



Proyecto Capstone

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

Autor: Ma. Belén Tipantuña

Índice

1. Introducción >
2. Revisión literaria >
3. Descripción del problema >
4. Objetivos del estudio >
5. Metodología utilizada >
6. Resultados >
7. Conclusiones >



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.



Revisión literaria - Métodos

Basado en contenido	Basado en filtrado colaborativo	Basado en grafos	Híbrido
<p>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</p>	<p>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</p>	<p>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</p>	<p>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</p>

Revisión literaria

Tipos de grafos según sus propiedades



Grafos dirigidos

Las conexiones son asimétricas, lo que significa que la secuencia de los nodos define la dirección de la arista correspondiente



Grafos no dirigidos

Las conexiones entre nodos carecen de dirección específica



Grafo cíclico

Es cuando incluye al menos un ciclo, que es una secuencia de vértices que comienza y termina en el mismo vértice



Grafo acíclico

Si un grafo no contiene ciclos, se denomina acíclico



Grafos ponderado

Es cuando se asignan pesos a los nodos o a sus conexiones. Se prioriza analizar la ruta más corta según el costo total en lugar del número de saltos.



Grafos no ponderado

Las conexiones carecen de peso



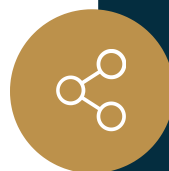
Revisión literaria

Tipos de grafos según su estructura



Grafo aleatorio

La característica fundamental de los grafos aleatorios es cuando la existencia o ausencia de una arista se decide de una manera totalmente independiente de las otras aristas. +



Grafo completo

Es un grafo completamente conectado en donde todos los pares de nodos están conectados por una arista, es decir, cada nodo presenta una conexión directa con todos los otros nodos. +



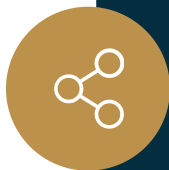
Grafo clique

Se considera como un subgrafo completo de un grafo más grande. Esto implica que si seleccionamos un subconjunto de nodos de un grafo completo y los conectamos entre sí, generamos un grafo clique. +



Grafo 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. +



Grafo bipartito

En este tipo de grafo los nodos se separan en dos grupos claramente definidos, en donde todas las aristas conectan nodos de un grupo con nodos del otro grupo, es decir, no hay conexiones de nodos dentro del mismo conjunto. +



Descripción 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.

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





General

El objetivo principal en este proyecto es desarrollar un Sistema de recomendación utilizando la técnica basada en grafos 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

Metodología utilizada

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. 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

2

Identificación de variables

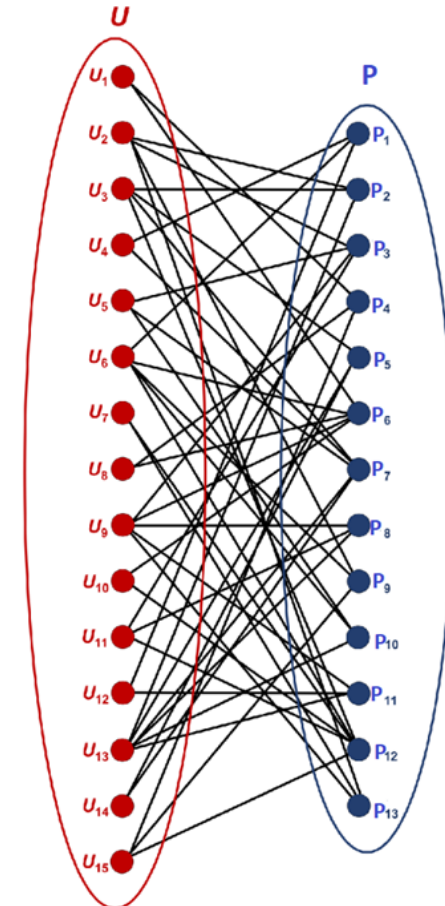
Este conjunto de datos tiene más de 7 millones de calificaciones de productos relacionados con la electrónica vendidos en su sitio web. En dichos registros constan la identificación única del usuario y del producto y la calificación del 1 al 5 dada por el usuario a un producto (Rating) donde 1 es una baja calificación y 5 la mejor calificación del producto. También se tiene la marca de tiempo (Unix) de la calificación dada.

