SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:

STATISTICAL PHYSICS, THERMODYNAMICS AND NONLINEAR DYNAMICS

APPLIED PHYSICS

SYSTEMS BIOLOGY

STATISTICS

Flavor network and the principles of food pairing

Yong-Yeol Ahn[1,2,3]*, Sebastian E. Ahnert[1,4]*, James P. Bagrow[1,2] & Albert-László Barabási[1,2]

[1]Center for Complex Network Research, Department of Physics Northeastern University, Boston, MA 02115, [2]Center for Cancer Systems Biology Dana-Farber Cancer Institute, Harvard University, Boston, MA 02115, [3]School of Informatics and Computing Indiana University, Bloomington, IN 47408, [4]Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

Introduction

- **Are there any general patterns that determine the ingredient combinations used in food today or principles that transcend individual tastes and recipes?**
- A flavor network that captures the flavor compounds shared by culinary ingredients
- **Food Paring Hypothesis**: “Ingredients sharing flavor compounds are more likely to taste well together than ingredients that do not”
- A systematic understanding of culinary practice.

    Number of recipes (cookpad): 10^6

    Number of potential recipes: > 10^{30}

    Number of recipes used for the studies: 56498

Summary

- The flavor network allows us to reformulate the food pairing hypothesis as a topological property: **Do we more frequently use ingredient pairs that are strongly linked in the flavor network or do we avoid them?**
- To test this hypothesis we need data on ingredient combinations preferred by humans, information readily available in the current body of recipes.
- For generality, 56,498 recipes, provided by two American repositories (epicurious.com and allrecipes.com), were used.
- To avoid a distinctly Western interpretation of the world's cuisine, we also used a Korean repository (menupan.com).
- The recipes are grouped into geographically distinct cuisines (North American, Western European, Southern European, Latin American, and East Asian).

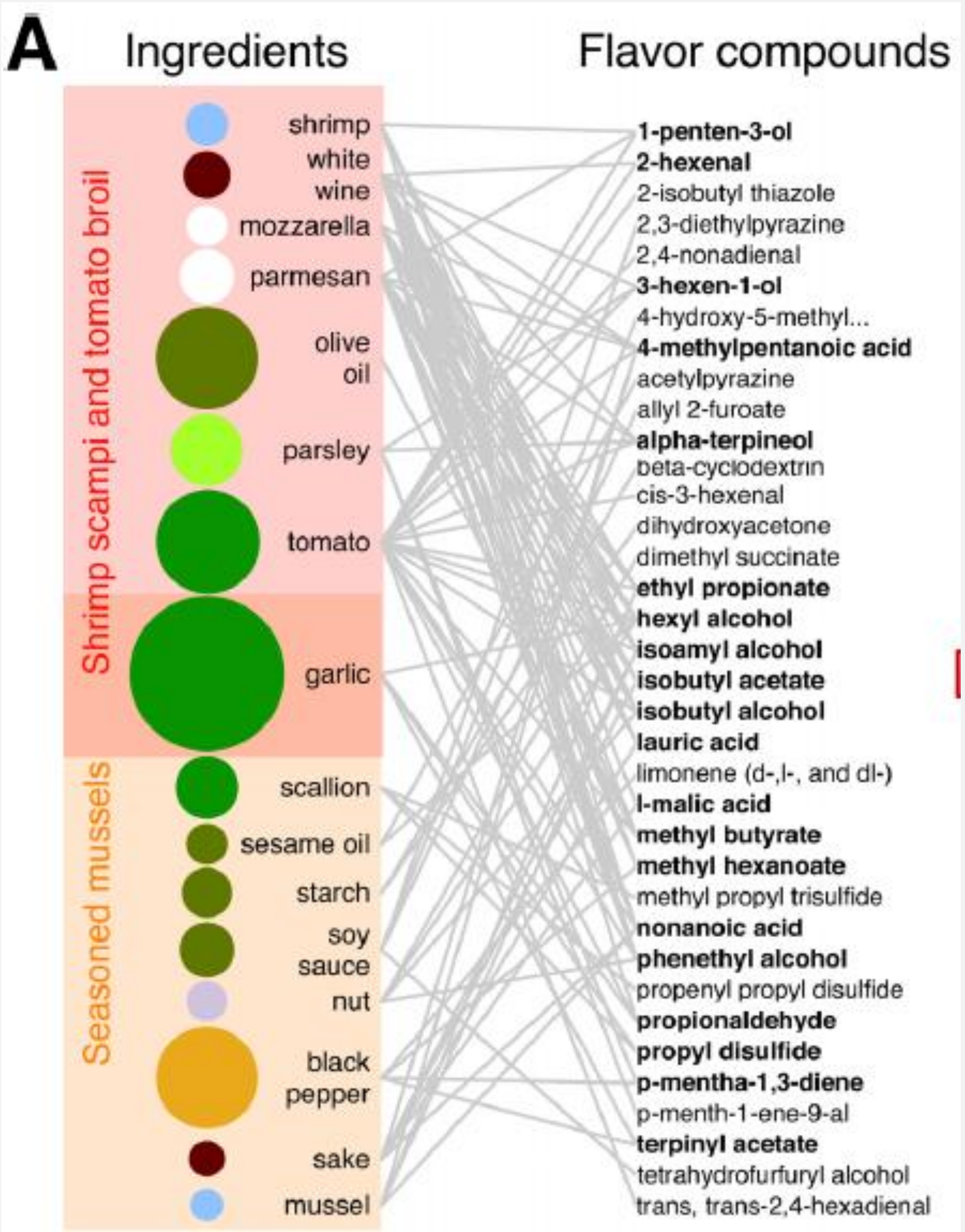
Basis of Flavor Network:

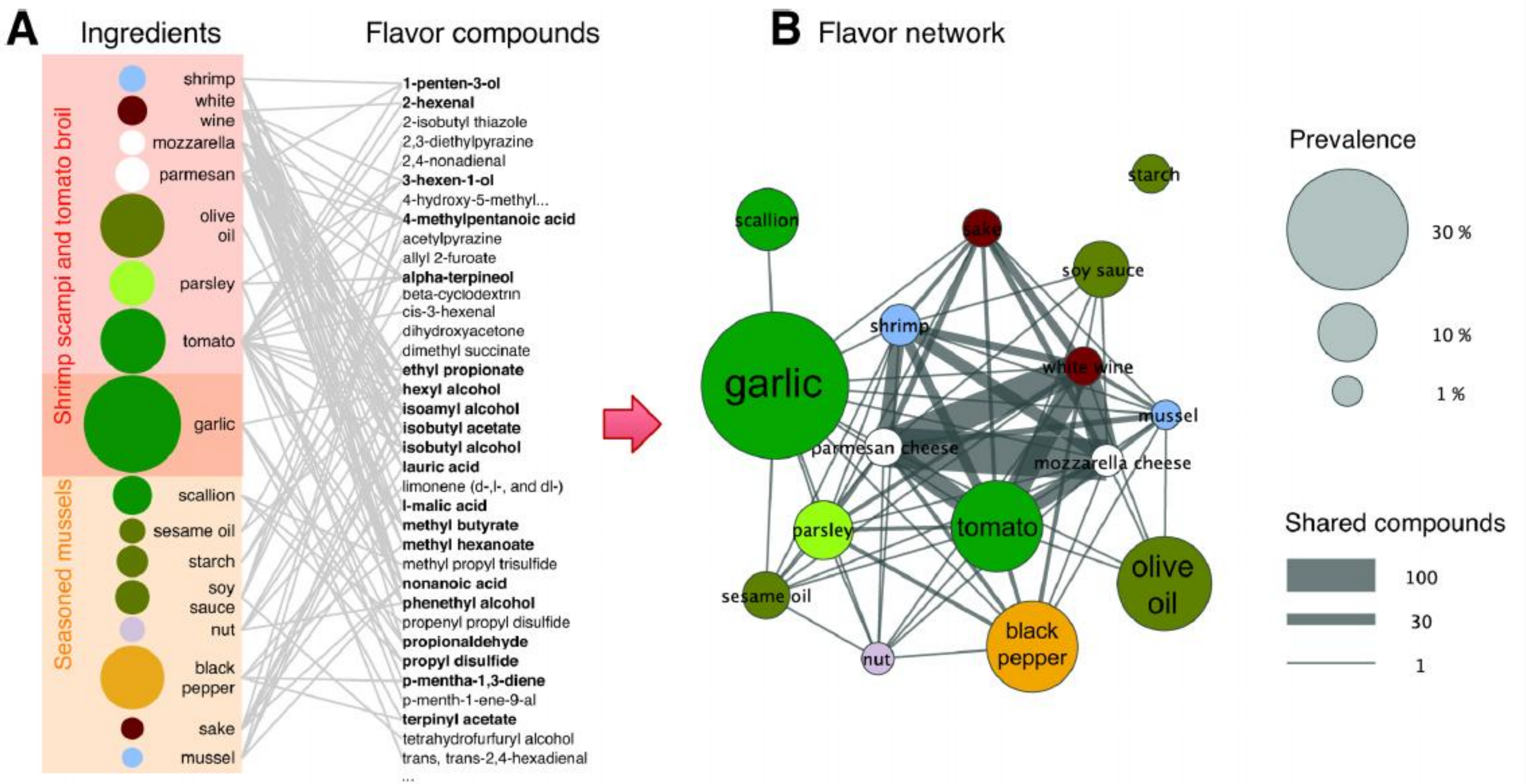
(i) 381 ingredients used in recipes throughout the world, and

