Chapter- IV

Results and Discussion

Chapter IV

4. Results and Discussion

This chapter presents the findings of the investigation conducted to find solutions to the research questions stated in the objectives. The findings are critically appreciated by relating them to theoretical understandings, and/or research findings of earlier workers, and are presented in the sequence of the objectives from sections 4.1 to 4.4

4.1 Characterising and analysing traditional food recipes for ingredient pairing

The strategy employed to fulfil objective 1 involves recipe data collection, pre-processing of the recipe data and ingredient flavour compound data collection as the first task as described in section 3.1.1 and the results are presented in subsection 4.1.1. During the pre-processing of the recipe data, redundant terms, phrases, and quantifiers are removed from the data in order to generate the ingredients required for the study. Subsection 4.1.2 deals with the statistical analysis of recipe size and frequency rank distribution which is carried out as a preliminary investigation. Accordingly, the results in the form of parameters such as the authenticity of ingredients, food pairing and the contribution of ingredients are discussed below.

4.1.1 Data on recipes and flavour compounds

4.1.1.1 Recipe data for regional cuisines

An extensive data curation process was used to collect traditional recipes from northeast regional cuisines as described in section 3.1.1. As a result of the curation process, we collected 702 recipes for the eight regional sub-cuisines: Assam, Arunachal, Manipur, Meghalaya, Mizoram, Nagaland, Tripura and Sikkim. It is interesting to note that cultures and geographical influences abound in the Northeast region of India. However, in contrast to earlier studies conducted on the Indian regional cuisines, there are limited data on Northeast regional cuisines, resulting in fewer recipes collected.

The statistics of the regional cuisine recipe data consisting of the recipe and its ingredients are listed in Table 4.1. A total of 126 ingredients were determined from the recipes after pre-processing as described in sub section 3.1.1.2. These ingredients were further divided into constituent categories based on their nature and origin. As a result, a total of 13 categories of ingredients were obtained viz., vegetables, spice, fruit, nut/seed, cereal/crop, plant derivative, dairy, pulse, fish/seafood, herb, meat, plant and animal product, listed in the Table 4.2. Ingredients from the categories of vegetables and spices accounted for the highest number of ingredients as compared to others.

Table 4.1 Regional cuisine statistics of recipe data

Regional cuisine	Number of recipes	Number of ingredients	Average recipe size
Assam	401	105	6.725
Arunachal	41	49	4.199
Manipur	38	56	7.148
Meghalaya	34	28	4.428
Mizoram	30	39	4.121
Nagaland	78	38	4.734
Sikkim	52	52	4.708
Tripura	28	37	5.045

Table 4.2 Number of ingredients in each category

Sl no.	Ingredient category	Number of ingredients in each category
1	vegetable	31
2	spice	22
3	fruit	17
4	cereal/crop	9
5	plant derivative	8
6	nut/seed	7
7	dairy	7
8	pulse	6
9	fish/seafood	5
10	plant	4
11	meat	4
12	herb	4
13	animal product	2

4.1.1.2 Datasets on edible ingredients and flavour compound

The data on ingredients and the flavour compound curated from the archived data as explained in section 3.1.1.3 along with the additional flavour compounds data of ingredients from Northeast were maintained consisting of its unique compound ID and the corresponding compound name. The information gathered from the dataset of Ahn et al. [5] consists of ingredients starting from ID 0 to 1531 and flavour compounds from ID 0 to 1107. The updated archive dataset was maintained by Jain et al. [39] with the addition of 50 ingredients from the Indian sub-cuisine, thus the ingredients ID tallied up to 1581 flavour compounds data tallied up to 1414 with additional 307 flavour compounds. The final list for the study was obtained after adding the ingredients from the Northeast to the dataset of Rakhi's which is the updated version of Ahn's dataset. The final dataset of ingredients ID tallied up to 1586 and flavour compound dataset tallied up to 1561. The basic statistics of the datasets are shown in Table 4.3. This final dataset after adding the additional ingredients from the Northeast regional cuisines were used for the study.

Table 4.3 Statistics of different datasets

	Number of ingredients	Ingredients dataset ID	Additional	Flavour compounds dataset ID	Additional
Ahn	381	0-1531	-	0-1107	-
Rakhi	194	0-1581	50	0-1414	307
Northeast	126	0-1586	5	0-1561	147

Further, five ingredients unique to Northeast cuisines were added to the archived list of Ahn's and Jain's both in the ingredient's dataset and the flavour compound dataset. The additional ingredients along with the source and the number of flavour compounds are listed in Table 4.4. The detail list of flavour compounds is given in Table-A1, A2, A3, A4 and A5 of appendix.

Table 4.4 List of additional ingredients from Northeast

Name of Ingredients	Source	No. of flavour compounds
Bamboo shoot	Garg et al. [6]	90
Fermented bamboo shoot	Fu et al. [5]	56
Fermented rice	Lee et al. [9]	29
Fermented soybean	Chung [4]	90
Fermented fish	Mohamed et al. [11]	72

4.1.2 Characteristics of the regional cuisines from the Northeast states

4.1.2.1 Average recipe size and ingredient frequency

An analysis was conducted on the statistical significance of recipes and the patterns of ingredient usage across the eight regional cuisines as per section 3.2.1. The statistical characteristic of the cuisine is highlighted by the size of the recipe and ingredient frequency rank distribution. The variation in recipe size indicates the richness of ingredient distribution. Northeast cuisine showed a bounded distribution of recipe sizes, ranging from 1 to 20, with an average recipe size of 7 see Fig. 4.1 (a), which is similar with the recipe size of Indian regional cuisines [38].

Ingredient frequency versus rank reveals bias in ingredient use by ordering ingredients according to their prevalence in a cuisine. As we sort the items in decreasing usage frequency, the pattern of ingredient distribution across each regional cuisine revealed an invariant pattern with three orders of magnitude variation following a scale-free distribution shown in Fig. 4.1b. This result is in line with studies reported on Indian regional cuisines [38]. Furthermore, we can observe that some specific ingredients are overused, indicating their popularity in the cuisine. In addition, the complementary cumulative degree distribution shown in Fig. 4.1(c) of the ingredients in the cuisine displays a pure power law with fit proportionate $k^{-0.77}$ indicating that an ingredient is found in more than *k* recipes.

4.1.2.2 Prevalence of ingredients

Findings of 4.1.2.1 indicated that some specific ingredients are used more frequently than others thereby indicating their popularity in the cuisine. For the identification of the popular ingredients, equation 3.2 of section 3.2.1.2 was used to analyse the prevalence of ingredients in the regional cuisines. The top five prevalent ingredients of each regional cuisine are listed in Table A7 (a).

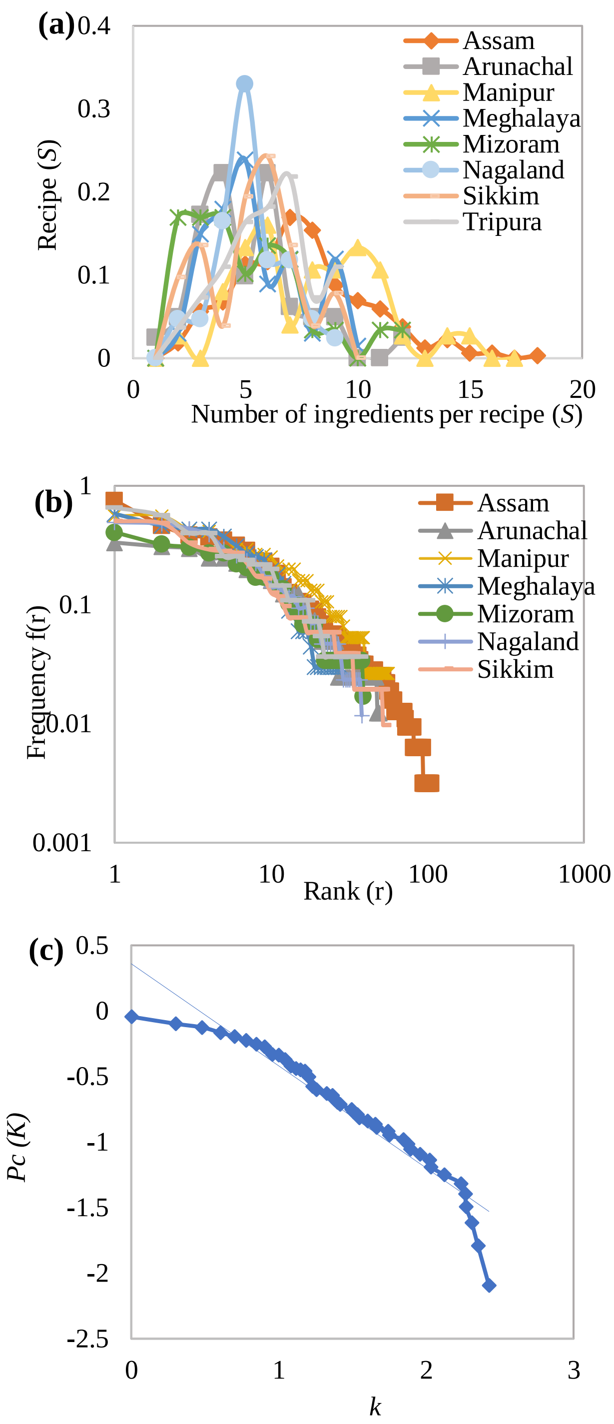


Fig. 4.1 a) Regional cuisine recipe size distribution (b) Regional cuisine frequency rank distribution (c) Regional cuisine complementary cumulative degree distribution plot with power-law fitting $k^{-0.77}$

From the Table-A7 (a), we can observe that ingredients such as black mustard seed oil, onion, cayenne, ginger, green bell pepper, garlic, turmeric, bay laurel, pork, rice and tomato are the most prevalent ingredients across the cuisines. Most of the prevalent ingredients are found to be from the category of spice. Further, black mustard seed oil and onion appear as prevalent ingredients in all the regional cuisines except for the cuisines from Nagaland.

For a better comprehension across all the cuisines and for a comparison with combined cuisines from all the eight regions, top-ten prevalent ingredients are listed in Table-A7 (b) along with region specific cuisines. As a whole for the NE region black mustard seed oil and onion appear as the most prevalent ingredients. Spices like bay laurel, and garlic and meat category ingredient viz., pork appearing in this list which appear in top 05 prevalent in one or two regional cuisines only. For an understanding about the contribution of these prevalent ingredients towards characterization of the cuisines, authenticity values of these ingredients within the cuisines were analysed.

4.1.3 Ingredient authenticity

4.1.3.1 Authentic ingredients from the Northeast cuisines

The taste palette of regional cuisine is characterized by its unique set of ingredients and ingredient combinations. Based on specific flavours, authentic ingredients of a cuisine illustrate the differences between regional cuisines, highlighting each cuisine's signature taste. List of authentic ingredients, authentic pairs and authentic triplets are determined as shown in Table-A8.

The majority of the authentic ingredients in Northeast regional cuisines as listed in Table A8 were found to be from the category of plant derivatives, spices and vegetables such as black mustard seed oil, green bell pepper, ginger, cayenne, bay laurel, garlic and turmeric, except for pork from meat category. We can observe that among all the ingredients listed mustard seed oil i.e., an edible oil seems to be widely used in the recipes. As expected, the column of authentic ingredients resonates with the list of prevalent ingredients.

To observe if the most prevalent ingredients contribute towards the authentic pairs and triplets, based on the data analysis, columns (2) and (3) (Col-2, Col-3) of Table A8 are prepared. In the list of the top five authentic pairs and triplets of the regional cuisine, it can be seen that in both cases the ingredients pairs and ingredients triplets are mostly between ingredients of spice, plant derivative and vegetable categories. Additionally, we observed that the most prevalent ingredients (of Table-A7 (a) and (b)) contribute to the authenticity of cuisines in pairs and triplets with the inclusion of a few more ingredients, viz., cabbage, carrot, bean, coriander, sesame seed, bamboo shoot, fish, fermented fish.

The pair of ingredients black mustard seed oil and onion appeared as authentic pairs, and as one of the common triplets (14 occurrences out of 40 sampled triplets). The ingredient Pork appeared as an authentic ingredient in one cuisine, authentic pair on one occasion and authentic triplet on 03 occasions. These can be seen as one distinct characteristic of the Northeast regional cuisine as compared to the Indian regional cuisines reported by Jain et al. [39].

Further, the ingredient usage pattern is almost similar across the Northeast regional cuisine. Geographical proximity may be a factor in the similarity of ingredients as the Northeast states are close to each other Zhu et al. [94] in their findings have also highlighted that geographical distance increases the usage of similar ingredients.

4.1.3.3 Flavour sharing among authentic ingredients

We have shown the number of shared flavour compounds between the five authentic pairs and triplets of the regional cuisine in Table-A9**.** In most cases, we observed that the ingredients share less or no flavour compounds. As compared to ingredients from vegetable categories, spice ingredients share fewer flavour compounds. The maximum number of shared flavour compounds is between tomato and onion with twenty number of common ingredients between them.

As reported in earlier works, the difference between the number of shared compounds in real cuisines and random cuisines yield a negative value when the ingredients in recipes constituting the cuisine do not share much flavour compounds [5,39]. This difference was used to explain the food pairing behaviour. Before applying the principle, we need to know if flavour compounds are common at the ingredient level. We have shown the number of shared flavour compounds between the five authentic pairs and triplets of the regional cuisine in Table-A9.

The table reveals that in most cases the ingredients of the authentic pairs/ triplets share less or no flavour compounds. Further, as compared to ingredients from vegetable categories, spice ingredients share fewer flavour compounds. The maximum number of shared flavour compounds is between tomato and onion with twenty number of common ingredients between them. The concept of flavour pyramid is used for a graphical representation of the flavour sharing behaviour of the authentic pairs and triplets.

4.1.3.2 Flavour pyramid of authentic ingredients

We determined the affinities toward ingredient pairs based on the number of shared flavour compounds for the authentic pairs and triplets. We organised the five most authentic single ingredients, ingredient pairs and ingredient triplets for the Northeast regional cuisines in a flavour pyramid. The illustration of the flavour pyramid of each regional cuisine is shown in Fig. 4.2. The size of the node indicates how prevalent an ingredient is in the particular cuisine's recipes and the flavour compounds shared are indicated by the link thickness between the node. The colour of each node represents the ingredient’s category.

The flavour pyramid reveals the links between the ingredient pairs and ingredient triplets are not very significant, indicating that the ingredients do not share much flavour compounds. The regional cuisines heavily rely on ingredients from plant derivatives such as black mustard seed oil, followed by spices-based ingredients such as green bell pepper, turmeric, ginger, bay laurel etc. This result is similar to the East Asian cuisine as reported in earlier studies by Ahn et al. [5]. However, it is in contrast with the studies reported on North American food as they rely mostly on dairy-based ingredients with a significant number of compounds shared between the ingredient’s pairs and triplets. Overall, it was evident that Arunachal cuisine contains a wider variety of ingredients than any other regional cuisine.

Thus, considering the fact that Northeast regional cuisines have less common ingredient combinations we discussed the ingredient pairing behaviours in the upcoming section. This is investigated across the recipes for obtaining the difference of the average value of the number of shared compounds.

