4
Spices form the Basis of Food Pairing in Indian Cuisine

In this Chapter, we study the food pairing pattern in Indian Cuisine and built computational models of the cuisine to identify features that explain the statistical properties of the cuisine. The flavor constitution of the Indian cuisine was explored for ingredient composition and food pairing at the levels of cuisine, sub-cuisines, recipes and ingredient pairs. We built controls to probe for the role of factors that may be crucial in shaping recipes and hence the cuisine. Our study illustrates the application of data analysis and modeling for exploring the chemical basis of a cuisine.

4.1 COMPUTING FOOD PAIRING AT RECIPE LEVEL

The composition of recipes in a cuisine could be studied in terms of the co-occurrence of ingredients [Ahn et al., 2011; Teng, Lin, and Adamic, 2011]. One of the notions associated with ingredient co-occurrence is food pairing hypothesis— ingredients sharing flavor compounds are more likely to taste well together than ingredients that do not [Ahn et al., 2011; Blumenthal, 2008]. While this hypothesis holds true for some cuisines (North American, Western European and Latin American), it does not hold for a few others (Southern European and East Asian) which show an opposite food pairing trend [Ahn et al., 2011]. Thus, beyond following generic statistical patterns in recipe sizes as well as ingredient use, skewed food pairing seems to be a unique feature representing the molecular basis of ingredient combinations dominant in a cuisine. Towards the aim of quantifying the pattern of ingredient composition of recipes, we studied food pairing (sharing of flavor compounds) in Indian cuisine.

The flavor sharing [Ahn et al. 2011] pattern was enumerated among the ingredients that co-occur in a recipe, starting from the set of 2,543 (N_R) traditional Indian recipes comprising of 194 (N_I) ingredients. We computed the average number of shared compounds in each recipe N_s^R and further calculated a representative average flavor sharing index $\overline{N_s}$ ($= \sum_R N_s^R / N_R$) of the cuisine. Figure 4.1 presents a graphic illustration of this procedure. For a recipe R with s ingredients N_s^R is defined as,

$$N_s^R = \frac{2}{s(s-1)} \sum_{i,j \epsilon R, i \neq j} \left| F_i \cap F_j \right|, \qquad (4.1)$$

where F_i represents the flavor profile of ingredient i (a set of compounds).

Average flavor sharing in Indian cuisine was compared with a corresponding randomized cuisine to assess its statistical relevance by computing $\Delta N_s = \overline{N}_s^{IC} - \overline{N}_s^{Rand}$, where 'IC' and 'Rand' indicate Indian cuisine and corresponding random cuisine, respectively. Five types of randomized cuisines were created by maintaining the recipe size distribution of the original Indian cuisine: a randomized control where ingredients were chosen uniformly (20,000 recipes); a frequency-preserved control in which frequency of use of ingredients was preserved (20,000 recipes); a category-preserved control in which while the category composition of the recipe was preserved, ingredients were randomly chosen from each constituent category (8 sets of control cuisines, 20,344 recipes); a frequency-and-category-preserved control where the category composition was maintained and each ingredient was chosen with probability consistent with its frequency in Indian cuisine (8 sets of control cuisines, 20,344 recipes); a frequency-preserved randomized control where the top 10 ranked ingredients in the Indian cuisine were randomly swapped with low ranked (rank ≥ 11) ingredients (10 sets of control cuisines, 200,000 recipes). The statistical significance of ΔN_s was measured with Z-score,

$$Z = \sqrt{N_{Rand}}\frac{(\bar{N}_s^{IC} - \bar{N}_s^{Rand})}{\sigma_{Rand}}. \quad (4.2)$$

Here N_{Rand} and σ_{Rand} represent total number of recipes in the randomized cuisine and standard deviation, respectively. The interrelationship among ingredients by virtue of shared flavor compounds could be represented as a flavor graph that illustrates the underlying topology of flavor sharing (Figure A.1.1 and A.1.2). The ingredients have dominant intra-category flavor sharing indicating a significant overlap of flavor profiles within the category (Figure A.1.3). We quantified flavor sharing in a recipe (N_s^R) and average flavor sharing of the cuisine ($\overline{N_s}$) by comparing profiles of ingredient pairs and their joint occurrence in recipes. Figure 4.1 illustrates this quantification procedure starting from data of recipes and flavor profiles.

