



**ESCUELA DE NEGOCIOS**

**MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS**

**SISTEMA DE RECOMENDACIÓN PARA ENCONTRAR LOS MEJORES  
PRODUCTOS DE AMAZON UTILIZANDO PROYECCIÓN DE REDES  
BIPARTITAS.**

**Profesor**

**Ing. Mario González. PH.D**

**Autor**

**María Belén Tipantuña Toapanta**

**2023**

## RESUMEN

El presente proyecto tiene como propósito fundamental desarrollar un sistema de recomendación utilizando la técnica basada en grafos donde se construye un grafo para generar recomendaciones de un producto con la finalidad de ayudar a las empresas a mejorar la experiencia de sus compradores en un sitio web y así dará como resultado una mejor adquisición y retención de clientes, en este caso se realiza un análisis con la base de datos que nos proporciona Amazon donde se tiene los registros de identificación única del usuario y del producto y la calificación del 1 al 5 dada por el usuario a un producto (Rating) donde 1 es una baja calificación y 5 la mejor calificación del producto.

Con el análisis exploratorio se identifica el número de productos de acuerdo con su calificación y sobre todo se identifica los 10 primeros productos más populares de acuerdo a la base de datos utilizada.

En cuanto al plan metodológico, se utiliza la técnica basada en grafos implementada en Python con la ayuda de la librería NetworkX para representar y analizar relaciones y la estructura de los datos, al igual que se obtendrá diversas métricas de los grafos para analizar las propiedades y características del mismo.

Los hallazgos presentados en el presente proyecto proporcionarán información valiosa ya que los grafos juegan un papel significativo en los sistemas de recomendación debido a su capacidad para representar de manera eficaz las relaciones complejas entre usuarios y productos, esto resulta en sugerencias más exactas y adaptadas a cada usuario

Palabras clave: Sistemas de recomendación, grafos, redes bipartitas, personalización

## **ABSTRACT**

The fundamental purpose of this project is to develop a recommendation system using the graph-based technique where a graph is built to generate recommendations for a product with the purpose of helping companies improve the experience of their buyers on a website and thus will result in better customer acquisition and retention, in this case an analysis is carried out with the database provided by Amazon where there are unique identification records of the user and the product and the rating from 1 to 5 given by the user to a product (Rating) where 1 is a low rating and 5 is the best rating of the product.

With the exploratory analysis, the number of products is identified according to their rating and, above all, the first 10 most popular products are identified according to the database used.

Regarding the methodological plan, the graph-based technique implemented in Python with the help of the NetworkX library is used to represent and analyze relationships and the structure of the data, as well as various metrics of the graphs to analyze the properties and characteristics of it.

The findings presented in this project will provide valuable information since graphs play a significant role in recommendation systems due to their ability to effectively represent complex relationships between users and products, this results in more accurate and tailored suggestions for each user

Keywords: Recommendation systems, graphs, bipartite networks, personalization

## ÍNDICE DEL CONTENIDO

RESUMEN .....	2
ABSTRACT .....	3
ÍNDICE DEL CONTENIDO .....	4
ÍNDICE DE TABLAS .....	5
ÍNDICE DE FIGURAS .....	6
INTRODUCCIÓN .....	1
REVISIÓN DE LITERATURA .....	2
Sistemas o motores de recomendación .....	2
Basado en contenido .....	3
Basado en filtrado colaborativo .....	4
Basado en grafos .....	5
Tipos de grafos según sus propiedades .....	10
Tipos de grafos según su estructura .....	11
Híbrido .....	17
IDENTIFICACIÓN DEL OBJETO DE ESTUDIO .....	19
PLANTEAMIENTO DEL PROBLEMA .....	20
OBJETIVO GENERAL .....	21
OBJETIVOS ESPECÍFICOS .....	21
JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA .....	22
RESULTADOS .....	29
DISCUSIÓN DE LOS RESULTADOS .....	37
PROPUESTA DE SOLUCIÓN .....	38
Implicaciones Organizacionales .....	38
CONCLUSIONES Y RECOMENDACIONES .....	40
Conclusiones .....	40
Recomendaciones .....	41
Referencias .....	42

## ÍNDICE DE TABLAS

Tabla 1. Conjunto de variables a utilizar .....	23
Tabla 2. Ejemplos de recomendaciones de productos .....	32

## ÍNDICE DE FIGURAS

Figura 1. Grafo formado por X vértices y X aristas.....	5
Figura 2. Ejemplo de un grafo dirigido y su matriz de adyacencia .....	6
Figura 3. Algoritmo de búsqueda en profundidad(DFS).....	7
Figura 4. Algoritmo de búsqueda en anchura (BFS).....	8
Figura 5. Tipos de grafo no dirigido y dirigido .....	10
Figura 6. Tipo de grafo cíclico y acíclico .....	11
Figura 7. Tipo de grafo ponderado y no ponderado.....	11
Figura 8. Grafo Aleatorio .....	12
Figura 9. Grafo completo.....	14
Figura 10. Grafo clique .....	14
Figura 11. Grafo árbol.....	15
Figura 12. Grafo bipartito .....	16
Figura 13. Grafo utilizado para un sistema básico de recomendación [9] .....	17
Figura 14. Método híbrido paralelo.....	18
Figura 15. Número de usuarios y productos únicos .....	25
Figura 16. Calificaciones dadas por los usuarios.....	25
Figura 17. Top 10 de los productos más populares .....	26
Figura 18. Frecuencia de Ratings para los top 10 de productos .....	26
Figura 19. Red Bipartita.....	28
Figura 20. Red de relaciones entre productos.....	29
Figura 21. Componente más grande del grafo .....	30
Figura 22. Recomendación de un fragmento del grafo .....	31
Figura 23. Herramienta interactiva.....	31
Figura 24. Grado de vecino promedio.....	35
Figura 25. Centralidad de grafo.....	35
Figura 26. Cercanía de centralidad .....	36
Figura 27. Centralidad de intermediación.....	36

## INTRODUCCIÓN

Actualmente el comercio electrónico está en auge, donde se puede acceder a una gran variedad de productos para consumir. La oferta es prácticamente ilimitada en la Web, lo que por un lado es interesante para los usuarios, quienes tienen más opciones de encontrar productos que están buscando. Sin embargo, esto también puede resultar agobiante, debido a la imposibilidad de localizar aquellos productos que sean más adecuados para ellos y es aquí donde surgen los sistemas de recomendación para resolver este problema.

Un sistema de recomendación es una herramienta que ayuda a los usuarios a encontrar los mejores productos según sus preferencias, necesidades o gustos. Los sistemas de recomendación utilizan diferentes técnicas para analizar la información de los usuarios y de los productos, y ofrecer sugerencias personalizadas que aumenten la satisfacción y la fidelidad de los clientes, así como las ventas y los beneficios de los vendedores.

En el presente proyecto se realiza un estudio de un sistema de recomendación basado en grafos para una base de datos de Amazon para así obtener sugerencias personalizadas de productos a los clientes basándose en valoraciones. Este sistema de recomendación busca ser más efectivo y preciso ya que al contar con una gran cantidad de datos de los usuarios y productos puede adaptarse a las necesidades individuales de cada cliente

## REVISIÓN DE LITERATURA

### SISTEMAS O MOTORES DE RECOMENDACIÓN

Un sistema de recomendación es una herramienta que establece un conjunto de criterios y valoraciones sobre los datos de los usuarios para realizar predicciones sobre recomendaciones de elementos que puedan ser de utilidad o valor para el usuario. Estos sistemas seleccionan datos proporcionados por el usuario de forma directa o indirecta, y procede a analizar y procesar información del historial del usuario para transformar estos datos en conocimiento de recomendación [1]

Los sistemas de recomendación trabajan con la información del perfil de usuario, la información está clasificada en dos tipos : explícita e implícita, cuando el usuario proporciona sus preferencias de manera consciente en su perfil, el sistema de recomendación hace el uso de la información explícita, esto quiere decir que solicitan al usuario una calificación sobre un ítem en particular o sobre un conjunto de ítems, ahora un sistema de recomendación que usa información implícita si deducen cuales son las preferencias del usuario a partir del comportamiento y del historial del mismo.

El objetivo principal de los sistemas de recomendación es proveer las preferencias de un usuario o grupo de usuarios con respecto a un producto o servicios ya que hoy en día las aplicaciones utilizan dichas recomendaciones para ofrecer a los usuarios opciones válidas de productos o servicios acorde a sus necesidades

Existen diversos algoritmos o métodos de recomendación que toman en cuenta diversas características, al momento de buscar productos o servicios relevantes para los usuarios. Por ejemplo, unos consideran más importante la relación que pueda existir entre usuarios categorizados como semejantes y hay sistemas de recomendación que sugieren productos dependiendo del contenido relacionado a las preferencias de cada usuario. Las técnicas o métodos más utilizados se

categorizan en cuatro tipos: basado en contenido, filtro colaborativo, híbrido y basado en grafos

### **Basado en contenido**

El objetivo principal de los sistemas de recomendación basados en contenido se centra en el producto o servicio para generar la predicción y luego su recomendación se basa en el usuario, donde se valida el historial de preferencias, construyendo un perfil y buscando similitudes con las características de diferentes productos.

Existen ventajas de la técnica ya que se recomienda por contenido y no por opiniones subjetivas de otros usuarios ya que el sistema se basa en el historial del usuario. Sin embargo también existen varias desventajas de este sistema de recomendación, una de ellas es que el usuario tiene que rankear un número suficiente de productos para que el sistema pueda aprender y entender sus preferencias, por eso se vuelve difícil para el sistema aprender a adaptarse a los cambios hasta no haber recolectado un número suficiente de ratings actualizados.

El proceso consta de tres fases:

- Representación del item: Consiste en organizar la información genérica, expresar las propiedades de los artículos mediante vectores para luego calcular la similitud con el perfil de los usuarios. Para ello, se puede emplear el método de representación TF-IDF en el caso de documentos. El TF-IDF (Frecuencia de término - Frecuencia inversa de documento) es un indicador estadístico que mide la relevancia de una palabra en un documento y en un conjunto de documentos.
- Perfil de aprendizaje: El propósito de esta etapa es construir el perfil del usuario usando algoritmos de extracción de temas como el LDA, que obtiene las posibles recomendaciones pertinentes para el usuario, evaluando la similitud de los vectores resultantes de los procesos anteriores.

