

**UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE
PORTO ALEGRE – UFCSPA
DEPARTAMENTO DE NUTRIÇÃO
TECNOLOGIA EM GASTRONOMIA**

Michele dos Santos da Silva Tanus

PRINCÍPIOS DO PAREAMENTO DE SABORES NA COZINHA BRASILEIRA

Porto Alegre

2023

Michele dos Santos da Silva Tanus

Princípios do pareamento de sabores na cozinha brasileira

Trabalho de Conclusão de Curso de graduação apresentado ao Departamento de Nutrição da Fundação Universidade Federal de Ciências da Saúde de Porto Alegre, como requisito parcial para a obtenção do grau de Tecnólogo em Gastronomia.

Orientador: Prof. Dr. Juliano Garavaglia

Porto Alegre

2023

Catlogação na Publicação

dos Santos da Silva Tanus, Michele
Princípios do Pareamento de Sabores na Cozinha
Brasileira / Michele dos Santos da Silva Tanus. -- 2023.
52 p. : il., tab. ; 30 cm.

Monografia (trabalho de conclusão de curso) --
Universidade Federal de Ciências da Saúde de Porto
Alegre, Curso de Gastronomia, 2023.

Orientador(a): Juliano Garavaglia.

1. Gastronomia Computacional. 2. Pareamento de
Sabores. 3. Mineração de Dados. I. Título.

Michele dos Santos da Silva Tanus

Princípios do pareamento de sabores na cozinha brasileira

Trabalho de Conclusão de Curso de graduação apresentado ao Departamento de Nutrição da Fundação Universidade Federal de Ciências da Saúde de Porto Alegre, como requisito parcial para a obtenção do grau de Tecnólogo em Gastronomia.

Aprovado em: 05 / 07 / 2023

BANCA EXAMINADORA

Prof. Dr. Juliano Garavaglia

Instituição: Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

Prof. Dra. Cecília Dias Flores

Instituição: Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

Prof. Dra. Graziela Bruschi Brinques

Instituição: Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

AGRADECIMENTOS

Agradeço ao meu país pela oportunidade de ter aprendido tanto através de um ensino público, gratuito e de qualidade, nutrindo sempre a esperança de que esse ensino esteja disponível para cada vez mais pessoas.

Sou grata à minha família que me estimula a dar o meu melhor, com amor e incentivo, sempre orgulhosa e celebrando cada pequeno passo.

Sem a enorme dedicação dos professores do curso de gastronomia da UFCSPA eu não teria chegado até aqui. Agradeço à professora Isabel Kasper, cujo brilho no olho desde os primeiros dias de aula me contagiou e que nos corrigiu e guiou de uma forma muito humana. Sou muito grata também à professora Cheila de Paula e à professora Graziela Brinques pelas contribuições dadas a este trabalho desde as fases iniciais de sua concepção. Finalmente, o apoio e esforço do meu orientador, professor Juliano Garavaglia, foi o que possibilitou este trabalho e a ele sou grata.

O aprendizado foi guiado pelos professores, mas vivenciado em conjunto com colegas que enriqueceram essa jornada. Agradeço a todos que contribuíram diariamente com seu companheirismo e experiências que seguirão me acompanhando ao longo da vida. Em especial, agradeço à Carolina Knevez que esteve ao meu lado em tantos trabalhos e deu muito apoio para que este fosse realizado.

RESUMO

A gastronomia é composta por sabores que variam enormemente com os costumes locais, dando origem às diferentes cozinhas regionais. Apesar dessa variedade, o número de receitas existentes é muito menor que o número de receitas possíveis. Existem princípios que guiam o desenvolvimento de receitas, como o pareamento de ingredientes. Entender como esses princípios afetam a gastronomia é fundamental para o projeto criativo de cardápios. O objetivo deste trabalho é aplicar a gastronomia computacional para auxiliar no entendimento de como esses pareamentos ocorrem na cozinha brasileira. Para isso, foram aplicadas técnicas de mineração de dados a um grande volume de receitas obtido no site de receitas mais popular do país, usado diariamente por milhões de pessoas. A partir desses dados foi construída uma rede de sabores que conecta os ingredientes mais usados de acordo com seus compostos de sabor. Nossos resultados mostram que o trio cebola, alho e tomate formam a base da cozinha nacional. Os ingredientes mais prevalentes compartilham um grande número de compostos de sabor, dando suporte à hipótese do pareamento de sabores. Considerando a grande variação no número de compostos de sabor mapeados para cada ingrediente e a inexistência de informações sobre as quantidades de cada composto nos ingredientes, não é possível afirmar que a origem dessas combinações é o compartilhamento de sabores. É provável que outros fatores como textura e disponibilidade de alimentos em determinado período histórico tenham contribuído para a formação de nossa cozinha. No entanto, os dados obtidos apontam que é um fator importante e que pode ser usado como ferramenta para descobrir combinações pouco usuais que tenham potencial de funcionar bem em uma receita ou cardápio.

Palavras-chave: pareamento de sabores, gastronomia computacional, mineração de dados.

ABSTRACT

Gastronomy is composed of flavors that vary enormously within cultures, giving birth to regional cuisines. In spite of this variety, the number of existing recipes is much smaller than the number of possible recipes. It's likely there are principles that guide recipe development, such as flavor pairing. Understanding how these principles impact food development is an important tool for creative menu development. The goal of this work is applying computational gastronomy in order to understand how these pairings happen in Brazilian cuisine. In order to do so, we apply data mining to a large recipe website, used by millions of people. We then construct a food network that connects the most commonly used ingredients according to their sharing of flavor compounds. Our results show that a few key ingredients such as onion, garlic and tomato are the signature of Brazilian cuisine. The most common ingredients mostly share a large amount of flavor compounds, supporting the food pairing hypothesis. Given the variation in the number of known compounds for each ingredient and the absence of information on quantities of each flavor in foods, it wouldn't be possible to attribute the frequent pairings exclusively to flavor sharing. Texture and historical period of ingredients' availability likely played an important role. However, our data supports it as an important factor and as a tool for spotting unusual combinations that are likely to work well in a given recipe or menu.

Keywords: food pairing, computational gastronomy, data mining.

LISTA DE ILUSTRAÇÕES

Figura 1.1 — Diferença entre aprendizado de máquina e programação tradicional..	13
Figura 3.1 — Perfil dos usuários do site Tudo Gostoso.....	17
Figura 4.1 — Categorias das receitas usadas nas análises.....	25
Figura 4.2 — Representação dos ingredientes em um espaço bidimensional.....	26
Figura 4.3 — Histograma do número de ingredientes por receita.....	29
Figura 4.4 — Número de vezes em que as cinco receitas mais frequentes aparecem no conjunto de dados.....	30
Figura 4.5 — Compostos de sabor compartilhados entre os ingredientes mais prevalentes.....	33
Figura 4.6 — Grafo de ingredientes com as arestas representando compostos de sabor compartilhados.....	34
Figura 4.7 — Grafo de ingredientes mais prevalentes com a espessura das arestas representando o número de compostos de sabor compartilhados.....	35
Figura 4.8 — Correlação entre a prevalência de um par de ingredientes em receitas e o número de compostos de sabor compartilhados.....	36

LISTA DE TABELAS

Tabela 3.1 — Origem dos acessos ao site Tudo Gostoso.....	16
Tabela 3.2 — Dados de interesse obtidos do FlavorDB para o ingrediente brócolis.	18
Tabela 3.3 — Dados de interesse para a molécula cumarina, presente no brócolis.	19
Tabela 4.1 — Número médio de ingredientes por receita.....	29
Tabela 4.2 — Dez ingredientes mais frequentes e suas prevalências.....	31
Tabela 4.3 — Exemplo de matriz de correlação entre ingredientes e receitas.....	32
Tabela 4.4 — Primeiro e segundo ingrediente mais similar de acordo com a coexistência em receitas.....	32

SUMÁRIO

1 INTRODUÇÃO.....	8
1.1 REFERENCIAL TEÓRICO.....	9
1.1.1 Gastrofísica.....	9
1.1.2 Gastronomia Computacional.....	10
1.1.3 Computação cognitiva e criativa.....	11
1.1.4 Mineração de dados e aprendizado de máquina aplicados à gastronomia.....	12
2 OBJETIVOS.....	15
3 METODOLOGIA.....	16
3.1 POPULAÇÃO ALVO.....	16
3.2 AMOSTRAGEM.....	16
3.3 COLETA DE DADOS.....	17
3.3.1 Obtenção das receitas.....	17
3.3.2 Obtenção dos compostos de sabor.....	17
3.4 TRATAMENTO DOS DADOS.....	19
3.4.1 Remoção de duplicatas.....	19
3.4.2 Extração da lista de Ingredientes.....	20
3.4.3 Separação dos ingredientes das quantidades.....	20
3.4.4 Diferenciação dos ingredientes usados como medidas.....	21
3.4.5 Agregação de ingredientes.....	22
3.4.6 Prevalência de ingredientes e construção da rede de sabores.....	23
4.1 SEPARAÇÃO DOS INGREDIENTES.....	25
4.1.1 Ingredientes sem as quantidades.....	25
4.1.2 Limitações impostas pela linguagem coloquial.....	27
4.2 NÚMERO DE INGREDIENTES POR RECEITA.....	28
4.3 PREVALÊNCIA DE RECEITAS.....	29
4.4 PREVALÊNCIA DE INGREDIENTES.....	30
4.5 COEXISTÊNCIA DE INGREDIENTES EM RECEITAS.....	31
4.6 COMPARTILHAMENTO DE COMPOSTOS DE SABOR.....	32
5 CONCLUSÕES.....	38
REFERÊNCIAS.....	40
ANEXO A - EXEMPLO DE DADOS DE RECEITA.....	46
ANEXO B - EXEMPLO DE DADOS DE COMPOSTOS DE SABOR.....	48
ANEXO C - LISTA DE INGREDIENTES USADOS NAS ANÁLISES.....	50

1 INTRODUÇÃO

Os seres humanos evoluíram em meio à difícil tarefa de identificar e coletar alimentos que satisfaçam suas necessidades nutricionais sem desenvolver doenças associadas a seu consumo. Esse processo contribuiu para nossa dieta atual, que é influenciada por fatores como preferências sensoriais, valor nutricional, facilidade de produção e clima.

Apesar da grande variedade de alimentos disponíveis, há uma recorrência frequente de determinadas combinações nas diferentes cozinhas regionais, o que mostra que estamos explorando uma pequena fração das potenciais combinações. Esses padrões indicam que existem princípios que norteiam nossas escolhas por certas combinações de ingredientes em detrimento de outras (AHN et al., 2011).

O sabor é a base do desenvolvimento de uma receita ou produto alimentício. Além disso, alimentos são frequentemente consumidos em conjunto com outros alimentos ou bebidas, sendo relevante o estudo de suas combinações.

Diversos estudos sugerem que um dos princípios que guiam o desenvolvimento de receitas é o pareamento de sabores. Esse pareamento ocorre quando combinamos no mesmo preparo ingredientes que compartilham compostos de aroma e sabor (KLEPPER, 2011).

O pareamento de sabores pode impactar a disposição dos alimentos no prato, a combinação entre diferentes pratos ao longo de uma refeição e sua interação com diferentes bebidas (GALMARINI, 2020).

Alguns trabalhos analisam livros de receitas em busca de ingredientes e combinações usadas em determinadas cozinhas (KINOUCI et al., 2008). No entanto, Ahn et al. (2011) mostraram que o número e variedade de ingredientes usados pela população em geral difere da descrita em livros de gastronomia.

A disponibilização de milhões de receitas na Internet por usuários das mais diversas culturas ao redor do mundo torna possível expandir essas análises para além dos livros. Ahn et al. (2011) utilizaram dados de sites populares de receitas estadunidenses e coreanos para mapear como ingredientes e suas combinações são usados em determinadas cozinhas. O trabalho usa compostos de sabor

presentes em cada ingrediente e verifica se há coexistência de ingredientes com compostos de sabor similares. O estudo mostra que a cozinha estadunidense vale-se de combinações de ingredientes que compartilham compostos de sabor, enquanto a asiática dá preferência a ingredientes que não tenham compostos de sabor em comum.

Simas et al. (2017) propõem uma hipótese alternativa ao pareamento de sabores: as pontes de sabor. Assumem que dois ingredientes que não compartilham afinidade molecular podem se tornar afins através de uma cadeia de afinidades. Por exemplo, damasco e whisky não compartilham compostos de sabor mas podem se tornar compatíveis, assumindo que o tomate é compatível com damasco e com whisky, criando uma cadeia de afinidades.

