Recipe Data Analytics

Evolution of Cuisines

- How did recipes evolve over time?
- The Panda cuisine
- Models for recipes evolution
  - Kinouchi et al, NJP (2008)
  - Jain & Bagler, Physica A (2018)
  - Tuwani et al., ICDE (2019)

- Simple models for recipe evolution
  - Ingredients and Categories

Evolution of Cuisines

- The oldest cuisines with appreciable complexity (cooking protocols as well as ingredient combinations) may be around 10-15 thousand years old.
- One of the most intuitional accessible mechanisms which would contribute to increase in repertoire of recipes of a cuisine/culture is replication of an existing recipe with a ‘modification’.
- Following are the three possible mechanisms of modifications:
  - Ingredient Mutation/Replacement
  - Ingredient Addition
  - Ingredient Deletion
- Let’s consider the simplest of these models: **Ingredient Mutation/Replacement**.

New Journal of Physics
The open-access journal for physics

The non-equilibrium nature of culinary evolution

Evolution of Cuisines

- **A model for cuisine evolution by copy-mutate model**: Kinouchi et al, ‘The non-equilibrium nature of culinary evolution’, New Journal of Physics, 10 (2008).
- **Data**: Brazilian, British, French and Medieval cookery books.
- A copy-mutate process to model culinary evolution that fits the empirical data very well.
- A cultural ‘founder effect’ is produced by the non-equilibrium dynamics of the model.

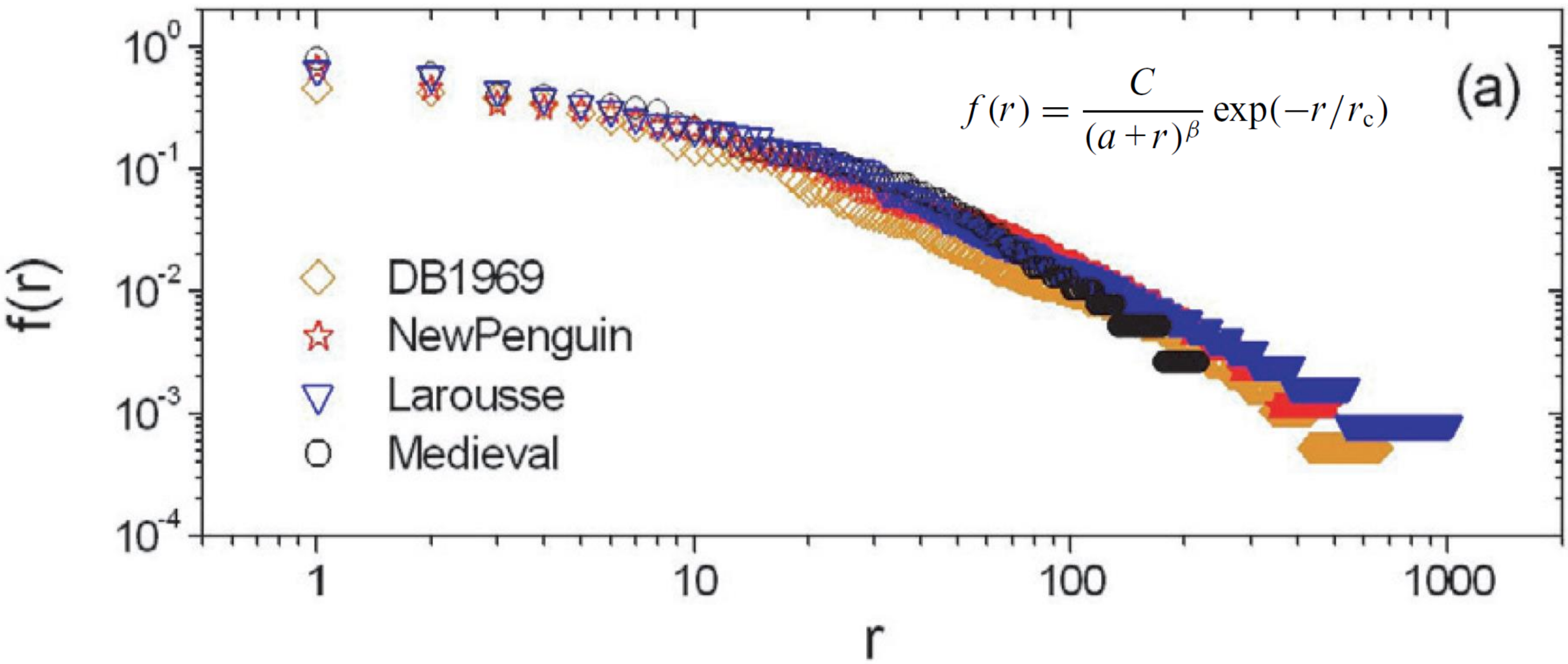
Evolution of Cuisines

Table 1. Numerical characteristics for the considered cookery books.

Cookery book	No. of recipes	No. of ingredients	Average recipe size
Dona Benta (1946)	1786	491	6.7
Dona Benta (1969)	1894	657	7.1
Dona Benta (2004)	1564	533	7.4
Larousse Gastronomique (2004)[a]	1200	1005	10.8
New Penguin Cookery Book (2001)	878	489	9.5
Pleyn Delit (Medieval)	380	219	9.7

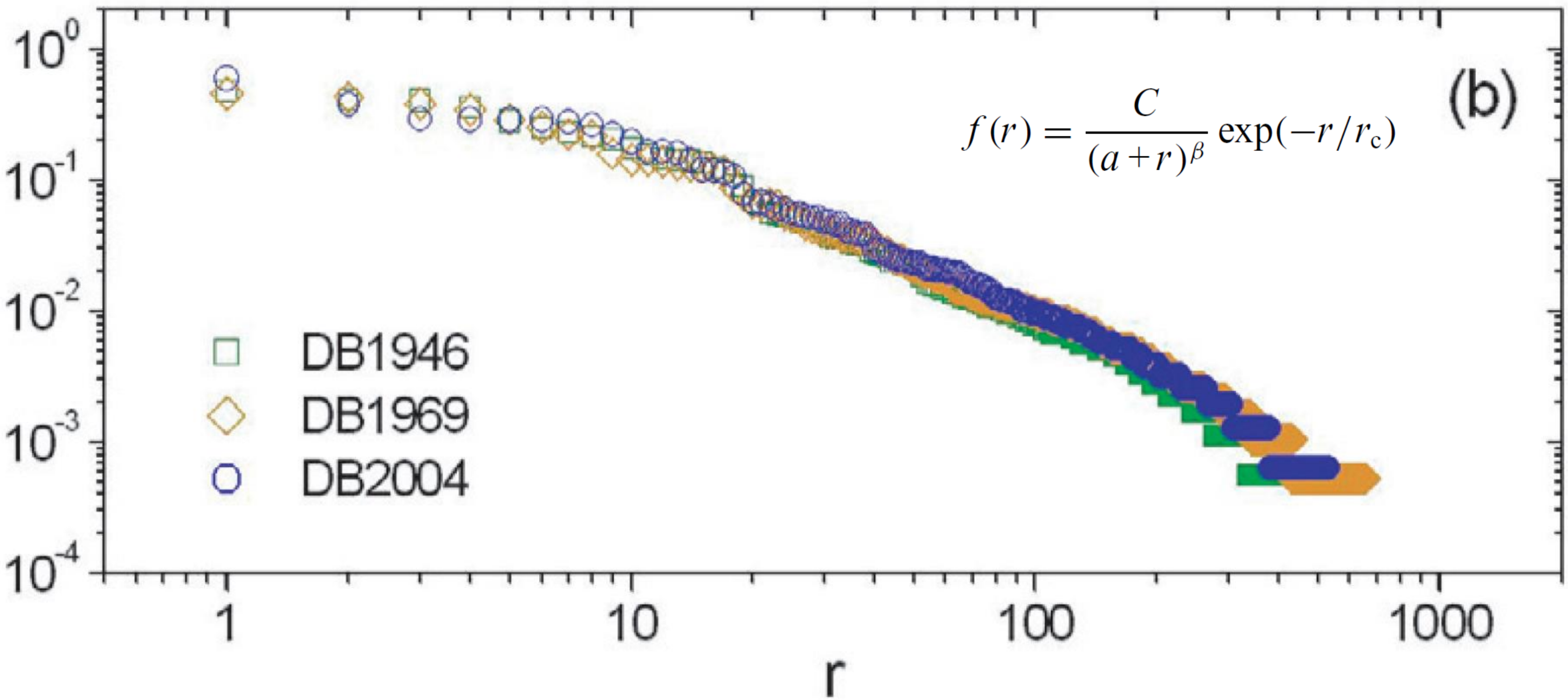
[a]Recipes have been sampled. The total number of recipes is above 3000.

