



Pontificia Universidad
JAVERIANA
Cali

Facultad de Ingeniería
y Ciencias

Acta de Correcciones al Proyecto de Grado Ingeniería de Sistemas y Computación

Fecha: 21 de junio 2023

Autores: Nicolas Alberto Ortiz Aristizabal

Nombre del Proyecto de Grado:

Reconocimiento y cuantificación de polarización ideológica en redes sociales.

Director: Dr. Frank Valencia

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.

Firma de Director(a) del Proyecto de Grado

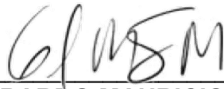
Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado
en cumplimiento de los requisitos exigidos por la
Pontificia Universidad Javeriana para optar el
título de Ingeniero de Sistemas y Computación.



Dr. CAMILO ROCHA

Decano de la Facultad de Ingeniería



ING. GERARDO MAURICIO SARRIA

Director Carrera Ingeniería Sistemas y Computación.



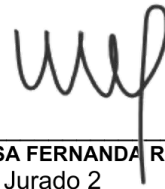
Dr. FRANK VALENCIA

Director Trabajo



ING. GERARDO MAURICIO SARRIA

Jurado 1



ING. LUISA FERNANDA RINCÓN

Jurado 2

Santiago de Cali, 1 de Junio 2023.

Señores

Pontificia Universidad Javeriana Cali

Dr. Gerardo Sarria

Director Carrera de Ingeniería de Sistemas y Computación

Cali

Cordial Saludo.

Por medio de la presente me permito informarle que el estudiante de Ingeniería de Sistemas y Computación Nicolas Ortiz Aristizabal (cod: 8925540) trabajó bajo mi dirección en el proyecto de grado titulado "Reconocimiento y cuantificación de polarización ideológica en redes sociales".

El trabajo anexo ha sido revisado y aprobado por mí, el cual cuenta con mi firma

Atentamente,

A handwritten signature in black ink, reading "Frank D. Valencia", enclosed in a thin black rectangular border.

Dr. Frank Valencia

Santiago de Cali, 1 de Junio 2023.

Señores

Pontificia Universidad Javeriana Cali

Dr. Gerardo Sarria

Director Carrera de Ingeniería de Sistemas y Computación

Cali

Cordial Saludo.

Me permito presentar a su consideración el proyecto de grado titulado “Reconocimiento y cuantificación de polarización ideológica en redes sociales” con el fin de cumplir con los requisitos exigidos por la Universidad para optar al título de Ingeniero de Sistemas y Computación.

Al firmar aquí, doy fe que entiendo y conozco las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería y Ciencias aprobadas el 26 de Noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.

Atentamente,



Nicolas Ortiz Aristizabal

Código: 8925540

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Ingeniería de Sistemas y Computación
Proyecto de Grado

Reconocimiento y cuantificación de polarización ideológica en redes sociales

Nicolas Ortiz Aristizabal

Director: Dr. Frank Valencia

21 de Junio 2023



Abstract

Social media has marked a new era for communication in the modern world, not only providing a path for sharing information but also for expressing individual opinions in civil discourse that can reach millions and millions of people. This behavior has led to a multitude of sociological and psychological studies on people's behavior and the phenomena that may appear from it, with evident massive impacts, for example in the use of advertising or political campaigns. Hence, the need to measure, analyze, and predict these impacts arises in order to understand their behavior within a massive information system.

The growth of communication and the spread of opinions within social networks have generated a special interest in understanding the social and psychological phenomena that arise from this medium. Additionally, it is a challenging task to comprehend public opinion, especially considering the influence of these phenomena on individuals' own beliefs. One of the most important factors in understanding public opinion and its behavior is polarization. It describes how a population of individuals exchanging ideas with differing or opposing opinions can have various consequences on their own opinion, some of which may be counterintuitive.

Throughout this work, it is described the objectives and approaches of computer science in the analysis of opinion on social networks. It also presents different mathematical and computational models that have been implemented to measure and simulate ideological polarization in social networks. Subsequently, a methodology is proposed to conduct tests with real data obtained from interactions on social networks using polarization models.

In addition, an algorithm is presented that quantifies the opinion of messages posted on social networks using a pre-trained artificial intelligence model. This allows opinions to be represented within the parameters of the polarization models. Different influence graphs are also implemented, which define various ways to represent opinion exchanges on social networks.

Finally, the document presents the result of the quantification of the opinions from a dataset from Reddit, and the polarization simulations that are generated using the model presented by the Avispa group. From these simulations, it presents the limitations and feedback of the current model and proposes new approaches that can enhance the simulation of opinion and polarization on social networks from the field of computer science.

Resumen

Las redes sociales han marcado una nueva era para la comunicación en el mundo moderno y ya no sólo abren un camino para compartir información, sino también para dar a conocer la opinión de cada uno en discursos civiles que pueden llegar a millones y millones de personas. Este comportamiento genera el nacimiento de una infinidad de estudios de sociología y psicología sobre el comportamiento de las personas y los fenómenos que pueden aparecer sobre ellos, teniendo impactos masivos que son evidentes, por ejemplo, en el uso de la publicidad o las campañas políticas. De ahí nace además la necesidad de medirlos, analizarlos y predecirlos, con el fin de entender su comportamiento en un sistema masivo de información.

Dicho crecimiento de la comunicación y la dispersión que tienen las opiniones dentro de las redes sociales, han generado un interés especial en comprender los fenómenos sociales y psicológicos que se originan a partir de este medio. Además, de la difícil tarea de conocer la opinión pública, aún con estos fenómenos actuando sobre las propias creencias de cada persona. Uno de los más importantes a la hora de conocer la opinión pública y su comportamiento es la polarización, pues describe como una población de personas que intercambian sus ideas, en el momento en el que estas tengan opiniones diferentes o contrarias, tendrán diferentes consecuencias en sus ideales, algunos de ellos contraintuitivos.

A lo largo de este trabajo se describen los objetivos y acercamientos de las ciencias de la computación en el análisis de la opinión en redes sociales, se exponen diferentes modelos matemáticos y computacionales que se han implementado para medir y simular la polarización ideológica en redes sociales y posteriormente se expone una metodología para realizar pruebas con datos reales, obtenidos a partir de las interacciones en redes sociales, en los modelos de polarización.

Además, se presenta un algoritmo que cuantifica la opinión de mensajes publicados en redes sociales empleando un modelo pre-entrenado de inteligencia artificial, de modo que las opiniones sean representadas en los parámetros de los modelos de polarización. También se implementan diferentes grafos de influencia que definen algunas formas como se pueden representar los intercambios de opinión en redes sociales.

Por último, se evalúan las opiniones de un conjunto de datos de Reddit y se generan las simulaciones de polarización empleando el modelo presentado por el grupo Avispa, a partir de las cuales se presentan límites y retroalimentaciones del modelo actual y se proponen nuevos enfoques que pueden favorecer a la simulación de la opinión y la polarización en redes sociales desde las ciencias de la computación.

Palabras Clave: Polarización, Redes sociales, Modelos de opinión, Simulación.

Índice general

1. Introducción	9
2. Descripción del Problema	11
2.1. Planteamiento del Problema	11
2.1.1. Formulación	12
2.1.2. Sistematización	12
2.2. Objetivos	13
2.2.1. Objetivo General	13
2.2.2. Objetivos Específicos	13
2.3. Justificación	13
3. Marco de Referencia	15
3.1. Marco de Referencia	15
3.1.1. áreas Temáticas	15
3.1.2. Marco Teórico	15
3.1.3. Trabajos Relacionados	17
3.1.4. Principios de compartir información	18
3.1.5. Sesgos al compartir información	19
3.1.6. Efectos y riesgos al realizar intercambios de información en comunidades	20
3.2. Fenómenos de polarización	20
3.2.1. Comportamiento de nivel micro	20
3.2.2. Factores externos	21
4. Casos reales de polarización	23
4.1. Acuerdo de paz de 2016 entre el gobierno y las Fuerzas Armadas Revolucionarias de Colombia	23
4.2. Medio ambiente	24
4.3. Corrupción	25
4.4. Metodología	26
5. Cuantificar la polarización con ciencias de la computación	27
5.1. Objetivos	27
5.2. Retos	27
5.3. Beneficios	28

6. Modelos y acercamientos al tema	29
6.1. Modelo presentado por el grupo Avispa [MSA21, MA60, MSAV20]	29
6.1.1. Propuesta	29
6.1.2. Distintivo	29
6.1.3. Parámetros del modelo	30
6.1.4. Medición de polarización en el modelo	31
6.1.5. Función de actualización	31
6.1.6. Pruebas	32
6.2. A model of opinion and propagation [HAP20]	33
6.2.1. Propuesta	33
6.2.2. Distintivo	33
6.2.3. Definición del modelo	34
6.2.4. Pruebas	35
7. Metodología para la obtención y clasificación de datos	37
7.1. Retos en la recolección de datos y realización de pruebas	37
7.1.1. Información limitada en las interacciones en las redes sociales	38
7.1.2. Interpretaciones y sesgos en la clasificación o medición de las intenciones o creencias de los mensajes o las interacciones entre individuos	38
7.1.3. Retos para definir o construir a un único individuo a partir de la agrupación de muchas de sus interacciones	38
7.2. Recolección de los datos	39
7.3. Preprocesamiento de datos	40
7.4. Cuantificación de la creencia de los individuos	41
7.5. Definir el grafo de influencia	45
7.6. Visualización de datos	45
7.7. Descripción del algoritmo	48
7.8. Validación del modelo con datos reales	51
8. Resultados	53
8.1. Evaluación en el modelo	53
8.2. Retroalimentación del modelo	57
8.3. Nuevos enfoques	58
9. Conclusión	59
9.1. Futuros pasos	59
9.2. Conclusión	59
Bibliografía	61

Introducción

Las redes sociales han generado un gran impacto no sólo en el mercado y la publicidad, sino también en campos diversos como la política. Dada la facilidad y los diferentes medios por los cuales compartir y difundir la opinión del individuo, los discursos civiles se han vuelto algo común para la sociedad. Debido a esta realidad, en los últimos años ha nacido una tendencia a estudiar y comprender el comportamiento de las interacciones sociales dentro de las redes sociales y cómo afectan estas a las creencias de cada persona y en general a la opinión pública.

En el presente documento se detalla un proyecto de investigación que tiene como objetivo analizar diferentes estudios sobre uno de los fenómenos psicológicos y sociológicos más famosos en las redes sociales, relacionado a la opinión, las creencias y la polarización ideológica. Además, desarrollar cómo este puede ser estudiado, medido y simulado desde las ciencias de la computación, empleando técnicas y estrategias basadas en modelos matemáticos, inteligencia artificial y modelos probabilísticos.

De formas más específicas, el proyecto tiene como objetivo analizar el modelo matemático planteado por el grupo de investigación Avispa, detallando así sus fundamentos teóricos y contrastarlos con estudios similares. Además emplear el mismo para realizar pruebas con datos reales basados en interacciones en redes sociales, con el fin de identificar las características del modelo que favorecen al análisis en cuestión, así como sus limitaciones.

Descripción del Problema

2.1. Planteamiento del Problema

Las redes sociales pueden facilitar el discurso civil, al hacer posible un rápido intercambio de hechos, historias y opiniones. También permiten e incluso sugieren que los usuarios se unan a grupos que comparten intereses o ideologías similares. Estas son algunas de las características atractivas que han hecho de las redes sociales uno de los medios más populares, si no el más popular, para la interacción y la comunicación en el mundo de hoy.

Sin embargo, las redes sociales también pueden filtrar y dirigir (o permitir que otros dirijan) información a millones de usuarios en función de su posición ideológica en redes sociales, conectividad con otras personas y preferencias o afiliación a grupos. Si bien, esto puede aumentar una exposición saludable a diversas perspectivas, lo que es beneficioso para nuestra sociedad, también puede aumentar las burbujas de segregación ideológica. El problema con esto es que las personas con puntos de vista opuestos tenderían a interpretar esta información con su propio sesgo, reforzando así sus creencias originales en lugar de cuestionarlas. Esto puede resultar en una separación aún más notoria de sus opiniones [Plo10]. Del mismo modo, en grupos con vistas uniformes, los usuarios pueden volverse más extremos al reforzar las opiniones de cada uno [Aro10b]. De hecho, la psicología social nos dice que los grupos pueden fortalecer sus creencias al interpretar la información a favor de la confirmación del sesgo o a oponerse firmemente si contradice sus puntos de vista (efecto contraproducente).

De hecho, las plataformas de redes sociales han impulsado la polarización en el proceso político. Los referendos como el Brexit y el acuerdo de paz Colombiano, así como las elecciones presidenciales de Brasil / Colombia / Estados Unidos son ejemplos convincentes de este fenómeno³. Estos temas han sido testigos del hecho de que los mensajes divisivos en las redes sociales con elementos de ideología extremista, información engañosa o demagogia, pueden influir en los procesos fundamentales de toma de decisiones. De hecho, hemos llegado a un punto crítico en el que la producción de información maliciosa para causar polarización en las redes sociales, se ha convertido en una fuente rentable de ingresos en algunos países, como Macedonia [Kir17].

Aunque existen modelos en economía y psicología social (por ejemplo, [Aro10a, ER94a, Sin18]) para razonar sobre este fenómeno de polarización grupal, no son adecuados para simular y predecir la dinámica de agentes que interactúan en sistemas de comunicación masivos mediante el intercambio de información epistémica. Esto se debe al gran número de agentes e interacciones presentes

en el sistema y a que este se desarrolla continuamente. Por esta razón, es necesario contar con herramientas dinámicas que sean capaces de analizar dichos volúmenes de datos y en base a estos generar una predicción.

En ciencias de la computación se han realizado algunos trabajos sobre el tema, donde se plantean métodos, principalmente desde la inteligencia artificial, para analizar este fenómeno en las redes sociales. Sin embargo, existen enfoques orientados hacia modelos matemáticos, como es el caso de los trabajos del grupo de investigación Avispa de la Universidad Javeriana y la Universidad del Valle, donde se desarrolló un modelo que describe el comportamiento de la polarización ideológica [MA60]. Este modelo se basa en la abstracción de fenómenos y teorías estudiadas desde la economía, sociología y psicología. Sin embargo, sus pruebas se realizaron con sistemas de ejemplo, como una distribución binaria, donde la mitad de los agentes apoyan una idea y la otra mitad está en desacuerdo con la misma, que no representan casos de la vida real. Dado a esto, el estudio no presenta pruebas que justifiquen su precisión dentro de sistemas de interacción social reales.

Por lo tanto y con el fin de argumentar sobre la efectividad del modelo y así continuar con la investigación de este fenómeno a partir de las ciencias de la computación, es necesario realizar pruebas y simulaciones donde se dé una evidencia objetiva de la precisión del modelo, con el fin de retroalimentar. Para este fin es necesario definir métodos o algoritmos que permitan abstraer la información de las interacciones en redes sociales en parámetros con los que el modelo pueda trabajar.

2.1.1. Formulación

¿Cómo medir, a partir de datos obtenidos en redes sociales, la polarización ideológica de un grupo de personas y simular su comportamiento bajo un modelo matemático y computacional?

2.1.2. Sistematización

- ¿Qué enfoques ha habido desde las ciencias de la computación sobre la polarización ideológica en redes sociales y estudios desde otros campos que pueden aportar al análisis de este fenómeno desde el punto de vista computacional?
- ¿Cómo abstraer o cuantificar la información presente en las interacciones de las redes sociales, de modo que sean interpretadas por un modelo matemático?
- ¿Qué métodos se pueden emplear para simular la polarización ideológica de las redes sociales desde un modelo matemático?
- ¿Qué tan buena es la precisión del modelo matemático frente al análisis de la polarización ideológica y cómo puede mejorarse?
- ¿Cuáles serían los retos y los posibles enfoques a futuro del análisis, medición y predicción de la polarización ideológica en redes sociales, desde las ciencias de la computación?

2.2. Objetivos

2.2.1. Objetivo General

El objetivo principal de este trabajo es realizar una investigación sobre la polarización en redes sociales desde el punto de vista social y psicológico, orientado a cómo abordarlo desde las ciencias de la computación, con el fin de realizar reconocimiento, análisis, simulación y predicción de dicho fenómeno a partir de las interacciones presentes en las redes sociales.

De forma más concreta, se planea analizar el modelo matemático planteado por el grupo de investigación Avispa, teniendo en cuenta sus características y fundamentos. Posteriormente, realizar pruebas de dicho modelo con información de redes sociales reales, que permitan dar un resultado objetivo de la eficacia del modelo, compararlo con estudios similares e identificar las características y limitaciones del mismo. Además, se deben desarrollar métodos o algoritmos que permitan al modelo interpretar los datos presentes en las redes sociales, de modo que estos sean consecuentemente los parámetros del modelo.

2.2.2. Objetivos Específicos

(O1) Analizar información actual sobre polarización y los enfoques que ha habido desde las ciencias de la computación, orientado principalmente a cuantificar dicho fenómeno.

(O2) Diseñar algoritmos que permitan transformar la información de los formatos presentes en las interacciones de algunas redes sociales, de modo que esta pueda ser interpretada por el modelo matemático establecido [MA60].

(O3) Realizar simulaciones y pruebas que den evidencia de la precisión del modelo matemático en casos reales de polarización en redes sociales.

