

3.1 INDIAN CUISINE—AN INTRODUCTION

Culinary practices around the world are influenced by climate, culture, history, and geography, which leave their imprint on the food preferences on regional cuisines. The study of the molecular composition of recipes in a cuisine reveals these food preference patterns. Indian culinary system has traditionally developed dietary practices where food has nutritional as well as medicinal value [Atridevji Gupt, 1948; Govindarajan, Vijayakumar, and Pushpangadan, 2005; Thatte and Dahanukar, 1986]. Ayurveda, the classic medicinal system of India, proposes that food has as much therapeutic value as drugs and even uses similar processing techniques for their preparation [Atridevji Gupt, 1948; Valiathan, 2010].

Beyond being orally transmitted, the traditional Indian recipes have been documented in the form of books [Appadurai, 1988] and lately through online repositories. We compiled the recipes data from TarlaDalal.com (<http://www.tarladalal.com>, November 2014), one of the largest, most extensive online repertoires of original Indian recipes. After curation, the Indian cuisine contained 2543 recipes from the following eight sub-cuisines: Bengali, Gujarati, Jain, Maharashtrian, Mughlai, Punjabi, Rajasthani and South Indian. These sub-cuisines span diverse geographies, climates and cultures of the Indian subcontinent. After aliasing the ingredient names, the recipes comprised of 194 unique ingredients. This chapter analyses the recipe composition as well as flavor compound constitution of the Indian cuisine in search of its quintessential features.

3.2. DATA MINING RECIPES IN INDIAN CUISINE

3.2.1. Data Source Selection

The data on Indian cuisine was obtained from one of the leading cookery websites in India, <http://tarladalal.com>. We cross-checked the data from other similar repositories for Indian Cuisine before deciding on Tarladalal.com. Upon comparison, TarlaDalal.com was found to be the best in terms of authentic recipes, cuisine annotations and coverage across major regional subtypes. Upon comparison, TarlaDalal.com was found to be the best in terms of authentic recipes, cuisine annotations and coverage across major regional subtypes. The website has 3330 recipes from 8 Indian cuisines. Among other online sources, Sanjeev Kapoor (<http://www.sanjeevkapoor.com>) has 3399 recipes from 23 Indian cuisines. NDTV Cooks (<http://cooks.ndtv.com>) has 667 Indian recipes across 15 cuisines. Manjula's Kitchen (<http://www.manjulaskitchen.com>) is restricted to 730 Indian vegetarian recipes across 19 food categories. Recipes Indian (<http://www.recipesindian.com>) has 891 recipes from around 16 food categories. All Recipes (<http://www.allrecipes.com>) has only 449 recipes from 6 food categories. In comparison to these sources, as of January 2015, Tarladalal.com was adjudged as the best online source of recipes for Indian cuisine.

3.2.2 Data Collection and Curation of Recipes and Ingredients

We started with the data containing 3330 recipes. This number was reduced after pruning to remove duplicates and other inconsistencies, to obtain 3037 recipes. Recipes containing any of the ingredients for which no flavor profile was available (349) were removed from the data,

leaving us with 2688 recipes. Finally, all recipes having a single ingredient were removed to obtain a final set of 2543 recipes.

Table 3.1: Number of ingredients in each category. Ten major categories with five or more ingredients are highlighted.

Sl. No	Ingredient category	No. of ingredients in the category
1	Spice	39
2	Vegetables	36
3	Fruit	33
4	Plant derivative	18
5	Nut/seed	15
6	Cereal/crop	12
7	Dairy	11
8	Plant	8
9	Pulse	6
10	Herb	5
11	Meat	3
12	Fish/seafood	2
13	Beverage	2
14	Animal product	2
15	Flower	1

In the next step, we isolated the ingredients from the collected recipes. The Indian cuisine data from Tarladalal.com comprised a total of 588 ingredients. For the purpose of mapping the ingredients to their flavor profiles (list of volatile flavor compounds present in that ingredient), these ingredient names were aliased to 339 source ingredients. These ingredients were categorized into seventeen ingredient categories based on the nature of ingredients. For the purpose of flavor sharing analysis, we removed ingredients from ‘snack’ and ‘additive’ categories, as their flavor profiles could not be determined. The rest of the ingredients belonged to the following 15 categories: spice, vegetable, fruit, plant derivative, nut/seed, cereal/crop, dairy, plant, pulse, herb, meat, fish/seafood, beverage, animal product, and flower (Table 3.1). Certain pulses (millets), lentils, spices and vegetables, seldom used in other cuisines, were common to Indian cuisine. Out of these 339 aliased ingredients, we could obtain flavor profiles for 194 ingredients. By aliasing, we mean, removing the redundant terms in ingredient names. For instance, *canned pineapple* was aliased as *pineapple*. Ingredient names in Hindi were translated to English. For instance, *anardaana* was renamed as *pomegranate*. Further, we aliased some of the ingredient names for the purpose of matching our ingredient names with the data of Ahn *et al.* [Ahn *et al.*, 2011] for ease of collection of flavor compounds. For instance, *bay leaf* was named *bay laurel*. By crosschecking with the dataset of Ahn *et al.* [Ahn *et al.*, 2011], ingredients unique to Indian cuisine were identified. The overview of data collection is illustrated in Figure 3.1.

