Artificial Intelligence In Medicine

Automated Identification of Healthier Food Substitutions through a combination of Graph Neural Networks and Nutri-Score

--Manuscript Draft--

Manuscript Number:	
Article Type:	Original research Paper
Keywords:	Graph neural networks; Nutri-Score; Healthier food choice; Nutritional profile; Ingredient substitution; Food similarity
Corresponding Author:	Julie Loesch, M.Sc. Maastricht University Maastricht, NETHERLANDS
First Author:	Julie Loesch, M.Sc.
Order of Authors:	Julie Loesch, M.Sc.
	Ilse van Lier, PhD student
	Alie de Boer, Dr.
	Jan Scholtes, Prof. Dr. Ir.
	Remzi Celebi, Dr.
Manuscript Region of Origin:	Europe
Abstract:	The increased incidence of chronic diseases globally has made a healthy diet increasingly important, as it is one of the factors that can support maintaining health. Unhealthy diets are associated with an increased risk of a variety of health issues and diseases. However, even though healthy, nutritious diets are known to be vital as one of most basic human needs, individuals often struggle to make healthy dietary choices. A potential solution to this problem is to develop a system that can recommend "healthier" alternative ingredients to ingredients in an individual's current diet to encourage people by changing their diets. To this end, our work applies the classical and most common variants of Graph Neural Networks (GNNs), which are GraphSAGE and Graph Attention Network (GAT), to automatically generate food substitutions. Our best results were achieved with GAT and by putting the recommended substitutes in the same food category as the query food. The specific results achieved were a Recall Rate of 0.647 and 0.733 for the top 5 and 10 results, respectively, which is good performance for a recommender system. We also employed the French nutritional rating system Nutri-Score to identify "healthier" food recommendations. Hence, our final recommendation algorithm consists of two parts. The first part automatically generates food substitutions using Graph Neural Networks; the second part identifies "healthier" food recommendations employing the Nutri-Score algorithm and ranks these "healthier" food options according to their highest Nutri-Score. Ultimately, we found out that by simply ranking our food recommendations by the highest Nutri-Score is enough to claim that the final substitutions have a better nutritional profile.
Suggested Reviewers:	Tome Eftimov tome.eftimov@ijs.si Tome Eftimov is a well-known researcher focusing on food and nutrition data analytics. His other areas of research include statistical data analysis, natural language processing, machine learning, and information theory.
	Robert Hoehndorf robert.hoehndorf@kaust.edu.sa Robert Hoehndorf is the leader of the Bio-Ontology Research Group at KAUST, one of the top research centers for ontology, biomedical informatics, and knowledge representation. He has extensive knowledge of food and health domain, as well as profound expertise in knowledge graphs and the deep learning models.

Dear Editor,

We would like to submit our manuscript entitled 'Automated Identification of Healthier Food Substitutions through a combination of Graph Neural Networks and Nutri-Score' to Artificial Intelligence in Medicine. Considering that your journal aims at supporting decision-based medical tasks through the use of artificial intelligence, we believe that it would be the perfect home for our paper.

Our study develops an automated approach that finds healthier food substitutions through a combination of Graph Neural Networks and Nutri-Scores. Whilst healthy diets are known to be important, individuals do not always make healthy dietary choices. There are many reasons why someone might not eat a healthy and nutritious diet, but one of the main issues seems to be selecting/choosing healthier food options.

Despite guidelines and recommendations for healthy eating, it is still proven difficult to adhere to these guidelines. One of the arguments raised is that these recommendations are insufficiently personalised. Personalising recommendations could be a way to stimulate making healthier food decisions. One approach to personalising dietary recommendations could be to assist people in choosing alternative ingredients that have a healthier nutrient profile. By developing a system that can recommend healthier alternative ingredients to ingredients in an individual's current diet, healthier choices can be made easier. In our paper, we describe how Graph Neural Networks and Nutri-Scores allow for identifying healthier food choices, which can support individuals in making more personalised decisions. We believe that this developed system offers interesting insights for a wide range of scholars, in particular those interested in supporting decision-based tasks.

With kind regards,
On behalf of all co-authors,

Julie Loesch

Highlights:

- Creation of a novel knowledge graph that leverages semantic information with the aim of identifying food substitutions.
- Providing an implementation that generates food substitutions using Graph Neural Networks.
- Developing an approach that employs the Nutri-Score algorithm along with food categorization to determine healthier food substitution options.

Automated Identification of Healthier Food Substitutions through a combination of Graph Neural Networks and Nutri-Score

Julie Loesch[a], Ilse van Lier[b], Alie de Boer[c], Jan Scholtes[a], Remzi Celebi[a,d,*]

[a] *Department of Advanced Computing Sciences, Maastricht University, Paul-Henri Spaaklaan 1, Maastricht, 6229 EN, Netherlands*
[b] *Chair Youth, Food, and Health, Maastricht University Campus Venlo, Nassaustraat 36, Venlo, 5911 BV, Netherlands*
[c] *Food Claims Centre Venlo, Campus Venlo, Faculty of Science and Engineering, Maastricht University, Nassaustraat 36, Venlo, 5911 BV, Netherlands*
[d] *Institute of Data Science, Maastricht University, Paul-Henri Spaaklaan 1, Maastricht, 6229 GT, Netherlands*

Abstract

The increased incidence of chronic diseases globally has made a healthy diet increasingly important, as it is one of the factors that can support maintaining health. Unhealthy diets are associated with an increased risk of a variety of health issues and diseases. However, even though healthy, nutritious diets are known to be vital as one of most basic human needs, individuals often struggle to make healthy dietary choices. A potential solution to this problem is to develop a system that can recommend "healthier" alternative ingredients to ingredients in an individual's current diet to encourage people by changing their diets. To this end, our work applies the classical and most common variants of Graph Neural Networks (GNNs), which are GraphSAGE and Graph Attention Network (GAT), to automatically generate food substitutions. Our best results were achieved with GAT and by putting the recommended substitutes in the same food category as the query food. The specific results achieved were a Recall Rate of 0.647 and 0.733 for the top 5 and 10 results, respectively, which is good performance for a recommender system. We also employed the French nutritional rating system Nutri-Score to identify "healthier" food recommendations. Hence, our final recommendation algorithm consists of two parts. The first part automatically generates food substitutions using Graph Neural Networks; the second part identifies "healthier" food recommendations employing the Nutri-Score algorithm and ranks these "healthier" food options according to their highest Nutri-Score. Ultimately, we found out that by simply ranking our food recommendations by the highest Nutri-Score is enough to claim that the final substitutions have a better nutritional profile.

Keywords: Graph neural networks, Nutri-Score, Healthier food choice, Nutritional profile, Ingredient substitution, Food similarity

1. Introduction

Adhering to a healthy diet is becoming more and more important in light of the rising rates of chronic diseases [1]. An unhealthy diet has been identified as one of the leading causes of health risks globally, and is associated with an increased risk of a range of health issues and diseases including, but not limited to, poor bone health, high blood pressure, type 2 diabetes, cardiovascular disease and various types of cancers [2, 3]. Health crises such as the COVID-19 pandemic also highlight the importance of a healthy diet, as dietary and health status have been shown to influence people's ability to prevent, combat, and recover from infections [4, 3]. Although they are not cures for infections or diseases, healthy diets are important to maintain an individual's immune system [5, 3].

Whilst healthy diets are known to be important, individuals do not always make healthy dietary choices. There are many reasons why someone might not eat a healthy and nutritious diet, but the most common and also obvious reason is the lack of affordability. In fact, approximately three billion people cannot afford a healthy diet, whilst more than three billion people suffer one or more manifestations of poor nutrition [6]. Those people often have several micronutrient deficiencies, meaning that they lack essential vitamins and min-

*Corresponding author
Email address: `remzi.celebi@maastrichtuniversity.nl` (Remzi Celebi)

erals required in small amounts by the body for proper growth and development.

