

# Odhalovanie tajomstiev kuchyne

Kristína Galková, Jana Poľašková, Simona Bednariková, Zuzana Kovačová

22. septembra 2025

Meno	podiel na projekte	popis práce
Kristína Galková	0.25	dokumentácia a vizualizácia prvej výskumnej otázky
Jana Poľašková	0.25	dokumentácia a vizualizácia druhej výskumnej otázky
Simona Bednariková	0.25	dokumentácia a vizualizácia tretej výskumnej otázky
Zuzana Kovačová	0.25	dokumentácia a vizualizácia štvrtej výskumnej otázky

Tabuľka 1: Rozdelenie práce

## 1 Úvod

Tento projekt vychádza z dát publikovaných v štúdii "The Cultural Diversity of Culinary Practice" ([Yong-Yeol Ahn et al., 2011](#)), ktorá skúma vzorce v kombináciach ingrediencií naprieč rôznymi regionálnymi kuchyňami. Štúdia zavádzá koncept tzv. „flavor network“ (chuťová sieť), ktorá zachytáva zdieľané chuťové zlúčeniny medzi ingredienciami. Tento prístup poskytuje nový pohľad na kulinárské praktiky a ich rôznorodosť. Z tejto štúdie sme si práve vybrali dataset, ktorý poskytuje vzťahy medzi ingredienciami a jednotlivými regiónmi.

## 2 Dáta

Dataset nám umožňuje odhaliť, ktoré ingrediencie sú globálne zdieľané, ktoré sú typické len pre určité oblasti a ako sú regióny kulinárne prepojené. Teda prepájame kultúru so sietovou analýzou.

Dáta obsahujú zoznam receptov vo formáte CSV. Prvý stĺpec predstavuje názvy regiónov, zatiaľ čo nasledujúce stĺpce obsahujú ingrediencie používané v konkrétnych receptoch. S dátami sme pracovali v programovacom programe Python a graf vytvorili pomocou knižnice NetworkX.

Kedže vzťahy sú definované výlučne medzi regiónmi a ingredienciami, dáta môžu byť reprezentované ako bipartitný graf. Jednu partíciu tvoria regióny a druhú partíciu ingrediencie. Ak sa daná ingredencia vyskytuje v recepte z určitého regiónu, je spojená s týmto regiónom hranou. Váha hrany vyjadruje počet receptov, v ktorých je ingredencia použitá.

### 2.1 Základné štatistiky

Základné štatistiky sú zobrazené v tabuľke 2 a štatistiky podľa regiónov sú uvedené v tabuľke 3. Tabuľka 3 ukazuje rozdelenie unikátnych a celkových ingrediencií podľa jednotlivých regiónov, spolu s počtom receptov z každého regiónu.

Bipartitný graf 1 zobrazuje vzťahy medzi regiónmi a ingredienciami na základe prítomnosti jednotlivých ingrediencií v receptoch. Na ľavej strane grafu sú zobrazené regióny, ktoré sú odlišené rôznymi farbami (spolu 11 regiónov). Na pravej strane grafu sa nachádzajú ingrediencie (spolu 381), ktoré sú zoradené do štyroch stĺpcov kvôli lepšej prehľadnosti.

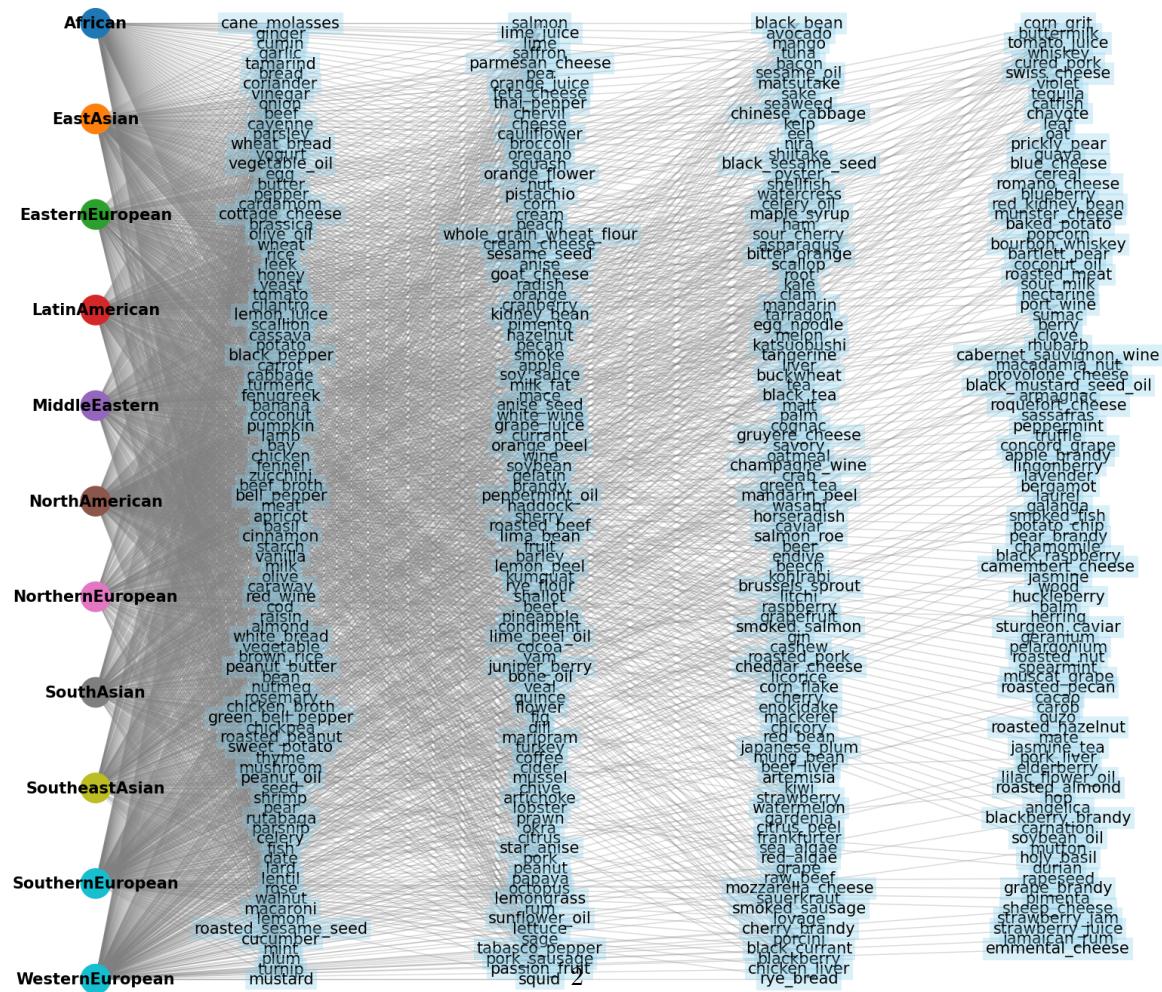
Celkový počet uzlov	392
Celkový počet hrán	2641
Počet regiónov (počet uzlov v partícií 0)	11
Počet ingrediencií (počet uzlov v partícií 1)	381
Hustota siete	0.6302
Priemerný stupeň (Regióny - partícia 0)	240.09
Priemerný stupeň (Ingrediencie - partícia 1)	6.93