Diversos estudos aplicam análises de pareamento de sabores a cozinhas regionais, como, por exemplo, a indiana (JAIN; RAKHI; BAGLER, 2015) e a árabe (TALLAB; ALRAZGAN, 2016). Apesar da existência de sites populares e colaborativos de receitas como Tudo Gostoso, com mais de 280 mil receitas, e Cybercook, com mais de 105 mil receitas, não existem análises similares realizadas no Brasil. Esses sites combinados recebem mais de 80 milhões de visitas mensais, mostrando sua relevância no dia-a-dia da cozinha de milhões de brasileiros.

Assim, este trabalho propõe o uso do grande volume de dados de receitas disponíveis para identificar como é feito o pareamento de sabores na cozinha brasileira. Através da coexistência de alimentos em receitas e dos compostos de sabor associados, identificar algumas bases que norteiam a cozinha popular no país.

1.1 REFERENCIAL TEÓRICO

1.1.1 Gastrofísica

Diversos fatores influenciam nossa experiência ao comer e beber, mas ainda não somos capazes de entender completamente a complexidade de sabor dos alimentos nem por que alguns estimulam mais nosso apetite que outros. O foco da gastronomia moderna está nos alimentos e em sua preparação. Por outro lado, a análise sensorial tem seu foco na percepção que as pessoas têm de atributos

sensoriais do que elas comem e bebem em laboratório: o quão doce é, o quão intenso é, se elas gostaram do sabor (SPENCE, 2017).

A gastrofísica nasceu com a ideia de ser uma nova técnica para medir e entender os fatores que influenciam as respostas de pessoas reais a comidas e bebidas reais, idealmente sob as condições mais naturais possíveis. Ela é o estudo científico dos fatores que influenciam a experiência multisensorial durante a degustação de alimentos e bebidas. O termo vem da junção de gastronomia e psicofísica: a gastronomia se concentra nas experiências culinárias que são alvo de inspiração para a pesquisa nessa área, enquanto a psicofísica tem seu foco no estudo científico da percepção (SPENCE, 2017).

Embora estudos em gastrofísica apresentem críticas ao pareamento de sabores, já que ressaltam a importância multifatorial, como textura, temperatura e sequencialidade, reconhecem sua potencialidade como ferramenta de criatividade (BLUMENTHAL, 2008).

Um dos trabalhos recentes em gastrofísica aborda a importância do posicionamento de elementos no prato de acordo com a harmonia de sabores. Spence et al. (2017) defendem que o empratamento vem priorizando a estética em detrimento da combinação de sabores. Segundo os autores, existe uma ordem temporal na percepção de sabores que é, muitas vezes, ignorada.

1.1.2 Gastronomia Computacional

Nas últimas duas décadas a análise de dados em larga escala revolucionou as ciências biológicas, que passaram por uma explosão na quantidade de dados experimentais provenientes de novas tecnologias, como o sequenciamento de DNA. Grandes volumes de dados também estão modificando metodologias nas ciências sociais em virtude dos dados gerados pela comunicação de dispositivos móveis e redes sociais. Até mesmo as artes e humanidades estão se adaptando à crescente digitalização de fontes literárias e históricas. A ideia de que a ciência dos alimentos pode se beneficiar de análise de dados em larga escala deu origem ao termo gastronomia computacional (AHNERT, 2013).

Vertentes da gastronomia computacional defendem que é possível produzir recomendações de pareamentos inovadores e com maiores chances de sucesso baseando-se na identificação de compostos de sabor em comum. Por exemplo, um composto químico como o acetal pode ser encontrado na maçã, no whisky e na beterraba (BURDOCK, 2010). Morango e vinho branco também compartilham compostos de sabor. Essa abordagem envolve mineração de dados e análise de redes, usando o grande conjunto de receitas publicamente disponíveis, e química dos sabores (AHNERT, 2013).

A gastronomia computacional começou com a hipótese de pareamento de sabores: se dois ingredientes compartilham compostos de sabor, eles produzirão bons resultados juntos. Uma série de combinações pouco usuais foram criadas, como chocolate branco e caviar (ambos contendo aminas) ou fígado de porco e jasmim (ambos contendo indol). Outras combinações bem sucedidas incluem ostras e maracujá ou bananas e salsa (SPENCE; WANG; YOUSSEF, 2017).

No entanto, experimentos mostraram que nem sempre esse pareamento fornece indícios das melhores combinações de ingredientes (KLEPPER, 2011). Ele pode ser usado como ferramenta de auxílio à criatividade, mas não substitui intuição e imaginação (BLUMENTHAL, 2008).

As pesquisas mais recentes na área enfatizam o uso de elementos culturais para entender formas em que os diferentes sabores são combinados. Esses estudos mostram que ingredientes que ocorrem juntos em muitas cozinhas ocidentais tendem a compartilhar compostos, enquanto na cozinha oriental ocorre o oposto.

1.1.3 Computação cognitiva e criativa

Criatividade computacional é uma área emergente da inteligência artificial que coloca os computadores no centro do processo criativo. De forma geral, a criatividade envolve um processo de geração que produz diversas ideias e um processo seletivo que determina quais são as melhores. A maior dificuldade está em obter-se um processo de seleção adequado (VARSHNEY et al., 2019).

Varshney et al. (2019) propuseram uma abordagem baseada em dados para o problema e apresentaram um sistema que produz receitas e menus, podendo

operar de forma autônoma ou semi-autônoma com interação humana. Os autores também avaliaram a velocidade, a variedade e a veracidade dos dados em computação criativa.

Cromwell et al. (2015) examinaram o potencial da criatividade nas artes culinárias. Os autores desenvolveram um chef artificial que propõe saladas a partir do desenvolvimento de um modelo estatístico para dar pontuação a receitas e de um algoritmo de busca para explorar o espaço amostral e descobrir novas combinações de ingredientes. As receitas do chef artificial foram comparadas através de testes cegos a receitas de chefs humanos, obtendo performance similar.

Outro exemplo de computação cognitiva aplicada à gastronomia é o chef Watson. Ele faz parte da missão da IBM de desenvolver aplicações de computação cognitiva que ajudem pessoas a descobrirem novas ideias, desde criar receitas surpreendentes até contribuir com a pesquisa médica. Aplicações cognitivas como o chef Watson podem analisar vastas quantidades de dados, usando técnicas de processamento de linguagem natural. Essa foi uma das razões para o Watson ter ganhado o concurso Jeopardy!, competição de conhecimentos gerais da televisão americana. O chef Watson processou a composição química de centenas de ingredientes diferentes e analisou milhares de receitas do site Bon Appétit. Através da combinação dos dados e da detecção de padrões, o chef artificial aprendeu a sugerir até quatro ingredientes diferentes que combinam entre si (BRANDT, 2017).

1.1.4 Mineração de dados e aprendizado de máquina aplicados à gastronomia

Cozinheiros e pesquisadores na área de alimentação estão usando a mineração de dados como uma ferramenta para explorar, entender e inovar. A partir de vastas quantidades de dados disponíveis, ela busca descobrir padrões e gerar *insights* para informar criações gastronômicas.

A mineração de dados como parte da gastronomia computacional envolve a extração de informações de grandes conjuntos de dados relacionados a diferentes aspectos dos alimentos, como ingredientes, receitas, técnicas culinárias, perfis de sabores e preferências de consumidores. Esse processo usa uma larga gama de técnicas e algoritmos da área de aprendizado de máquina e estatística para descobrir relações nos dados (GOEL; BAGLER, 2022).

Aprendizado de máquina é uma área da inteligência artificial cujo foco é o desenvolvimento de algoritmos ou modelos que possibilitem predições sem a necessidade da especificação de regras. Como pode ser visto na Figura 1.1, na programação tradicional, o desenvolvedor de software escreve um conjunto de regras de negócio que serão aplicadas aos dados de entrada. No aprendizado de máquina, o programador fornece um grande número de exemplos e saídas correspondentes e o programa infere as regras a partir dos dados. Isso permite que regras que não são conhecidas por nós sejam descobertas.

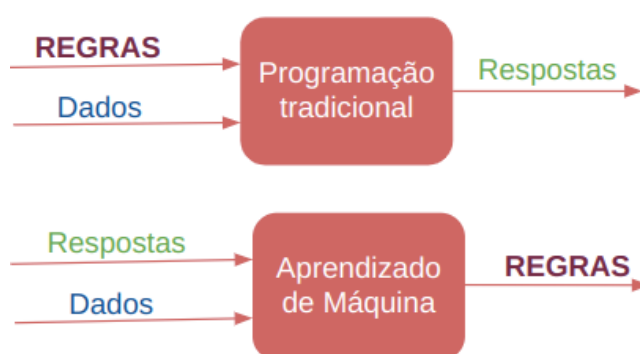


Figura 1.1. Diferença entre aprendizado de máquina e programação tradicional.

Por exemplo, suponha que desejamos classificar automaticamente uma receita como sobremesa ou salada. Podemos escrever algumas regras como “se contiver alface é uma salada”, “se incluir morango é uma sobremesa”. Porém, além de tedioso, dificilmente conseguiremos modelar todas as possibilidades. Usando aprendizado de máquina, passamos ao computador listas de ingredientes e classificações conhecidas (normalmente atribuídas anteriormente por pessoas) de um grande número de receitas. Durante o aprendizado, o programa recebe essas entradas e ajusta suas regras de forma que elas gerem as saídas esperadas. A partir daí, ele é capaz de classificar uma receita nunca antes vista.

O aprendizado de máquina é tradicionalmente dividido em supervisionado e não supervisionado. O aprendizado supervisionado se enquadra no exemplo da classificação de receitas que vimos acima: é necessário fornecer uma entrada e uma saída para que as regras sejam inferidas. No aprendizado não supervisionado não temos uma resposta previamente conhecida para fornecer ao modelo, mas podemos usar informações intrínsecas aos dados (BISHOP; NASRABADI, 2006). Os modelos de linguagem frequentemente se enquadram nessa categoria, já que existe uma

estrutura implícita em nossa comunicação. Por exemplo, na frase “depois do almoço esqueci de escovar os _____” conseguimos preencher a lacuna facilmente. Os modelos de linguagem são treinados em grandes volumes de texto e, portanto, também são capazes de aprender sequências naturais em nossa língua.

2 OBJETIVOS

O objetivo geral deste trabalho é identificar como os compostos de sabor são pareados em preparos do site de receitas mais acessado do Brasil.

Os objetivos específicos são:

1. Identificar os principais ingredientes que compõem as receitas do site Tudo Gostoso;
2. Correlacionar o uso dos principais ingredientes de acordo com seus compostos de sabor;
3. Identificar os ingredientes que diferenciam a cozinha brasileira, representada pelas receitas obtidas do site, de outras cozinhas descritas na literatura.

3 METODOLOGIA

3.1 POPULAÇÃO ALVO

A população alvo é formada pelos usuários de sites de receitas nacionais. O site Tudo Gostoso, o mais popular site de receitas do país e nono do mundo, tem mais de 280 mil receitas e cerca de 60 milhões de visitas mensais. Nele são compartilhadas receitas por mais de 12 milhões de usuários.

Embora seja um site nacional, alguns dos acessos são feitos de fora do país, como pode ser visto na Tabela 3.1.

Tabela 3.1. Origem dos acessos ao site Tudo Gostoso. Fonte: ALEXA, 2022.

País	Percentual de acessos
Brasil	89,2%
Angola	3,3%
Portugal	2,2%
Japão	2,0%
Estados Unidos	1,1%

O perfil dos usuários que acessam o site é bastante variado, conforme ilustrado na Figura 3.1. A maioria possui curso superior e há uma predominância de usuários do sexo masculino. Esse perfil é traçado de forma automática, de acordo com outros sites acessados pelo usuário e sua confiabilidade é moderada.

3.2 AMOSTRAGEM

São usados dados de todas as receitas de todos os usuários do site. No entanto, a demografia dos usuários de internet não é a mesma do país. Os sites não disponibilizam informações detalhadas sobre usuários e, portanto, não é possível realizar a amostragem dentro dos diferentes segmentos da população.