(O4) Retroalimentar el modelo matemático establecido, a partir del análisis de los resultados de la simulación y la información recolectada.

(O5) Proponer, a partir de los resultados obtenidos, posibles nuevos enfoques sobre los que se puede abordar la polarización en redes sociales por medio de las ciencias de la computación.

2.3. Justificación

El reciente desarrollo de las redes sociales presenta problemas críticos e impactos sociales por la difusión de información errónea y creencias radicales. Las redes sociales se han vuelto omnipresentes y contienen grandes cantidades de información personal y confidencial, incluidas opiniones y datos; esta información se comparte y se utiliza para cálculos de formas complejas con efectos impredecibles, como la radicalización y la polarización, la difusión de noticias falsas, las violaciones de la

privacidad y la filtración de información. Los acontecimientos recientes han demostrado que estos sistemas no se comprenden bien y que tienen un gran impacto en la democracia y en la sociedad en general. Por ejemplo, los efectos de las redes sociales en la opinión pública se vieron reflejados recientemente en el Brexit, las elecciones presidenciales de 2016 en Estados Unidos, el plebiscito de paz colombiano y el movimiento de Gilets Jaunes en Francia. Estos efectos no se previeron de antemano y algunos resultaron problemáticos, dado que se basaron, en parte, en la difusión de información errónea y la radicalización de las opiniones de los usuarios. Lo cual deja en evidencia la necesidad actual de comprender y analizar la difusión de la opinión y los fenómenos sociales que esta trae consigo.

Actualmente, la investigación sobre la polarización grupal y sobre las creencias y opiniones en las redes sociales se ha centrado en enfoques psicológicos y filosóficos [ML76, JSG13, CH15, ZCP16]. Estos enfoques presentan algunas limitaciones a la hora de realizar análisis específicos para obtener respuestas concretas, referente a los efectos causados por la polarización en comunidades específicas. Esto se debe principalmente a la velocidad con la que evolucionan los comportamientos de los individuos en espacios de comunicación masivos, que dificulta establecer teorías estáticas sobre estas, sin tener en cuenta las condiciones particulares de cada grupo o red de comunicación. Por lo cual, estos estudios se centran en la descripción del comportamiento de los individuos dentro de estos sistemas y los diferentes fenómenos que aparecen, así como también explican, cómo estos pueden afectar a las opiniones de las personas que hacen parte de él. Se ve necesario establecer un estudio o herramienta que permita, a partir de los planteamientos realizados desde la psicología y sociología, analizar grupos sociales concretos, con sus numerosas interacciones, con el fin de identificar la aparición de fenómenos como la polarización y llegar a predecir el impacto que estos tienen en la opinión pública.

Debido a la gran cantidad de información que debe tenerse en cuenta para llegar a una conclusión consecuente con el reconocimiento de polarización grupal, se hace oportuno optar por el estudio de este fenómeno desde las ciencias de la computación, con herramientas como modelos matemáticos e inteligencia artificial, que permiten el procesamiento de grandes volúmenes de información con el fin de obtener una aproximación a un resultado en concreto.

Es importante mencionar que, desde las ciencias de la computación se han realizado muchos estudios alrededor del tema de las redes sociales, y estos generalmente se han centrado en cuestiones como la topología de la red [LNK07, AMB07, KHH18], la ubicación [ECL11] y la privacidad [GA05]. Sin embargo, se han realizado pocos enfoques o acercamientos en la opinión y la difusión de la información, sobre todo teniendo como objetivo el análisis de fenómenos psicológicos y sociológicos, a pesar del impacto social que dichos fenómenos están teniendo sobre diferentes campos, como son la política o la economía.

Marco de Referencia

3.1. Marco de Referencia

3.1.1. áreas Temáticas

- Theory of computation Models of computation Concurrency
- Theory of computation Formal languages and automata theory Formalisms
- Theory of computation Design and analysis of algorithms Graph algorithms analysis
- Theory of computation Design and analysis of algorithms Algorithm design techniques
- Theory of computation Logic AbstractionNetworks Network algorithms
- Mathematics of computing Probability and statistics Probabilistic representations
- Mathematics of computing Mathematical analysis Functional analysis Approximation
- Computing methodologies Modeling and simulation

3.1.2. Marco Teórico

Cada persona tiene sus propias opiniones e ideologías, las cuales se forman a partir de sus experiencias personales, la información que obtiene de diferentes medios, así como la interpretación que este le dé a las ideas, basada en su personalidad. Debido a esto, existen tantas opiniones como personas hay en el mundo y del mismo modo interacciones entre las personas, conversaciones, discursos, críticas, libros, entre otras. Esta interacción permite a los receptores adquirir nueva información y por consiguiente, evolucionar sus propias creencias. Actualmente, con la evolución de las redes sociales y la posibilidad de difundir información de forma masiva, se han generado diversos fenómenos y circunstancias que hacen inciertos los resultados de estas interacciones sociales en las ideologías de las personas que participan en estas. Causando así, acontecimientos sin precedentes como los que se han descrito a lo largo de este documento. Uno de estos fenómenos, que además está en auge en estos momentos, es la polarización ideológica.

El fenómeno social de la polarización fue definido rigurosamente por primera vez por los economistas Esteban y Ray [ER94b]. Su medida de polarización es influyente, y es la que adoptamos en este artículo. Li et al. [Li13] fueron los primeros en modelar el consenso y la polarización en las redes

sociales. Sin embargo, como la mayoría de los otros trabajos, no cuantifican la polarización, sino que se centran en cuándo y bajo qué condiciones una población llega a un consenso. Proskurnikov et al. [Pro16] investigó la formación de consenso o polarización en las redes sociales, pero consideró la polarización de las habitaciones como una falta de consenso, en lugar de un fenómeno en sí mismo.

En ciencias sociales, la polarización grupal se refiere a la tendencia natural de los grupos a tomar decisiones más extremas que los individuos. Es decir, que el fenómeno nace del comportamiento de cada individuo, con sus creencias y opiniones, a la hora de interactuar en un grupo con otras personas, donde estas pueden compartir su creencia o estar en desacuerdo con la misma. Además, define cómo a raíz de esta variedad de opiniones compartidas se generan efectos en las propias personas y en la opinión pública, tal que las decisiones grupales pueden ser más radicales que las propias de cada individuo.

Este fenómeno también se ha estudiado en los últimos 50 años en el contexto de la psicología social y la economía [RS15]. No obstante, se han realizado pocos estudios de esto, abordado desde las ciencias de la computación. A pesar de que las redes sociales han sido un gran foco para los estudios matemáticos y computacionales, debido al sin número de elementos importantes e influyentes en la sociedad contemporánea, pocos de estos se han centrado en analizar los fenómenos vistos por la psicología y sociología, en especial optando por un análisis de la opinión como un componente epistémico, es decir, estudios que se centren en las creencias del individuo y cómo estas se desarrollan en una comunidad.

Al abordar este problema desde las ciencias de la computación, se encuentran diferentes estrategias que favorecen al análisis de los datos y el cálculo de los resultados. Técnicas como las referentes a inteligencia artificial, la simulación, los modelos matemáticos y computacionales son algunas de las que suelen relacionarse en trabajos similares a este. Este proyecto, parte de un modelo desarrollado por el grupo de investigación Avispa, el cual se centra en adaptar las teorías sociales y psicológicas afines al reconocimiento de la polarización en las redes sociales, de modo que este se pueda observar como una medida cuantitativa. Más concretamente, el presente proyecto busca enfatizar en las metodologías que permitan transformar los datos presentes en las interacciones de las redes sociales, los cuales son totalmente cualitativos (tales como mensajes y suscripciones), a elementos que sean entendidos por el modelo matemático y computacional planteado por Avispa. Del mismo modo, se piensa enfatizar en estrategias de simulación, probabilidad y predicción basadas en las proyecciones dadas por el mismo modelo o un agente planteado desde la inteligencia artificial.

Estas estrategias han sido trabajadas en proyectos similares, como Sirbu et al. [Sir19] el cuál utiliza un modelo algo similar al que se espera analizar en este proyecto, que se actualiza probabilísticamente. En él, se investigan los efectos del sesgo algorítmico en la polarización, contando el número de grupos de opinión, interpretando un solo grupo de opinión como consenso, en lugar de medir directamente la polarización en sí. Leskovec et al. [Gar16] desarrolla redes sociales simuladas y observa la formación de grupos a lo largo del tiempo. Sin embargo, su trabajo no representa una

medida formal de polarización.

Se ha trabajado además en el análisis de datos de redes sociales legales y funciona con otros usuarios humanos para comprender la polarización y los fenómenos relacionados. Crandall et al. [Cra08] estuvieron entre los primeros en investigar los cambios en el acuerdo del usuario a lo largo del tiempo en dichos sistemas, basados en datos de Wikipedia y LiveJournal. Aunque no mencionan explícitamente la polarización, sus preocupaciones son similares a las de este proyecto, ya que se centran en que los usuarios se vuelvan más similares a lo largo del tiempo y en la formación de comunidades distintas, denominadas "Balkanization". Leskovec et al. [Les10] profundiza en medir y representar el estado de los usuarios y las opiniones positivas o negativas entre ellos en las redes sociales, utilizando datos de Epinions, Slashdot y Wikipedia. El estado y las opiniones mutuas no son lo mismo que la influencia, pero comparten algunas similitudes. Flaxman et al. [Fla18] utiliza datos sobre las visitas de los lectores a los sitios web de noticias para medir su segregación política, la cual define como la distancia promedio entre las creencias políticas de dos usuarios seleccionados al azar. Fletcher and Nielsen [Fle17] analizan el acceso de los usuarios a los medios de comunicación, tanto en como fuera de línea, en diferentes países. Sus ideas sobre la fragmentación son relevantes para nuestro trabajo y ayudaron a desarrollar escenarios de prueba en relación con los medios de comunicación.

3.1.3. Trabajos Relacionados

- Toward a Formal Model for Group Polarization in Social Networks [MA60]:

Este trabajo fue realizado por integrantes del grupo de investigación Avispa y es el trabajo principal sobre el que se va a desarrollar la investigación. El trabajo presenta un primer acercamiento a la medición de polarización de grupos en redes sociales, a partir del estudio de diferentes fenómenos sociales y psicológicos. De modo que, plantea un modelo matemático basado en la interacción de agentes en un sistema. Además, se basa en parte de teoría económica para desarrollar la influencia que tiene cada agente en la medición de la polarización.

Inicialmente en el desarrollo del trabajo se habla de las diferentes causas o interacciones que tienen influencia en la ideología de cada participante del grupo. Cuando 2 personas interactúan entre ellas, dependiendo de algunas variables como su conocimiento del otro y sus propias creencias, pueden tener diferentes tipos de comportamientos, ya sean de reforzar su propia creencia sobre el tema o por el contrario de dudar sobre ella. Posteriormente, se analizan cada uno de los fenómenos y se determinan como agentes, para describir un modelo matemático donde se cuantifica cada agente y se le da un valor de influencia en la medición a cada uno.

Finalmente, se prueba el modelo, simulando el comportamiento de grupos con parámetros específicos (de extrema diferencia de ideologías, ideologías similares o ideologías normales) con el fin de dar una prueba superficial al modelo matemático y comprobando los comportamientos

básicos del mismo. Sin embargo, no se realizan pruebas en conjuntos de datos reales de alguna red social.

- A model of opinion and propagation structure polarization in social media [HAP20]:

En este artículo se propone un modelo dinámico para el análisis de opinión, fundamentado en el cambio continuo de opiniones de un individuo al interactuar con otros. Los parámetros que se tienen en cuenta para este modelo son la opinión de cada individuo sobre un tema y el grado de conexión o confianza que se tiene cada individuo con los demás.

Además, se realizan simulaciones sobre la propagación de múltiples opiniones a lo largo de una red social con múltiples individuos y se observa la evolución de los parámetros del modelo. A partir de los resultados de dichas simulaciones, se encontraron patrones de comportamiento de la polarización en grupos y se determinó lo rápido que esta genera un cambio en la opinión de los individuos que participan. Además, se realizaron simulaciones con datos obtenidos en “Twitter” con el fin de desarrollar pruebas con datos reales.

Existen además algunos trabajos sobre el análisis de interacciones en las redes sociales empleando enfoques basados en la lógica y a pesar de no compartir el enfoque concreto de este trabajo, presentan técnicas que aportan al desarrollo del mismo. Los más relevantes son:

- Liu et al. [Liu14] usa ideas de la lógica doxástica y la lógica epistémica dinámica para modelar cualitativamente la influencia y el cambio de creencias en las redes sociales.
- Christoff [Chr16] desarrolla varias lógicas para redes sociales y a pesar de que este trabajo tampoco tiene un enfoque cuantitativo, se refiere a problemas relacionados con la polarización como las cascadas de información.
- Seligman et al. [Sel11] introduce una "lógica de Facebook" básica. Esta lógica no es cuantitativa, pero su punto interesante es que las posibles redes de un agente son diferentes redes sociales, este es un enfoque prometedor para el modelado formal de problemas epistémicos en las redes sociales.
- Hunter [Hun17] introduce una lógica de actualizaciones de creencias en las redes sociales donde los agentes más cercanos en la red social son más confiables y, por lo tanto, más influyentes. Si bien las creencias en esta lógica no son cuantitativas, incluye una noción cuantitativa de influencia entre los usuarios.

3.1.4. Principios de compartir información

Una parte fundamental de la individualidad de cada ser humano está en la capacidad de formar ideas u opiniones propias sobre un tema. Esta opinión puede ser dictada por un sin número de factores como sus experiencias, conocimientos, posición social, política o económica frente al tema,

sus principios o valores, entre muchas otras, haciendo de la opinión de cada persona sobre cada tema una característica realmente compleja de comprender, medir o analizar.

Al añadir diferentes temas al problema, encontramos aún más complejidad en la opinión de cada persona. Esto, debido a que no solo las opiniones previas sobre temas similares crean una clara influencia en la construcción de nuevas ideas u opiniones, sino que también, la importancia que le da cada persona a cada uno de los temas varía, teniendo opiniones más o menos construidas sobre algunos temas frente a otros, así como con mayor o menor impacto o flexibilidad en la forma como influyen su criterio.

Habiendo dado una base de la complejidad e individualidad de la opinión de una persona, hay que mencionar que el análisis de esta opinión carece de sentido si se realiza a cada persona independientemente. Debido a la naturaleza social del ser humano, no sólo la opinión se ve influenciada por la posición social en la que se encuentra y la forma como cada tema influencia su vida, sino que está también cambia a través del tiempo debido al intercambio de información con otras personas.

La comunicación se define como la acción consciente de intercambiar información u opiniones entre múltiples individuos. Esta se puede dar de forma oral, escrita u en cualquier otro medio por el cual se transmita información. Además, la comunicación no siempre debe ser bidireccional. Medios como los discursos públicos o las transmisiones televisivas también transmiten ideas hacia los receptores. Esta exposición de nuevas ideas sobre un tema tiene un impacto sobre la opinión del individuo. Al recibir nueva información y argumentos que pueden apoyar o contradecir sus opiniones personales generan un cambio en la opinión personal.

Entender como el intercambio de información puede afectar la opinión de cada individuo es un tema ampliamente estudiado desde la psicología y se han determinado muchos fenómenos que pueden causar resultados contra intuitivos. Esto es debido a que a pesar de que la intuición dicta que cada persona reevalúa su opinión con base a la nueva información recibida, la realidad es que hay muchos factores adicionales. Tales como la influencia que tenga la persona que transmitió esa idea, su confianza en ella o su afinidad a la idea original, que pueden afectar en gran medida la forma y magnitud cómo ésta influirá el cambio sobre la opinión de la persona.[ML76, JSG13]

3.1.5. Sesgos al compartir información

Los factores que hacen parte de una conversación o cualquier otro tipo de intercambio de información son muy variados y afectan a todos los elementos del intercambio de ideas. Desde la intención del mensaje que quiere transmitir el emisor, la forma como el receptor interprete y asuma la información e incluso el mensaje que se transmite en sí mismo, se ven afectados. Entre ellos, podemos evidenciar factores personales como la cercanía o confianza entre los individuos, factores sociales como una jerarquía laboral o académica, e incluso factores que hacen parte del propio tema sobre el que se discute, por ejemplo, tener mayor recepción con mensajes que van acorde a las

propias opiniones de la persona.

Diferentes estudios y experimentos desde la psicología, teniendo en cuenta los factores nombrados anteriormente, han definido varios de los fenómenos que ocurren durante la comunicación, los cuales influyen en la forma como se transmite la información entre los individuos. Algunos de los fenómenos más conocidos y estudiados también desde las ciencias de la computación son enumerados y explicados más adelante en este trabajo.

Todos estos fenómenos que generan sesgos en la comunicación y el intercambio de ideas son de vital importancia a la hora analizar la forma como evolucionan las opiniones de cada individuo al interactuar con otros. Del mismo modo, son estos los que conducen a los fenómenos sociales que observamos en nuestro día a día como las cámaras de eco o la polarización.[HAP20]

3.1.6. Efectos y riesgos al realizar intercambios de información en comunidades

Las redes sociales han definido un nuevo rumbo para el comportamiento de todas las personas, así como la interacción entre ellas. Masificar el intercambio de información, anonimizar a los individuos y obviar las distancias geográficas o culturales, son solo algunos de los aspectos que ha traído el internet a la forma como se intercambian cada día en nuestra sociedad actual.