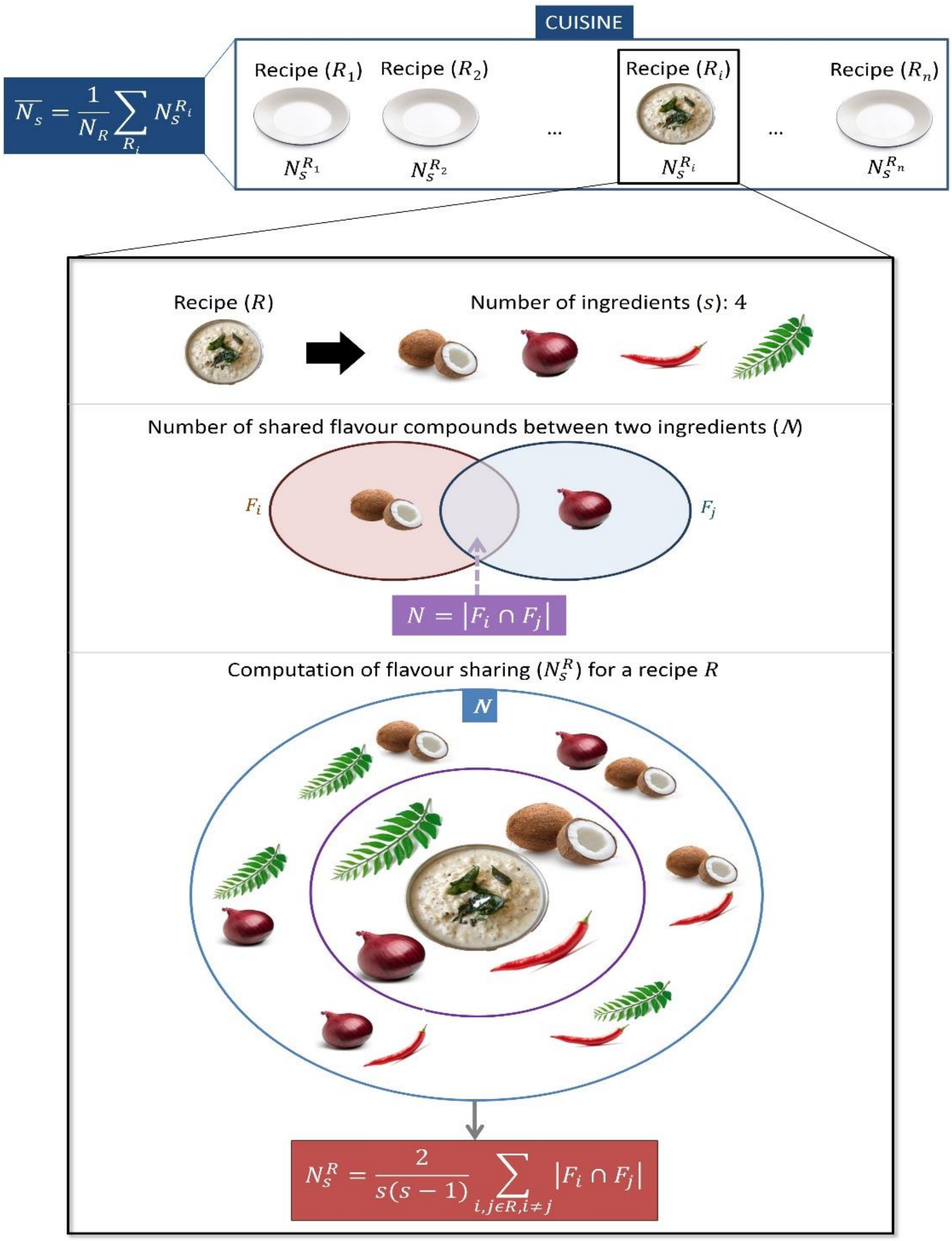

Figure 4.1: Illustration of procedure used for computation of average food pairing of a cuisine. Starting from the cuisine data and flavor profiles of ingredients, average number of shared compounds in

each recipe was computed. The average food pairing of a recipe set was further computed to enumerate flavor sharing.

