# Food Synergy: The Right Combination Matters

Nivedita Sethiya, Shahid Shafi Dar
nivedita.21csz0019@iitrpr.ac.in, shahid.21csz0029@iitrpr.ac.in

**Abstract**

In the emergence of modern nutrition research, the discovery of nutrients and the study of their bioactivity were fundamental advances. The concept of food synergy (the idea that looking at foods yields more information than looking at single food components) is used to define the function of the food matrix(the composite of naturally occurring food components) in human biological systems. Despite our daily interactions with food and a large amount of food data available on the internet, food pairing has not progressed to its full potential. Rather than scientific understanding or statistical learning, the complimentary food pairings documented so far were generated by creative chefs' intuition. We offer a one-of-a-kind technique that combines many NLP approaches to capture the diverse food combinations. Food is significantly more sophisticated than drugs, but as food or pattern, it is largely unstudied. Food synergy is a notion that encourages innovative thinking in nutrition science and can help with the creation of reasonable nutrition policy and future nutrition research planning.

Keywords

Food Synergy, Food matrix, relation extraction, Flavour Graph, Food-Nutrient Graph

## 1 .INTRODUCTION

Food pairing hasn't been advanced fully until this period. Despite working with food daily and having access to a vast amount of food information on the internet. The harmonious food pairings were not discovered through statistical learning or scientific understanding, but rather through the natural wisdom of skilled chefs, researchers, and gastronomists. Exploration of novel and fascinating food pairings is a typical technique for product innovation and distinction among restaurant chefs and food product pairings. For many food professionals, however, researching and designing ideal food pairings that are well appreciated by consumers is a difficult task. Recipes have been utilised as sources of information by social scientists to examine the relationship between food and broader economic and socio-cultural phenomena.

Food pairing has been one of the heated topics in food science and is currently a critical task for cooking practices. In the Culinary world, instead of so many efforts by chefs, and researchers to discover new food pairings, there are still food pairings that have yet to be explored. Understanding food is a challenging task because of many distinct features like flavour, texture, colour etc. Also, we do not have a thorough understanding of food composition, and certain impacts may be caused by unrecognised components. To get optimal health benefits, the combination of food components must take into account. Food components must survive digestion to reach the human system in such a way that the person can experience the mutual effects of the various components.

Humans have long struggled to find and acquire food that meets nutritional demands while avoiding foodborne illnesses. This process has shaped human diets today, which are influenced by a variety of elements including an evolved predilection for sugar and fat, nutritional value, culture, ease of production, and environment .The small number of recipes in use ($10^6$) compared to the enormous number of potential recipes ($10^{15}$) combined with the frequent occurrence of particular combinations in various regional cuisines, suggests that we are only utilising a small portion of the potential combinations. Foods that share flavour components are more likely to taste good together than ingredients that do not, according to a notion that has made news among chefs and food scientists over the last decade .This food matching concept has been utilised to find new ingredient combinations, prompting some modern restaurants to combine white chocolate and caviar, which share trimethylamine and other flavour compounds, or chocolate and blue cheese, which have at least 73 flavour compounds in common.

### 1.1 BENEFITS OF FOOD SYNERGY

a) Buffer Effect:-The effect of a significant intake of a particular nutrient may differ depending on whether it is consumed in concentrated form or as part of a food matrix. The European Union Recommended Dietary Allowance for iron, 14 mg, can be obtained in a bolus from a single tablet rather than 677 g roast beef, 879 g spinach, or 210g cornflakes. The food matrix also slows nutrient absorption, reducing the chance of a bolus effect.
b) Nutrients that affect each other's absorption, such as copper-zinc and manganese-iron, are another example of synergy. In the presence of iron, vitamin C can behave as a pro-oxidant, and alcohol can disturb iron homeostasis by affecting iron-binding proteins like ferritin and transferring. Buffering and competing effects are examples of control over component entry into the human system.
c) The third facet of synergy is whether the ingredients were created through technological or biological processes. Trans fat, which is formed in ruminant animals and in food processing during hydrogenation of vegetable oils, is a good example.