Luego de presentar las complejidades y retos que se encuentran al analizar las opiniones de un individuo y cómo estas evolucionan al interactuar con otros, es momento de integrar el internet y las redes sociales a este sistema.

3.2. Fenómenos de polarización

Si bien hay múltiples factores que se han atribuido como la causa detrás de las cámaras de eco y la polarización, estos factores se pueden clasificar en dos: primero, el comportamiento de nivel micro de los individuos, que causan colectivamente un fenómeno de nivel macro, como lo es la polarización. Segundo, los factores externos que aparecen como elementos fuera de la comunicación entre individuos e influyen en la interacción y la evolución de opiniones colectivas.[HAP20, ML76]

3.2.1. Comportamiento de nivel micro

Los comportamientos que hacen parte de la categoría micro, son los referentes a la forma como la opinión del individuo es sesgada por la información que recibe en comparación a su idea original y la influencia que tiene el emisor sobre él.

Los fenómenos implicados en la forma como la opinión del emisor influye en la personalidad, pueden verse originados por el fenómeno de la disonancia cognitiva. Un fenómeno ampliamente estudiado dentro del campo de la psicología, y se refiere al malestar que experimentan las personas cuando se les presenta información que desafía sus creencias o decisiones. Este fenómeno fomenta

la aparición de otros fenómenos en la comunicación que podemos evidenciar como causas directas en la polarización ideológica.[MA60]

Entre estas conductas, una de las que más se atribuye como causa de la polarización es la exposición selectiva (selective exposure). Esta es la tendencia de los individuos a exponerse solo a la información que se alinea con sus creencias preexistentes.[HAP20] La investigación sobre la exposición selectiva es anterior a la adopción masiva de Internet; el comportamiento se ha observado en estudios de control en los que se pidió a los participantes que examinaran información después de tomar una decisión/opinión inicial. Más recientemente, en el contexto de las redes sociales en línea, también se ha observado empíricamente una exposición selectiva en Facebook.

Otro comportamiento se denomina sesgo de confirmación (confirmation bias) o asimilación sesgada (biased assimilation)[Plo10]. Este concepto está relacionado con la asimilación o interpretación de la información, donde independientemente de cuál sea la información real (contradictoria o de apoyo), los individuos tienen una mayor tendencia a interpretarla como de apoyo, reforzando sus creencias. De manera similar a la exposición selectiva, la investigación sobre la asimilación sesgada existía antes de Internet en forma de experimentos controlados.

Adicional a los anteriormente nombrados, hay otro fenómeno que puede catalogarse como una causa y a la vez consecuencia de una comunicación polarizada, estas son las cámaras de eco. Estas se refieren, más que a un fenómeno social a un estado en el que las interacciones en línea se llevan a cabo en un patrón polarizado. Dentro de una red social en línea, existen grupos que corresponden a una determinada creencia (por ejemplo, izquierda o derecha en la política), y la mayoría de las interacciones sobre este tema en los usuarios ocurren en el interior de este grupo.

3.2.2. Factores externos

La opinión popular, guiada por el constante intercambio de información entre diferentes individuos de la sociedad, tiene un gran impacto en nuestro mundo moderno. La opinión hacia una marca o empresa influye en gran medida sobre el rendimiento económico de la misma, sin mencionar además los intereses políticos sobre la perspectiva que tienen frente a las masas.

Este interés en buscar una opinión pública favorable promueve la creación de medios artificiales para manipularla. Agentes externos a la comunicación que influyen en la opinión de las personas de manera inflexible, impulsados únicamente por su objetivo, se denominan factores externos. Incluso antes de la aparición del internet, los discursos políticos o medios informativos televisivos, ya hacían gran parte de los elementos que influyen la opinión de las personas.

Con la aparición del internet y las redes sociales, el intercambio de información masivo se convirtió en un sistema más complejo, así como un pilar principal dentro de cada aspecto que compone nuestra sociedad. Ya sea política, social, cultural o económicamente, la opinión pública de cada

entidad se tornó una prioridad. Por lo anterior, y aprovechando el anonimato e impersonalización característico de las redes sociales, se empezaron a desarrollar diferentes mecanismos artificiales que influyen en la opinión de las personas.[HAP20]

Un factor que se ha relacionado con la polarización es la personalización algorítmica empleada por muchos sitios de redes sociales. Como es bien sabido, los contenidos que aparecen en la página de inicio de un usuario de Facebook o Twitter son seleccionados por un algoritmo cuyo objetivo es maximizar la participación o la relevancia. La idea general es que estos contenidos relevantes suelen ser contenidos que se alinean con la predisposición de un usuario, creando una cadena de publicaciones que pueden reforzar opiniones personales. Otro factor es la existencia de agentes 'obstinados', cuyas opiniones son estáticas y extremas, por ejemplo, bots políticos de Twitter, fuentes de noticias partidistas. La idea es que, a través del intercambio de opiniones con estos agentes obstinados, las opiniones inevitablemente cambiarán a los extremos del espectro, ya que las opiniones de estos agentes son inmutables.

Casos reales de polarización

Las redes sociales han creado el escenario idóneo para la aparición de fenómenos sociales como la polarización ideológica, a causa del masivo intercambio de información y opiniones. Además, de facilitar la aparición de factores externos en masa como bots o cadenas anónimas.

Es posible evidenciar casos de polarización en nuestro día a día. Muchos de estos han tenido una influencia global como el tratado del Brexit, las elecciones presidenciales de Estados Unidos en 2016 o las vacunas frente al Covid 19. Leer opiniones extremistas, grupos de conversación centrados en un único lado del debate o bots de apoyo u oposición frente a una idea específica, son solo algunos de los ejemplos a los que estamos expuestos cada día en redes sociales como Twitter o Instagram.

Pero no es necesario ir tan lejos, la polarización es algo que vemos en la actualidad en Colombia más que nunca. Basta con ingresar a alguna red social para ver todo tipo de fenómenos de polarización en un sin número de temas, entre ellos podemos destacar el Covid 19, la vacunación, la inmigración y principalmente la política.

En el trabajo “Estudio Nacional Sobre Xenofobia Y Polarización Para Lograr Inclusión Y Cohesión En Colombia” realizado por el laboratorio social Movilizadorio, en conjunto con Ford Foundation, Fundación Santo Domingo y Open Society Foundations[Mov20], se estudió la polarización presente en redes sociales enfocado en Colombia, frente a diferentes temas controversiales que hemos vivido en los últimos años. Entre ellos se destacan principalmente:

4.1. Acuerdo de paz de 2016 entre el gobierno y las Fuerzas Armadas Revolucionarias de Colombia

En el estudio, Movilizadorio presenta los hallazgos sobre el análisis de este tema. En él destaca, cómo se pueden distinguir 2 grupos polarizados, un grupo que reúne posiciones políticas conservadoras (“verde”) y otro que nuclea a posiciones más progresistas (“violeta”). Ambos, se presentan en burbujas separadas, mostrando una tendencia extremadamente polarizada. Ver Figura 4.1

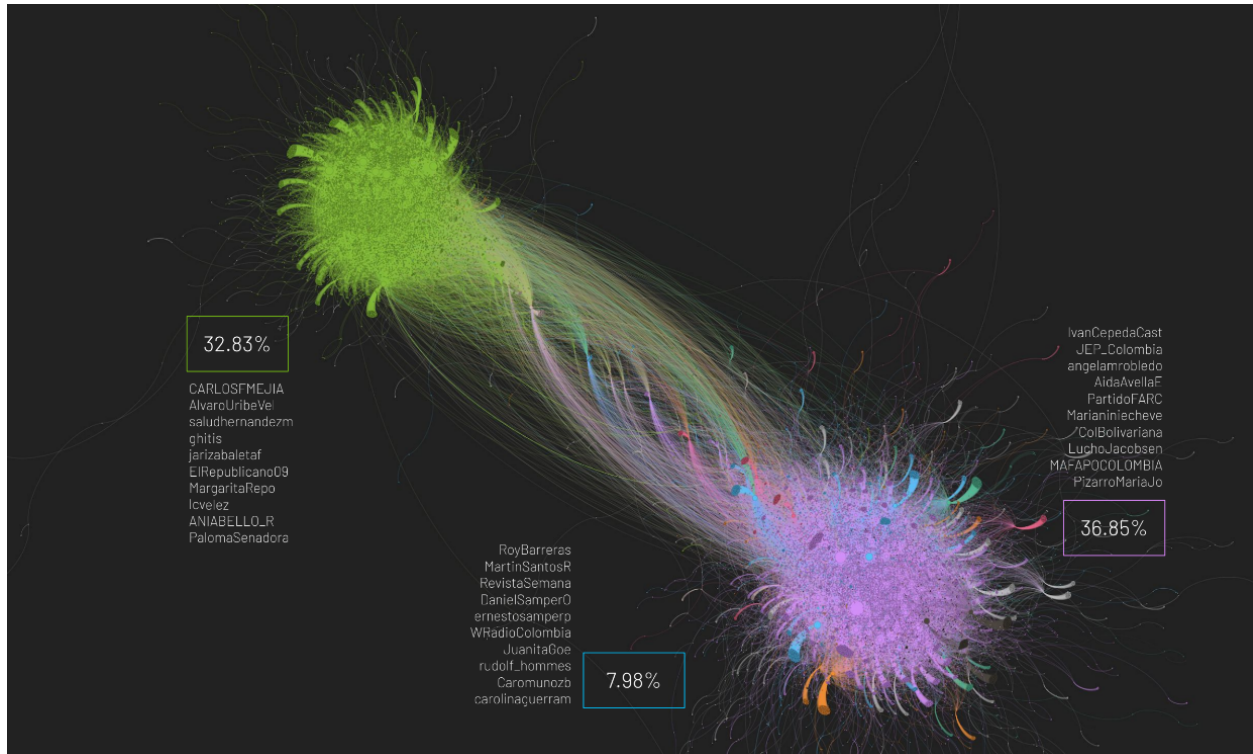


Figura 4.1: Caso de análisis de polarización con el tema del acuerdo de paz colombiano de 2016 [Mov20]

4.2. Medio ambiente

Los resultados encontrados por Movilizadorio frente al tema del cuidado del medio ambiente, presentan un claro ejemplo real de un tema con una opinión más integrada y donde la interacción entre los nodos se encuentra menos fraccionada en grupos que compartan una opinión. Ver Figura 4.2

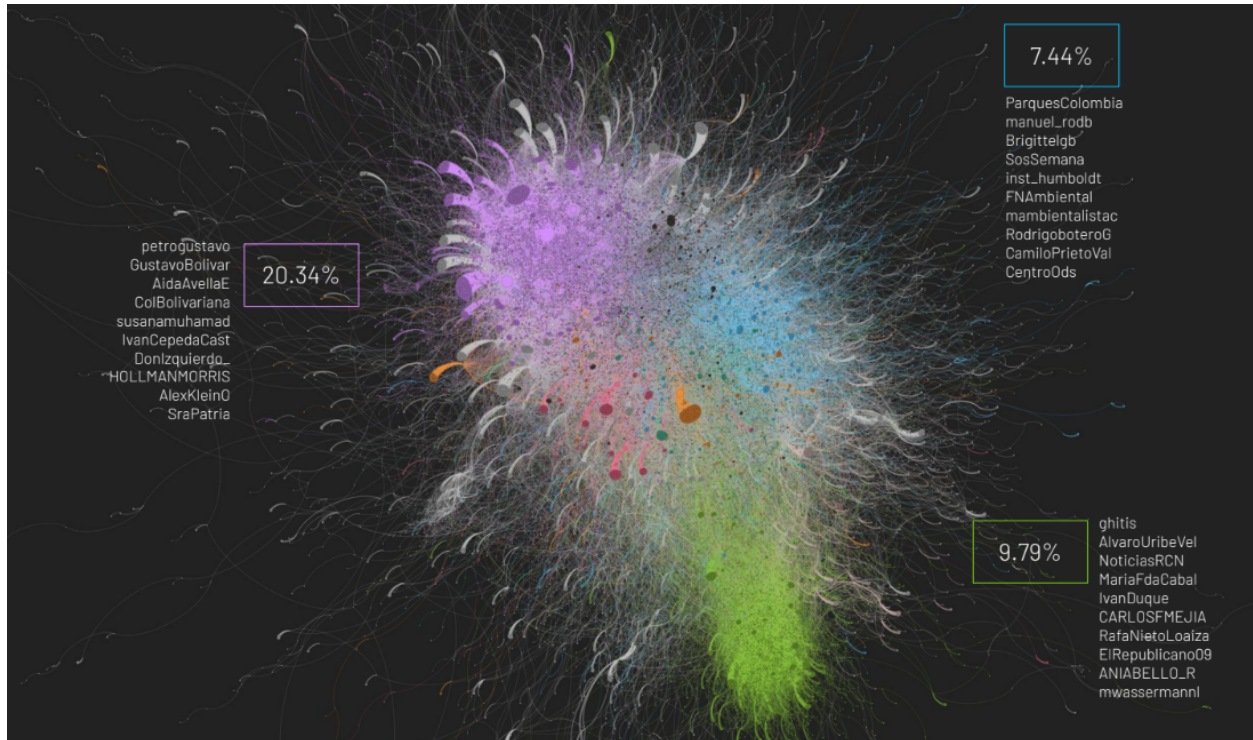


Figura 4.2: Caso de análisis de polarización con el tema del medio ambiente [Mov20]

4.3. Corrupción

En el estudio sobre el tema de la corrupción se identificaron 2 fenómenos interesantes relacionados con la polarización. El primero es que el resultado es el grafo con mayor cohesión del estudio y sin embargo se exhiben pocas interacciones entre nodos verdes y no verdes de la red, que representan dos grupos con opiniones antagónicas, ilustrando un claro ejemplo de las cámaras de eco expuestas anteriormente. El segundo es el caso particular de un usuario que tiene un número de conexiones mucho mayores a las de otros nodos, pero únicamente conectados a uno de los grupos, además este presenta una opinión opuesta a este, por lo que podría tratarse de un factor externo como una granja de bots, como lo afirma el estudio. Ver Figura 4.3.

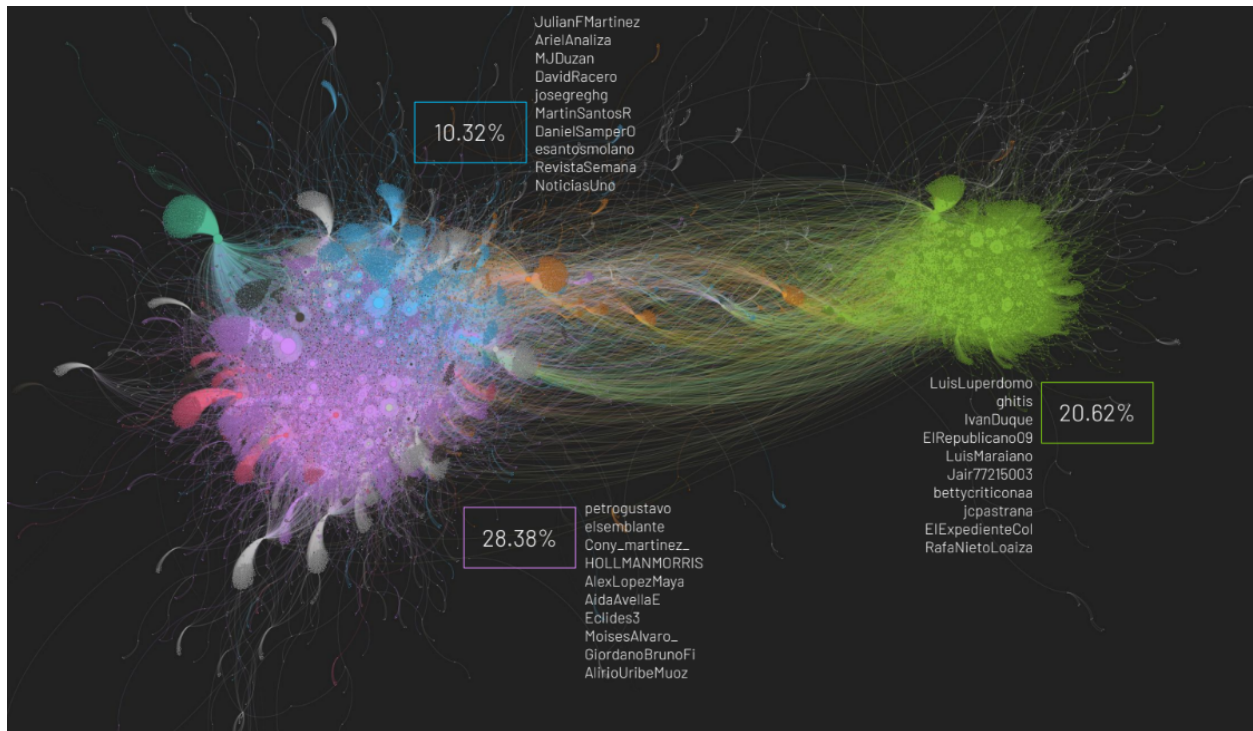


Figura 4.3: Caso de análisis de polarización con el tema de corrupción en Colombia [Mov20]

4.4. Metodología

Para la realización del estudio presentado por Movilizadorio, se realizó una recolección de datos sobre los diferentes temas en la red social Twitter. Se emplearon técnicas de análisis de sentimiento con machine learning y posteriormente se graficó la opinión de los twits conectados en relación al hilo que responden, de modo que un agente (twit) está conectado con otro si este es una respuesta al primero. Posteriormente realizaron un análisis cualitativo con base al gráfico generado.