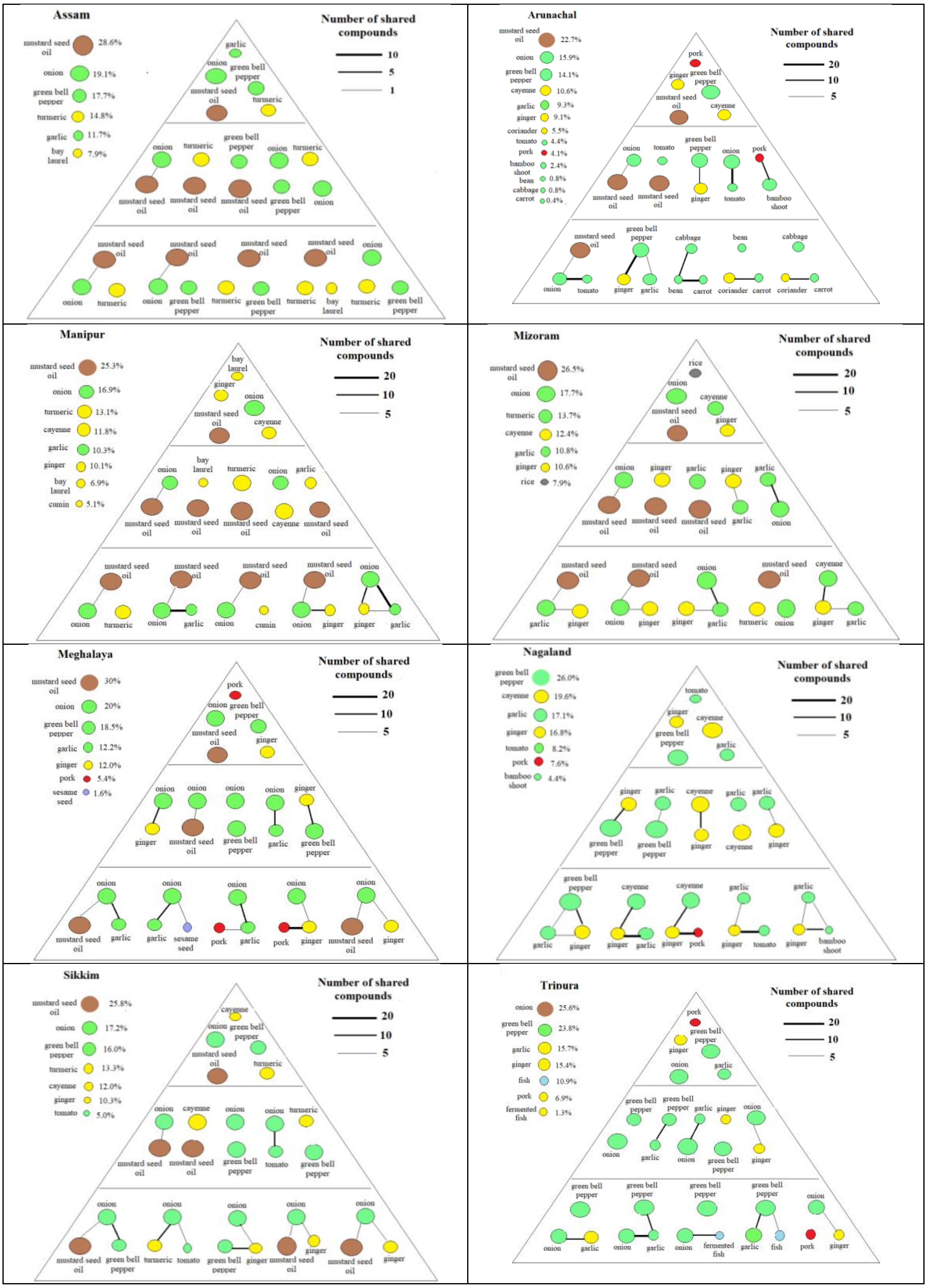


Fig. 4.2 Flavour pyramids for the Northeast regional cuisines

4.1.4 Flavour profiles and sharing of compounds among ingredients

The Flavour pyramids revealed that the most authentic ingredient combinations of the Northeast recipes share very little in terms of flavour compounds, the question is to what extent their flavour profile influence the ingredient pairing behaviour. For this the cuisine level pairing behaviour is studied and then the influence of these ingredients is estimated.

Ingredient pairing behaviour based on flavouring compounds were studied at the level of recipes, ingredients and cuisines. By studying these cuisines at multiple levels, it was possible to gain a comprehensive understanding of their ingredient usage patterns. The subsequent analysis revealed the role of ingredients and ingredient categories in determining food pairings of the regional cuisines by identifying the features that contribute to food pairing. We determine the systematic difference in the recipes of regional cuisines by comparing these with randomized recipe dataset, in terms of flavour sharing behaviour.

4.1.4.1 Characteristics of randomized cuisines

We attempted to explore the potential factors that could explain the negative relationship between regional cuisines and food pairings. The criteria used is an estimation of the difference of average count of shared flavour compounds across the cuisine with respect to a randomized cuisine ΔN_S . Four randomly generated cuisines were used. As explained in the section 3.1.2 the uniform selection of ingredients provided the first model of random control. The second model was generated by selecting an ingredient while keeping its frequency in mind. The third model was generated by considering the ingredient while keeping the category in mind. The final model was generated by choosing ingredients while keeping the category and frequency in mind. The size of each of the random cuisines were 10,000. Before the computation of the average count of shared flavour compounds across the cuisine, the random cuisines are compared with the original regional cuisines by the rank frequency plots.

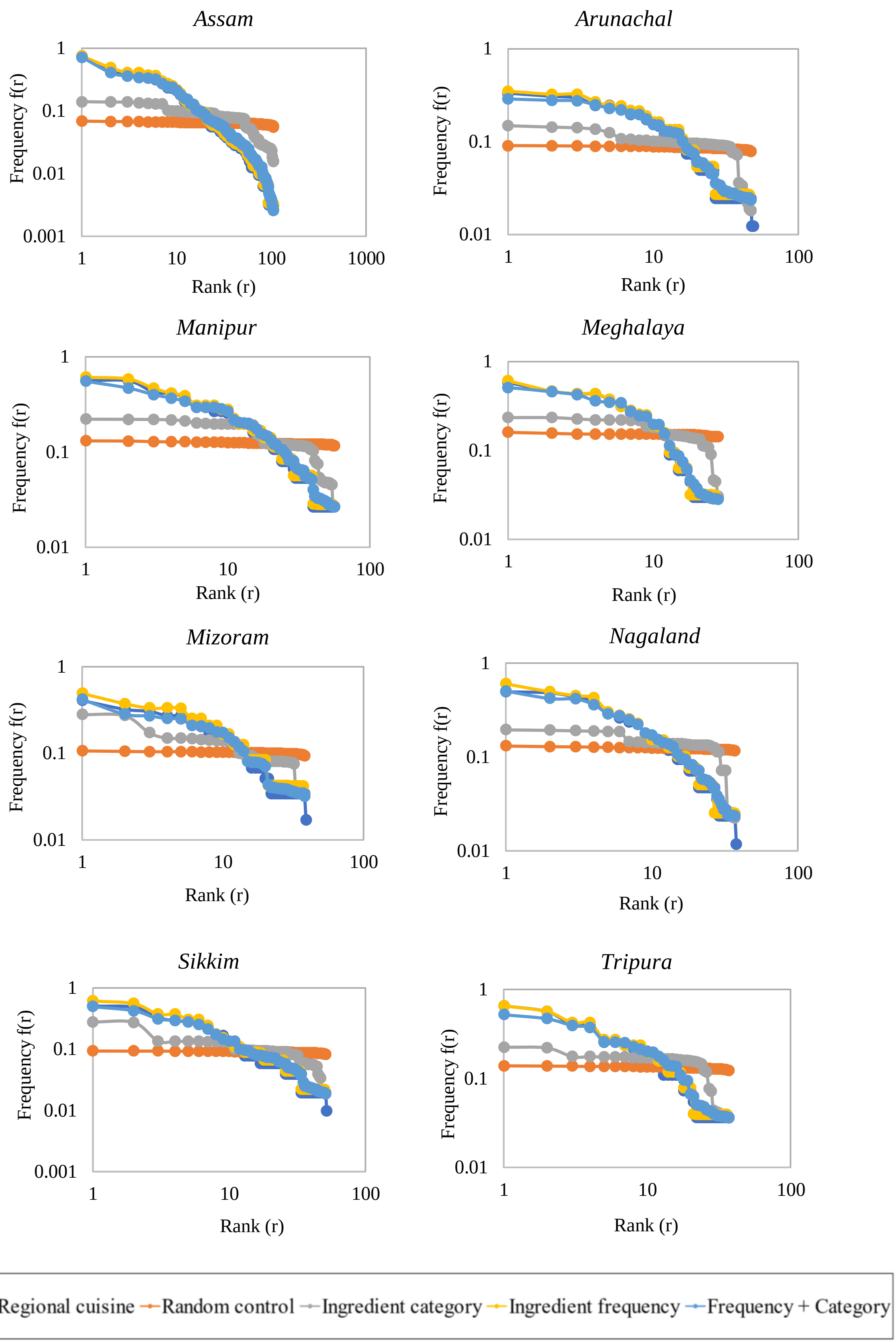


Fig. 4.3 Frequency rank distribution with its corresponding random models

The plots for each of these regional cuisines are shown in Fig 4.3. The plots reveal that as expected the rank-frequency plots of second model and the fourth model (frequency preserving and frequency + category preserving) almost reproduces same pattern as the original cuisines (the correlation co-efficient > 0.99). However, the plots for the first model (random) and third model (category preserving) differ with large deviations (correlation co-efficient < 0.9). In spite of a highly correlated rank frequency plot, there is no reason to assume that the ingredient wise frequency distribution is similar. While a similar rank-frequency plot may prompt to expect a value of ΔN_S close to 0.0, a bias towards a specific ingredient combination in the real cuisine is likely to yield a non-zero value for ΔN_S. Outcome of such an estimation and analysis is included in following section.

4.1.4.2 Flavour sharing among ingredients of recipes

At the level of recipes, ingredients and cuisines, the pattern of food pairing was examined. By studying these cuisines at multiple levels, it was possible to gain a comprehensive understanding of their ingredient usage patterns. The subsequent analysis revealed the role of ingredients and ingredient categories in determining food pairings of the regional cuisines by identifying the features that contribute to food pairing.

We determine the systematic difference in the recipes of regional cuisines by comparing the randomized recipe dataset with the observed number of shared compounds characterizing the regional cuisines (Eq.3.5). The art of culinary science varies across the regional cuisines as we observe distinct differences in the choice of ingredient usage and ingredient combination which stands unique to the particular region. Fig. 4.4 illustrates the statistics of the shared compound hypothesis at the regional cuisine recipe level. The extent of bias in the eight regional cuisines viz, *Assam, Arunachal, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura*, when compared with the corresponding randomized cuisine showed uniform negative food pairing behaviour. In comparison, *Sikkim* showed the most negative food pairing with ΔN_s value of -3.188 while, *Assam* showed the least negative with ΔN_s value of -0.726 (Table 4.5). As a result, the regional cuisine dishes use fewer compound sharing combination than one might expect by chance. This result correlates with earlier research on Indian regional cuisine, where the more flavour compounds two ingredients share, the less likely they are used together [39].

However, it differs from North American recipes as ingredients are more likely to be used together if they share flavour compounds [5].

Table 4.5 Statistics of food pairing behaviour

Cuisine	$\Delta N_S = \overline{N_S}$ *(real)* - $\overline{N_S}$ *(rand)*
Assam	-0.726
Arunachal Pradesh	-1.187
Manipur	-0.927
Meghalaya	-1.448
Mizoram	-2.799
Nagaland	-2.095
Sikkim	-3.188
Tripura	-2.414

4.1.5 Ingredient contribution to food pairing behaviour

In Northeast regional cuisine, ingredient pairs tend to share fewer flavour compounds than one might expect by chance. This result is in contrast with North American dishes as they tend to use more compound-sharing ingredient pairs [5]. As a result, we investigated the underlying process causing these variations. Therefore, we quantified the contribution χ_i (Eq. 3.6) of each ingredient, estimating to what extent its presence impacts the shared compound effect (ΔN_s) in a given cuisine c. We presented a scatter plot for each ingredient of the Northeast regional cuisine in Fig. 4.5. The size of the circles represents its prevalence. If an ingredient lies on the 0 axis (χ_i = 0) it indicates that the contribution of the particular ingredients is negligible. We observed that few frequently used ingredients of the Northeast regional cuisine lie predominantly in the negative χ_i region. This result is similar with studies reported on East Asian cuisine but it is contrast with the North American cuisines where most of the ingredients tend to lie in positive χ_i region. According to our findings, the majority of the ingredients that made a substantial impact on food pairing were from the spice category. This result suggests that a few outliers commonly used in a particular cuisine probably account for the negative food pairing effect, e.g., black mustard seed oil, green bell pepper, cayenne, onion, garlic, turmeric, bay laurel,

