# Modularity Based Recipe Generator

Eren Arkangil

## I. INTRODUCTION

The combination of total number of recipe can be generated from ingredients is approximately on the order of 10 to the 15th power and the actual number of current recipes humans prefer to eat is around one million. This information can be easily derived with a quick calculation, yet the point we reached here is worth to research, the number of current used recipes is a very small fraction of all combinations.That means, there is an ongoing challenge for chefs is to discover any pattern or laws that ends up with tasty combinations and use them to create new recipes to be experienced by consumers.

## II. DATASET AND DISTRIBUTION

Albert-László Barabási and his coworkers discovered an important principle of flavor combination by studying foods with complex networks in ‘‘Flavor network and the principles of food pairing” article[1]. Nodes were ingredients, where edges represented shared flavors in the paper. This new insight could help create novel recipes.

### A. Dataset

The dataset consist of 9945 recipes with its ingredient information. The Id is the identifier of each recipe. Yummly, the Personalized Recipe Recommendations and Search website is the provider of the dataset [2].
Our data has 2380 nodes as ingredients and 7811 edges between these nodes. Its average clustering is 0.093 and degree 6.56.

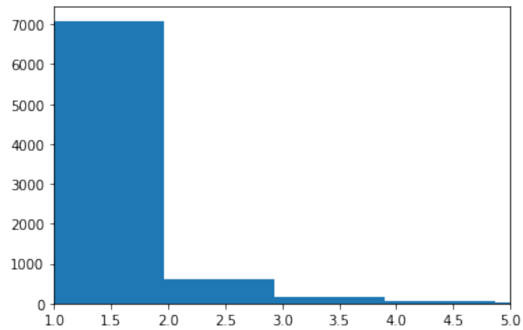


Fig. 1: x axis: the degree, y axis: number of nodes

The dataset is possessed of information recipes scraped from internet as json files. Therefore it needs to be converted into a dictionary first to be available for the Project. Then for simplicity i created a csv file to work with pandas. Once the network structure is set up the network has 2380 nodes where each of them is different ingredient. There are 7811 edges between these nodes. The graph is undirected, furthermore its average degree is 6.563 whereas its average clustering is 0.0936. Clustering coefficient is the measurement of the overall level of clustering in a network and gives the average of the local clustering coefficients of all the nodes. Therefore this number gives an information of how much does the network inclined to have local clusters. In our case it is approximately 0.1, which emphasizes that there is % 10 chance that two neighbors of a node are linked. Average degree is simply the average number of edges per node in the graph. There is power distribution as you can see most of the nodes has degree 1, the nodes such as salt has more weights since almost every food includes salt

### B. Degree distribution

The degree distribution $p_k$ provides the probability that a randomly selected node in the network has degree of k. For any integer $k \geq 0$, the quantity $p_k$ is the fraction of nodes having degree $k$. This is also the probability that a randomly chosen node in the network has degree $k$. The quantities $p_k$, for $k \geq 0$, represent the degree distribution of the network.

The following plot below illustrates the Degree distribution on y axis and the number of Degrees on the x axis.

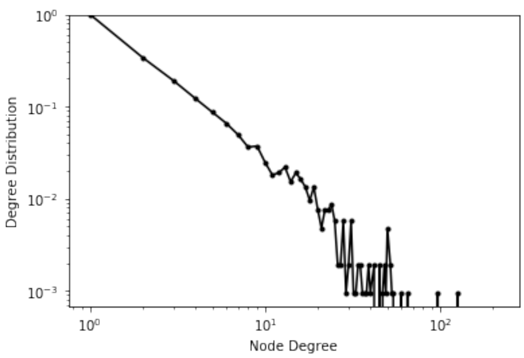


Fig. 2: x axis: fraction of degree distribution y axis: node degree

As it can be seen the probability of a node to have a high degree is less likely than having a low one, this is, most are highly right-skewed, meaning that a large majority of nodes have low degree but a small number have high degree. What we can deduce from the graph above that our network follows power law distribution.

### C. Clustering Coefficient

The clustering coefficient captures the degree to which the neighbors of a given node link to each other. For a node i with degree $k_i$ the local clustering coefficient is defined as $C_i = 2L_i/(k_i)(k_i - 1)$, where $L_i$ represents the number of links between the $k_i$ neighbors of node i. The average clustering of the network is obtained in the following cell, which is 0.1. It is the overall level of clustering in a network, and it basically calculates the average of the local clustering coefficients of all the vertices, and therefore informs of how much does the network tend to create local clusters. In this case it is close to 0.1, which implies that there is about % 10 chance that two neighbors of a node are linked.

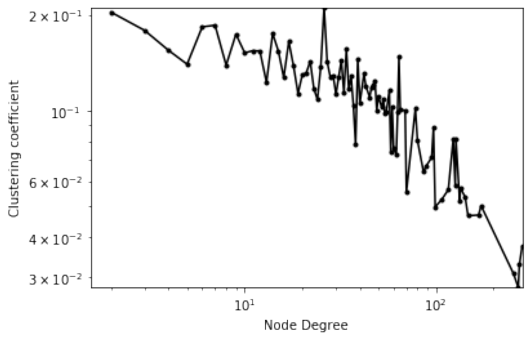


Fig. 3: x axis: clustering coefficient y axis: node degree

Our average clustering is 0.963. It seems intuitive due to the nature of the network that ingredients with a very low degree, therefore small amount of recipes will tends to create clusters with nearby ingredients, given that they will be less interconnected with other common ingredients. This is what is reflected in the Clustering coefficient vs degree graph, where the clustering coefficient decreases the higher the degree of a node. In order to travel from one of these low degree nodes, to another cluster several jumps will be needed, thus one should be in the right side nodes of the graph. Common ingredients however are more connected with other common ingredients in the world cuisine, and not so much with some ingredients, which might only be connected with some other 2 or 3 other ingredients.

