

# Análisis de Redes Bipartitas en Series de Anime

Alejandro Urrego Lopéz  
Cesar Prieto Sarmiento

Departamento de Estadística  
Facultad de Ciencias  
Universidad Nacional de Colombia.



## 1. Introducción

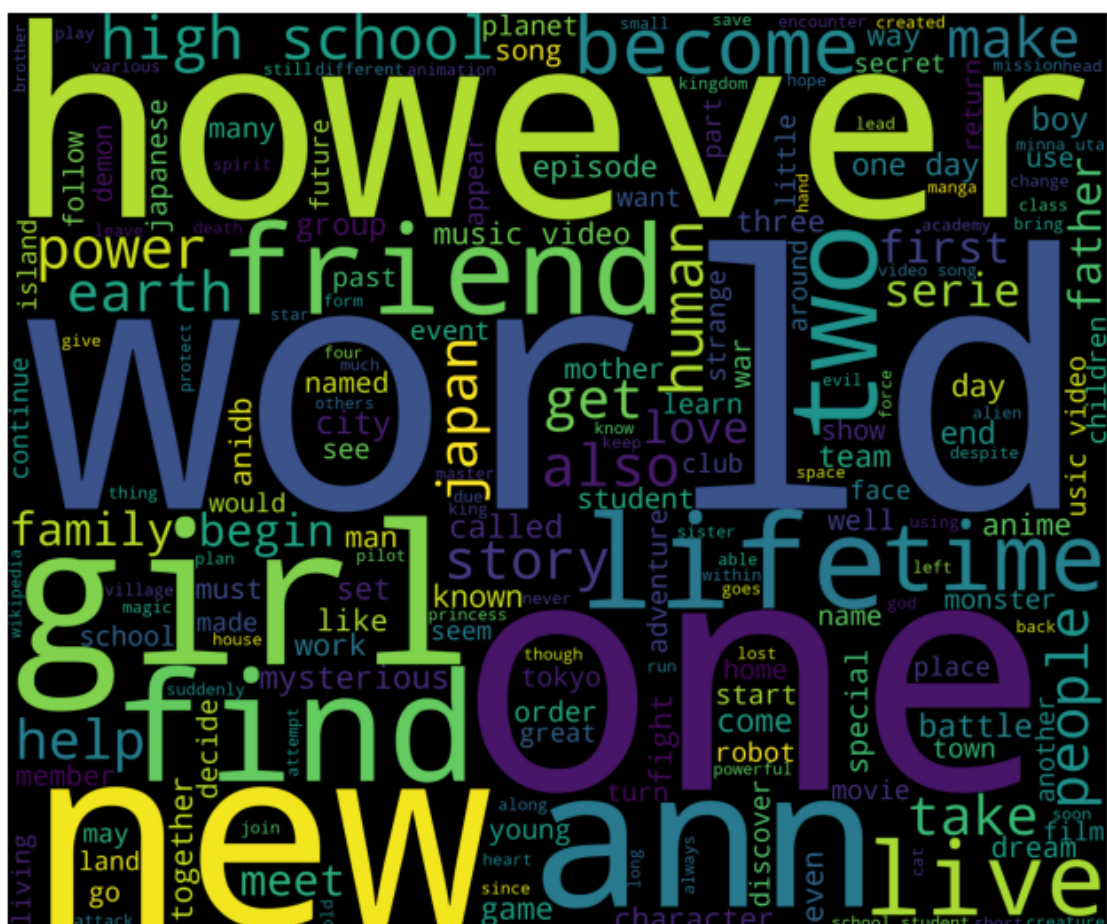
El análisis de **redes bipartitas** es útil para estudiar las **relaciones** entre **usuarios** y **productos**, permitiendo explorar estas interacciones mediante **modelos estadísticos** utilizando toda la información nodal disponible. Las **descripciones** de productos, analizadas con técnicas de procesamiento de texto como el análisis de **bigramas**, son cruciales para identificar características y patrones en dichas descripciones. Estas descripciones pueden impactar significativamente las **redes usuario-producto**. Para evaluar esta influencia, se puede analizar la red **producto-producto** generada a partir de la proyección de la matriz de adyacencia y aplicar modelos de grafos aleatorios exponenciales (**ERGM**).

