



Universidad Internacional de La Rioja

Escuela Superior de Ingeniería y

Tecnología

Máster Universitario en Inteligencia artificial

Generación de Redes Complejas y Dinámica de Opiniones con Modelos Clásicos y LLMs

Trabajo fin de estudio presentado por:	Carlos García Alvarado
Tipo de trabajo:	Comparativa de soluciones
Director/a:	Xiomara Patricia Blanco Valencia
Fecha:	16-02-2025

A mi familia y amigos.

Especial mención a mi tía Tere, por su generosidad inagotable y por ser un ejemplo de trabajo, perseverancia y ayuda desinteresada. A mi tía Carmen, por su lucidez y sentido común, cuya compañía y relatos me brindan paz y perspectiva. A mi amigo Nacho, compañero de mil batallas, por su nobleza, lealtad y por haber compartido conmigo tantos momentos. A mi amigo y colega Stefano Restelli, por su tesón, su energía inagotable y por ser un estímulo intelectual permanente, siempre dispuesto a desafiar ideas y explorar nuevos caminos.

Resumen

En este trabajo estudio la formación de redes sociales complejas mediante de redes clásicas (i.e grafo aleatorio y red libre de escala) y se realiza un estudio numérico de la formación de opiniones en un conjunto social mediante las reglas de mayoría clásicas y modificadas, donde cada individuo tiene una opinión binaria sobre un tema determinado y puede cambiar de opinión para coincidir con la opinión de la mayoría. Asimismo, se estudia la formación de redes generadas con Large Language Models (LLMs) donde cada persona (generada con LLMs siguiendo las características demográficas de USA) actúa como un agente y forma una red social compleja, estudio las características de estas redes complejas donde las comparamos con las redes clásicas y redes sociales reales, así mismo, se describen fenómenos como la homofilia y la polarización en estas redes, y se explora su dinámica utilizando modelos de mayoría, tanto en sus versiones clásicas como modificadas.

Palabras clave: Ciencia de redes, redes sociales complejas, dinámica social, modelos de lenguaje (LLMs)

Abstract

In this work, I study the formation of complex social networks using classical networks (i.e., random graphs and scale-free networks) and conduct a numerical analysis of opinion dynamics in a social group using both classical and modified majority rules. In this model, each individual holds a binary opinion on a given topic and can change their opinion to align with the majority view. Additionally, we examine the formation of networks generated with Large Language Models (LLMs), where each person (generated by LLMs based on U.S. demographic characteristics) acts as an agent forming a complex social network. We analyze the features of these complex networks, including density, clustering, community structure, and degree distribution, and compare them to classical networks and real-world social networks. Furthermore, we describe phenomena such as homophily and polarization in these complex social networks and explore their dynamics using majority models, both in their classic and modified versions.

Keywords: Network science, complex social networks, social dynamics, LLMs

Índice de contenidos

1.	Introducción	1
1.1.	Motivación	1
1.2.	Planteamiento del trabajo	2
1.3.	Estructura del trabajo	4
1.3.1.	Metodología	4
1.3.2.	Resultados	5
1.3.3.	Discusión	5
2.	Contexto y estado del arte	7
2.1.	Contexto del problema	7
2.2.	Estado del arte	8
2.3.	Conclusiones	13
3.	Objetivos concretos y metodología de trabajo	15
3.1.	Objetivo general	15
3.2.	Objetivos específicos	15
3.3.	Metodología del trabajo	17
3.3.1.	Dinámica social: reglas mayoritarias para la formación de opiniones en redes complejas clásicas	17
3.3.2.	Generación de redes con LLMs y análisis	23
4.	Desarrollo específico de la contribución	41
4.1.	Formación de opiniones en redes complejas clásicas: aplicación de la regla de mayoría clásica y modificada	41
4.2.	Redes generadas con LLMs. Análisis.	56
5.	Conclusiones y trabajo futuro	66
5.1.	Conclusiones	66
5.2.	Líneas de trabajo futuro	68
	Referencias bibliográficas	71
Anexo A.	Código fuente y datos analizados	75
Anexo B.	Costes de Generar las Redes Sociales con LLMs	75
Anexo C.	Temperatura de los LLMs	76

Índice de figuras

Figura 1. Ejemplo de Red Erdős-Rényi.	41
Figura 2. Ejemplo de Red Erdős-Rényi con opiniones	41
Figura 3. Evolución de opinión en Red Erdős-Rényi	42
Figura 4. Estado inicial y final de opiniones en la Red Erdős-Rényi	44
Figura 5. Evolucion de S+ frente al tiempo en la red Erdős-Rényi	45
Figura 6. Efecto del parámetro f en la evolución de S+ en la red Erdős-Rényi	46
Figura 7. Efecto del parámetro $\langle k \rangle$ en la evolución de S+ en la red Erdős-Rényi.....	47
Figura 8. Evolucion de S+ frente al tiempo en la red Barabási-Albert.....	48
Figura 9. Efecto del parámetro f en la evolución de S+ en la red Barabási-Albert	48
Figura 10. Efecto del parámetro $\langle k \rangle$ en la evolución de S+ en la red Barabási-Albert	49
Figura 11. Evolución de S+ en la red Erdős-Rényi con modelo de opinión modificada	51
Figura 12. Evolución del parámetro f en la red Erdős-Rényi con modelo de opinión modificada.....	52
Figura 13. Evolución del parámetro $\langle k \rangle$ en la red Erdős-Rényi con modelo de opinión modificada.....	52
Figura 14. Evolución de S+ en la red Barabási-Albert con modelo de opinión modificada....	53
Figura 15. Evolución de S+ en la red Barabási-Albert con modelo de opinión modificada en función de f	55
Figura 16. Evolución de S+ en la red Barabási-Albert con modelo de opinión modificada en función de $\langle k \rangle$	55
Figura 17. Propiedades de las redes generadas con LLMs.....	57
Figura 18. Homofilia de genero, etnia, religión y politica.....	58
Figura 19. Metricas estructurales de los modelos de OpenAI	59
Figura 20. Homofilia por edad.....	61
Figura 21. Polarizacion de los LLMs	62

Índice de tablas

Tabla 1. <i>Distribución de personas por su afiliación política.</i>	24
Tabla 2. <i>Distribución de personas por su afiliación etnia.</i>	24
Tabla 3. <i>Distribución de personas por su religión.</i>	24
Tabla 4. <i>Distribución de personas por su genero.</i>	24
Tabla 5. <i>Distribución de personas por su rango de edad</i>	25
Tabla 6. <i>Descripcion de los LLMs usados.</i>	27
Tabla 7. <i>Coste de las simulaciones finales usando LLMs.</i>	75

1. Introducción

1.1. Motivación

La Ciencia de Redes es un campo interdisciplinario que analiza las relaciones e interacciones entre los elementos de un sistema, representados como nodos y enlaces. Su utilidad abarca desde el estudio de redes sociales, donde permite identificar comunidades e influencers; la biología, con el análisis de redes metabólicas para identificar rutas clave, y redes neuronales para mapear conectomas y estudiar trastornos; la epidemiología, mediante la modelización de redes de contagio, que permite predecir la propagación de enfermedades e identificar superpropagadores en epidemias; la economía, a través del análisis de redes de comercio global para identificar mercados clave, y la modelización de redes financieras para evaluar riesgos sistémicos; la tecnología, con el análisis de redes de Internet para mejorar la conectividad y la identificación de vulnerabilidades en sistemas cibernéticos; la logística, mediante el diseño y optimización de redes de transporte para garantizar una distribución eficiente; y el análisis de redes eléctricas, donde se evalúan y optimizan infraestructuras energéticas para garantizar su estabilidad y eficiencia. Este enfoque ofrece herramientas esenciales para comprender las dinámicas estructurales y su impacto en los sistemas que modelan.

En este trabajo nos centramos en el estudio de redes sociales, sus dinámicas de opinión y el emergente uso de LLMs para su generación, con un enfoque en los sesgos presentes en las redes generadas mediante estos modelos de lenguaje.

He mencionado anteriormente las aplicaciones de la ciencia de redes en distintos campos. Aquí profundo un poco más en la aplicación de la generación de redes sociales realistas. La aplicación de generación de redes sociales realistas es esencial para numerosos campos, especialmente cuando no es posible acceder a la red social real debido a restricciones como la privacidad o cuando se requiere un entorno ficticio para modelar interacciones. Por ejemplo, en el contexto de las epidemias, las redes sintéticas permiten a los investigadores simular la transmisión de enfermedades al mapear contactos entre individuos (Barrett et al., 2009; Block et al., 2020). Además, estas redes se utilizan ampliamente para estudiar el comportamiento en redes sociales digitales (Pérez-Rosés y Sebé, 2015; Sagduyu et al., 2018)

o analizar fenómenos como la polarización y los cambios en las opiniones colectivas (Dandekar et al., 2013; Das et al., 2014). También son muy útiles en disciplinas científicas, Boccaletti et al, 2006 describen principales conceptos y resultados alcanzados recientemente en el estudio de la estructura y dinámica de redes complejas, y resumen las aplicaciones relevantes de estas ideas en muchas disciplinas diferentes, que abarcan desde la ciencia no lineal hasta la biología, desde la mecánica estadística hasta la medicina y la ingeniería.

Los métodos modernos basados en aprendizaje profundo suelen necesitar grandes volúmenes de datos de entrenamiento específicos del dominio para generar redes (You et al., 2018). Esto supone una limitación significativa cuando las redes reales no están disponibles. Por otro lado, los modelos tradicionales ofrecen soluciones más sencillas, pero con limitaciones evidentes al depender de supuestos simplificados. Por ejemplo, los modelos de Erdős–Rényi asumen que cada conexión se forma con una probabilidad constante p (Erdős y Rényi, 1959). Aunque modelos como el de mundo pequeño (Watts y Strogatz, 1998) o el de bloques estocásticos (Holland et al., 1983) ofrecen mejoras, su capacidad para capturar la complejidad de las interacciones sociales sigue siendo limitada debido a la rigidez de sus parámetros predefinidos.

Una alternativa interesante para abordar estas limitaciones es el uso de modelos de lenguaje de gran escala (LLMs), que ofrecen nuevas posibilidades en la generación de redes sociales con un mayor grado de realismo.

1.2. Planteamiento del trabajo

Primeramente, comenzaré realizando simulaciones numéricas de dinámicas de formación de opiniones en redes complejas, basándome en el modelo clásico de regla de la mayoría (CMR) (Krapivsky & Redner, 2003). Inicialmente, trabajaré con redes Erdős–Rényi para analizar cómo la estructura de la red, la conectividad promedio y la distribución inicial de opiniones afectan el estado final del sistema. Para ello, asignaré opiniones aleatorias a los nodos y simularé su evolución según la mayoría de sus vecinos hasta alcanzar un estado estable. Posteriormente, implementaré redes escala-libre, generadas con el algoritmo de Barabási-

Albert (Barabási & Albert, 1999), para comparar los resultados y explorar cómo una estructura más representativa de redes sociales reales influye en el proceso. También incluiré una versión modificada del modelo, donde la opinión propia del nodo se incorpora al conteo, lo que introduce un sesgo de auto refuerzo (Shao, Havlin, & Stanley, 2009). Estas simulaciones me permitirán entender mejor cómo las dinámicas sociales son moldeadas por la estructura de las redes. Queremos entender cómo los parámetros de balance inicial de opiniones (f) y conectividad promedio (k) influyen en la dinámica del sistema. Exploraremos cómo f afecta la evolución y distribución de opiniones, y cómo k impacta la propagación y estabilidad de las interacciones en la red. Esto nos permitirá identificar patrones clave y relaciones entre estructura y comportamiento en redes sociales simuladas. Se utilizarán las ambas redes y ambos modelos de opinión para este estudio.

Como segunda parte del trabajo, el objetivo principal de este trabajo es comparar la capacidad de los modelos tradicionales de generación de redes sociales con la de los modelos de lenguaje de gran escala (LLMs) para producir redes que reflejen la estructura y dinámica de las interacciones sociales reales. Mientras que los modelos clásicos, como los de Erdős-Rényi, Watts-Strogatz y el modelo de Barabási-Albert, ofrecen aproximaciones simplificadas de la realidad, este trabajo investiga cómo los LLMs, entrenados con grandes volúmenes de datos, pueden generar redes sociales con estructuras más complejas y realistas.

A través de un análisis comparativo, se evaluarán las redes generadas por ambos enfoques en términos de sus propiedades estructurales, como la distribución de grados, el coeficiente de agrupamiento, su modularidad, la asortatividad, la eficiencia, su grado de Gini y su entropía PageRank. En particular, se pondrá a prueba la capacidad de los LLMs para capturar patrones sociales complejos que los modelos tradicionales no pueden, como la formación de comunidades o la dinámica de influencia dentro de la red.

Sin embargo, al igual que en estudios previos, como el mencionado en el artículo de Chang et al. (2024), también se explorará una posible limitación de los LLMs: la tendencia a sobreestimar ciertos sesgos, como la homofilia y la polarización de opiniones política, que puede no estar tan marcada en redes reales. El trabajo se centrará en cómo estos sesgos

pueden surgir en redes generadas por LLMs y cómo se comparan con las redes tradicionales y reales en términos de la representación de tales fenómenos.

El planteamiento de esta segunda parte del trabajo busca, en última instancia, profundizar en la generación de redes con LLMs y si esta generación de redes con LLMs representa una mejora significativa sobre los enfoques tradicionales en la generación de redes sociales sintéticas, y en qué contextos pueden ofrecer ventajas o presentar limitaciones en comparación con los modelos clásicos.

1.3. Estructura del trabajo

El trabajo se estructurará en dos partes:

- **Dinámicas de opiniones en los modelos clásicos de generación de redes:**
Descripción de los modelos tradicionales de redes (Erdős-Rényi, Watts-Strogatz, bloques estocásticos, etc.) y sus características. Se estudiará su dinámica bajo la regla de la mayoría y la regla de la mayoría modifica.
- **Modelos de lenguaje de gran escala (LLMs):** Explicación de qué son los LLMs, cómo funcionan y por qué se consideran una herramienta prometedora para generar redes sociales más realistas. Asimismo se introducirán las métricas clave utilizadas para evaluar redes sociales, como la distribución de grados, el coeficiente de agrupamiento, modularidad, etc. Después se analizarán sus sesgos como la homofilia, así como la polarización de la red. Finalmente se analizará su dinámica bajo las reglas de mayoría clásica y modificada.

1.3.1. Metodología

- **Generación de redes tradicionales y estudio de su dinámica:** Descripción del proceso de generación de redes utilizando los modelos clásicos y cómo se implementarán en el estudio. Se comenzará con simulaciones numéricas del modelo clásico y modificado de regla de la mayoría en redes complejas. Se analizarán redes Erdős-Rényi y escala-libre para estudiar cómo la estructura de la red y la conectividad influyen en la formación de opiniones y en el estado final del sistema.
- **Generación de redes con LLMs y su análisis:**

- **Generación:** Explicación de cómo se utilizarán los LLMs para generar redes, incluyendo detalles sobre el modelo de lenguaje elegido, datos de entrenamiento y las metodologías aplicadas.
- **Comparación entre las redes y sesgos:** Explicación de cómo se llevará a cabo la comparación entre las redes tradicionales y las generadas por LLMs en función de las métricas y propiedades definidas en la sección teórica.
- **Validación y evaluación:** Métodos de validación de los resultados, cómo se comprobarán las redes generadas y cómo se medirá su "realismo" en comparación con redes sociales reales.
- **Simulación de la dinámica de opiniones de las redes generadas con LLMs.** Al igual que en la primera parte si simularan bajo las dinámicas de opiniones de la regla de la mayoría clásica y modificada.

1.3.2. Resultados

- **Simulación con redes clásicas:** cómo la estructura de la red y parámetros como la conectividad promedio y la distribución inicial de opiniones afectan la dinámica de formación de consenso y los estados finales en modelos de regla de la mayoría (*Majority Rule*). Además, se analizarán las diferencias clave entre redes aleatorias y escala-libre, así como entre las versiones clásica y modificada del modelo, para entender mejor su aplicabilidad a redes sociales reales
- **Comparación de las redes generadas mediante LLMs:** Presentación de los resultados obtenidos en términos de las métricas seleccionadas, con comparaciones entre los modelos tradicionales, los LLMs y la sociedad real.
- **Análisis de los sesgos (si se encuentran):** Discusión de cualquier sesgo observado en las redes generadas por LLMs (por ejemplo, sobreestimación de la homofilia política) y comparación con las redes tradicionales.

1.3.3. Discusión

- **Interpretación de los resultados de las dinámicas de opinión de las redes clásicas:** Se discutirá los resultados obtenidos en la modelización del modelo de mayoría y del modelo de mayoría modificado aplicados a las redes ER y BA. Se evaluará el impacto

de factores como la conexión de nodos K y el parámetro f_i que representa la fracción de nodos que siguen una determinada regla o comportamiento, en la convergencia de opiniones dentro de estas redes.

- **Interpretación de los resultados de las simulaciones con LLMs:** se analizarán las implicaciones de los resultados obtenidos, comparando cómo las distintas estructuras de red y modificaciones en la dinámica afectan los procesos de formación de opinión.
- **Limitaciones del estudio:** Discusión sobre las limitaciones de los enfoques utilizados y posibles áreas de mejora.
- **Aplicaciones prácticas y futuras líneas de investigación:** Reflexión sobre cómo los resultados pueden aplicarse en distintos contextos y qué investigaciones futuras podrían surgir de este trabajo.

2. Contexto y estado del arte

2.1. Contexto del problema

La generación de redes sociales sintéticas es un área clave de investigación en diversas disciplinas, incluyendo la sociología, la informática y la epidemiología, entre otras. Estas redes permiten modelar interacciones y comportamientos sociales sin necesidad de recurrir a datos reales, los cuales pueden ser difíciles de obtener debido a preocupaciones sobre la privacidad o a la imposibilidad de observar ciertos escenarios hipotéticos. Por ejemplo, en estudios de propagación de enfermedades, los investigadores suelen emplear redes sociales sintéticas para modelar cómo se propagan las infecciones entre los individuos, sin poner en riesgo la privacidad de las personas involucradas (Barrett et al., 2009; Block et al., 2020).

En este contexto, los modelos tradicionales de generación de redes, como los de Erdős-Rényi, Watts-Strogatz y los modelos de Barabási-Albert (BA), han sido ampliamente utilizados. Estos enfoques, aunque eficaces para modelar algunas propiedades fundamentales de las redes, se ven limitados por su simplicidad estructural y la rigidez de sus supuestos. Por ejemplo, los modelos clásicos suelen asumir que las conexiones entre los nodos se forman de manera aleatoria o según reglas predefinidas, lo que puede no reflejar adecuadamente la complejidad de las interacciones sociales reales, donde los individuos se agrupan en comunidades y las relaciones no siempre siguen patrones simples.

Estas redes clásicas se han utilizado para el estudio de la formación de opiniones, que es fundamental para comprender cómo las sociedades toman decisiones, adoptan normas y responden a desafíos colectivos. Las opiniones individuales y su propagación influyen en fenómenos como la polarización, la cohesión social, la difusión de información o desinformación, la adopción de innovaciones y comportamientos, y las dinámicas políticas y económicas. Además, la estructura de las redes sociales desempeña un papel clave en la velocidad y naturaleza del cambio de opiniones, afectando la propagación de ideas y la estabilidad social. Modelos como el de la Regla de la Mayoría (CMU) permiten simular estas dinámicas, ofreciendo herramientas para analizar cómo las interacciones locales generan patrones globales y cómo factores como la conectividad o el balance inicial de opiniones impactan en estos procesos.

Con el avance de la inteligencia artificial y el procesamiento del lenguaje natural, los modelos de lenguaje de gran escala (LLMs) han emergido como una herramienta poderosa para generar redes sociales de forma más realista. Los LLMs, entrenados con grandes volúmenes de datos textuales, tienen el potencial de capturar las dinámicas sociales complejas y las interacciones no lineales que los modelos tradicionales no logran reproducir. Sin embargo, el uso de LLMs para la generación de redes sociales también plantea nuevos desafíos, como la posibilidad de que estos modelos generen sesgos o representaciones distorsionadas de ciertos fenómenos sociales, como la homofilia política, que puede no reflejar la diversidad de opiniones de una red social real.

La segunda parte de este trabajo se enmarca en este contexto y aborda la necesidad de comparar la capacidad de los modelos tradicionales frente a los LLMs para generar redes sociales que no solo sean estructuralmente realistas, sino que también representen con mayor precisión las complejidades inherentes a las interacciones sociales.

2.2. Estado del arte

Antecedentes y evolución de los modelos tradicionales de generación de redes sociales y estudio de la propagación de opiniones.

