

PRUEBA DE EVALUACIÓN CONTINUA 1 - PEC1

En busca de usuarios semejantes: homofilia

Ana Solbas Casajús
asolbas2@alumno.uned.es

26 de mayo de 2025



Minería de Datos de los Medios Sociales
Curso 2024-2025

Máster en Ingeniería y Ciencia de Datos
Universidad Nacional de Educación a Distancia (UNED)

Índice

| | |
|---|----------|
| 1. Introducción | 1 |
| 2. Objetivos | 2 |
| 3. Materiales y métodos | 2 |
| 3.1. Selección de pódcasts | 2 |
| 3.2. Extracción de datos | 4 |
| 3.3. Análisis estático y dinámico | 4 |
| 3.3.1. Preprocesamiento de la red | 5 |
| 3.3.2. Visualización de la red | 5 |
| 3.3.3. Análisis estático | 5 |
| 3.3.4. Análisis dinámico | 6 |
| 3.4. Modularidad | 6 |
| 3.5. Aplicación de técnicas de aprendizaje automático | 7 |
| 3.6. Estudio de la homofilia | 8 |
| 4. Resultados y discusión | 9 |
| 4.1. Análisis exploratorio de la red | 9 |
| 4.2. Análisis de la centralidad | 12 |
| 4.3. Análisis dinámico de la red | 15 |
| 4.4. Análisis de comunidades basadas en la optimización de la modularidad | 18 |
| 4.5. Análisis de comunidades basadas en métricas de centralidad | 19 |
| 4.6. Estudio de la homofilia | 22 |
| 4.6.1. Semejanza estructural entre los usuarios de Bluesky | 22 |
| 4.6.2. Semejanza estructural entre los miembros de una comunidad | 23 |
| 4.6.3. Semejanza estructural entre los miembros de un clúster | 26 |
| 4.6.4. Homofilia funcional entre los miembros de un clúster | 27 |
| 4.6.5. Análisis comparativo de la homofilia | 29 |

| | |
|--|-----------|
| 5. Conclusiones | 31 |
| 6. Referencias | 32 |
| 7. Anexo | 34 |
| 7.1. Material suplementario | 34 |
| 7.2. Métricas de centralidad | 34 |

1. Introducción

En la actual era digital, las redes sociales se han consolidado como espacios centrales para la difusión y el consumo de contenidos culturales, informativos y políticos (Bird, 2011; Curiel, 2011; Marcos-García et al., 2021). En concreto, las plataformas de *microblogging*, tradicionalmente dominadas por X (anteriormente Twitter) y, más recientemente por Bluesky, se han convertido en espacios de divulgación donde los contenidos publicados no solo se consumen, sino que también se comparten y comentan colectivamente. Estas dinámicas favorecen la creación de comunidades implícitas de usuarios que se agrupan en torno a intereses compartidos, valores comunes o identidades culturales afines (Ito et al., 2018; Marcos-García et al., 2021). Por este motivo, medios alternativos como los podcasts independientes, el periodismo digital o los medios de divulgación crítica, social y política encuentran en las redes sociales un canal privilegiado para conectar con la audiencia, al margen de los grandes medios tradicionales, los cuales no siempre favorecen la participación y colaboración entre usuarios (Broersma y Eldridge II, 2019; Chen y Keng, 2024; Wendland, 2024).

En este contexto, una de las principales fuerzas sociales que estructuran las interacciones digitales es la *homofilia*, es decir, la tendencia de los usuarios a interactuar con otros perfiles similares (Bisgin et al., 2012). En el ámbito de las redes sociales, esto se traduce en patrones de conexión entre usuarios que comparten atributos como la ideología política o los intereses culturales (Basov, 2020). Los usuarios no solo siguen contenidos por su calidad informativa o entretenimiento, sino también porque refuerzan identidades compartidas. Este fenómeno contribuye a la creación de cámaras de eco, espacios donde las opiniones, creencias e intereses se ven amplificados por la repetición constante de voces afines, mientras que se minimiza la exposición a puntos de vista divergentes. Aunque estas cámaras de eco pueden fomentar un sentimiento de pertenencia y cohesión, también pueden limitar la diversidad informativa y reforzar la polarización de la sociedad (Poongundranar, 2023).

El análisis de la homofilia en redes sociales, por tanto, puede ayudar a comprender cómo se configuran estos espacios y cuáles son las dinámicas que articulan la interacción entre usuarios. Debido al tamaño y la complejidad de las redes sociales, en este trabajo se ha optado por centrar el análisis en un pequeño conjunto de cuentas de podcasts culturales, sociales y políticos de habla hispana, ya que este formato audiovisual representa una forma de contenido sociocultural que combina la divulgación informativa con las narrativas personales, en un tono más informal y accesible que el empleado por medios tradicionales como la radio. Esta proximidad favorece una mayor implicación de la audiencia y puede contribuir a la creación de comunidades especialmente activas y comprometidas en plataformas como Bluesky (Wendland, 2024). En los últimos años, el formato podcast ha experimentado una notable expansión, con un aumento de producciones independientes y creadores amateurs impulsado por su bajo coste de producción, su capacidad de distribución global y su estrecha vinculación con las redes sociales. Aunque solo una minoría de estas producciones logra alcanzar audiencias masivas, numerosos creadores han logrado consolidarse en microcomunidades de usuarios a través de la publicación de contenidos en nichos muy específicos (McGregor, 2022). Estos factores convierten a los podcasts en candidatos idóneos para estudiar el fenómeno de la homofilia en redes sociales.

2. Objetivos

El objetivo general del trabajo es estudiar el fenómeno de la homofilia en una red formada por los usuarios de Bluesky que han interactuado con cinco cuentas de pódcast que difieren en tamaño, alcance y temática, con el fin de explorar si estas diferencias pueden influir en la configuración de comunidades y afinidades dentro de la red y, al mismo tiempo, ofrecer una representación del ecosistema audiovisual actual. Se espera que el análisis permita identificar si los seguidores de estos espacios culturales tienden a formar comunidades cerradas y homogéneas, formadas por usuarios semejantes, o si, por el contrario, existen interacciones entre usuarios más diversos que favorezcan una mayor pluralidad. Para ello, se plantean los siguientes objetivos específicos:

- Construir una red de interacciones de usuarios de la red social Bluesky asociados a pódcasts de carácter sociocultural y político.
- Realizar un análisis estático y dinámico de la red generada, identificando las figuras más influyentes y estudiando patrones de interacción.
- Calcular la assortatividad de la red a partir de las relaciones entre los usuarios.
- Detectar comunidades de usuarios con un algoritmo de modularidad y analizar su cohesión interna en términos de homofilia.
- Aplicar técnicas de aprendizaje automático no supervisado y evaluar la homofilia de los usuarios en función de su equivalencia estructural y el rol que desempeñan en la red.
- Realizar un análisis comparativo de los resultados obtenidos.

3. Materiales y métodos

3.1. Selección de pódcasts

Para el estudio de la homofilia en Bluesky, se seleccionaron cinco cuentas que estuvieran vinculadas a pódcast de habla hispana. Estos pódcasts debían cumplir una serie de criterios definidos para garantizar la relevancia y coherencia del análisis. En concreto, debían i) abordar temas de carácter cultural, social o político, ii) tener una audiencia superior a 1000 seguidores para asegurar un nivel mínimo de alcance e influencia, y iii) estar actualmente activos en Bluesky para garantizar la representatividad de la red de interacciones resultante. En aquellos casos en los que el pódcast no disponía de una cuenta oficial en la plataforma, se dio por válida la cuenta personal de su locutor principal.

Las cinco cuentas seleccionadas cumplen con los requisitos establecidos y representan una diversidad de enfoques dentro del ámbito cultural, informativo y político:

- **Carne Cruda** (@carnecrudaradio.bsky.social): pódcast independiente producido por la República Independiente de la Radio y ganador del Premio Ondas al Mejor Programa en

2012. Dirigido por Javier Gallego, es conocido por tratar temas de crítica social, cultura, humor y música. Su cuenta oficial de Bluesky cuenta actualmente con más de 19 mil seguidores y 350 publicaciones.

- **Juanlu Sánchez** ([@juanlusanchez.bsky.social](#)): periodista y conductor del pódcast *Un tema al día*, vinculado al medio digital *eldiario.es*, del que también es cofundador y subdirector. Este pódcast, ganador en 2023 del Premio Ondas al pódcast Revelación, se centra en explicar la actualidad desde el periodismo informativo. Dado que el pódcast no tiene una cuenta oficial en Bluesky, hubo que seleccionar la cuenta personal de su locutor, que supera los 11.000 seguidores y funciona como principal canal de difusión del programa.
- **El Orden Mundial** ([@elordenmundial.bsky.social](#)): medio independiente especializado en análisis internacional y creador del pódcast semanal *No es el fin del mundo*, ganador del Ondas a Mejor Pódcast Conversacional en 2025. En este espacio se analizan los temas que mueven la política internacional con un enfoque divulgativo a la vez que riguroso. En la actualidad, este medio cuenta con más de 11 mil seguidores en Bluesky y 600 publicaciones.
- **Sara Riveiro** ([@sarariveiro.bsky.social](#)): periodista cultural, comunicadora audiovisual y cofundadora del pódcast *Coquetas y Bravas*, que realiza junto con Marina Grandoso. Este pódcast, de carácter costumbrista, aborda temáticas de actualidad y análisis sociocultural en un formato que emula una conversación entre amigas. Aunque se trata de un proyecto reciente, la cuenta personal de Sara en Bluesky, empleada como principal canal de difusión del pódcast en esta red social, supera los 4.000 seguidores y presenta una gran actividad, con más de 800 publicaciones.
- **Punzadas Sonoras** ([@punzadas.bsky.social](#)): pódcast producido por Inés García Hernández y Paula Ducay en colaboración con Radio Primavera Sound. En él, Paula e Inés conversan sobre filosofía, amor y cultura a partir del análisis de ficciones y obras artísticas procedentes de la literatura, el cine o las series. Aunque en Bluesky solo cuentan con 1.700 seguidores y menos de 100 publicaciones, tienen una presencia consolidada en otras plataformas como Spotify, y han colaborado con medios especializados como El Orden Mundial.

De esta manera, la selección incluye programas vinculados a medios consolidados, como *Un tema al día*, asociado a *eldiario.es*, junto con producciones completamente independientes como *Coquetas y Bravas*. En cuanto a los contenidos, abarcan desde programas con un enfoque generalista como *Carne Cruda* hasta propuestas centradas en nichos más específicos como *El Orden Mundial*, centrado en geopolítica, o *Punzadas Sonoras*, orientado al análisis de la filosofía contemporánea. Asimismo, se contemplan diferencias sustanciales en términos de audiencia: algunos de estos espacios, como *Carne Cruda*, *Un tema al día* o *El Orden Mundial*, todos ellos galardonados con el Premio Ondas, cuentan con una amplia base de oyentes, mientras que los otros se dirigen a una audiencia más reducida. Esta variedad permite analizar cómo el tamaño, la temática y el grado de institucionalización de los pódcasts pueden influir en la configuración de comunidades y en los patrones de homofilia en redes sociales.

3.2. Extracción de datos

Las interacciones entre usuarios de Bluesky y las cinco cuentas seleccionadas fueron a través de *Communalytic* (Gruzd y Mai, 2025), una herramienta computacional diseñada para el análisis en ciencias sociales que permite estudiar comunidades en línea y el discurso público en varias redes sociales, incluida Bluesky.

En concreto, los datos fueron recopilados con la herramienta *Bluesky User's Timeline Collector*, una utilidad integrada en *Communalytic* que permite extraer las publicaciones de un conjunto de cuentas de interés, junto con sus republicaciones, respuestas y menciones por parte de otras cuentas públicas. Esta información es de acceso público, por lo que para extraerla solo fue necesario introducir las direcciones de las cinco cuentas seleccionadas en el buscador de la herramienta.

La versión educativa de *Communalytic*, accesible de forma gratuita, permite la descarga de hasta 30.000 registros que representan publicaciones, republicaciones, menciones o respuestas. Esta limitación no supuso un obstáculo para el estudio, ya que las cuentas seleccionadas presentaban una actividad relativamente reciente en la red y, por lo tanto, un volumen moderado de publicaciones e interacciones con otras cuentas. No obstante, dicha restricción sí condicionó la elección de los cinco perfiles analizados, ya que hubo que excluir cuentas o medios de mayor tamaño como *El País* o *Cadena SER*, que también han producido varios podcasts. Además, la versión gratuita solo permitió centrar el análisis en un máximo de cinco cuentas, de ahí la selección.

La extracción de los datos se llevó a cabo el 27 de abril de 2025 a las 11:36 a.m. y retornó un total de 11.959 registros que se descargaron en formato CSV. En dicho archivo, cada fila corresponde a una publicación, *repost* o respuesta, e incluye variables como la fecha publicación, el nombre y usuario del autor, el contenido textual del mensaje, los usuarios referenciados y diversas métricas que miden el impacto de las publicaciones como el número de *likes*, respuestas o *reposts*, útiles para evaluar la influencia de los usuarios en la red.

Asimismo, también se exportó la red de interacciones entre los usuarios de Bluesky en formato GEXF, compatible con herramientas de análisis de redes como *Gephi*. Esta red fue construida automáticamente con la herramienta *Network Analyzer* de *Communalytic* a partir de las respuestas y menciones de las publicaciones de las cinco cuentas seleccionadas. El grafo resultante incluye un total de 6.424 nodos, cada uno correspondiente a un usuario único, y 8.234 enlaces dirigidos que representan las interacciones entre ellos, ya sea en forma de mención (*quote*) o respuesta a una publicación.

3.3. Análisis estático y dinámico

El análisis de la red de interacciones se llevó a cabo utilizando Gephi v10.1, un software de código abierto orientado a la visualización y el análisis estructural de redes (Bastian et al., 2009). Esta red fue importada en Gephi a partir del archivo en formato GEXF generado por *Communalytic* y se visualizó utilizando la pestaña de *Vista General*.

3.3.1. Preprocesamiento de la red

Al importar la red en Gephi, se identificó un número considerable de parejas nodos aislados o agrupados en pequeñas comunidades no conectadas a ninguna de las cinco cuentas principales, ya fuera de manera directa o indirecta. Dado que estos nodos introducían ruido y no aportaban ningún valor al análisis, al no estar aparentemente relacionados con las cuentas de interés, se decidió eliminarlos de la red.

Para ello, se ejecutó el algoritmo de componentes conexos (Tarjan, 1972), disponible en el panel de *Estadísticas* de Gephi, que agrupa en un mismo componente *débil* a todos los nodos conectados entre sí por algún camino, sin tener en cuenta la dirección de las interacciones. Los cinco nodos principales, correspondientes a las cuentas de partida, se encontraban agrupados en el mismo componente, lo que simplificaría el proceso de filtrado posterior. A continuación, se aplicó un filtro para conservar únicamente aquellos nodos pertenecientes a dicha componente. Este filtro se aplicó desde el panel de *Filtros*, donde se seleccionó el filtro de igualdad de atributos para el *Component ID*, y se especificó el identificador compartido por los cinco nodos principales del estudio, que era el 0.

El filtrado generó un subgrafo de 6.086 nodos interconectados a través de 7.939 aristas. Este procedimiento permitió eliminar los nodos aislados, asegurando que el posterior análisis de la red se centrara exclusivamente en las cuentas asociadas con los cinco pódcast seleccionados.

3.3.2. Visualización de la red

Para facilitar la visualización de la red, se aplicó el algoritmo ForceAtlas 2, configurado específicamente para evitar el solapamiento entre los nodos. Este algoritmo utiliza un modelo de fuerzas físicas para distribuir los nodos de manera que aquellos más conectados se agrupan, mientras que los menos conectados se dispersan, favoreciendo la interpretación de la red. La opción para evitar el solapamiento entre los nodos se activó tras haber distribuido los nodos con la opción de *Aproximar Repulsión*, ya que ambas opciones son incompatibles simultáneamente. Esta distribución generó una representación más clara y comprensible de las relaciones dentro de la red, permitiendo distinguir todos los actores.

Todas las visualizaciones generadas para la memoria fueron exportadas desde la ventana de *Previsualización* en formato SVG. Este formato asegura la preservación de la calidad de las imágenes, lo que facilita la exploración de los resultados obtenidos.

3.3.3. Análisis estático

El análisis estático de la red se realizó mediante el cálculo de estadísticas globales y métricas locales de centralidad:

- Las métricas globales seleccionadas fueron el grado medio de la red, el diámetro, la densidad del grafo y el coeficiente medio de *clustering*. Para calcular las métricas, se consideró el

peso de las interacciones, aunque en el caso del grado medio también se obtuvo su valor no ponderado. Todas estas métricas se encuentran disponibles en el panel de *Estadísticas* de Gephi.

- Para evaluar la importancia de los nodos dentro de la red, se calcularon las métricas de centralidad de grado, de intermediación, de cercanía y del vector propio, junto con los algoritmos *HITS* (Kleinberg, 1999) y *PageRank* (Page et al., 1999).

3.3.4. Análisis dinámico

En un siguiente análisis, se estudió la evolución temporal del grafo. Para convertir la red estática en dinámica, fue necesario definir un intervalo de tiempo que indicara la fecha de publicación de cada uno de los mensajes. Este atributo se creó desde la ventana de *Laboratorio de datos*, seleccionando la tabla de aristas y copiando los valores de la columna *TimeSet*, que contenía el intervalo de tiempo proporcionado por *Communalytic*, a la variable *Interval*, que es reconocida automáticamente por Gephi para la construir la línea temporal. El intervalo de tiempo se estableció únicamente con la fecha de publicación debido a que *Communalytic* no permite extraer el instante de tiempo preciso en el que tuvo lugar cada interacción.

A continuación, se activó la línea temporal en la ventana de *Vista general* y se configuró para que se mostrase el tiempo en formato fecha, en lugar de usar valores numéricos, con el fin de facilitar su interpretación. Asimismo, se acotó el intervalo temporal desde octubre de 2024 hasta mayo de 2025, que corresponde al período en el que se registró actividad en la red.

Con esta configuración, se creó una animación para visualizar la evolución de la red a lo largo del tiempo utilizando una ventana de una semana. Se escogió este tamaño de ventana para poder observar de manera clara los cambios en las interacciones entre los nodos sin sobrecargar la visualización con demasiados detalles, ya que permitía equilibrar la frecuencia de las interacciones con la claridad de los patrones emergentes. La animación resultante se grabó con la aplicación Microsoft Clipchamp para Windows y se exportó en formato mp4. Este archivo se puede encontrar como material complementario.

Por último, el análisis de la red dinámica se intentó complementar con el cálculo de estadísticas temporales, como el número de nodos y aristas activos en cada intervalo de tiempo, el grado medio y el coeficiente de *clustering* por ventana temporal. Estas métricas permiten observar cómo varía la actividad y la estructura local de la red a lo largo del tiempo, proporcionando una visión más completa de su evolución. No obstante, el elevado tamaño de la red supuso una limitación significativa, ya que el tiempo de cómputo requerido para obtener estas métricas resultó excesivo y no se pudieron calcular.