4.2 INDIAN CUISINE IS CHARACTERIZED WITH STRONG NEGATIVE FOOD PAIRING

The food pairing hypothesis is tested by computing the average number of flavor molecules shared in pairs of ingredients in the real cuisine as compared to that in their corresponding random cuisine [Ahn et al., 2011]. For Indian cuisine, we found that average flavor sharing was significantly lesser than expected by chance (Figure 4.2a). To distinguish it from the already observed phenomenon of food pairing in Western cuisines like North American and Western European, we refer to this trend as 'negative food pairing'. When computed for all recipes in the cuisine, average flavor sharing for Indian cuisine was found to be 5.876, as compared to that of 9.442 for a randomized cuisine, which was constructed by randomly picking the ingredients while maintaining the recipe size distribution. This reflects a strong signature of non-random ingredient co-occurrence ($\Delta N_s = \bar{N}_s^{IC} - \bar{N}_s^{Rand}$ = -3.566 and Z-score of -54.727) between pairs of ingredients and their co-occurrence in the cuisine (Figure 4.3a). More the extent of flavor sharing between any two ingredients in the Indian cuisine, lesser is their co-occurrence. The extent of food pairing bias in the Indian cuisine is much stronger than reported earlier for any other cuisine [Ahn et al., 2011; K. R. Varshney et al., 2013] and is persistent regardless of the recipe size (Figure 4.2a). Our analysis also showed that each of the sub-cuisines independently displayed negative food pairing, highlighting it as an invariant feature of the Indian cuisine (Figure 4.2b). Thus, we conclude that Indian cuisine is characterized by a strong negative food pairing.

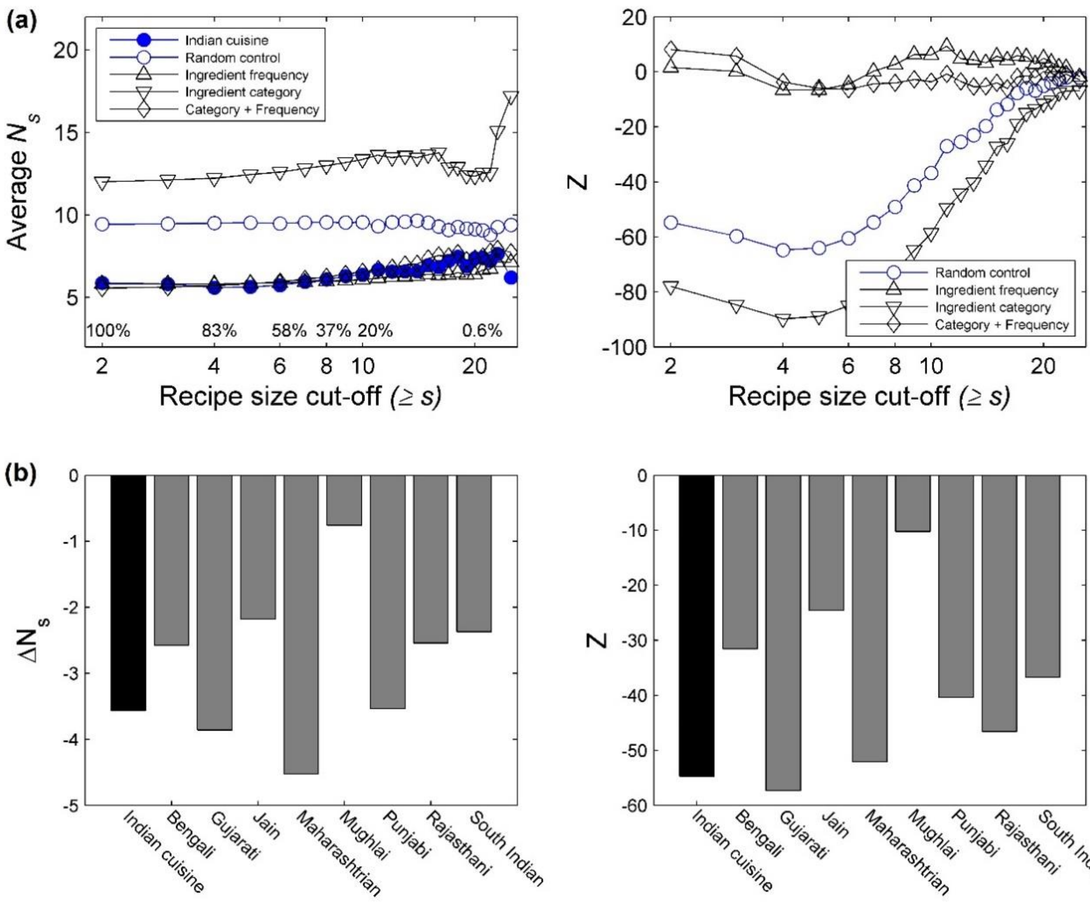

Figure 4.2 : Strong negative food pairing in Indian cuisine and constituent sub-cuisines. (a) The Indian cuisine is characterized by strong negative flavor sharing when compared to its random control. The pattern of negative food pairing is independent of the recipe size (s) and is statistically significant. While all 2,543