Metodología utilizada

3

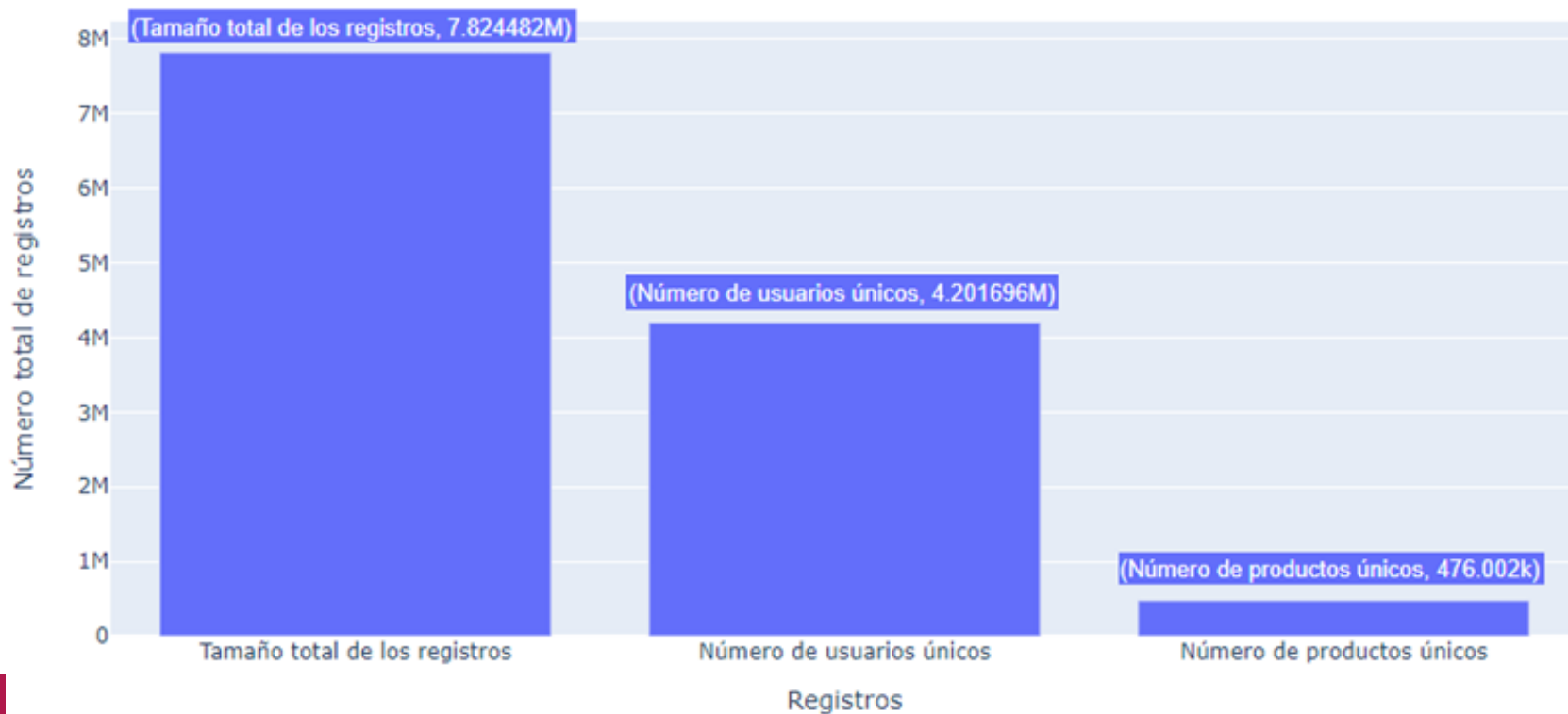
Selección de modelo estadístico

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. 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