However, it should also be pointed out that even people who can afford a healthy diet struggle to choose healthy food options. Despite guidelines and recommendations for healthy eating, such as the guidelines set by the Dutch National Health Council [7], it is still proven difficult to adhere to these guidelines. One of the arguments raised is that these recommendations are insufficiently personalised [8]. Personalising recommendations could be a way to stimulate making healthier food decisions [9]. One approach to personalising dietary recommendations could be to assist people in choosing alternative ingredients that have a healthier nutrient profile. By developing a system that can recommend "healthier" alternative ingredients to ingredients in an individual's current diet, healthier choices can be made easier.

One of the first efforts to automate the selection of food substitutions have leveraged semantic information; Shirai et al. [10] developed a heuristic called *DIISH* to automatically rank plausible ingredient substitutions, by using semantic information from FoodKG [11], a knowledge graph of recipe and ingredient information. However, their methods were tested on validation data that had not been verified by domain experts. In addition, it offers "healthy" ingredient substitution options, taking into account only certain dietary restrictions (e.g., vegetarian diets and allergies) and limiting the consumption of certain foods. In our earlier study, we developed an automated approach that finds alternative food items with comparable nutritional profiles that fall within a similar product category using knowledge graph embeddings (KGE) [3]. In the same study, we also addressed the low quality of the food substitution dataset that Shirai et al. [10] used, by curating an expert-verified dataset for the evaluation. Although embedding-based methods are known to have some limitations on complex logical reasoning, some recently proposed Graph Neural Networks (GNNs) based methods are very promising for handling complex reasoning on knowledge graphs [12]. With the help of GNNs, it is possible to extract both entity characteristics and relations from knowledge graphs, which is an essential factor for food-knowledge-graph-based applications, such as prediction of compound-food relationship and food recommendation [12]. To this extent, this work can help people make "healthier" food choices, which have similar nutritional characteristics than the query food, through the use of Nutri-Scores and GNNs.

This study presents an automated approach that identifies "healthier" food substitutions using a nutrient profiling system along with semantic similarity and Graph Neural Networks. Our contributions are as follows.

- Creation of a novel knowledge graph that leverages semantic information with the aim of identifying food substitutions.
- Providing an implementation that generates food substitutions using Graph Neural Networks.
- Developing an approach that employs the Nutri-Score algorithm along with food categorization to determine "healthier" food substitution options.

Consequently, in this work, we apply the classical and most common variants of Graph Neural Networks, which are GraphSAGE [13] and Graph Attention Network (GAT) [14], to automatically generate food substitutions. We obtained the best results with GAT [14] by restricting the recommended substitutes to be in the same food category than the food in question. The specific results achieved were a Recall Rate of 0.647 and 0.733 for the top 5 and 10 results, respectively, showing significant performance improvement compared to previous KGE-based approaches. We also employed the French nutritional rating system Nutri-Score to identify "healthier" food recommendations [15]. Our experiments show that the Nutri-Score algorithm can be used to identify food options which have a better nutritional profile compared to the query food.

2. Related Work

Researchers have leveraged a wide range of resources on datasets for analyzing foods, dishes or recipes similarity. To name a few, FlavorDB collects information such as flavor molecules [16], FooDB collects data on compounds and nutrients[1] and FoodOn [17] on food categorization. Other resources to collect recipes and related information have also been developed, such as RecipeDB [18], Recipe1M [19], AllRecipes[2], and Yummly[3]. OpenFoodFacts[4], a food product database, is another resource that collects features on food products. Furthermore, efforts have been made to bring together various sources of information, such as the work of Haussmann et al. [11], which presents FoodKG, a knowledge graph that incorporates the ingredients from

[1] https://foodb.ca/
[2] https://www.allrecipes.com/
[3] https://www.yummly.com/
[4] https://world.openfoodfacts.org/

USDA[5], food ontology from FoodOn [17], and recipes from Recipe1M [19].

Some studies have focused on embedding approaches to represent high dimensional data with low-dimensional vector representations. For instance, Eftimov et al. [20] created vector representations of foods (namely FoodEx2Vec) from the FoodEx2 data, which is a comprehensive system for classifying and describing food items developed by the European Food Safety Authority (EFSA) [21]. The work of Pan et al. [22] further employed embedding based similarity techniques to identify food substitutions. They collected recipe data about different cuisine styles from a website hosting thousands of recipes (e.g., Spoonacular[6]), from which they generated ingredient and recipe embeddings. Specifically, the authors obtained the embeddings for all ingredients by performing the Skip-gram model on collected recipes. The authors then used NLP and cosine similarity to search for ingredient substitutes; however, no formal evaluation of the generated substitutes was provided.

Previous studies have largely employed word embeddings which compute distributional vector space representation of words, aiming to group words that are frequently used together in the text corpus. For example, Shirai et al. [10] identifies food substitutions based on similarities in word embeddings of ingredients. In their approach, the authors used two sources of latent semantic information about ingredients as word embeddings based on ingredient names: the Word2Vec model [23] and the spaCy's word embedding model [24]. Subsequently, each word embedding model was used to compute the cosine similarity between ingredient names separately. Their final heuristic *DIISH* is a combination of the four scores where other two scores capture the similarity between how two ingredients relate to recipes and the similarity of ingredients that co-occur in recipes alongside the target ingredient. The authors obtained excellent results by achieving a Recall Rate of 0.625 and 0.764 for the top 5 and top 10 results, respectively. In addition, the authors used a filtering strategy to filter out unhealthy food substitutes. The authors considered two categories of dietary restrictions, namely restrictions on the types of ingredients that may be consumed (e.g., replacing meat-based ingredients with vegetarian alternatives or replacing allergens such as peanuts), as well as a second constraint of limiting the consumption of certain nutrients (e.g., replacing high-carb ingredients with low-carb alternatives). The work of Tansey et al. [25] presented diet2vec, which is a scalable and robust approach for modelling nutritional diaries from smart phone apps. The authors analyzed massive amounts of nutritional data generated by 55k active users of a diet tracking app called LoseIt[7]. To encode the foods, the authors first ran Word2Vec [23] on the names of the food and subsequently ran weighted k-means to cluster the foods into $5,000$ food words, placing 20% of the weight on the name and 80% of the weight on the nutrients. The clusters would be used to find these substitutes. However, no formal evaluation of the results was provided. Another relevant study is the work of Kazama et al. [26], which proposed a novel system that can transform a recipe into any selected regional style using recipe data from Yummly[3]. For instance, the authors demonstrated how their proposed system can create a French version of a traditional Japanese recipe, Sukiyaki. To do so, the authors first needed to build a neural network model that learns to identify a recipe's regional cuisine style mixture. Moreover, their system can suggest ingredient substitutions that create a recipe that is most authentic to a regional cuisine. The authors accomplished this through an extended Word2Vec [23] model.

While embedding-based methods are known to have some limitations on complex logical reasoning, Graph Neural Networks (GNNs) based methods are very promising for handling complex reasoning on knowledge graphs [12]. As a result, Tang et al. [27] proposed a three-part algorithm by employing GraphSAGE [13] with a focus on providing healthier and tastier recipe alternatives. The first part of their algorithm consists of developing a nutritional model using linear regression to understand the overall nutritional content of a recipe when given its ingredients, while the second part predicts a user's rating scores using GraphSAGE [13] on ingredient-recipe and recipe-user bipartite graphs. For this, the authors used the recipe data from Kaggle[8], which consists of about $50,000$ recipes scraped from AllRecipes[2]. Then, they created a recipe-user bipartite graph with users and recipes as nodes, and directed weighted edges representing the rating that users give for recipes. Moreover, the researchers constructed an ingredient-recipe bipartite graph, in which each recipe is linked to its component ingredients. Finally, in the third part, the authors combined these two models, which allowed them to evaluate the health and taste levels of novel recipes according to users' preferences.