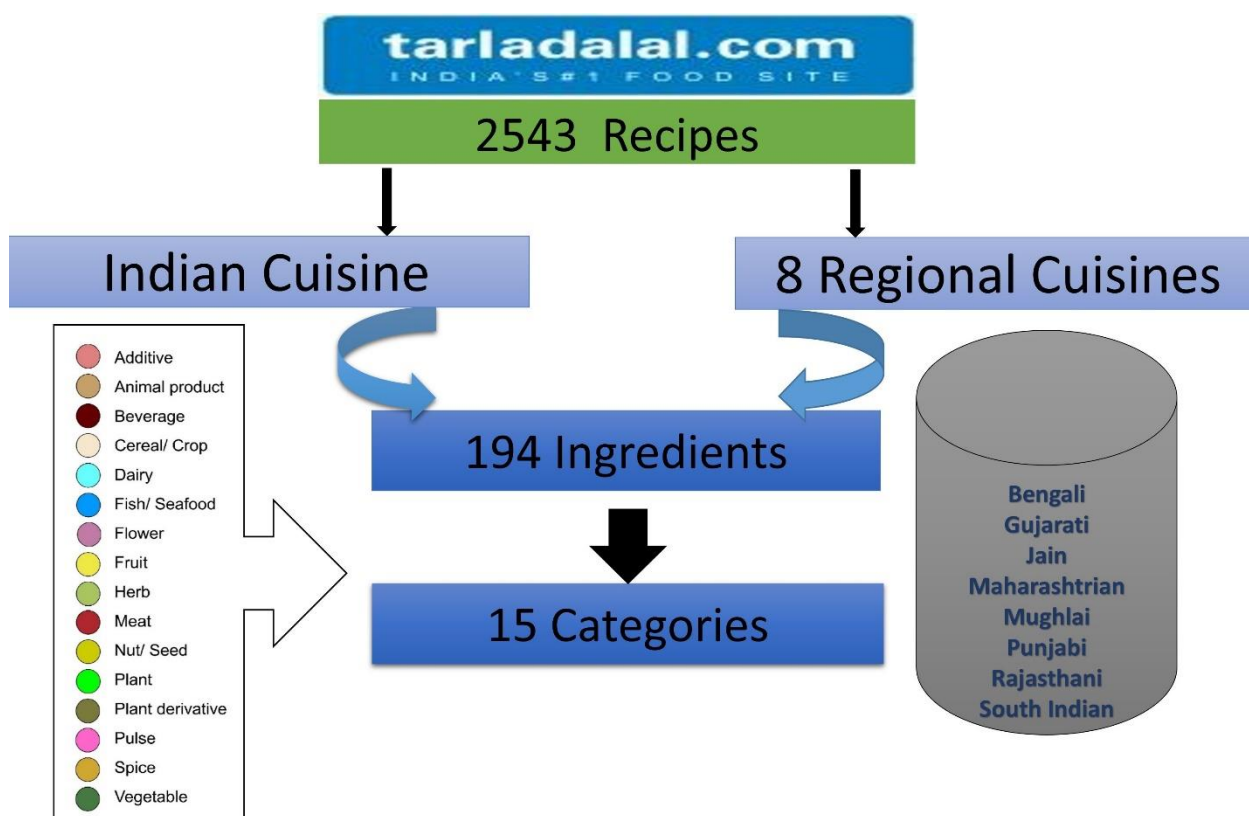


Figure 3.1: Overview of Indian Cuisine Data

3.2.3 Compilation of the Flavor Profiles Data

The flavor of an ingredient is an important factor for its incorporation in recipes. Flavor compounds are chemical compounds whose chief function in food is to impart flavor [Burdock, 2010]. They include natural compounds found in food ingredients as well as synthetic compounds used to enhance the palatability of food. A common example of a flavor compound is that of citric acid found in citrus fruits that give them their sour taste. Similarly, diallyl sulfide imparts the pungent taste to the spice asafetida. The data on flavor compounds are difficult to obtain and is not freely available.