La generación de redes sociales ha sido un área de estudio activo durante varias décadas. Los primeros modelos clásicos de redes sociales, como el modelo de Erdős-Rényi (1959), se caracterizan por su simplicidad, en los cuales cada par de nodos tiene una probabilidad constante de estar conectado. Aunque estos modelos han sido fundamentales para entender las propiedades globales de las redes, como la conectividad, las redes generadas por estos modelos a menudo no capturan características más complejas de las redes sociales reales, como la presencia de comunidades o estructuras de dependencia locales.

Más adelante, surgieron otros modelos como el de Watts-Strogatz (1998), conocido por su capacidad para generar redes "pequeñas" (small-world networks), que presentan un alto coeficiente de agrupamiento y caminos pequeños entre nodos, reflejando mejor las redes sociales reales. El modelo de Barabási-Albert (1999) introdujo el concepto de redes libres de

escala (scale-free networks), que se caracterizan por una distribución de grado que sigue una ley de potencias. Este modelo captura la idea de crecimiento preferencial, donde los nodos con más conexiones tienden a atraer nuevas conexiones, una característica común en redes sociales reales donde algunos individuos o entidades actúan como hubs altamente conectados.

Por otro lado, los modelos de bloques estocásticos (Holland et al., 1983) ofrecen un enfoque más estructurado al modelar comunidades o grupos dentro de la red, donde las conexiones entre nodos dentro de un grupo son más frecuentes que entre nodos de diferentes grupos. Sin embargo, estos modelos siguen siendo limitados por su necesidad de predefinir estructuras y parámetros, lo que reduce su flexibilidad para capturar la complejidad de las interacciones sociales reales.

El estudio de la dinámica de opiniones en redes sociales ha sido una línea de investigación fundamental en la ciencia de redes y sistemas complejos. Los primeros modelos clásicos, como el modelo de consenso de DeGroot (1974), exploraron cómo las opiniones individuales convergen en una única opinión común mediante interacciones repetidas en una red. Sin embargo, estos modelos simplifican las dinámicas sociales al asumir una homogeneidad de influencia y la ausencia de estructuras complejas.

Estos modelos se han utilizado para estudiar la dinámica de opiniones, el modelo clásico de Majority Rule (Galam, 2002) introdujo una perspectiva diferente al modelar el cambio de opiniones basado en la regla de la mayoría dentro de grupos locales. Según este enfoque, los nodos (representando individuos) actualizan sus opiniones en función de la opinión mayoritaria dentro de sus vecinos directos, reflejando comportamientos observados en contextos sociales reales, como el voto político o la adopción de ideas. Este modelo permite capturar fenómenos clave como el efecto de dominancia de grupos pequeños con alta cohesión interna o la aparición de transiciones abruptas en las opiniones globales de la red, conocidas como bifurcaciones.

Además, cuando se combina con redes complejas, como las redes "small-world" de Watts-Strogatz o las redes libres de escala de Barabási-Albert, el modelo ***Classic Majority Rule*** revela cómo las propiedades estructurales de la red influyen en el proceso de formación de consensos, permitiendo comprender mejor fenómenos como la persistencia de minorías

influyentes o la polarización en redes sociales reales. Estos avances destacan la importancia de considerar tanto la dinámica local de interacción como la estructura global de la red para el estudio de las opiniones en sistemas complejos.

Modelos de lenguaje de gran escala (LLMs) en la generación de redes sociales

Con el auge de los modelos de lenguaje de gran escala (LLMs), como GPT (Generative Pretrained Transformer), se han abierto nuevas posibilidades para la generación de redes sociales más realistas. Los LLMs, entrenados sobre grandes corpus de texto, son capaces de aprender patrones de interacción y generar estructuras de redes sociales que reflejan con mayor precisión la complejidad de las relaciones sociales. A diferencia de los modelos tradicionales, los LLMs no se basan en una estructura predeterminada, sino que pueden generar redes que adaptan dinámicamente las conexiones entre los nodos en función de las características de los datos de entrada.

Estudios recientes han comenzado a explorar el uso de los LLMs para simular redes sociales. Por ejemplo, Pérez-Rosés y Sebé (2015), así como Sagduyu et al. (2018), analizaron cómo los LLMs pueden ser utilizados para modelar plataformas de redes y fenómenos sociales como la polarización. Un aspecto destacado de los LLMs es su capacidad para capturar no solo las interacciones entre nodos, sino también las dinámicas subyacentes de influencia social, como los efectos de la homofilia (Dandekar et al., 2013; Das et al., 2014). Sin embargo, algunos estudios han advertido que los LLMs tienden a sobrestimar ciertos sesgos, como la homofilia política, un fenómeno donde individuos con opiniones similares tienden a conectarse más entre sí, algo que no siempre refleja la diversidad de las interacciones en redes sociales reales (Yasseri et al., 2015).

En trabajos relacionados los LLMs han demostrado su potencial para simular interacciones humanas y generar estructuras de red (Aher et al., 2023; Park et al., 2023). Sin embargo, muchos estudios asumen estructuras simplistas o dependen de intervención humana (Chuang et al., 2023; Gao et al., 2023). En contraste, trabajos recientes exploran cómo estos modelos generan redes escala-libres (Marzo et al., 2023) o analizan homofilia en sociedades simuladas (He et al., 2023). Papachristou y Yuan (2024) destacan principios como el apego preferencial (es decir, los nodos (i.e. personas) en una red tienden a conectarse preferentemente con nodos que ya tienen muchas conexiones). Este trabajo sigue un

enfoque en el que se generan los individuos con LLMs usando datos reales, integrando métricas, demografía y el estudio de homofilias, en línea con desarrollos recientes en la literatura (Chang et al., 2024).

Dinámicas de Formación de Opiniones en Redes Complejas tradicionales

El fenómeno de la formación de opiniones en redes sociales ha captado la atención de investigadores debido a su relevancia en dinámicas sociales reales, como procesos electorales, difusión de información y polarización política (Barrat et al., 2008). En este contexto, las redes complejas se han convertido en un marco teórico esencial, permitiendo modelar estructuras de interacción social y comprender cómo las opiniones individuales se ven influenciadas por la topología de estas redes (Newman, 2010). El modelo de regla de la mayoría clásica (Classic Majority Rule, CMR) es un enfoque ampliamente estudiado que asume que los individuos adoptan la opinión predominante entre sus vecinos, lo que ilustra cómo las conexiones sociales determinan la convergencia de opiniones en sistemas dinámicos (Krapivsky & Redner, 2003).

Sin embargo, la topología de la red tiene un papel crucial en la dinámica de este modelo. Por ejemplo, las redes aleatorias de Erdős-Rényi, caracterizadas por conexiones homogéneas, presentan dinámicas distintas a las redes libres de escala, donde los nodos con alta conectividad (hubs) tienen una influencia desproporcionada en la propagación de opiniones (Barabási & Albert, 1999). Además, la inclusión de modificaciones en el modelo, como considerar la opinión previa de los nodos en la regla de la mayoría, introduce variaciones significativas en los resultados, reflejando escenarios más realistas de persistencia de opiniones (Shao et al., 2009).

Este trabajo explorará cómo la estructura de la red y las dinámicas de decisión local afectan los patrones globales de formación de opiniones. Mediante simulaciones numéricas, se espera comprender cómo las redes complejas capturan comportamientos observados en sistemas sociales reales, como la formación de consenso, la fragmentación de opiniones y la influencia de los hubs en la estabilidad del sistema. Este enfoque permitirá contextualizar las

dinámicas de opinión en un marco computacional robusto, abriendo la puerta a estudios más específicos sobre fenómenos de interacción social en el mundo real.

Enfoques existentes

Mientras que los modelos tradicionales como Erdős-Rényi, Watts-Strogatz o bloques estocásticos siguen siendo populares debido a su simplicidad y facilidad de implementación, su capacidad para capturar la complejidad de las interacciones sociales es limitada. Los modelos tradicionales tienden a suponer que las relaciones entre individuos son homogéneas y no reflejan dinámicas complejas como la formación de comunidades o la variabilidad en la intensidad de las conexiones.

En contraste, los LLMs permiten una mayor flexibilidad, ya que pueden aprender de vastos volúmenes de datos y generar redes que se ajusten más estrechamente a las características observadas en redes sociales reales. Sin embargo, a pesar de su potencial, los LLMs requieren grandes cantidades de datos de entrenamiento y un procesamiento computacional considerable, lo que puede limitar su aplicabilidad en ciertos contextos. Además, la interpretación y validación de las redes generadas por los LLMs es un área aún en desarrollo, y los sesgos introducidos por los modelos de lenguaje deben ser cuidadosamente evaluados.

En términos de herramientas específicas, existen plataformas como NetworkX (Hagberg et al., 2008) para la creación de redes basadas en los modelos clásicos, mientras que para la generación de redes a partir de LLMs se utilizan grandes modelos preentrenados para tareas de NLP (o multimodales) que pueden adaptarse para generar redes. Sin embargo, aún no existe una herramienta estándar que integre ambos enfoques de manera eficiente, lo que representa un desafío para la investigación futura.

Conclusiones del estado del arte

Los modelos clásicos se han utilizado para comprender la dinámica de opiniones. El **Classic Majority Rule (CMR)** con su capacidad para modelar la formación de consensos y la influencia local en redes complejas ha demostrado ser un enfoque valioso para entender la dinámica de opiniones en sistemas sociales. Sin embargo, el estado del arte también indica

que los modelos basados en CMR tienden a simplificar aspectos clave de las interacciones sociales, como la heterogeneidad en la influencia de los nodos, los efectos de memoria en las decisiones o el impacto de estructuras más dinámicas. Así como obvia la opinión propia del nodo.

Por otro lado, a pesar de los avances en la generación de redes sociales, sigue existiendo una brecha entre las redes generadas por modelos clásicos y aquellas producidas por LLMs. Los primeros son más simples y carecen de la capacidad de modelar relaciones sociales complejas, mientras que los LLMs tienen el potencial de generar redes más realistas, pero con limitaciones derivadas de los sesgos y la dependencia de grandes volúmenes de datos. Esta comparativa sugiere que la investigación futura debería centrarse en cómo combinar lo mejor de ambos enfoques, aprovechando la estructura matemática de los modelos tradicionales y la flexibilidad de los LLMs para mejorar la calidad de las redes generadas.

2.3. Conclusiones

El análisis de los modelos tradicionales y los enfoques basados en LLMs para la generación de redes sociales muestra una evolución significativa en la capacidad para modelar interacciones sociales complejas. Los modelos clásicos, como Erdős-Rényi, Watts-Strogatz y los modelos de bloques estocásticos, han sido fundamentales para la comprensión de las propiedades estructurales de las redes, pero su capacidad para reflejar dinámicas sociales reales es limitada debido a sus supuestos simplificados. Estos modelos tienden a pasar por alto factores importantes como la formación de comunidades o las complejidades en la intensidad de las relaciones, lo que puede resultar en redes que no capturan la riqueza de las interacciones sociales observadas en el mundo real.

Para estudiar la dinámica de opiniones en modelos clásicos, se ha utilizado el *Classic Majority Rule (CMR)* que es un modelo fundamental en el estudio de dichas dinámicas de opiniones en redes sociales, ofreciendo una representación intuitiva y poderosa del proceso de formación de consensos. Su simplicidad, basada en la actualización de opiniones según la mayoría dentro de un grupo local, ha permitido analizar fenómenos sociales clave, como la dominancia de opiniones mayoritarias, la influencia de estructuras de red en la propagación

de ideas, y la resiliencia de minorías cohesivas. Además, cuando se aplica a redes complejas, el CMR ha demostrado ser útil para entender cómo las propiedades estructurales de estas redes influyen en el comportamiento colectivo.

Por otro lado, los modelos de lenguaje de gran escala (LLMs) presentan un avance significativo en la generación de redes sociales, ya que pueden aprender patrones complejos de interacción y generar redes más realistas que reflejan mejor la estructura y las dinámicas sociales. Sin embargo, su uso no está exento de desafíos. A pesar de su capacidad para generar redes más ajustadas a la realidad, los LLMs pueden introducir sesgos, como la sobreestimación de la homofilia política, lo que limita su aplicabilidad para ciertos estudios y contextos. Además, la necesidad de grandes volúmenes de datos de entrenamiento y de recursos computacionales plantea barreras para su uso en investigaciones de menor escala.

Una de las principales conclusiones es que no existe un enfoque único que sea ideal para todas las aplicaciones. Los modelos tradicionales siguen siendo útiles para generar redes simples y estudiar propiedades estructurales, mientras que los LLMs pueden ser más apropiados para contextos donde se busca capturar interacciones más complejas y realistas. No obstante, ambos enfoques presentan limitaciones que deben ser consideradas cuidadosamente al seleccionar la herramienta adecuada para cada caso.

La investigación futura en este campo debería centrarse en mejorar la integración de ambos enfoques. Una posible dirección es el desarrollo de modelos híbridos que combinen la simplicidad y la estructura matemática de los modelos clásicos con la flexibilidad y la capacidad de aprendizaje de los LLMs. Además, será crucial avanzar en la validación de redes generadas por LLMs, evaluando los sesgos que estos modelos pueden introducir y proponiendo metodologías para mitigar estos efectos. En este sentido, se abre una oportunidad significativa para explorar cómo los avances en el procesamiento del lenguaje natural pueden mejorar la generación de redes sociales, ofreciendo una mayor precisión y aplicabilidad en áreas como la simulación de plataformas de redes sociales, la investigación en fenómenos sociales y el modelado de propagación de enfermedades.

3. Objetivos concretos y metodología de trabajo

3.1. Objetivo general

El objetivo general de este trabajo es i) **estudiar la formación de opiniones en redes sociales complejas mediante las reglas de mayoría clásicas y modificadas** y ii) **generar redes sociales con LLM, comparar las redes generadas con LLMs con redes reales y modelos clásicos y estudiar las dinámicas de opinión en estas redes generadas con LLMs**, evaluando su capacidad para replicar características estructurales reales de las redes sociales, especialmente en términos de dinámicas de interacción y homofilia política. Esta comparación permitirá identificar las fortalezas y debilidades de cada enfoque, y explorar cómo los LLMs pueden ofrecer una alternativa para modelar interacciones sociales complejas de forma más realista.

3.2. Objetivos específicos

Objetivo 1: Análisis de la Formación de Opiniones en Redes Complejas clásicas: Aplicación de la Regla de Mayoría Clásica y Modificada

El estudio de la dinámica social y la formación de opiniones en los modelos clásicos, como la regla de mayoría (CMR), que se utilizan para describir cómo las opiniones se propagan y evolucionan a través de una red de individuos interconectados. Sin embargo, la simplicidad de estos modelos puede limitar su capacidad para capturar la complejidad de las redes sociales reales. Con el fin de superar estas limitaciones, se exploran modificaciones del modelo clásico que incluyen variaciones en las reglas de actualización de las opiniones y la estructura de la red. Este trabajo se enfoca en el análisis y la simulación de la formación de opiniones en redes complejas utilizando tanto la regla de mayoría clásica como su versión modificada, con el objetivo de estudiar la influencia de la conectividad de la red y la estructura de los nodos en el proceso de consenso. Se llevará a cabo un análisis numérico utilizando redes aleatorias de Erdős-Rényi y redes libres de escala, con el fin de evaluar las diferencias en los resultados y sus implicaciones para la modelización de dinámicas sociales.

Objetivo 2: Generar redes sociales con LLMs, identificar sesgos y estudiar sus dinámicas de opinión:

1. **Evaluar las redes sociales generadas por LLMs**, comparándolas tanto con los modelos tradicionales como con las redes sociales reales. El análisis se centrará en la capacidad de los LLMs para replicar dinámicas sociales complejas, como la formación de comunidades o la propagación de opiniones.
 - Generar redes sociales utilizando LLMs y compararlas con redes reales y las generadas por modelos tradicionales.
 - Medir la similitud estructural entre las redes generadas por LLMs, redes tradicionales y redes reales utilizando métricas de redes sociales como la densidad, el coeficiente de agrupamiento y el diámetro medio.
 - Implementar un sistema de generación de redes utilizando un modelo LLM preentrenado y aplicar métricas estándar de comparación.
 - Evaluar cómo los LLMs mejoran o no la capacidad para replicar las interacciones sociales reales y cómo se comparan con los modelos tradicionales.
2. **Identificar los sesgos presentes en las redes generadas por LLMs**, especialmente en lo relacionado con la sobreestimación de la homofilia política y otras dinámicas sociales.
 - Estudiar cómo los LLMs pueden generar redes sesgadas en términos de homofilia y polarización política y otros factores sociales que afectan las interacciones.
 - Evaluar el grado de homofilia y polarización política utilizando medidas de concentración de opiniones y polarización en las redes generadas por LLMs.
3. **Estudiar la dinámica de opinión en redes formadas con LLMs aplicación de la Regla de Mayoría Clásica y Modificada:** De manera análoga a lo realizado para las redes clásicas ER y BA, aplicar el modelo de mayoría clásico y el modificado para el estudio de la dinámica de opinión de las redes generadas con LLMs.

3.3. Metodología del trabajo

3.3.1. Dinámica social: reglas mayoritarias para la formación de opiniones en redes complejas clásicas

Como hemos descrito anteriormente, primeramente, generaremos redes clásicas, estudiaremos brevemente sus propiedades y sus dinámicas de opinión. Nos enfocaremos en las redes aleatorias (generadas con el modelo Erdős-Rényi) y la libre de escala (generada con el modelo Barabási-Albert).

Red aleatoria (modelo Erdős-Rényi)

El modelo **Erdős-Rényi** es uno de los enfoques más estudiados en la teoría de redes y representa un punto de partida fundamental para la comprensión de sistemas complejos. Este modelo genera **redes aleatorias** partiendo de un conjunto de N nodos, donde cada par de nodos tiene una probabilidad constante p de estar conectado. Esta probabilidad define la densidad de la red y garantiza que las conexiones se distribuyan de manera independiente, sin introducir estructuras predeterminadas.

La conectividad promedio de la red, $\langle k \rangle$, es un parámetro clave que describe el número promedio de conexiones por nodo. En el modelo $G(n, p)$, esta conectividad está relacionada directamente con la probabilidad de conexión p mediante la fórmula $\langle k \rangle = p \cdot (N - 1)$. Por lo tanto, al ajustar p , es posible controlar la densidad de conexiones en la red, pasando de redes dispersas a redes más densas.

Desde un punto de vista probabilístico, la distribución del grado de los nodos en una red Erdős-Rényi sigue una distribución binomial en redes finitas (ecuación 1):

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (1)$$

Sin embargo, en el caso de redes grandes ($N \rightarrow \infty$) y para valores pequeños de p , esta distribución se aproxima a una distribución de Poisson. Esto implica que la mayoría de los nodos tendrán un número de conexiones cercano a la conectividad promedio, mientras que los nodos con grados significativamente mayores o menores son menos comunes, lo que difiere de las distribuciones de grado observadas en muchas redes reales.

El modelo también presenta propiedades emergentes interesantes. Por ejemplo, existe un umbral crítico de conectividad por encima del cual la red se vuelve globalmente conectada, es decir, existe un camino entre cualquier par de nodos. Este umbral está relacionado con la probabilidad $p \ln(N)/N$. Además, las redes generadas por este modelo suelen tener caminos promedio cortos entre nodos, lo que significa que, aunque las conexiones sean aleatorias, la distancia típica (i.e. el número de saltos) entre dos nodos es pequeña.

Sin embargo, a pesar de su simplicidad y utilidad teórica, el modelo Erdős-Rényi presenta limitaciones al intentar modelar redes sociales reales. Estas redes suelen tener características que el modelo no captura, como un alto coeficiente de agrupamiento o la existencia de nodos altamente conectados (*hubs*). Además, no considera la dinámica evolutiva de muchas redes reales, donde las conexiones cambian con el tiempo.

En resumen, el modelo Erdős-Rényi es una herramienta teórica poderosa para estudiar las propiedades fundamentales de las redes aleatorias. Aunque carece de la capacidad para modelar estructuras complejas presentes en redes reales, sigue siendo un punto de partida esencial para entender cómo las propiedades globales emergen de reglas locales simples.

Red libre de escala (modelo Barabási–Albert)

El modelo **Barabási–Albert** (BA) es un enfoque teórico desarrollado para explicar la aparición de distribuciones de grado que siguen leyes de potencia en redes complejas, una característica comúnmente observada en redes sociales. Introducido en 1999, este modelo introdujo dos conceptos clave que lo diferencian de los modelos clásicos: el crecimiento de la red y la conexión preferencial. Estos mecanismos lo convierten en un modelo fundamental para generar **redes libres de escala** (scale-free networks).

En el modelo BA, la red comienza con un pequeño número de nodos iniciales conectados entre sí. A medida que la red crece, se añaden nuevos nodos que se conectan a los existentes siguiendo la regla de conexión preferencial: la probabilidad de que un nuevo nodo se conecte a un nodo existente es proporcional al grado de este último. Es decir, los nodos que ya tienen muchas conexiones tienen una mayor probabilidad de recibir nuevas

conexiones. Este proceso iterativo genera redes donde la distribución del número de conexiones (k) o sigue una ley de potencia $P(k) \sim k^{-\gamma}$, con un exponente γ , típicamente con valor entre 2 y 3.