## 2. Análisis

### 2.1 Descripción de los datos

Este estudio se basa en un conjunto de datos de **MyAnimeList** que contiene información sobre 17,562 animes y las preferencias de 325,772 usuarios. Debido al alto **costo computacional** de analizar una red bipartita completa con todos los **usuarios y animes**, se decidió realizar un muestreo en la base de datos de animes.

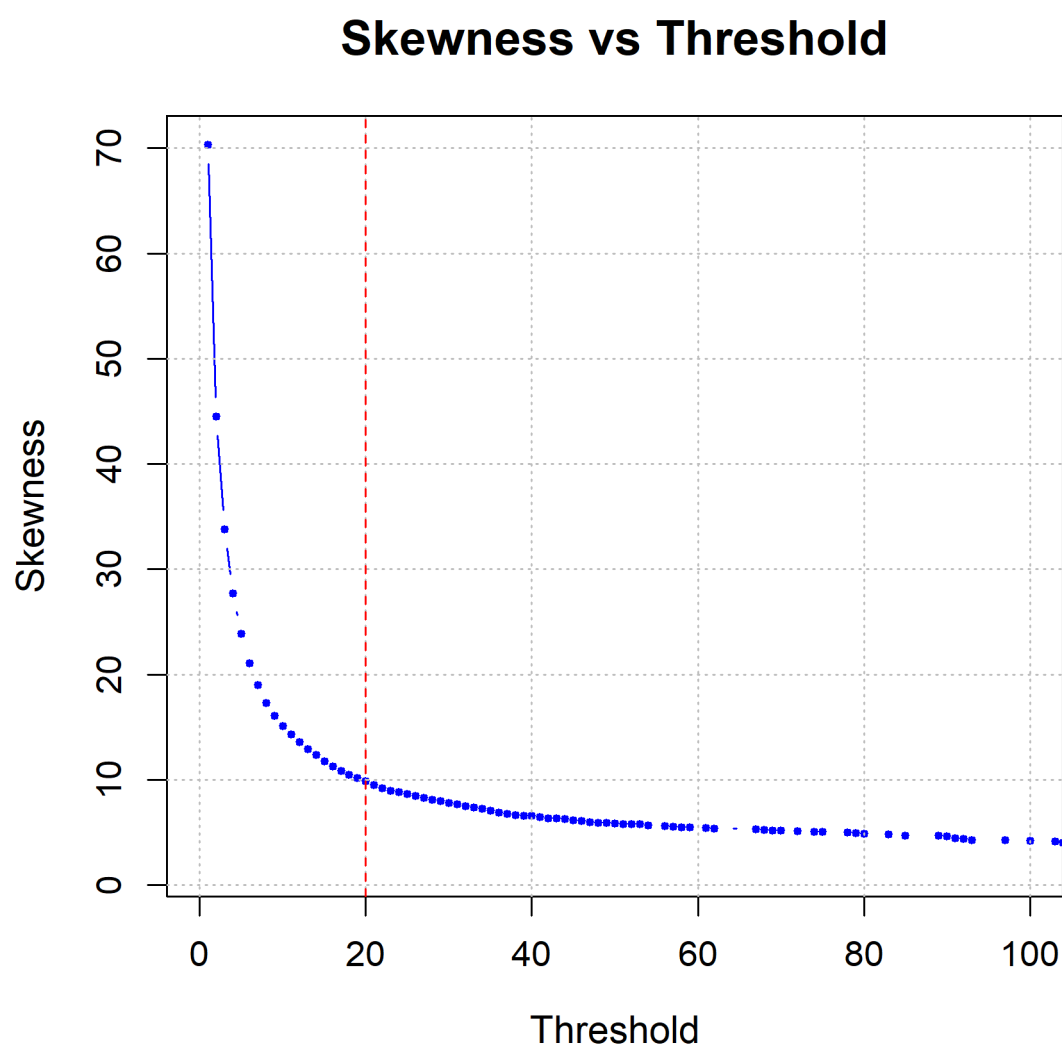
### 2.2 Nube de Palabras



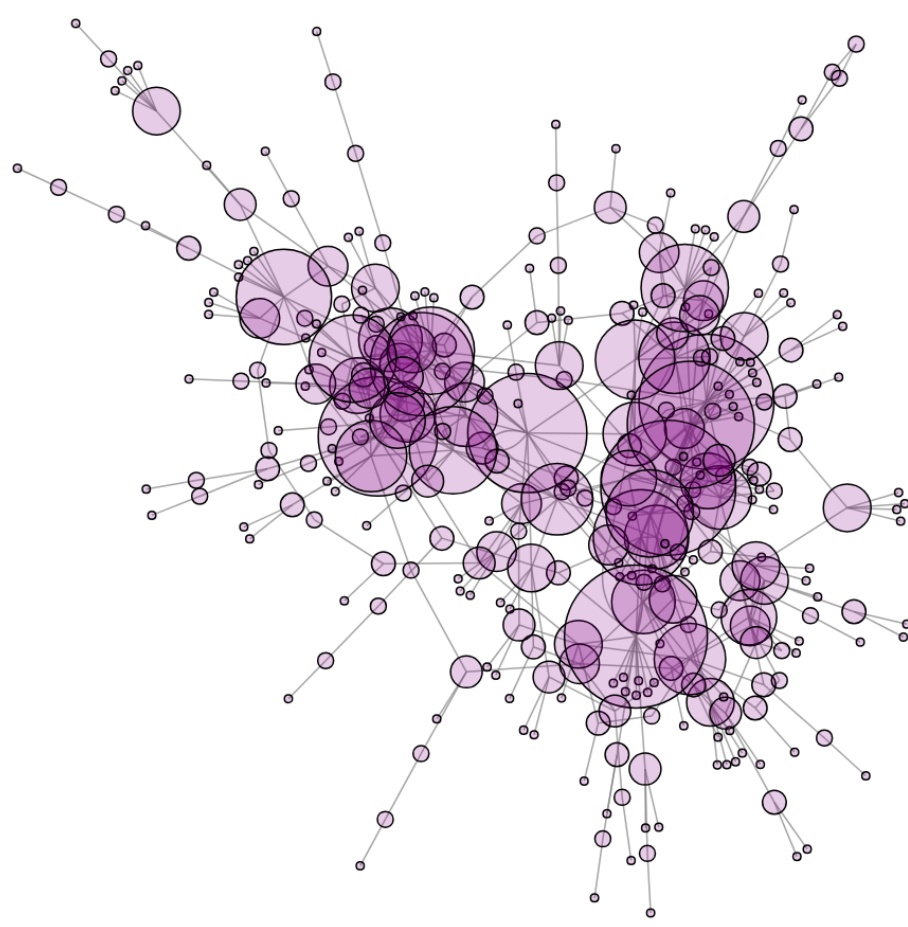
La **nube de palabras** nos ofrece una visión clara de las **palabras** más frecuentes en las descripciones, facilitando la rápida **identificación de temas** y elementos predominantes en los animes. Las palabras más prominentes revelan una mezcla de **contextos**: Escolares, relaciones personales, mundos ficticios, y elementos de aventura y fantasía.

### 2.3 Bigramas y selección del umbral

Para realizar los **bigramas**, se tomaron dos palabras consecutivas. Se eliminaron aquellos que contenían el carácter especial que separaba las descripciones. Se utilizó un umbral que **minimiza el sesgo** y que conservara la mayor cantidad de bigramas posibles, con el objetivo de mejorar la **capacidad computacional** y poder generalizar más los grupos formados más adelante para todas las descripciones. Para construir la red, se seleccionaron los bigramas con una frecuencia superior o igual a un **umbral de 20**.