El algoritmo LDA (Latent Dirich Allocation) es un modelado de temas y es utilizado para clasificar texto de un documento en un tema particular. El

algoritmo construye un tema por cada modelo de cada documento y se utiliza este algoritmo para la separación e interpretación de temas de un documento

### **Basado en filtrado colaborativo**

El filtro colaborativo es uno de los algoritmos de aprendizaje automático más utilizado para recomendación de productos o servicios. Empresas que brindan servicios alrededor del planeta utilizan esta técnica como parte de su sistema de recomendación, desarrolladores usan este algoritmo para sugerir a los compradores en función de los gustos y disgustos de compradores similares.

El algoritmo de filtro colaborativo se apoya en las valoraciones o acciones realizadas en los artículos por otros usuarios que tengan un perfil parecido. Al igual que la técnica basada en contenido, este método requiere conocer los intereses del usuario, es decir el sistema de recomendación de filtro colaborativo es el proceso de recomendar productos usando el criterio de otros usuarios cercanos. Para este sistema se emplean diferentes algoritmos de aprendizaje como agrupamientos o redes neuronales. [2]

Los sistemas de recomendación de filtrado colaborativo se clasifican en dos tipos según el modo de hacer las predicciones: basado en usuarios y basado en ítems

- Basado en usuarios: Es cuando usan el perfil de otros usuarios parecidos para hacer la predicción. Para este caso los usuarios se agrupan en categorías o grupos, donde los usuarios comparten intereses similares en el mismo grupo.
- Basado en ítems: Se centra en la relación que existe entre los ítems, en lugar de las relaciones entre los usuarios. Si los usuarios dan calificaciones positivas sobre algunos artículos, el sistema recoge los artículos candidatos a ser elegidos basándose en el rating histórico de calificaciones del usuario.

El mayor inconveniente de este sistema es el inicio en frío, ya que para los ítems nuevos no tienen una calificación asignada y por lo tanto no pueden ser recomendados. [3]

### Basado en grafos

Un grafo está formado por un conjunto de vértices que también son denominados nodos, y representan en este ejemplo a los productos, los mismos que tienen interconexión entre ellos dentro de la estructura del grafo y  $A$  es el conjunto de aristas que representan estas conexiones entre los vértices, reflejando así las relaciones existentes dentro del grafo.

Matemáticamente, se establece la definición de un grafo  $G$  a través de tuplas, como se visualiza en la siguiente ecuación:

$$G = (V, A)$$

En la siguiente figura presenta un caso ilustrativo de un grafo compuesto por  $X$  aristas y  $X$  vértices.

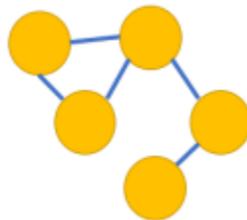


Figura 1. Grafo formado por  $X$  vértices y  $X$  aristas.

Sea  $G = (V, A)$  un grafo  $(p, q)$ , Considerar la matriz:  $M = [M_{ij}]$  con dimensiones  $p \times q$  en el cual cada columna y fila de la matriz  $M$  se relaciona con un vértice único en el conjunto  $V$ . La matriz de adyacencia  $G$  describe la conexión entre los nodos del grafo. Esta matriz se utiliza de manera efectiva para almacenar esta

información ya que proporciona una representación eficiente que facilita un acceso rápido a los datos de las conexiones entre los vértices o nodos.

La representación de la matriz de adyacencia se presenta como una matriz cuadrada, donde tanto las filas como las columnas corresponden a los vértices del grafo. Si dos nodos son conectados por una arista, el valor es 1 o algún otro valor distinto de cero en la matriz. Por otro lado, si no existe una conexión entre dos vértices a través de una arista, el valor se vuelve 0 en la matriz. En situaciones de grafos ponderados, los valores de la matriz pueden indicar los pesos relacionados con las aristas. En la figura 2 se muestra la matriz de adyacencia y cómo se establece la relación entre un grafo.

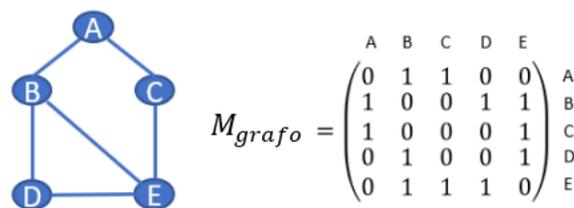


Figura 2. Ejemplo de un grafo dirigido y su matriz de adyacencia

Una vez que disponemos de la matriz de adyacencia, podemos llevar a cabo el análisis de nuestro grafo de forma eficiente y varias operaciones. Como ejemplo, podemos verificar la existencia entre dos vértices de una arista o identificar los vecinos de un nodo específico. Para esto último, basta con observar la fila o columna correspondiente a dicho nodo en la matriz, donde los valores igual a 1 indicarán las conexiones con otros vértices, lo que representa sus vecinos.

Otra de las aplicaciones clave de la matriz de adyacencia es su capacidad para facilitar la implementación de los algoritmos de búsqueda y recorrido en los grafos de una manera estructurada. Los algoritmos de búsqueda de caminos o pathfinding se utilizan para determinar el camino más corto entre dos nodos, ya sea en términos del número menor de saltos o peso en la ruta. Los dos principales algoritmos para realizar esto son:

Algoritmo de búsqueda en profundidad o Depth First Search (DFS) es un algoritmo para explorar un grafo y funciona de la siguiente manera: comienza en un nodo inicial y se expande de manera recursiva, visitando cada nodo que encuentra, avanzando de un nodo de nivel superior a uno de nivel inferior. Luego, cuando se ha recorrido todo el camino y no hay nodos por explorar, repite el procedimiento con cada uno de los vecinos del nodo actual. La figura 3 a continuación ilustra cómo se lleva a cabo este proceso de exploración de manera recursiva. Este algoritmo es particularmente útil en situaciones como la resolución de laberintos, donde se busca encontrar una solución.

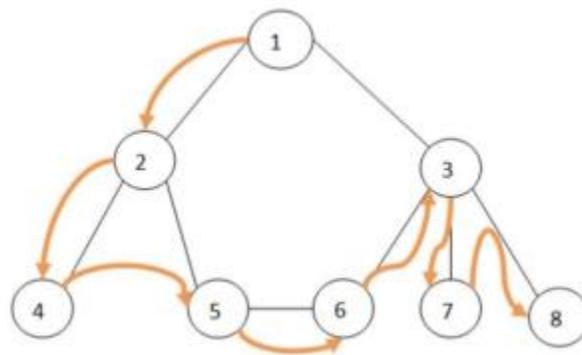


Figura 3. Algoritmo de búsqueda en profundidad [4]

Algoritmo de búsqueda en anchura o como Breath First Search (BFS) está diseñado para explorar los nodos del grafo. Inicia eligiendo un vértice o nodo de origen en el grafo, al que se denomina raíz y se etiqueta con un valor 0. Luego, procede a explorar todos los nodos vecinos de la raíz, etiquetándolos con un valor 1. A continuación, se exploran los nodos vecinos de estos nodos de nivel 1, incrementando el valor en 1 en cada salto. Este proceso se repite hasta que se haya recorrido todo el grafo de manera exhaustiva, buscando al mismo tiempo el camino más corto entre dos nodos,  $i$  y  $j$ . Si no existe un camino directo entre  $i$  y  $j$ , se asume que la distancia entre estos dos nodos es considerada como infinita.

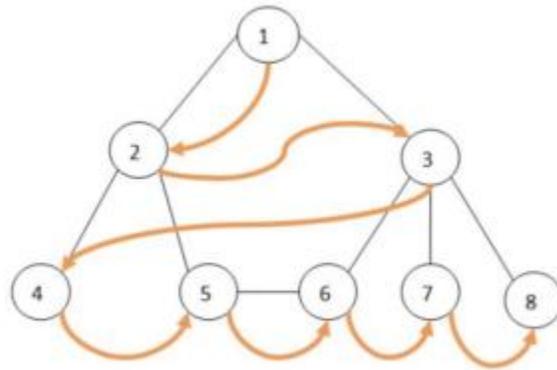


Figura 4. Algoritmo de búsqueda en anchura [4]

En un grafo no dirigido el grado de un nodo hace referencia a la cantidad de conexiones que se conectan a dicho nodo y se encuentra determinado por la siguiente fórmula:

$$\sum_j k_j = \sum_{i,j} M_{i,j} = 2L$$

En donde la variable  $L$  denota el doble del número de aristas debido a que es un grafo no dirigido el número de aristas se duplica, razón por la cual el número de enlaces es multiplicado por 2. Por otro lado para el cálculo del grado medio de un grafo se utiliza la siguiente fórmula:

$$\langle k \rangle = \frac{2L}{N}$$

En el tipo de grafo dirigido, la definición de grado se divide en dos categorías: el grado entrante de un nodo se define como la cantidad de aristas que se dirigen hacia ese nodo, y el grado saliente de un nodo se relaciona con la cantidad de aristas que se originan en ese nodo, mientras que la variable  $N$  representa el total de conexiones en el grafo.

El grado de las aristas se define como los nodos enlazados por una arista. En los grafos no dirigidos, este valor es constantemente igual a 2, ya que dos nodos están vinculados por la misma arista. En un grafo dirigido, el grado puede ser 1 o 2, ya

que puede haber conexiones unidireccionales o bidireccionales entre dos nodos. La siguiente fórmula ilustra cómo se realiza este cálculo, y en este caso, ya no es necesario realizar la multiplicación por 2 previamente, ya que dos nodos no necesariamente están conectados de manera recíproca.

$$\langle k^{in} \rangle = \langle k^{out} \rangle = \frac{L}{N}$$

También es importante explicar la densidad de un grafo que es una métrica que refleja el nivel de conexiones en ese grafo, es otras palabras, se trata de determinar la cantidad de aristas en relación con el número total de nodos. Para calcular esta métrica, se divide el total de aristas del grafo para el máximo de aristas de un grafo completo. Se debe tomar en cuenta que la densidad varía entre 0 y 1 en los grafos no dirigidos, y la densidad puede superar en 1 en los grafos dirigidos, ya que se tiene aristas con direcciones específicas entre los vértices. La fórmula siguiente se utiliza para calcular la densidad de un grafo con un orden de  $N$ .

$$D = \frac{size}{MaxSize} = \frac{L}{\binom{N}{2}}$$

Siendo  $MaxSize$  representa el número máximo de nodos por  $N$  nodos y se obtiene mediante el siguiente cálculo:

$$MaxSize = \binom{N}{2} = \frac{N(N-1)}{2}$$

Es esencial comprender la importancia de la matriz de adyacencia, debido a que representa de una manera precisa las relaciones dentro de un grafo. Esto facilita la ejecución eficaz de operaciones fundamentales y algoritmos avanzados en grafos, y desempeña un papel crítico en múltiples disciplinas, como por ejemplo la teoría de redes, optimización y planificación de rutas.