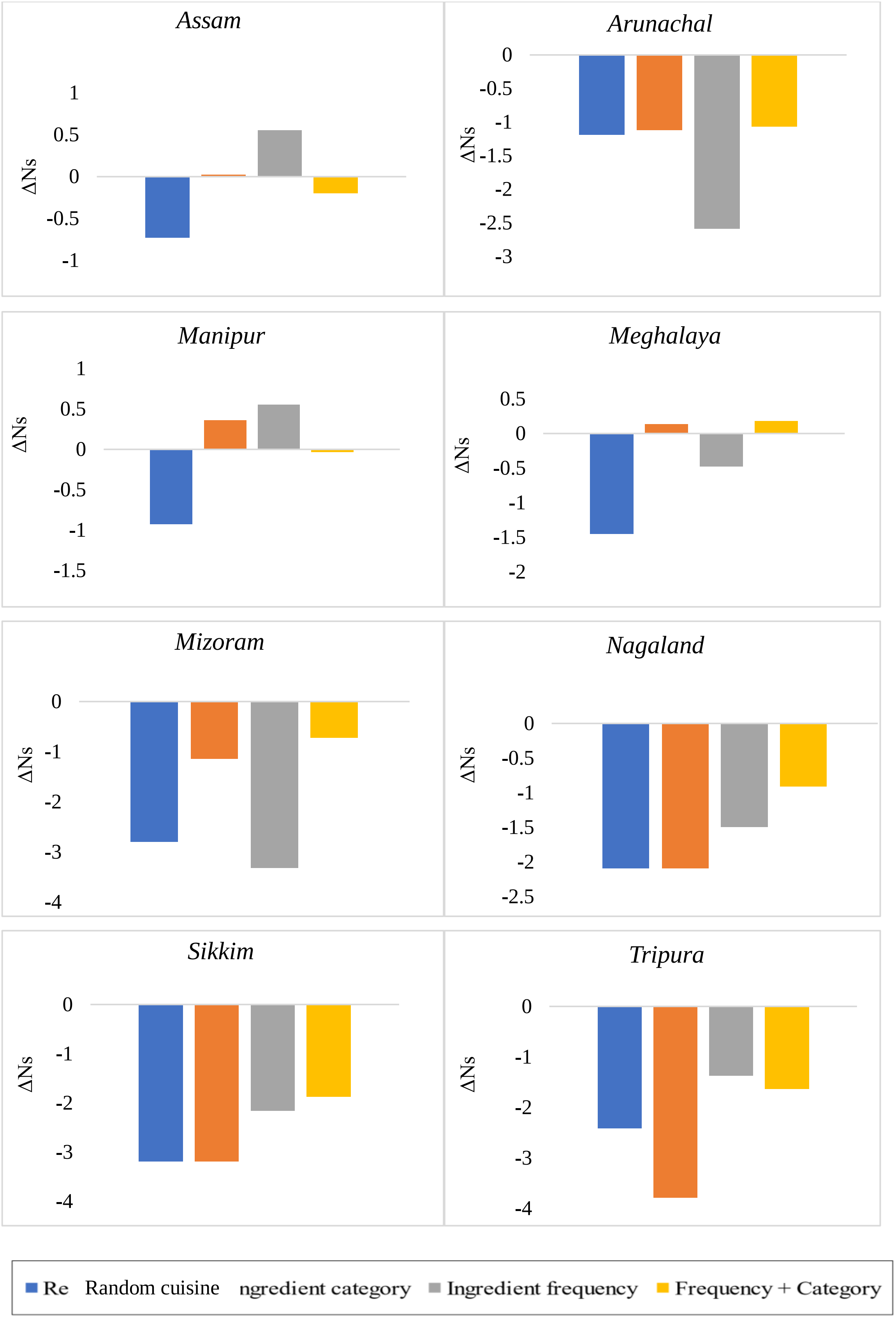
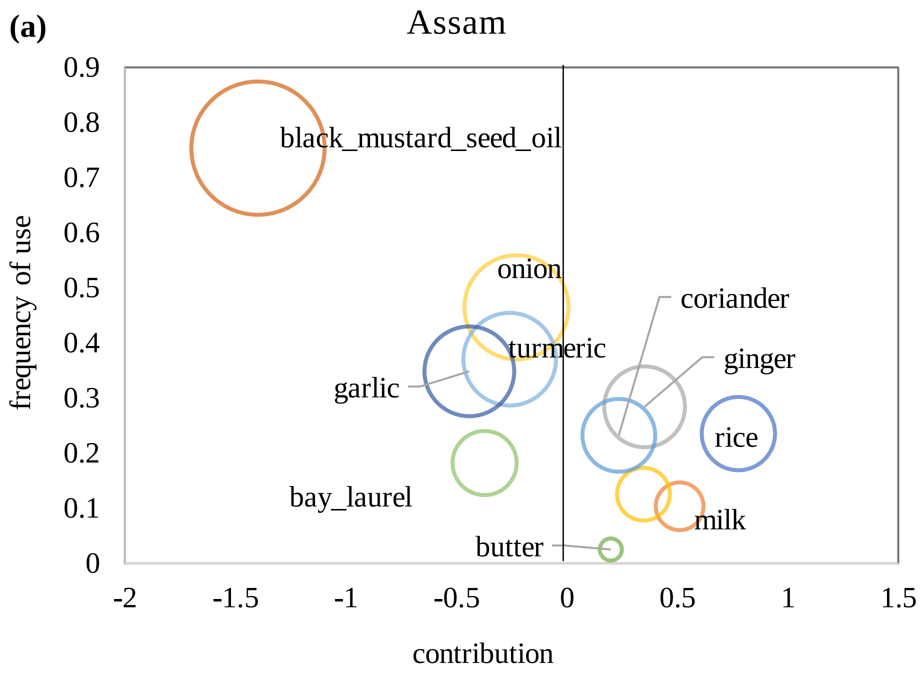
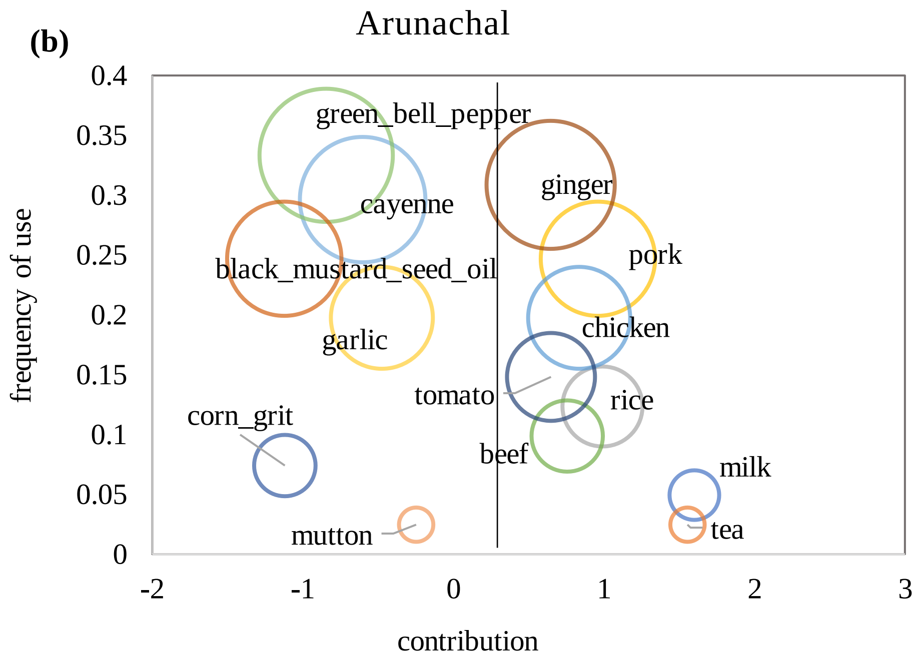
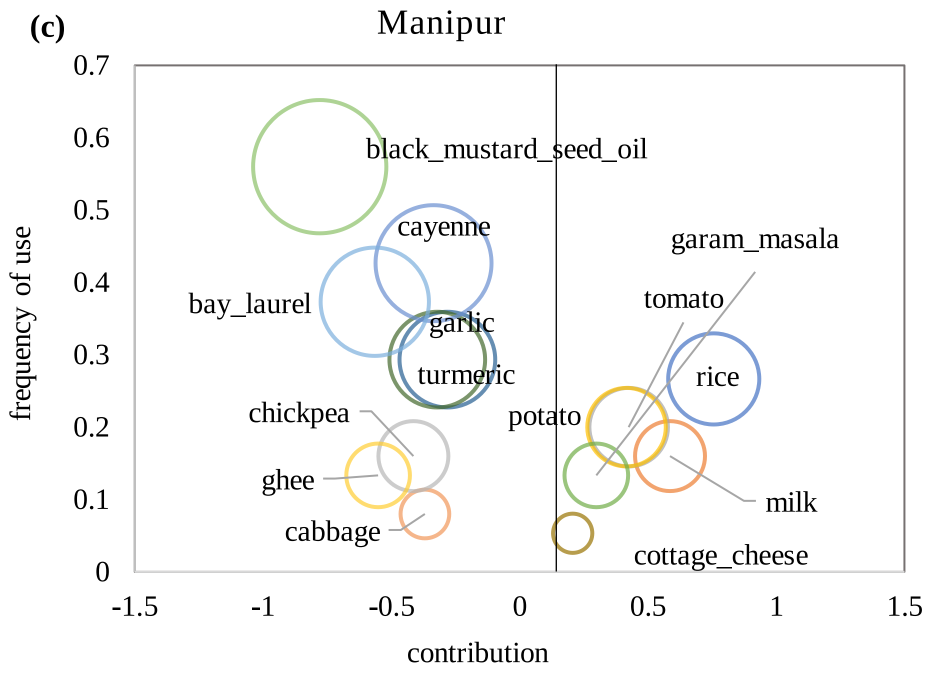


Fig. 4.4 Statistical significance of Δ*Ns*, which indicates the extent of bias in food pairings of regional cuisines with their random models


(a)
Assam
frequency of use
contribution
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
black_mustard_seed_oil
onion
turmeric
garlic
coriander
ginger
rice
bay_laurel
milk
butter

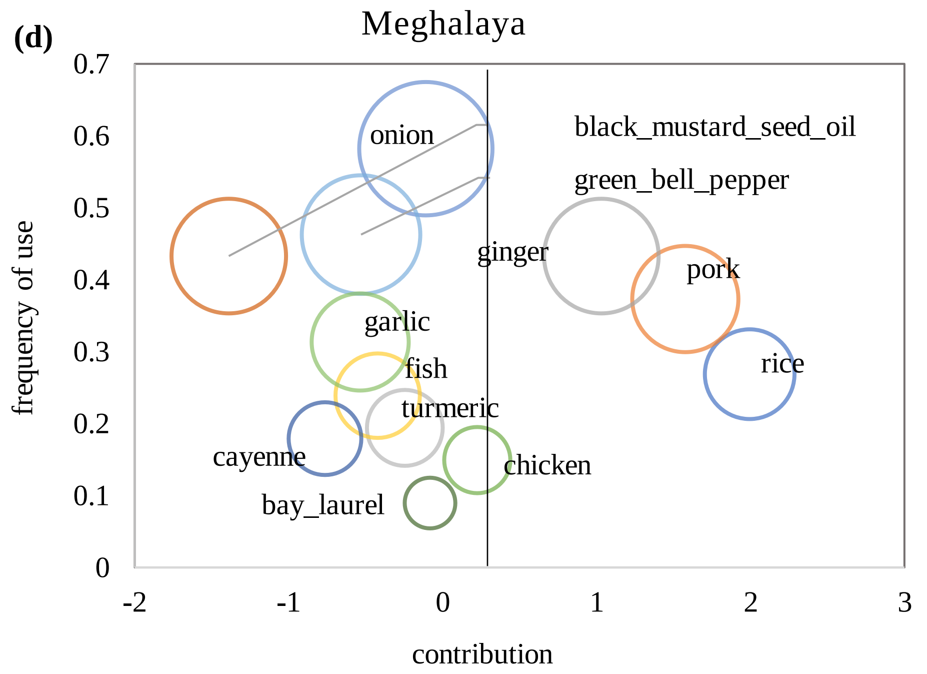

(b)
Arunachal
frequency of use
contribution
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-2
-1
0
1
2
3
green_bell_pepper
cayenne
ginger
black_mustard_seed_oil
pork
garlic
chicken
tomato
rice
corn_grit
beef
milk
mutton
tea

(c)



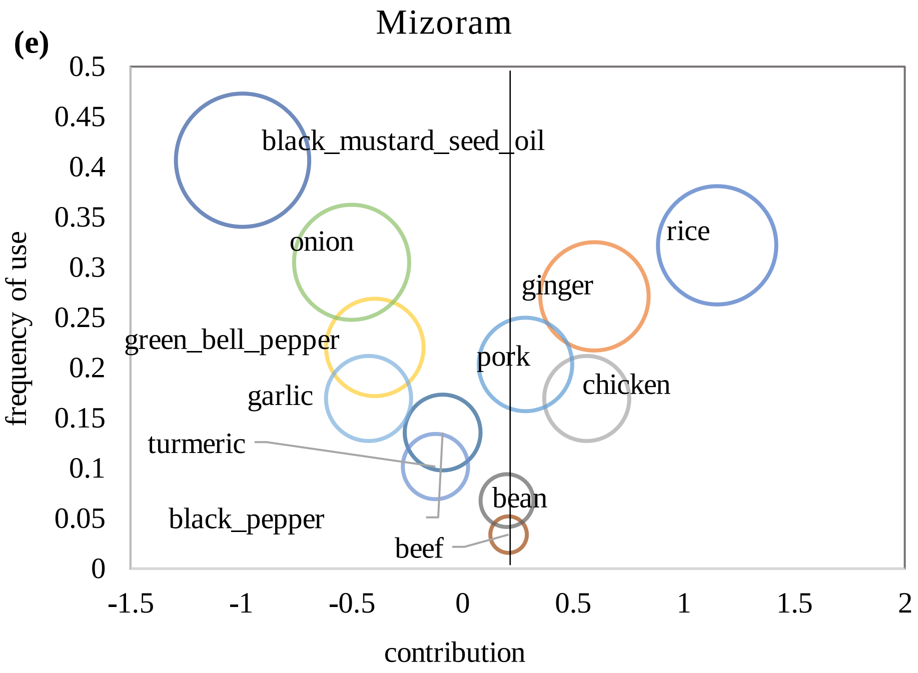
Manipur
frequency of use
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1.5
-1
-0.5
0
0.5
1
1.5
contribution
black_mustard_seed_oil
cayenne
bay_laurel
garlic
turmeric
chickpea
ghee
cabbage
garam_masala
tomato
rice
potato
milk
cottage_cheese

(d)

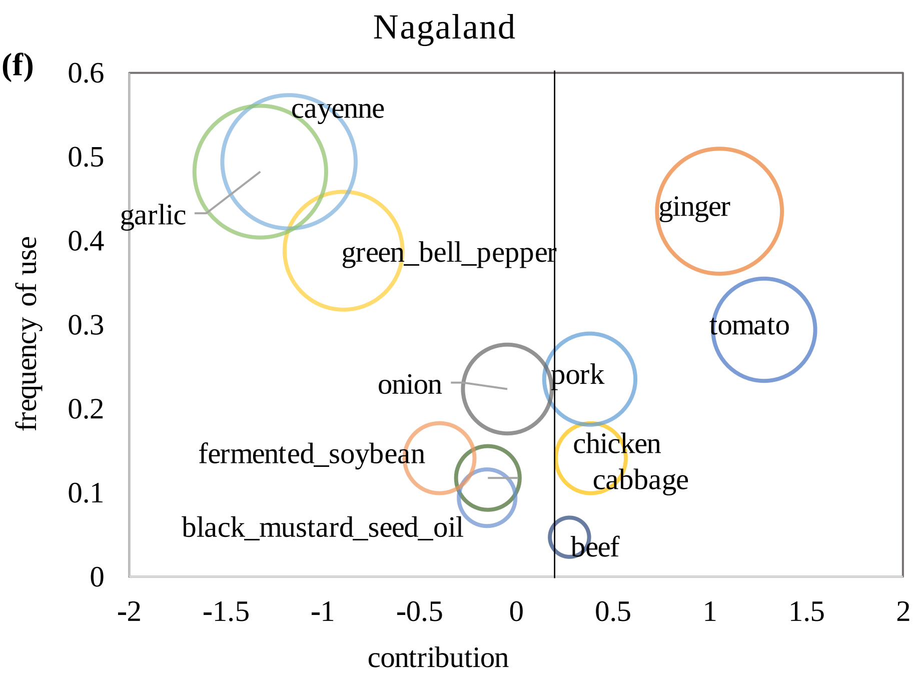


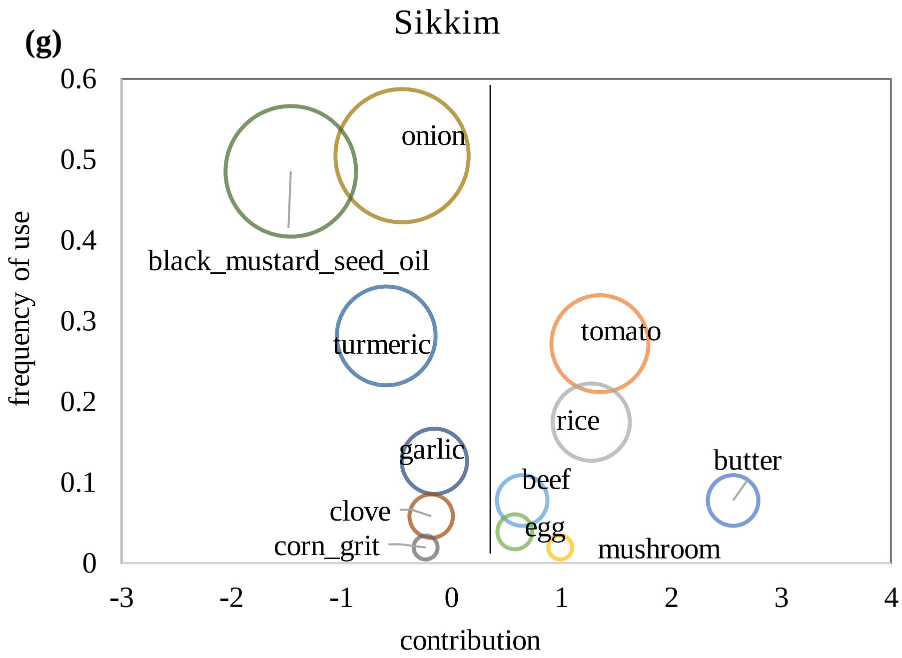
Meghalaya
frequency of use
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2
-1
0
1
2
3
contribution
onion
black_mustard_seed_oil
green_bell_pepper
ginger
pork
garlic
fish
rice
turmeric
cayenne
chicken
bay_laurel

(e)


Mizoram
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
frequency of use
black_mustard_seed_oil
onion
green_bell_pepper
garlic
turmeric
black_pepper
beef
bean
pork
ginger
rice
chicken
-1.5
-1
-0.5
0
0.5
1
1.5
2
contribution

(f)


Nagaland
0.6
0.5
0.4
0.3
0.2
0.1
0
frequency of use
cayenne
garlic
green_bell_pepper
ginger
tomato
onion
pork
fermented_soybean
chicken
cabbage
black_mustard_seed_oil
beef
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
contribution



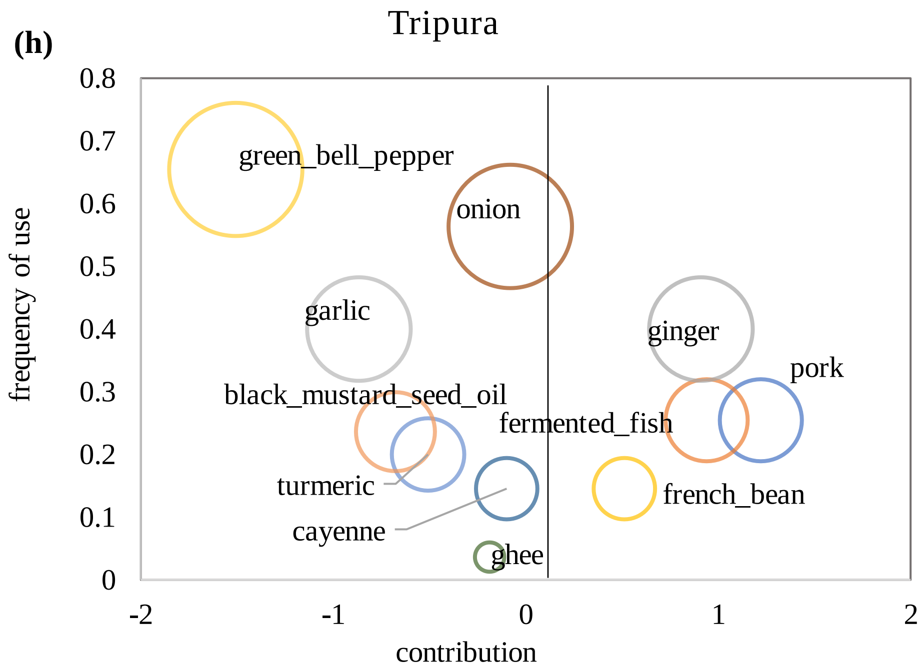


Fig. 4.5 Ingredient contribution: flavour pairing pattern vs. frequency of use in the regional cuisine (a) Assam (b) Arunachal (c) Manipur (d) Meghalaya (e) Mizoram (f) Nagaland (g) Sikkim (h) Tripura

Further, to understand the mechanism underlying the type of food pairing behaviour and shared compound hypothesis we studied the role of each ingredient in the regional cuisines if any alteration in the ingredients is to be made. The change in the average food paring index $\overline{N}_s$, increases when the least contributing ingredients (the most negative contributing ingredients to food pairing) were removed shown in Fig. A1. As a result, we can conclude that the contribution of the ingredient is negative. Additionally, we observed that when we removed those ingredients which significantly contribute to the shared compound effect ΔN_s, the original pattern drastically changes even after the removal of one ingredient, changing the food pairing pattern shown in Fig. A2. Overall, we observed that, the measure of the average food pairing index (N_S) and the shared compound effect (ΔNs) increased significantly even after the removal of one ingredient. Black mustard seed oil, green bell pepper, cayenne, onion, garlic, turmeric, bay laurel were found to contribute significantly towards the negative food pairing while rice and milk contribute towards the positive pairing. Additionally, spice, both as an individual and as a category, plays an important role in determining the food pairing behaviour. A previous study on eight Indian regional cuisines reported similar findings [39].

