

A network approach to cooking

Project definition:

We want to try to find patterns in the ingredients of the existing recipes in order to recommend a machine-made recipe. Given the importance of food in our society and daily life, we came up with different questions we would like to explore and analyze, for example: How are ingredients related? Which ingredients are often used together? Can we create new recipes using a network approach? What ingredients should not be used together? Can we replace unhealthy ingredients with healthy ones? We will try to answer all these questions using the Recipes 1M dataset.

Dataset:

We will use the dataset Recipes 1M, that contains circa 1M recipes, with 13M images depicting such recipes. Unfortunately for technical constraints we do not plan to use them yet. We can extend the dataset by using OpenFood dataset, which is an open-source dataset containing information about nutritional facts of ingredients, allergens, and all the information you can find on product labels.

Plan:

Creating the graph: We will use ingredients as nodes. Two ingredients will be connected if they appear in common in a number of recipes higher than a threshold.

Cleaning the data: As it is extracted from text, the name of the ingredients is very noisy, so we can assume that ingredients appearing only once are noise, and as such we could either remove them from the graph or assume that they are the valid ingredient with the closest name.

Data Exploration: We will study the cleaned graph, to determine which ingredients are hubs. We will determine the diameter (the farthest ingredients can be interesting; can we still try to put them together in a recipe? Do recipes with these two ingredients exist?). We will try to use the graph to do recipes recommendations and to create new recipes. We will also do classification on the ingredients and try to replace some unhealthy ingredients by better ones.

Execution: We plan to use spectral clustering and others community detection algorithms, i.e Louvain or other, in order to recognize clusters and identify the most popular ingredients. We then think to use an ensemble method of the different clustering algorithms in order to come up with better recommendations.

Objective: Our final goal for this project is to test some new recipes created with this network approach!