Tabuľka 2: Základné štatistiky grafu

Región	Unikátne ingredience	Celkové ingredience	Počet receptov
African	197	3674	351
EastAsian	242	22498	2512
EasternEuropean	198	3196	381
LatinAmerican	260	27360	2917
MiddleEastern	227	5411	645
NorthAmerican	354	330618	41524
NorthernEuropean	175	1706	250
SouthAsian	205	6388	621
SoutheastAsian	184	5172	457
SouthernEuropean	290	37038	4180
WesternEuropean	309	21341	2659

Tabuľka 3: Štatistiky podľa regiónov

Bipartitný graf: Regióny vs. Ingredience



### 3 Hlavné otázky

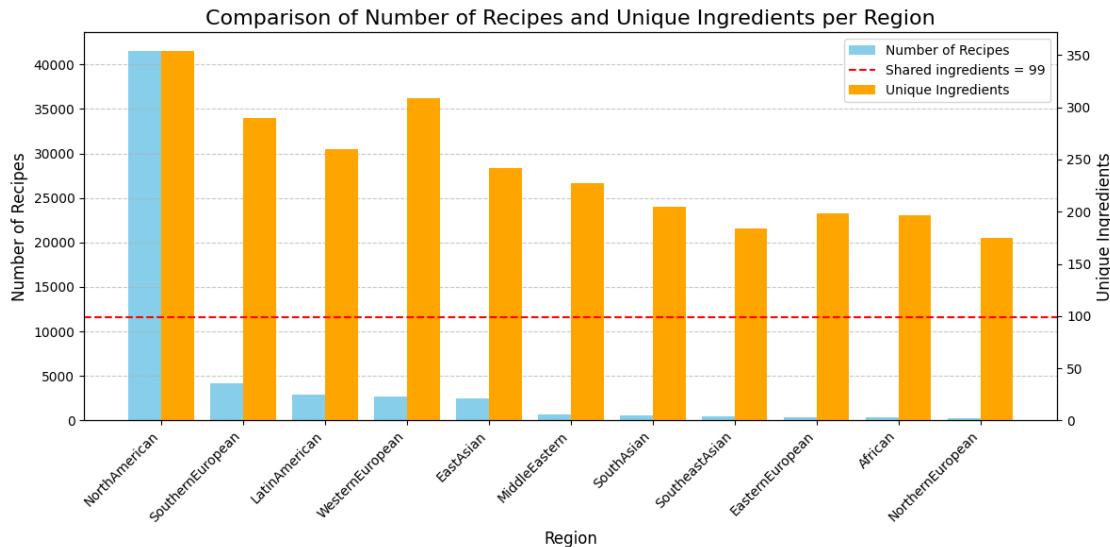
Na základe vykreslenia grafu a základných štatistik sme si mohli položiť niekoľko výskumných otázok, ktoré nás zaujímajú a na ktoré sa pokúsime odpovedať.

#### Otázka č. 1: Aké sú dôvody, prečo Severná Amerika využíva väčšie množstvo ingrediencií a receptov v porovnaní s inými regiónymi?

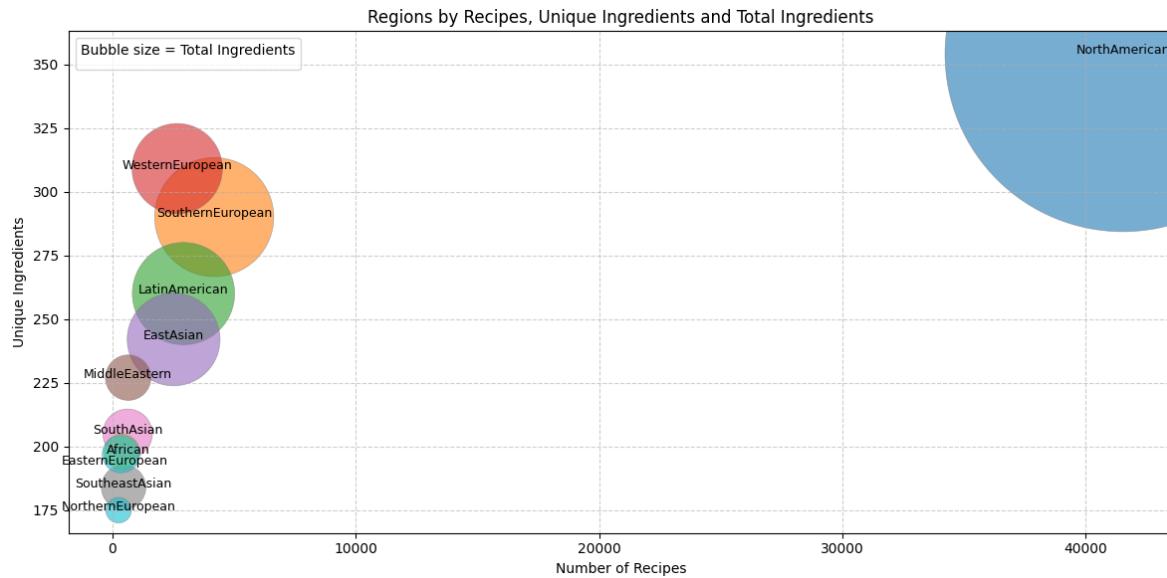
Pri pohľade na bipartitný graf a základné štatistiky regiónov a ingrediencií je zrejmé, že Severná Amerika vyniká medzi svetovými regiónmi nielen počtom receptov, ale aj rozmanitosťou ingrediencií. Za hlavné faktory tejto rozmanitosti považujeme nasledovné:

- Prvým dôvodom je kultúrna diverzita ako výsledok pristáhovalectva. Severná Amerika – predovšetkým USA a Kanada – je rajom pristáhovalcov z celého sveta. Každá komunita si priniesla svoje tradičné recepty, suroviny a chutové preferencie. Výsledkom je gastronomická fúzia, kde vedľa seba existujú ingrediencie z Ázie, Európy, Latinskej Ameriky aj Afriky. Táto jedinečná zmes kultúr sa prirodzene odráža v počte a rôznorodosti ingrediencií, ktoré sa v regióne bežne používajú.
- Druhým dôvodom môžu byť ich menej prísne potravinové regulácie. V USA platí princíp, že látka môže byť používaná, pokiaľ neexistuje dôkaz o jej škodlivosti (tzv. GRAS – generally recognized as safe). Naopak, Európska únia uplatňuje princíp predbežnej opatrnosti, kde sa látka musí preukázať za bezpečnú, aby mohla byť schválená. Tento rozdiel sa netýka samotných ingrediencií, ale chemických látok, ktoré sa v nich môžu nachádzať – napríklad konzervanty, farbivá či sladidlá. Tak sa môže stať, že určitá ingredencia (napr. konkrétny druh sladička alebo dochucovadla) je v Európe obmedzená alebo zakázaná práve kvôli zložkám, ktoré obsahuje, zatiaľ čo v USA je bežne povolená. To môže viesť k väčšiemu počtu rôznych ingrediencií evidovaných v amerických receptoch. ([Time, 2024](#)), ([CBS News, 2023](#))
- Tretím možným faktorom je samotný dataset. Je možné, že dataset obsahuje výrazne viac severoamerických receptov v porovnaní s ostatnými regiónmi, čo prirodzene zvyšuje aj zaznamenaný počet tamojších ingrediencií. Tento dátový nepomer môže byť spôsobený vyššou digitalizáciou, popularitou amerických receptov online či angličtinou ako dominantným jazykom v dostupných zdrojoch.

Na nasledujúcich grafoch (Obr. 2 a Obr. 3) vidíme porovnanie počtu receptov, unikátnych ingrediencií a ingrediencií pre jednotlivé regióny. Severná Amerika tu jasne dominuje vo všetkých kategóriách:



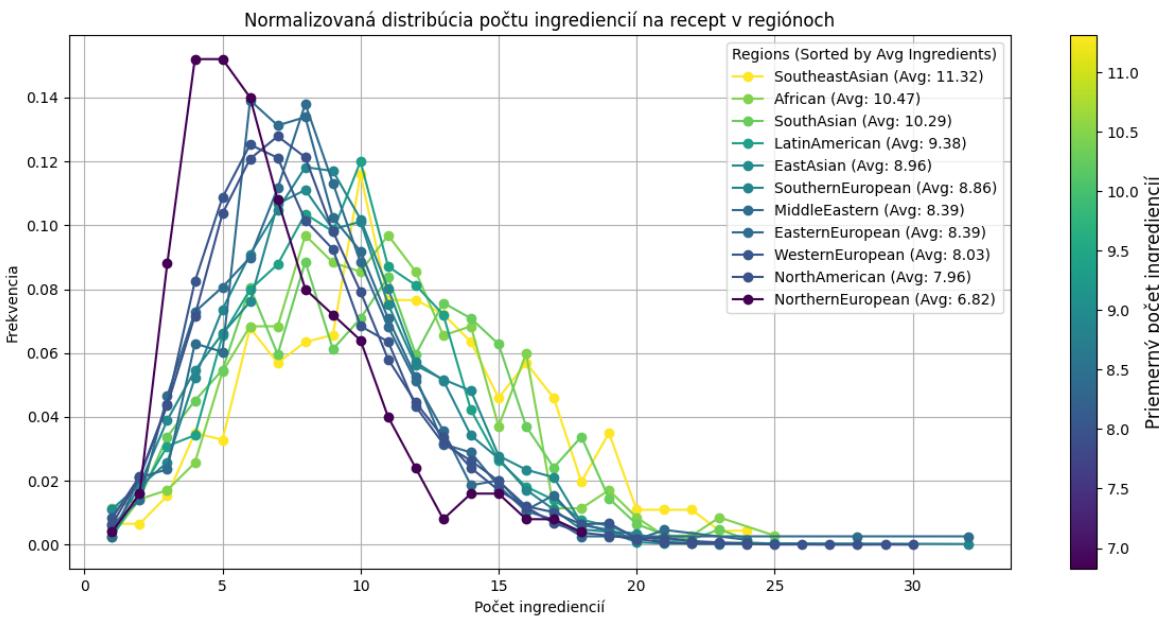
Obr. 2: Porovnanie počtu receptov a unikátnych ingrediencií pre jednotlivé regióny



Obr. 3: Regióny na základe počtu receptov a unikátnych ingrediencií a celkového počtu ingrediencií