Table-A8 presents the top five ingredients contributing to negative and positive pairing behaviour along with their prevalence values. It reveals that among the prevalent ingredients, black mustard seed oil, onion, garlic, turmeric, green bell pepper and cayenne are consistently contribution to the negative pairing behaviour. Here the ingredient black mustard seed oil is the most prevalent ingredient with only one flavouring compound. Hence its contribution for negative pairing is as expected. Other five ingredients are from the spices category and are recognized to be contributor for negative pairing.

On the other end, as expected the dairy category product milk gets enlisted as one of the ingredients contributing to the positive pairing. Also, the meat category ingredients appear as. the ingredients contributing to the positive pairing. Rice, tea, and coconut are also some of the ingredients contributing to the positive pairing. In-spite of ginger being spice it was found to contribute to positive pairing behaviour.

4.1.6 Summary on food pairing behaviour

A rank frequency plot reveals the bias for few ingredients as favoured ingredients. Based on the frequency of their use, the prevalent ingredients were identified as black mustard seed oil, onion, cayenne, ginger, green bell pepper, garlic, turmeric, bay laurel, pork, rice and they also feature as the ingredients in the authentic pairs and triplets of ingredients from the cuisines. However, these pairs and triplets share very few favour compounds, even none in many instances. This leads to the possibility of a negative pairing behaviour of the cuisines. Cuisine level flavour pairing behaviour against randomly created cuisines validates the negative pairing behaviours. The prevalent ingredients are analysed for their contribution towards negative and positive pairing behaviours.

The study revealed that the ingredients from the spice category were the most common ingredients that made a significant impact on the food pairing behaviour in the food recipes from the Northeast regional cuisines. Ingredients such as black mustard seed oil, green bell pepper, onion, cayenne, ginger, garlic, rice, turmeric, bay laurel, tomato and pork were listed as the most authentic ingredients across the regional cuisines. In terms of pairing behaviour, the difference comes down to the question of how closely the flavour compounds are shared, i.e., either relatively low or high (negative and positive). Among the ingredients, cayenne, garlic, turmeric, and bay laurel contribute to the negative pairing. At the same time, ingredients such as ginger, tomato, and pork contribute to the positive pairing. Pork meat is identified as one of the most authentic ingredients in the Northeast regional cuisine which is not the case for other Indian regional cuisines. This, along with the wider use of black mustard seed oil are some of the distinct characteristics of the Northeast regional cuisine. The frequency of ingredient usage showed a generic culinary pattern conforming to the unique taste palate of the region.

We observed that the ingredient usage pattern is almost similar across the Northeast regional cuisine. Zhu et al. [94] in their study has reported that geographic distance promotes the use of similar ingredients. As a result, the geographic distance may be a factor in the similarity of the regional cuisine because the Northeast states are so close together. In summary, our work presents a scientific validation of the existing trend of ingredient combination which forms a food recipe for a cuisine used in a limited geographical location.

4.2 Application of data-driven analysis for intra- and inter-cuisine similarities

For studying the similarities, the uniqueness of the ingredients used is investigated first. While the prevalent ingredients are used to characterize the cuisines, based on their frequency of use in recipes, it does not imply any exclusivity. A data -driven clustering analysis is applied to find whether these ingredients stand out from cuisines from other regions. t-SNE clustering was applied both at ingredient level and at the flavour-compound level. Further the technique of visualization by graph networks of flavour compounds is used to analyse the similarity at cuisine level as well as at recipe levels. Finally, the inter-cuisine and intra-cuisine level similarity of recipes are analysed by the estimation of cosine similarity values.

4.2.1 t-SNE clustering

t-SNE clustering of the Northeast regional cuisine was carried out to visualize multi-dimensional data and to examine whether there is any overlap in the choice of ingredients and the flavour profile within the regional cuisines. Fig. 4.6 shows the t-SNE clustering of the Northeast regional cuisines, we can observe significant overlap both in ingredient and the flavour space, geographic proximity can be the reason which increases the usage of similar ingredients. We further compared Indian cuisines and cuisines from other countries. Fig. 4.7 shows the t-SNE clustering of Northeast regional cuisines with Indian cuisine, we observed that in the ingredient space the Northeast ingredient formed a significant cluster, as a result, we do not observe overlap with the other regional cuisines. However, in case of flavour profile though distinct clusters are formed we can observe some overlap as compared to the ingredient profile. Fig. 4.8 shows the t-SNE clustering of Northeast cuisine with cuisines from other countries (*American, Italian and Chinese)* we observed that distinct clusters are formed both in ingredient space and flavour space. However, we can observe clusters overlap in flavour profile with the other regional cuisines which are not observed in case of ingredient profile. This highlights the possibility of using an ingredient which may be an alternative to an unavailable ingredient in a recipe to attain a similar flavour profile. E.g., ginger can be substituted with cinnamon, mace, and nutmeg which is akin to that of ginger considering the similar flavour profile.

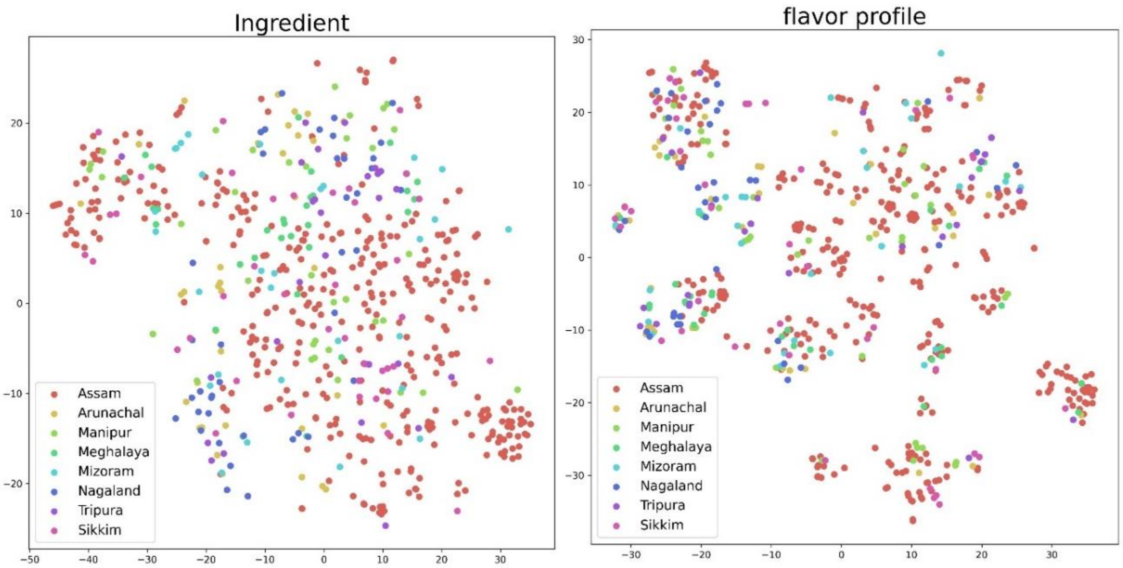


Fig. 4.6 t-SNE clustering of Northeast regional cuisine in ingredient space and flavour space

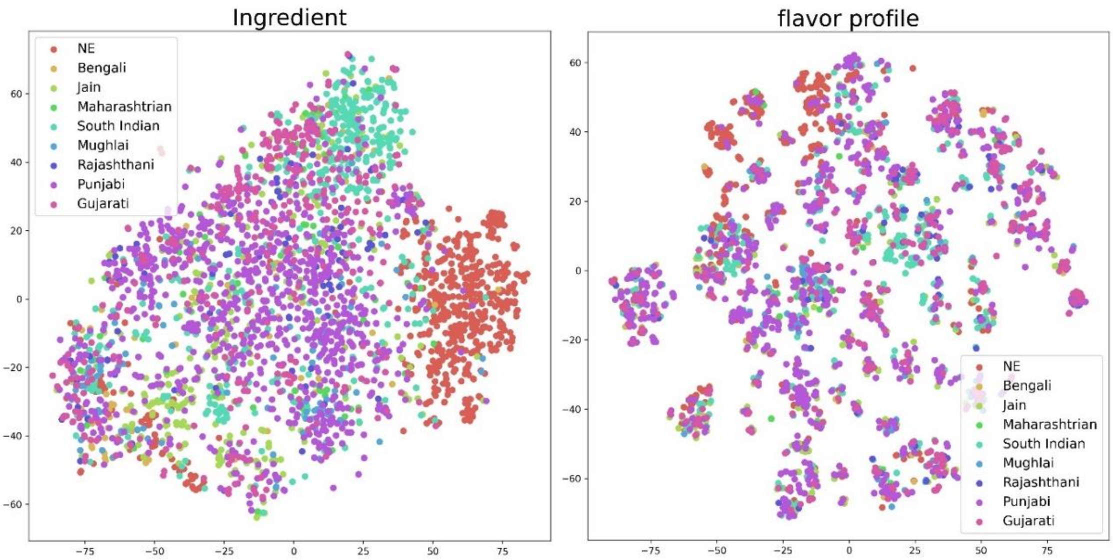


Fig.4.7 t-SNE clustering of Northeast regional cuisine and Indian cuisine in ingredient space and flavour space

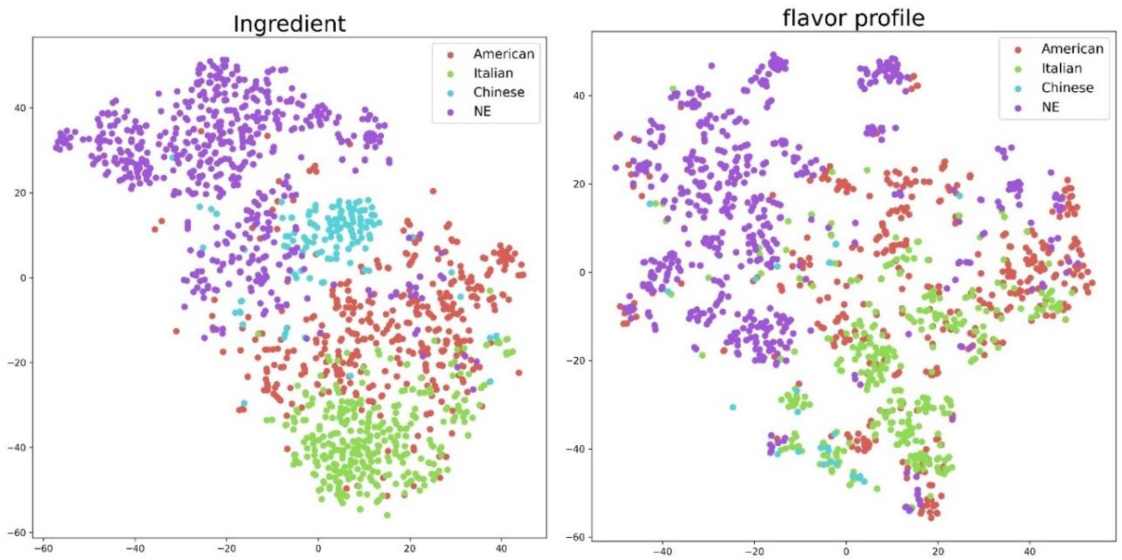


Fig. 4.8 t-SNE clustering of Northeast regional cuisine and other countries in ingredient space and flavour space

4.2.2 Flavour network-based visualization of similarities

4.2.2.1 Ingredient- compound bipartite network

The ingredient-compound bipartite network was constructed for the 126 ingredients used in the Northeast recipes (refer to Table 4.2) and 1149 flavour compounds that are known to contribute to the flavour of each of these ingredients. The ingredient-compound bipartite network determines the relationship between the ingredients and the flavour compounds. It helps us to identify the number of shared flavour compounds between the ingredients. The ingredient-compound bipartite network consists of two nodes which are the food ingredients and flavour compounds and the link between the nodes signifies the relationship in terms of shared flavour compound, Fig. 4.9. Flavour compounds are linked to the ingredients that contain them, forming a bipartite network. Some flavour compounds are shared by more than one ingredient. Further, an ingredient-compound bipartite network can be projected into the ingredient space which results in the formation of a flavour network, where ingredients are connected if they share at least one flavour compound.

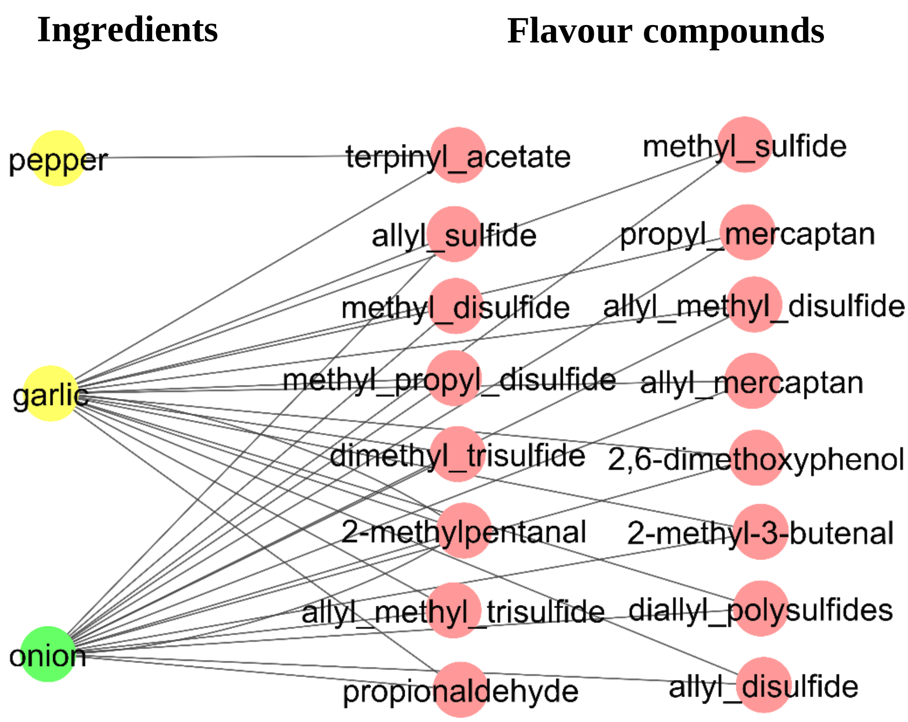


Fig. 4.9 The ingredient-compound bipartite network of an authentic ingredient triplet

4.2.2.2 Ingredient flavour network

The flavour network is the projection of the bipartite network in the ingredient space where ingredients are linked if they share at least one flavour compound. However, it is hard to directly visualize the flavour network since several flavour compounds are shared by several ingredients. To circumvent the density of the network for clear visualization, backbone extraction was carried out to avoid the network's density and for clear visualisation. The flavour graph backbone of 126 ingredients of the Northeast regional cuisines is shown in Fig. 4.10 in order of their prevalence in the recipes. Each node represents an ingredient, and an edge represents a shared flavour compound. sizes of the nodes are scaled based on the frequency of use of ingredients, while edges are scaled based on the number of flavour compounds shared.