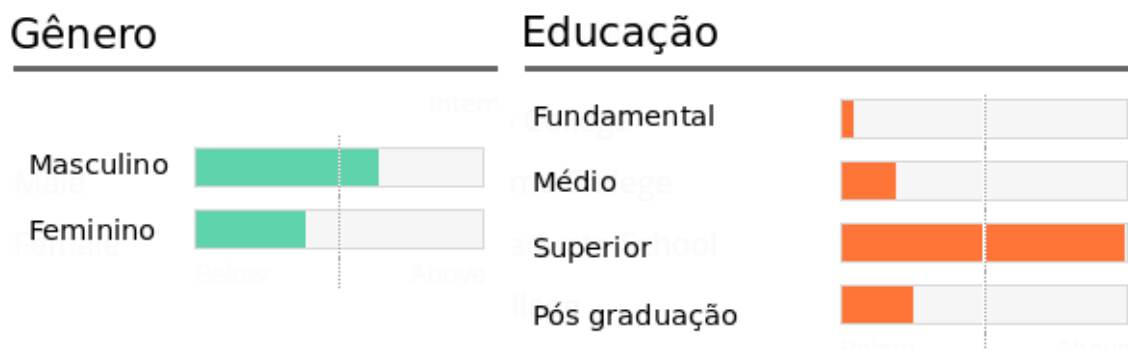


Figura 3.1. Perfil dos usuários do site Tudo Gostoso. Fonte: ALEXA, 2022.

3.3 COLETA DE DADOS

3.3.1 Obtenção das receitas

Um total de 185.107 receitas foram obtidas a partir do site Tudo Gostoso através de um script de computador criado como parte deste trabalho. Essas receitas foram obtidas no formato de arquivos JSON, frequentemente usado para troca de informações entre programas. Dentre as informações relevantes para esta pesquisa, estão inclusos os seguintes dados:

- **Id:** identificador único da receita, composto por dígitos e letras. Permite diferenciar receitas diferentes de mesmo título.
- **Título:** definido pelo usuário que criou a receita.
- **Avaliação:** nota de 0 a 100 dada pelos usuários do site.
- **Número de comentários:** comentários dos usuários na receita.
- **Número de favoritos:** número de usuários que favoritaram a receita.
- **Categoria:** categoria da receita, como, por exemplo, “doces e sobremesas”
- **Lista de ingredientes:** Ingredientes usados na receita.

O Anexo A apresenta um exemplo completo dos dados obtidos do site.

3.3.2 Obtenção dos compostos de sabor

Existem diferentes bases de dados de moléculas em alimentos. Porém, poucas têm foco específico nos compostos de aroma e sabor. Para este trabalho foi selecionado o FlavorDB (GARG et al., 2018) por ter publicações e ser aberto ao uso pela comunidade científica.

O FlavorDB disponibiliza compostos de sabor de 937 itens alimentícios. Os dados disponibilizados foram baixados por meio de um programa criado pelos autores deste trabalho para este fim. Eles também estão no formato JSON e contêm as seguintes informações de interesse para este trabalho:

1. **Nome:** nome do alimento.
2. **Categoria:** categoria do alimento. Por exemplo, laticínios ou frutas.
3. **URL do alimento:** um link para a descrição do alimento na Wikipedia.
4. **Moléculas:** uma lista contendo moléculas de sabor com dados associados:
 - a. **Nome comum:** nome popular da molécula.
 - b. **ID no FoodDB e no Pubchem:** identificador na molécula em diferentes bancos de dados de moléculas.
 - c. **Nome IUPAC:** nome dado pela União Internacional de Química Pura e Aplicada.
 - d. **InChI:** Identificador Químico Internacional da IUPAC é um identificador textual para substâncias químicas.
 - e. **SMILE:** representação textual da estrutura da molécula.
 - f. **Amargo:** indicador (verdadeiro ou falso) de se a molécula é amarga.
 - g. **Perfil de sabor:** descritivo textual do sabor associado à molécula: doce, frutado, amargo etc.

Outras informações químicas sobre as moléculas como peso molecular e volume tridimensional estão disponíveis, mas não foram usadas neste trabalho. As tabelas 3.2 e 3.3 detalham os campos com exemplos. O Anexo B inclui um exemplo completo dos dados disponíveis no FlavorDB.

Tabela 3.2 - Dados de interesse obtidos do FlavorDB para o ingrediente brócolis.

Dado	Valor original	Valor traduzido
Nome	Broccoli	Brócolis
Categoria	Cabbage	Couve (Brassica oleracea)
URL	https://en.wikipedia.org/wiki/Broccoli	https://pt.wikipedia.org/wiki/Br%C3%B3colis

Tabela 3.3 - Dados de interesse para a molécula cumarina, presente no brócolis.

Dado	Valor
Nome comum	Coumarin (Cumarina em português)
ID no Pubchem	323 (https://pubchem.ncbi.nlm.nih.gov/compound/323)
ID no FoodDB	FDB030742 (https://foodb.ca/compounds/FDB030742)
Nome IUPAC	chromen-2-one
InChI	InChI=1S/C9H6O2/c10-9-6-5-7-3-1-2-4-8(7)11-9/h1-6H
Smile	<chem>C1=CC=C2C(=C1)C=CC(=O)O2</chem>
Amargo	Sim

Para a associação entre os ingredientes encontrados nas receitas e aqueles fornecidos pelo banco de dados de moléculas de sabor, foi feita uma tradução dos nomes dos ingredientes.

3.4 TRATAMENTO DOS DADOS

3.4.1 Remoção de duplicatas

Para evitar que receitas muito populares, compartilhadas múltiplas vezes por diferentes usuários, afetem as análises de prevalência dos ingredientes, foram removidas aquelas com o mesmo título. Como critérios para a seleção, foram usados, nesta ordem:

1. Nota dada pelos usuários;
2. Número de comentários;
3. Número de usuários que colocaram a receita em seus favoritos.

Esses critérios foram adotados de acordo com a frequência em que estavam disponíveis: a maioria das receitas têm nota, boa parte tem comentários e algumas estão nos favoritos. A literatura não define uma metodologia para tratar esses casos, até pela diferença nos dados fornecidos pelos diferentes sites de receitas.

Não foi feita uma tentativa de remoção de receitas cujo título é diferente mas usam os mesmos ingredientes. O modo de preparo pode, a partir dos mesmos produtos, gerar resultados completamente distintos, não servindo como único fator para detectar duplicações.

3.4.2 Extração da lista de Ingredientes

O site fornece uma lista de ingredientes para cada receita. Por ser de formato livre, muitos usuários colocam também instruções e equipamentos necessários para a reprodução da receita.

A seguir um exemplo de lista de ingredientes para a receita “Canjica Divina”:

2 xícaras de canjica
2 1/2 litros de água
1 caixinha de leite condensado
1 vidro de leite de coco
200 gramas de coco ralado bem fininho
A medida da caixinha do leite condensado de leite
2 pedaços de canela
10 cravos-da-índia
canela em pó para decorar

Como pode ser observado, os ingredientes vêm acompanhados de suas quantidades e, em alguns casos, descritivos como “para decorar”.

3.4.3 Separação dos ingredientes das quantidades

Para uma extração apenas dos ingredientes de cada item foi usada uma técnica de processamento de linguagem natural. Ela consiste em um modelo grande de linguagem (LLM, *Large Language Model*) similar ao usado pelo ChatGPT (OPENAI, 2023). O algoritmo é treinado em *corpus* de texto para aprender as palavras mais prováveis em uma determinada sequência. Neste trabalho foi usado um modelo pré-treinado e disponibilizado pela Meta para pesquisa científica denominado LLaMA (TOUVRON et. al, 2023).

A seguir um exemplo de suas entradas e saídas:

Entrada: "Retorne lista de ingredientes sem as quantidades:
100g de queijo ralado, 3 ovos, 1/2 xícara de óleo"
Saída: Queijo ralado, ovos, óleo

Detalhes técnicos sobre o modelo de linguagem utilizado estão disponíveis no Apêndice A.

3.4.4 Diferenciação dos ingredientes usados como medidas

Embalagens de alguns dos ingredientes são usadas como medidas para outros, o que pode dificultar a identificação do ingrediente correto. Na receita de canjica da seção 3.4.2 temos o exemplo “a medida da caixinha do leite condensado de leite”.

Não é possível definir um padrão fixo para detectar esses casos, como pode ser observado nas frases a seguir.

1. Um copo de iogurte de morango;
2. um copo de iogurte de conhaque.

No primeiro caso, o ingrediente é o iogurte de morango, enquanto no segundo, o ingrediente é o conhaque e sua quantidade é medida no copo de iogurte.

Em diversos casos o ingrediente cuja embalagem está sendo usada também está presente na receita, como é frequentemente o caso da “caixa de creme de leite” ou da “lata de milho”. Nesses casos, uma análise que os incluísse não seria incorreta. Porém, determinadas medidas se parecem com ingredientes como é o caso de “uma colher de café”. Outras se referem a embalagens de ingredientes que não estão na receita mas que são frequentemente reaproveitadas, como é o caso do “copo de requeijão”.

A fim de mitigar esse problema e fazer a completa extração dos ingredientes das receitas, foi criado e treinado um novo modelo de inteligência artificial. Seu

objetivo é, a partir de uma entrada como “um copo de iogurte de conhaque”, prever o ingrediente “conhaque”.

Esse algoritmo é baseado no conceito de *embedding* (MIKOLOV et al., 2013): a partir de um texto de entrada, definir um valor para cada uma das palavras de acordo com o contexto em que elas aparecem. Por exemplo, num texto listando ingredientes temos várias vezes “uma rodela de limão” e “uma rodela de laranja”, mas dificilmente “uma rodela de feijão”. Dessa forma, o modelo dá valores parecidos para limão e laranja.

Para aprendizado do modelo de predição do ingrediente final, uma lista de ingredientes similar ao do exemplo anterior “canjica divina” foi anotada de forma a incluir o resultado desejado. Por exemplo:

<p>Entrada: A medida da caixinha do leite condensado de leite</p> <p>Anotação: Leite</p>
--

Foram rotulados manualmente 10.000 de um total de 62.932 itens para o treinamento. Detalhes técnicos sobre o treinamento e utilização deste modelo de aprendizado de máquina estão descritos no Apêndice A.

3.4.5 Agregação de ingredientes

Para as análises, ingredientes cujos sabores são similares, como amaretto e amêndoa foram unificados. Ingredientes com a mesma base, porém com funções diferentes foram mantidos. Por exemplo, glicose de milho, amido de milho, farinha de milho e milho foram considerados de forma independente.

Alguns itens podem ou não ser acompanhados de sabor adicional. Nesses casos, o sabor também é adicionado à lista de ingredientes quando presente. Por exemplo, quando o item é “sorvete de maracujá”, maracujá também é incluído. Não foi feita distinção entre aromas naturais e artificiais, já que essa informação não está disponível.

Certos ingredientes são mencionados tanto de forma mais geral quanto em subcategorias. Por exemplo, 100 g de queijo ou 100 g de parmesão. As subcategorias mais prevalentes foram mantidas, as demais foram integradas à

categoria mãe. Queijo minas ocorre em muitas receitas e foi mantido como ingrediente. Já o queijo comté foi incluído como queijo.

Para a avaliação dos compostos de sabor, certas subcategorias foram unidas pois não havia informações para todos os tipos de queijos, pimentas ou feijões.

Bebidas, quando possível, foram representadas pelo sabor do seu componente principal. Por exemplo, o licor de cacau foi representado por cacau, o refrigerante guaraná foi representado pelo fruto guaraná. Outras bebidas, como whisky ou vermute, que não têm um único ingrediente base e que estão catalogadas em bancos de dados de compostos de sabor foram mantidas.

3.4.6 Prevalência de ingredientes e construção da rede de sabores

Para avaliar a prevalência foi usada a metodologia proposta por Ahnet et al. (2013):

$$P_i^c = n_i^c / N_c$$

n é o número de receitas que contêm o ingrediente i em uma dada cozinha c e N é o número total de receitas naquela cozinha.

A prevalência é um valor entre 0 e 1 que reflete o quanto determinado alimento aparece em um conjunto de receitas. Quando multiplicada por cem, pode ser interpretada como o percentual de receitas em que esse ingrediente aparece.

Para melhor visualização do compartilhamento dos compostos de sabor entre os ingredientes mais prevalentes no conjunto de receitas, foi criado um gráfico de rede de sabores similar ao proposto em (AHN et al., 2011). A rede de sabores é um grafo em que os ingredientes são representados por nodos e as conexões entre eles indicam compostos de sabor em comum. Cada uma dessas conexões tem um peso associado de acordo com o número de moléculas de sabor compartilhadas entre os dois alimentos. Por exemplo, cebola e alho têm 120 compostos em comum, originando uma conexão de peso 120, enquanto cebola e ovo dão origem a uma conexão de peso 22.

Para o desenho desse grafo é usada a visualização do tipo *Spring* (BERTAULT, 2000), ou massa-mola, em que os pesos das conexões são usados como se representassem a força de uma mola que atrai os ingredientes entre si. Quanto maior a força da mola (número de ingredientes compartilhados), mais próximos eles estarão na visualização.