## III. Problem statement and formulation

This project provides an analysis of ingredients from various recipes including the connections between ingredients nodes and centralities. Therefore, it aims to create recommendations for new recipes.

Ingredient network is represented In the bipartite form, an ingredient-recipes network consist of node which the ingredients used in the recipes. The links connecting different types of nodes are undirected, represent certain compound occur in each ingredient. The edge between two ingredients shows that they were used in the same recipe.

An important feature that the ingredient network will show is the principle of food pairing. A well-known hypothesis states that recipes sharing ingredient compounds are more likely to taste well together than recipes that do not.

First we must introduce some basic ideas about complex networks. As its name indicates, a network is a group of nodes connected between them by links. When the network is big enough, some complex behavior appears; for instance, the small-world effect: in a big net, the mean distance between any pair of nodes can be very small. Following are some properties that are useful when studying a recommendation system.

### *A. Centrality*

There have been studies in the concept of centrality to quantify the importance of a node in a network. Many measures have been defined in this context; the simplest is the degree: the more a node is connected with other nodes, the more that node should be central for the network. A more complex approach to evaluate the centrality of a node is the PageRank algorithm [3], developed to quantify the importance of web pages. According to degree centrality most central nodes in the network are salt: 0.118 , olive oil: 0.114 , eggs: 0.055 and garlic: 0.053. Therefore basically this measure is mainly informing of the most common ingredients, given that most of the recipes includes salt and oil in it.

### *B. Bipartite graphs*

Bipartite graphs,where is no edge that connects vertices of same set a strong tool when an analysis is required for food pairing [4]. I build a bipartite network consisting of nodes which are ingredients of the recipes. The ingredients are connected if they are used in the same recipe.

## IV. Generator Model

The proposed solution to generate new edible recipes needs to detect communities according to the similarities between nodes in the network. Firstly the Gırwan-Newman algorithm is tried but since the network is so huge it computationally takes too much time. Hence, i used an heuristic algorithm called Clauset-Newman-Moore greedy modularity maximization. Greedy modularity maximization begins with each node in its own community and joins the pair of communities that most increases modularity until no such pair exists and yields sets of nodes, one for each community.

## V. Experimental Results

The generated recipes look unexpected good and this work is worth to proceed further. The following illustrations show the power of complex network approach for creating and recommending new recipes. The ingredients of new generated data gives the hint that the algorithm has the notion of type of food and its cuisine.

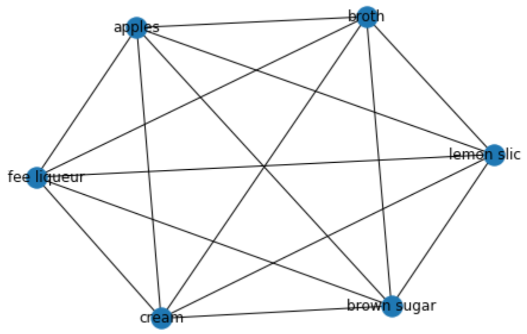


Fig. 4: A dessert example

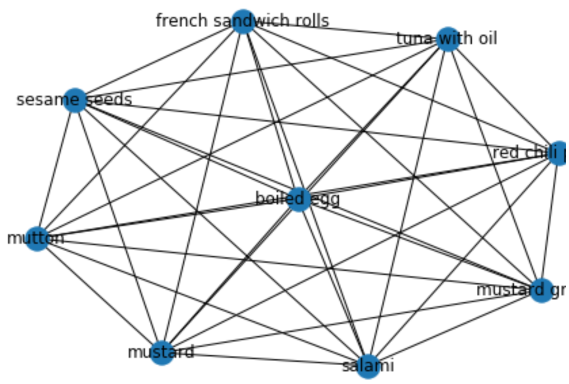


Fig. 5: A lunch recipe

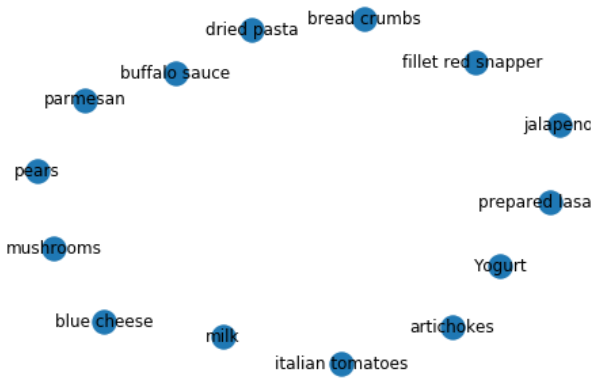


Fig. 6: Italian cuisine recipe