### Otázka č. 2: Aké sú rozdiely v receptoch medzi regiónmi, najmä v počte ingrediencií na recept?

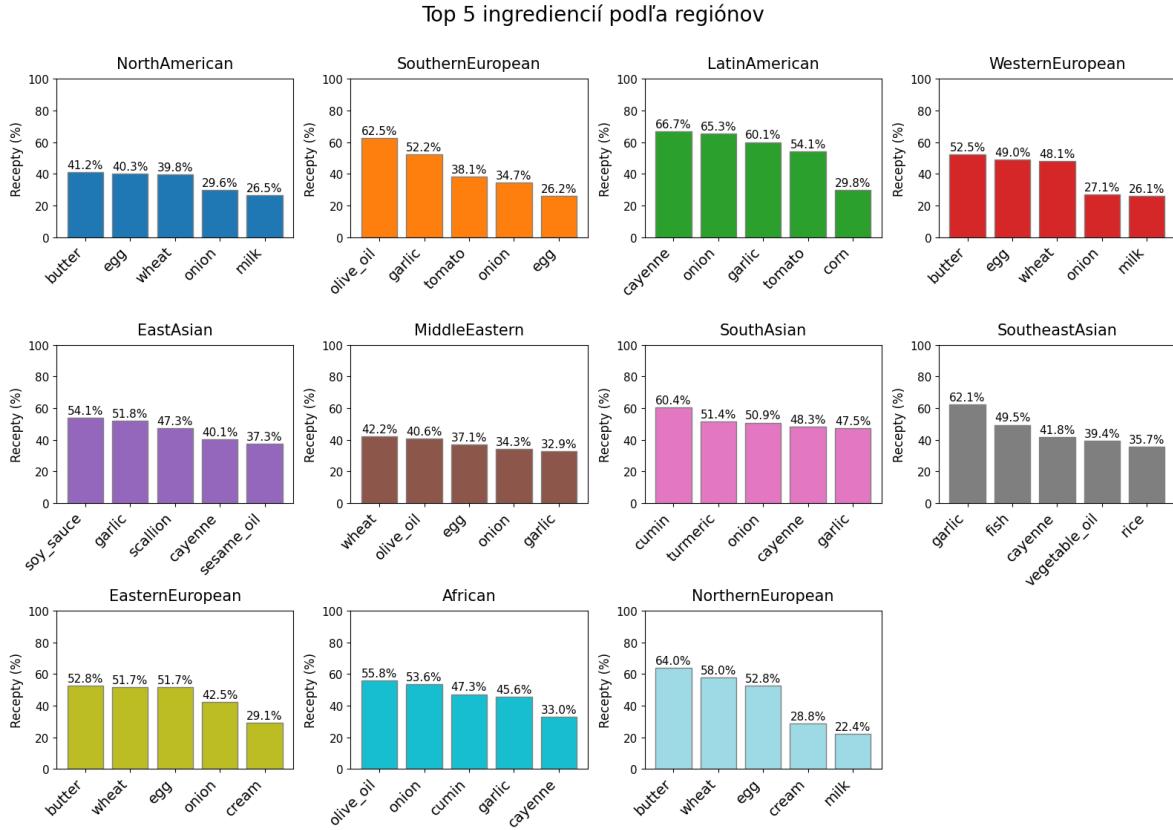
Graf 4 zobrazuje normalizovanú distribúciu počtu ingrediencií na recept v rôznych regiónoch. Na osi x pozorujeme počet ingrediencií na recept, zatiaľ čo os y zobrazuje frekvenciu výskytu daného počtu ingrediencií v danom regióne. Každá krivka zodpovedá inému regiónu a je zafarbená podľa priemeru počtu ingrediencií v danom regióne. Tento graf poskytuje prehľad o tom, ako sa lísi počet ingrediencií používaných v receptoch medzi rôznymi regiónmi.



Obr. 4: Normalizovaná distribúcia počtu ingrediencií na recept v regiónoch

Graf bol vytvorený tak, že sme pre každú krajинu spočítali počet ingrediencií, ktoré sú uvedené v receptoch a z tohto sme vypočítali normalizovanú frekvenciu (výskyt) - teda každú frekvenciu sme vydelili súčtom všetkých frekvencií v danom regióne.

Pozorujeme, že regióny Juhovýchodná Ázia a Afrika majú tendenciu používať v priemere väčší počet ingrediencií na recept, pričom ich distribúcie sú širšie, s výraznejšími vrcholmi v oblasti vyšších počtov ingrediencií. Naopak, regióny ako Severná Európa a Západná Európa majú tendenciu používať menší počet ingrediencií. Tieto rozdiely odrážajú rozdielne zvyky medzi regiónmi. Teplejšie regióny disponujú širším spektrom byliniek, korenín a lokálnych plodín, čo vedie k častejšiemu pridávaniu rôznorodých ingrediencií. Túto domnieku tiež potvrzuje skutočnosť, že top 5 používaných surovín pre Afirku (Obr. 5) sú prevažne bylinky ako rímsky kmín, kurkuma a kajenské korenie.



Obr. 5: Top 5 ingrediencií naprieč regiónmi. Na osi x sú zobrazené najpoužívanejšie ingrediencie pre každý región, zatiaľ čo na osi y je uvedený percentuálny podiel receptov z daného regiónu, v ktorých sa príslušná ingredencia vyskytuje.

### Otázka č. 3: Ktoré kuchyne si sú navzájom podobné a aké suroviny sú pre jednotlivé kuchyne jedinečné?

Z bipartitného grafu región–ingredencia sme najprv vytvorili váženú jednomódovú (iba jednej množine uzlov – v tomto prípade regióny) projekciu na uzloch regiónov. Hrana medzi dvoma regiónmi  $r_i, r_j$  má hmotnosť rovnú počtu ingrediencií, ktoré zdieľajú. Aby sme kvantifikovali ich vzájomnú kulinárnu podobnosť, definovali sme *Jaccardovu vzdialenosť* pre množiny ingrediencií  $I_i$  a  $I_j$  takto:

$$d_J(r_i, r_j) = 1 - \frac{|I_i \cap I_j|}{|I_i \cup I_j|},$$

**Podobnosť medzi regiónmi:** Ak majú dva regióny veľmi podobné ingrediencie, prienik ich množín bude veľký, a zjednotenie týchto množín bude len o trochu väčšie. Takto bude Jaccardova vzdialenosť blízka 0, čo naznačuje, že kuchyne sú veľmi podobné.