Mesmo restringindo a visualização aos cem alimentos mais prevalentes, o grafo resultante apresentou mais de duas mil conexões, o que gerou muita sobreposição e inviabilizou sua utilização para identificação dos relacionamentos entre ingredientes.

Com objetivo de mitigar este problema, foi empregada a técnica denominada extração de *backbone* (SERRANO et al., 2009). Ela foi desenvolvida para visualização de redes complexas e simplifica o grafo de acordo com a relevância estatística de suas conexões.

4 RESULTADOS

As análises a seguir foram feitas para 113.194 receitas, resultantes da remoção de duplicatas, e foram considerados 473 ingredientes. As receitas são divididas pelo site em categorias, como pode ser visualizado na Figura 4.1.

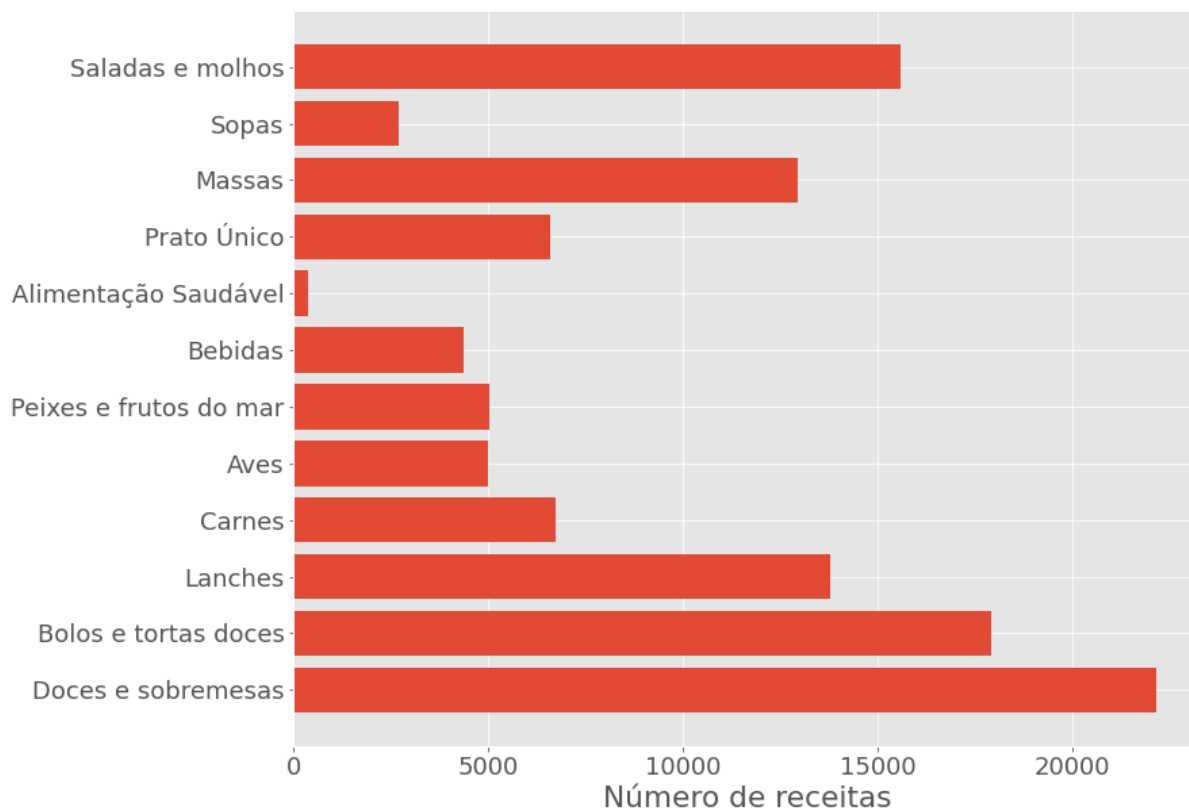


Figura 4.1. Categorias das receitas usadas nas análises.

Como as receitas doces, em grande quantidade no site, têm um perfil diferente das salgadas, optou-se por fazer também análises por categoria.

4.1 SEPARAÇÃO DOS INGREDIENTES

4.1.1 Ingredientes sem as quantidades

Os ingredientes aparecem com suas quantidades e usos, como, por exemplo, "uma colher de sopa de cacau para decorar". Com a utilização do modelo de linguagem (LLM), foram extraídos apenas os alimentos listados.

Apesar de permitir uma filtragem de grande parte das receitas, ele não foi capaz de obter a totalidade dos ingredientes devido a ambiguidades que ocorrem na descrição das receitas.

O saída a seguir exemplifica uma das falhas:

Entrada: "Retorne a lista de ingredientes sem as quantidades:
100g de queijo ralado, 3 ovos, 1/2 copo de requeijão de óleo"
Saída: Queijo ralado, ovos, *requeijão de óleo*

O modelo de linguagem não extraiu todos os ingredientes corretamente, obtendo um “requeijão de óleo”.

Para resolver esse problema, o modelo de inteligência artificial treinado neste trabalho transformou cada um dos ingredientes em representações numéricas. A Figura 4.2 exemplifica o espaço de ingredientes em duas dimensões. O limão, indicado no gráfico, tem um ponto logo ao seu lado que é a laranja e está próximo a outras frutas e mais distante de verduras.



Figura 4.2. Representação dos ingredientes em um espaço bidimensional.

Em alguns casos ingredientes de diferentes categorias aparecem próximos, porém pode ser um artefato da visualização 2D. Se estivessem em 3D, alguns poderiam estar mais para frente e outros mais para o fundo da imagem, não necessariamente colidindo. O modelo usado tem 100 dimensões (cada palavra é representada por 100 valores) e portanto não é possível visualizá-lo diretamente.

O modelo resultante apresentou acurácia de 89% nos dados de validação (frases não usadas durante o treinamento) e foi usado para extrair os demais ingredientes.

Cerca de 10% dos ingredientes resultantes passaram por uma inspeção manual a fim de identificar erros. A maioria dos erros ocorreu em virtude de algum ingrediente não presente na amostra de treinamento, como diferentes tipos de pimentas, que tiveram de ser adicionados ao modelo final. O uso dessa abordagem levou a um grande número de ingredientes inicialmente desconhecidos, o que reforça a necessidade de uma abordagem baseada em dados em vez de tentar manualmente formular uma lista exaustiva de ingredientes. Como exemplo, alguns dos peixes descobertos pelo modelo: albacora, guaricema, gurijuba mapará, paru e surubim.

Como o resultado de qualquer modelo de inteligência artificial, nem todos os ingredientes previstos estão corretos. Porém, com a estimativa de acurácia obtida e o grande volume de dados de entrada, as análises ainda são capazes de produzir resultados generalizáveis.

4.1.2 Limitações impostas pela linguagem coloquial

Na linguagem do dia-a-dia, alguns ingredientes diferentes com a mesma base são usados de forma intercambiável e não foi possível eliminar essas ambiguidades. Por exemplo, “trigo até dar o ponto” muitas vezes refere-se à farinha de trigo.

Outro desafio são ingredientes distintos conhecidos pelo mesmo nome. Por exemplo, a palavra pimenta quando usada sozinha pode se referir à pimenta do reino, fruto seco da Piperaceae (em bolinhas), ou à pimenta dedo de moça (Capsicum). Elas têm características sensoriais distintas. Este trabalho assumiu que a palavra pimenta, quando não especificada, designa a Piperaceae.

A palavra azeite frequentemente refere-se ao de oliva, mas, muitas vezes, é intercambiável com óleo vegetal e não é possível distinguir somente pelo texto da receita.

Foi possível notar também que muitas vezes manteiga e margarina são confundidas no texto. Por exemplo, a expressão “manteiga qualy” aparece quando, na verdade, está se referindo a margarina (não existe uma manteiga desta marca).

O uso do plural e singular algumas vezes indica diferenças importantes entre os ingredientes. Por exemplo, cebolinha frequentemente é o tempero verde enquanto a palavra cebolinhas normalmente indica cebolas. Bananinha é um corte de carne enquanto a palavra bananinhas é normalmente o diminutivo de bananas. Não foi possível detectar quando cebolinha referia-se a cebolas pequenas.

Certos ingredientes são frequentemente associados a marcas, como catupiry para requeijão, pomarola e salsaretti para molho de tomate ou ajinomoto e sabor ami para glutamato monossódico. Esses itens foram associados manualmente ao ingrediente correto.

O Anexo C apresenta uma lista completa dos ingredientes resultantes usados nas análises.

4.2 NÚMERO DE INGREDIENTES POR RECEITA

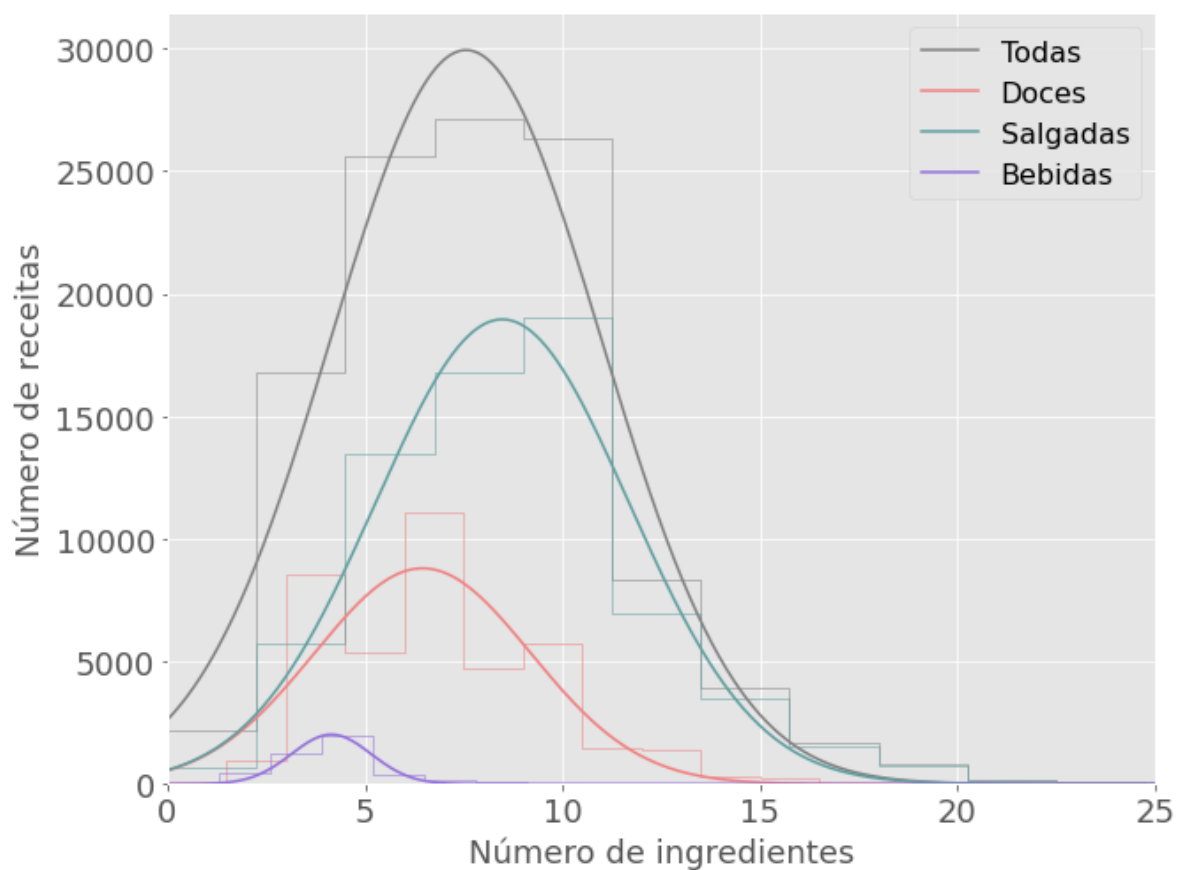
A média de ingredientes nas receitas é de 7,8. As receitas salgadas têm mais ingredientes que as doces e ambas têm mais que as bebidas.

O número de ingredientes por receita encontrado é menor que 9,38, valor reportado para receitas da América Latina por Ahn et al. (2011). Cabe ressaltar que nesse trabalho os autores usaram receitas publicadas em sites dos Estados Unidos que estavam indicadas como sendo receitas típicas da América Latina e não de sites dos próprios países.

As distribuições do número de ingredientes para cada uma das categorias podem ser observadas na Tabela 4.1 e Figura 4.3. Poucas receitas têm mais de 15 ingredientes. Uma inspeção manual dessas receitas com muitos ingredientes indicou uma preponderância de diferentes pré-preparos, como, por exemplo, a massa e o recheio de uma empada.