The flavour network is the result of the backbone extraction (*p*-value 0.07) whereby we retained only the significant edges/link. However, for the overall analysis, the entire network is taken into account. The flavour network shows the relationship between different food categories based on the flavour compounds they share, which has resulted in estimating those food categories that are closely related. It further enables us to understand whether we choose those ingredients that share a significant link or we avoid them.

In the flavour network graph, Fig. 4.11 (a) size of each node represents the prevalence of the ingredients and the width of the edges represents the number of shared flavour compounds between the ingredients. From the flavour network, we could observe that the ingredients from the spice (yellow) are close to vegetables (green) and fruits (pink), while dairy (white) ingredients are close to meat (red), fruits (pink) and cereals (grey). Additionally, we could observe that the ingredients from the spice categories are isolated as they share flavour compounds only with other spice ingredients. The link between the ingredients of spice and vegetables is thinner as compared to dairy, meat and other categories. In order of their prevalence (size of the node) we observe that most of the ingredients are from the vegetables and spice category as they have larger node sizes.

In the flavour network graph, Fig. 4.11 (b) size of each node represents the number of flavour compounds of the ingredients, and the width of the edges represents the shared flavour compounds between the ingredients. We could observe a clear difference between the two figures in terms of the difference in the number of flavour compounds. The ingredients with the most flavour compounds are mostly from the category of dairy, meat and nut/seed. We have observed from Fig 4.11 (a) that the most prevalent ingredients are mostly from the category of spice as they have larger node sizes when compared to other category ingredients. As a result, we can conclude that spice category ingredients though it is more prevalent it has fewer flavour compounds. This could be the reason behind the thin edges between the ingredients of the spice category. Overall, the flavour network of the Northeast regional cuisines showed that the most prevalent ingredient is mostly from the spice and vegetable categories.

Further, we constructed a flavour graph of each ingredient category as we could not represent all the ingredients in the flavour network graph. The flavour graph is constructed to examine the nature of flavour sharing within the ingredients of each category, Fig. A3. We can observe that the link/edges between ingredients from the spice categories though they have more ingredients are not as significant as compared to other categories such as dairy, cereal/crop and meat. This could be the reason behind the negative food pairing behaviour which has been observed across all the Northeast regional cuisines.

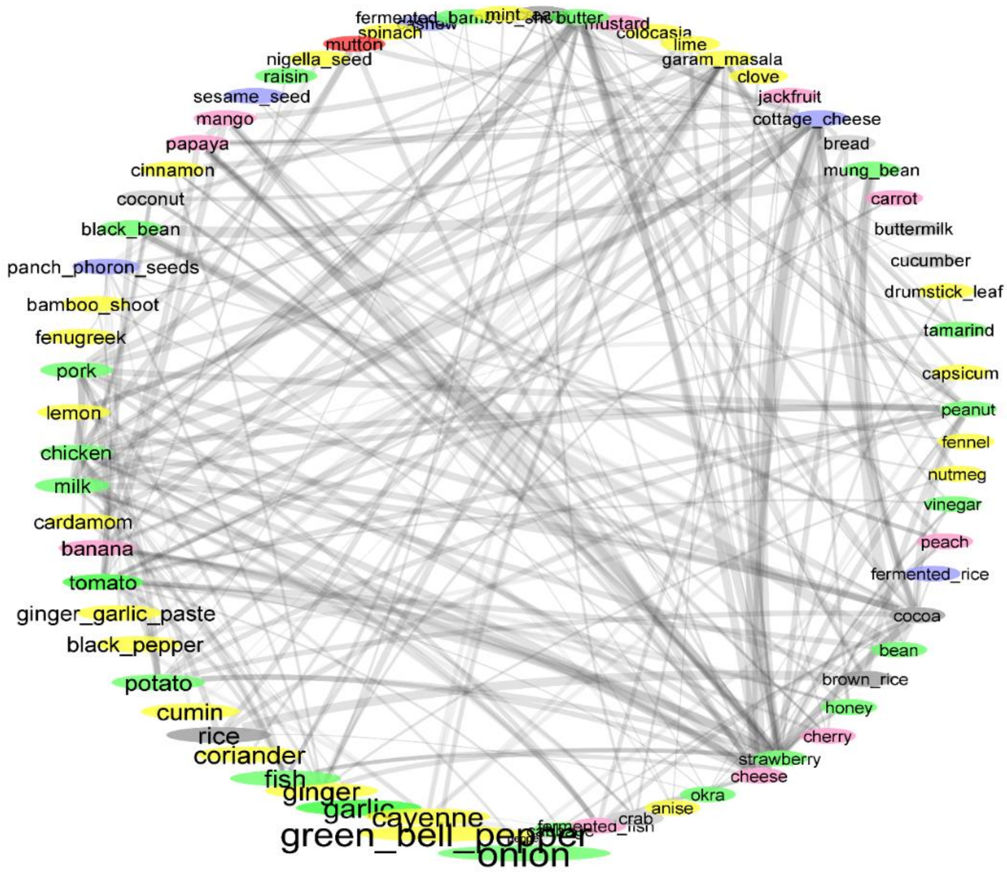


Fig. 4.10 The flavour graph backbone of 126 ingredients of Northeast cuisine. A node represents each of the 126 ingredients, and an edge represents a shared flavour compound. sizes of the nodes are scaled based on the frequency of use of ingredients, while edges are scaled based on the amount of flavour compounds shared.

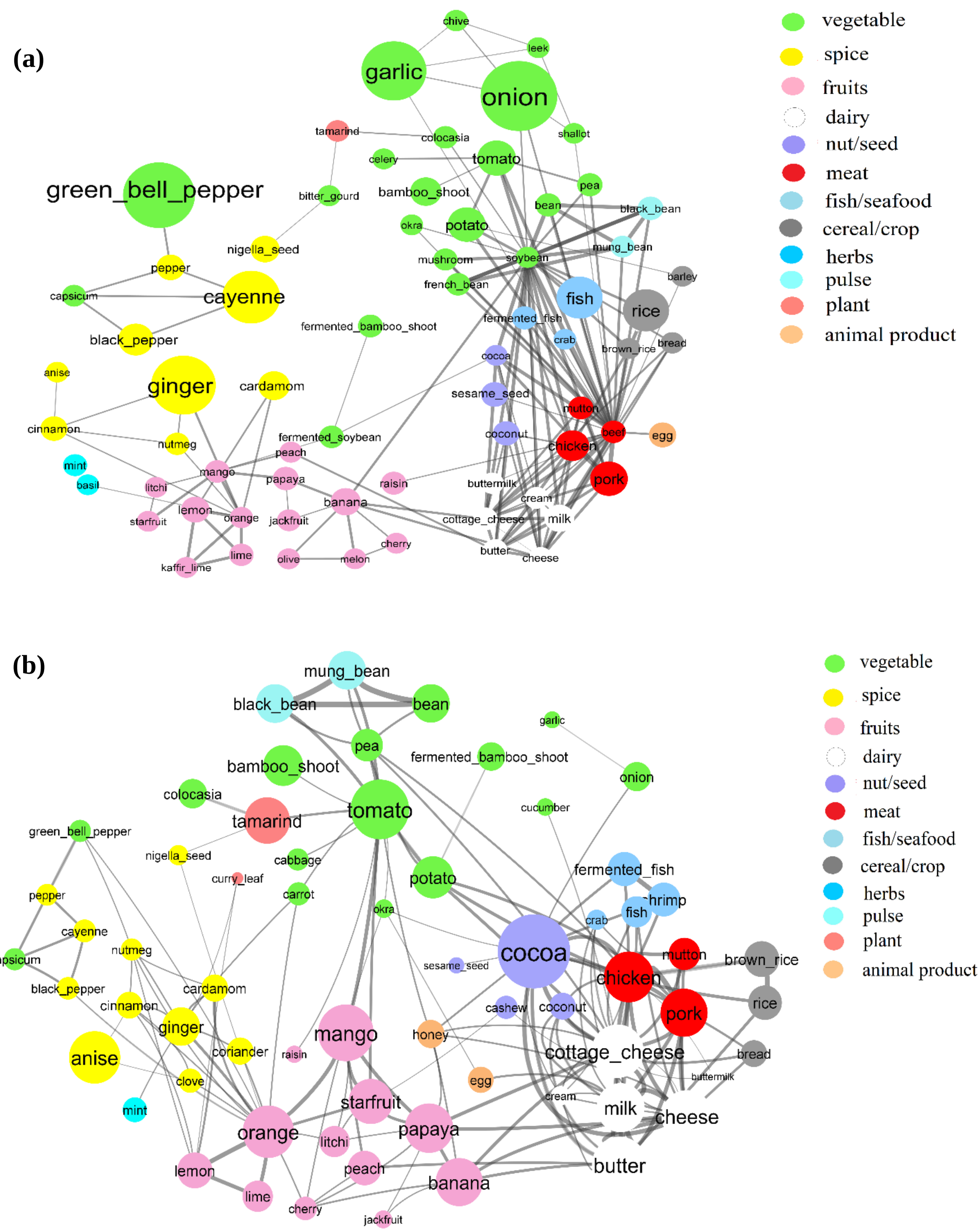


Fig. 4.11 The Flavour network: (a) Size of each node represents the prevalence of the ingredients, the width of edges represents the number of shared flavour compounds between the ingredients and the colour of the node represents the food category (b) Size of each node represents the number of flavour compounds of the ingredients, the width of edges represents the shared flavour compounds between the ingredients and the colour of the node represents the food category

4.2.3 Cosine similarities for recipes of the regional cuisines

Regional cuisines vary from one region to another, and the difference in food choice is the result of the differences in flavour preferences. Differences in the geographical environment greatly affect people's eating habits and the choice in flavour preferences. However, there are flavour similarities in regional cuisines that are geographically adjacent to each other. The dietary preferences of users are influenced by many factors, such as hereditary, geographical environment, cultural environment, current health needs, dietary balance and some other factors [94]. Recipe recommendation for restaurants needs proper planning to launch dishes to meet the preference of customers [50]. The results obtained from the cosine similarity can be used as a dish recommender system to find similar dishes across various regional cuisines which is based on flavour similarity.

4.2.3.1 Cosine similarity analysis within Northeast regional cuisine

To determine the degree of similarity and dissimilarity among the recipes within Northeast regional cuisines, a cosine similarity analysis was conducted, as per section 3.3.2.2. A detailed analysis of every recipe in each regional cuisine was conducted based on the ingredients used. We estimated the similarity percentage of each regional cuisine to determine whether they are more similar or more less similar, the total recipe considered was 640 (recipe size > 2). Our analysis revealed the Tripura recipe as the least similar followed by Mizoram (Table 4.6). This result reflects the uniqueness of regional cuisine in terms of the ingredients used in their recipes. It is however found that Assamese recipes share many similarities with other regional cuisines. Additionally, we have listed some of the top ten recipes with cosine similarity values close to 1.0 in Table-A11.

Table 4.6 Similarity percentage of recipes across regional cuisine

Regional cuisine	Similarity percentage (%)
Assam	62.21
Sikkim	6.65
Meghalaya	6.40
Nagaland	6.12
Arunachal	5.06
Manipur	5
Mizoram	4.78
Tripura	3.75

It was observed that recipes from the same region have the highest degree of similarity. The recipes that are found to be similar have a difference of at least one ingredient or none at all as shown in Table A11. The difference can be observed in case if an ingredient gets replaced with another or addition of new ingredients in the recipe set. We observed in the case of a recipe where the main ingredient belongs to the meat or fish category, the replacement of ingredients happens within the ingredients from the same category itself, one such example is *ari fish fry* (ari fish, black mustard seed oil, turmeric, bay laurel) and *chital fish fry* (chital fish, black mustard seed oil, turmeric, bay laurel) the only difference in these two recipes is the type of fish where ari fish is replaced with chital fish rest of the ingredients remains the same. The replacement of ingredients is within the category of fish. Another example is *duck with potato* (duck, onion, turmeric, green bell pepper, black mustard seed oil, ginger garlic paste, black pepper, cumin, bay laurel, potato), and *pigeon with potato* (pigeon meat, onion, turmeric, green bell pepper, black mustard seed oil, ginger garlic paste, black pepper, cumin, bay laurel, potato) where the only difference in the recipe set is the type of meat. Further, Table-A12 shows the list of top ten recipes with cosine similarity value ranging from 0.9 to 0.8. We observed that more than two ingredients are either replaced or added, as a result, the similarity index is lower than those compared to recipes with cosine similarity value 1.

4.2.3.2 Cosine similarity analysis with Indian regional cuisine

We carried out a comparative study of Northeast cuisine with the other Indian regional cuisines to determine the extent of similarities and dissimilarities between the regional cuisines. The recipe dataset of the regional cuisine of India was obtained from the archive data of Jain et al. [39]. The dataset consists of eight regional cuisines of India viz Bengali, Gujarati, Jain, Maharashtrian, Mughlai, Punjabi, Rajasthani and South India. The total number of recipes considered for the other Indian regional cuisine was 2916 recipes to be compared with 640 recipes of the Northeast regional cuisine. We analysed the similarity of Northeast regional cuisine across the Indian regional cuisines in terms of their ingredient usage. The number of similar recipes among regional cuisines is shown in Table 4.7.

In terms of similar recipes, Assam regional cuisine shares the most affinities with Indian regional cuisine with a total of 519 similar recipes out of 2916 recipe of the other Indian regional cuisine. Possibly, it is because there are more recipes for Assamese cuisine as compared to other regional cuisines, which have fewer recipe data

Additionally, Tripura recipes were found to be least similar as only 4 recipes are found to be similar. In terms of similarities, we can observe that the northeast regional cuisines do not show much similarities with the Indian cuisines. Consequently, the Northeast regional cuisines have a distinct advantage in terms of their uniqueness.

Table 4.7 Number of similar recipes from the Northeast and other Indian regional cuisines

Northeast regional cuisines	Number of similar Indian recipes
Assam	519
Arunachal	35
Manipur	41
Meghalaya	15
Mizoram	17
Nagaland	24
Sikkim	60
Tripura	4

4.2.3.3 Cosine similarity analysis with western countries and east Asian countries

We conducted similarity analysis of the Northeast regional cuisine with a positively paired regional cuisine and a negatively paired regional cuisine to determine the degree of similarity and dissimilarity. Considering the criteria, we selected western cuisines and east Asian cuisines as studies have been reported that western cuisine exhibits positive food pairing pattern while east Asian cuisine exhibit negative food pairing pattern [5]. The number of similar recipes is listed in Table 4.8.

Table 4.8 Number of similar recipes of the Northeast regional recipes compared to western and east Asian

Northeast regional cuisine	No of similar east Asian recipes	No of similar western recipes
Assam	149	78
Arunachal	32	18
Manipur	8	4
Meghalaya	35	8
Mizoram	42	13
Nagaland	27	8
Sikkim	26	18
Tripura	26	6

The recipe dataset of the western and east Asian countries was obtained from the archive data of Ahn et al. [5]. In terms of ingredient usage, the recipe dataset served as a tool to compare Northeast regional cuisine with western cuisine with total recipe 995 and east Asian with total recipe 951. The list of top five recipes which are found to be similar to the western and east Asian cuisine are listed in Table-A12. The regional cuisine with highest number of similar recipes both in the case of western and east Asian cuisine was found to with Assam and the least similar was with Manipur. Compared to Northeast and Indian cuisines, the number of similar recipes in case of western and east Asian is relatively low. Additionally, east Asian cuisine are found to be more similar than the western cuisine as they have more similar recipe.