La estructura generada por este modelo refleja de manera más precisa los patrones de conectividad observados en redes sociales reales en comparación con las redes aleatorias del modelo Erdős-Rényi. En una red libre de escala, unos pocos nodos (denominados hubs) tienen un número desproporcionadamente alto de conexiones, mientras que la mayoría de los nodos tienen solo unas pocas conexiones. Este patrón jerárquico, característico de las redes libres de escala, permite que la red sea altamente eficiente para el transporte de información o recursos, al tiempo que introduce una vulnerabilidad inherente a la eliminación de los *hubs*.

El modelo también presenta propiedades interesantes en términos de dinámica global y resiliencia. Por ejemplo, las redes BA son altamente robustas frente a fallos aleatorios de nodos, ya que la mayoría de las conexiones están concentradas en los *hubs*. Sin embargo, son vulnerables a ataques dirigidos, en los cuales la eliminación de los nodos más conectados puede fragmentar rápidamente la red. Estas propiedades han sido fundamentales para entender fenómenos como la propagación de información, epidemias y fallos sistémicos en redes reales.

A pesar de su utilidad, el modelo Barabási–Albert tiene limitaciones importantes. Una de las más notables es que asume que el crecimiento de las redes siempre sigue el principio de conexión preferencial, lo cual no es cierto en todos los contextos. Además, no considera factores adicionales como el *clustering* (agrupamiento local), la aparición de comunidades, o la evolución dinámica de las conexiones, lo que limita su capacidad para representar ciertos aspectos de redes sociales o biológicas.

En conclusión, el modelo Barabási–Albert proporciona un marco poderoso para estudiar redes complejas y capturar la naturaleza jerárquica y libre de escala de muchas de ellas. Su capacidad para reproducir patrones de conectividad reales, como la concentración de conexiones en hubs, lo convierte en una herramienta invaluable en el estudio de redes sociales y otros sistemas complejos. Sin embargo, sus limitaciones en la representación de estructuras más complejas subrayan la necesidad de integrar este modelo con enfoques más

avanzados que incluyan dinámicas locales y globales, así como factores externos que afectan la evolución de las redes. Estas mejoras permitirían ampliar su aplicabilidad y profundizar en la comprensión de las propiedades emergentes de las redes complejas.

Estudio del Majority Rule Model en las redes Erdős-Rényi y Barabási-Albert

El modelo **Majority Rule (MR)** es una aproximación teórica fundamental para analizar la dinámica de formación de opiniones en redes sociales. Este modelo asume que la opinión de un individuo está influenciada por sus vecinos dentro de la red social, en un proceso iterativo en el que prevalece la opinión mayoritaria a nivel local. En este enfoque, cada nodo representa un individuo, y las conexiones entre nodos representan relaciones sociales o de influencia.

Inicialmente, a cada nodo se le asigna de manera aleatoria un estado u opinión. Con una probabilidad f , el nodo adopta la opinión $\sigma_+ = +1$, mientras que con probabilidad $1-f$ adopta la opinión $\sigma_- = -1$ (o 0 , similar al modelo de Ising para el estudio del comportamiento de metales ferromagnéticos). En cada paso temporal de la simulación, cada nodo actualiza su estado para reflejar la opinión mayoritaria de sus vecinos directos (los nodos conectados a él). En caso de empate, el nodo mantiene su opinión actual. Las actualizaciones se realizan simultáneamente en todos los nodos durante cada paso de la simulación. El sistema se considera que ha alcanzado un estado estable cuando ya no ocurren cambios en las opiniones de los nodos.

En este estudio, el modelo Majority Rule se implementará en dos tipos de redes bien diferenciadas: redes **Erdős-Rényi (ER)**, caracterizadas por una conectividad homogénea y aleatoria, y redes **Barabási-Albert (BA)**, que presentan una estructura jerárquica con distribución de grado siguiendo una ley de potencia. Estas redes se analizarán para explorar cómo las diferencias estructurales influyen en la dinámica del modelo.

El análisis incluye los siguientes experimentos:

1. **Evolución temporal de las opiniones:** Usando parámetros $N=10,000$, $\langle k \rangle=10$, y $f=0.5$ (opinión inicial balanceada), se ejecutará el programa hasta alcanzar un estado estable. Para obtener resultados robustos, se promediarán al menos 100

simulaciones independientes. Se graficará la densidad de nodos en estado $\sigma_+ = +1$ (S_+) frente al número de pasos temporales de la simulación, observando la evolución hacia el consenso

2. **Efecto del balance inicial de opiniones (f):** Manteniendo $N=10,000$ y $\langle k \rangle=10$, se variará el balance inicial de opiniones (f) en el rango $[0.1, 0.9]$. Esto permitirá estudiar cómo la distribución inicial de opiniones influye en el resultado final de la simulación. En particular, se analizará la relación entre f y la probabilidad de alcanzar un consenso en la opinión inicial mayoritaria.
3. **Efecto de la conectividad promedio ($\langle k \rangle$):** Usando $N=10,000$ y un balance inicial de opiniones $f=0.5$, se explorará el efecto de la conectividad promedio de la red ($\langle k \rangle$) en el rango $[2, 100]$. Este experimento buscará comprender cómo la densidad de conexiones afecta la rapidez de convergencia al consenso y la capacidad del sistema para alcanzar estados estables.

Los resultados obtenidos de estas simulaciones permitirán un análisis comparativo entre redes Erdős-Rényi (ER) y Barabási-Albert (BA). En las redes ER, se espera que la propagación de opiniones sea más uniforme debido a su distribución homogénea de conexiones. Por otro lado, en las redes BA, se prevé que los *hubs* (nodos altamente conectados) puedan desempeñar un papel clave en la aceleración del proceso de consenso, aunque también podrían amplificar la polarización si adoptan opiniones opuestas.

En conclusión, este estudio explorará cómo las características estructurales de las redes afectan la dinámica del modelo Majority Rule, proporcionando una visión más profunda de los procesos de formación de opiniones en redes complejas y su relación con factores como la conectividad y la distribución inicial de opiniones.

Estudio del Majority Rule Model con una modificación en las redes Erdős-Rényi y Barabási-Albert

Introduzco una modificación al modelo de **Majority Rule (MR)** con el objetivo de explorar cómo pequeños cambios en las reglas locales de decisión pueden afectar la dinámica global de la formación de opiniones en redes complejas. En este nuevo *approach*, cuando un nodo i evalúa la opinión mayoritaria de sus vecinos, **también incluye su propia opinión en el conteo**. Esta variación refleja escenarios sociales donde los individuos no solo se ven

influenciados por sus conexiones, sino que también **consideran activamente su opinión inicial** como parte del proceso de decisión.

El estudio se llevará a cabo recalculando los resultados obtenidos en el modelo clásico Majority Rule con esta modificación en el código. Al igual que en el análisis anterior, se utilizarán redes **Erdős-Rényi (ER)** y **Barabási-Albert (BA)** para investigar el impacto de esta regla modificada en diferentes configuraciones estructurales. Los experimentos incluirán los mismos conjuntos de parámetros:

1. $N=10,000$, $\langle k \rangle=10$, y $f=0.5$, para analizar la evolución temporal de las opiniones y graficar la densidad de nodos con estado $\sigma_+ = +1$ (S_+) frente al número de pasos temporales.
2. Variación de f en el rango $[0.1, 0.9]$, para estudiar cómo el balance inicial de opiniones afecta el resultado final.
3. Variación de $\langle k \rangle$ en el rango $[2, 100]$, para evaluar cómo la conectividad promedio influye en la convergencia al consenso.

El objetivo principal será identificar las diferencias clave entre los resultados del modelo clásico y el modelo modificado. Se espera que la inclusión de la propia opinión en el conteo reduzca la velocidad de cambio de estado en las primeras etapas del proceso, ya que los nodos con una opinión inicial fuerte podrían mostrar mayor resistencia al cambio. Además, esta modificación podría generar una mayor persistencia de opiniones minoritarias, especialmente en redes con hubs, como las de tipo BA, donde los nodos altamente conectados podrían amplificar la estabilidad de sus propias opiniones.

En términos de modelado de dinámicas sociales, esta variación tiene implicaciones importantes. Refleja situaciones donde los agentes sociales tienden a priorizar su perspectiva individual junto con las influencias externas, lo cual puede ser representativo de sistemas más realistas. Comparar los resultados de ambos modelos ayudará a comprender mejor cómo las pequeñas modificaciones en las reglas de interacción local pueden transformar significativamente los patrones globales de consenso y polarización en redes complejas.

Generación de redes con LLMs

El proceso para generar redes sociales utilizando Large Language Models (LLMs) se estructura en dos etapas principales. En la primera etapa, se construye con LLMs un conjunto de "personas" o perfiles individuales que representan nodos en la red. En la segunda etapa, nuevamente utilizando un LLM se utiliza para generar las conexiones sociales entre las personas, creando así la red social completa. Este proceso aprovecha la capacidad de los LLMs para capturar patrones complejos, lo que permite modelar relaciones entre nodos de manera más realista y contextualizada. A diferencia de los modelos tradicionales, como los de Erdős-Rényi o Barabási-Albert, los LLMs tienen el potencial de incorporar atributos adicionales y factores externos en la generación de redes, lo que los hace especialmente útiles para estudiar sistemas sociales más ricos y dinámicos. El enfoque abre nuevas posibilidades para modelar redes sociales con una granularidad y realismo nunca antes posibles de alcanzar, pero también presenta desafíos importantes, como la posible introducción de sesgos y la necesidad de diseñar criterios claros para evaluar la calidad de las redes generadas.

3.3.2. Generación de redes con LLMs y análisis

Generación de redes con LLMs: generacion de personas

La metodología de generación de personas es similar a la propuesta por Chang et al., 2024. Las personas que forman los nodos de la red social fueron generadas utilizando **ChatGPT 3.5 Turbo**. Cada persona incluye características clave como género, edad, raza, religión y afiliación política, dimensiones que influyen en cómo se forman las conexiones en redes sociales reales (McPherson et al., 2001).

Para asignar estas características, utilizamos datos del Censo de EE.UU. (2023) para reflejar la distribución conjunta de género, edad y raza. Luego, asignamos la religión y la afiliación política en función de estas características, basándose en estudios demográficos previos (Statista, 2016; Pew Research Center, 2024). Este enfoque garantiza que las personas generadas sean representativas y relevantes para el estudio de redes sociales.

Por lo que generamos personas con las siguientes características de género, raza, edad, religion y afiliacion politica, que representan una muestra de la sociedad americana. Un ejemplo en formato .json

```
{"gender": "Woman", "race/ethnicity": "White", "age": 1, "religion": "Catholic",  
"political affiliation": "Democrat"}.
```

Este enfoque permite crear un conjunto de personas representativas y realistas, asegurando consistencia en la generación de perfiles para el análisis de redes sociales.

Tabla 1. Distribución de personas por su afiliación política.

Afiliación Política	
Demócrata	50
Republicano	49
Independiente	1

Tabla 2. Distribución de personas por su afiliación etnia.

Raza/Etnia	
White	65
Hispanic	21
Black	8
Asian	5
American Indian/Alaska Native	1

Tabla 3. Distribución de personas por su religión.

Religión	
Protestant	38
Unreligious	34
Catholic	27
Hindu	1

Tabla 4. Distribución de personas por su género

Género	
Woman	51
Man	48
Nonbinary	1

Tabla 5. *Distribución de personas por su rango de edad*

Rango de Edad	Conteo
1-10	12
10-20	7
20-30	12
30-40	18
40-50	13
50-60	12
60-70	15
70-80	7
80-90	4
90-100	0

Generación de redes con LLMs: generación de redes

El paper de Serina Chang et al. propone tres metodologías principales para generar redes sociales a partir de personas previamente generadas. En este trabajo, utilizamos la metodología conocida como “*secuencial*”, ya que, según la literatura, es la que genera redes más realistas al aprovechar al máximo las capacidades de los LLMs.

Pasos a seguir:

- **Asignación de Personas:** En cada paso, el LLM recibe una persona (nodo de la red) junto con la lista de todas las demás personas ya generadas. Esto permite al modelo decidir con quién formar conexiones de manera informada y progresiva.
 - **Información sobre la Red:** Además de la lista de personas, se proporciona información sobre la red construida hasta el momento. El modelo conoce únicamente el grado, es decir, el número total de amigos de cada persona, sin detalles de quiénes son esas personas.

- **Formación de Conexiones:** El LLM utiliza la información proporcionada (características demográficas de las personas y estructura de la red) para decidir las nuevas conexiones.

La elección de la metodología secuencial se basa en evidencia previa de que este enfoque genera redes más cercanas a las observadas en sistemas reales. A diferencia de otros métodos, este permite que las conexiones se formen de manera iterativa y contextualizada, integrando tanto las características de las personas como la estructura emergente de la red. Este nivel de detalle hace que las redes resultantes sean más útiles para estudiar dinámicas sociales complejas.

Cuando una persona se une a la red, se le presenta el siguiente *prompt*¹:

“You are joining a social network. You will be provided a list of people in the network *#list_of_people* followed by *#current_number_of_friends*. Which of these people will you become friends with? Provide a list of YOUR friends in the format of IDs.”

Así, el LLM que actúa como un agente (persona), decide con quién crear una conexión en la red. El modelo tiene la capacidad de decidir no conectar con nadie (no enlaces) o conectar con todos los nodos de la red, dependiendo de la información disponible y las dinámicas internas del modelo.

LLMs utilizados

Para generar las redes, se utilizan los siguientes LLMs: **gpt-3.5-turbo**, **gpt-4o-mini**, **gpt-4o** (OpenAI), **grok-2-latest** (xAI) y **llama3.1-70b** (Meta). Estos modelos se comparan en la Tabla 6 para evaluar su desempeño en la creación de redes sociales.

¹ Todos los *prompts* y datos para la generación de la LLMs serán en inglés ya que está demostrado que es en lenguaje donde mejores predicciones generan

Tabla 6. *Descripción de los LLMs usados.*

Modelo	Organización	Características	Tamaño
gpt-3.5-turbo	OpenAI	Fast model for simple tasks, superceded by GPT-4o-mini	Pequeño
gpt-4o-mini	OpenAI	Fast, affordable, small model focused tasks	Pequeño
gpt-4o	OpenAI	Versatil, high intelligence, OpenAI flagship model	Grande
grok-2-latest	xAI	XAI frontier language model with state-of-the-art reasoning capabilities	Grande
llama3.1-70b	Meta	new state-of-the-art model from Meta (available in 8B, 70B and 405B parameter sizes)	Grande

Métricas utilizadas

En este trabajo, calculamos métricas estructurales tanto a nivel de grafo como a nivel de nodo para evaluar las redes generadas. Todas las redes son tratadas como **no dirigidas** para garantizar consistencia en el análisis. Dado que el número de nodos puede variar entre redes, seleccionamos métricas que sean comparables entre grafos de diferentes tamaños. Aquellas métricas que dependen del tamaño de la red se escalan según su comportamiento esperado en un grafo aleatorio de Erdős-Rényi (Erdős y Rényi, 1959), asegurando una evaluación justa y uniforme.

Este enfoque de escalado permite realizar comparaciones significativas entre redes generadas bajo diferentes condiciones o modelos, identificando patrones que reflejan estructuras más realistas o complejas en comparación con redes aleatorias.

Densidad:

La densidad es una de las propiedades más fundamentales de una red, ya que describe cuán conectados están sus nodos en relación con el máximo número de conexiones posibles. En términos generales, las redes sociales suelen ser dispersas, lo que significa que su densidad tiende a ser baja (Wong et al., 2006).

La densidad se calcula como la proporción de conexiones observadas (E) respecto al número total de conexiones posibles en la red. En la ecuación 2, vemos como se expresa matemáticamente:

$$\text{Densidad} = \frac{2E}{N(N-1)} \quad (2)$$

Donde N es el número de nodos en la red y E el número de conexiones (o aristas). Este cálculo nos permite comparar redes de diferentes tamaños, ya que ajusta la conectividad observada a la escala máxima teórica para una red con el mismo número de nodos. Una baja densidad refleja la naturaleza dispersa de las redes sociales, en las que solo una fracción de todas las conexiones posibles están realmente presentes.

En el caso extremo en el que todos los nodos estén conectados con todos los demás (grafo completo), la densidad será igual a 1. Por otro lado, si no hay conexiones en la red, $E=0$ y, por lo tanto, la densidad será 0. Estas condiciones límite ayudan a interpretar la densidad observada en redes reales, donde suele encontrarse en valores intermedios, reflejando una conectividad parcial.

Coeficiente de agrupamiento promedio

El coeficiente de agrupamiento es una métrica clave para evaluar la cohesión local en redes sociales, capturando la tendencia de los amigos de un nodo a ser también amigos entre sí. Este fenómeno, conocido como agrupamiento o *clustering*, es una característica común en las redes sociales reales (Alizadeh, 2017).

Para un nodo i , su coeficiente de agrupamiento mide la proporción de conexiones existentes entre sus vecinos frente al total de conexiones posibles. Matemáticamente, se calcula como vemos en la ecuación 3:

$$C_i = \frac{N_i(N_i - 1)}{2E_i} \quad (3)$$

Donde N_i es el número de vecinos (grado) del nodo i y E_i es el número de conexiones entre esos vecinos. Este valor describe qué tan conectado está el entorno local del nodo i .

El **coeficiente de agrupamiento promedio** se obtiene calculando el promedio de los coeficientes de agrupamiento de todos los nodos de la red, lo vemos en la ecuación 4:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (4)$$

Esta métrica proporciona una visión global de la cohesión local en la red y permite comparaciones entre redes con diferentes estructuras y tamaños.

En los casos extremos, si todos los vecinos de cada nodo están completamente conectados entre sí (entornos totalmente conectados), el coeficiente de agrupamiento promedio será igual a 1. Por el contrario, si no existen conexiones entre los vecinos de ningún nodo, el coeficiente de agrupamiento promedio será igual a 0.

Proporción de Nodos en el Componente Conexo Más Grande (LCC)

La proporción de nodos en el **componente conexo más grande** (*Large connected component*, i.e. LCC) es una métrica fundamental para evaluar la conectividad global de una red. En redes sociales, es común observar una conectividad elevada, donde la gran mayoría de los nodos (a menudo más del 99%) pertenecen al LCC. Este componente representa el subgrafo más grande en el cual todos los nodos están conectados directa o indirectamente entre sí (Ugander et al., 2011).

Matemáticamente (ecuación 5), esta métrica se calcula como la fracción de nodos que forman parte del LCC respecto al total de nodos en la red:

$$\text{Proporción LCC} = \frac{N_{\text{LCC}}}{N} \quad (5)$$

Donde N_{LCC} es el número de nodos en el LCC y N es el número total de nodos en la red.

El valor de esta métrica permite identificar qué tan bien conectada está la red. Una alta proporción indica que la mayoría de los nodos son accesibles entre sí dentro de un único componente, lo que es característico de redes sociales reales.

En los casos extremos, si la red es completamente desconectada, es decir, no hay conexiones entre nodos, el LCC estará compuesto por un solo nodo y la proporción será igual a $1/N$. Por el contrario, en una red completamente conectada (grafo completo), todos los nodos pertenecen al LCC y la proporción será igual a 1.

Camino Más Corto Promedio en el LCC

En redes sociales, no solo es importante que los nodos estén conectados, sino también que puedan alcanzarse entre sí a través de caminos relativamente cortos. Este fenómeno, conocido como la "pequeña dimensión del mundo", es una característica típica de las redes sociales reales (Watts y Strogatz, 1998; Alizadeh, 2017). Para capturar esta propiedad, se mide el camino más corto promedio entre todos los pares de nodos dentro del **componente conexo más grande (LCC)**, ya que la distancia entre nodos desconectados sería infinita.

El **camino más corto promedio en el LCC** se define como:

Donde $d(i, j)$ es la distancia más corta entre los nodos i y j , y N_{LCC} es el número de nodos en el LCC. Esta métrica evalúa la eficiencia con la que los nodos pueden comunicarse dentro del componente principal de la red.

Para comparar esta métrica entre redes de diferentes tamaños, se normaliza dividiendo el camino más corto promedio por $\log(N_{\text{LCC}})$, ya que en grafos aleatorios de **Erdős-Rényi**, los caminos más cortos escalan de forma logarítmica con el número de nodos (Watts y Strogatz, 1998), véase la ecuación 6:

$$L_{\text{promedio}} = \frac{1}{|N_{\text{LCC}}|(|N_{\text{LCC}}| - 1)} \sum_{i,j \in N_{\text{LCC}}}^d (i,j) \quad (6)$$

En los casos extremos:

- En una red completamente conectada (grafo completo), el camino más corto promedio será igual a 1, ya que todos los nodos están directamente conectados.
- Por otro lado, en una red donde el LCC es una cadena lineal de nodos, el camino más corto promedio será proporcional al número de nodos en el LCC, reflejando una conectividad mínima.

