"Ss. Cyril and Methodius" University in Skopje
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

Analyze and recreate the "Flavor network and the principles of food pairing" study

Unstructured databases and XML

Bojana Cvetkovska 14045

Atanas Kostovski 143035

04.02.2019

Introduction

The purpose of this project is to try to recreate the results from a previous study, “Flavor network and the principles of food pairing”.

The diversity in regional cuisines, raises the question of whether there are is a pattern that determines which ingredients are combined with one another and whether there are different patterns between regions.

A hypothesis, which over the past decade has received attention among some chefs and food scientists, states that ingredients sharing flavor compounds are more likely to taste well together than ingredients that do not. This study researches how the flavor compounds of each ingredient impact the combination of ingredients in recipes.

Methodology

1. Flavor compounds in ingredients

Each ingredient has a number of flavor compounds that contribute to its flavor. If we present each ingredient as a node and the compounds, they share with each other as a weighted link, we get a weighted network.

The compound concentration in each ingredient is not considered due to lack of information.

2. Extracting the backbone

Since several flavor compounds are shared by a large number of ingredients, the weighted flavor network is too dense for visualization. Therefore, a backbone extraction method is used to remove insignificant links between nodes.

The flavor network helps us reformulate the food pairing hypothesis: do we more frequently use ingredient pairs that are strongly linked in the flavor network, or do we avoid them?

3. Food pairing

To test the hypothesis above, we need to see which ingredient combinations humans prefer. That information is readily available in the form of recipes. In order to avoid a Western interpretation of the world's cuisine, the recipes are

provided from different repositories (epicurious.com, allrecipes.com and menupan.com). They are grouped into geographically distinct cuisines (North American, Western European, Southern European, Latin American, and East Asian). By doing this separation, we can see which ingredients are preferred by the distinct cuisines. We can also see how many ingredients per recipe each cuisine uses.

4. Difference in cuisines

To check if humans prefer ingredients that share flavor compounds, we examine the ingredients used in each recipe. Since recipes are grouped in distinct cuisines, we can see if there are any differences in preferences between different regions.

5. The flavor principle

The differences between regional cuisines can be reduced to a few key ingredients with a specific flavor. Can we identify

the ingredients responsible for the taste palette of a cuisine? To do this we measure the authenticity of each ingredient. Then we can see the shared compounds between each authentic ingredient. The shared compounds between the ingredients are shown as a weighted link.

6. Shared ingredients

The authentic ingredients of each cuisine allow us to further explore the ingredient similarity between regional cuisines.

For this, 6 of the most authentic ingredients of each cuisine are selected and combined in a diagram that illustrates the authentic ingredients shared by various cuisines. It also singles out those that are unique to a particular region.

Results

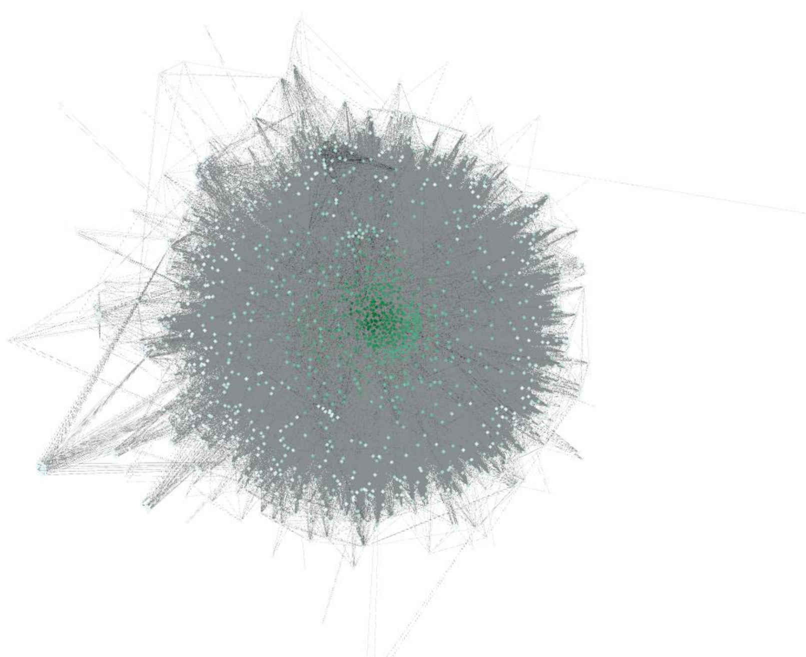
1. The Datasets

To recreate the results in the study, we had access to 2 datasets in csv format:

- containing ingredient links and their shared compounds (ingr1, ingr2, shared_compounds)
- containing recipes grouped by regions (country, ingr1, ingr2, ingr3)

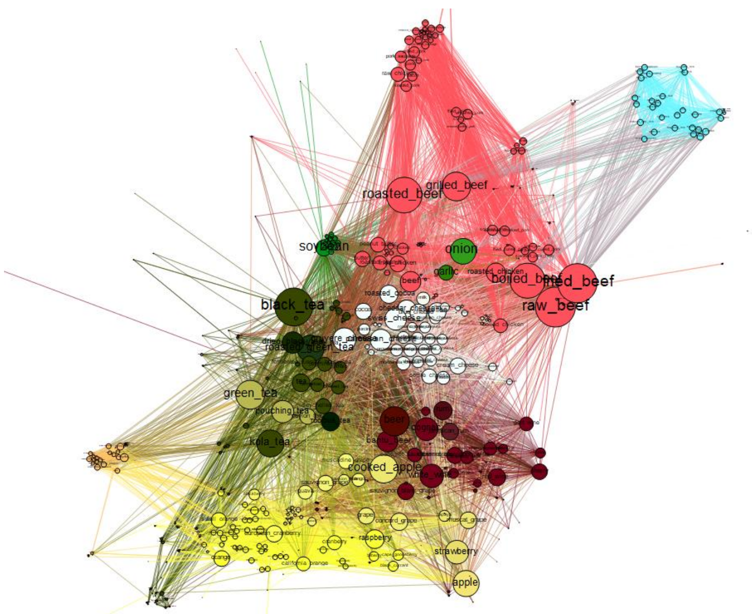
2. Flavor compounds network and extracting the backbone

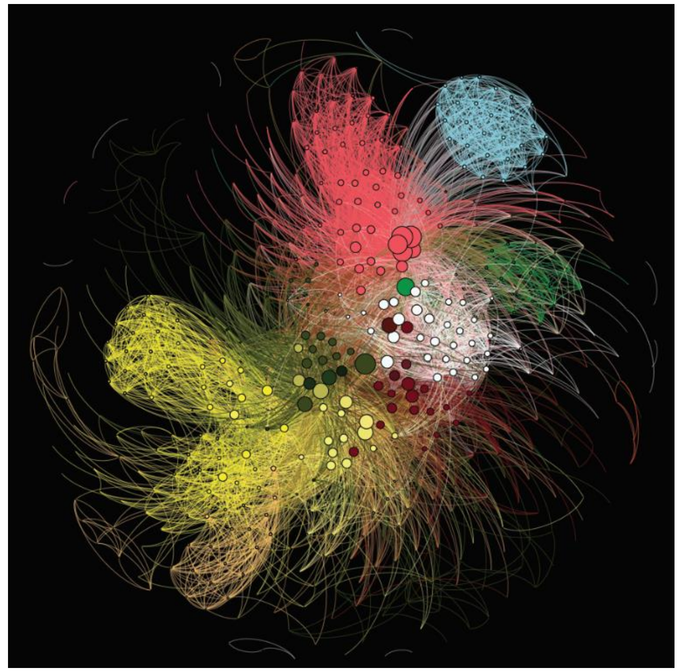
Transforming the first dataset into a weighted network created a very dense, hard to visualize network.



To extract the backbone, we used a disparity filter. The disparity filter gives each node a value of significance (alpha value).

Then it removes every node that has a smaller value than the alpha value that we choose. The study we were recreating used an alpha value of A=0.04, therefore we used the same value. This resulted in a more readable, easy to visualize network.