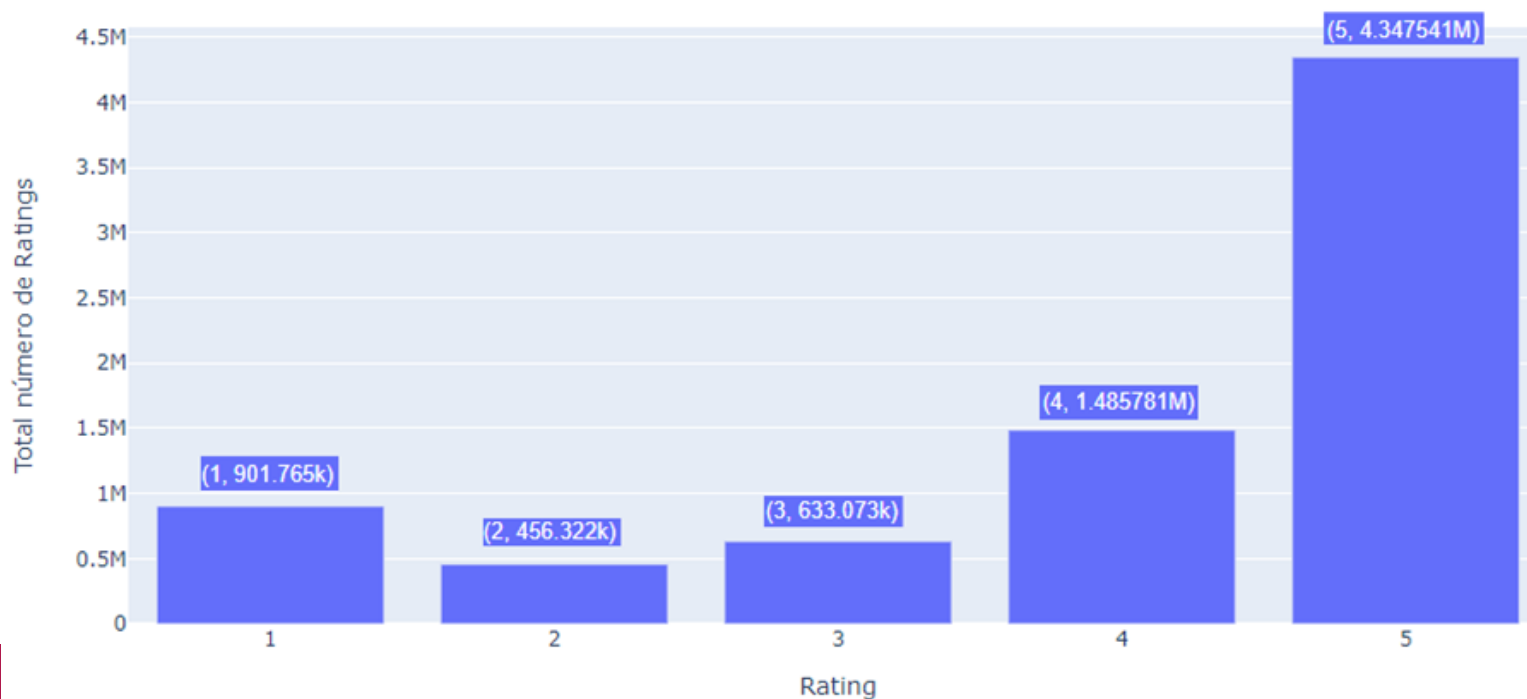
Visualización de variables



Número de usuarios y productos únicos



Visualización de variables



Calificaciones dadas por los usuarios



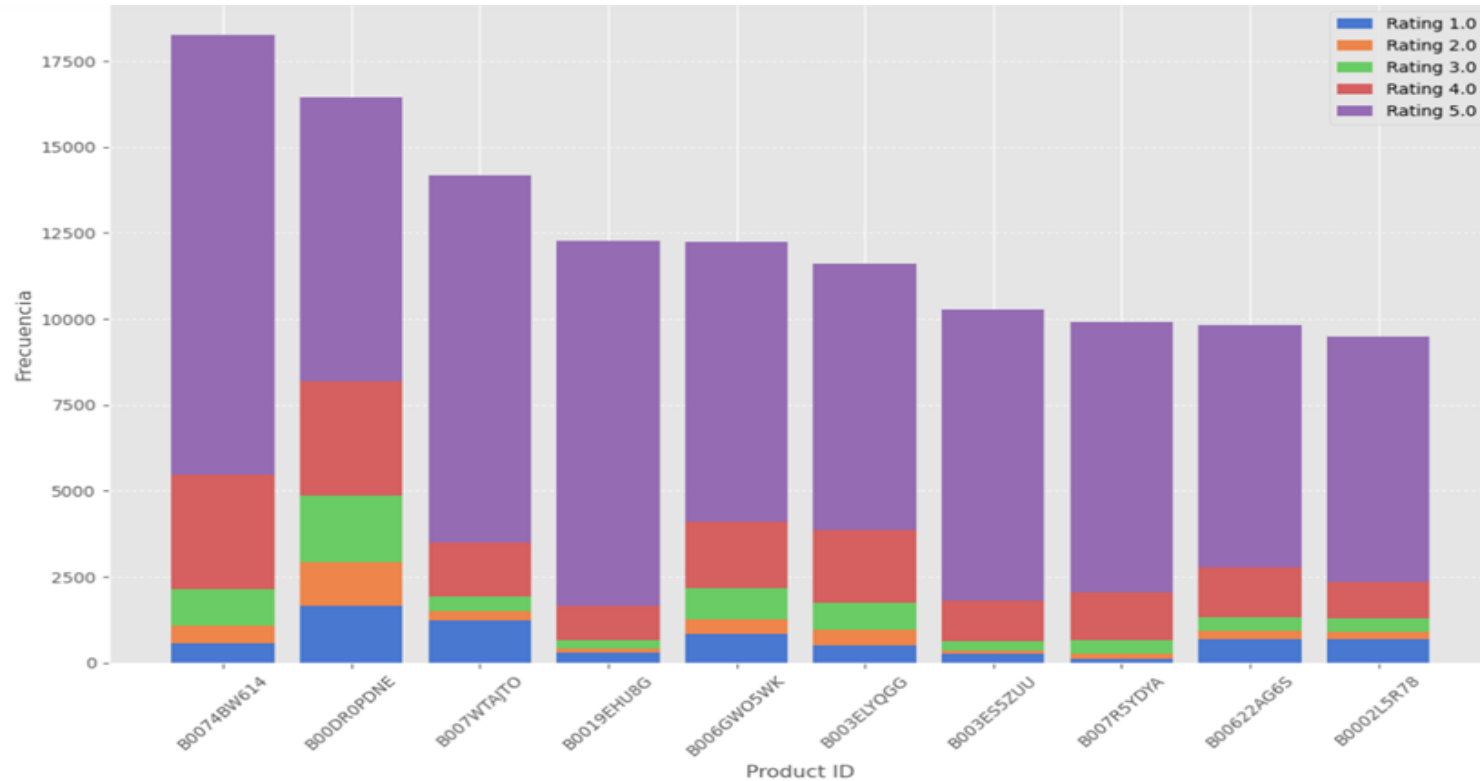
Visualización de variables