Tabela 4.1. Número médio de ingredientes por receita.

Receitas	Número médio de ingredientes	Desvio padrão
Todas	7,8	3,4
Doces	6,7	2,7
Salgadas	8,6	3,5
Bebidas	4,1	1,6

**Figura 4.3.** Histograma do número de ingredientes por receita.

4.3 PREVALÊNCIA DE RECEITAS

Excepcionalmente para essa análise foram usadas todas as receitas, incluindo duplicatas. Os títulos das receitas foram normalizados com a remoção de maiúsculas e acentos de forma que “Empadão Goiano” e “empadao goiano” representam a mesma receita.

A receita mais comum foi a da torta de limão, com 408 ocorrências. Dentre as salgadas, uma receita tipicamente brasileira, pão de queijo, com 372 receitas listadas. A Figura 4.4 lista as cinco receitas mais frequentes em cada categoria.

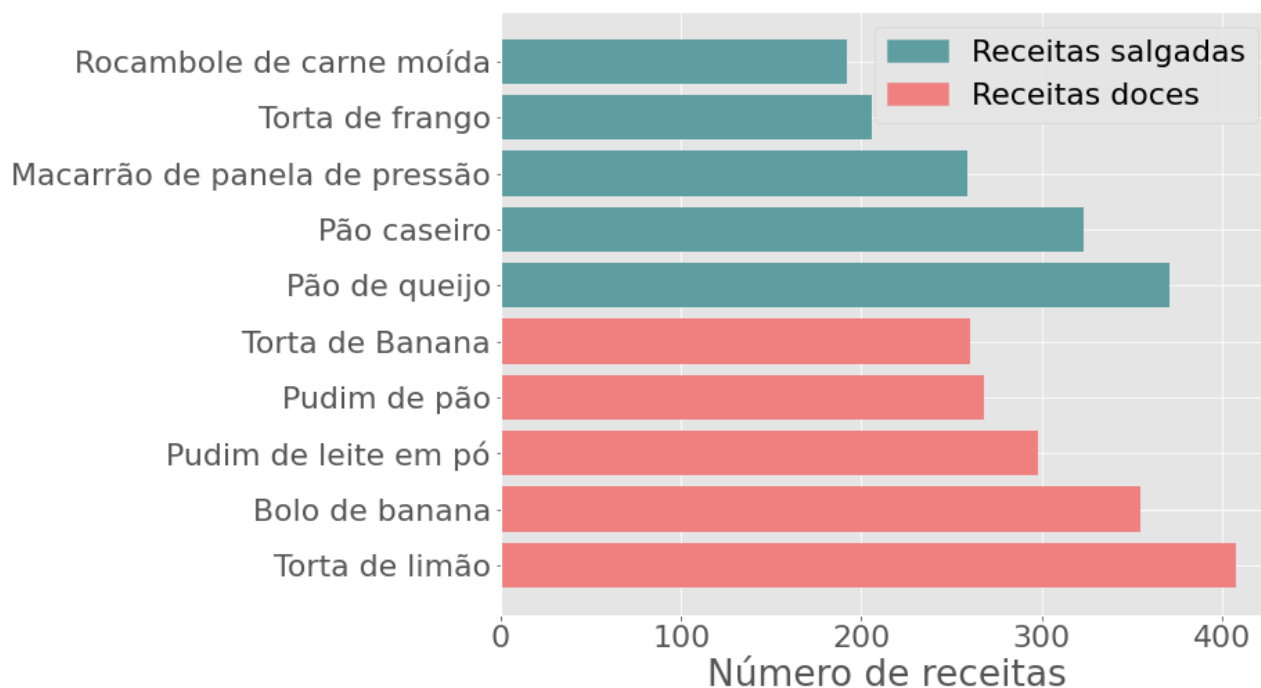


Figura 4.4. Número de vezes em que as cinco receitas doces e salgadas mais frequentes aparecem no conjunto de dados.

4.4 PREVALÊNCIA DE INGREDIENTES

O ingrediente mais prevalente foi o ovo, aparecendo em quase 45% das receitas. Ele ocorre em 67% das receitas doces, sendo o ingrediente mais comum nessas receitas, na frente até mesmo do açúcar – 65% (provavelmente pela alta prevalência do leite condensado – 47%).

Nas receitas salgadas, depois do sal que é listado em 64% das receitas (o que pode ser explicado pelo uso de tempero pronto ou pelo esquecimento em listar um ingrediente que quase sempre é incluído nos pratos), temos cebola que está presente em mais de 50% das receitas. Os dez ingredientes mais prevalentes de cada categoria estão listados na Tabela 4.2.

Tabela 4.2. Dez ingredientes mais frequentes e suas prevalências.

Receitas salgadas		Receitas doces	
Ingrediente	Prevalência	Ingrediente	Prevalência
Sal	0,64	Ovo	0,67
Cebola	0,51	Açúcar	0,65
Tomate	0,37	Leite	0,54
Alho	0,36	Leite Condensado	0,47
Ovo	0,35	Creme de leite	0,43
Carne bovina	0,29	Chocolate	0,42
Farinha de trigo	0,26	Farinha de trigo	0,31
Óleo vegetal	0,26	Fermento químico	0,29
Leite	0,25	Margarina	0,22
Frango	0,25	Biscoito doce	0,20

Já que ingredientes com subcategorias, como queijos, feijões, pimentas e peixes foram contados separadamente, foi feita uma nova análise em que foram incluídos de forma conjunta. Feijão carioca e fradinho, por exemplo, entraram na análise como feijão. Mesmo com essa modificação, não houve alteração nos dez ingredientes mais prevalentes mostrados na Tabela 4.2.

4.5 COEXISTÊNCIA DE INGREDIENTES EM RECEITAS

Antes de avaliar os compostos de sabor dos ingredientes, é interessante verificar quais ingredientes costumam aparecer juntos em receitas. Com esse objetivo, foi construída uma matriz que tem como linhas cada um dos 473 ingredientes e como colunas as 113.194 receitas. Essa matriz é esparsa, apresentando zeros na maior parte das posições, porém apresentando o valor 1 quando o ingrediente é usado em determinada receita. Essa matriz permite identificar facilmente os ingredientes mais parecidos, bem como as receitas mais similares usando-se a similaridade do cosseno (SINGHAL et al., 2001).

No exemplo hipotético da Tabela 4.3, cheesecake e bolo de goiaba são mais similares por apresentarem um ingrediente em comum. A similaridade de cosseno

entre os ingredientes queijo (representado como 0, 0, 1 – primeira linha na tabela) e chocolate (representado como 1, 0, 0) é 0, enquanto a similaridade entre os ingredientes queijo e goiaba (representado como 0, 1, 1) é de 0,71.

Tabela 4.3. Exemplo de matriz de correlação entre ingredientes e receitas.

Ingrediente/ Receita	Brigadeiro	Bolo de goiaba	Cheesecake
Queijo	0	0	1
Chocolate	1	0	0
Goiaba	0	1	1

A Tabela 4.4 apresenta alguns ingredientes das receitas e o ingrediente que mais foi encontrado em conjunto. Além dos ingredientes mais prevalentes como alho, cebola e tomate, foram listados alguns menos comuns.

Tabela 4.4. Primeiro e segundo ingrediente mais similar de acordo com a coexistência em receitas.

Ingrediente	# 1	# 2
Cebola	Alho	Tomate
Beterraba	Cenoura	Alface
Peixe	Limão	Cebola
Gengibre	Molho de soja	Cravo
Maçã	Canela	Uva passa

Foi possível identificar alguns pares usuais, como maçã com canela ou peixe com limão. Isso reforça a tese de que apesar das inúmeras possibilidades de combinações, as receitas tendem a usar um número pequeno delas, conforme a estimativa de Ahn et al. (2011). Como um exemplo, apenas quatro receitas incluem conjuntamente peixe e canela.

4.6 COMPARTILHAMENTO DE COMPOSTOS DE SABOR

Alguns ingredientes que aparecem frequentemente juntos compartilham mais compostos de sabor que outros. A Figura 4.5 ilustra o compartilhamento de compostos de sabor e aroma entre os cinco ingredientes mais prevalentes em receitas salgadas.

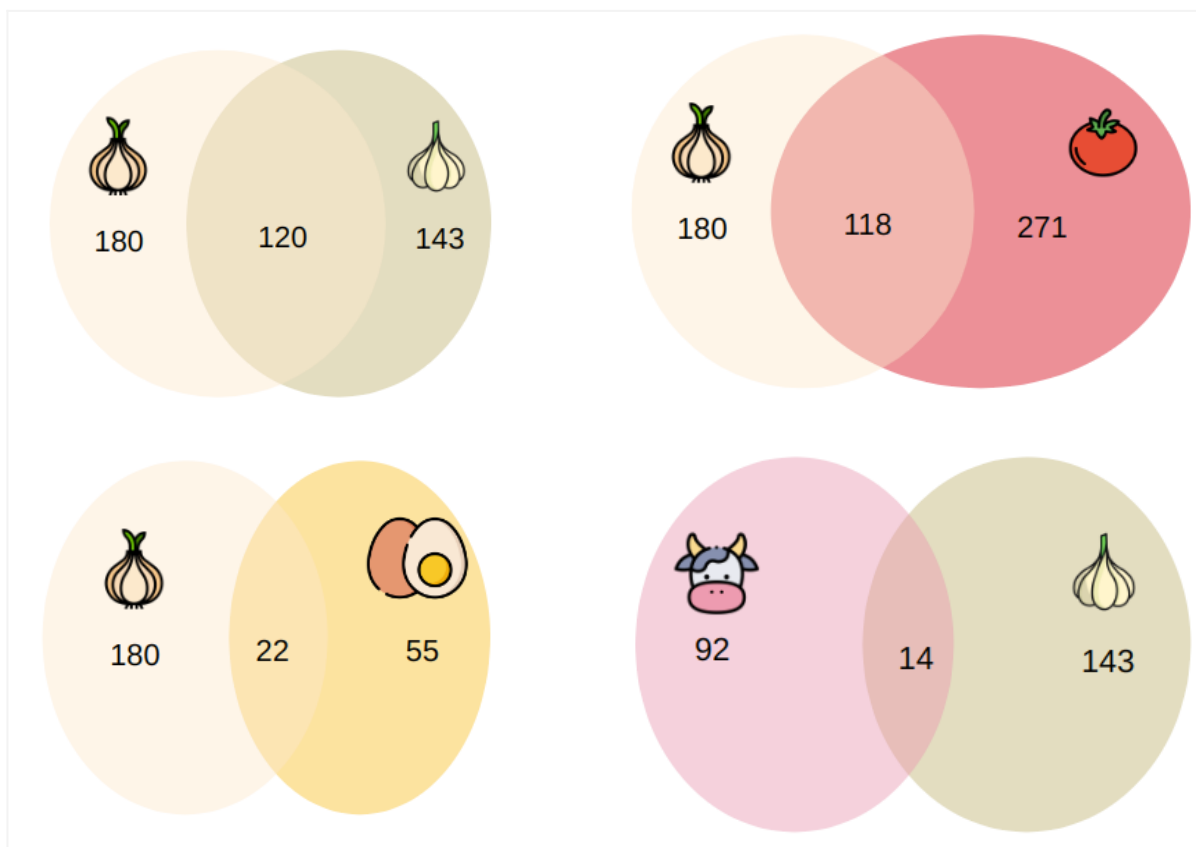


Figura 4.5. Compostos de sabor compartilhados entre os ingredientes mais prevalentes. A linha superior apresenta ingredientes com maior compartilhamento.

Nota-se que os ingredientes que mais frequentemente aparecem juntos apresentam um grande número de compostos de sabor em comum. Porém, é importante ressaltar que existe uma diferença relevante no número de compostos de sabor mapeados para cada ingrediente. Por exemplo, o tomate tem 271 compostos catalogados, enquanto o ovo apenas 55. Essa diferença em moléculas de sabor compartilhadas poderia ocorrer em virtude de compostos que ainda não são conhecidos. Caso fosse possível obter mais compostos para o ovo, por exemplo, é possível que ele tivesse mais compostos em comum com a cebola.

Por outro lado, algumas combinações pouco usuais de fato apresentaram poucos compostos de sabor em comum. Peixe tem 47 compostos catalogados, canela 182 e somente 9 desses compostos são comuns aos dois.

A ausência de quantidades limita a aplicabilidade dessa correlação. Ainda que fosse possível usar as medidas dos ingredientes listadas nas receitas, não seria possível obter a quantidade de cada composto de sabor nos alimentos, apenas sua presença.