Food pairing can never be completely precise because it depends on different factors such as texture, colour, and flavour. The chemical composition of food pairings varies by country. For example, if we combine cheese and peas in India, it would have a distinct boosting character than in other places. Our method consisted of attempting to extract information on food pairing from prior study papers, which can be biased. As a result, before confronting the problem of food pairing, we must evaluate different solutions.

## 2. Literature Review

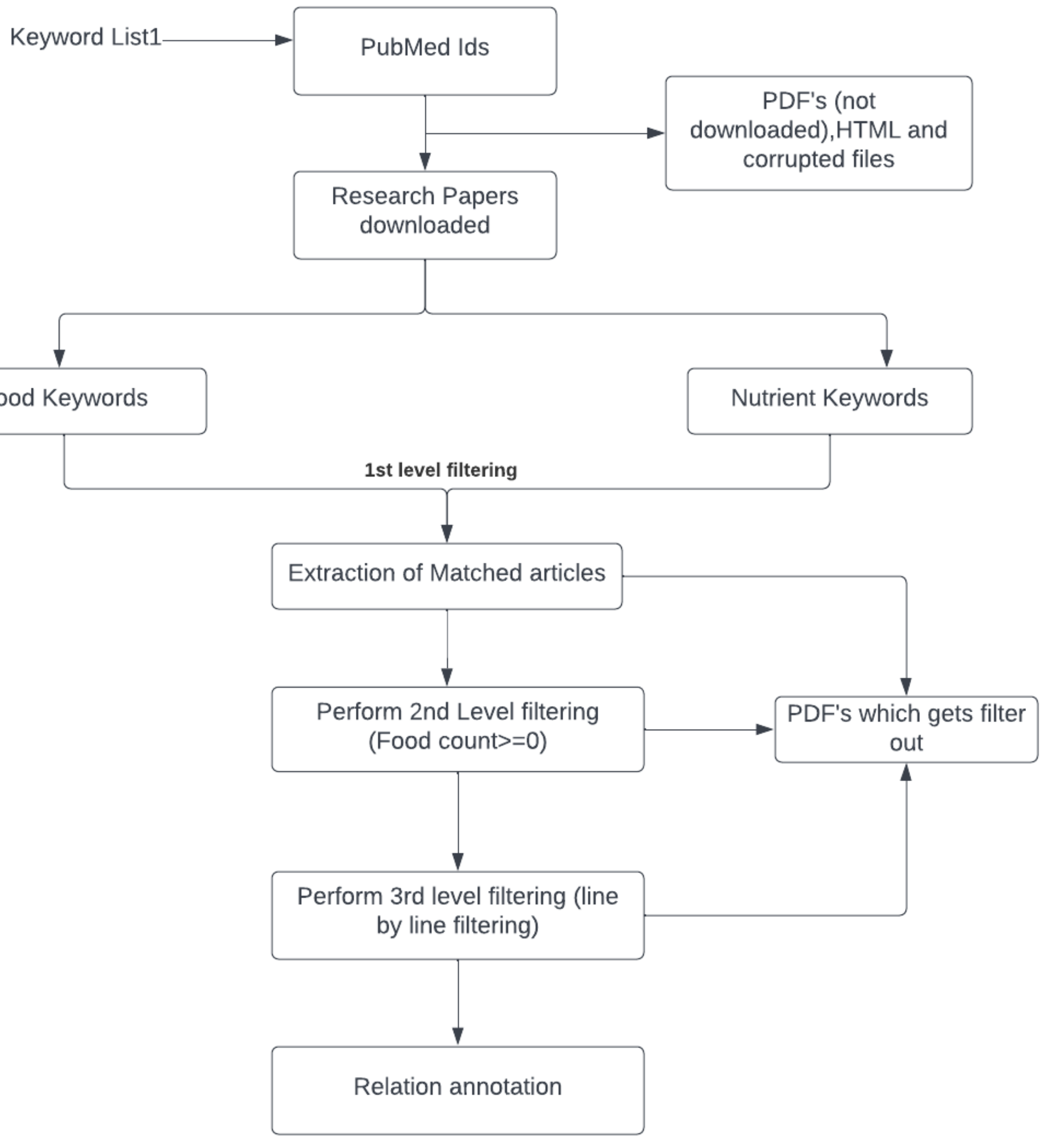
In the domain of food and nutrition, the efforts to develop relation extraction (RE) systems have been far more limited in comparison to biomedical relations. For the detection of food pairings, numerous methods such as remote supervision, pattern-matching, and the use of co-occurrence metrics have been investigated. Various chemical-based ways to improve food pairing have been explored in previous studies. Ahn et al. [1, 2] proposed a flavor network in which the network edges are based on the number of flavor compounds that culinary components share. FlavorDB [3] combines current food repositories to create a larger database with an interactive interface for users. When the similarity in the chemical composition of two ingredients is modest, then food-bridging [4] strengthens the flavor network by establishing extra bridges between two ingredients through a chain of pairwise affinities. However, one major drawback of chemical-based techniques is that they can be studied for a limited variety of foods. Experiments in food chemistry (such as Gas Chromatography) are quite costly. Furthermore, because there are numerous qualities of food (e.g., flavour, colour, texture, smell) and different types of the same meal, it is challenging to effectively represent the chemical constituents of foods in a manner that can be saved digitally. Food pairing requires the incorporation of flavour compound information. Food pairing requires the incorporation of flavor compound information. Chemical-based techniques, on the other hand, find it challenging to generate correct food representations in food pairing challenges due to a lack of available chemical information.