Our analysis revealed that the cosine similarity values of the similar recipes were comparatively low in western cuisine as compared to east Asian cuisine. Additionally, they are found to be similar to a limited number of recipes, not more than ten. Most similar recipes have ingredients derived from the dairy category, as we can see from the list of similar recipes. The same holds true for western cuisines, where dairy-based ingredients contribute to the positive food pairing behaviour of western cuisine. However, in the case of east Asian cuisine we can observe that ingredient such as cayenne, ginger, and rice are found to be listed among the similar ingredients. In conclusion, we did not observe much significant similarities between the recipes of Northeast cuisines and western cuisines.

4.2.3.4 Flavour network plot for higher cosine similarity value at recipe level

For visualization, we plotted a flavour network to determine the relationship of ingredients concerning the number of flavour compounds they share. This was done to highlight the difference in the choice of ingredients and combination forming a recipe. As an example, we selected a few similar recipes with cosine similarity values close to 1. Each node represents the ingredient and the colour of the node represents the ingredient category. Nodes with a link represent that they share a flavour compound, and their thickness indicates their number of flavour compounds. The thicker the link, the more flavour compounds are shared.

As it pertains to Northeast recipes, we selected the recipe *chital fish fry* and *ari fish fry* (refer to Table-A11) to plot the flavour network between the ingredients. The two recipes only differed by one ingredient between them and they each had four ingredients. We observed that none of the ingredients shared flavour compounds as shown in Fig. 4.12 (a) as there is no link between the nodes. Similarly, we selected a recipe with more than one ingredient difference, *black gram with ou tenga* and *black gram curry*, the recipes differed by two ingredients green bell pepper and cayenne. Additionally, we observed that the number of shared flavour compounds in between the ingredients is similar even after replacement with the other ingredients in the recipe, Fig. 4.12 (b). We further selected the similar recipe of Northeast and western with higher cosine similarity values, see Table-A13. *Caramel pudding* with five ingredients and *western_92* recipe with nine ingredients were selected. The recipes shared three similar ingredients. We observed from the flavour network, Fig. 4.12 (c) the difference between the two recipes concerning the number of flavour compounds shared between the ingredients. As observed the ingredients between dairy products share a greater number of flavour compounds. This has been reported in earlier studies for western cuisines as they tend to use more dairy-based ingredients, which have a greater number of flavour compounds as compared to ingredients from other categories. We further constructed the flavour network graph of the similar Northeast recipe and east Asian recipe. We observed that the ingredients do not share as many flavour compounds between the ingredients as compared to western recipes, Fig. 4.12 (d).

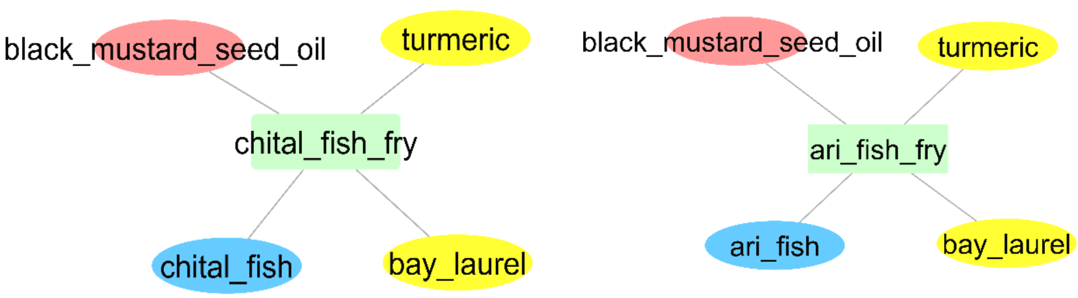


Fig. 4.12 (a) Flavour network plot of recipe chital fish fry and ari fish fry with a difference of one ingredient in between them

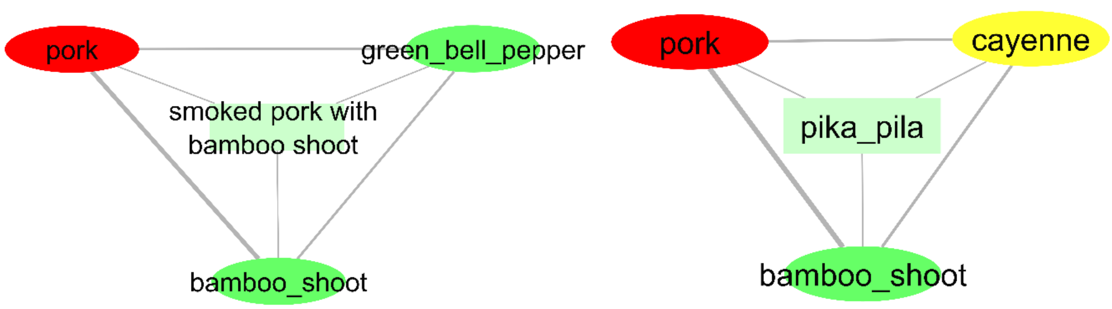


Fig. 4.12 (b) Flavour network plot of recipe black gram with ou tenga and black gram curry with a difference of one ingredient in between them

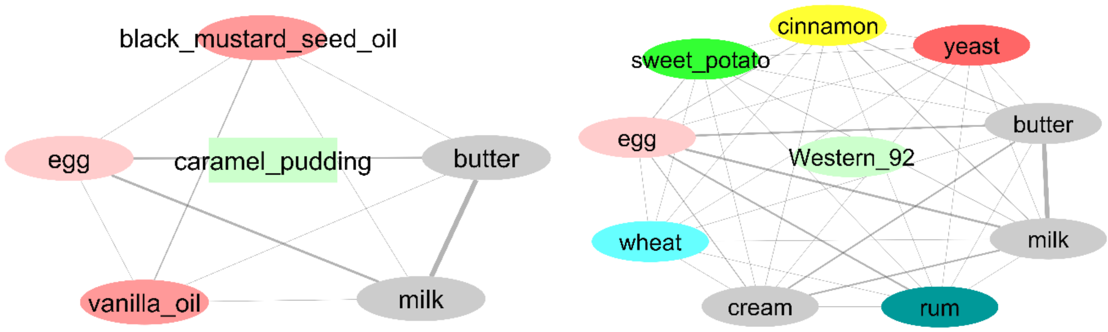


Fig. 4.12 (c) Flavour network plot of caramel pudding (Northeast recipe) and western_92 (western recipe)

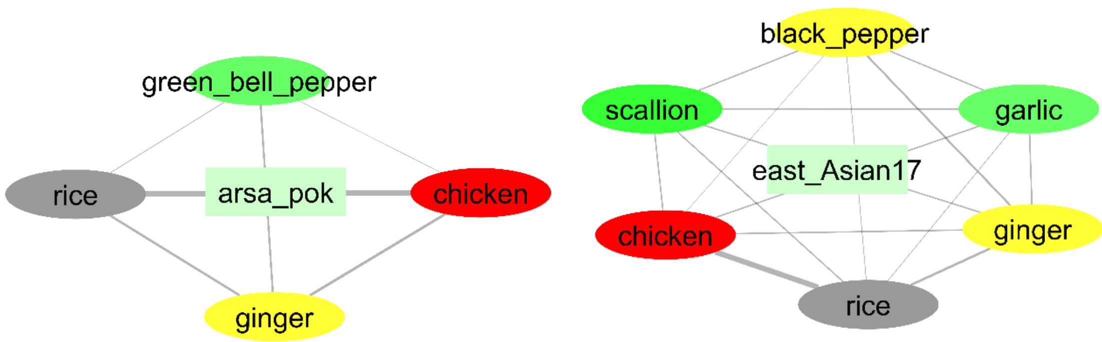


Fig. 4.12 (d) Flavour network plot of arsa pok (Northeast recipe) with east Asian17 (east Asian recipe)

4.2.4 Summary of findings from similarity analysis

t-SNE clustering of the Northeast regional cuisine showed distinct differences as compared to the other regional cuisines. As a result, Northeast regional cuisines differ significantly in both their choice of ingredients and their flavour when compared with regional cuisines of other geographic regions. However, we can observe that the distinct clusters are formed mostly in the case of the ingredient profile but not in the flavour profile. The similarity in flavour profile further highlights the possibility of using the ingredient as a substitute in situations where ingredients are not readily available for a recipe or product development. E.g., ginger can be substituted with cinnamon, mace, and nutmeg which is akin to that of ginger considering the similar flavour profile. The flavour network-based analysis of food pairing applied to the sub-cuisines from the Northeast regional cuisines showed that the link/edges between ingredients from the spice categories are not as significant as compared to the other categories such as dairy, cereal/crop/ and meat. This can be the reason behind the negative food pairing behaviour which has been observed in the regional cuisines. Based on the comparative study done using cosine similarity we can conclude that the proposed algorithm can be used as an effective tool in determining the similarity across various regional cuisines

4.3 Generation of recipe composition based on identified consumer preference for ingredient pairing

4.3.1 Recipe generation based on flavour network of pre-existing recipe data

4.3.1.1 Theory

Food pairing theory states that food that shares a higher number of flavour compounds tastes well together. However, this theory stands applicable only in the case of western cuisines and not in the Asian and Indian cuisines in particular. As our study has found out that Northeast regional cuisines show strong negative food pairing behaviour (section 4.1.4.2) and with the help of flavour network analysis (section 4.2.2.2), we created a new food recipe by evaluating the ingredients to be selected based on the number of shared flavour compounds.

4.3.1.2 Approach

To create a new food recipe in the Northeast region, flavour network analysis was used to evaluate the ingredient pairs based on the number of compounds they have in common. The selection of food ingredients was based on the outcome of the flavour network of the recipe selected. To provide a diversity of tastes, we selected eight recipes from different meal groups such as main dishes, side dishes, salads, and desserts consumed in Northeast region. Ultimately, a matrix chart was created based on common flavour compounds among the ingredients. Among the listed ingredients, the ones that shared the least flavour compounds were used to develop the new dish.

A total of thirty-six ingredients from the eight recipes were scanned in order to achieve the goal. The top ten ingredients that share the least number of flavour compounds are listed in Table 4.9 in ascending order. Out of the thirty-six ingredients we selected the top 7 ingredients to be used for the new recipe, since the average recipe size of Northeast was found to be seven (section 4.1.2.1). The ingredients selected were black mustard seed oil, bay laurel, turmeric, garlic, cumin, cashew, and green bell pepper.

Table 4.9 Average shared compounds of ingredients

Rank	Ingredients	Average shared compounds
1.	black mustard seed oil	0.058
2.	bay laurel	0.235
3.	turmeric	1.235
4.	garlic	1.588
5.	cumin	2.235
6.	cashew	3.323
7.	green bell pepper	3.411
8.	bamboo shoot	3.911
9.	nigella seed	3.941
10.	cream	4.382

4.3.1.3 New recipe preparation

We design a new recipe considering the ingredients with the least shared flavour compounds in preparation for a dish. We considered a non-veg preparation such as mock chicken or analogue meat assuming that the mixture of the ingredients should taste well when prepared together. The mock chicken was purchased online and two criteria have been set for the preparation.

a) We prepare the dish as instructed in the labels given in the packet using the same spices as included. The spices mix consists of thirteen ingredients as listed in Table 4.10, spices (a).
b) We prepare the dish following the same procedure as instructed but we used the spices mix generated from the flavour network. We considered the seven least flavour-sharing ingredients as the average recipe size is seven for the northeast cuisine (section 4.3.1.2). The list of the spices mix is shown in Table 4.10, spices (b).

Table 4.10 List of spices for new recipe

Commercially purchased mock chicken

List of spices (a)	List of spices (b)
Cashew nuts	Mustard seed
Red chillies	Bay laurel
Turmeric	Turmeric
Fenugreek	Garlic
Black pepper	Cumin
Cumin seeds	Green chilli
Coriander	Cashew
Black cardamom	
Cinnamon	
Mace	
Long pepper	
Ginger	
garlic	

4.3.1.4 Sensory analysis of new recipe

The recipe developed was named Recipe A and Recipe B. Recipe A was prepared with the list of spices (a) and Recipe B was prepared with the list of spices (b) (Table 4.10). spices (b) had an additional eleven unique ingredients such as red chillies, fenugreek, black pepper, cumin seeds, coriander, black cardamom, cinnamon, mace, long pepper and ginger.

The hedonic rating test sheet for sensory analysis is given in appendix Table-A6. Both recipes were served to 15 volunteer students in the department with a mixture of students from the Northeast and other parts of India. A nine-point hedonic scale of the sensory evaluation was used to measure the consumer preference for the new dish [57]. Each student was asked to fill out a questionnaire that included questions about their taste preferences, and if their local tastes were similar. The nine-point hedonic score of the sensory analysis is shown in Fig. 4.13.

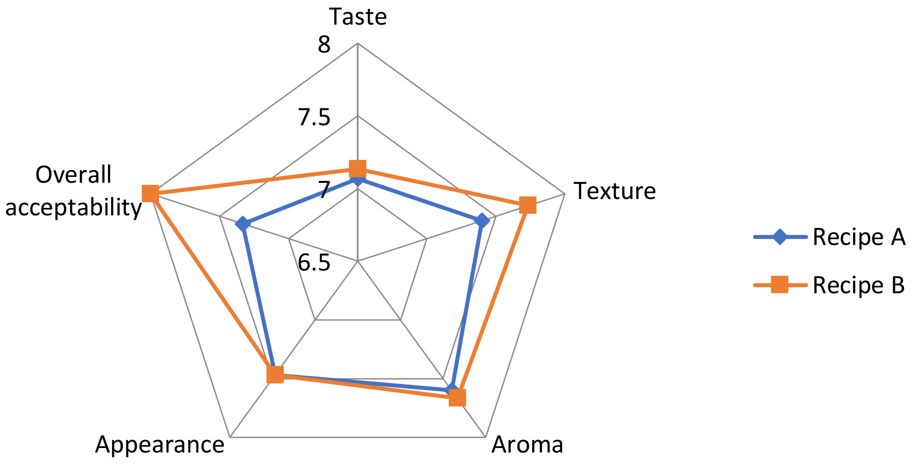


Fig.4.13 Nine-point hedonic score of the new recipe sensory analysis

The result obtained from sensory study showed that there is not much difference in the taste, texture, aroma and appearance in both the recipe. However, the overall acceptability was higher for recipe B which was prepared using the spice mix generated from the flavour network. There is a possibility that the high acceptability of recipe B was due to the preference for the taste with less spices among the Northeast region panellists.

We observed that the panel from the northeast state preferred the meal prepared using the spices mix (b) of flavour network which is of less spices mix while the rest preferred the meal prepared out of default spices mix (a). Through this, we can conclude that Northeast cuisine do not use much spices as compared to the other regional cuisines of India, as a result, the preference would be biased towards the less mixture of spices.

4.3.1.5 Similarity analysis of new recipe using cosine similarity

To determine the similarity of recipes across regional cuisines, Recipe A and Recipe B (derived from the flavour network of Northeast recipe) were compared to pre-existing recipe data (Northeast and other Indian regional cuisine). We observed that Recipe A is similar to mostly recipes of Punjabi, Rajasthani and South Indian while Recipe B was found to be similar with mostly Assam recipes which is a part of Northeast regional cuisine, Fig. 4.14. As a result, we can conclude that analysis based on regional food data is useful in developing alternative dishes for people who are accustomed to consume similar ingredients.