Ainda assim, é interessante observar as correlações entre os componentes de sabor dos alimentos. A Figura 4.6 permite visualizar o compartilhamento de sabores entre os alimentos mais prevalentes nas receitas coletadas.

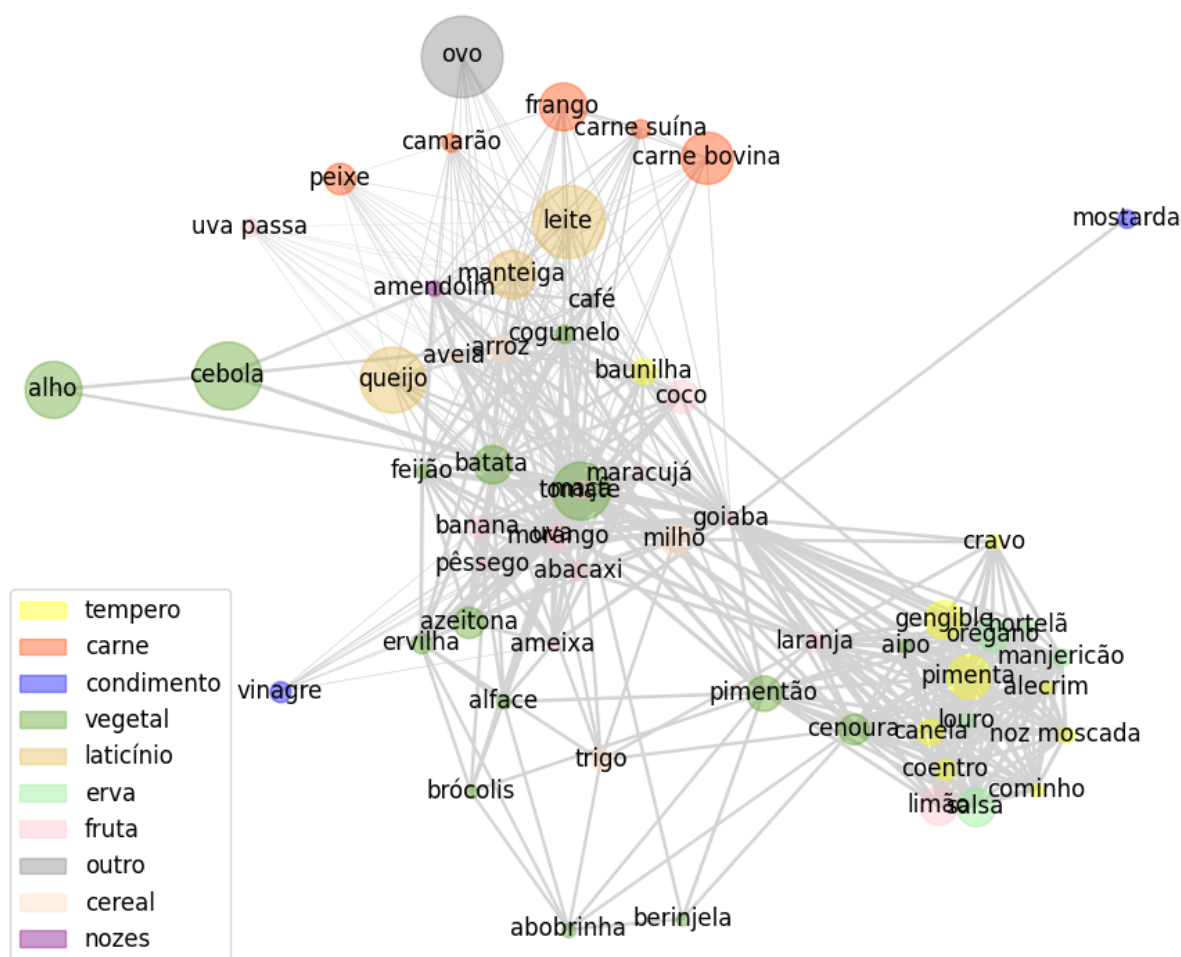


Figura 4.6. Grafo de ingredientes com as arestas representando compostos de sabor compartilhados. O tamanho do círculo representa a prevalência do ingrediente e a espessura da linha o número de compostos compartilhados. O desenho do grafo aproxima ingredientes com maior compartilhamento.

Este gráfico foi construído usando os cem alimentos mais recorrentes nas receitas coletadas e as moléculas de sabor compartilhadas entre eles. Para possibilitar a visualização, apenas as relações mais relevantes foram mantidas, conforme detalhado na Seção 3.4.6.

Na Figura 4.6, alimentos que têm a mesma função, como carne suína e bovina, compartilham um número maior de compostos de sabor, aparecendo na mesma região do gráfico. Alho e cebola, a combinação mais frequente nas receitas salgadas, também aparecem próximos.

É possível identificar outras relações interessantes, como limão mais próximo de temperos e especiarias do que de outras frutas, concordando com seu uso frequente em pratos salgados.

Alimentos como a mostarda aparecem um pouco mais afastados, o que indica o pouco compartilhamento de sabores com os demais ingredientes mais comuns nas receitas. Novamente cabe ressaltar que essa análise depende dos compostos de sabor disponíveis no FlavorDB e, portanto, pode ser que a mostarda tenha compostos em comum com outros ingredientes que ainda não foram catalogados.

A Figura 4.7 apresenta uma visualização mais detalhada da rede de sabores apenas para os ingredientes mais prevalentes em receitas salgadas.

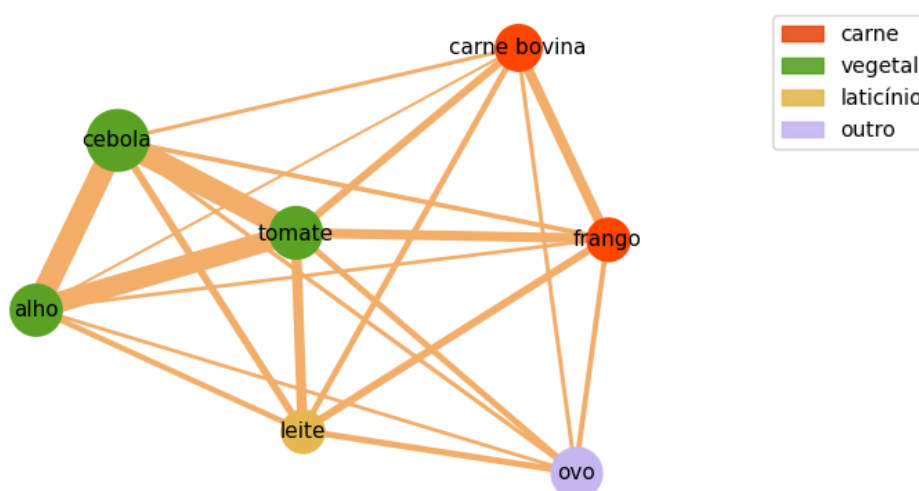


Figura 4.7. Grafo de ingredientes mais prevalentes com a espessura das arestas representando o número de compostos de sabor compartilhados.

Comparando-se os ingredientes com os encontrados por Ahn et al. (2011) para as diferentes cozinhas do mundo, temos cebola, tomate e alho como sendo característicos da América Latina. Diferentemente do encontrado para essa região, no entanto, não temos milho figurando entre os principais ingredientes das receitas compartilhadas no site Tudo Gostoso. Em vez disso, temos maior presença de carnes, principalmente a bovina e de frango.

Já leite e ovos, que aparecem muito em nossas receitas doces, são ingredientes prevalentes na cozinha norte-americana.

Para avaliar de forma global se o compartilhamento de compostos de sabor está correlacionado ao número de receitas em que os alimentos aparecem juntos, foram criados pares com todas as combinações de ingredientes usados neste trabalho. Para cada par, foi obtido o número de compostos compartilhados e a prevalência do par em receitas, conforme ilustrado na Figura 4.8.

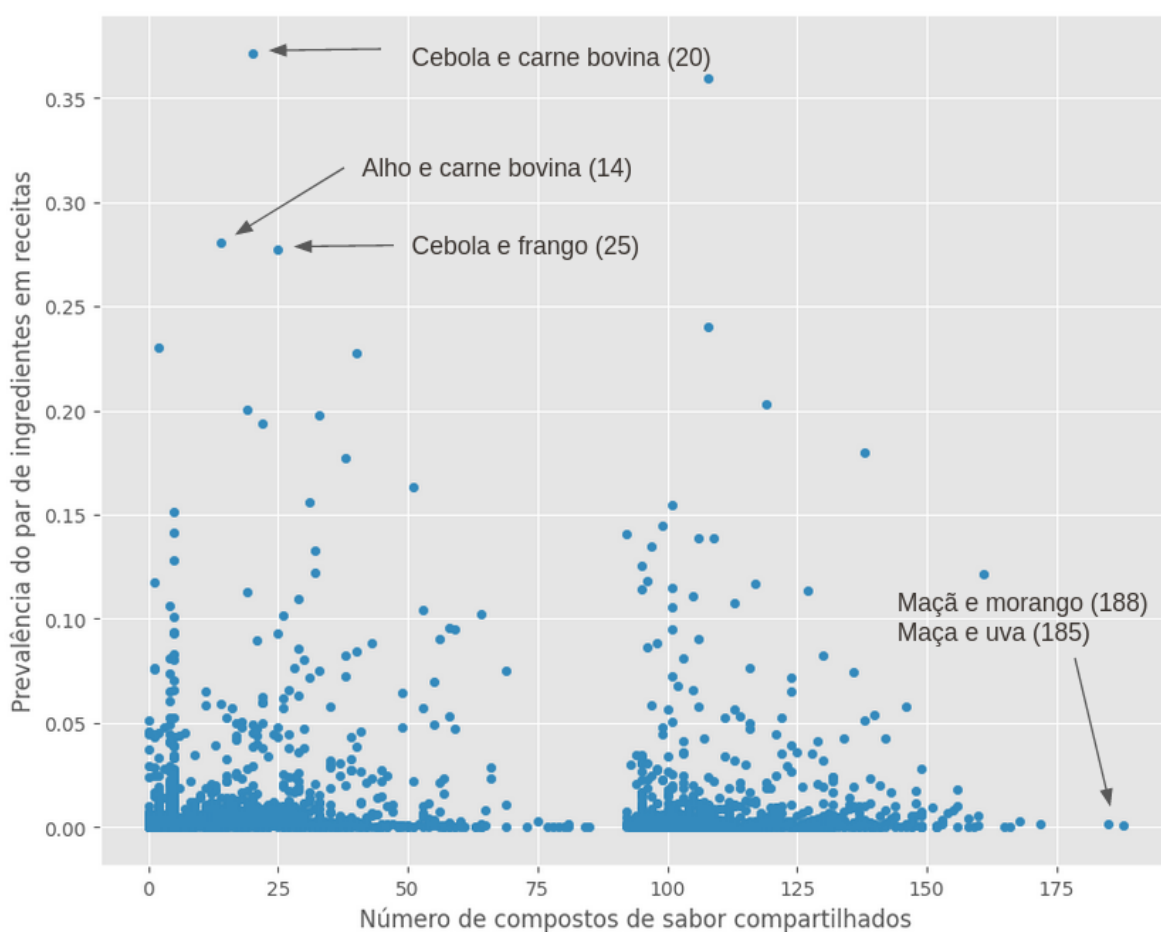


Figura 4.8. Correlação entre a prevalência de um par de ingredientes em receitas e o número de compostos de sabor compartilhados.

É possível observar que no conjunto de todos os ingredientes não existe uma correlação forte entre número de compostos de sabor compartilhados e prevalência da aparição de um par de ingredientes em receitas. Alguns *outliers* se destacam, por aparecerem em um grande número de receitas mesmo com baixo compartilhamento de sabores, como carnes, cebola e alho. Outros, por terem muitos compostos de sabor compartilhados mas raramente aparecerem juntos em receitas, como maçã e morango.

A correlação de Pearson, mesmo com a remoção desses *outliers*, é próxima a zero, mostrando que para o conjunto mais amplo de ingredientes não foi possível confirmar a hipótese do pareamento de sabores.

5 CONCLUSÕES

Este trabalho teve como objetivo usar dados de receitas do principal site de culinária do país para descobrir padrões sobre o uso de ingredientes e suas combinações no Brasil.

A pesquisa realizada indicou algumas receitas muito populares como torta de limão, com mais de 400 versões, e pão de queijo, com mais de 370 receitas de diferentes usuários catalogadas no site.

Alho, cebola e tomate foram os ingredientes mais prevalentes nas receitas salgadas avaliadas, o que é similar ao que Ahn et al. (2011) identificou na América Latina. O ovo foi um ingrediente muito prevalente tanto em receitas salgadas como doces.