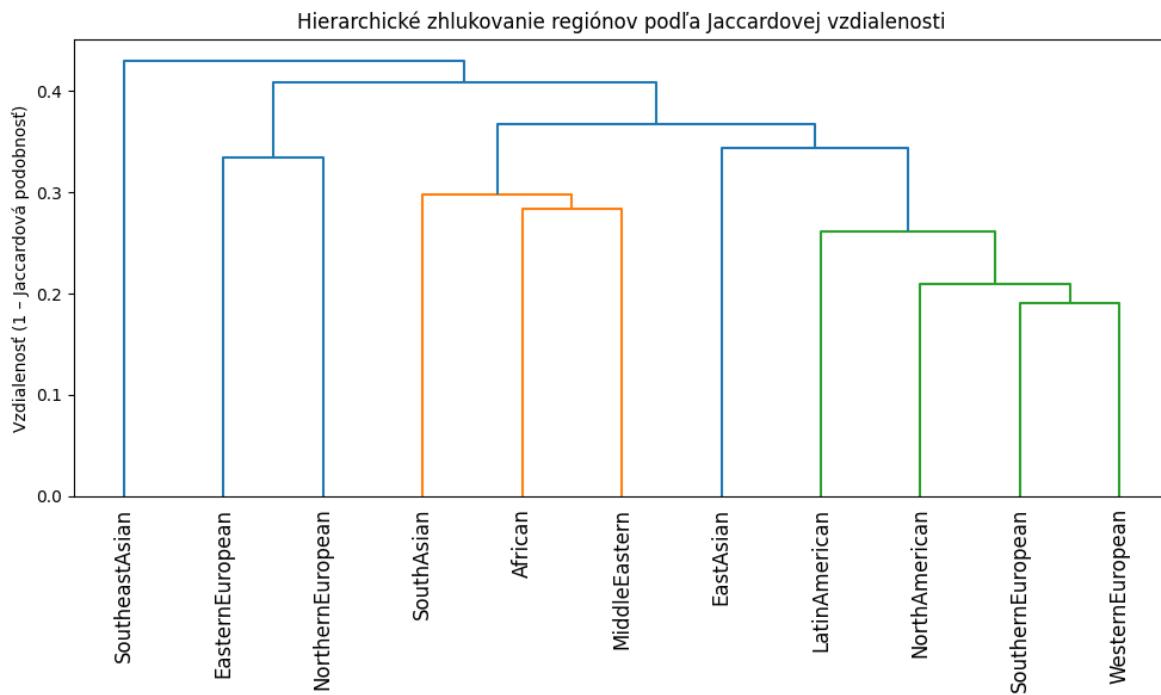
**Rozdielnosť medzi regiónmi:** Ak majú regióny len málo spoločných ingrediencií (napríklad niektoré ingrediencie sa používajú len v jednom regióne), prienik bude malý a zjednotenie bude veľké, čo

spôsobí vysokú hodnotu blízko 1. To znamená, že tieto kuchyne sú veľmi odlišné.

Ciže  $d_J = 0$  znamená úplnú zhodu surovín a  $d_J = 1$  ich úplnú odlišnosť. Na matice týchto vzdialenosťí sme potom aplikovali hierarchické zhlukovanie, ktoré spája klastry podľa priemernej Jaccardovej vzdialenosťi.

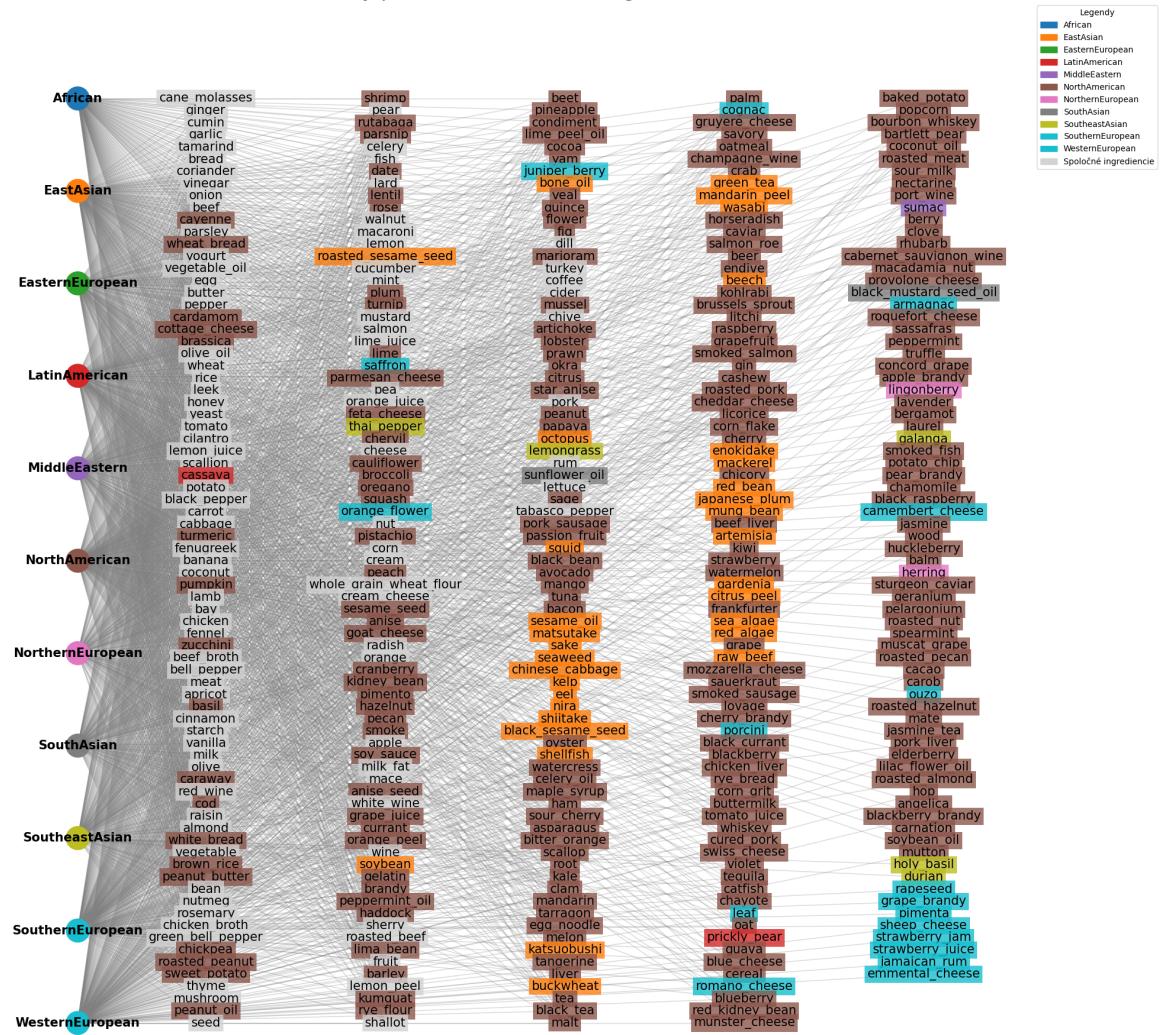
Výsledný dendrogram (Obr. 6) odhalil tri prirodzené skupiny kuchýň. Suroviny nachádzajúce sa v konkrétnej kuchyni pozorujeme na obrázku 7 a 8 :