Top 10 de los productos más populares



Visualización de variables



Frecuencia de Ratings para los top 10 de productos



Resultados

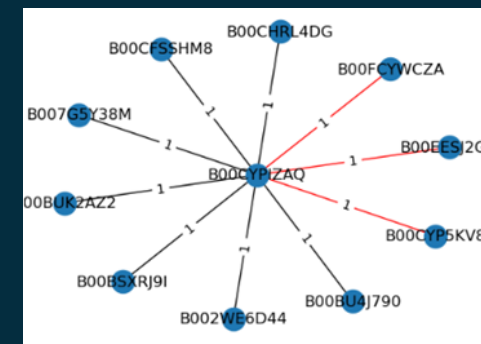
B007CVTQJG

Nº de
recomendaciones:
2



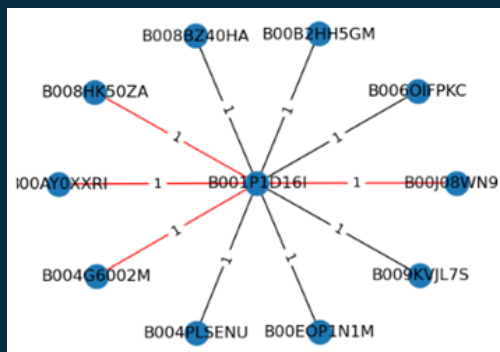
