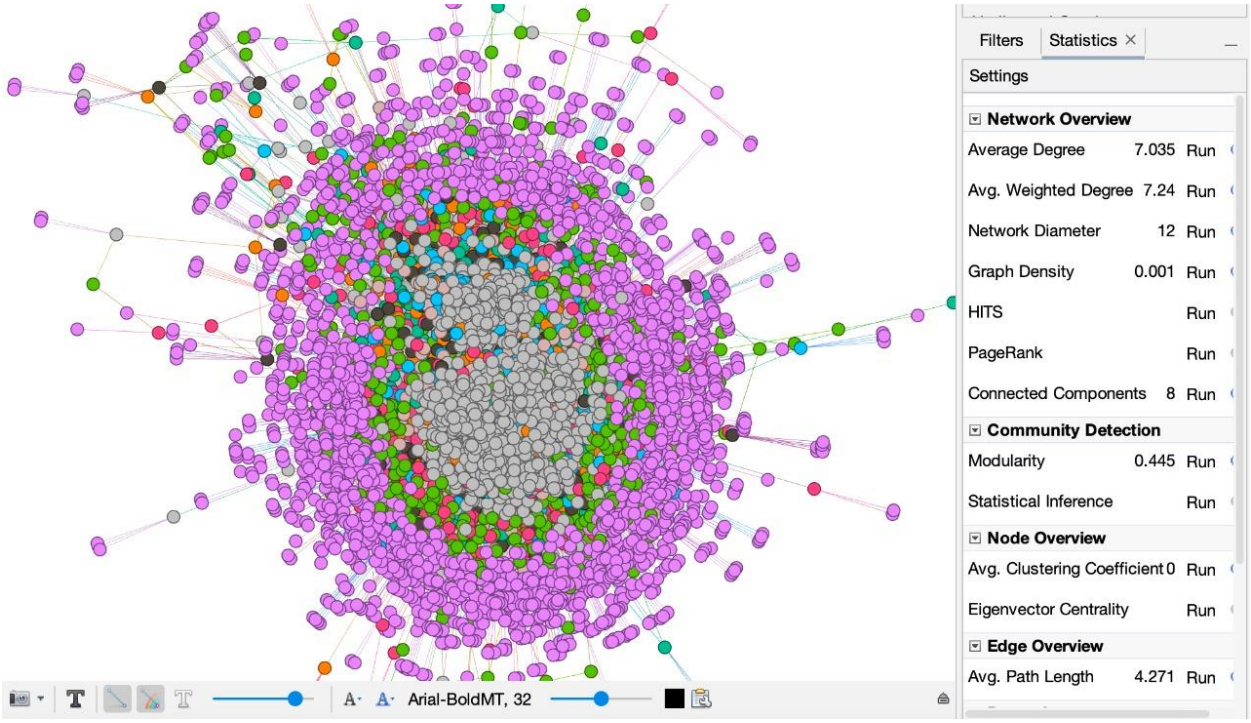

Figure.3. Network Statistics of the recipe-ingredient dataset.

TOOLS USED

NetworkX

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and function of complex networks. It is used to study large complex networks represented in the form of graphs with nodes and edges. Using networkx we can load and store complex networks. We can generate many types of random and classic networks, analyse network structure, build network models, design new network algorithms and draw networks.

NLTK

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for application in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, lemmatizing, tagging and semantic reasoning. It also includes graphical demonstrations and sample data sets as well. *Tokenization* is used in natural language processing to split paragraphs and sentences into smaller units that can be more easily assigned meaning. *Semantic Reasoning* helps understand the meaning of the sentences. It does this by taking into account the context, logical structuring of the sentences and grammar roles. *Lemmatization* is the process of grouping together the different inflected forms of a word so they can be analyzed as a single item. Lemmatization is similar to stemming but it brings context to the words. So it links words with similar meanings to one word.

SCIKIT-LEARN

In the Python environment, Scikit-learn, an open source data analysis toolkit, is considered to be the pinnacle of machine learning. Important ideas and traits include: algorithms for making decisions, such as: Data are identified and categorized by classification based on patterns. *SVD* : Organizing comparable things into groups is the work of clustering. It is a method of unsupervised machine learning. Most of us associate clustering with K-Means Clustering, a straightforward yet effective technique. It's not always the most accurate, though.