3.4. Modularidad

La modularidad de la red se calculó mediante el algoritmo de detección de comunidades propuesto por Blondel et al., 2008, y disponible en el panel de *Estadísticas* de Gephi. Este algoritmo

emplea un enfoque heurístico para identificar comunidades dentro de una red, agrupando nodos en regiones que presentan una densidad de conexiones internas significativamente mayor que con el resto de la red, optimizando la modularidad. El valor de modularidad global obtenido, Q , se encuentra normalizado en el rango [0,1], donde valores altos indican una mayor estructura comunitaria. Además, el algoritmo asignó a cada nodo una etiqueta correspondiente a la comunidad a la que había sido asignado.

3.5. Aplicación de técnicas de aprendizaje automático

Con el objetivo de identificar grupos de nodos que desempeñan un papel similar en la red, con independencia de su posición en la misma, se aplicó una técnica de aprendizaje automático no supervisado para agrupar las cuentas en base a sus valores de centralidad. El *clustering* se llevó a cabo en WEKA versión 3.8.6, una aplicación de código abierto para tareas de aprendizaje automático (Frank et al., 2016).

Para la segmentación, se utilizaron como atributos las siguientes métricas de centralidad previamente calculadas en Gephi: grado de entrada y salida ponderados, centralidad de cercanía, de intermediación y del vector propio, y *PageRank*. Estas métricas se escogieron porque capturan distintos aspectos del rol que desempeña un nodo en una red social. Por lo tanto, al combinarse pueden proporcionar una caracterización robusta de la funcionalidad e importancia de cada nodo, permitiendo identificar cuentas con funciones similares dentro de la red de Bluesky. Los valores de las métricas seleccionadas para cada nodo se extrajeron de la ventana de *Laboratorio de datos*, desde la que se exportaron en formato CSV.

Una vez importada la tabla de nodos en WEKA, se realizó un preprocesamiento de los datos. En primer lugar, se seleccionaron únicamente los atributos relevantes para el estudio - correspondientes a las métricas de centralidad mencionadas- que se normalizaron desde la pestaña de *Preprocess* utilizando el filtro *Normalize* con los siguientes parámetros: S 1.0 T 0.0. Este filtro escala todos los atributos al rango [0,1], lo cual es esencial cuando se trabaja con algoritmos de *clustering* basados en medidas de distancias. Esta normalización asegura que cada atributo contribuya equitativamente al cálculo de la distancia, evitando sesgos provocados por diferencias en las unidades de medida.

El *clustering* se realizó desde la pestaña de *Cluster* en WEKA. Para este proceso, se seleccionó el algoritmo *Expectation-Maximization (EM)* (Dempster et al., 1977) que puede aprender la distribución subyacente de los datos sin requerir un número predefinido de clústeres. Se optó por este algoritmo frente a otros métodos de aprendizaje automático no supervisado como *k-means* porque permite inferir automáticamente el número de clústeres y asignar probabilísticamente cada instancia a uno de los grupos. A diferencia de k-means, que asume particiones esféricas y equitativas, EM resulta más adecuado para datos con distribuciones más complejas y estructuras menos simétricas, como las que suelen derivarse de métricas de redes sociales (Ibrahem Hafez et al., 2013).

A continuación, se muestra el código con los parámetros utilizados para el entrenamiento del

modelo en WEKA para facilitar la replicación del trabajo:

```
01 | weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M  
1.0E-6 -K 10 -num-slots 1 -S 100
```

La partición resultante se exportó desde el panel de *Visualize cluster assignments*, disponible al realizar clic derecho sobre el modelo entrenado. Los resultados se guardaron en formato ARFF, que luego se convirtió a CSV, conservando los atributos seleccionados, correspondientes al identificador de los nodos y a las métricas de centralidad normalizadas, junto con el clúster asignado a cada nodo.

3.6. Estudio de la homofilia

El grado de homofilia de una red se puede estudiar mediante el cálculo de la asortatividad, una métrica que cuantifica la tendencia de los nodos a conectarse con otros con los que comparten características similares con mayor frecuencia que con nodos distintos. Para realizar este análisis, se utilizó el lenguaje de programación Python (versión 3.14). El código correspondiente se ha incluido como material suplementario.

El análisis se planteó desde dos enfoques, en función del criterio utilizado para calcular la semejanza entre los pares de nodos:

- **Similitud estructural:** grado en que dos nodos comparten conexiones similares dentro de la red. En este estudio, la equivalencia estructural entre los pares de nodos se calculó a partir de la matriz de adyacencia ponderada que, a su vez, se construyó a partir del fichero CSV con las aristas de la red extraído del *Laboratorio de datos* de Gephi. La matriz de adyacencia permitió considerar el número de vecinos comunes entre dos nodos junto con la intensidad de las conexiones, permitiendo identificar usuarios con patrones de conexión similares.
- **Similitud de rol:** semejanza funcional entre dos nodos en la red, independientemente de su vecindario. Para estimarla, se emplearon las seis métricas de centralidad empleadas para el *clustering*. Estas métricas capturan distintos aspectos del posicionamiento de los nodos dentro de la red, permitiendo comparar su influencia o papel estructural en términos globales.

En ambos casos, la similitud se cuantificó utilizando el coeficiente de correlación de Pearson (Pearson y Galton, 1895) aplicado en paralelo sobre las matrices de adyacencia (semejanza estructural) y las métricas de centralidad (semejanza regular). La matriz de correlaciones resultante se utilizó como punto de partida para estudiar la homofilia en la red e identificar parejas de usuarios semejantes. La homofilia global de la red se definió como la media de los coeficientes de correlación entre pares de nodos, lo que permite interpretar el resultado como un indicador agregado de cohesión estructural o funcional, según el tipo de similitud considerado.

Además del estudio de la homofilia en la red completa, se repitió el análisis para cada una de las comunidades detectadas mediante el algoritmo de modularidad y de los clústeres obtenidos por

modelo de expectación-maximización, con el fin de evaluar el grado de homogeneidad interna en cada agrupación. El coeficiente de correlación entre los pares de nodos dentro de cada comunidad o clúster se comparó con el coeficiente observado entre nodos de diferentes grupos, utilizando la prueba U de Mann-Whitney (Mann y Whitney, 1947) para evaluar diferencias significativas en la homogeneidad interna.

4. Resultados y discusión

4.1. Análisis exploratorio de la red

La extracción de datos mediante *Communalytic* permitió recopilar todas publicaciones, menciones, respuestas y republicaciones en Bluesky asociadas a cinco cuentas de pódcast que tratan temas socioculturales en español. Tal como se muestra en la Figura 1, la actividad en torno a estas cuentas comienza a ser significativa a partir de noviembre de 2024 y presenta un par de picos a mediados de noviembre y finales de enero. Estos cambios temporales en la actividad de las cuentas tendrán un impacto en la red de interacciones generada, por lo que se estudiarán en mayor profundidad con la red dinámica (sección 4.3).

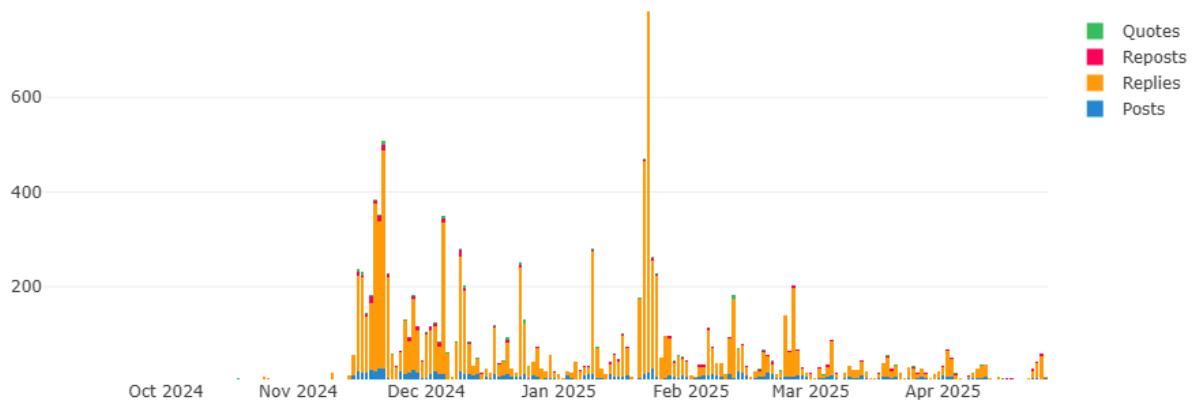


Figura 1: Distribución temporal del número de publicaciones, menciones, republicaciones y respuestas obtenidas mediante *Communalytic* entre octubre de 2024 y abril de 2025. El histograma muestra la evolución de la actividad en torno a las cinco cuentas analizadas.

Por otro lado, se observa que la mayoría de los registros obtenidos corresponden a respuestas de publicaciones asociadas a las cuentas analizadas. Estas respuestas, junto con las menciones o citas (*quotes*), fueron los elementos utilizados para construir la red de interacciones, dado que representan los vínculos explícitos entre usuarios. Al centrarse en estas interacciones, el análisis de la red permite explorar las dinámicas conversacionales y relacionales que se generan en torno a pódcasts socioculturales. Esta aproximación resulta especialmente relevante para el objetivo del trabajo, que se centra en estudiar la semejanza entre los nodos de una red, ya que estas conexiones pueden relejar afinidades temáticas o de estatus entre los usuarios de una red social.

En la Figura 2, se representa la red de interacciones resultante. Este grafo muestra las interacciones entre 6086 usuarios de Bluesky vinculados de manera directa o indirecta al ámbito de los

pódcasts. Debido al elevado número de nodos y aristas, la visualización presenta una estructura densa y difícil de interpretar. Por ese motivo, se han resaltado los nodos con mayor centralidad de grado no ponderado, es decir, aquellas cuentas que han interactuado con un mayor número de usuarios distintos sin tener en cuenta la dirección de las relaciones. Estos nodos corresponden a los principales actores en la red en términos de popularidad y socialibilidad, siendo los principales focos de interacción entre usuarios de la red social.

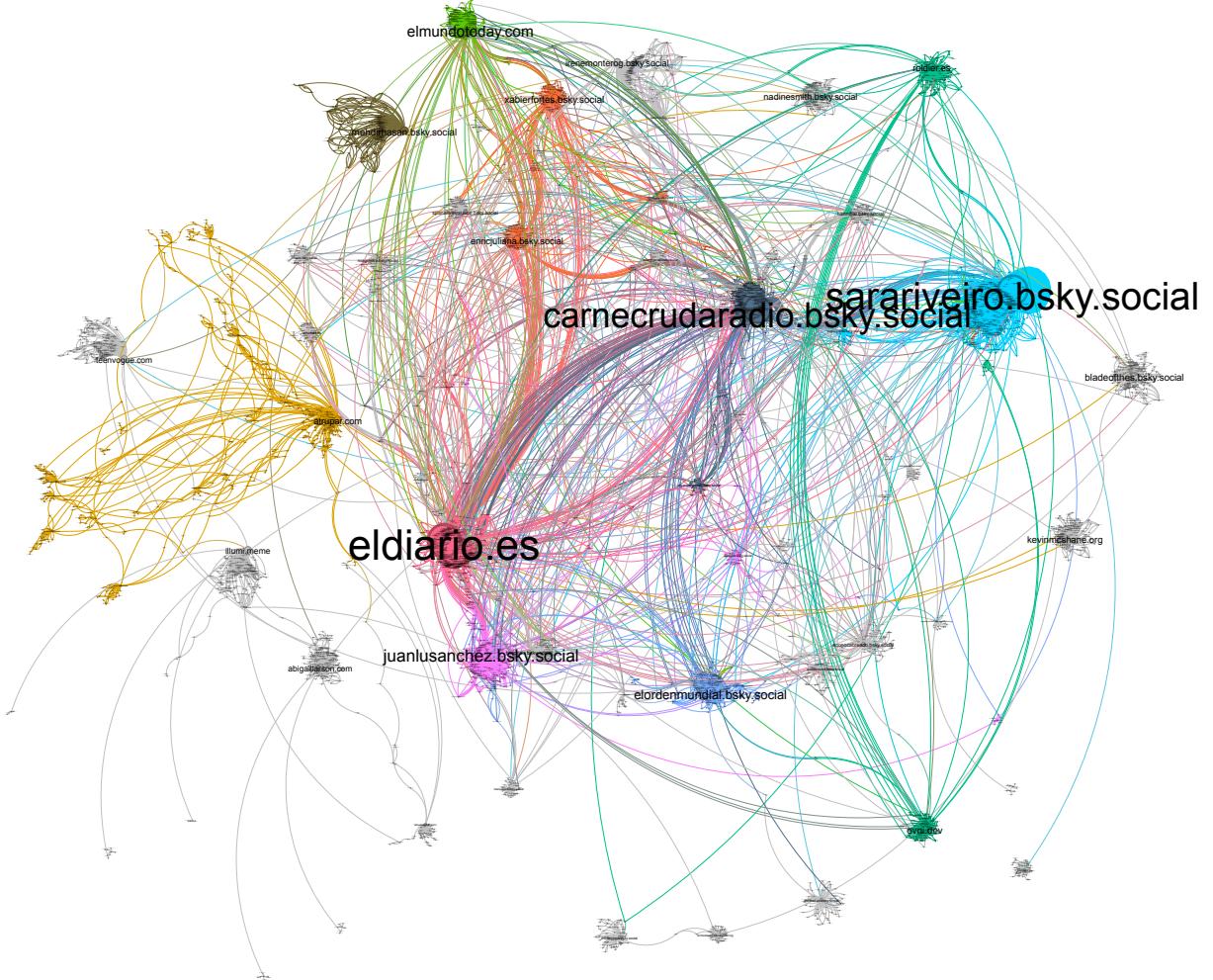


Figura 2: Red de interacciones en Bluesky entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado, mientras que el color indica la comunidad a la que han sido asignados por el algoritmo de detección de comunidades. Solo se han coloreado las 10 comunidades con mayor porcentaje de miembros. El color de los enlaces corresponde al del nodo de origen, el tamaño indica el número de interacciones entre los usuarios, y la curvatura señala la dirección de la interacción. La distribución espacial de los nodos se obtuvo aplicando el algoritmo ForceAtlas2.

Un primer aspecto destacable es que las cinco cuentas de pódcasts seleccionadas para el estudio se encuentran dentro de un mismo componente conectado, lo cual indica que, a pesar de su diversidad temática y su diferente alcance, forman parte de un ecosistema de interacciones común. Estos resultados sugieren que el contenido cultural y social puede fluir entre distintos nichos a través de usuarios y medios de comunicación compartidos.

Desde una perspectiva estructural, la red presenta un diámetro *dirigido* de 13, lo que implica que cualquier usuario puede ser alcanzado desde otro en un máximo de 13 pasos. Además, la

longitud media del camino más corto entre dos nodos es de 4,2, lo que refleja una conectividad razonable que indica que la información puede fluir fácilmente a través de la red, a pesar de su baja densidad global, próxima a 0. Esta baja densidad y la elevada estructura comunitaria de la red, que tiene un valor de modularidad global de 0,77, son debidas a que la mayoría de usuarios solo mantienen una o dos conexiones, generalmente con alguna de las cuentas principales. Por este mismo motivo, el coeficiente de clustering global es muy bajo, siendo inferior a 0,04, lo que indica que la tendencia de los usuarios a formar triángulos de interacciones es muy baja, típico de redes descentralizadas donde las relaciones están centradas en cuentas influyentes, sin que el resto de usuarios interactúen necesariamente entre ellos. Este hecho se ve acentuado por la metodología de extracción de datos utilizada, basada en seleccionar las respuestas y menciones asociadas a un pequeño número cuentas de Bluesky.

En la red también se observa que las cuentas que interactúan con un mayor número de usuarios distintos son la del *eldiario.es*, la cuenta personal de Sara Riveiro -cofundadora del pódcast *Cocquetas y Bravas-* y la de *Carne Cruda*. Estos resultados sugieren que estos usuarios desempeñan un papel central en la dinámica de la red. En particular, destaca la relevancia de *eldiario.es*, ya que no se trata de una de las cinco cuentas seleccionadas para el análisis y tampoco pertenece a un pódcast, sino a un medio de comunicación consolidado con una fuerte presencia digital. Su elevada centralidad de grado sugiere que actúa como un nodo clave en la conversación pública, posiblemente favoreciendo el debate dentro de la comunidad de Bluesky, amplificando contenidos e interactuando activamente con perfiles vinculados al ecosistema de los pódcasts socioculturales. Esta posición central podría deberse a que trata temas similares a los de los pódcasts seleccionados, al estar especializado en análisis social y político, así como a su vinculación con proyectos como *Un tema al día*, conducido por Juanlu Sánchez -cofundador y subdirector *eldiario.es*-, *Carne Cruda*, que es un medio asociado a este periódico, con el que se alinea políticamente, o incluso con Sara Riveiro quien colaboró en el pasado como redactora. En conjunto, estos resultados sugieren que tanto medios consolidados como alternativos pueden tener una amplia presencia en redes sociales emergentes como Bluesky, articulando comunidades y contenidos de carácter político, social y cultural.

Por otro lado, tanto *El Orden Mundial* como Juanlu Sánchez presentan una centralidad de grado algo menor, aunque ambos parecen mantener un volumen significativo de interacciones con usuarios dentro de sus respectivos círculos. Asimismo, su conexión con nodos centrales como *eldiario.es* o *Carne Cruda* sugiere que pueden tener un papel relevante en la red, aunque no interactúen con tantos usuarios.

En contraste, *Punzudas Sonoras*, el otro pódcast seleccionado para el análisis, aparece relativamente aislado, formando una pequeña comunidad periférica conectada al resto de la red de forma indirecta a través de algunos usuarios comunes con el *El Orden Mundial*. Esta menor conectividad puede explicarse por la orientación filosófica de su contenido, que podría atraer a un público más reducido y menos alineado con el enfoque político y social de las otras cuentas analizadas, limitando la interconexión de sus audiencias. Estas diferencias estructural dentro de la red refuerza el interés por explorar la homofilia desde una perspectiva funcional y estructural, ya que sugiere que la afinidad de intereses o los roles de los usuarios pueden traducirse en

patrones concretos de interacción.

4.2. Análisis de la centralidad

Las métricas de centralidad permiten identificar a aquellos usuarios que desempeñan un papel destacado en la red, ya sea por su elevada actividad, su estatus dentro de la comunidad o su posición estratégica para facilitar la circulación de información. En la Figura 3, se puede observar que distintas cuentas adquieren diferente importancia en función de la métrica considerada, lo que pone de manifiesto la diversidad de roles que pueden desempeñarse en una red social. Como es de esperar, debido al mayor número de interacciones posibles, la mayoría de las cuentas destacadas pertenecen a comunidades de mayor tamaño, que son las que aparecen destacadas en la imagen. Esta figura resume los principales resultados, pero los grafos completos con los valores de centralidad para todos los nodos se pueden consultar en el Anexo (sección 7.2).