[5] `https://fdc.nal.usda.gov/index.html`

[6] `https://spoonacular.com/`

[7] `lhttps://www.loseit.com/`

[8] `https://www.kaggle.com/datasets/nguyentuongquang/all-recipes`

3. Methodology

This section introduces the proposed algorithm to find food substitutes that have similar nutritional characteristics to the query food. Thus, Subsection 3.1 explains step by step how we used Graph Neural Networks to extract food substitutes. The pipeline of the steps to train a GNN-based link prediction model is shown in Figure 1[9].

3.1. GNN-based Link Prediction Model

3.1.1. Dataset

The input graph to our GNN-based model is the knowledge graph of food. We gathered various sources of information about ingredients from USDA[5], products from OpenFoodFacts[4], flavor molecules from FooDB[1], so-called *Food Tags* [3], which indicate the presence of rich mineral or vitamin content, and a ground truth food substitution dataset [3] that is necessary to perform a semi-supervised link prediction task. The final knowledge graph of food is presented in Figure 2 and Table 1 shows some statistics (e.g., # of triples, # of unique food items, etc.) about each dataset that was used to build the final knowledge graph.

3.1.2. Graph Neural Network

The second step in the pipeline involves learning the graph representation through a message passing and aggregation scheme. Figure 1 shows how a target node (e.g., yellow node A) aggregates all the messages from its neighbors in *one* GNN layer. This implies that the information that is *one* hop away from target node A is transferred to the target node A. The three small gray boxes, that refer to Figure 1, take those feature vectors as input and aggregate them to generate the output represented by the big gray box. Technically speaking, the l^{th} layer computes new node representations (e.g., X_A) by aggregating vector representations from the nodes that are l hop(s) away from the target node A, where X_A is a feature vector of target node A. GNN initializes the activation units first by setting $h_v^{(0)} = X_v$, $\forall v \in \{A, B, C, D\}$. Then, nodes B, C and D will create a message (incorporating the feature vectors) and target node A will aggregate all the messages from the node itself $m_A^{(1)}$ and the neighboring nodes $m_u^{(1)}, u \in \{B, C, D\}$ using this formula $m_u^{(1)} = MSG^{(1)}(h_u^{(0)}), u \in \{B, C, D\}$.

During this step, different variants of Graph Neural Networks, such as GraphSAGE [13], or GAT [14], can be applied. They differ in the ways they aggregate the information across the layers:

- GraphSAGE: $h_v^{(l)} = \sigma(W^{(l)} \cdot CONCAT(h_v^{(l-1)}, AGG(\{h_u^{(l-1)}, \forall u \in N(v)\})))$.
- GAT: $h_v^{(l)} = \sigma(\sum_{u \in N(v)} \alpha_{uv} W^{(l)} h_u^{(l-1)})$, where α_{uv} are the attention weights.

However, as our input data is a knowledge graph/heterogeneous graph which has multiple node or edge types, we extended the model to be able to work with multiple relation types. The idea is that the GAT layer [14] or GraphSAGE layer [13] should treat each edge type differently. The process takes an existing GNN model and duplicates the message functions to work on each relation type individually.

3.1.3. Node Embeddings

By repeating the second step for each node $v \in V$, we obtained node embeddings $h_v^{(l)}$, $\forall v \in V$, as shown in the third small graph in Figure 1.

3.1.4. Prediction Head

In this work, we are interested in edge-level predictions in which predictions are made using pairs of node embeddings. To be able to predict whether an edge between two nodes u and v exists, a common approach for scoring pairs is to calculate the dot product between two node embeddings $h_u^{(l)}$ and $h_v^{(l)}$. In other words, a GNN-based link prediction model computes the likelihood of a connectivity between two nodes u and v as a function of $h_u^{(l)}$ and $h_v^{(l)}$.

3.1.5. Predictions and Labels

In this study, we formulated the problem as a semi-supervised link prediction task, which is an approach that combines a small amount of labeled data with a large amount of unlabeled data during training. Here, the ground truth food substitution dataset [3] is used as the labeled data. Therefore, after computing the likelihood of a link between two nodes u and v, we optimized the loss between the predicted values and the true values.

Precisely, the loss was calculated by the Mean Squared Error (MSE), which basically computes the difference between the predicted values and the true values. Formally, $MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, where n is the number of data points, Y_i is the true value of data point i, and $\hat{Y}_i$ is the predicted value of data point i. The squaring is necessary to remove any negative signs and the metric is called the *mean* squared error because it

[9] `https://web.stanford.edu/class/cs224w/slides/08-GNN-application.pdf`

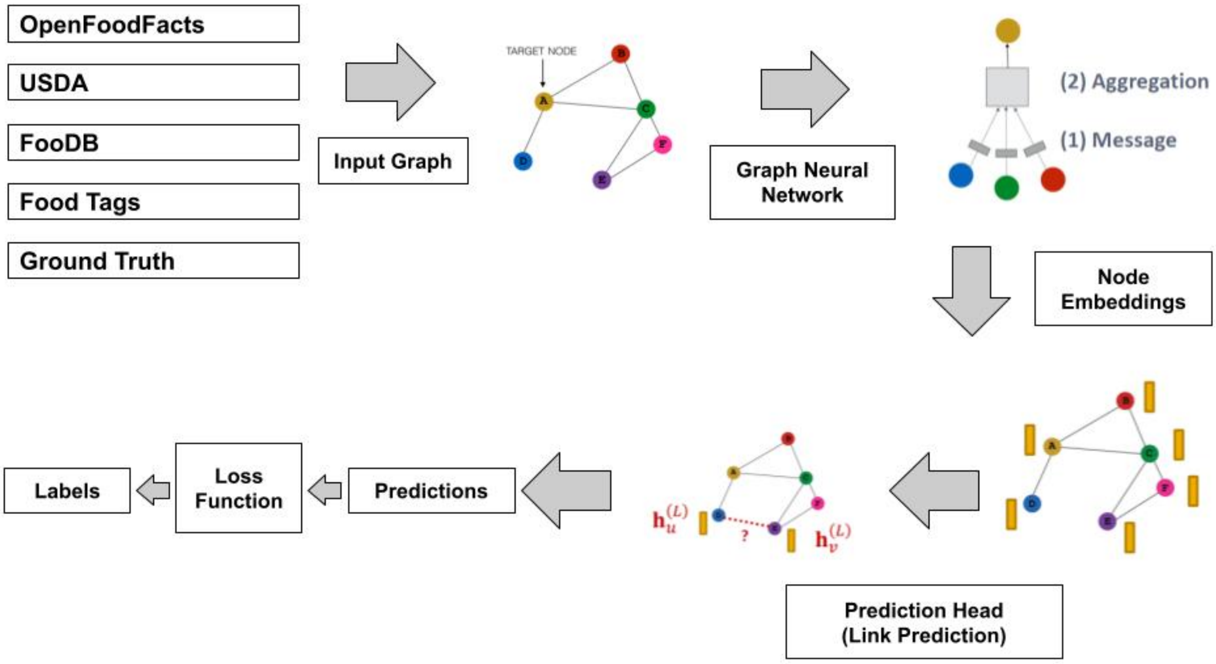

Figure 1: Pipeline of GNN-based link prediction model.

