A Network Tour of Kitchen Science: Graph-Based Ingredient Suggestion and Replacement

Erick Maraz, Lara Orlandic, Niccolo Stefanini, Valentin Viennot

I. Story

Adopting a vegetable-based diet can greatly reduce an individual's carbon footprint as studies show [1]. However, replacing meat products with plant-based ones can be challenging. What if you could make any recipe vegetarian or vegan with a magic wand? Similarly, what if you could automatically replace an ingredient to which you are allergic with the best possible alternative?

As the magic wand, we propose a tool to remove certain ingredients from a recipe and output the best ingredients to replace them. To do so, we use regularization on a similarity graph of ingredients to get suggestions. We then judge the likeliness of obtained recipes using geometric deep learning.

II. Data Acquisition

We will use the Recipe1M+ dataset, containing over one million recipes from various cooking websites. The features include ingredients, nutrition facts, preparation instructions, and health scores [2].

Since the ingredient detection is noisy, we might need to cross the data with a finite list of possible ingredients, such as the the U.S. Agricultural Research Service dataset [3]. Furthermore, to reduce computational cost, we may use smaller subsets of the dataset.

III. Data Exploration

We will build a weighted adjacency matrix of ingredients based on which ingredients frequently appear together in recipes. We will then determine whether or not there are "hub" ingredients by analyzing the clusters and degree distribution of the graph.

This network should result in a strongly connected giant component, since there are more than one million recipes and an average of 10 ingredients per recipe. Therefore, we will apply an RBF kernel to sparsify the matrix. This will also enable us to perform dimensionality reduction to create an alternative ingredient clustering.

We could then take our graph exploration a step further and determine whether ingredients with similar health statistics tend to cluster together, and compute the Fourier transform of the graph based on nutrition information to determine how quickly the ingredients' health statistics change from one node to the next. This will give us insights on how to implement suggestions of healthier ingredients.

IV. Data Exploitation

A straightforward method of ingredient replacement suggestion would be to use K-NNs. However, this may oversimplify the problem and produce illogical answers (e.g. replacing chicken with olive oil which are often used together). Alternatively, we could use the node2vec [4] library to identify ingredient similarities, or link prediction between unconnected ingredients.

For this project, we propose the following:

1) Project a recipe on the graph as a boolean or ingredient quantity signal
2) Remove one or more ingredients: set binary value or quantity to 0
3) Use GSP smoothing on the signal (Tikhonov, GTV)
4) The ingredients with the highest values are the most likely replacements

A further step could utilize GCNN pattern recognition as in [5] to validate our ingredient suggestions by classifying the thousands of possibly generated recipe signals as valid or impossible recipes. Furthermore, instead of using binary recipe signals, we could potentially construct a similarity graph using the quantity information of each ingredient, as well as its nutritional information. This would allow suggesting ingredients that have similar nutritional values as the replaced ingredients. Finally, clustering of ingredients based on diet or user preferences could be done to automate specific removals.

REFERENCES

[1] A. Rosi, P. Mena, N. Pellegrini, S. Turroni, E. Neviani, I. Ferrocino, R. Di Cagno, L. Ruini, R. Ciati, D. Angelino, J. Maddock, M. Gobbetti, F. Brighenti, D. Del Rio, and F. Scazzina, "Environmental impact of omnivorous, ovo-lacto-vegetarian, and vegan diet," *Scientific Reports*, vol. 7, no. 1, 12 2017.

[2] "Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images - MIT." [Online]. Available: http://pic2recipe.csail.mit.edu/

[3] "USDA Food Descriptions." [Online]. Available: https://www.ars.usda.gov/

[4] "node2vec." [Online]. Available: https://snap.stanford.edu/node2vec/

[5] "rusty1s/pytorch_geometric: Geometric Deep Learning Extension Library for PyTorch." [Online]. Available: https://github.com/rusty1s/pytorch$_geometric$