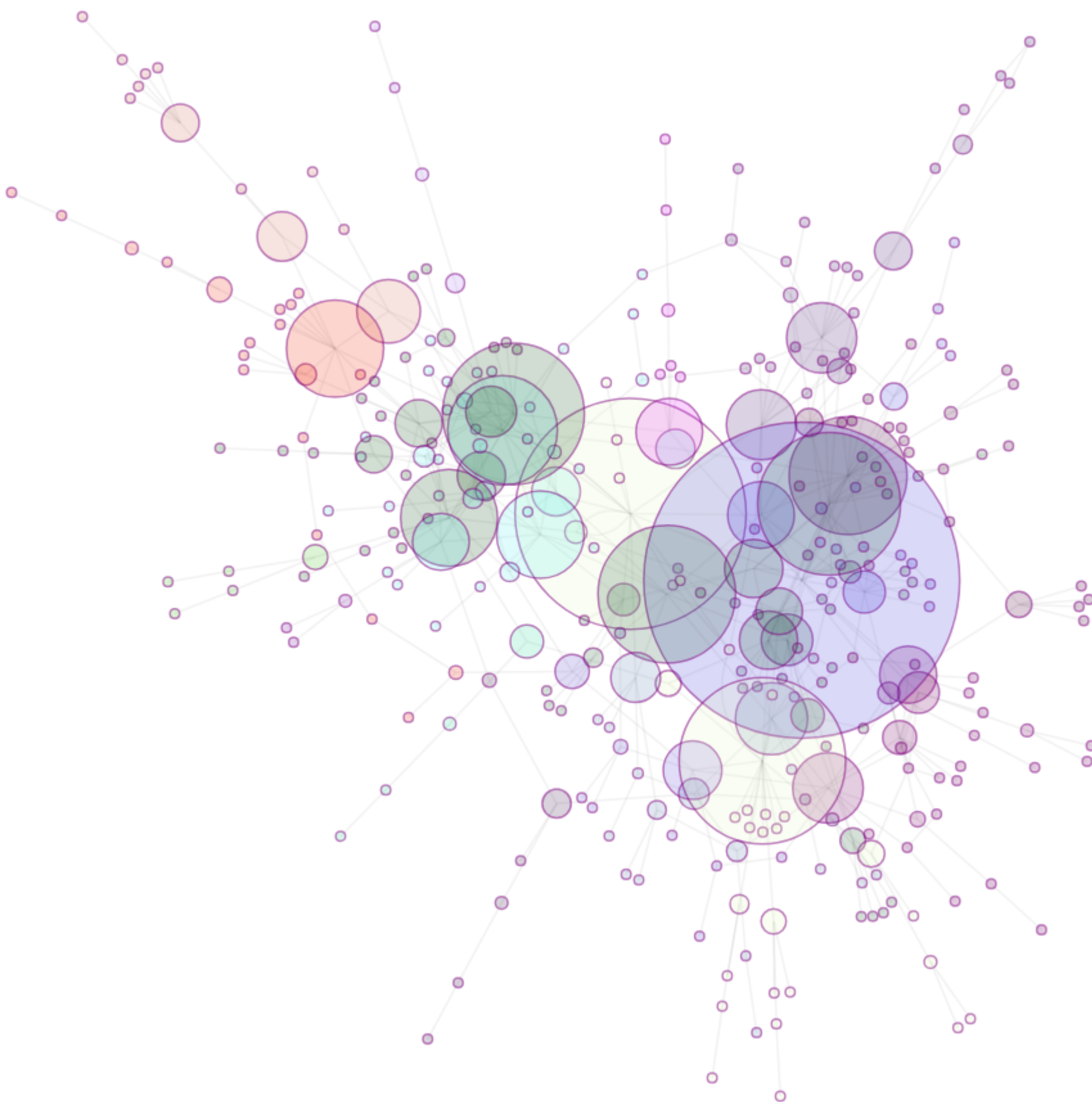
### 2.4 Análisis de la red de Bigramas



Métrica	Valor
Distancia media	5.295
Grado medio	2.326
Grado desviación	0.043
Número clan	3
Densidad	0.004
Transitividad	0.048
Asortatividad	0.043

La **componente conexa** de la red de bigramas revela una estructura compleja con varios **nodos clave** que actúan como puntos centrales en la red. El **grado** y la presencia de clanes indican que **existen temas recurrentes** en las descripciones de los animes.

Por otra parte, para hacer la partición de la red y encontrar sus respectivos **grupos de palabras**, se emplearon varios **algoritmos de agrupación jerárquica** implementados en igraph, y se escogió el que tuviera una mayor **modularidad**, esta fue de 0.74.