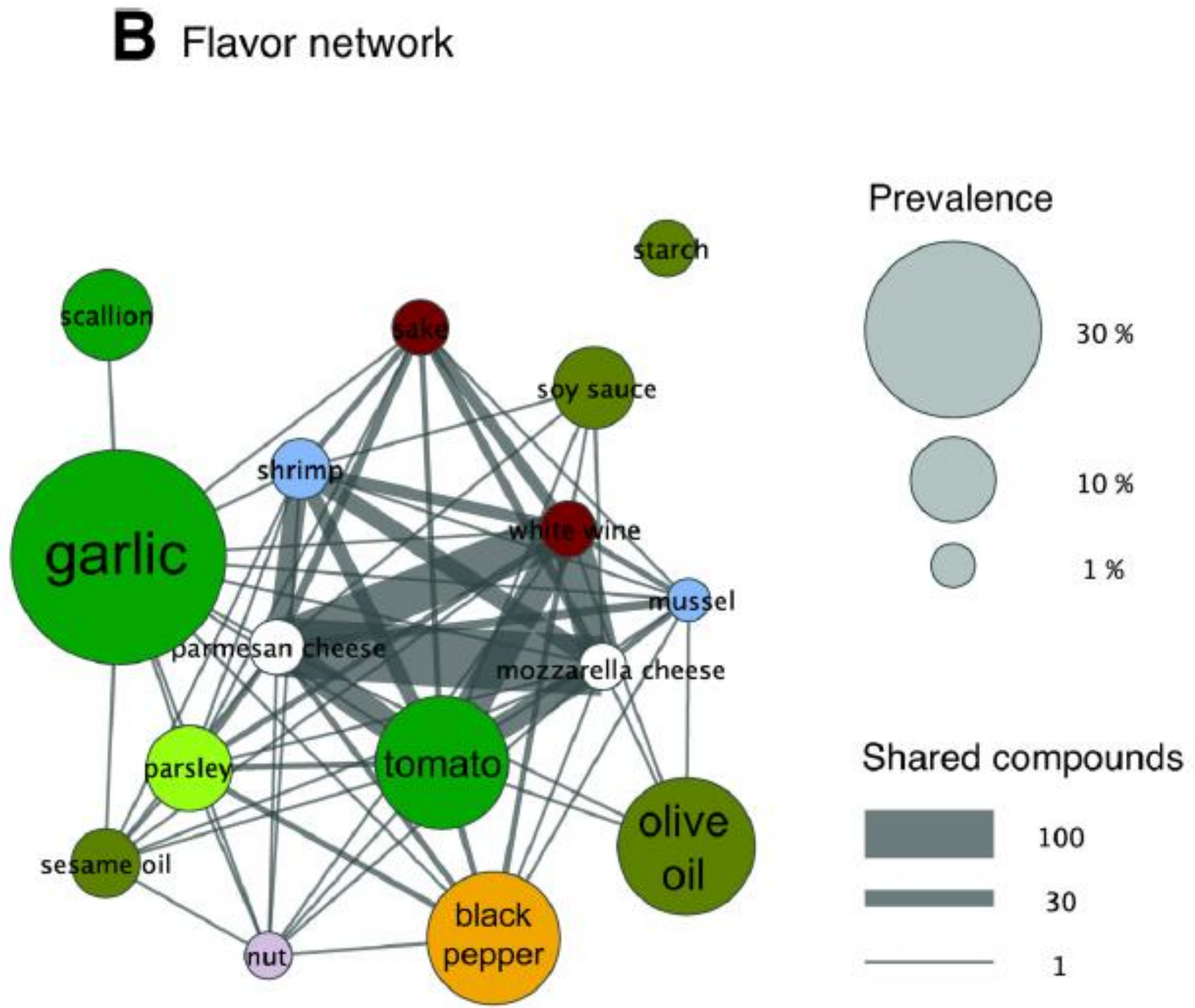
(ii) 1,021 flavor compounds that are known to contribute to the flavor of each of these ingredients.