- We can see which ingredient falls into certain food group, and which food groups are closer to one another as well as the important ingredients that group them

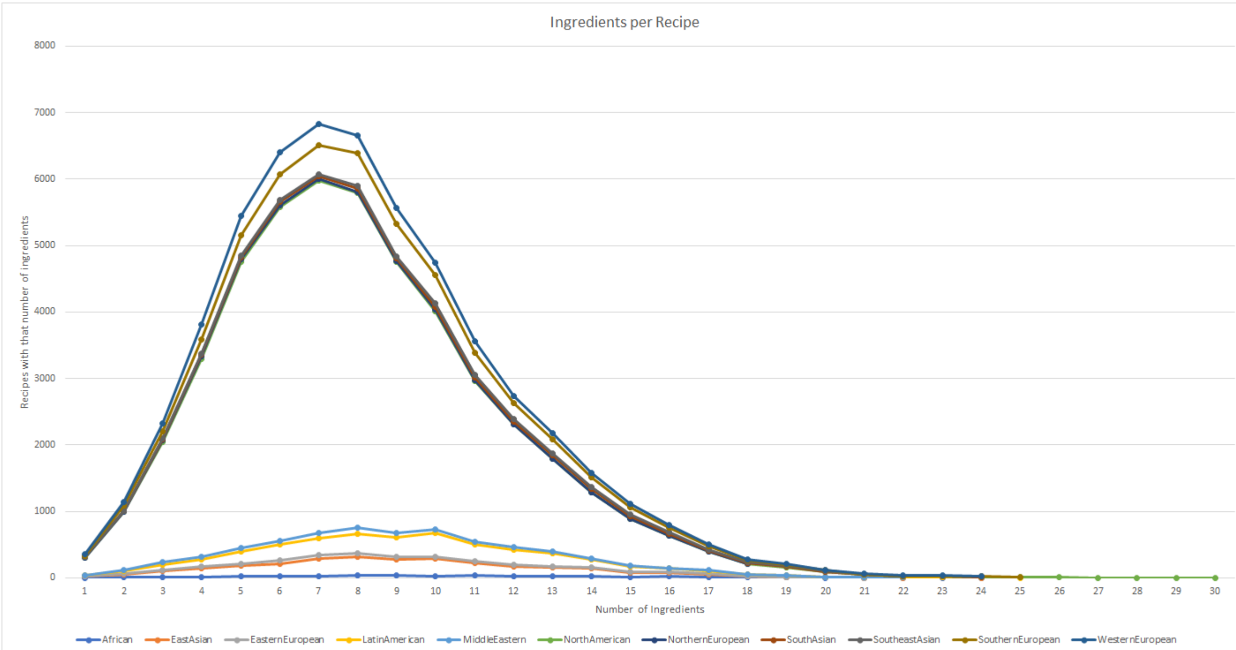
3. Food pairing

To get information on regional recipes, we used the second dataset. We extracted the data into dictionaries which we used to draw charts, making the information easier to visualize.

a. Ingredients per cuisine

To get all the values for numbers of ingredients per cuisine, we parsed the data of the second dataset into a dictionary:

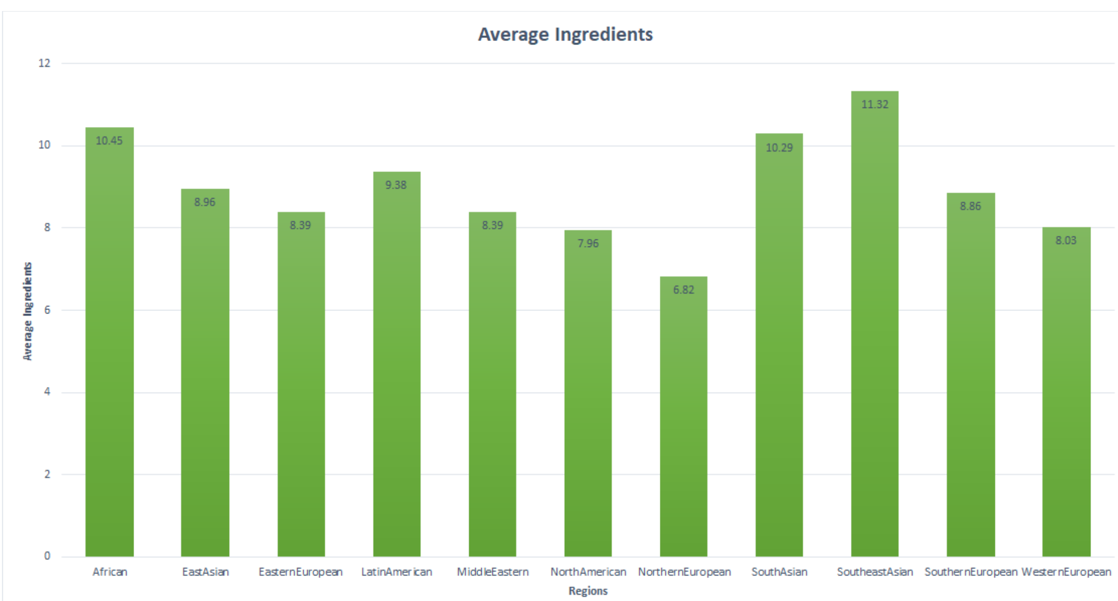
That we turned into a chart:



b. Average ingredients per cuisine

To see the average of ingredients per cuisine, we parsed the data of the second dataset into our own .csv dataset (region, avg_num)

Turning the results into a dataset made it easier to create a chart for visualization



4. Shared compounds

In order to see which cuisine prefers ingredients with shared compounds and which one doesn't, we used the second dataset to examine all the ingredients in the recipes in a cuisine, and the first dataset to extract the number of compounds between each ingredient used in every recipe.

For a recipe with three ingredients, we count the number of shared compounds in every possible pair of ingredients and divide it by the number of possible pair of ingredients. Then we get the mean value for every cuisine.

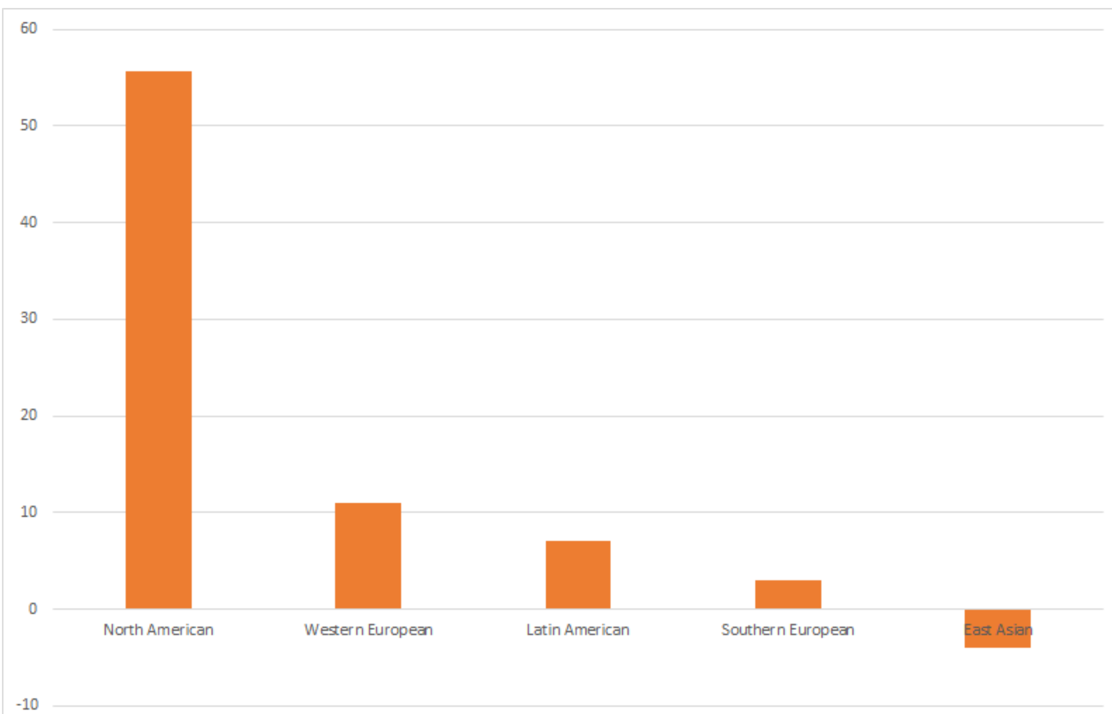
Sol. For each row in dataset2 we get the sum of shared compounds between every ingredient.

Then, we divide that sum by the number of possible links between ingredients in the recipe/row.