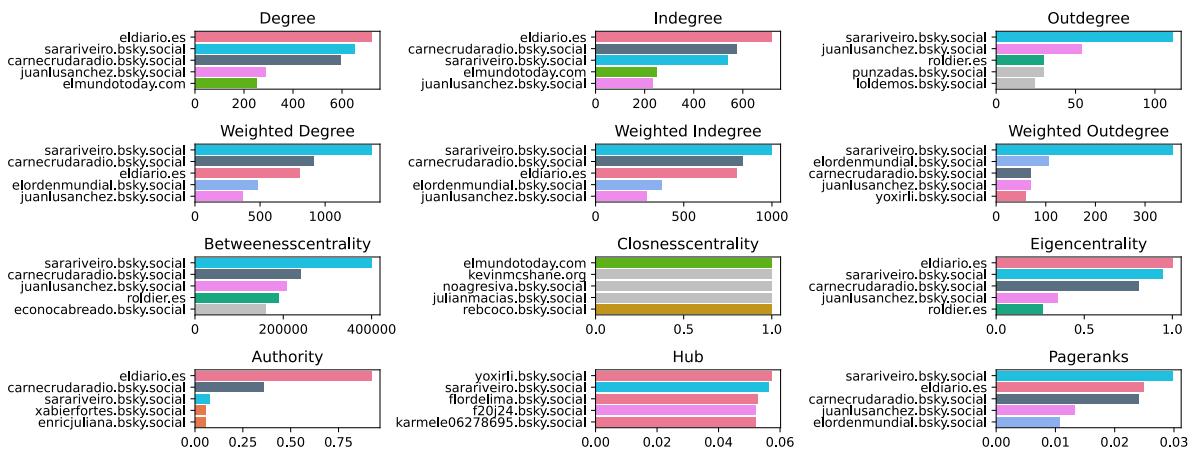


Figura 3: Comparativa de las cinco cuentas más destacadas según distintas métricas de centralidad en la red de interacciones de Bluesky. Los colores indican la comunidad a la que han sido asignados por el algoritmo de modularidad. Las comunidades con menos de 200 miembros se muestran en color gris.

Los resultados revelan que *eldiario.es*, *Carne Cruda*, Sara Riveiro y, en menor medida, Juanlu Sánchez y *El Orden Mundial* son centrales en la red según varias de las métricas consideradas. Esto sugiere que actúan como nodos clave dentro del ecosistema digital de los podcasts, ya sea facilitando la difusión de contenidos, articulando comunidades o promoviendo la interacción entre distintos grupos de usuarios.

En concreto, Sara Riveiro destaca como la usuaria más activa en la red de Bluesky, tal como lo evidencia su elevado grado ponderado. Este indicador refleja que mantiene un volumen particularmente alto de interacciones superando a cuentas más conocidas. Aunque *eldiario.es* es la cuenta que concentra la mayor cantidad de interacciones entrantes desde usuarios únicos, lo cual se traduce en una mayor centralidad de grado de entrada no ponderada, Sara Riveiro presenta un mayor grado de salida, lo que indica que es quien más “habla” o participa en conversaciones con otros usuarios. Esta distinción pone en relieve diferentes formas de participación en la red: mientras unos actores reciben más atención, otros impulsan activamente la comunicación. Estas conclusiones se acentúan al considerar el peso de las interacciones, donde *Sara Riveiro* lidera

el ranking tanto en grado de entrada como de salida, lo que refuerza su rol central dentro de la red. Aunque no es el nodo más “popular” en términos de número de usuarios únicos que interactúan con ella, sí recibe un mayor volumen de respuestas o menciones que otras cuentas, lo que sugiere una participación más intensa y sostenida. Además, su elevada centralidad de intermediación sugiere que también actúa como puente entre diferentes comunidades de usuarios dentro la red, desempeñando un papel clave la circulación de la información al facilitar la conexión entre distintos grupos de usuarios.

Los resultados obtenidos cobran sentido al analizar el perfil específico de Sara Riveiro, una periodista y comunicadora audiovisual joven especializada en redes sociales y creación de contenidos en plataformas como *TikTok* o *X*. Esta familiaridad con las nuevas tecnologías y los entornos digitales emergentes puede explicar su alta actividad en *Bluesky*. A diferencia de cuentas institucionales más consolidadas mediáticamente, como *eldiario.es*, que deben cuidar su presencia en la red, Sara Riveiro puede mantener un trato más directo y cercano con sus seguidores, favoreciendo una interacción más dinámica y activa. Un fenómeno similar se observa en *Punzadas Sonoras*, un podcast conducido por dos filósofas jóvenes que, aunque cuenta con una comunidad de seguidores reducida como se puede deducir por su bajo grado de entrada, presenta un grado de salida no ponderado mayor que el de otras cuentas de la red. Además, en el caso de Sara Riveiro, sus interacciones con cuentas de gran autoridad como el *eldiario.es* y *Carne Cruda*, con el que colabora esporádicamente como invitada, junto con su elevada actividad en la red, pueden contribuir a situarla en una posición estratégica dentro del ecosistema digital analizado a pesar de no tener un nombre tan reconocido. De hecho, la cuenta de Sara Riveiro tiene el mayor valor de *PageRank* de todas las analizadas. Este alto valor de *PageRank* no solo refleja su nivel de actividad e importancia en la red, sino también la relevancia de los nodos con los que interactúa.

Por otro lado, *eldiario.es* se posiciona como el nodo más “popular” de la red, al ser la cuenta que mayor cantidad de respuestas y menciones recibe por parte de usuarios distintos, aunque no muestre una actividad especialmente destacada si se considera el grado de salida ponderado. Estos resultados sugieren que sus publicaciones incentivan el debate y fomentan la participación, probablemente debido a que abordan temas de interés general propios de un medio de comunicación generalista, lo que contribuye a atraer a un público más amplio. Además, es la cuenta con mayor autoridad dentro de la red, con una diferencia notable con respecto al resto de nodos, lo que refuerza la idea de que su contenido es percibido como relevante y valioso para la comunidad. Por otro lado, su elevada centralidad del vector propio también indica que es un actor clave en la difusión de la información, ya que está bien conectado con otras cuentas influyentes, por lo que sus publicaciones tendrán un gran alcance. No obstante, resulta llamativo que este medio no figure entre los nodos con mayor centralidad de intermediación, lo que sugiere que no actúa como un puente destacado entre diferentes subgrupos o comunidades dentro de la red. Este rol está más asociado a las cuentas personales de periodistas como Sara Riveiro, Javier Gallego (*Carne Cruda*), Juanlu Sánchez, o Yago Álvarez Barba (*econocabreado*).

Otros actores centrales en la red son *Carne Cruda* y Juanlu Sánchez, destacando especialmente el primero de ellos por sus elevados grados de autoridad y *PageRank*. Ambas cuentas parecen desempeñar un papel central en la difusión de información de la red, reflejado en su alta centralidad.

lidad de intermediación. Además, su elevada centralidad de grado de entrada -con y sin pesos- sugiere que sus publicaciones son de interés general para el público de Bluesky, mientras que su grado de salida ponderado indican una alta actividad y una interacción directa con otros usuarios. Esta dinámica puede deberse a no ser cuentas estrictamente institucionales, que les permite mantener una mayor cercanía con su audiencia, a la vez que generan contenido de interés. En el grafo de la Figura 2, también veíamos que estas cuentas están fuertemente conectadas con otros actores importantes como *eldiario.es*.

Finalmente, los resultados correspondientes a *El Orden Mundial* sugieren que este medio mantiene una actividad considerable en Bluesky, aunque interacciona con un menor número de usuarios. Esta dinámica se observa al comparar los grados de entrada y salida ponderados con sus equivalentes no ponderados que no son tan destacados, lo que sugiere que mantienen una interacción activa con su audiencia, posiblemente formada por un núcleo reducido pero fiel de usuarios. A diferencia de otros podcasts como *Carne Cruda* que también aborda temas sociales y políticos, *El Orden Mundial* se especializa en su análisis geopolítico, lo que explica que el círculo de usuarios con los que interactuan sea más específico y reducido.

La centralidad de cercanía también se ha representado en la Figura 3, pero su análisis resulta poco informativo, ya que se calculó sobre la red dirigida, lo que implica que muchos nodos quedaron excluidos del cálculo al no existir trayectorias dirigidas hacia ellos desde ciertos puntos de la red debido a la baja densidad de la misma. Como consecuencia, un número considerable de nodos presentan un valor de cercanía igual a 1, lo que sugiere que pueden estar aislados en términos de accesibilidad estructural, muchos de ellos siendo nodos que solo están conectados a un único usuario. Estas cuentas “centrales” suelen mostrar una baja relevancia en el resto de métricas de centralidad, lo que refuerza su carácter periférico dentro del ecosistema analizado. Un caso ilustrativo es el de *El Mundo Today*, un medio de comunicación satírico que, a siendo una de las cuentas con mayor grado de entrada -no ponderado-, presenta una cercanía de 1. Esto se debe a que las interacciones se concentran en menciones hacia su cuenta, sin que esta actúe necesariamente como emisora, limitando su capacidad para acceder a otras regiones de la red de forma dirigida.

Estas limitaciones ponen de manifiesto la escasa utilidad de la centralidad de cercanía en redes dirigidas débilmente conectadas cuando se pretende identificar a los principales actores estructurales. Por este motivo, se repitió el análisis considerando la red como no dirigida. En este nuevo enfoque, los resultados se vuelven más coherentes con el resto de métricas (Figura 26 del Anexo), destacando cuentas como Sara Riveiro, *eldiario.es* y *Carne Cruda*, apoyando la hipótesis de que actúan como actores estratégicos para acceder y difundir información dentro de la red. También sobresalen las cuentas del periodista Aaron Rupar (*atrupar*) y *Teen Vogue*, que también tienen valores altos de intermediación en la red no dirigida, lo que refleja su capacidad para mantenerse conectados a distintos sectores de la red.

En conjunto, el análisis de las métricas de centralidad permite identificar a los actores más influyentes dentro del ecosistema digital analizado. Estas métricas permiten detectar nodos que desempeñan papeles similares dentro de la red, junto con las cuentas que favorecen la circulación de contenidos y la interconexión de distintos grupos de usuarios. Esta capacidad de conectar

comunidades diversas y promover la difusión de información puede influir en los patrones de homofilia observados en la red. Por un lado, las figuras influyentes pueden atraer y poner en contacto a usuarios con perfiles similares que comparten intereses e ideas, pero por otro lado, también pueden incrementar la exposición a discursos heterogéneos dentro de Bluesky y promover, por tanto, la interacción entre usuarios con intereses parcialmente distintos, disminuyendo así la creación de cámaras de eco.

4.3. Análisis dinámico de la red

Para estudiar la evolución temporal de la red, se ha creado una animación que muestra las interacciones semanales entre los distintos usuarios a lo largo del periodo de estudio, y se ha incluido como material suplementario. Esta visualización permite identificar los momentos de mayor actividad y observar cómo se reconfiguran los patrones de interacción a lo largo del tiempo.

En la animación, se aprecia un incremento notable de las interacciones durante la segunda quincena de noviembre (Figura 4), momento en el que emergen como actores centrales cuentas de periodistas y medios consolidados como *eldiario.es*, *Carne Cruda*, Sara Riveiro, *El Mundo Today* y Juanlu Sánchez. Este periodo se caracteriza por una alta concentración de interacciones en torno a estos nodos, que además interaccionan entre ellos, lo que sugiere un momento de gran actividad periodística y mayor cohesión informativa que puede haber sido impulsado por acontecimientos relevantes que estimularon el debate dentro de Bluesky. Esta reorganización temporal de la red también parece favorecer un aumento en la conectividad entre comunidades previamente más aisladas, lo que puede tener un impacto en la homofilia de la red, ya que podría estar poniendo en contacto a usuarios con intereses similares previamente más aislados.

A partir de la segunda semana de diciembre, la red experimenta un descenso progresivo de actividad que se ve interrumpido por un breve repunte en enero (Figura 5). En este periodo, emergen nuevos actores como el periodista Aaron Rupar (*atrupar*), que concentra la mayoría de las conexiones durante esa semana, aunque no dejan de destacar cuentas como *eldiario.es* o *Carne Cruda*. Las poblaciones de usuarios asociadas a estas dos cuentas parecen interaccionar entre ellas más que con otras comunidades. A partir de febrero, la actividad de la red entra en una fase de declive marcado, con una caída sostenida del número de interacciones que no logra recuperarse en los meses siguientes.

Esta evolución temporal es coherente con la tendencia observada en la Figura 1. En conjunto, estos resultados sugieren una dinámica episódica de la conversación, con momentos puntuales de alta intensidad, en noviembre y enero, seguidos de períodos de menor actividad. Los incrementos en la actividad pueden estar asociados a migraciones significativas de usuarios desde X (anteriormente Twitter) hacia Bluesky, un fenómeno que ha sido documentado en estudios recientes como el de Arroyo-Machado et al., 2024. Este cambio de plataforma podría explicarse por una combinación de factores como la adopción de nuevas políticas de moderación de contenidos en X, la integración de sistemas de inteligencia artificial generativa, la pérdida de privacidad de los usuarios o el incremento de los *bots*, así como el repunte de figuras controvertidas como Donald

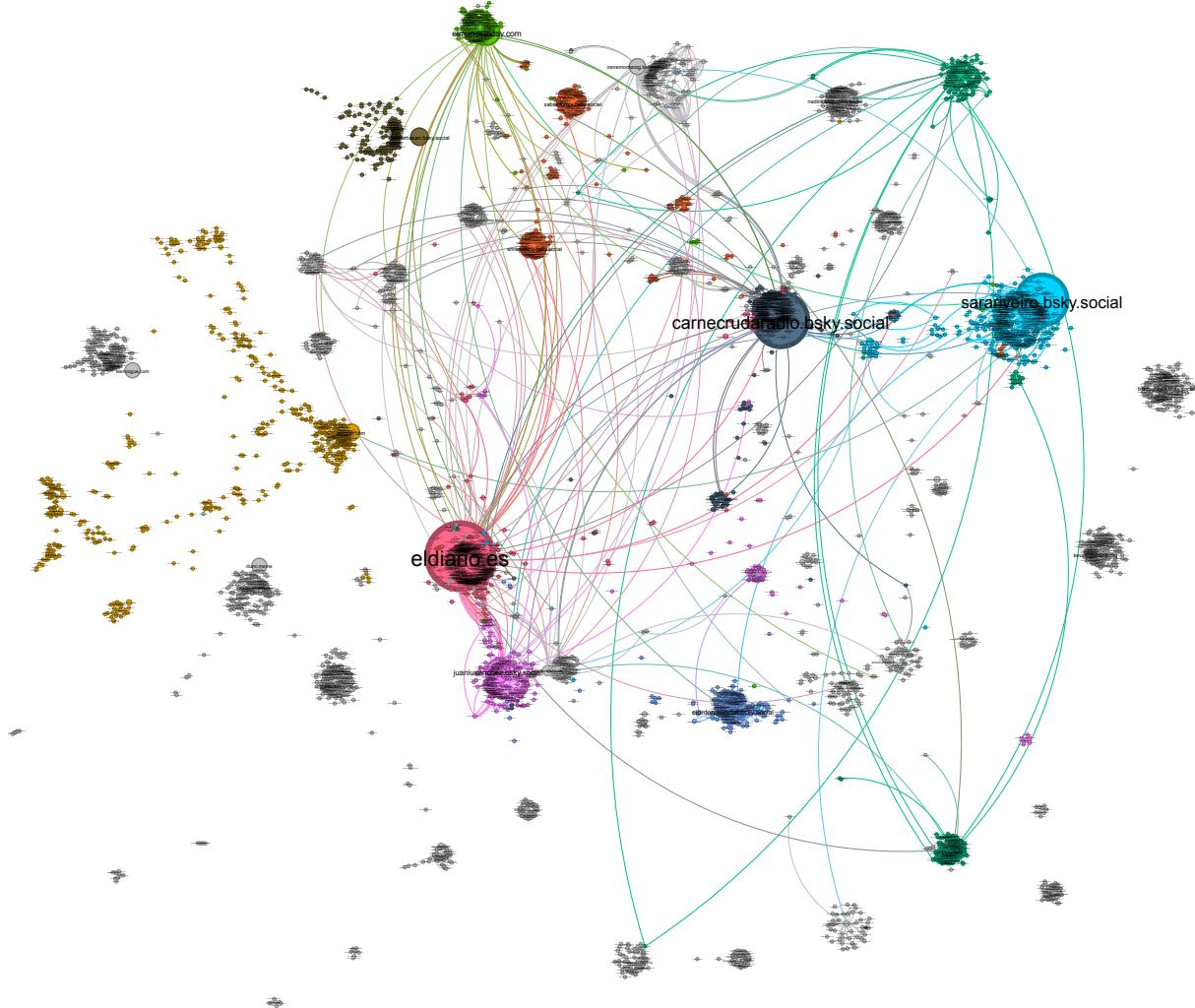


Figura 4: Snapshot de la red de interacciones en Bluesky entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana en la semana del 15 de noviembre de 2024. El tamaño de los nodos es proporcional a su centralidad de grado, mientras que el color indica la comunidad a la que han sido asignados por el algoritmo de detección de comunidades. Solo se han coloreado las 10 comunidades con mayor porcentaje de miembros. El color de los enlaces corresponde al del nodo de origen, el tamaño indica el número de interacciones entre los usuarios, y la curvatura señala la dirección de la interacción.

Trump tras su victoria en las elecciones presidenciales estadounidenses. Además, la estrecha relación entre Trump y Elon Musk, actual propietario de X, pudo generar preocupación sobre la neutralidad de esta red, lo que pudo llevar a muchos usuarios a explorar alternativas más descentralizadas (Arroyo-Machado et al., 2024; Ittefaq, 2025; Pascual, 2024). En este contexto, es de esperar que estas migraciones masivas de usuarios atrajeran también a medios de comunicación, artistas y cuentas socioculturales hacia Bluesky, ampliando así su ecosistema y reforzando su papel como espacio alternativo para la conversación pública.

Asimismo, eventos con alto impacto político, como las elecciones de Estados Unidos del 5 de noviembre de 2024 y la investidura de Trump el 20 de enero de 2025, que coinciden con los dos picos observados, probablemente contribuyeron a un aumento puntual de la visibilidad y la actividad de cuentas culturales y sociopolíticas que cubrieran esos eventos y fomentasen el debate en esta nueva plataforma.