## Grafos basados en sus propiedades

La orientación de un grafo es una característica fundamental que se utiliza para clasificar los grafos en dos categorías principales: grafos dirigidos y no dirigidos. En grafos no dirigidos, las conexiones entre los nodos carecen de dirección específica. Por otro lado, en los grafos dirigidos, las conexiones son asimétricas, lo que significa que la secuencia de los nodos define la dirección de la arista correspondiente. Un ejemplo ilustrativo de estos dos tipos de grafos sería el análisis de amistades, donde en un grafo no dirigido se asume que la conexión es recíproca, mientras que un grafo dirigido resulta beneficioso para examinar las vías de una ciudad, dado que la mayoría presenta un único sentido de circulación. [5]

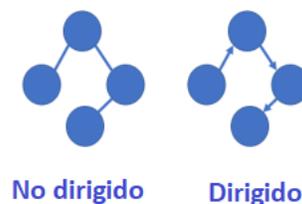


Figura 5. Tipos de grafo no dirigido y dirigido

La siguiente propiedad que se utiliza para clasificar un grafo es su ciclicidad. Se considera cíclico un grafo cuando incluye por lo menos un ciclo, que es una secuencia de nodos que comienza y finaliza en el mismo vértice. Por otro lado se considera acíclico si un grafo no tiene ciclos. Para saber si un grafo es cíclico o no, se pueden aplicar diversas metodologías de recorrido. En el caso de un grafo no dirigido, en ocasiones se utiliza el algoritmo de búsqueda en profundidad para explorar los nodos y aristas. Es importante identificar un nodo previamente visitado, y no es el padre del nodo actual, señala la presencia de un ciclo en el grafo. Cuando un grafo contiene un ciclo lleva a la conclusión de que el grafo es cíclico ya que es imposible establecer una secuencia topológica. [6]

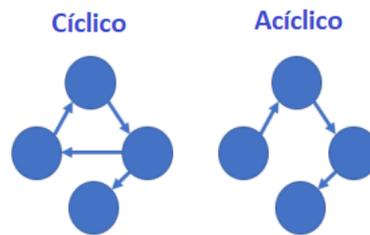


Figura 6. Tipo de grafo cíclico y acíclico

La tercera propiedad utilizada para clasificar un grafo es la ponderación. Cuando se asignan pesos a los nodos o a sus conexiones se llaman grafos ponderados. Por otro lado, en el caso de grafos no ponderados, las conexiones carecen de peso. Cuando se examina la ponderación entre dos nodos, se suele utilizar un algoritmo que determina la ruta más corta en función del número de saltos. En cambio, en un grafo ponderado, se prioriza analizar la ruta más corta según el costo total en lugar del número de saltos.

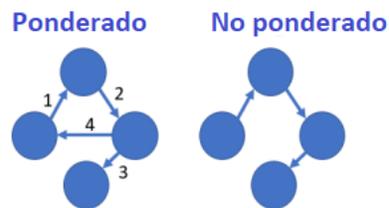


Figura 7. Tipo de grafo ponderado y no ponderado

### Grafos basados en su estructura

**Grafo aleatorio:** Se utiliza este modelo de grafo para representar redes aleatorias donde se parte de un conjunto que consta de  $N$  nodos, y la relación entre cada par de vértices o nodos se establece mediante  $p$  que es una probabilidad para la existencia de una arista que los conecte. La creación del grafo aleatorio se realiza escogiendo de manera independiente, para cada par de nodos, si estarán conectados o no, con una probabilidad  $p$ .

La característica fundamental de los grafos aleatorios es cuando una arista existe o no y eso se decide de una manera totalmente independiente de las otras aristas. Por lo tanto, se obtiene una distribución de grado que sigue una distribución binomial, esto implica que la probabilidad de que un vértice tenga  $k$  conexiones se ajusta a una distribución binomial. Esto es muy relevante al buscar la modularidad de un grafo, que mide la calidad de la división de vértices en comunidades o grupos, se debe tomar en cuenta que los grafos aleatorios, generalmente no muestran una fuerte modularidad ya que sus conexiones son aleatorias. En su mayoría, las conexiones tienden a distribuirse de manera uniforme, lo que complica la detección de comunidades claramente definidas. La modularidad suele ser baja a menos que la densidad de aristas sea considerablemente alta. La probabilidad de conexión se normaliza mediante la siguiente fórmula.

$$p_{uv} = \frac{k_u k_v}{2L - 1}$$

En el caso de superar este número indica la existencia de comunidades. Un grafo aleatorio se visualiza en la figura 8.

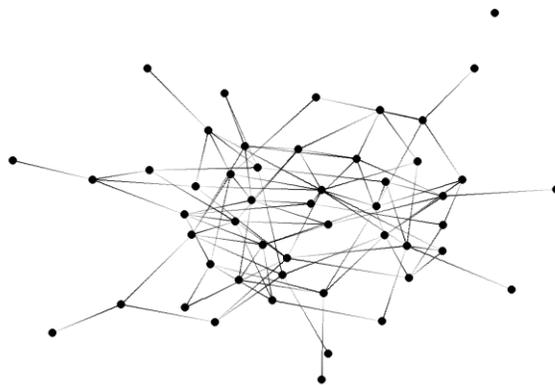


Figura 8. Grafo Aleatorio

**Grafo completo:** Es un grafo completamente conectado en donde todos los pares de vértices están interconectados, es decir, cada vértice presenta una conexión directa con todos los otros nodos en el grafo completo.

Para identificar las aristas de un grafo completo con  $N$  nodos, se utiliza la siguiente ecuación:

$$\frac{N * (N - 1)}{2}$$

Donde cada vértice se interconecta con  $N - 1$  vértices distintos y es dividido por 2 para que no exista duplicidad de aristas. Con eso los grafos completos pueden tener una alta densidad en cuanto a las conexiones entre sus nodos.

Usualmente, se representa al grafo completo como  $K_n$ , donde  $N$  indica la cantidad de vértices en el grafo.

La simetría es la particularidad fundamental de los grafos completos, esto implica que si el vértice A está conectado al vértice B, entonces el vértice B también está conectado al vértice A debido a que no tienen orientación las aristas en un grafo completo, estableciendo una relación simétrica entre los nodos. Por lo tanto, cuando el número de vértices aumenta, la cantidad de aristas también crece significativamente. Por lo que puede resultar que este tipo de grafos sean difíciles de manejar en métodos de almacenamiento y cálculos, y sobre todo cuando se tienen muchos nodos. Entonces en la práctica, se trabaja con distintos tipos de grafos con menor densidad o estructuras específicas para representar relaciones entre vértices en procedimientos más complicados. En la figura 9 se puede visualizar un grafo completo.

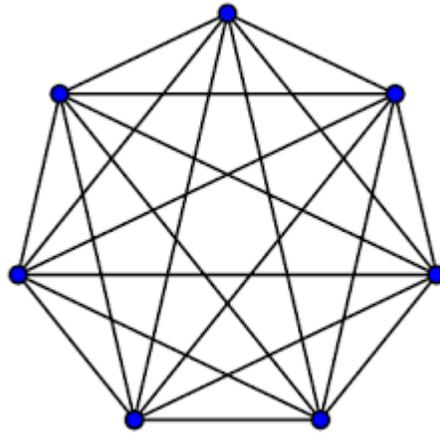


Figura 9. Grafo completo

**Grafo clique:** es una variante particular de un grafo completo donde todos los vértices están conectados entre sí, es decir, cada par de vértices se encuentra interconectado mediante una arista. Este grafo se considera como un subgrafo completo de un grafo más grande. Esto implica que al elegir un grupo de nodos de un grafo completo y establecer conexiones entre ellos, generamos un grafo clique. En la figura 10 se visualiza un grafo clique.

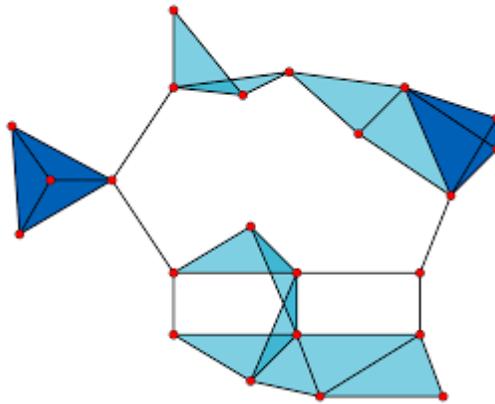


Figura 10. Grafo clique

**Grafo árbol:** Se compone de nodos unidos por aristas, y cumple con una condición importante: existe un único camino entre cada par de nodos en el grafo, lo que significa que no contiene ciclos. En un árbol, cada nodo, a excepción de uno denominado "raíz," posee un solo nodo llamado padre, pero el nodo raíz carece de

un padre. Sin embargo, un nodo puede tener varios hijos. La jerarquía se refleja en la estructura de un árbol, ya que establece una relación de nivel entre los vértices, en otras palabras, siempre los nodos hijos están vinculados directamente a sus nodos padres correspondientes, creando una estructura de ramificación. Además, se caracteriza por la ausencia de ciclos, lo que significa que en un grafo árbol, no es posible seguir una sucesión de aristas y retornar al nodo inicial. En la figura 11 se presenta un ejemplo de un grafo árbol.

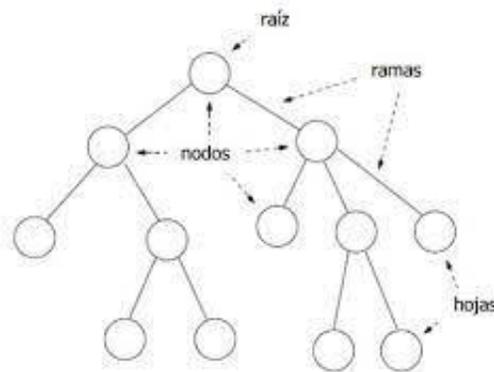


Figura 11. Grafo árbol

**Grafo bipartito:** En este tipo de grafo los nodos se separan en dos grupos claramente definidos, y todas las aristas establecen conexiones entre nodos de un conjunto y nodos del otro conjunto. Por lo tanto, se define el grafo bipartito  $G = (U, V)$  como un grafo donde todos los nodos de  $U$  están conectados a nodos de  $V$ .

Un grafo bipartito se representa visualmente mediante dos conjuntos de nodos dispuestos en columnas, con las aristas conectando nodos de una columna con nodos de la otra columna. La característica distintiva del grafo radica en la partición de los nodos en dos grupos bien definidos, como se muestra en la figura 12.