B007CVTQJG

Nº de
recomendaciones:
3



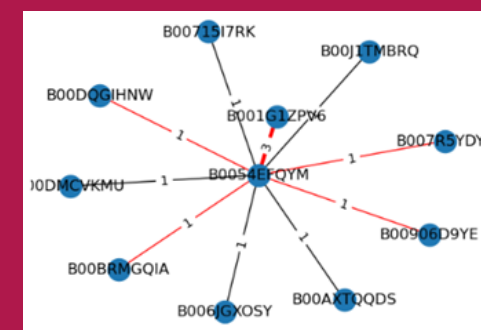
B007CVTQJG

Nº de
recomendaciones:
4



B007CVTQJG

Nº de
recomendaciones:
5





Propiedades de la red

Resultados



Resultado: 11

Diametro de la red

+ info



Resultado: 3.6721

Longitud de ruta
promedio de la red

+ info



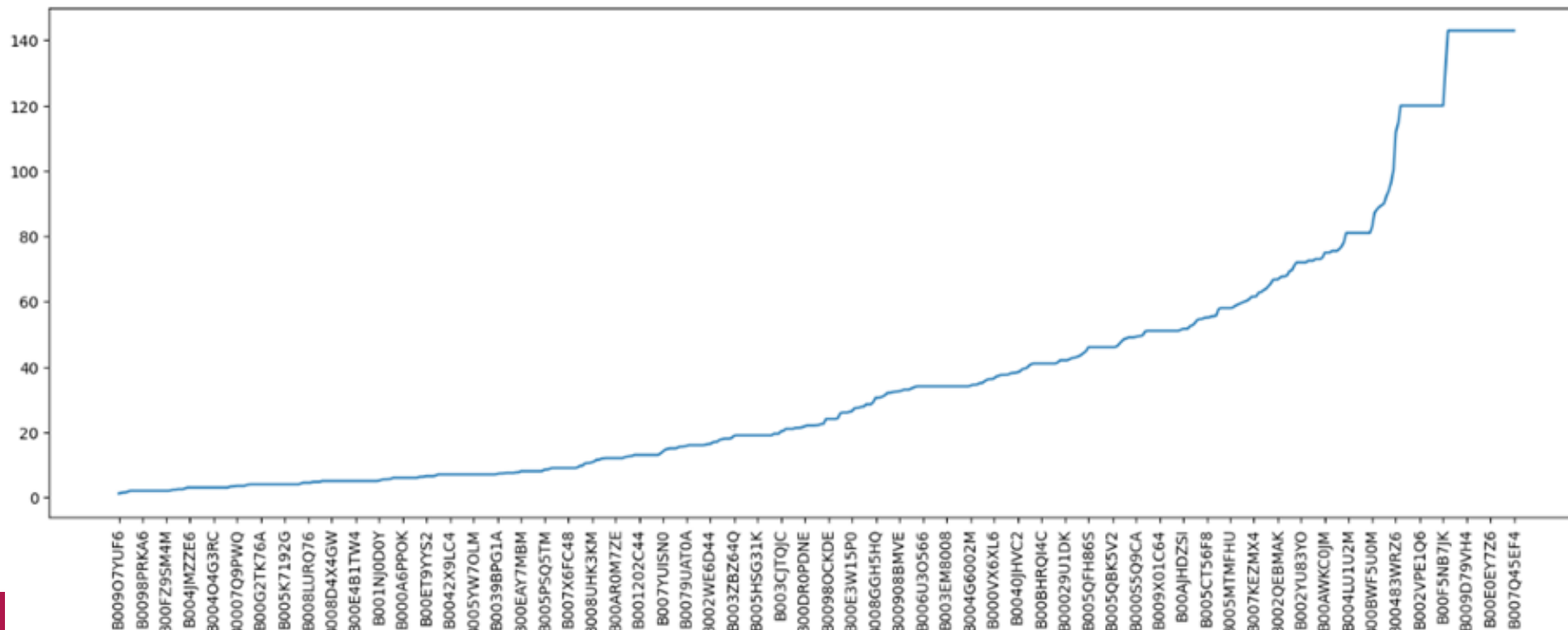
Resultado: 0.12

Transitividad de la
red

+ info



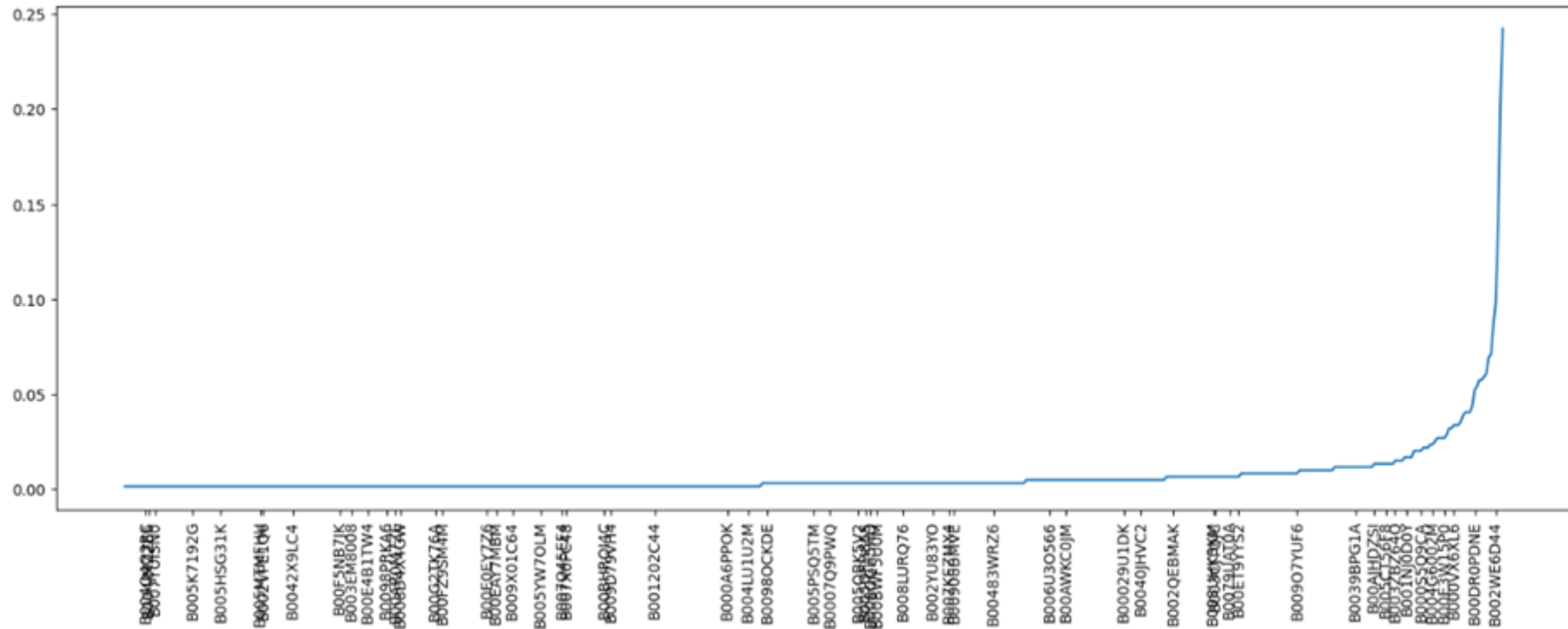
Resultados



Grado de vecino promedio



