

# Análisis de redes complejas en el mundo de la cocina: una exploración de las relaciones entre ingredientes y recetas

Jorge Morgado

[JORGE.MORGADOV@GMAIL.COM](mailto:JORGE.MORGADOV@GMAIL.COM)

Roberto García

[ROBEGR42@GMAIL.COM](mailto:ROBEGR42@GMAIL.COM)

## Resumen

El análisis de redes complejas es una técnica matemática que se utiliza para estudiar sistemas complejos que pueden ser representados mediante nodos y conexiones. Esta técnica se aplica en una amplia variedad de campos, como la biología, la física, la sociología y la informática. En el ámbito de las recetas de cocina, el análisis de redes complejas se utiliza para analizar las relaciones entre los ingredientes y las recetas. Estos estudios han permitido identificar patrones, tendencias en la forma en que se utilizan los ingredientes, las combinaciones de sabores más populares y las recetas más comunes.

## Abstract

Complex network analysis is a mathematical technique used to study complex systems that can be represented by nodes and connections. This technique is applied in a wide variety of fields such as biology, physics, sociology and computer science. In the field of cooking recipes, complex network analysis is used to analyze the relationships between ingredients and recipes. These studies have allowed the identification of patterns and trends in the way ingredients are used, the most popular flavor combinations, and the most common recipes.

**Palabras Clave:** Red compleja, grafo, ingredientes

**Tema:** Análisis de redes complejas

## 1. Introducción

El análisis de redes complejas es un campo interdisciplinario que se enfoca en el estudio de las propiedades y el comportamiento de sistemas complejos que pueden ser modelados como un grafo, cuyos nodos y enlaces representan entidades y relaciones en sistemas como redes sociales, sistemas de transporte, redes de comunicación, redes de información, etc. [3]

El análisis de redes complejas utiliza herramientas matemáticas y computacionales para estudiar la estructura, la dinámica y la función de las mismas. Algunas de las medidas importantes que se utilizan en el análisis de redes complejas son la centralidad, la modularidad, la conectividad y la robustez. [5]

El análisis de redes complejas se ha utilizado para analizar las relaciones entre los ingredientes y las recetas. Esta técnica ha permitido identificar patrones y tendencias en la forma en que se utilizan los ingredientes y las combinaciones de sabores más populares. Además, ha permitido entender mejor la historia y la evolución de las recetas, al analizar cómo los ingredientes y las combinaciones de sabores han cambiado con el tiempo y cómo se han difundido a través de diferentes culturas y regiones. [2, 6]

En este trabajo, se explorará el análisis de redes complejas en el ámbito de las recetas de cocina. Se realizarán varios estudios con respecto a distintos grafos obtenidos, entre los cuales se representan relaciones entre ingredientes y recetas y relaciones entre recetas buscando un comportamiento de similitud.

El informe prosigue con la siguiente estructura: primero una sección dedicada al tratamiento y extracción de los datos 2, una sección que aborda la creación de los grafos analizados así como

su representación 3 y luego una sección sobre la aplicación creada para la interacción de usuarios 4. Posteriormente, se presentan las conclusiones finales 5 y algunas recomendaciones para estudios futuros 6.

## 2. Extracción de datos

Para la realización de este proyecto se usaron dos fuentes de datos: el libro *Cocina al minuto* de la escritora Nitza Villapol y el *dataset* Recipes5k creado en la Universidad de Barcelona [4].

La extracción de las recetas e ingredientes del libro *Cocina al minuto* se realizó de forma manual. Posteriormente, estos datos fueron traducidos al inglés con el objetivo de mantener la consistencia con el resto de las fuentes. En el caso del *dataset* Recipes5k se seleccionaron los nombres de las recetas y los ingredientes simplificados.

Finalmente todos estos datos fueron unificados en un solo archivo de formato `json`. La estructura del mismo se basa en un diccionario donde cada receta representa una llave y su valor es una lista de ingredientes.

## 3. Grafos implementados

En el proyecto se implementaron grafos utilizando la biblioteca *NetworkX* de *Python*. El objetivo principal del proyecto fue extraer información de los grafos para mostrarla en una aplicación. Para lograr este objetivo, se crearon grafos a partir de los datos previamente mencionados. Una vez creados los grafos, se utilizaron las funciones de *NetworkX* para analizarlos y extraer la información deseada. Dichos grafos serán explicados en las siguientes secciones.

### 3.1 Grafo de Ingredientes y Recetas

Una de las interrogantes más simples que un usuario puede tener es saber qué recetas se pueden cocinar dado una lista de ingredientes. Para ello, se construyó un grafo bipartito donde cada vértice representa una receta o un ingrediente y cada arista  $(u, v)$  existe si y solo si el ingrediente  $u$  es necesario en la receta  $v$  (o vice versa).

Luego, sea  $nb(x)$  los vecinos de un nodo  $x$ , y dado una lista de ingredientes  $A = a_1, a_2, \dots, a_n$ , el conjunto de recetas que se pueden preparar está dado por:

$$nb(a_1) \cap nb(a_2) \cap \dots \cap nb(a_n)$$

Expresado de otra manera, el conjunto de recetas  $u$  tal que existe al menos una arista en  $E$  (conjunto de aristas del grafo) entre algún ingrediente de  $A$  y la receta  $u$ :

$$\{v \mid (a, v) \in E \wedge a \in A\}$$

### 3.2 Grafo de Similitud de Recetas

Otra pregunta interesante es saber que recetas son similares entre sí. Para ello se construyó un grafo ponderado donde cada vértice representa una receta, y las aristas entre recetas indican el nivel de similitud de las mismas.