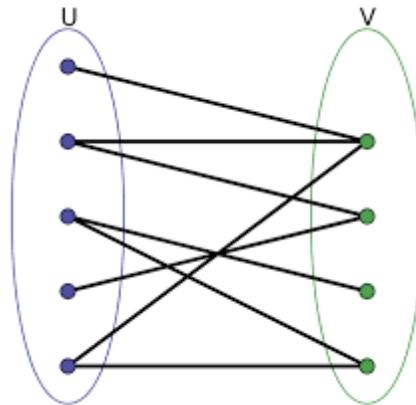


Figura 12. Grafo bipartito

Los grafos bipartitos encuentran aplicación en la modelación de relaciones entre dos conjuntos. Se puede emplear para poder representar las relaciones entre usuarios y eventos o grupos en redes sociales.

Dentro de estos grafos, existen subgrafos que son proyecciones que se consiguen de las aristas del grafo original en uno de los conjuntos de nodos. Concretamente, existen dos proyecciones: la primera en el conjunto de nodos izquierdo y la segunda en el conjunto de nodos derecho.

La proyección en el conjunto de nodos izquierdo implica la eliminación de las aristas que conectan nodos dentro del mismo conjunto, manteniendo solamente las aristas que vinculan nodos de diferentes conjuntos, es otras palabras, aquellas que parten desde nodos izquierdos hacia nodos derechos y viceversa.

Cuando se realiza la proyección de un grafo bipartito, se obtiene uno nuevo que ilustra las conexiones entre los vértices de un conjunto definido, excluyendo las conexiones internas en ese conjunto. Por lo tanto, resulta ventajoso para comprender y analizar las relaciones entre los elementos de un conjunto particular.

[7]

La técnica de recomendación basada en grafos se enfoca en la construcción de un grafo para generar las predicciones y recomendaciones donde cada usuario/artículo

puede verse como entidad o nodo y las posibles relaciones que existen entre ellos se pueden representar como aristas [8]. La figura 13 muestra un grafo utilizado como sistema de recomendación de artículos, donde los usuarios y los productos se representan como nodos.

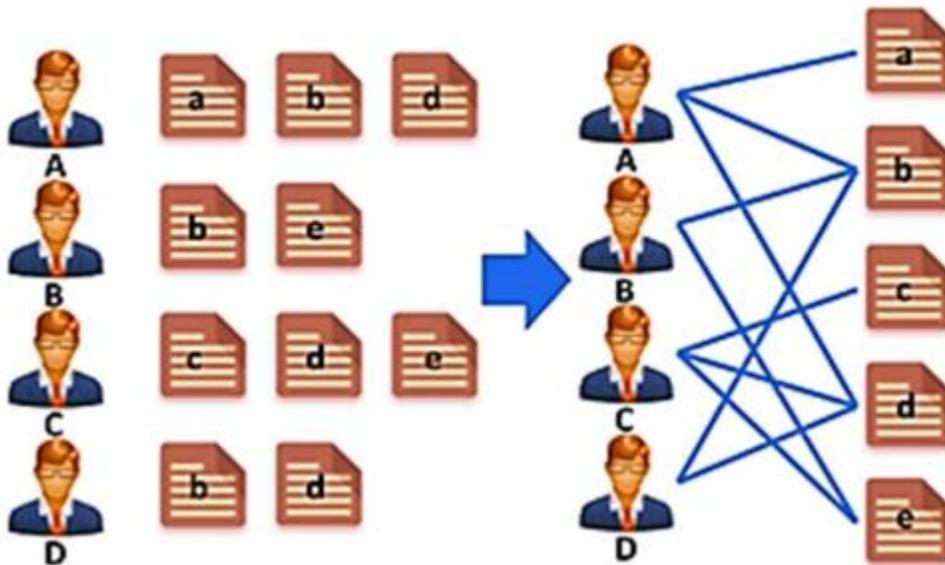


Figura 13. Grafo utilizado para un sistema básico de recomendación [9]

Para crear la sugerencia, generalmente se recurre a algoritmos de exploración o búsqueda, como el algoritmo de caminata aleatoria con reinicio. Este método evalúa la probabilidad de llegar desde un punto de partida  $i$  a un punto de destino  $j$ , pasando por nodos adyacentes. Es importante tener en cuenta la posibilidad de regresar al punto de inicio  $i$ , lo que justifica el término "reinicio". [10]

### Híbrido

El método híbrido se propone fusionar las características de dos o más enfoques con el fin de complementar sus respectivas limitaciones. Esto se debe a que, por separado, estos métodos tienen sus ventajas y desventajas, ya que no son capaces

de abordar completamente un amplio espectro de opciones. Por lo tanto, el método híbrido se encarga de combinar diversas técnicas de recomendación y recopilar información de múltiples fuentes.

Como se ilustra en la figura 14, un sistema híbrido se asemeja a la combinación de sistemas basados en contenido y sistemas de filtrado colaborativo. De esta manera, se tiene en cuenta el perfil del usuario, que se crea a partir de sus intereses, así como las calificaciones que el usuario otorga a diversos productos. Al final, se realiza una recomendación basada en la coincidencia de las dos listas de productos diferentes que resulten relevantes para el usuario.



Figura 14. Método híbrido paralelo

## IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

Los métodos de compra más antiguos donde los productos se producían en grandes cantidades para un mercado y audiencia únicos ya no es viable ya que los mercados basados en largos ciclos de desarrollo y productos ya no pueden sobrevivir, es por eso que para seguir siendo competitivos, los mercados necesitan ofrecer diferentes productos y servicios a diferentes clientes con diferentes necesidades. El cambio a las compras en línea ha hecho que corresponda a los productores y minoristas personalizar las compras según las necesidades de los clientes y, al mismo tiempo, ofrecer más opciones de las que antes no eran posibles.

En la tendencia reciente de la economía digital, una mayor interacción entre las organizaciones y sus clientes se reconoce como la característica clave del entorno de marketing, y las empresas enfrentan el desafío de proporcionar productos y servicios diferenciados a cada cliente en función de la información que recopilan de los clientes por lo que el objeto de estudio es realizar un sistema de recomendación para que sea de ayuda a las empresas a mejorar la experiencia de sus compradores en el sitio web y poder diseñar una metodología de recomendaciones personalizadas y relevantes para tener como resultado una mejor adquisición, fomentar la fidelidad de clientes, impulsar las ventas y mejorar la eficiencia en la toma de decisiones.

Cuando un nuevo cliente sin historial de compras visita el sitio web de comercio electrónico por primera vez, se le recomiendan los productos más populares vendidos en el sitio web de la empresa. Una vez que realiza una compra, el sistema de recomendaciones se actualiza y recomienda otros productos en función del historial de compras y las calificaciones proporcionadas por otros usuarios en el sitio web.

## PLANTEAMIENTO DEL PROBLEMA

Al paso de los años se ha visto un crecimiento exponencial en el mundo web y el uso de Internet, es por esto que ha crecido de igual manera los datos y la información que existe en el mundo y cada vez son más complicadas las búsquedas de información. Es por lo que a raíz de este problema se ha ido consolidando herramientas que ayudan a disminuir la sobrecarga de información y entre estas herramientas existen los Sistemas de recomendación, los cuales filtran los ítems de la información a partir de técnicas que identifican los principales atributos que más se acomodan a las preferencias de los usuarios, recomendaciones que se generan a partir de gustos de otros usuarios o del contenido de la información [11].

Un “sistema recomendador” analiza y procesa información histórica de los usuarios (edad, compras previas, calificaciones), de los productos o de los contenidos (marcas, modelos, precios, contenidos similares) y la transforma en conocimiento accionable, es decir, predice qué producto puede ser interesante para el usuario y para la empresa. Los recomendadores, además, tienen cierto nivel de autonomía a la hora de presentar las recomendaciones al usuario final. Estos sistemas se han hecho muy populares en los últimos años.

Existen una gran variedad de SR en el mundo Web, Amazon cuenta con un algoritmo personalizado de recomendación que ellos llaman filtrado colaborativo ítem a ítem es decir utiliza el historial de compras y navegación del usuario para sugerir productos que se adapten a sus gustos y necesidades. La empresa espera que esta herramienta mejore la experiencia del usuario y aumente sus ventas. Con este lanzamiento, Amazon se posiciona como una empresa que se preocupa por ofrecer la mejor experiencia de compra en línea a sus clientes [12].

## **OBJETIVO GENERAL**

El objetivo principal en este proyecto es desarrollar un Sistema de recomendación utilizando la técnica basada en grafos y técnicas de aprendizaje automático para mejorar la precisión y relevancia de las recomendaciones ofrecidas a los usuarios ya que los grafos ofrecen una representación efectiva de las relaciones y conexiones entre usuarios y productos.

## **OBJETIVOS ESPECÍFICOS**

1. Utilizar el dataset proporcionado por Amazon para tener la información que se desea utilizar para un sistema de recomendación
2. Realizar la actividad de limpieza de datos para usar en el algoritmo a modelar.
3. Recomendar artículos a los usuarios según el historial de compras utilizando la técnica basada en grafos
4. Encontrar productos para un usuario en particular mediante la identificación de patrones basados en las preferencias de múltiples datos de usuarios
5. Análisis de los resultados con el método de recomendación que se utilizó para encontrar el objetivo principal del estudio

## JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

En la actualidad, el uso de nuevas tecnologías en todos los aspectos de nuestras vidas, el consumo en línea ha experimentado un rápido avance debido a que en el comercio electrónico siendo este uno de los principales canales, los usuarios tienen acceso a una multitud de productos en diversas categorías y la selección de productos que se ajusten a las necesidades y preferencias individuales puede resultar desafiante debido a la abrumadora cantidad de opciones disponibles en estas plataformas. Los sistemas de recomendación han surgido como una solución a esta problemática. Su objetivo principal radica en identificar aquellos productos que mejor se adecúan a los intereses del usuario al que va dirigida la recomendación, con el fin de facilitar la toma de decisiones apropiadas.

A continuación, se detalla el proceso aplicado para el análisis de un sistema de recomendación:

### **1. Recolección de datos**

La base de datos utilizada para el análisis del presente proyecto es de las calificaciones de productos electrónicos de Amazon que es una plataforma conocida de comercio en línea que permite a los clientes comprar una amplia gama de productos. Este conjunto de datos tiene más de 7 millones de calificaciones de productos relacionados con la electrónica vendidos en su sitio web. En dichos registros constan la identificación única del usuario y del producto y la calificación del 1 al 5 dada por el usuario a un producto (Rating) donde 1 es una baja calificación y 5 la mejor calificación del producto. También se tiene la marca de tiempo (Unix) de la calificación dada.