Frequency-Rank Plot



Cultural invariance. Frequency-rank plot for different cookery books: *Pleyn Delit* (circles), *Dona Benta* 1969 (squares), *New Penguin* (stars) and *Larousse Gastronomique* (triangles).

Frequency-Rank Plot



Temporal invariance. Frequency-rank plot for *Dona Benta* 1946 (squares), 1969 (lozenges) and 2004 (circles).

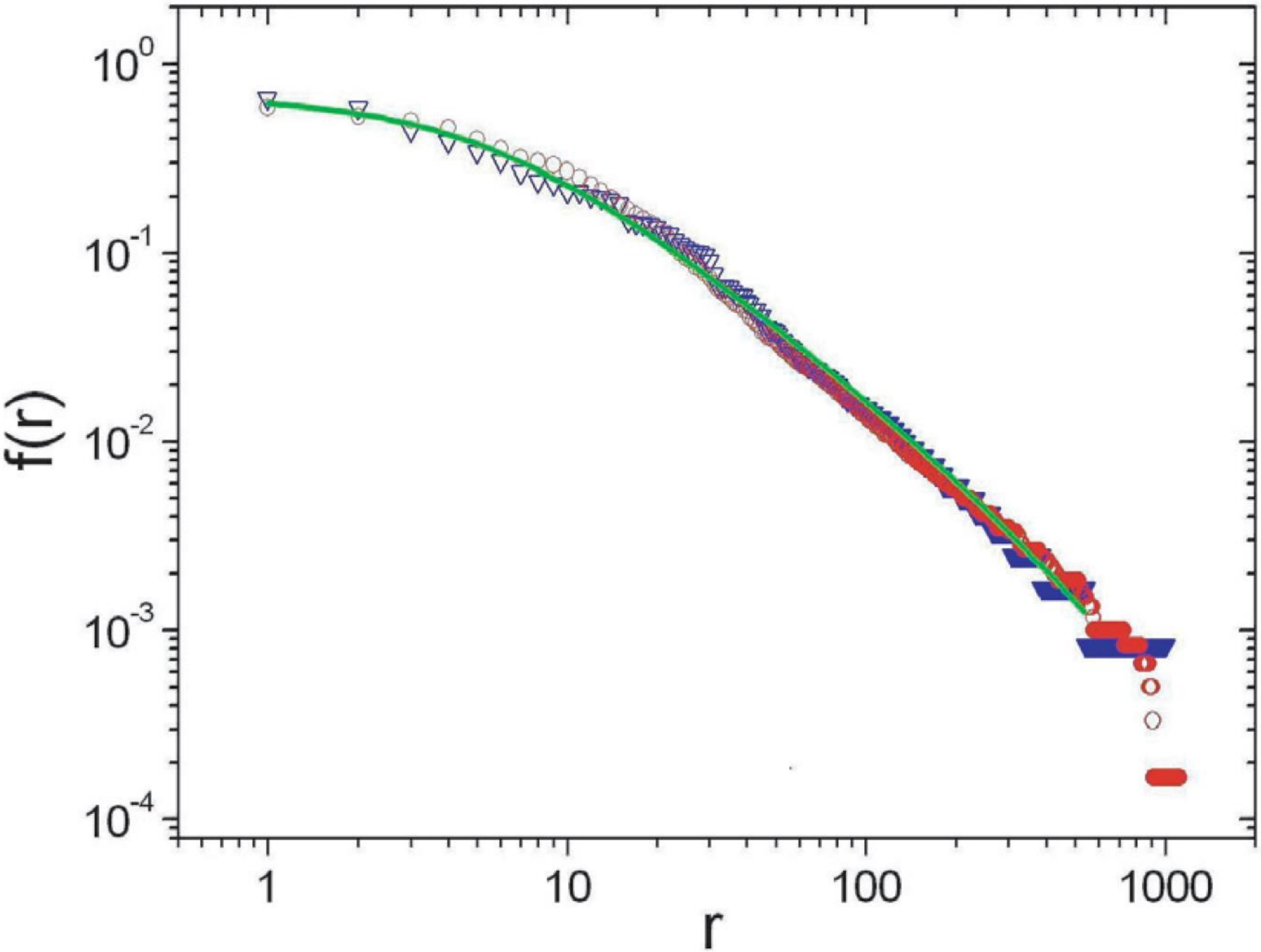
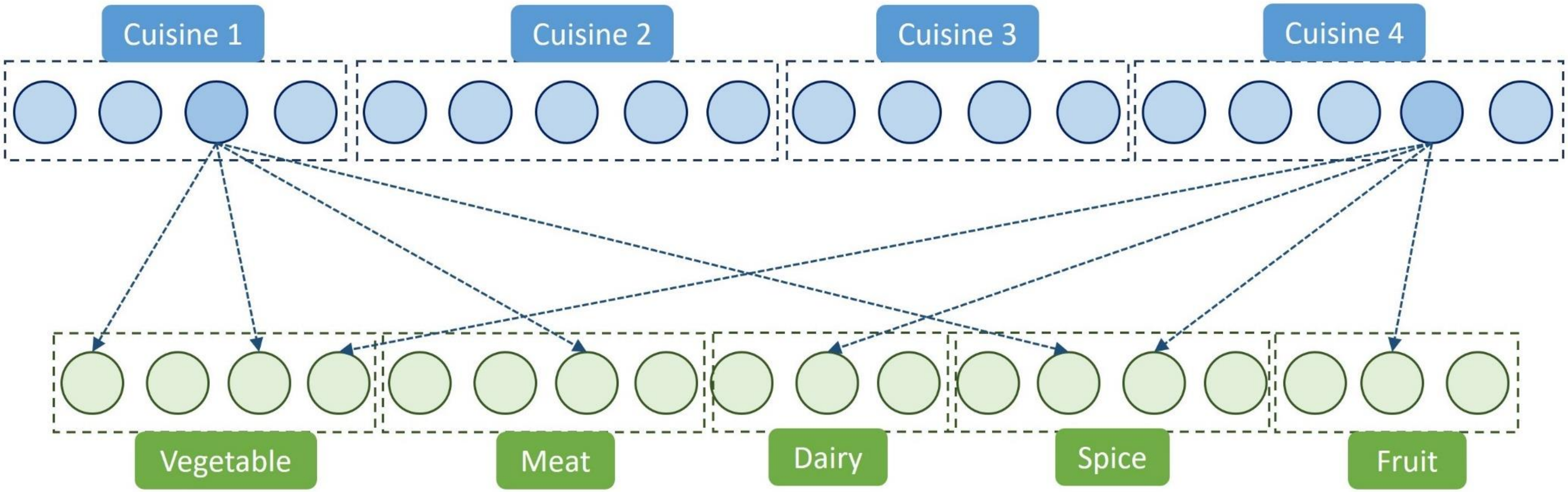


Figure 3. Frequency-rank plot for the Laurousse cookery book (triangles) fitted by a Zipf–Mandelbrot curve (solid) and by the model (circles, average of five runs).