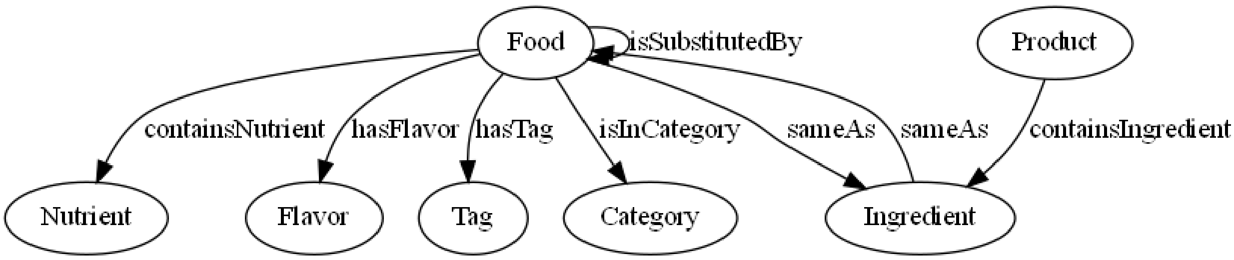

Figure 2: Knowledge graph of food.

Table 1: A breakdown of data sources used to build the knowledge graph.

Source	# of triples	# of unique food items	# of unique
USDA	$300,523$	$8,618$	nutrients: $63,883$
OpenFoodFacts	$890,789$	$125,130$	products: $71,777$
FooDB	$11,167$	125	flavor molecules: 272
Food Tags	$17,746$	$5,404$	tags: 25
Ground Truth	$1,841$	370	substitutions: 704

finds the average of a set of errors. With the attention of avoiding overfitting (e.g., model loses the ability to generalize) or underfitting (e.g., model fails to capture the pattern in the data), we repeated all the steps of the pipeline (e.g., trained our model) until the loss between the predicted values and true values stopped improving.

4. Experiments

We experimented with different variants of Graph Neural Networks (GNNs) and investigated whether simply ranking our food recommendations by the highest Nutri-Score will be enough to claim that the final substitutions are healthier than the query food from a nutritional point of view (the Nutri-Score Experiment). The experimental setup can be found in Subsection 4.1. In addition, we evaluated the performance of the different variants of GNNs using Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and Recall Rate at k (RR@k). Moreover, to assess whether the Nutri-Score Experiment is statistically significant, we computed Cohen's Kappa coefficient and employed a Paired Sample T-test. Hence, Subsection 4.2 describes each evaluation metric.

4.1. Experimental Setup

4.1.1. Food Substitution Experiment

We applied GraphSAGE [13] and GAT [14] on different subsets of the knowledge graph and evaluated the performance of the models using Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and Recall Rate at k (RR@k).

However, due to the fact that the dataset is randomly split into training (80%), validation (10%), and test sets (10%) and the ground truth food substitution dataset itself is not very big, we evaluated our model by using k-fold cross-validation. Cross-validation is a resampling method that uses different portions of the data to train and test a model on different iterations. Concretely, k-fold cross-validation splits the data into k equal folds, using $k-1$ folds as the training set and 1 fold as the test set. Then, we fit and evaluate the model k times by always choosing a different fold as the test set. The main purpose of cross-validation is to test the model's ability to predict unseen noisy data that was not used for training, to flag overfitting or recognize the label bias problem, and to give insights on how the model will generalize to new data.

To describe it more precisely, we fitted our model $k = 5, 10$ times by using different training and test sets each time. Then, we took the average from the evaluation metrics obtained to give a final estimate of the model's performance.

4.1.2. Nutri-Score Experiment

The aim of this work is not only to generate food substitutes, but also provide recommendations for "healthier" food options. To this end, we made use of the French nutritional rating system Nutri-Score [28].

We first calculated the Nutri-Scores for all the ingredients from the USDA[5] dataset. The French Nutri-Score, also known as the 5-Colour Nutrition Label, is a front-of-pack nutrition label which simplifies the results of an overall scoring of the nutrient profile of a food product into an alphabetical scoring with associated colours: a dark green A resembles a relatively positive nutritional profile, the dark red E highlights that a product has a less favourable nutritional profile [29]. These letters follow from an overall calculation of different nutrients per 100 g: *negative* or *unfavourable* nutrients, being (i) energy, (ii) saturated fat, (iii) total sugar and (iv) sodium, are rewarded positive points. Similarly, *positive* or *favourable* nutrients receive points per 100 g product: (i) the percentage of fruit, vegetables, legumes and nuts, (ii) fibre, and (iii) protein. The final score is calculated by subtracting the points attributed to positive nutrients from those points for negative nutrients.

In our work, we make the assumption that the Nutri-Score algorithm can discriminate between healthy and unhealthy food ingredients. However, the Nutri-Score scoring system has been the subject of ongoing debate in Europe. For instance, the work of [30] investigates how well Nutri-Score complies with the dietary guidelines of the Dutch Health Council. The authors recommended that the Nutri-Scores classification increases the consumption of fruits and vegetables, legumes, and unsalted nuts, however, they found that the algorithm is less in line with the recommendations to limit sugared (dairy) beverages and to reduce the consumption of lean meats. As exemplified in that paper [30], while both whole-grain and non-whole-grain breads can have the same Nutri-Score, whole-grain bread is nutritionally superior.

Therefore, we investigated whether simply ranking our food recommendations by the lowest Nutri-Value would be sufficient to claim that the final substitutions are healthier than the query food with respect to its nutritional value. In addition, we investigated whether a Nutri-Score, that was initially developed for *processed* foods, can be applied to *unprocessed* foods as well.

After using our GNN-based link prediction model, we retrieved all the food substitutions and ranked them based on their likelihood of being a food substitute. Then, for each query food, we added the Nutri-Values to the most likely food substitutes. Finally, we output two substitutes such that there is always an option with a better Nutri-Score opposed to the other substitution option.

Thus, in the end, the dataset contains one substitute which has a lower Nutri-Value (or better Nutri-Score) than the other substitute. Two domain experts (nutrition scholars and co-authors AdB and IvL) annotated a sample of the dataset by mentioning which of the two substitutes is a healthier food option from a nutritional point of view.

Specifically, the two domain experts first looked at the macronutrients (protein content, carbohydrates, lipids, and energy) for scoring the healthiness of a product. In the event there was still doubt about the score, they compared the micronutrients of the food item.

4.2. Evaluation

4.2.1. Mean Average Precision (MAP)

The Mean Average Precision (MAP) is the mean of the average precision. To calculate the average precision, we need to compute how many retrieved ingredients are relevant, and then find the mean. In other words, the average precision calculates the precision at the position of every relevant element and takes the mean of these numbers. Specifically, MAP tries to measure: *how many retrieved items are relevant?*.

4.2.2. Mean Reciprocal Rank (MRR)

The Mean Reciprocal Rank (MRR) is the average of the reciprocal rank. To calculate the reciprocal rank, we need to compute the rank of the first relevant ingredient that was retrieved, and take the reciprocal of it. Specifically, MRR tries to measure: *where is the first relevant item?*.

4.2.3. Recall Rate at k (RR@k)

The Recall Rate at k (RR@k) is the proportion of relevant ingredients found in the top-k recommended food substitutions.

4.2.4. Inter-annotator Agreement

In order to assess the inter-rater reliability score between the two domain experts for the Nutri-Score Experiment, we employed the Cohen's Kappa coefficient, which computes how often the raters may agree by chance. The value for Kappa can also be negative. Specifically, a score of 0 means that there is a random agreement among the judges, whereas a score of 1 means that there is a complete agreement between the raters. Thus, a score that is negative means that the inter-rater reliability is less than random chance.