That value is stored in a dictionary whose keys represent the regions and each value is the compounds value per recipe that we get above:

Ex. { 'African' : [3.0, 14.44, ...] , ...}

After getting all the values for each recipe, we can simply get the average compounds per recipe per cuisine



The results imply that in North American recipes, the more compounds are shared by two ingredients, the more likely they appear in recipes. However, this doesn't prove that humans prefer to combine ingredients that share compounds because the results also imply that in East Asian cuisine, the more flavor compounds two ingredients share, the less likely they are used together.

5. Authentic ingredients

To further study the shared compounds connection in recipe creation, we extract the authentic ingredients of each cuisine and check the shared compounds between them. The study we were trying to recreate studied the difference between North American and East Asian cuisine, so we did the same.

We took the 6 most authentic ingredients of each cuisine and compared the compounds between them. To find out the authenticity we calculated the prevalence of each ingredient in every cuisine.

Sol. Pi = n/N -where n is the number of recipes in a cuisine that contain the particular ingredient and N is the total number of recipes in a cuisine. We wrote down the values in a dictionary:

Ex. { 'turnip': {'African': 0.025, 'EastAsian': 0.002, ...} }

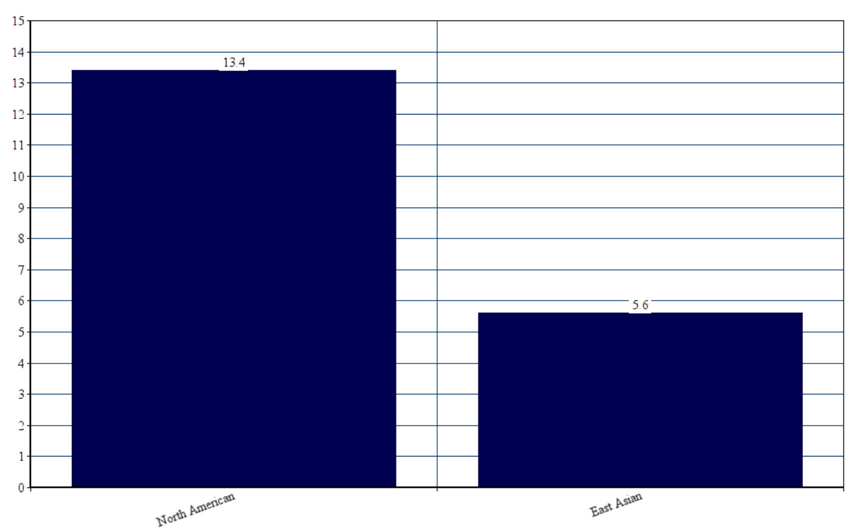
Then we took the ingredients with the highest values for North American and East Asian cuisine. We wrote down the values in small .csv datasets:

(Ingredient, prevalence)

Ingredient	Prevalence
rice	0.294
red_bean	0.152
milk	0.055
green_tea	0.041
butter	0.041
peanut	0.038

Ingredient	Prevalence
milk	0.529
butter	0.511
cocoa	0.377
vanilla	0.239
cream	0.154
cream_cheese	0.154

If we compare the shared compounds now, we can still see that North American cuisine prefers ingredients that share compounds, whereas East Asian cuisine avoids them.



Conclusion

In this project we managed to recreate the results from the study, Flavor network and the principles of food pairing. We showed that flavor compounds in ingredients might affect the way ingredients are combined.

However, we also showed that while some regions prefer ingredients that share more compounds, other regions prefer the opposite. The reason as to why this happens is not researched in this project. To improve these results, the concentration of the compounds in every ingredient should be taken into account.

Also, this research doesn't consider the importance of recipes in a cuisine. Calculating the authenticity of ingredients is a good starting point, but it can be improved if there was a dataset created where each region would have nationally approved recipes that are characteristic for that region. This would improve the chances of calculating the right ingredients for each region.

Another interesting study would be to choose the most famous recipes, approved and liked by people in every region. Then we would be able summarize whether people from around the prefer recipes that have ingredients with shared compounds or not.

References

- Flavor network and the principles of food pairing
  Yong-Yeol Ahn, Sebastian E. Ahnert, James P. Bagrow & Albert-László Barabási
  - https://www.nature.com/articles/srep00196
- Network
  https://networkx.github.io/
- numpy
  http://www.numpy.org/
- Github repository with the project code for parsing the data
  https://github.com/BonneC/flavor-RM