The data on flavor compounds were obtained from various sources (Figure 3.2). Our main source of flavor compounds was the data made available in the work by Ahn *et al.* [Ahn et al., 2011] and the 6th edition of Fenarolis Handbook of Flavor Compounds [Burdock, 2010]. Out of 194 ingredients with flavor profiles, 146 were obtained from the dataset made available by Ahn *et al.* [Ahn et al., 2011]. Ingredients for which the flavor data could not be obtained from either of these sources were compiled through a literature survey. Our data of Indian cuisine as well as flavor compounds is not exhaustive. The original ingredient number had to be drastically reduced as flavor profiles of some of the ingredients could not be identified.

The 31 ingredients which were found to be unique for Indian cuisine, their flavor profiles were compiled by an extensive literature survey (See Annexure A.2 for a complete list of references). All the flavor profiles were cross-checked with those in the 6th edition of Fenaroli's Handbook of Flavor Compounds [Burdock, 2010] so that their names are used consistently. Chemical Abstract Service numbers were also used as unique identifiers to bring consistency in the nomenclature of flavor molecules. Refer to Figure 3.3 for the statistics of distribution of flavor profiles across categories.

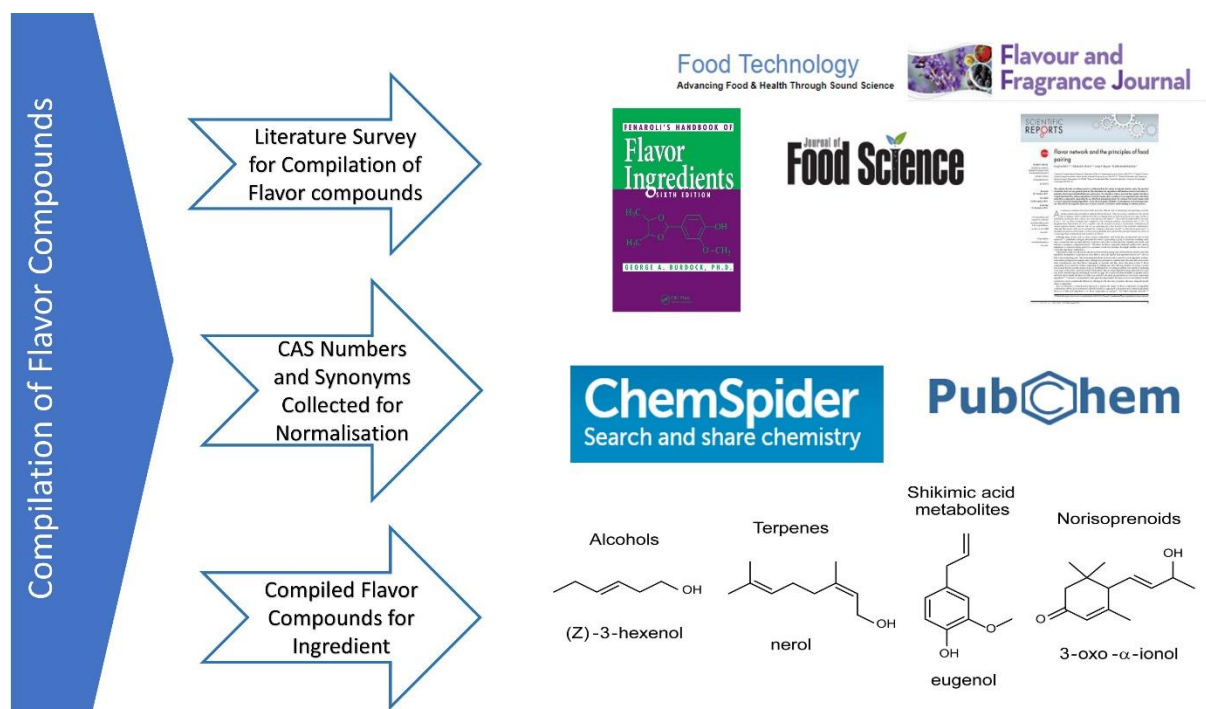


Figure 3.2: Compilation of Flavor compounds Data

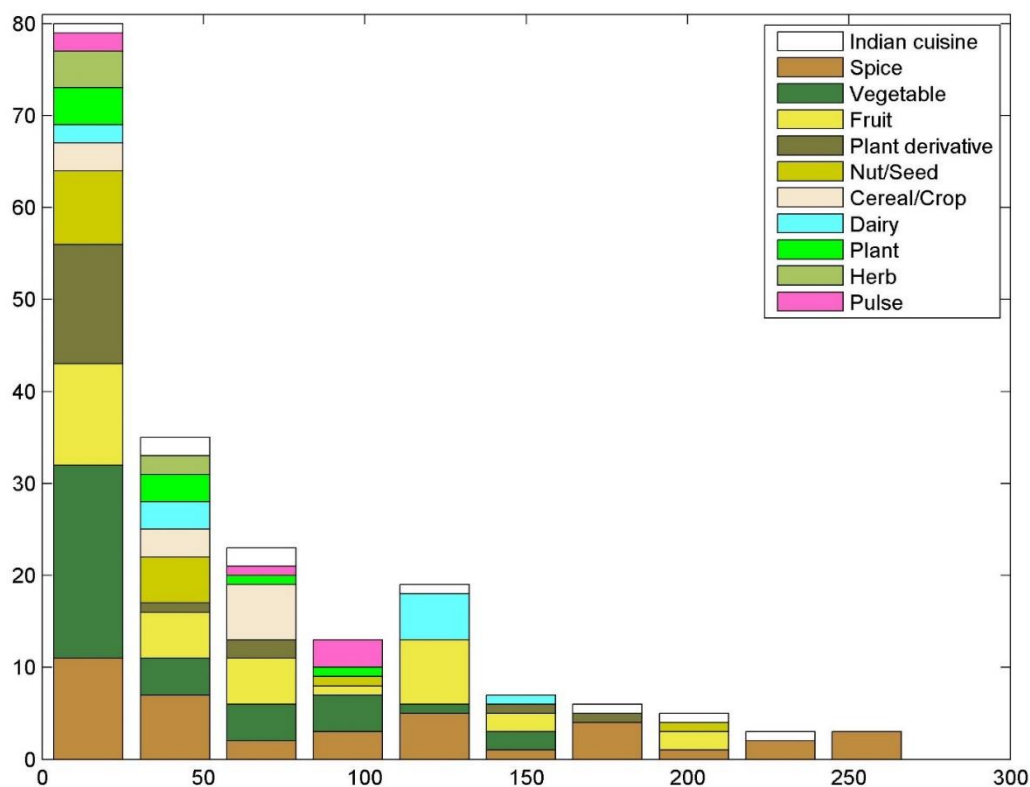


Figure 3.3: The number of flavor molecules across all ingredients in Indian Cuisine. The white box stands for the number of flavor molecules across all ingredients in the Indian cuisine.

3.3. STATISTICS OF RECIPE SIZE AND INGREDIENT FREQUENCY

Recipe size distribution reflects the ingredient richness of recipes in the cuisine. It captures the number of ingredients per recipe. It has been shown that cuisines across the world show a similar distribution in recipe size coincidentally, meaning that world cuisines use on an average of 7-8 ingredients in their recipes. They follow a bounded distribution. Similar to other cuisines [Ahn et al., 2011; Kinouchi et al., 2008; Zhu et al., 2013], the recipe size distribution of Indian cuisine is also bounded, varying between 1 and 40 with an average size of 7 (Figure 3.4a). Mughlai sub-cuisine, with a royal lineage, had a bimodal distribution with exceptionally large recipe sizes.