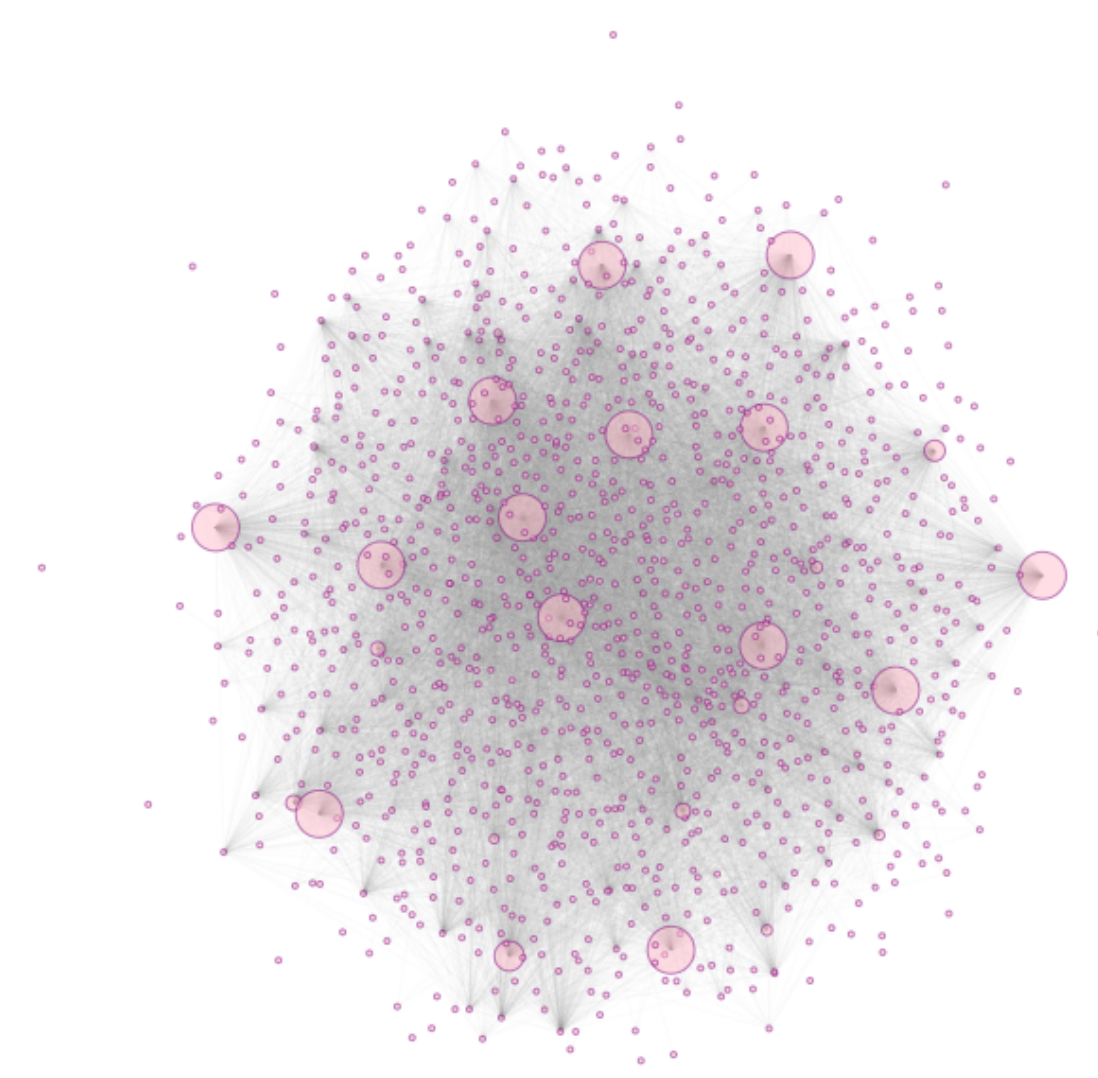


Adicionalmente, con ayuda de **inteligencia artificial**, se analizan las palabras por grupo y se intenta dar un significado a este, obteniendo los siguientes resultados:

Cluster	Tópico Principal
Cluster 0	Lanzamientos de episodios y medios
Cluster 1	Relaciones y experiencias personales
Cluster 2	Música y medios digitales
Cluster 3	Temporalidad y distancias
Cluster 4	Personajes y relaciones familiares
Cluster 5	Videos promocionales y juegos
Cluster 6	Narrativas épicas y mundos fantásticos
Cluster 7	Viajes y misiones
Cluster 8	Vida escolar y cotidiana
Cluster 9	Adaptaciones y formatos de medios
Cluster 10	Amistad y relaciones escolares
Cluster 11	Diseño y desarrollo de personajes
Cluster 12	Participación y decisiones
Cluster 13	Información pendiente
Cluster 14	Contextos históricos y de ciencia ficción
Cluster 15	Técnicas de animación y estudios
Cluster 16	Libros infantiles y dramas
Cluster 17	Bandas y música tradicional

### 2.5 Red Bipartita

La **red bipartita** tiene 8,710,328 enlaces que indican si un usuario ha visto alguno de los animes del muestreo. Luego, se proyecta la red premultiplicando la **matriz de adyacencia** A por su transpuesta, obteniendo una matriz de 1264 x 1264 donde cada entrada representa el número de usuarios que han visto ambos animes. Para aplicar el **ERGM**, la red se binarizó, convirtiendo en uno cada entrada de la matriz de proyección si el número de usuarios que han visto ambos animes es **mayor o igual al 75%** del total de usuarios que vieron el anime, garantizando que varios usuarios han visto ambos animes.