Cuisine as a bipartite network



Cuisine as a bipartite network

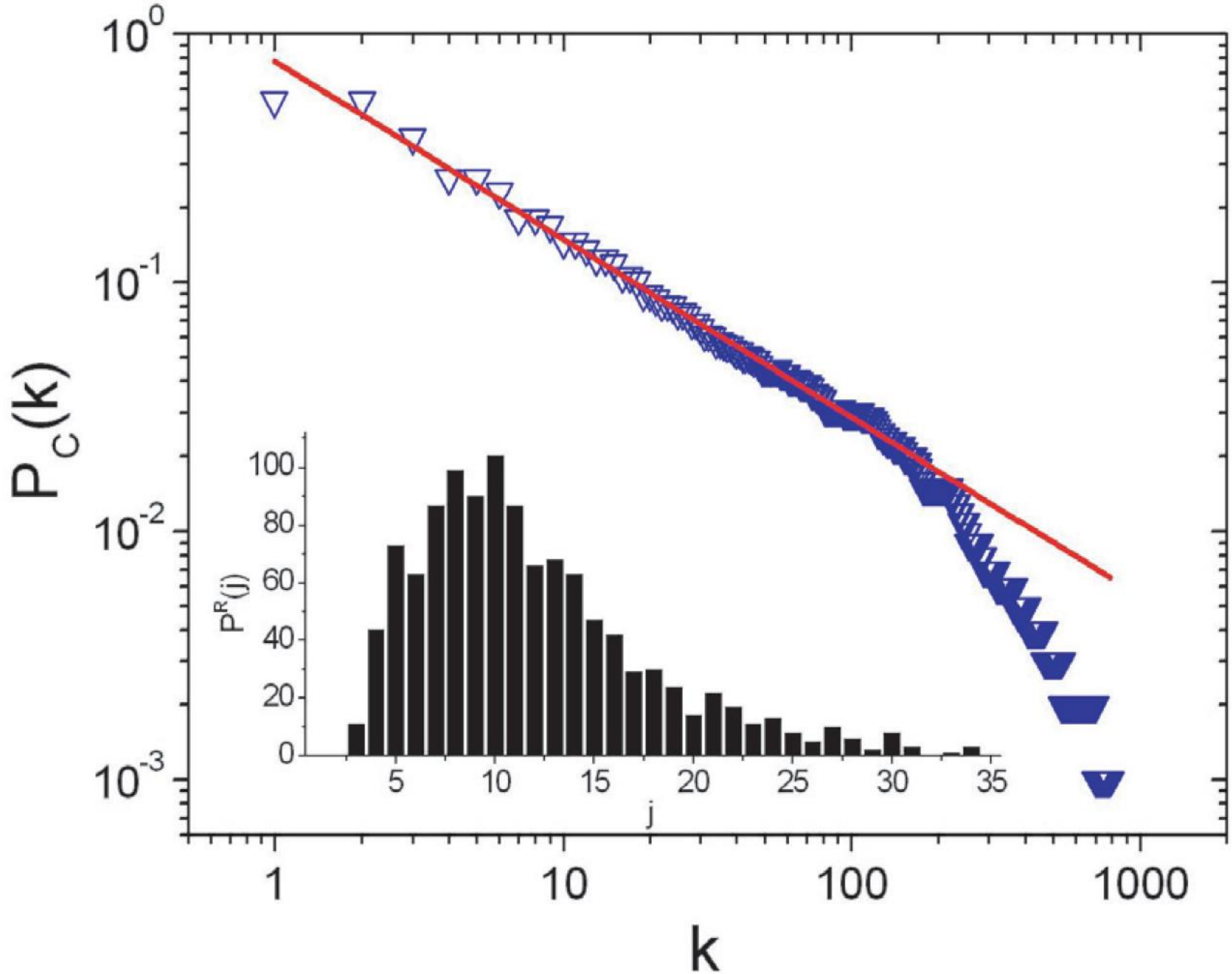
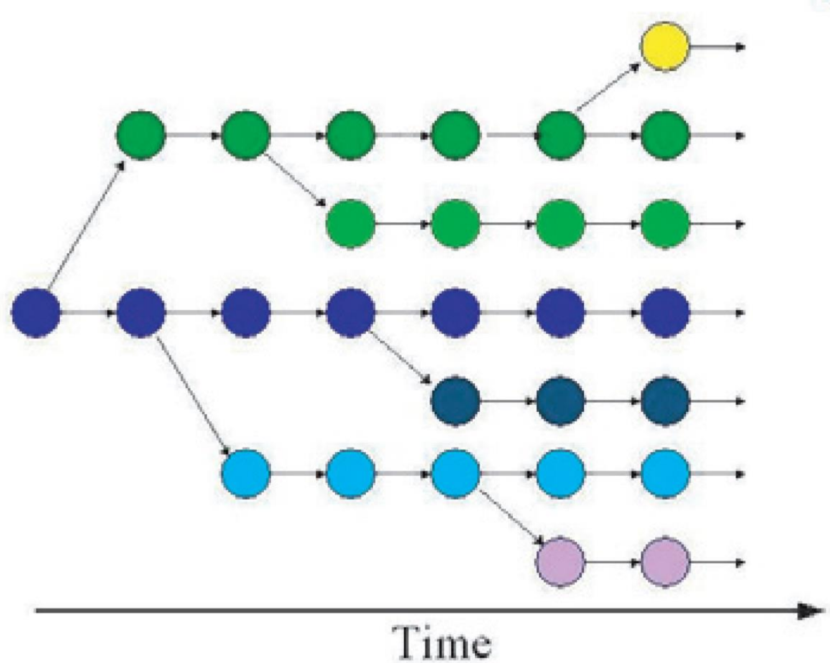


Figure 2. Complementary cumulative degree distribution $P_C(k)$ for ingredients with power law fitting $k^{-0.73}$ (solid line). Inset: degree distribution $P^R(j)$ for French recipes.

A SIMPLE COPY-MUTATE MODEL FOR THE EVOLUTION OF THE CUISINES

On the basis that culinary recipes are examples of cultural replicators, or 'memes' [25, 26], we propose a copy-mutate algorithm to model culinary evolution as a branching process (figure 3 inset). Our attempts to fit all the frequency distributions have shown that our model needs at least five parameters to work: the number T of generations (iterations or 'time') until the process is halted, the number K of ingredients per recipe (where K is compared to the average number $\langle K \rangle$ of the cookery book), the number $L(\neq K)$ of ingredients ('loci') in each recipe to be mutated, the number R_0 of initial recipes in the cuisine and M, the ratio between the sizes of the pool of ingredients and the pool of recipes. Hence, starting with $t_0 = R_0$, at each time step there are $R(t) = t$ recipes and $MR(t)$ ingredients available to be used.