recipes are included for enumeration at the cut-off of two, only around 3% (80) and 0.6% (15) recipes are considered at the cut-off of 15 and 20, respectively. While the recipes set controlled only for ingredient category did not explain the negative food pairing, controlling for frequency of use of ingredient reproduces the characteristic profile. (b) Strong negative food pairing emerged as an invariant feature of all sub-cuisines as measured in terms of average food pairing and its statistical significance (Recipe size, s ≥ 2).

We further explored the origin of this characteristic pattern by controlling for category and frequency of use of ingredients. The former is a recipe-level control that generates a cuisine by preserving the category composition of each recipe, whereas the latter is a cuisine-level control that generates recipes by preserving the frequency of occurrence of each ingredient. Interestingly, we observed that controlling only for the ingredient frequency leads to a food pairing pattern similar to that of real-world cuisine (Figure 4.2a, Figure 4.3b and Figure A.1.4). Controlling only for the ingredient category, on the other hand, led to a pattern similar to that of a randomized cuisine. A randomized control that combines category-composition, as well as ingredient frequency, also reproduced the food pairing pattern. Thus, ingredient frequency emerged as the dominant factor specifying the characteristic flavor sharing pattern of Indian cuisine. Considering the biased use of ingredients, we investigated the role of top-ranked ingredients by randomly swapping the top ten ingredients with the rest. We found that, indeed, the highly ranked ingredients play a key role in shaping the negative food pairing pattern of the cuisine, in contrast to ingredients with poor ranking (Figure A.1.5).

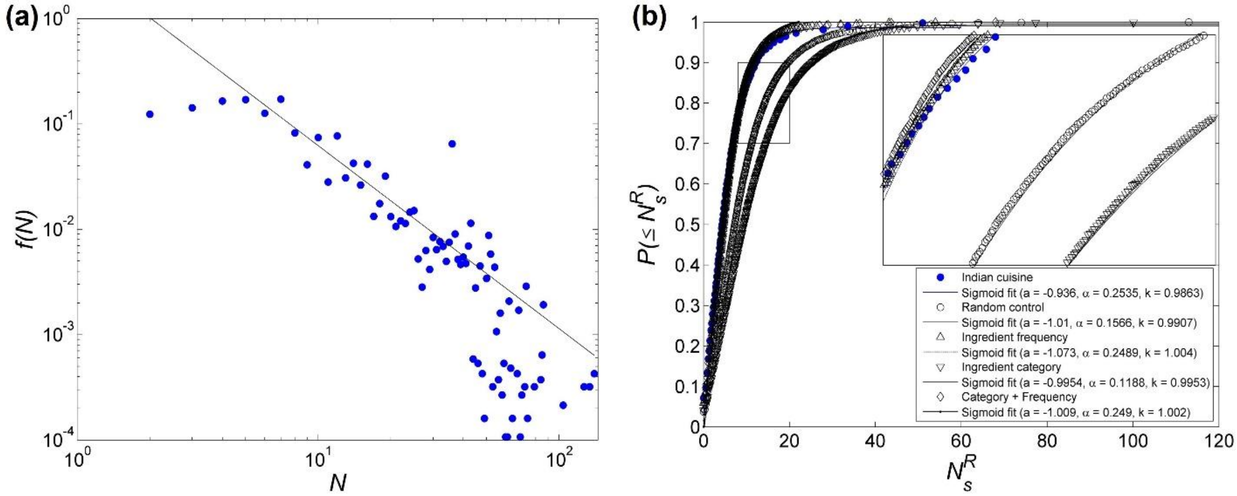

Figure 4.3 : Negative food pairing at ingredient level and investigation of food pairing with recipe-level statistics. (a) Fraction of ingredient pairs' frequency (f(N)) with increasing number of shared flavor compounds (N). The figure shows that more the flavor sharing between two ingredients, the less is their pairing in the cuisine. The frequency of ingredient co-occurrence falls as a power law (with an exponent of -1.74). (b) Cumulative distribution of 'average number of shared flavor compounds of recipes' in a cuisine. Cumulative distribution of N_s^R $P(\le x) = a + \frac{(k-a)}{1+e^{-\alpha x}} N_s^R$ frequency follows an exponential distribution. These results corroborate the observation that frequency of use of ingredients is a key contributor to the food pairing pattern.