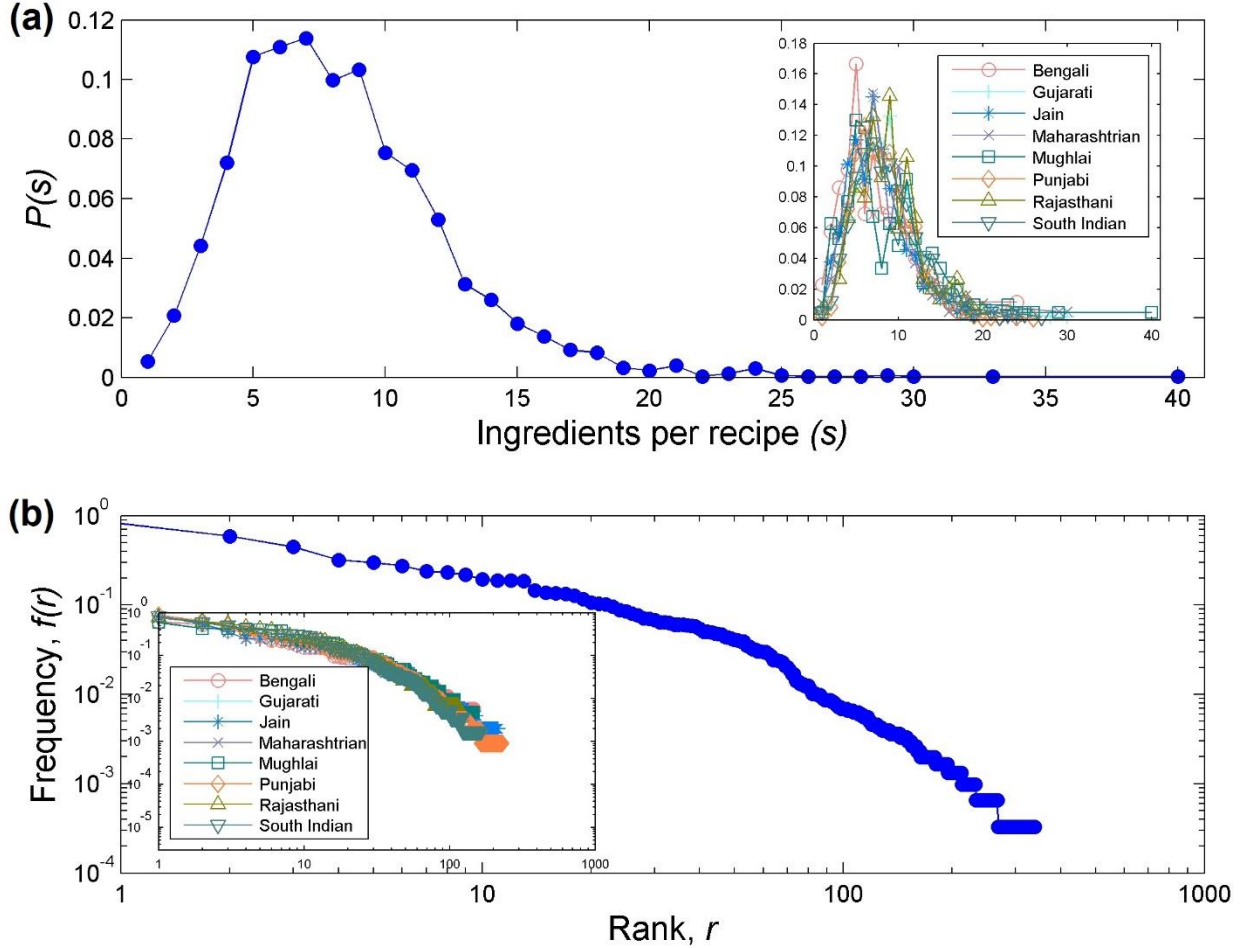


Figure 3.4: Statistics of Indian cuisine. (a) The recipe size distribution of the Indian cuisine exhibits a bounded distribution similar to other cuisines [Ahn et al., 2011; Kinouchi et al., 2008; Zhu et al., 2013]. The inset shows the distribution for constituent sub-cuisines. (b) Frequency rank plot of use of ingredients, in Indian cuisine and its sub-cuisines (inset), shows preferential use of few ingredients. The more frequently an ingredient is used in the cuisine, the better is its rank.

Ingredient frequency, when plotted against rank, by ordering ingredients according to their prevalence in the cuisine, reflects bias in the use of ingredients. In contrast to Recipe size distribution, the Ingredient frequency of several cuisines varies over orders of magnitude, illustrating large differences in frequency of use of ingredients in cuisines. This power law distribution suggests that some ingredients are used more than others. This is not only seen in the case of Indian Cuisine but is a universal property followed by other world cuisines as well as exemplified in Ahn et al. [Ahn et al., 2011]. It has been observed that in every cuisine, there is a small number of ingredients that are used very frequently in a cuisine. Upon preliminary investigation, we observed that ingredients like Ghee, Asafetida, Garam Masala, Curry leaf, etc. are the most frequently used ingredients.

The frequency-rank distribution of the Indian cuisine varies over three orders of magnitude following a scale-free distribution (Figure 3.4b). All sub-cuisines have strikingly similar profiles, indicative of generic culinary growth mechanisms (Figure 3.4b, inset). A few ingredients are excessively used, indicating either their inherent 'fitness' or possible accidental lock-in [Kinouchi et al., 2008]. Knowing that many cuisines share this property [Ahn et al., 2011; Kinouchi et al., 2008; Zhu et al., 2013], the question is whether the pattern of use of ingredients has any role in rendering the profile of a cuisine.

3.4 CONCLUSION

This chapter sums up the sources from which data was collected for the study of Indian Cuisine. It also illustrates the statistics of recipes, ingredients and flavor profile data of the traditional Indian Cuisine.

...