Formally, we have $Kappa = \frac{P(O)-P(E)}{1-P(E)}$, where $P(O) = \frac{Number\ in\ Agreement}{Total}$ is the proportion of time that the annotators agree with each other (e.g., observed agreement) and $P(E)$ is the probability of random agreement (e.g., expected or chance agreement).

4.2.5. Identification of Healthier Foods with Nutri-Scores

To reach a conclusion for our Nutri-Score Experiment, we applied a hypothesis test in which we computed a *Paired Sample T-test*. Concretely, we can easily compute the mean of the Nutri-Value scores that can be obtained based on the annotations agreed by both experts, say sample 1. In other words, sample 1 then represents all the food options that are healthier from a nutritional perspective (e.g., considering macro- and micro-nutrients). While, for sample 2, we added all the Nutri-Value scores of the food recommendations that are not the healthiest food option according to our two domain experts. Thus, we want to statistically test whether the mean of sample 1 is significantly different from the mean of sample 2. By doing so, we can investigate whether simply ranking our food recommendations by the lowest Nutri-Value will be enough to claim that the final substitutions are "healthier" than the query food from a nutritional point of view.

Our findings were a null hypothesis $H_0 : \mu_1 = \mu_2$ and alternative hypothesis $H_1 : \mu_1 \neq \mu_2$. In other words, the null hypothesis H_0 states that there is no significant difference between the means of the two groups, while the alternative hypothesis H_1 suggests that there is a significant difference between the two population means, and that this difference is unlikely to be caused by a sampling error or chance.

5. Results

5.1. GNN-based Link Prediction Model

We performed the link prediction experiments using k-fold cross-validation ($k = 5, 10$). The results can be seen in Table 2a and Table 3a, without filtering, and Table 2b and Table 3b, with filtered ranking using food categorization. Similar to our previous work [3], we achieved the best results by restricting the recommended substitutes to be in the same food category as the query food. Moreover, we also noticed that utilizing the entire knowledge graph of food compared to only make use of the ingredient dataset (that consists of $8,618$ different food items and provides the information on both macro- and micro-nutrients) improves the prediction performance.

The results using 5-fold cross-validation can be seen in Table 2a, without filtering, and Table 2b, with filtered ranking using food categorization. The best results are obtained using GAT [14] and the entire food knowledge graph.

Specifically, looking at Table 2b, we can observe that the MAP and MRR values are relatively low, namely 0.345 and 0.549, respectively. Recall that the Mean Average Precision (MAP) gives an indication of how many retrieved ingredients are relevant, while the Mean Reciprocal Rank (MRR) gives an impression of the rank order for which the first relevant ingredient was retrieved.

On the other hand, the Recall Rate for the top 5 and 10 results achieve a value of 0.680 and 0.757, respectively. In other words, this means that our recommender system finds more than 65% of the relevant ingredients in the top-5 recommended food substitutions and 75% of the relevant ingredients in the top-10 recommended food substitutions, showing significant results for a recommender system.

Similarly, the results using 10-fold cross-validation can be found in Table 3a, without filtering, and Table 3b, with filtered ranking using food categorization. Again, the best results are obtained using GAT [14] and the entire food knowledge graph. We obtained a MAP and MRR of 0.353 and 0.512, respectively. While the Recall Rate for the top 5 and 10 results achieved a value of 0.647 and 0.733, respectively.

It is notable that there is not a huge discrepancy between the results that were obtained by 5-fold cross-validation and 10-fold cross-validation, meaning that our GNN-based link prediction model is reasonably robust. As k increases, the difference in size between the training set and the resampling subsets gets smaller, and therefore, as the difference decreases, the bias becomes smaller [31]. In summary, there is a bias-variance trade-off associated with the choice of k in k-fold cross-validation [32]. Typically, one performs k-fold cross-validation using $k = 5$ or $k = 10$, as these values have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance [32], which can also be observed here.

5.2. Nutri-Score Experiment

5.2.1. Inter-annotator Agreement

The data for the paired ratings on a nominal scale of three categories can easily be displayed by an 3×3 contingency table, as shown in Table 4. This table shows the annotations from our two domain experts who assessed 465 food items independently. Cell a indicates the number of food options for which both raters agree on to be *Sub A*, which is the substitution candidate that has a lower Nutri-Score as opposed to the other substitution option. Cell c presents the number of food items for which both raters agree on as *Sub B*, which is the substitution candidate with the better Nutri-Score. Cell i shows the number of food ingredients for which both experts agree on to be *No Choice* and the off-diagonal cells represent disagreement.

The sum of the frequencies from the main diagonal cells was calculated to yield the frequency of observed agreement. Therefore, dividing by n gives the proportion of time that the judges agree with each other:

$$P(O) = \frac{a + c + i}{n} = 0.8129$$

The proportion of expected or chance agreement can be calculated as follows:

$$P(E) = \frac{(\frac{f_1 \times g_1}{n}) + (\frac{f_2 \times g_2}{n}) + (\frac{f_3 \times g_3}{n})}{n} = 0.3622$$

Leading to a decent Kappa score of:

$$Kappa = \frac{P(O) - P(E)}{1 - P(E)} = 0.7066$$

5.2.2. Paired Sample T-test

The p-value obtained of 0.2391 of the two-tailed paired sample T-test is very close to 0, suggesting that there is strong evidence to reject the null hypothesis H_0. Thus, the mean of sample 1 deviates significantly from the mean of sample 2.

We also investigated whether the mean of sample 1 is significantly lower than the mean of sample 2. Thus, we employed a left-tailed paired sample T-test (instead of a two-tailed one), and got null hypothesis H_0: $\mu_1 \geq \mu_2$ and alternative hypothesis H_1: $\mu_1 < \mu_2$.

The obtained p-value 0.1196 of the left-tailed paired sample T-test is also very close to 0, suggesting that there is strong evidence to reject the null hypothesis H_0. Thus, the mean of sample 1 is significantly lower in comparison to the mean of sample 2. Hence, the sample that incorporates all the food items that were annotated as being healthier from a nutritional perspective (e.g., in a comparison of macro- and micro-nutrients) has significantly lower Nutri-Values (or higher Nutri-Scores) than the other sample that includes all the food items that were annotated as unhealthier. Consequently, according to the experts, if a food recommendation has a lower Nutri-Value (or better Nutri-Score) in comparison of two food items, then that food recommendation has a better nutritional profile.

Table 2: Results of experiments for 5-fold cross-validation.

(a) Without filtering

	Method	MAP	MRR	RR@5	RR@10
USDA + Tags + OpenFoodFacts (Previous work [3])	RDF2Vec	0.133	0.234	0.330	0.400
USDA	GraphSAGE	0.044	0.166	0.221	0.322
	GAT	0.112	0.319	0.387	0.466
USDA + Tags + OpenFoodFacts + Flavor + Ground Truth	GraphSAGE	0.048	0.165	0.227	0.313
	GAT	0.116	0.342	0.422	0.484

(b) With filtered ranking using food category

	Method	MAP	MRR	RR@5	RR@10
USDA + Tags + OpenFoodFacts (Previous work [3])	RDF2Vec	0.154	0.259	0.359	0.438
USDA	GraphSAGE	0.277	0.466	0.638	0.749
	GAT	0.326	0.522	0.657	0.740
USDA + Tags + OpenFoodFacts + Flavor + Ground Truth	GraphSAGE	0.274	0.460	0.577	0.693
	GAT	0.345	0.549	0.680	0.757

6. Discussion

Our best results were achieved with GAT and by putting the recommended substitutes in the same food category as the query food. The specific results achieved were a Recall Rate of 0.647 and 0.733 for the top 5 and 10 results, respectively. Moreover, we found out that ranking the food recommendations by the lowest Nutri-Value mostly allowed for providing final substitutions of ingredients with better nutritional profiles.