4.3 SPICES ARE UNIQUELY PLACED IN THE RECIPES

Negative food pairing in Indian cuisine is a cumulative result of individual ingredient contributions by virtue of pairing with other ingredients in recipes. To investigate the importance of individual ingredients and their categories in the composition of recipes, we randomized ingredients of each category independently, while maintaining the category as well as the frequency of occurrence of the rest. We found that randomizing ingredients in any of the categories, except spices, does not affect the negative food pairing pattern, thereby implying their insignificance (Figure 4.4a). Spices, on the other hand, when swapped selectively, randomize the negative food pairing significantly (Figure 4.4a and b, $\Delta N_s^{spice} = 4.229$ and Z-score of -61.524). This implies that each of the spices is uniquely placed in its recipe to shape the flavor sharing pattern with the rest of the ingredients and is sensitive to replacement even with other spices, which is noteworthy given that the extent of flavor sharing is high among spices (Figure A.1.3 A).

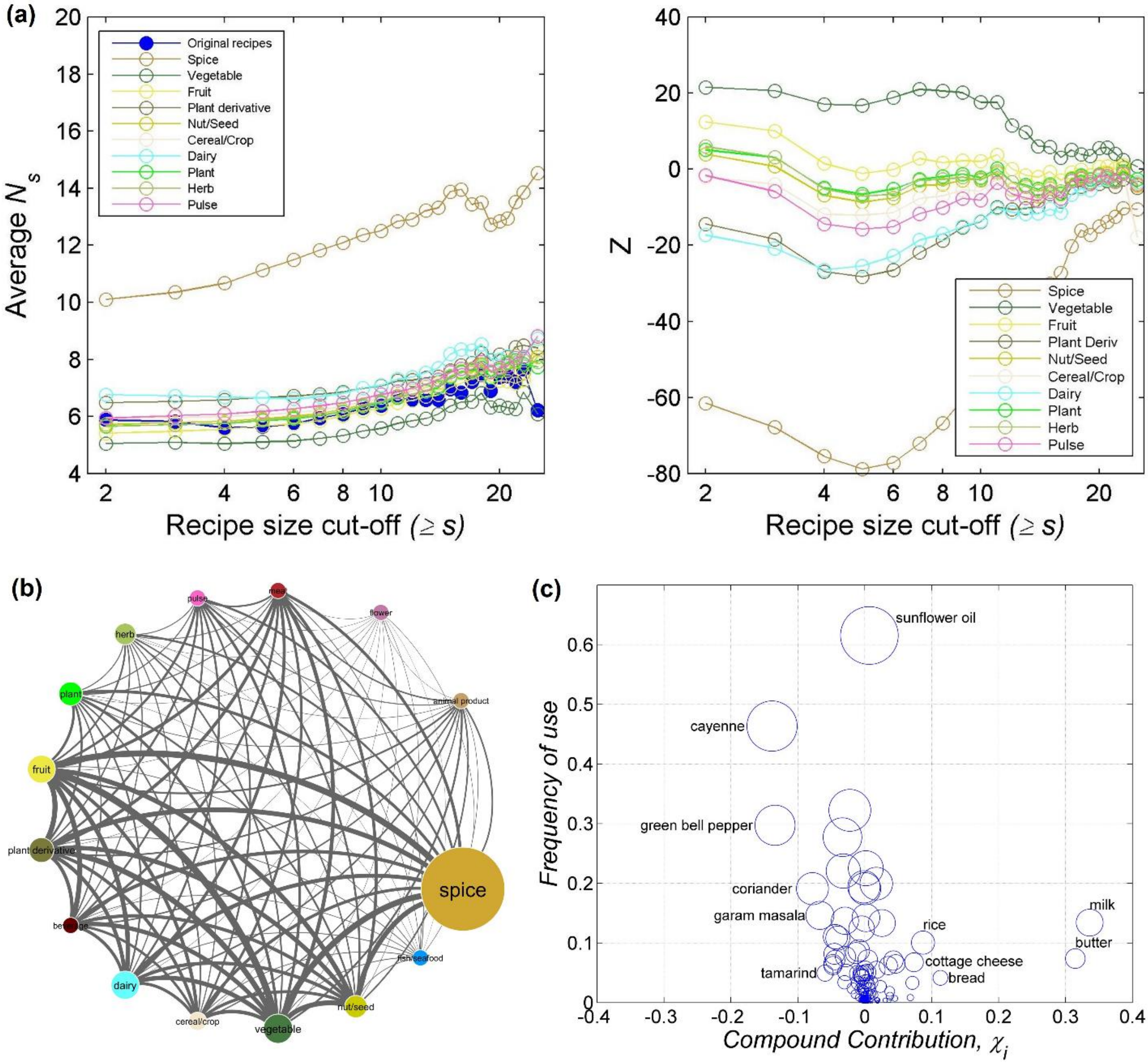

Figure 4.4: Spices are critical contributors to the negative food pairing in Indian cuisine. (a) Average food pairing of Indian cuisine when each ingredient of a given category is randomly replaced with another ingredient of the same category and its statistical significance. Such intra-category randomization reflects the uniqueness of the ingredient in recipes knowing that ingredients tend to have similar flavor profiles within the category (Figure A.1.4). Spices are uniquely placed in the recipes, and when randomly replaced by another spice, the flavor sharing pattern was drastically randomized. For a similar random intra-category replacement of ingredients of other categories, the flavor pattern showed no significant change. (b) Flavor sharing among ingredient categories. Size of circles denotes the extent of change that the category makes when its ingredients are randomly shuffled (ΔN_s^{cat}) reflecting its importance in flavor sharing profile. (c) The relevance of individual ingredient enumerated in terms of the extent of its contribution towards positive or negative food pairing (χ_i) and frequency of use. Spices emerge as the most significant contributors to negative food pairing.