Los métodos de recolección de datos usados para este estudio, así como las herramientas para cuantificar y analizar la polarización ideológica sobre cada tema son profundizados más adelante en este trabajo, junto con otros métodos propuestos en diferentes proyectos.

Cuantificar la polarización con ciencias de la computación

La gran escala de datos que se presentan en las redes sociales dentro de las interacciones entre los usuarios le ha dado a las ciencias de la computación un espacio de trabajo, donde la automatización y el análisis de grandes volúmenes de datos aparentan ser la única solución viable.

Con el fin de analizar un fenómeno sociológico como la polarización, es necesario primero cuantificar el problema. Para esto se debe intentar modelar un sistema que permita representar los elementos que se consideran influyentes en las interacciones sociales, así como las variables y parámetros, los eventos que modifican el sistema y los diferentes agentes. Además, es necesario definir la medida en la que cada uno de los fenómenos que se quieren explorar, cómo los sesgos o los factores externos, se interpretan dentro del sistema.[\[HAP20, MA60\]](#)

5.1. Objetivos

- Definir un modelo con base en los factores y fenómenos que se quieren tener en cuenta
- Definir la forma como se va a medir o analizar la polarización
- Optimizar los parámetros generales explorando los impactos que tienen en los resultados del modelo
- Establecer una estructura en la que se interpretan y cuantifican los datos del mundo real dentro del modelo

5.2. Retos

- Cada suposición, estructura o algoritmo, que se defina a partir de la intuición de cómo se comporta el modelo a nivel psico-social, supone un potencial sesgo en el modelo.
- Al no haber un estándar en los modelos existentes, cada modelo puede cimentarse a partir de diferentes principios, bases y finalidades, ocasionando que el análisis o comparativa entre ellos deba recaer en un análisis cualitativo de sus resultados.

- Al igual que los elementos del modelo, la forma como se interpretan los datos del mundo real para ser analizados por el modelo, son propensos a sesgos o cambios que pueden tener repercusiones en la lealtad de sus resultados.
- La limitada información que se puede abstraer de las redes sociales, no siempre satisfacen la cantidad de información requerida por el modelo. Un claro ejemplo de ello es la influencia que tiene un individuo en otro, ya sea por confianza, amistad, jerarquía social, etc.

5.3. Beneficios

Lograr definir un modelo preciso para medir, detectar o analizar la polarización ideológica en redes sociales, nos permitirá analizar las masivas interacciones de usuarios frente a un tema específico, y de esta manera:

- Detectar la aparición de la polarización y realizar un análisis de la topología de la misma.
- Identificar factores que están favoreciendo a la aparición de la polarización.
- Incluso en modelos más específicos, se evidencia un primer acercamiento a la detección de cámaras de eco o factores externos, que, de ser identificados, pueden ser intervenidos con el fin de prevenir la polarización.

Modelos y acercamientos al tema

Hasta el momento se ha abordado el tema de la polarización en ciencias de la computación con diferentes modelos. Cada uno de ellos explorando múltiples formas de interpretar o cuantificar los datos y las interacciones presentes en las redes sociales, así como las características y eventos de los modelos, con el objetivo de simular de la forma más precisa la evolución de las opiniones en las redes sociales y detectar la polarización. Estos modelos se diferencian entre ellos principalmente en:

- El espectro de opinión (Discreto o continuo)
- El sistema de interacción entre los agentes.
- Los parámetros variables y estáticos del modelo.
- Los fenómenos considerados dentro del mismo.
- Mecanismo de selección y actualización de los nodos.

6.1. Modelo presentado por el grupo Avispa [MSA21, MA60, MSAV20]

6.1.1. Propuesta

La propuesta de este trabajo es desarrollar un modelo de opinión entre diferentes agentes a través del tiempo, de forma que se simula las interacciones de cada agente con todos los demás que tengan contacto con él, para cada instante de tiempo. Las opiniones de un individuo sobre un tema en particular son representadas como un valor numérico entre $[0, 1]$ y se actualizan después de cada interacción con otro agente. Además, se tiene en cuenta la influencia que tienen ciertos agentes sobre otros. Esta influencia está representada por un grafo de influencia.

6.1.2. Distintivo

Como se mencionaba anteriormente, las interacciones entre personas y la forma como estas influyen en sus opiniones no necesariamente es sencilla e intuitiva, debido a que existen fenómenos psicosociales que afectan de forma directa estas interacciones y suelen generar sesgos tanto en la opinión de las personas como en la interpretación de los modelos.

Por esta razón este modelo intenta representar los sesgos que existen en el intercambio de ideas, que además se han estudiado desde la sociología, que son causantes directos de polarización. En

específico este modelo intenta representar el impacto de los sesgos de confirmación y exposición selectiva.

6.1.3. Parámetros del modelo

Los elementos pertenecientes al modelo desarrollado por el grupo Avispa se pueden dividir en 2, elementos estáticos y dinámicos.

6.1.3.1. Elementos estáticos

El objetivo de estos elementos es representar el estado general del modelo que se mantendrá durante toda la simulación. El modelo incluye los siguientes elementos estáticos:

- Un conjunto finito A de agentes en el sistema
- Un tema o proporción p sobre el cual los agentes tendrán una opinión.
- Una configuración de la opinión que tendrá cada agente. Para cada agente en el modelo le corresponde una opinión entre 0 y 1, donde los valores extremos 0 y 1 representan estar totalmente en contra o totalmente a favor respectivamente de la proposición p .
- Una medida de polarización, la cual calcula según la opinión de todos los elementos del sistema en cualquier instante, una medida numérica sobre el nivel de polarización presente en ese instante en la simulación. En este trabajo se centran en la medida de polarización “influential measure” propuesta por Esteban y Ray.

6.1.3.2. Elementos dinámicos

Estos elementos del modelo capturan la información necesaria para simular la evolución de la opinión de los agentes al interactuar entre ellos. No todos los elementos se usan para medir el nivel de polarización, sin embargo, permiten al modelo simular el comportamiento de las interacciones sociales para así evolucionar el modelo a través del tiempo. Los elementos dinámicos del modelo son:

- El grafo de influencia I . Tal que para cada par de agentes hay un valor entre 0 y 1 que representa la influencia que tiene la opinión de uno sobre el otro.
- El conjunto de todos los instantes de tiempo T
- El conjunto de opiniones de todos los agentes para cada instante de tiempo B_t .
- Una función de actualización, la cual describe la forma como influye la opinión de cada agente en todos los demás en los que tenga influencia, basado en su propia opinión y la influencia que tenga con cada agente, para cada instante de tiempo. Esta función intenta describir la forma como los agentes incorporan nueva evidencia (transmitida por los agentes con los que está en contacto) a su propio razonamiento sobre el tema.

6.1.4. Medición de polarización en el modelo

Como se mencionaba anteriormente, en este modelo se usa la media de polarización de Esteban-Ray, que es explicada en más detalle en el artículo [MA60]. La medida está definida en la imagen. Donde Y es un conjunto de valores de tamaño K , π_i corresponde a la frecuencia del valor y_i en la distribución y α una constante comúnmente asignada a 1.6.

$$\rho_{ER}(\pi, y) = K \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \pi_i^{1+\alpha} \pi_j |y_i - y_j|$$

Figura 6.1: Función para el cálculo de la polarización de Esteban-Ray [MSAV20]

La medida de Esteban-Ray da un índice de polarización basado en la distancia de todos los valores del sistema agrupados por los elementos en Y . De esta forma para medir la polarización basado en las variables del modelo Avispa, se deben definir los valores de la opinión de todos los agentes del sistema para un instante de tiempo dentro de un conjunto finito de valores. Debido a que la medida de Esteban-Ray funciona con un conjunto discreto de datos y la polarización en el modelo se define en un rango constante de 0 a 1, se deben devenir K rangos de valores de forma que se puedan agrupar los valores de las creencias en estos y calcular su frecuencia para cada instante de la simulación.

6.1.5. Función de actualización

La función de actualización definida en este modelo se calcula usando la opinión y el gráfico de influencia de los agentes. Para esto, para calcular la creencia de un individuo en un instante $t+1$ se realiza la función en la figura. Se realiza un promedio del cambio de la opinión del individuo i , con la interacción con cada uno de los agentes con los que está relacionada. Esta interacción con cada agente j se calcula a su vez como la suma entre la opinión inicial de i y la diferencia entre la opinión de ambos, multiplicado por la influencia del agente j en i .

$$B_i^{t+1} = \frac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} B_{i|j}^{t+1}, \quad \text{where} \quad B_{i|j}^{t+1} = B_i^t + \mathcal{I}_{j,i} (B_j^t - B_i^t)$$

Figura 6.2: Definición de la función de actualización del modelo Avispa [MSAV20]

6.1.6. Pruebas

En este trabajo se generaron pruebas sintéticas con diferentes combinaciones de redes de interacción e influencias, así como diferentes opiniones iniciales, las cuales permiten experimentar como el modelo evoluciona con diferentes configuraciones iniciales. Así como también permite probar diferentes valores y parámetros de modo que se optimicen para representar de la forma más precisa el comportamiento esperado. Con este propósito se emplearon diferentes tipos de configuraciones para los datos iniciales. En el caso de la opinión, se usó variaciones de distribuciones uniformes y extremas expuestas en la figura 6.3. Además para la red de influencia se emplearon diferentes estructuras ordenadas, como se puede ver en la figura 6.4.

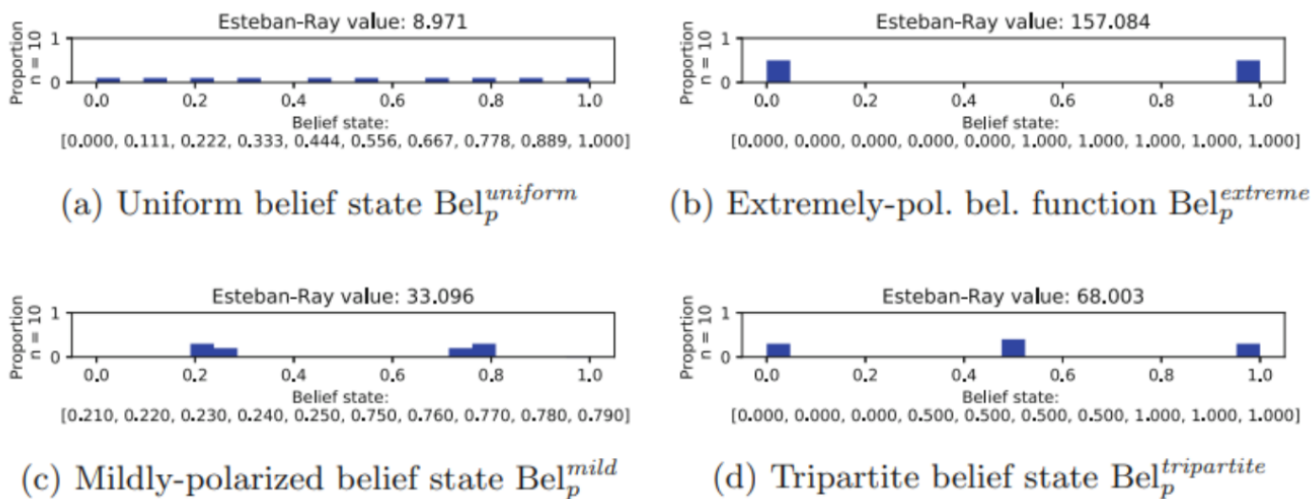


Figura 6.3: Tipos de distribuciones de opinión para las redes sintéticas [MA60]

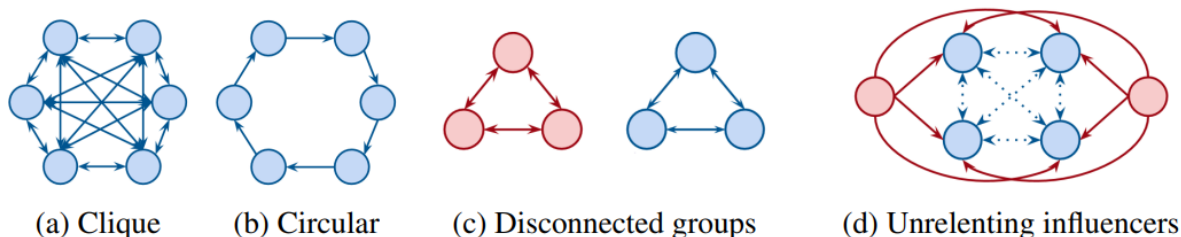


Figura 6.4: Tipos de grafos sintéticos de influencia explorados en el trabajo [MSA21]

6.2. A model of opinion and propagation [HAP20]

6.2.1. Propuesta

La propuesta de este trabajo consiste en un modelo de opinión dinámico que se centra en la idea de que los cambios de opinión son provocados por la exposición a noticias o nueva información. El modelo incluye parámetros para las opiniones y la fuerza de cada conexión, o cuánta influencia tiene dicha conexión en la opinión del individuo, las cuales se actualizan mediante la propagación de noticias.

Se propone un modelo centrado en la propagación de noticias donde los cambios en las opiniones se producen como resultado del intercambio de éstas entre diferentes agentes. Al adoptar esta perspectiva, es viable capturar tanto la polarización de las opiniones como la segregación en la estructura de propagación, con el fin de ser capaces de detectar fenómenos en la topología de las conexiones como las cámaras de eco.

Este escenario además les permite describir algunos comportamientos evidenciados en las redes sociales debido a dos razones principales: la primera es que estudios recientes han descubierto que las redes sociales son los principales medios de noticias hoy en día, al ser los más consumidos por la sociedad. La segunda razón se relaciona con el impacto que se ha comprobado que tienen las noticias partidistas (es decir, las noticias que se presentan para favorecer un punto de vista particular en lugar de ser imparciales) en las opiniones individuales.

6.2.2. Distintivo

El modelo propuesto se centra en el concepto de que la propagación de noticias influye en la dinámica de opinión de los individuos, combinando conocimientos de las áreas de investigación de análisis de opinión y propagación de redes. El modelo permite la observación de la segregación de la estructura de propagación.

Este difiere principalmente de otros en la forma como están planteadas las interacciones. El problema con los intercambios de opinión en modelos donde un par de nodos aleatorios interactúan cada instante de tiempo, es que se convierten en una serie de episodios desconectados, lo que no refleja cómo se difunden las opiniones en las redes sociales. En realidad, las opiniones se transmiten y propagan a través de la red por medio de noticias y la propagación exitosa de esta conduce a una mayor propagación.

Adicionalmente, este trabajo también modela la fuerza de conexión entre diferentes individuos de forma dinámica, representando la confianza o la influencia que tienen las noticias que son recibidas por dicho individuo en su creencia personal.

6.2.3. Definición del modelo

La base de este modelo es la idea de que las personas expresan sus opiniones compartiendo noticias en las plataformas de redes sociales. En otras palabras, un cambio en el punto de vista puede ocurrir como consecuencia de la difusión de noticias a través de los usuarios en la red.

El modelo Polarizing Independent Cascade (PIC), que se propone, se basa en el escenario de propagación de noticias del modelo IC. Sin embargo, hay tres diferencias significativas en el modelo PIC: la existencia de parámetros del modelo, una nueva fórmula para la probabilidad de propagación y un proceso de actualización de los parámetros. La representación algorítmica del proceso de propagación del modelo se puede encontrar en el Algoritmo 6.5.