6.1. GNN-based Link Prediction Model

Table 5a and Table 5b display examples of substitution options for *Cheese, Colby* produced by our approach with and without filtered ranking using food category, respectively. The ground truth substitution options do not have any ranked order while our approach here presents the top five ranked substitution recommendations.

With filtered ranking using food category, four of the top five ranked ingredients matched with those from the ground truth food substitution dataset, as illustrated in Table 5a. For the fifth ranked ingredient, our algorithm gave the nonsensical suggestion of substituting *Cheese, Colby* with *Butter, with Salt*, which is incoherent because these ingredients are used in different food contexts and also have distinct functions. One plausible explanation is that butter and cheese are made of the same ingredient, which is milk and thus, our algorithm considers both food items to be "similar". In addition, since our GNN-based link prediction model's aim was to find food substitutes that have nutritional characteristics similar to those of the food in question, it is still logical to consider the option of butter to be "relevant".

Without filtered ranking using food category, three of the top five ranked ingredients corresponded to those from the ground truth food substitution dataset, as shown in Table 5b. For the third and fourth ranked ingredient, our algorithm recommends the substitution of *Cheese, Colby* with *Lingcod* and *Shortening*, which, however, is not reasonable because the recommended food items have distinct functionalities, are used in different food contexts, and have very dissimilar nutritional characteristics compared to the query food.

Filtering food substitutions using food categorization is a practical method for simple identification of options such as discovering different varieties of cheese. In addition, this process may automatically filter out (most

Table 3: Results of experiments for 10-fold cross-validation.

(a) Without filtering

	Method	MAP	MRR	RR@5	RR@10
USDA + Tags + OpenFoodFacts (Previous work [3])	RDF2Vec	0.133	0.234	0.330	0.400
USDA	GraphSAGE	0.040	0.143	0.202	0.280
	GAT	0.116	0.297	0.373	0.472
USDA + Tags + OpenFoodFacts + Flavor + Ground Truth	GraphSAGE	0.057	0.193	0.257	0.353
	GAT	0.111	0.306	0.401	0.474

(b) With filtered ranking using food category

	Method	MAP	MRR	RR@5	RR@10
USDA + Tags + OpenFoodFacts (Previous work [3])	RDF2Vec	0.154	0.259	0.359	0.438
USDA	GraphSAGE	0.254	0.417	0.595	0.714
	GAT	0.348	0.501	0.651	0.764
USDA + Tags + OpenFoodFacts + Flavor + Ground Truth	GraphSAGE	0.279	0.451	0.624	0.746
	GAT	0.353	0.512	0.647	0.733

Table 4: Kappa value in contingency matrix.

		Domain Expert 2 (AdB)			
		Sub A	Sub B	NC	**Total**
Domain Expert 1 (IvL)	Sub A	$a = 94$	$b = 9$	$c = 20$	$g_1 = 123$
	Sub B	$d = 9$	$e = 99$	$f = 17$	$g_2 = 125$
	NC	$g = 17$	$h = 15$	$i = 185$	$g_3 = 217$
	Total	$f_1 = 120$	$f_2 = 123$	$f_3 = 222$	$n = 465$

of the times) incorrect substitution options such as recommendations of substitutions of fish or shortening for cheese. We consider the use of categorization for similarity to be an appropriate method of finding simple and closely related food substitution options sharing similar nutritional characteristics with the query food.

On the other hand, this approach will be less effective in discovery of good substitutions that are not closely related [10]. Specifically, this technique is limited to vegan substitution options for cheese, such as cashew or almond products, because they are in different food categories. In addition, in our results, we noticed that for some cases our filtering strategy could not always successfully omit undesirable food substitution options (e.g., substituting cheese with butter). Specifically, we used 10-fold cross-validation and Graph Attention Network to re-assess the performance of our model *per food categories* to analyze the food groups for which our algorithm performs poorly against the ground truth food substitution dataset.

Table 6 shows that our algorithm performs poorly for *Dairy and Egg Products* and *Cereal, Grains and Pasta*, and satisfactorily for *Fats and Oils* and *Sausages and Luncheon Meats*.

The main reason that food categorization similarity does not always succeed in excluding undesirable food substitution options from the same food categories is that the food groups might be too comprehensive and general. For instance, the food group *Dairy and Egg Products* includes butter, cheese, (sour) cream, eggnog, sour dressing, (chocolate) milk, dessert topping, yogurt, eggs, ice cream, and so on, leading to a diverse food

Table 5: Comparison between with and without filtered ranking using food category.

(a) With filtered ranking using food category

Target Ingredient	Ground Truth Substitutes	GAT's Top 5 Ranked Substitutes
Cheese, Colby (01011)	Cheese, Fontina Cheese, Feta Cheese, Parmesan Cheese, Blue Cheese, Limburger	Cheese, Fontina Cheese, Feta Cheese, Parmesan Cheese, Blue Butter, with Salt

(b) Without filtered ranking using food category

Target Ingredient	Ground Truth Substitutes	GAT's Top 5 Ranked Substitutes
Cheese, Colby (01011)	Cheese, Fontina Cheese, Feta Cheese, Parmesan Cheese, Blue Cheese, Limburger	Cheese, Fontina Cheese, Feta Lingcod Shortening Cheese, Parmesan

Table 6: Error analysis.

Food Group	MAP	MRR	RR@5	RR@10
Cereal, Grains and Pasta	0.261	0.396	0.652	0.756
Dairy and Egg Products	0.346	0.631	0.788	0.855
Fats and Oils	0.732	0.842	0.898	0.922
Sausages and Luncheon Meats	0.632	0.784	0.929	0.980

category. Indeed, many food products consist of either milk or eggs or both. Thus, there is a relatively high chance that they are similar as regards their nutritional content. However, as already mentioned earlier, one would hardly substitute cheese with butter in practice because they are used in different food contexts and have distinct functionalities.

Table 7 reveals another example, with filtered ranking using food category, that performs poorly against the ground truth food substitution dataset. For *Egg Noodles* the results are very poor. The main issue is that *Egg Noodles* are contained in the food category *Cereal, Grains and Pasta*, which is a too broad food group and even contains the ingredients that are utilized to make pasta. In fact, we can see that the top five ranked options for *Egg Noodles* are different varieties of wheat, grains or flour, which are basically unprocessed variants of products like pasta. Thus, in some sense, the proposed food substitution options are similar to *Egg Noodles*, however, again the functionality of the recommended ingredients is totally distinct compared to the food in question and they are also used in different food contexts than the query food. Our results therefore show that, while filtering food substitutions using food categorization can provide a good start for recommender systems, it has its shortcomings.

Conversely, the food categories *Fats and Oils* or *Sausages and Luncheon Meats* are much more narrow. For example, *Fats and Oils* incorporates food items such as oil, lard, salad dressing, shortening, margarine, and butter, which would also be substitutes for each other in practice. This suggests that the use of food categorization similarity can be a successful method if the food groups are neither too wide-ranging nor widespread.

6.2. Nutri-Score Experiment

To test whether simply ranking our food recommendations by the lowest Nutri-Value implies that the final

Table 7: Example of substitution options that performed poorly against the ground truth due to the use of food categorization similarity.

Target Ingredient	Ground Truth Substitutes	GAT's Top 5 Ranked Substitutes, with filtered ranking using food categorization
Egg Noodles (20109)	Rice Noodles Spaghetti Chinese Noodles Wheat Durum	Spelt Wheat Flour Sorghum Grain Rice Flour Tapioca

substitutions are indeed healthier than the query food, we conducted the Nutri-Score Experiment. This scoring system is adapted from the UK Food Standards Agency nutrient profile model [33]. The validity of this scoring system in countries outside France has however been subject of debate, in particular due to questions regarding its alignment with national dietary recommendations [34, 35]. We therefore wanted to assess whether the Nutri-Value in itself implies that a food is indeed healthier when the calculated score is lower. Specifically, we asked two domain experts to choose between two food substitutions, where one substitution option has a lower Nutri-Value (or better Nutri-Score) than the other.