4.4 SPICES ARE KEY CONTRIBUTORS TO THE NEGATIVE INGREDIENT PAIRING

Beyond global statistical features, we identified the ingredients that make key contributions towards the food pairing by computing the extent to which their presence affects the magnitude of average food pairing (χ_i). We found that the key ingredients that contribute to negative food pairing of Indian cuisine were spices (Figure 4.4c). Among the top ten ingredients whose presence bias flavor sharing pattern of the Indian cuisine towards negative pairing, nine were spices: cayenne, green bell pepper, coriander, garam masala, tamarind, ginger garlic paste, ginger, clove, and cinnamon (See Annexure Table A.2.3). We surmise that this pivotal role of spices carries evidence of the historical practice of a health-centric diet in the Indian subcontinent.

4.4.1 Ingredient Uniqueness

The uniqueness of an ingredient of a given category by virtue of flavor sharing pattern with other ingredients in the recipe was computed by replacing it with a randomly chosen ingredient from the same category. Deviation in the average flavor sharing of the randomized recipes (8 sets of control cuisines, 20,344 recipes) from that of the original cuisine was measured for ten major categories (depicted in Figure A.1.4).

$$\Delta N_s^{cat} = |\bar{N}_s^{IC} - \bar{N}_s^{cat}| \ \forall\ s \geq 2\,, \tag{4.3}$$

Here, *cat* stands for the ingredient category. This index enumerates the contribution of ingredients of a given category towards the flavor sharing pattern of the cuisine. The statistical significance of ΔN_s^{cat} was measured with the Z-score.

4.4.2 Ingredient Contribution

The contribution of each ingredient (χ_i) to the flavor sharing pattern of the cuisine was quantified [Ahn et al. 2011] in terms of the extent to which its presence biases the flavor pairing.

$$\chi_i = \left(\frac{1}{N_R} \sum_{R \ni i} \frac{2}{s(s-1)} \sum_{j \neq i (j.i \in R))} |F_i \cap F_j|\right) - \left(\frac{2f_i}{N_R \langle s \rangle} \frac{\sum_{j \in c} f_j |F_i \cap F_j|}{\sum_{j \in c} f_j}\right), \tag{4.4}$$

Here, f_i is the frequency of occurrence of ingredient i.

This value is an indicator of how an ingredient's presence affects the net positive (negative) food pairing shown by the cuisine, depending on whether the value is positive (negative).

...