### **2. Limpieza, pre-procesamiento y/o transformación de datos.**

Tanto para la limpieza y pre-procesamiento de datos se está usando la herramienta Jupyter Notebook con el lenguaje de programación Python ya que es versátil y ampliamente utilizado ya que su sintaxis es clara, legible y sobre todo cuenta con

una biblioteca estándar rica en módulos y funciones que abarcan una amplia gama de tareas. Esto facilita el desarrollo de aplicaciones sin necesidad de escribir código desde cero, ya que muchas de las funcionalidades comunes están disponibles de inmediato.

En primera instancia, importamos las librerías necesarias para análisis de la base de datos como son: numpy, pandas, matplotlib, seaborn, networkx. Posteriormente importamos la base de datos en forma de un Dataframe para utilizar las funciones de pandas. Con ayuda de dichas funciones primero se verifica si existen valores nulos en cada campo, dando como resultado que no existen valores nulos en nuestra base de datos.

```

UserId      0
ProductId   0
Rating      0
Timestamp   0

```

Por fines académicos se ha escogido los registros de calificaciones mayores a 4 ya que al tener demasiados registros se dificulta realizar los grafos porque se requiere mayor procesamiento del equipo de computación donde se está realizando el respectivo análisis y sobre todo no se visualizaría de la mejor manera los nodos y relaciones de los grafos. Eliminando los registros de calificación menores a 4 quedan 4347541 registros.

### 3. Identificación y descripción de variables

Existen 4 variables en la base de datos, en la tabla 1 se presenta cada una de ellas.

**Tabla 1.** Conjunto de variables a utilizar

Variable	Tipo	Descripción
UserId	Númerica	Identificación única del usuario
ProductId	Númerica	Identificación única del producto

Rating	Númerica	Calificación en el rango de 1 a 5 dada por el usuario a un producto donde 1 es una baja calificación y 5 la mejor calificación del producto.
Timestamp	Númerica	Marca de tiempo (Unix) de la calificación dada

#### 4. Visualización de variables

La visualización de las variables de la base de datos proporcionará una comprensión de como están conformadas cada una de ellas y poder describir el comportamiento de las mismas.

Se debe tomar en cuenta que para esta parte del análisis de los datos, se tomará en cuenta todos los registros, sin aún eliminar las calificaciones menores a 4. Por lo tanto, en primer lugar se verifica en la figura 15 el número de usuarios, productos únicos y el tamaño total de los registros.

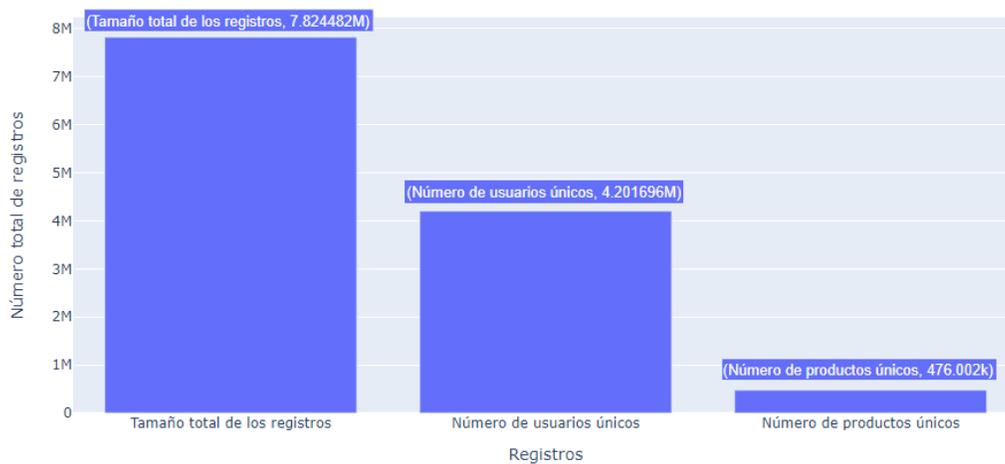


Figura 15. Número de usuarios y productos únicos

En segundo lugar, se verifica en la figura 16 el número total de las calificaciones (rating) distribuidos en 5 valores que es el rango de las calificaciones dadas por el usuario. Donde se puede observar que el mayor número se encuentra concentrado en el valor 5 (mayor calificación). Debido a esta concentración de datos en el rating 5 es también una de las razones por la cual se ha trabajado con esta parte de la base de datos.



Figura 16. Calificaciones dadas por los usuarios

En tercer lugar, se puede verificar en la figura 17 el top 10 de los productos más populares en orden descendente ya que se realiza la cuenta de los productos que se tienen mayor venta por lo que fueron valorados. Por ejemplo, se puede observar que el producto con ID # B0074BW614 tiene ventas por más de 18000 usuarios

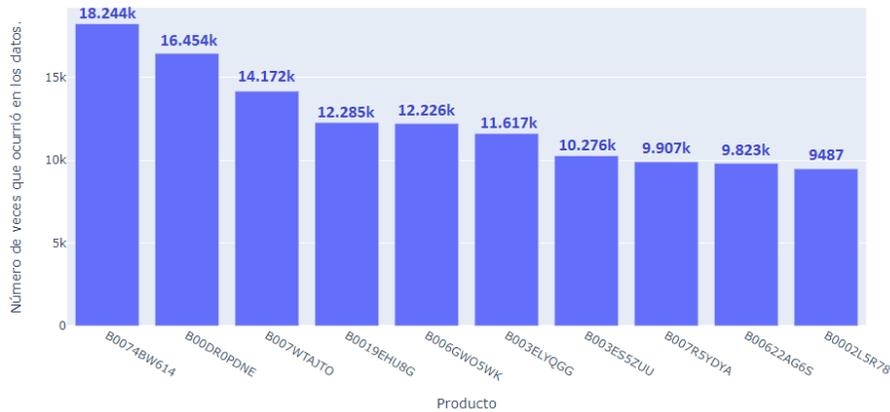


Figura 17. Top 10 de los productos más populares

Finalmente, en la figura 18 se puede observar la frecuencia de ratings para los top 10 de productos más vendidos, donde se verifica que en la mayoría de productos prevalece el rating 5 (buena calificación).

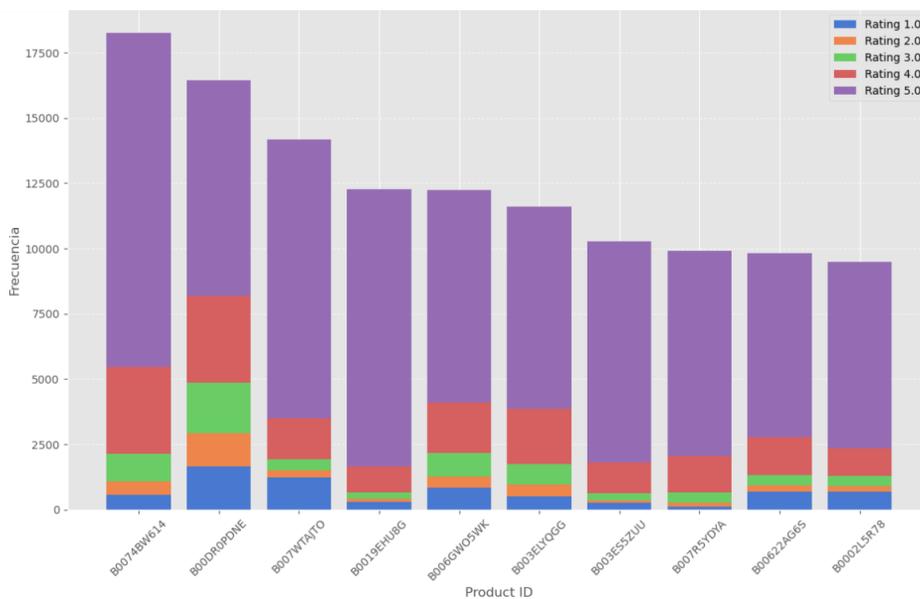


Figura 18. Frecuencia de Ratings para los top 10 de productos

## 5. Selección de modelo estadístico

En esta etapa se define el modelo o técnica que se va a utilizar para el sistema de recomendación del presente proyecto. Existen numerosos algoritmos y enfoques para implementar sistemas de recomendación en el contexto de e-commerce y como se mencionó en la revisión literaria existen 3 categorías principales: recomendación basada en contenido, recomendación colaborativa, basado en grafos y por último los híbridos.

La finalidad del proyecto es utilizar la técnica basado en grafos para una mejor comprensión de las relaciones entre productos y usuarios para proporcionar sugerencias más exactas y pertinentes, lo que puede mejorar la experiencia del usuario y aumentar las ventas.

La técnica basada en grafos emplea nodos y aristas

- **Nodos:** Estos elementos representan entidades, que pueden ser individuos, compañías u otros elementos que pueda ser rastreado, y se asemejan a una entrada en una base de datos relacional.
- **Aristas:** Son los enlaces que unen los nodos de un grafo, simbolizando las conexiones entre ellos. Constituyen el elemento fundamental en las bases de datos orientadas a grafos y constituyen una abstracción que no se encuentra en otros sistemas.[10]

Para modelar las compras de productos de los clientes se va a utilizar un grafo bipartito donde los dos conjuntos de nodos los del usuario y los productos, y los bordes indican que un usuario  $U$  compró un producto  $P$  como se puede observar en la figura 19.