Distribución de Grado

La distribución de grado es una métrica clave en el análisis de redes sociales, ya que describe cómo se distribuye el número de conexiones k entre los nodos. Las redes sociales reales suelen seguir una distribución de grado tipo ley de potencia, representada como $P(k) \sim k^{-\gamma}$, donde γ es una constante (Barabási y Albert, 1999). Esta distribución de cola larga (heavy right tail) refleja que, aunque la mayoría de los nodos tienen pocas conexiones, algunos nodos (hubs) tienen muchas más conexiones que el promedio.

Para medir la distribución de grado en una red, se calcula el grado de cada nodo (k_i), es decir, el número de conexiones que tiene cada nodo. Para hacer que el grado sea comparable entre redes de diferentes tamaños, se normaliza dividiéndolo por el número total de nodos en la red (N), tal y como vemos en la ecuación 7:

$$k_{\text{normalizado}} = \frac{k_i}{N} \quad (7)$$

Esto permite comparar las redes independientemente de su tamaño y escala.

Para analizar y resumir la distribución de grado en un conjunto de redes (por ejemplo, todas las redes generadas), se agrupan los grados normalizados de todos los nodos en un único conjunto de datos. A partir de este conjunto combinado, se calcula la distribución agrupada utilizando intervalos de 0.05, cubriendo el rango de valores de 0 a 1. Esta representación proporciona una visión general de cómo se distribuyen las conexiones en las redes analizadas.

En los casos extremos:

- Si todos los nodos tienen el mismo número de conexiones, la distribución de grado será uniforme y concentrada en un solo valor.
- En una red completamente conectada (grafo completo), todos los nodos tendrán un grado normalizado igual a $k_{\text{normalizado}} = \frac{k_i}{N}$
- En redes dispersas o poco conectadas, la distribución de grado estará concentrada cerca de 0, reflejando la baja conectividad de los nodos.

Homofilia

La **homofilia** es una propiedad fundamental de las redes sociales, reflejando la tendencia de las personas con características similares a formar conexiones entre sí. Este fenómeno ha sido ampliamente documentado en estudios de redes sociales (McPherson et al., 2001; Kossinets y Watts, 2009).

Para medir la homofilia en una red, calculamos la proporción observada de conexiones entre nodos del mismo grupo (por ejemplo, el mismo género) y la comparamos con la proporción esperada bajo una distribución aleatoria de conexiones. Muestro la definición de esta métrica en la ecuación 8:

$$H = \frac{S_{\text{exp}}}{S_{\text{obs}}} \quad (8)$$

Donde:

- En la ecuación 9 muestro como se calcula la S observada, que es la cantidad observada de conexiones entre nodos del mismo grupo. Se calcula como:

$$S_{\text{obs}} = \sum_{i,j} A_{ij} \cdot 1[g_i = g_j] \quad (9)$$

A es el valor en la matriz de adyacencia (es decir, 1 si los nodos i y j están conectados, 0 en caso contrario), y g_i indica el grupo al que pertenece el nodo i

- S esperada es la cantidad esperada de conexiones entre nodos del mismo grupo.

Vemos la formula matemática en la ecuación 10:

$$S_{\text{exp}} = \frac{1}{N(N-1)} \sum_g N_g(N_g - 1) \quad (10)$$

Donde N_g es el número de nodos en el grupo g , y N es el número total de nodos en la red.

Un valor de homofila $H > 1$ indica que hay más conexiones entre nodos del mismo grupo de lo esperado (homofilia). Por el contrario, un valor $H < 1$ indica **heterofilia**, donde las conexiones entre nodos de diferentes grupos son más frecuentes que lo esperado (como, por ejemplo, ocurre en las apps heterosexuales de citas).

Casos:

- En una red homofílica, los nodos se tienden a conectar con nodos de su mismo grupo, y H será significativamente mayor que 1.
- En una red aleatoria, donde las conexiones no están influenciadas por los grupos, H estará cerca de 1.
- En una red heterofílica, los nodos tienden a evitar conectarse con otros de su mismo grupo, H será menor que 1.

Polarización

La polarización en redes sociales mide la tendencia de los nodos (usuarios) a interactuar principalmente con otros nodos que comparten una orientación similar, especialmente en contextos como opiniones políticas. Este fenómeno refleja la segmentación de una red en comunidades homogéneas con poca interacción entre ellas, lo que puede tener implicaciones significativas en la propagación de información y la formación de opiniones. Para cuantificar la polarización, se utiliza una métrica que evalúa el sesgo de conexión de cada nodo en relación con un balance ideal. Dado un nodo i , se calcula su polarización como muestra la ecuación 11:

$$p_i = 2 \cdot \left| 0.5 - \frac{\alpha}{\alpha + \beta} \right| \quad (11)$$

Donde:

- α es el número de conexiones de i hacia nodos con una orientación específica (por ejemplo, izquierda, Partido Demócrata en nuestro caso).
- β es el número de conexiones de i hacia nodos con la orientación opuesta (por ejemplo, derecha, Partido Republicano en nuestro caso).

El valor de p captura la desviación de un balance perfecto ($\alpha=\beta$), donde 0.5 representa una distribución equilibrada. La polarización total de la red se calcula en la ecuación 12 como el promedio de los valores p_i de todos los nodos:

$$P_{\text{promedio}} = \frac{1}{N} \sum_{i=1}^N p_i \quad (12)$$

Valores más altos indican niveles más altos de polarización en la red, mientras que valores más bajos sugieren una mayor mezcla entre orientaciones.

Casos extremos:

- **Red completamente polarizada:** Si todos los nodos solo están conectados con otros de su misma orientación, $\alpha=0$ o $\beta=0$, entonces $p_i=1$ para todos los nodos, y $P_{\text{promedio}} = 1$, indicando polarización máxima.
- **Red completamente equilibrada:** Si cada nodo tiene el mismo número de conexiones hacia ambas orientaciones ($\alpha = \beta$), entonces $p_i=0$ para todos los nodos, y $P_{\text{promedio}} = 0$, indicando ausencia de polarización.

Coeficiente de Asortatividad por Grado

El coeficiente de asortatividad por grado mide la tendencia de los nodos en una red a conectarse con otros nodos que tienen un grado similar al suyo. Esta métrica captura patrones de homofilia estructural en términos de conectividad: los nodos altamente conectados (*hubs*) tienden a conectarse con otros *hubs*, mientras que los nodos con pocas conexiones tienden a vincularse con otros de bajo grado.

Matemáticamente, el coeficiente de asortatividad por grado se calcula como la correlación de Pearson entre los grados de los nodos en ambos extremos de una arista (ecuación 13):

$$r = \frac{\sum_{(i,j) \in E} (k_i - \bar{k})^2}{\sum_{(i,j) \in E} (k_i - \bar{k})(k_j - \bar{k})} \quad (13)$$

Donde:

- k_i y k_j son los grados de los nodos i y j conectados por una arista.
- \bar{k} es el grado promedio de la red
- E es el conjunto de todas las aristas en la red.

El coeficiente r toma valores en el rango $[-1,1]$:

- Un valor $r>0$ indica asortatividad, donde los nodos tienden a conectarse con otros nodos de grados similares.

- Un valor $r < 0$ indica disasortatividad, donde los nodos tienden a conectarse con otros nodos de grados muy diferentes.
- Un valor $r = 0$ indica que no hay correlación entre los grados de los nodos conectados

El coeficiente de asortatividad por grado es especialmente relevante en redes sociales, donde una alta asortatividad puede indicar la formación de comunidades cohesivas, mientras que una disasortatividad puede sugerir estructuras jerárquicas o centralizadas. Esta métrica ofrece una perspectiva estructural complementaria para analizar cómo se organiza la conectividad en redes complejas.

Eficiencia Global

La **eficiencia global** es una métrica que mide qué tan eficientemente fluyen las conexiones o la información en toda la red. Está basada en las distancias más cortas entre todos los pares de nodos, evaluando qué tan bien los nodos están conectados en términos globales. Una red con alta eficiencia global tiene rutas cortas que conectan a la mayoría de los nodos, mientras que una red con baja eficiencia global implica que muchos nodos están separados por largas distancias o incluso desconectados.

La fórmula 14 muestra como calcular la eficiencia global de una red G se define como:

$$E_{\text{global}} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d(i,j)} \quad (14)$$

Donde:

- N es el número total de nodos en la red.
- $d(i,j)$ es la distancia más corta (geodésica) entre los nodos i y j .
- $1/d(i,j)$ es la eficiencia de la conexión entre i y j ; si i y j están desconectados, su eficiencia es 0 .

Casos extremos:

- **Red completamente conectada (grafo completo):** En una red donde todos los nodos están directamente conectados entre sí, la eficiencia global será máxima (i.e. Eficiencia=1).
- **Red completamente desconectada:** En una red donde no hay conexiones entre los nodos, la eficiencia global será 0, ya que no existen caminos entre pares de nodos.
- Otro caso extremo es el de una **red de cadena lineal**, es una red donde los nodos están conectados en una línea (un camino), la eficiencia global será baja (cercana a 0), ya que la mayoría de los nodos estarán separados por distancias largas.

coeficiente de Gini del grado

El **coeficiente de Gini del grado** mide la desigualdad en la distribución de los grados de los nodos en una red. Esta métrica, originalmente desarrollada para medir la desigualdad económica, se adapta al análisis de redes para evaluar si las conexiones están distribuidas de manera uniforme o concentradas en unos pocos nodos (*hubs*).

El coeficiente de Gini varía entre 0 y 1:

- Un valor de **0** indica que todos los nodos tienen el mismo grado (distribución perfectamente uniforme).
- Un valor cercano a **1** indica que la mayoría de las conexiones están concentradas en unos pocos nodos, mientras que muchos nodos tienen pocos o ningún enlace.

El coeficiente de Gini del grado se calcula como se puede ver en la formula 15:

$$G = \frac{2 \sum_{i=1}^N (i \cdot k_i)}{N \sum_{i=1}^N k_i} - \frac{N + 1}{N} \quad (15)$$

Redes según su desigualdad:

- **Distribución perfectamente uniforme:** Si todos los nodos tienen el mismo grado, el coeficiente de Gini será $G=0$, indicando igualdad total.

- Distribución altamente desigual: Si uno o unos pocos nodos concentran todas las conexiones y el resto tiene grado 0, el coeficiente de Gini será cercano a $G=1$, reflejando máxima desigualdad.

En una red aleatoria de Erdős-Rényi, el coeficiente de Gini generalmente tendrá un valor bajo, ya que los grados tienden a estar distribuidos de manera más uniforme. Sin embargo, en una red libre de Barabási-Albert el coeficiente de Gini será generalmente alto ya que se forman *hubs* (nodos altamente conectados).

El coeficiente de Gini del grado es una herramienta valiosa para identificar estructuras jerárquicas en redes y entender cómo se distribuyen las conexiones, particularmente en sistemas donde los *hubs* desempeñan un papel central.

Entropía Normalizada de PageRank

El **PageRank**, desarrollado por Larry Page (fundador de Google) como parte del algoritmo de búsqueda original de Google Search, mide la importancia relativa de los nodos en una red basándose en la estructura de enlaces. En el contexto de redes sociales y otras redes complejas, PageRank evalúa qué nodos son más influyentes considerando no solo el número de conexiones directas, sino también la importancia de los nodos conectados.

La **entropía normalizada de PageRank** describe la uniformidad de esta distribución de importancia en la red. Si la influencia está concentrada en unos pocos nodos, la entropía será baja. Por el contrario, si la influencia está distribuida uniformemente entre los nodos, la entropía será alta. Este análisis permite identificar redes altamente jerárquicas o bien distribuidas.

El cálculo de la entropía normalizada de PageRank sigue los siguientes pasos:

1. Se calculan los valores de PageRank para todos los nodos de la red (PR_i).
2. Los valores de PageRank se normalizan para formar una distribución de probabilidad, véase la ecuación 16:

$$P_i = \frac{PR_i}{\sum_j PR_j} \quad (16)$$

3. En la ecuación 17 calculo la entropía de esta distribución se calcula como:

$$H = - \sum_i p_i \cdot \log_2(p_i) \quad (17)$$

4. Para normalizar la entropía, se divide por la entropía máxima posible, que ocurre cuando todos los nodos tienen la misma importancia (ecuación 18):

$$H_{\max} = \log_2(N) \quad (18)$$

5. Por lo tanto, vemos en la ecuación 19 el valor de la entropía normalizada es:

$$H_{\text{norm}} = \frac{H}{H_{\max}} \quad (19)$$

Potenciales casos de la entropía normalizada de PageRank:

- **Red con distribución uniforme de PageRank:** Si todos los nodos tienen el mismo valor de PageRank ($P=1/N$), la entropía será máxima ($H_{\text{norm}} = 1$), indicando una red completamente equitativa.
- **Red con concentración extrema en un nodo:** Si un único nodo concentra todo el PageRank ($P=1$ para un nodo y $P=0$ para el resto), la entropía será mínima ($H_{\text{norm}} = 0$), reflejando una jerarquía extrema.
- **Red aleatoria:** En redes con grados distribuidos homogéneamente, la entropía normalizada tenderá a valores intermedios, mostrando una dispersión moderada en la importancia de los nodos.

Comparativa

Para realizar los experimentos, seleccionamos un conjunto fijo de individuos/personas $N=100$. Estas personas fueron generadas previamente y se utilizaron de manera consistente en todos los experimentos realizados con los diferentes modelos de lenguaje (LLMs). En la **Tabla 1-5**, se detalla la composición demográfica de estas 100 personas.

Para cada método de generación basado en LLMs, se construyeron **11 redes distintas**. Esto nos permitió capturar la varianza introducida por diferentes factores, como el orden de las

personas en el *prompt* y las respuestas del modelo. La repetición del experimento asegura una representación más robusta de las propiedades de las redes generadas.

En cuanto al análisis de las métricas a nivel de grafo, presentamos los resultados visualizando tanto la **media** como el **error estándar** de las métricas calculadas, representado los **intervalos de confianza** de las mismas. Además, mostramos los puntos individuales de los datos para proporcionar una visión más completa de la variabilidad de los resultados. Este enfoque combina la inferencia estadística con la observación directa de la dispersión de los datos, tal como recomiendan Zhang et al. (2023).

Esta configuración experimental garantiza un análisis robusto, permitiendo comparar el desempeño de los distintos modelos y métodos de generación de redes de manera transparente y estadísticamente informada.

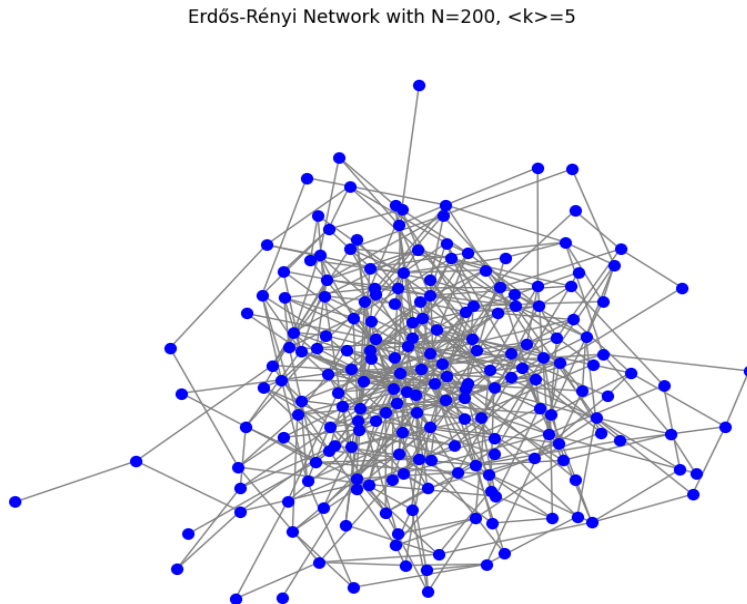
4. Desarrollo específico de la contribución

4.1. Formación de opiniones en redes complejas clásicas: aplicación de la regla de mayoría clásica y modificada

Red Erdős-Rényi

Primeramente, se ilustra cómo generar una red aleatoria utilizando el modelo Erdős-Rényi $G(N,p)$. En este caso, la red consta de $N=200$ nodos y una conectividad promedio $\langle k \rangle=5$. La probabilidad de conexión p se calculó como $p = \langle k \rangle / (N-1)$. La red generada se utilizó como base para las simulaciones, y su visualización inicial es la siguiente:

Figura 1. *Ejemplo de Red Erdős-Rényi.*

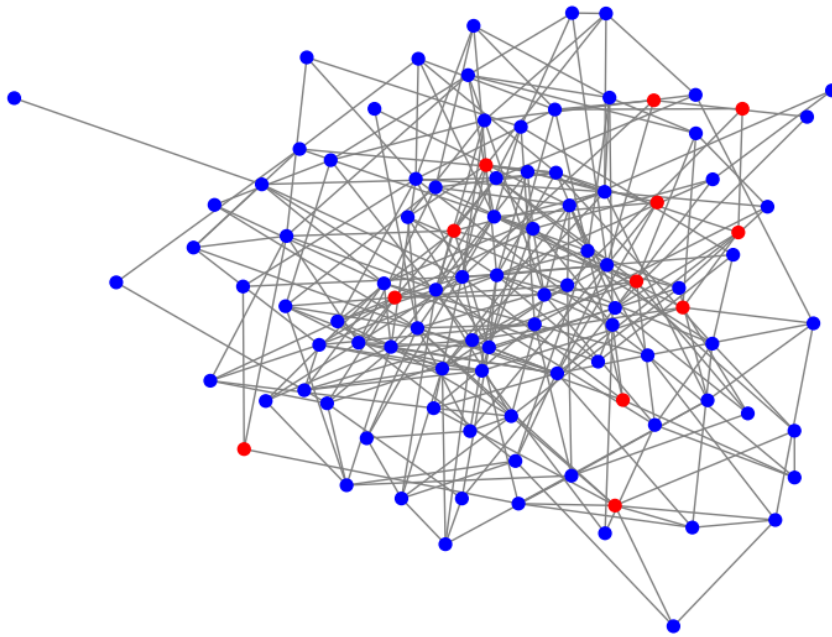


A continuación, se asigna aleatoriamente un estado u opinión a cada nodo de la red generada. Con probabilidad f un nodo es asignado a la opinión $\sigma_+ = 1$, mientras que con probabilidad $1-f$, se le asigna la opinión $\sigma_- = -1$. Para este caso, se emplearon $f=0.9$, $N=100$ nodos, y una conectividad promedio de $\langle k \rangle=6$.

La red visualizada refleja los nodos con opiniones $+1$ en azul y -1 en rojo, lo que permite observar la distribución inicial de estados. Este paso inicial es crucial para estudiar cómo las dinámicas de opinión evolucionan en función de la topología de la red.

Figura 2. *Ejemplo de Red Erdős-Rényi con opiniones*

Erdős-Rényi Network with Opinions ($f=0.9$)



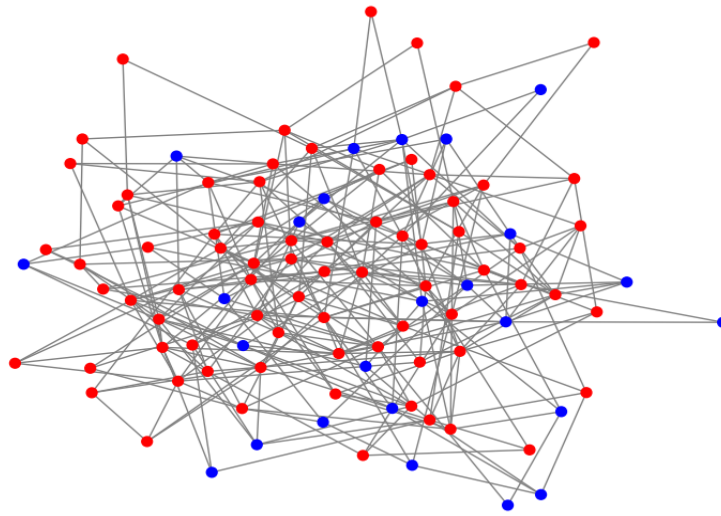
Resultados: Dinámica de Opiniones en la Red

Una vez asignadas las opiniones iniciales, se implementó una dinámica en la que, en cada paso de simulación, los nodos actualizan simultáneamente sus opiniones siguiendo la regla de la mayoría local. Específicamente, cada nodo adopta la opinión sostenida por la mayoría de sus vecinos. En caso de empate, el nodo conserva su opinión actual. Este proceso permite observar cómo las opiniones evolucionan en el tiempo bajo la influencia de la estructura de la red.

En el caso simulado, se utilizó una red Erdős-Rényi con $N=100$ nodos y una conectividad promedio de $\langle k \rangle=6$. Se ejecutaron 10 pasos de simulación, partiendo de una distribución inicial en la que cada nodo tenía probabilidad $f=0.5$ de iniciar con opinión “+1”. La figura 3 generada muestra la configuración final de opiniones, destacando las regiones de consenso y posibles áreas de conflicto. Este análisis proporciona una base para estudiar la propagación de opiniones en redes sociales bajo diferentes configuraciones iniciales y reglas de actualización.