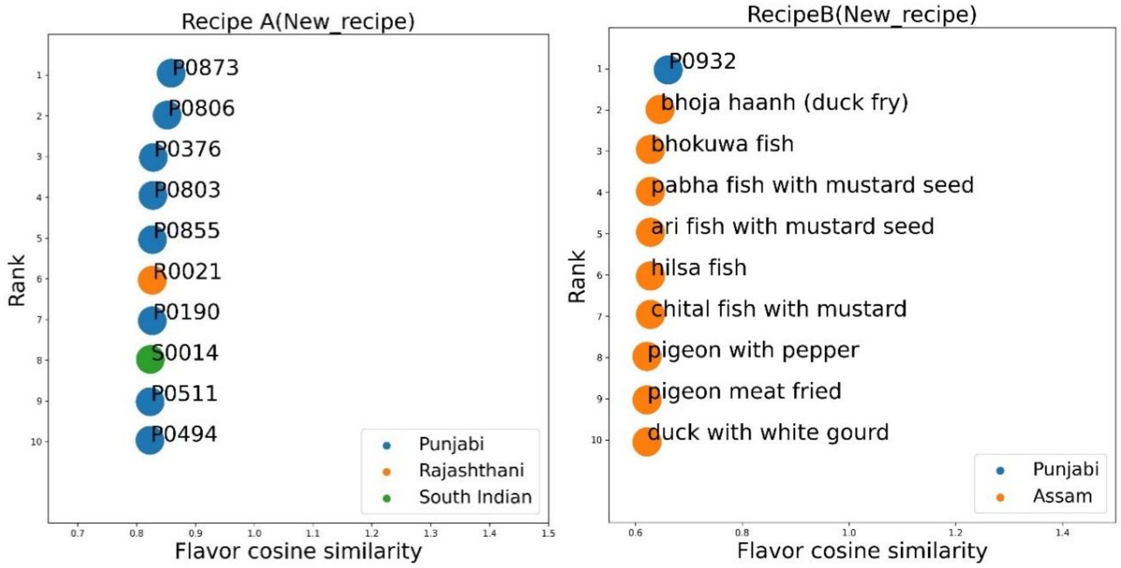


Fig. 4.14 Cosine similarity of new recipe and pre-existing recipe

4.3.2 Recipe generation from flavour network of khichdi recipe

We collected 27 khichdi recipes each from 27 different states of India which were listed in the book "Khichdi: simple soulful and soothing" [15]. We constructed a flavour network of 57 ingredients collected from the khichdi recipe. The most prevalent ingredient was found to be mostly from the categories of pulse and spices as listed in Table 4.11. As reported in earlier studies on Indian regional cuisines the presence of spices in the recipe contributes to negative food pairing behaviour. A similar trend can be observed as depicted in the flavour network of the khichdi recipe as the edges formed between the ingredients are not very significant, shown in Fig. 4.15.

Table 4.11 Top 10 most prevalent ingredients in khichdi recipe

Rank	Most prevalent ingredients	Prevalence
1	Rice	0.92
2	Cumin	0.57
3	Green bell pepper	0.50
4	Ginger	0.50
5	Mung bean	0.46
6	Onion	0.46
7	Cayenne	0.42
8	Coriander	0.38
9	Garlic	0.34
10	Bay laurel	0.34

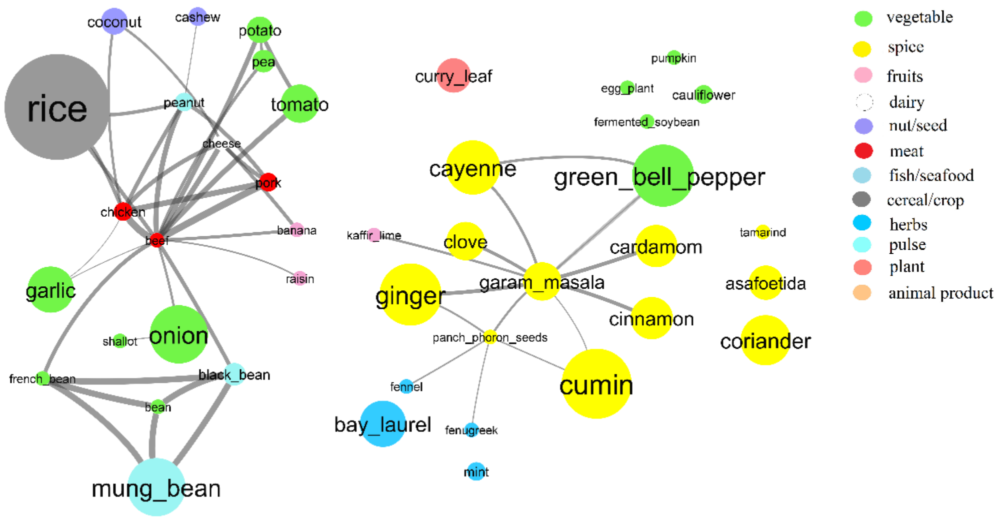


Fig. 4.15 Flavour network of Khichdi recipe; the size of the node represents the prevalence of the ingredients and the size/thickness of the link between the ingredients represents the number of shared flavour compounds, colour of the nodes represents the ingredients categories

we created a new recipe set from the outcome of the flavour network considering the most prevalent ingredients along with the addition of authentic ingredients from the regional cuisines of the Northeast shown in Table 4.12. Further, we compared the cosine similarities of the new recipe. We maintained the recipe size of the recipe as the average size of the original khichdi recipe which was found to be 10.

Table 4.12 List of authentic ingredients to be added to the set of prevalent ingredients of khichdi recipe for new recipe generation

Regional cuisine	Khichdi prevalent ingredient	Authentic ingredient
Assam	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, turmeric, bay laurel
Arunachal	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, tomato, garlic
Manipur	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, bay laurel, garlic
Meghalaya	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, sesame seed, garlic
Mizoram	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, turmeric, garlic
Nagaland	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	pork, bamboo shoot, garlic
Sikkim	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	black mustard seed oil, turmeric, tomato
Tripura	Rice, Cumin, Green bell pepper, Ginger, Mung bean, Onion, Cayenne +?	garlic, pork

4.3.2.1 New recipe validation using Cosine similarities

The similarities of the new recipe created were measured using cosine similarity. A comparison was done considering two criteria

1. Comparison with the original khichdi recipe with the most prevalent ingredients selected from the flavour network shown in Fig. 4.16.
2. Comparison with the original khichdi recipe with top seven prevalent ingredients selected from the flavour network in addition to three authentic ingredients from the Northeast regional cuisine shown in Fig. 4.17 (a-h).

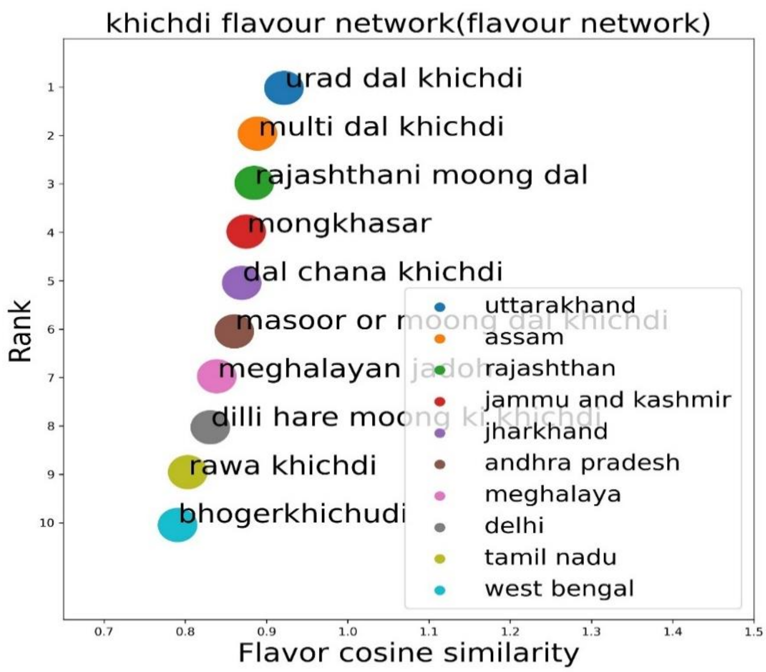
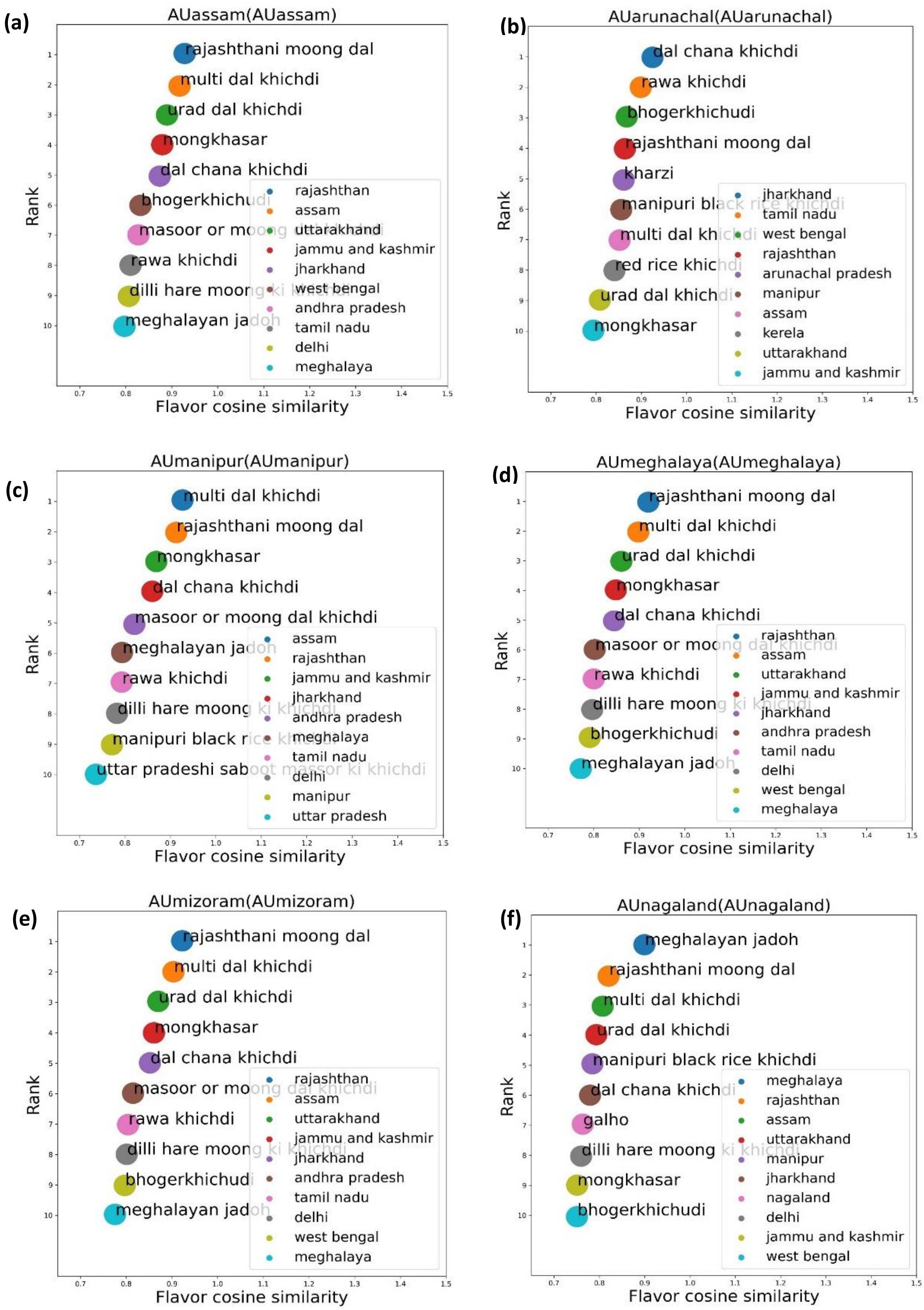


Fig. 4.16 Flavour cosine similarity of recipes across regional cuisines which is the result of flavour network


(a)
AUassam(AUassam)
rajashthani moong dal
multi dal khichdi
urad dal khichdi
mongkhasar
dal chana khichdi
bhogerkhichudi
masoor or moong dal khichdi
rawa khichdi
dilli hare moong ki khichdi
meghalayan jadoh
rajashthan
assam
uttarakhand
jammu and kashmir
jharkhand
west bengal
andhra pradesh
tamil nadu
delhi
meghalaya
Rank
Flavor cosine similarity
(b)
AUarunachal(AUarunachal)
dal chana khichdi
rawa khichdi
bhogerkhichudi
rajashthani moong dal
kharzi
manipuri black rice khichdi
multi dal khichdi
red rice khichdi
urad dal khichdi
mongkhasar
jharkhand
tamil nadu
west bengal
rajashthan
arunachal pradesh
manipur
assam
kerela
uttarakhand
jammu and kashmir
Rank
Flavor cosine similarity
(c)
AUmanipur(AUmanipur)
multi dal khichdi
rajashthani moong dal
mongkhasar
dal chana khichdi
masoor or moong dal khichdi
meghalayan jadoh
rawa khichdi
dilli hare moong ki khichdi
manipuri black rice khichdi
uttar pradeshi saboot masoor ki khichdi
assam
rajashthan
jammu and kashmir
jharkhand
andhra pradesh
meghalaya
tamil nadu
delhi
manipur
uttar pradesh
Rank
Flavor cosine similarity
(d)
AUmeghalaya(AUmeghalaya)
rajashthani moong dal
multi dal khichdi
urad dal khichdi
mongkhasar
dal chana khichdi
masoor or moong dal khichdi
rawa khichdi
dilli hare moong ki khichdi
bhogerkhichudi
meghalayan jadoh
rajashthan
assam
uttarakhand
jammu and kashmir
jharkhand
andhra pradesh
tamil nadu
delhi
west bengal
meghalaya
Rank
Flavor cosine similarity
(e)
AUmizoram(AUmizoram)
rajashthani moong dal
multi dal khichdi
urad dal khichdi
mongkhasar
dal chana khichdi
masoor or moong dal khichdi
rawa khichdi
dilli hare moong ki khichdi
bhogerkhichudi
meghalayan jadoh
rajashthan
assam
uttarakhand
jammu and kashmir
jharkhand
andhra pradesh
tamil nadu
delhi
west bengal
meghalaya
Rank
Flavor cosine similarity
(f)
AUnagaland(AUnagaland)
meghalayan jadoh
rajashthani moong dal
multi dal khichdi
urad dal khichdi
manipuri black rice khichdi
dal chana khichdi
galho
dilli hare moong ki khichdi
mongkhasar
bhogerkhichudi
meghalaya
rajashthan
assam
uttarakhand
manipur
jharkhand
nagaland
delhi
jammu and kashmir
west bengal
Rank
Flavor cosine similarity

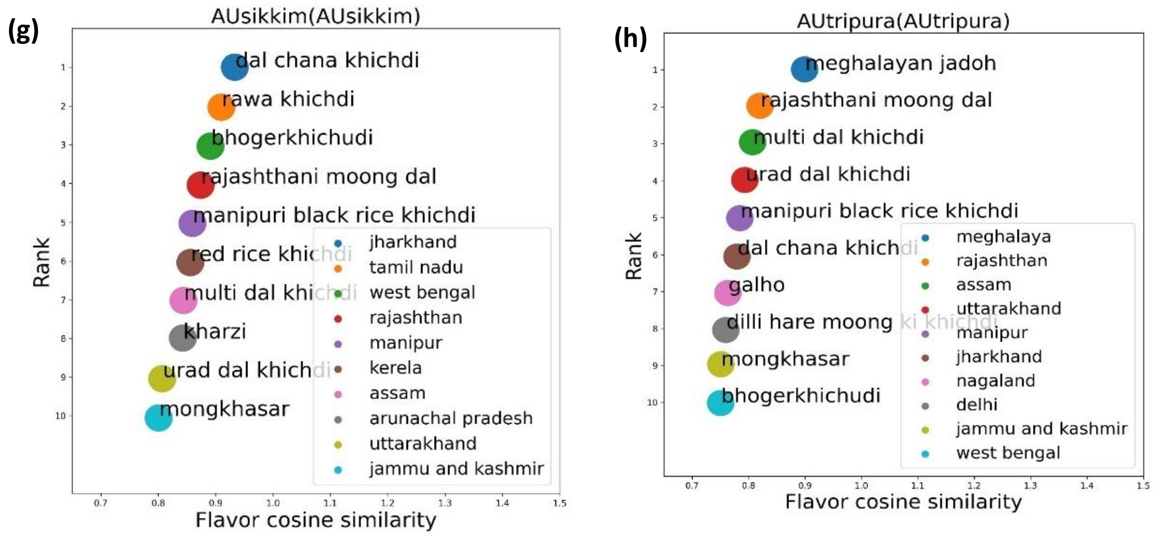


Fig. 4.17 (a-h) Flavour cosine similarity of recipes across regional cuisines with top seven prevalent ingredients selected from the flavour network, AU denotes authentic ingredients

4.3.3 Recipe generation based on data-driven models

Recipe and flavour data were analysed using canonical correlation analysis. The data-driven models suggest, for a given set of ingredients, those ingredients that can best be combined with all of the given ingredients. It helps us to identify ingredients which can be best paired with the desired set of ingredients where the ingredients generated are considered to be the best candidate to use for recipe completion purposes.