Consequently, the mean of sample 1 is significantly lower than the mean of sample 2. Thus, the sample that incorporates all the food items that were annotated as being healthier from a nutritional point of view has significantly lower Nutri-Values (or higher Nutri-Scores) than the sample that includes all the food items that were annotated as unhealthier. Hence, simply ranking our food recommendations by the lowest Nutri-Value is enough to claim that the final substitutions have a better nutritional profile.

To illustrate how the final algorithm works, Table 8a displays the top 5 food substitutions for ham, with and without filtered ranking using Nutri-Scores (e.g., filters out the food items that have higher Nutri-Values than the target food). For ham, four of the top five ranked ingredients corresponded to those from the ground truth food substitution dataset. Suggesting that our GNN-link prediction model, with the help of food categorization similarity, recommends *relevant* food substitutions. Moreover, with filtered ranking using Nutri-Scores, the recommended substitutions are not only relevant but also are also *healthier than the query food*.

However, it should be noted that the comparison between the two proposed food substitutions was not always "fair". For example, it is not useful to do a side-by-side comparison of 100g of a vegetable with 100g of a herb because the latter is used in smaller portions and also for a different purpose.

Nor is it useful to make comparisons between unprepared (or raw) products and prepared products because there is no information provided about how the processed products have been prepared, which highly affects the evaluation of the "healthfulness" level of a food product (e.g., how long did a fried chicken stay in the frying pan/fat?).

In addition, to compare the two recommended food options, the two domain experts mainly focused on macronutrients and salt on the principle that lower calories, lower lipids, lower carbohydrates, higher fiber, and higher protein are already accepted as healthy. However, this approach still makes the annotations challenging. For instance, for some substitutions, one option was higher in lipids whereas the other option contained more sugar (e.g., carbohydrates), making it hard to select the healthiest food option. Similarly, the lipid profile was not considered in this manual comparison of macronutrients. Hence, there is not necessarily one type of nutrient that promotes a healthy diet. Instead, achieving a *balanced diet* is considered ideal. Specifically, making a determination of which of the two food options is the healthiest depends on the situation (e.g., other parts of the diet). Therefore, we would like to stress that, although we are looking at the healthiest food options, it also depends on the other foods that are eaten during the meal and within the full diet.

Because the comparison between the two proposed food substitutions was not always "fair", the two domain experts often opted for the third option *NC*, representing *No Choice*. In addition, it should be noted that the raters usually were undecided when rating fruits and vegetables because they could not decide which option was healthier, as both (raw) products are by their nature

Table 8: Top 5 substitutes ranked by our approach for *Ham*.

(a) Ham (07030) (15)

Ground Truth Substitutes	GAT's Top 5 Ranked Substitutes, with filtered ranking using food categorization	GAT's Top 5 Ranked Substitutes, with filtered ranking using food categorization and Nutri-Scores
Blood Sausage	Blood Sausage (16)	Oscar Mayer Chicken Breast (6)
Smoked Link Sausage	Smoked Link Sausage (16)	Meatballs (10)
Salami	Salami (17)	Roast Beef (11)
Turkey Bacon	Frankfurter Beef Sausage (18)	Pickle and Pimento Loaf (12)
Bologna	Bologna (14)	Bologna (14)

healthy. There were a total of 185 *NC* options that nutrition scholars agreed upon. Indeed, the concerned food categories, which were the hardest to annotate, were the ones that are already essentially healthy, such as vegetables, fruits, spices and herbs. Furthermore, spices cannot be compared based on Nutri-Scores because you use them purely for flavoring and in very small amounts.

The two nutrition scholars also observed that not all proposed food substitutions were useful. For instance, the substitution of a soup with a sauce or vice versa is impractical, suggesting that our GNN-based link prediction model is unable to capture the food context and functionality of the food products. Furthermore, the use of food categorization similarity is a plausible method to find simple substitutions (e.g., finding different varieties of cheese). However, this approach will be insufficient to discover good substitutions that are not closely related (e.g., a vegan substitution for meat will have vastly different food categorization) [10] or this process will even mislead us to irrelevant substitutions (e.g., substituting soup by sauce). On the other hand, it should also be noted that the primary goal of annotating was not to look at which substitution was best but whether the Nutri-Score algorithm can help us in finding "healthier" food substitutions. Still, the overall effectiveness of the use of these substitutions in developing a healthier dietary pattern requires further research.

Moreover, it should be mentioned that the two domain experts did not annotate all the food categories. For example, baby food is a difficult food group since it is subject to strict regulation and no Nutri-Scores are used in analyzing their overall healthiness. It is therefore not possible and (especially) not appropriate to include baby food into the experiment.

Finally, a balanced diet outweighs these other considerations. We now have performed comparisons of individual products, which can be very helpful in selecting a healthier option from a nutritional point of view (e.g., finding a healthier option for fried potatoes). However, at the same time, this is also very limiting: replacing one vegetable with another vegetable is not necessarily leading to a "healthier" diet, if it leads to a less varied diet. Including a variety of healthy foods in the diet is important to ensure sufficient intake of all macro- and micronutrients essential for the body.

7. Conclusion

The final algorithm, which identifies "healthier" food substitutions using a nutrient profiling system along with semantic similarity and deep learning methods, consists of two parts. The first part finds food substitutes that have similar nutritional characteristics than the query food using Graph Neural Networks. While the second part identifies "healthier" food recommendations employing the Nutri-Score algorithm and ranks these "healthier" food options according to their highest Nutri-Score.

The food substitutions having a better nutritional profile suggested by the recommender system were mostly recognized by two independent domain experts, who had an inter-rater reliability score, of 0.71. Ultimately, we found out that by simply ranking our food recommendations by the highest Nutri-Score is enough to claim that the final substitutions have a better nutritional profile.

A possible significant direction for future work would be to target a specific population, for instance, disadvantaged people that cannot afford a nutritious, healthy diet. In fact, approximately three billion people cannot afford a healthy diet, and more than three billion people suffer one or more manifestations of poor nutrition [6]. For that reason, it would be much more outstanding to not only recommend *healthier* food options but also *healthier* and *cheaper* food options.

References

[1] M. L. Neuhouser, The importance of healthy dietary patterns in chronic disease prevention, Nutr Res 70 (2018) 3–6.

[2] J. Jin, Dietary Guidelines for Americans, JAMA 315 (2016) 528–528. URL: https://doi.org/10.1001/jama.2016.0077. doi:10.1001/jama.2016.0077.

[3] J. Loesch, L. Meeckers, I. van Lier, A. de Boer, M. Dumontier, R. Celebi, Automated identification of food substitutions using knowledge graph embeddings, CEUR Workshop Proceedings 3127 (2022) 19–28.

[4] Maintaining a healthy diet during the COVID-19 pandemic, FAO, 2020. URL: http://www.fao.org/documents/card/en/c/ca8380en. doi:10.4060/ca8380en.

[5] M. J. Butler, R. M. Barrientos, The impact of nutrition on covid-19 susceptibility and long-term consequences, Brain, Behavior, and Immunity 87 (2020) 53–54. URL: https://www.sciencedirect.com/science/article/pii/S0889159120305377. doi:https://doi.org/10.1016/j.bbi.2020.04.040.