SURPRISE

A Python package called Surprise lets you build and test rate prediction algorithms. The scikit-learn API, which should be recognisable to users familiar with the Python machine learning ecosystem, served as the inspiration for its creation. For assessing predictions, Surprise contains a collection of estimators (or prediction algorithms). The implementation includes traditional methods like the primary similarity-based algorithms and matrix factorization algorithms like SVD and NMF. Additionally, it contains methods for selecting the best model, such as grid search and randomized search, as well as tools for evaluating the model, such as cross-validation iterators and scikit-built-in learnt metrics. Due to simple primitives and a light API, users may create their own recommendation method with less coding.

CHALLENGES FACED

- Visualization of the dataset remained a daunting process due to its huge size. Thus, evaluation of various metrics such as the average path length remained computationally infeasible. Therefore, it required fetching a sub-portion of the dataset without compromising its real world characteristics such as the degree distribution as per the power law.
- Choosing of the optimal collaborative filtering machine learning model was challenging as the choice required analysis of various existing frameworks such as KNNwithMeans, NMF, SVD etc. Altering the hyper parameter of such models to further improvise their accuracy was tedious.
- Development of a machine learning model to predict the FSA score of the recipe using the ingredients was hard. The dataset provided the nutrition content of the recipes, out of which fetching the nutrition provided by a single ingredient was perplexing. Thus, we were required to come up with a novel ML regression algorithm that predicts the FSA based on the ingredients provided.
- When provided with an unhealthy recipe, the exact ingredient that must be removed to improve the FSA score was unknown. Thus, we came up with the solution to replace the element that least occurs in the recipes among all the ingredients.

CONTRIBUTION OF TEAM MEMBERS

Roll No.	Name	Contribution
19Z304	Aditya Sriram	Report & Performance Analysis of Recommendation models
19Z307	Barath Kumar G	Analysis of Recommendation model and dataset visualization
19Z324	Kamalraj D	Ingredient to FSA mapping ML model & ingredient replacement code
19Z357	T S Swaminathan	Report & Performance Analysis of Recommendation model
19Z358	Tarun Visva R	Ingredient to FSA mapping ML model & ingredient replacement code

ANNEXURE I : CODE

Code link: https://github.com/BarathKumarBK-15/Food-Recipe-Recommendation-System

ANNEXURE II : SNAPSHOTS OF OUTPUT

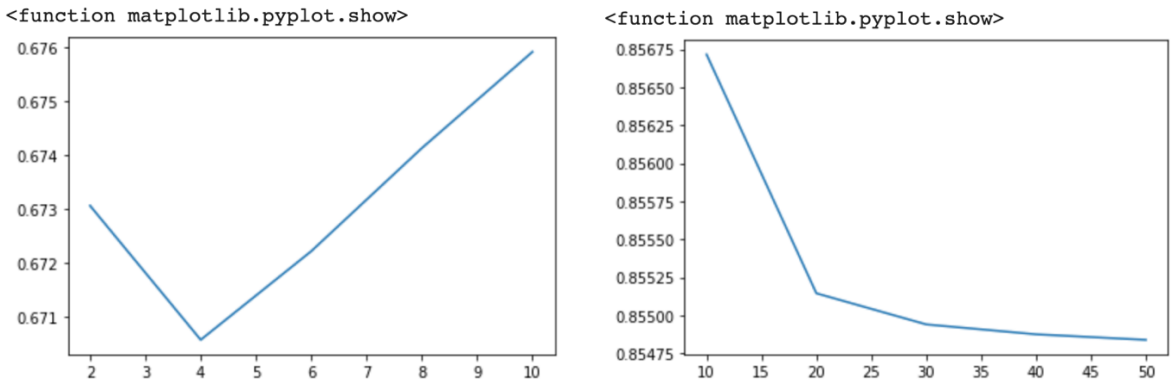

Figure.4. Variation of MAE with respect to the number of neighbours in KNNBasics(L) and KNNwithMeans(R).

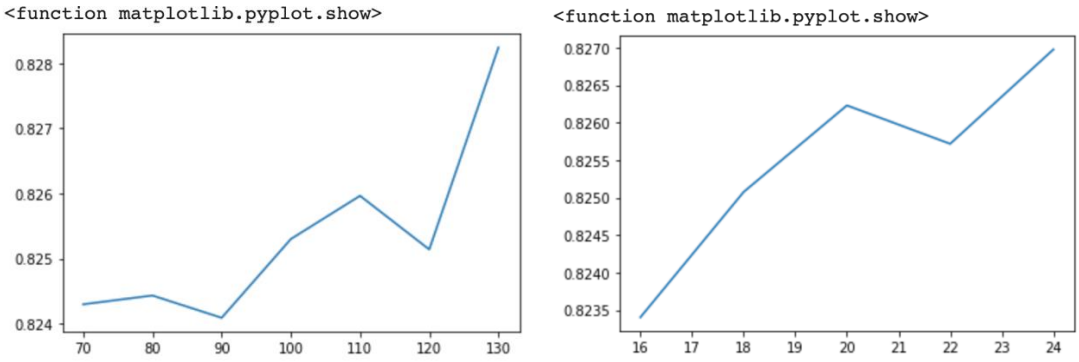

Figure.5. Variation of MAE in SVD with respect to the n_factors(L) and n_epochs(R).