Table 4.13 Three best paired Ingredient of the regional cuisines

Region	Best paired ingredient	
Assam	[black mustard seed, garlic, green bell pepper]	[fermented bamboo shoot, ginger, rice]
Arunachal	[green bell pepper, ginger, garlic]	[cabbage, bean, carrot]
Manipur	[onion, ginger, garlic]	[onion, black mustard seed oil, cumin]
Meghalaya	[onion, pork, garlic]	[onion, garlic, sesame seed]
Mizoram	[black mustard seed oil, ginger, garlic]	[black mustard seed oil, onion, ginger]
Nagaland	[cayenne, garlic, ginger]	[garlic, ginger, green bell pepper]
Sikkim	[onion, black mustard seed oil, green bell pepper]	[onion, turmeric, tomato]
Tripura	[green bell pepper, onion, ginger]	[green bell pepper, onion, fermented fish]

We estimated three sets of the best-paired ingredient for each regional cuisine based on the co-occurrence and the shared flavour compounds. Using the RLS models we were further able to recommend ingredients which can be paired with the three sets of ingredients forming a recipe. The list of three best-paired ingredients for regional cuisine is listed in Table 4.13. further, a two-step RLS model is constructed using the three sets of ingredients.

Additionally, Table 4.14 shows the recommended ingredient sets and the most popular list of ingredients pairing recommendations.

Table 4.14 Pairing recommendation to complete the set of three best paired ingredients

Best paired ingredients	Pairing recommendation
green bell pepper + ginger + garlic +….	coriander
cabbage + bean + carrot +….	black bean, cauliflower, colocasia, cucumber, eggplant, mutton
onion + ginger + garlic +….	black bean, garcinia indica, ghee
onion + black mustard seed oil + cumin +….	anise, basil, black bean, cauliflower, chinese cabbage, french bean, garcinia indica, ghee, pork
onion + pork + garlic +….	basil, black bean, garcinia indica, ghee
onion + garlic + sesame seed +….	basil, black bean, black mustard seed oil, cauliflower
black mustard seed oil + ginger + garlic +….	egg, fish, ginger garlic paste, prawn
black mustard seed oil + onion + ginger +….	black mustard seed oil, colocasia, fermented soybean
cayenne + garlic + ginger +….	fermented fish
garlic + ginger + green bell pepper +….	cauliflower, colocasia, ginger garlic paste, okra
onion + black mustard seed oil + green bell pepper +….	cauliflower, ghee
onion + turmeric + tomato +….	ash gourd, black bean, cauliflower, fermented soybean, garcinia indica, ghee
green bell pepper + onion + ginger +….	ash gourd, beef, garcinia indica, ghee, ginger garlic paste, mung bean
green bell pepper + onion + fermented fish +….	black bean, garcinia indica, ghee, ginger garlic paste
black mustard seed + garlic + green bell pepper +….	cauliflower, fenugreek, garcinia indica, ghee, ginger garlic paste
fermented bamboo shoot + ginger + rice +….	ash gourd, bamboo shoot, black bean, cauliflower, fermented rice, garam masala, tomato

The recommended list of ingredients suggested appears to be acceptable as most of the recommended ingredients are from the same categories itself. For example, coriander, an ingredient from the category of spice was recommended to complete the set of ingredients green bell pepper (vegetable) + ginger (spice) + garlic (vegetable). Coriander has a total of 52 flavour compounds out of which the maximum number of shared compounds was found to be with ginger with 86 flavour compounds. They share a total of 24 flavour compounds which is the highest number of shared compounds in both ingredients when compared to others ingredients from the same category itself.

The number of shared flavour compounds between the other ingredients was found to be, ginger + coriander = 24, garlic + coriander= 2, green bell pepper + coriander = 12.

However, in some cases we can observe the occurrence of ingredients from different categories, these ingredients can be eliminated if we select the category. These ingredients may be suggested because of their high co-occurrence or high number of shared flavour compounds. We further validated the set of ingredients forming a new recipe if it is similar to any pre-existing set of recipes considering if it does then the chances of acceptability by consumers would be high, shown in Table-A14.

4.3.4 Summary of findings on recipe generation

Three approaches were considered for recipe generation. In the first instance with the attempt to generate recipe for a meat analogue, the ingredient combinations with least number of shared flavour compounds were used. For this, attempt was made to create a recipe by combining the least flavour sharing ingredients from the Northeast recipes, in place of the commercial pre-mix supplied along with the meat analogue. The created recipe with Northeast specific ingredients scored better among the consumers from the region. This initial finding paved the way for extending the criteria to an established pan-Indian recipe. The documented recipes of Khichdi were used for this validation - authentic regional ingredients were combined with a base combination of most prevalent ingredients of khichdi recipes from all over India to generate region specific khichdi recipes. Cosine similarity analysis of the generated khichdi recipes against the khichdi recipes available in open sources indicated higher similarity values in almost all cases. Based on the similarity with recipes published in open domain, these generated recipes were considered as acceptable to consumers. The reliability of the selection of prevalent and authentic ingredients as consumer accepted ingredients for recipe generation, the data driven model was tested for completing/ suggesting ingredients in a recipe. The RLS models were successfully applied to complete sets of three-ingredients by converting them into a recipe each. The recipe generated when validated for similarity using cosine similarity with pre-existing recipe data showed a similarity score close to 1. As a result, this proposed algorithm can be used as an effective tool for recipe completion or recipe recommendation. Further, it can also be used as a tool for new product development to estimate the best-paired ingredients.

4.4 Development of alternative recipe by ingredient replacement for customised specifications with the application of flavour network theory

4.4.1 Ingredient combinations for a food product

Recipe recommendation for alternative ingredients plays an important role to address issue such as limitation in dietary intake and unavailability of certain ingredients considering the similarity and compatibility of ingredients. Humans have a natural tendency to crave restricted foods. A person diagnosed with a diet-restricted disease often loses satisfaction with their taste and appetite and as a result, they may be unable to enjoy their food. The development of modern technology allows for the production of multiple dietary-restricted food items that appeal to health-conscious individuals and those with dietary restrictions as a result of specific religious beliefs. Observations from such products show that they tend to taste different from the prototype they were modelled after.

Flavour perception in humans is quite complicated taking into account a variety of sensory inputs and psychological states. An individual's taste depends not only on the five basic flavour properties but also on surrounding conditions, smells, textures, and memories associated with a particular dish, which all differ from person to person. Thus, expertise such as chefs and food material scientist need to develop a suitable technological solution by manipulating formulations and technological variables to enhance the functional properties of ingredients [3].

Our work presents a scientific claim which shows that flavour compounds play an important role in determining the choice of ingredients in a cuisine, thereby determining the food pairing pattern. Thus, the principle of food pairing can be used as a basic algorithm in designing innovative food products and in generating new recipes as a thorough grasp of customer-based knowledge perceptions, expectations, and attitudes toward new food products is required for the development of innovative food products. Statistical methods are used to create new ingredient combinations that mimic the flavour characteristics of the initial item being replaced. By generating a profile of flavour similar to that of a selected item people with dietary restrictions e.g., diabetes can take pleasure in a particular sweet dish, A, regardless of their dietary restrictions without exceeding the sugar and calorie requirements.

Table 4.15 Alternative ingredients for pork in pork with bamboo shoot curry

		Meat		**Sea Food**		**Vegetable**	
Rank	Original ingredient	Ingredient Name	Opponent	Ingredient Name	Opponent	Ingredient Name	Opponent
1	pork	chicken	0.434654	crab	0.2500	bean	0.9120
2	bamboo shoot	mutton	0.565346	Fermente d fish	0.2500	carrot	0.0482
3	onion			prawn	0.2500	radish	0.0191
4	cayenne			shrimp	0.2500	cauliflower	0.0157
5	ginger garlic paste			fish	0.0000	cabbage	0.0046
6	coriander					lentil	0.0001
7	black mustard seed oil					spinach	0.0001
8						black bean	0.0000
9						bottle gourd	0.0000
10						ash gourd	0.0000

Table 4.16 Alternative ingredients for bamboo shoot in khorisar lagot gahori mangxo

		Meat		**Sea Food**		**Vegetable**	
Rank	Original ingredient	Ingredient Name	Opponent	Ingredient Name	Opponent	Ingredient Name	Opponent
1	pork	chicken	0.9877	crab	0.0000	potato	0.89037
2	bamboo shoot	mutton	0.0123	fermented fish	0.0000	tomato	0.10393
3	onion	pork	0.000	prawn	1.0000	lentil	0.00207
4	cayenne			shrimp	0.0000	black bean	0.00142
5	ginger garlic paste			fish	0.0000	pumpkin	0.00098
6	coriander					eggplant	0.00025
7	black mustard seed oil					spinach	0.00045
8						bottle gourd	0.00004
9						fermented bamboo shoot	0.00020
10						mustard	0.00018

To test our algorithm, we selected a few recipes from each regional cuisine considering if a particular ingredient needs to be replaced. The algorithm was tested only for the Northeast recipe dataset which consists of only 126 ingredients. The recommendation is based on the ingredients available in the recipe dataset. As a result, the recommended ingredients may be fewer. If we wish to have a larger set of ingredients recommendations, we can add more recipe datasets to expand the ingredient list. The alternatives for ingredient replacements of pork in pork with bamboo shoot curry are shown in Table 4.15. An alternative of bamboo shoot in khorisar lagot gahori mangxo is shown in Table 4.16.

Table 4.17 Alternative ingredients for soybean in Dawlrep bai (Mizo stew)

		Meat		Sea Food		Vegetable	
Rank	Original ingredient	Ingredient Name	Opponent	Ingredient Name	Opponent	Ingredient Name	Opponent
1	pork	chicken	1.000	fish	1.000	potato	0.4537
2	cayenne	mutton	0.000	fermented fish	0.000	cauliflower	0.3025
3	garlic	pork	0.000	prawn	0.000	eggplant	0.1512
4	ginger			shrimp	0.000	pumpkin	0.0766
5	soybean			crab	0.000	bean	0.0137
6						tomato	0.0020
7						carrot	0.0000
8						cucumber	0.0000
9						pea	0.0000
10						radish	0.0000

Table 4.18 Alternative ingredients for potato in Ironba (Manipur chutney)

		Meat		Sea Food		Vegetable	
Rank	Original ingredient	Ingredient Name	Opponent	Ingredient Name	Opponent	Ingredient Name	Opponent
1	potato	chicken	1.0000	fish	0.9	spinach	0.3124
2	french bean	mutton	0.0000	prawn	0.1	pumpkin	0.1562
3	cabbage	pork	0.0000	shrimp	0.00	pea	0.1116
4	cayenne			crab	0.00	tomato	0.1046
5	fermented fish			fermented fish	0.00	cauliflower	0.0503
6						bean	0.0503
7						lentil	0.0026
8						carrot	0.0000
9						cucumber	0.0000
10						radish	0.0000

4.4.2 Validation using cosine similarity

We validated the alternative ingredients by replacing the ingredients to be replaced using the cosine similarity shown in Fig. 4.18 (a) and Fig. 4.18 (b). The study showed that the recipe generated is similar to a pre-existing recipe. As a result, we can conclude that the algorithm can be used for recommending alternative ingredients for any target recipe.

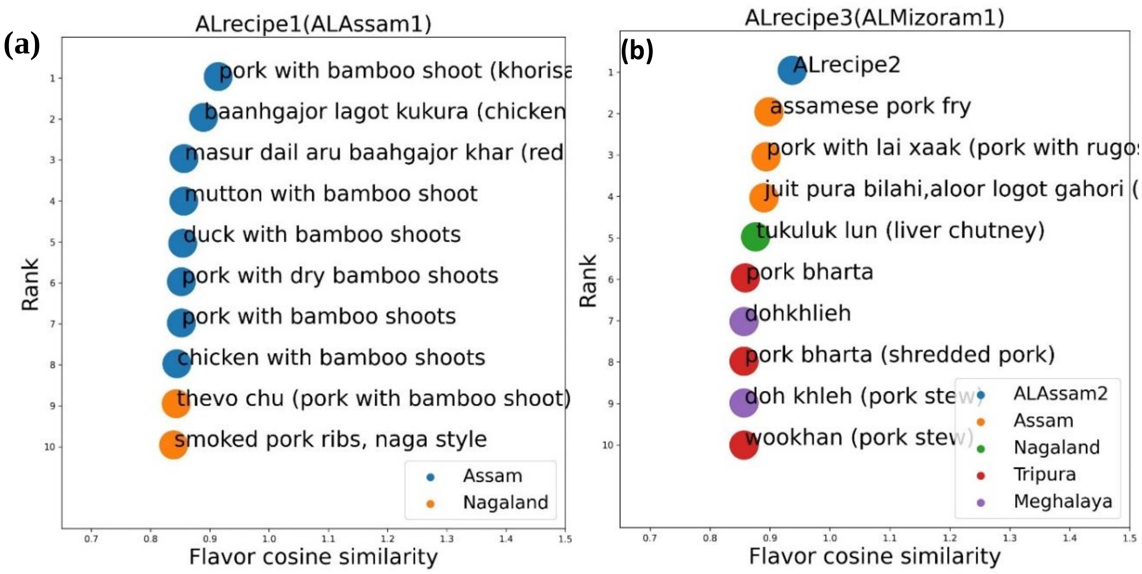


Fig. 4.18 (a) Cosine similarity with original recipe pork replaced with mutton (b) Cosine similarity with original recipe soybean replaced with potato

4.5 Chapter summary

The regional recipes were characterized by the average size of recipes and the rank-frequency plot for the ingredients. Based on the revelation from the exponential form of rank-frequency that few ingredients are more used more frequently used in the cuisine, the prevalent ingredients were determined. It was observed that ingredients such as black mustard seed oil, onion, cayenne, ginger, green bell pepper, garlic, turmeric, bay laurel, pork, rice, and tomato are the most prevalent ingredients across the cuisines. Additionally, ingredients such as black mustard seed oil, green bell pepper, onion, cayenne, ginger, garlic, rice, turmeric, bay laurel, tomato and pork were found to be the most authentic ingredients across the regional cuisines. The identification of pork meat as one of the most authentic ingredients in the Northeast regional cuisine is found to be a distinctive feature as compared to other Indian regional cuisines. The data driven clustering analysis at ingredient level did not indicate any separate cluster for the ingredients used in the Northeast regional cuisines. The comparative study of Northeast cuisine with the other Indian regional cuisines revealed that Assam regional cuisine shares the most affinities with Indian regional cuisine with a total of 519 similar recipes out of 2916 recipes of the other Indian regional cuisine and Tripura recipes were found to be least similar as only 4 recipes are found to be similar. The ingredient-based cosine similarity analysis at recipe level indicated that recipes within the cuisine are quite like each other. In many instances recipes vary by one ingredient, due to a replacement ingredient from the same category. This finding has led to testing of few ideas of recipe generation or recipes completion in the third objective.

Flavour network-based analysis of recipes revealed preference for combinations of ingredients based on sharing of flavours among the ingredients in a recipe. Ingredients such as cayenne, garlic, turmeric, and bay laurel contributed to the negative pairing and ingredients such as ginger, tomato, and pork contributed to the positive pairing. Overall, all the regional cuisines from Assam, Arunachal, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura, when compared with the corresponding randomized cuisine showed uniform negative food pairing behaviour. The majority of the ingredients that made a substantial impact on the food pairing were from the spice category. Based on the negative pairing behaviour and contributions of spice category ingredients towards negative pairing of ingredients, generation or creation of recipes based on Northeast cuisines were attempted with prevalent and/or authentic ingredients, primarily from the spice category, in the works against the third and fourth objectives. It revealed that with the development of an algorithm for alternative ingredients considering customer specifications, new ingredient combinations can be created that mimic the flavour characteristics of the initial item being replaced. The recipe generated through the algorithm when compared to the pre-existing recipe considering the flavour properties was closely related. As a result, we can consider the proposed algorithm to be used for future recipe development purposes.