For food pairing tasks, several recipe-based techniques involving recipe libraries have previously been presented. Teng et al. [5] suggested a technique for recipe recommendation that relies on ingredient networks to determine whether a food item is required in a recipe. This method uses two different recipe networks to determine which elements pair well or can be substituted for superior dishes. There has been research in China [6] and India [7] on the examination of food preference and food pairing based on regional features. Wagner et al. [8] and Abbar et al. [9] studied personal food preferences utilising online user data in their studies on food preferences. The restaurant recommendation was presented by Zhang et al. [10] to advise dining preferences based on the user's food history. Other approaches [11,12] for automatic recipe generation that incorporates case-based reasoning and deep learning have been introduced. KitcheNette [13] predicts meal pairing scores using deep Siamese neural networks trained on a huge recipe dataset. By referencing similar food representations, the hidden representations from KitcheNette's shared embedding layer are utilised to predict the co-occurrence of food items in recipes and contribute to the discovery of novel food pairings. Most recently, Reciptor [18] suggested a set transformer-based model for obtaining recipe embeddings and used a knowledge graph (KG) 19 derived triplet sampling strategy to improve the acquired embeddings. Node2vec [20] has been used to create node representations from network data in graph-based embedding methods. Random walks are generated by Node2vec based on its network's relations, and the walks are equivalent to sentences in word2vec. Metapath2vec [21] generates random walks termed metapaths, which are used to create similar representations of heterogeneous nodes based on often related nodes. These data-driven and graph-based techniques that embed conceptual representations with rich domain-specific information may improve food pairing recommendations because they can be used to build food representations based on the relationships between different foods and chemical compounds.

Using the approaches of previous studies, we introduce a novel method of creating a food pairing dataset from a large corpus of PDFs which were extracted from PubMed using NLP techniques. Our approach is different from the others that we did not consider either chemical composition or other features of food pairings. We just focus on inhibiting and enhancing the characteristics of foods involved in food pairing.

## 3. Methodology

We describe the proposed relation extraction (RE) technique, starting with the extraction of PubMed Ids (23316) from PubMed via keyword matching. Retrieved PubMed Ids are used to download all of the PubMed articles. Extracting 23316 PubMed items was time-consuming and required several hours. We use batching on PubMed Ids to make this process a little easier. We were able to retrieve 7546 PubMed articles, including some HTML and corrupted files. We were then left with 5559 articles that could aid us in achieving our goal.




First level Filtering:-

First level filtering was done on the remaining PubMed articles using two keyword lists (Food and Nutrient). Again batching was done to make our process fast. After parsing of these articles, the count decreases to 5285. At this level, filtering was done on whole article and wherever there was keyword match, that article was kept otherwise not.