- **Východná a Severná Európa (EasternEuropean + NorthernEuropean)** (cca 0.33) — spoločné endemity ako brusnice (angl. lingonberry), sled' - druh ryby ktorá sa vyskytuje v oblastiach okolo Severnej Európy (angl. herring) či sladké drievko (licorice) tvoria charakter tejto kuchyne. Juhovýchodná Ázia (SoutheastAsian): (vzdialosť > 0.4) — najunikátnejšia kuchyňa. Používa suroviny špecifické pre svoju geografickú oblasť ako durian, galanga či holy basil.
- **Afrika, Juh Ázie, Blízky východ (African, SouthAsian, MiddleEastern)** (cca 0.28–0.30) — spájajú ich koreniny, oleje a obilniny (kokosový olej, ryža, olej z čiernej horčice).
- **Latinská Amerika a Severná Amerika (LatinAmerican + NorthAmerican) Južná Európa a Západná Európa (SouthernEuropean + WesternEuropean)** (cca 0.26) — základné plodiny a mliečne produkty (kukurica, fazuľa, paradajka, mliečne výrobky). Južná Európa a Západná Európa (SouthernEuropean + WesternEuropean): (cca 0.20) — typická mediteránno-európska báza: olivový olej, cesnak, vína a syry.

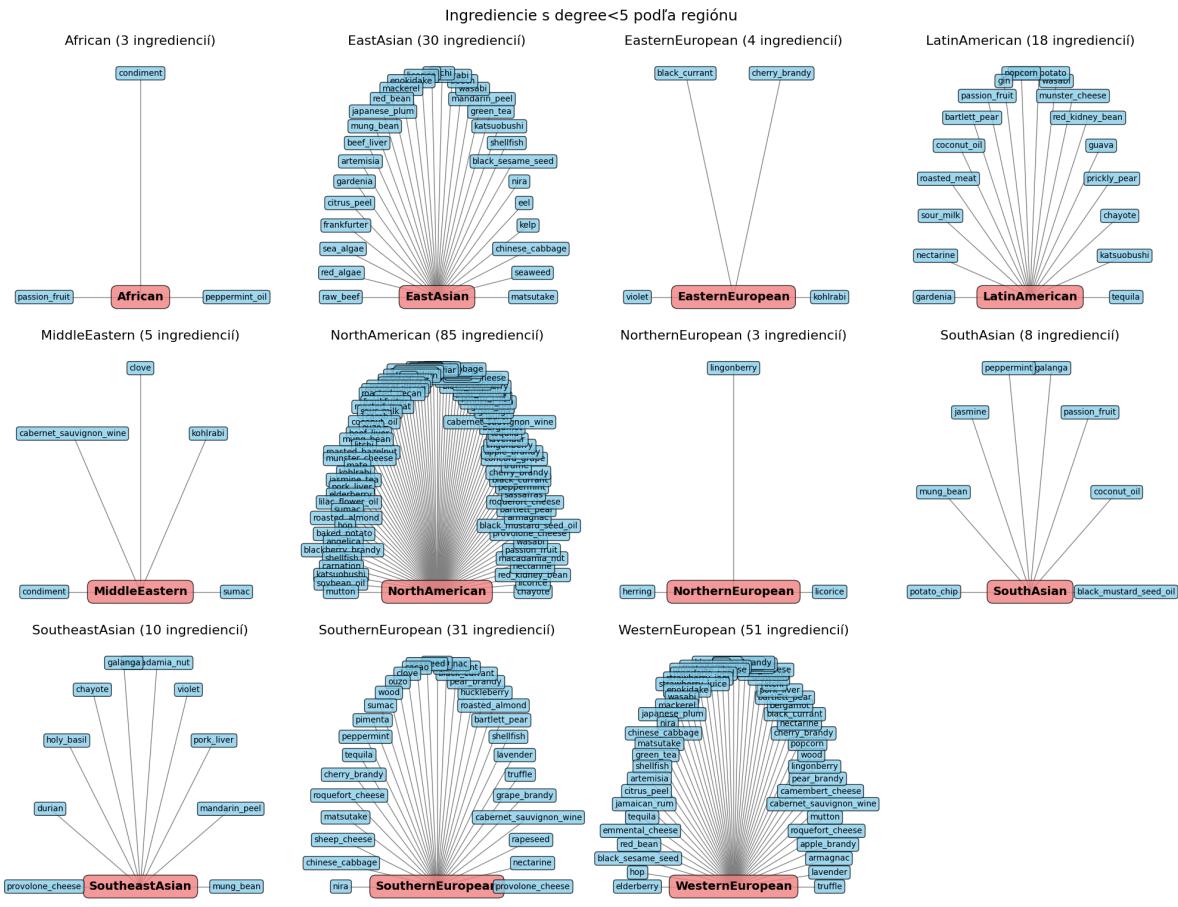


Obr. 6: Hierarchické zhlukovanie regiónov podľa Jaccardovej vzdialenosťi ingrediencií.