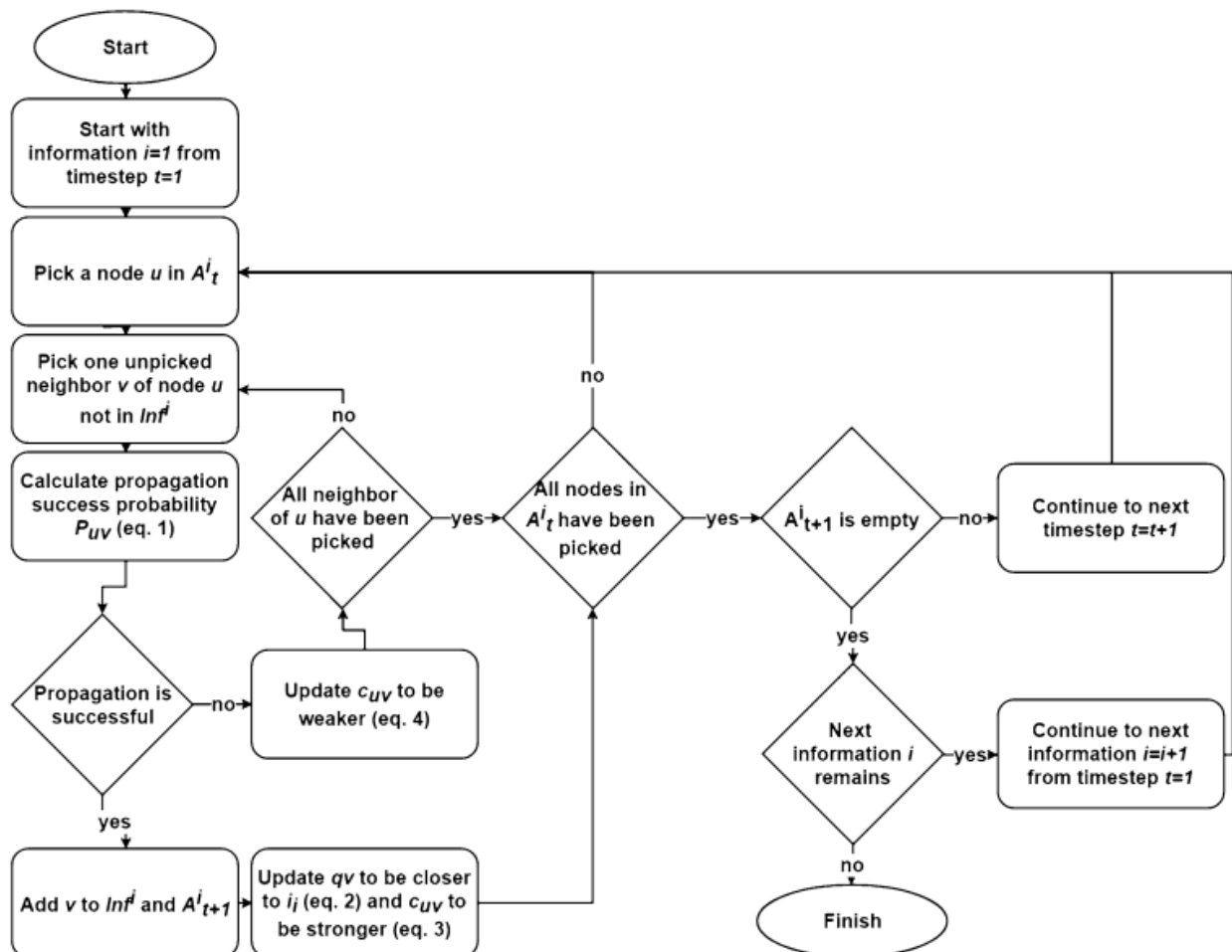


Figura 6.5: Diagrama de flujo de Polarizing Independent Cascade (PIC) [HAP20]

6.2.4. Pruebas

Las pruebas realizadas sobre este modelo consisten en propagar secuencialmente T elementos que representan noticias en la red social y observar los cambios que ocurren en las creencias de los individuos y el grafo de influencia de la red. Sin embargo, si todas las T noticias tienen el mismo puntaje de opinión es claro que las opiniones eventualmente convergen. Por lo tanto, el escenario particular de interés en este trabajo es cuando se propagan noticias cuyos puntajes de opinión se organizan bajo distribuciones variadas.

En este trabajo, se emplearon dos tipos de redes con diferentes propósitos: sintéticas y reales. En primer lugar, se utilizan redes sintéticas para comprender el modelo a través de observar su comportamiento bajo diferentes situaciones y una amplia exploración de parámetros. Posteriormente, el modelo se simula sobre redes reales de una base de datos obtenida de Twitter para determinar la consistencia de su comportamiento en casos reales.

Metodología para la obtención y clasificación de datos

7.1. Retos en la recolección de datos y realización de pruebas

Después de definir un modelo de polarización y como en cualquier otro ámbito, basado en la cuantificación e interpretación de algún fenómeno o sistema del mundo actual, es necesario comprobar cómo se comporta el modelo dependiendo del estado inicial de las variables que se le asignen y los valores que se usan para sus parámetros. En el caso de los modelos de polarización que se han discutido, es comúnmente necesario definir la opinión inicial de todos los individuos, el grafo de influencia, además de algunos parámetros específicos que definen factores de probabilidad entre otros.

Las pruebas de los modelos tienen 2 objetivos principales, inicialmente realizar un ajuste empírico de los parámetros, evaluando el modelo con datos conocidos, usualmente sintéticos, con el fin de identificar los valores que permitan un resultado más preciso o cercano a la respuesta esperada para los casos definidos. Con el fin de realizar estas pruebas o ajustes de los parámetros, se emplean usualmente redes sintéticas, que siguen algún tipo de estructura específica (lo cual nos permite escalar fácilmente el número de individuos en la red). Un ejemplo de esto lo podemos encontrar en los trabajos “Polarization and Belief Convergence of Multi-Agents Systems under Confirmation Bias” y “A model of opinion and propagation”, donde se definieron redes con estructuras Clique, Circulares o subgrupos desconectados, para sus redes de influencia y definen la creencia inicial de los individuos entre diferentes opciones, como uniforme (todos los individuos mantenían la misma creencia), polarizada (se divide los individuos en 2 grupos, cada uno con opiniones diferentes), entre otros.

El segundo objetivo de las pruebas en los modelos es evaluar su precisión e identificar el comportamiento del modelo en un escenario real. Con el fin de obtener conclusiones no solo de la calidad, sino también sobre cómo pueden estar relacionados los fenómenos definidos, que se tomaron como base para la creación del modelo, con el comportamiento de los individuos. Como se mencionó al inicio de este trabajo, el proceso del intercambio de opiniones en una comunidad es sumamente complejo y tiene muchas variables y fenómenos psicosociales que se involucran. Por esto, cuando se define un modelo, se suele enfocar en representar un conjunto parcial de estos comportamientos, además de que el modelo genera una solución cuantitativa para un problema de naturaleza cualitativa, como son las opiniones personales y públicas. Por consiguiente, verificar la precisión o

validez del modelo no es un problema completamente objetivo y las conclusiones observables, tales como, identificar las relaciones que hay entre los diferentes componentes del modelo y cómo estos influyen la red de creencias y la formación de polarización, es donde se encuentra el valor que aportan estos modelos al estado actual del problema de medir la polarización ideológica en redes sociales.

Con base en lo descrito anteriormente es clara la importancia de la realización de pruebas con datos reales en los modelos. Sin embargo, la variedad de interpretaciones o acercamientos a los diferentes comportamientos que describen este fenómeno dan pie a la aparición de muchos retos o dificultades a la hora de obtener datos relevantes y que sean empleados de forma consistente en los diferentes modelos.

7.1.1. Información limitada en las interacciones en las redes sociales

Cada uno de los modelos o proyectos que se trabajan sobre el tema, definen parámetros y variables que permitan representar los estados de su sistema en cada momento, así como los eventos, interacciones y fenómenos presentes en el mismo. Sin embargo, en muchas ocasiones esta información no está presente en la estructura de las redes sociales o no es de fácil acceso. Un ejemplo frecuente de este problema es la identificación de las redes de influencia, que se han mencionado en diversas ocasiones a lo largo de este trabajo. Que consiste en medir la influencia o confianza que hay entre los individuos que comparten sus opiniones en redes sociales.

7.1.2. Interpretaciones y sesgos en la clasificación o medición de las intenciones o creencias de los mensajes o las interacciones entre individuos

Las opiniones que podemos evidenciar en redes sociales, así como las interacciones entre los usuarios, se realizan por medio de mensajes de texto que deben ser cuantificados, con el fin de obtener valores que permitan la creación de las redes de opinión necesarias para la evaluación de los modelos. Esta cuantificación de los mensajes es un problema en sí mismo que ha tenido diferentes acercamientos desde las ciencias de la computación y más particularmente la inteligencia artificial. Sin embargo, las soluciones a este problema no son del todo precisas y presentan sesgos que son transferidos a los resultados del modelo. Especialmente teniendo en cuenta que muchas interacciones de los usuarios son sarcásticas, irónicas o no muy claras, dificultando su correcta interpretación.

7.1.3. Retos para definir o construir a un único individuo a partir de la agrupación de muchas de sus interacciones

Twitter es una de las redes sociales más famosas del momento y es en ella donde se presenta la mayor cantidad de exposición pública de opiniones controversiales, políticas, sociales y económicas. Llegando al punto que muchos medios de comunicación oficial, representación de gobierno y empresas, así como muchas otras figuras públicas, están haciendo parte activa del discurso distribuido dentro de esta. Sin embargo, el formato de interacción dentro de esta red social y muchas otras está basado en un foro, donde cualquier usuario puede responder a otro de forma pública, generando que

la mayor parte de las interacciones sean la puesta de opinión arbitraria de muchos individuos dentro de la red. Haciendo así, que la tarea de identificar múltiples opiniones de un mismo individuo sea compleja y en la mayoría de los casos, toda la información que se puede obtener sobre un tema y usuario específico es una interacción. Dificultando y sesgando la interpretación del individuo dentro del modelo basado únicamente en la información accesible.

7.2. Recolección de los datos

La polarización puede estar presente en cualquier medio en donde múltiples individuos intercambien sus opiniones sobre algún tema. Por lo que cualquier conjunto de interacciones sobre algún tema en específico que esté grabado o almacenado de alguna forma, es un conjunto de datos potencial para el análisis de la opinión y el reconocimiento de la polarización ideológica. De forma consecuente a lo anterior, es claro que las interacciones presentes en las diferentes redes sociales son el principal foco para la recolección de los datos.

Existen diferentes modelos o formatos de redes sociales: plataformas de chat como Whatsapp, por publicaciones como Instagram o Facebook y otras tipo foro como lo son Twitter o Reddit. Así mismo cada una de sus plataformas tienen sus propias limitaciones a la hora de conseguir información. En el caso de Whatsapp, todas las discusiones realizadas dentro de la plataforma son completamente privadas. En el caso de Instagram o Facebook, existe cierta cantidad de información que se puede acceder de forma pública, sin embargo, siempre va a estar limitada por las publicaciones y sus comentarios. Por último, las redes sociales tipo foro son las más usadas para la recolección y el análisis de opiniones. Esto se debe a que además de su estructura orientada a las discusiones sobre temas particulares, que facilita la agrupación de interacciones sobre un mismo tema, también ofrecen la mayor parte de las interacciones de forma pública para realizar estudios y análisis sobre estos. Algunos de los trabajos sobre la polarización en redes sociales discutidos en este documento emplean bases de datos obtenidos de Twitter.

Debido a la popularidad reciente en la obtención de datos de Twitter para la realización de estudios sociales, existen varios conjuntos de herramientas para facilitar la obtención de los twitts deseados con base en un hilo, un usuario o un tema en específico. Todas estas inician en un API de pago desarrollado y publicado por Twitter. Presentándose así, como una excelente opción para proyectos financiados, debido a la popularidad de la red social y lo amplia que es la funcionalidad del API, permitiendo realizar búsquedas complejas. Una estrategia común para recuperar muchas interacciones sobre un mismo tema, es obtener todos los twitts que posean un #hashtag específico, como vimos además en los trabajos de Movilizadorio [Mov20] y [HAP20].

Sin embargo Twitter tiene varias restricciones respecto a la distribución de los datos, a pesar de que todos los mensajes son públicos y pueden ser utilizados para cualquier estudio, se le prohíbe a los trabajos publicar las bases de datos de tweets obtenida por medio de la API, más allá de la lista de ids de cada interacción.

Para este trabajo se emplea una base de datos obtenida por medio de Kaggle, una plataforma enfocada en la ciencia de datos, que presenta bases de datos gratuitas, trabajos de análisis de datos e inteligencia artificial, así como competencias en este mismo campo. En este caso, los mensajes obtenidos son de la red social Reddit, por lo que son de libre acceso, es una colección de hilos sobre la polarización política en USA, titulado “Liberals vs Conservatives on Reddit” [Gaj22] posee 13000 interacciones diferentes de 15 sub hilos diferentes. Adicionalmente a eso, todas las interacciones están previamente clasificadas en 2 categorías Liberales o Conservadores dependiendo de la inclinación política de cada mensaje.

7.3. Preprocesamiento de datos

Antes de iniciar el análisis de las opiniones y polarización en la base de datos, es importante preparar, analizar y visualizar los datos que se van a utilizar. Para esto y al igual que el resto del desarrollo de este proyecto, se utilizará Python en conjunto con las librerías de manejo y visualización de datos presentes en el software de distribución libre orientada a ciencia de datos, Anaconda.

Inicialmente podemos evidenciar cómo están distribuidas las interacciones en los diferentes subreddits o hilos. Ver figura. Como se puede ver tienen una cantidad uniforme de interacciones por hilo a excepción de un único tema, que no se usará para este trabajo. También se puede ver en la figura, como hay una mayor cantidad de interacciones clasificadas como liberales que conservadores, con una diferencia de 64.7% liberales a 35.3% conservadores.

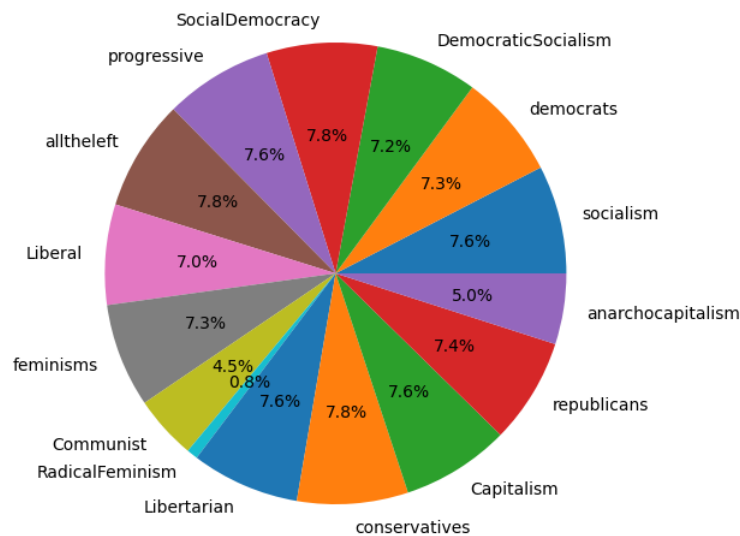


Figura 7.1: Proporción de interacciones por hilo en la base de datos

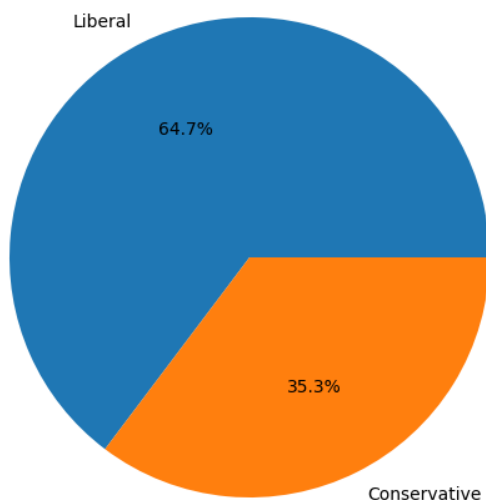


Figura 7.2: Proporción de interacciones por orientación política

7.4. Cuantificación de la creencia de los individuos

Como se mencionaba en capítulos anteriores, la tarea de identificar el sentimiento o creencia que quiere transmitir un mensaje es un problema ampliamente investigado y trabajado por la inteligencia artificial. Siendo una de los problemas más famosos el análisis de sentimiento, que consiste en identificar si la intención de un mensaje es positiva o negativa frente a un tema en específico.[Cha12] Existen muchos modelos entrenados ampliamente conocidos, que solucionan este problema cada vez con mejores resultados, teniendo aún dificultades con temas como el sarcasmo o la ironía. Sin embargo, para el análisis de sentimiento se buscaba un acercamiento que brinde un análisis del mensaje que no sea binario, sino que permita identificar y diferenciar opiniones intermedias, así como extremas sobre un mismo tema.

Con este fin y basándonos en los trabajos [Mov20, HAP20], se decidió utilizar un acercamiento al modelo circuplejo de las emociones. Este cataloga un mensaje de forma continua en la presencia de 7 emociones: rabia, disgusto, miedo, alegría, tristeza, sorpresa y neutral. Este modelo además tiene en consideración la correlación de estas emociones de modo que asigna un número entre 0 y 1 para cada una de las emociones, de forma que la suma de todas ellas sea 1.

En este trabajo se usó un modelo de análisis circuplejo de las emociones pre-entrenado y de libre distribución por medio de la plataforma huggingface. El modelo pertenece a la categoría “text-classification” y se llama “emotion-english-distilroberta-base” [Har22]. Podemos ver un ejemplo de este en las figuras 7.3 y 7.4. El modelo fue evaluado en el trabajo [Har22] resultando en una precisión

del 66 %, en comparación al 14 % de un modelo aleatorio.