Second level Filtering: - The matched articles were parsed again using some different condition like food count should be greater than zero. After this stage the number of articles left were 5080.This level of filtering was performed on sections of articles.

Third level Filtering: - At this level, line by line filtering was done and wherever there is match the above three lines and below three lines were taken. This process was done on 5080 articles and due to which we got 1113 entries in our dataset.

Data cleaning:-

Data cleaning is an important step in every machine learning model, but it is especially important in NLP. Without the cleaning procedure, the dataset is frequently a jumble of words that the machine is unable to comprehend. We'll go over the procedures involved in cleaning data as follows:-

Step 1:  Lemmatize or Stem

Stemming and Lemmatizing is the process of reducing a word to its root form. The main purpose is to reduce variations of the same word, thereby reducing the corpus of words included in the model. The difference between stemming and lemmatizing is that, stemming chops off the end of the word without taking into consideration the context of the word. Whereas, Lemmatizing considers the context of the word and shortens the word into its root form based on the dictionary definition. Stemming is a faster process compared to Lemmatizing. Hence, it a trade-off between speed and accuracy.

Step 2:- Remove numbers, hashtags, URLS and special characters from text.

After all these steps our dataset contains 1113 entries which contains inhibiting and enhancing entities. The dataset can be used for so many purposes in future and it helps us in identifying new food pairings.

## 5. Results

We performed preliminary analysis on the large amount of PubMed articles which were extracted from PubMed. We first examined the relevancy between foods and nutrient lists. Next, we looked at the enhancing and inhibiting characteristics of food pairings. After formation of dataset we created Food-Nutrient graph on both inhibiting and enhancing food pairings. We created two graphs

A) Food-Nutrient enhancing Graph
B) Food-Nutrient inhibiting Graph

The food combinations that have boosting qualities on each other are shown in Figure. 1. Also, Figure 2 depicts the inhibitory effects of food pairings on one another. We discovered that several meal pairings were clearly defined, such as rice-tomato, rice-beans, rice-meat, and rice-potato, which we discovered in our dataset. These combinations complement each other, but the meat and egg pairings we found in our dataset had inhibitory qualities, which is also true in real life.