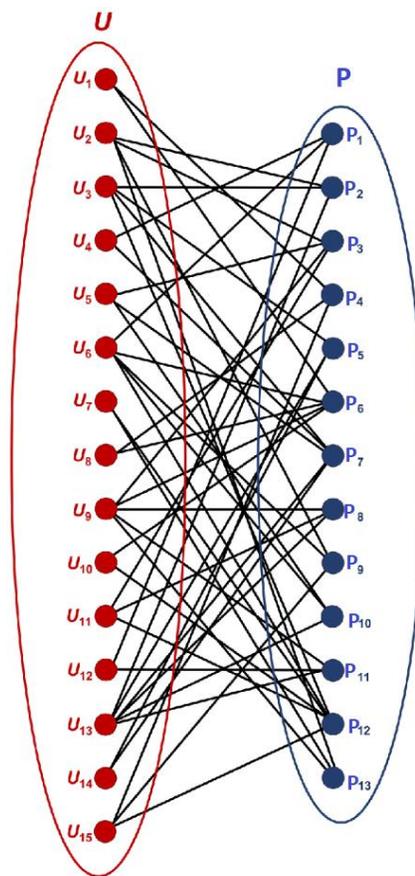


Figura 19. Red Bipartita

## RESULTADOS

Para construir una proyección ponderada de un grafo en NetworkX, se utilizó la función `weighted_projected_graph`. Esta función toma un grafo bipartito y proyecta sus nodos en uno de los conjuntos en una nueva red, donde las aristas están ponderadas según la cantidad de vecinos comunes entre los nodos proyectados. El resultado es un grafo ponderado donde las aristas están ponderadas en función de la cantidad de vecinos comunes que tienen los nodos proyectados, como se muestran en la figura 20

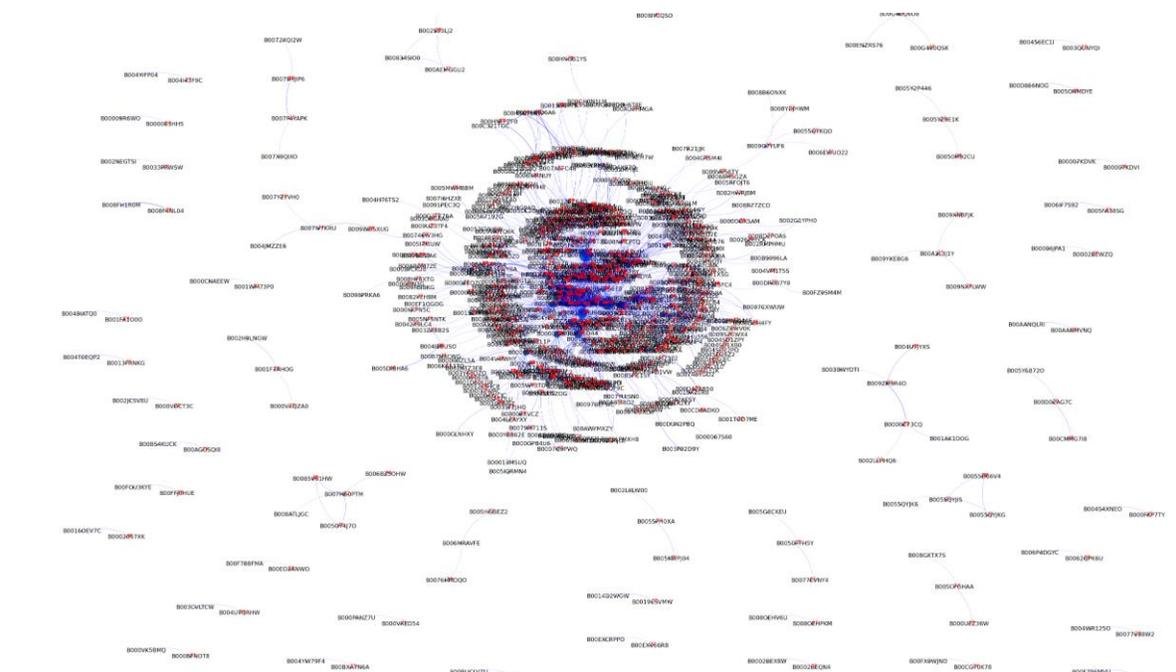


Figura 20. Red de relaciones entre productos

Para encontrar la componente más grande en un grafo utilizando NetworkX en Python, se usa la función `connected_components` para obtener todas las componentes conexas del grafo. El resultado de nuestra base de datos es 591 productos (nodos) en el componente más grande, como se puede observar en la figura 21.

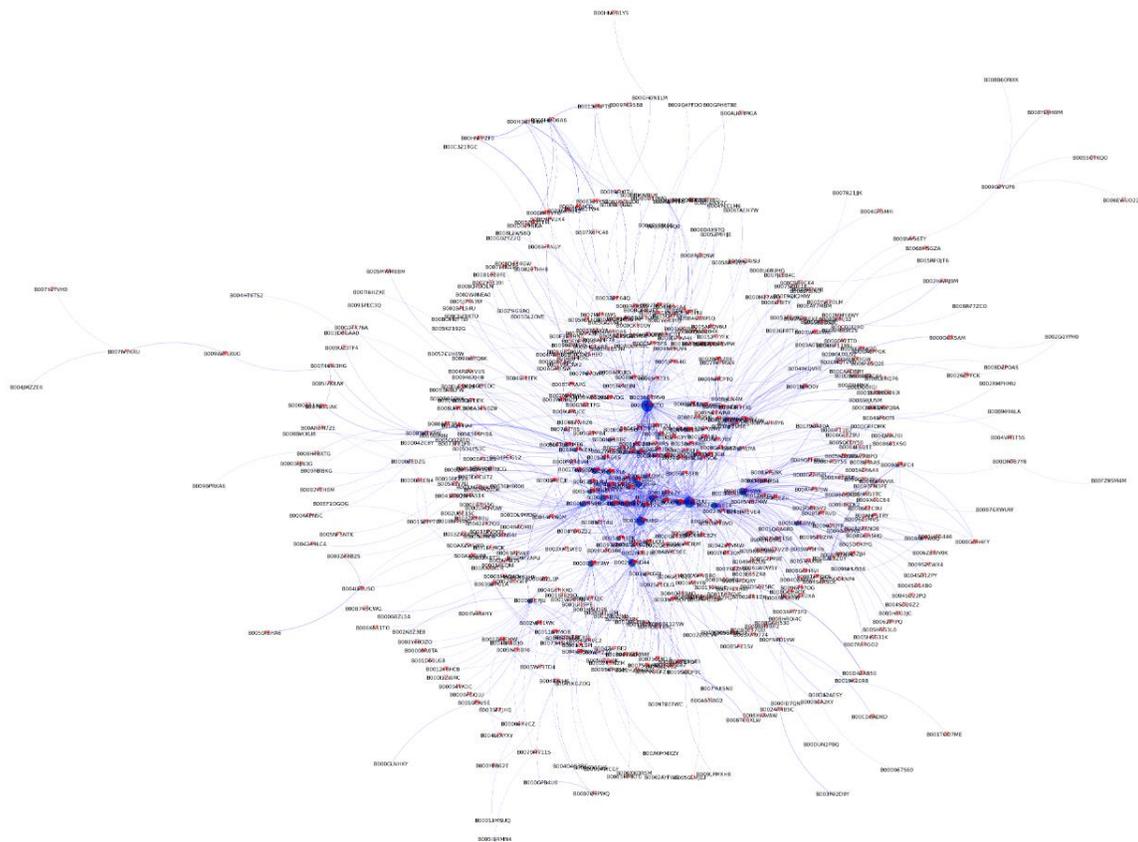
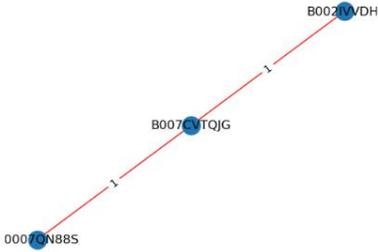
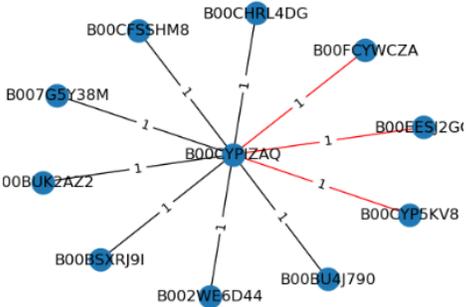
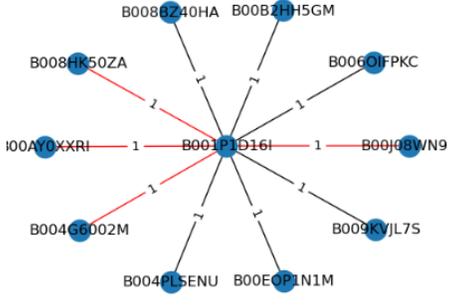
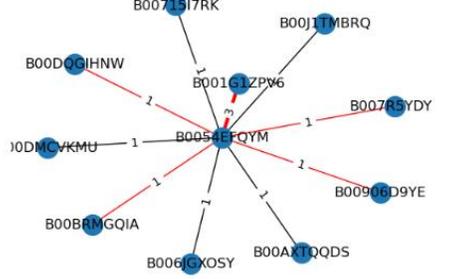


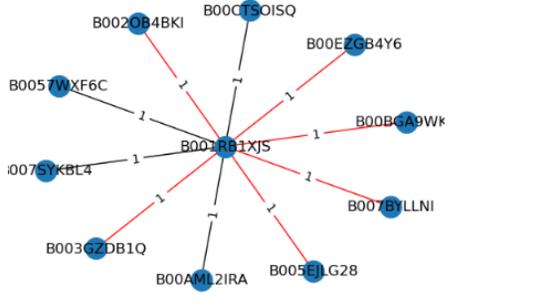
Figura 21. Componente más grande del grafo

En la figura 22, se puede observar una recomendación en un fragmento del grafo, se toma como criterio el producto B0074BW614 con su ponderación respectiva. En donde se observa los 4 productos conectados con mayor ponderación, es decir estos productos son recomendados a los usuarios que valoraron al producto B0074BW614



**Tabla 2.** Ejemplos de recomendaciones de productos

Producto	N° de recomendaciones	Productos recomendados
B007CVTQJG	2	 <p>A network diagram with a central node B007CVTQJG. It is connected to B007CVVDH (top right) and B007CVN88S (bottom left) by red lines. All edges are labeled with the number '1'.</p>
B00CYPITZAQ	3	 <p>A network diagram with a central node B00CYPITZAQ. It is connected to 11 other nodes: B00CHRL4DG, B00CFSSH8, B007G5Y38M, B00BK2AZZ, B00BSXRJ9I, B002WB6D44, B00BUJ790, B00FCYWCZA, B00EES12G, and B00CP5KV8. Red lines connect the central node to B00FCYWCZA, B00EES12G, and B00CP5KV8. All other edges are black. All edges are labeled with the number '1'.</p>
B001P1D16I	4	 <p>A network diagram with a central node B001P1D16I. It is connected to 10 other nodes: B0086Z40HA, B00BZH5GM, B008HK50ZA, B006GJFKC, B00AY0XXRI, B009G8WN9, B004G6002M, B009KVJL75, B004FL5ENU, and B00EOP1N1M. Red lines connect the central node to B0086Z40HA, B008HK50ZA, B00AY0XXRI, and B004G6002M. All other edges are black. All edges are labeled with the number '1'.</p>
B0054EFQYM	5	 <p>A network diagram with a central node B0054EFQYM. It is connected to 10 other nodes: B007LS17RK, B00J1M8RQ, B00DQGIHNW, B001G1ZPW6, B007R5YDY, B00MCVKMU, B00906D9YE, B00BEMGQIA, B006GXOSY, and B00AKTQDS. Red lines connect the central node to B007LS17RK, B00DQGIHNW, B00MCVKMU, B007R5YDY, and B00906D9YE. All other edges are black. All edges are labeled with the number '1'.</p>