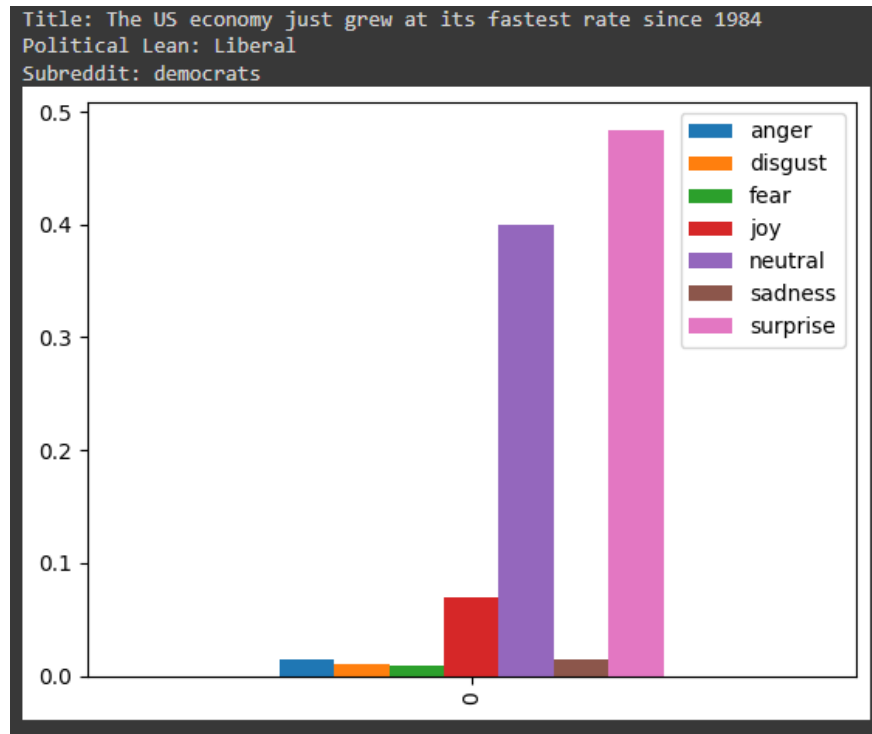


Figura 7.3: Ejemplo de la clasificación del modelo circunplejo de emociones

Como se puede observar, el modelo identifica diferentes sentimientos entre negativos y positivos en cada mensaje. Además, estos están relacionados entre sí, resultando que, un mensaje que posea altos sentimientos negativos como el odio o el disgusto no presenta casi nada de alegría. Con base en esto, se abordó la estrategia para definir la intención del mensaje basándose en los resultados del modelo circunplejo. Se exploró la posibilidad de usar un único sentimiento como la rabia, el disgusto o la alegría para medir la opinión de un mensaje sobre un tema en específico, pero estos no mostraban la disparidad presente en los mensajes. Debido a que dos mensajes que tuvieran un nivel bajo de rabia podían ser completamente diferentes cuando tomamos en cuenta el disgusto o el miedo. Por medio de la experimentación, la mejor solución en donde se podía evidenciar la disparidad entre las opiniones de diferentes mensajes fue la de tomar una combinación lineal de todas las emociones negativas (rabia, decepción, miedo y tristeza), lo que nos entregaba para cada mensaje un nivel de emoción negativa frente al tema entre 0 y 1. En las figuras se muestran los resultados de esta medida en todas las interacciones de los sub-hilos “Capitalism” 7.7, “Republicans” 7.5 y “AllTheLeft” 7.6 respectivamente.

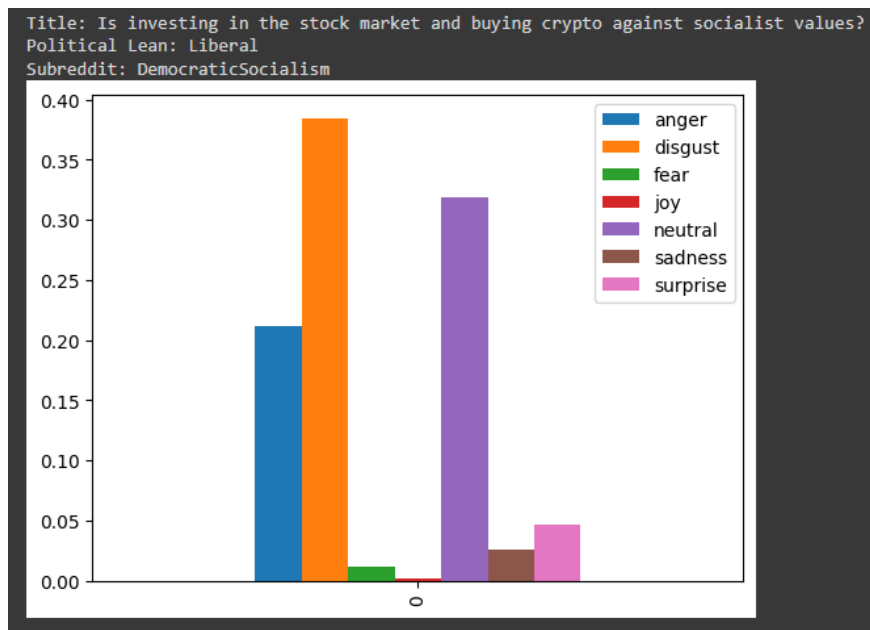


Figura 7.4: Ejemplo de la clasificación del modelo circunplejo de emociones

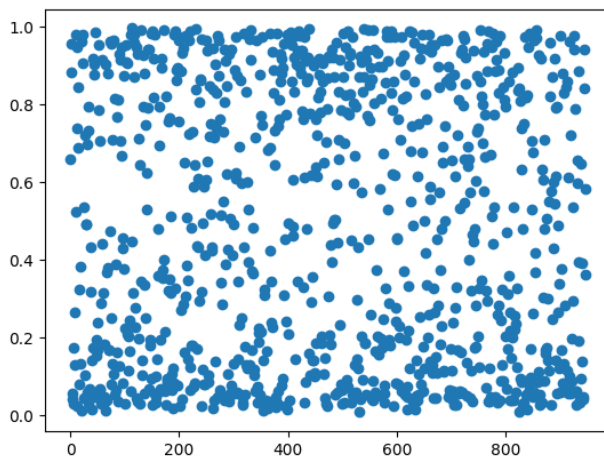


Figura 7.5: Distribución de creencias de las interacciones en el hilo de republicans

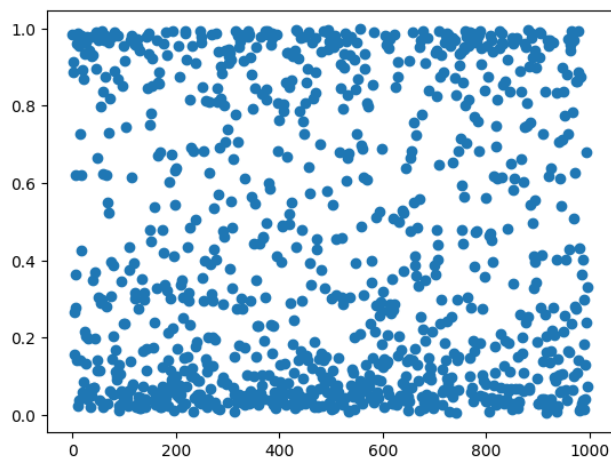


Figura 7.6: Distribución de creencias de las interacciones en el hilo de alltheleft

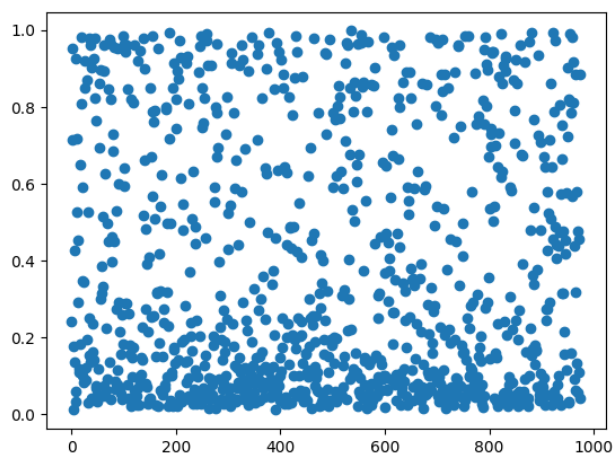


Figura 7.7: Distribución de creencias de las interacciones en el hilo de capitalism

7.5. Definir el grafo de influencia

Como se discutía en el capítulo anterior, uno de los grandes retos a la hora de definir y estructurar los datos para llevarlos a los modelos de polarización, es la limitada información que se puede recuperar de las redes sociales, lo cual requiere de una interpretación o suposición sobre cómo representar diferentes aspectos o variables en el modelo que genera de por sí un sesgo en los datos. Uno de los casos expuestos era el de definir el grafo de influencia, el cual representa la confianza o impacto que tiene la opinión de un individuo en otro.

Para este caso, con los datos que tenemos, se exploraron 4 alternativas:

- Dos individuos están relacionados si hacen parte del mismo sub-hilo en reddit y la confianza entre todos ellos es uniforme
- Dos individuos están relacionados si hacen parte del mismo sub-hilo en reddit y la confianza entre todos ellos depende de su inclinación política. Siendo la influencia entre todas las personas de una misma inclinación política la misma.
- Dos individuos están relacionados si la distancia de sus opiniones es menor a una constante específica.
- Dos individuos están relacionados si ambos hacen parte de la misma clase de opinión, siendo estas a fervor, neutral o en contra.

Para definir el grafo de influencia con los nodos y las aristas se empleó la librería de Python “networkx” que permite una representación y manejo del grafo de forma clara y fácil de leer. Además de ser compatible con las librerías utilizadas para procesar los datos y visualizarlos.

7.6. Visualización de datos

Teniendo los datos separados por los diferentes sub-hilos, cada una de las opiniones de las interacciones cuantificadas por medio del modelo circunplejo y el grafo de influencia definido, es posible realizar una visualización de datos más elaborada, donde permita evidenciar la presencia de conjuntos de opiniones similares donde se puede evidenciar la polarización.

Para este objetivo, se empleó la herramienta Gephi, un software open-source de análisis y visualización de redes. En él se permite analizar las estructuras del grafo, diferentes medidores estadísticos, identificar conjuntos e incluso correr algoritmos topológicos para identificar estructuras internas en base a las aristas y sus respectivos pesos.

En este trabajo se definió el grafo con los nodos y el análisis de opinión expuestas anteriormente para cada uno de los sub-hilos presentes en los datos y se definió el peso entre las aristas de cada par de individuos como 1 menos la diferencia absoluta de las opiniones de cada uno. De este modo se presenta una topología donde frente al mismo tema, dos nodos están más cerca uno del otro si

presentan la misma opinión.

Dentro de la herramienta Gephi se empleó el algoritmo de Modularidad de “fast unfolding of communities in large networks” [VDB08]. Para clasificar los nodos en conjuntos en base a su opinión, se le asignó a cada clase un color y por último se corrió el algoritmo de ForceAtlas2 para diferenciar visualmente la topología del grafo, acercando los nodos que poseen un mayor peso y alejando los demás. Los resultados de este análisis se encuentran en las figuras.

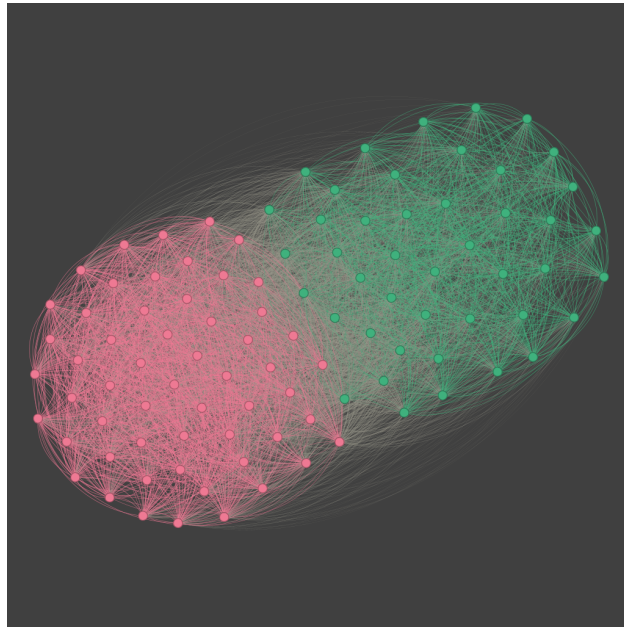


Figura 7.8: Topología del grafo de opinión e influencia sobre el hilo republicans

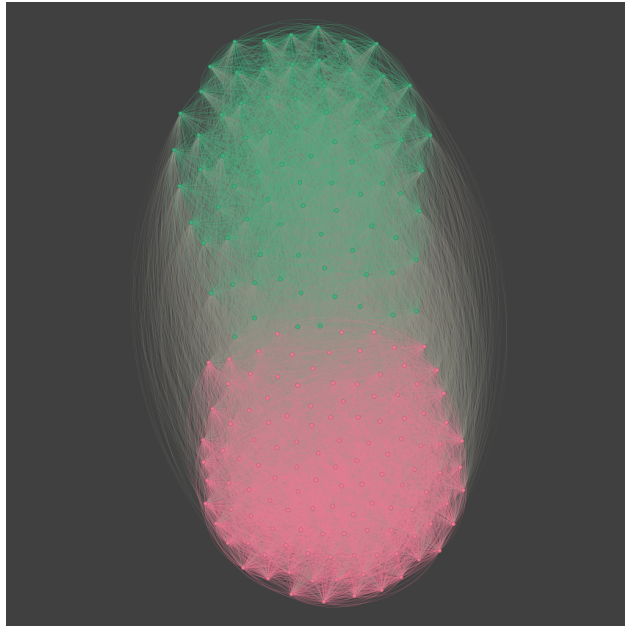


Figura 7.9: Topología del grafo de opinión e influencia sobre el hilo alltheleft

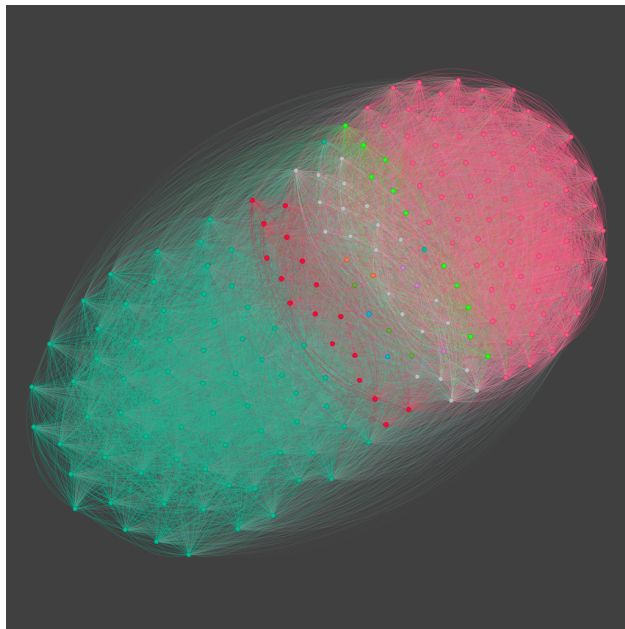


Figura 7.10: Topología del grafo de opinión e influencia sobre el hilo capitalism

7.7. Descripción del algoritmo

La metodología que se siguió para la preparación y transformación de los datos obtenidos del conjunto de datos de Reddit, de modo que se cuantificara la creencia de los mensajes y se dispusiera un grafo de influencia que permitiera correr las simulaciones, se puede dividir en los siguientes pasos:

1. Extraer los mensajes usando la librería Pandas y separarlos en conjuntos de mensajes que se refieran al mismo tema. En el caso de la base de datos empleada en este trabajo, se separaron en base a el hilo o subreddit al que hacen parte.
2. Descargar y preparar el modelo de análisis de sentimiento que se va a emplear y definir las operaciones necesarias para leer y analizar la predicción del mismo. Como se mencionaba anteriormente se decidió usar el modelo pre entrenado de las emociones circuplejo, para ellos se usó la interfaz de huggingface en Python por medio de la librería Transformers. Esta permite importar fácilmente modelos pre entrenados por la comunidad. Además, se debe etiquetar correctamente las predicciones del modelo, las cuales son la medición entre 0 y 1 de las 7 emociones presentadas en el capítulo anterior. Ver figura 7.11
3. Clasificar la creencia de los mensajes basado en los resultados del modelo circuplejo. Para esto se siguió la estrategia expuesta en el capítulo anterior, en la cual se realiza una combinación lineal de las emociones negativas, el cual dará siempre como resultado un valor entre 0 y 1, el cual se tomará como la creencia del agente frente al tema. De modo que, para cada uno de los mensajes de un tema en el conjunto de datos, se corre el modelo y se calcula la creencia. Ver figura 7.12

```

1 from transformers import pipeline
2
3 def classifier2dataFrame(obj):
4     dic = {}
5     for emotion in obj:
6         dic[emotion['label']] = emotion['score']
7     return pd.DataFrame(dic, index = [0])
8
9 classifier = pipeline(
10     "text-classification",
11     model="j-hartmann/emotion-english-distilroberta-base",
12     return_all_scores=True
13 )

```

Figura 7.11: Código en Python para importar el modelo circuplejo usando la interfaz de huggingface

4. Generar el grafo de influencia basado en la estrategia a seguir. Como se describió anteriormente, definir lo que naturalmente sería la confianza o el impacto de la opinión de una persona en

```

1 predictions = textSubreddit.map(classifier)
2 ▾ subredditScores = predictions.map(Lambda x:
3     x[0][0]['score'] + # Valor correspondiente a la rabia
4     x[0][1]['score'] + # Valor correspondiente al disgusto
5     x[0][2]['score'] + # Valor correspondiente al miedo
6     x[0][5]['score']   # Valor correspondiente a la trsiteza
7 )

```

Figura 7.12: Código en Python para calcular la opinión en base a los resultados del modelo circunplejo

otra, es un gran reto a la hora de extraer la información de una red social y significa un sesgo para los datos dentro del modelo. Por ello para este trabajo se optaron por diferentes tipos de estrategias para definir el peso o la influencia entre cada par de agentes. Ver figura 7.13

- Basado en la inclinación política dada como parte de los datos, donde los agentes que pertenezcan a una misma inclinación política tienen un mayor valor de influencia (0.5) en comparación a los demás (0.1).
- Basado en una distribución densa y uniforme, donde todos los agentes se relacionan con todos los demás participantes del foro con un mismo valor de influencia (0.1).
- Basada en la distancia de las creencias entre los agentes, donde todos los usuarios que participan de un mismo foro están relacionados, pero la confianza inicial de unos y los otros depende de la distancia entre sus opiniones. Siendo 0.5 en caso de que sean cercanas y 0.05 en caso contrario.
- Basada en la distancia de los datos, donde únicamente los usuarios con una opinión similar están conectados entre ellos (0.5).
- Basado en clases según la opinión de los agentes, se clasifican los agentes entre extremo negativo [0, 0.3], neutral (0.3, 0.7) y extremo positivo [0.7, 1]. Todos los agentes que participan del foro están relacionados y la influencia de los nodos que pertenecen al mismo extremo es mayor. Siendo 0.5 para los nodos que pertenecen a las clases extremas y 0.05 para todos los demás.
- Basado en clases según la opinión de los agentes y clasificado igual que en la estrategia anterior, pero los agentes únicamente se relacionan con aquellos que pertenecen a su misma clase. Siendo 0.4 para los nodos que pertenecen a las clases extremas y 0.1 aquellos que pertenecen a la clase neutral.