<k> = 214

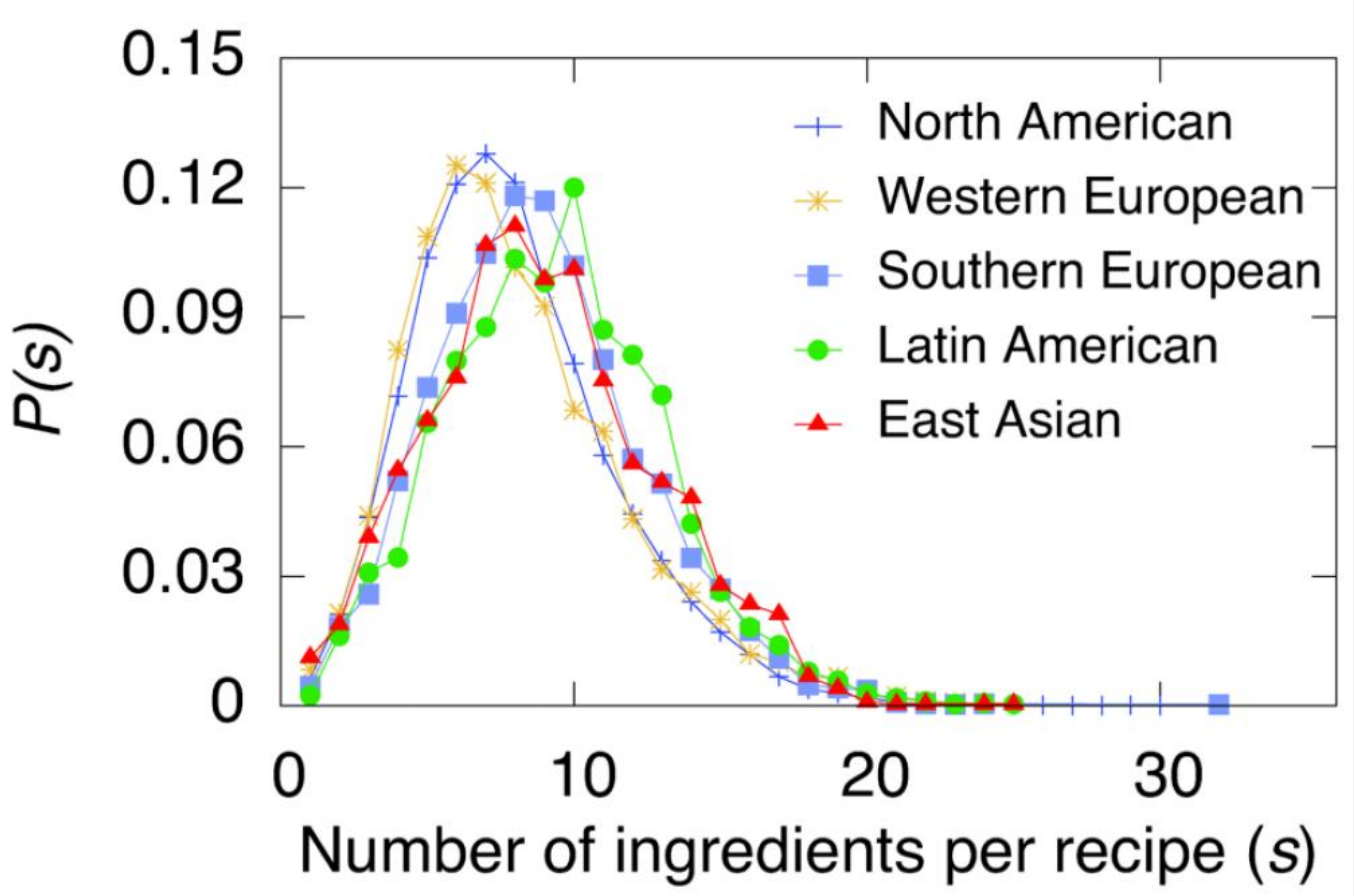
www.epicurious.com
www.allrecipes.com



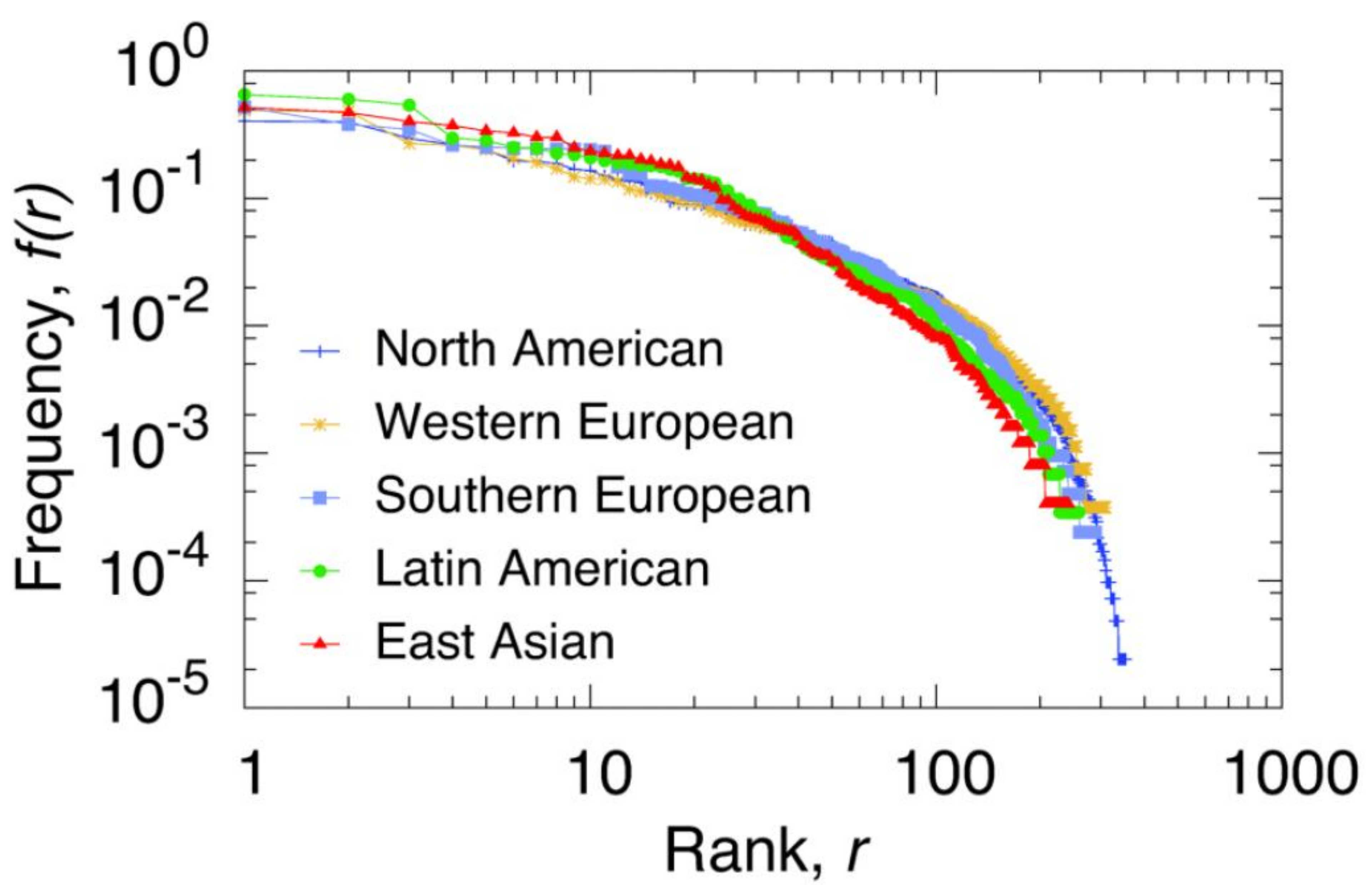

A
Ingredients
Flavor compounds
Shrimp scampi and tomato broil
Seasoned mussels
shrimp
white wine
mozzarella
parmesan
olive oil
parsley
tomato
garlic
scallion
sesame oil
starch
soy sauce
nut
black pepper
sake
mussel
1-penten-3-ol
2-hexenal
2-isobutyl thiazole
2,3-diethylpyrazine
2,4-nonadienal
3-hexen-1-ol
4-hydroxy-5-methyl...
4-methylpentanoic acid
acetylpyrazine
allyl 2-furoate
alpha-terpineol
beta-cyclodextrin
cis-3-hexenal
dihydroxyacetone
dimethyl succinate
ethyl propionate
hexyl alcohol
isoamyl alcohol
isobutyl acetate
isobutyl alcohol
lauric acid
limonene (d-,l-, and dl-)
l-malic acid
methyl butyrate
methyl hexanoate
methyl propyl trisulfide
nonanoic acid
phenethyl alcohol
propenyl propyl disulfide
propionaldehyde
propyl disulfide
p-mentha-1,3-diene
p-menth-1-ene-9-al
terpinyl acetate
tetrahydrofurfuryl alcohol
trans, trans-2,4-hexadienal
...
B
Flavor network
starch
scallion
sake
soy sauce
shrimp
white wine
garlic
mussel
parmesan cheese
mozzarella cheese
parsley
tomato
olive oil
sesame oil
black pepper
nut
Prevalence
30 %
10 %
1 %
Shared compounds
100
30
1

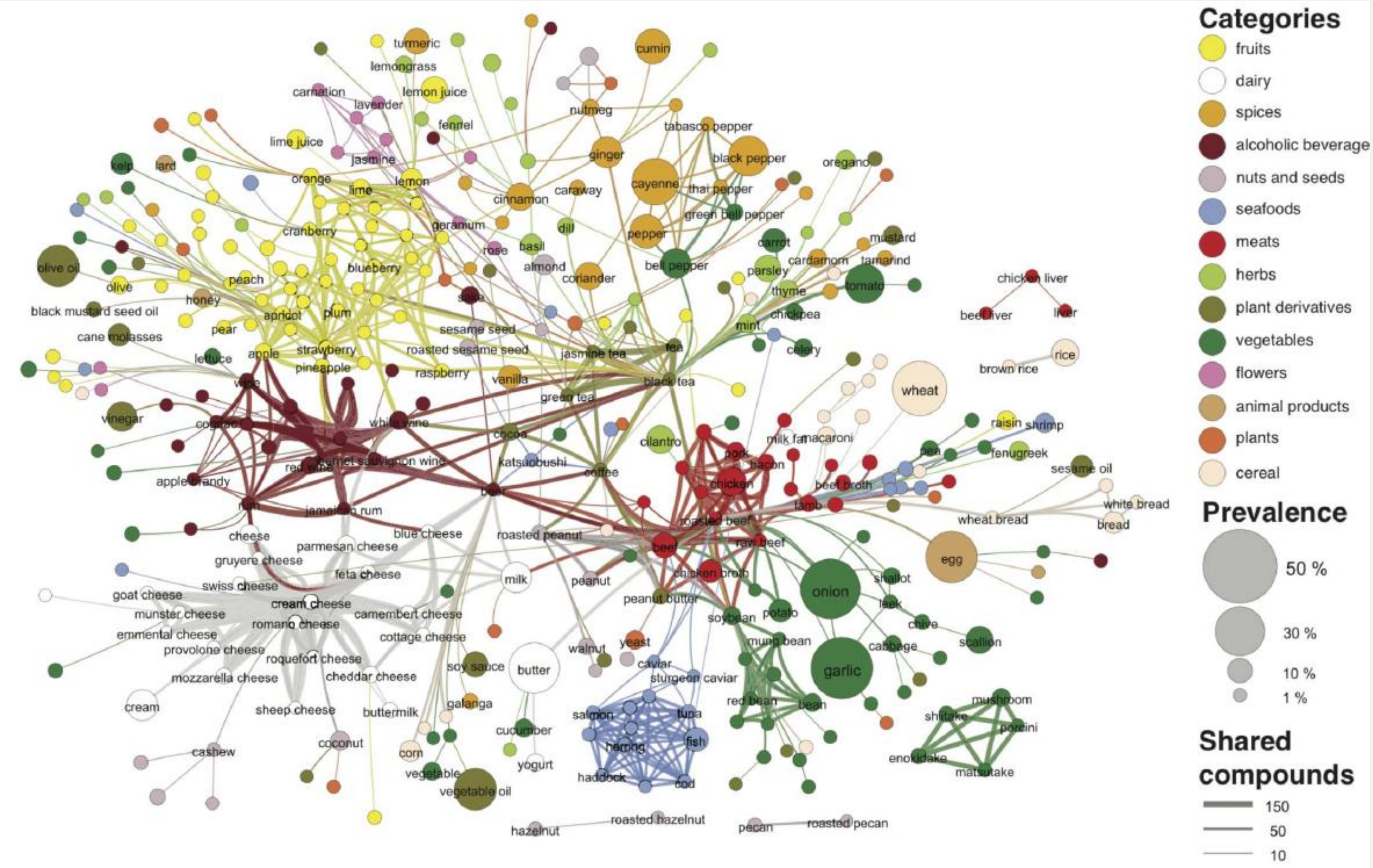

B Flavor network
starch
scallion
sake
soy sauce
shrimp
white wine
garlic
mussel
parmesan cheese
mozzarella cheese
parsley
tomato
olive oil
sesame oil
black pepper
nut
Prevalence
30 %
10 %
1 %
Shared compounds
100
30
1

The average number of ingredients used in a recipe is around eight, and the overall distribution is bounded, indicating that recipes with a very large or very small number of ingredients are rare.



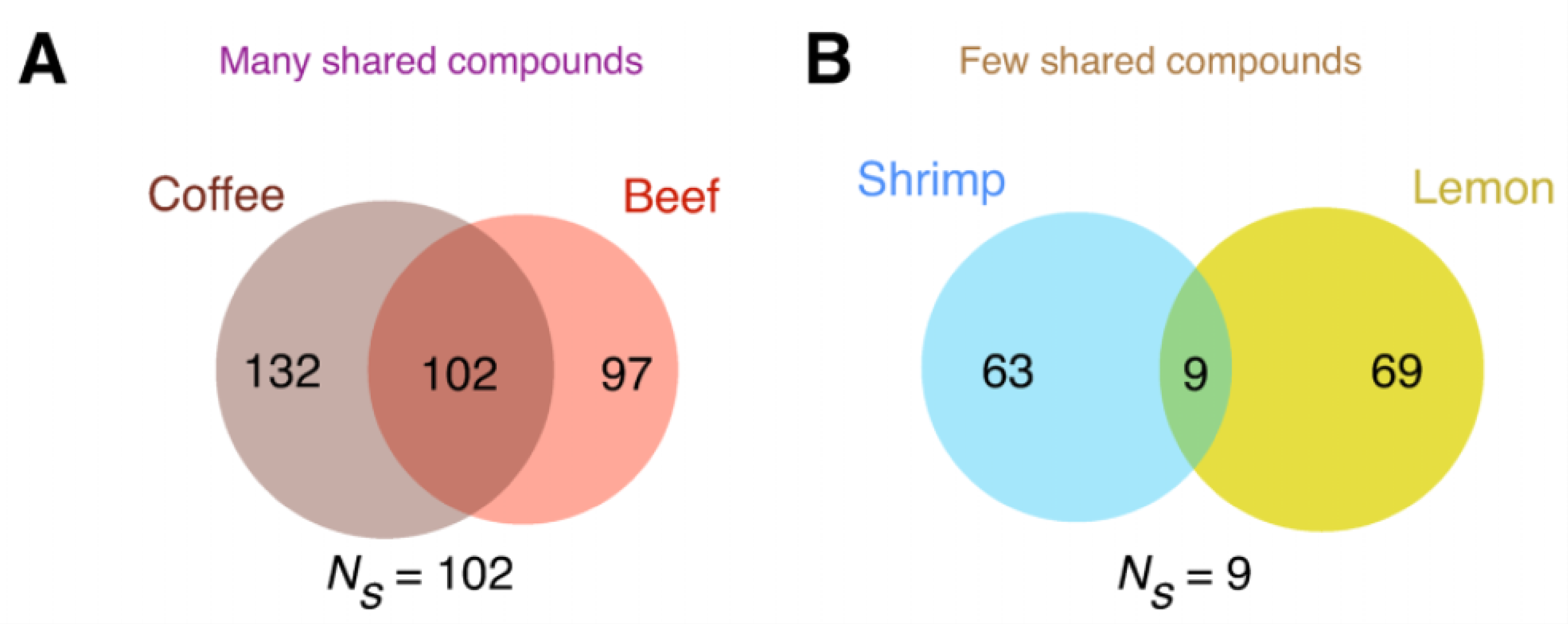
Ingredients Jasmine tea, Jamaican rum, and 14 other ingredients are used only in one recipe.