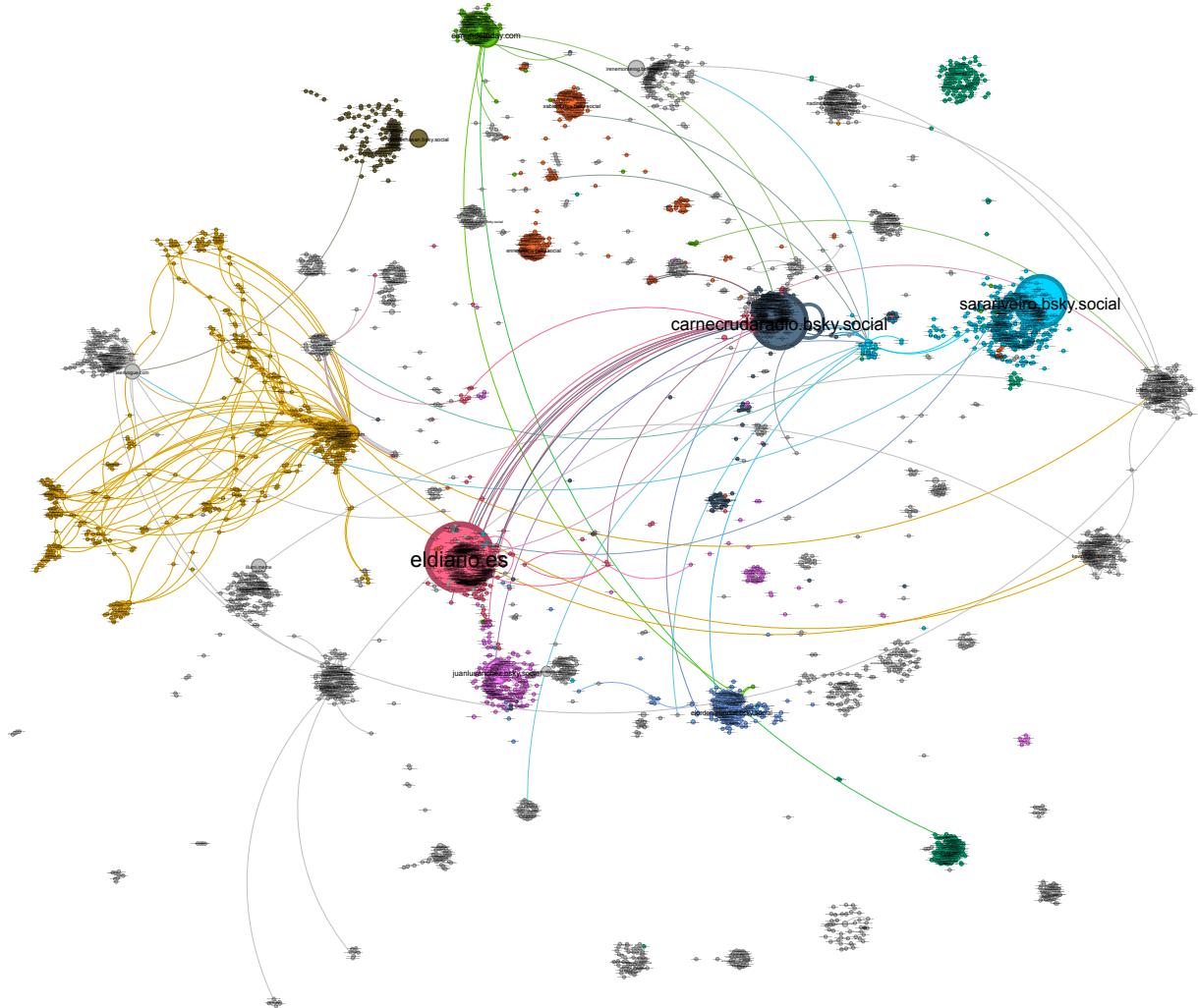


Figura 5: Snapshot de la red de interacciones en Bluesky entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana en la semana del 20 de enero de 2025. El tamaño de los nodos es proporcional a su centralidad de grado, mientras que el color indica la comunidad a la que han sido asignados por el algoritmo de detección de comunidades. Solo se han coloreado las 10 comunidades con mayor porcentaje de miembros. El color de los enlaces corresponde al del nodo de origen, el tamaño indica el número de interacciones entre los usuarios, y la curvatura señala la dirección de la interacción.

Por último, las tres cuentas que se mantienen más activas a lo largo del tiempo coinciden con aquellas que presentaban una mayor centralidad en la red: *eldiario.es*, *Carne Cruda* y *Sara Riveiro*, como era de esperar. Resulta especialmente interesante observar que, durante los picos de actividad, estas cuentas y sus comunidades interactúan tanto entre ellas como con usuarios pertenecientes a otras comunidades. En cambio, en los períodos de menor actividad, sus interacciones se concentran en un grupo más limitado de seguidores de su propia comunidad. Este patrón apoya la hipótesis de que los momentos de mayor actividad en la red pueden estar vinculados a eventos de especial relevancia que movilizan a un conjunto más amplio de usuarios en Bluesky, lo que a su vez fomenta que las publicaciones lleguen a un público cada vez más amplio. Este patrón de interacción temporal puede vincularse con la dinámica de la homofilia en la red. Durante los picos de gran actividad, la mayor interacción entre diferentes comunidades y actores sugiere que se favorece la conexión entre grupos diversos promoviendo un intercambio más amplio de ideas. En contraste, los períodos de menor actividad, con interacciones más

concentradas dentro de comunidades, podrían reflejar una tendencia hacia la homofilia, donde los usuarios tienden a interactuar preferentemente con perfiles con intereses más similares a los suyos. De esta manera, la evolución temporal de la red sugiere los eventos globales y el contexto sociopolítico pueden influir la apertura o el cierre de los espacios conversacionales.

4.4. Análisis de comunidades basadas en la optimización de la modularidad

La red de interacciones presenta una marcada estructura comunitaria, con un valor de modularidad de 0,77. Este valor indica una clara segmentación de la red de Bluesky en subgrupos de usuarios que están densamente conectados entre sí y tienen pocas conexiones con el resto de la red, un fenómeno común en redes sociales. En particular, el algoritmo de detección de comunidades de Gephi, basado en el cálculo de la modularidad, identifica un total de 33 comunidades, como se muestra en la Figura 6. Muchas de esas comunidades tienen un tamaño reducido, con menos de 200 usuarios, y se han representado de color gris para centrar la atención en las comunidades más grandes.

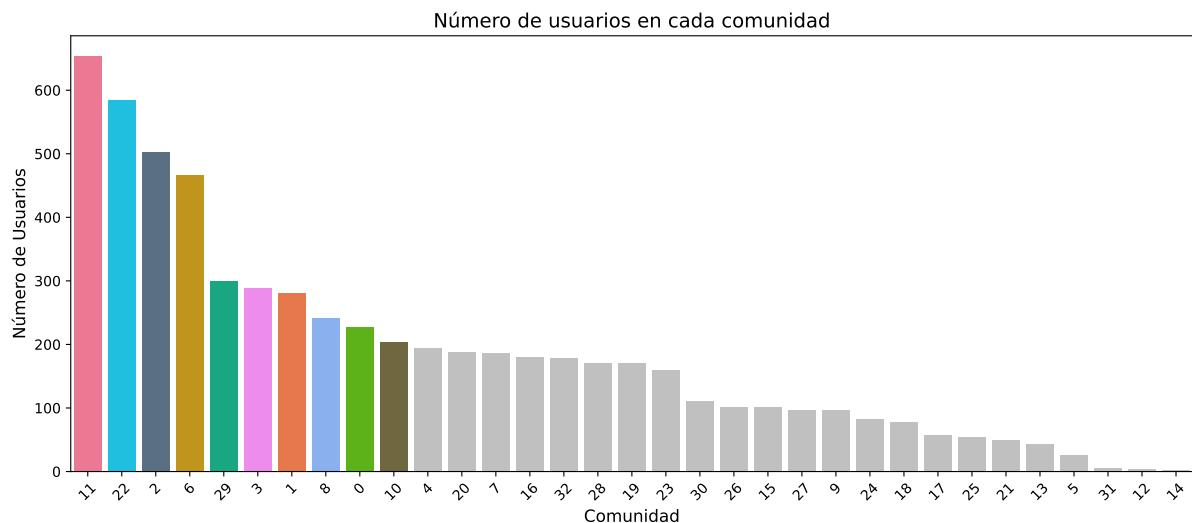


Figura 6: Número de usuarios por comunidad, según la partición obtenida mediante el algoritmo de detección de comunidades basado en modularidad. Cada barra representa una comunidad, y se han coloreado únicamente las diez comunidades con mayor número de usuarios.

En el grafo presentado en la Figura 2, se observaba que muchas de las comunidades identificadas están articuladas en torno a una cuenta central que concentra la mayoría de las interacciones. En la mayoría de los casos, el resto de los usuarios de la comunidad interactúan de forma exclusiva y unilateral con la cuenta central, sin que haya muchas interacciones entre ellos. Esta estructura radial sugiere que las comunidades se generan en torno a figuras influyentes, que actúan como polos de atracción que organizan la actividad de sus seguidores.

En particular, las cuatro comunidades de mayor tamaño están centradas en torno a cuentas con gran relevancia dentro de la red según las métricas de centralidad. En concreto, *eldiario.es*, Sara Riveiro, *Carne Cruda* y *atrupar*, asignados a las comunidades 11, 22, 2 y 6, respectivamente, parecen desempeñar un papel central en sus respectivos grupos, todos ellos superando los 400

usuarios. En el top 10 también destacan otras comunidades lideradas por figuras relevantes, principalmente periodistas y medios de comunicación con una fuerte presencia digital como Juanlu Sánchez (comunidad 3), *El Orden Mundial* (comunidad 8) o *El Mundo Today* (comunidad 0). Este patrón confirma que las cuentas con mayor visual e influencia dentro del ámbito periodístico estructuran la mayor parte de las dinámicas de interacción de la red.

En contraste, el pódcast *Punzadas Sonoras* aparece vinculado a un grupo mucho más reducido (comunidad 13), con solo 44 usuarios. Este resultado es coherente con su perfil, previamente analizado, ya que se trata de un pódcast que aborda temas muy especializados, lo que parece traducirse en una comunidad más pequeña y aislada. La presencia de comunidades periféricas de menor tamaño refleja la diversidad de perfiles en la red y pone en evidencia como los nichos temáticos pueden generar agrupaciones más reducidas, pero bien cohesionadas.

Por lo tanto, la modularidad parece ser una herramienta útil para detectar posibles agrupaciones de nodos similares estructuralmente, es decir, usuarios que interactúan con las mismas cuentas formando comunidades implícitas. Esta hipótesis se fundamenta en que el algoritmo de detección de comunidades optimiza la densidad de enlaces internos frente a externos, favoreciendo la formación de grupos con alta cohesión interna, por lo que sus miembros tenderán a compartir un mayor número de vecinos que con otras comunidades. Según las teorías de homofilia social e influencia, es más probable que se formen conexiones entre nodos similares que entre nodos distintos, lo que refuerza la idea de que la modularidad puede ayudar a identificar grupos de usuarios con características similares. No obstante, es posible que nodos con perfiles estructurales similares queden asignados a diferentes comunidades debido a la naturaleza del proceso de partición de la red. Asimismo, tampoco significa que todos los miembros de un grupo comparten necesariamente las mismas características, ya que la modularidad no mide directamente la similitud entre nodos. Por ese motivo, es necesario complementar este estudio con un análisis cuantitativo que permita cuantificar la similitud entre los pares de nodos pertenecientes a una misma comunidad. Este análisis se presenta en la sección 4.6.

4.5. Análisis de comunidades basadas en métricas de centralidad

El modelo de aprendizaje automático entrenado identifica tres clústeres de usuarios en base a sus métricas de centralidad. Estos grupos están desbalanceados en cuanto a su tamaño, ya que hay un primer clúster que contiene al 82,2% usuarios (clúster 1), un segundo con un 16,2% (clúster 0) y un último grupo más reducido que contiene el 1,6% restante (clúster 2).

En la Figura 7, se aprecia que los clústeres están razonablemente bien caracterizados en función de las métricas de centralidad, aunque existe cierto solape entre algunos grupos, especialmente entre el 0 y el 2. Este resultado es el esperado, al haber utilizado los valores de las métricas de centralidad normalizados para realizar el *clustering*, por lo que se puede afirmar que el modelo logra separar razonablemente bien los distintos grupos por su importancia y posición estratégica en la red. Consecuentemente, esta segmentación permite identificar usuarios con roles diferenciados dentro de la red, desde nodos periféricos hasta actores clave con gran influencia y capacidad de intermediación.

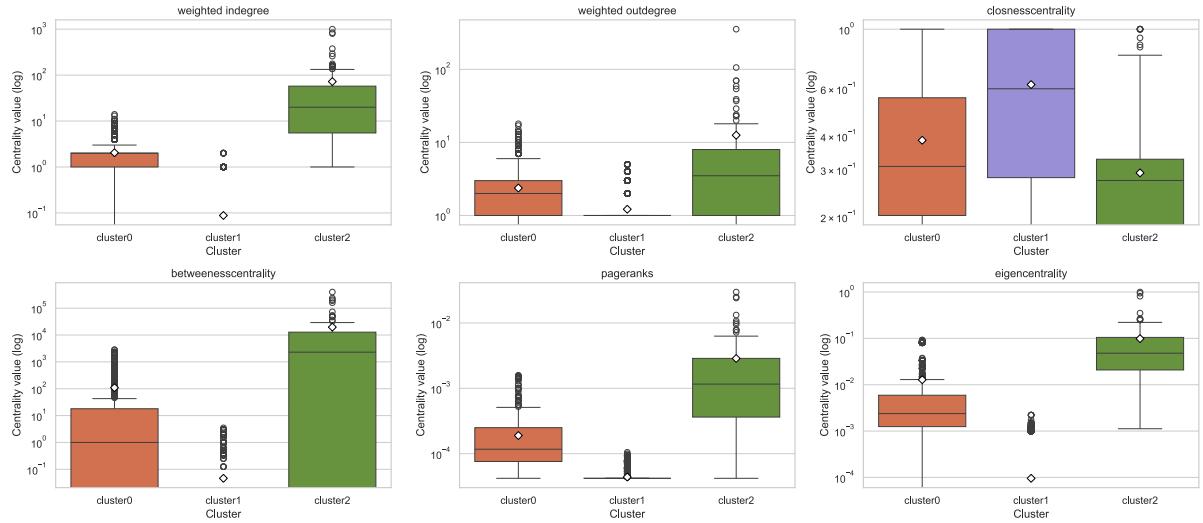


Figura 7: Distribución de los valores de las métricas de centralidad para cada clúster. Esta distribución se ha representado con diagramas de cajas y bigotes donde los valores medianos se indican con una línea negra en el interior de cada caja, los valores medios con un rombo blanco y los valores extremos con un círculo blanco.

El clúster 2, compuesto por apenas 100 usuarios, agrupa a las cuentas con mayor estatus dentro de la red, lo que se refleja en sus valores altos de centralidad en casi todas las métricas, excepto en la de cercanía. En contraste, el clúster 0 está formado por un mayor número de usuarios con valores intermedios, aunque más próximos a cero. Ambos clústeres presentan una notable variabilidad interna, especialmente el grupo de menor tamaño (clúster 2), que además cuenta con un número considerable de *outliers* con valores extremadamente altos en métricas como la centralidad de grado, intermedicación, vector propio o *PageRank*, en comparación con otros miembros de su propio grupo. Es probable que estos valores correspondan a las principales figuras destacadas observadas en la Figura 3, como *eldiario.es*, *Sara Riveiro* o *Carne Cruda*. Esta elevada heterogeneidad interna podría explicar el bajo valor del coeficiente de Silhouette obtenido para la partición generada, que se aproxima a 0.022, lo que sugiere que los clústeres no están claramente separados. A pesar de ello, no se ha logrado mejorar este resultado con un número distinto de clústeres.

Por el contrario, el clúster 1, a pesar de ser el de mayor tamaño, presenta menor variabilidad con la mayoría de las cuentas concentradas en valores próximos a cero, lo que sugiere una posición periférica de los nodos y una escasa influencia dentro de la red. Este patrón es coherente con la estructura típica de las redes sociales, donde existe una distribución desigual del estatus o centralidad con un pequeño número de cuentas que concentra la mayor visibilidad y capacidad de influencia, mientras que la mayoría de los usuarios ocupa posiciones periféricas, con conexiones limitadas y bajo alcance.

En el caso de la centralidad de cercanía, debido a los motivos expuestos en la sección 4.2, los patrones se invierten. Los usuarios más relevantes en la red presentan valores más bajos, ya que están conectados a un mayor número de nodos, lo que reduce su distancia media al resto de la red, mientras que los nodos más aislados presentan valores más altos debido a su limitada conectividad .

Para profundizar en el papel de los distintos grupos en la dinámica de la red, se han representado los clústeres identificados sobre el grafo de interacciones (Figura 8). Esta visualización permite explorar la estructura comunitaria de la red y los patrones de interacción entre los usuarios pertenecientes a los diferentes grupos. A través de esta representación, es posible identificar no solo la distribución espacial de los clústeres, sino también su grado de conexión con otros nodos, lo que aporta información relevante sobre su función.

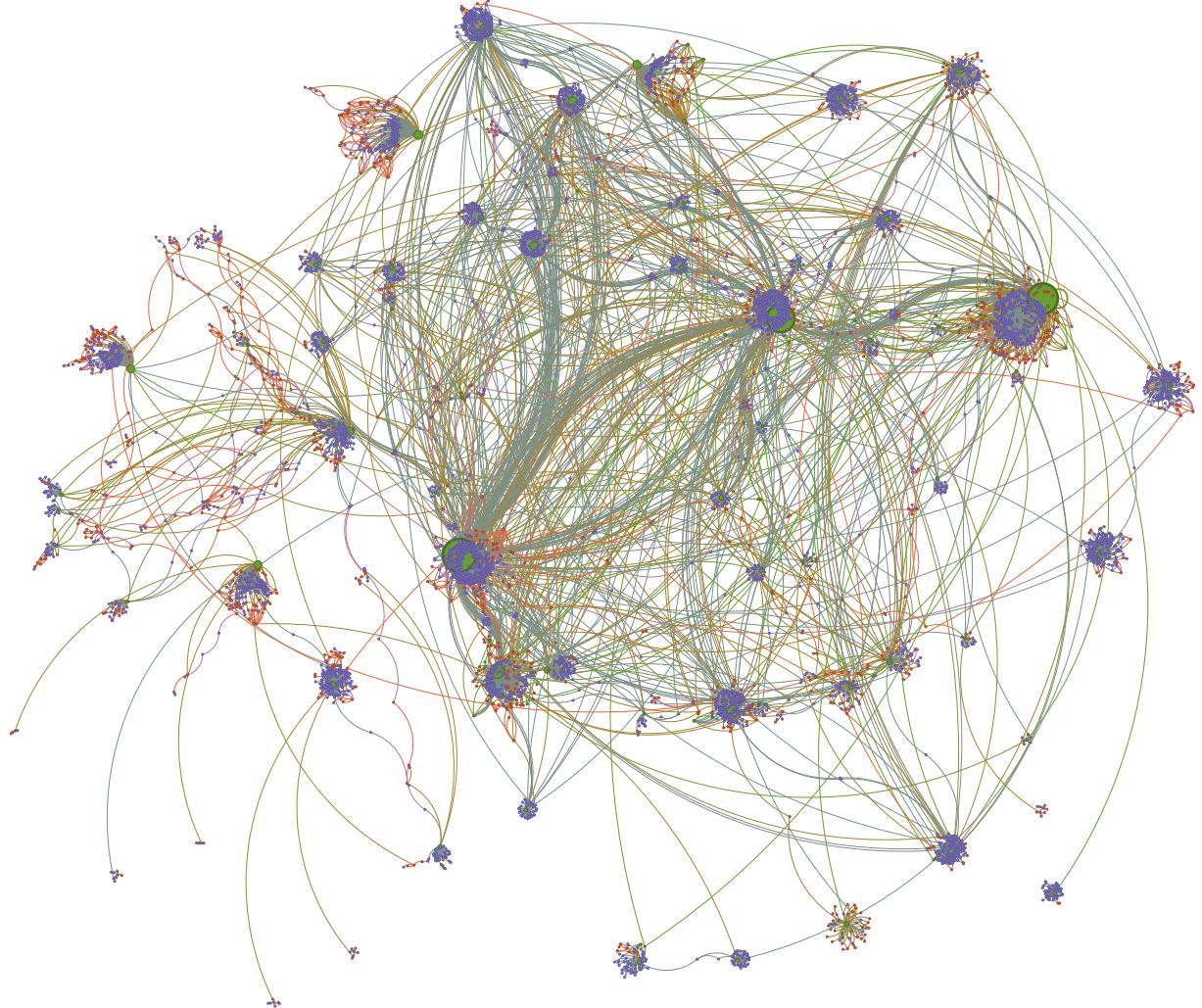


Figura 8: Red de interacciones en Bluesky entre usuarios vinculados a cinco podcasts socioculturales de habla hispana . El tamaño de los nodos es proporcional a su centralidad de grado, mientras que el color indica la comunidad a la que han sido asignados por el algoritmo de aprendizaje automático no supervisado: naranja para el clúster 0, morado para el 1 y verde para el 2. El color de los enlaces corresponde al del nodo de origen, el tamaño indica el número de interacciones entre los usuarios, y la curvatura señala la dirección de la interacción. Las etiquetas de los nodos no se muestran para facilitar la visualización de los clústeres.