Se realizaron múltiples simulaciones con cada uno de los grafos con el fin de optimizar los valores de las influencias para cada caso, de forma que se logra evidenciar de forma clara las gráficas de la simulación de cada uno y sus impactos en el desarrollo de las opiniones de los agentes a través del tiempo.

```

1 def get_class(score):
2     if score <= 0.3:
3         return 0
4     elif score <= 0.7:
5         return 1
6     else:
7         return 2
8
9 class InfluenceGraphType(Enum):
10     POLITICAL_LEAN = 0
11     UNIFORM = 1
12     DISTANCE = 2
13     DISCONNECTED_DISTANCE = 3
14     CLASSES = 4
15     DISCONNECTED_CLASSES = 5
16
17 def buildInfluenceGraph(num_agents, influenceGraphType: InfluenceGraphType, politicalLean = None):
18
19     if ( influenceGraphType == InfluenceGraphType.POLITICAL_LEAN ):
20         if ( not politicalLean ): raise Exception("politicalLean undefined")
21         influence = np.array([
22             [ 0.5 if politicalLean[i] == politicalLean[j] else 0.1 for i in range(num_agents) ]
23             for j in range(num_agents)])
24
25     elif ( influenceGraphType == InfluenceGraphType.UNIFORM ):
26         influence = np.array([
27             [ 0.1 for _ in range(num_agents) ]
28             for _ in range(num_agents)])
29
30     elif ( influenceGraphType == InfluenceGraphType.DISTANCE ):
31         influence = np.array([
32             [ 0.5 if abs(i - j) < 0.1 else 0.05 for i in scores ]
33             for j in scores])
34
35     elif ( influenceGraphType == InfluenceGraphType.DISCONNECTED_DISTANCE ):
36         influence = np.array([
37             [ 0.5 if abs(i - j) < 0.1 else 0.0 for i in scores ]
38             for j in scores])
39
40     elif ( influenceGraphType == InfluenceGraphType.CLASSES ):
41         influence = np.array([
42             [ 0.5 if get_class(i) == get_class(j) and get_class(i) != 1
43               else 0.05 for i in scores ]
44             for j in scores])
45
46     elif ( influenceGraphType == InfluenceGraphType.DISCONNECTED_CLASSES ):
47         influence = np.array([
48             [ (0.4 if get_class(i) != 1 else 0.1) if get_class(i) == get_class(j)
49               else 0.0 for i in scores ]
50             for j in scores])
51
52     else:
53         raise Exception("influenceGraphType wrongly defined")
54
55     np.fill_diagonal(influence, 1.0)
56     return influence

```

Figura 7.13: Algoritmo en Python para construir diferentes grafos de influencia

7.8. Validación del modelo con datos reales

Uno de los grandes retos para medir o simular la polarización ideológica u otros modelos de opinión está en la poca estructura o estándares a la hora de clasificar o medirla de una forma universalmente correcta. Existe una gran subjetividad dentro de la emisión y la interpretación de un mensaje, así como, de la propia opinión de un individuo. Incluso muchos estudios dentro de la psicología y sociología han abordado el problema de la polarización desde diversas perspectivas, como se presentó al inicio de este trabajo, de modo que no existe una única medida que represente correctamente la presencia del fenómeno o la opinión de cada individuo. Teniendo en cuenta lo anterior, el objetivo de los modelos de polarización presentado en el trabajo consiste no en generar valores específicos para la opinión de un individuo que correspondan a la realidad, pero de identificar a gran escala patrones que puedan permitirnos analizar e identificar la aparición de polarización, así como identificar fenómenos que puedan ser causa/consecuencia de la misma y llegar a identificar estructuras de comunicación que están impulsando la aparición de este fenómeno.

Con base en lo anterior, se puede analizar la validez de un modelo de opinión de forma empírica, tal como se presenta en el artículo "Opinion dynamics in social networks: From models to data" [AFP22]. En este se expone el análisis del comportamiento de opinión popular en votaciones o elección políticas, donde se busca recopilar datos previos a las elecciones en las redes sociales y analizar patrones de comportamiento relacionados, entre lo que se percibe como usuario de las redes sociales y los resultados en las votaciones y las simulaciones realizadas por el modelo. Como se menciona en el artículo, estas observaciones no suelen ser información suficiente para determinar una relación directa entre el modelo y la polarización real, sin embargo, permiten identificar algunos patrones de comportamiento relacionados a fenómenos de la comunicación, diferentes mecanismos de interacción entre los agentes y distribuciones en la opinión que pueden influir en la aparición de la polarización ideológica.

Resultados

8.1. Evaluación en el modelo

Para la evaluación del modelo desarrollado por el grupo de investigación Avispa se empleó los scripts presentados en el artículo [MA60], para esto se utilizaron los subreddits: “republicans”, “Capitalism” y “alltheleft”. Para el análisis de los resultados del modelo se realizaron 2 comparativas diferentes, la evolución de la creencia u opinión de todos los agentes a través del tiempo y la evolución de la polarización a través del tiempo para cada uno de los grafos de influencia en cada uno de los temas.

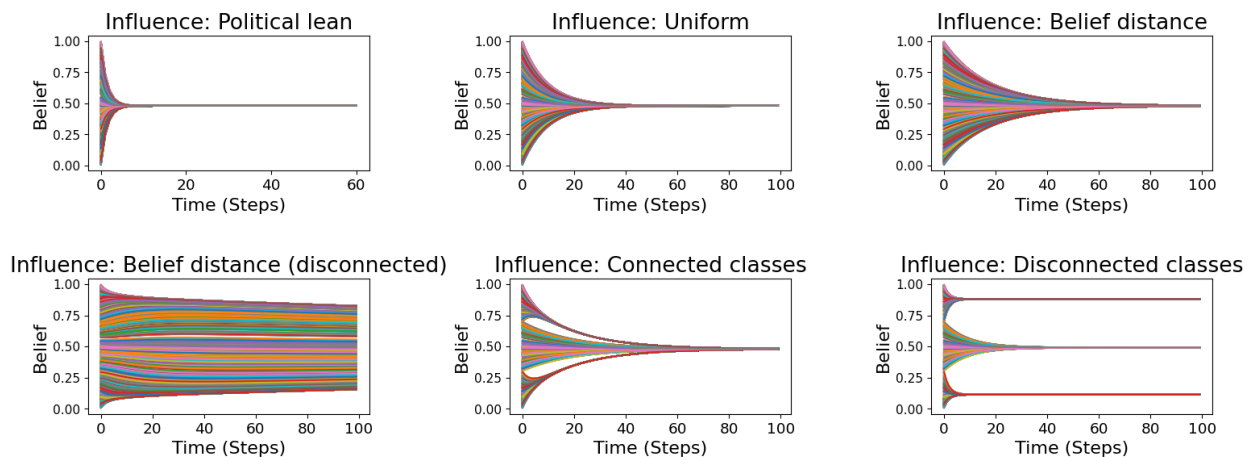


Figura 8.1: Simulación de opiniones de los agentes en el tema Republicans

Inicialmente al comparar los resultados de los diferentes temas o subreddits no es posible identificar características que supongan una variación al modelo o su evolución, más allá de que para el caso de las simulaciones donde la opinión popular converge a un único punto, estas varían el tiempo o pasos que tardan en converger. Además, en las gráficas del capitalismo, es posible ver como una mayor cantidad de agentes que tienen una opinión cercana a 0 u a algún extremo, hace que la convergencia tarde más pasos y el valor al que converga se desvíe ligeramente hacia ese extremo. A parte de lo mencionado, los tres temas presentan resultados muy similares, lo que implica que la diversidad que tienen las opiniones del conjunto de datos utilizado es suficiente para que el modelo las interprete de forma similar, sin exponer un resultado diferente entre ellas.

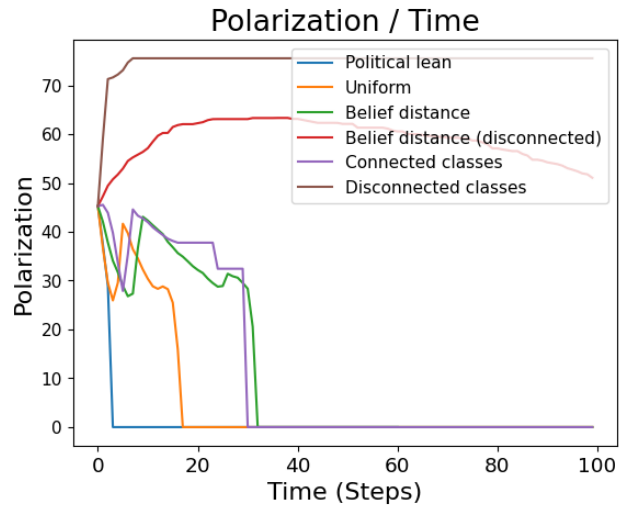


Figura 8.2: Simulación de polarización de la opinión en el tema Republicans

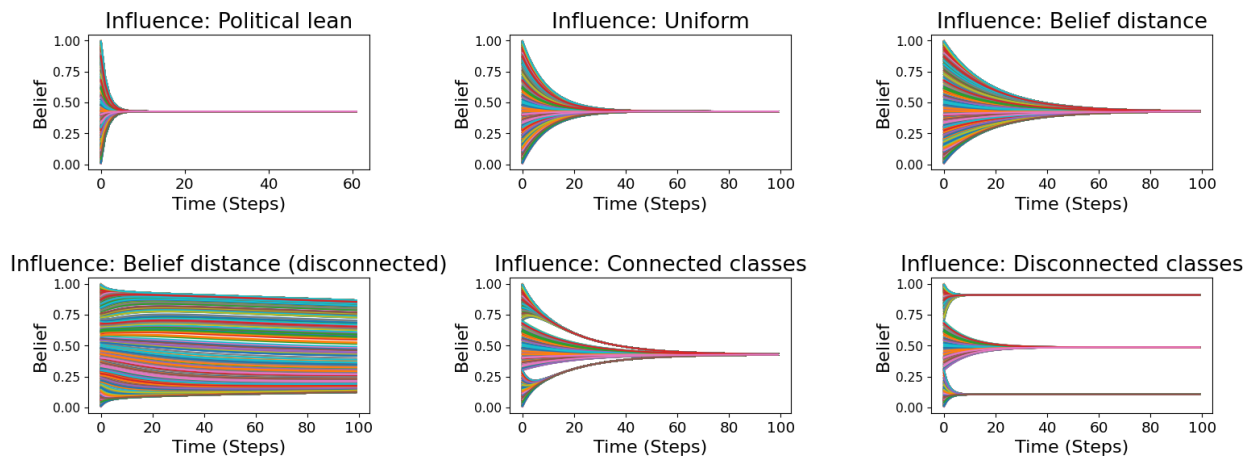


Figura 8.3: Simulación de opiniones de los agentes en el tema Alltheleft

Por otra parte, la comparativa entre los diferentes tipos de influencia tiene grandes resultados en la forma como se simula la polarización a lo largo del tiempo. Para los grados de inclinación política, uniforme, distancia de opinión y clases conectadas, podemos ver como la opinión converge en un punto cercano a la media 0.5. Entre ellos varía principalmente la cantidad de pasos que toma converger al punto medio, siendo el caso de la inclinación política donde menos tiempo se tarda y la distancia de opinión donde más. A partir de esto, podemos evidenciar una relación inversa entre cuán fuertemente conexo sea el grafo de influencia y lo alto que sean las influencias entre los agentes con el tiempo que se tarda en llegar al punto de convergencia.

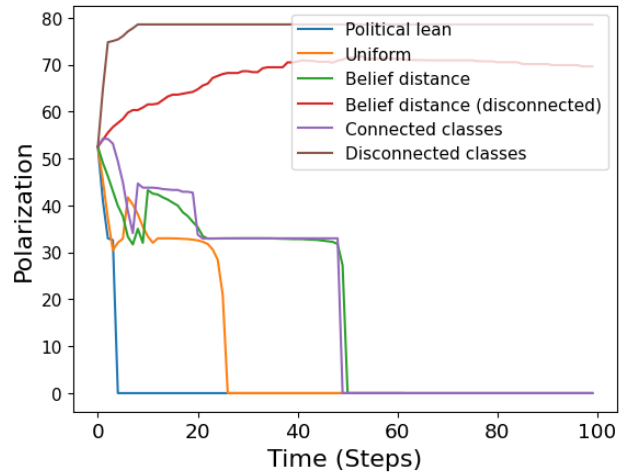


Figura 8.4: Simulación de polarización de la opinión en el tema Alltheleft

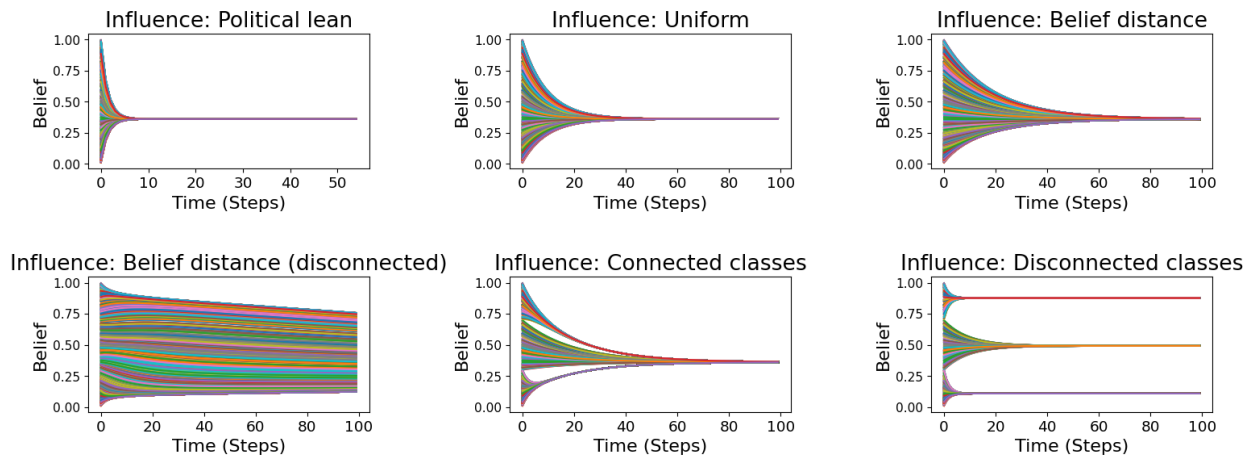


Figura 8.5: Simulación de opiniones de los agentes en el tema Capitalism

En el caso de los grafos de influencia desconectados (distancia de opinión desconectada y clases desconectadas) no se llega al punto de convergencia durante la simulación. En el análisis de la distancia de opinión desconectada, se presenta una pequeña conversión entre los polos de la creencia para luego formarse muchas clases muy conectadas entre ellas y poco conectadas con las demás, de modo que tienen de converger muy lentamente. Por otro lado, el grafo de influencia de clases desconectadas muestra rápidamente como las 3 clases (Extremo negativo, neutro y extremo positivo) convergen individualmente en la media de cada uno, además debido a que las clases de opinión extrema tienen una mayor influencia, estas convergen antes que la neutral.

Un análisis interesante que destacar es el de las clases conectadas, debido a la estructura del grafo de influencia. Donde los agentes que pertenecen a una misma clase tienen mayor influencia

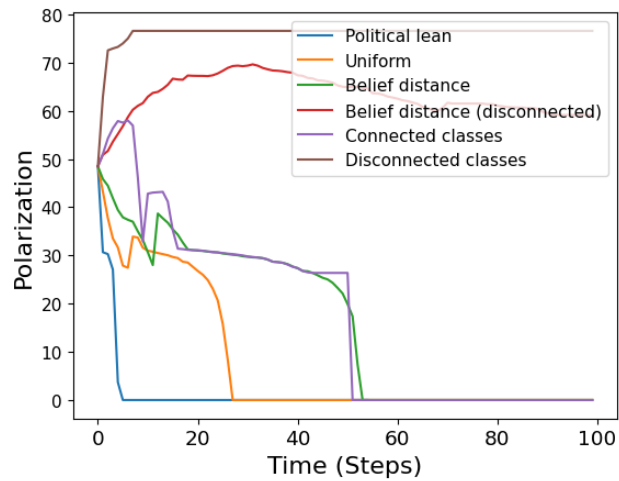


Figura 8.6: Simulación de polarización de la opinión en el tema Capitalism

entre ellos que con los demás, inicialmente las opiniones de cada clase convergen, para luego converger la opinión de las tres clases en la media de todos los agentes.