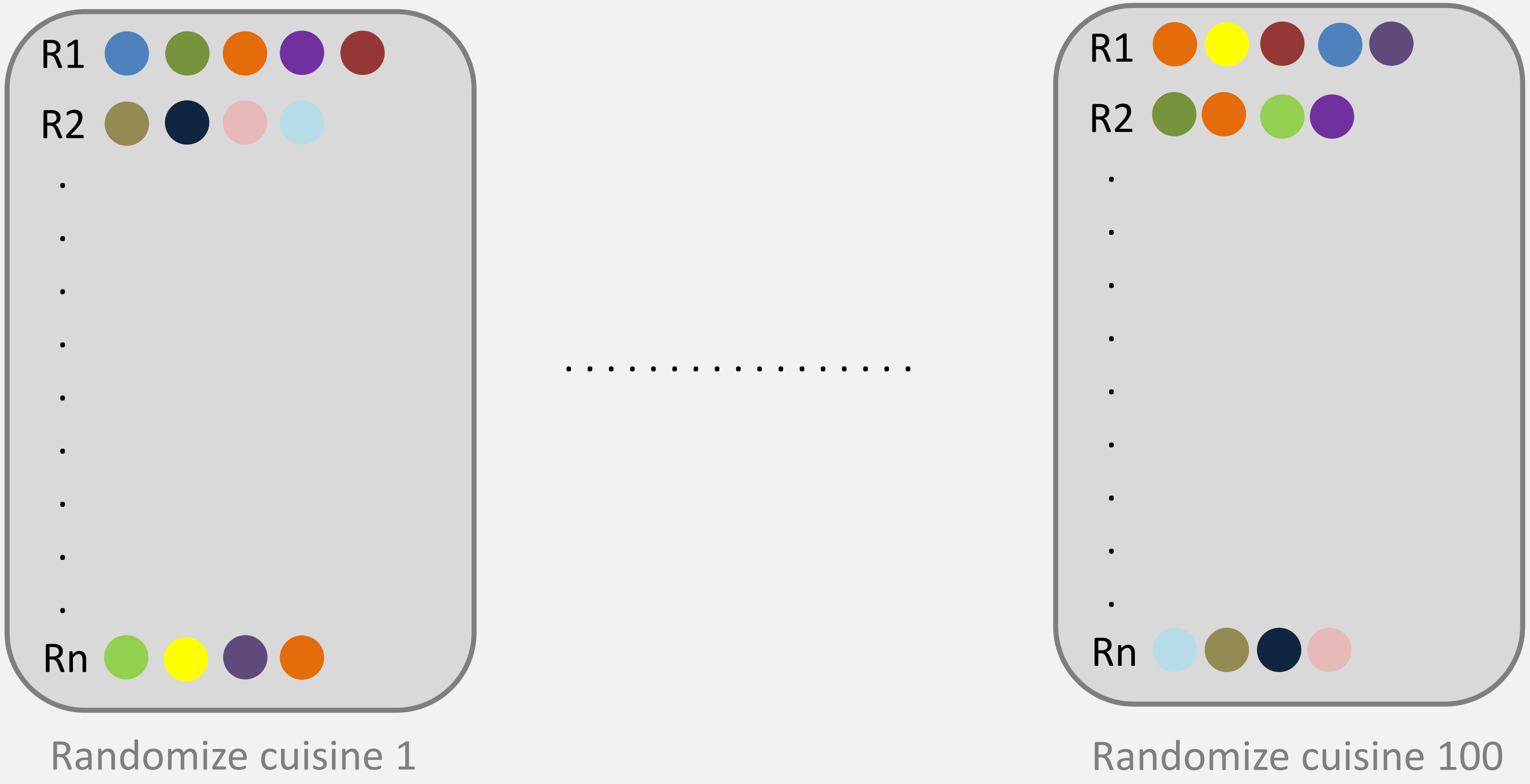
Backbone extraction algorithm

Shared flavor molecules in ingredient pairs



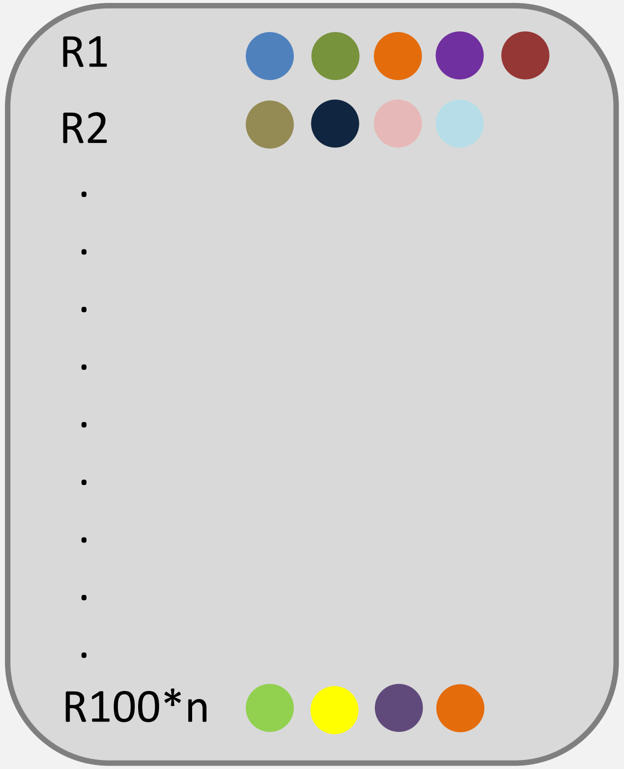
How to create a random cuisine? (Strategy – 1)

(1) Corresponding to every recipe, create its size-controlled randomized version by randomly sampling ingredients from the ‘Ingredients Basket’ (without replacement).
(2) Having thus created ‘a randomized cuisine’, create a large number of such cuisines (say, 100) for statistical analysis.