En términos generales, la mayoría de las comunidades identificadas mediante el algoritmo de modularidad (véase Figura 2) presentan una estructura similar, con una composición heterogénea en la que conviven usuarios de los tres clústeres. Estas comunidades tienden a articularse en torno a un nodo central (representado en color verde), que actúa como eje principal de las interacciones tanto internas como externas. El mayor tamaño de estos nodos en comparación con el resto refleja sus altos valores de centralidad de grado. En su entorno inmediato, predominan

usuarios del clúster que presentaba niveles bajos de centralidad (color morado), cuya actividad se limita en muchos casos a una única interacción con el nodo central o alguno de sus vecinos, lo que explica sus valores cercanos a 0 para todas las métricas y la baja variabilidad de este grupo. A su vez, también se identifican usuarios con valores intermedios de centralidad (color naranja), que muestran un patrón de interacción más activo con un número más amplio de usuarios. Aunque sus interacciones se concentran principalmente dentro de sus propias comunidades, en ocasiones también establecen vínculos con usuarios externos, contribuyendo a favorecer la conectividad entre grupos.

Un caso curioso es el de *Punzadas Sonoras*, que forma una pequeña comunidad periférica en la parte inferior derecha del grafo de la Figura 8. A pesar de su número reducido de usuarios, la mayoría ha interactuado con más de un miembro del grupo y se han agrupado dentro del clúster intermedio (clúster 0). Este patrón sugiere que en el entorno del pódcast se ha creado un grupo de usuarios cohesionado con un alto grado de afinidad temática, lo que resulta coherente con la naturaleza del pódcast, que está centrado en un nicho muy específico. A pesar de su escasa visibilidad dentro de la red global, la estructura de esta comunidad sugiere que los intereses compartidos pueden generar vínculos estrechos que formen comunidades activas. Además, el hecho de que el pódcast no tenga tanto alcance en la sociedad puede favorecer que los usuarios tiendan a buscar refugio en redes sociales como Bluesky, para conectar con otras personas con intereses similares.

En conjunto, estos resultados ponen de manifiesto la estructura jerárquica y modular de la red, donde la posición de los usuarios parece estar estrechamente relacionada con su nivel de actividad e influencia. Esta estructura sugiere la posible existencia de homofilia en la red, ya que los usuarios tienden a agruparse dentro de comunidades bien caracterizadas que podrían sugerir intereses compartidos. Además, la presencia de nodos que enlazan distintos grupos sugiere la existencia de cuentas que facilitan la comunicación entre distintos conjuntos de usuarios. En la siguiente sección, se profundizará en el análisis de la homofilia en la red de forma cuantitativa.

4.6. Estudio de la homofilia

4.6.1. Semejanza estructural entre los usuarios de Bluesky

En esta sección, se estudia la presencia de homofilia en la red, entendida como la tendencia de los nodos a conectarse con otros que comparten características similares. En este caso, el análisis se centra en la equivalencia estructural entre los nodos, definida como la similitud en los patrones de conexión de los usuarios. Para ello, se ha calculado el coeficiente de correlación de Pearson entre pares de usuarios, siguiendo la metodología descrita en la sección 3.6. La matriz de correlación de Pearson se ha calculado sobre la matriz de adyacencia de la red, por lo que pares de nodos con un coeficiente de correlación alto tendrán un patrón similar de conexiones. Los resultados se muestran en la Figura 9.

Debido al elevado número de nodos, la matriz de correlaciones resulta difícil de interpretar, aunque es posible extraer algunas conclusiones relevantes. En primer lugar, la mayoría de las

parejas de usuarios presentan coeficientes de correlación próximos a cero, lo que indica una baja equivalencia estructural. Este patrón se justifica por la baja densidad de la red, su reducido grado medio y la fuerte estructura comunitaria observada previamente (véase la sección 4.1). No obstante, también se identifican pares usuarios con valores elevados de similitud estructural, superiores a 0,5 y, en la mayoría de los casos próximos a 1. Una posible explicación es que estos usuarios pertenezcan a una misma comunidad, donde los usuarios suelen estar conectados a una misma figura central, por lo que comparten vecindario.

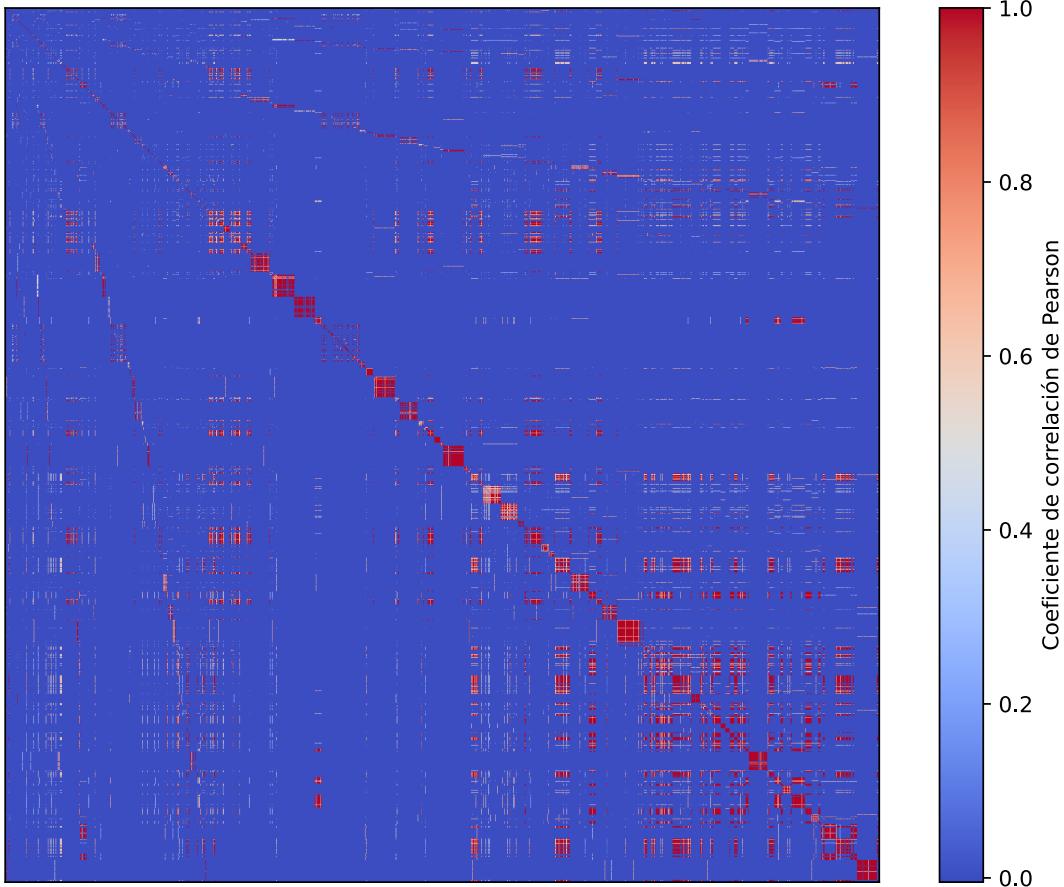


Figura 9: Similitud estructural entre los pares de usuarios de la red. Los valores de las celdas representan el coeficiente de correlación de Pearson calculado a partir de la matriz de adyacencia.

4.6.2. Semejanza estructural entre los miembros de una comunidad

Para contrastar la hipótesis anterior de que la similitud estructural es mayor entre los usuarios de una misma comunidad, se calculó la matriz de correlación de Pearson para cada una de las comunidades detectadas mediante el algoritmo de modularidad (Figura 10). Los resultados muestran de forma clara que, en términos generales, los usuarios pertenecientes a una misma comunidad presentan una elevada equivalencia estructural, lo que era de esperar, ya que el algoritmo de detección de comunidades se basa en optimizar la modularidad. De hecho, el test de la U de Mann-Whitney confirma que la similitud estructural dentro de las comunidades es significativamente mayor que entre grupos (Figura 11). No obstante, también se observan ciertas

variaciones internas que indican que no todos los pares de usuarios dentro de una misma comunidad son similares. Esto sugiere que pueden existir diferencias en los patrones de interacción que afecten a la dinámica interna de cada grupo.

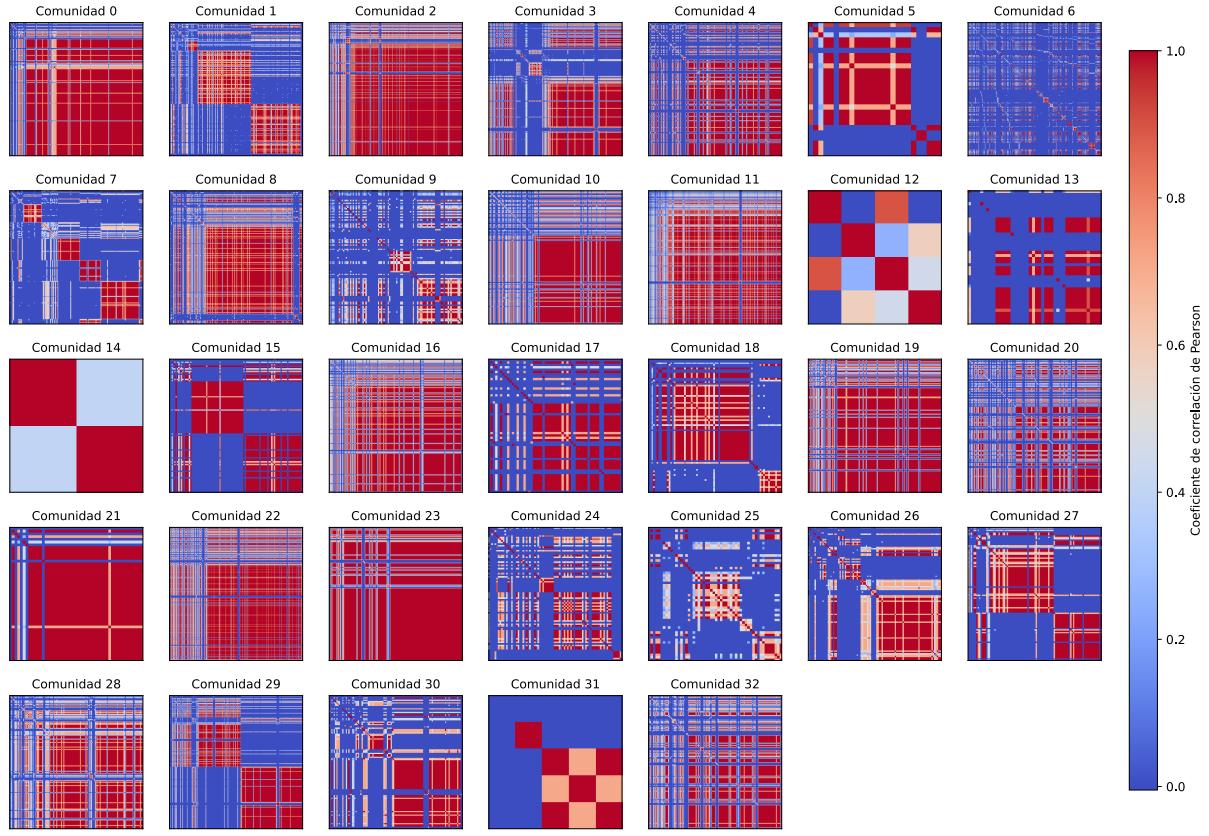


Figura 10: Similitud estructural entre los pares de usuarios de las comunidades identificadas por el algoritmo de detección de comunidades basado en modularidad. Los valores de las celdas representan el coeficiente de correlación de Pearson calculado a partir de la matriz de adyacencia.

Una de esas excepciones se observa en la comunidad 6, que es la que menor similitud estructural interna presenta, con una media inferior a 0,1, y cuya estructura era más heterogénea que la del resto de comunidades (véase Figura 2). A diferencia de otras comunidades, que presentan una estructura radial, con un número variable de usuarios posicionados alrededor de un nodo central, la comunidad 6 cuenta con múltiples focos de interacción interconectados. Uno de los nodos más relevantes de este grupo es la cuenta del periodista Aaron Rupar (atrupar.com), con el que interactúa un número elevado de usuarios de dicha comunidad, pero también existen otros núcleos periféricos que contribuyen a una mayor variabilidad estructural interna. Esta variabilidad se traduce en una menor semejanza entre los miembros de la comunidad, que tendrán distintos vecinos.

Las dos comunidades con mayor homogeneidad estructural interna son de tamaño reducido y están relativamente aisladas en la red, lo que probablemente favorezca esa mayor homogeneidad. En ambos casos, las comunidades se estructuran en torno a una figura central: en la comunidad 21, alrededor de *popbase.tv*, una cuenta dedicada a noticias de entretenimiento y cultura pop, y en la comunidad 23 en torno a Nadine Smith, una reconocida activista por los derechos LGBT. Entre las comunidades con mayor similitud interna también se encuentran comunidades de mayor

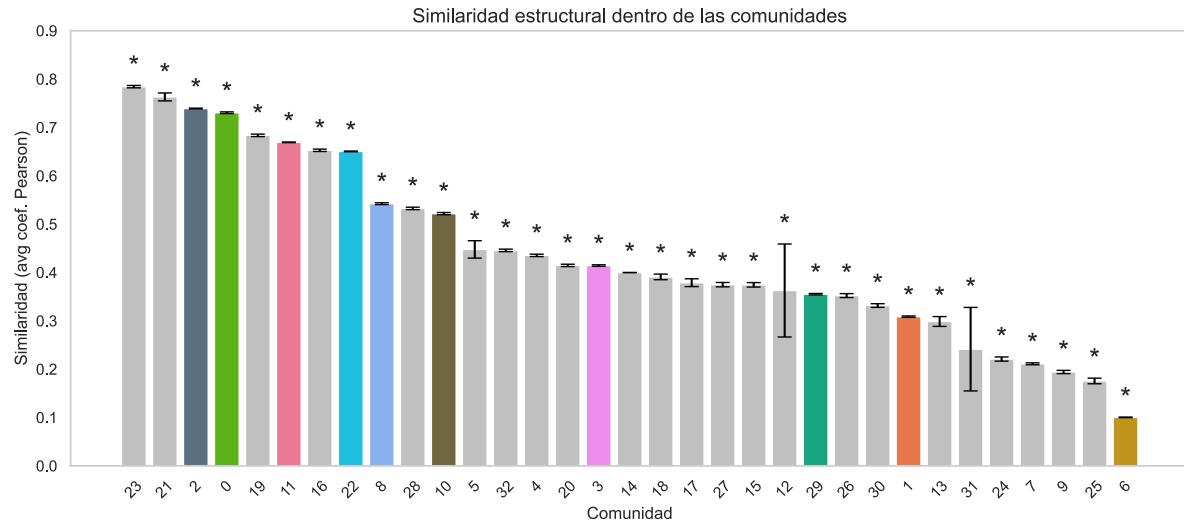


Figura 11: Similitud estructural entre los pares de usuarios asignados a las comunidades identificadas por el algoritmo de modularidad. Se han añadido barras de error que representan el error estándar de la media. Los asteriscos indican la presencia de diferencias estadísticamente significativas ($p < 0,05$) en comparación con la media entre clústeres, según la prueba U de Mann-Whitney.

tamaño como las de *Carne Cruda* (comunidad 2), *El Mundo Today* (comunidad 0), *eldiario.es*, Sara Riveiro (comunidad 22) o *El Orden Mundial* (comunidad 8). Este patrón podría explicarse por el hecho de que se trata de cuentas grandes y ampliamente reconocidas, que atraen a un mayor número de usuarios. Al concentrar un volumen elevado de interacciones, es más probable que los usuarios que siguen o interactúan con estas cuentas comparten vecinos en la red, lo que incrementa la similitud estructural dentro de sus respectivas comunidades. Todas ellas tienen una estructura bien definida con un conjunto de usuarios centrados alrededor de una figura central de gran importancia en la red y muchos de ellos solo interaccionan con dicha cuenta, por lo que serán estructuralmente equivalentes.

Por otro lado, se observa que, en algunas comunidades de tamaño reducido (e.g. comunidades 12 o 31), hay una gran variabilidad de la similitud estructural entre sus miembros, lo que sugiere una menor cohesión interna. Si se analiza la red, se puede observar que, a diferencia de otras comunidades, estas no presentan una estructura radial centrada en una figura central clara, lo que explica su mayor variabilidad estructural.

En conclusión, la semejanza estructural en la red parece ser más pronunciada en aquellas comunidades bien estructuradas, y con fuerte cohesión interna. Asimismo, los resultados indican que el algoritmo de detección de comunidades ha logrado una buena partición de la red, ya que logra identificar subconjuntos de nodos que están densamente conectados y son estructuralmente semejantes. Dentro de estas comunidades, los usuarios tienden a interactuar con cuentas comunes, lo que sugiere que sus miembros comparten intereses o inquietudes. Por lo tanto, estas comunidades implícitas podrían entenderse como espacios sociales definidos por afinidades temáticas o relaciones dentro de la red global. Esta hipótesis toma como base la teoría de la correlación social, según la cual las entidades sociales tienden a establecer vínculos con aquellos usuarios con los que comparten determinadas características o intereses. Por ejemplo, la comunidad 8,

centrada en torno a *El Orden Mundial* probablemente esté compuesta por usuarios interesados en la geopolítica y las relaciones internacionales, mientras que los de la comunidad 0, a la que pertenece *El Mundo Today*, se pueden agrupar en torno al humor satírico y la crítica social, lo que sugiere afinidades temáticas muy distintas. De hecho, estas dos comunidades se encuentran bastante separadas en el grafo.

4.6.3. Semejanza estructural entre los miembros de un clúster

En un segundo análisis, se evaluó la similitud estructural dentro de los clústeres identificados por un algoritmo de aprendizaje automático no supervisado (Figura 12). A diferencia del caso anterior, estos grupos muestran una baja equivalencia estructural interna, lo cual era esperable, ya que los clústeres agrupaban usuarios con roles similares dentro de la red, independientemente de su vecindario.