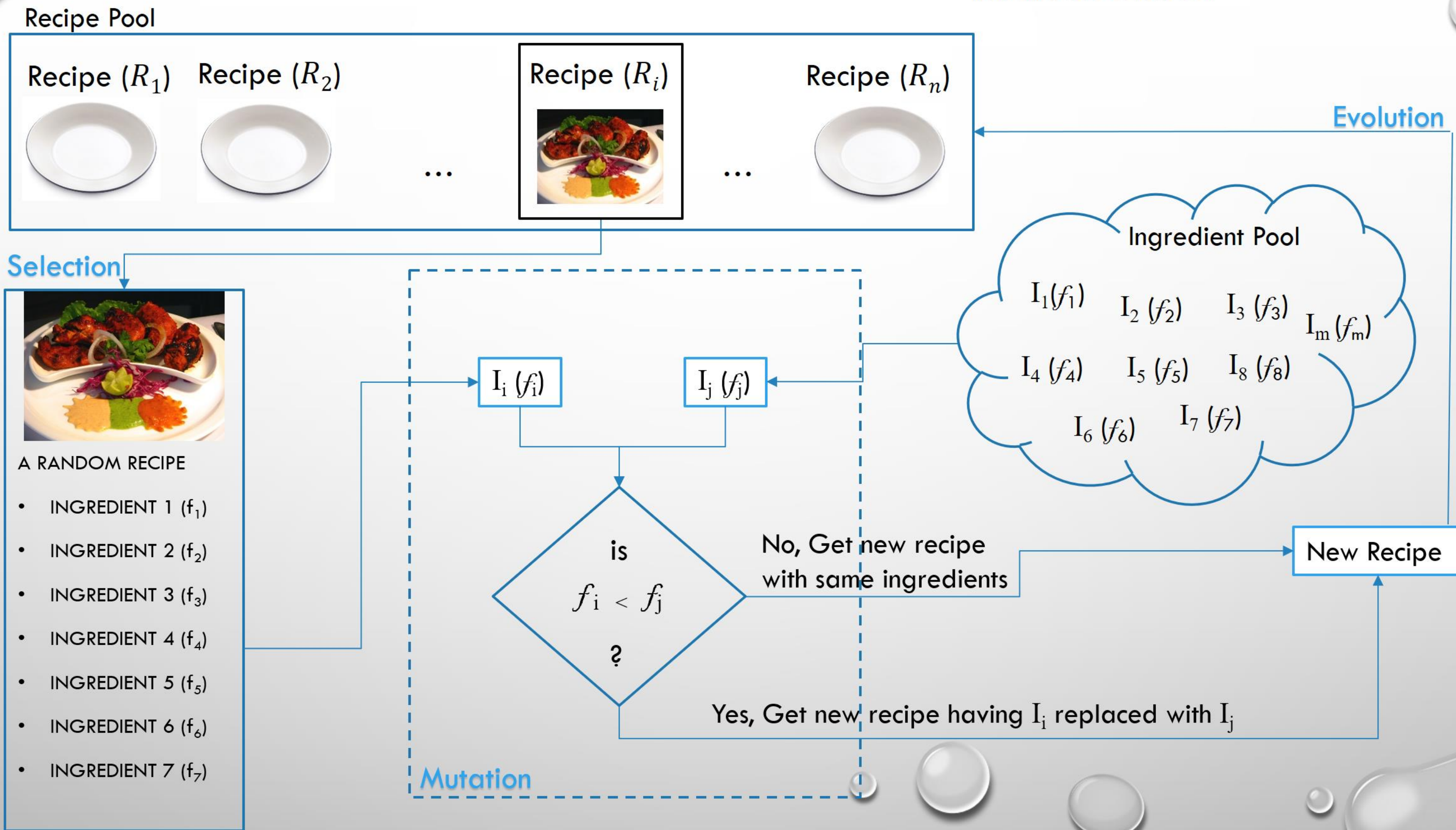


A SIMPLE COPY-MUTATE MODEL FOR THE EVOLUTION OF THE CUISINES

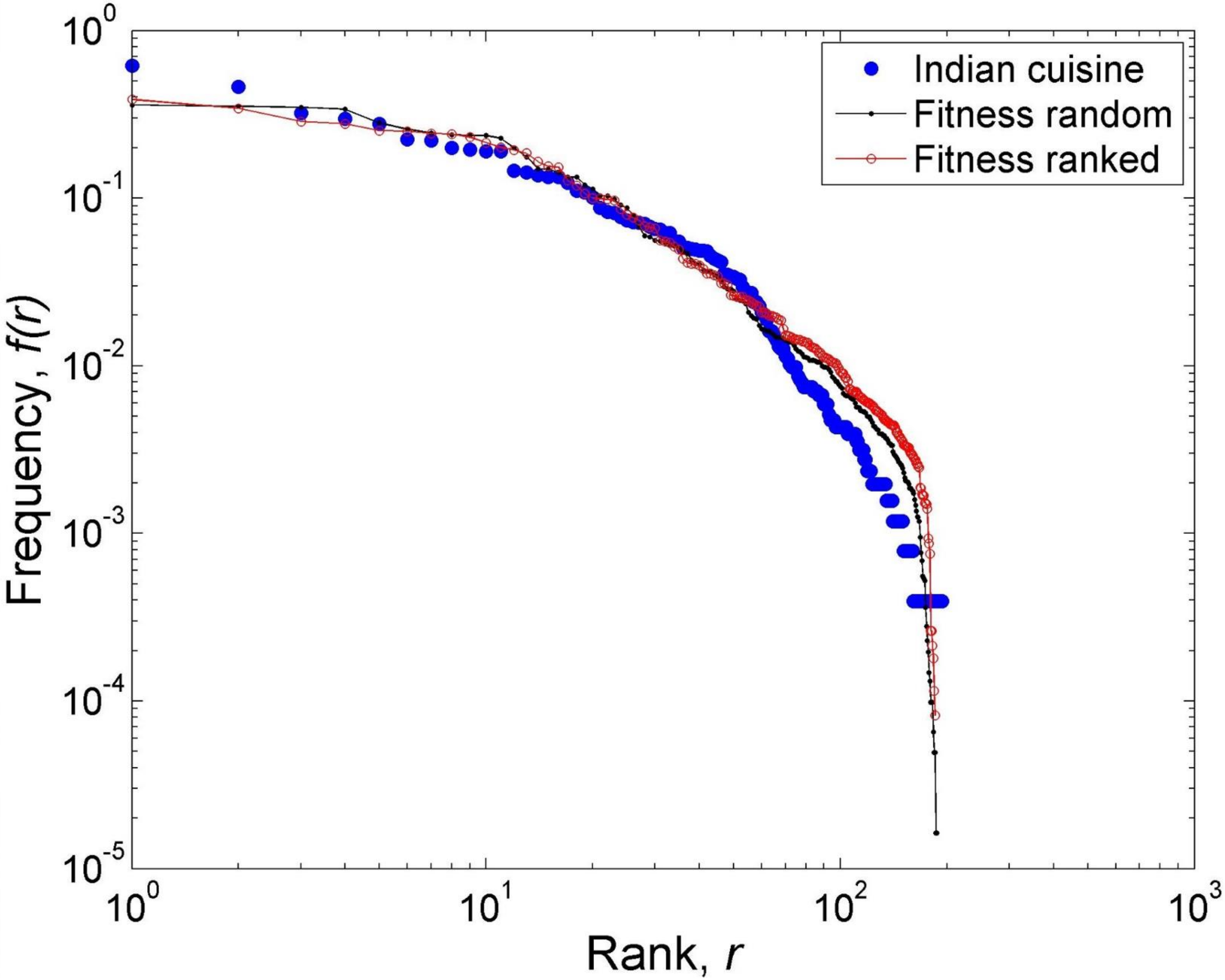
We tried models without the concept of fitness, with no success. So, the present model ascribes to each ingredient i a random fitness f_i with values uniformly distributed in the interval [0,1]. We interpret this fitness as related to intrinsic ingredient properties such as nutritional value, aspect, flavour, versatility, cost and availability. At each iteration, we randomly chose a recipe (a 'mother') and copied it. Within this copy, we randomly chose an ingredient (with fitness f_i) to be compared with an ingredient also randomly chosen from the ingredients pool: if $f_j > f_i$, where f_j is the fitness of the ingredient from the pool, we replace ingredient i by ingredient j.