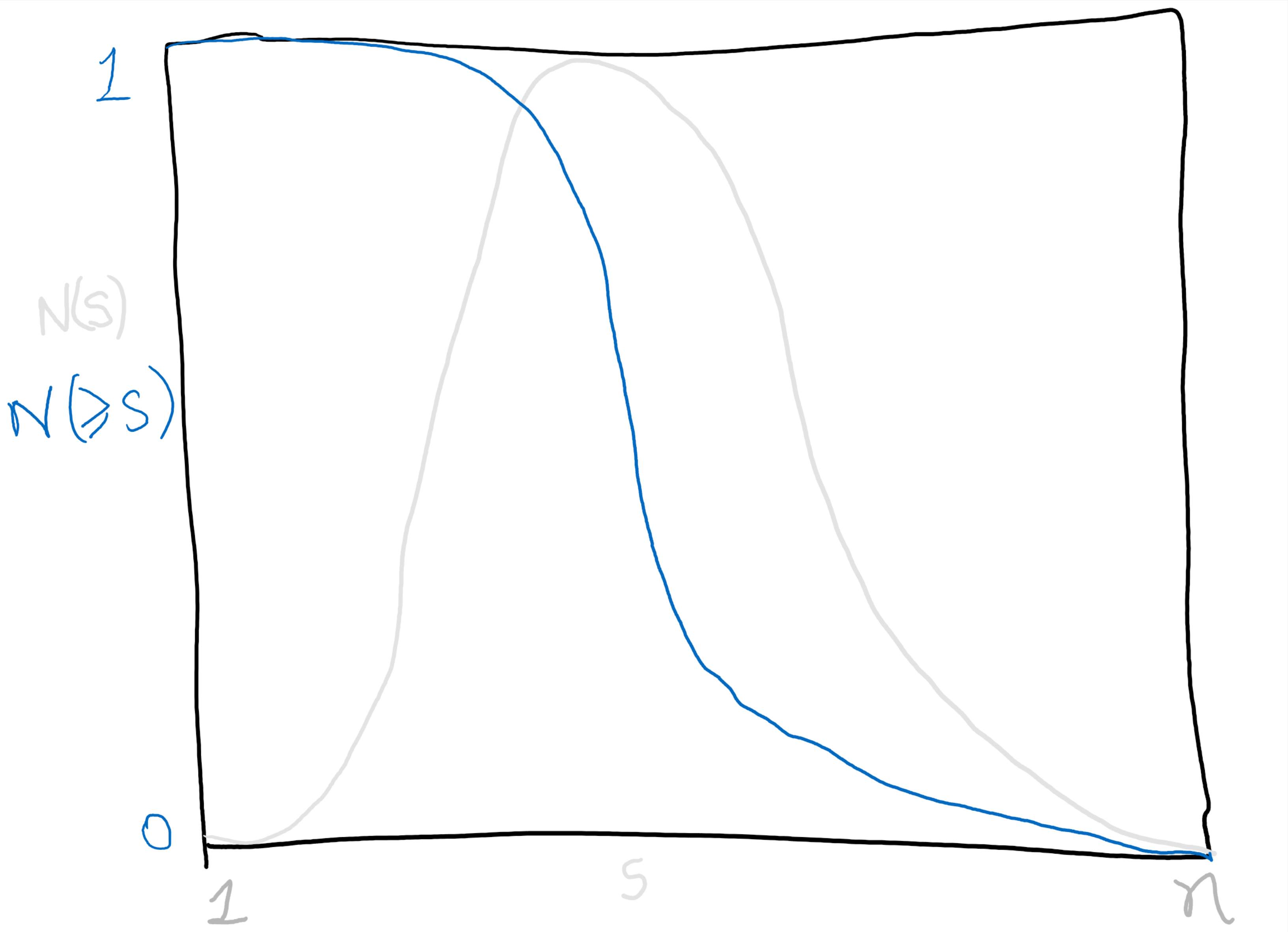


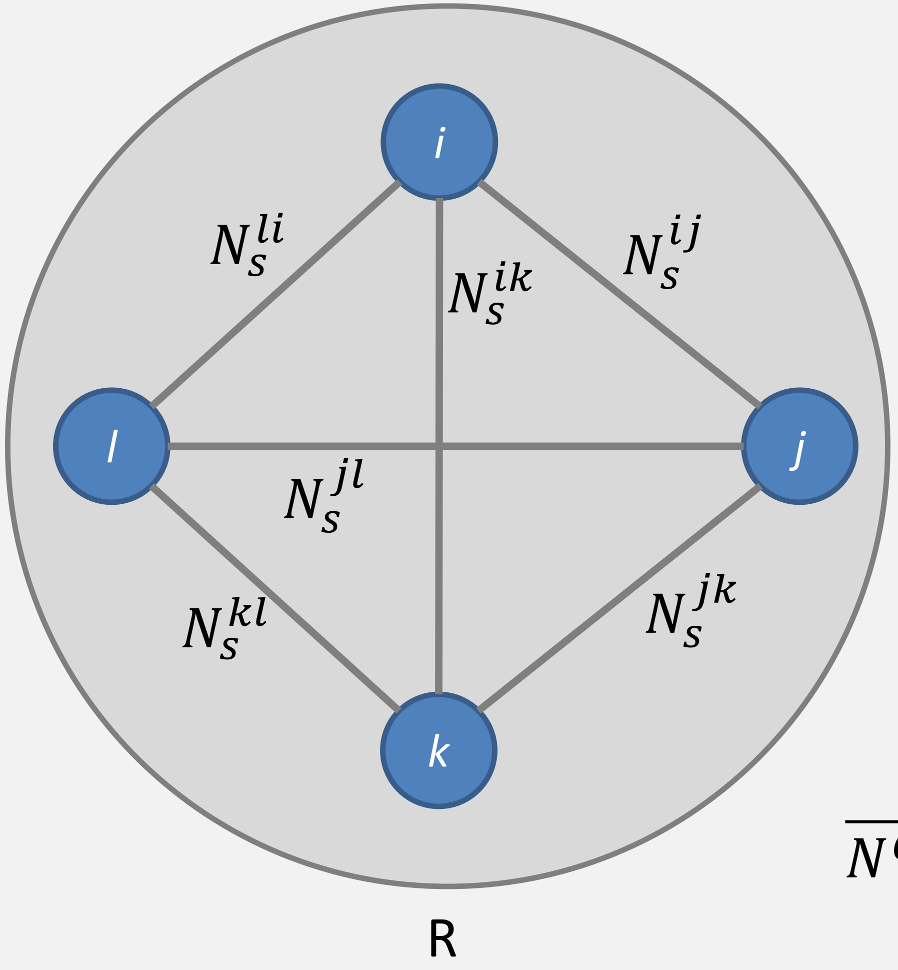
How to create a random cuisine? (Strategy – 2)

(1) Find the ‘recipe size distribution’ of the cuisine.
(2) Find the ‘cumulative recipe size distribution’ of the cuisine.
(3) Generate a random number to pick a recipe size.
(4) Create the recipe by randomly sampling ingredients from the ‘Ingredients Basket’ (without replacement).
(5) Generate a large number of recipes (say, 100 times the total number of recipes) for statistical analysis.



Generating the desired recipe size distribution of the cuisine





N_s^{ij} Number of shared flavor molecules between ingredients *i* and *j*

$\overline{N_s^{R}}$ Average number of shared flavor molecules across all pairs of ingredients in the recipe R

$\overline{N^{Cuisine}}$ Average number of shared flavor molecules across all recipes in a cuisine

Ideas

- **A theoretical model for food pairing**.
- Starting with a ‘basket of ingredients and their flavor profile, implement strategies for ‘pairing’ and observe the cuisine architecture.

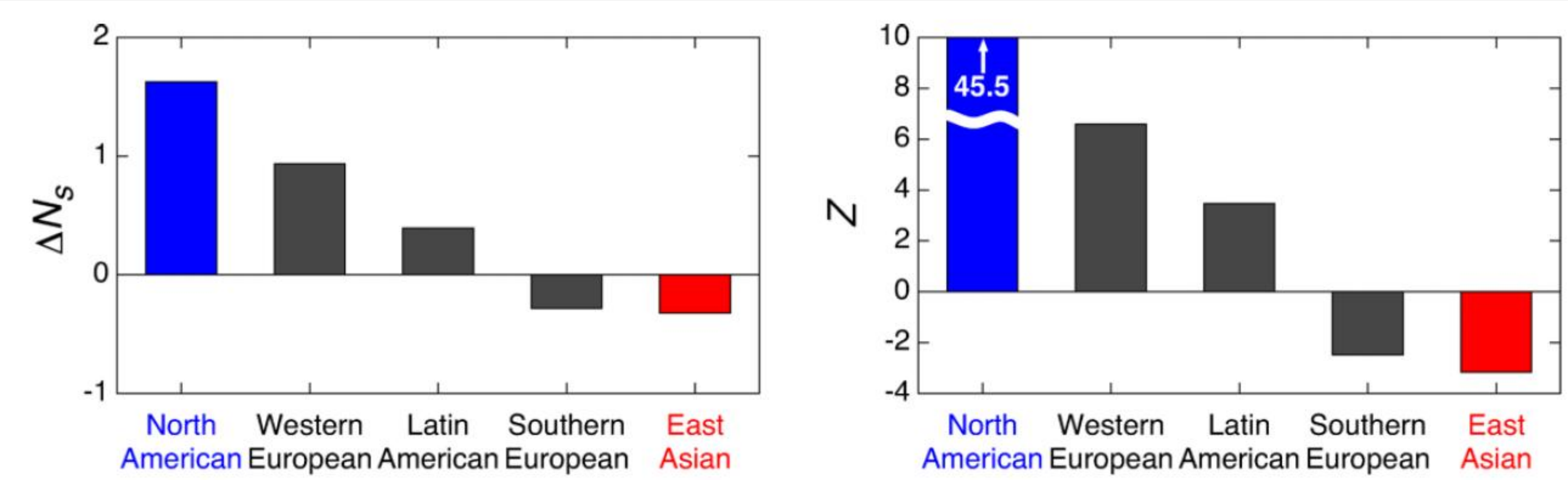
- **Create a python library for the analysis of cuisine, starting with cuisine data imported in a certain format.**
- Recipe size distribution, ingredient popularity stats
- Category composition analysis
- Itemset analysis
- Food pairing analysis
- Random cuisine generation

- **Create Computational Gastronomy Kaggle challenges**
- For a given set of recipes, predict their cuisine.
- Given the flavor profile of an (unlabeled) ingredient, predict its cuisine.
- Given a molecule (SMILE format), predict its taste (bitter, sweet, tasteless).
- and more.

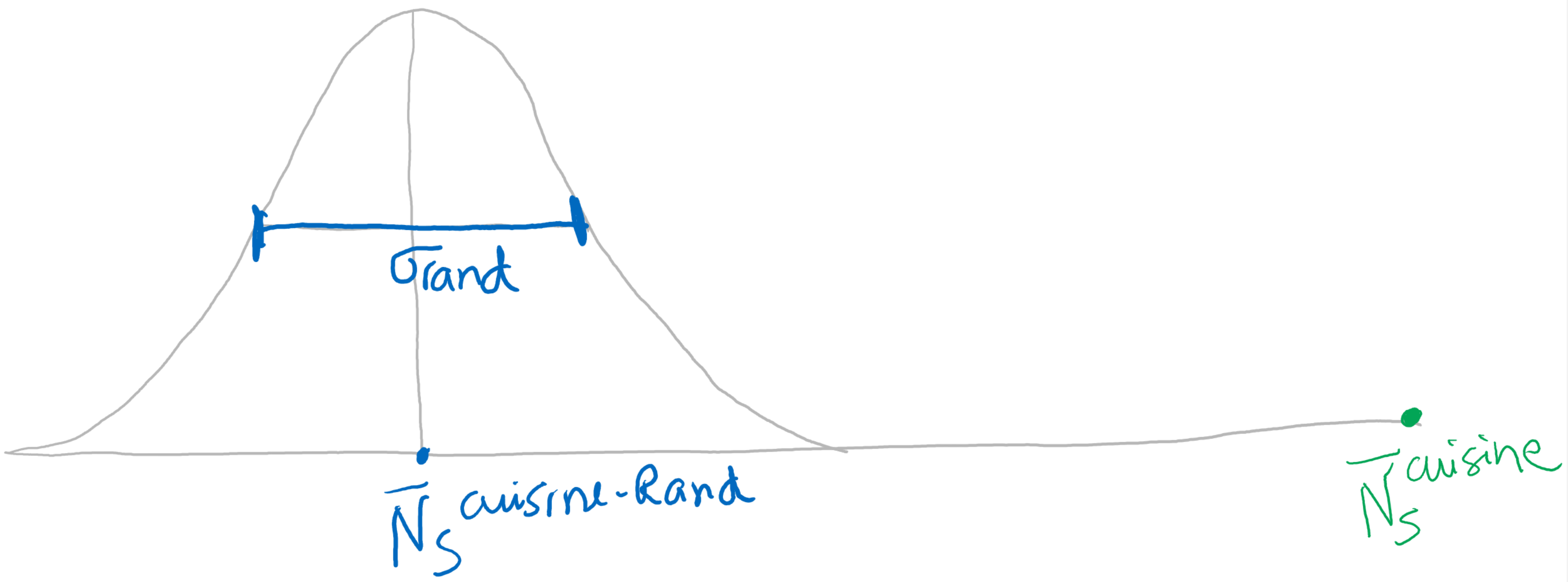
Random Cuisine and comparison of food pairing

North American and Western European cuisines exhibit a statistically significant tendency towards recipes whose ingredients share flavor compounds.