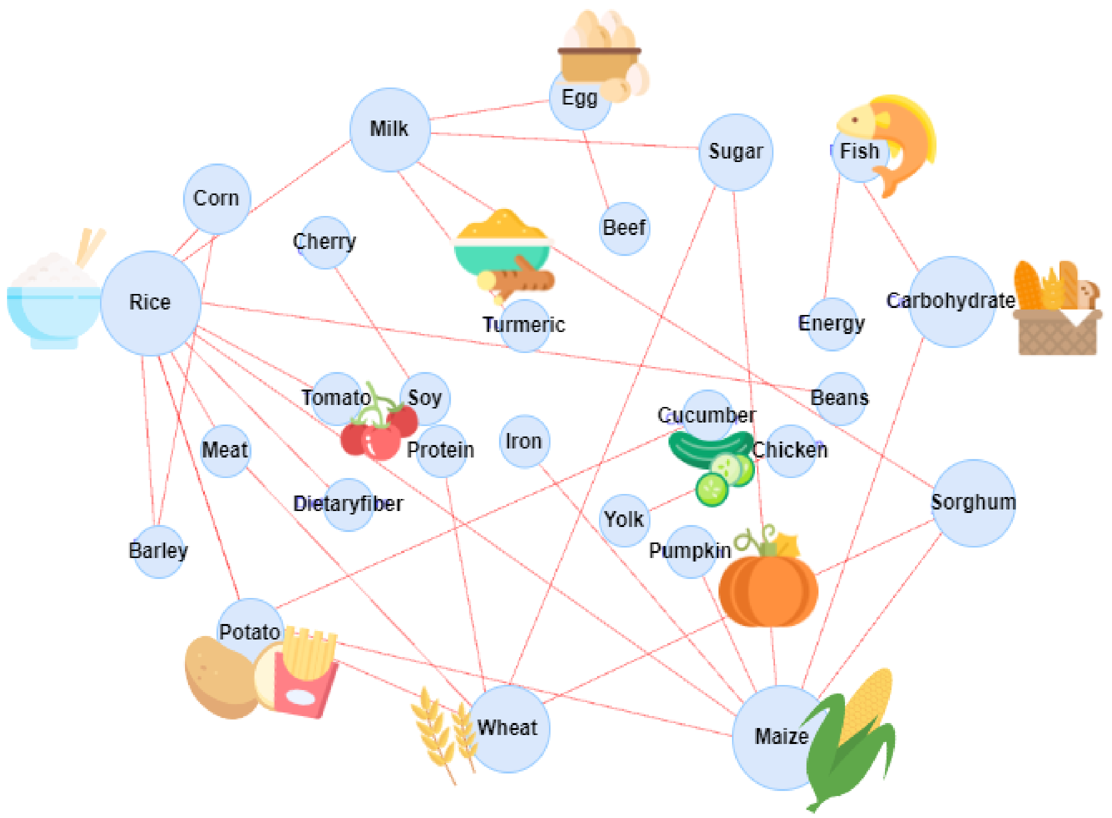



Fig. 1 shows enhancing properties of food pairings

The above graph was designed from the food pairings of our created dataset. The graph depicts pairings that have a favourable impact on one another.

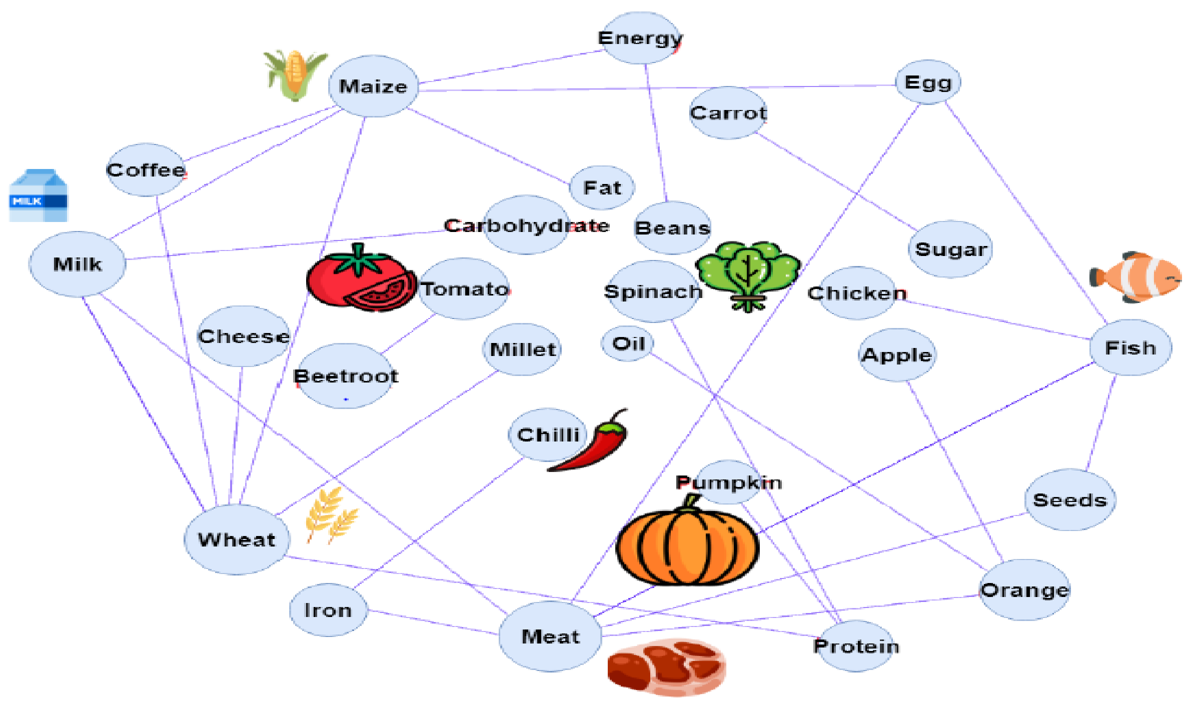



Fig.2 shows inhibiting property of food pairings

The graph in Fig.2 shows Food pairings that have a negative impact on each other, indicating that they are not ideal for mixing.