This process is repeated L times, thus generating a 'daughter' recipe that is added to the recipes pool (it is possible that the daughter remains identical to the mother, which we interpret as a new recipe that differs from the previous one in the cooking procedures but not in their ingredients). Finally, at each time step, new ingredients are introduced in the ingredients pool to maintain the ratio M fixed. A somewhat similar algorithm has been proposed by Ramezanpour [27], but without the introduction of (the very important) fitness-based selection process.

THE COPY-MUTATE MODEL
METHODOLOGY



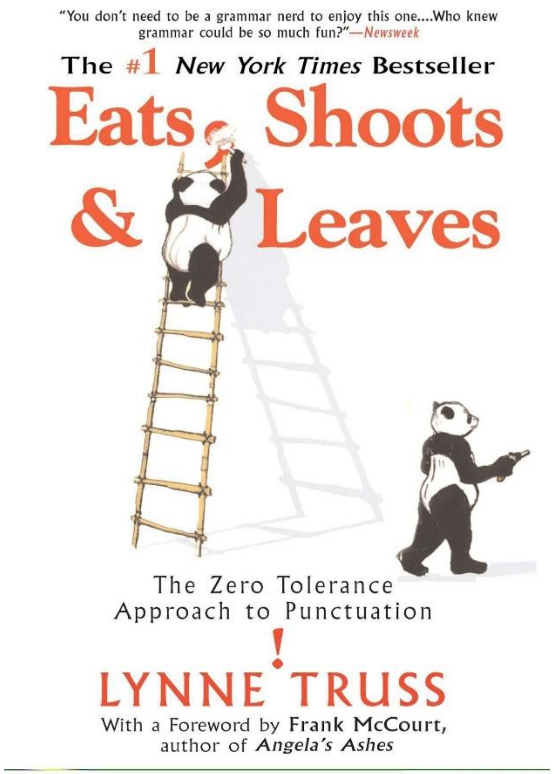
THE COPY-MUTATE MODEL
RESULTS



A SIMPLE COPY-MUTATE MODEL FOR THE EVOLUTION OF THE CUISINES

1. Let's assume that all that the nature had to offer was available as a 'Nature Basket (NB)'. The primordial cuisine used a subset of these ingredients as part of their culinary repertoire—the Kitchen Basket (KB). Here, $KB \subseteq NB$.
2. First, **create a random primordial cuisine** with following parameters.
   - Size of recipes = (say) 10
   - Number of recipes at $t = 0$, $N_R^{t=0} = 500$
   - Size of Kitchen Basket = 50
   - $N_R^{t=0} \gg$ Size of KB
3. Declare Epoch=1
4. Pick a random recipe for 'modification'.
5. Pick an ingredient randomly from the chosen recipe & one from KB compare the ingredients.
6. If the 'KB ingredient' is not the same as 'recipe ingredient' replace the latter with the former.
7. **Repeat 5** until the Recipe gets modified.
8. For every (Final recipes/5, say) **declare the next Epoch** and expand the KB to match with the finally expected (Total Number of recipes/KB)

A SIMPLE COPY-MUTATE MODEL FOR THE EVOLUTION OF THE CUISINES



- Beyond the **panda cuisine**…
- Need to accommodate the **variation in the size of recipes**
- What does the **'fitness of an ingredient'** represent?
- Accommodating the **flavor/food-pairing** aspect.
- Cuisine evolution, where **'ingredient pairs'** are **the units of replacement/deletion**.

Computational models for the evolution of world cuisines

Rudraksh Tuwani[1†], Nutan Sahoo[1,2†], Navjot Singh[1] and Ganesh Bagler[1*]

[1]*Center for Computational Biology, Indraprastha Institute of Information Technology (IIIT-Delhi), New Delhi, India*

[2]*Sri Venkateswara College, University of Delhi, New Delhi, India*

[*]*Corresponding author: Ganesh Bagler,* bagler@iiitd.ac.in

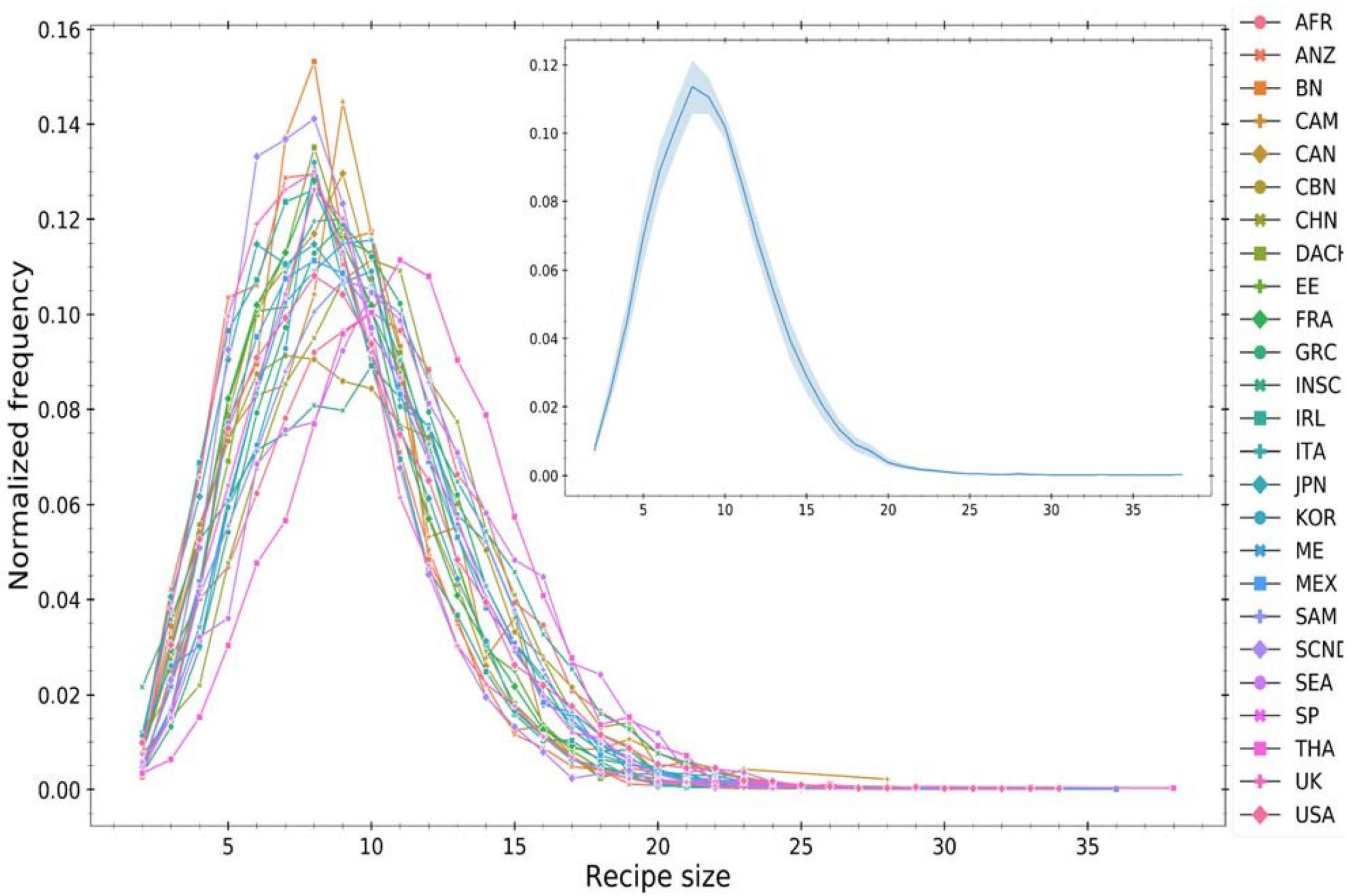
Cooking is a unique endeavor that forms the core of our cultural identity. Culinary systems across the world have evolved over a period of time in the backdrop of complex interplay of diverse sociocultural factors including geographic, climatic and genetic influences. Data-driven investigations can offer interesting insights into the structural and organizational principles of cuisines.