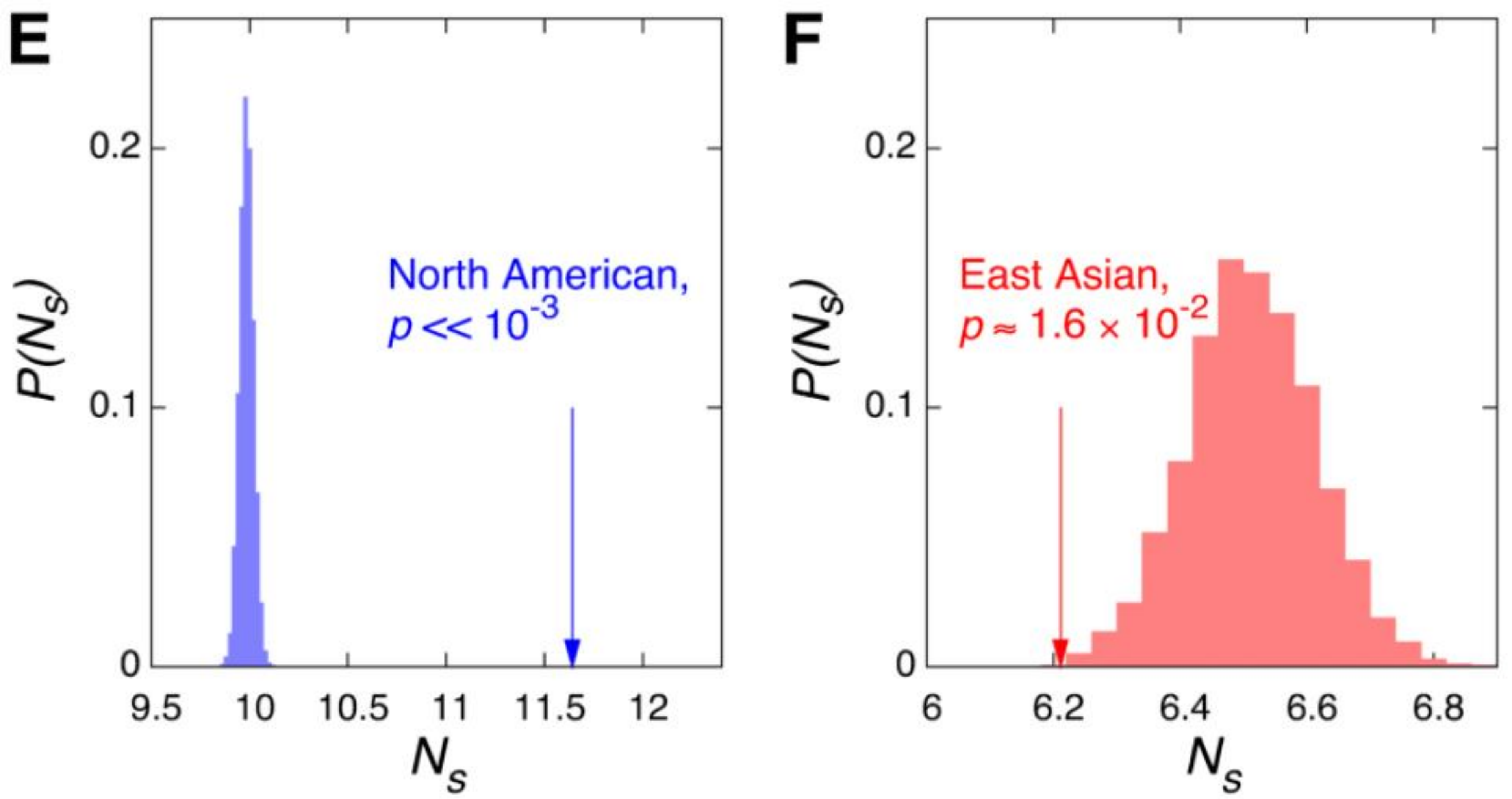
By contrast, East Asian and Southern European cuisines avoid recipes whose ingredients share flavor compounds.



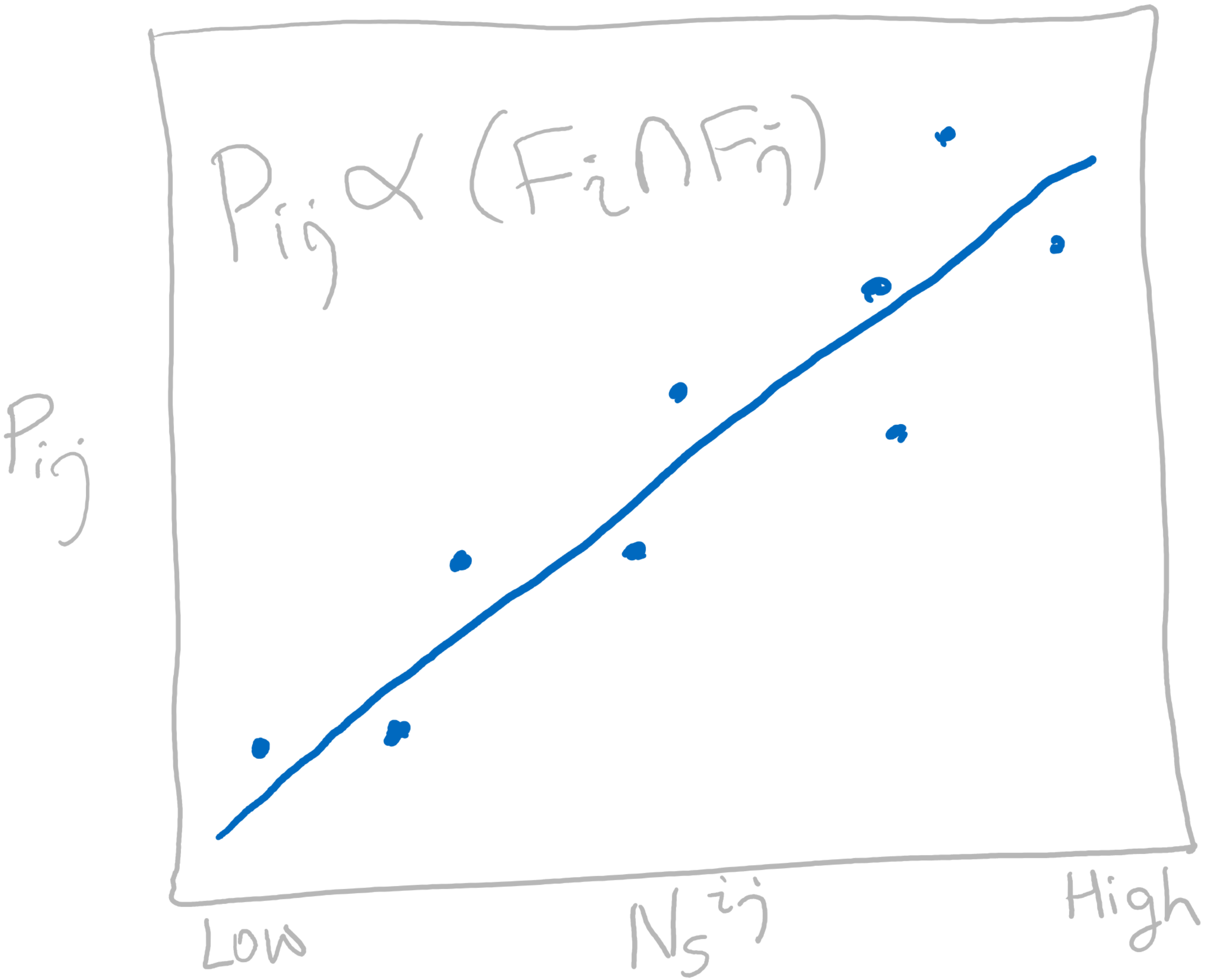
$$Z = \frac{\bar{N}_s^{cuisine} - \bar{N}_s^{cuisine\text{-}Rand}}{\sigma_{Ran}}$$