Name	Closeness	Betweenness	Eigenvector
Mahou Shoujo Madoka Magica	1.00	-	1.00
Bishoujo Senshi Sailor Moon	1.00	-	1.00
Charlotte	1.00	-	1.00
Shinmai Maou no Testament	1.00	69.58	-
Boku no Hero Academia 4th Season	-	69.58	-
Digimon Tamers	-	69.58	-
Pokemon	-	69.58	-
Steins;Gate	0.85	24,225.40	1.00
Natsume Yuujinchou San	0.80	15,552.56	0.98
Zombie-Loan	0.79	14,127.91	0.97
Trigun	0.72	-	-
One Piece: Episode of Nami	-	6,830.04	-
Neon Genesis Evangelion	-	-	0.95

Name	Closeness	Betweenness	Eigenvector
Little Village People	0.99	402,989.42	-
Baolie Feiche II: Xing Neng Juexing	0.81	141,314.74	-
Baolie Feiche 3: Shou Shen Heti	0.81	141,314.74	-
Kouya no Kotobuki Hikoutai Kanzenban	0.59	16,990.44	-
Closers: Side Blacklamb	0.56	19,071.02	-
Death Note	-	-	1.00
Sen to Chihiro no Kamikakushi	-	-	0.78
Elfen Lied	-	-	0.78
Kimi no Na wa.	-	-	0.77
Kiseijuu: Sei no Kakuritsu	-	-	0.70

La comparación entre la **red binarizada** como la **no binarizada** muestra cómo las métricas de **centralidad** pueden variar significativamente dependiendo de la consideración del peso de las conexiones

### 2.6 Aplicación del modelo ERGM

Cluster	Estimado	Error Estándar	p-valor
edges	-1.608	0.007	$1 \times 10^{-4}$
cluster 0	-0.008	0.001	$1 \times 10^{-4}$
cluster 1	-0.010	0.001	$1 \times 10^{-4}$
cluster 2	0.016	0.003	$1 \times 10^{-4}$
cluster 3	-0.014	0.002	$1 \times 10^{-4}$
cluster 4	-0.012	0.001	$1 \times 10^{-4}$
cluster 5	-0.010	0.001	$1 \times 10^{-4}$
cluster 6	0.002	0.001	0.005
cluster 7	-0.019	0.001	$1 \times 10^{-4}$
cluster 8	0.011	0.001	$1 \times 10^{-4}$
cluster 9	0.007	0.001	$1 \times 10^{-4}$
cluster 10	0.005	0.001	0.001
cluster 11	0.018	0.001	$1 \times 10^{-4}$
cluster 12	0.024	0.002	$1 \times 10^{-4}$
cluster 13	0.019	0.002	$1 \times 10^{-4}$
cluster 14	-0.133	0.004	$1 \times 10^{-4}$
cluster 15	0.091	0.004	$1 \times 10^{-4}$
cluster 16	0.015	0.002	$1 \times 10^{-4}$
cluster 17	-0.025	0.002	$1 \times 10^{-4}$

Al ajustar un modelo de grafos aleatorios exponenciales (**ERGM**) a la red analizada, se obtuvieron **resultados significativos** en cuanto a la **influencia** de ciertos grupos de palabras (clusters) en la **formación de enlaces**. Los parámetros estimados y sus respectivas pruebas de **significancia estadística** proporcionan una visión clara sobre qué clusters afectan positivamente o negativamente la creación de conexiones dentro de la red.

## 3. Resultados y discusión

El presente estudio demuestra que el **análisis de redes bipartitas**, en combinación con técnicas de **procesamiento de lenguaje natural**, es una herramienta eficaz para comprender la **influencia de las descripciones textuales** en la formación de comunidades de usuarios de anime. Estas técnicas tienen el potencial de ser implementadas en **sistemas de recomendación**, y **mejorar las descripciones** de este tipo de producto en particular. Todo esto se puede extender a diferentes plataformas de **streaming** conocidas. Basado en los hallazgos del estudio, se sugieren las siguientes recomendaciones:

- Implementar en un Sistemas de **Recomendación**
- **Explorar** otros tipos de entretenimiento
- Problemas en el establecimiento de las **temáticas de los Clusters**