Como se podía apreciar en la Figura 8, dentro de una misma comunidad de usuarios pueden coexistir perfiles de nodos muy distintos, cada uno con un papel específico. Este patrón se puede interpretar a partir de la teoría del estatus social que plantea que, en redes sociales, los usuarios con menor estatus tienden a conectarse con nodos de mayor estatus. Así, estos usuarios menos influyentes tienden a compartir vecindario con otras cuentas de su misma comunidad a través de su vínculo con una figura central, pero no con cuentas de otras comunidades, por lo que presentarán una baja similitud estructural con otras cuentas de su mismo estatus en distintas comunidades.

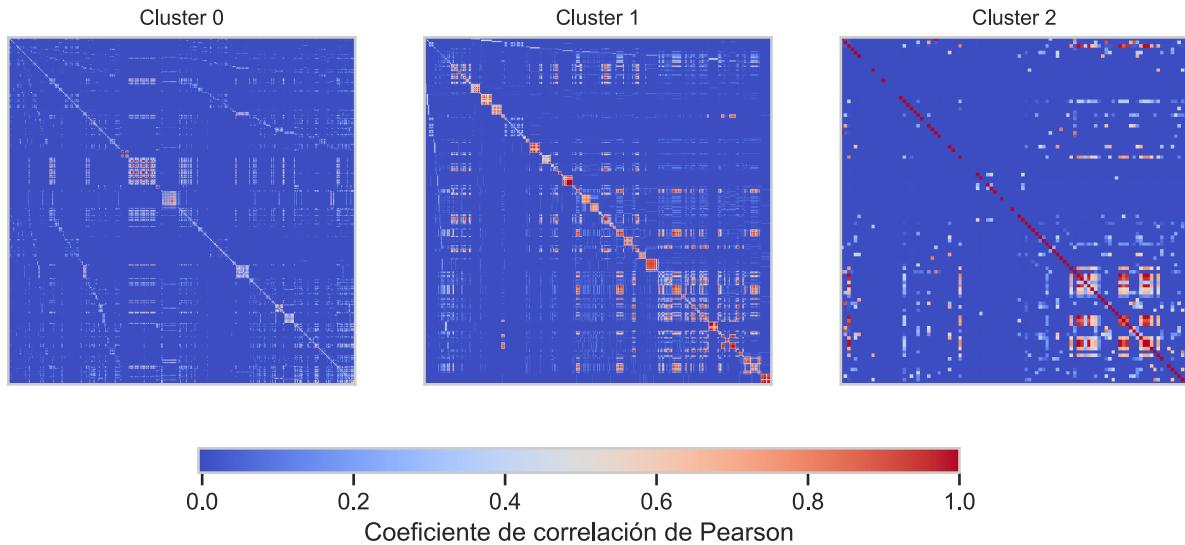


Figura 12: Similitud estructural entre los pares de usuarios de los clústeres identificados por el algoritmo de aprendizaje automático no supervisado. Los valores de las celdas representan el coeficiente de correlación de Pearson calculado a partir de la matriz de adyacencia.

De los tres clústeres identificados, el clúster 1 es el que mayor similitud estructural interna presenta, aunque su coeficiente de correlación de Pearson medio no supera el 0.05 (Figura 14a). Este clúster, el más numeroso con más de 5000 miembros, agrupa principalmente las cuentas de menor estatus que tienden a interactuar solo una vez con alguna figura influyente. Por este

motivo, las cuentas de este grupo tendrán una similitud estructural elevada con aquellas cuentas que estén conectadas a la misma figura influyente, pero no con el resto. Este patrón sugiere cierta homofilia temática, ya que las figuras centrales compartidas pueden representar intereses compartidos. Aunque las interacciones directas entre estos usuarios de menor estatus son escasas, el hecho de compartir un mismo referente podría actuar como un punto de convergencia que favorezca la formación de vínculos en el futuro, especialmente si la actividad en torno a dicha figura continúa o se intensifica.

Por su parte, el clúster 2, el más pequeño, agrupa las cuentas más importantes en la red según sus métricas de centralidad. Sus miembros mantienen múltiples interacciones, tanto dentro como fuera de sus comunidades, lo que implica vecindarios más diversos y, en consecuencia, a una menor similitud estructural entre ellos. Aun así, debido a que se ha considerado la intensidad de las relaciones y a que estas cuentas tienden a interactuar con otras cuentas de mayor estatus, sus miembros pueden compartir algunos vecinos, por lo que se pueden observar algunas parejas de cuentas semejantes.

En contraposición, el clúster 0, se caracteriza por una composición heterogénea y una baja conectividad tanto interna como externa, lo que se traduce en una menor similitud estructural entre sus miembros que tienden a no compartir vecindario ni siquiera con otros usuarios de su misma comunidad. Esto sugiere una baja homofilia tanto estructural como temática, ya que esos usuarios pueden mostrar intereses más variados y no necesariamente alineados con una única comunidad.

Estos resultados evidencian que, aunque la homofilia estructural dentro de los clústeres es en general baja, existen diferencias según la composición y el rol que desempeñan los nodos en cada clúster. Dado que los clústeres se generaron en base a métricas de centralidad, se espera que sus integrantes desempeñen funciones similares dentro de la red y tengan círculos sociales similares, aunque su vecindario sea distinto. En otras palabras, se espera que los clústeres presenten una mayor similitud funcional que estructural.

4.6.4. Homofilia funcional entre los miembros de un clúster

Para comprobar la hipótesis anterior, se calculó el coeficiente de correlación de Pearson entre los nodos de cada clúster, esta vez en base a las métricas de centralidad (Figura 13). En este caso, sí se observa una alta similitud dentro de los grupos, con un mayor coeficiente de correlación medio (Figura 14b), lo que confirma que los nodos agrupados comparten características relacionadas con su posición e influencia en la red, lo que a su vez sugiere que probablemente desempeñen un rol similar en sus respectivas comunidades. Estos resultados ponen de manifiesto la utilidad de los métodos de aprendizaje automático no supervisados para identificar nodos funcionalmente semejantes y detectar figuras centrales en la red.

Desde el punto de vista funcional, el clúster 1 vuelve a destacar por presentar la mayor similitud interna, en este caso reflejando que sus miembros comparten valores muy similares en las métricas de centralidad, con escasa variabilidad interna. Esto evidencia que las cuentas de este

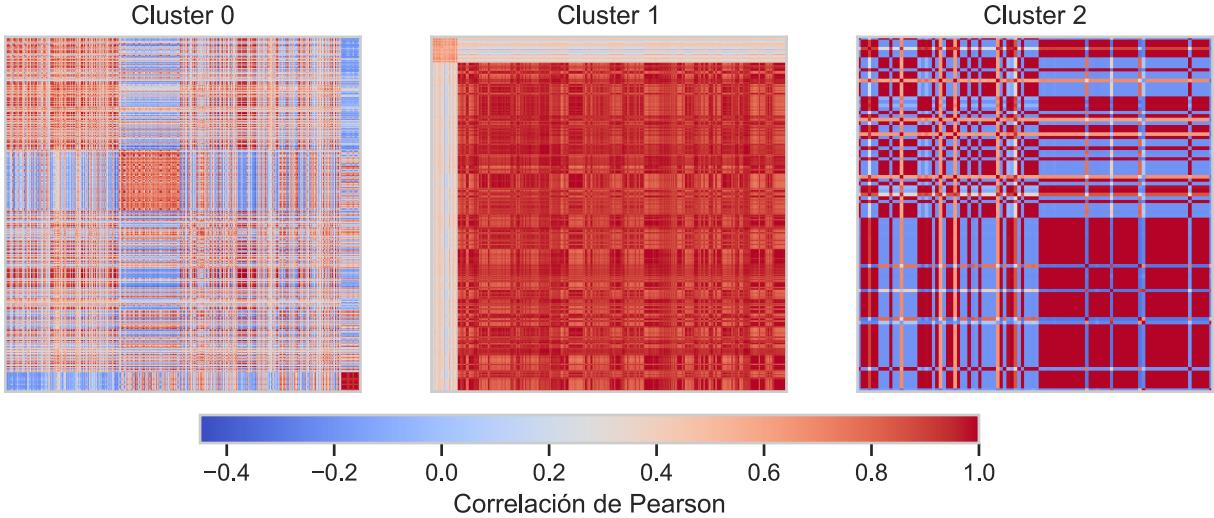
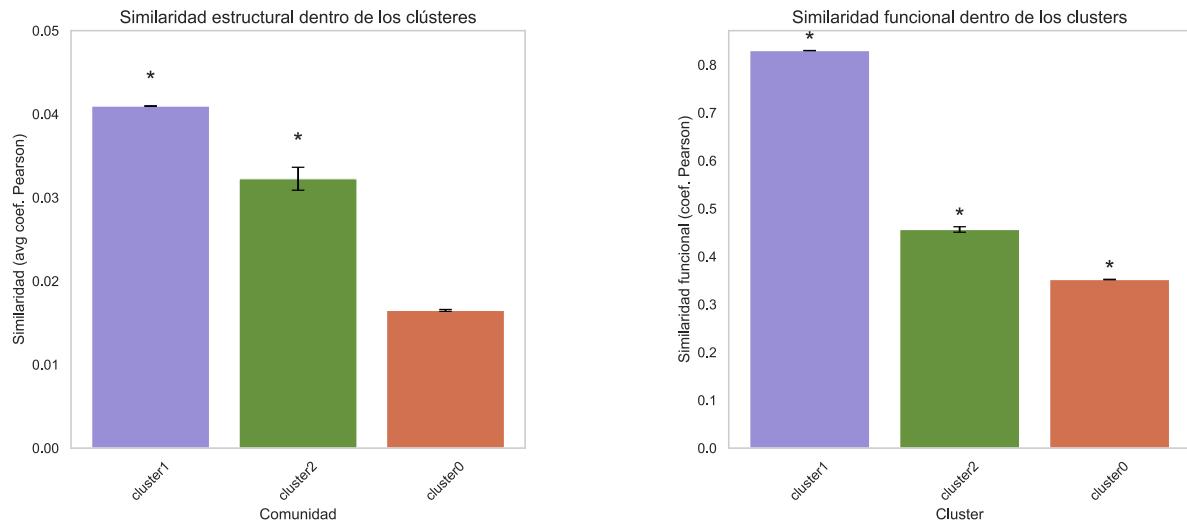


Figura 13: Similitud funcional entre los pares de usuarios de los clústeres identificados por el algoritmo de aprendizaje automático no supervisado. Los valores de las celdas representan el coeficiente de correlación de Pearson calculado a partir de las métricas de centralidad.

clúster tienden a desempeñar roles similares dentro de la red, principalmente interactuando casi exclusivamente con la figura central de su respectiva comunidad, lo que las sitúa en roles periféricos muy similares entre sí. En cambio, las cuentas de estatus intermedio, al tener una mayor variedad de conexiones, tanto hacia figuras centrales como hacia otros nodos, desempeñan roles más variados y complejos, lo que reduce la homogeneidad funcional dentro de su grupo.

En un punto intermedio se encuentra el clúster 2, aunque con un coeficiente de correlación medio que es prácticamente la mitad que el del clúster 1, reflejando una mayor diversidad en los roles funcionales dentro del grupo. Si bien este clúster agrupa a las figuras más influyentes en la red, no todas ocupan posiciones de igual peso, lo que se traduce en diferencias notables en sus métricas de centralidad. Por ejemplo, no tienen el mismo estatus *elDiario.es*, un medio consolidado con una gran audiencia y conexiones frecuentes con otras cuentas influyentes, que *Punzadas*, un podcast de alcance más limitado y con una comunidad reducida relativamente aislada del resto de la red, aunque dentro de su comunidad también ocupe un papel central. Esta disparidad funcional entre los miembros del mismo clúster explica la menor homogeneidad observada en relación a sus métricas de centralidad.

La mayor similitud estructural que funcional dentro de los clústeres demuestran la existencia de equivalencia regular. Aunque no comparten vecinos, los miembros de un clúster tienen círculos similares debido a su estatus similar. Por ejemplo, los actores principales de la red según las métricas de centralidad suelen tener círculos formados por usuarios de bajo de estatus, aunque también pueden conectarse con cuentas de su mismo nivel. Por el contrario, los usuarios de bajo estatus suelen estar conectados a una única cuenta de mayor estatus, mientras que los usuarios del clúster intermedio suelen presentar círculos sociales más heterogéneos.



(a) Similitud estructural media entre los pares de usuarios pertenecientes a cada clúster.

(b) Similitud estructural media entre los pares de usuarios pertenecientes a cada clúster.

Figura 14: Similitud entre los pares de usuarios pertenecientes a los clústeres identificados por el algoritmo de aprendizaje no supervisado en base a métricas de centralidad. Se han añadido barras de error que representan el error estándar de la media. Los asteriscos indican la presencia de diferencias estadísticamente significativas ($p < 0,05$) en comparación con la media entre clústeres, según la prueba U de Mann-Whitney.

4.6.5. Análisis comparativo de la homofilia

En el estudio de redes sociales, que suelen estar compuestas por miles de usuarios e interacciones entre ellos, el análisis de la red y, en particular, de la homofilia puede volverse especialmente complejo debido al gran volumen y densidad de datos, que dificulta la interpretación de los resultados. Si bien el análisis de la red en su conjunto permite extraer conclusiones generales acerca de la semejanza de los usuarios de la red, no permite identificar usuarios semejantes de manera sencilla ni tampoco analizar si estos se encuentran conectados.

Ante esta dificultad, la aplicación de métodos de segmentación previos, como puede ser la detección de comunidades que optimicen la modularidad o de agrupaciones de nodos que desempeñen funciones similares, se presenta como una estrategia útil para simplificar el análisis y facilitar el descubrimiento e interpretación de las dinámicas internas. Estas técnicas permiten identificar subgrupos de usuarios similares, lo que a su vez posibilita un análisis más detallado de los patrones de similitud y conexión entre los usuarios.

Si consideramos la assortatividad como una mayor tendencia de los nodos a conectarse con otros similares —y obviamente fenómenos como la influencia social o factores externos—, esta puede emplearse como una medida indirecta de la homofilia en la red. A su vez, la modularidad permite identificar comunidades de individuos que interactúan más entre sí que con otros grupos, por lo que comparten un mayor número de vecinos, como se confirma mediante el coeficiente de correlación de Pearson. Esta proximidad estructural sugiere que los usuarios pueden tener intereses comunes al interactuar con las mismas cuentas, aumentando también la probabilidad de que se relacionen entre sí en el futuro. Así, la formación de estas comunidades compactas refleja

un patrón de homofilia estructural, donde los usuarios con características o comportamientos similares tienden a conectarse preferentemente, aunque no siempre de forma directa. Estas comunidades suelen estructurarse en torno a cuentas de medios de comunicación, podcasts u otras figuras influyentes del ámbito sociocultural, que actúan como *hubs* temáticos dentro de la red, ya que la mayoría de los usuarios de la comunidad interactúan con ellos aunque no se relacionen entre sí. No obstante, este enfoque tiene limitaciones: al centrarse únicamente en la estructura de conexiones, no capta completamente las motivaciones subyacentes o afinidades temáticas reales entre los usuarios, que solo se podrían tratar de deducir estudiando el ámbito específico en el que se mueven las figuras centrales o realizando un análisis semántico de las publicaciones de estos usuarios. Por ejemplo, *Punzadas Sonoras* es un podcast especializado en filosofía, por lo que es de esperar que muchos de los usuarios que interactúan con ellas comparten un interés por dicho tema.

En contraste, el algoritmo de *clustering* basado en métricas de centralidad, no parece ser el método más adecuado para estudiar la assortatividad de la red, ya que mientras que en la detección de comunidades se consideran los enlaces entre individuos, en el clustering los usuarios no están embebidos en una red. De hecho, los miembros de un mismo clúster no suelen estar directamente conectados entre sí, sino que las conexiones parecen establecerse mayoritariamente entre usuarios pertenecientes a distintos clústeres, es decir, entre cuentas con distinto estatus. En concreto, la mayoría de las interacciones son unidireccionales, desde cuentas de menor a mayor estatus. Aun así, este enfoque permite identificar usuarios semejantes de acuerdo al papel que desempeñan en la red, lo que revela patrones valiosos sobre la estructura y la dinámica de la red. Por ejemplo, muestra que las cuentas de menor estatus tienden a vincularse con figuras de mayor estatus, generalmente de forma unidireccional, mientras que estas últimas también interactúan entre sí, promoviendo el intercambio de información entre distintos grupos de usuarios. Este comportamiento puede favorecer que se formen nuevas conexiones entre usuarios pertenecientes a comunidades distintas.

Por lo tanto, ambas aproximaciones permiten agrupar nodos semejantes partiendo de distintos criterios. Mientras que el algoritmo de modularidad tiende a formar comunidades de nodos con mayor similitud estructural, el algoritmo de clustering agrupa nodos que desempeñan roles funcionales similares dentro de la red. En este sentido, el algoritmo de modularidad resulta más adecuado para el estudio de la assortatividad, ya que revela grupos más compactos en los que los nodos tienden a compartir vecindario. Esta proximidad estructural facilita que los usuarios tengan intereses comunes al interactuar con las mismas cuentas, aumentando la probabilidad de que se relacionen entre sí. Así, la formación de estas comunidades compactas refleja un patrón de homofilia estructural, donde los usuarios con características o comportamientos similares tienden a conectarse preferentemente, reforzando la cohesión y la segregación dentro de la red.

Es necesario resaltar que, en este estudio, se ha llevado a cabo una estimación aproximada de la homofilia de la red utilizando las comunidades detectadas por el algoritmo de modularidad, dado que en estas comunidades los usuarios están más interconectados entre sí que con los de otros grupos. Sin embargo, para obtener una medida más precisa de la assortatividad de la red, habría sido necesario evaluar la similitud exclusivamente entre pares de nodos directamente conectados.

Esta aproximación fue seleccionada debido a que la red analizada representa solo una muestra del ecosistema digital de Bluesky, enfocándose exclusivamente en las interacciones alrededor de cinco cuentas de pódcast que poseen un determinado estatus social y funcionan como puntos focales dentro de sus respectivas comunidades. Como consecuencia, es posible que no se hayan capturado las interacciones entre usuarios de menor estatus que hubieran permitido profundizar en el análisis de la homofilia.

5. Conclusiones

El análisis realizado permite afirmar que la homofilia es un fenómeno que estructura las dinámicas de interacción entre usuarios de Bluesky en torno a pódcasts socioculturales y políticos en español. A través de la construcción y estudio de una red de más de seis mil usuarios conectados directa o indirectamente con cinco cuentas de pódcast, se ha evidenciado que las comunidades digitales tienden a organizarse en función de afinidades estructurales y funcionales, confirmando así una fuerte orientación hacia la conexión entre usuarios semejantes.

Metodológicamente, la combinación de análisis estático y dinámico de redes, algoritmos de detección de comunidades, medidas de centralidad y técnicas de aprendizaje automático no supervisado ha proporcionado una perspectiva multidimensional de la red. Este análisis reveló que la mayoría de los usuarios de menor estatus se agrupan alrededor de cuentas de mayor estatus, como las de reconocidos periodistas, pódcasts o medios de comunicación, incluyendo *eldiario.es*, *Carne Cruda* y Sara Riveiro. Adicionalmente, la aplicación del algoritmo de detección de comunidades, basado en la optimización de la modularidad, desveló una estructura comunitaria bien definida, donde los usuarios mantienen una alta densidad de conexiones internas y muestran poca interacción con nodos de otras comunidades. En particular, se observó que las comunidades centradas en nichos temáticos o cuentas menos conocidas tienden a estar más aisladas dentro de la red. Por el contrario, las cuentas de mayor estatus presentan un público más amplio y están más interconectadas, desempeñando un papel central en la difusión de contenidos. Finalmente, el uso de clustering basado en métricas de centralidad permitió identificar usuarios con roles y círculos sociales similares.