Distribution of Ns

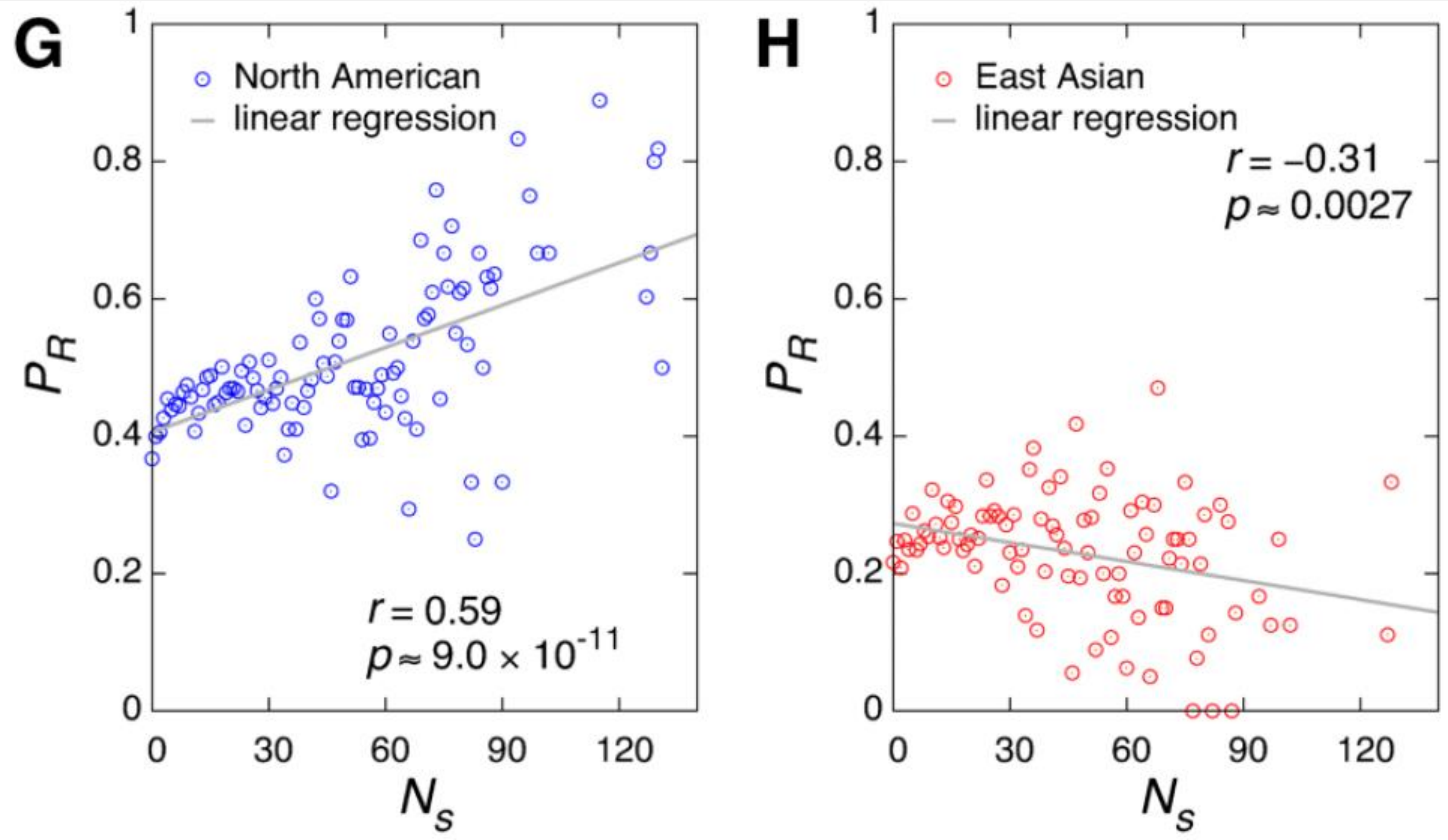


***Ns*: observed number of shared compounds characterizing the cuisines.**

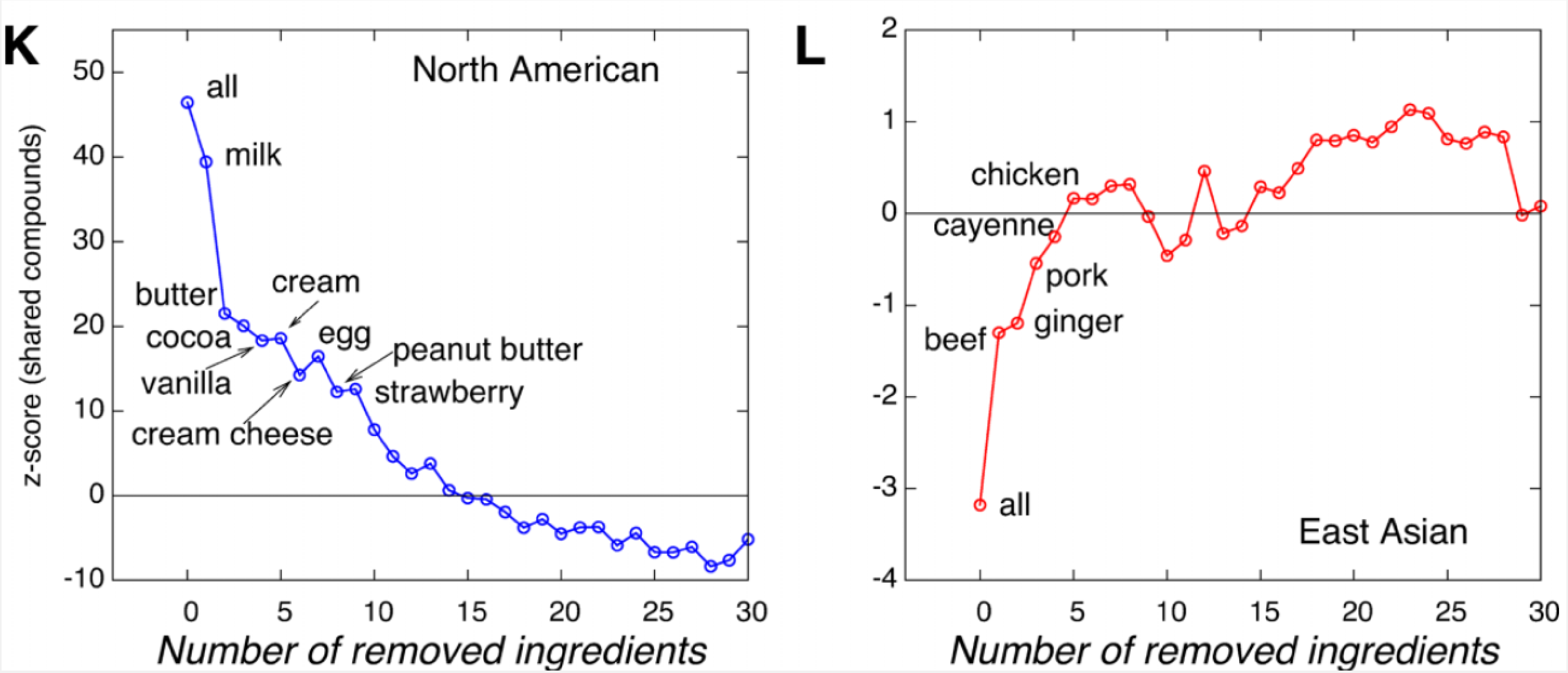


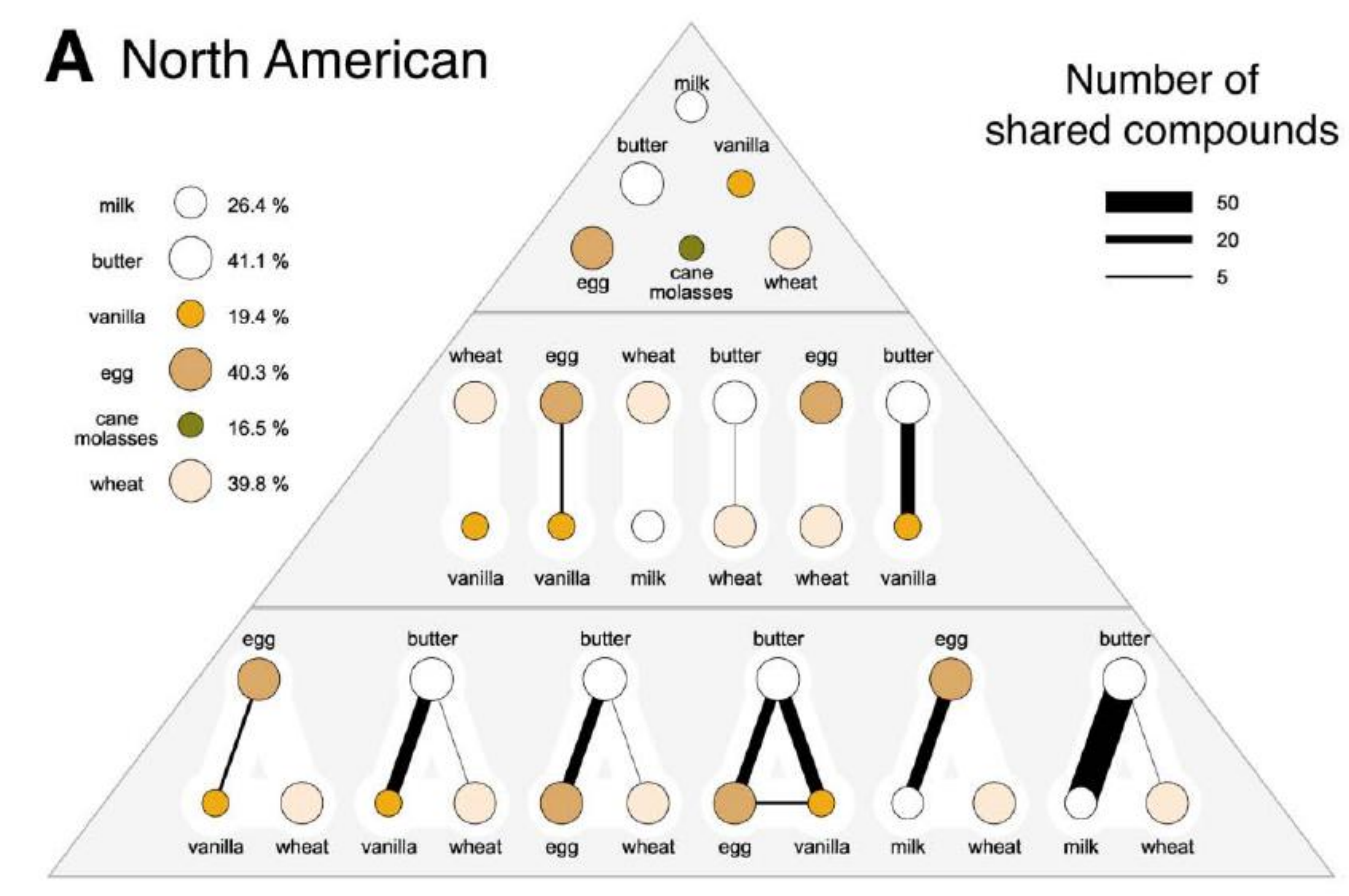
$P_{ij} \propto (F_i \cap F_j)$
P_{ij}
Low
N_s^{ij}
High

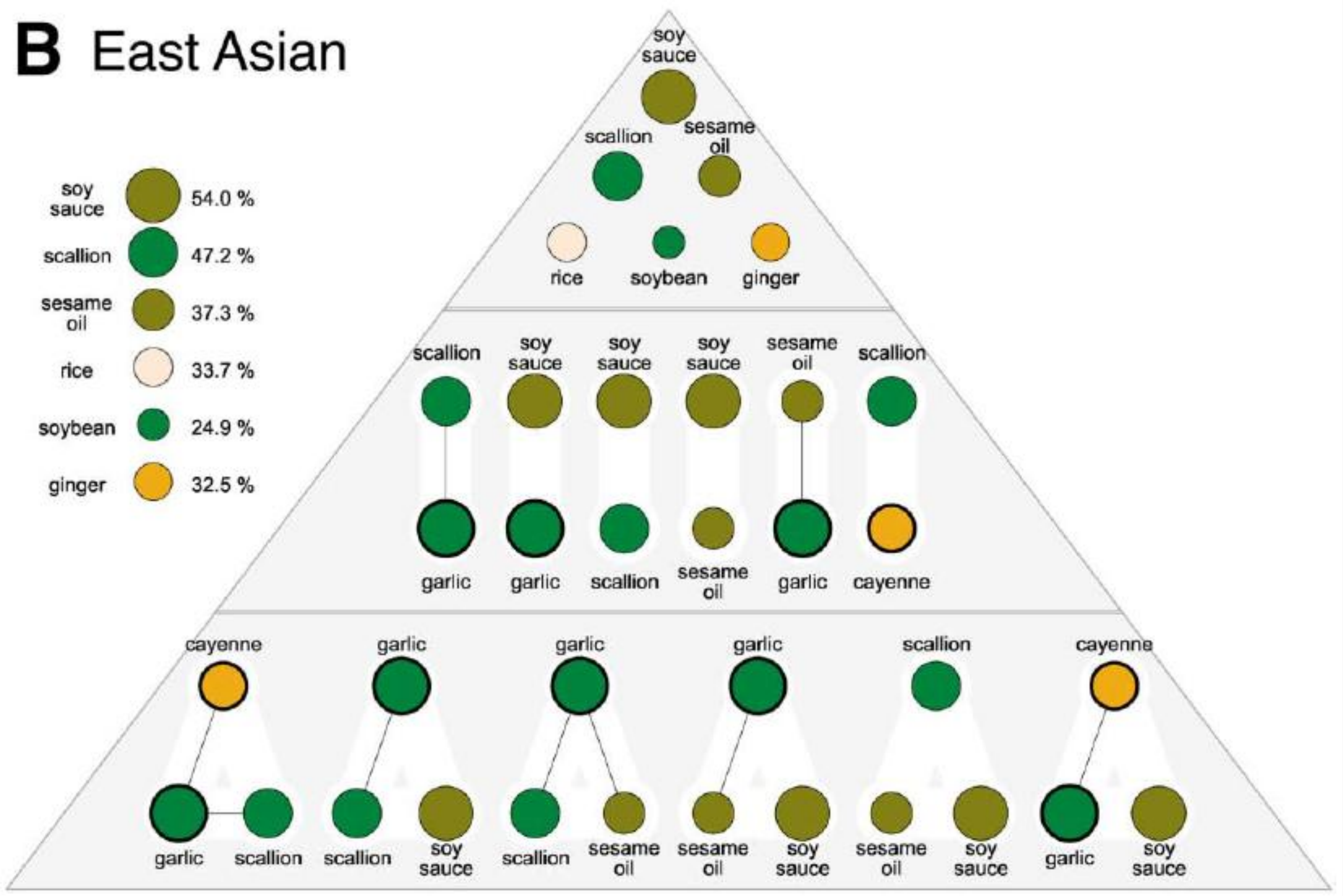
East Asian (Korean) cuisine the more flavor compounds two ingredients share, the less likely they are used together.