Bipartitný graf: Regióny vs. Ingredience  
(farby podľa dominantného regiónu)



Obr. 7: Bipartitný graf regióny vs. ingrediencie (farby podľa dominantného regiónu).



Obr. 8: Suroviny v jednotlivých regiónoch.

#### Otázka č. 4: Aké sú vzťahy medzi top 5 najpoužívanejšími ingredienciami z každého regiónu?

Na odpovedanie tejto otázky sme vykreslili projekciu bipartitného grafu, ktorý vizualizuje vzťahy medzi ingredienciami na základe ich spoločného výskytu medzi top 5 ingredienciami v rôznych regiónoch (Obr. 9).

- Uzly reprezentujú ingrediencie, pričom veľkosť uzla je úmerná celkovej váhe hrán, ktoré k nemu vedú. Väčšie uzly znamenajú ingrediencie, ktoré sú častejšie používané v kombinácii s ostatnými surovinami medzi regiónmi.
- Hrana medzi dvomi uzlami reprezentuje spoločný výskyt dvoch ingrediencií v rovnakých regiónoch. Hrubá hrana je úmerná počtu regiónov, v ktorých obe suroviny patria medzi top 5.
  - Hrubé hrany spájajú cesnak s viacerými ingredienciami, ako je cibuľa a paradajky. To naznačuje ich častý spoločný výskyt v regiónoch.
  - Tensie hrany sa objavujú medzi ingredienciami, ktoré sú menej často kombinované, ako je napríklad sójová omáčka a sezamový olej.

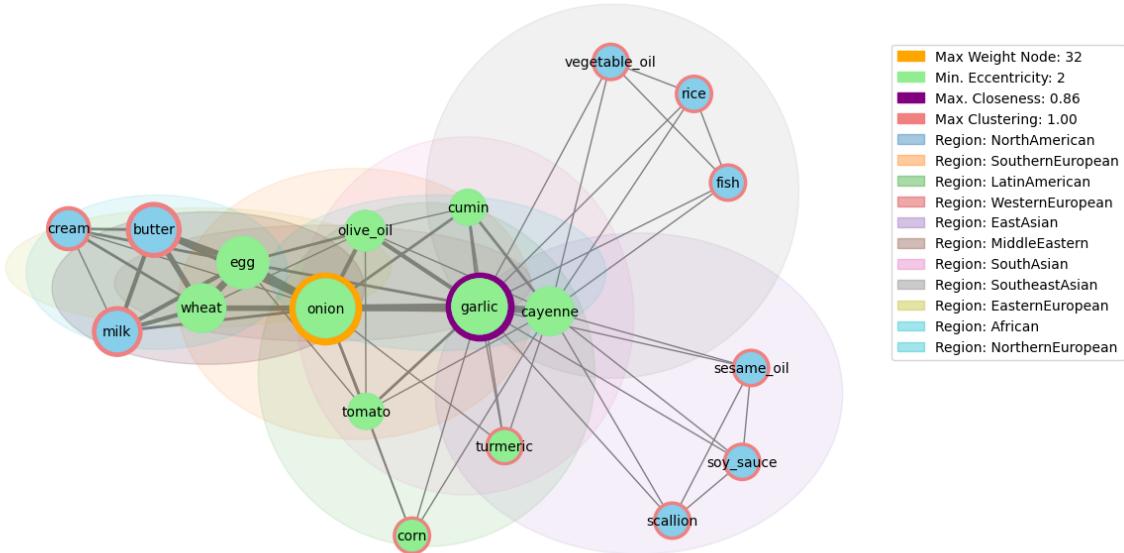
Ingrediencie (uzly) so špecifickými vlastnosťami sú vykreslené nasledovnými farbami:

- **Excentricita (zelená):** Excentricita vyjadruje najdlhšiu vzdialenosť uzla k iným uzlom v grafe. Zelené uzly sú ingrediencie, ktoré sú relatívne centrálne a dobre dostupné vo vzťahu k iným uzlom, ako **cibuľa, pšenica, vajce, paradajky, olivový olej a kurkuma**. Naznačuje to, že sa kombinujú s inými surovinami vo väčšine regiónov a môžeme ich považovať za základné

suroviny. Hodnota minimálnej centrality je 2, čo znamená, že medzi týmito ingredienciami a inými surovinami v grafe nie je vždy priamy vzťah (neobjavujú sa v rovnakých regiónoch), ale stále existuje minimálne jedna surovia, ktorá ich môže spojiť. (napríklad paradajky a maslo sa nevyskytujú vo výbere top 5 ingrediencií v rovnakom regióne, ale vajíčko ich premostí).