Si se define  $Ing(r)$  como el conjunto de ingredientes necesarios para la preparación de una receta  $r$ , el coeficiente de similitud entre dos recetas se puede definir como:

$$c(r_1, r_2) = \frac{|Ing(r_1) \cap Ing(r_2)|}{|Ing(r_1) \cup Ing(r_2)|}$$

Debido a la densidad de este grafo, solo se almacenaron las aristas con un coeficiente  $c > 0.3$ .

Luego, dado una receta  $r$ , se pueden obtener las recetas más similares a ella analizando el peso de las aristas de la forma  $(r, u)$  en el grafo.

### 3.3 Grafo de IF-IRF

Finalmente, siguiendo las ideas del artículo *Chocolate chips and fish sauce: a network analysis and visualization in ingredient pairings*[1], se utilizó una idea similar a TF-IDF (*Term Frequency - Inverse Document Frequency*) pero en este caso entre recetas e ingredientes. Se computó el coeficiente IF-IRF (*Ingredient Frequency - Inverse Recipe Frequency*) que denota la importancia que tiene cada par de ingredientes en las recetas de forma general:

$$w_{ij} = (1 + \log E_{ij}) \log\left(\frac{n_i + n_j}{N}\right)$$

Donde  $E_{ij}$  denota la cantidad de recetas que contienen a los ingredientes  $i$  y  $j$ ,  $n_i$  y  $n_j$  la cantidad de recetas con el ingrediente  $i$  y  $j$  respectivamente y  $N$  la cantidad de recetas en total.

Luego, analizando el peso de las aristas de un ingrediente determinado, se puede realizar una recomendación a los usuarios sobre posibles ingredientes que puede usar en conjunto con los que tiene.

## 4. Aplicación de Streamlit

Como medio de muestra de los estudios realizados en este trabajo se realizó una aplicación de Streamlit. En la aplicación se visualizan los algoritmos sobre los distintos grafos propuestos. De esta forma se observan todas las recetas del conjunto de datos, los ingredientes que la conforman, así como un sistema de filtrado de recetas de acuerdo a una lista de ingredientes que la conformen o no.

También se integró un pequeño modelo de *scraping* que muestra varios enlaces a sitios en internet que muestran como elaborar una receta dada, dichos enlaces son producto de una búsqueda en *Google*.

## 5. Conclusiones

En este trabajo, se mostraron cómo las técnicas de análisis de redes complejas se utilizan para entender las relaciones entre los ingredientes y las recetas. Se identificaron patrones y tendencias en la forma en que se utilizan los ingredientes y las combinaciones de sabores.

Se logró buscar una similitud entre tipos de recetas, con resultados basados en los datos analizados. Dichos datos representaban como ingredientes de características parecidas podían usarse indistintamente en una misma receta o analizando el porcentaje de ingredientes comunes entre las mismas. Con este resultado se obtiene versatilidad a la hora de hacer una elaboración deseada sin tener completamente el conjunto de ingredientes que la conforman.

Se obtuvo, además, una herramienta que permita analizar que ingredientes se suelen usar con otros y que al mismo tiempo dicha combinación sea representativa, la cual, junto a la herramienta de *scraping* integrada en la aplicación, se logra un modelo completo de ayuda al usuario, ya que proporciona los medios y los pasos de elaboración.

En resumen, el análisis de redes complejas es una técnica de análisis esencial para entender mejor y explorar la complejidad de las relaciones entre los ingredientes y las recetas en el mundo de la cocina.

## 6. Recomendaciones

A continuación, se presentan algunas recomendaciones a tener en cuenta para futuros trabajos:

- Agregar más información a las redes(grafos) analizados así como la exploración de otras variantes, como la creación de nuevos grafos.
- Insertar conocimiento nutricional para analizar estilos saludables de alimentación
- Insertar conocimiento geográfico que permita un análisis regional agregando, además, recetas de diferentes culturas.

- Agregar información sobre métodos de cocción que sirvan de métrica de comparación en funciones de similaridad entre recetas.

## Referencias

- [1] Chocolate chips and fish sauce: A network analysis and visualization in ingredient pairings. <https://sites.northwestern.edu/msia/2019/05/21/chocolate-chips-and-fish-sauce-a-network-analysis-and-visualization-in-ingredient-pairings/>.
- [2] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. Flavor network and the principles of food pairing. *Scientific reports*, 1(1):1–7, 2011.
- [3] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.
- [4] Marc Bolaños, Aina Ferrà, and Petia Radeva. Food ingredients recognition through multi-label learning. In *New Trends in Image Analysis and Processing–ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers 19*, pages 394–402. Springer, 2017.
- [5] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242, 2007.
- [6] Juan CS Herrera. The contribution of network science to the study of food recipes. a review paper. *Appetite*, 159:105048, 2021.