Este estudio pone de relieve el potencial de las plataformas digitales emergentes, como Bluesky, y las técnicas de análisis de redes sociales para investigar fenómenos sociales complejos. A medida que estas plataformas continúen creciendo y atrayendo más usuarios, será esencial seguir explorando la configuración de las comunidades y el impacto de estas estructuras en la circulación de información, la participación política y la construcción de identidades colectivas. Sin embargo, este análisis no permite cuantificar directamente la homofilia, ya que solo se han considerado las interacciones relacionadas con los pódcasts seleccionados y sus comunidades, sin abarcar todas las conexiones entre los usuarios de la red. Por lo tanto, puede haber relaciones entre usuarios con similitudes estructurales que no se reflejan en este análisis. Además, la mayoría de los usuarios estructuralmente similares no están conectados directamente entre sí, sino a través de una figura central que actúa como un hub temático. Aun así, este enfoque permite identificar usuarios similares desde una perspectiva estructural y funcional, que también podrían compartir

intereses afines o desempeñar roles similares en la red, facilitando una estimación de la homofilia de la misma.

Para futuras líneas de investigación, sería interesante ampliar el análisis incorporando todas las interacciones posibles dentro de la red, con el objetivo de obtener una medida más completa y precisa de la homofilia, basada únicamente en usuarios conectados directamente. Según la teoría de la correlación social, usuarios con similitudes podrían establecer conexiones directas en el futuro debido a intereses compartidos. Esta información podría ser valiosa para diseñar sistemas de recomendación de usuarios o contenidos en plataformas sociales como Bluesky, facilitando la creación de vínculos basados en intereses comunes. Además, se podría complementar el estudio con información sobre los intereses de los usuarios, lo que permitiría validar las afinidades estructurales observadas con afinidades basadas en intereses compartidos. Esto podría lograrse mediante un análisis temático de las publicaciones y mensajes intercambiados, extrayendo las palabras más frecuentes en cada comunidad y evaluando si están relacionadas con un tema específico y coherente.

6. Referencias

- Arroyo-Machado, W., Robinson-Garcia, N., & Torres-Salinas, D. (2024). Are there stars in Bluesky? A comparative exploratory analysis of altmetric mentions between X and Bluesky. <https://doi.org/10.5281/ZENODO.14290752>
- Basov, N. (2020). The ambivalence of cultural homophily: Field positions, semantic similarities, and social network ties in creative collectives. *Poetics*, 78, 101353.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An Open Source Software for Exploring and Manipulating Networks*. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Bird, S. E. (2011). Are we all producers now? *Cultural Studies*, 25(4-5), 502-516.
- Bisgin, H., Agarwal, N., & Xu, X. (2012). A study of homophily on social media. *World Wide Web*, 15(2), 213-232.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Broersma, M., & Eldridge II, S. A. (2019). Journalism and social media: redistribution of power? *Media and Communication*, 7(1), 193-197. <https://doi.org/https://doi.org/10.17645/mac.v7i1.2048>
- Chen, Y.-H., & Keng, C.-J. (2024). Understanding audience willingness to keep subscriptions in digital podcast: A dual identification perspective. *Journal of Marketing Communications*, 30(8), 926-947. <https://doi.org/10.1080/13527266.2023.2193191>
- Curiel, E. H. (2011). El periodismo en el siglo de las redes sociales. *Vivat Academia*, 1113-1128.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22. <https://doi.org/https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>

- Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques".
- Gruzd, A., & Mai, P. (2025). Communalytic: A no-code computational social science research tool for studying online communities and public discourse on social media [Accessed: 01-05-2025].
- Ibrahem Hafez, A., Hassanien, A. e., Fahmy, A. A., & Tolba, M. (2013). Community detection in social networks by using Bayesian network and Expectation Maximization technique. *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, 209-214. [http://doi.org/10.1109/HIS.2013.6920484](https://doi.org/10.1109/HIS.2013.6920484)
- Ito, M., Martin, C., Pfister, R. C., Rafalow, M. H., Salen, K., & Wortman, A. (2018). *Affinity online: How connection and shared interest fuel learning*. New York University Press.
- Ittefaq, M. (2025). Digital disconnective practice: Online platform migration and technology non-use in the age of emerging social media and polarized societies. *Media, Culture & Society*, 47(4), 815-823. <https://doi.org/10.1177/01634437251326397>
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5), 604-632. <https://doi.org/10.1145/324133.324140>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60.
- Marcos-García, S., Alonso-Muñoz, L., & López-Meri, A. (2021). Periodismo y nuevas narrativas. Storytelling como formato de difusión informativa en redes sociales. *Estudios sobre el mensaje periodístico*, 27(2), 553-567.
- McGregor, H. (2022, junio). Podcast Studies. <https://doi.org/10.1093/acrefore/9780190201098.013.1338>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. (Technical Report N.º 1999-66) (Previous number = SIDL-WP-1999-0120). Stanford InfoLab. Stanford InfoLab. <http://ilpubs.stanford.edu:8090/422/>
- Pascual, M. G. (2024). *Más allá de The Guardian y La Vanguardia: Elon Musk y Donald Trump precipitan la migración de usuarios de X a Threads y Bluesky*. Consultado el 15 de noviembre de 2024, desde <https://www.lanacion.com.ar/tecnologia/mas-allá-de-the-guardian-y-la-vanguardia-elon-musk-y-donald-trump-precipitan-la-migración-de-nid15112024/>
- Pearson, K., & Galton, F. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347-352), 240-242. <https://doi.org/10.1098/rspl.1895.0041>
- Poongundranar, T. (2023). Wars of Position: A Social Network Analysis of Political Podcasts. <https://doi.org/https://doi.org/10.6082/uchicago.6173>
- Tarjan, R. (1972). Depth-First Search and Linear Graph Algorithms. *SIAM Journal on Computing*, 1(2), 146-160. <https://doi.org/10.1137/0201010>
- Wendland, J. (2024). Building a Better Participatory Culture and Enhancing Sense of Community in Podcasts – Systematic Literature Review. *Journal of Radio & Audio Media*, 1-23. <https://doi.org/10.1080/19376529.2024.2347609>

7. Anexo

7.1. Material suplementario

Además de la memoria principal, se proporcionan los siguientes archivos suplementarios para facilitar la reproducción y comprensión del análisis realizado:

- **Datos:** archivos CSV con las publicaciones de Bluesky obtenidas desde Communalytic. También se proporciona el archivo GEFX que representa la red en formato de grafo, el proyecto completo de Gephi, y los archivos CSV con las aristas y los nodos.
- **Código del Análisis:** Jupyter notebook con el código utilizado para el análisis exploratorio y la evaluación de la homofilia en la red. También se ha incluido el archivo CSV que contiene la paleta de colores utilizada para representar las comunidades detectadas.
- **Resultados del Clustering:** resultados del clustering realizado en WEKA.
- **Animación de la Red Dinámica:** animación en formato MP4 que ilustra la evolución temporal de la red.

7.2. Métricas de centralidad

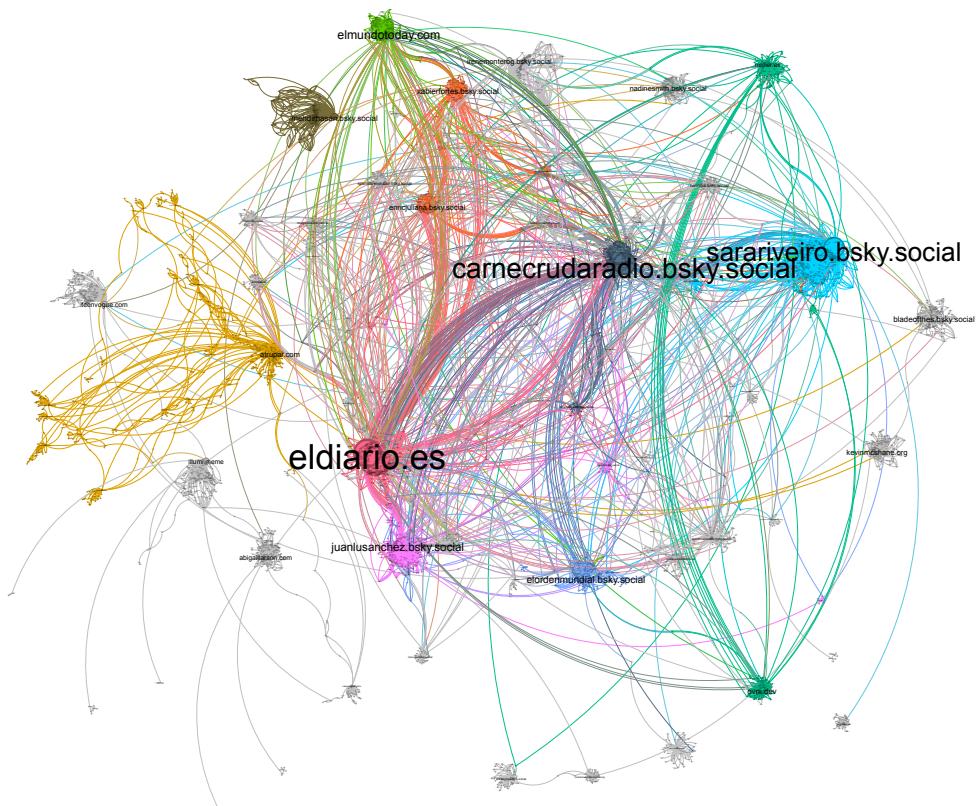


Figura 15: Red de interacciones entre usuarios vinculados a cinco podcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado de entrada.

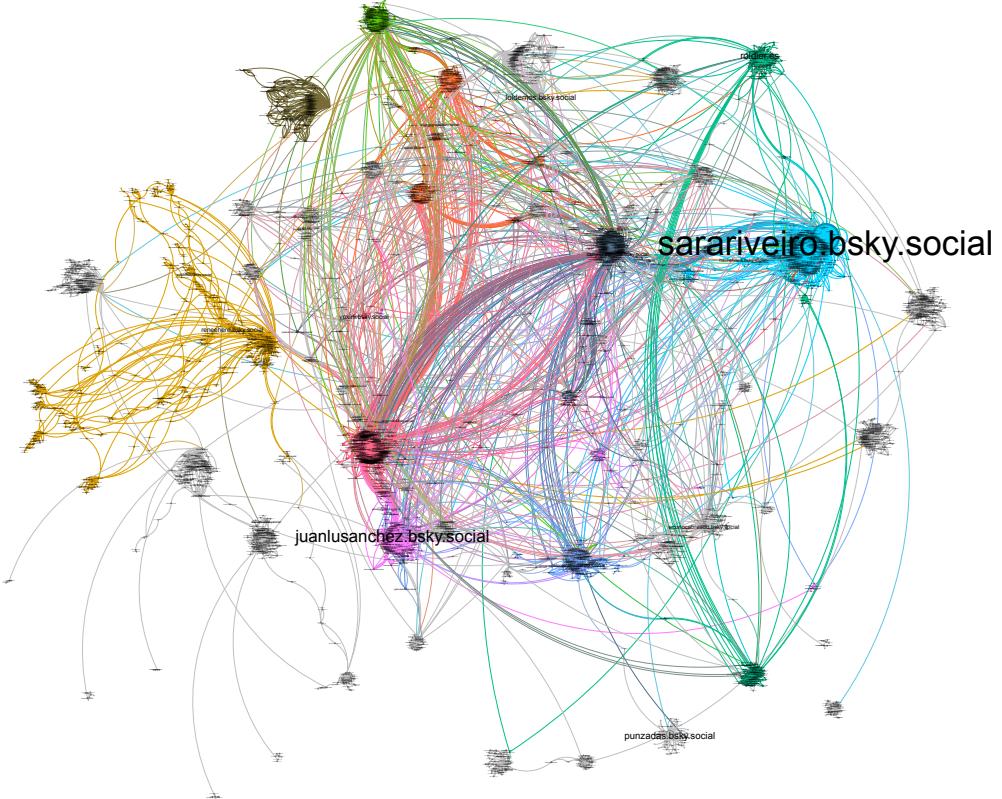


Figura 16: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado de salida.

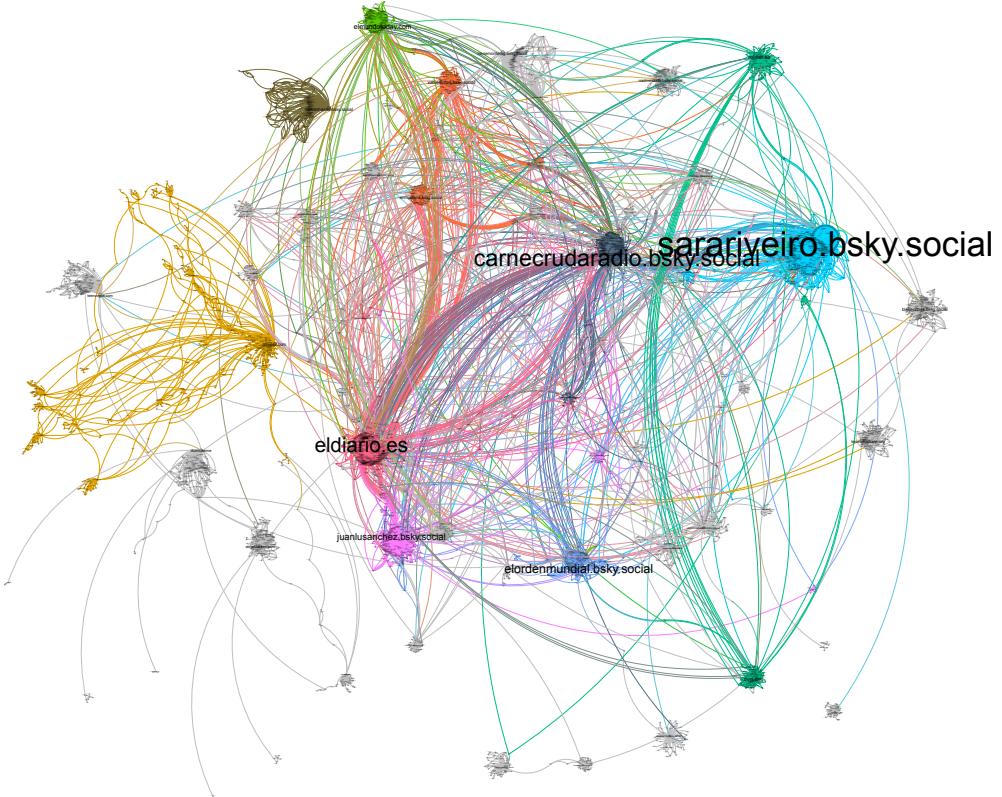


Figura 17: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado ponderado.

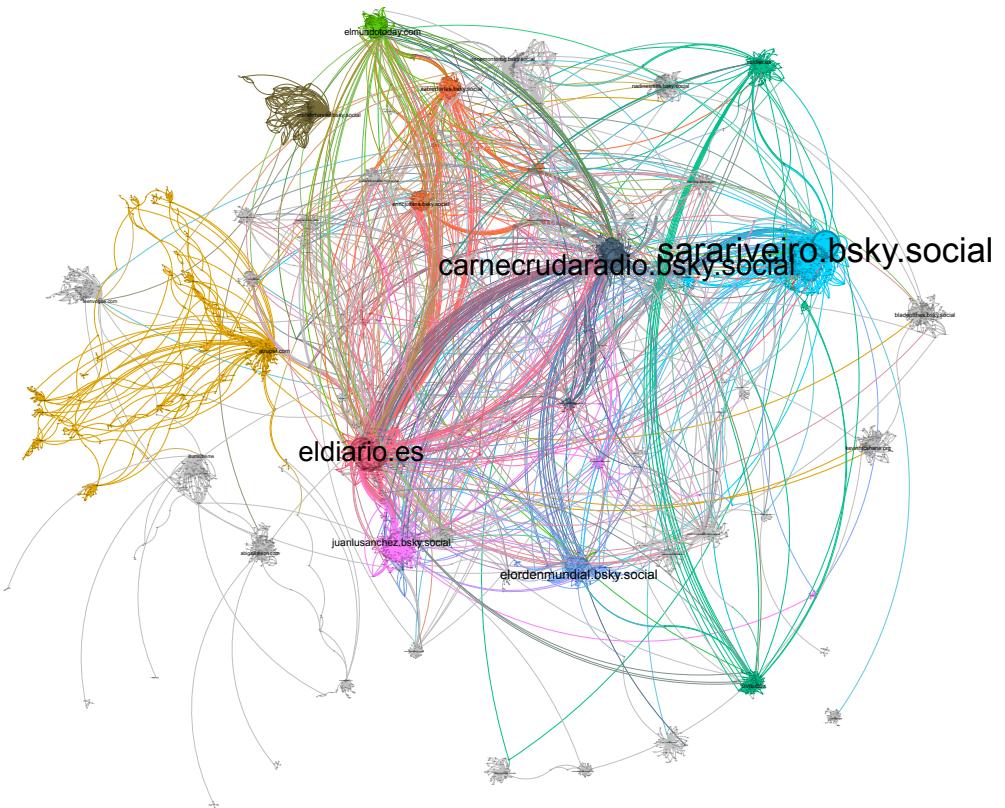


Figura 18: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado de entrada ponderado.

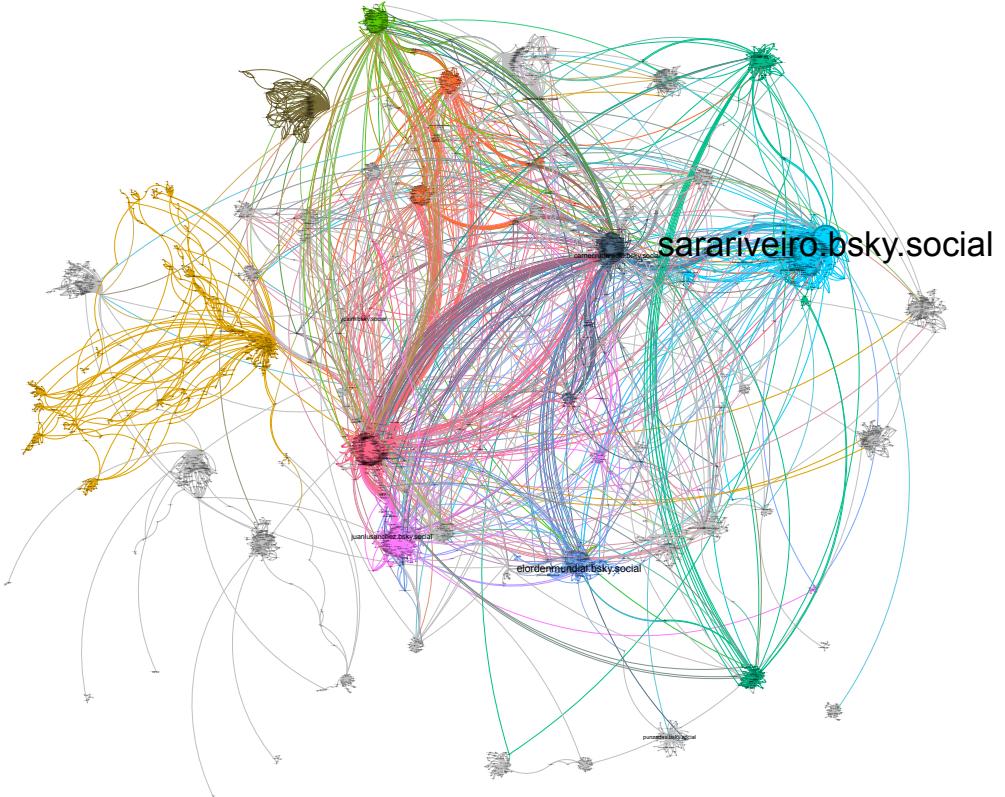


Figura 19: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de grado de salida ponderado.