Herein, we use a comprehensive repertoire of 158544 recipes from 25 geo-cultural regions across the world to investigate the statistical patterns in usage of ingredients and their categories. Further, we develop computational models for the evolution of cuisines. Our analysis reveals copy-mutation as a plausible mechanism of culinary evolution. As the world copes with the challenges of diet-linked disorders, knowledge of the key determinants of culinary evolution can drive the creation of novel recipe generation algorithms aimed at dietary interventions for better nutrition and health.

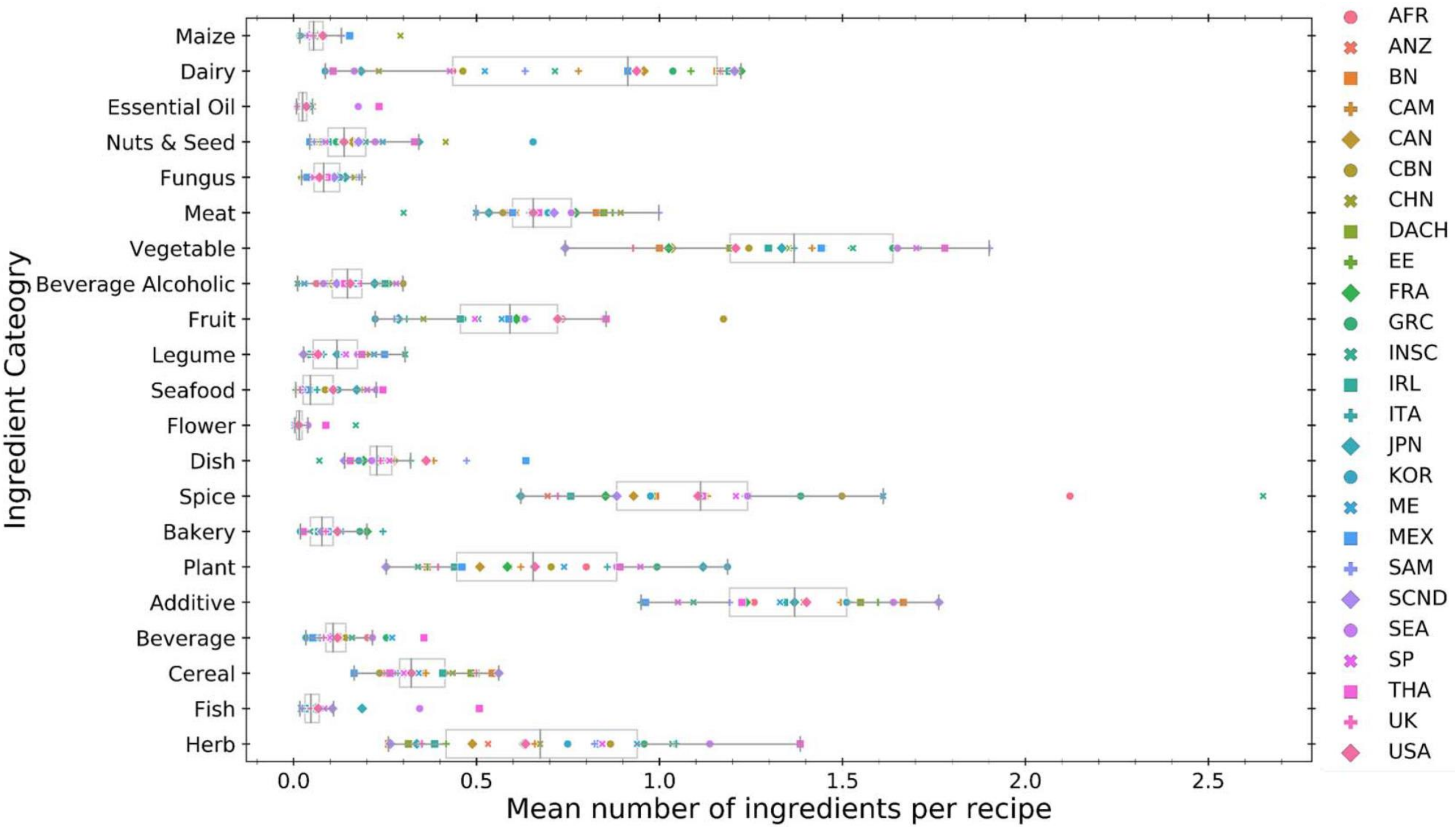
TABLE I. STATISTICS OF NUMBER OF RECIPES AND INGREDIENTS AS WELL AS TOP 5 OVERREPRESENTED INGREDIENTS IN EACH WORLD CUISINE.