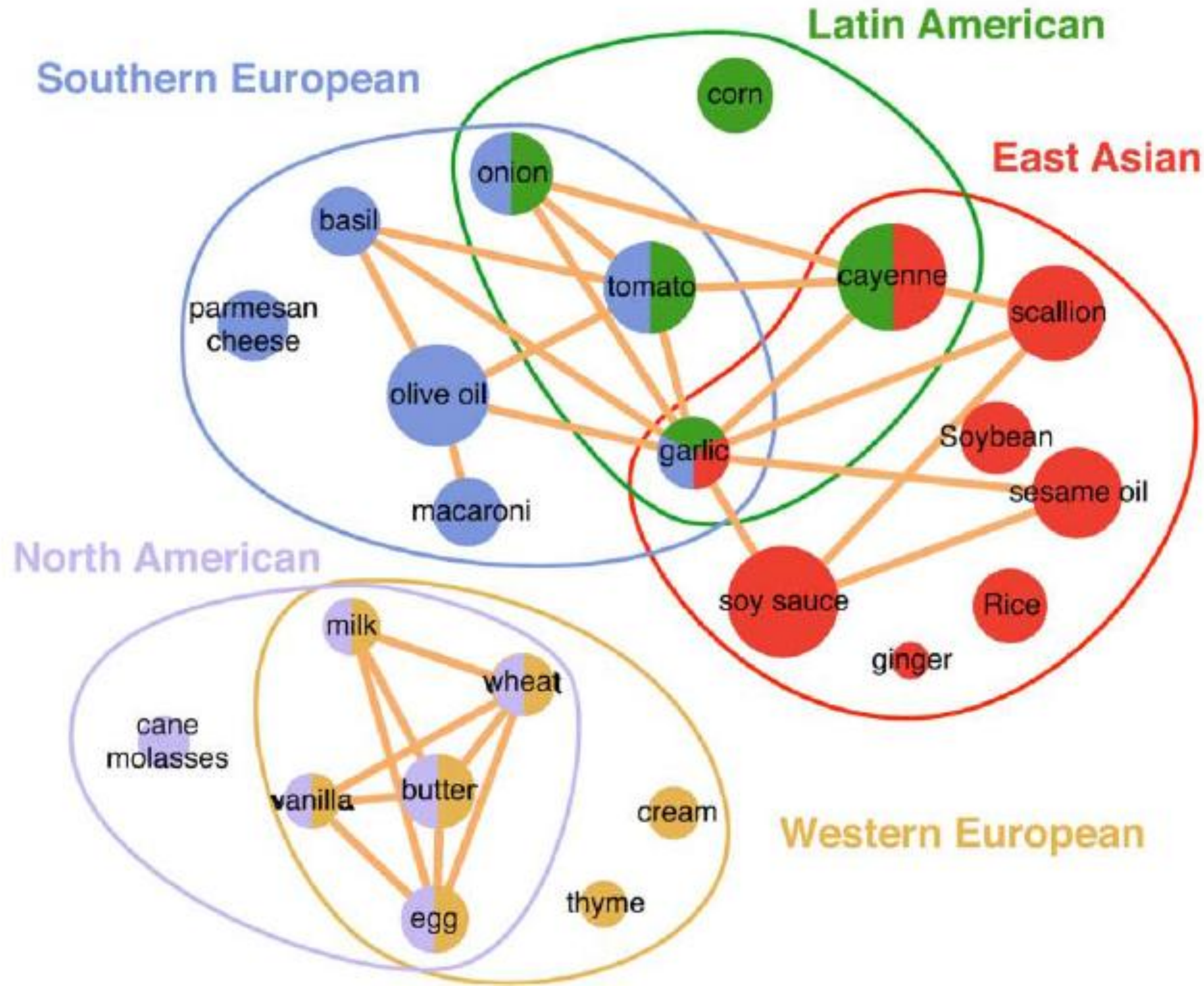
Contribution of individual ingredients towards the observed food pairing phenomena




A North American
Number of shared compounds
50
20
5
milk 26.4 %
butter 41.1 %
vanilla 19.4 %
egg 40.3 %
cane molasses 16.5 %
wheat 39.8 %
milk
butter
vanilla
egg
cane molasses
wheat
wheat
egg
wheat
butter
egg
butter
vanilla
vanilla
milk
wheat
wheat
vanilla
egg
butter
butter
butter
egg
butter
vanilla
wheat
vanilla
wheat
egg
wheat
egg
vanilla
milk
wheat
milk
wheat

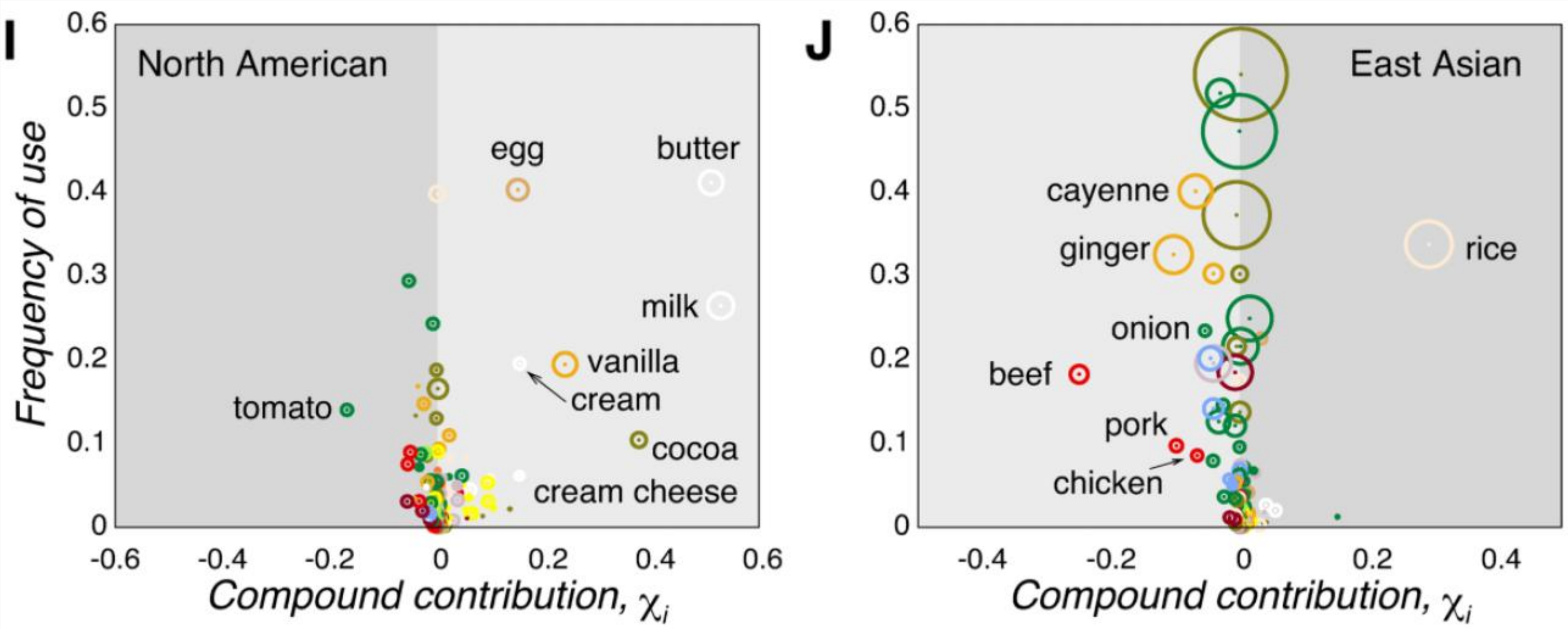

B East Asian
soy sauce
54.0 %
scallion
47.2 %
sesame oil
37.3 %
rice
33.7 %
soybean
24.9 %
ginger
32.5 %
soy sauce
scallion
sesame oil
rice
soybean
ginger
scallion
soy sauce
soy sauce
soy sauce
sesame oil
scallion
garlic
garlic
scallion
sesame oil
garlic
cayenne
cayenne
garlic
garlic
garlic
scallion
cayenne
garlic
scallion
scallion
soy sauce
scallion
sesame oil
sesame oil
soy sauce
sesame oil
soy sauce
garlic
soy sauce

C Co-occurrence in recipes



- Western Cuisines use ingredient pairs that share many flavor compounds
- By contrast, Korean cuisine tend to avoid compound sharing ingredients.

Contribution of an ingredient towards the food pairing pattern



$$\chi_i \left(\frac{1}{N_c} \sum_{R \ni i} \frac{2}{n_R(n_R-1)} \sum_{j \neq i (j, i \in R)} \left| C_i \cap C_j \right| \right) - \left(\frac{2 f_i}{N_c \langle n_R \rangle} \frac{\sum_{j \in c} f_j \left| C_i \cap C_j \right|}{\sum_{j \in c} f_j} \right)$$