Figura 3. Evolución de opinión en Red Erdős-Rényi

Opinion Dynamics at Step 10

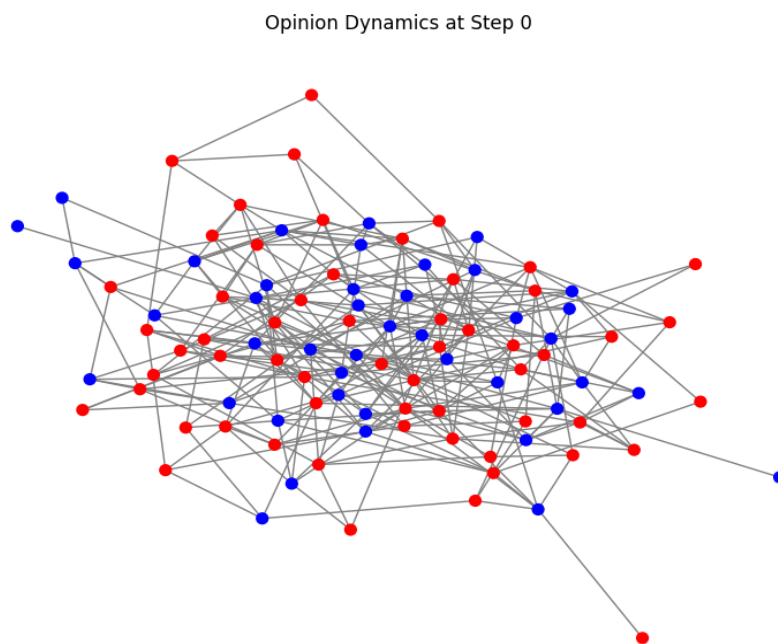


El sistema se considera en estado estable cuando las opiniones de los nodos dejan de cambiar en iteraciones sucesivas. Para analizar este comportamiento, se simuló la evolución de opiniones en una red Erdős-Rényi ($N=100$, $\langle k \rangle=6$) partiendo de una distribución inicial donde cada nodo tenía probabilidad $f=0.5$ de iniciar con la opinión $+1$.

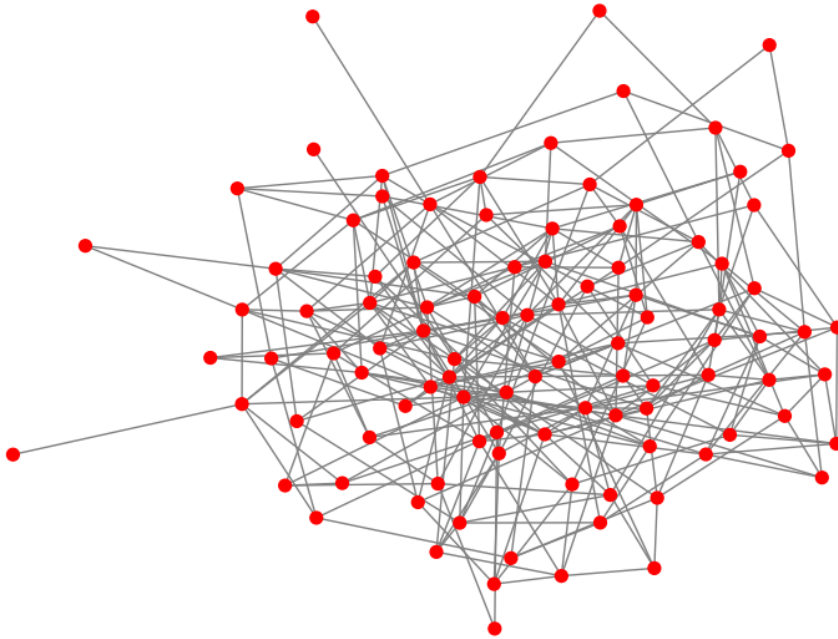
La simulación se detuvo cuando no se registraron cambios en las opiniones o se alcanzó un número máximo de 1000 pasos. El progreso se visualizó en intervalos regulares y al alcanzar el estado estable. Este enfoque permite observar cómo se distribuyen las opiniones finales en la red y evaluar el tiempo necesario para lograr consenso o coexistencia de estados.

En el caso simulado, el sistema alcanzó un estado estable después de 9 pasos, confirmando que la dinámica converge bajo las reglas definidas. La figura 4 generada muestra la configuración final de opiniones en la red.

Figura 4. Estado inicial y final de opiniones en la Red Erdős-Rényi



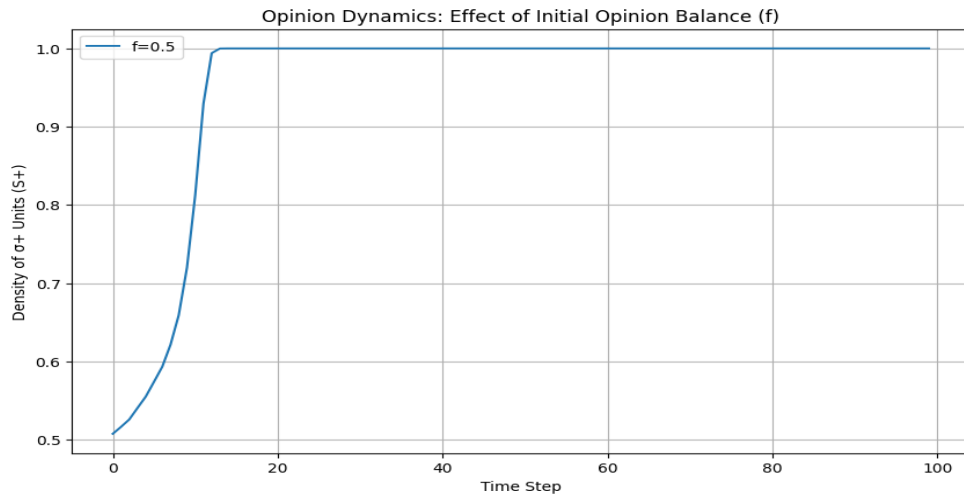
Opinion Dynamics at Step 9



Se simuló la dinámica de opiniones en una red Erdős-Rényi ($N=10,000$ y $\langle k \rangle=10$) hasta alcanzar un estado estable. La densidad de nodos $\sigma_{+}=+1$ se midió en cada paso temporal y se promediaron los resultados de 100 instancias para obtener estabilidad en las medidas.

El gráfico 5 muestra la evolución de S_{+} frente al tiempo, destacando cómo el balance inicial ($f=0.5$) influye en la convergencia hacia un consenso de opiniones. Se tarda 15 pasos en llegar a la estabilidad de la red.

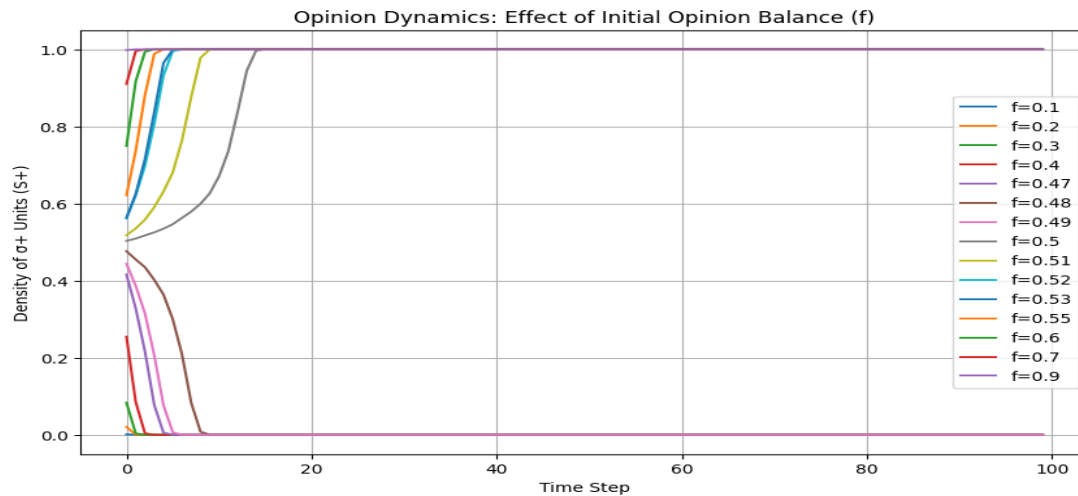
Figura 5. Evolucion de S_{+} frente al tiempo en la red Erdős-Rényi



Resultados: Efecto de f en la Dinámica de Opiniones

Se evaluaron valores de f en el rango $[0.1, 0.9]$ sobre una red Erdős-Rényi ($N=5000$ y $\langle k \rangle=10$).

Figura 6. Efecto del parámetro f en la evolución de $S+$ en la red Erdős-Rényi



La simulación alcanzó un estado estable en diferentes pasos según el valor inicial de f . Los tiempos de convergencia observados variaron entre 2 y 16 pasos, mostrando que:

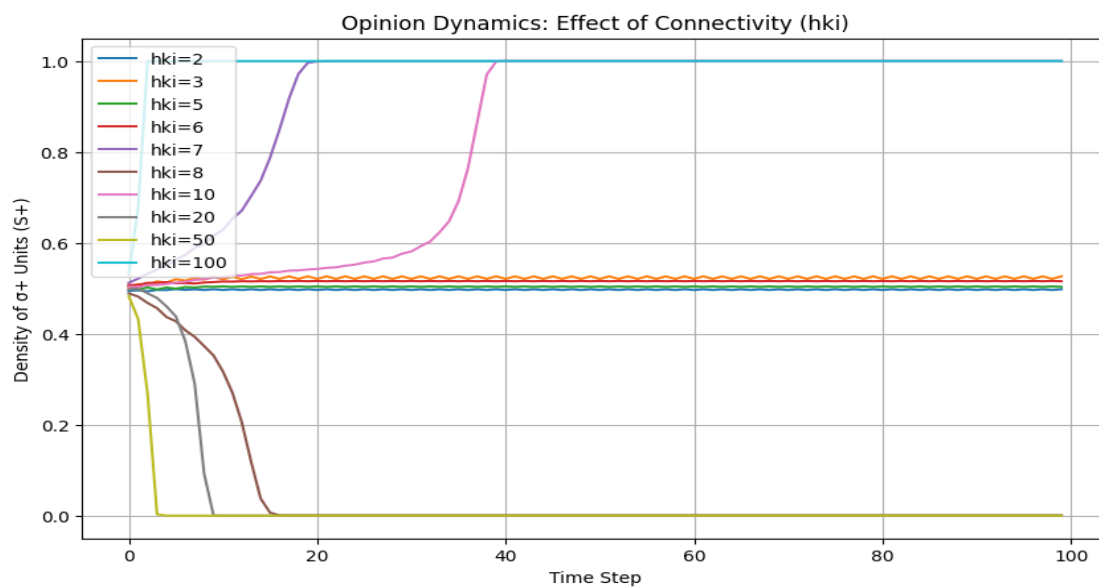
- **Valores extremos de f** (f lejos de 0.5) tienden a alcanzar un estado estable en pocos pasos, debido a la rápida dominancia de una opinión.
- **Valores cercanos a $f=0.5$** requieren más pasos para estabilizarse, reflejando una mayor competencia entre opiniones antes de llegar al consenso o a una coexistencia estable.

La distribución de tiempos de convergencia destaca la influencia del balance inicial en la dinámica del sistema.

Resultados: Efecto de la Conectividad $\langle k \rangle$ en la Dinámica de Opiniones:

Se analizaron diferentes valores del grado promedio $\langle k \rangle$ en el rango $[2, 100]$ en una red Erdős-Rényi ($N=10,000$) con $f=0.5$.

Figura 7. Efecto del parámetro $\langle k \rangle$ en la evolución de S_+ en la red Erdős-Rényi



Se observa que:

- **Baja conectividad ($\langle k \rangle \leq 5$):** El sistema tarda más en alcanzar un estado estable. En algunos casos, no converge en los 100 pasos máximos, mostrando configuraciones locales dominantes que no se propagan globalmente.
- **Conectividad intermedia ($6 \leq \langle k \rangle \leq 20$):** El sistema converge más rápidamente debido a una mayor influencia de la mayoría local.
- **Alta conectividad ($\langle k \rangle \geq 50$):** Convergencia muy rápida (~ 5 pasos) hacia el consenso, ya que los nodos están fuertemente conectados y las opiniones de la mayoría dominan rápidamente.

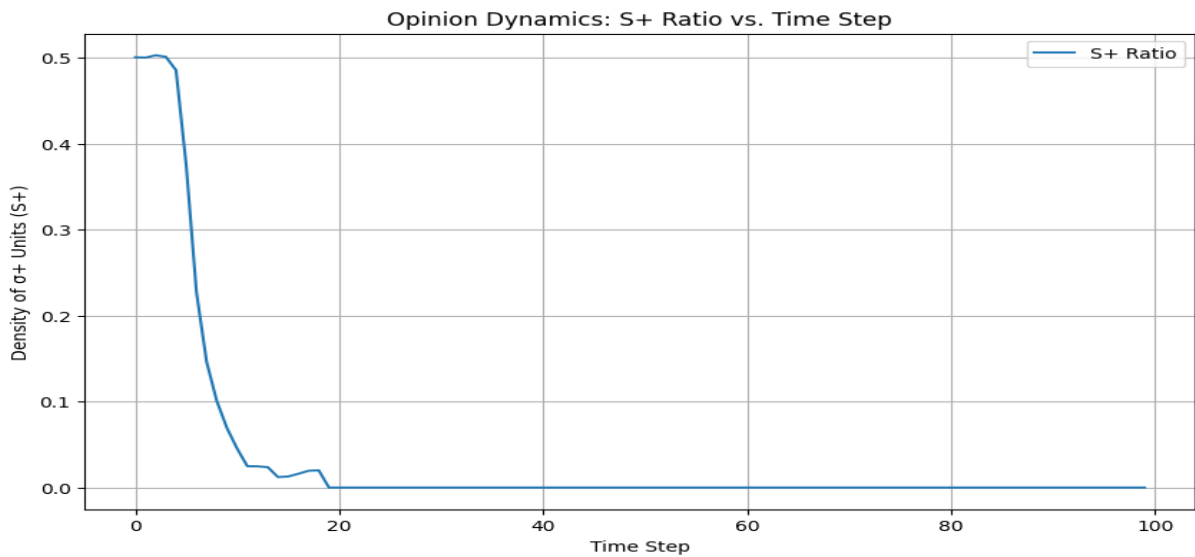
Redes Barabási-Albert (BA)

De manera análoga, se analizará la formación y dinámica de opiniones en redes libres de escala generadas mediante el modelo Barabási-Albert (BA). Este modelo crea redes con propiedades emergentes características, como hubs altamente conectados, que pueden influir significativamente en la propagación de opiniones y en la convergencia hacia estados estables.

Se implementarán las mismas reglas dinámicas: inicialización de opiniones con probabilidad f , actualización según la mayoría local, y evaluación de la evolución de la densidad $S+$ ($\sigma+ = +1$) hasta alcanzar un estado estable. Esto permitirá estudiar cómo la topología libre de escala impacta el tiempo de estabilización, el consenso, y las diferencias con las redes aleatorias Erdős-Rényi.

Dinámica de Opiniones en Redes Libres de Escala BA: Evolución del Ratio $S+$ con ($N=10,000$, $f=0.5$, $m=10$). En la figura 8 vemos la evolución Promediada del Ratio $S+$ en 100 simulaciones.

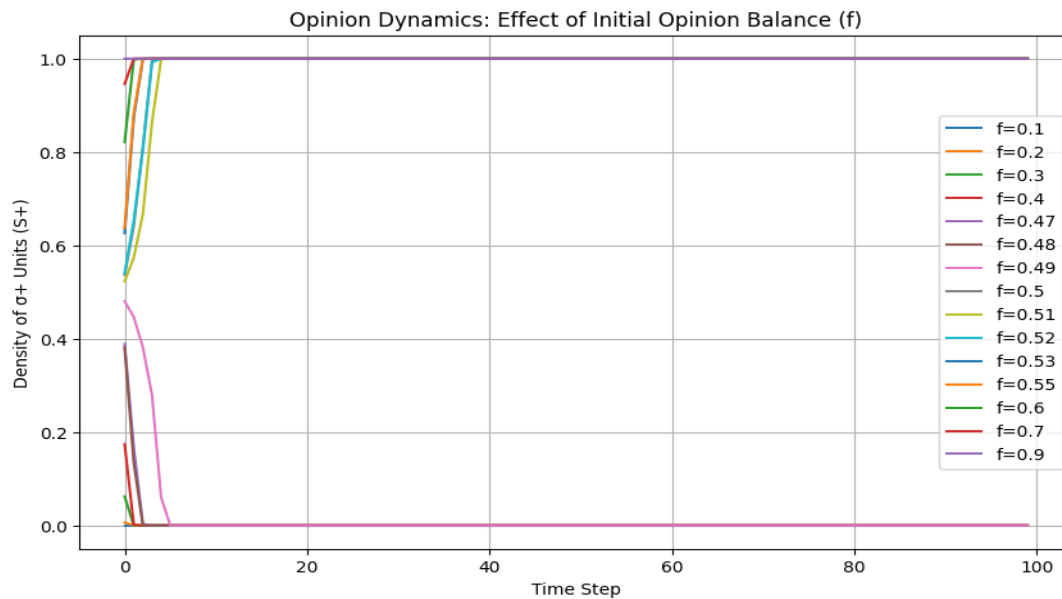
Figura 8. Evolucion de $S+$ frente al tiempo en la red Barabási-Albert



Dinámica de Opiniones en Redes BA con Diferentes Valores de f

Se analiza cómo el balance inicial de opiniones (f) afecta la evolución del sistema en redes Barabási-Albert ($N=10000$, $k=10$).

Figura 9. Efecto del parámetro f en la evolución de $S+$ en la red Barabási-Albert



La simulación muestra que:

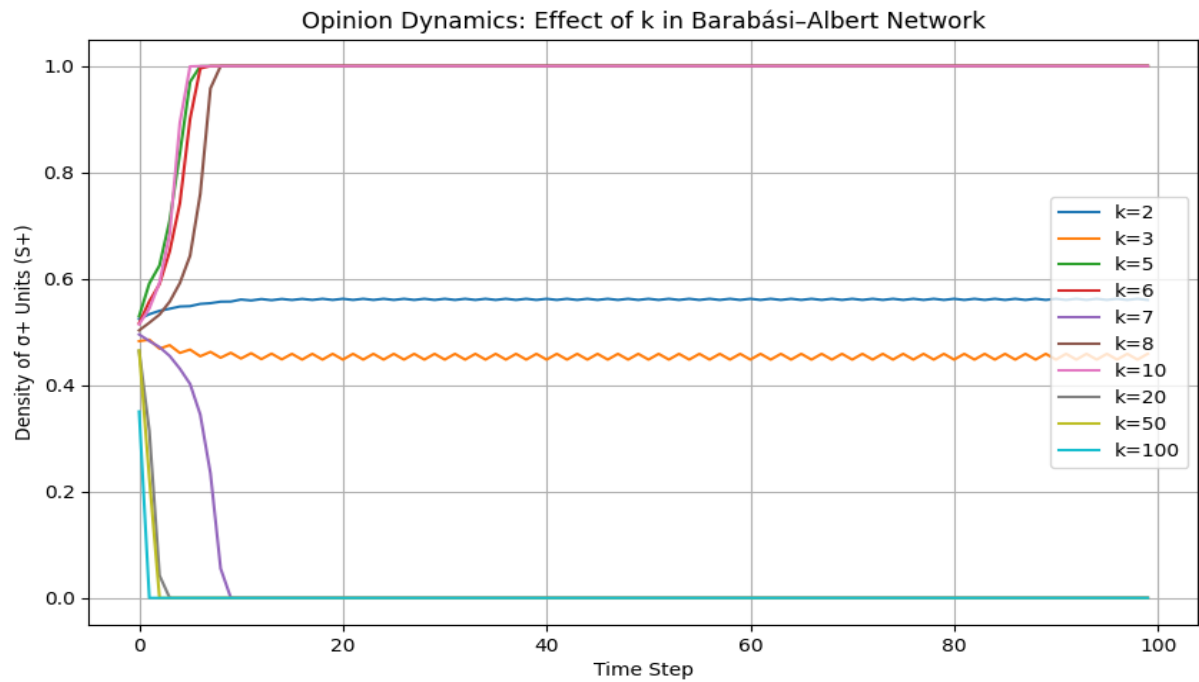
- **Valores extremos de f ($f \leq 0.3$ o $f \geq 0.7$)** conducen a una rápida convergencia hacia estados homogéneos debido a la dominancia de una opinión.
- **Valores cercanos a $f=0.5$** requieren más pasos para estabilizarse, reflejando competencia entre opiniones antes de alcanzar el equilibrio.