Region (Code)	Recipes	Ingredients	Overrepresented Ingredients
Africa (AFR)	5465	442	Cumin, Cinnamon, Olive, Cilantro, Paprika
Australia & NZ (ANZ)	6169	463	Butter, Egg, Sugar, Flour, Coconut
Republic of Ireland (IRL)	2702	378	Potato, Butter, Cream, Flour, Baking Powder
Canada (CAN)	7725	483	Baking Powder, Sugar, Butter, Flour, Vanilla
Caribbean (CBN)	3887	417	Lime, Rum, Pineapple, Allspice, Thyme
China (CHN)	7123	442	Soybean Sauce, Sesame, Ginger, Corn, Chicken
DACH Countries (DACH)	4641	430	Flour, Egg, Butter, Sugar, Swiss Cheese
Eastern Europe (EE)	3179	383	Flour, Egg, Butter, Cream, Salt
France (FRA)	9590	511	Butter, Egg, Vanilla, Milk, Cream
Greece (GRC)	5286	405	Olive, Feta Cheese, Oregano, Lemon juice, Tomato
Indian Subcontinent (INSC)	10531	462	Cayenne, Turmeric, Cumin, Cilantro, Ginger, Garam Masala
Italy (ITA)	23179	506	Olive, Parmesan Cheese, Basil, Garlic, Tomato
Japan (JPN)	2884	382	Soybean sauce, Sesame, Ginger, Vinegar, Sake
Korea (KOR)	1228	291	Sesame, Soybean sauce, garlic, Sugar, Ginger
Mexico (MEX)	16065	467	Tortilla, Cilantro, Lime, Cumin, Tomato
Middle East (ME)	4858	423	Olive, Lemon juice, Parsley, Cumin, Mint
Scandinavia (SCND)	3026	377	Sugar, Flour, Butter, Egg, Milk
South America (SAM)	7458	457	Beef, Onion, Pepper, Garlic, Mushroom
South East Asia (SEA)	2523	361	Fish, Sugar, Soybean sauce, Garlic, Lime
Spain (SP)	4154	413	Olive, Paprika, Garlic, Tomato, Parsley
Thailand (THA)	3795	378	Fish, Lime, Cilantro, Coconut Milk, Soybean sauce
USA (USA)	16026	592	Butter, Sugar, Vanilla, Flour, Mustard
Belgium-Netherlands (BN)	1116	323	Butter, Flour, Egg, Sugar, Milk
Central America (CAM)	470	294	Salt, Tomato, Onion, Macaroni, Celery
United Kingdom (UK)	5380	456	Butter, Flour, Egg, Sugar, Milk

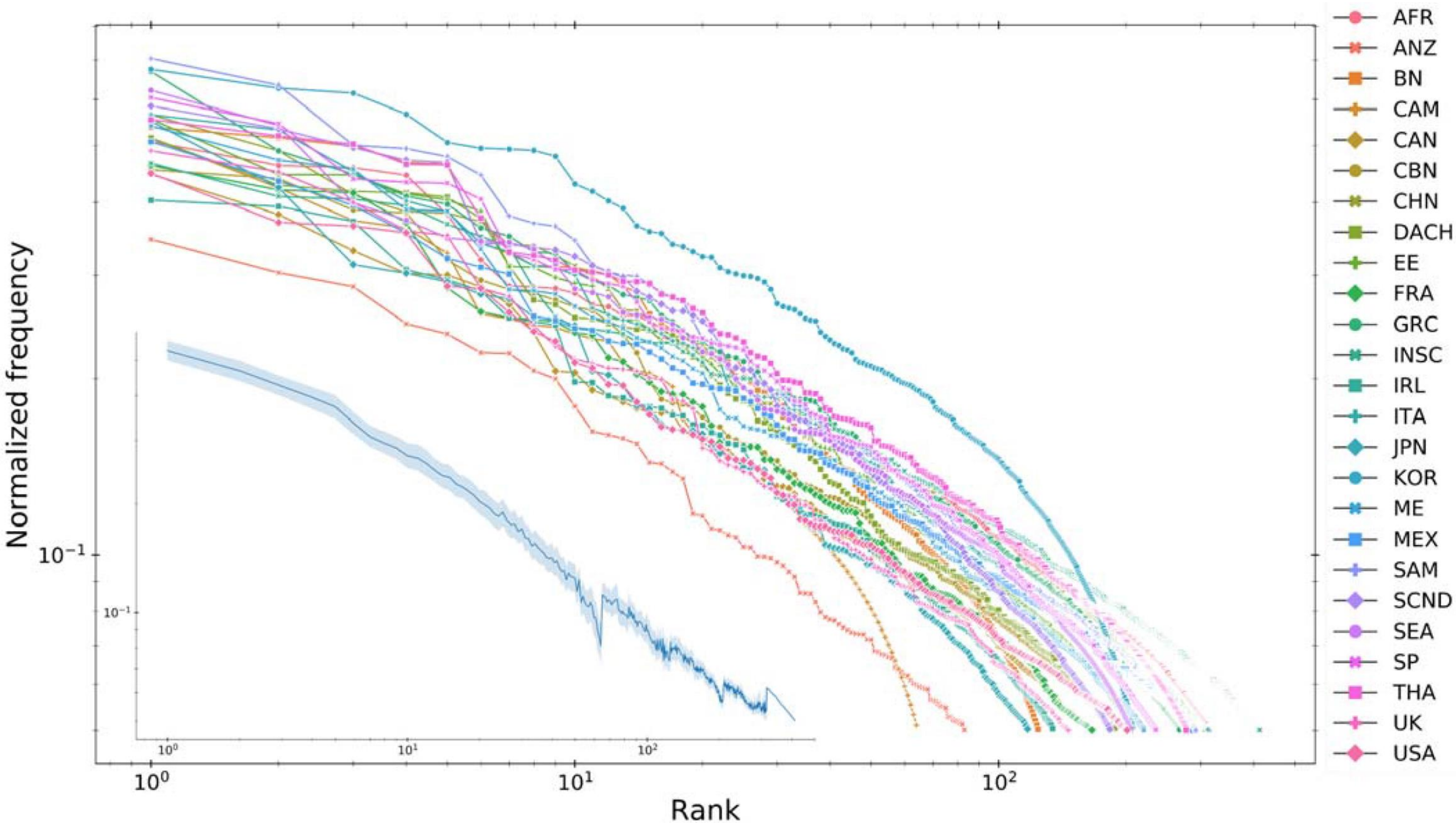
Computational models for the evolution of the world cuisines



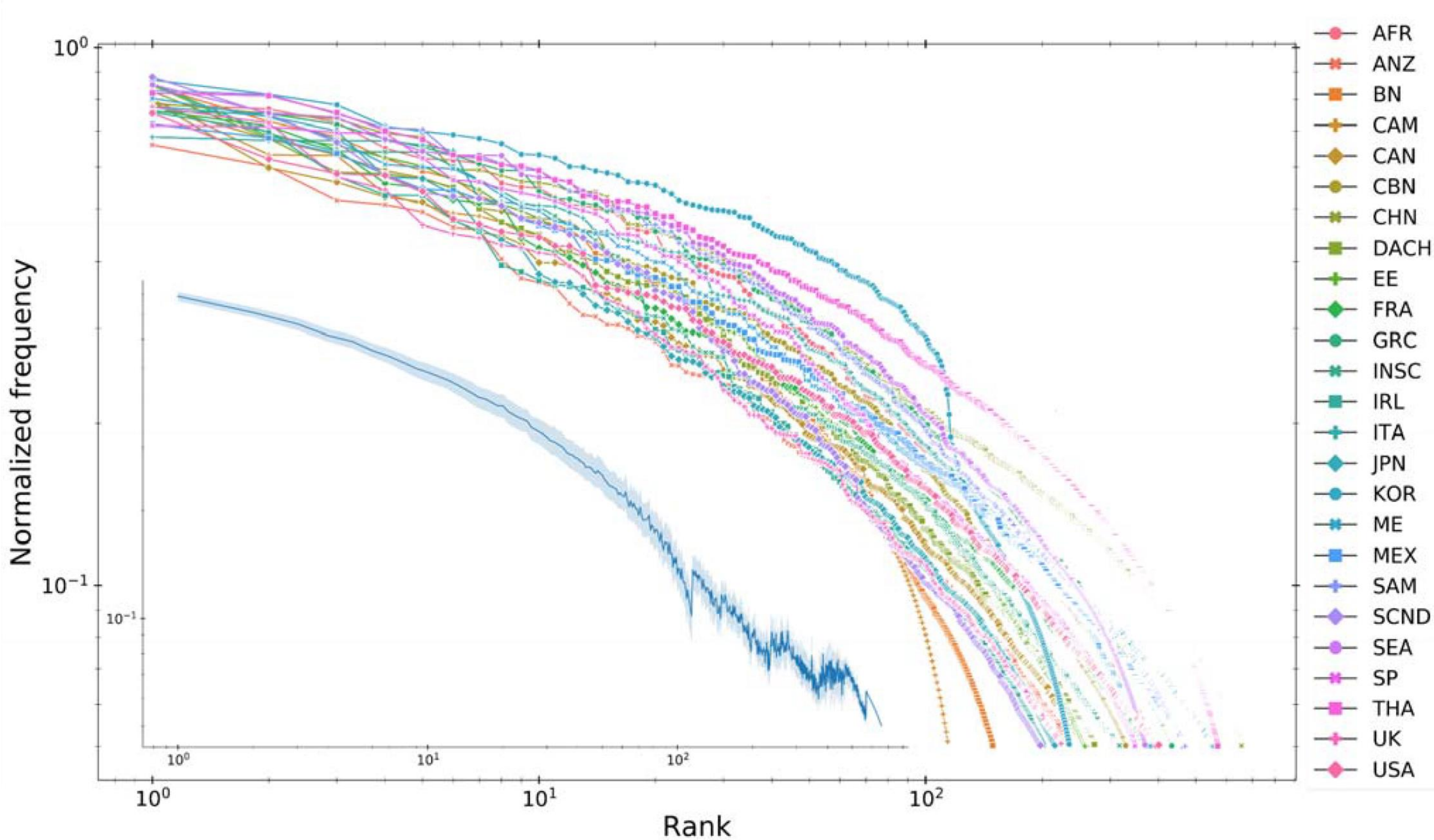
Computational models for the evolution of the world cuisines