[6] FAO, IFAD, UNICEF, WFP, WHO, The state of food security and nutrition in the world (sofi) (2020). doi:https://doi.org/10.4060/ca9692en.

[7] D. Kromhout, , C. J. K. Spaaij, J. de Goede, R. M. Weggemans, The 2015 dutch food-based dietary guidelines, European Journal of Clinical Nutrition 70 (2016) 869–878. URL: https://doi.org/10.1038/ejcn.2016.52. doi:10.1038/ejcn.2016.52.

[8] J. M. Ordovas, L. R. Ferguson, E. S. Tai, J. C. Mathers, Personalised nutrition and health, BMJ 361 (2018). URL: https://www.bmj.com/content/361/bmj.k2173. doi:10.1136/bmj.k2173.

[9] K. M. Livingstone, C. Celis-Morales, S. Navas-Carretero, R. San-Cristobal, H. Forster, C. Woolhead, C. B. O'Donovan, G. Moschonis, Y. Manios, I. Traczyk, T. E. Gundersen, C. A. Drevon, C. F. M. Marsaux, R. Fallaize, A. L. Macready, H. Daniel, W. H. M. Saris, J. A. Lovegrove, M. Gibney, E. R. Gibney, M. Walsh, L. Brennan, J. A. Martinez, J. C. M. and, Personalised nutrition advice reduces intake of discretionary foods and beverages: findings from the food4me randomised controlled trial, International Journal of Behavioral Nutrition and Physical Activity 18 (2021). URL: https://doi.org/10.1186/s12966-021-01136-5. doi:10.1186/s12966-021-01136-5.

[10] S. S. Shirai, O. Seneviratne, M. E. Gordon, C.-H. Chen, D. L. McGuinness, Identifying ingredient substitutions using a knowledge graph of food, Frontiers in Artificial Intelligence 3 (2021) 111. URL: https://www.frontiersin.org/article/10.3389/frai.2020.621766. doi:10.3389/frai.2020.621766.

[11] S. Haussmann, O. Seneviratne, Y. Chen, Y. Ne'eman, J. Codella, C.-H. Chen, D. McGuinness, M. J. Zaki, Foodkg: A semantics-driven knowledge graph for food recommendation, in: SEMWEB, 2019.

[12] W. Min, C. Liu, L. Xu, S. Jiang, Applications of knowledge graphs for food science and industry, Patterns 3 (2022) 100484. URL: https://www.sciencedirect.com/science/article/pii/S2666389922000691. doi:https://doi.org/10.1016/j.patter.2022.100484.

[13] W. L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 1025–1035.

[14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations, 2018. URL: https://openreview.net/forum?id=rJXMpikCZ.

[15] C. Julia, F. Etilé, S. Hercberg, Front-of-pack nutri-score labelling in france: an evidence-based policy, The Lancet Public Health 3 (2018) e164. URL: https://doi.org/10.1016/s2468-2667(18)30009-4. doi:10.1016/s2468-2667(18)30009-4.

[16] N. Garg, A. Sethupathy, R. Tuwani, R. NK, S. Dokania, A. Iyer, A. Gupta, S. Agrawal, N. Singh, S. Shukla, K. Kathuria, R. Badhwar, R. Kanji, A. Jain, A. Kaur, R. Nagpal, G. Bagler, FlavorDB: a database of flavor molecules, Nucleic Acids Research 46 (2017) D1210–D1216. URL: https://doi.org/10.1093/nar/gkx957. doi:10.1093/nar/gkx957.

[17] D. Dooley, E. Griffiths, G. Gosal, P. L. Buttigieg, R. Hoehndorf, M. Lange, L. Schriml, F. Brinkman, W. Hsiao, Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration, npj Science of Food 2 (2018). doi:10.1038/s41538-018-0032-6.

[18] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh, R. Tuwani, G. Bagler, RecipeDB: a resource for exploring recipes, Database 2020 (2020). URL: https://doi.org/10.1093/database/baaa077. doi:10.1093/database/baaa077, baaa077.

[19] J. Marín, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, A. Torralba, Recipe1m: A dataset for learning cross-modal embeddings for cooking recipes and food images, CoRR abs/1810.06553 (2018). URL: http://arxiv.org/abs/1810.06553. arXiv:1810.06553.

[20] T. Eftimov, G. Popovski, E. Valenčič, B. K. Seljak, Foodex2vec: New foods' representation for advanced food data analysis, Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 138 (2020) 111169. URL: https://doi.org/10.1016/j.fct.2020.111169. doi:10.1016/j.fct.2020.111169.

[21] E. F. S. A. (EFSA), The food classification and description system foodex 2 (revision 2), EFSA Supporting Publications 12 (2015) 804E. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2015.EN-804. doi:https://doi.org/10.2903/sp.efsa.2015.EN-804.

[22] Y. Pan, Q. Xu, Y. Li, Food recipe alternation and generation with natural language processing techniques, in: 2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW), 2020, pp. 94–97. doi:10.1109/ICDEW49219.2020.000-1.

[23] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, Curran Associates Inc., Red Hook, NY, USA, 2013, p. 3111–3119.

[24] M. Honnibal, I. Montani, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, 2017. To appear.

[25] W. Tansey, E. W. L. J. au2, J. G. Scott, Diet2vec: Multi-scale analysis of massive dietary data, 2016. arXiv:1612.00388.

[26] M. Kazama, M. Sugimoto, C. Hosokawa, K. Matsushima, L. R. Varshney, Y. Ishikawa, A neural network system for transformation of regional cuisine style, Frontiers in ICT 5 (2018). URL: https://www.frontiersin.org/

article/10.3389/fict.2018.00014. doi:10.3389/fict.2018.00014.
[27] Y. S. Tang, A. H. Zheng, N. Lai, Healthy recipe recommendation using nutrition and ratings models, 2019.
[28] C. Julia, S. Hercberg, Nutri-score: Evidence of the effectiveness of the french front-of-pack nutrition label, 2018.
[29] C. R. Julià, S. Hercberg, Development of a new front-of-pack nutrition label in france : the five-colour nutri-score, 2017.
[30] S. ter Borg, E. Steenbergen, I. E. J. Milder, E. H. M. Temme, Evaluation of nutri-score in relation to dietary guidelines and food reformulation in the netherlands, Nutrients 13 (2021). URL: https://www.mdpi.com/2072-6643/13/12/4536. doi:10.3390/nu13124536.
[31] M. Kuhn, K. Johnson, Applied predictive modeling (2013). URL: http://www.amazon.com/Applied-Predictive-Modeling-Max-Kuhn/dp/1461468485/.
[32] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2013.
[33] D. L. M. van der Bend, M. van Eijsden, M. H. I. van Roost, K. de Graaf, A. J. C. Roodenburg, The nutri-score algorithm: Evaluation of its validation process, Frontiers in Nutrition 9 (2022). URL: https://www.frontiersin.org/articles/10.3389/fnut.2022.974003. doi:10.3389/fnut.2022.974003.
[34] C. van Tongeren, L. Jansen, Adjustments needed for the use of nutri-score in the netherlands: Lack of selectivity and conformity with dutch dietary guidelines in four product groups, International Journal of Nutrition and Food Sciences 9 (2020) 33. URL: https://doi.org/10.11648/j.ijnfs.20200902.11. doi:10.11648/j.ijnfs.20200902.11.
[35] V. Braesco, E. Ros, A. Govindji, C. Bianchi, L. Becqueriaux, B. Quick, A slight adjustment of the nutri-score nutrient profiling system could help to better reflect the european dietary guidelines regarding nuts, Nutrients 14 (2022). URL: https://www.mdpi.com/2072-6643/14/13/2668. doi:10.3390/nu14132668.

Conflict of Interest

The authors declare that they have no conflict of interest.