B001RB1XJS	6	
------------	---	--

También se verifica la densidad de la red en el contexto de grafos que se refiere a la proporción de conexiones existentes en un grafo en comparación con el número total de conexiones que existen en un grafo completo. Cuanto más densa sea una red, más conexiones existen entre sus nodos en relación con el número máximo de conexiones posibles. En este caso el cálculo de la densidad es 0.000177 por lo que se verifica que tiene menos conexiones en relación con el número máximo de conexiones posibles

En la siguiente tabla se puede verificar algunas de las propiedades de la red

**Tabla 2.** Métricas de la red

Propiedad	Resultado	Descripción
Diámetro de la red	11	<p>Se refiere a la distancia más larga entre dos nodos en el grafo. En otras palabras, es la longitud del camino más largo que se puede tomar para ir desde un nodo a otro en la red.</p> <p>Este valor es alto debido la cantidad de nodos, las conexiones y la naturaleza de las relaciones representadas en el grafo. No hay un valor común específico para el diámetro de una red, ya que depende de las características particulares de la red en cuestión.</p>

Longitud de ruta promedio de la red	3,6721	Es una medida que indica la longitud promedio de los caminos más cortos entre los pares de nodos en el grafo. En otras palabras, es un promedio ponderado de las distancias entre los nodos en la red. Una longitud de ruta promedio más pequeña suele indicar una mayor eficiencia en la comunicación o acceso en la red.
Transitividad de la red	0,12	Indica qué tan conectados están los vecinos de un nodo en una red. En otras palabras, mide la probabilidad de que si un nodo A está conectado a un nodo B y un nodo B está conectado a un nodo C, entonces también existe una alta probabilidad de que el nodo A esté conectado al nodo C. Cuanto mayor sea el valor de transitividad, mayor será la probabilidad de conexiones transitivas en la red.

También se puede observar el grado de vecino promedio que es una medida que proporciona información sobre cómo se relacionan los nodos con sus vecinos en la red. Cuanto mayor sea el valor del grado de vecino promedio, mayor será el grado promedio de los vecinos de los nodos en la red. Esto puede indicar una mayor conectividad y redundancia en la red, lo que puede ser importante en diversos análisis de redes.

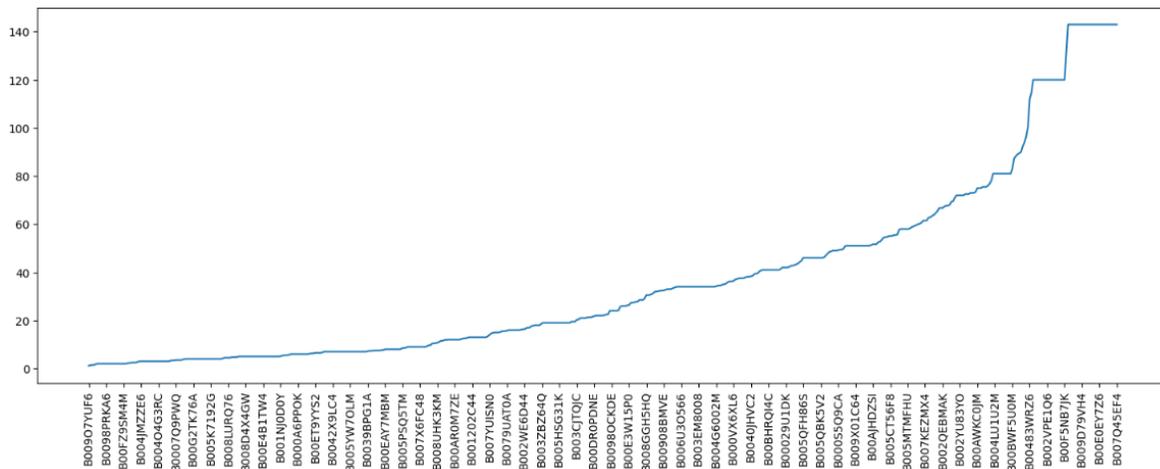


Figura 24. Grado de vecino promedio

Adicional se puede observar la centralidad de grafo ya que calcula el grado de importancia de un nodo en una red de acuerdo a la cantidad de conexiones directas que tiene. Los nodos con un alto grado de centralidad son aquellos que están más conectados con otros nodos en la red.

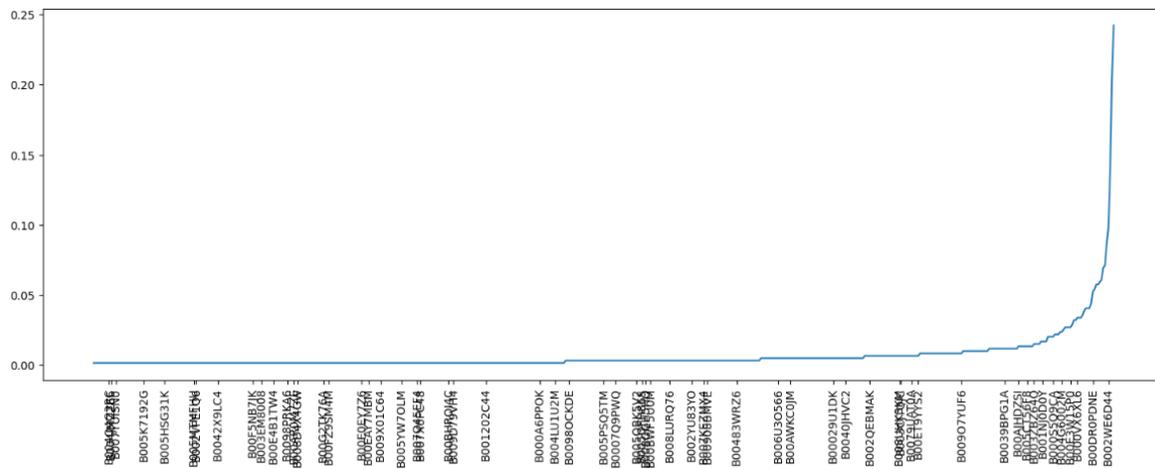


Figura 25. Centralidad de grafo

La cercanía de centralidad para evaluar el grado de importancia de un nodo de acuerdo a su distancia promedio a todos los demás nodos en la red. Cuanto más cercanos estén los nodos a todos los demás nodos en términos de distancia

geodésica, mayor será su centralidad de cercanía. Los nodos con alta centralidad de cercanía son aquellos que pueden comunicarse o alcanzar rápidamente otros nodos en la red

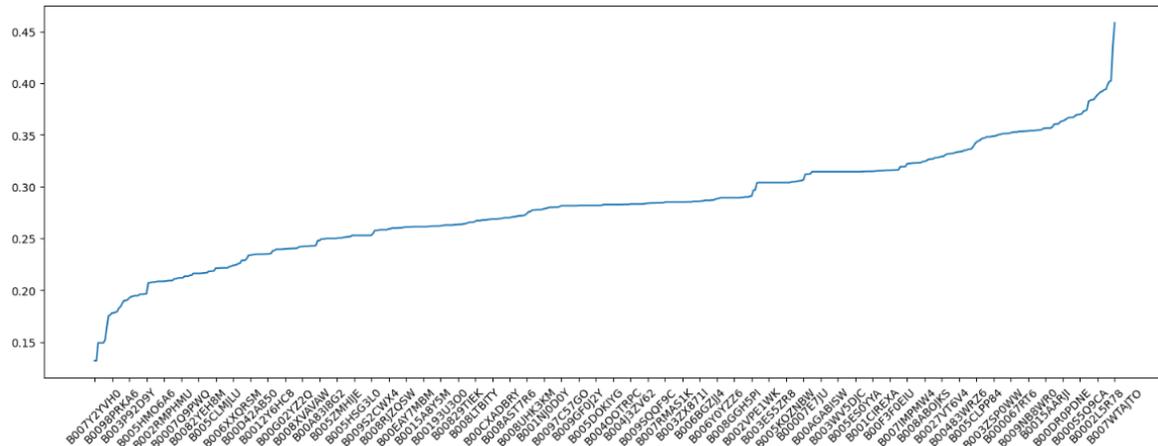


Figura 26. Cercanía de centralidad

La centralidad de intermediación evalúa la importancia de un nodo en función de la cantidad de veces que ese nodo se encuentra en el camino más corto entre otros pares de nodos en la red.

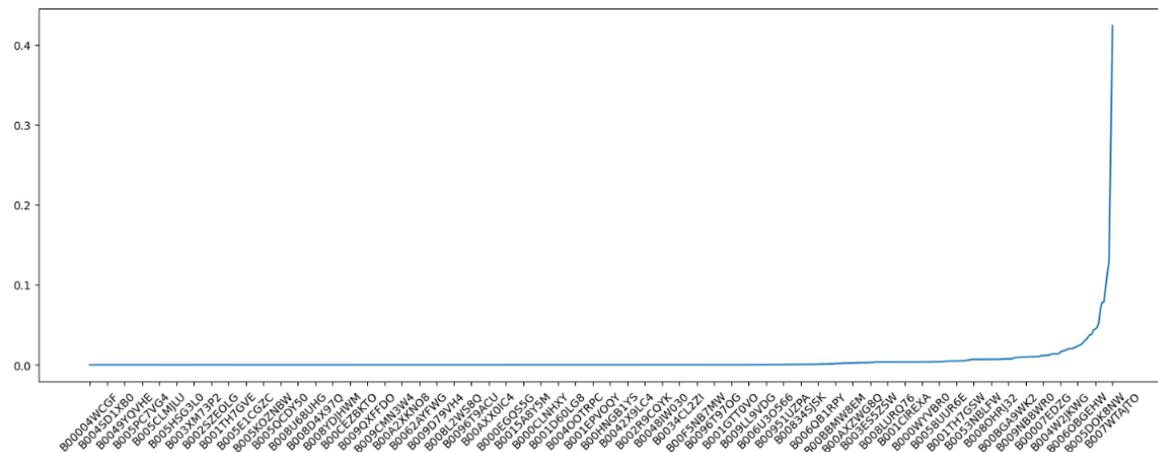


Figura 27. Centralidad de intermediación

## DISCUSIÓN DE LOS RESULTADOS

En los resultados expuestos en el apartado de anterior donde se realizaron pruebas para evaluar la técnica basado en grafos para el desarrollo del sistema de recomendación se fue introduciendo determinados productos y se intentó recomendar los productos en base a las ponderaciones dadas de otros productos, adicional se realizó un análisis de la red para determinando las métricas y propiedades de la misma.