Computational models for the evolution of the world cuisines



Computational models for the evolution of the world cuisines



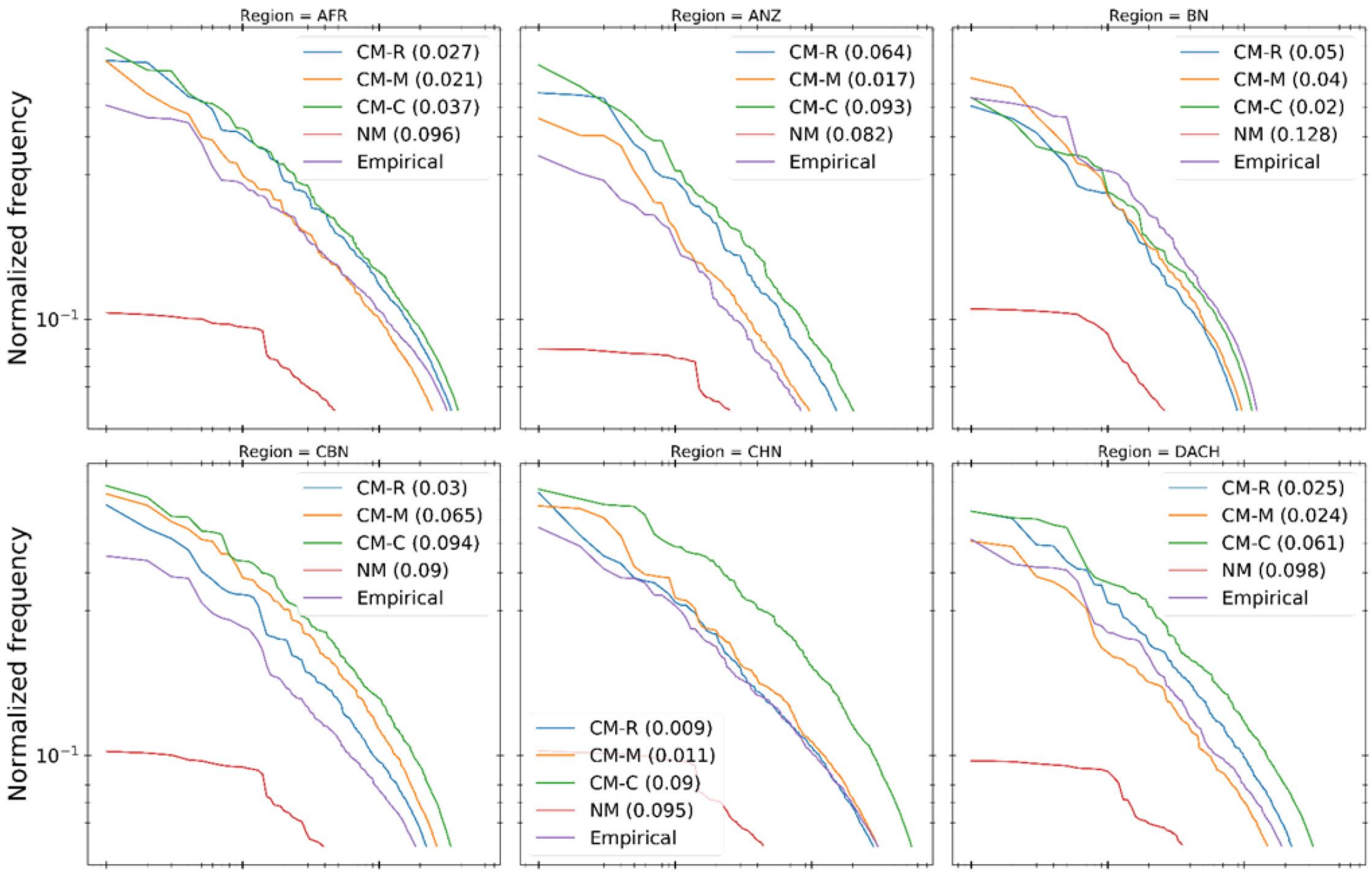
Algorithm 1 Algorithm for copy-mutate model

Input: List of ingredients in a cuisine (I), average recipe size of a cuisine ($\bar{s}$), size of initial recipe pool (n), size of initial ingredient pool (m), total number of recipes in cuisine (N), number of mutations (M), and ratio of the total number of ingredients to the total number of recipes in the cuisine ($\emptyset$).

Output: N mutated recipes

for all ingredients i in I **do**
  sample a value from $Uniform\ (0,1)$
  assign it to i
end for
$I_0 \leftarrow$ randomly sample (without replacement) m ingredients from I
$I \leftarrow I - I_0$
$R_0 \leftarrow$ randomly sample $\bar{s}$ ingredients n times from I_0
for $l = 1$ to $N - n$ **do**
  $\partial \leftarrow m/n$
  if $\partial \geq \emptyset$ **then**
    $r \leftarrow$ randomly choose a recipe from R_0
    for $g = 1$ to M **do**
      sample an ingredient i from r
      sample an ingredient j from I_0
      if fitness of $j >$ fitness of i **then**
        replace i with j in r
      end if
    end for
    $R_0 \leftarrow R_0 + r$
    $n \leftarrow n + 1$
  else
    choose an ingredient p randomly from I
    $I_0 \leftarrow I_0 + p$
    $m \leftarrow m + 1$
    $I \leftarrow I - p$
  end if
end for

Computational models for the evolution of the world cuisines



Ideas

- **Collecting/scraping data (from websites/books) and making a structured data repository**:

- Preservatives used in processed food
- Food Additives
- Fermented food
- Cocktails/Mocktails
- Allergic foods and allergens
- Nutritional profile of (say) Indian food (NIN)
- Fortified food (list and statistics)
- *Think of other sources of culinary data.*