## 6. Conclusion

We propose a new Relation Extraction (RE) approach for the detection of enhance and inhibit relations between food and nutrient entities from raw text. We create the Food pairing dataset, which consists of 1115 entries (sentences from 7546 PubMed articles) for the existence of enhancing and inhibiting relations between food and nutrient items. To the best of our knowledge, this dataset is the first annotated relation extraction dataset of such kind in the food domain. Our work will significantly improve understanding of food pairings. However, our work has some limitations; first, more food-related information is needed to better understand food pairings. This work only focuses on food-nutrient information. Second, we are extracting relations from PubMed articles (research papers) which necessarily need not be precise. Nevertheless, we believe that this food pairing dataset can be employed for a better understanding and uses of food.

In the future, we need to take into account various characteristics as well as the chemical composition of food pairings. Apart from PubMed, we need to pull articles from other repositories. We should use new NLP techniques to extraction relations from raw data. We have several obstacles in comprehending food and its pairings, thus this topic is open to investigation.

## 7. References


*1. Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P. & Barabási, A.-L. Flavor network and the principles of food pairing. Sci. Rep. 1, 196 (2011).*

*2. Ahn, Y.-Y. & Ahnert, S. The flavor network. Leonardo 46, 272–273 (2013).*

*3. Garg, N. et al. Flavordb: a database of flavor molecules. Nucl. Acids Res. 46, D1210–D1216 (2017).*

*4. Simas, T., Ficek, M., Diaz-Guilera, A., Obrador, P. & Rodriguez, P. R. Food-bridging: a new network construction to unveil the principles of cooking. Front. ICT 4, 14 (2017).*

*5. Teng, C.-Y., Lin, Y.-R. & Adamic, L. A. Recipe recommendation using ingredient networks. In Proceedings of the 4th Annual ACM Web Science Conference, 298–307 (ACM, 2012).*

*6. Zhu, Y.-X. et al. Geography and similarity of regional cuisines in China. PLoS ONE 8, e79161 (2013).*

*7. Jain, A., Rakhi, N. & Bagler, G. Analysis of food pairing in regional cuisines of india. PLoS ONE 10, e0139539 (2015).*

*8. Wagner, C., Singer, P. & Strohmaier, M. The nature and evolution of online food preferences. EPJ Data Sci. 3, 1–22 (2014).*

*9. Abbar, S., Mejova, Y. & Weber, I. You tweet what you eat: studying food consumption through twitter. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 3197–3206 (2015).*

*10. Zhang, F. et al. Exploiting dining preference for restaurant recommendation. In Proceedings of the 25th International Conference on World Wide Web, 725–735 (2016).*

*11. Grace, K., Maher, M. L., Wilson, D. C. & Najjar, N. A. Combining cbr and deep learning to generate surprising recipe designs. In International Conference on Case-Based Reasoning, 154–169 (Springer, 2016).*

*12. Grace, K. & Maher, M. L. Surprise-triggered reformulation of design goals. In Thirtieth AAAI Conference on Artificial Intelligence (2016).*

*13. Park, D., Kim, K., Park, Y., Shin, J. & Kang, J. Kitchenette: predicting and ranking food ingredient pairings using siamese neural networks. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (2019).*

*14. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv :1301.3781 (2013).*

*15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 3111–3119 (2013).*

*16. Salvador, A. et al. Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017).*

*17. Salvador, A., Drozdzal, M., Giro-i Nieto, X. & Romero, A. Inverse cooking: recipe generation from food images. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019).*

*18. Li, D. & Zaki, M. J. Reciptor: An effective pretrained model for recipe representation learning. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1719–1727 (2020).*

*19. Haussmann, S. et al. Foodkg: a semantics-driven knowledge graph for food recommendation. In International Semantic Web Conference, 146–162 (Springer, 2019).*

*20. Grover, A. & Leskovec, J. node2vec: scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864 (ACM, 2016).*

*21. Dong, Y., Chawla, N. V. & Swami, A. metapath2vec: scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 135–144 (ACM, 2017).*