Los pasos necesarios para alcanzar el estado estable varían entre 1 y 6, destacando cómo la estructura de la red y los valores de f interactúan para determinar la velocidad de convergencia.

Efecto del Número de Enlaces (k) en Redes BA y la Dinámica de Opiniones

Se investigó cómo el número de enlaces (k) que cada nodo nuevo establece en una red Barabási-Albert ($N=10,000$) influye en la dinámica de opiniones.

Figura 10. Efecto del parámetro $\langle k \rangle$ en la evolución de S_+ en la red Barabási-Albert



Se observa que:

1. **Baja conectividad ($k \leq 5$):** El sistema tiende a estabilizarse lentamente, con algunos casos donde no se alcanzó un estado estable (no converge) en los pasos máximos establecidos.
2. **Conectividad moderada ($6 \leq k \leq 10$):** La estabilización ocurre más rápidamente, en 7–11 pasos, debido a una mejor propagación de las opiniones a través de la red.
3. **Alta conectividad ($k \geq 20$):** El sistema converge rápidamente hacia el estado estable en pocos pasos (3–4), ya que la red altamente conectada facilita la dominancia de la mayoría.

El gráfico obtenido muestra claramente cómo k afecta la velocidad de convergencia y la evolución del ratio $S+$ en la red.

Análisis Comparativo: Redes ER y BA con Modelo de Majority Rule

Modificado

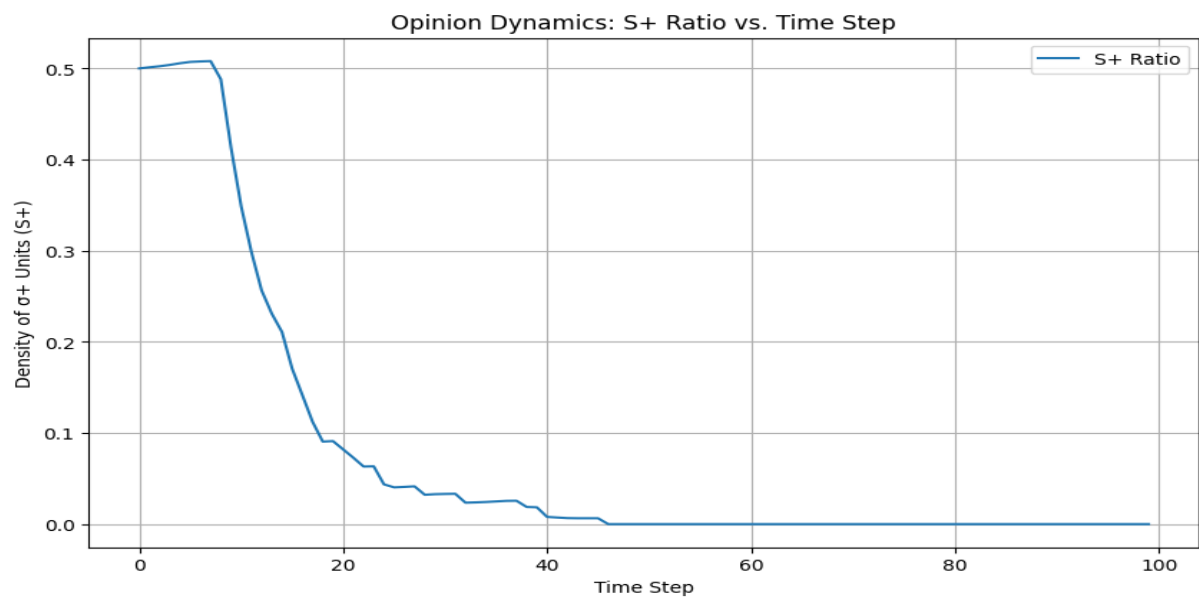
En esta sección se estudiará la dinámica de opiniones en redes Erdős-Rényi (ER) y Barabási-Albert (BA) utilizando un modelo modificado de *majority rule*. Este enfoque ajustado permitirá observar cómo afecta esta modificación del modelo de *majority rule* afectan la

propagación de opiniones en las diferentes topologías estudiadas y el tiempo necesario para alcanzar un estado estable (convergencia de opiniones).

Dinámica de Opiniones en Redes Erdős-Rényi con Regla de Mayoría Modificada"

El código simula la evolución de opiniones en redes Erdős-Rényi ($N=5000$), conectividad promedio $\langle k \rangle=10$, considerando la opinión del nodo en la regla de mayoría. Los resultados promedian 100 simulaciones, analizando la densidad de $S+$ frente al tiempo hasta alcanzar la estabilidad.

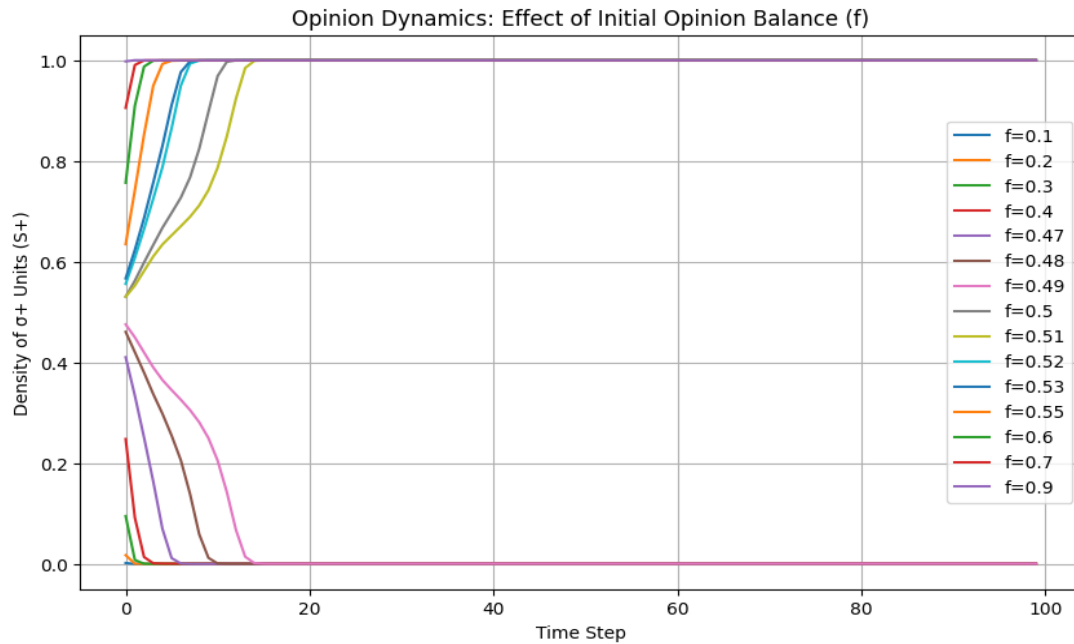
Figura 11. Evolución de $S+$ en la red Erdős-Rényi con modelo de opinión modificada



Impacto de f en la red ER modificada

Se estudió el efecto del balance inicial de opiniones (f) en redes Erdős-Rényi ($N=10,000$), $\langle k \rangle=10$), utilizando una regla de mayoría modificada. Se analizaron valores de f en el rango $[0.1, 0.9]$, midiendo la evolución de la densidad $S+$ hasta alcanzar un estado estable.

Figura 12. Evolución del parámetro f en la red Erdős-Rényi con modelo de opinión modificada

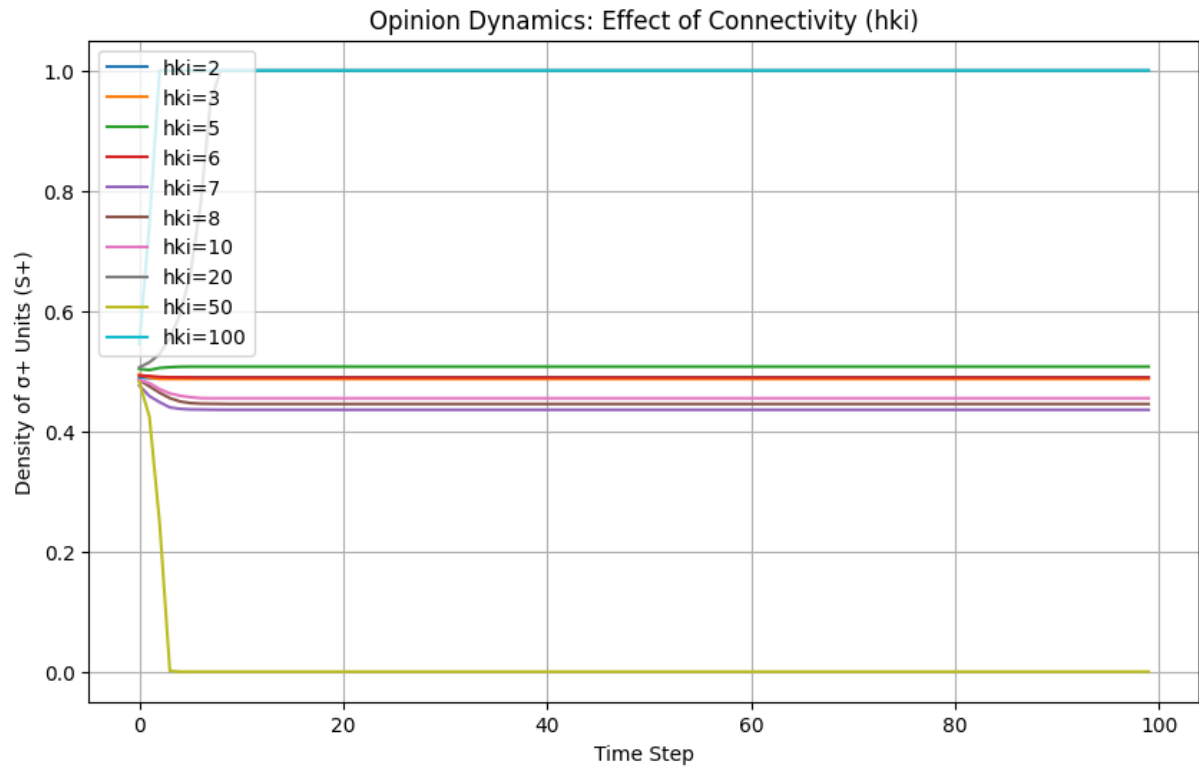


- **Valores extremos de f ($f \leq 0.3$ o $f \geq 0.7$):** El sistema converge rápidamente a un consenso dominado por una única opinión ($\sigma+ = +1$).
- **Valores cercanos a $f=0.5$:** La estabilización es más lenta debido a la competencia entre opiniones, requiriendo hasta 17 pasos para alcanzar el equilibrio.
- **Dinamismo estructural:** La topología aleatoria de la red facilita la propagación uniforme de las opiniones, pero el balance inicial (f) define el tiempo de convergencia.

Impacto de K en la red ER modificada

El impacto de la conectividad promedio ($\langle k \rangle$) en la dinámica de opiniones en redes Erdős-Rényi ($N=5000$) utilizando una regla de mayoría modificada. Los valores de $\langle k \rangle$ variaron entre $[2, 100]$, y se midió la evolución de $S+$ ($\sigma+ = +1$) hasta alcanzar un estado estable.

Figura 13. Evolución del parámetro $\langle k \rangle$ en la red Erdős-Rényi con modelo de opinión modificada



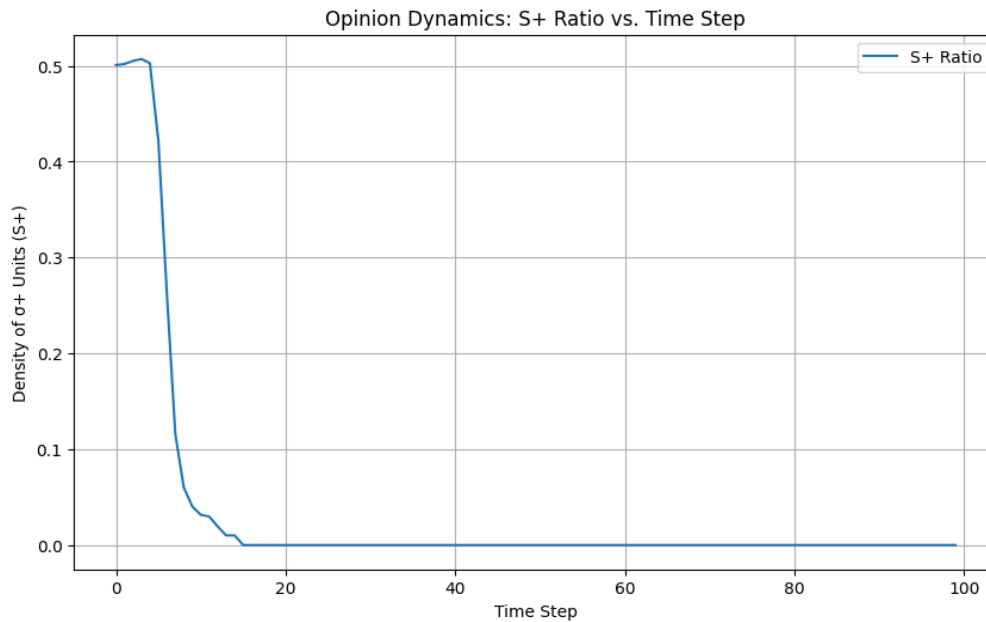
- **Baja conectividad ($\langle k \rangle \leq 5$):** El sistema tarda más en estabilizarse debido a la limitada interacción entre nodos.
- **Conectividad moderada ($6 \leq \langle k \rangle \leq 20$):** Se observa una estabilización más rápida, en 7 a 10 pasos, por una mejor propagación de opiniones.
- **Alta conectividad ($\langle k \rangle \geq 50$):** El consenso se alcanza en menos de 5 pasos, ya que la conectividad facilita la dominancia de la mayoría.

El aumento de $\langle k \rangle$ reduce el tiempo de estabilización y asegura un consenso más eficiente en la red.

Convergencia de la red BA con opinión modificada

Se analiza la dinámica de opiniones en redes libres de escala generadas con el modelo Barabási-Albert ($N=5000$ y $m=10$), utilizando una regla de mayoría modificada que incluye la opinión del nodo. Se promedian los resultados de 100 simulaciones para evaluar la evolución de S_+ ($\sigma_+ = +1$) frente al tiempo hasta alcanzar un estado estable.

Figura 14. Evolución de S_+ en la red Barabási-Albert con modelo de opinión modificada



Se observa que:

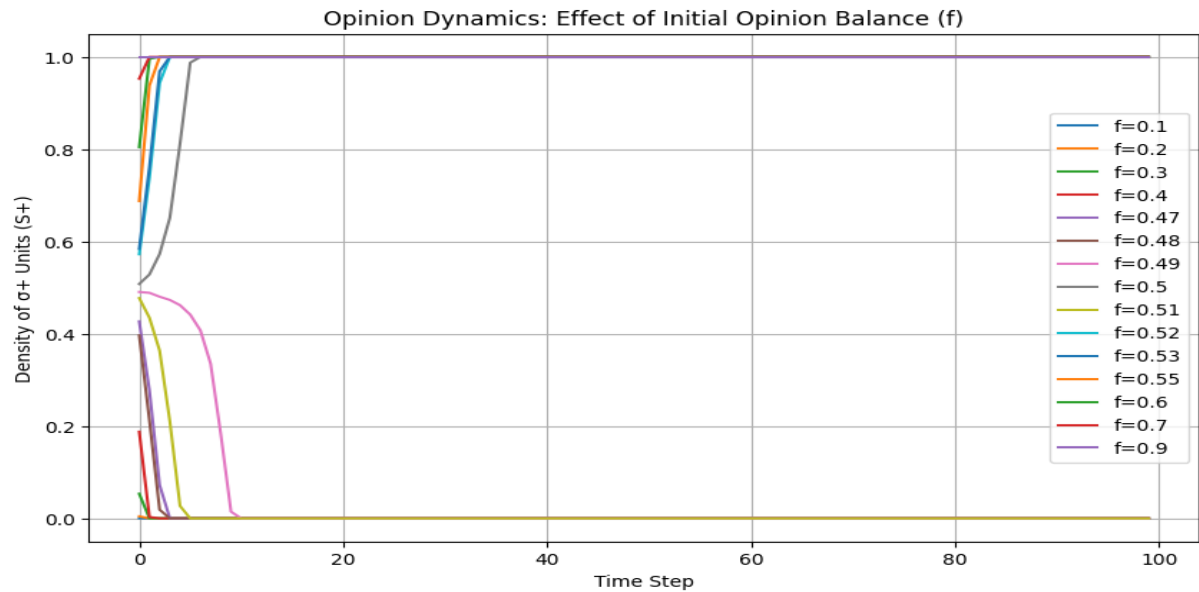
- **Estructura de *hubs*:** La presencia de nodos altamente conectados acelera la estabilización del sistema, ya que los *hubs* propagan rápidamente la opinión dominante.
- **Estabilización rápida:** En la mayoría de los casos, el sistema alcanza el estado estable en menos de 10 pasos, reflejando la eficiencia de la topología libre de escala para lograr consenso.
- **Competencia inicial ($f=0.5$):** Cuando las opiniones están balanceadas inicialmente, los hubs desempeñan un papel crítico en determinar la opinión final dominante.

El gráfico demuestra cómo la estructura jerárquica de la red impacta en la rapidez y la naturaleza de la estabilización del sistema.

Efecto de f en BA generación modificada

Se analizó el efecto del balance inicial de opiniones (f) en redes Barabási-Albert ($N=5000, k=10$), utilizando una regla de mayoría modificada que incluye la opinión del nodo. Los valores de f variaron entre $[0.1, 0.9]$, midiendo la evolución del ratio S^+ ($\sigma^+ = +1$) hasta alcanzar un estado estable.

Figura 15. Evolución de S_+ en la red Barabási-Albert con modelo de opinión modificada en función de f



Se observa que:

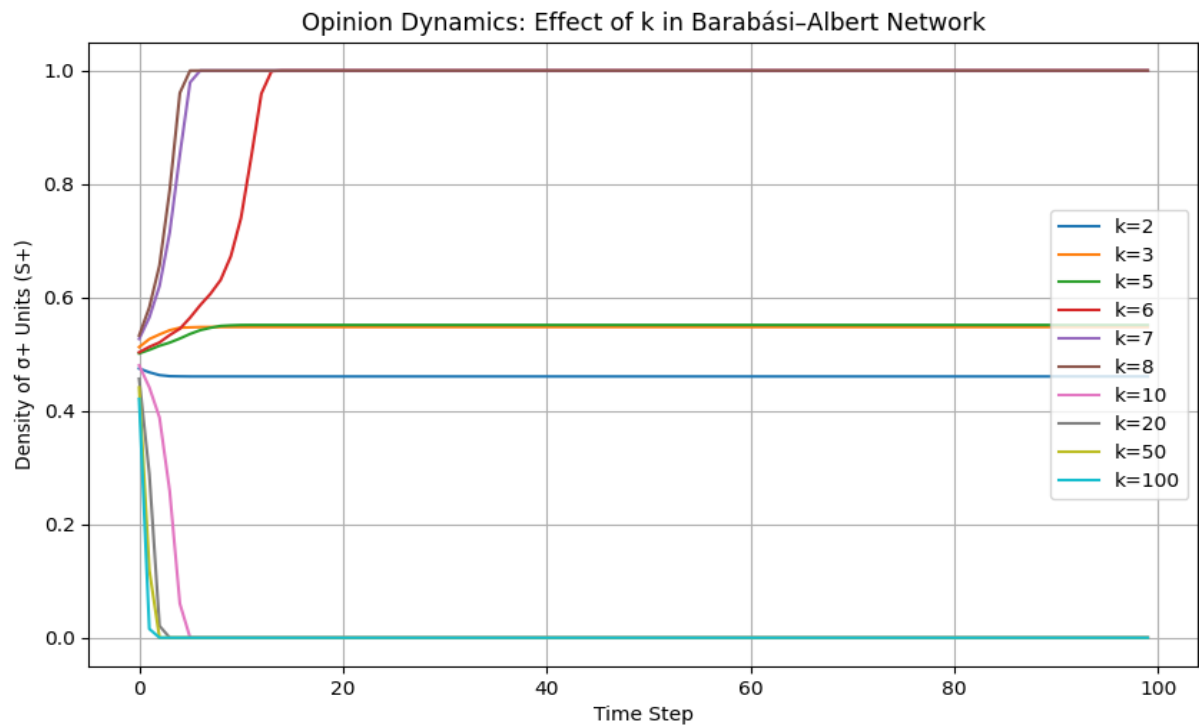
- **Valores extremos de f ($f \leq 0.3$ o $f \geq 0.7$):** El sistema converge rápidamente ($1-3$ pasos) hacia un consenso completo, dominado por una opinión.
- **Valores cercanos a $f=0.5$:** Se requiere más tiempo para estabilizarse ($6-11$ pasos), reflejando competencia inicial entre opiniones.
- **Impacto de la estructura de la red:** Los *hubs* facilitan la propagación rápida de opiniones dominantes, acelerando la convergencia.