Quando avaliamos o compartilhamento de compostos de sabor entre os ingredientes mais prevalentes, como cebola e alho, ele parece dar suporte à hipótese do pareamento de sabores. Porém, não foi possível validar essa hipótese para o conjunto de todos os ingredientes. Limitações nos dados utilizados, como diferente número de compostos mapeados para ingredientes distintos e falta das quantidades de cada composto nos ingredientes, fazem com que não seja possível afirmar que esse é o principal fator por trás dessas combinações.

Nos dados coletados, não foi possível prever a coexistência em receitas apenas a partir do compartilhamento de moléculas de sabor. Porém, apesar de não ser o único fator, o pareamento de sabores é uma das ferramentas disponíveis para o cozinheiro que busca se desafiar.

O grande volume de dados coletados durante este trabalho pode dar origem a outros estudos. Seguem alguns dos itens já mapeados para trabalhos futuros:

- Consultar diferentes bases de dados e bibliografias de compostos de sabor a fim de reduzir as discrepâncias entre o número de compostos para cada ingrediente.
- Realizar análises sensoriais com ingredientes que não costumam aparecer juntos, mas que têm muitos compostos de sabor em comum.

- Avaliar outros modelos de aprendizado de máquina a fim de reduzir a taxa de erro na predição do ingrediente em cada um dos itens da lista de ingredientes.

REFERÊNCIAS

- AHN, Yong-Yeol et al. Flavor network and the principles of food pairing. **Scientific reports**, Springer,v.1, n. 1, p. 1–7, 2011.
- AHNERT, Sebastian E. Network analysis and data mining in food science: the emergence of computational gastronomy. **Flavour**, BioMed Central, v. 2, n. 1, p. 1–3, 2013.
- ALEXA INTERNET. **Market Share, Revenue and Traffic Analytics**. Disponível em: <http://www.alexacom/>. Acesso em: jan. 2022.
- BERTAULT, François. A force-directed algorithm that preserves edge-crossing properties. **Information Processing Letters**, Elsevier, v. 74, n. 1-2, p. 7–13, 2000.
- BISHOP, Christopher M; NASRABADI, Nasser M. **Pattern recognition and machine learning**. Springer, 2006. v. 4.
- BRANDT, R. **What's cooking, Chef Watson?** Disponível em: <https://www.ibm.com/blogs/think/nl-en/2017/06/06/whats-cooking-chef-watson>. Acesso em: jun. 2023.
- COVER, Thomas; HART, Peter. Nearest neighbor pattern classification. **IEEE transactions on information theory**, IEEE, v. 13, n. 1, p. 21–27, 1967
- BLUMENTHAL, Heston.**The big fat duck cookbook**. Bloomsbury, 2008.
- BURDOCK, GA. Flavor Ingredients, 6th Edn Boca Raton. **FL: Fenaroli's Handbook CRC Press**, 2010.
- CROMWELL, Erol; GALEOTA-SPRUNG, Jonah; RAMANUJAN, Raghuram. Computational creativity in the culinary arts. In: **The Twenty-Eighth International Flairs Conference**, 2015.
- FOODB. **World's largest and most comprehensive resource on food constituents**. Disponível em: <http://foodb.ca>. Acesso em: jun. 2023.
- GALMARINI, Mara Virginia. The role of sensory science in the evaluation of food pairing. **Current Opinion in Food Science**, Elsevier, v. 33, p. 149–155, 2020.

GARG, Neelansh et al. FlavorDB: a database of flavor molecules. **Nucleic acids research**, Oxford University Press, v. 46, n. D1, p. d1210–d1216, 2018.

GOEL, Mansi; BAGLER, Ganesh. Computational gastronomy: A data science approach to food. **Journal of Biosciences**, Springer, v. 47, n. 1, p. 12, 2022

JAIN, Anupam; NK, Rakhi; BAGLER, Ganesh. Analysis of food pairing in regional cuisines of India. **PloS one**, Public Library of Science San Francisco, CA USA, v.10, n. 10, e0139539, 2015.

KINOUCI, Osame et al. The non-equilibrium nature of culinary evolution. **New Journal of Physics**, IOP Publishing, v. 10, n. 7, p. 073020, 2008.

MIKOLOV, Tomáš et al. Efficient Estimation of Word Representations in Vector Space. **Proceedings of the 1st International Conference on Learning Representations**, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, 2013.

OPENAI. GPT-4 Technical Report. **arXiv preprint**: 2303.08774, 2023.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **The Journal of Machine Learning Research**, JMLR. org, v.12, p. 2825–2830, 2011.

ŘEHŮŘEK, RADIM, et al. Software Framework for Topic Modelling with Large Corpora. **Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks**. ELRA, 2010.

SERRANO, M Ángeles; BOGUNÁ, Marián; VESPIGNANI, Alessandro. Extracting the multiscale backbone of complex weighted networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 106, n. 16, p. 6483–6488, 2009.

SIMAS, Tiago et al. Food-bridging: a new network construction to unveil the principles of cooking. **Frontiers in ICT**, Frontiers Media SA, v. 4, p. 14, 2017.

SINGHAL, AMIT. Modern information retrieval: A brief overview. **IEEE Data Eng. Bull.**, v 24, n.4, p. 35-43, 2001.

SPENCE, Charles. **Gastrophysics: The new science of eating**. Penguin UK, 2017.

SPENCE, Charles; WANG, Qian Janice; YOUSSEF, Jozef. Pairing flavours and the temporal order of tasting. **Flavour**, Springer, v.6, n. 1, p. 4, 2017.

TALLAB, Shahad Talal; ALRAZGAN, Muna Saleh. Exploring the food pairing hypothesis in Arab cuisine: A study in computational gastronomy. **Procedia Computer Science, Elsevier**, v. 82, p. 135–137, 2016.

TOUVRON, Hugo et al. Llama: Open and efficient foundation language models. **arXiv preprint:2302.13971**, 2023.

VARSHNEY, Lav R et al. A big data approach to computational creativity: The curious case of Chef Watson. **IBM Journal of Research and Development**, IBM, v.63, n. 1, p. 7–1, 2019.

KLEPPER, M. Food pairing theory: A European fad. **Gastronomica: Journal of Food Culture**, University of California Press USA, v. 11, n. 4, p. 55–58, 2011.

APÊNDICE A - USO E TREINAMENTO DOS MODELOS DE APRENDIZADO DE MÁQUINA

Modelo de linguagem usado na separação de ingredientes e quantidades

Para extrair um ingrediente, por exemplo, “cominho” a partir do item “uma colher de café de cominho” foi usado um grande modelo de linguagem (LLM, Large Language Model) pré treinado denominado LLaMA (Large Language Model Meta AI). Esse modelo foi treinado pela Meta em trilhões de palavras de diferentes línguas e está disponível em múltiplos tamanhos (TOUVRON et al., 2023).

Neste trabalho foi usado a versão de 7 bilhões de parâmetros em virtude dos recursos computacionais disponíveis. Os pesos do modelo pré-treinado foram obtidos através de um formulário para uso em pesquisa submetido à Meta. Para executar o modelo foi utilizada a biblioteca transformers da HuggingFace, versão 4.30.2.

O código fonte para predição usado foi o disponibilizado pela Meta no Github em <https://github.com/facebookresearch/llama>. Para obter melhores resultados no uso em Português, foi usada uma extensão chamada Alpaca-LoRA-PTBR, disponível em <https://github.com/DominguesM/alpaca-lora-ptbr-7b>.

Foram testadas diferentes entradas para que o modelo extraísse uma lista de ingredientes de sua entrada. Por exemplo, “quais são os ingredientes da receita?”, “faça uma lista de compras para os ingredientes a seguir” e “me dê a lista de ingredientes sem as quantidades”. A partir de uma avaliação qualitativa a última entrada foi selecionada. O modelo foi executado para a lista de ingredientes de cada receita e seus resultados foram salvos para análise posterior.

Modelo de aprendizado de máquina para predição do ingrediente final

Transformação das palavras em valores (Word2Vec)

Durante a extração dos ingredientes usando o modelo de linguagem, alguns ingredientes foram gerados incorretamente. Por exemplo, “um copo de requeijão de ervilha”, resultou em “requeijão”. Além disso, dependendo do contexto, a técnica não foi capaz de corrigir erros de digitação. A expressão “uma lata de revilha” foi traduzida corretamente para ervilha, mas a expressão “uma xícara de revilha” retornou uma resposta “não familiarizado com o ingrediente”.

Para corrigir esses problemas, foi treinado um modelo de inteligência artificial especificamente para a tarefa de extrair os ingredientes.

Para que palavras como cebolinha, *cebolina* e *cebolhinha* sejam consideradas o mesmo ingrediente, elas devem ter usos similares no texto. Por exemplo, dificilmente veremos “uma lata de cebolinha”, e muitas vezes veremos “cebolinha picada”. A partir desses contextos representados no texto é possível atribuir automaticamente valores às palavras. Num exemplo hipotético, a palavra cebolinha seria 1, a palavra *cebolina* seria 2 e a palavra milho seria 34.

Para realizar esse mapeamento entre palavras e valores foi empregada a técnica Word2Vec (MIKOLOV et al., 2013), em que uma rede neural é usada para atribuir diferentes valores a palavras (ou *tokens*, pois não necessariamente são palavras válidas na nossa língua). O treinamento desse modelo foi realizado usando-se CBOW (Continuous Bag-of-Words) em que para aprender um valor que representa determinada palavra são usadas n palavras que a antecedem e n palavras que a sucedem em um conjunto de frases.

Neste trabalho o n usado foi 5, selecionado através de uma avaliação qualitativa dos resultados. O número de dimensões de cada palavra foi definido como 100. Quanto maior o número de dimensões, mais fácil separar ingredientes similares.

Para o treinamento foi usada a biblioteca gensim (ŘEHŮŘEK et al., 2010), versão 4.2.0, e a entrada foi o conjunto de ingredientes de todas as receitas.

Extração dos ingredientes das frases

Após a transformação das palavras em valores numéricos, foi possível treinar um segundo modelo para prever a saída desejada a partir de uma frase. Por exemplo, “um copo de leite condensado de leite” deve gerar “leite” com resultado.

Para que isso seja possível, primeiro a frase deve ser transformada para uma representação numérica. Cada palavra dá origem a um valor e esses valores juntos servem de entrada para o modelo.

Como dados de entrada foram usados 10.000 frases cujos resultados desejados foram manualmente anotados. O modelo utilizado foi o K vizinhos mais próximos (KNN – K Nearest Neighbors) (COVER; HART, 1967), em que o resultado final é obtido a partir dos resultados das K frases mais parecidas. Dessa forma “dois maços de salsinha”, cuja saída “salsinha” estava no treinamento estará próxima da

frase “um maço de *salsina*”. O modelo retorna uma confiança na resposta, permitindo a revisão ou descarte de respostas pouco confiáveis. Por exemplo “forma de bolo” não estava próxima de nenhum item no treinamento e, portanto, não gerou nenhuma inclusão na lista de final de ingredientes. Para realizar o treinamento foi usada a biblioteca scikit-learn (PEDREGOSA, 2011), versão 0.24.2.

