

Objectif : mesurer la **complexité** et la **similarité** entre des recettes, en utilisant l'entropie. Puis construire une carte de réseau de recettes par similarité.

Intro : les infos de difficulté/temps indiquées dans les recettes sont subjectives. Avec notre travail on propose de faire une évaluation objective basée sur la science.

Etat de l'art : il y a d'autres travaux pour trouver les similarités entre recettes / les ingrédients « complémentaires », qui produisent des graphes de représentation à la fin aussi.

Méthode :

- extraction des ingrédients : ils chopent la liste des ingrédients et virent les informations non nécessaires. Ils virent certains mots/symboles avec des regex, et ensuite ils nettoient manuellement ce qui reste (voir tableau ci dessous)

Step	Work	Method	Type of work	Contents
1	Data removal	Pattern matching	Removing the word, not the ingredient name	<ul style="list-style-type: none">Words that mean the amount: 'little', 'one', or 'two spoon'.Words that mean the shape: 'mashed', 'cut', or 'frozen'Symbols: '+', '=', or '-'.
2	Data refinement	Manually	Refining the word, not matching the name	<ul style="list-style-type: none">Same ingredients but different grammatical words.Same ingredients, but similar words: 'Anchovy Soup', 'Anchovy Broth', 'Anchovy Sauce'
			Refining the word attached an adjective	<ul style="list-style-type: none">Ingredients attached the adjectives: 'Mashed Garlic', 'Sliced Garlic', or 'Whole Bulb of Garlic' is refined as 'Garlic'.
			Classifying the similar names	<ul style="list-style-type: none">Similar ingredients are classified as an ingredient prototype: 'Rice Cooking Wine', 'Clear Strained Rice Wine', and 'Cooking Wine' are classified as the 'Rice Wine'.

- extraction des « cooking verbs »

—> pas d'infos données sur comment ils extraient, ils disent juste qu'ils ont viré les « informations non nécessaires pour filtrer e ne garder que les cooking verbs » mdr
Du coup je pense qu'on peut être générer une liste de tous les verbes présent dans le corpus, puis la regarder à la main et virer tous les verbes qui sont pas des cooking verbs, et sauvegarder le tout comme un « lexique des cooking verbs »

- analyse statistique des fréquences : quels sont les ingrédients qui reviennent le plus ? le moins? idem pour les cooking verbs.

- mesure de l'entropie : Utilisation de l'entropie pour ne pas se baser uniquement sur « est-ce que les recettes partagent bcp d'ingrédients? » parce que si ces ingrédients sont « sel » ou « ail », ça ne veut pas dire grand chose. Avec l'entropie on prend en compte le degré d'information de l'ingrédient (s'il est rare ou pas)

- **mesure de la complexité** : Ils séparent en 2 sortes de complexités : « complexité de préparation » (ingrédients) « complexité de la procédure » (cooking verbs)
entropie de la préparation = somme des entropies de tous les ingrédients.

Si l'entropie est basse : les ingrédients sont communs ou peu nombreux

—> La recette n'est pas complexe

Si l'entropie est haute : les ingrédients sont peu communs ou nombreux :

—> La recette est complexe

Puis le même chose pour les verbes :

entropie de la procédure = somme des entropies de tous les cooking verbes

Si l'entropie est basse : les actions sont habituelles ou peu nombreuses

—> La recette n'est pas complexe

Si l'entropie est haute : les actions sont peu habituelles ou nombreuses:

—> La recette est complexe

Puis ils font un graphe avec x=entropie préparation et y = entropie procédure :

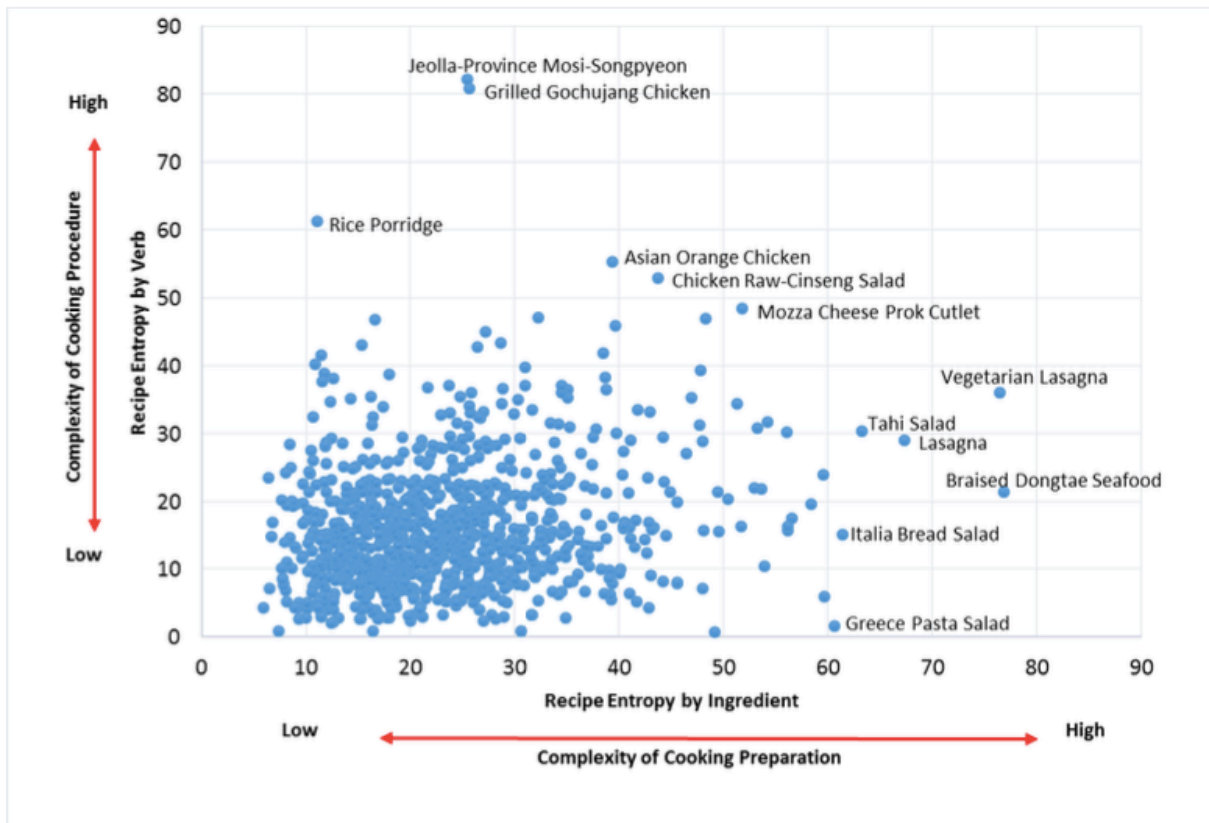


Figure 3. Distribution of recipe entropies by ingredient and cooking verb

- construction d'un réseau de similarité : mesure de la similarité entre recettes sur la base de l'entropie et construction d'un réseau.

Ma conclusion : leur méthode est intéressante mais c'est pas la méthode qui nous est demandée. Par contre ça vaut le coup d'en parler dans l'état de l'art pour dire que des travaux ont essayé de mesurer la complexité de recettes avec d'autres méthodes : par exemple par entropie des ingrédients et entropie des actions de la recette, mais que nous on veut tenter une approche inspirée du calcul de complexité des algorithmes.