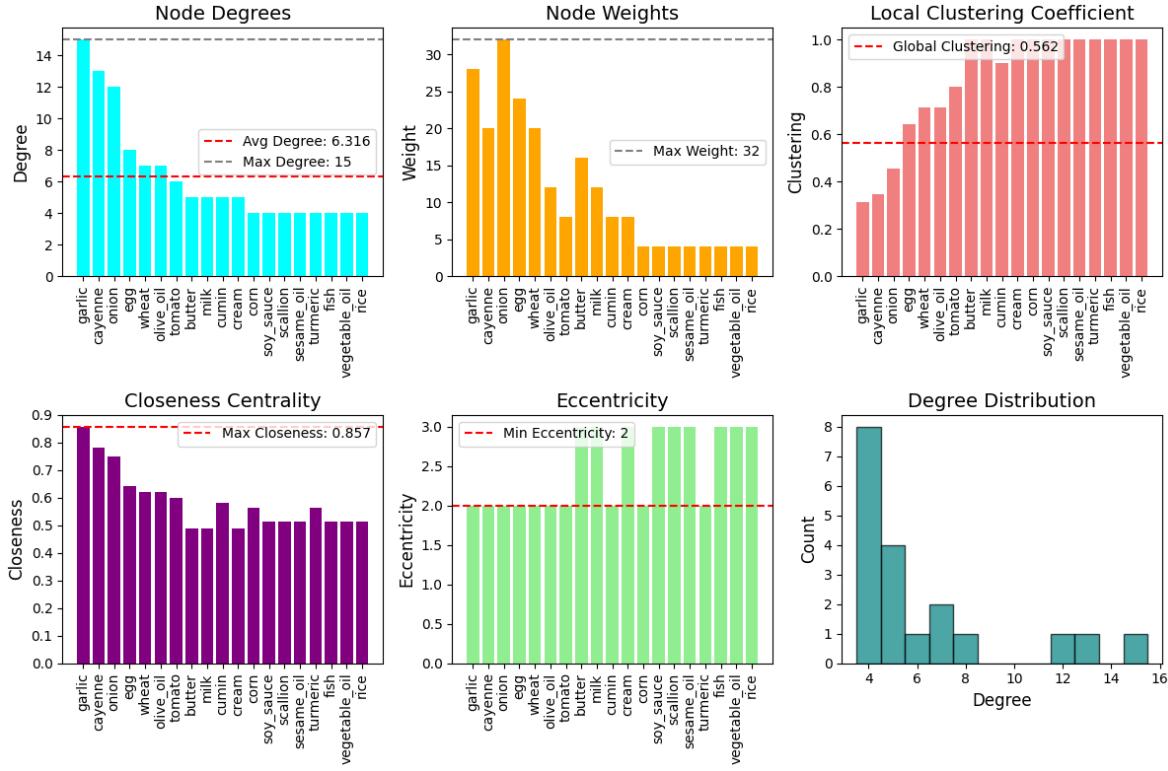
- **Centralita blízkosti (fialová):** Centralita blízkosti meria, ako blízko je uzol k ostatným uzlom v grafe. Ide o prevrátenú hodnotu priemernej vzdialenosť medzi uzlom a všetkými ostatnými uzlami – čím kratšia je priemerná vzdialosť, tým vyššia je centralita blízkosti. Graf ukazuje, že **cesnak** (garlic) má najvyššiu mieru centrality blízkosti, čo naznačuje, že je častokrát spájaný s inými ingredienciami, čím sa stáva centrálnou ingredienciou v rôznych kombináciach. Takisto ide o vrchol s najvyšším stupňom ( $d=15$ ).
- **Lokálny zhlukový koeficient (červené okraje):** Niektoré ingrediencie, ako **vegetable oil**, **rice**, a **fish**, majú vysoký zhlukový koeficient (1), čo znamená, že ich susedia sú kompletne prepojení a v regiónoch sa používajú spolu. V tomto prípade však ide o vrcholy s nízkym stupňom, ktoré sú typické len pre konkrétné regióny. Globálny zhlukový koeficient (0.5616) ukazuje, že viac ako polovica uzlov v grafe je prepojená v trojuholníkoch.

Projekcia bipartitného grafu na ingredience  
(zvýraznené vrcholy so špeciálnymi vlastnosťami)



Obr. 9: Projekcia bipartitného grafu na ingredience.

Hodnoty použitých metrík sme pre jednotlivé vrcholy vypísali aj do stĺpcových grafov (Obr. 10): stupeň vrchola, váha vrchola (súčet väh hrán, s ktorými vrchol je incidentný), lokálny zhlukový koeficient, centralita blízkosti, excentricita, rozdelenie stupňa vrchola.



Obr. 10: Prehľad štatistických metrič pre jednotlivé vrcholy.

## 4 Záver

V praci ukazujeme, že zvyklosti jednotlivých regiónov nie sú len otázkou chuti, ale aj dôsledkom historických, kultúrnych a geografických okolností. Severná Amerika vyniká svojou diverzitou receptov a ingrediencií, čo odráža prístahovalecké vplyvy a liberálnejší prístup k potravinovým prísadám. Naopak, regióny ako Južná Európa alebo Juhovýchodná Ázia majú silne definované charakteristiky založené na lokálnych surovinách a tradičných receptúrách.

Regióny ako Afrika a Ázia majú tendenciu používať najväčší počet ingrediencií na jeden recept, čo možno vysvetliť najmä vysokým zastúpením rôznych druhov korenín. Tieto koreniny sú často špecifické pre daný región.

Najčastejšie kombinovanou surovinou v rámci analyzovaných receptov sa stal cesnak. Táto surovina zohráva kľúčovú úlohu naprieč viacerými kuchyňami sveta vďaka svojej výraznej chuti a univerzálnemu použitiu. Vysoká miera jeho centrality v sieťovej analýze poukazuje na to, že cesnak slúži ako chutový most medzi rôznymi kulinárnymi tradíciami – od európskej, cez ázijskú až po latinskoamerickú kuchyňu.

Taktiež, sme odhalili ingrediencie, ktoré pôsobia ako mosty medzi rôznymi kultúrnymi tradíciami. Cesnak, cibuľa, olivový olej a ryža sa ukázali ako univerzálné suroviny prepájajúce veľkú časť svetových kuchýň. Aj keď mnohé regióny používajú podobné ingrediencie, spôsob ich kombinácie a miera ich zastúpenia v receptoch sa značne líši.