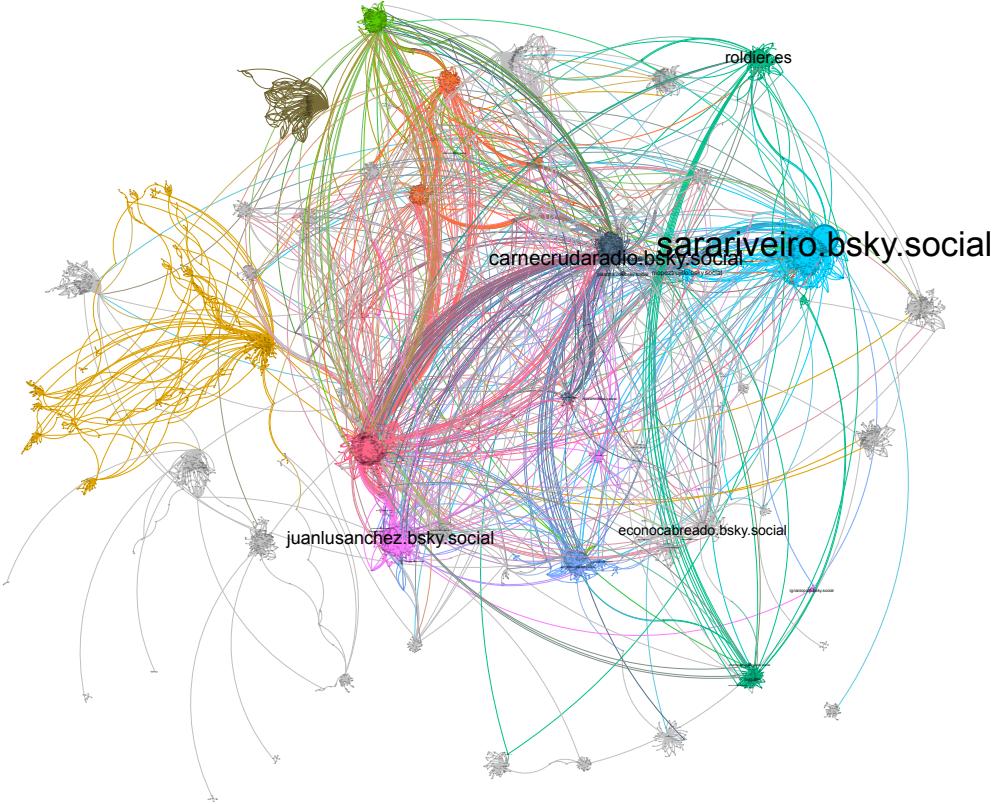


Figura 20: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de intermediación.

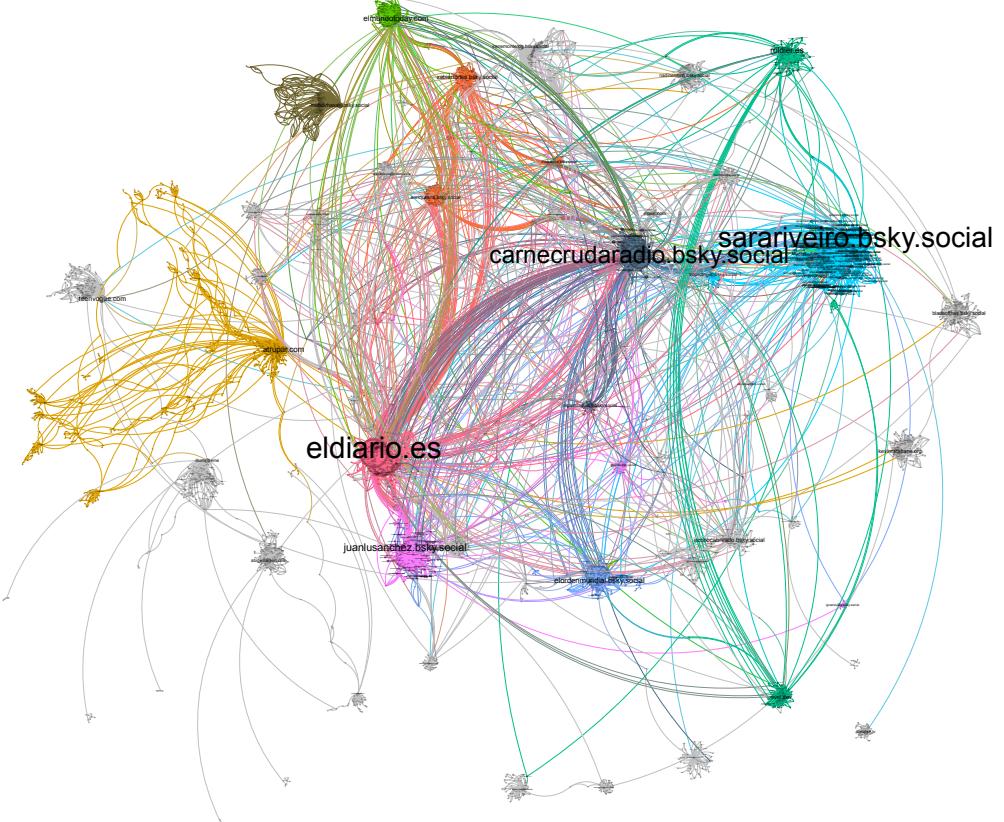


Figura 21: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad del vector propio.

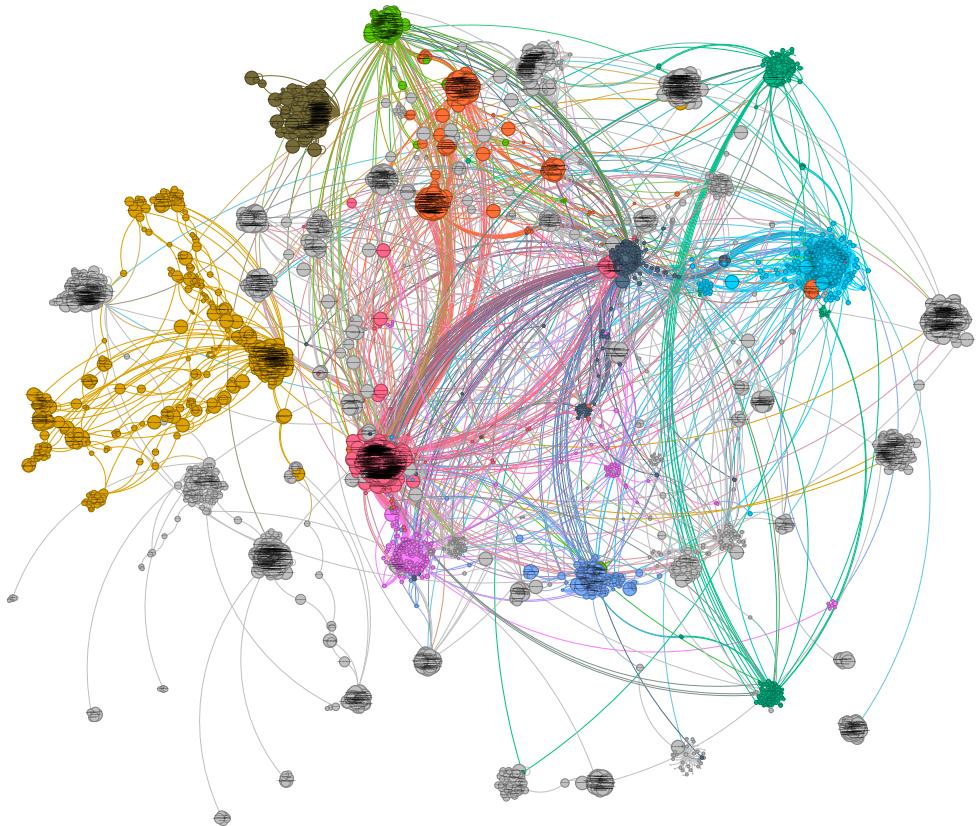


Figura 22: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su centralidad de cercanía.

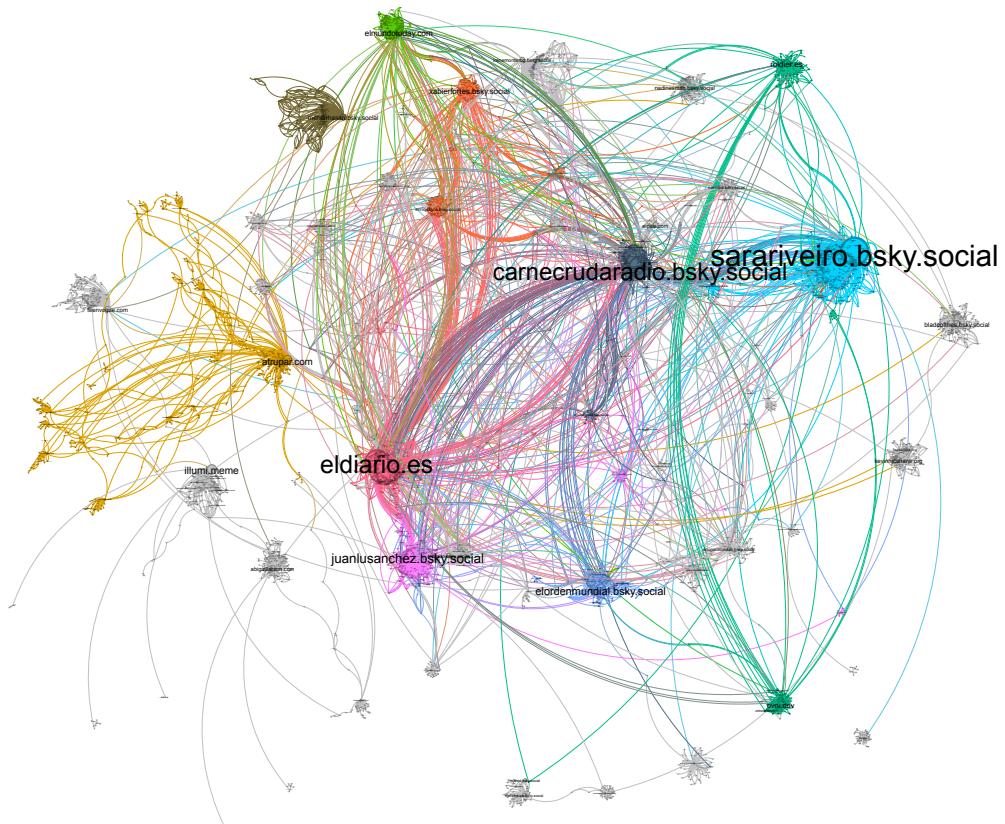


Figura 23: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su *PageRank*.

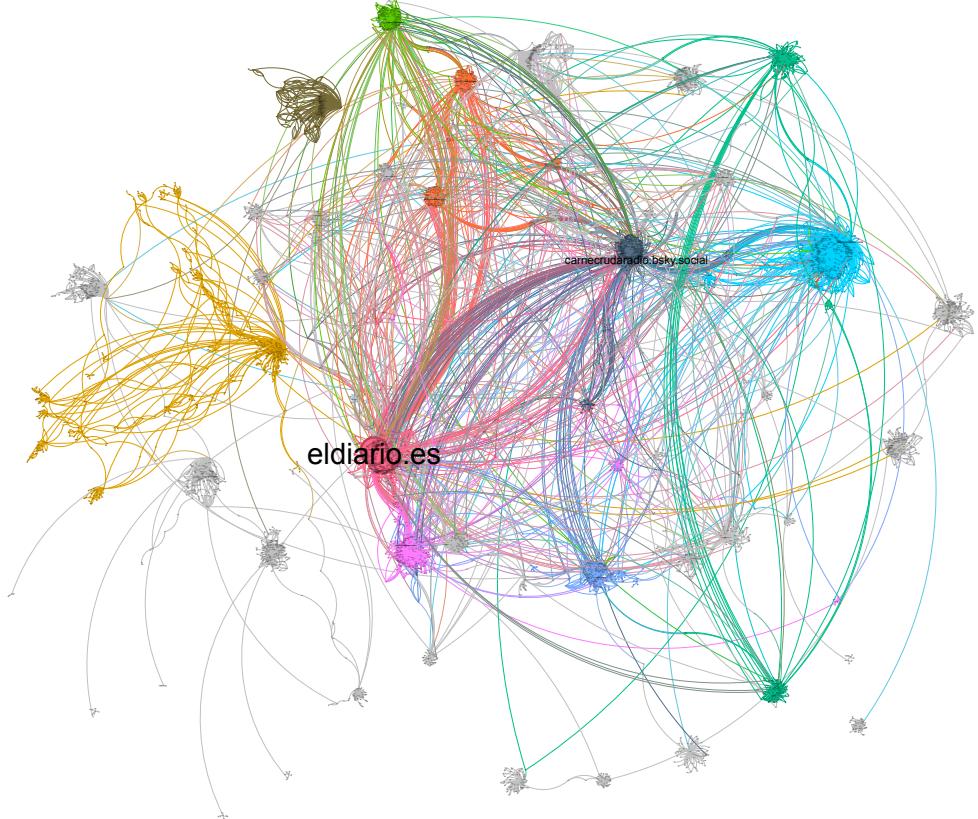


Figura 24: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su valor de autoridad.

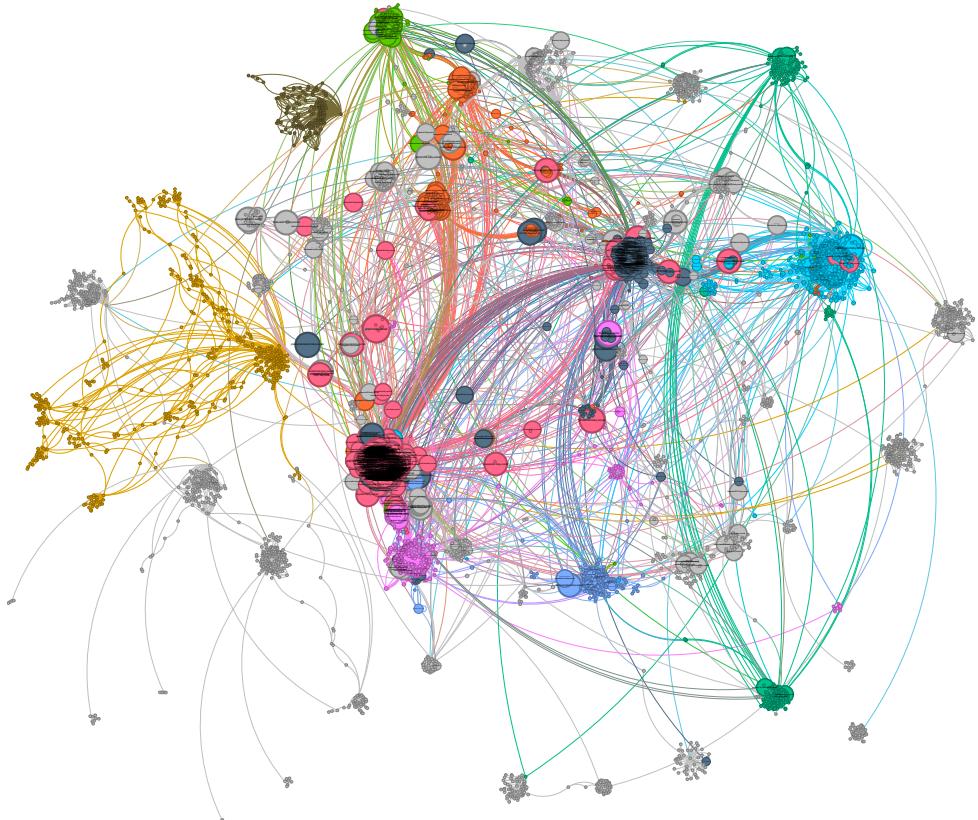


Figura 25: Red de interacciones entre usuarios vinculados a cinco pódcasts socioculturales de habla hispana. El tamaño de los nodos es proporcional a su valor de hub.

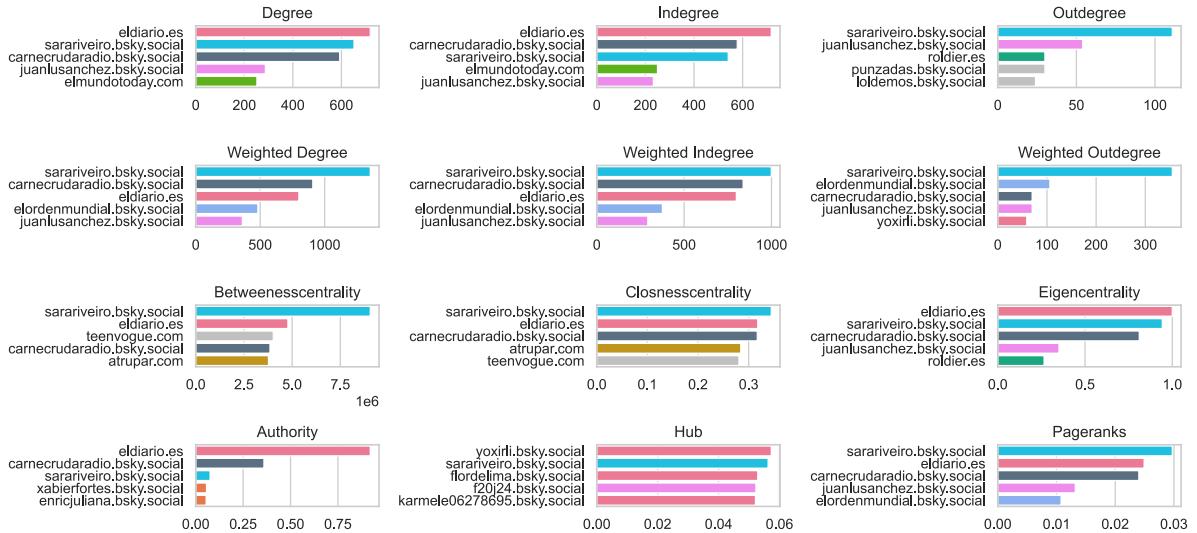


Figura 26: Comparativa de las cinco cuentas más destacadas según distintas métricas de centralidad en la red de interacciones de Bluesky. Las métricas relacionadas con el diámetro de la red se han computado de forma no dirigida. Los colores indican la comunidad a la que han sido asignados por el algoritmo de modularidad.