El gráfico ilustra cómo f influye en la dinámica de estabilización en redes libres de escala.

Efecto de k en BA generación modificada

Se analiza el efecto de la conectividad (k) en la dinámica de opiniones de redes Barabási-Albert ($N=10,000$) utilizando una regla de mayoría modificada. Los valores de k variaron entre $[2, 100]$, y se evaluó el ratio S_+ ($\sigma_+ = +1$) hasta alcanzar un estado estable.

Figura 16. Evolución de S_+ en la red Barabási-Albert con modelo de opinión modificada en función de $\langle k \rangle$



- **Baja conectividad ($k \leq 5$):** La estabilización es más lenta (12–15 pasos), ya que los nodos están menos conectados, dificultando la propagación de la opinión dominante.
- **Conectividad moderada ($6 \leq k \leq 10$):** El sistema converge en 6–8 pasos, reflejando una mayor eficiencia en la difusión de opiniones.
- **Alta conectividad ($k \geq 20$):** El consenso se logra rápidamente (3–4 pasos), gracias a la fuerte conectividad que facilita la dominancia de la mayoría.

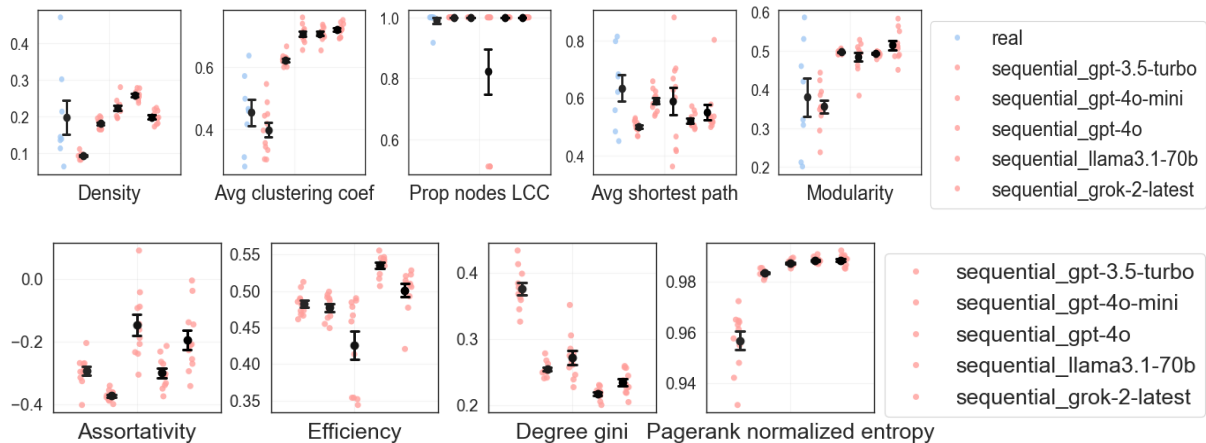
El gráfico destaca cómo k influye significativamente en la rapidez y la naturaleza de la convergencia en redes libres de escala.

4.2. Redes generadas con LLMs. Análisis.

Modelos LLMs: propiedad y comparativa con redes reales.

Se comparan las redes generadas por modelos de lenguaje (LLMs) entre si y con redes reales. Los gráficos comparan propiedades estructurales de redes reales con redes generadas por LLMs (primera imagen) y entre redes generadas exclusivamente por LLMs (segunda imagen).

Figura 17. Propiedades de las redes generadas con LLMs



Insights clave:

1. Primera imagen (red real vs. redes generadas):

- Densidad y coeficiente de agrupamiento:** Los modelos generan redes con densidad similar con la red real, excepto GPT-3.5-Turbo que presenta una menor densidad. En el coeficiente de agrupamiento GPT-3.5-Turbo es la que mejor se aproxima a la red real, mientras que los otros cuatro modelos lo sobrestiman.
- Proporción en la LCC:** en general, las redes generadas reproducen adecuadamente la conectividad global de la red real, excepto GPT-4o que presenta un valor menor y con mucha varianza.
- Camino mas corto:** las redes generadas tienen un camino mas corto ligeramente menor a la red real.
- Modularidad:** Los LLMs generan redes con valores razonables, aunque con dispersión entre modelos. El modelo GPT-3.5-turbo es el que mas se aproxima y el resto la sobrestiman.

2. Segunda imagen (comparación solo entre LLMs):

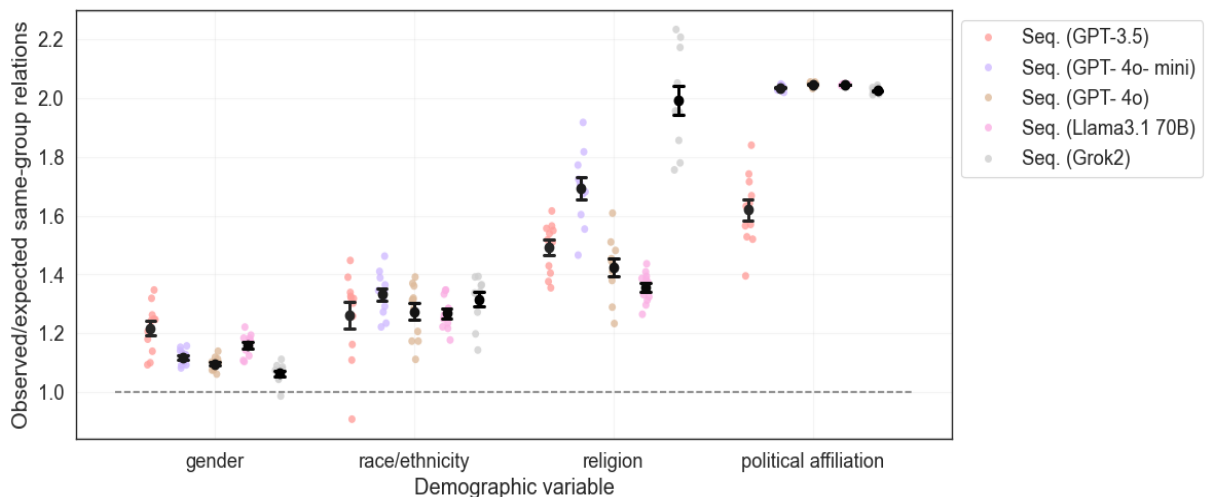
- Asortividad:** Modelos como gpt-4o y grok-2 presentan una mayor asortividad que el resto de los modelos.
- Eficiencia:** GPT-4-o presenta una eficiencia menor que el resto de LLMs.
- Grado de Gini:** El grado de Gini de la red GPT-3.5-turbo es mayor al resto de modelos.

- d. **Entropía Page Rank:** La entropía de PageRank es próxima a 1 en todos los modelos, pero en el modelo es ligeramente menor al resto.

En la imagen 17 se observa que los LLMs logran replicar parcialmente. Sin embargo, hay variaciones significativas en algunas de las métricas. Estas diferencias reflejan fortalezas y limitaciones en cómo los modelos generan redes estructurales.

El gráfico 18 muestra la relación entre las interacciones observadas y las esperadas dentro del mismo grupo para diversas variables demográficas en redes generadas por diferentes modelos de lenguaje (LLMs).

Figura 18. Homofilia de género, etnia, religión y política



Se observa que:

1. **Género:** Los modelos generan relaciones cercanas al valor esperado (relación ~ 1), lo que sugiere un equilibrio entre interacciones intra e intergrupales para esta variable. Se observa que el modelo *GPT-3.5* es el que más se aleja del valor esperado.
2. **Raza/etnicidad:** Hay una ligera tendencia hacia relaciones intra-grupales más altas que las esperadas, lo que indica una cierta preferencia por conexiones dentro del mismo grupo. Todos los modelos tienen comportamiento similares.
3. **Religión:** Los valores aumentan significativamente, con relaciones intra-grupales mucho mayores que las esperadas, mostrando una fuerte agrupación por esta variable. Especialmente *Grok-2* y, en menor medida *GPT-4o-mini*.

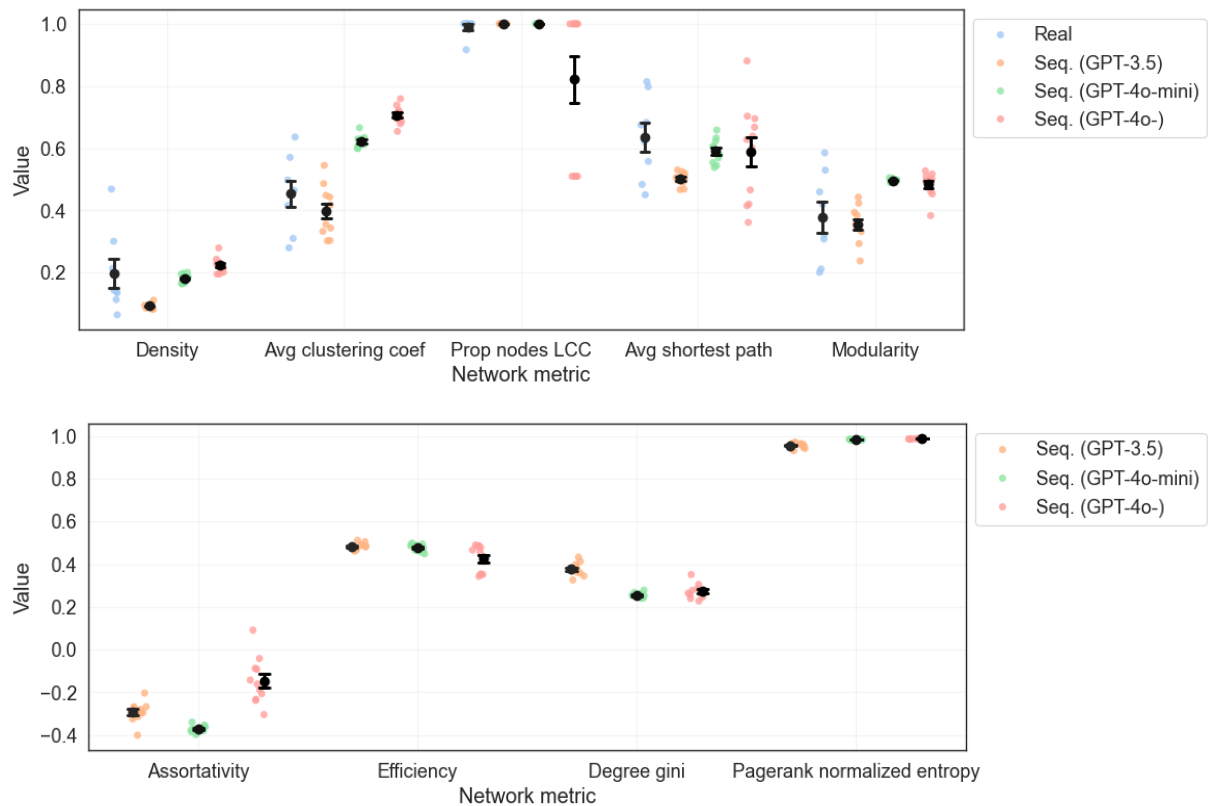
4. **Afiliación política:** Esta categoría muestra los valores más altos ($h \approx 2$), reflejando una notable homofilia, donde los nodos tienden a conectarse de manera predominante con otros del mismo grupo político. El único modelo que no presenta $h \approx 2$ es *GPT-3.5* que su valor es en torno a 1.6.

Los LLMs replican patrones de homofilia con variaciones entre las variables demográficas. Mientras que el género muestra conexiones balanceadas, la religión y la afiliación política destacan por una fuerte agrupación intra-grupal, lo que resalta diferencias en cómo los modelos capturan la segregación social según la categoría. Estas tendencias podrían reflejar sesgos subyacentes en los datos de entrenamiento o en las estructuras generadas.

Comparaciones de modelos de OpenAI

El gráfico compara métricas estructurales de redes generadas por diferentes versiones de GPTs (GPT-3.5, GPT-4o-mini, GPT-4o) frente a una red real.

Figura 19. *Métricas estructurales de los modelos de OpenAI*



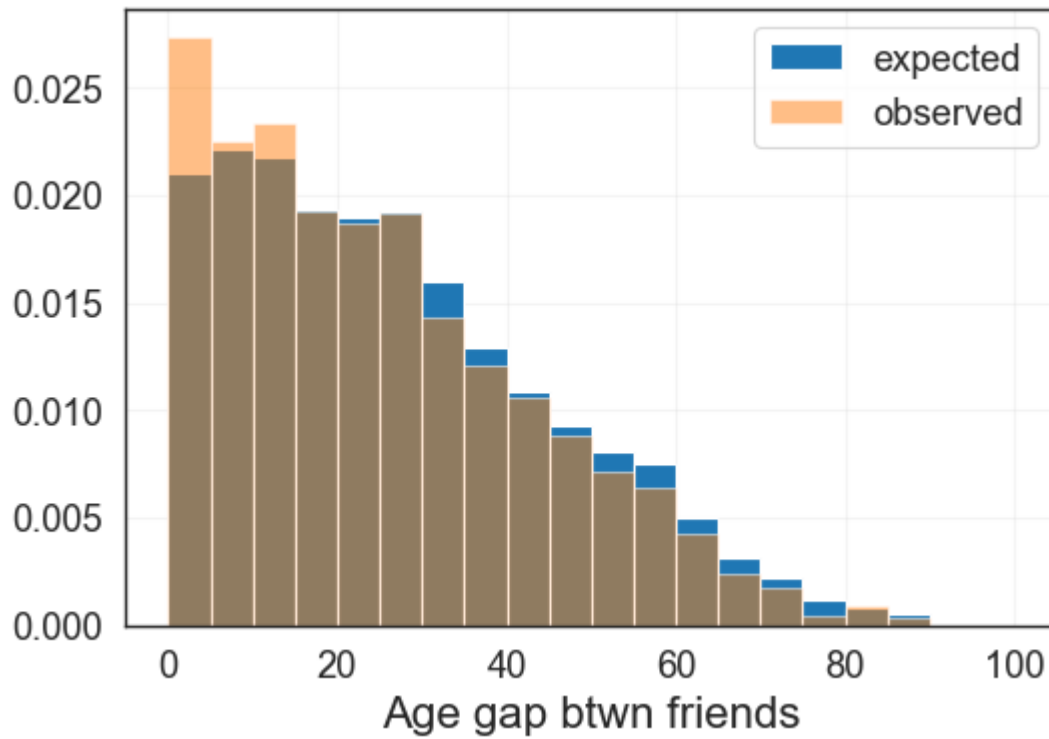
En términos de **densidad**, GPT-4o y GPT-4o-mini generan redes más densas que GPT-3.5, este último tiene valores menores respecto a la red real. Para el **coeficiente de agrupamiento promedio**, GPT-4o y GPT-4o-mini sobreestima los niveles observados en la red real, pero GPT-3.5-turbo muestra mejores aproximaciones, replicando parcialmente la formación de clústeres. La **proporción de nodos en la LCC** es reproducida con gran precisión por todos los modelos (excepto GPT-4o), logrando una conectividad consistente. En cuanto al **camino promedio más corto**, los valores generados son cercanos a los reales con ligeras variaciones entre los modelos. Respecto al modularidad, GPT-4o y GPT-4o-mini genera redes con mayor estructura comunitaria en comparación con GPT-3.5 y la red real.

En métricas más específicas, como **asortatividad**, todos los modelos presentan valores negativos similares, indicando que las conexiones tienden a ser entre nodos de diferentes grados. Para **eficiencia** y **grado Gini**, las diferencias entre los modelos son menores. Finalmente, la **entropía normalizada del PageRank** es consistente entre modelos siendo esta próxima a 1.

GPT-4o destaca por su capacidad para aproximarse mejor a propiedades estructurales complejas, como agrupamiento y modularidad, mientras que GPT-3.5 y GPT-4o-mini muestran un desempeño más limitado. A pesar de ello, ningún modelo logra replicar completamente todas las métricas de la red real, especialmente en la formación de clústeres, indicando margen de mejora en la modelación de redes realistas.

Homofilia por edad:

Figura 20. *Homofilia por edad*

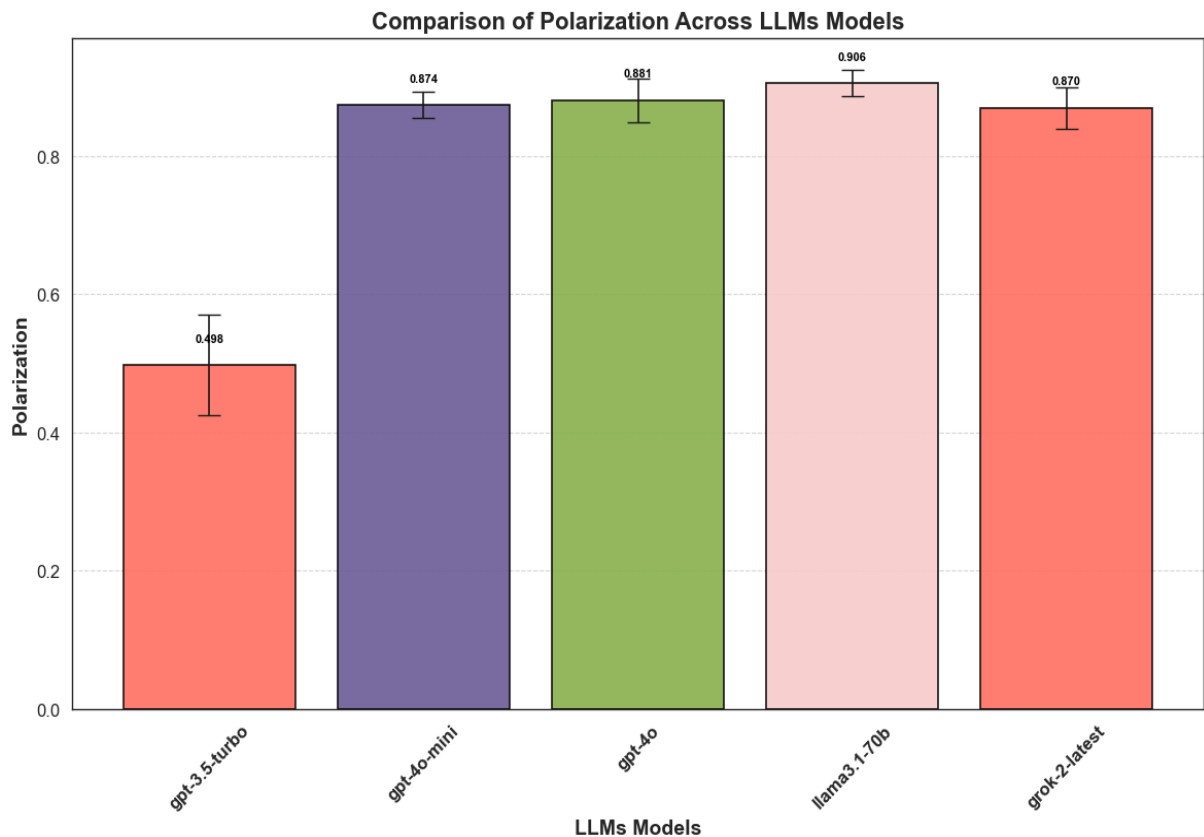


El gráfico 20 evidencia una homofilia por edad en las conexiones sociales. Las relaciones observadas son significativamente más altas que las esperadas para diferencias de edad pequeñas (0-5 años), indicando una fuerte preferencia por conectar con personas de edades similares. A partir de los 20 años de diferencia, las frecuencias observadas se aproximan a los valores esperados, mostrando que las conexiones con grandes diferencias de edad son menos comunes, pero se ajustan mejor a lo esperado. Esto refleja tanto patrones sociales como estructurales en la red analizada.

Comparación de Polarización entre Modelos LLM

El gráfico 21 muestra la polarización generada por diferentes modelos de lenguaje (LLMs) y la compara con la polarización observada en redes reales como Twitter, que oscila entre 0.33 y 0.42.

Figura 21. Polarización de los LLMs



Observamos que la polarización es

1. Polarización LLM cercana a la de Twitter:

- gpt-3.5-turbo, con un valor de polarización de 0.498, es el modelo que más se aproxima a los niveles de polarización observados en Twitter. Aunque está ligeramente por encima del rango real, sigue siendo relativamente cercano.

2. Polarización LLM sensiblemente superior a la real de Twitter:

- Los modelos de OpenAI más recientes **gpt-4o-mini** (polarización de 0.874) y **gpt-4o** (polarización de 0.881) muestran un aumento considerable en la polarización, generando redes más divididas en comparación con Twitter.
- llama3.1-70b**, con un valor de polarización de 0.906, es el modelo que genera las redes más polarizadas, seguido de grok-2 (polarización de 0.870).

Esto sugiere que llama3.1-70b tiende a crear estructuras sociales más segmentadas y menos interconectadas.