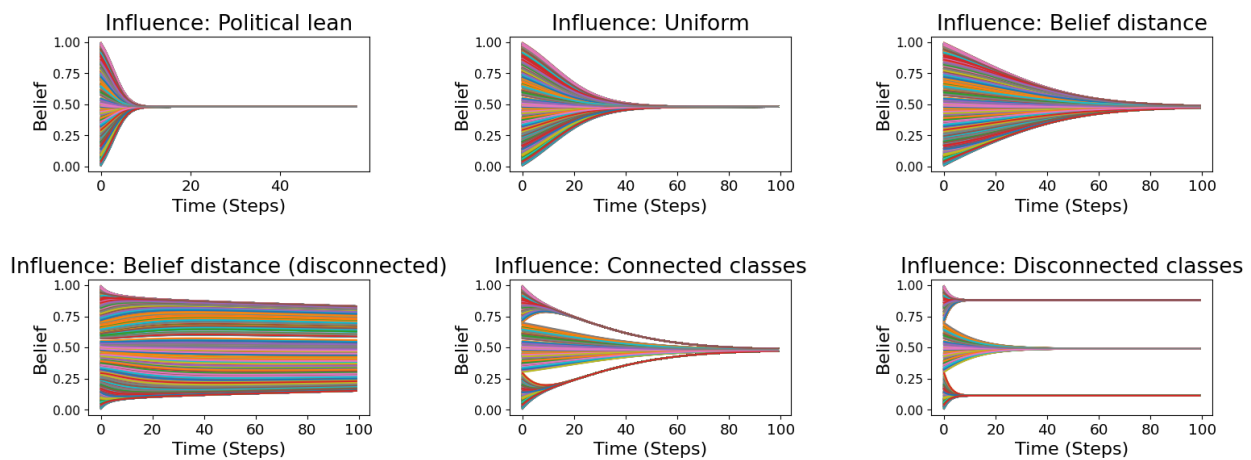


Figura 8.7: Simulación de opiniones de los agentes en el tema Republicans empleando la función de actualización Confirmation Bias

Por último, se realizaron todas las simulaciones del tema “republicans” empleando la función de actualización del modelo “Avispa” que intenta representar el fenómeno de sesgo de confirmación. Debido a la definición de esta función, donde se ve reflejado el impacto que tiene una interacción de un agente sobre otro, dado que esta está fuertemente afectada por su propia opinión, causando que las interacciones con agentes con opiniones similares a él tengan más peso en su propia opinión, la opinión de los agentes tarda más en converger, además se puede observar en las gráficas como

la curva de convergencia de las opiniones más extremas es menos inclinada que con el método de actualización clásico.

8.2. Retroalimentación del modelo

Con base a los resultados obtenidos anteriormente podemos analizar algunas características del modelo presentado por el grupo de investigación Avispa:

- Las opiniones tienden a la media o a converger siempre que el grafo de influencias sea conexo. En otras palabras, siempre que haya algún camino entre cada par de agentes la polarización tiende a desaparecer. Esta conclusión también es presentada por el grupo de investigación en el artículo [MSAV20].
- Al representar el grafo de influencia de los datos obtenidos usando diferentes estrategias y evidenciando como únicamente en las que se presentan subgrafos desconectados se mantiene la polarización a través del tiempo, se puede afirmar la existencia de agentes en las redes sociales que o ignoran o no se ven influenciados por las interacciones con otros, incluso al participar de las mismas discusiones o interactuar entre ellos, lo cual corresponde al fenómeno psicológico llamado exposición selectiva. En consecuencia, la interpretación de los datos donde se definen las conexiones o influencia de los agentes con base en la interacción entre ellos no es del todo acertada y deben considerarse nuevas estrategias que permitan representar dicho comportamiento.
- Como la distribución de las creencias obtenidas en los datos reales no presenta una estructura extrema (a pesar de que se pueden identificar polos, también hay opiniones intermedias) el modelo no permite resaltar esas opiniones extremas, y actúa de manera similar a una distribución de opinión uniforme, como la presentada en los datos sintéticos.
- Desde el modelo podríamos concluir que siempre que el grafo de influencia sea conexo, la opinión del sistema debería de converger en algún momento. Sin embargo, esto parece negar la experiencia empírica que se vive hoy en día frente a los temas evaluados, por lo que se podría investigar nuevos fenómenos o funciones de actualización que permitiera representar o indagar sobre el tipo de comportamiento que evita que la opinión popular pueda converger.
- La representación de los datos en redes sociales tiene limitación a la hora de traducirse al modelo, lo cual limita lo expresivo que este pueda ser o lo cercano a la representación adecuada de las interacciones en las redes sociales. Como se mencionaba anteriormente, la representación del grafo de influencia o la constancia de las interacciones, no se asemeja a la realidad de las redes sociales. El modelo plantea que para cada instante de tiempo todos los agentes interactúan con todos los demás de forma síncrona. Sin embargo, una simulación precisa de las interacciones sociales debería de considerar como la comunicación de un único agente modifica la de algunos otros sin influir en su propia opinión, siguiendo el comportamiento de redes sociales como Twitter o Reddit basado en foros o publicaciones.

8.3. Nuevos enfoques

Teniendo en cuenta las características expuestas anteriormente, el análisis realizado sobre los fenómenos que inducen a la polarización y los resultados obtenidos en este trabajo, se proponen los siguientes enfoques que se podrían explorar en busca de favorecer la simulación de la polarización en redes sociales reales del modelo y la representación de los diferentes fenómenos presentes en el intercambio de opiniones.

- Explorar fenómenos psico-sociales diferentes que dificultan el cambio de opinión de agentes que presentan una opinión extrema, de modo que al acercarse a una opinión extrema estos se vean muy poco influenciados por otras opiniones. O incluso influenciados de forma negativa como es el caso del efecto contraproducente. [MA60]
- Planear diferentes modelos que se alineen con el tipo de estructura de los datos. Grafos fuertemente conexos como los presentes en un foro o débilmente conexos como los presentes en una conversación tienen comportamientos, actualizaciones e influencias diferentes, que podrían ser tratados con mayor precisión si son estructurados de formas específicas para simular su comportamiento.
- Explorar nuevas metodologías de actualización de la opinión a través del tiempo. Como se mencionaba en la sección anterior, las características de las redes sociales basadas en publicaciones o foros son difícilmente representables por un método de actualización donde todos los agentes influyen en aquellos que tienen contacto por cada instante en el tiempo. Explorar métodos basados en la difusión de publicaciones o las interacciones entre grupos de nodos permitirá una mejor representación del comportamiento presente en las interacciones de las redes sociales.
- Con base en los resultados obtenidos en este trabajo, se podría considerar que la actualización dinámica de la influencia presentada por el modelo debería permitir que los agentes eliminen completamente la conexión que tienen unos con otros, de modo que los estados de divergencia que se encuentran en los grafos desconectados puedan ser representados como una consecuencia de la polarización, en lugar de únicamente una causa de ella. Además, permitir que el modelo actualice no solo la influencia entre los agentes, sino también la conexión entre ellos puede permitir explorar la topología del grafo generada durante la simulación del modelo, de modo que se puedan detectar otros fenómenos determinantes de la polarización como son las cámaras de eco.
- Permitir al modelo simular agentes obstinados. Los cuales mantienen una opinión extrema y constante, sin ser influenciados por otros.

Conclusión

9.1. Futuros pasos

- Definir un estándar en los modelos de polarización ideológica, que permita unificar los diferentes estudios que se han realizado al respecto, realizar comparativas cuantitativas frente a la precisión de cada uno de ellos frente e identificar de forma cuantitativa las ventajas que ofrece el tener en cuenta diferentes fenómenos sociales.
- Recolectar datos sobre temas de la actualidad y el contexto colombiano. Con el fin de analizar las crecientes problemáticas sociales que vive Colombia en estos momentos y el papel que juega la polarización ideológica en él.
- Continuar la investigación en diferentes métodos para cuantificar las intenciones o creencias de un individuo con base en un mensaje, de modo que se minimicen los sesgos que nacen de los errores, en especial en mensajes con una estructura compleja como el sarcasmo o la ironía.
- Explorar por medio de los resultados y simulaciones obtenidos en los modelos de polarización actuales, métodos o estrategias para identificar factores externos que estén promoviendo la aparición de polarización, como bots o agentes obstinados.
- Desarrollar aplicaciones reales que pongan en uso la detección y medición de la polarización en redes sociales, como la detección de agentes obstinados, bots u factores que puedan estar influyendo en la aparición de la polarización. Además de definir alarmas que permitan identificar la aparición de polarización en temas donde la opinión popular juega un papel relevante.

9.2. Conclusión

Se exploró la teoría detrás de la polarización ideológica como un fenómeno de los intercambios de opiniones entre individuos, los diferentes fenómenos psico-sociales ampliamente estudiados que promueven a la aparición de estos polos extremos de opinión, del mismo modo como los factores externos a la conversación que se presentan con el fin de alterar la opinión popular frente a un tema y que se han visto fuertemente impulsados con el intercambio masivo de opiniones presente en las redes sociales. Además de analizar cada uno de estos factores y como las ciencias de la computación y el análisis de datos puede favorecer al estudio, simulación y reconocimiento de este fenómeno en las redes sociales, así como los acercamientos que se han realizado al tema y los retos presentes en el mismo. Se analizaron los retos que tienen que afrontar los modelos de polarización, tanto para

representar la información presente en las redes sociales, como para simular de forma precisa el comportamiento de la opinión de múltiples individuos al interactuar entre sí de forma precisa, teniendo en cuenta los múltiples fenómenos conocidos que afectan de forma, en ocasiones, contraintuitiva el impacto que tiene una opinión sobre otra.

Sin embargo, los diferentes modelos que se han realizado hasta el momento permiten evidenciar algunos comportamientos de la polarización en redes sociales que vivimos día a día. Así como explorar y sacar algunas conclusiones sobre el impacto que pueden estar teniendo los fenómenos sociales en la comunicación dentro de las redes sociales, además de evidenciar algunas tendencias y perspectivas gracias al análisis computacional, como es las redes topológicas o las cámaras de eco y como estas se hacen presentes y evolucionan fuertemente relacionadas a la polarización.

Se presentó además una metodología y algoritmo para la recolección, análisis, clasificación y visualización de las redes de opinión a partir de las interacciones en redes sociales que siguen un formato tipo foro. El cual puede contribuir a las pruebas de los modelos de polarización con datos reales, así como la integración futura de dichos modelos para lograr aplicaciones como la medición de polarización o la detección de factores externos que promueven la aparición de este fenómeno.

Además, a partir de las pruebas realizadas en el modelo de polarización presentado por el grupo Avispa con conjunto de datos obtenido de la red social Reddit. Se logró evidenciar el comportamiento del modelo en casos reales e identificar retos en la representación de estos, como son la actualización de las conexiones dentro de la red por medio de la difusión de publicaciones en lugar de la influencia de todos los agentes simultáneamente, la representación de personas con opiniones extremistas e inflexibles o la inclusión de factores externos, como bots o agentes obstinados. También se logró evidenciar características presentadas por dicho trabajo, como la relación entre la aparición de polarización en el modelo y el grado de conexión del grafo de influencia.

Por último basado en la investigación realizada sobre la polarización ideológica, diferentes acercamientos al tema que se han desarrollado en los últimos años y las conclusiones obtenidas de los resultados del modelo con datos reales, se presentaron diferentes propuestas que pueden permitir a futuros modelos de polarización una mejor representación de las interacciones presentes en las redes sociales, los fenómenos que son causa/consecuencia de la polarización y la representación de estructuras dentro de influencia que promueven la aparición de este fenómeno social.

Bibliografía

- [AFP22] Gerardo Iniguez Antonio F. Peralta, Janos Kertesz. Opinion dynamics in social networks: From models to data. 2022.
- [AMB07] K. P. Gummadi P. Druschel A. Mislove, M. Marcon and B. Bhattacharjee. Measurement and analysis of online social networks. *7th ACM SIGCOMM*, 2007.
- [Aro10a] Elliot Aronson. Social psychology. *Upper Saddle River*, 2010.
- [Aro10b] Wilson T. Akert Aronson, E. *Social Psychology*. Prentice Hall, 2010.
- [CH15] Z. Christoff and J. U. Hansen. A logic for diffusion in social networks. *J. Applied Logic*, 2015.
- [Cha12] Springer Cham. *Sentiment Analysis and Opinion Mining*. Springer Cham, 2012.
- [Chr16] et al.: Dynamic logics of networks Christoff, Z. information flow and the spread of opinion. *University of Amsterdam*, 2016.
- [Cra08] Cosley D. Huttenlocher D. Kleinberg J. Suri S Crandall, D. Feedback effects between similarity and social influence in online communities. *ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008.
- [ECL11] S. A. Myers E. Cho and J. Leskovec. Friendship and mobility: User movement in location-based social networks. *In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011.
- [ER94a] J. Esteban and D. Ray. The measurement of polarization. *Econometrica*, 1994.
- [ER94b] J. Esteban and D. Ray. The measurement of polarization. *Econometrica*, 1994.
- [Fla18] Goel S. Rao J.M Flaxman, S. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 2018.
- [Fle17] Nielsen R.K Fletcher, R. Are news audiences increasingly fragmented? a crossnational comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication*, 2017.
- [GA05] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. *ACM Workshop on Privacy in the Electronic Society*, 2005.
- [Gaj22] Neel Gajare. Liberals vs conservatives on reddit [13000 posts]. <https://www.kaggle.com/datasets/neelgajare/liberals-vs-conservatives-on-reddit-13000-posts>, 2022.

- [Gar16] Gandica Y Gargiulo, F. The role of homophily in the emergence of opinion controversies. *preprint arXiv*, 2016.
- [HAP20] Tsuyoshi Murata Hafzh A. Prasetya. A model of opinion and propagation structure polarization in social media. *Computational social networks*. Springer, 2020.
- [Har22] Jochen Hartmann. Emotion english distilroberta-base. <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/>, 2022.
- [Hun17] A. Hunter. Reasoning about trust and belief change on a social network: A formal approach. *International Conference on Information Security Practice and Experience*. Springer, 2017.
- [JSG13] F. Liu J. Seligman and P. Girard. Facebook and the epistemic logic of friendship. *CoRR*, 2013.
- [KHH18] S. Soundarajan K. He, Y. Li and J. E. Hopcroft. Hidden community detection in social networks. *Information Sciences*, 2018.
- [Kir17] E Kirby. The city getting rich from fake news. *BBC News*, 2017.
- [Les10] Huttenlocher D. Kleinberg J. Leskovec, J. Signed networks in social media. *SIGCHI Conference on Human Factors in Computing Systems*, 2010.
- [Li13] Scaglione A. Swami A. Zhao Q Li, L. Consensus, polarization and clustering of opinions in social networks. *IEEE Journal on Selected Areas in Communications*, 2013.
- [Liu14] Seligman J. Girard P Liu, F. Logical dynamics of belief change in the community. *Synthese*, 2014.
- [LNK07] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 2007.
- [MA60] F. Valencia M. Alvim, S. Knight. Toward a formal model for group polarization in social networks. *Lecture Notes in Computer Science series*, 11760, 1960.
- [ML76] D. G. Myers and H. Lamm. The group polarization phenomenon. *Psychological bulletin*, 1976.
- [Mov20] Fundacion Santo Domingo Open Society oundations Movilizadorio, Ford Foundation. Xenofobia y polarizacion para lograr inclusion y cohesion en colombia. 2020.
- [MSA21] Sophia Knight Santiago Quintero Frank Valencia Mario S. Alvim, Bernardo Amorim. Polarization and belief convergence of multi-agents systems under confirmation bias. *Lecture Notes in Computer Science series*, 2021.

- [MSAV20] Sophia Knight Santiago Quintero Mario S. Alvim, Bernardo Amorim and Frank Valencia. Polarization and belief convergence of agents in strongly-connected influence graphs. 2020.
- [Pl010] S Plous. *The Psychology of Judgment and Decision Making*. McGraw-Hill, 2010.
- [Pro16] Matveev A.S. Cao M Proskurnikov, A.V. Opinion dynamics in social networks with hostile camps: Consensus vs. polarization. *IEEE Transactions on Automatic Control*, 2016.
- [RS15] Nicolas Roux and Joel Sobel. Group polarization in a model of information aggregation. *American Economic Journal: Microeconomics*, 2015.
- [Sel11] Liu F. Girard P Seligman, J. Logic in the community. *Springer*, 2011.
- [Sin18] Bramson A Grim P Singer, D.J. Rational social and political polarization. *Philos Stud*, 2018.
- [Sir19] Pedreschi D. Giannotti F. Kertesz J Sirbu, A. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PLOS ONE*, 2019.
- [VDB08] Renaud Lambiotte Etienne Lefebvre Vincent D Blondel, Jean-Loup Guillaume. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008.
- [ZCP16] J. U. Hansen Z. Christoff and C. Proietti. Reflecting on social influence in networks. *Journal of Logic, Language and Information*, 2016.