ANEXO A - EXEMPLO DE DADOS DE RECEITA

Exemplo de receita obtida do site Tudo Gostoso (ingredientes em negrito).

```
{'title': 'Bolo de banana com abacaxi',
'rating': 0,
'votes_count': 'nenhum voto',
'comments_label': '12 comentários',
'servings_label': '12 porções',
'has_comments': True,
'author_name': 'Thaís Santos',
'has_tags': False,
'has_extras': False,
'extras': '',
'youtube_id': None,
'users_favorited_count': '1.126',
'ldjson': '{"@context":"http://schema.org/","@type":"Recipe","name":"Bolo de banana com abacaxi","author":{"@type":"Person","name":"Thaís Santos"},"url":"http://api.tudogostoso.com.br/receita/175542-bolo-de-banana-com-abacaxi.html","image":"http://img.itdg.com.br/tdg/images/recipes/000/175/542/196631/196631_original.jpg","description":"Receita de Bolo de banana com abacaxi. Enviada por Thaís Santos e demora apenas 20 minutos.","aggregateRating":{"@type":"AggregateRating","ratingValue":0,"reviewCount":0,"bestRating":"5","worstRating":"0"},"prepTime":"PT20M","totalTime":"PT20M","recipeYield":12,"recipeIngredient":["3 bananas nanicas","3 ovos","3 xícaras (chá) de farinha de trigo","1/2 xícara de leite","1 colher (sopa) de fermento em pó","2 xícaras (chá) de açúcar","1 pitada de sal","1 colher (café) de canela em pó"],"recipeInstructions":["No liquidificador, coloque os ovos, a banana, o leite, o açúcar e a canela, bata bem","Coloque a mistura em um recipiente e incorpore a farinha","Por fim, acrescente o fermento em pó","Montagem:","Caramelize 1 xícara de açúcar, na forma que irá assar o bolo","Assim que ficar pronto, coloque as rodellas de abacaxi de preferência não muito grossas","Despeje a massa sobre o abacaxi com o caramelo","Não precisa untar a forma","Coloque no forno a 180°C, até dourar ou colocar o palito e o mesmo sair limpo","Desenforme e sirva como quiser"]},
'id': '175542-bolo-de-banana-com-abacaxi',
'image_url':
'http://img.itdg.com.br/tdg/images/recipes/000/175/542/196631/196631_highlight.jpg',
'original_url':
'http://www.tudogostoso.com.br/receita/175542-bolo-de-banana-com-abacaxi.html?nomobile=true',
'tags': [],
'category_name': 'Bolos e tortas doces',
'category_id': 1000,
```

```

'ads_targeting':
'channel=1000;tag=1070;gender=;content_id=175542;portion=12;rating=0.0;comments=
12;title=Bolo;title=de;title=banana;title=com;title=abacaxi;ingr=bananas;ingr=nanicas;ingr
=ovos;ingr=xícaras;ingr=cha;ingr=de;ingr=farinha;ingr=de;ingr=trigo;ingr=xicara;ingr=de;i
ngr=oleo;ingr=colher;ingr=sopa;ingr=de;ingr=fermento;ingr=em;ingr=po;ingr=xícaras;ingr
=cha;ingr=de;ingr=acucar;ingr=pitada;ingr=de;ingr=sal;ingr=colher;ingr=cafe;ingr=de;ingr
=canela;ingr=em;ingr=po;special_id=;timespend=20;user_id=5647464;',
'preparation_time': '20 MIN',
'campaign_id': None,
'campaign_name': None,
'campaign_icon': None,
'campaign_highlight': None,
'campaign_url': None,
'campaign_recipes_count': None,
'ingredient_lists': [{'has_title': False,
'title': None,
'members': ['3 bananas nanicas',
'3 ovos',
'3 xícaras (chá) de farinha de trigo',
'1/2 xícara de óleo',
'1 colher (sopa) de fermento em pó',
'2 xícaras (chá) de açúcar',
'1 pitada de sal',
'1 colher (café) de canela em pó']}],
'preparation_lists': [{'has_title': False,
'title': None,
'members': ['No liquidificador, coloque os ovos, a banana, o óleo, o açúcar e a canela,
bata bem',
'Coloque a mistura em um recipiente e incorpore a farinha',
'Por último, acrescente o fermento em pó']},
{'has_title': True,
'title': 'Montagem:',
'members': ['Caramelize 1 xícara de açúcar, na forma que irá assar o bolo',
'Assim que ficar pronto, coloque as rodela de abacaxi de preferência não muito
grossas',
'Despeje a massa sobre o abacaxi com o caramelo',
'Não precisa untar a forma',
'Coloque no forno a 180°C, até dourar ou colocar o palito e o mesmo sair limpo',
'Desenforme e sirva como quiser']}],
'user': {'id': 5647464,
'name': 'Thaís Santos',
'avatar':
'http://img.itdg.com.br/tdg/assets/default/users\_avatars/user\_avatar\_original.png'}

```

ANEXO B - EXEMPLO DE DADOS DE COMPOSTOS DE SABOR

Exemplo de uma saída obtida do FlavorDB para o ingrediente Brócolis incluindo um dos compostos de sabor (os demais foram omitidos, mas podem ser encontrados em https://cosylab.iitd.edu.in/flavordb/entity_details?id=351)

```
{
  "category": "vegetable-cabbage",
  "category_readable": "Cabbage",
  "entity_alias": "broccoli",
  "entity_alias_basket": "broccoli, broccoli-cooked",
  "entity_alias_readable": "Broccoli",
  "entity_alias_synonyms": "Broccoli",
  "entity_alias_url": "https://en.wikipedia.org/wiki/Broccoli",
  "entity_id": 351,
  "molecules": [
    {
      "atom_stereo_count": 0,
      "bitter": 1,
      "bitterdb_id": "49",
      "bond_stereo_count": 0,
      "cas_id": "91-64-5",
      "charge": 0,
      "common_name": "coumarin",
      "complexity": 196.0,
      "covalently_bonded_unit_count": 1,
      "defined_atom_stereocenter_count": 0,
      "defined_bond_stereocenter_count": 0,
      "exact_mass": 146.037,
      "fema_flavor_profile": "",
      "fema_number": "",
      "fenoroli_and_os": 1,
      "flavor_profile": "sweet@new mown hay@green@tonka@bitter",
      "flavornet_id": 1,
      "fooddb_flavor_profile": "new mown hay@bitter@green@sweet@tonka",
```

```
"fooddb_id": "FDB030742",
  "functional_groups": "oxo(het)arene@aromatic compound@heterocyclic
compound",
  "hba_count": 2,
  "hbd_count": 0,
  "heavy_atom_count": 11,
  "inchi": "InChI=1S/C9H6O2/c10-9-6-5-7-3-1-2-4-8(7)11-9/h1-6H",
  "isotope_atom_count": 0,
  "iupac_name": "chromen-2-one",
  "molecular_weight": 146.145,
  "monoisotopic_mass": 146.037,
  "natural": 1,
  "num_rotatablebonds": 0,
  "odor": "pleasant, fragrant odor resembling that of vanilla beans.@hay-like
bittersweet odor",
  "pubchem_id": 323,
  "smile": "C1=CC=C2C(=C1)C=CC(=O)O2",
  "super_sweet": "Sweet",
  "supersweetdb_id": "10853",
  "synthetic": 0,
  "taste": "bitter undertone@nut-like flavor on dilution@bitter,
aromatic@burning taste",
  "topological_polar_surfacearea": 26.3,
  "undefined_atom_stereocenter_count": 0,
  "undefined_bond_stereocenter_count": 0,
  "unknown_natural": 0,
  "volume3d": "116.3",
  "xlogp": "1.4"
}
```

ANEXO C - LISTA DE INGREDIENTES USADOS NAS ANÁLISES

Abacaxi, abobrinha, abricó, absinto, abóbora, acelga, acerola, adobo, adoçante, agar agar, agave, agrião, aipo, alcachofra, alcaparra, alcarávia, alecrim, alface, alfafa, alfarroba, alga, alginato de sódio, alho, alho poró, almeirão, amaranto, amarula, ameijoa, ameixa, amendoim, amido de milho, amora, amêndoa, angostura, anis, araruta, araticum, arraia, arroz, aspargo, assafétida, atemoia, aveia, avelã, azedinha, azeite, azeite de dendê, azeitona, açafração, açai, açúcar, açúcar mascavo, babaçu, bacalhau, bacon, bacupari, bacuri, bala, banana, bananeira, banha, bardana, batata, batata doce, baunilha, berbigão, berinjela, bertalha, beterraba, bicarbonato de sódio, biscoito, biscoito doce, biscoito salgado, bolo, brandy, broto de bambu, broto de feijão, brócolis, buriti, butiá, cacau, cachaça, café, caju, cajá, calabresa, camarão, cambuquira, camomila, campari, canela, capim limão, capuchinha, caqui, carambola, caramelo, caranguejo, cardamomo, carne bovina, carne de cabra, carne de coelho, carne de jacaré, carne de ovelha, carne suína, carne vegetal, cará, cassis, castanha de baru, castanha de caju, castanha do pará, castanha portuguesa, catuaba, catupiry, cavalinha, caviar, cebola, cebola roxa, cebolinha, cenoura, centeio, cerefólio, cereja, cerveja, cerveja preta, cevada, cheiro verde, chia, chicória, chimichurri, chocolate, chocolate branco, chuchu, chá, cidra, coalhada, coco, codorna, coentro, cogumelo, cogumelo trufa, colorau, cominho, conhaque, couve, couve de bruxelas, couve flor, cranberry, cravo, creme de leite, creme de soja, cremor de tártaro, cupuaçu, curaçu, curry, cuxá, cúrcuma, damasco, dill, doce de leite, emulsificante, endívia, erva cidreira, erva doce, ervas de provence, ervilha, ervilha torta, escarola, espinafre, espumante, estragão, farinha de arroz, farinha de centeio, farinha de mandioca, farinha de milho, farinha de rosca, farinha de trigo, farinha de trigo sarraceno, farinha lactea, farinha panko, feijão, feijão azuki, feijão branco, feijão carioca, feijão de corda, feijão fradinho, feijão moyashi, feijão preto, feijão verde, feno grego, fermento biológico, fermento químico, figo, flocos de arroz, flocos de milho, foie gras, framboesa, frango, fruta, fruta pão, frutas cristalizadas, fumaça líquida, funcho, fécula de batata, fécula de mandioca, gabirola, gelatina, gengibre, gergelim, gin, glicose, glutamato monossódico, glúten, goiaba, goma xantana, gordura vegetal, granola, graviola, groselha, groselha, grão de bico, guaraná, germen de trigo, hibisco, hortelã, inhame, iogurte, jabuticaba, jaca, jambo, jambú, jatobá, jenipapo, jiló, ketchup, kiwi, lagosta, lagostim, lambreta, laranja,

laranjeira, lavanda, legumes, leite, leite condensado, leite de coco, leite de soja, leite vegetal, lentilha, levedo de cerveja, lichia, liga neutra, lima, limoncello, limão, linguiça, linguiça calabresa, linhaça, louro, lula, macarrão, maionese, malte, mamão, mandioca, mandioquinha, manga, mangaba, maniva, manjeriço, manjerona, manteiga, manteiga de cacau, maracujá, marapuama, margarina, maria mole, marisco, marmelada, marshmallow, massa, massa folhada, massala, mate, maxixe, maçã, maçã verde, mel, melado, melancia, melão, mexilhão, milho, mirtilo, missô, molho barbecue, molho branco, molho de soja, molho inglês, molho teriyaki, morango, mortadela, mostarda, mostarda dijon, mostarda em grãos, mussarela, nabo, nectarina, noz moscada, noz pecan, nozes, ora pro nobis, orégano, ostra, ovo, painço, palmito, papoula, pastrami, pato, peixe, peixe aliche, peixe atum, peixe bacalhau, peixe cação, peixe linguado, peixe merluza, peixe salmão, peixe sardinha, peixe tainha, peixe tilápia, pepino, pequi, pera, perdiz, peru, pickles, pimenta, pimenta adobo, pimenta biquinho, pimenta bode, pimenta branca, pimenta caiena, pimenta calabresa, pimenta cambuci, pimenta chili, pimenta chipotle, pimenta comari, pimenta da jamaica, pimenta de cheiro, pimenta dedo de moça, pimenta do reino, pimenta jalapeño, pimenta malagueta, pimenta piri piri, pimenta rosa, pimenta síria, pimenta tabasco, pimenta verde, pimentão, pinhão, pinoli, pipoca, pisco, pistache, pituaia, pitanga, polvilho, polvilho azedo, polvilho doce, polvo, pomelo, presunto, própolis, psyllium, páprica, páprica defumada, páprica doce, páprica picante, pão, pêssigo, queijo, queijo brie, queijo camembert, queijo cheddar, queijo coalho, queijo cottage, queijo cream cheese, queijo emmental, queijo feta, queijo gorgonzola, queijo minas, queijo mussarela, queijo parmesão, queijo parmesão, queijo prato, queijo provolone, queijo ricota, quiabo, quinoa, rabanete, radite, raiz forte, rapadura, refrigerante, repolho, repolho roxo, requeijão, romã, rosa, ruibarbo, rum, rúcula, sagu, sal, sal amoníaco, salame, salsicha, salsinha, saquê, sardinha, segurelha, semente de girassol, semolina, seriguela, siri, soja, sorvete, suspiro, sálvia, taioba, tamarindo, tangerina, tapioca, tequila, tofu, tomate, tomilho, trigo, tucupi, tâmara, umbu, uva, uva passa, vagem, vermute, vieira, viinagre de vinho branco, vinagre, vinagre balsâmico, vinagre branco, vinagre de arroz, vinagre de maçã, vinagre de vinho branco, vinagre de vinho tinto, vinagre de álcool, vinho, vinho branco, vinho do porto, vinho doce, vinho madeira, vinho rosé, vinho tinto, vodka, vôngole, whisky, xarope de bordo, zaatar, zimbro, ácido cítrico, álcool, óleo de dendê, óleo de gergelim, óleo vegetal.