Comparación de Redes Generadas por GPT-3.5-turbo con Modelos Clásicos

El análisis compara las redes generadas por GPT-3-Sequential (el modelo que tiene menos homilía y polarización política, y parece que ha generado estructuras de redes sociales más realistas) con tres modelos clásicos de generación de redes: Erdős-Rényi (ER), Barabási-Albert (BA) y Watts-Strogatz (WS), considerando diversas métricas estructurales clave.

Principales observaciones:

- **Densidad:** GPT-3.5-turbo tiene una desviación media mayor (mean_diff_norm = 0.857) que ER (0.002), BA (0.540), y WS (0.624), indicando que las redes generadas son más densas que las esperadas en modelos clásicos.
- **Coeficiente de agrupamiento promedio:** GPT-3.5-turbo muestra una desviación menor (0.502) en comparación con BA (1.926) y ER (2.291), pero es mayor que WS (0.006), lo que sugiere que las redes generadas por GPT son más agrupadas que ER y BA, pero menos que WS.
- **Proporción de nodos en la componente conexa más grande (LCC):** GPT-3.5-turbo y todos los modelos clásicos muestran un comportamiento similar (mean_diff_norm = 0.378), indicando que la mayoría de los nodos se agrupan en una única componente en todos los casos.
- **Camino promedio más corto:** GPT-3.5-turbo (1.072) presenta un valor intermedio, con menor desviación que ER (1.919) y BA (1.569), pero mayor que WS (0.926).
- **Modularidad:** GPT-3.5-turbo (0.183) genera redes con menor desviación en modularidad que BA (1.326), ER (1.627), y WS (1.156), mostrando una estructura comunitaria más consistente.

Las redes de formadas por ChatGPT-3.5 tienen propiedades más densas y estructuradas que ER y BA, pero menos agrupadas que WS. Este equilibrio refleja la capacidad del modelo para capturar características más complejas que los modelos clásicos, aunque con ciertos sesgos hacia una mayor densidad y modularidad.

Dinámica de la opinión política de las redes generadas con LLMs mediante el Modelo de Majority Rule simple y modificado

De manera similar a lo realizado al inicio del presente trabajo con las redes de Erdős-Rényi y Barabási-Albert, he aplicado el modelo de Majority Rule y su versión modificada a redes generadas mediante LLMs.

Sin embargo, el procedimiento difiere notablemente debido a las características intrínsecas de las redes generadas con LLMs. En estas redes, los nodos ya tienen asignada una opinión binaria (en este caso, demócratas o republicanos), lo que elimina la necesidad de realizar un paso previo para asignar opiniones a los nodos o determinar el porcentaje de prevalencia de cada una de ellas. Además, tampoco es necesario especificar el número de conexiones que debe tener cada nodo, ya que el propio LLM, a través del *prompt* utilizado, genera las conexiones de manera autónoma según su diseño interno. Por lo tanto, las redes creadas por LLMs no requieren la consideración explícita de variables estructurales del modelo, como ocurre en los modelos clásicos de generación de redes.

El único paso necesario en este caso consiste en realizar la simulación en la que los nodos intentan convencer a sus vecinos de cambiar de opinión siguiendo las reglas del modelo de Majority Rule y del modelo de Majority Rule modificado. Posteriormente, se analiza el tiempo necesario para que la red alcance un estado estable, es decir, el momento en el que ya no se producen más cambios de opinión entre los nodos.

Para este análisis, se seleccionaron cinco redes distintas, cada una generada con un modelo de lenguaje específico: gpt-3.5-turbo, gpt-4o-mini, gpt-4o, grok-2-latest y llama3.1-70b. Se realizaron las simulaciones correspondientes a ambas variantes del modelo de Majority Rule.

El resultado obtenido es sorprendente: **todas las redes formadas por LLMs analizadas son estáticas**. Tanto con el modelo de Majority Rule simple como con su versión modificada, no se observó ningún cambio en las opiniones de los nodos. Esto significa que, en las redes generadas con los cinco modelos de lenguaje anteriormente mencionados, todos sus nodos tienen al menos la mitad de sus conexiones con otros nodos que comparten la misma opinión política. Por tanto, podemos afirmar con rotundidad que estas redes son inherentemente estáticas bajo los modelos de opinión considerados.

5. Conclusiones y trabajo futuro

5.1. Conclusiones

Primeramente, he realizado un estudio de dos tipos de redes: una red aleatoria generada con el algoritmo de Erdős-Rényi y una red libre de escala generada con el algoritmo de Barabási-Albert. En la red de Erdős-Rényi, los nodos siguen una distribución de Poisson, lo que implica que la mayoría de los nodos tienen un número de conexiones similar. Por otro lado, la red libre de escala Barabási-Albert sigue una distribución de potencia, caracterizada por la existencia de pocos nodos altamente conectados (hubs) y muchos nodos con pocas conexiones. La presencia de estos hubs reduce significativamente el promedio de distancias entre nodos. A ambas redes se les aplicaron dos modelos: el modelo de Majority Rule y una versión modificada de este. En el estudio se analizaron las dinámicas de ambos modelos, específicamente:

- El tiempo necesario para que las redes converjan a uno de los estados de opinión: Las redes Barabási-Albert (BA) convergen más rápido que las Erdős-Rényi debido a la presencia de hubs (nodos altamente conectados) que actúan como amplificadores de opinión y pueden influenciar simultáneamente a muchos otros nodos. La estructura jerárquica y la distribución de ley de potencia característica de las redes BA permite una propagación más eficiente de la opinión mayoritaria a través de estos nodos centrales.
- El impacto del parámetro f , porcentaje inicial de cada una de las dos opiniones binarias en la red: En la red Erdős-Rényi, la convergencia es más lenta cerca de $f=0.5$ debido al equilibrio inicial de opiniones, mientras que valores extremos de f facilitan una rápida uniformidad. En la red Barabási-Albert, los *hubs* aceleran la difusión, permitiendo una convergencia más rápida incluso para valores intermedios de f .
- La influencia del número de nodos en la red sobre el comportamiento observado: En la red Erdős-Rényi, un mayor grado medio (k) acelera la convergencia, mientras que, para grados bajos, las opiniones se dividen y el proceso es más lento o, incluso, no llega a una convergencia de opiniones. En la red Barabási-Albert, incluso con grados bajos, los *hubs* aseguran una difusión de la opinion eficiente, y el aumento de k acelera aún más la convergencia.

- Comparación entre el modelo Majority Rule y el modelo Majority Rule Modificado:
Con el modelo Majority Rule la opinión converge más rápido en ambas redes, con difusión uniforme en Erdős-Rényi y acelerada por los hubs en Barabási-Albert. En cambio, el modelo modificado introduce mayor resistencia al cambio al incorporar opiniones individuales, ralentizando la convergencia, especialmente en Erdős-Rényi con f cercanos a 0.5 o grados bajos. En Barabási-Albert, los hubs mitigan parcialmente este efecto, pero la diversidad inicial persiste más tiempo. También en el modelo de Majority Rule modificado, la resistencia al cambio por opiniones individuales se amplifica con un mayor número de nodos, ralentizando la convergencia, especialmente en redes Erdős-Rényi.

En la segunda parte del trabajo, se generaron redes utilizando LLMs, se analizó su estructura, se estudiaron sesgos como la homofilia y la polarización, y se simuló su dinámica bajo el modelo Majority Rule y su versión modificada:

- Redes Generadas: Estructura de las redes generadas mostraron una fuerte estructura de comunidades (alta modularidad), consistente con redes reales. Muestra un alto *clustering* (los amigos de un nodo suelen ser amigos entre sí), asegura una conectividad sólida con la mayoría de nodos en la componente más grande, mantiene caminos cortos entre nodos (estructura de "mundo pequeño") y crea comunidades claras con modularidades similares a redes reales, aunque ligeramente más bajas.
- Homofilia y polarización. Las redes generadas con LLMs presentan una clara homofilia política, exagerando las conexiones entre individuos de la misma opinión política y una polarización mayor a la observada en redes reales. Además, este sesgo indica posibles estereotipos políticos presentes en los datos de entrenamiento del modelo.
- Diferencias entre modelos: Sorprendentemente, el modelo más pequeño, GPT-3.5-turbo, es el que mejor reproduce las características estructurales de las redes sociales reales, superando a GPT-4o-mini (la evolución) de GPT-3.5-turbo, a los modelos más grandes (GPT-4o-mini, llama3.1-70b y grok-2). Parece que cuanto mejor es el modelo en los *benchmarks* convencionales para evaluar LLMs, más sesgadas son las redes que genera.

- LLMs vs redes clásicas: el LLM analizado refleja la capacidad del modelo para capturar características más complejas que los modelos clásicos, aunque con ciertos sesgos hacia una mayor densidad y modularidad.
- Dinámica de las redes bajo modelos de Majority Rules: También es sorprendente que las redes sociales generadas son estáticas desde el punto de vista de la dinámica de opiniones, tanto al usar el modelo estándar de Majority Rule como su versión modificada. Esto se debe a que ya existe una dinámica no aleatoria en la formación de la red, donde las personas con opiniones binarias similares (en este caso políticas) ya han formado nexos entre sí, lo que da lugar a una red estable en su estado de opinión.

5.2. Líneas de trabajo futuro

Dentro del campo de ciencia de redes, este trabajo ha realizado un análisis exhaustivo del análisis de la propagación de opiniones y las propiedades de las redes generadas mediante algoritmos clásicos y LLMs. Sin embargo, creo que algunas líneas de investigación y posibles mejoras posibles de temas relacionados con el trabajo son las siguientes:

- Dinámicas de opinión en redes dinámicas (que la estructura de red evoluciona en el tiempo). La mayoría de los estudios de propagación de opiniones asumen redes estáticas, donde las conexiones entre nodos no cambian. Sin embargo, en redes sociales reales, las interacciones evolucionan con el tiempo: nuevas conexiones se crean, otras desaparecen y la intensidad de las relaciones varía. Incorporar esta temporalidad permitiría analizar cómo estos cambios estructurales afectan la difusión de opiniones y la estabilidad del consenso (Holme & Saramäki, 2012). Estudios previos han demostrado que redes dinámicas pueden alterar significativamente los patrones de propagación, generando efectos no observables en modelos estáticos (Masuda & Lambiotte, 2016; Perra et al., 2012).
- Análisis de múltiples opiniones: Expandir el modelo para considerar dinámicas con más de dos opiniones, que representan escenarios más cercanos a contextos reales (Xia, Wang, & Zheng, 2011).

- **Simular con distintas modificaciones del modelo Majority Rule:** Incorporar nuevas características al modelo, como la confianza en las opiniones, la resistencia al cambio en función del tiempo y la capacidad de minorías persistentes de influir, podría ofrecer una dinámica más sofisticada y rica de las dinámicas reales de opinión (Mobilia, 2003).
- **Usar otros LLMs y evaluar su capacidad para generar redes sociales reales:** Evaluar modelos de DeepSeek, Claude y Gemini podría aportar nuevas estrategias para mejorar la generación de redes. DeepSeek-R1 ha optimizado el razonamiento en LLMs (DeepSeek-AI, 2025). Claude 3, de Anthropic, se ha centrado en la generación de modelos con valores humanos y la reducción de sesgos (Anthropic, 2024). Gemini (desarrollado por Google DeepMind) se utiliza para tareas de generación y modelado de estructuras complejas, lo que puede influir en la formación de redes más realista (Gemini Team, 2023).
- **Evaluación de diferentes arquitecturas de LLMs:** En vez de centrarse en el estudio de los últimos modelos de propósito general, se podría estudiar las arquitecturas de los modelos, -como los Transformers estándar (Vaswani et al., 2017) o variantes como Sparse Transformers (Child et al., 2019)-, influyen en las propiedades estructurales y los sesgos de las redes generadas. Comparar estas arquitecturas puede ofrecer nuevas perspectivas para mejorar la calidad de las redes.
- **Fine-tuning y reducción de sesgos en LLMs:** Aplicar fine-tuning dirigido específicamente a la tarea de generación de redes podría mitigar los sesgos observados. Al ajustar el modelo con datos diseñados para reflejar estructuras de red más realistas y menos polarizadas, se podría mejorar significativamente la calidad de las redes generadas (Solaiman & Dennison, 2021).
- **Mitigación del sesgo de homofilia política:** dado que los LLMs tienden a exagerar la homofilia y la polarización política en las redes generadas, sería valioso explorar métodos para reducir este sesgo. Esto podría incluir el diseño de *prompts* (*prompt engineering*) que prioricen otros tipos de homofilia y reduzcan la homofilia política.
- **Mejorar la diversidad y la escalabilidad de las redes generadas con LLMs.** Están suelentes tener características muy uniformes, como grados similares entre nodos, lo que no refleja la variedad encontrada en redes sociales reales. Por ejemplo, en una red real, algunos nodos (personas) pueden tener cientos de conexiones mientras que

otros tienen muy pocas, formando una distribución más compleja. Además, las redes generadas enfrentan problemas al aumentar de tamaño debido a las limitaciones del modelo. Desarrollar técnicas que permitan representar esta diversidad y generar redes más grandes sería un paso importante hacia modelos más realistas

- **Mejoras en la diversidad y la escalabilidad en las redes generadas con LLM:** estas redes tienen menos variabilidad en las métricas estructurales que las redes sociales reales, como la distribución de grados o la modularidad, lo que limita su capacidad para representar la complejidad de redes reales. el método utilizado requiere listar todas las personas (nodos) en el *prompt*, lo que no es práctico para redes grandes debido a las restricciones de contexto de los LLMs, como las ventanas limitadas de tokens.
- **Aplicar las redes generadas para realizar simulaciones reales:** Analizar campañas de (des)información, campañas políticas, propagación de epidemias o estrategias de marketing digital.

Referencias bibliográficas

- Galam, S. (2002). Minority opinion spreading in random geometry. *The European Physical Journal B*, 25(4), 403–406. <https://doi.org/10.1140/epjb/e20020045>
- DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121. <https://doi.org/10.2307/2285509>
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512. <https://doi.org/10.1126/science.286.5439.509>
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137. [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6, 290–297.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442. <https://doi.org/10.1038/30918>
- You, J., Ying, R., Ren, X., Hamilton, W., & Leskovec, J. (2018). GraphRNN: Generating realistic graphs with deep auto-regressive models. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)* (pp. 5708–5717). <https://proceedings.mlr.press/v80/you18a.html>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5), 175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>
- Das, A., Gollapudi, S., & Munagala, K. (2014). Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)* (pp. 403–412). <https://doi.org/10.1145/2556195.2556233>
- Dandekar, P., Goel, A., & Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15), 5791–5796. <https://doi.org/10.1073/pnas.1217220110>
- Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2015). Dynamics of conflicts in Wikipedia. *PLOS ONE*, 7(6), e38869. <https://doi.org/10.1371/journal.pone.0038869>

- Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. En Proceedings of the 7th Python in Science Conference (SciPy2008) (pp. 11–15).
- Sagduyu, Y. E., Grushin, A., & Shi, Y. (2018). Synthetic social media data generation. IEEE Transactions on Computational Social Systems, 5(3), 605–620.
<https://doi.org/10.1109/TCSS.2018.2856442>
- Pérez-Rosés, H., & Sebé, F. (2015). Synthetic generation of social network data with endorsements. Journal of Simulation, 9(4), 279–286.
<https://doi.org/10.1057/jos.2014.37>
- Meysam Alizadeh. 2017. Generating and analyzing spatial social networks. Computational and Mathematical Organization Theory, 23:362–390.
- Newman, M. E. J. (2010). Networks: An introduction. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nature Human Behaviour, 4(6), 588–596.
<https://doi.org/10.1038/s41562-020-0898-6>
- Marios Papachristou and Yuan Yuan. 2024. Network formation and dynamics among multi-LLMs. arXiv preprint arXiv:2402.10659.
- Barrat, A., Barthelemy, M., & Vespignani, A. (2008). Dynamical processes on complex networks. Cambridge University Press.
- Barrett, C. L., Beckman, R. J., Khan, M., Kumar, V. S. A., Marathe, M. V., Stretz, P. E., Dutta, T., & Lewis, B. (2009). Generation and analysis of large synthetic social contact networks. En Proceedings of the 2009 Winter Simulation Conference (WSC'09) (pp. 1003–1014)
- Chang, S., Chaszczewicz, A., Wang, E., Josifovska, M., Pierson, E., & Leskovec, J. (2024). LLMs generate structurally realistic social networks but overestimate political homophily.
- Barrat, A., Barthelemy, M., & Vespignani, A. (2008). Dynamical processes on complex networks. Cambridge University Press.
- Shao, J., Havlin, S., & Stanley, E. (2009). Dynamic networks and their properties. Physical Review Letters, 103(1), 018701. <https://doi.org/10.1103/PhysRevLett.103.018701>

- Krapivsky, P. L., & Redner, S. (2003). Dynamics of growing networks. *Physical Review Letters*, 90(23), 238701. <https://doi.org/10.1103/PhysRevLett.90.238701>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Technical Report, Stanford InfoLab. Retrieved from <http://ilpubs.stanford.edu:8090/422/>
- Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97-125.
- Masuda, N., & Lambiotte, R. (2016). A guide to temporal networks. World Scientific Publishing.
- Perra, N., Gonçalves, B., Pastor-Satorras, R., & Vespignani, A. (2012). Activity driven modeling of time varying networks. *Scientific Reports*, 2, 469.
- Xia, H., Wang, C., & Zheng, X. (2011). The public opinion dynamics model based on the social impact theory. *Physica A: Statistical Mechanics and its Applications*, 390(4), 910-921.
- Mobilia, M. (2003). Does a single zealot affect an infinite group of voters? *Physical Review Letters*, 91(2), 028701. <https://doi.org/10.1103/PhysRevLett.91.028701>
- Solaiman, I., & Dennison, C. (2021). Process for adapting language models to society (PALMS) with values-targeted datasets. *Advances in Neural Information Processing Systems*, 34, 5861–5873.
- DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. <https://arxiv.org/abs/2501.12948>.
- Anthropic. (2024). *Claude 3: Technical report*. https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf
- Gemini Team, Google. (2023). Gemini: A Family of Highly Capable Models for Complex Task Generation. <https://arxiv.org/abs/2312.11805>
- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998-6008
<https://doi.org/10.48550/arXiv.1706.03762>
- Child, R., Gray, S., Radford, A., & Sutskever, I. (2019). Generating Long Sequences with Sparse Transformers. arXiv preprint arXiv:1904.10509.
<https://doi.org/10.48550/arXiv.1904.10509>

Anexo A. Código fuente y datos analizados

Por favor, encuentre en el siguiente repositorio <https://github.com/CarlosGarcia9/-Social-Network-Science> el código utilizado para realizar las simulaciones.

- En la carpeta llamada *“Social dynamics: majority rules for opinion formation in complex networks”* se encuentra el código correspondiente a la primera parte del trabajo, relacionado con la dinámica de opiniones mediante el modelo de mayoría y el modelo de mayoría modificado aplicado a redes de Erdős-Rényi y Barabási-Albert.
- En la carpeta llamada *“Network Generation with LLMs”* se encuentra el código generado para la creación y análisis de las redes producidas mediante LLMs.

Anexo B. Costes de Generar las Redes Sociales con LLMs

En la inteligencia artificial generativa, un factor importante es el coste de generar las redes. A continuación, muestro en una tabla el coste de cada uno de las 11 redes generadas por cada modelo.

Tabla 7. Coste de la generación de redes usando LLMs.

Modelo	Coste (en dolares)
GPT 4.0	\$5.85
GPT 4.0 - mini	\$0.42
GPT 3.5 - turbo	\$0.29
Grok-2	\$9.98
llama3.1-70b	\$5.35

Podemos ver que los modelos grandes (GPT-4.0, Grok-2 y llama3.1-70b) son notablemente más caros (especialmente grok-2) que los modelos más pequeños (GPT 4.0 - mini y GPT 3.5 - turbo).

Insistir que este es el coste de la simulación final de los 11 modelos. El coste total durante el desarrollo del trabajo ha sido mayor debido a que el consumo de las diferentes APIs debido a pruebas y errores durante el proceso.

Anexo C. Temperatura de los LLMs

La temperatura es un parámetro en los modelos de lenguaje (LLMs) que controla el nivel de aleatoriedad en las respuestas generadas. Valores más bajos de temperatura (por ejemplo, 0.3) hacen que el modelo sea más conservador, favoreciendo las opciones con mayor probabilidad, mientras que valores más altos (como 1.3) introducen mayor diversidad y creatividad en las respuestas.

En el presente trabajo, he probado diferentes valores de temperatura en el rango [0.8, 1.1], con el objetivo de observar si este ajuste influía significativamente en las redes generadas por los LLMs. Sin embargo, los resultados no mostraron diferencias apreciables entre las redes obtenidas, lo que sugiere que el rango seleccionado no afecta de manera sustancial la estructura generada en este caso. El trabajo ha sido realizado a una temperatura $T=0.9$.