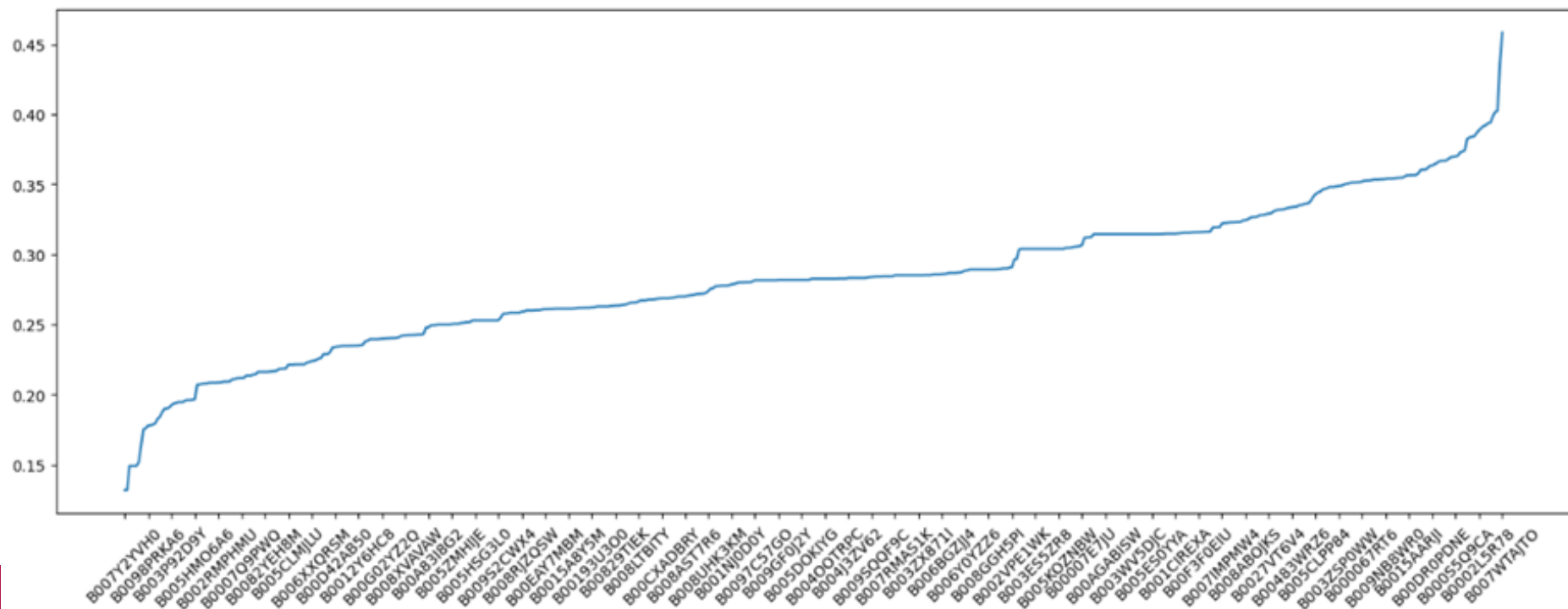
Resultados



Centralidad del grafo

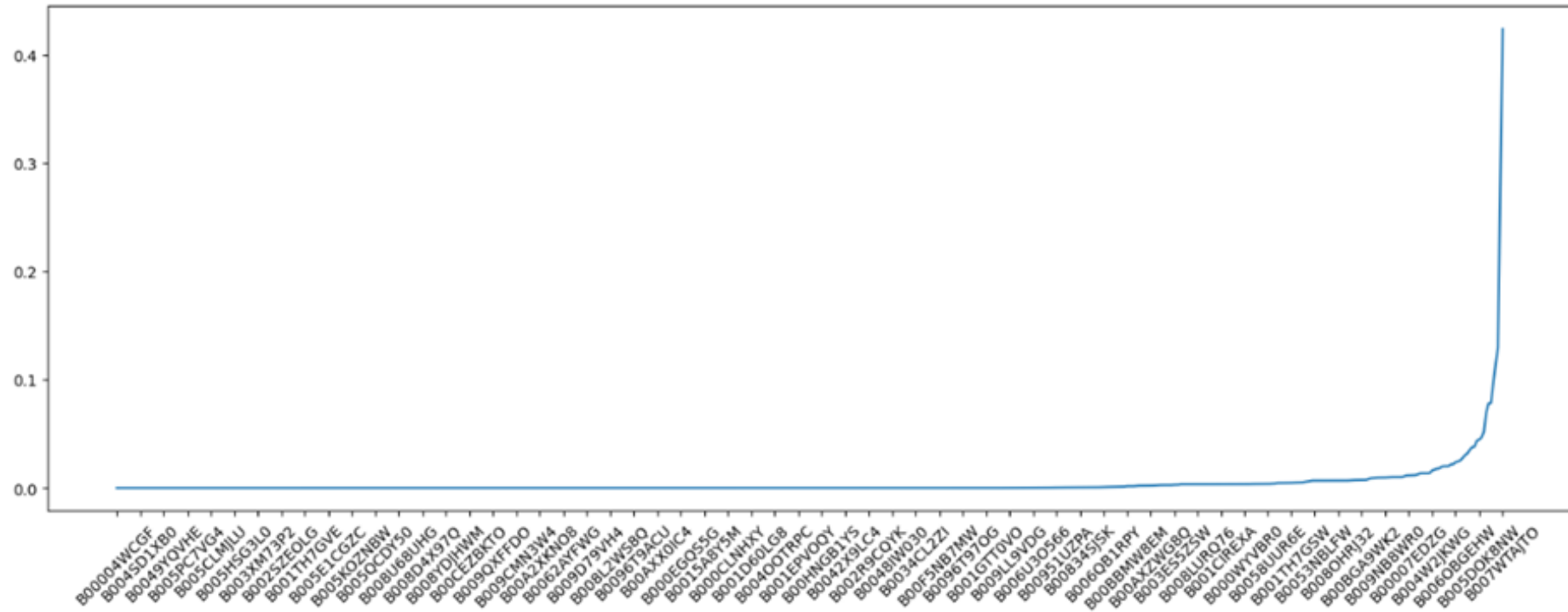


Resultados



Cercanía de centralidad

Resultados



Centralidad de intermediación



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.