En la tabla 2, se puede observar las métricas que al ser una red bastante grande se tienen valores como el diámetro de la red que es un valor alto debido a la cantidad de nodos y conexiones. En la longitud de ruta promedio de la red no se tiene valores pequeños debido al tamaño de la red y su complejidad. Para la transitividad de la red se tiene un valor bajo lo que nos indica que la mayoría de los productos no implican la conexión de todos los nodos entre sí.

En la figura 23 se observa que el producto B007Q45EF4 tiene un alto grado de vecino promedio, significa que está fuertemente conectado con otros elementos en el grafo, lo que puede influir en la generación de recomendaciones basadas en conexiones directas. Realizando la búsqueda de los productos electrónicos más populares en Amazon, este ID de producto puede ser Fire TV stick

En la figura 24 se puede observar la importancia del producto B002WE6D44 en el grafo de la red de recomendación. Al realizar la búsqueda de los productos electrónicos de Amazon con el ID, se indica que es una tarjeta de memoria. En este sistema de recomendación este producto con un alto grado de centralidad podría ser recomendado como un nodo influyente o popular.

En la figura 25 y 26 se puede observar que el producto B007WTAJTO tanto en la centralidad de intermediación y la cercanía de centralidad tiene un alto valor y podría actuar como un "conector" entre grupos de productos y podría ser recomendado a otros usuarios que estén cerca de él en términos de conexiones directas.

## PROPUESTA DE SOLUCIÓN

### Implicaciones Organizacionales

Este proyecto busca plantear una solución diferente y eficiente a los sistemas de recomendación comunes que considerando la llegada del comercio electrónico y la compra de productos, contenidos o servicios mediante internet es esencial que las páginas web desarrolladas dispongan de un sistema que les permita comprender los intereses y elecciones de sus clientes o usuarios y se pueda aprovechar plenamente el análisis de datos al tomar decisiones:

- **Mayor satisfacción del cliente:** Un sistema de recomendación basado en grafos puede mejorar la experiencia del usuario al proporcionar recomendaciones precisas y relevantes, lo que puede aumentar la satisfacción del cliente. Esto puede tener un impacto positivo en la retención de clientes y la lealtad a la marca.
- **Mayor participación del usuario:** Al proporcionar recomendaciones personalizadas, los usuarios son más propensos a interactuar con la plataforma, lo que puede llevar a una mayor participación y, en última instancia, a un mayor compromiso con los productos o servicios ofrecidos por la organización.
- **Aumento de las ventas y los ingresos:** Los sistemas de recomendación basados en grafos pueden ayudar a impulsar las ventas al sugerir productos o servicios adicionales que los clientes pueden estar interesados en comprar. Esto puede conducir a un aumento de los ingresos y una mayor rentabilidad.

Los sistemas de recomendación basado en grafos serían de gran ayuda para muchos negocios y sobre todo para los supermercados en el país ya que se genera varios beneficios:

- **Diversidad de productos:** Los supermercados suelen ofrecer una amplia gama de productos, desde alimentos frescos hasta productos de cuidado personal. La implementación de un sistema de recomendación basado en grafos requerirá una modelización precisa de la estructura de los productos y la interconexión entre ellos. Esto puede ser un desafío organizacional, ya que es necesario garantizar que todos los productos estén representados de manera adecuada en el sistema de recomendación.
- **Logística y abastecimiento:** La implementación de un sistema de recomendación puede influir en la gestión de la cadena de suministro y el abastecimiento de productos. Los supermercados deben asegurarse de que los productos recomendados estén disponibles en el inventario en cantidades suficientes y en la ubicación adecuada en la tienda.
- **Experiencia del cliente en la tienda:** Los supermercados pueden utilizar sistemas de recomendación para mejorar la experiencia del cliente en la tienda física. Esto podría incluir la incorporación de señales de ubicación y la navegación en la tienda para ayudar a los clientes a encontrar los productos recomendados de manera eficiente.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

Con el trabajo realizado se observó que los grafos son una de las herramientas que se pueden emplear para implementar sistemas de recomendación, sin embargo, esta técnica requiere de un mantenimiento constante para asegurarse de que los datos y las relaciones en el grafo estén actualizados y sean relevantes.

Se debe tomar en cuenta que por temas académicos, el estudio y análisis se realizó con las valoraciones (ratings) mayores a 4 debido a la gran cantidad de datos que se tiene en la base de datos que se obtuvo por parte de Amazon. Y sobre todo se realizó con las ponderaciones entre 20 y 100 que permiten realizar análisis más sofisticados y precisos, como encontrar la ruta más corta y las métricas que se presentó en el apartado de los resultados. Y así destacar varios beneficios con la técnica utilizada:

- Los sistemas de recomendación basados en grafos pueden ser eficientes para generar recomendaciones personalizadas, ya que utilizan relaciones y conexiones entre elementos para predecir preferencias del usuario. Esto puede llevar a recomendaciones más precisas y relevantes en comparación con métodos más simples.
- Es importante mencionar que la técnica basada en grafos nos permite tener una mayor escalabilidad para los sistemas de recomendación sobre todo porque se manejó con grandes cantidades de datos y usuarios, por lo tanto la optimización y el diseño eficiente de algoritmos son esenciales para garantizar que el sistema pueda manejar grandes volúmenes de información.

## Recomendaciones

Para llevar a cabo el uso eficiente del análisis de datos y transformar el método tradicional que la mayoría de las empresas hoy en día utilizan, se pueden tomar varias medidas como se recomienda a continuación:

- Recopilar los datos detallados de productos y compras de los clientes que incluya información sobre productos, precios, categorías, ubicaciones en la tienda, fechas y preferencias del cliente. Cuantos más datos tenga, mejor se podrá mejorar el sistema de recomendación
- Utilizar un enfoque de grafo para modelar la relación entre productos. Esto implica la creación de nodos para cada producto y la definición de relaciones entre ellos, como productos relacionados, productos que se compran juntos con frecuencia y categorías similares. Esta estructura de grafo servirá como base para las recomendaciones.
- El sistema de recomendación debe estar integrado también con el sistema de gestión de inventario para garantizar que el sistema pueda tener en cuenta la disponibilidad de productos y así poder evitar recomendaciones de productos agotados.
- Al igual como se trabajó en este proyecto, se debe retroalimentar continuamente la información donde se incluya la clasificación y calificación de productos recomendados y adicional se puede trabajar con los comentarios y sugerencias de los clientes para poder mejorar el sistema de recomendación ya que cuanto más se tenga información de las preferencias individuales de los clientes, mejor podrá personalizar las recomendaciones. Se puede utilizar datos de compras pasadas, historiales de navegación en línea y otras fuentes de información
- Evaluar regularmente el rendimiento del sistema de recomendación y realice ajustes según sea necesario. Los patrones de compra y las preferencias de los clientes pueden cambiar con el tiempo, por lo que es importante mantener el sistema actualizado.

## REFERENCIAS

- [1] GraphEverywhere, E. (2019, December 2). Sistemas de recomendación | Que son, tipos Y ejemplos. GraphEverywhere. <https://www.grapheverywhere.com/sistemas-de-recomendacion-que-son-tipos-y-ejemplos>
- [2] Galán, S. (2007). Filtrado Colaborativo y Sistemas de Recomendación. 8. Madrid, España: Universidad Carlos III. Obtenido de <https://www.it.uc3m.es/jvillena/irc/practicas/06-07/31.pdf>
- [3] Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J., Miller, B., & Riedl, J. (1998). Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. Proceedings of the 1998 ACM conference on Computer supported cooperative work. <https://doi.org/10.1145/289444.289509>
- [4] M. L. Mamani, "DFS vs BFS. Ancora," 25 Mayo 2020. [Online]. Available: <https://www.encora.com/es/blog/dfs-vs-bfs>. [Accessed 31 Mayo 2023].
- [5] M. N. & A. E. Hodler, A Comprehensive Guide to Graph Algorithms un Neo4j, neo4j, 2021.
- [6] S. G. Prado, "Algoritmos para Big Data: Grafos y PageRank," Universidad de Valladolid, Valladolid, 2017.
- [7] S. Cruces, Graph and Network Analytics, Universidad de Sevilla, 2022.
- [7] I. Popescu, K. Portello, C. Anagnostopoulos y N. Ntarmos, "The case for graph-based recommendations", Diciembre 2017. IEEE International Conference on Big Data (Big Data), vol.4, no. 3
- [8] Wang, X., Zhang, Y., & Yamasaki, T. (2020). Earn more social attention: User popularity based tag recommendation system [Paper presentation]. <https://doi.org/10.1145/3366424.3383543>

[9] Tkalcic, M., Odic, A., Kosir, A., & Tasic, J. (2013). Affective labeling in a content-based recommender system for images. *IEEE Transactions on Multimedia*, 15(2), 391-400. <https://doi.org/10.1109/tmm.2012.2229970>

Collado Sánchez, A. (2016). *Sistema de recomendación de recursos basado en filtrado colaborativo para la plataforma edX* (Bachelor's thesis).

[10] E. Herrera-Viedma, L. Olvera, E. Peis, C. Porcel. "Revisión de los sistemas de recomendaciones para la recuperación de información. Tendencias de investigación en organización del conocimiento (Trends in knowledge organization research)" José Antonio Frías, Ed. Crispulo Travieso, Universidad de Salamanca, (2003), 507-513.

[11] Así funcionan las recomendaciones de Amazon. (2014, January 31). Genbeta - Software, descargas, aplicaciones web y móvil, desarrollo. <https://www.genbeta.com/web/asi-funcionan-las-recomendaciones-de-amazon>

[12] Valecillos Girand, O. A. (2019). Desarrollo de un sistema de recomendaciones para un sitio de comercio electrónico (Doctoral dissertation).

## ANEXOS

Se comparte la URL de Github con los códigos utilizados en este proyecto

- [https://github.com/Belen1112/Proyecto\\_capstone\\_Maestr-a\\_Inteligencia\\_Negocios\\_Ciencia\\_Datos.git](https://github.com/Belen1112/Proyecto_capstone_Maestr-a_Inteligencia_Negocios_Ciencia_Datos.git)