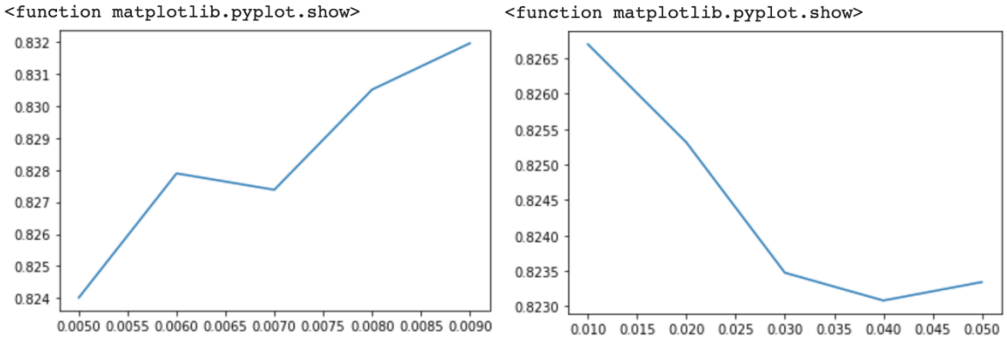

Figure.6. Variation of MAE in SVD with respect to the lr_all(L) and reg_all(R).

```
user 481224 recommendations :
Cinnamon Rolls II with ingredients ['butter', 'confectioner sugar', 'cream cheese', 'egg', 'flour', 'ground cinnamon', 'milk', 'salt', 'sugar', 'van
illa extract', 'yeast', 'biscuit quarter']
Apple Crisp III with ingredients ['apple', 'flour', 'ground cinnamon', 'sugar', 'herb-and-spice blend']
Baby BLT with ingredients ['bacon', 'cherry tomato', 'mayonnaise', 'onion', 'parsley']
```

Figure.7. The result of recommending healthy recipe along with the ingredient for a user

```
Replacing "brown sugar" with "biscuit quarter" in the ingredient set ['brown sugar', 'butter', 'confectioner sugar', 'cream cheese', 'egg', 'flour',
'ground cinnamon', 'milk', 'salt', 'sugar', 'vanilla extract', 'yeast']  to get ['butter', 'confectioner sugar', 'cream cheese', 'egg', 'flour', 'gr
ound cinnamon', 'milk', 'salt', 'sugar', 'vanilla extract', 'yeast', 'biscuit quarter']
Old fsa  = 8
New fsa = 4.174786175809525
Recipe  Cinnamon Rolls II
Replacing "butter" with "herb-and-spice blend" in the ingredient set ['apple', 'butter', 'flour', 'ground cinnamon', 'sugar']  to get ['apple', 'flo
ur', 'ground cinnamon', 'sugar', 'herb-and-spice blend']
Old fsa  = 8
New fsa = 5.137705117132132
Recipe  Apple Crisp III
```

Figure.8. Result of replacing an ingredient in the unhealthy recipe to create a new healthy recipe

REFERENCES

- Dataset: https://www.kaggle.com/datasets/elisaxxygao/foodrecsysv1
- http://snap.stanford.edu/class/cs224w-2019/project/26410425.pdf
- https://gephi.org/users/
- https://networkx.org/
- https://www.nltk.org/
- https://scikit-learn.org/stable/
- https://surpriselib.com/
- https://www.tutorialspoint.com/scikit_learn/index.htm
- https://medium.com/swlh/a-tutorial-on-networkx-network-analysis-in-python-part-i-43c1d35830b6
- https://github.com/NicolasHug/Surprise
- https://en.wikipedia.org/wiki/Traffic_light_rating_system
- https://github.com/NicolasHug/Surprise/blob/master/examples/top_n_recommendations.py
- https://surprise.readthedocs.io/en/stable/matrix_factorization.html
- https://albertauyeung.github.io/2017/04/23/python-matrix-factorization.html
- https://surprise.readthedocs.io/en/stable/knn_inspired.html